

# Aplicació de diferents metodologies d'anàlisi i representació de dades per a l'avaluació de la qualitat ambiental d'una conca hidrogràfica

Marta Terrado Casanovas

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tesisenxarxa.net](http://www.tesisenxarxa.net)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tesisenred.net](http://www.tesisenred.net)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tesisenxarxa.net](http://www.tesisenxarxa.net)) service has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized neither its spreading and availability from a site foreign to the TDX service. Introducing its content in a window or frame foreign to the TDX service is not authorized (framing). This rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

# **APLICACIÓ DE DIFERENTS METODOLOGIES D'ANÀLISI I REPRESENTACIÓ DE DADES PER A L'AVALUACIÓ DE LA QUALITAT AMBIENTAL D'UNA CONCA HIDROGRÀFICA**

**Marta Terrado Casanovas  
Barcelona, 2009**



Universitat de Barcelona  
Facultat de Geologia

Departament de Geoquímica, Petrologia i  
Prospecció Geològica



Consell Superior d'Investigacions Científiques (CSIC)  
Institut de Diagnosi Ambiental i Estudis de l'Aigua  
(IDÆA)

Departament de Química Ambiental



---

# **Aplicació de diferents metodologies d'anàlisi i representació de dades per a l'avaluació de la qualitat ambiental d'una conca hidrogràfica**

Memòria realitzada per Marta Terrado Casanovas sota la direcció dels Doctors Romà Tauler Ferré i Damià Barceló Cullerés, i la tutoria del Doctor Manel Viladevall Solé, al Departament de Geoquímica, Petrologia i Prospecció Geològica de la Universitat de Barcelona, per a optar al grau de Doctor en Ciències Geològiques.

Programa de Doctorat: Ciències de la Terra (Bienni 2005-2007)

Barcelona, Setembre de 2009

Doctorand:

**Marta Terrado Casanovas**

Directors de Tesi:

**Dr. Romà Tauler Ferré**

Departament de Química Ambiental de l'Institut de Diagnosi Ambiental i Estudis de l'Aigua (IDAEA – CSIC)

**Dr. Damià Barceló Cullerés**

Departament de Química Ambiental de l'Institut de Diagnosi Ambiental i Estudis de l'Aigua (IDAEA – CSIC)

Tutor:

**Dr. Manel Viladevall Solé**

Departament de Geoquímica, Petrologia i Prospecció Geològica  
Facultat de Geologia, Universitat de Barcelona (UB)

---





# ÍNDEX

1. INTRODUCCIÓ .....	9
1.1. Caracterització de l'àmbit d'una conca hidrogràfica .....	9
1.2. Problemàtiques que afecten la qualitat ambiental d'una conca hidrogràfica ...	13
1.2.1. Legislació europea .....	15
1.2.2. La contaminació de les aigües superficials .....	16
1.2.3. La contaminació de les aigües subterrànies .....	23
1.2.4. La contaminació del sòl.....	24
1.2.5. La contaminació dels sediments .....	27
1.3. Metodologies aplicades a l'anàlisi i avaluació de la qualitat ambiental d'una conca hidrogràfica .....	29
1.3.1. Metodologies empíriques .....	29
1.3.2. Metodologies deterministes .....	35
1.3.3. Els SIG com a eina d'anàlisi ambiental .....	44
1.4. Justificació de la tesi.....	48
1.5. Objectius.....	50
2. METODOLOGIES APLICADES A L'ANÀLISI DE DADES AMBIENTALS.....	53
2.1. Mètodes estadístics i quimiomètrics .....	53
2.1.1. Mètodes univariants d'anàlisi de dades .....	54
2.1.2. Mètodes multivariants d'anàlisi de dades.....	55
2.1.2.1. Estructura de les taules de dades ambientals .....	56
2.1.2.2. Tractament preliminar de les dades .....	58
2.1.2.3. Pretractament de les dades.....	59
2.1.2.4. Anàlisi de components principals (PCA).....	64
2.1.2.5. Resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS) .....	67
Article 1: Quality assessment of the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method for the investigation of environmental pollution patterns.....	89
2.1.2.6. Resultats de l'avaluació del mètode MCR-ALS per a la investigació de pautes de contaminació en aigües superficials .....	91
2.1.2.7. Conclusions .....	99
2.1.3. Mètodes geostatístics .....	100
2.2. Mètodes deterministes.....	106
2.2.1. Model PHYSITEL .....	106
2.2.2. Model HYDROTEL.....	109
2.2.3. Sistema GIBSI .....	120

3. APLICACIÓ DE METODOLOGIES D'ANÀLISI DE DADES AMBIENTALS A L'ESTUDI DE DIFERENTS CONQUES HIDROGRÀFIQUES.....	127
3.1. Introducció .....	127
3.2. Conca hidrogràfica del riu Ebre .....	128
3.2.1. Resultats de l'aplicació de mètodes quimiomètrics a l'estudi de la contaminació a la conca del riu Ebre .....	133
Article 2: Identification and distribution of contamination sources in the Ebro River basin by chemometric modelling coupled to geographical information systems .....	141
Article 3: MCR-ALS modelling of organic pollution patterns in the Ebro River surface water-groundwater-sediment-soil system .....	143
Anàlisi de fàrmacs en aigües superficials.....	145
3.3. Delta del Riu Ebre .....	155
3.3.1. Resultats de l'aplicació de mètodes quimiomètrics i geostatístics a l'estudi de la contaminació al delta del Riu Ebre.....	158
Article 4: Use of chemometric and geostatistical methods to evaluate pesticide pollution in the irrigation and drainage channels of the Ebro River delta during the rice-growing season.....	165
3.4. Conca hidrogràfica del Riu Llobregat .....	166
3.4.1. Resultats de l'aplicació de mètodes quimiomètrics i deterministes a l'estudi de la contaminació a la conca del riu Llobregat.....	169
Article 5: Distribution and assessment of surface water contamination by application of chemometric and deterministic models .....	188
3.5. Conclusions .....	189
4. UTILITZACIÓ D'ÍNDEXS PER A L'AVALUACIÓ DE LA QUALITAT DE L'AIGUA EN UNA CONCA HIDROGRÀFICA.....	191
4.1. Introducció .....	191
4.1.1. Preselecció d'índexs per a dades de xarxes de control automàtiques....	193
4.2. Índex CMME WQI de Canadà .....	197
Article 6: Surface water quality indices for the analysis of data generated by automated sampling networks.....	199
4.2.1. Resultats de l'aplicació del CCME WQI a l'anàlisi de dades generades per xarxes de control automàtiques .....	200
4.3. Índex prioritari de risc de contaminació (basat en el procediment COMMPS) a les conques hidrogràfiques de Catalunya.....	203
4.4. Conclusions .....	214
5. CONCLUSIONS GENERALS.....	216

Referències.....	220
Annex I .....	228
Annex II .....	230
Annex III .....	232
Annex IV.....	240

# ÍNDEX DE TAULES

Taula 1.1. Models empírics de tractament de dades ambientals .....	31
Taula 1.2. Models deterministes de qualitat a escala de conca hidrogràfica.....	38
Taula 1.3. Models deterministes de qualitat per aigües subterrànies .....	41
Taula 1.4. Models deterministes de qualitat per aigües superficials .....	41
Taula 1.5. Models deterministes d'escolament urbà.....	43
Taula 2.1. Processos i algorismes disponibles per les simulacions hidrològiques amb HYDROTEL .....	110
Taula 3.1. Característiques principals de les àrees estudiades en aquesta Tesi.....	127
Taula 3.2. Resum de les pressions i els impactes ambientals existents sobre el territori de la conca del riu Ebre .....	131
Taula 3.3. Identificació dels punts de presa de mostra de la RCSP .....	134
Taula 3.4. Identificació dels punts de presa de mostra d'aigües superficials i sediments .....	137
Taula 3.5. Identificació dels punts de presa de mostra d'aigües subterrànies i sòls .....	138
Taula 3.6. Identificació dels fàrmacs analitzats en aigües superficials de la conca del riu Ebre i grup terapèutic al qual pertanyen .....	148
Taula 3.7. Identificació de les mostres recollides a la conca del Riu Ebre per a l'anàlisi de fàrmacs classificades en diferents tipus .....	148
Taula 3.8. Identificació de les estacions de presa de mostra al delta de l'Ebre .....	159
Taula 3.9. Identificació de les variables analitzades a les mostres del delta de l'Ebre.....	159
Taula 3.10. Identificació, localització i altitud de les estacions meteorològiques utilitzades en el model HYDROTEL per a la simulació de la dinàmica fluvial a la conca del riu Llobregat....	175
Taula 3.11. Identificació, localització i altitud de les estacions hidromètriques utilitzades en el model HYDROTEL per a la simulació de la dinàmica fluvial a la conca del riu Llobregat .....	176
Taula 3.12. Etapes del procés de calibratge proposades per Turcotte et al. 2003.....	179
Taula 3.13: Paràmetres utilitzats durant el procés de simulació hidrològica amb HYDROTEL la conca del riu Llobregat .....	182
Taula 3.14. Cabals observats i simulats a les diferents estacions hidromètriques mitjançant el model HYDROTEL i error del model.....	184
Taula 3.15. Valors de sorpció, solubilitat i vida mitjana dels productes fitosanitaris simulats amb el model PS.....	186
Taula 4.1. Factors de control utilitzats per al càlcul de PNECs.....	207
Taula 4.2. Puntuació assignada als efectes aquàtics indirectes .....	207
Taula 4.3. Frases R i puntuació assignada .....	208
Taula 4.4. Llista de compostos prioritari classificats a les conques catalanes .....	209

# ÍNDEX DE FIGURES

Figura 1.1. El cicle de l'aigua.....	9
Figura 1.2. Activitats de les conques hidrogràfiques que poden afectar la qualitat de les aigües superficials.....	14
Figura 1.3. Cicle d'un contaminant a la hidrosfera.....	16
Figura 1.4. Superposició de capes de diferent geometria en un SIG.....	45
Figura 1.5. Representació ràster i vectorial de la realitat.....	46
Figura 1.6. Unió de dades $\alpha$ -numèriques i cartogràfiques.....	46
Figura 2.1. Esquema d'un diagrama de caixa (a) i mapa de correlació de variables (b).....	54
Figura 2.2. Tipus d'estructures i models de dades ambientals.....	57
Figura 2.3. Diagrames de caixa d'un conjunt de dades ambientals.....	61
Figura 2.4. Nous eixos ortogonals creats a partir del conjunt de dades originals.....	64
Figura 2.5. Descomposició per PCA (model bilineal) de la matriu de dades D per un nombre de components $N = 3$ .....	65
Figura 2.6. Descomposició de valors propis d'una matriu de dades per a la determinació del nombre de components.....	66
Figura 2.7. Diagrama d'implementació del mètode MCR-ALS amb l'exemple d'aplicació de tres restriccions diferents: no negativitat, normalització dels <i>loadings</i> i trilinearitat.....	70
Figura 2.8. Diferents estructures i ordenacions de dades.....	78
Figura 2.9. Descomposició d'un conjunt de dades en tres direccions seguint un model trilineal.....	81
Figura 2.10. Descomposició de la matriu de dades augmentada $D_{aug}$ per MCR-ALS aplicant la restricció de trilinearitat.....	82
Figura 2.11. <i>Loadings</i> i <i>scores</i> (amb estructura no trilineal i estructura trilineal) de les pautes de contaminació simulades durant 4 mesos.....	92
Figura 2.12. Comparació dels <i>loadings</i> de les tres fonts de contaminació obtingudes a partir de l'anàlisi de les dades.....	94
Figura 2.13. Comparació dels perfils de <i>loadings</i> teòrics, els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades, i els perfils obtinguts desescalant els perfils escalats (dades simulades amb soroll).....	95
Figura 2.14. Comparació dels perfils de <i>loadings</i> teòrics escalats i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades amb i sense soroll.....	97
Figura 2.15. Comparació dels perfils de <i>loadings</i> teòrics i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS amb la restricció de trilinearitat per dades trilineals i dades no trilineals (dades simulades amb soroll).....	98
Figura 2.16. Paràmetres utilitzats per a l'ajust del semivariograma empíric a la funció matemàtica més idònia.....	103
Figura 2.17. Exemple de representació d'una xarxa fluvial en format matricial, indicant les seccions de riu, nodes i el node exutori de la conca.....	108
Figura 2.18. Operacions efectuades mitjançant l'aplicació del software PHYSITEL a la conca del riu Llobregat.....	109
Figura 2.19. Diferents processos simulats sobre una UHRH.....	111
Figura 2.20. Exemple d'un fitxer *.hyd amb les dades hidrològiques de cabal diari mesurat a l'estació de Sant Joan Despí (Barcelonès).....	112

Figura 2.21. Exemple d'un fitxer *.met amb les dades meteorològiques diàries de l'estació de La Quar (Berguedà) .....	112
Figura 2.22. Subdivisió del balanç vertical en tres capes dins del submodel BV3C d'HYDROTEL.....	117
Figura 2.23. Representació dels components i de l'estructura del sistema GIBSI .....	120
Figura 2.24. Factors que afecten l'erosió del sòl .....	122
Figura 2.25. Simulació del transport de productes fitosanitaris del sòl al riu.....	124
Figura 3.1. Mapa de situació de la conca del riu Ebre i comunitats autònomes dins de l'àmbit de la conca .....	129
Figura 3.2. Mapa amb la identificació del riu Ebre i els principals afluents que aporten una major quantitat d'aigua.....	129
Figura 3.3. Mapa de localització de les principals pràctiques agrícoles i dels centres industrials i urbans més importants a la conca hidrogràfica del riu Ebre.....	132
Figura 3.4. Localització dels punts de presa de mostra de la RCSP a la conca de l'Ebre.....	134
Figura 3.5. Localització dels punts de presa de mostra d'aigües superficials i sediments a la conca del riu Ebre .....	136
Figura 3.6. Localització dels punts de presa de mostra d'aigües subterrànies i sòls a la conca del riu Ebre .....	137
Figura 3.7. <i>Loadings</i> dels components principals seleccionats amb la contribució dels diferents grups de fàrmacs investigats a les aigües de riu de la conca de l'Ebre.....	149
Figura 3.8. <i>Scores</i> descrivint les diferents contribucions de les mostres d'aigua superficial de la conca de l'Ebre durant el període 2006-08.....	150
Figura 3.9. Representació dels <i>scores</i> de PC2 respecte a PC3 de les mostres de fàrmacs agrupats per grups terapèutics en aigües superficials de la conca del riu Ebre durant el període 2006-08.....	151
Figura 3.10. Representació dels <i>scores</i> de PC1 respecte a PC2 de les mostres de fàrmacs en diferents tipus d'aigües de riu i de depuradora analitzades a la conca del riu Ebre durant el període 2006-08.....	152
Figura 3.11. <i>Loadings</i> i <i>scores</i> de PC1, descrivint la composició i la contribució de la primera pauta de contaminació d'antibiòtics identificada a les mostres d'aigua superficial de la conca de l'Ebre durant el període 2006-08 .....	154
Figura 3.12. Mapa de situació del delta de l'Ebre amb els principals nuclis urbans i zones humides d'interès natural .....	156
Figura 3.13. Situació de les estacions de presa de mostra en els canals d'irrigació i drenatge del delta del riu Ebre .....	159
Figura 3.14. <i>Loadings</i> (gràfics de l'esquerra) i <i>scores</i> (gràfics de la dreta) identificats pel primer component principal (PC) mitjançant l'aplicació de PCA .....	161
Figura 3.15. Comparació de la composició de les principals pautes de contaminació de productes fitosanitaris identificades als canals del delta de l'Ebre mitjançant l'aplicació de PCA i MCR-ALS.....	162
Figura 3.16. Situació de la conca del riu Llobregat i els seus embassaments .....	167
Figura 3.17. Distribució dels diferents usos agrícoles, industrials i urbans sobre el territori de la conca del riu Llobregat .....	169
Figura 3.18. Exemple de fitxer de tipus de sòl a la conca del riu Llobregat en format ascii.....	172
Figura 3.19. Informació d'entrada per al model PHYSITEL .....	173
Figura 3.20. Representació de les UHRHs calculades dins de PHYSITEL.....	174

Figura 3.21. Localització de les estacions hidrològiques i meteorològiques sobre el territori de la conca hidrogràfica del riu Llobregat.....	176
Figura 3.22. Exemple de fitxer d'índexs foliars utilitzats en la simulació hidrològica a la conca del riu Llobregat .....	177
Figura 3.23. Divisió de la conca del riu Llobregat en 6 subconques (grups d'UHRHs) per al procediment de calibratge HYDROTEL.....	178
Figura 4.1. Exemple de càlcul de l'índex diari de qualitat d'una massa d'aigua mitjançant la funció wqindex en MATLAB .....	201
Figura 4.2. Representació gràfica de l'índex prioritari i els índexs d'exposició i d'efecte obtinguts per 50 compostos orgànics a les conques hidrogràfiques catalanes.....	211
Figura 4.3. Representació gràfica de l'índex prioritari de contaminació a les localitzacions pel territori de Catalunya.....	213



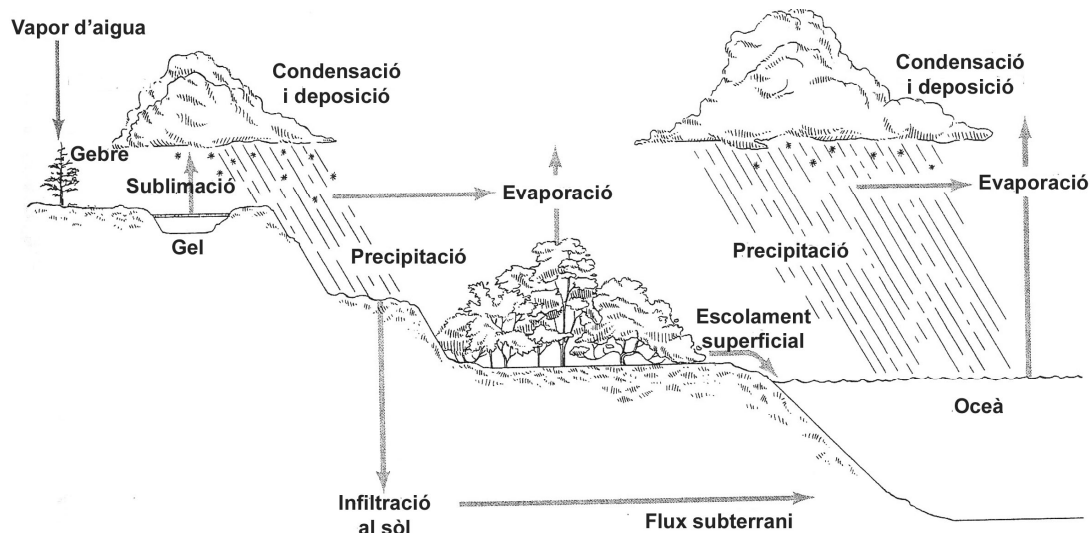


# 1. INTRODUCCIÓ

## 1.1. Caracterització de l'àmbit d'una conca hidrogràfica

Des de sempre, l'aigua ha estat sinònim de vida. Els rius són les venes de la terra; així que la xarxa fluvial constitueix la base de la vida continental de la mateixa manera que la xarxa limfàtica constitueix la base de la vida dels humans.

L'aigua al món es troba distribuïda entre els compartiments oceànic, terrestre i atmosfèric (Figura 1.1). La major part, un 97,2%, forma part dels oceans, mentre que només un 0,0001% es troba present en els corrents fluvials (*U.S. Geological Survey*). El cicle de l'aigua és un procés imprescindible d'intercanvi entre els diferents compartiments de la terra. Es tracta d'un sistema mundial enorme impulsat per l'energia solar, en el qual l'atmosfera proporciona un nexa vital entre els oceans i els continents. No obstant, al mateix temps que l'aigua és transferida d'un compartiment a l'altre, el mateix passa amb les substàncies contaminants que aquesta pugui contenir.



**Figura 1.1.** El cicle de l'aigua (Moran et al., 1986)

Aquella part de l'aigua que no s'infiltra dins de la terra o que no és evaporada, esdevé escolament superficial. Aquest escolament, es desplaça en forma de rius i torrents des de la terra fins al mar. De la mateixa manera, l'aigua subterrània, les fonts i la precipitació caiguda directament sobre els cursos fluvials, també contribueixen al seu cabal. L'àrea de territori que aboca l'aigua, sediments i substàncies dissoltes en un punt comú s'anomena conca hidrogràfica. La regió geogràfica d'una conca hidrogràfica es troba definida pel riu principal i pels afluents que desemboquen en aquest. El clima, la vegetació i la topografia d'una conca,

## Capítol 1

---

juntament amb les alteracions causades pels processos naturals i per les activitats humanes, tindran una afectació tant sobre la quantitat com sobre la qualitat de l'aigua de la conca.

En general, els rius han estat tipificats morfològicament en sentit longitudinal, dividint-los en el que s'anomena curs alt, curs mitjà i curs baix, o el que és el mateix, zona de capçalera, zona de transferència i zona de deposició. No obstant, avui en dia també es considera la seva morfologia des del punt de vista transversal, on en la majoria de casos es troben tres elements principals: el llit fluvial, per on discorre l'aigua al llarg de tot o part de l'any; la plana al·luvial o d'inundació, que en períodes de crescuda del riu i, amb freqüència irregular, pot quedar coberta per les aigües fluvials; i finalment es parla de la zona de transició, que es troba tocant a la plana d'inundació i que constitueix una connexió entre el riu i el paisatge o zona d'influència que l'envolta.

En el curs alt d'un sistema fluvial s'hi troba el naixement o zona fontinal, que acostuma a situar-se en àrees de muntanya o al marge de les valls. Quan l'aigua es mou pendent avall, la seva energia potencial es converteix en energia cinètica, i és una porció d'aquesta energia la responsable dels processos d'erosió. L'energia cinètica addicional és utilitzada per al transport de materials sòlids així com altres substàncies que entren en el riu, com poden ser residus industrials i domèstics. La zona del curs alt acostuma a estar formada per un llit de còdols o graves gruixudes i tendeix a tenir una gran concentració d'oxigen i una temperatura baixa de l'aigua. Es tracta d'aigua poc mineralitzada ja que majoritàriament rep aportacions de la precipitació atmosfèrica i de la fusió dels gels d'alta muntanya.

En el curs mitjà del riu, l'erosió disminueix a causa de la menor pendent. El llit sol estar format per graves fines o per sorres més o menys gruixudes, i comencen a assentar-s'hi plantes aquàtiques anomenades macròfits com a conseqüència de la reducció del transport de materials. En aquesta zona, la turbulència de les aigües fa que encara existeixi una incorporació força elevada d'oxigen atmosfèric. Aquí l'aigua conté una major concentració d'espècies químiques, fruit dels processos d'erosió i dissolució que tenen lloc al llarg del seu recorregut pel llit del riu. També es reben aportacions d'aigües subterrànies i aportacions antropogèniques, la qual cosa incrementa encara més la seva mineralització.

Al curs baix del riu es donen els fenòmens de deposició de sediments fins i gruixuts que han anat erosionant-se als cursos alt i mitjà, donant lloc a les grans planes al·luvials o extensions de la conca hidrogràfica inundades de manera periòdica a causa de les crescudes del riu. Al llarg del recorregut de la zona de deposició, la temperatura de l'aigua tendeix a augmentar. Aquest fet, possibilitarà l'existència d'una major concentració de nutrients la qual, al seu torn, fomentarà el desenvolupament d'algues. En aquells casos en què el creixement d'algues sigui excessiu, es poden originar fenòmens de dèficit d'oxigen.

Depenent de la dinàmica del riu i del material geològic per on discorre el seu llit, la desembocadura al mar pot donar lloc a un estuari, influït per les marees, o a un delta, formació

sovint triangular i formada per al·luvions del riu. A Catalunya hi trobem dos grans exponents d'aquesta darrera formació de la desembocadura: els deltes del riu Ebre i del riu Llobregat.

La dinàmica fluvial és una de les forces més poderoses de la natura, la qual ha anat modelant molts dels paisatges de la terra. Rius i torrents, més que cap altre element del paisatge terrestre, són vies de connexió naturals entre els sistemes més diversos des del seu naixement fins a la seva desembocadura al mar. A més, els rius també han estat artèries reguladores dels assentaments humans. La importància de l'aigua per a la vida humana ha fet que des de temps remots la humanitat hagi cercat la proximitat al medi aquàtic per a establir-se, degut a l'existència de diversos factors: sòls més productius, un subministrament garantit de l'aigua, un mitjà de transport econòmic, un potencial de pesca i recreatiu i, finalment, per raons estètiques. No obstant, convé no oblidar que les inundacions són fenòmens naturals que afecten a molts rius. Durant períodes de descàrrega anormalment llargs, el canal fluvial no és capaç d'incorporar tota l'aigua que li arriba, de manera que el seu excedent s'escampa ràpidament sobre la plana d'inundació. Lamentablement, molts assentaments humans sovint s'estenen dins el territori que constitueix aquesta zona, la qual cosa pot provocar seriosos problemes en períodes d'inundació.

En termes quantitius, es fa difícil establir recomanacions per a la predicció de la quantitat d'aigua disponible en àrees que presenten una forta dependència dels recursos aquàtics. Durant els períodes secs, en els quals l'escolament superficial cap al riu és proper a zero, la recàrrega del curs fluvial es realitza directament a partir de les aigües subterrànies o bé de les aigües que lentament s'alliberen des de les zones humides. S'anomena cabal base d'un riu a aquell subministrament d'aigua mínim necessari per al manteniment de les característiques essencials i de l'estructura de la comunitat natural aquàtica que s'hi troba present, basant-se en criteris estrictament hidrològics i propis de cada ambient fluvial considerat. Una gestió política adequada hauria de planificar l'aprofitament dels recursos aquàtics en funció d'aquest nivell de cabal base. Tot i això, convé estar preparats davant la possibilitat de llargs períodes de sequera, on fins i tot els nivells garantits de cabal base poden disminuir, la qual cosa requerirà la implementació de mesures de conservació de l'aigua.

Els materials que han estat rentats des de la superfície terrestre i incorporats al riu, juntament amb els sediments que han estat erosionats del canal, constitueixen la càrrega del riu. Aquesta càrrega és transportada en suspensió, en solució, i al llarg del llit del riu mitjançant la força del corrent. La composició química de l'aigua d'un riu pot ser molt variable, essent el resultat de la interacció entre l'aigua de la precipitació, el sòl i les roques, a la qual cosa cal afegir també les aportacions d'origen antropogènic. Aquesta composició, però, pot variar de manera estacional.

## Capítol 1

---

Així per exemple, durant la tardor i l'hivern es produeix habitualment un augment de la concentració de compostos orgànics com a resultat de la degradació de la biomassa acumulada al sòl i de l'arrossegament de material húmic. Per altra banda, a l'estiu, la disminució de les pluges sovint comporta una reducció del cabal del riu i, com a conseqüència d'això, un augment de la concentració de les espècies dissoltes.

Molts dels problemes que actualment tenen relació amb la contaminació de l'aigua tenen el seu origen en els diferents usos que existeixen dins el territori de la conca hidrogràfica (veure Figura 1.2). En termes generals, els usos presents en una conca es poden agrupar en tres categories diferents: i) àrees protegides per a la vida animal i/o recreatives, les quals es mantenen en un estat natural i relativament inalterat; ii) àrees destinades a activitats rurals o agrícoles (entre elles, les zones cultivades); iii) àrees que contenen activitats industrials i urbanes.

Cadascun d'aquests usos introduirà una barreja diferent de contaminants dins del medi d'una conca hidrogràfica, amb la conseqüent afectació sobre els seus diferents compartiments ambientals.

## 1.2. Problemàtiques que afecten la qualitat ambiental d'una conca hidrogràfica

A finals del segle XIX, quan moltes ciutats dels països desenvolupats experimentaven un fort creixement com a resposta al procés de la Revolució Industrial, les deixalles que s'acumulaven als carrers van començar a esdevenir un problema per a la població. La brutícia, juntament amb una manca de costums sanitaris que actualment es troben normalment assumits, va portar a la dispersió de malalties i epidèmies arreu del món. Molts dels recursos d'aigua utilitzats a les ciutats s'extreien directament de les aigües superficials continentals, les quals solien estar contaminades. Els esforços per tal de controlar les malalties, van començar amb la filtració de l'aigua de consum públic a través de sorres [1]. Més tard, a principis del segle XX, el clor va ser afegit a l'aigua municipal abans de ser conduïda a través de la xarxa de distribució. Aquest fet va fer que moltes de les malalties que s'havien escampat a través de l'aigua comencessin a esdevenir cada vegada menys importants. No obstant això, durant la primera meitat del segle XX, un altre tipus de contaminació, aquest cop d'origen químic, va començar a experimentar un important increment [1]. Degut a l'avenç de la tecnologia, que esdevenia cada vegada més sofisticada, les tipologies de residus generats també es van anar fent cada vegada més variats. Actualment, els recursos aquàtics són atacats per diferents productes de naturalesa química com són fertilitzants, productes fitosanitaris, detergents, quantitats traça de metalls pesants, efluents àcids procedents de mines, i substàncies radioactives, a més d'una enorme varietat de productes químics industrials. Els contaminants d'origen químic, constitueixen avui en dia una de les principals amenaces a la qualitat de les aigües (veure Figura 1.2).

Com ja anunciava l'escriptora Rachel Carson en la seva reconeguda obra *Silent Spring* (1964), *"la intoxicació que ha penetrat en les nostres aigües prové de molts orígens: residus radioactius procedents de reactors, laboratoris i hospitals; contaminacions causades per les explosions nuclears; residus domèstics en ciutats i pobles; residus químics de les fàbriques...A això cal afegir-hi una nova forma de contaminació: les polvoritzacions químiques aplicades a terrenys de cultiu, jardins, boscos i camps. Molts dels agents químics imiten i augmenten els perniciosos efectes de les radiacions i, dins dels propis grups de productes químics, existeixen interaccions sinistres i poc conegudes, transformacions i acumulacions d'efectes. Des de que els químics van començar a fabricar substàncies que la natura no va inventar mai, el problema de la purificació de l'aigua ha esdevingut complex i el perill dels usuaris de l'aigua ha augmentat"* [2].



**Figura 1.2.** Activitats de les conques hidrogràfiques que poden afectar la qualitat de les aigües superficials (Moran et al., 1986)

En les conques naturals, l'aigua (1) dissol i erosiona lentament la roca, i les zones humides (2) alliberen sediments que aniran a parar a l'aigua. En conques urbano-industrials, les plantes de tractament d'aigües residuals (3) no poden eliminar tots els residus de l'aigua; l'aigua de la pluja (4) conté residus que provenen del rentat dels carrers de les ciutats; les aigües residuals industrials (5) contenen una gran varietat de contaminants; aigües àcides s'escapen de la mineria a cel obert (6); aigua escalfada és alliberada de les plantes energètiques (7); gasos industrials són rentats de l'atmosfera constituint la pluja àcida (8); el transport de petroli (9) pot causar abocaments, incloent petits vessaments durant la seva descàrrega; i l'abocament no controlat de residus (10) pot contaminar l'aigua subterrània. En conques agrícoles, l'aplicació de productes fitosanitaris (11) incorpora aquests productes a l'escolament superficial; els fertilitzants aplicats de manera incorrecta (12) es dissolen a l'aigua d'escolament; i els residus animals (13) són rentats de les superfícies agrícoles.

Els efectes que la contaminació química comportarà a llarg termini encara no es troben prou ben entesos per la societat en el moment actual. Tot i això, els efectes a curt termini són prou visibles - la mortalitat de peixos que acompanya episodis accidentals d'abocaments de substàncies químiques, platges inutilitzables, o llacs eutrofitzats pel creixement massiu de plantes aquàtiques, entre d'altres - .



Actualment, moltes indústries i municipalitats actuen d'acord amb les directrius establertes a la legislació vigent, les quals requereixen uns mínims estàndards de qualitat per als efluent. No obstant, les directrius per al control de petites descàrregues de substàncies amb una elevada toxicitat encara no han estat totalment establertes. Tot i que actualment s'estan elaborant diverses propostes per tal de regular la presència d'aquelles substàncies considerades potencialment perilloses pel medi ambient, a la pràctica, la seva inclusió en un text legal passa per procediments que encara són massa lents.

### 1.2.1. Legislació europea

La legislació en matèria d'aigües actualment vigent dins el territori de la Unió Europea es troba regulada per la Directiva 2000/60/CE o Directiva Marc de l'Aigua (DMA) [3]. Aquesta Directiva ha estat posteriorment modificada per una altra recentment aprovada pel Consell Europeu, la Directiva 2008/105/CE, que estableix els nivells màxims per a les substàncies prioritàries per tal de complir els estàndards de qualitat en aigües de riu.

Pel que fa a les aigües subterrànies, la Directiva més recent és la 2006/118/CE [4], específicament elaborada per a aigües de subsòl. Aquesta Directiva no modifica els nivells màxims permesos de plaguicides en aigües respecte a la Directiva anteriorment vigent, la 98/83/CE [5], relativa a les aigües de consum humà. Per hidrocarburs aromàtics policíclics, continuen estant igualment en vigor els límits establerts per la Directiva de l'any 1998.

La legislació referent a productes fitosanitaris d'ús agrícola, experimenta modificacions contínuament. Sovint s'elaboren directrius específiques per a un sol compost, com és el cas de l'atrazina en la Decisió 2004/248/CE [6]. També cal destacar algunes Directives que regulen altres tipus de compostos com, per exemple, els alquilfenols, substàncies que constitueixen una amenaça per al medi ambient i que són extensament utilitzades en la indústria i en les formulacions de productes fitosanitaris. Els alquilfenols es troben regulats des del 1986 per la Directiva 86/278/CEE [7] pel que fa a la seva presència en fangs de depuradora, i des del 2003 es regula la seva comercialització i ús en els estats membres a través de la Directiva 2003/53/CE [8].

Fins al moment, els compostos estudiats en aquesta Tesi no es troben legisllats en sòls. Existeixen només alguns acords no vinculants nascuts del Conveni de Rotterdam (1998), el qual proposa la promoció de la responsabilitat compartida en el comerç internacional de certes substàncies químiques perilloses per tal de protegir la salut humana i el medi ambient; i del Conveni d'Stockholm (2001), que pretén regular la contaminació per contaminants orgànics persistents (POPs) en el medi. Aquest darrer, estableix el que s'anomena la "dotzena bruta" (*dirty dozen*). En el cas dels sediments, de la mateixa manera que ocorre amb els sòls, no



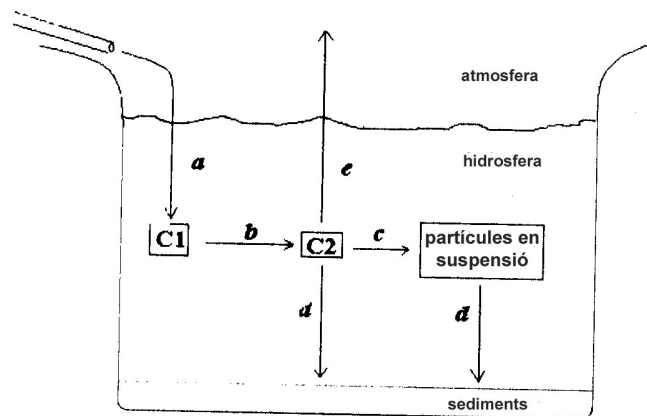
existeixen límits legiscats a Europa. Per tal de complir la legislació vigent, simplement és necessària una millora del nivell de contaminació respecte a l'any anterior.

### 1.2.2. La contaminació de les aigües superficials

Qualsevol episodi de contaminació, sigui del tipus que sigui, estarà constituït pels següents components:

- 1) una font de producció del contaminant,
- 2) un medi de transport i de transformació física i química, i
- 3) un medi receptor

La font d'origen de la contaminació determina la naturalesa química del contaminant emès, el medi on aquest s'aboca i el flux d'emissió. Un cop un contaminant (contaminant primari C1, veure Figura 1.3) és introduït en el medi, aquest es dispersa associant-se a les diferents fases que constitueixen el medi en funció de la seva naturalesa química. La solubilitat d'un contaminant serà el paràmetre que determinarà la seva dissolució en el medi hídric. Dins de l'aigua, el contaminant pot ser transportat a través de múltiples mecanismes. L'advecció, mecanisme de transport pel qual el contaminant es desplaça incorporat al moviment de la massa d'aigua, n'és un dels principals (Figura 1.3a). Altres mecanismes són la difusió molecular i la difusió turbulenta. Aquest últim, és el mecanisme més rellevant que es dona a les capçaleres dels rius, on els pendents són més acusats i, per tant, predomina el règim turbulent.



**Figura 1.3.** Cicle d'un contaminant a la hidrosfera (Domènech, 1999). C1) contaminant primari, C2) contaminant secundari, a) transport, b) transformació química, c) sorpció, d) sedimentació i e) volatilització

Les transformacions químiques més rellevants que el contaminant pot experimentar en medi hídric són les reaccions d'hidròlisi, l'oxidació deguda a la presència d'oxigen dissolt, els processos de reducció en reservoris profunds i on la concentració d'oxigen és molt baixa i, finalment, les reaccions fotoquímiques, especialment en zones superficials on l'exposició a la

llum solar és més alta (Figura 1.3b). En el cas de transformació química, s'obindrà un contaminant secundari C2 (veure Figura 1.3). D'altra banda, el contaminant també pot desaparèixer del medi per volatilització (Figura 1.3e), procés que afecta a les espècies més volàtils, les quals aniran a parar a l'atmosfera, o bé pot acabar dipositant-se en els sediments en cas que es produeixin reaccions de precipitació (Figura 1.3d). A més, en el medi hídric s'hi troben partícules sòlides en suspensió amb les quals el contaminant es pot associar (Figura 1.3c) i posteriorment sedimentar. En la majoria de casos, les partícules en suspensió provenen del sòl, arrancades a través de processos d'arrossegament. Es tracta de partícules de tipus argilós, amb una elevada superfície específica i capaces d'adsorbir i retenir espècies químiques. A l'aigua s'hi troba també material orgànic particulat, procedent del sòl i generat pels propis organismes aquàtics, el qual és capaç de retenir compostos químics hidròfobs.

### **Fonts de contaminació puntual i difosa**

Les fonts de contaminació descrites per diferents aportacions procedents de punts que es troben distribuïts dins una àrea determinada, i no a partir d'una única font clara i definida com seria, per exemple, una canonada, s'anomenen fonts de contaminació difosa. Contràriament a les fonts de contaminació puntual, que consisteixen en residus concentrats, provinents per exemple, de la indústria o de plantes de tractament d'aigües residuals, la contaminació per fonts difoses és especialment difícil i cara de controlar ja que malgrat trobar-se a concentracions relativament baixes, els seus volums en el medi ambient són enormes.

Les conques industrials i urbanes contenen una gran quantitat de fonts de contaminació puntual. Tot i que la quantitat de residus no tractats que van a parar directament als cursos fluvials d'aquestes àrees ha disminuït en els últims temps, la xarxa de clavegueram encara representa una problemàtica important pel que fa a la contaminació de l'aigua superficial, ja que no totes les ciutats disposen de sistemes de tractament de les aigües residuals, i no totes les indústries netegen l'aigua que aboquen directament a la llera. Els efluents d'algunes indústries poden ser tractats per les plantes municipals, però de vegades aquests són tan tòxics que les plantes no els accepten. És per aquest motiu, que les indústries que envien els seus efluents a les plantes de tractament han de limitar-ne el nivell de toxicitat mitjançant processos de pretractament. El problema és que no totes les fàbriques estan equipades de manera adequada per tal de dur a terme aquest tipus de processos i, a banda d'això, també poden existir abocaments accidentals. Com a resultat de l'ajust dels estàndards de qualitat de l'aire i de l'aigua, moltes indústries i municipalitats s'han vist obligades a enterrar els residus més sòlids enlloc de processar-los com a efluents. Es tracta d'un procés que normalment protegeix les aigües superficials, però que constitueix una amenaça per a les aigües subterrànies quan els emplaçaments no estan ben escollits, construïts o bé gestionats.

### **Contaminants naturals**

Les àrees naturals, com els boscos, pantans, prats, etc., generalment contribueixen amb petites quantitats de materials als cursos fluvials. Les aigües superficials que s'escolen a través d'aquests sistemes contenen concentracions "normals" de substàncies dissoltes que són característiques d'aquests medis. D'altra banda, la composició dels sediments varia considerablement i, com a resultat, l'aigua que ha estat en contacte amb aquests presentarà un contingut diferent de material mineral en funció de la naturalesa del sediment. Però no tota l'aigua que prové de les àrees naturals presenta un nivell de qualitat bo. Per exemple, aquells torrents que flueixen a través de zones pantanoses presentaran una coloració més groguenca i un pH més àcid, ja que hauran incorporat àcids orgànics generats per organismes descomponedors presents en els medis pantanosos. Un altre fenomen diferent s'observa en aquells rius que s'escolen a través d'àrees constituïdes per materials rics en calci, magnesi i bicarbonats. Encara que aquests compostos no representin un perill per a la vida aquàtica, provocaran un augment de la duresa de l'aigua del riu. Així, per exemple, a més duresa de l'aigua, la quantitat de detergents que haurà de ser aplicada per a la neteja, haurà de ser també força més elevada i esdevindrà, al seu torn, un altre problema de contaminació ambiental.

### **Contaminants antropogènics**

Una àmplia varietat de contaminants generats per l'activitat humana existent dins de l'àmbit d'una conca hidrogràfica seran detectats a les aigües superficials. A continuació es descriuen aquells contaminants que apareixen amb més freqüència en les problemàtiques ambientals de les conques hidrogràfiques estudiades en aquesta Tesi.

#### **▪ Metalls pesants**

Els elements metàl·lics són utilitzats en grans quantitats en la indústria per a una variada gamma d'aplicacions (catalitzadors, pintures, detergents, materials de construcció, additius, pesticides, etc.). Bona part d'aquests elements s'incorpora al medi en forma iònica i, per tant, soluble. Encara que molts elements metàl·lics són micronutrients, és a dir, nutrients necessaris en petites quantitats pel desenvolupament dels organismes vius, quan aquests excedeixen una determinada concentració esdevenen perjudicials. Els metalls que contenen arsènic, cadmi, crom, coure, plom, níquel, platí, zinc, tal·li, vanadi i or, estan classificats com a metalls pesants i considerats perillosos. Alguns com l'arsènic, cadmi, níquel i algunes formes del crom, es consideren, a més, carcinògens. Molts metalls pesants actuen com a inhibidors d'enzims, els quals afecten als processos metabòlics dels organismes, i tenen el perill afegit que no poden ser degradats a formes no tòxiques. Així, quan els medis aquàtics són contaminats per algun d'aquests compostos, esdevenen una amenaça potencial durant anys.

### ▪ Nutrients

La presència de nutrients com el nitrogen i el fòsfor en quantitats excessives a l'aigua, provoca un ràpid creixement de les algues i altres plantes verdes, les quals recobreixen la superfície de les aigües i impedeixen que la llum solar arribi a les capes inferiors. A conseqüència de la descomposició de la biomassa generada, es consumeix l'oxigen de l'aigua i això dóna lloc al fenomen de l'eutrofització.

Existeixen diverses fonts de nutrients que contribueixen a l'eutrofització. Una d'aquestes fonts són els residus domèstics. Aproximadament el 50% del fòsfor contingut en aquest tipus de residus prové dels detergents. L'escolament urbà també constitueix una font important de nitrogen i fòsfor. Aquests materials provenen de la fertilització de la gespa, de material fecal d'origen animal, i de productes de la combustió, entre d'altres. Les entrades de nutrients procedents de la indústria són molt variables. Les indústries papereres contribueixen amb enormes quantitats de substàncies que presenten grans demandes d'oxigen però alliberen baixes concentracions de nutrients per a les plantes. D'altra banda, aquelles indústries que disposen d'extenses àrees que necessiten ser netejades, com ara els rentadors de cotxes o les fàbriques de productes làctics, utilitzen grans quantitats d'agents de neteja que contenen fòsfor. L'explotació minera de fosfats és una altra font de contaminació important. Finalment, en àrees agrícoles on es realitza una aplicació important de fertilitzants, la quantitat de nutrients que poden anar a parar a les aigües també és significativa.

### ▪ Compostos orgànics

La contaminació del medi hídric per compostos orgànics és un fenomen cada vegada més important degut a l'augment de la producció de substàncies sintètiques. Les fonts de contaminació més rellevants per compostos orgànics són:

- aigües industrials i municipals, que contenen una variada gamma de compostos orgànics sintètics entre els que destaquen els hidrocarburs aromàtics (benzè, fenol i hidrocarburs aromàtics policíclics), compostos halogenats (alifàtics i aromàtics), bifenils policlorats, detergents, etc...
- aigües d'escolament, les quals no corresponen a abocaments puntuals sinó que arrosseguen contaminants generats en grans zones, com àrees rurals (productes fitosanitaris, residus animals, fertilitzants) i àrees urbanes.
- aigües de precipitació, les quals en algunes circumstàncies poden contenir compostos orgànics volàtils (dissolvents, hidrocarburs, productes fitosanitaris, etc...) o substàncies amb un elevat pes molecular adsorbides a les partícules sòlides (bifenils policlorats, dioxines, etc...).

## Capítol 1

---

Els residus industrials i municipals acostumen a contenir concentracions elevades de substàncies orgàniques. La seva presència potencia el creixement dels organismes descomponedors, que redueixen la quantitat d'oxigen dissolt de l'aigua, i a conseqüència d'això, altres membres de la comunitat aquàtica, especialment peixos i mol·luscs, es troben mancats d'oxigen i fins i tot poden arribar a morir.

Si en algun moment s'arriba a una eliminació completa de l'oxigen del riu, ocorre un canvi en el tipus d'organismes descomponedors, i es passa dels descomponedors aerobis (que requereixen oxigen) a un tipus de descomponedors anaerobis (que requereixen absència d'oxigen). Els productes d'aquests descomponedors també difereixen. Mentre que els aerobis generen sobretot diòxid de carboni, aigua, nitrats i sulfats, els quals generalment no representen cap perill, els anaerobis es caracteritzen per la formació de metà, amoni i sulfur d'hidrogen, que són gasos perillosos. Sota condicions anaeròbies, les aigües residuals es converteixen en ambients pútrids i tèrbols, condicions sota les quals cap organisme (a excepció de les bacteries anaeròbies) és capaç de sobreviure.

La quantitat d'oxigen dissolt que necessiten els descomponedors per tal de descompondre els materials orgànics presents en un volum determinat d'aigua s'anomena Demanda Biològica d'Oxigen (DBO). Així, la DBO serà una mesura del nivell de contaminació orgànica que es troba present en aigües residuals.

### **Productes fitosanitaris**

Els productes fitosanitaris són substàncies naturals o sintètiques que serveixen per combatre els agents responsables de les plagues (fongs, insectes, males herbes, etc.). S'introdueixen al medi hídric per mitjà de l'escolament superficial, infiltracions a través del sòl cap a les aigües subterrànies, introducció directa en el moment de la seva aplicació i abocaments per part de les indústries que els produeixen, a través de les aigües residuals. En general, es tracta de substàncies que presenten una baixa solubilitat en aigua i, per tant, una elevada tendència a adsorbir-se a les partícules sòlides i a dipositar-se en els sediments. Així, la seva concentració en el medi aquós sol ser baixa. Normalment, els plaguicides organoclorats (el més conegut, el DDT) són els que presenten una major resistència a la degradació. D'entre tots els productes fitosanitaris que van a parar a les aigües superficials i que posteriorment s'acumulen a les cadenes tròfiques, els plaguicides organoclorats són els que representen un major risc per a les espècies i el medi ambient en general. L'Agència de Protecció Ambiental (EPA) ha prohibit l'ús general de molts compostos organoclorats, com per exemple, el DDT, aldrí, dieldrí, heptaclor, clordà, i toxafè, entre d'altres.

### **Bifenils policlorats (PCBs)**

Alguns compostos clorats industrials com els PCBs també es bioacumulen. Com que es tracta de compostos resistents al foc i estables a elevades temperatures, s'utilitzen com a materials aïllants en diferents tipus d'aparells i en la fabricació d'alguns plàstics. Com a resultat del seu ús extens, els PCBs poden entrar als ecosistemes aquàtics mitjançant diverses vies. Amb la combustió de paper i plàstic que conté PCBs, per exemple, aquests es vaporitzen i condensen en les partícules de pols que després es dipositaran al sòl o bé a les aigües superficials o que seran rentades de l'atmosfera amb les precipitacions. A més, aquests compostos també poden entrar als ecosistemes aquàtics per escolament superficial, evaporació a partir d'abocaments accidentals, o com a efluents de les plantes de tractament, etc.

### **Alquilfenols (APs)**

Es tracta de substàncies que presenten excel·lents propietats surfactants utilitzades en aplicacions industrials, agrícoles i domèstiques. Es troben en els productes de neteja, tèxtils, petroli, resines fenòliques, vernissos, en la indústria del paper i en la formulació de diversos pesticides. Una de les principals fonts d'alquilfenols (APs) en el medi ambient són les entrades directes d'efluents urbans o industrials, o bé a través dels efluents de les estacions depuradores d'aigües residuals (EDARs), les quals en moltes ocasions no són capaces d'eliminar completament aquestes substàncies durant els processos de tractament de les aigües, de manera que els nivells d'APs al sistema aquàtic incrementen. Són compostos que actuen com a disruptors endocrins (capaços d'alterar el funcionament normal del sistema endocrí dels animals). El nonilfenol (NP), per exemple, és el producte de degradació majoritari dels anomenats alquilfenols etoxilats, i presenta més toxicitat i més potencial estrogènic que els seus compostos parentals.

### **▪ Substàncies farmacèutiques**

Les substàncies farmacèutiques són un tipus dels anomenats nous contaminants emergents que, en els darrers anys, han anat adquirint importància degut a l'augment de la seva presència en el medi. Els productes farmacèutics en la seva forma original o bé com a metabòlits són contínuament introduïts a les aigües residuals, sobretot a través de l'excreció, l'eliminació de medicaments sense utilitzar o bé caducats, o directament a partir d'abocaments per part de les indústries farmacèutiques. Aquests compostos es troben distribuïts per tot el medi aquàtic, ja que alguns no són eliminats de manera eficient durant els processos de tractament d'aigües residuals, de manera que poden arribar a afectar les aigües superficials i subterrànies i, per tant, l'aigua de beguda. Pel fet de ser contínuament introduïts al medi ambient a través de les

aigües residuals, aquests productes poden ser classificats com a contaminants pseudopersistentes.

### ▪ Altres compostos tòxics

Alguns compostos químics que s'utilitzen en la indústria comporten perills directes tant per als humans com per a la vida aquàtica. Alguns exemples en són l'amoní, cianurs, sulfurs, fluorurs, àcids forts i alcalins. Molts d'aquests compostos s'utilitzen en les seves formes concentrades, que són directament tòxiques per als organismes però indispensables per a la indústria. La descàrrega accidental d'aquests residus tòxics, generalment fruit de l'error humà, pot arribar a eliminar les comunitats aquàtiques que es troben presents riu avall de l'abocament.

L'exposició d'un organisme a substàncies tòxiques pot ser agut (com a resultat d'una sola dosi forta) o crònic (com a resultat de dosis més petites però repetides en el temps). El perill que comporta la fuga de petites quantitats de materials tòxics al medi ambient, encara que d'entrada sembli no representar una amenaça immediata per a la salut humana, és un problema tant important com podria ser-ho un gran abocament industrial. Es tracta de substàncies que tenen la capacitat de bioacumular-se en els organismes i que per tant, seran posteriorment transmeses a través de la cadena tròfica.

De la mateixa manera que els nutrients passen d'un nivell tròfic al següent, el mateix passa amb les substàncies tòxiques persistents. Si aquestes substàncies no són excretades o eliminades biològicament per l'organisme, es queden retingudes en els seus teixits. En el cas que l'organisme continuï ingerint aliment contaminat durant un cert període de temps, la concentració de l'agent contaminant va augmentant dins de l'organisme. És per això que els contaminants persistents que entren a les cadenes tròfiques es van fent cada vegada més concentrats a mesura que passen a un nivell tròfic superior. D'aquesta manera, els organismes que se situen en els nivells superiors de la cadena (entre ells els humans) es troben exposats a unes concentracions més elevades d'aquestes substàncies. Es poden identificar algunes propietats químiques que afavoreixen el procés de bioacumulació. Per exemple, aquelles substàncies més liposolubles tindran més tendència a bioacumular-se, així com els compostos que siguin més químicament estables, o sigui, capaços de persistir en el medi durant períodes de temps més llargs. Per tal de ser persistent, un compost ha de ser resistent a la biodegradació per part dels organismes, a la reacció amb aigua o amb la humitat (hidròlisi), o bé a la degradació per la llum solar (fotòlisi).

### 1.2.3. La contaminació de les aigües subterrànies

Les mateixes activitats que provoquen la contaminació de les aigües superficials d'una conca hidrogràfica, també poden arribar a contaminar les aigües subterrànies. L'abocament de residus, les activitats agrícoles, les fosses sèptiques, les llacunes que emmagatzemen residus industrials, els pous d'injecció subterrània (en els quals els residus industrials són bombejats a pressió per sota de les capes de roca permeable), i la producció de petroli i gas natural, s'identifiquen com a possibles amenaces a la qualitat de les aigües subterrànies. Cal tenir en compte que aquesta qualitat també variarà depenent de quina sigui la font d'aigua subterrània, a més de les interaccions que existeixin entre aquesta aigua i el sòl i materials rocosos a través dels quals circuli.

#### ▪ Nutrients

La infiltració de l'aigua a través del sòl és un mecanisme de purificació bastant eficient. De fet, en el cas de substàncies com els fosfats, aquests s'uneixen a les partícules de la superfície del sòl i són eliminats (per tant, no passen a les aigües subterrànies). No obstant, a diferència dels fosfats, els nitrats presenten una problemàtica major. Donat que són substàncies molt solubles en aigua, no s'adsorbeixen a les partícules del sòl, i són principalment perillosos quan van a parar a aigües dels pous. Les concentracions de nitrats més elevades en aigües subterrànies es detecten en àrees on existeix una gran aplicació de fertilitzants agrícoles.

#### ▪ Compostos orgànics

##### Hidrocarburs

Un altre problema de contaminació que amenaça freqüentment les aigües residuals és el degoteig de tancs d'emmagatzemament de combustible. Les estacions de servei, les granges i moltes indústries i institucions emmagatzemen aquestes substàncies en tancs d'acer ubicats sota terra, els quals es troben exposats a la corrosió i per tant, a fuites del seu contingut. Els sòls no retenen ni la benzina ni altres substàncies derivades del petroli, i les petites fuites no acostumen a ser detectades fins que no es descobreix algun pou contaminat. Però arribat aquest moment, el problema ja es considera seriós, pel fet que una petita quantitat de gasolina és capaç de contaminar un gran volum d'aigua.

##### Residus municipals, industrials i miners

Els residus municipals, industrials i miners també poden contaminar les aigües subterrànies. Es tracta de residus emmagatzemats normalment sota terra, on l'aigua pot infiltrar-se i dissoldre'n els compostos solubles, incorporant aquest tipus de contaminació. L'aigua que passa a través d'una zona de recàrrega contaminada pot arribar a contaminar les reserves d'aigua subterrània



durant un període de centenars i fins i tot milers d'anys, ja que dins de l'aqüífer gairebé no existeix dilució ni tampoc eliminació de contaminant. Les activitats mineres també poden contaminar les aigües subterrànies de diverses maneres. Algunes alliberen metalls pesants tòxics, altres augmenten l'acidesa de l'aigua, i n'hi ha que canvien els nivells i patrons de circulació de l'aigua subterrània. L'element clau per a la protecció de l'aigua subterrània consisteix, doncs, en evitar qualsevol opció de contaminació.

### **Productes fitosanitaris**

Els productes fitosanitaris no haurien de ser utilitzats en zones de recàrrega on els sòls no són capaços d'aturar-ne la migració. Tot i que es tracta de productes necessaris per a l'agricultura, alguns d'ells, i sobretot els més solubles, han estat freqüentment detectats a les aigües subterrànies de les àrees irrigades.

## **1.2.4. La contaminació del sòl**

El sòl ha estat, des de sempre, l'abocador per excel·lència de bona part dels residus originats per l'home, degut a la seva fàcil accessibilitat. Tan sols la pressió demogràfica ha fet que la recerca de terrenys suficientment allunyats dels nuclis urbans resulti cada cop més difícil i inciti a buscar altres medis naturals (com són la hidrosfera i l'atmosfera).

A diferència del que ocorre en el medi hídric, un contaminant té, en principi, poca mobilitat en el sòl. A més, els efectes de dilució característics d'altres medis tan sols juguen un paper marginal en aquest compartiment. És per aquest motiu que la probabilitat de que un contaminant s'incorpori a la cadena tròfica a través de la seva ingestió per part d'algun organisme que tingui el sòl com a substrat vital, és molt elevada.

Ja s'ha esmentat el paper depurador que presenten determinats components del sòl. En aquest sentit, cal destacar la funció dels materials húmics, els quals són capaços de fixar una elevada varietat de substàncies dissoltes en l'aigua edàfica, en particular, cations polivalents de metalls pesants. Per tant, la persistència dels contaminants tendeix a ser major en sòls amb nivells superiors de matèria orgànica.

### **▪ Nutrients**

Els residus de les granges contribueixen a les entrades d'ió amoni al sòl. El nitrogen orgànic, que és la forma sota la qual està present aquest element en els residus animals, es transforma en ió amoni sota l'acció catalítica de determinats bacteris anomenats amonificants. Altres aportacions de nitrogen són degudes a l'ús de fertilitzants, emissions de la indústria i abocaments d'aigües residuals domèstiques.

Fins fa poc, la generació i posterior aplicació de residus animals no suposava cap tipus de problemàtica ambiental. Històricament, els fems s'han aprofitat per a la fertilització del sòl gràcies al seu elevat valor nutritiu per les espècies vegetals, ja que no només contenen nutrients minerals sinó també una elevada proporció de matèria orgànica. No obstant, actualment es produeixen quantitats tan grans de residus animals que el sòl és incapaç d'assimilar-les, sobretot quan aquests són aplicats de manera incontrolada, sense tenir en compte ni l'època de l'any ni la quantitat, tan sols guiant-se per la necessitat de buidar les fosses. L'aplicació al sòl d'un excés de fertilitzant pot originar, a la llarga, problemes de contaminació de l'entorn.

En certes ocasions, els residus animals contenen antibiòtics o bé altres productes farmacèutics que procedeixen de la seva ingestió per part del bestiar en cas de malaltia i que no han estat assimilats per l'animal. En el cas dels antibiòtics, quan aquests passen al sòl poden alterar-ne significativament l'ecologia microbiana i, com a conseqüència, afectar els processos bioquímics que hi tenen lloc.

L'aplicació de dosis elevades de residus animals al sòl pot conduir a una salinització excessiva d'aquest, de manera que el creixement de les plantes es vegi limitat. A més, un excés de sals pot produir una desestabilització del sòl que ocasioni una disminució de la seva capacitat d'infiltració.

### ▪ **Metalls pesants**

El procés de biometilació origina la mobilització de metalls pesants tòxics, ja que produeix una reacció entre ions metàl·lics i determinades substàncies orgàniques naturals per tal de formar compostos organometàl·lics. En aquesta situació, les propietats físico-químiques del metall canvien radicalment. Els compostos resultants solen ser molt liposolubles, de manera que travessen amb facilitat les membranes biològiques i, en particular, la pell. Un dels processos de biometilació més importants és l'enllaç entre un catió metàl·lic i el grup metil (-CH<sub>3</sub>). Aquest és el principal mecanisme natural de mobilització de metalls pesants. Els metalls que ofereixen més afinitat a la biometilació són el mercuri, el plom, l'arsènic i el crom. Aquests metalls es troben presents en molts tipus de sòls a concentracions molt baixes, de manera que normalment es tracta d'un fenomen de poca transcendència. El problema esdevé més important quan aquests metalls són introduïts al sòl de manera externa, en forma d'abocaments incontrolats. En aquests casos, la probabilitat que aquests metalls s'incorporin a la cadena tròfica és bastant elevada.

### ▪ **Compostos orgànics**

#### **Productes fitosanitaris**

## Capítol 1

---

Els productes fitosanitaris poden ser aplicats directament sobre el sòl o bé en forma d'esprai, ruixant la superfície de les fulles de les plantes. En aquest darrer cas, una part del producte persisteix absorbit sobre la superfície de la planta mentre que la resta es dissol o és arrossegat per l'aigua de la pluja fins a incorporar-se al sòl.

Un cop al sòl, el producte pot migrar a través d'aquest i arribar fins a les aigües subterrànies. La facilitat amb què es porti a terme aquest desplaçament dependrà de la naturalesa i propietats químiques del producte aplicat, de la seva formulació, del mètode d'aplicació, de l'entorn on s'aplica (existència de pendents, grau de drenatge del terreny, etc...), així com del tipus d'espècie cultivada. En principi, aquells productes fitosanitaris que siguin poc solubles en aigua, poc volàtils i químicament estables, romandran al sòl durant un període de temps llarg.

Els productes fitosanitaris que més s'utilitzen en l'actualitat són els compostos organoclorats, els organofosforats i els carbamats. Els plaguicides organoclorats van ser els primers en ser comercialitzats i són, de fet, els responsables d'un impacte més negatiu sobre el medi ambient, ja que el seu efecte letal es segueix mantenint durant anys després de la seva aplicació. El DDT, que és un compost relativament fàcil de sintetitzar i econòmic en relació a altres compostos d'igual efectivitat, mata els insectes per simple contacte, per alteració del seu sistema nerviós. D'altra banda, els compostos organofosforats són menys estables que els seus anàlegs organoclorats i, per tant, menys persistents en el medi. No es bioacumulen, de manera que són menys nocius i presenten una menor probabilitat de passar a la cadena tròfica. No obstant, alguns com el metilparatió, són molt tòxics. Els carbamats també actuen sobre el sistema nerviós però són més innocus que els pesticides organofosforats.

L'aplicació indiscriminada de productes fitosanitaris al sòl, pot originar problemes greus per a l'equilibri ecològic del medi. Amb el pas del temps, si es continua aplicant el mateix producte i noves generacions de la plaga es van desenvolupant, són cada vegada més els individus que hi esdevenen resistents. Això obliga a l'agricultor a aplicar el producte amb més freqüència i en majors quantitats, de manera que l'efecte negatiu del compost es multiplica. A més, hi ha agricultors que utilitzen una bateria de productes fitosanitaris, és a dir, un conjunt de productes de naturalesa química diferent però que s'apliquen al mateix temps. En aquest cas el risc d'acumulació de residus continua essent elevat.

Un altre problema relacionat amb l'aplicació d'aquests productes és que tenen poca selectivitat, és a dir, que són poc específics respecte a una espècie determinada. Com a conseqüència d'això, quan s'aplica el producte fitosanitari existeixen moltes possibilitats de que altres espècies que cohabituen juntament amb les causants de la plaga resultin també afectades.

Tot i que existeix una necessitat real d'aplicar productes fitosanitaris per millorar el rendiment de les collites, és necessari un canvi en les pràctiques agrícoles, de manera que es vagi tendint

cap a una agricultura menys intensiva, però més selectiva. Es fa necessari el desenvolupament de noves metodologies compatibles amb el medi natural. En aquest sentit, el producte ideal seria aquell que presentés una elevada toxicitat respecte a l'espècie a eliminar, que persistís en el medi el temps just per tal d'assolir una elevada taxa d'eliminació, i que no afectés a la resta d'organismes vius de l'entorn.

### **Hidrocarburs aromàtics policíclics (PAHs)**

Els PAHs es poden formar de manera natural (formació geològica de combustibles fòssils), mitjançant processos de combustió, tant natural com antropogènica (incendis forestals, transports, incineradores, centrals elèctriques...), o per síntesi industrial. Solen acumular-se a la matèria orgànica del sòl, però depenent de la seva massa molecular, també poden tendir a evaporar-se. Degut a la seva difícil degradació (especialment els més pesants) es poden classificar com a contaminants ubics. La via més habitual d'eliminació al sòl és mitjançant la degradació microbiana. Tot i que els animals metabolitzen aquest tipus de substàncies, es pot donar bioacumulació de PAHs a les plantes.

En molts casos, els residus que es troben al sòl provenen d'abocaments indirectes, com és el cas dels òxids de sofre i de nitrogen emesos principalment a causa de la combustió de carburants fòssils. Bona part d'aquests òxids finalitzen el seu recorregut dipositant-se al sòl. El sòl rep també altres aportacions aèries com són, per exemple, les partícules sòlides que, quan provenen d'àrees urbanes i industrials, poden transportar una càrrega química complexa [9].

## **1.2.5. La contaminació dels sediments**

Quan les gotes de pluja cauen sobre el sòl desprotegit (sense vegetació) arrenquen les partícules de sòl. Aquestes partícules són transportades a través de l'escolament superficial cap a rius, llacs o oceans i són dipositades allí com a sediments. La part grollera del sediment, com són les sorres, es diposita força ràpidament. No obstant, les partícules més fines com les argiles, poden continuar en suspensió durant mesos, afavorint una major torbesa de l'aigua. Aquells sòls que es deixen sense cultivar, ja sigui per una rotació de conreus (guaret), la tala de fusta, la mineria a cel obert, la sobrepastura, la construcció de carreteres o altres activitats de construcció, estan subjectes a elevades taxes d'erosió.

Sovint, els sediments acaben emmagatzemant els contaminants hidròfobs que van a parar a les aigües superficials i que, després d'adsorbir-se a les partícules en suspensió, acaben sedimentant. Aquest seria el cas de substàncies com els productes fitosanitaris, els hidrocarburs aromàtics i els alquilfenols, entre d'altres. Els sediments constitueixen el compartiment ambiental que aporta un registre més fiable de la contaminació que afecta al riu,

## Capítol 1

---

ja que a diferència de l'aigua superficial, que flueix contínuament, el sediment és un medi bastant estàtic.

## 1.3. Metodologies aplicades a l'anàlisi i avaluació de la qualitat ambiental d'una conca hidrogràfica

L'estudi de l'àmbit d'una conca hidrogràfica inclou un ampli ventall d'aspectes, ja que aquesta es troba constituïda per diversos compartiments ambientals amb característiques molt diferents. Al llarg del temps, s'han anat desenvolupant diverses maneres d'estimar la càrrega de substàncies transportada per un riu, o bé l'escolament superficial que existeix en una determinada conca. En els estudis ambientals, és important separar aquelles contribucions que provenen de fonts diferents, a més de distingir entre la variabilitat natural i l'impacte antropogènic. Això permetrà realitzar un control ambiental més eficient i seleccionar les tècniques de gestió més adients en cada cas.

Seria difícil, a més de força car, obtenir una bona descripció dels diferents fenòmens que es donen dins del sistema d'una conca hidrogràfica basant-se només en mesures directes. És per aquest motiu que es fa necessari l'aplicació d'algun tipus de metodologia que permeti l'anàlisi i avaluació general de la qualitat ambiental d'una conca. En aquesta Tesi es presenten, a nivell general, dos tipus de metodologies per tal de dur a terme aquesta tasca d'avaluació: metodologies empíriques i metodologies deterministes.

Els models de tipus determinista [10] són mètodes numèrics basats en lleis generals o principis teòrics que permeten l'estimació de la qualitat de diferents compartiments ambientals d'una conca i la simulació del transport de substàncies a través de la xarxa hidrogràfica. Pressuposen que un sistema o procés opera de tal manera que un conjunt determinat d'esdeveniments porta a un resultat únic. D'altra banda, els models de tipus empíric [11-12] consideren la incertesa estadística tant en les condicions inicials com en les prediccions obtingudes pel model, incorporant-hi una component aleatòria.

Històricament, les comunitats científiques que han utilitzat metodologies deterministes són totalment diferents de les que han utilitzat metodologies empíriques, i existeix un gran desconeixement mutu de les possibilitats que ofereixen unes i altres. No obstant, en aquesta Tesi s'intenta apostar per la integració d'ambdós tipus de metodologies, per a la qual ja existeixen propostes incipients.

### 1.3.1. Metodologies empíriques

Les dades ambientals es caracteritzen normalment per una elevada variabilitat. Els principals orígens d'aquesta variabilitat en l'àmbit d'una conca hidrogràfica poden ser influències de tipus geogènic, hidrològic i meteorològic, però també hi tenen un paper important les fonts de tipus

## Capítol 1

---

antropogènic. Les incerteses resultants del procés de presa de mostres, dels processos de conservació i pretractament d'aquestes mostres, de les mesures analítiques i fins i tot de la fase de calibratge, es troben contingudes en les dades finals. Per aquest motiu, no és sorprenent que els resultats analítics de qualsevol estudi realitzat sobre una conca hidrogràfica continguin aquesta variabilitat produïda per les incerteses, a més de les principals característiques que influencien el medi estudiat. Per tal de resoldre el problema de les incerteses, cal que es treballi amb un elevat nombre de mostres. Això implica un gran volum de dades, les quals portaran associada molta informació redundant i latent.

Amb l'objectiu de descobrir les pautes i fonts de variació existents en les sèries i taules de dades multivariants obtingudes dels programes de seguiment ambiental sobre una conca hidrogràfica, es proposa l'aplicació de metodologies de tipus empíric. L'assumpció bàsica d'aquestes metodologies és que cadascun dels paràmetres mesurats experimentalment en una mostra particular (per exemple, la concentració d'un contaminant en un tram de riu) es troba afectat per diferents contribucions de fonts independents (que poden ser tant puntuals com difoses) d'orígens diversos.

Les metodologies empíriques es basen en la descripció i anàlisi de la variància observada en les dades experimentals. Ofereixen eines eficients per a l'extracció d'informació útil a partir de grans conjunts de dades, i són capaces d'identificar i capturar aquells paràmetres clau que condicionen el comportament dels sistemes ambientals. Ja que molts sistemes ambientals involucren processos que no es coneixen completament i per als quals no existeixen models formals que els descriguin, es presenta la necessitat d'utilitzar metodologies complementàries a les metodologies de tipus determinista. Dins del grup de metodologies empíriques existeixen, entre altres, les que es recullen a la Taula 1.1. En aquesta Tesi no es pretén fer una classificació exhaustiva de les diferents metodologies empíriques que existeixen per al tractament de les bases de dades ambientals, sinó tan sols citar-ne alguns exemples.

---

### Metodologies empíriques de tractament de dades ambientals

---

Mètodes d'anàlisi de dades ( <i>data mining</i> )	Mètodes de l'estadística i àlgebra multivariant	Tècniques d'anàlisi de variància (ANOVA) i anàlisi discriminant <sup>1</sup> Anàlisi de factors ( <i>factor analysis</i> ) <sup>2</sup> Anàlisi de components principals ( <i>principal component analysis, PCA</i> ) <sup>3</sup> Reconeixement de pautes ( <i>pattern recognition</i> ) <sup>4</sup> Anàlisi d'agrupacions ( <i>cluster analysis</i> ) <sup>5</sup> <i>k</i> -veí més proper ( <i>k-nearest neighbor</i> ) <sup>6</sup> SIMCA <sup>7</sup> Anàlisi de sèries temporals i d'autocorrelació ( <i>time series analysis</i> ) <sup>8</sup> ; models ARIMA <sup>9</sup>  Mètodes de regressió multivariant en les seves diferents variants, entre els que actualment destaquen
---	---	--

	els mètodes de regressió per mínims quadrats parcials ( <i>partial least squares, PLS</i> ) <sup>10</sup>
	Xarxes neuronals ( <i>artificial neural networks</i> ) <sup>11</sup>
	Xarxes de decisió basades en mètodes probabilístics o baiesians ( <i>bayesian decision networks</i> ) <sup>12</sup>
	<i>Support vector machines</i> <sup>13</sup>
	Algoritmes genètics ( <i>genetic algorithms</i> ) <sup>14</sup>
	Lògica difosa ( <i>fuzzy logic</i> ) <sup>15</sup>

Mètodes geostatístics<sup>16</sup>; desenvolupats fonamentalment des de la hidrologia i les ciències de la terra

<sup>1</sup> McLahlan 1992. Wiley, Chichester

<sup>2</sup> Malinowski 1991. Wiley, New York

<sup>3</sup> Jolliffe 2002. Springer, Berlin

<sup>4</sup> Breerton 1992. Elsevier Science, New York

<sup>5</sup> Massart and Kaufman 1983. Wiley, New York

<sup>6</sup> Fix and Hodges 1951. Report Number 4, USAF, Texas

<sup>7</sup> Wold and Sjostrom 1977. ACS Symposium Series 52

<sup>8</sup> Chatfield 1989. Chapman and Hall, London

<sup>9</sup> Abraham and Ledolter 1983. John Wiley and Sons, NJ

<sup>10</sup> Geladi and Kowalski 1986. Anal Chim Acta 185:1-17

<sup>11</sup> Zupan and Gasteiger 1991. Anal Chim Acta 248: 1-30

<sup>12</sup> Jensen 1996. Springer-Verlag, New York

<sup>13</sup> Cristianini and Shawe-Taylor 2000, Cambridge University Press, Cambridge

<sup>14</sup> Goldberg 1989. Addison-Wesley, Reading, MA

<sup>15</sup> Bandemer and Näther 1992. Kluwer, Dordrecht

<sup>16</sup> Cressie 1991. Wiley, New York

**Taula 1.1.** Models empírics de tractament de dades ambientals

## Mètodes d'anàlisi de dades

Els mètodes d'anàlisi de dades que s'han utilitzat en aquesta Tesi corresponen exclusivament al grup dels mètodes de l'estadística i àlgebra multivariant. És per aquest motiu que la resta de mètodes, tot i que es troben llistats a la Taula 1.1, no seran descrits en aquest apartat.

L'anàlisi estadística multivariant s'aplica a dades que consisteixen en múltiples mesures d'un nombre d'individus, objectes o mostres. La mesura i l'anàlisi de la dependència entre les variables és fonamental en aquest tipus de metodologia [13], ja que es pretén l'agrupació i classificació dels objectes o, en alguns casos, la modelització de les relacions que existeixen entre aquests. Dins d'aquest conjunt de mètodes hi trobem diferent tipus de tècniques com són: les tècniques d'anàlisi de la variància i anàlisi discriminant, l'anàlisi de factors, l'anàlisi de components principals (PCA), o l'anàlisi d'agrupacions, entre d'altres.

En el cas de les **tècniques d'anàlisi de la variància (ANOVA)** i de l'anàlisi discriminant, es parla de mètodes d'anàlisi supervisada de dades. En aquests, la identitat de les categories o factors és coneguda d'entrada i la finalitat és determinar si la mitjana d'una variable difereix significativament entre els diferents grups. L'extensió multivariant de l'anàlisi de la variància dona lloc a la denominada MANOVA, on l'interès rau en determinar si el conjunt de mitjanes de



les múltiples variables són diferents entre si. D'altra banda, quan es tracta de determinar si les mitjanes d'un conjunt de dades difereixen degut a diversos factors o combinació de factors (més de dos), s'utilitza la generalització d'ANOVA, anomenada N-way ANOVA.

**L'anàlisi de factors** és un mètode de l'estadística multivariant que s'utilitza per a la identificació d'un nombre reduït de factors a partir d'un conjunt de dades amb variables correlacionades. El conjunt original de dades (amb  $n$  dimensions) es projecta sobre un sistema de coordenades de dimensions més reduïdes (1, 2 o 3 dimensions). Com que cada factor pot afectar diverses variables, aquests s'anomenen factors comuns. S'assumeix que cada variable depèn de la combinació lineal dels factors comuns, i els coeficients s'anomenen *loadings*. Les variables mesurades inclouen també una component deguda a la variabilitat aleatòria, anomenada variància específica, perquè és específica a cada variable. Es tracta d'una eina útil per a extreure informació latent de les dades, com ara aquelles relacions no directament observables que existeixen entre elles.

**L'anàlisi de components principals (PCA)** és probablement el mètode més àmpliament utilitzat dins de la família de mètodes de l'anàlisi de factors. S'aplica PCA per tal de realitzar una anàlisi exploratòria de les dades, condensant el màxim d'informació sobre el sistema utilitzant el menor nombre possible de variables. El PCA aconsegueix aquesta simplificació de la informació substituint grups de variables per una nova variable. El mètode genera un nou conjunt de variables anomenades components principals, on cada component és combinació lineal de les variables originals. Tots els components principals són ortogonals i no contenen informació redundant. El primer component principal és un eix a l'espai. Quan cadascuna de les mesures originals es projecta sobre aquest eix, els valors resultants defineixen una nova variable. I la variància d'aquesta variable és la màxima d'entre totes les possibles alternatives d'aquest nou eix. El segon component principal és, de la mateixa manera, un altre eix a l'espai, el qual serà perpendicular al primer. La projecció de les mesures originals sobre aquest segon eix, defineix una nova variable, la variància de la qual també serà màxima. El conjunt total de components principals serà tant gran com el conjunt original de variables. No obstant, només els primers components principals, els quals descriuen la major part de la variància de les dades originals, seran seleccionats per explicar les dades de manera simplificada. A la secció 2.1.2.4 d'aquesta Tesi es presenta una descripció detallada del mètode PCA.

Normalment, la terminologia i els objectius per descriure l'anàlisi de factors i l'anàlisi de components principals se superposen, de manera que es fa difícil fer una distinció entre els dos mètodes. Es tracta, en tots dos casos, de tècniques de reducció de la dimensionalitat de les dades, en el sentit que poden ser utilitzades per substituir un conjunt gran de variables mesurades per un conjunt més reduït de noves variables. No obstant, els dos mètodes es

diferencien en els seus objectius i en els seus models subjacents. En general, PCA pot ser utilitzat quan es vol simplificar o realitzar una aproximació de les dades utilitzant un menor nombre de dimensions (per exemple, per tal de visualitzar-les). En canvi, l'anàlisi de factors serà utilitzat quan es faci necessària l'obtenció d'un model que expliqui les correlacions que existeixen entre les dades.

La finalitat de l'**anàlisi d'agrupacions** és la identificació de grups d'objectes relativament similars pel que fa a les seves propietats mesurades. Es tracta de crear grups d'objectes o *clusters*, de manera que els perfils dels objectes dins del mateix grup siguin molt semblants mentre que els perfils dels objectes de diferents grups siguin diferents.

Els **mètodes de reconeixement de pautes** es descriuen com a processos de formulació de regles de classificació. Partint d'un conjunt de classes donat, cadascuna de les quals conté un nombre d'observacions representat per multitud de variables, es desenvolupen pautes i regles per tal de poder classificar les noves observacions com a semblants o bé diferents als membres de les classes existents. Es parla de dos tipus generals de classificació: supervisada i no supervisada. En el reconeixement de pautes no supervisat, el nombre i tipus de classes és desconegut i caldrà que sigui determinat. En el reconeixement de pautes supervisat, per contra, l'objectiu és assignar un objecte determinat a una classe prèviament definida.

El **k-veí més proper** és un mètode de classificació no supervisada en el qual es calcula, en primer lloc, la distància euclidiana d'un objecte a la resta d'objectes. Posteriorment, la distància mínima és seleccionada per tal d'assignar l'objecte a la classe corresponent. Normalment, el nombre d'objectes veïns  $k$  es fixa, i com a resultat de l'anàlisi s'obtenen separacions força flexibles entre grups d'objectes. No obstant, la classificació obtinguda dependrà del nombre d'objectes fixats per cada classe, i en el cas de classes superposades, aquest s'assignarà a la classe amb el major nombre d'objectes.

**SIMCA** és un mètode de classificació supervisada on les classes de les mostres del conjunt de calibratge són conegudes d'antuvi. Aquestes classes o grups d'objectes són aleshores descrits mitjançant un model matemàtic independent de la resta de classes. El mètode SIMCA construeix models de components principals diferents per a cadascuna de les classes a analitzar. Cadascun dels models pot identificar un nombre diferent de components principals, de manera que poden representar línies, plans, cubs o hipercubs en més de tres dimensions. Aquests models poden ser posteriorment utilitzats per eliminar objectes aberrants, per estimar la capacitat de modelatge d'una variable particular o bé per a classificar nous objectes.

## Capítol 1

---

Una **sèrie temporal** està formada per un conjunt d'observacions mesurades de manera seqüencial en el temps. Qualsevol procés discret o bé continu que sigui mesurat regularment pot ser representat per una sèrie temporal. Les correlacions que existeixen dins de les sèries temporals són descrites en forma d'autocorrelació o autocovariància i de les seves funcions derivades. Dues sèries temporals diferents es caracteritzen per la seva correlació creuada (*cross-correlation*). La informació que s'extreu normalment d'aquest tipus de models és: i) la desviació, ii) el soroll, iii) la periodicitat dels processos, i finalment, iv) una predicció dels valors futurs en funció de la tendència presentada per la sèrie.

En els **mètodes de regressió multivariant** s'utilitzen conjunts de dades que poden ser disposades en dos blocs diferents (X i Y). Tant si aquesta relació és lineal com si no ho és, serà necessari validar adequadament els resultats obtinguts. Entre els mètodes de regressió multivariant, destaca el mètode de mínims quadrats parcials (*partial least squares*, PLS). Aquest mètode fa la regressió a partir de components que adopten un compromís entre l'explicació de la variància de X i la capacitat predictiva de Y (maximitzen la covariància entre els dos blocs de dades).

## Mètodes geostatístics

Els mètodes geostatístics consideren les variables de tipus ambiental com a funcions en les que se superposa una component aleatòria amb una component determinista de variació lenta. Com que visitar cadascuna de les localitzacions existents en una àrea d'estudi per tal de mesurar una alçada o bé la concentració d'una variable determinada és normalment difícil o bé massa car, una possible alternativa, és la mesura directa del fenomen estudiat en algunes localitzacions disperses de manera estratègica i la posterior predicció dels valors corresponents a la resta de localitzacions. Una de les principals aplicacions de la geostatística és en el camp de la cartografia automàtica, tant pel que fa a la predicció en determinades localitzacions dels valors de la variable estudiada (ja siguin puntuals o mitjans), com de la incertesa associada a aquests. S'han descrit altres aplicacions com són la simulació de processos ambientals, l'objectiu de la qual és analitzar la variabilitat espacial, o el disseny òptim de xarxes de control i presa de mostra.

Segons Matheron (1963) [14], pare de la geostatística en la seva forma actual, la geostatística és l'aplicació de la Teoria de les Variables Regionalitzades a l'estimació de processos o fenòmens geològics a l'espai. Aquesta teoria assumeix que la variació espacial d'una variable representada a localitzacions específiques, és estadísticament homogènia a través d'una superfície determinada. Així doncs, des del punt de vista matemàtic, una variable regionalitzada és simplement una funció  $z(s)$  que adopta un valor per a cada punt  $s$  a l'espai.

### 1.3.2. Metodologies deterministes

Les metodologies deterministes permeten la simulació de diferents processos basant-se en lleis físiques, químiques i biològiques. Un model determinista de la qualitat ambiental consta típicament d'una sèrie d'equacions diferencials que incorporen paràmetres o relacions matemàtiques teòriques, així com diversos coeficients mesurats, estimats o suposats, per tal de descriure el comportament de les pressions (agents que provoquen un impacte sobre el medi) i les relacions entre aquestes pressions i les característiques del medi. Un dels usos més típics de les metodologies deterministes és la quantificació de com un sistema modelat particular, per exemple una massa d'aigua, respon a pressions concretes, com ara a la presència de determinats compostos tòxics.

Les equacions utilitzades per construir el model, sovint simplifiquen la realitat de manera significativa però necessària. Aquest tipus de metodologies s'ha convertit en una eina útil per als organismes competents en matèria de gestió de l'aigua, ja que permeten la realització de simulacions mitjançant diferents escenaris. Partint d'un escenari de referència (generalment l'estat real de la conca), aquest es compara amb un escenari simulat (en unes condicions diferents, en molts casos un cop una mesura correctora ja ha estat implementada). Dita mesura pot consistir tant en una regulació introduïda per al control de la contaminació, com en el canvi d'una determinada pràctica que tingui lloc sobre l'àmbit de la conca estudiada. La comparació entre escenaris pot ser de gran ajuda en els processos de decisió. L'anàlisi d'escenaris també pot anar associada a un model de tipus predictiu que permeti la descripció del transport de substàncies a través de la xarxa hidrogràfica.

Els models deterministes es poden dividir en 1) estacionaris i 2) dinàmics, depenent de si incorporen o no una dimensió temporal [10]. Els models estacionaris no inclouen una component temporal variable sinó que descriuen les condicions temporals mitjanes que es donen en el període d'estudi. D'altra banda, els models dinàmics presenten una dimensió de temps amb velocitats específiques pels diferents processos, donant lloc a sèries temporals. Els models estacionaris es poden dividir, al seu torn, segons quin sigui l'inici espacial pels seus càlculs, en: 1.1) models d'immissió, en els que el transport o concentració estimats d'una substància a la sortida de la conca es troben relacionats amb les característiques que es donen corrent amunt; i 1.2) els models d'emissió, els quals resumeixen els coeficients d'escolament i/o dades d'emissió empíriques per diferents classes de contribucions que es troben presents en una conca, per tal d'establir les condicions a la sortida. Els models estacionaris poden o no estar basats en els resultats d'un model hidrològic. D'altra banda, els models dinàmics de transport es basen generalment en models hidrològics, ja que el flux d'aigua constitueix el medi de transport, de manera que molta de la variabilitat en el transport de substàncies és un efecte

## Capítol 1

---

de la variabilitat hidrològica. Temporalment, els càlculs acostumen a ser repetits a intervals de temps horaris o bé diaris. Existeixen models basats en fenòmens puntuals que són capaços de simular, per exemple, el transport de substàncies durant un episodi aïllat de tempesta, mentre que n'existeixen d'altres de tipus continu que poden realitzar simulacions per determinades estacions, anys o fins i tot dècades. Espacialment, les simulacions poden anar des d'una fins a tres dimensions.

Degut al balanç que existeix entre la complexitat d'un model i el seu poder predictiu, existeixen dues categories de models dinàmics: 2.1) mecanicistes, o sigui, basats en lleis físiques, químiques i biològiques, i 2.2) conceptuals, en els quals només aquells processos més importants són descrits pel model o en els quals diversos processos s'agrupen sota una sola equació. El principal avantatge d'aquests darrers, és que requereixen un volum menor d'informació d'entrada així com un esforç computacional inferior, la qual cosa els fa adequats per a estudis a gran escala o per conques per les quals la informació de base disponible és limitada.

L'Agència de Protecció Ambiental dels Estats Units (*Environmental Protection Agency, EPA* [www.epa.gov](http://www.epa.gov)) treballa, conjuntament amb les autoritats estatals i locals, per tal de portar a terme projectes de seguiment i control de la qualitat del medi fluvial. El seu estudi es centra cada vegada més en la interacció que existeix entre els diferents compartiments ambientals dins d'una mateixa conca hidrogràfica. Aquest enfocament, permet una cada vegada major identificació dels canvis en la qualitat de l'aigua, el descobriment de nous problemes emergents i l'avaluació de diferents mesures de control de la contaminació, l'eficiència de la seva aplicació i la seva resposta davant de situacions d'emergència. Existeixen diversos models deterministes proposats per la EPA. Alguns d'aquests models, i altres, han estat utilitzats dins de l'àmbit de la Unió Europea.

Recentment, s'ha portat a terme el projecte *Benchmark Models for the Water Framework Directive (BWW)* per tal d'oferir consell als professionals del sector de l'aigua en la selecció dels models més adients per tractar una problemàtica ambiental determinada. A la literatura hi apareixen una gran quantitat de models deterministes amb noms diferents però que sovint utilitzen equacions bàsiques semblants, tot i que la seva interfase així com les rutines d'entrada i sortida, poden diferir considerablement. Existeixen molts models que no poden ser classificats específicament dins d'una categoria concreta. Sovint, alguns d'ells poden acoblar-se per a la seva implementació conjunta, de vegades sota un mateix nom i de vegades sota un nom diferent, la qual cosa pot donar lloc a confusió.

La major part d'inconvenients relacionats amb els models de tipus determinista recauen en la gran quantitat de dades d'entrada que es necessiten per al seu funcionament. Aquesta entrada de dades inicial requereix, per norma general, la inversió d'una quantitat molt important de

temps. A més, cal que l'usuari tingui un coneixement exhaustiu de l'àrea d'estudi i també molta experiència en l'aplicació del model per tal que aquest sigui utilitzat de manera eficient. Moltes vegades el model requereix l'entrada de valors de diversos paràmetres com són constants, coeficients o índexs. L'elecció dels valors correctes per aquests paràmetres requereix un coneixement i una experiència en el modelat de la qualitat de l'aigua que de vegades l'usuari no té i a més, desafortunadament, moltes vegades no es troben recomanacions per aquests paràmetres en els manuals del model. Quan un model ha estat dissenyat per ser aplicat en unes condicions determinades, el fet d'aplicar-lo en condicions diferents requerirà una modificació d'aquest, adaptant-lo a les noves característiques de l'àrea d'estudi. Finalment, el fet que les descripcions dels processos en un model determinista hagin de ser simplificades, és la causa d'incerteses en les prediccions que posteriorment donarà com a resultat el model.

Pel que fa a la modelització determinista de la qualitat de les aigües superficials dins de l'àmbit europeu, l'informe elaborat per Arheimer i Olsson (2003) [10] recull alguns dels models que s'han anat utilitzant, classificats segons el medi físic al qual està dirigida la seva aplicació. A continuació es presenta un recull no exhaustiu de diversos models deterministes que han estat aplicats amb la finalitat de modelitzar la dinàmica de diferents conques hidrogràfiques o dels compartiments ambientals que les constitueixen.

### 1. Models de conca hidrogràfica

<b>Model</b>	<b>Nom complet</b>	<b>Origen</b>	<b>Propòsit/substàncies modelitzades</b>
BASINS <sup>1</sup>	<i>Better Assessment Science Integrating point and Nonpoint Sources</i>	<i>European Protection Agency (EPA), 1996</i>	Sistema de control ambiental de les fonts de contaminació puntual i difosa a l'aigua
WCS <sup>2</sup>	<i>Watershed Characterisation System</i>	EPA, 2001	Variables físiques, variables de qualitat de l'aigua i càrregues de diferents fonts (nutrients, productes fitosanitaris, sediments, etc)
LSPC <sup>3</sup>	<i>Loading Simulation Program C++</i>	EPA, 2001	Cabal, transport de sediments i qualitat de l'aigua
AGNPS <sup>4</sup>	<i>AGricultural Non-Point Source pollution model</i>	<i>United States Department of Agriculture (USDA), 1987</i>	Nutrients i productes fitosanitaris
HBV-N <sup>5</sup>		<i>Swedish Meteorological and Hydrological Institute</i>	Control de l'eutrofització i transport de nitrogen

		(SMHI), 1994	
INCA <sup>6</sup>	<i>Integrated Nitrogen in Catchments</i>	Universitat de Reading, UK, 1998	Control de l'eutrofització i transport de nitrogen
MIKE <sup>7</sup> BASIN <sup>8</sup>		<i>Danish Hydraulic Institute</i> (DHI), anys 90	Control de l'eutrofització, transport de nitrogen i contaminants
SHETRAN <sup>9</sup>		Universitat de Newcastle, 1996	Control de contaminants i transport de nitrogen i sediments
SWAT <sup>10</sup>	<i>Soil and Water Assessment Tool</i>	USDA, 1993	Control de l'eutrofització, productes fitosanitaris, sediments i nutrients
GIBSI <sup>11</sup>	<i>Gestion Intégrée des Bassins versants à l'aide d'un Système Informatisé</i>	<i>Institut National de la Recherche Scientifique</i> (INRS), 1998	Nutrients i productes fitosanitaris, contaminació urbana (EDAR) <sup>a</sup>

<sup>1</sup>Whittemore 1998. *Water Environ Technol* 10:57-61

<sup>2</sup>USEPA 2001. *Watershed Characterization System data*, UST Tech, Region IV, Georgia

<sup>3</sup>Shen et al. 2005. *J Environ Sci Health A Tox Hazard Subst Environ Eng* 40:1791-1807

<sup>4</sup>Young et al. 1989. *J Soil Water Conserv* 44(2):168-173

<sup>5</sup>Arheimer and Wittgren 1994. *Ambio* 23(6):378-386

<sup>6</sup>Whitehead et al. 1998. *Sci Total Environ* 210/211:547-583

<sup>7</sup>Xevi et al. 1997. *Water Resour Mahage* 11:219-242

<sup>8</sup>Jha and Das Gupta 2003. *Water Int* 28:27-35

<sup>9</sup>Lunn et al. 1996. *J Hydrol* 174:285-304

<sup>10</sup>Arnold et al. 1998. *J Am Water Res Assoc* 34:73-89

<sup>11</sup>Villeneuve et al. 1998. *Rapport du Projet GIBSI, Tome I R-462, INRS-Eau*

<sup>a</sup>estació depuradora d'aigües residuals

**Taula 1.2.** Models deterministes de qualitat a escala de conca hidrogràfica

### **BASINS (Better Assessment Science Integrating Point and Nonpoint Sources)**

Model que té com a objectius (i) caracteritzar la qualitat de les aigües, (ii) identificar les fonts de contaminació existents en el medi i (iii) avaluar i localitzar aquelles masses d'aigua que no compleixen els estàndards de qualitat definits, per a una posterior discussió de les alternatives de gestió a aplicar. Combina un sistema d'informació geogràfica (SIG), informació diversa sobre les conques hidrogràfiques i eines d'avaluació i modelització ambiental ([www.epa.gov/athens/wwqtsc/html/basins.html](http://www.epa.gov/athens/wwqtsc/html/basins.html)).

### **WCS (Watershed Characterization System)**

Eina que permet realitzar una caracterització de les conques hidrogràfiques. Efectua una avaluació de la qualitat de l'aigua i de les fonts de contaminació potencials per les quals pot ser afectada. Com a resultat de l'aplicació del model, es generen 3 tipus d'informes: un informe físic de la conca (amb la seva delimitació, usos del sòl, clima, cabal, fonts puntuals, recursos, etc.), un informe de la qualitat de l'aigua, i un informe de les càrregues de fonts puntuals,



nutrients, productes fitosanitaris, sediments, etc. ([www.epa.gov/athens/wwqtsc/html/WCS-toolbox.pdf](http://www.epa.gov/athens/wwqtsc/html/WCS-toolbox.pdf))

### **LSPC (Loading Simulation Program C<sup>++</sup>)**

Eina que permet la simulació de la hidrologia i del transport de contaminants en una conca hidrogràfica. És capaç de representar càrregues de contaminants provinents de fonts tant puntuals com difoses. Es tracta d'un dels pocs models a nivell de conca que simula processos que afecten al sòl i al flux d'aigua simultàniament ([www.epa.gov/athens/wwqtsc/html.lspc.html](http://www.epa.gov/athens/wwqtsc/html.lspc.html))

### **AGNPS (AGricultural Non-Point Source pollution model)**

Model desenvolupat per avaluar la qualitat de l'aigua quan aquesta es troba afectada per l'erosió del sòl a causa de les pràctiques agrícoles. El model consta de 3 components principals: (i) component hidrològic, que proporciona una predicció del volum d'escolament i del cabal màxim, (ii) component d'erosió del sòl, que inclou tant l'erosió com la sedimentació, i (iii) component de contaminació per nutrients, que analitza les concentracions de nitrogen, fòsfor i la demanda química d'oxigen, tant en el material erosionat com en el propi sediment. Model recomanat principalment per a la descripció d'episodis puntuals ([www.wsi.nrcs.usda.gov/products/w2q/h&h/tools\\_models/agnps/index.html](http://www.wsi.nrcs.usda.gov/products/w2q/h&h/tools_models/agnps/index.html))

### **HBV-N**

Model que simula la pèrdua de nitrogen i el transport d'aquest cap als sistemes d'aigües subterrànies, rius i llacs. Es tracta d'un model conceptual que descriu de manera simplificada la hidrologia de la conca, sense incloure una descripció física exacta dels diferents processos. Utilitza rutines separades per la simulació del nitrogen inorgànic i orgànic ([www.smhi.se/sgn0106/if/hydrologi/hbv\\_np.htm](http://www.smhi.se/sgn0106/if/hydrologi/hbv_np.htm))

### **INCA (Integrated Nitrogen in CAatchments)**

Model que considera les diferents fonts de nitrogen que existeixen a la conca hidrogràfica estudiada, i que processa aquesta informació, després de la interacció d'aquestes fonts amb els cultius, la vegetació semi-natural, els microbis i el sòl, per tal de predir la quantitat de nitrogen que anirà a parar al riu o corrent d'aigua que constitueixi el seu destí final. Basat en balanços de massa i reaccions cinètiques, el model simula els principals mecanismes que operen en el cicle del nitrogen, incloent la mineralització, la immobilització, la nitrificació i la desnitrificació. Si el nivell de nitrogen que arriba finalment al flux d'aigua és massa elevat (superior als 50 mg/L de nitrats establerts per la Unió Europea), diferents mesures poden ser adoptades per tal de reduir-lo ([www.rdg.ac.uk/INCA](http://www.rdg.ac.uk/INCA))



## Capítol 1

---

### **MIKE BASIN**

Model utilitzat per identificar masses d'aigua en risc, les fonts de contaminació més importants que les afecten, i el càlcul de les reduccions necessàries per tal d'acomplir els objectius de qualitat definits a la normativa vigent per aquestes masses d'aigua. El model ha estat utilitzat en diversos estudis i aplicat a una gran varietat de condicions climàtiques i humanes. ([www.dhigroup.com/Software/WaterResources/MIKEBASIN.aspx](http://www.dhigroup.com/Software/WaterResources/MIKEBASIN.aspx))

### **SHETRAN**

Eina útil per a l'estudi de l'impacte ambiental de l'erosió del sòl, la contaminació i els efectes dels canvis d'usos del sòl i climàtics. També utilitzat per a l'estudi dels recursos d'aigua superficial i subterrània i la seva gestió, així com per la modelització del transport de nitrogen i sediments ([www.ceg.ncl.ac.uk/shetran](http://www.ceg.ncl.ac.uk/shetran))

### **SWAT (Soil and Water Assessment Tool)**

Model capaç de predir l'impacte de diverses pràctiques de gestió de l'aigua, sediments, nutrients i pesticides en conques grans i complexes, amb una gran varietat d'usos del sòl, textures de sòl i diferents condicions de gestió al llarg d'extensos períodes de temps. Ideal per a la predicció de canvis en la gestió dels usos del sòl (clima i canvis en la vegetació, pràctiques agrícoles, gestió d'embassaments, transferència d'aigua, extracció d'aigües subterrànies, etc.) ([www.brc.tamus.edu/swat](http://www.brc.tamus.edu/swat))

### **GIBSI (Gestion Intégrée des Bassins versants à l'aide d'un Système Informatisé)**

Sistema dissenyat com a eina d'avaluació de l'impacte que els diferents usos del sòl i dels recursos hídrics tenen sobre la qualitat de l'aigua d'una conca hidrogràfica. Inclou diferents models físics de simulació (hidrològic, erosió del sòl, transport de nutrients i productes fitosanitaris, i qualitat de l'aigua), així com diferents models de gestió (fonts de contaminació puntual, fonts de contaminació difosa i gestió d'embassaments) ([www.inrs-ete.quebec.ca/activites/modeles/gibsi](http://www.inrs-ete.quebec.ca/activites/modeles/gibsi)). Per una descripció més detallada del sistema GIBSI, veure secció 2.2.3. En aquesta Tesi, el sistema és aplicat per a la modelització de la hidrologia i el transport de productes fitosanitaris a la conca hidrogràfica del riu Llobregat.

2. Models per aigües subterrànies

Model	Nom complet	Origen	Propòsit/substàncies modelitzades
ASM <sup>1</sup>	<i>Aquifer Simulation Model</i>	<i>Swiss Federal Institute of Technology (ETH), 1986</i>	Dispersió de la contaminació
MODFLOW <sup>2</sup>		<i>US Geological Survey (USGS), 1988</i>	Flux d'aigua subterrània i transport de soluts

<sup>1</sup> Kinzelbach 1986. Elsevier, Amsterdam

<sup>2</sup> Ashley 1994. J Inst Water Environ Manage 8(6):635-645

Taula 1.3. Models deterministes de qualitat per aigües subterrànies

**ASM (Aquatic Simulation Model)**

Modelització del transport de contaminants i del flux d'aigua subterrània en dues dimensions (horitzontal i vertical) ([www.baug.ethz.ch](http://www.baug.ethz.ch))

**MODFLOW**

Model utilitzat per a la simulació del flux d'aigua subterrània i dels contaminants dins de l'aquífer (aquífer confinat, no confinat o bé una combinació d'ambdues opcions) ([www.modflow.com](http://www.modflow.com))

3. Models per aigües superficials

Model	Nom complet	Origen	Propòsit/substàncies modelitzades
WASP <sup>1</sup>	<i>Water quality Analysis Simulation Program</i>	EPA, anys 70	Paràmetres d'eutrofització, substàncies químiques orgàniques, mercuri
MIKE11 <sup>2</sup>		DHI, 1999	Qualitat de l'aigua i transport de sediments
PC-QUASAR <sup>3</sup>		<i>Center for Ecology and Hydrology, UK, 1997</i>	Cabal i qualitat de l'aigua
QUAL2E QUAL2K <sup>4</sup>	<i>River and Stream Water Quality Model</i>	EPA, 1987	Cabal i qualitat de l'aigua

<sup>1</sup> Nikolaidis et al. 2006. J Mar Sys 60:51-62

<sup>2</sup> Hanley et al. 1998. J Environ Manage 52(3):211-225

<sup>3</sup> Lewis et al 1997. Sci Total Environ 194/195:399-418

<sup>4</sup> Azzellino et al. 2006. Sci Total Environ 371:214-222

Taula 1.4. Models deterministes de qualitat per aigües superficials

## Capítol 1

---

### **WASP (Water quality Analysis Simulation Program)**

Programa per a la modelització dels diferents compartiments ambientals que conformen els ecosistemes aquàtics, incloent la columna d'aigua i el bènctic subjacent. D'una banda, el model simula els paràmetres convencionals que causen l'eutrofització, mentre que de l'altra, simula també la dinàmica dels contaminants tòxics (substàncies químiques orgàniques i mercuri). El principal avantatge és la flexibilitat de la seva aplicació a gairebé qualsevol tipus de massa d'aigua, així com a molts dels problemes més convencionals pels quals pot estar afectada la qualitat d'aquesta aigua ([www.epa.gov/athens/wwqtsc/html/wasp.html](http://www.epa.gov/athens/wwqtsc/html/wasp.html))

### **MIKE11**

Es tracta d'un paquet de software dirigit principalment al camp de l'enginyeria, per a la simulació del cabal, la qualitat de l'aigua i el transport de sediments en estuaris, rius, sistemes d'irrigació, canals i altres masses d'aigua. Conté un mòdul hidrodinàmic que s'utilitza per al càlcul del cabal en rius i estuaris; un mòdul d'advecció-dispersió per simular el transport dels materials dissolts i, finalment, un mòdul de qualitat de l'aigua que, juntament amb el mòdul d'advecció-dispersió, simula els processos de reacció que existeixen en els sistemes multicomponent, incloent la degradació de la matèria orgànica, la fotosíntesi i respiració de les plantes, la nitrificació i l'intercanvi d'oxigen amb l'atmosfera. Addicionalment, el software consta d'un mòdul dedicat a l'estudi del transport de sediments i de les condicions morfològiques dels rius ([www.dhigroup.com/Software/WaterResources/MIKE11.aspx](http://www.dhigroup.com/Software/WaterResources/MIKE11.aspx))

### **PC-QUASAR**

Model que descriu els canvis en la qualitat de l'aigua i el cabal a la xarxa fluvial al llarg d'un període de temps determinat. Proporciona un mètode de comparació entre l'estat actual del riu i el que existiria després d'un canvi, planificat o no, a la xarxa fluvial. Quan s'utilitza amb finalitats de planificació, el model és capaç de proporcionar distribucions del cabal i dades de la qualitat de l'aigua en punts determinats (punts clau), permetent d'aquesta manera, fixar els nivells màxims d'efluent per tal de mantenir una bona qualitat ecològica del riu. Poden ser modelats els següents paràmetres de la massa d'aigua: cabal, amoníac, pH, nitrats, temperatura, E.coli, DBO, oxigen dissolt i múltiples contaminants ([www.ceh.ac.uk/products/software/CEHSoftware-PC-QUASAR\\_000.htm](http://www.ceh.ac.uk/products/software/CEHSoftware-PC-QUASAR_000.htm))

### **QUAL2E/QUAL2K (River and Stream Water Quality Model)**

QUAL2K és la versió moderna del conegut model QUAL2E. Es tracta d'un model continu aplicat a contaminants convencionals en els cursos fluvials. Pot ser utilitzat per avaluar l'impacte provocat per abocaments residuals puntuals sobre la qualitat de l'aigua així com per a la identificació de les característiques quantitatives i qualitatives dels abocaments de tipus

difós. El model ha estat àmpliament utilitzat per a la simulació de la concentració d'oxigen dissolt en els rius. A més, també pot ser aplicat per descriure les principals reaccions que existeixen en els cicles de nutrients, la generació d'algues, la demanda de carboni, i el reaireig atmosfèric, així com els efectes d'aquestes reaccions sobre el balanç d'oxigen dissolt ([www.epa.gov/athens/wwqtsc/html/qual2k.html](http://www.epa.gov/athens/wwqtsc/html/qual2k.html))

#### 4. Models per a l'escolament urbà

Model	Nom complet	Origen	Propòsit/substàncies modelitzades
MOUSE <sup>1</sup>	<i>Modelling Of Urban SEwers</i>	DHI, anys 80	Qualitat de l'aigua i transport de sediments per sistemes de drenatge urbà, desaignes urbans i sanitaris
SWMM <sup>2</sup>	<i>Storm Water Management Model</i>	EPA, anys 70	Tots els aspectes del cicle hidrològic urbà, incloent les precipitacions, l'escolament superficial i subsuperficial, recorregut de l'aigua a través de la xarxa de drenatge, emmagatzemament i tractament

<sup>1</sup> Jensen and Linde-Jensen 1992. Water Sci Technol 25(9):101-113

<sup>2</sup> Peterson and Wicks 2006. J Hydrol 329:294-305

Taula 1.5. Models deterministes d'escolament urbà

#### **MOUSE (Modelling Of Urban SEwers)**

Sistema per a la modelització global de l'escolament superficial en àrees urbanes, el cabal en els canals oberts, el cabal dins de les canonades, la qualitat de l'aigua i el transport de sediments en els sistemes de drenatge urbà, els sistemes receptors de l'aigua de pluja i els receptors dels efluents municipals ([www.dhi.es/Software/Urbano/MOUSE.aspx](http://www.dhi.es/Software/Urbano/MOUSE.aspx))

#### **SWMM (Storm Water Management Model)**

Model global per a l'anàlisi de problemes relacionats amb la quantitat i la qualitat de l'escolament superficial. És capaç de simular tots els aspectes dels cicles urbans hidrològics i de qualitat de l'aigua, com són les pluges, el desgel, l'escolament superficial i subsuperficial, el flux d'aigua a través de la xarxa de drenatge, l'emmagatzemament i el tractament. Permet avaluar el problema de l'escolament superficial i buscar possibles solucions per a la seva reducció. D'altra banda, conté també un bloc estadístic que pot ser utilitzat per a l'anàlisi de freqüències d'hidrogrames i pol-lutogrames de llargues sèries temporals, i per a la identificació d'episodis hidrològics individuals que puguin ser d'especial interès. S'ha observat un bon funcionament del model quan aquest s'aplica a àrees urbanitzades sense drenatge, tot i que aquest també ha estat àmpliament aplicat sobre altres zones amb característiques diferents (<http://cce.oregonstate.edu/water/swmm>)

### 1.3.3. Els SIG com a eina d'anàlisi ambiental

En els darrers anys, l'activitat en el camp dels sistemes d'informació geogràfica [15-17] (SIG, o en anglès GIS, per *Geographical Information Systems*), ha evolucionat de la pura recerca universitària a una gran popularització, fins al punt que SIGs senzills ja han aparegut sota la forma de guies urbanes de diferents ciutats i municipis o d'instruments d'ajuda a la circulació automobilística.

Segons la definició de Burrough i McDonnell (1998) [15], un SIG és un conjunt d'eines per a la captació, emmagatzematge, recuperació, transformació i visualització de dades espacials del món real per a unes finalitats determinades. La representació del món real es fa en base a les coordenades geogràfiques, els atributs i la topologia dels elements representats.

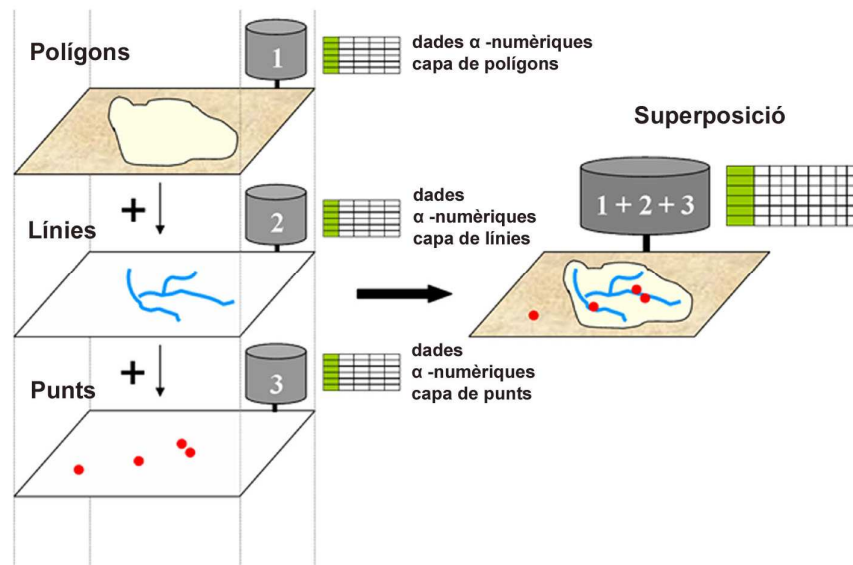
Un SIG està compost per tres elements bàsics. El primer és el *hardware* sobre el qual descansa el sistema i que determinarà les capacitats de processament de què es disposa. El segon és el *software* amb què es manegen les dades, que determinarà les capacitats funcionals amb què l'usuari es pot adreçar a aquestes dades. I finalment, hi ha les dades en si, tant si són de tipus purament espacial (per exemple, els límits d'una conca hidrogràfica) com si són dades alfanumèriques de naturalesa geogràfica (per exemple, el cabal mesurat a cada tram d'una xarxa fluvial).

Més enllà de la definició relativa a l'estructura i funcionalitat d'un SIG, la característica més distintiva d'aquest és la seva capacitat d'interrelacionar o integrar diferents conjunts d'informació sobre fenòmens o localitzacions en virtut de la seva localització i relacions en l'espai. És habitual que un SIG contingui diferents capes d'informació relatives a un determinat paràmetre (altitud, usos del sòl, concentració de contaminants, tipus de textura del sòl, direcció del flux d'aigua, etc.), i que es consulti aquesta informació ja sigui de manera individualitzada, ja sigui per encreuament o superposició entre capes (Figura 1.4). Una possible consulta podria ser "quines textures del sòl emmagatzemen generalment una concentració més elevada de contaminants?".

Els principals tipus de consultes que es poden realitzar en un SIG corresponen a les anomenades "consulta per localització" i "consulta per atributs". A mode d'exemple, una consulta per localització podria ser "a quina alçada es troba una coordenada concreta?". En canvi, com a exemples d'una consulta per atributs es podria indicar "mostrar totes les zones on l'ús del sòl sigui industrial". No obstant, el tipus de qüestions plantejables en un SIG permet altres possibilitats, com ara la consulta de relacions espacials de proximitat, "a quina distància dels sòls de tipus industrial han estat detectades les concentracions més elevades d'un determinat contaminant?", la consulta de relacions espacials de veïnatge, "es troben en contacte les estacions depuradores d'aigües residuals i les concentracions més elevades de

carboni orgànic dissolt al riu?”, o la consulta de relacions espacials de contenció, “les mostres més contaminades es troben dins de l'àmbit d'una determinada comarca?”.

També es poden demanar les característiques espacials (àrea, perímetre, forma, etc.) d'un fenomen, o bé fer consultes sobre quina és la ruta més curta sobre una xarxa de corredors biològics. A més, si es disposa d'informació relativa a dos moments històrics diferents, es poden abordar qüestions sobre l'evolució que han patit diversos atributs en el temps: “en quina proporció ha disminuït la superfície d'irrigació durant els darrers 10 anys?”.



**Figura 1.4.** Superposició de capes de diferent geometria en un SIG

Fonamentalment, existeixen dues maneres de representar la informació digital en un SIG: els sistemes vectorials i els sistemes ràster o matricials (Figura 1.5). Un sistema vectorial permet representar les entitats espacials mitjançant punts, línies o polígons a base d'emmagatzemar les coordenades dels punts o vèrtexs. Aquests objectes gràfics poden portar associada informació diversa. Per exemple, a la Figura 1.4, els vectors tipus polígon representen el contorn de les diferents conques hidrogràfiques, els vectors tipus línia, el recorregut de la xarxa fluvial, i els vectors tipus punt la localització de les diferents mostres analitzades. Cadascun d'aquests objectes gràfics està associat amb una base de dades alfanumèrica que conté informació addicional (Figura 1.6). La utilitat d'aquest format es basa en la possibilitat de realitzar consultes de la base de dades, però amb l'interès afegit d'obtenir els resultats representats a l'espai.

El sistema matricial o ràster representa la realitat mitjançant cel·les o píxels (unitats elementals d'informació del territori). Aquestes cel·les contenen valors numèrics però, a diferència d'una matriu numèrica en el sentit matemàtic, un ràster està georeferenciat (té coordenades que tenen equivalència dins un determinat sistema de projecció).

## Capítol 1

Pel fet de ser matrius numèriques, els ràsters també permeten que es puguin fer operacions matemàtiques i lògiques entre diverses capes d'informació. Des de sempre, hi ha hagut una gran preocupació sobre quin model de dades és el millor per representar l'espai geogràfic. Tanmateix, actualment se sol acceptar la necessitat que els programes donin suport a les dues aproximacions, ja que depenent del tipus de dades o aplicacions, serà més convenient adoptar-ne una o altra.

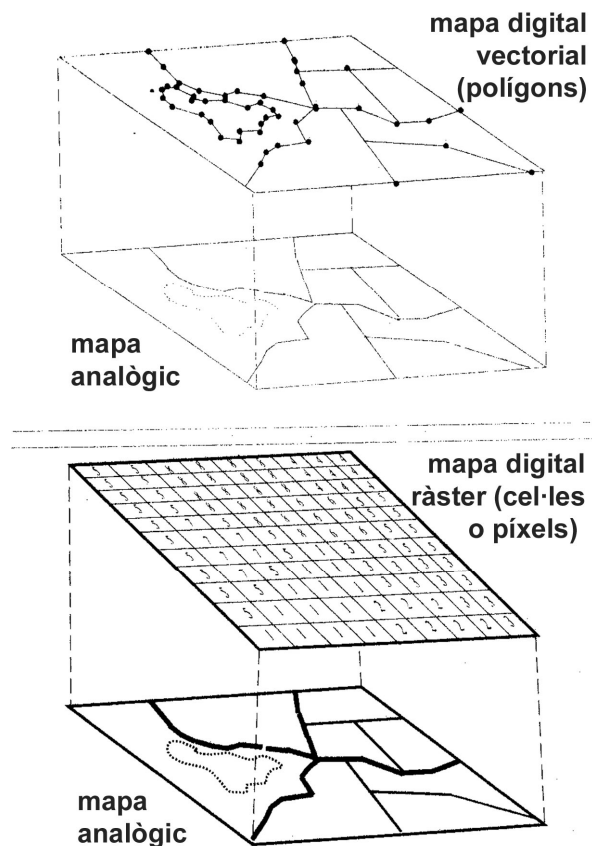


Figura 1.5. Representació ràster i vectorial de la realitat (Comas i Ruiz, 1993)

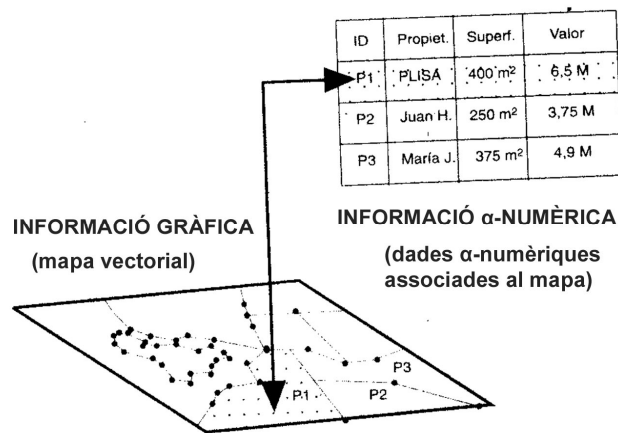


Figura 1.6. Unió de dades α-numèriques i cartogràfiques (Comas i Ruiz, 1993)

Tot i que la majoria de SIGs són bidimensionals, molts d'ells permeten el tractament del relleu a través d'aproximacions espacials. La més utilitzada és el model digital d'elevacions (MDE). Un MDE sol ser un fitxer ràster que recull, per al centre de cada cel·la, la seva altitud. Donada la naturalesa de la distribució de les dades altimètriques, l'altitud en qualsevol altre punt que no correspongui al centre d'una cel·la es pot obtenir utilitzant tècniques espacials d'interpolació. Els MDE són de gran utilitat ja que, a més de facilitar la informació altimètrica, permeten obtenir informació derivada, així com superposar-hi altres capes d'informació per tal de generar representacions tridimensionals. Els sistemes estrictament vectorials solen adoptar un model de representació que simplifica el terreny en zones triangulars irregulars de pendent i exposició similars, on la densitat de triangles és més alta en les zones de relleu complex. Aquest model s'anomena TIN (*Triangulated Irregular Network*).

La utilització dels SIGs no suposa cap disjunció respecte de l'ús dels mapes convencionals (en paper), sinó més aviat un enriquiment. Algunes de les limitacions del mapa clàssic com a sistema d'informació, com són la reducció del volum d'informació per imperatius de representabilitat i legibilitat (amb enorme dependència de l'escala escollida), la dificultat d'extreure i combinar informació entre mapes de diferent temàtica, de diferent moment temporal i de diferent full, i la dificultat d'actualització de la informació continguda en el mapa, es veuen superades gràcies als avenços de la cartografia digital. El suport informàtic dels SIGs permet una major facilitat de manteniment i recuperació de dades, l'anàlisi de canvis temporals, així com una llarga sèrie de funcionalitats que faciliten tant l'anàlisi científica i tècnica com la representació cartogràfica tradicional. A més, cal destacar que la presa de decisions utilitzant els SIGs facilita la feina i sol fer-la molt menys propensa a errors.

El punt fort dels sistemes SIG es que són capaços de tractar el que en principi semblen tipus molt diferents de dades, integrant informació que pertany a diferents àmbits mitjançant una característica comuna: la component geogràfica. La seva utilització implica la necessitat d'abstraccions o assumpcions per tal de representar la realitat geogràfica, ja que la informació ha de ser susceptible de ser processada per un ordinador.

Tot i les potencialitats dels SIGs, existeixen també alguns inconvenients que caldrà tenir en compte de cara a la seva utilització. I es que pot arribar a ser una perillosa forma de treballar allunyada de la realitat, en casos en què no es recorri al treball de camp perquè es té una inconscient fe cega en les dades que subministra l'ordinador. Així, el professional del medi natural ha de veure la informació que proporciona un SIG com una eina complementària al seu treball, potencialment tant subjecta a errors com qualsevol altre producte humà [16].



### 1.4. Justificació de la tesi

En els darrers anys, han anat proliferant un important nombre de treballs científics dedicats a la descripció de les problemàtiques que afecten la qualitat ambiental de les conques hidrogràfiques. Els entorns fluvials han patit una important degradació en tots els nivells, anant des de la contaminació química d'aigües superficials i subterrànies, sediments i sòls, a la modificació de les condicions hidromorfològiques, detracció de cabals, etc., amb la consegüent afectació que aquestes alteracions tenen sobre l'ecosistema de la conca.

Aquesta Tesi ha estat elaborada amb una intenció principalment metodològica. No té la finalitat d'abordar l'estudi específic d'una problemàtica concreta, sinó de proporcionar un conjunt d'eines adequades per a l'estudi de diverses problemàtiques de caràcter ambiental, sovint enteses des de diferents disciplines científiques. Aquestes eines són:

- ❖ Els mètodes estadístics i quimiomètrics per a l'extracció d'informació rellevant a partir de l'anàlisi de taules de dades ambientals. Aquests mètodes permeten la identificació de les principals fonts de contaminació que afecten una determinada regió, tant pel que fa a la composició química d'aquestes fonts com a la seva distribució espacial i temporal. Els mètodes estadístics s'utilitzen en totes les ciències experimentals. En particular, han adquirit força importància en l'àmbit de la química analítica.
- ❖ Els mètodes deterministes per a la modelització de les dinàmiques existents dins d'una determinada conca hidrogràfica. Aquests han estat utilitzats per a modelar la hidrologia de la conca, així com el transport de contaminants químics provinents de fonts de contaminació a través de la xarxa fluvial. Els mètodes deterministes s'utilitzen en diversos camps de les ciències ambientals i de les ciències de la terra, de la geologia, la biologia, l'ecologia, la química, la física, etc.
- ❖ Els sistemes d'informació geogràfica (SIGs) es proposen com a eines de representació de dades i de resultats de la seva anàlisi, ja que a més de la informació pròpiament continguda a les taules de dades, permeten la incorporació de la seva component espacial mitjançant l'addició de coordenades geogràfiques. Dins d'aquest apartat de representació espacial de dades, se situen també els mètodes d'interpolació geostatística, que permeten l'estimació de determinats atributs en localitzacions on aquests no han estat mesurats. L'aproximació geostatística s'acostuma a associar als camps de la geologia i la mineria.

- ❖ El càlcul d'índexs de qualitat per a la caracterització d'una determinada massa d'aigua, es proposa com un mètode per tal de resumir, de manera concisa, la informació més rellevant provinent dels programes científics de seguiment ambiental i transmetre-la de manera intel·ligible tant als organismes gestors com al públic en general. En aquest sentit, s'ha posat especial atenció en l'aprofitament d'aquelles dades procedents de les xarxes automàtiques de presa de mostra.

### 1.5. Objectius

Dins de cadascun dels tipus de metodologies aplicades en aquesta Tesi, s'han desenvolupat objectius concrets:

#### ❖ MÈTODES ESTADÍSTICS I QUIMIOMÈTRICS

1- Aplicació de mètodes estadístics i quimiomètrics a l'anàlisi de taules de dades ambientals per a la identificació de les principals fonts de contaminació així com de les seves tendències espacials i temporals:

- i) Estudi de la contaminació històrica i actual a la conca del riu Ebre a partir de dades de la *Confederación Hidrográfica del Ebro* (CHE, dades històriques) i del Projecte Europeu AquaTerra (dades recents)
- ii) Estudi de les fonts de contaminació presents a la conca del riu Llobregat
- iii) Estudi de l'impacte del cultiu de l'arròs sobre la qualitat de l'aigua dels canals d'irrigació i drenatge del delta del riu Ebre

2- Validació del mètode de resolució multivariant de corbes (MCR-ALS) per a la investigació de fonts de contaminació en aigües superficials

3- Estudi de l'efecte que diversos factors relacionats amb el procediment d'anàlisi de dades (com el mètode de pretractament de les dades, el soroll, l'ambigüitat rotacional, l'estructura de les dades o les restriccions aplicades durant el procés d'anàlisi) tenen sobre la recuperació de les fonts de contaminació

#### ❖ MÈTODES DETERMINÍSTICS

1- Calibratge del model GIBSI (hidrologia, erosió i transport de contaminants) per a la seva aplicació a la conca hidrogràfica del riu Llobregat

2- Modelització de la hidrologia de la conca del riu Llobregat

3- Simulació del transport i distribució de contaminants procedents de fonts agrícoles (aplicats sobre els sòls de l'àmbit de la conca hidrogràfica) a través de la xarxa fluvial del riu Llobregat

#### ❖ SIG I GEOSTADÍSTICA

1- Representació de dades espacials (àmbit de la conca d'estudi, estacions de presa de mostra, xarxa hidrogràfica, usos del sòl, etc.), mitjançant la cartografia digital

- 2- Representació contínua de la distribució de la contaminació sobre l'àrea d'estudi a partir de l'interpolació geostatística utilitzant el mètode *kriging*
- 3- Representació gràfica de resultats obtinguts a partir de l'anàlisi multivariant de dades

### ❖ INDEXS DE QUALITAT DE L'AIGUA

- 1- Estudi dels principals tipus d'índexs existents a la literatura utilitzats per a l'avaluació de la qualitat de les masses d'aigua. Establiment dels requisits per tal d'obtenir una bona categorització de la qualitat. Comparació de diferents índexs.
- 2- Validació de l'índex de Canadà (CCME WQI) per a l'anàlisi de dades fisicoquímiques procedents de xarxes automàtiques de presa de mostra. Aplicació de l'índex sobre diferents escenaris ambientals
- 3- Adaptació del procediment *combined monitoring-based and modelling-based priority setting scheme* (COMMPS) a les conques hidrogràfiques de Catalunya. Càlcul de l'índex prioritari per les diferents substàncies que presenten un risc per als ecosistemes aquàtics i per l'home via els ecosistemes aquàtics. Càlcul d'un índex prioritari de contaminació de localitzacions



## 2. METODOLOGIES APLICADES A L'ANÀLISI DE DADES AMBIENTALS

Les taules de dades ambientals s'obtenen generalment a partir d'estudis o programes de seguiment ambiental (monitoratge). Poden contenir paràmetres de diferent naturalesa (físics, químics, biològics, toxicològics, etc.), i les seves dades es caracteritzen per una gran variabilitat que prové de fonts molt diverses i, normalment, desconegudes. La variabilitat existent en les dades ambientals es pot assignar a tres tipus diferents de fonts o orígens [18]:

- (a) fonts naturals sense fortes influències de contaminació, les quals varien geogràficament i temporalment d'una manera estocàstica com a conseqüència dels fenòmens naturals;
- (b) fonts antropogèniques, que són causades per activitats humanes i interfereixen de manera significativa amb la variabilitat de tipus natural;
- (c) fonts de variabilitat produïdes per errors de tipus experimental i de mesura, causades a la vegada per les diferents etapes que conformen el procés analític: presa de mostra, pretractament i mesura d'aquesta i, finalment, pel procés d'anàlisi de les dades. La contribució de l'error experimental a la variabilitat de les dades augmentarà quan les concentracions determinades siguin baixes i també quan la magnitud de la mostra disminueixi.

### 2.1. Mètodes estadístics i quimiomètrics

Normalment, l'obtenció de mesures selectives de determinats compostos o de les seves fonts resulta difícil. Apart del soroll de fons, les dades experimentals es veuen afectades per interferències químiques i/o físiques procedents de la mostra o bé del procés de mesura utilitzat. En general, l'aplicació de tècniques d'anàlisi univariant de dades acostuma a no ser suficient i cada vegada existeix una major utilització de les tècniques d'anàlisi multivariant. Les dades de tipus univariant s'obtenen a partir de senyals analítiques mesurades com a funció d'una única variable de control (per exemple, la longitud d'ona, el temps, etc.). Òbviament, la quantitat d'informació que pot ser extreta a partir de senyals univariants és limitada. D'altra banda, els objectius de l'anàlisi ambiental acostumen a ser de tipus multivariant. Diversos patrons de contaminació, procedents de fonts diferents i sovint desconegudes, poden existir en el medi. Gran quantitat d'informació ambiental pot ser extreta d'un conjunt de dades multivariant si s'apliquen les eines d'anàlisi multivariant adients.

## 2.1.1. Mètodes univariants d'anàlisi de dades

Un primer anàlisi estadístic de les dades es pot dur a terme a partir de la utilització de:

- (a) Les eines de l'**estadística descriptiva** univariant que proporcionen diversos fulls de càlcul. En el cas del software EXCEL (Microsoft, Redmond, WA, USA), existeix l'opció *Anàlisi de dades* → *Estadística descriptiva*, mitjançant la qual es poden obtenir els següents paràmetres estadístics, agrupant les dades per variables i/o mostres: mínim, màxim, mitjana, mediana (valor central de la distribució), moda (valor que es repeteix més), variància (mesura de la dispersió de valors), desviació estàndard (arrel quadrada de la variància), suma i recompte del nombre total de mostres.
- (b) Els **diagrames de caixa**, una representació gràfica de les característiques principals d'una distribució de dades. En aquest tipus de diagrama, cada variable està representada per una caixa, la qual conté el 50% de les dades centrals de la distribució (veure Figura 2.1a). Els extrems inferior i superior de la caixa indiquen el 1er i 3er quartils, o el que és el mateix, el percentils 25% i 75%. Aquests indiquen les posicions a partir de les quals hi ha el 25% i 75% dels valors mesurats en la distribució després de ser ordenats de menor a major. El rang entre aquests dos quartils és el que es coneix com a rang interquartil. La línia a l'interior de la caixa indica el valor de la mediana, o sigui, el valor central de la distribució (valors ordenats de menor a major). El fet que la mediana no estigui localitzada en el centre de la caixa, indica que es tracta d'una distribució esbiaixada. Els límits superior i inferior de les línies de fora de la caixa (*whiskers*), serveixen per a identificar els valors atípics. Generalment se situen a 1.5 vegades la distància interquartil del 1er i 3er quartils (valor que se suma al 3er quartil i que es resta al 1er quartil). Els valors fora d'aquest rang i representats per signes positius (+) es consideren valors atípics o *outlayers*.

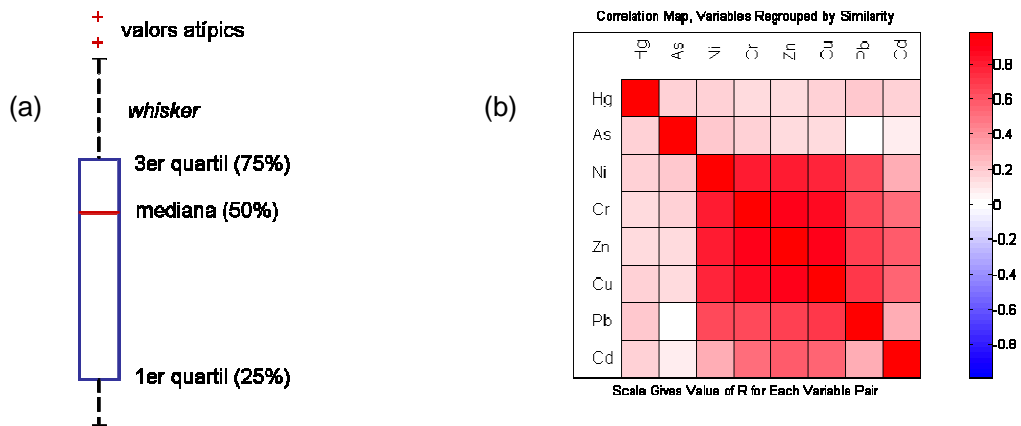


Figura 2.1. Esquema d'un diagrama de caixa (a) i mapa de correlació de variables (b)

- (c) **L'anàlisi de correlació** per parelles de variables a partir del càlcul dels coeficients de correlació entre les variables obtinguts de les dades experimentals. Aquesta anàlisi permet una primera investigació sobre la possible presència i distribució de fonts de contaminació comunes. En l'exemple de la Figura 2.1b, les variables Ni, Cr, Zn i Cu són les que presenten una correlació més elevada entre elles (una major similitud). En canvi, es veu com l'Hg, As i Cd varien de forma més independent i no es troben correlacionades amb la resta de variables analitzades.

### 2.1.2. Mètodes multivariants d'anàlisi de dades

Un dels objectius de la recerca ambiental actual és obtenir informació el més objectiva possible sobre els múltiples i complexos processos que tenen lloc en el medi ambient. El desenvolupament de mètodes quimiomètrics que ha tingut lloc durant les darreres dècades, ha aportat una eina poderosa per a la resolució d'aquest tipus de problemes.

La quimiometria ambiental és una disciplina relativament recent. El seu propòsit general consisteix en l'aplicació de mètodes estadístics i matemàtics amb l'objectiu de trobar una manera òptima de resoldre problemes químics ambientals i extreure la màxima informació possible de les dades experimentals. Així, l'aplicació de mètodes quimiomètrics en el camp de la química ambiental és necessari i útil degut a les següents raons:

- (a) Els processos naturals i antropogènics que ocorren en el medi ambient es troben principalment caracteritzats per canvis multivariable en les concentracions de múltiples compostos i/o contaminants químics en els diferents compartiments ambientals i en les seves propietats i paràmetres fisicoquímics i biològics,
- (b) Les emissions de tipus antropogènic s'acostumen a caracteritzar pel seu caràcter multicomponent o multielement. A més, aquestes substàncies poden reaccionar de manera sinèrgica o competitiva, és a dir, que les interaccions entre contaminants i entre aquests i els components naturals del medi seran complexes i també hauran de ser considerades,
- (c) Les dades experimentals obtingudes en estudis ambientals contenen informació sobre els diferents processos ambientals i també sobre la variància causada per l'error analític i altres incerteses experimentals. Un dels propòsits d'aplicació de les metodologies de tipus quimiomètric en l'àmbit ambiental serà la minimització o bé, en cas que sigui possible, l'eliminació d'aquestes incerteses experimentals i de l'error analític associat a les mesures ambientals,



(d) Els darrers anys han estat caracteritzats per un gran desenvolupament dels instruments de química analítica i del processament de la informació obtinguda a partir d'ells. Aquests canvis han permès un seguiment més detallat i precís del medi ambient i de les reaccions de substàncies químiques en diversos medis. La necessitat de gestionar l'enorme quantitat de dades obtingudes a partir dels estudis de seguiment i monitoratge ambiental és una altra de les principals raons de l'aplicació dels mètodes quimiomètrics. El principal objectiu es centra en l'extracció i simplificació de la informació rellevant pel que fa a l'estat de la contaminació, els canvis espacials i temporals d'aquesta i la identificació de les principals fonts. La modelització del comportament dels contaminants dins i entre els diferents compartiments ambientals, el seu impacte sobre l'ésser humà i l'avaluació dels riscos ambientals i toxicològics, també requereix l'aplicació de mètodes quimiomètrics.

En resum, la quimiometria aplicada als estudis mediambientals (quimiometria ambiental) és una eina que permet realitzar una investigació més rigorosa i objectiva dels diferents processos que ocorren en el medi ambient.

### 2.1.2.1. Estructura de les taules de dades ambientals

Les dades obtingudes a partir dels estudis de seguiment i monitoratge ambiental es poden classificar, segons la seva complexitat, en dades ordenades en una direcció (*one-way data*), en dues direccions (*two-way data*), en tres direccions (*three-way data*) i en múltiples direccions (*multiway data*) [19-20]. Tal com es mostra en la Figura 2.2, els valors numèrics escalars (per exemple, la mesura d'una única variable en una mostra) correspondrien a dades no direccionades o ordenades en zero direccions (*zero-way*), mentre que les dades vectorials (per exemple, mesures de diverses variables en una mostra) es troben ordenades en una direcció. Quan es mesuren diverses variables en diferents mostres, s'obtenen dades que es poden ordenar en dues direccions, és a dir, en una taula o matriu de dades. Finalment, la recopilació de varies matrius de dades o bé de dades ordenades en tres direccions, dóna lloc al que s'anomena un cub de dades.

Una matriu de dades constitueix l'estructura més comunament trobada en els estudis de seguiment ambiental. En aquestes taules o matrius de dades, les diferents mostres analitzades es col·loquen en les diferents files de la matriu, i les diferents variables mesurades (concentracions de compostos químics, paràmetres fisicoquímics, etc.) es col·loquen en les diferents columnes de la matriu. Les tècniques estadístiques necessàries per al processament

multivariant d'aquestes dades agrupades en forma de taula o matriu utilitzen les eines, formulacions i notacions de l'àlgebra lineal. Aquestes taules o matrius de dades ambientals seran representades per una lletra majúscula en negreta. A mode d'exemple,  $\mathbf{D}(i,j)$  seria la notació utilitzada per indicar una matriu de dades ambientals constituïda d'  $i$  files (per exemple, el nombre de mostres ambientals analitzades) i  $j$  columnes (per exemple, el nombre de paràmetres ambientals analitzats). Quan es passa de dades en una direcció a dades ordenades en tres direccions, la quantitat d'informació augmenta, però augmentarà també la complexitat de la seva anàlisi.

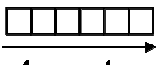
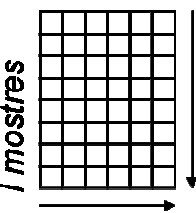
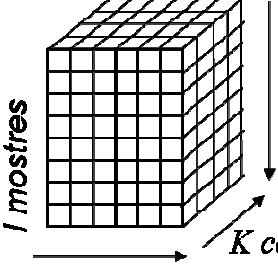
	<b>0 direccions</b>	<b>2 direccions</b>	<b>3 direccions</b>
<b>ESTRUCTURA</b>	<p><b>1 variable</b></p> <p>□</p> <p><b>1 mostra</b></p> <p><b>1 direcció</b></p> <p><b>J variables</b></p>  <p><b>1 mostra</b></p>	<p><b>J variables</b></p>  <p><b>i mostres</b></p>	<p><b>J variables</b></p>  <p><b>i mostres</b></p> <p><b>K condicions</b></p>
<b>MODEL</b>	<p><b>-Models lineals</b></p> <p><b>-Models no lineals</b></p>	<p><b>-Models bilineals</b></p> <p><b>-Models no bilineals</b></p>	<p><b>-Models trilineals</b></p> <p><b>-Models no trilineals</b></p>

Figura 2.2. Tipus d'estructures i models de dades ambientals

Depenent de l'estructura que presentin les dades a analitzar, es podran aplicar diferents tipus de models. Així, quan es tracta de dades en una direcció, els tipus de models utilitzats podran ser els models lineals (veure Equació 2.1) i no lineals (veure Equació 2.2) univariants. Per dades ordenades en dues direccions, es podran aplicar models bilineals (veure Equació 2.3) o models no bilineals. Finalment, per les dades ordenades en tres direccions, es podran aplicar models trilineals (si les dades presenten una estructura trilineal) (veure Equació 2.4) o bé models no trilineals.

$$y_i = b_0 + x_i \quad \text{Equació 2.1}$$

$$y_i = f_{\text{no lineal}}(x_i) \quad \text{Equació 2.2}$$

$$x_{ij} = \sum_{n=1}^N g_{in} f_{nj} + e_{ij} \quad \text{Equació 2.3}$$

$$x_{ijk} = \sum_{n=1}^N g_{in} f_{jn} z_{kn} + e_{ijk}$$

Equació 2.4

La descomposició de valors singulars (*Singular Value Decomposition*, SVD [21]) pot ajudar a esbrinar quina és l'estructura de les dades i el model més adequat per analitzar-les. Es dona més informació sobre els diferents tipus de models aplicats a les matrius de dades ambientals i sobre la SVD a l'apartat 2.1.2.5 i a l'article 1 de la present Tesi.

### 2.1.2.2. Tractament preliminar de les dades

Abans de poder aplicar l'anàlisi quimiomètrica, les taules de dades ambientals necessiten una preparació prèvia, que sovint acostuma a ser la part més laboriosa de tota l'anàlisi. Aquesta preparació comportarà una feina d'estructuració, homogeneïtzació i transformació de les dades, que dependrà, en cada cas, del format en què aquestes dades hagin estat inicialment obtingudes, del seu volum, etc.

Entre els problemes més comuns a resoldre, es troben:

(a) La presència de **valors buits** (*missing values*). Es tracta de valors que no es coneixen o que estan en blanc i on els compostos considerats no han estat o no han pogut ser analitzats. En el cas que el nombre de valors buits sigui petit i es trobin distribuïts d'una manera aleatòria dins de la taula de dades, aquests poden ser estimats numèricament tenint en compte les variacions observades a les variables (columnes de la matriu) i a les mostres (files de la matriu). L'estimació d'aquests valors es pot fer mitjançant mètodes quimiomètrics, com per exemple, l'anàlisi de components principals [22] o altres algorismes iteratius [23-25]. Un cop realitzada l'estimació dels valors buits, és aconsellable comprovar que les estimacions realitzades siguin plausibles i no discordants. S'ha de comprovar que els valors estimats per aquests procediments tinguin sentit des d'un punt de vista ambiental (per exemple, que no s'obtinguin valors negatius en el cas de variables que només poden ser positives) o bé que la magnitud no s'aparti excessivament d'allò que seria d'esperar tenint en compte la variació observada per a la resta de mesures. En el cas que el valor que s'ha de predir fós en realitat un valor atípic, és evident que la seva estimació a partir d'aquests procediments no seria possible a no ser que correspongués a una variació sistemàtica prèviament detectada per a la resta de dades. En el cas d'aquesta Tesi, l'estimació dels valors buits s'ha fet mitjançant les funcions d'imputació per PCA de la PLS Toolbox (Eigenvector Research Ltd., Manson, WA, USA), dins l'entorn de programació i visualització del software MATLAB (The Mathworks, MA, USA). Aquest tipus d'estimació es realitzarà substituint inicialment els valors buits de la matriu de

dades per zeros i fent la seva anàlisi per PCA. Posteriorment, els zeros són substituïts pels nous valors predits pel model PCA per un nombre de components determinat i, amb aquests nous valors predits, es torna a calcular un nou model PCA. Aquest procés es va repetint fins que els valors buits estimats per PCA convergeixen (s'acaben obtenint sempre els mateixos valors). La selecció del nombre de components principals es fa de la forma usual, intentant explicar una proporció considerable de la variància de les dades, sense incloure el soroll experimental. Quan es fa una sobreestimació del nombre de components, s'acostumen a predir valors no raonables, com ara valors negatius en el cas de concentracions de compostos químics.

(b) Valors **sota el límit de detecció** ( $<LD$ ). Aquests valors se substitueixen per la meitat del límit de detecció ( $LD/2$ ) considerat, ja que s'ha comprovat que aquesta és una bona estimació, millor que la simple substitució per zeros o pel mateix valor del límit de detecció, a més d'implacar poc esforç computacional [26]. Altres propostes com la substitució d'aquests valors pels obtinguts directament a partir de les mesures instrumentals es consideren encara millors (dades no censurades). Tot i això, no sempre es disposa dels valors dels senyals mesurats experimentalment a partir dels instruments emprats en les anàlisis [12], ja sigui perquè el senyal instrumental de la substància no ha estat detectat, o bé pel fet que tot i detectar el senyal, aquest ha estat censurat per l'analista en considerar-lo poc fiable estadísticament.

### 2.1.2.3. Pretractament de les dades

Per tal de facilitar l'extracció d'informació ambiental útil de les dades experimentals, sovint es fa necessària l'aplicació de tècniques de preprocessament o pretractament de dades. La selecció de les tècniques de pretractament més adients dependrà de diversos factors inherents a cada conjunt de dades, com per exemple, de la manera com aquestes han estat obtingudes, del tipus d'informació ambiental que es busca, o del coneixement previ que es tingui sobre la seva naturalesa. No existeix un únic mètode de pretractament de dades que pugui ser generalitzat a totes les matrius de dades ambientals. L'elecció del mètode més adequat dependrà principalment del tipus de fenomen o problema ambiental que es pretén posar de manifest.

El fet que generalment les dades de tipus ambiental no portin associades les incerteses en la seva determinació, en dificulta un pretractament òptim. Aquest pretractament implicaria la utilització de mètodes de càlcul de sospesament de dades que tinguessin en compte aquests errors o incerteses, com es fa en el mètode anomenat factorització positiva de la matriu (*Positive Matrix Factorization*, PMF [27]) o bé, el desenvolupat més recentment, mètode de mínims quadrats alternats sospesats (*Weighted Alternating Least Squares*, WALs [28]).

## Capítol 2

---

Els mètodes de pretractament de dades que més s'han aplicat al llarg d'aquesta Tesi són l'escalat i l'autoescalat. Aquests mètodes, juntament amb altres possibles tipus de pretractament, es descriuen a continuació:

### Centrat amb la mitjana de les dades

El centrat consisteix en restar a cada valor de les variables originals el valor de la seva mitjana. D'aquesta manera, s'aconsegueix que cada variable de la nova matriu de dades (matriu centrada) quedi amb una mitjana igual a zero.

$$z_{ij} = x_{ij} - \bar{x}_j \quad i = 1, \dots, I \quad j = 1, \dots, J \quad \text{Equació 2.5}$$

$$\bar{x}_j = \frac{\sum_{i=1}^I x_{ij}}{I} \quad \text{Equació 2.6}$$

on  $z_{ij}$  és el nou valor centrat,  $x_{ij}$  és el valor original de la variable, i  $\bar{x}_j$  és la mitjana dels  $x_i$  valors corresponents a la variable  $j$ . El centrat amb la mitjana de les dades realitza una translació de l'origen de coordenades de les dades des de zero fins al valor de la mitjana de les dades. Així, es fa més fàcil visualitzar les variacions respecte al valor mitjà (veure Figura 2.3b). Es descarta, per tant, aquella informació de les variables que no varia i que es pot considerar constant a la matriu (*offsets*) ja que interessa emfatitzar les variacions. No obstant, cal remarcar que el centrat farà més difícil distingir l'aportació (*apportionment*) o assignació quantitativa proporcional de les diferents fonts de contaminació sobre les mostres. El centrat de les dades és un procediment recomanat per a mètodes quimiomètrics com l'anàlisi de components principals (PCA), però es desaconsella per als mètodes de resolució multivariant [29], el PMF [27] o el PARAFAC (*Parallel Factor Analysis* [30]), quan s'apliquen restriccions de no negativitat.

### Escalat

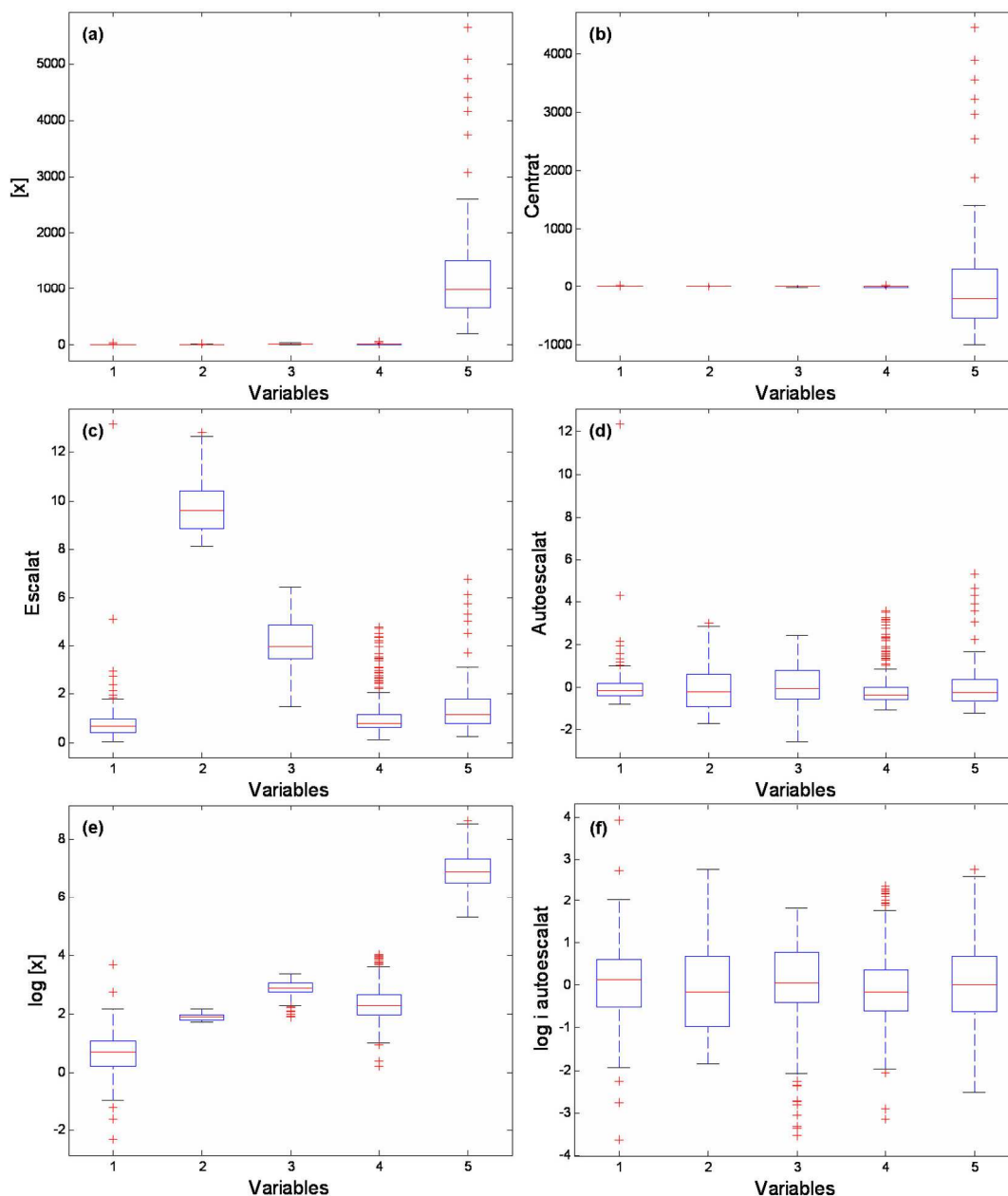
L'escalat consisteix en dividir cada valor de les variables per la seva desviació estàndard.

$$z_{ij} = \frac{x_{ij}}{s_j} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad \text{Equació 2.7}$$

on  $z_{ij}$  és el nou valor escalat,  $x_{ij}$  el valor original i  $s_j$  el valor de la desviació estàndard de la variable:

$$s_j = \sqrt{\frac{\sum_{i=1}^I (x_{ij} - \bar{x}_j)^2}{I-1}} \quad \text{Equació 2.8}$$

On  $\bar{x}_j$  és la mitjana dels valors de la variable  $j$ . L'escalat és una forma de pretractament de dades força utilitzada en el cas de conjunts de dades ambientals, sobretot quan aquestes han estat mesurades per variables, l'escala i les unitats de les quals varien. Les variables escalades que s'obtenen, presenten una distribució més homogènia, cosa que les fa més fàcilment comparables entre si (veure Figura 2.3c).



**Figura 2.3.** Diagrames de caixa d'un conjunt de dades ambientals (a) originals, (b) centrades, (c) escalades, (d) autoescalades, (e) amb transformació logarítmica i (f) amb transformació logarítmica i escalades

A més, la desviació estàndard de les variables un cop escalades és, per totes, igual a 1, de manera que s'iguali la seva importància relativa. No obstant, cal anar en compte en el cas de

## Capítol 2

---

tenir variables en les que tots els seus valors siguin molt baixos i constants en gairebé totes les mesures, ja que aleshores el seu escalat pot augmentar de manera artificial la seva importància relativa (perquè es divideixen per una desviació estàndard molt petita) quan en realitat no la tenen i, per tant, això pot portar a resultats i conclusions totalment errònies. En aquests casos, és millor eliminar directament aquestes variables o, en el cas que es consideri necessària la seva presència en l'anàlisi, donar-les-hi un pes petit (per exemple, dividint tots els valors de la variable per un valor gran, com ara 20, enlloc de dividir-los per la desviació estàndard de la variable).

L'escalat de les dades és un pretractament especialment útil quan les dades de variables diferents es troben expressades en unitats i magnituds diferents. En aquests casos, les variables amb valors més grans dominen de forma no desitjable la variància observada, i per tant, les variables amb els valors més baixos es troben emmascarades. En aquestes situacions és preferible escalar les dades per tal de disminuir l'efecte exagerat d'aquelles variables que presenten magnituds majors. En canvi, quan les variables analitzades tenen les mateixes unitats de mesura, no es veu tan clara la conveniència del seu escalat. Com ja s'ha esmentat, quan s'escala es dona la mateixa importància a totes els variables. Per exemple, existeixen casos en què certs compostos químics minoritaris tenen tanta o més importància que els majoritaris, de manera que amb concentracions més baixes, els seus efectes poden ser molt més greus, donant lloc a una contaminació més important. Totes aquestes consideracions caldrà tenir-les en compte a l'hora de decidir si s'aplica l'escalat o no de les dades, així com en decidir si aquest escalat s'aplica a totes o a només algunes de les variables.

En els casos en què es disposi de diverses matrius de dades on les variables analitzades coincideixen (per exemple, matrius d'anys diferents o bé corresponents a diferents compartiments ambientals), l'escalat pot ser aplicat sobre cadascuna de les matrius de dades individuals o conjuntament sobre la matriu formada a partir de totes les matrius de dades individuals (matriu augmentada, veure més endavant, secció 2.1.2.5) al mateix temps. L'elecció d'una opció o altra caldrà que es faci en funció de la finalitat de l'estudi. Per exemple, en aquells casos en què s'analitzin matrius obtingudes en diferents anys, caldria escalar-les conjuntament si el que interessa observar són les diferències temporals (entre anys diferents) dels components identificats. Si, en canvi, el que es pretén posar de manifest són les diferències que existeixen entre les localitzacions geogràfiques, atribuint una importància menor a les diferències entre anys, serà millor optar per l'escalat individual i per separat de cadascuna de les matrius de dades abans de la seva anàlisi conjunta. Tot això es discuteix amb més detall a l'article 4 d'aquesta Tesi.

Com ja s'ha esmentat anteriorment, les dades ambientals porten associades incerteses que no han estat determinades. Caldrà també anar en compte amb aquestes incerteses, ja que es poden veure amplificades de manera considerable a l'hora d'aplicar l'escalat.

### Autoescalat

L'autoescalat consisteix en l'aplicació simultània del centrat i de l'escalat:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad \text{Equació 2.9}$$

on  $z_{ij}$  és el nou valor autoescalat,  $x_{ij}$  el valor original,  $\bar{x}_j$  la mitjana dels valors de la variable (descrita a l'Equació 2.6), i  $s_j$  la desviació estàndard dels valors corresponents a la variable  $j$  (descrita a l'Equació 2.8). La distribució dels valors de les variables obtingudes en aplicar el tractament d'autoescalat és similar al cas de l'escalat, però al mateix temps les dades experimenten una translació del seu origen degut al centrat amb la mitjana de les dades (veure Figura 2.3d). L'autoescalat és el pretractament preferit en l'aplicació de mètodes com PCA, on l'interès fonamental recau en la descripció i investigació de la variància de les dades. No obstant, tal com ocorre en el cas del centrat, amb l'autoescalat apareixeran valors negatius a la nova matriu de dades autoescalada, la qual cosa impedirà l'aplicació dels procediments de resolució multivariant o d'altres quan s'apliquen restriccions de no negativitat (veure més endavant, secció 2.1.2.5).

### Transformació logarítmica

La transformació logarítmica de les dades actua, com l'escalat, augmentant el pes d'aquelles variables que presenten uns valors més baixos respecte a aquelles que presenten els valors més alts. És útil en aquells casos en què existeixen variables amb diferents magnituds, escales o unitats de mesura. Es tracta d'un procediment que sovint es troba recomanat en la literatura per conjunts esbiaixats de dades, com ara les dades d'estudis ambientals, on la major part dels valors de les variables són baixos i existeix un menor nombre de valors alts. Amb l'aplicació del logaritme, s'obté una distribució més simètrica de les dades (veure Figura 2.3e). No obstant, la transformació logarítmica també pot donar lloc a efectes indesitjats, ja que l'estructura interna de les dades pot quedar afectada i augmentar la seva possible no linealitat i el nombre de components necessari per explicar la seva variància. Es per això que es considera convenient la comparació del nombre de components obtinguts per PCA o SVD (descrits més avall) que es necessiten per explicar la mateixa quantitat de variància, abans i després de la seva transformació logarítmica. Posteriorment a la transformació logarítmica de les dades, és també possible aplicar addicionalment qualsevol dels altres pretractaments descrits anteriorment



(centrat, escalat o autoescalat (Figura 2.3f), obtenint diferents distribucions dels valors de les variables.

### 2.1.2.4. Anàlisi de components principals (PCA)

L'anàlisi de components principals (*Principal Component Analysis*, PCA) [22, 31] és un mètode d'estadística multivariant freqüentment utilitzat en l'anàlisi de taules de dades obtingudes en estudis de seguiment i monitoratge ambiental. Parteix de la hipòtesi que en el grup de dades original existeixen un conjunt reduït de factors o components dominants (fonts de variació) que influeixen de manera important sobre la variància observada en les dades, i que aquests factors o components no són mesurats directament (es tracta de factors no explícits o amagats, *hidden factors*), ja que no existeixen sensors específics per a ells o no són observables com a tals experimentalment.

Generalment, una part important de la informació continguda en les variables mesurades experimentalment és redundant (correlacionada amb les altres variables) o bé irrellevant. El PCA permet transformar aquest conjunt de variables en un nou conjunt de variables no correlacionades entre si, que és més senzill d'interpretar, fent més evident aquella informació subjacent que probablement és més rellevant. Aquestes noves variables, factors o components principals són doncs ortogonals. Això significa que el que explica un component, no ho explica l'altre. A més, en PCA, el primer component es troba en la direcció que expliqui el màxim de variància, i el mateix pel segon component per la variància residual, el tercer, i així successivament. La finalitat del PCA és trobar, a partir de la combinació lineal de les variables originals, un nou conjunt d'eixos ortogonals de coordenades sobre els que es poden projectar les dades originals i proporcionar una millor i més fàcil interpretació dels fenòmens o fonts que estan causant la variància observada en les dades ambientals mesurades (veure Figura 2.4).

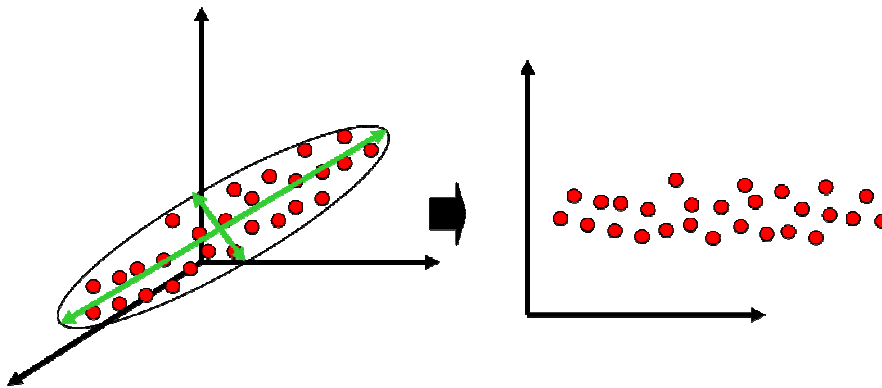


Figura 2.4. Nous eixos ortogonals creats a partir del conjunt de dades originals

El mètode PCA pretén resumir aquella informació rellevant que es troba continguda a la matriu de dades analitzada (separant-la de la part que correspon al soroll o error). Matemàticament, la matriu original de dades **D** es descompon, seguint un model bilineal, en el producte de dues matrius ortogonals, **X** i **Y<sup>T</sup>** (veure Figura 2.5)

$$\mathbf{D} = \mathbf{X} \mathbf{Y}^T + \mathbf{E} = \mathbf{D}^* + \mathbf{E} \quad \text{Equació 2.10}$$

On **D** és la matriu de dades experimentals, **X** la matriu d'*scores* (mapa de mostres), **Y<sup>T</sup>** la matriu de *loadings* (mapa de variables) i **E** la matriu d'error (Equació 2.10). El producte de les matrius d'*scores* i de *loadings* reproduirà la matriu original de dades per un nombre determinat *N* de components (nombre de columnes en **X** i de files en **Y<sup>T</sup>**). Degut a què els components en PCA s'obtenen en ordre decreixent de la variància explicada, aquella informació més rellevant del conjunt de dades es troba sempre concentrada en els primers components principals generats. És per aquest motiu que els components posteriors podran ser omesos sense una pèrdua significativa de la variància o d'informació rellevant, ja que només explicaran una petita part de la variància la qual estarà relacionada amb el soroll i l'error experimental.

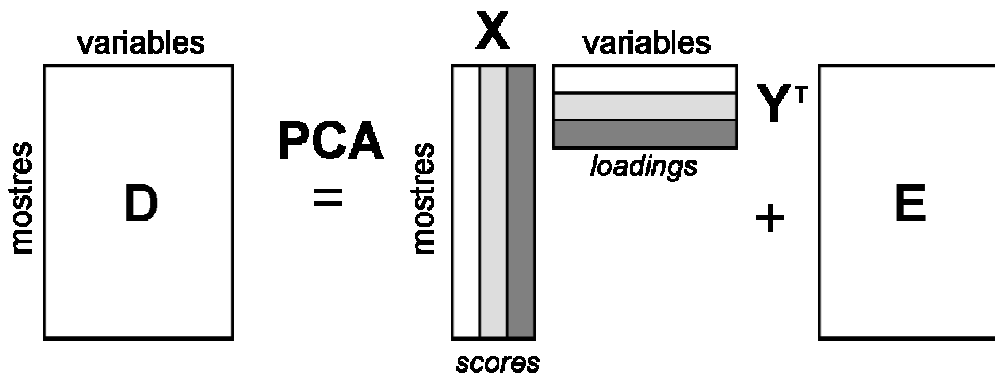
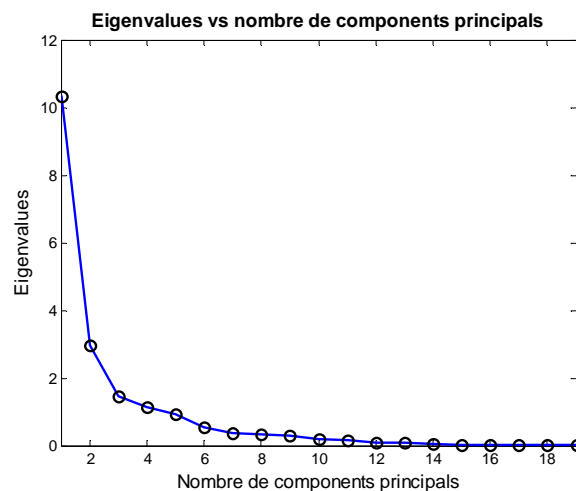


Figura 2.5. Descomposició per PCA (model bilineal) de la matriu de dades **D** per un nombre de components  $N = 3$

El mètode utilitzat freqüentment per realitzar la descomposició PCA de la matriu **D** és el de la descomposició en valors singulars (SVD [21]). Una primera manera de veure quants components cal incloure a l'anàlisi és mitjançant l'observació dels valors propis (*eigenvalues*) associats a aquests components, ja que aquests indiquen la quantitat de variància retinguda o explicada per cada component considerat. Els valors propis corresponen al quadrat dels valors singulars (*singular values*) trobats per SVD, i la quantitat de variància retinguda per a cada component, va disminuint successivament fins a un nivell petit i que ja varia poc per a nous components. Generalment, el nombre de components *N* s'escull fins que l'addició d'un nou component ja no proporciona informació addicional rellevant dins del context del problema estudiat o, dit d'una altra manera, quan ja només explica el soroll experimental. Aquells components que expliquen proporcions de variància petites no són investigats, i s'assumeix

que es troben principalment associats a petites contribucions de fons o al soroll i error experimental. El nombre escollit de components principals es relacionarà amb el nombre de fonts o pautes de variació (contaminació) independents que es troben presents en les dades analitzades. Per exemple, en el cas de la descomposició de la matriu de la Figura 2.6, el nombre de components aconsellat en funció de la representació dels valors propis seria de fins a tres (tot i que també caldria mirar el % de variància observada, que podria ser que amb dos components ja fos força elevada). A partir del tercer component, els valors propis disminueixen ja molt poc de mida i tendeixen a fer-se constants. En funció de la magnitud de la variància retinguda i de la contribució que les variables originals tenen a cada component, es pot deduir el significat ambiental dels diferents components trobats i pot determinar-se també el nivell de soroll aproximat de les dades experimentals. En aquest context, són extremadament útils les representacions de les matrius d'*scores* (matriu  $\mathbf{X}$ ) i de *loadings* (matriu  $\mathbf{Y}^T$ ) obtingudes directament de la descomposició PCA o SVD de la matriu de dades original  $\mathbf{D}$ .



**Figura 2.6.** Descomposició de valors propis d'una matriu de dades per a la determinació del nombre de components

Els valors dels *loadings* indiquen quina és la magnitud de la contribució de cadascuna de les variables originals a cada component o factor (nous eixos de representació). Quan es tracta de dades ambientals, els *loadings* poden indicar quina és la composició química de les fonts de contaminació identificades. Les variables amb valors grans de *loadings* sobre el mateix component, covarien. Si tenen el mateix signe, covarien positivament mentre que si tenen signes oposats, ho fan negativament (covarien inversament). La representació de les mostres en el nou espai definit pels nous components o factors són els *scores*. Aquestes representacions d'*scores* permeten agrupar les mostres segons la seva similitud i detectar la presència de mostres que presenten valors extrems. Des d'un punt de vista ambiental, els *scores* contenen informació sobre la distribució de les fonts de contaminació a les mostres (ja sigui informació geogràfica, temporal, per tipus de compartiment ambiental, etc.). La matriu

d'error conté aquella part de la variació de les dades que no es troba explicada pel model de PCA i que, per tant, es deu al soroll experimental. Generalment, la interpretació correcta de les diferents agrupacions de mostres obtingudes en els gràfics d'*scores* requereix que aquests siguin analitzats conjuntament amb els gràfics corresponents de *loadings*.

En determinats casos, per a una millor interpretació dels resultats obtinguts per PCA, és recomanable l'aplicació d'algun tipus de rotació dels components. Les tècniques d'anàlisi de factors utilitzen diferents tipus de rotacions. En el cas de les rotacions ortogonals, els nous eixos són també ortogonals entre ells, però els *loadings* de les variables que contribueixen a cada component es veuen modificats per tal de facilitar la interpretació de les dades. A més, la quantitat de variància explicada per cada component individual també variarà. La rotació Varimax [22, 32] intenta maximitzar la variància dels *loadings* de les variables dominants per rotació seqüencial dels *loadings* originals (obtinguts per PCA). El resultat són uns perfils de *loadings* simplificats i una redistribució de la variància, això sí, conservant-se la mateixa quantitat de variància total explicada pels factors o components identificats.

### 2.1.2.5. Resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS)

La resolució multivariant de corbes (*Multivariate Curve Resolution*, MCR) descriu una família de procediments quimiomètrics utilitzats amb la finalitat d'identificar i resoldre les contribucions existents en un conjunt de dades. Aquests procediments, sovint han estat aplicats per a la resolució de múltiples components químics en mescles investigades mitjançant tècniques d'anàlisi espectroscòpica (per exemple, per a la resolució d'espectres i perfils de concentració en mescles de compostos químics de composició desconeguda, sota diverses assumpcions) [33].

Darrerament, els procediments MCR han vist múltiples aplicacions a la literatura [34]. Alguns exemples de processos i sistemes analitzats mitjançant procediments MCR són l'anàlisi de mescles de compostos químics a partir de mesures espectrals o voltamperomètriques, l'estudi de processos i reaccions químiques (de cinètica i equilibri químic), les anàlisis cromatogràfiques complexes de mostres naturals amb detecció multicanal espectral (LD-DAD, LC-MS i GC-MS, etc.), l'anàlisi espectroscòpica de processos industrials i l'anàlisi d'imatges espectroscòpiques, entre d'altres. A més a més, aquests procediments també han estat aplicats per a la resolució de problemes de naturalesa ambiental. Aquest últim cas és el que es presenta en aquesta Tesi.

En primer lloc, per tal d'utilitzar els procediments de resolució multivariant de corbes, es requereix disposar, com a mínim, d'un conjunt de dades ordenades en dues direccions en una taula o matriu de dades **D**, per exemple un conjunt de *J* variables o concentracions de compostos químics (columnes de la matriu de dades) mesurades en *I* mostres (files de la matriu de dades) diferents. El mètode MCR es basa en la descomposició de la informació continguda en la matriu de dades **D**, seguint un model bilineal similar a l'emprat en el cas de PCA, format també per la suma de les contribucions de cadascun dels components (model additiu). Així, el model bilineal de resolució de corbes també s'expressarà com:

$$\mathbf{D} = \mathbf{X} \mathbf{Y}^T + \mathbf{E} \quad \text{Equació 2.11}$$

On  $\mathbf{Y}^T$  és la matriu de factors (anomenada de *loadings* en la notació de PCA), amb la contribució de cadascuna de les variables a les diferents fonts de contaminació (que expressarà la composició de les fonts de contaminació identificades), **X** és la matriu de factors (anomenada d'*scores* en la notació de PCA), amb la contribució de les fonts de contaminació identificades en cadascuna de les mostres analitzades (que expressarà la distribució de les fonts de contaminació sobre les mostres), i **E** conté l'error de les dades (part de la variància de les dades no explicada pel model). En aquesta Tesi, les matrius de factors **X** i  $\mathbf{Y}^T$  també s'anomenaran matrius d'*scores* i de *loadings*, conservant doncs la mateixa notació utilitzada en PCA.

Tot i que el mètode MCR utilitzi el mateix model bilineal additiu que el mètode PCA descrit anteriorment, la descomposició de la matriu de dades es fa de forma totalment diferent i donant lloc a resultats també diferents. Mentre que en PCA les matrius de factors, matriu de *loadings* i d'*scores*, són ortogonals i es troben en la direcció de màxima variància explicada (a més de que els *loadings* es troben normalitzats), en el cas dels procediments MCR, les matrius de factors es busquen de manera que aquests tinguin unes propietats similars a les que realment causen la variació experimental observada. Per exemple, en el cas ambiental, la composició química de les fonts de contaminació i la seva distribució geogràfica a les mostres només pot presentar valors positius o iguals a zero, però no negatius. Aquesta propietat no s'assolirà, en general, en el cas dels factors ortogonals trobats per PCA.

Existeixen diferents algorismes MCR de descomposició bilineal, entre els quals destaquen els procediments iteratius de resolució. Aquests mètodes de resolució iteratius recuperen el model bilineal a través d'una optimització iterativa de **X** i/o  $\mathbf{Y}^T$  o bé a través de l'optimització d'una matriu de transformació **R**. Així, el model bilineal bàsic de l'Equació 2.11, també es pot escriure com:

$$\mathbf{D} = (\mathbf{X}_0 \mathbf{R}) (\mathbf{R}^{-1} \mathbf{Y}_0^T) + \mathbf{E} \quad \text{Equació 2.12}$$

$$\text{On } \mathbf{X} = \mathbf{X}_0 \mathbf{R} \quad \text{Equació 2.13}$$

i

$$\mathbf{Y}^T = \mathbf{R}^{-1} \mathbf{Y}_0^T \quad \text{Equació 2.14}$$

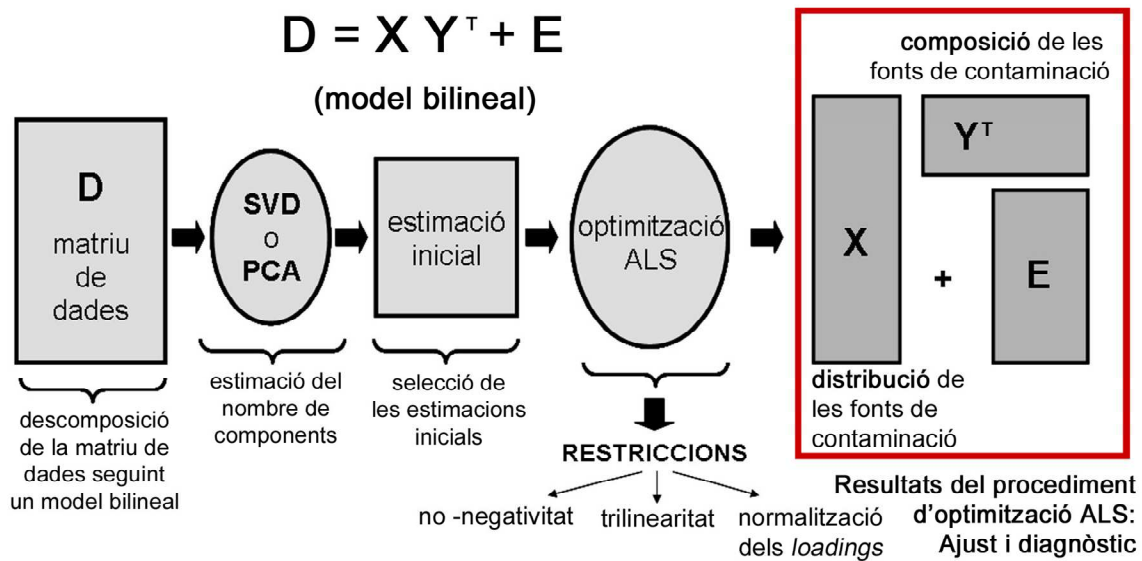
On  $\mathbf{X}_0$  i  $\mathbf{Y}_0^T$  són les estimacions inicials de  $\mathbf{X}$  i  $\mathbf{Y}^T$  i poden ser factors abstractes o bé factors reals, depenent del mètode de resolució utilitzat. Tots els mètodes de resolució per iteracions comencen el procés d'optimització amb les estimacions inicials d'algunes de les matrius del model bilineal ( $\mathbf{X}_0$ ,  $\mathbf{Y}_0^T$  o bé  $\mathbf{R}_0$ ) i aquestes estimacions inicials són posteriorment modificades de manera iterativa sota l'acció de diferents restriccions. Les restriccions forcen als factors  $\mathbf{X}$  i /o  $\mathbf{Y}^T$  a obeir unes propietats químiques o matemàtiques prèviament definides, i són les característiques principals dels mètodes iteratius. Les restriccions constitueixen, al mateix temps, una manera d'introduir tota la informació matemàtica i química que es troba disponible d'entrada, per tal d'assegurar l'obtenció de solucions interpretables. A més, la correcta selecció i aplicació de les restriccions, minimitza o fins i tot pot arribar a eliminar l'ambigüitat relacionada amb els mètodes de resolució multivariant (descrita més endavant).

La diferència entre els diversos mètodes iteratius de resolució que existeixen a l'actualitat es pot relacionar amb el tipus de factors que acostumen a ser optimitzats de manera iterativa, les estimacions inicials utilitzades, la naturalesa i aplicació de restriccions, o bé el tipus d'estructura de dades necessària per tal de que l'anàlisi sigui possible.

Entre els mètodes iteratius de resolució multivariant de corbes destaca el mètode de resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS). Aquest mètode utilitza una implementació específica dels mínims quadrats alternats (ALS) que ha estat proposada i desenvolupada pel grup de recerca dins del qual s'ha realitzat aquesta Tesi doctoral [29, 35-36]. El mètode bàsic del procediment MCR-ALS es basa també en el model bilineal que s'ha descrit a l'Equació 2.11. El mètode troba, de manera iterativa, les matrius dels factors  $\mathbf{X}$  i  $\mathbf{Y}^T$  a través de l'optimització d'estimacions inicials d'aquestes per un procediment de mínims quadrats alternats, que fa mínima una funció minimoquadràtica definida a partir dels residuals entre els valors de les dades experimentals i els valors corresponents estimats com a resultat de l'optimització de les matrius de factors  $\mathbf{X}$  i  $\mathbf{Y}^T$  i, cosa molt important, utilitzant determinades restriccions durant el procés d'optimització. En l'MCR-ALS, cap de les dues matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$  té prioritats respecte a l'altra, i les dues seran utilitzades en cada cicle d'iteració. El procediment

general utilitzat per MCR-ALS inclou els passos que es descriuen a continuació (veure Figura 2.7):

1. Determinació del nombre de components de la matriu de dades experimentals  $\mathbf{D}$
2. Generació de les estimacions inicials (per exemple,  $\mathbf{X}_0$ , per la matriu  $\mathbf{X}$ )
3. Càlcul de la matriu  $\mathbf{Y}^T$  sota l'aplicació de determinades restriccions
4. Càlcul de la matriu  $\mathbf{X}$  sota l'aplicació de restriccions
5. Obtenció de la matriu  $\mathbf{D}$  mitjançant el producte de  $\mathbf{X}$  i  $\mathbf{Y}^T$
6. Càlcul de la funció minimoquadràtica dels residuals definits a partir de la diferència entre la matriu experimental  $\mathbf{D}$  i la seva estimació iterativa  $\mathbf{X}\mathbf{Y}^T$ ,  $\|\mathbf{D} - \mathbf{X}\mathbf{Y}^T\|$ , o arrel quadrada de la suma dels residuals al quadrat, definits com a diferència entre els valors experimentals i els estimats per l'aplicació del model bilineal (Equació 2.11) amb les estimacions de  $\mathbf{X}$  i  $\mathbf{Y}^T$  a la iteració considerada
7. Retorn al pas número 3 fins que s'assoleix la convergència, és a dir, fins que els residuals definits en el pas 6, ja pràcticament no canviïn



**Figura 2.7.** Diagrama d'implementació del mètode MCR-ALS amb l'exemple d'aplicació de tres restriccions diferents: no negativitat, normalització dels *loadings* i trilinearitat

### Determinació del nombre de components

El nombre de components ( $M$ ) de la matriu  $\mathbf{D}$  pot ser inicialment estimat utilitzant PCA o bé pot ser ja conegut d'entrada. En qualsevol cas, es recomana resoldre el sistema utilitzant diferent nombre de components, per tal de comparar els resultats i escollir aquell nombre de components que millor descriu les dades. Dit d'una altra manera, aquell nombre de components que proporcioni un error minimoquadràtic mínim i que doni solucions que compleixin les restriccions imposades i siguin plausibles dins del context del problema que es

vol resoldre. En el cas de les dades analitzades en aquesta Tesi (dades ambientals),  $N$  es referirà a les fonts o pautes de variació (contaminació ambiental) més importants i diferenciades. Un cop determinat el nombre de components, existirà un percentatge significatiu de variància no explicada provinent del soroll experimental i possiblement també de múltiples fonts de contaminació naturals i no naturals desconegudes, que tenen cadascuna d'elles una contribució a la variància petita, i que es trobaran agrupades dins de la matriu de residuals  $\mathbf{E}$ .

En els models bilineals, els percentatges de variància explicada per cada component poden ser calculats a partir de l'expressió següent:

$$R^2 = 1 - \frac{\sum_{i=1}^I \sum_{j=1}^J e_{i,j}^2}{\sum_{i=1}^I \sum_{j=1}^J d_{i,j}^2} \quad e_{i,j} = d_{i,j} - \hat{d}_{i,j} \quad \text{Equació 2.15}$$

On  $d_{i,j}$  és el valor de la variable  $j$  a la mostra  $i$  de la matriu de dades experimentals i  $\hat{d}_{i,j}$  és el valor corresponent calculat utilitzant MCR.  $R^2$  pot ser calculada per tot el model (pels  $N$  components considerats conjuntament) així com també per a cadascun dels components individuals d'un model determinat per separat.

### Selecció de les estimacions inicials

Una vegada realitzada l'estimació del nombre de components, qualsevol model de resolució iterativa basat en el model bilineal descrit a l'Equació 2.11, requereix l'estimació inicial d'una de les dues matrius,  $\mathbf{X}$  o bé  $\mathbf{Y}^T$  (veure Figura 2.7), per tal de començar el procés iteratiu d'optimització. Posteriorment, les matrius dels dos factors seran modificades pas a pas durant els cicles iteratius consecutius fins que s'assoleixi la convergència final.

Existeixen múltiples possibilitats per tal de generar les estimacions inicials, no obstant, en el context de MCR, és millor evitar les estimacions aleatòries. Enlloc d'aquestes, l'ús d'estimacions inicials que siguin plausibles (que compleixin les restriccions imposades) i tinguin un significat ambiental disminueix el temps computacional i minimitza els problemes de convergència i de mínims locals, ja que el conjunt de factors inicial obeeix les mateixes propietats que tenen els factors reals i, per tant, aquests factors inicials es troben més propers de la solució final real. Apart de la presència de mínims locals diferents, estimacions inicials diferents poden donar lloc a solucions diferents ambientalment però equivalents matemàticament degut a l'existència dels problemes d'ambigüitat rotacional (descrits més endavant). No obstant, totes aquestes solucions obtingudes presentaran un ajust òptim de la



## Capítol 2

---

funció minimoquadràtica i seran ambientalment possibles segons el coneixement previ que es tingui del sistema i les restriccions imposades.

Les estimacions inicials no aleatòries poden ser obtingudes directament a partir del conjunt de les dades sense processar, o bé poden ser el resultat de l'aplicació de mètodes quimiomètrics auxiliars. Així, la selecció d'un conjunt de columnes o d'un conjunt de files a partir de la matriu de dades original  $\mathbf{D}$ , igual en nombre als components estimats en el sistema, pot proporcionar una estimació inicial de les matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$  respectivament. Aquesta estimació es pot obtenir seleccionant directament aquelles files (per estimar  $\mathbf{Y}^T$ ) o bé aquelles columnes (per estimar  $\mathbf{X}$ ) de la matriu  $\mathbf{D}$  que siguin més diferents entre elles i, per tant, més pures o selectives. Aleshores, si el nombre de components escollit és correcte, qualsevol fila o columna de la matriu de dades serà combinació lineal d'aquestes files o columnes més pures.

En alguns casos, el simple coneixement previ del problema ambiental a analitzar pot ser suficient per tal de realitzar una selecció encertada de les estimacions inicials. D'altra banda, també existeixen mètodes quimiomètrics auxiliars quan es disposa de poca informació sobre el conjunt de dades, o bé simplement quan es prefereix obtenir estimacions inicials basades en procediments matemàtics.

Un dels mètodes quimiomètrics que permet obtenir estimacions inicials per qualsevol conjunt de dades és el mètode SIMPLISMA (*SIMPL*e-to-use *I*terative *S*elf-*M*odeling *A*nalysis [37-38]). Aquest mètode selecciona, de manera seqüencial, aquelles variables, en la direcció de les columnes o bé en la direcció de les files, que menys informació tenen en comú amb les seleccionades prèviament. El mètode SIMPLISMA és el que s'ha utilitzat per realitzar les estimacions inicials en els casos estudiats en aquesta Tesi. Altres mètodes utilitzats per aconseguir estimacions inicials es troben descrits en diversos treballs [39-40].

### Optimització ALS

Després de l'obtenció de les estimacions inicials, s'inicia el procediment de mínims quadrats alternats, el qual consisteix en resoldre les dues equacions següents sota les restriccions adjients:

$$\min_{\hat{\mathbf{X}}} \left\| \mathbf{D}_{\text{PCA}} - \hat{\mathbf{X}}\hat{\mathbf{Y}}^T \right\| \quad \text{Equació 2.16}$$

$$\min_{\hat{\mathbf{Y}}^T} \left\| \mathbf{D}_{\text{PCA}} - \hat{\mathbf{X}}\hat{\mathbf{Y}}^T \right\| \quad \text{Equació 2.17}$$

En les Equacions 2.16 i 2.17, la norma dels residuals (arrel quadrada de la suma dels residuals al quadrat) entre les dades reproduïdes per PCA,  $\mathbf{D}_{PCA}$ , utilitzant el nombre seleccionat de components, i les dades reproduïdes per l'ALS utilitzant les estimacions per mínims quadrats de les matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$ ,  $\hat{\mathbf{X}}$  i  $\hat{\mathbf{Y}}^T$ , (sota determinades restriccions) és alternativament minimitzada mantenint constant  $\hat{\mathbf{X}}$  o bé  $\hat{\mathbf{Y}}^T$ .

Es comença calculant la solució mínimoquadràtica inicial per a la matriu per la qual no s'ha realitzat l'estimació inicial. Així, si per exemple es disposa d'una estimació inicial dels factors per  $\mathbf{X}$ , la millor solució per mínims quadrats (en aquest cas, sense restriccions) de la matriu  $\mathbf{Y}^T$  serà calculada tal com s'indica a continuació:

$$\hat{\mathbf{Y}}^T = \hat{\mathbf{X}}^+ \mathbf{D}_{PCA} \quad \text{Equació 2.18}$$

On  $\hat{\mathbf{X}}^+$  és la matriu pseudoinversa de  $\mathbf{X}$  [21], definida com aquella matriu que multiplicada per  $\mathbf{X}$  dóna la matriu identitat. Si, en canvi, es disposa d'una estimació inicial dels factors de concentració  $\mathbf{Y}^T$ , la millor solució es calcularà de la següent forma:

$$\hat{\mathbf{X}} = \mathbf{D}_{PCA} (\hat{\mathbf{Y}}^T)^+ \quad \text{Equació 2.19}$$

Els dos problemes de mínims quadrats de les Equacions 2.16 i 2.17, es resolen de manera seqüencial en cadascun dels cicles iteratius, de manera que la matriu  $\mathbf{Y}^T$  és calculada i posteriorment utilitzada per obtenir la matriu  $\mathbf{X}$ . Cal remarcar que la matriu emprada durant la minimització mínimoquadràtica del procediment d'optimització és, segons les Equacions 2.16 a 2.19, la matriu calculada a partir d'un model de PCA amb un nombre determinat de components,  $\mathbf{D}_{PCA}$ . Es tracta, per tant, d'una matriu a la qual ja s'ha eliminat una part important del soroll experimental i que per tant, tot i mantenir pràcticament tota la informació química rellevant continguda en les dades originals, facilita el procés d'optimització i l'assoliment més fiable de la convergència dels factors optimitzats i de l'obtenció de les solucions finals.

El criteri de convergència en l'optimització ALS es basa en la comparació de l'ajust obtingut en dues iteracions consecutives. Quan la diferència d'ajust relativa mesurada a partir de la desviació estàndard dels residuals es troba per sota d'un valor lliandar establert (normalment fixat al 0,1%), el procés d'optimització es dóna per finalitzat. Altres possibilitats consisteixen en definir un nombre màxim de cicles iteratius com a criteri de finalització, o bé en comparar la forma dels factors resolts en iteracions consecutives. Sempre que s'utilitzin les restriccions adequades durant els processos d'optimització, MCR-ALS proporcionarà solucions possibles,

encara que les ambigüitats rotacionals i/o d'intensitat poden continuar-hi presents depenent de l'estructura, la complexitat i el coneixement previ del sistema sota estudi. Una versió MATLAB de l'algoritme corresponent al procediment MCR-ALS que inclou una interfase gràfica per a facilitar la seva implementació, es pot trobar disponible i pot ser descarregada lliurement des de la pàgina web [www.ub.edu/gesq/mcr/mcr.htm](http://www.ub.edu/gesq/mcr/mcr.htm) [41].

### Resultats de l'ALS. Ajust i diagnòstic

Com a resultat del procés d'optimització ALS s'obté (veure Figura 2.7) una matriu  $\mathbf{Y}^T$  o matriu de *loadings*, amb la descripció de les variables que formen part dels components que, en el cas ambiental, correspondrà a la composició de les fonts de contaminació identificades, una matriu  $\mathbf{X}$  o matriu d'*scores*, amb la distribució d'aquestes fonts a les mostres, i finalment una matriu  $\mathbf{E}$ , amb l'error o part de la variància no explicada pel model.

Ja que el nombre de components no és un paràmetre conegut d'entrada, és important comparar l'ajust obtingut per diferent nombre de components així com investigar el sentit de les solucions obtingudes en cada cas. De la mateixa manera, quan s'analitzen simultàniament múltiples conjunts de dades ordenades en matrius de mides similars, caldrà comparar els resultats obtinguts amb l'aplicació de diferents tipus de models (bilineal o trilineal) quan no s'estigui del tot segur del model que segueixen les dades.

### Incerteses inherents al mètode MCR-ALS

Les principals fonts d'incertesa associades als resultats obtinguts amb els mètodes de resolució multivariant són les ambigüitats (rotacionals i d'escala o d'intensitat) dels components recuperats i la propagació del soroll experimental [36, 42]. Per tant, l'estudi d'aquests dos efectes és necessari quan es vol avaluar la qualitat dels resultats obtinguts i es vol comparar aquests resultats amb els obtinguts mitjançant altres procediments quimiomètrics. El desenvolupament de metodologies per quantificar les incerteses no és només un tema d'interès creixent en la literatura quimiomètrica actual sinó també una necessitat per tal d'estendre la utilització d'aquests mètodes de resolució de forma més eficient.

El problema a resoldre en la resolució multivariant de corbes, pot ser descrit matemàticament de la següent manera:

Donada una matriu de dades  $\mathbf{D}$ :

- (1) caldrà trobar  $N$ , el nombre de components causant de la variància observada a les dades,

- (2) trobar la contribució dels components a les diferents mostres analitzades (perfils d'scores) o matriu  $\mathbf{X}$
- (3) trobar la contribució de les variables als components identificats (perfils de *loadings*) o matriu  $\mathbf{Y}^T$

No obstant, les solucions obtingudes per  $\mathbf{X}$  i  $\mathbf{Y}^T$  a partir de l'Equació 2.11 (que correspon a la descomposició de la matriu de dades  $\mathbf{D}$  seguint un model bilineal), serà ambigua si no es disposa d'informació addicional, o dit d'una altra manera, existirà, en les solucions obtingudes de l'Equació 2.11, una llibertat rotacional i d'escala. Aquest problema sovint rep el nom de problema d'ambigüitat de l'anàlisi de factors. Les ambigüitats rotacionals i d'escala, es mostren de manera simplificada a l'Equació 2.20:

$$\mathbf{D} = \mathbf{X}_o \mathbf{Y}_o^T = (\mathbf{X}_o \mathbf{T})(\mathbf{T}^T \mathbf{Y}_o^T) = \mathbf{X}_k \mathbf{Y}_k^T \quad \text{Equació 2.20}$$

On  $\mathbf{X}_o$  i  $\mathbf{Y}_o^T$  corresponen a matrius inicials d'scores i *loadings* obtingudes de la descomposició bilineal de la matriu  $\mathbf{D}$ , i  $\mathbf{X}_k$  i  $\mathbf{Y}_k^T$  són matrius d'scores i *loadings* obtingudes a partir del producte de les matrius inicials d'scores i *loadings* per una matriu  $\mathbf{T}$  (anomenada matriu de rotació) [43]. D'aquí se'n dedueix que, si no s'imposa cap tipus de restricció, existiran un nombre infinit de solucions a l'Equació 2.20 per qualsevol matriu invertible  $\mathbf{T}$ . Diverses combinacions lineals de les solucions poden ser obtingudes a partir de matrius de rotació apropiades  $\mathbf{T}$ .

L'existència de les ambigüitats d'intensitat o d'escala, comporta que els components estimats per MCR-ALS (o per qualsevol altre mètode de descomposició bilineal) estiguin escalats per un factor desconegut, el qual haurà de ser fixat ja sigui en base al coneixement extern o bé per algun tipus de normalització (en molts casos, els perfils de *loadings* es normalitzen entre 0 i 1), per tal d'evitar que surtin fora d'escala durant l'optimització per ALS. D'altra banda, existeix un altre tipus d'ambigüitat, anomenada ambigüitat rotacional, que és un problema persistent i més seriós que l'ambigüitat d'escala. Aquesta apareixerà quan hi ha dos o més components que se superposen i, a conseqüència d'això, és possible obtenir un conjunt de solucions diferents que ajusten igualment bé les dades i que compleixen correctament les restriccions posades al sistema (la solució no és única). És a dir, les solucions obtingudes són equivalents des d'un punt de vista d'optimització matemàtica, però són diferents des d'un punt de vista fisicoquímic. Quan això passa, no hi ha manera d'estar segur de si la solució obtinguda és la solució veritable. No obstant, normalment és possible reduir considerablement aquest nombre infinit de solucions mitjançant l'aplicació de restriccions derivades de la naturalesa i del coneixement previ del problema que s'estudia. Una discussió més detallada d'aquest tipus d'ambigüitats cau fora de l'abast d'aquesta Tesi i constitueix un dels temes actuals de recerca (veure, per exemple, Abdollahi et al. (2009) [44]).

## Capítol 2

---

A banda de les ambigüitats d'intensitat i rotacionals, existeix un altre problema que sempre es troba present en l'anàlisi de dades experimentals, que és el problema de la propagació de soroll. Fins i tot en absència d'ambigüitats rotacionals, l'error, inevitablement present en les dades reals, es propaga a les solucions obtingudes. Aquesta font d'incertesa afecta els resultats de tot tipus de mètodes d'anàlisi de dades. En els casos senzills (com el calibratge univariant o multivariant lineal), la determinació de l'efecte del soroll es pot estimar a partir de formules matemàtiques analítiques senzilles. Però l'error associat als mètodes de resolució multivariant és un problema complex, degut a l'elevat nombre de paràmetres no lineals que han de ser calculats i a l'aplicació de les restriccions. S'han proposat procediments d'estimació dels errors associats a les incerteses a partir de l'aplicació de procediments de simulació dels efectes del soroll utilitzant tècniques de Monte Carlo i *boot-straping* [42, 45]. Per altra banda, la disminució i eventual eliminació dels efectes del soroll o error experimental sobre els paràmetres es pot aconseguir a partir de l'aplicació de mètodes basats en aproximacions de màxima versemblança (*maximum likelihood*) com ara els de sospesament de mínims quadrats alternats (*Weighted Alternating Least-Squares*, WALs [28, 46]) proposats recentment. Una discussió de tots aquests aspectes cau també fora de l'àmbit d'aquesta Tesi, però poden consultar-se en els diversos treballs citats en aquest apartat i d'altres [47-48].

### Restriccions

Les restriccions són propietats químiques o matemàtiques generals que les matrius de factors  $X$  i /o  $Y^T$  han de complir, les quals actuen dirigint el procediment d'optimització iteratiu cap a una solució amb significat físic i interpretable en termes ambientals. La utilització adequada de restriccions és una condició imprescindible en tots els procediments de resolució multivariant i permet disminuir l'ambigüitat de les solucions finals obtingudes. Les restriccions poden ser classificades d'acord amb la seva naturalesa o bé d'acord amb la seva utilització. Quan s'aplica una determinada restricció, el coneixement matemàtic i ambiental sobre els factors ha de ser traduït a condicions matemàtiques.

Una restricció o condició pot ser imposada de dues maneres: forçant el factor (o alguns dels seus elements) a ser iguals a determinats valors preestablerts (restriccions d'igualtat, *equality constraints*), o bé forçar-lo a ser superior o inferior a ells (restriccions de desigualtat, *inequality constraints*). D'aquesta manera, si a un factor se li aplica la restricció de no negativitat, els elements negatius d'aquest factor són forçats a ser positius (o zero). En el cas de dades ambientals, la restricció de no negativitat pot ser aplicada, per exemple, sobre concentracions químiques (valors de les variables en els *loadings*) o bé sobre les contribucions de les fonts de contaminació sobre les mostres (valors dels *scores*), ja que en els dos casos només tenen sentit si són positives [49]. Hi ha diferents formes d'aplicar la restricció de no negativitat, com la

substitució directa dels valors negatius per zero durant els diferents passos iteratius o els mètodes més rigorosos, basats en procediments per mínims quadrats no negatius (*non-negative least squares*, nnls, [49]) i les seves variants més ràpides, com per exemple la *fast non-negative least squares*, fnnls [50]. En aquesta Tesi s'ha utilitzat la basada en aquest darrer mètode de mínims quadrats no negatius ràpids (fnnls). A més, la matriu de *loadings*  $\mathbf{Y}^T$ , es normalitza per tal d'evitar problemes d'ambigüitats d'escala [36] (veure apartat de les incerteses inherents al mètode MCR-ALS).

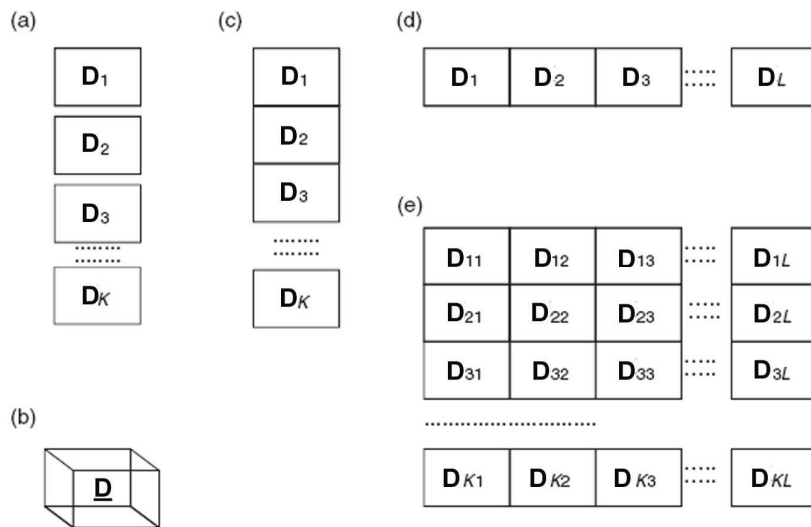
A banda de les restriccions de no negativitat i de normalització (sovint aplicades per a dades ambientals), el mètode MCR-ALS també permet l'aplicació d'altres restriccions majorment utilitzades en altres camps de la química analítica, com són l'unimodalitat (permet als perfils tenir només un màxim), la de sistema tancat (degut, per exemple, al compliment de balanços de massa), les de selectivitat o rang local, o les de modelatge rígid (*hard-modeling*), sovint utilitzades en el cas de sistemes químics que segueixen parcial o totalment models cinètics o d'equilibri. La descripció de totes aquestes possibilitats cau fora de l'àmbit d'aquesta Tesi, i es troben recollides a la bibliografia d'aplicació del procediment MCR-ALS [36, 47, 51-54].

Finalment queda un tercer tipus de restricció, la restricció de trilinearitat, que ha estat emprada en l'àmbit d'aquesta Tesi. Aquesta restricció és aplicada en l'anàlisi simultània de diverses matrius de dades (ambientals i no ambientals), i es troba descrita en el següent apartat, que descriu l'anàlisi simultània de dades ordenades en tres o més direccions (*multiway data*).

És important, a l'hora d'aplicar qualsevol restricció, considerar que aquesta estigui plenament justificada des d'un punt de vista matemàtic o bé ambiental. El fet d'utilitzar una restricció inadequada, pot donar lloc a resultats erronis i pitjors que si aquesta no s'apliqués. De vegades, poden aparèixer dubtes sobre l'aplicació o no d'una determinada restricció. En aquests casos, és millor anar introduint les restriccions de manera gradual durant el procés de resolució, i anar comprovant l'efecte que té la seva introducció a partir del nivell d'ajust obtingut en la reproducció del conjunt de dades original i l'efecte sobre les solucions obtingudes. El decreixement significatiu de l'ajust i l'aparició de residuals amb tendències no aleatòries lligats a la introducció d'una restricció determinada, és una indicació de que probablement les dades no la compleixen. Quan l'aplicació de les restriccions porta a solucions que no són raonables des d'un punt de vista físic (ambiental), caldrà comprovar si la seva aplicació ha estat massa estricta, i ampliar els nivells de tolerància per al seu incompliment. Finalment, cal remarcar que existeixen restriccions que, un cop aplicades sobre les dades, permeten eliminar l'ambigüitat rotacional. Aquest és el cas de les restriccions de trilinearitat, de les de model rígid o de les de selectivitat i rang local.

**Extensió del mètode MCR-ALS a l'anàlisi simultània de múltiples conjunts de dades**

Un dels majors avantatges del mètode MCR-ALS és que pot ser fàcilment estès a l'anàlisi simultània de diversos conjunts de dades obtinguts en experiments, situacions o condicions diferents. Això és possible a partir de l'augmentació o concatenació de les matrius de dades individuals corresponents als experiments individuals. Una matriu de dades pot ser augmentada per files (*row-wise*), per columnes (*column-wise*) o bé per files i columnes (*row-and column-wise*) per tal de formar una nova estructura multimatriu on les matrius amb el mateix nombre de files, columnes o ambdues a la vegada, són concatenades en la direcció apropiada [48]. El primer avantatge que cal destacar d'aquesta augmentació de matrius de dades és que aquelles característiques de resolució millors per una o algunes de les matrius analitzades, sempre tindran un efecte positiu sobre la resolució d'aquelles altres matrius més complexes. Tal com es mostra a la Figura 2.8, hi ha tres tipus possibles d'augmentació o concatenació de les matrius de dades, les quals poden ser indicades de manera concisa utilitzant la notació de l'entorn de programació MATLAB. Les matrius de dades augmentades per columnes,  $D_{aug}$  (veure Figura 2.8c), s'escriuen comon el símbol “;” s'utilitza per indicar que les diferents matrius de dades  $D_k$ ,  $k=1,\dots,K$ , (on  $k$  indica les diferents experiments, condicions o situacions, les quals poden correspondre per exemple, a diferents temps, a diferents compartiments ambientals, etc.) es troben concatenades per columnes, una sobre l'altra, mantenint el mateix nombre de columnes que les matrius individuals. En aquesta estructura de dades, se suposa que les diferents matrius comparteixen el seu espai vectorial per columnes.



**Figura 2.8.** Diferents estructures i ordenacions de dades (Tauler et al., 2009); (a) matrius individuals, (b) conjunt de dades multiway (cub de dades), i (c)-(e) disposició de múltiples matrius. (c) matrius augmentades per columnes, (d) matrius augmentades per files, i (e) matrius augmentades per files i per columnes

Per aquest tipus d'augmentació de les dades, el model bilineal es pot escriure mitjançant la següent equació:

$$\begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_k \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k \end{pmatrix} Y^T + \begin{pmatrix} E_2 \\ E_2 \\ E_3 \\ \vdots \\ E_k \end{pmatrix} = X_{aug} Y^T + E_{aug} \quad \text{Equació 2.21}$$

O, d'una forma més compacta (notació MATLAB):

$$[D_1; D_2; D_3; \dots; D_k] = [X_1; X_2; X_3; \dots; X_k] Y^T + [E_1; E_2; E_3; \dots; E_k] = X_{aug} Y^T + E_{aug} \quad \text{Equació 2.22}$$

En el model bilineal descrit a les Equacions 2.21 i 2.22 només es formula una sola matriu  $Y^T$ , equivalent a la matriu de *loadings*, que en el cas de les taules de dades ambientals descriurà la composició de les diferents fonts de contaminació identificades, comuna a totes les matrius  $D_k$  considerades. D'altra banda, la matriu augmentada d'*scores*  $[X_1; X_2; X_3; \dots; X_k]$ , descriurà (també pel cas de taules ambientals) els canvis en la distribució de cadascun dels components o fonts de contaminació identificades en cadascuna de les mostres contingudes a les diferents matrius  $D_k$ . Així, el model bilineal de les Equacions 2.21 i 2.22 assumeix que els components identificats a les matrius de dades  $D_k$  inclosos a la matriu augmentada per columnes, comparteixen els mateixos perfils de composició, mentre que la seva distribució a les mostres és diferent en cadascuna de les  $k$  condicions o situacions.

Les matrius de dades augmentades per files (veure Figura 2.8d) es col·loquen una al costat de l'altra per a la seva concatenació. En aquest cas, es considera que les diferents matrius comparteixen l'espai vectorial de les seves files. Les matrius de dades també poden ser augmentades per columnes i per files al mateix temps (veure Figura 2.8e). Això significa que algunes de les matrius comparteixen l'espai vectorial de les seves columnes mentre que d'altres el comparteixen de les files. L'augmentació de matrius per files i en ambdues direccions (files i columnes) no es descriuen amb més detall en aquest treball pel fet que no han estat utilitzades en aquesta Tesi.

El model bilineal descrit a l'Equació 2.22 pot ser aplicat de diferents maneres partint de la informació continguda a les matrius de dades originals. En MCR, només es consideraran aquelles solucions del model que tinguin un significat físic i interpretable en termes ambientals. En aquest punt, MCR es diferencia d'altres aproximacions basades en models bilineals, com



## Capítol 2

---

ara PCA, que també pot ser estesa a l'anàlisi de matrius de dades augmentades utilitzant equacions similars. El mètode PCA busca solucions que siguin ortogonals i que expliquin la màxima variància de les dades (sense superposició, no compartida). És per aquest motiu que les solucions obtingudes de PCA no tenen un significat físic directe. En MCR, s'apliquen diferents tipus de restriccions per tal d'obtenir uns perfils dels components que tinguin un sentit físic i que, per tant, puguin ser fàcilment interpretables en termes ambientals. La implementació de restriccions naturals (com la no negativitat i la normalització, en el cas de les dades ambientals), seguirà la mateixa estratègia adoptada per a l'anàlisi de les matrius individuals. Tot i això, cal tenir en compte que les restriccions seran aplicades individualment i per separat per a cadascuna de les submatrius en les matrius augmentades  $\mathbf{X}_{\text{aug}}$  (si es realitza l'augmentació per columnes).

El problema de mínims quadrats resolt per MCR-ALS quan aquest s'aplica a conjunts de dades ordenats en múltiples direccions, és el mateix que quan aquest és aplicat a l'anàlisi d'una matriu de dades individual; la única diferència és que ara  $\mathbf{D}$  i  $\mathbf{X}$  i/o  $\mathbf{Y}^T$  són matrius augmentades. No obstant, l'aplicació d'MCR-ALS a aquest conjunt més complex de dades no hauria de ser considerada equivalent a la descomposició d'una matriu de dades individual, degut a la possible aplicació de les restriccions a nivell de submatriu. El primer avantatge clar de l'anàlisi simultània de múltiples matrius de dades és que, en general, aquesta genera estimacions per mínims quadrats que són més robustes i precises.

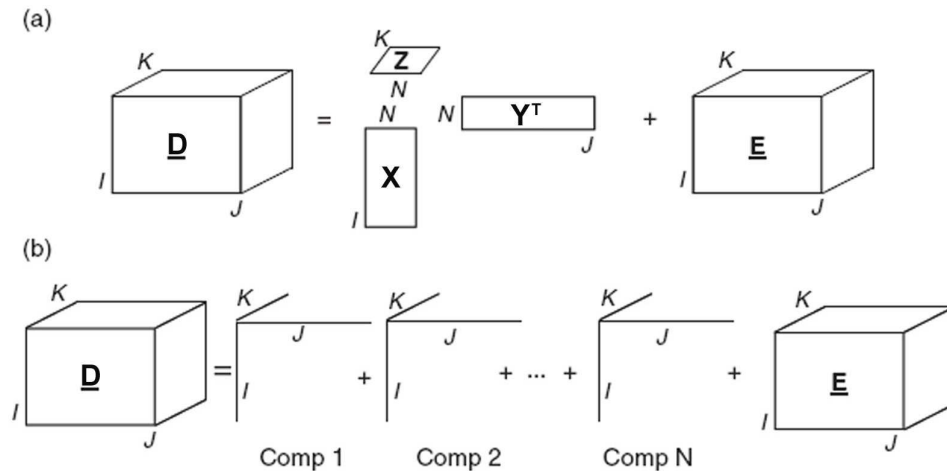
Una altra característica interessant de l'anàlisi de múltiples matrius de dades és la possibilitat d'aplicar noves restriccions relacionades amb les estructures de dades anomenades *multiway*. Aquest és el cas quan, per exemple, les diferents matrius de dades analitzades simultàniament tenen les mateixes mides (files i columnes), i es pot construir amb elles un cub de dades. En aquests casos, el model bilineal es pot utilitzar també sobre aquestes noves estructures de dades, però aquest model bilineal es pot estendre igualment a un model trilineal. Aquí, la restricció aplicada en el mètode MCR-ALS s'anomena de trilinearitat, i implica que els perfils dels factors d'un mateix component en les diferents matrius de dades analitzades simultàniament tinguin la mateixa forma i que només variïn entre si en un factor d'escala.

L'expressió del model trilineal (també anomenat anàlisi de factors paral·lels o *parallel factor analysis*, PARAFAC [30, 55]) utilitzada en aquesta Tesi per descriure la descomposició d'un conjunt de dades ordenat en tres direccions (*three-way data set*) es dona a l'Equació 2.23.

$$\mathbf{D}_k = \mathbf{XZ}_k\mathbf{Y}^T + \mathbf{E}_k$$

Equació 2.23

On  $\mathbf{X}$ ,  $\mathbf{Z}$  i  $\mathbf{Y}^T$  són les matrius de factors (perfils dels components en les tres direccions, modes o ways) i  $\mathbf{E}$  la matriu de residuals (part de la variància de les dades no explicada pel model). Aquesta descomposició trilineal d'un conjunt de dades en tres direccions  $\mathbf{D}$  es troba també representada gràficament a la Figura 2.9.



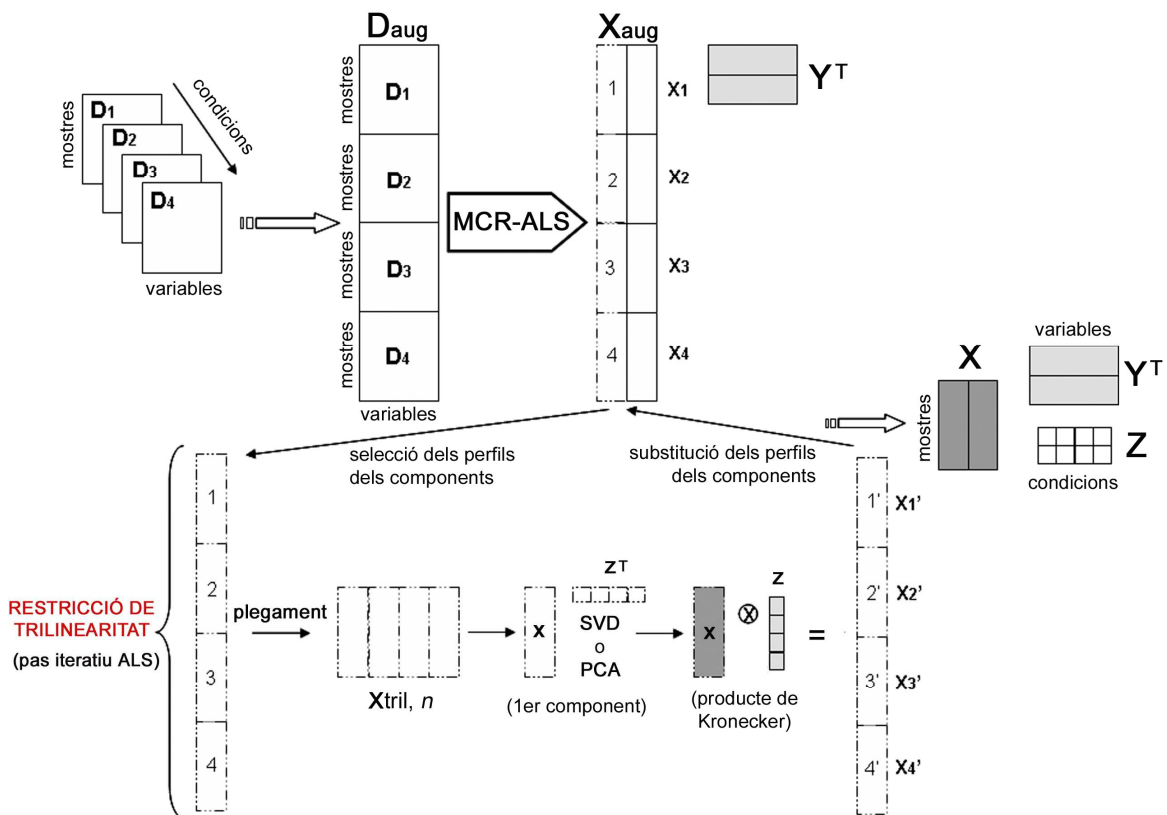
**Figura 2.9.** Descomposició d'un conjunt de dades en tres direccions seguint un model trilineal (Tauler et al., 2009) (a) expressat mitjançant la descomposició en matrius de *loadings* i (b) expressat com la suma dels components purs

L'anàlisi d'un conjunt de dades en tres direccions mitjançant un model trilineal genera directament tres conjunts de matrius de factors amb els perfils dels components (*loadings*), amb el mateix nombre de components en els tres modes o direccions. A la Figura 2.9a, la matriu  $\mathbf{X}$  correspon als *loadings* comuns en el primer mode (també anomenada *scores* en el cas dels models bilineals de PCA quan les dades s'ordenen en dues direccions) de totes les matrius analitzades, i conté la contribució dels diferents components sobre les mostres, que en termes ambientals correspon a la distribució geogràfica o temporal de les fonts de contaminació identificades. La matriu  $\mathbf{Y}^T$  són els *loadings* comuns en el segon mode de totes les matrius analitzades, i conté la contribució de les diferents variables en els components, que en el cas ambiental correspon a la composició química de les fonts de contaminació identificades. La nova matriu de dades  $\mathbf{Z}_k$  de *loadings* en el tercer mode o direcció, és una matriu diagonal, amb la contribució relativa de cada component en cadascuna de les matrius de dades considerades  $\mathbf{D}_k$ . Tal com es descriu en l'Equació 2.23, els elements de la matriu  $\mathbf{Z}_k$  seran els únics valors en el model de descomposició trilineal que canvien per a les diferents matrius  $\mathbf{D}_k$ . Les matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$  són úniques (comunes) i invariants (les fonts o components identificats tenen la mateixa composició i la mateixa distribució a les mostres) per a cada component individual en totes les matrius  $\mathbf{D}_k$  analitzades simultàniament. La trilinearitat és doncs una restricció forta, que només es compleix en determinades ocasions, però per altra banda i com a conseqüència de la seva

## Capítol 2

aplicació, permet l'obtenció de solucions úniques per a les tres matrius de factors i evita la presència de les ambigüitats rotacionals associades a models bilineals.

El mètode MCR-ALS és adequat tant per a l'anàlisi de dades ordenades en tres direccions, formades per l'augmentació de matrius individuals de diferents mides i que no presenten una estructura trilineal, com per a l'anàlisi de conjunts de dades trilineals, constituïts per diferents matrius de dades amb el mateix nombre de files i columnes (cubs de dades). En aquest darrer cas, MCR-ALS aprofita els avantatges d'aquesta estructura i pot forçar les solucions obtingudes (les matrius de factors en els diferents modes) a que aquestes compleixin les premisses del model trilineal. L'estructura trilineal s'incorpora com una restricció opcional durant l'optimització ALS dels perfils de la matriu  $\mathbf{X}_{aug}$ . A més de ser opcional, aquesta restricció pot ser aplicada per separat sobre cadascun dels components. A continuació es mostra un esquema de la seva implementació en el mètode MCR-ALS [56]:



**Figura 2.10.** Descomposició de la matriu de dades augmentada  $\mathbf{D}_{aug}$  per MCR-ALS aplicant la restricció de trilinearitat.  $\mathbf{D}_{aug}$  formada per la concatenació de les matrius individuals ( $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4$ ).  $\mathbf{X}_{aug}$  és la matriu augmentada d'scores (mostres en les diferents matrius).  $\mathbf{Y}^T$  és la matriu de *loadings* en el segon mode (variables).  $\mathbf{X}$  és la matriu de *loadings* en el primer mode (mostres).  $\mathbf{Z}$  és la matriu de *loadings* en el tercer mode (condicions o situacions, equivalents a períodes de temps o compartiments en el cas de dades ambientals).

Tal com es mostra a la Figura 2.10, quan s'aplica la restricció de trilinearitat, els perfils de cada component en el primer mode (*scores* o bé *loadings* en el primer mode) de les diferents matrius  $\mathbf{X}_k$  es veuen forçats a adoptar exactament la mateixa forma (només variarà la seva escala per les diferents matrius de dades) durant cada iteració de l'optimització ALS. Per tal de fer això, en primer lloc, els perfils d'*scores* corresponents a un component en les diferents matrius es col·loquen un al costat de l'altre (plegament) per tal de generar una nova matriu,  $\mathbf{X}_{\text{tril},n}$ , amb  $I$  files (igual al nombre de files de cada matriu de dades) i  $K$  columnes (igual al nombre de matrius analitzades simultàniament),  $n$  indicant el component sobre el qual s'aplica la restricció. En el cas de l'aplicació d'un model totalment trilineal, cada component donarà una matriu  $\mathbf{X}_{\text{tril},n}$  ( $n = 1, \dots, N$ ), que es descomposarà bilinealment emprant només un component (mitjançant, per exemple, la descomposició de valors singulars (SVD)), tal com s'indica a continuació:

$$\mathbf{X}_{\text{tril},n}(I,K) = \mathbf{x}(I,1) \mathbf{z}^T(1,K) \quad \text{Equació 2.24}$$

On  $\mathbf{x}$  és el nou vector columna que conté el perfil de *loadings* comú (mitjà) en el primer mode (*scores*) del component  $n$  en les diferents matrius  $K$ , i  $\mathbf{z}^T$  és un nou vector fila amb el perfil de *loadings* en el tercer mode, que expressa les quantitats relatives d'aquest component en les diferents matrius  $K$ . El producte de Kroneker [57] d'aquests dos nous vectors de *loadings* dona lloc a un nou vector d'*scores* augmentats, el qual substitueix la columna corresponent de la matriu augmentada  $\mathbf{X}_{\text{aug}}$ . Els perfils  $\mathbf{x}$  i  $\mathbf{z}$  en cada iteració ALS donen la darrera estimació dels perfils de *loadings* en el primer i tercer mode per al component considerat. Així, quan la restricció de trilinearitat s'imposa durant el procediment d'optimització iterativa ALS, es força a que la forma dels vectors de *loadings* en el primer mode (distribució de les fonts de contaminació sobre les mostres en el cas de taules de dades ambientals) sigui la mateixa en totes les condicions (períodes de temps o compartiments ambientals). A més a més, la variació relativa de la contribució del component considerat a les diferents matrius, es captura en els *loadings* del tercer mode, mostrant les diferències d'aquest component entre les diferents taules de dades ambientals analitzades simultàniament, per exemple, en períodes de temps o compartiments ambientals diferents.

Les solucions obtingudes mitjançant l'aplicació del model trilineal són úniques i, a diferència del que ocorre en el cas dels models bilineals, no presenten ambigüitats rotacionals [36]. L'assumpció d'un model trilineal simplifica considerablement la interpretació dels resultats obtinguts ja que, en termes ambientals, això significa que no només la composició (matriu  $\mathbf{Y}^T$ ) de les fonts de contaminació identificades per MCR-ALS amb l'aplicació de trilinearitat serà comuna a totes les matrius analitzades simultàniament sinó que la seva distribució a les

## Capítol 2

---

mostres (distribució geogràfica) també serà la mateixa per a una mateixa font (component) en les diferents condicions.

Els resultats obtinguts mitjançant MCR-ALS amb la restricció de trilinearitat aplicada sobre tots els components del sistema, serien pràcticament els mateixos que s'obtidrien amb l'aplicació d'altres mètodes basats en models trilineals (com per exemple, pel mètode PARAFAC [30]) sobre el mateix sistema. No obstant, el principal avantatge del mètode MCR-ALS és que permet l'aplicació de la restricció de trilinearitat de manera independent i opcional sobre cadascun dels components del conjunt de dades, donant més flexibilitat a l'anàlisi i permetent la utilització de models trilineals totals o parcials, segons sigui convenient. Això fa que aquest mètode es distingeixi de PARAFAC, on tots els components obtinguts han de complir la restricció de trilinearitat. Aquesta flexibilitat permet un modelatge més representatiu d'algunes situacions reals, com per exemple, les d'aquells sistemes (com ocorre en determinats problemes ambientals) en què els perfils d'alguns components es poden comportar de manera trilineal mentre que els perfils d'altres components poden no comportar-s'hi.

L'estructura trilineal d'un conjunt de dades pot ser comprovada aplicant i comparant l'anàlisi SVD de les matrius de dades augmentades en els diferents modes o direccions (augmentació per files i per columnes) i mitjançant la comparació de l'ajust obtingut a partir de la utilització de models bilineals i trilineals, tal com s'explica amb més detall a continuació.

### **Rang i estructura de les dades**

Per un sistema de dades determinat ordenat en una taula o matriu de dades, s'anomena rang (representat per  $N$ ) al nombre de components linealment independents en absència de soroll que són necessaris per tal de representar la variació observada a les dades. Aquesta definició no coincidiria amb la de rang matemàtic d'una matriu de dades, la qual es referiria al nombre de files o columnes linealment independents de la matriu i tindria en compte també les variacions degudes al soroll i error experimental. El rang matemàtic serà igual al nombre de files o columnes (el que sigui inferior dels dos) de la matriu de dades, ja que sempre existirà una component aleatòria associada a l'error experimental. No obstant, en aquest cas no interessarà el rang matemàtic, sinó el nombre de causes de variació sistemàtica no associades amb el soroll ni l'error experimental. Normalment, aquest nombre de components  $N$ , no és conegut d'entrada. Tot i que existeixen diverses aproximacions basades en criteris estadístics per tal de determinar el rang d'un conjunt de dades, a la pràctica s'acostumen a escollir nombres diferents de components per tal de comparar els resultats obtinguts per MCR. En general, l'ajust millora a mesura que s'incrementa el rang, però si s'escull un rang massa elevat, els

resultats obtinguts no són raonables i s'obté un sobreajust de les dades, o en alguns casos fins i tot un empitjorament de les dades (degut a la dificultat de complir les restriccions imposades).

En el cas d'estudis de sistemes de reacció química i d'altres, és freqüent que es presenti la situació de deficiència de rang [54]. Aquest problema es presenta quan en un sistema tancat es detecta un nombre de components linealment independents inferior al nombre vertader de components existents en el sistema, és a dir, que per exemple es detectarien menys fonts de contaminació de les que realment es troben presents en el medi estudiat. Aquest problema és conseqüència de la presència de dependències lineals en els perfils dels components en un dels modes. Existeixen diferents maneres d'eliminar els problemes de deficiència de rang, una de les més importants consisteix en emprar estratègies d'augmentació de matrius de dades com les abans descrites. Quan les diferents matrius simultàniament analitzades aporten informació independent dels diferents components presents, aleshores molt probablement serà possible eliminar aquest tipus de deficiència [54, 58-59].

En el cas de l'anàlisi de matrius augmentades, el càlcul del rang també és útil a l'hora de determinar el model que segueixen les dades (només bilineal o també trilineal). El rang de les matrius augmentades es pot calcular en les seves diferents direccions o modes possibles d'augmentació, per files, per columnes, i en profunditat (en cas de tenir-ne suficients, és a dir, que el nombre de matrius simultàniament analitzades sigui més gran que el nombre de components del sistema), segons es disposin una matriu al costat de l'altra (augmentació o desplegament en la primera direcció), una a sobre de l'altra (augmentació o desplegament en la segona direcció) o es desplegui cada matriu en un vector llarg i es posin una al costat de l'altra (augmentació o desplegament en la tercera direcció). Quan en les tres direccions de desplegament o augmentació s'obté el mateix nombre de components o rang, el conjunt de dades en les tres direccions podrà descomposar-se adequadament mitjançant un model trilineal i, en canvi, quan els seus rangs en les tres direccions de desplegament siguin diferents, el conjunt de dades no s'explicarà bé segons un model trilineal, però molt probablement encara podrà ser descrit correctament amb un model bilineal. Aquest tipus de comprovació és molt senzilla i útil per tal de comprovar l'estructura del conjunt de dades. En el cas que el nombre de matrius analitzades simultàniament sigui petit (igual o inferior al nombre de components del sistema), només caldrà comprovar el rang de les matrius augmentades per files i per columnes (en els dos primers modes).

### PCA respecte a MCR-ALS

Els algoritmes o mètodes de descomposició bilineals utilitzats en MCR i PCA són totalment diferents, de manera que les matrius resultants d'*scores* i *loadings* també ho seran. De fet,

## Capítol 2

---

degut a la presència de les anomenades ambigüitats rotacionals (a més de les d'escala) associades a l'anàlisi de factors [36], existeixen un nombre indeterminat de possibles descomposicions matemàtiques d'una matriu de dades que reproduïxen igualment bé la matriu de dades original. Dit d'una altra manera, mentre que en PCA, les solucions obtingudes són úniques degut a les restriccions d'ortogonalitat, màxima variància i normalització, en el cas de MCR no ho són sempre, ja que les restriccions que s'imposen són d'una altra mena. Les solucions obtingudes per PCA són solucions úniques que tenen unes propietats matemàtiques òptimes (ortogonalitat i màxima variància explicada en ordre decreixent), però generalment, no tenen un significat físic i per tant, no són directament interpretables. No obstant, la informació continguda en les dades continua essent la mateixa que en les solucions obtingudes per MCR, amb la diferència que aquesta informació està expressada de manera diferent. Les solucions buscades per MCR es troben sotmeses a una sèrie de restriccions que els proporcionen significat físic. En canvi, aquestes es trobaran correlacionades entre si (no són linealment independents). A partir del coneixement físic del sistema, es podran restringir les solucions de manera que compleixin un determinat nombre de condicions conegudes. Així es podrà limitar de forma considerable el nombre de solucions possible i obtenir, eventualment, les solucions "vertaderes". Recentment, s'han proposat diversos mètodes alternatius per tal d'avaluar el grau d'ambigüitat associat a unes determinades solucions MCR [43].

Quan s'apliquen restriccions d'ortogonalitat, com en el cas del PCA, les variàncies  $R^2$  pels components individuals d'un model particular són additives i la seva suma és igual a la  $R^2$  total explicada per tot el model. Això és totalment diferent en el cas de models on no s'utilitzen restriccions d'ortogonalitat, com ara MCR-ALS, en els quals la suma de les variàncies  $R^2$  calculades individualment per cada component excedeix la variància explicada total del model, calculada incloent simultàniament tots els components. La raó d'aquesta diferència és que les variàncies explicades dels components individuals no ortogonals se superposa, i inclou les covariàncies existents entre aquests components, les quals són diferents de zero. Això és també una conseqüència de la correlació que pot existir entre els perfils dels components purs reals en els sistemes ambientals i, per tant, de la correlació existent entre els perfils dels components trobats per MCR. En aquest cas, la superposició de perfils porta associada també una superposició de variància explicada, que es pot mesurar, per exemple, a partir de la diferència entre la suma de les variàncies explicades per cada component individualment en el model de la variància explicada i la variància dels components considerats conjuntament en el model.

En aquesta Tesi, l'anàlisi quimiomètrica de les dades per PCA s'ha realitzat utilitzant la PLS Toolbox (Eigenvector Research Ltd., Manson, WA, USA), i per MCR-ALS a partir d'una

## **Metodologies aplicades a l'anàlisi de dades ambientals**

col·lecció de rutines quimiomètriques essencials i avançades [34, 41] que treballen també dins de l'entorn de programació, anàlisi numèrica i visualització de dades del software MATLAB (The Mathworks, MA, USA).





## **Article científic 1**

**Article 1: Quality assessment of the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method for the investigation of environmental pollution patterns**

Marta Terrado, Damià Barceló and Romà Tauler

*Environmental Science and Technology* 43 (2009) 5321-5326



### 2.1.2.6. Resultats de l'avaluació del mètode MCR-ALS per a la investigació de pautes de contaminació en aigües superficials

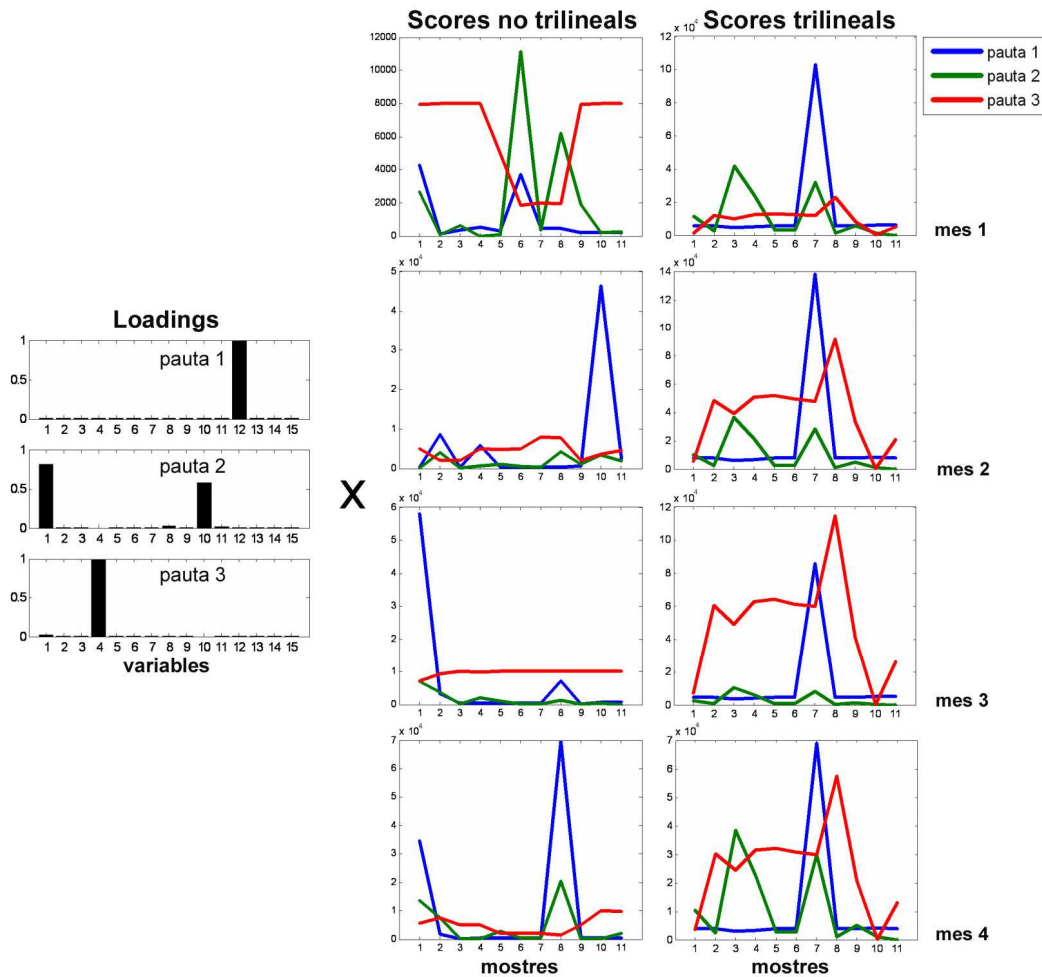
A l'article científic número 1, s'ha presentat l'avaluació de la qualitat dels resultats obtinguts mitjançant el mètode MCR-ALS en el cas de la investigació de fonts o pautes de contaminació en aigües superficials a partir de dades de monitoratge ambiental. A més, diversos conjunts de dades simulades han estat dissenyats per tal de poder realitzar la validació del mètode per a les diverses possibles estructures de dades obtingudes (simplement bilineal o també trilineal) a partir de diferents estudis de seguiment ambiental. A partir de les simulacions utilitzades, també ha estat possible contrastar l'efecte de l'anàlisi directa de les dades originals (sense cap pretractament) respecte a l'anàlisi de dades que han estat prèviament pretractades, per exemple per procediments d'escalat. Finalment, també s'ha estudiat l'efecte de la propagació del soroll sobre les fonts de contaminació resoltes, per tal d'establir diferències entre aquesta propagació i l'efecte de les possibles ambigüitats rotacionals inherents als mètodes de resolució multivariant de corbes.

Per a la preparació de les matrius de dades utilitzades en la present avaluació, s'han pres com a referència les dades de l'estudi experimental sobre la contaminació de productes fitosanitaris als canals del delta del riu Ebre durant la principal estació de creixement de l'arròs (veure article científic número 4). S'han simulat, per tant, 4 matrius de dades individuals (corresponents als 4 mesos de l'aplicació d'aquests productes). Aquestes matrius han estat creades assumint el model bilineal descrit a l'Equació 2.11, descrita a l'apartat anterior. Per a cada matriu individual  $\mathbf{D}_k$ ,  $k=1,2,3,4$ , s'ha suposat una composició determinada de les fonts de contaminació de productes fitosanitaris (files de la matriu  $\mathbf{F}^T$  o matriu de *loadings*) i una distribució d'aquestes fonts a les mostres (columnes de les matrius  $\mathbf{G}_k$  o matriu d'*scores*), diferent per a cada matriu de dades. D'aquesta manera, cada matriu  $\mathbf{D}_k$  expressarà les concentracions dels diferents productes fitosanitaris a cada una de les mostres analitzades durant les 4 campanyes. Tal com s'observa en el gràfic de *loadings* de la Figura 2.11, en aquest cas les pautes de contaminació són bastant ortogonals (no existeix gaire superposició de variables en les 3 fonts de contaminació simulades). En aquesta figura, la contribució de les variables minoritàries als gràfics de *loadings* gairebé no es percep, però aquestes sí que s'hi troben presents, amb contribucions que tot i ser baixes, són diferents de zero. Aquesta és una situació freqüentment trobada a la pràctica, i per aquest motiu ha estat utilitzada en la preparació de les dades d'aquest estudi.

S'han simulat dos tipus diferents de matrius d'*scores*  $\mathbf{G}_k$ , una d'elles seguint una estructura no trilineal i l'altra amb una estructura trilineal. En el cas de la matriu d'*scores* no trilineal de la

Figura 2.11, la distribució de les tres pautes de contaminació té unes característiques diferents per a cada matriu individual. En canvi, quan la matriu d'scores presenta una estructura trilineal, la distribució de les pautes de contaminació a les mostres és invariant en el temps, canviant només en la seva escala relativa però no en la seva forma. Com ja s'ha explicat en la descripció del mètode MCR-ALS, per tal de comprovar si les dades presenten una estructura trilineal, només cal aplicar la descomposició de valors singulars (SVD) a les matrius de dades augmentades per files i per columnes, i comprovar si el nombre de components principals coincideix. Si és així, les dades segueixen un model trilineal.

Del producte de la matriu de *loadings* amb les diferents matrius d'scores, se n'obtenen les diferents matrius que contenen la concentració de productes fitosanitaris a les diferents mostres en quatre condicions de temps diferents (mes 1 a 4), les quals seran posteriorment analitzades mitjançant el mètode MCR-ALS.



**Figura 2.11.** *Loadings* i *scores* (amb estructura no trilineal i estructura trilineal) de les pautes de contaminació simulades durant 4 mesos. Variables: 1, MCPA; 2, mecoprop; 3, 2,4-D; 4, bentazona; 5, simazina; 6, isoproturon; 7, clortoluron; 8, atrazina; 9, diuron; 10, propanil; 11, molinat; 12, alaclor; 13, metolaclor; 14, fenitrotion; 15, diazinon. Mostres: 1, canal d'entrada 1; 2-9, canals d'irrigació 1 al 8; 10, llac de l'Encanyissada; 11, Badia dels Alfacs

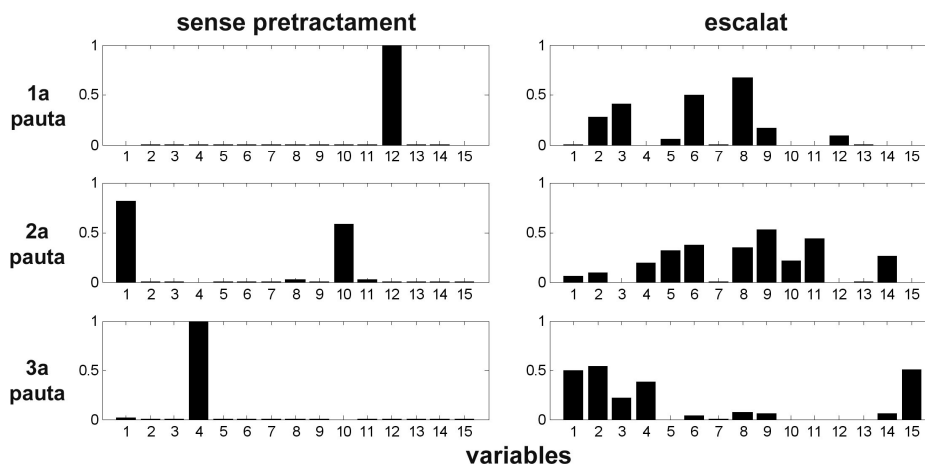
Ambdós tipus de conjunts de dades (trilineal i no trilineal) han estat preparats per duplicat. En un d'ells s'ha addicionat soroll mentre que a l'altre no. Tal com es descriu a l'article científic número 1, la matriu de soroll  $\mathbf{E}_k$ ,  $k=1,2,3,4$ , presenta la contribució de dos tipus diferents d'error per cadascuna de les matrius de dades: un tipus d'error que és constant per cada variable (en aquest cas del 10% de la màxima concentració de la variable corresponent) i un altre que és proporcional al valor mesurat (també del 10%). El valor final de l'error segueix una distribució aleatòria Gaussiana amb un valor mitjà igual a zero i desviació estàndard igual a l'arrel quadrada de la suma dels dos tipus d'error (constant i proporcional) calculats per cada valor particular de la matriu de dades  $\mathbf{D}_k$ . Ja que l'addició de l'error estimat d'aquesta manera pot donar lloc a valors negatius, els quals no tenen sentit per aquest tipus de dades ambientals, en aquest cas els valors negatius han estat substituïts per un valor positiu molt baix i proper a zero. Aquest valor presenta un efecte negligible sobre els resultats finals obtinguts.

Les 4 matrius  $\mathbf{D}_k$  finalment generades (amb estructura trilineal i no trilineal, amb soroll i sense soroll) han estat analitzades mitjançant el mètode MCR-ALS, i els seus resultats han estat posteriorment comparats per tal de veure quin és l'efecte que tenen sobre el resultat final (perfils d'*scores* i *loadings* obtinguts), els diferents tipus de restriccions i de pretractaments aplicats a les dades. Els principals pretractaments investigats han estat: (1) no pretractament, (2) escalat de les dades, (3) escalat de les dades i posterior desescalat dels perfils obtinguts (*loadings*) per als factors. El desescalat dels *loadings* consisteix en aplicar l'operació inversa de l'escalat sobre les dades (és a dir, es multipliquen cadascun dels valors de les variables en els *loadings* per la desviació estàndard prèviament estimada pels valors de la mateixa variable a les dades originals), i s'aplica per tal que els perfils obtinguts quan s'escalen les dades siguin comparables amb els perfils obtinguts quan no s'escalen les dades. L'escalat sobre les variables de les dades originals només hauria d'afectar als *loadings* i no als *scores*. Pel que fa a la comprovació de l'efecte de les restriccions, l'efecte principalment estudiat ha estat el de la restricció de trilinearitat durant l'anàlisi MCR-ALS: (4) aplicació de la restricció de trilinearitat utilitzant dades brutes o sense pretractament i (5) aplicació de la restricció de trilinearitat utilitzant dades escalades. L'efecte dels diferents factors s'ha investigat a partir de la comparació entre els perfils d'*scores* i *loadings* simulats (perfils teòrics) i els obtinguts després de l'aplicació d'MCR-ALS per cadascuna de les opcions 1 a 5 descrites en aquest paràgraf.

### Efecte de l'escalat de les dades

En aquells casos en què les variables presenten diferències en les seves magnituds, és aconsellable realitzar algun tipus d'escalat abans d'aplicar l'anàlisi quimiomètrica. En cas contrari, només aquelles variables que dominen la variància de les dades presentaran una

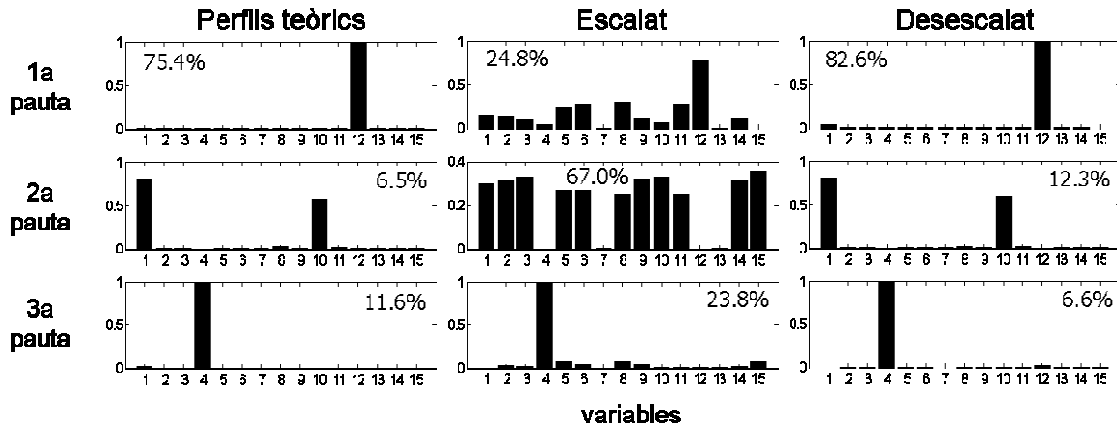
elevada contribució significativa en els perfils de *loadings* obtinguts. En el cas de la Figura 2.12, per exemple, es mostren els resultats obtinguts de l'aplicació de l'anàlisi MCR-ALS sobre la matriu augmentada de dades experimentals (no simulades) amb les concentracions de productes fitosanitaris al delta del riu Ebre durant l'any 2005, en el cas en què les dades no han estat pretractades i en el cas en què les dades han estat escalades. Quan no es realitza cap pretractament a les dades, només les variables alaclor (1a pauta o font de contaminació), MCPA i propanil (2a font de contaminació) i bentazona (3a font de contaminació) presenten una contribució significativa en els perfils de *loadings*, mentre que la resta queden amagades. És doncs en casos com aquest, que l'escalat de les variables és particularment important, ja que facilita la identificació de les contribucions de les diferents variables menors (amb concentracions baixes dels compostos químics) en els perfils de *loadings* i, conseqüentment, contribuirà a millorar la interpretació ambiental de les diferents fonts de contaminació identificades.



**Figura 2.12.** Comparació dels *loadings* de les tres fonts de contaminació obtingudes a partir de l'anàlisi de les dades experimentals de productes fitosanitaris al delta del riu Ebre, dades sense pretractament i dades escalades. Variables: 1, MCPA; 2, mecoprop; 3, 2,4-D; 4, bentazona; 5, simazina; 6, isoproturon; 7, clortoluron; 8, atrazina; 9, diuron; 10, propanil; 11, molinat; 12, alaclor; 13, metolaclor; 14, fenitrotion; 15, diazinon

No obstant això, a l'hora d'escalar les variables, cal anar en compte amb aquelles que presenten valors massa petits i que a més gairebé no varien, ja que en escalar-los, es corre el risc d'augmentar innecessàriament la seva importància relativa. Això té lloc quan s'inclouen mesures de variables amb valors per sota del límit de detecció en gairebé tots els casos. Aquest efecte es presenta en el cas d'algunes de les variables de la segona font de contaminació obtinguda a la Figura 2.12 després de l'aplicació de l'anàlisi quimiomètrica sobre les dades prèviament escalades. A l'apartat 2.1.2.3 d'aquesta Tesi, s'han exposat diferents alternatives que permeten evitar aquest efecte negatiu, per exemple, a partir de la divisió dels valors petits i pràcticament constants d'aquestes variables per un valor numèric relativament gran (com ara 20) que fa que finalment la contribució d'aquestes variables continuï sent petita a

les dades un cop escalades. És important també tenir en compte que l'escalat és una operació matemàtica que, en principi, només afectarà a les variables (ja que només s'aplica sobre les columnes de la matriu de dades) mentre que aquest pretractament no ha de tenir cap efecte sobre les mostres (sobre les files de la matriu de dades). Tot i això, algunes de les diferències observades en la recuperació dels perfils d'*scores* seran degudes als efectes del soroll, inherent a les dades de tipus ambiental. L'escalat també presentarà altres efectes, com per exemple un repartiment més uniforme de la variància de les dades en les fonts de contaminació identificades quan aquestes són prèviament escalades (veure Figura 2.13).



**Figura 2.13.** Comparació dels perfils de *loadings* teòrics, els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades, i els perfils obtinguts desescalant els perfils escalats (dades simulades amb soroll)

A la Figura 2.13 es poden observar els tres casos estudiats. En primer lloc, es mostren els perfils de *loadings* simulats o teòrics (bastant ortogonals i amb una contribució petita de la majoria de les variables), en segon lloc, es mostren els perfils de *loadings* obtinguts aplicant el mètode MCR-ALS sobre la matriu de dades augmentada (pels 4 mesos) prèviament escalada. En aquest cas, la importància d'algunes de les variables que presentaven poca contribució en els perfils de *loadings* teòrics, s'ha vist incrementada per l'escalat. També s'observa clarament l'efecte de la redistribució de la variància, la qual en els perfils teòrics es trobava majoritàriament concentrada en la primera font de contaminació i, en les dades escalades, es troba més repartida entre les tres fonts identificades. Finalment, es mostren els perfils recuperats un cop els perfils de *loadings* obtinguts a partir de les dades escalades han estat posteriorment desescalats. L'operació de desescalat, es mostra en forma matemàtica a les següents equacions:

$$\mathbf{X} = \{x_{i,j}\} \quad \text{Equació 2.25}$$

$$s_j = \sqrt{\frac{\sum_{i=1}^I (x_{i,j} - \bar{x}_j)^2}{I-1}} \quad \text{Equació 2.26}$$



$$sX_{ij} = \frac{x_{ij}}{s_j} \quad \text{Equació 2.27}$$

$$\mathbf{SX} = \{sx_{ij}\} \quad \text{Equació 2.28}$$

$$lx_{n,j} = lSX_{n,j}s_j \quad \text{Equació 2.29}$$

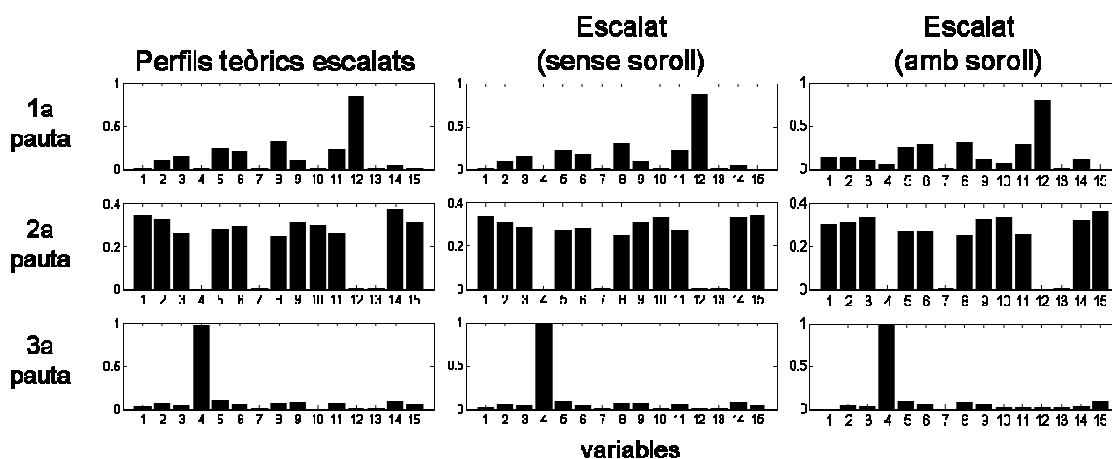
A les Equacions 2.25 a 2.29,  $x_{ij}$  correspon a la mesura individual de la variable  $j$  en una mostra  $i$ . El conjunt de les mesures de les diferents variables a les diferents mostres donen lloc a la matriu de dades  $\mathbf{X}$  (Equació 2.25). El paràmetre  $s_j$  (Equació 2.26) correspon a la desviació estàndard (descrita a l'apartat 2.1.2.3 d'aquesta Tesi) de la variable  $j$  (per a totes les mostres), i  $sx_{ij}$  són els valors de les mesures originals escalades, els quals donen lloc a la matriu  $\mathbf{SX}$  (Equació 2.28). El valor dels *loadings* desescalats per la variable  $j$  en el component  $n$ , ve donat per  $lx_{n,j}$ , un cop els *loadings* obtinguts per les dades escalades,  $lSX_{n,j}$ , s'han multiplicat per  $s_j$ , la desviació estàndard obtinguda a partir dels valors originals de les mesures de la variable (Equació 2.29).

És interessant comprovar que l'operació de l'escalat no comporta cap pèrdua d'informació significativa, ja que quan els perfils de contaminació obtinguts a partir de l'anàlisi de les dades escalades (perfils escalats) es desescalen, s'obtenen altra vegada uns perfils molt similars als teòrics (veure Figura 2.13 i els valors de recuperació dels perfils de *loadings* escalats i desescalats de la Taula 2 en l'article científic número 1). Finalment, com a resultat de l'estudi de les simulacions realitzades mitjançant MCR-ALS, s'observa que sovint l'escalat té tendència a disminuir el sobreajust inherent al model MCR-ALS. A la Taula 2 de l'article científic número 1, es mostra com en la majoria dels casos de dades simulades amb soroll, el sobreajust obtingut analitzant dades sense pretractar (diferència entre la variància explicada,  $R^2$ , amb l'aplicació del model a dades sense pretractar i la  $R^2$  teòrica), és superior al sobreajust que s'obté mitjançant l'anàlisi de dades escalades (diferència entre la  $R^2$  obtinguda amb l'aplicació del model a dades escalades i la  $R^2$  teòrica).

### Efecte del soroll de les dades

Els efectes del soroll existent en els conjunts de dades ambientals poden ser avaluats comparant els resultats obtinguts en l'aplicació del mètode MCR-ALS sobre dades simulades en presència i en absència de soroll. En absència de soroll, les petites diferències detectades en la recuperació dels perfils seran degudes fonamentalment a les ambigüitats rotacionals inherents als mètodes de resolució de corbes. Com ja s'ha explicat anteriorment en l'apartat de descripció del mètode MCR-ALS, aquest tipus d'ambigüitat ocorre quan dos o més components se superposen, i una combinació lineal d'aquests ajusta igualment les dades i compleix

igualment bé les restriccions del sistema, de manera que les solucions no són úniques. A més, en presència de soroll, és més complicat distingir aquells components que expliquen la variància estructural de les dades d'aquells que només expliquen soroll.



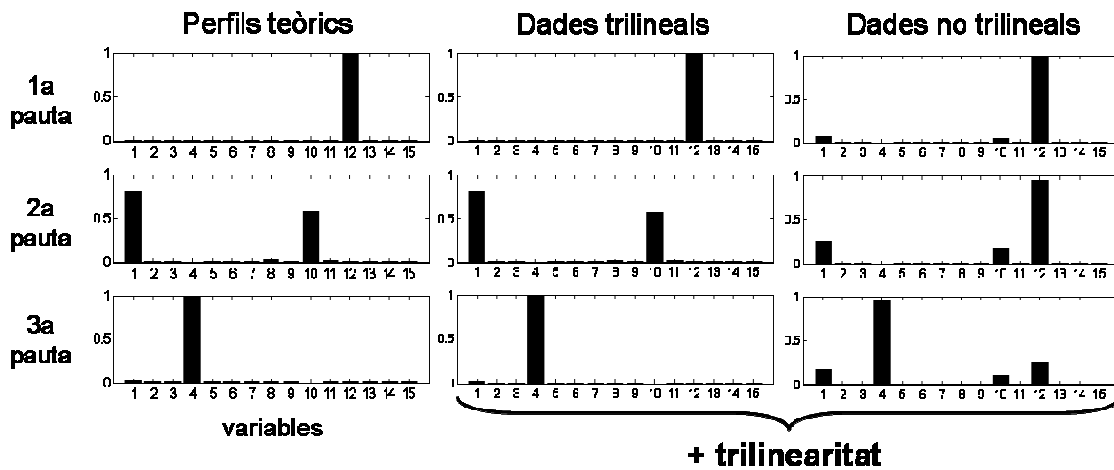
**Figura 2.14.** Comparació dels perfils de *loadings* teòrics escalats i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades amb i sense soroll

A la Figura 2.14, s'observa com en el cas de les dades amb soroll, els resultats obtinguts són pitjors i apareixen més dificultats en la correcta recuperació dels perfils originals o teòrics. El soroll és propagat i amplificat des de les dades experimentals originals fins als resultats de l'MCR-ALS com a conseqüència de les diferents operacions matemàtiques realitzades (començant, per exemple, amb l'escalat de les dades), la qual cosa fa més difícil la correcta identificació especialment d'aquells components més minoritaris, que presenten una relació senyal/soroll més petita. A l'article científic número 1, es dona una taula (Taula 2) amb els valors de la recuperació dels perfils teòrics de *loadings* ( $r^2$ ) un cop les dades amb i sense soroll han estat analitzades aplicant l'escalat. Les recuperacions dels perfils *d'scores* es poden consultar a la informació suplementària d'aquest article. S'observa que les recuperacions sempre són lleugerament més bones en els casos sense soroll per als tres perfils de *loadings*. En aquesta mateixa taula també es pot observar com afecta el soroll en els casos en què s'aplica el desescalat als perfils de *loadings* obtinguts de l'anàlisi de la matriu escalada. Les recuperacions sempre són molt bones en el cas del desescalat dels perfils en absència de soroll ( $r^2 = 1$ ), mentre que per les dades amb soroll, aquest entra a dins dels perfils obtinguts mitjançant l'anàlisi de les dades escalades i els seus efectes es mantenen també durant el desescalat d'aquests perfils de *loadings*.

**Efecte de la restricció de trilinearitat**

L'efecte que té l'aplicació de la restricció de trilinearitat sobre els resultats de l'MCR-ALS s'ha avaluat a partir de la comparació d'aquests resultats quan la restricció s'aplica sobre matrius de dades amb estructura trilineal així com quan s'aplica sobre matrius de dades amb estructura no trilineal.

L'assumpció d'un model trilineal simplifica considerablement la interpretació ambiental de les pautes de contaminació identificades, ja que no només la composició de les fonts de contaminació (*loadings* en el segon mode) és la mateixa per totes les situacions o condicions (per exemple, períodes de temps o compartiments ambientals), sinó que la seva distribució espacial també es manté constant (*loadings* en el primer mode o *scores*), variant només en un factor d'escala (*loadings* en el tercer mode).



**Figura 2.15.** Comparació dels perfils de *loadings* teòrics i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS amb la restricció de trilinearitat per dades trilineals i dades no trilineals (dades simulades amb soroll)

A la Figura 2.15, es mostren, en primer lloc, els perfils de *loadings* teòrics simulats, i a continuació els perfils de *loadings* obtinguts un cop s'ha aplicat l'anàlisi MCR-ALS sobre les dades amb soroll, aplicant la restricció de trilinearitat: (i) per a dades que segueixen un model trilineal i (ii) per a dades que no segueixen un model trilineal (veure els perfils d'*scores* trilineals i no trilineals representats a la Figura 2.11).

Com a resultat de l'aplicació de l'MCR-ALS amb la restricció de trilinearitat sobre el conjunt de dades amb estructura trilineal, els perfils de *loadings* obtinguts coincideixen força bé amb els perfils teòrics (algunes petites diferències que poden existir, seran degudes principalment a l'efecte del soroll). En canvi, quan l'anàlisi MCR-ALS amb la restricció de trilinearitat és aplicat sobre dades que no segueixen el model trilineal, els resultats obtinguts són erronis, ja que les dades es veuen forçades a seguir una estructura que en realitat no posseeixen. En el cas mostrat a la Figura 2.15, per exemple, les tres fonts de contaminació simulades inicialment

(perfils teòrics), no es recuperen per separat sinó de manera barrejada en els tres perfils de *loadings* resultants de l'anàlisi de les dades no trilineals. Per tant, els resultats obtinguts de l'assumpció d'un model trilineal per aquest tipus de dades no trilineals no seran correctes, de manera que les fonts o pautes de contaminació identificades a arrel de l'anàlisi quimiomètrica seran errònies (o no coincidents amb els perfils teòrics). Aquestes recuperacions es troben quantificades a l'article científic número 1, dins de la Taula 2 pel que fa a les recuperacions dels perfils de *loadings* i a la informació suplementària d'aquest mateix article pel que fa a les recuperacions dels perfils d'*scores*.

Com ja s'ha explicat dins de l'apartat de descripció del mètode MCR-ALS, l'ambigüitat rotacional pot ser completament eliminada mitjançant l'aplicació de la restricció de trilinearitat durant l'anàlisi MCR, de manera que les solucions obtingudes en aquests casos siguin úniques. És per això que és recomanable aplicar aquesta restricció en tots aquells casos en què sigui possible i adequat fer-ho. No obstant, en els estudis de resolució amb MCR-ALS, cal tenir en compte que aquesta restricció només té sentit d'ésser imposada quan les dades presenten una estructura trilineal. En aquest últim cas, a més d'eliminar l'ambigüitat rotacional, la resolució de les fonts o pautes de contaminació s'aconsegueix amb un menor sobreajust del model (la diferència entre la variància explicada  $R^2$  mitjançant l'aplicació de l'MCR-ALS i la  $R^2$  teòrica disminueix quan s'aplica la trilinearitat). I aquest possible sobreajust encara disminueix més quan les dades es troben prèviament escalades (veure resultats obtinguts per  $R^2$  a la Taula 2 de l'article científic número 1).

### 2.1.2.7. Conclusions

Les conclusions principals extretes de la secció de mètodes d'anàlisi multivariant, es recullen a continuació:

- Es posa de manifest el potencial dels mètodes quimiomètrics PCA i MCR-ALS per a l'anàlisi de taules de dades ambientals
- Es valida la utilització del mètode MCR-ALS per a l'anàlisi de dades i per a la identificació de les principals fonts o pautes de contaminació que afecten les aigües superficials (i, per extensió, altres compartiments ambientals)

## Capítol 2

---

- Es comprova que l'aplicació de qualsevol restricció durant l'anàlisi de dades multivariant ha d'estar d'acord amb les característiques i l'estructura interna de les dades (sinó, els resultats obtinguts poden ser erronis)
- S'identifica que l'escalat de les dades (i l'autoescalat, en el cas de PCA) és un pretractament adient i recomanable per tal de facilitar la interpretació ambiental dels resultats obtinguts, especialment quan les magnituds o les unitats de les variables difereixen. No obstant, cal anar en compte amb aquelles variables que presenten valors petits que gairebé no varien, ja que es poden veure exagerades per aquesta operació
- S'aconsella l'aplicació de la restricció de trilinearitat per a l'anàlisi de dades en tres direccions o modes sempre que sigui possible, ja que així les solucions obtingudes són úniques (s'elimina l'ambigüitat rotacional) i són més fàcilment interpretables des del punt de vista ambiental. A més, s'observa que l'aplicació de la trilinearitat també permet disminuir el possible sobreajust del model, associat als models bilineals (sobretot quan els dades es troben prèviament escalades)
- Es confirma que per a la identificació d'una determinada pauta o font de contaminació, el senyal o contribució de les variables que descriuen aquesta font ha de ser superior al nivell de soroll contingut a les dades (sinó, el senyal quedarà tapat pel soroll)

### 2.1.3. Mètodes geostatístics

Com ja s'ha comentat, les dades ambientals es troben típicament distribuïdes en l'espai i en el temps. Per tant, el coneixement del valor d'un atribut, com per exemple la qualitat d'un mineral o bé la concentració d'un determinat contaminant, presenta molt poc interès quan no es coneix la seva localització o bé el moment de la seva mesura, i això no és tingut en compte durant el procediment d'anàlisi de les dades. És en aquest sentit que la geostatística [11, 14, 60-61] proporciona un conjunt d'eines estadístiques que permeten incorporar les coordenades espacials i temporals de les observacions durant el processament de les dades. D'aquesta manera, es pot descriure la possible continuïtat espacial de qualsevol fenomen natural. A més d'interpol·lar el valor d'un determinat atribut en aquelles localitzacions on aquest no ha estat o no ha pogut ser mesurat, els mètodes geostatístics també permeten obtenir una modelització de les incerteses que existeixen en la predicció dels valors d'aquest atribut.

La geostatística està basada en el concepte de funcions estocàstiques, en el qual el conjunt de valors mesurats es considera com un conjunt de variables aleatòries espacialment dependents. Una variable aleatòria és una variable que pot prendre una sèrie de valors diferents en funció d'una distribució de probabilitats. En les funcions aleatòries, cada realització de la funció pot ser concebuda com la suma d'una component estructurada i una altra aparentment erràtica o estocàstica. La component estructurada és la que permet assegurar que els valors observats tinguin una variació sistemàtica, és a dir, que si ens trobem, per exemple, en una zona on s'han realitzat diverses mesures per sobre del que és normal, el més probable és que les mesures addicionals també siguin elevades. D'altra banda, la component aleatòria és la que dificulta predir amb exactitud el valor d'aquestes mesures hipotètiques, ja que aquestes presenten fluctuacions impredecibles.

Amb la finalitat de predir el valor de determinats atributs en localitzacions no mesurades, existeixen diversos algoritmes, com és el cas del mètode anomenat *kriging* [62].

### **Kriging**

El mètode *kriging* està basat en la teoria de les variables regionalitzades [63], la qual assumeix que la variació espacial d'una variable presentada a localitzacions específiques és estadísticament homogènia a través d'una superfície definida. Es tracta d'una aproximació geostatística que quantifica l'estructura de correlació espacial entre les diferents localitzacions en funció de la seva distància de separació. Tot i que existeixen diferents tipus de *kriging*, el *kriging* ordinari és el que acostuma a utilitzar-se més, i el que serà descrit en aquest apartat. La predicció del valor d'un atribut en una localització determinada es pot descriure de la següent manera:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad \text{Equació 2.30}$$

On  $\hat{Z}(s_0)$  és el valor de l'atribut a la localització predita  $s_0$ ,  $Z(s_i)$  són els valors de l'atribut a les localitzacions mesurades  $s_i$ ,  $i=1, \dots, n$ , on  $n$  és el nombre de valors per als quals es disposa d'una mesura, i  $\lambda_i$  correspon als pesos que es donen a cada valor mesurat, i que dependran de:

- (i) el semivariograma empíric (explicat a sota)
- (ii) les distàncies a la localització predita
- (iii) les relacions espacials entre els valors mesurats al voltant de la localització que es vol predir

Les equacions utilitzades durant el procediment de predicció de valors en el mètode *kriging* no dependran dels valors mesurats de les variables sinó només de la seva posició i del semivariograma. Els passos que es segueixen per portar a terme la tasca de predicció es descriuen a continuació:

- 1. Estimació del semivariograma empíric.** La funció bàsica que descriu la variabilitat espacial del fenomen d'interès es coneix com el semivariograma empíric (inversa de l'anomenat diagrama de correlació experimental). El semivariograma  $\hat{\gamma}(h)$  s'utilitza per quantificar l'estructura de correlació espacial o covariància, o sigui per a explorar les relacions que existeixen entre les dades. Es tracta doncs d'una mesura de la dissimilaritat mitjana existent entre parells de dades separades per una distància o vector  $h$ , i es calcula de la següent manera:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(s_i + h) - z(s_i)]^2 \quad \text{Equació 2.31}$$

On  $N(h)$  és el nombre de parells de mesures separades per una distància  $h$  que presenten un valor mesurat de  $z(s_i)$ .

- 2. Determinació del model que segueixen les dades.** Tot i que el semivariograma empíric proporciona força informació (discreta) sobre el comportament espacial d'una variable, és convenient ajustar una funció contínua que faciliti la quantificació del grau i l'escala de la variació espacial i que permeti realitzar, posteriorment, estimacions acurades en les localitzacions no mesurades. En geostatística existeixen diversos models d'ajust de la funció, entre els quals els més utilitzats són l'esfèric, l'exponencial, el gaussià i el lineal. L'ajust de la funció permet obtenir una sèrie de paràmetres que són els que seran utilitzats per a realitzar la interpolació geostatística (*kriging*) i que defineixen el grau i l'escala de la variació espacial. Aquests paràmetres són el rang ( $A_0$ ), el *nugget* ( $C_0$ ), el *sill* ( $C_0+C$ ) i la proporció de la variància explicada ( $C/C_0+C$ ), sovint expressada en tant per cent (veure Figura 2.16). El rang ( $A_0$ ) és el valor de  $h$  (distància entre parells de dades) dins del qual té lloc l'estabilització, i per sobre del qual la semivariància ja deixa d'augmentar. El *nugget* ( $C_0$ ) és la variància no explicada pel model, calculada com la intercepció amb l'eix Y quan es fa la representació dels valors del semivariograma  $\hat{\gamma}(h)$  respecte a la separació entre parells de dades. També es

coneix com a variància error. La màxima semivariància identificada entre parells de punts s'anomena *sill* i descriu la variància que existeix entre el conjunt de mostres, que és el resultat de la suma de la variància error ( $C_0$ ) i la variància estructural explicada de les dades ( $C$ ). Per valors elevats d' $h$ , el semivariograma empíric s'estabilitza, indicant que ja no existeix cap correlació entre les localitzacions.

En general, al voltant d'uns 30 punts són necessaris per tal de generar un semivariograma raonable. La part més important del semivariograma és la seva forma al voltant de l'origen, ja que els punts més propers són els que reben més pes en el procés d'interpolació.

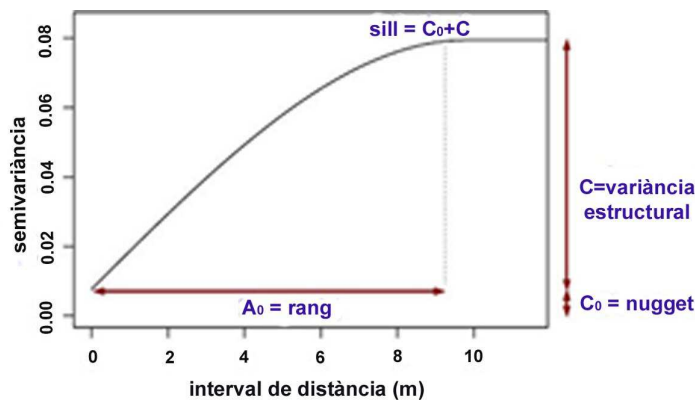


Figura 2.16. Paràmetres utilitzats per a l'ajust del semivariograma empíric a la funció matemàtica més idònia

3. **Creació de les matrius de pesos per al *kriging*.** Les equacions per al *kriging* ordinari es troben contingudes en matrius i vectors que depenen de l'autocorrelació espacial entre els valors a les localitzacions mesurades i predites. Caldrà calcular, en primer lloc, la matriu de covariància **C** entre els diferents parells de localitzacions  $i$ , posteriorment, una matriu **D** amb les distàncies entre les localitzacions mesurades i aquelles que es volen predir. D'aquesta manera, la matriu de pesos  $\lambda$  es calcula mitjançant el producte entre la matriu de distàncies i la matriu de covariàncies transposada:

$$\lambda = C^{-1}D \quad \text{Equació 2.32}$$

4. **Predicció.** A partir dels pesos assignats a les diferents localitzacions mesurades, es calcula una predicció per aquelles localitzacions per les quals no es coneixen els valors. Per al càlcul del valor de l'atribut en un punt en què aquest no ha estat mesurat, es realitza el producte de la matriu de pesos obtinguda i dels diferents valors a les localitzacions mesurades (veure Equació 2.30).



## Capítol 2

---

Abans de generar la superfície final, caldrà tenir una idea de la qualitat de la predicció obtinguda (com de bé el model prediu el valor de l'atribut a les localitzacions no mesurades). El mètode de la validació creuada (*cross-validation leave one out*) utilitza totes les dades menys una per tal d'estimar el model d'autocorrelació que s'utilitza per predir el valor prèviament no considerat. Després torna a posar la dada extreta i en treu una altra de nova per tal de tornar a predir el seu valor. I així successivament per a totes les dades del conjunt. Cada vegada es calcula la diferència entre el valor predit i el valor mesurat, i d'aquesta manera s'obté un resum estadístic amb els errors de la predicció, el qual pot ser utilitzat a mode de diagnòstic.

L'anàlisi SIG així com l'anàlisi geostatística de les dades s'ha dut a terme utilitzant el software ArcGis (ESRI, Redlands, CA, USA), que conté les següents aplicacions: (i) ArcMap, que permet visualitzar, crear i consultar mapes així com editar-ne la seva base de dades, (ii) ArcCatalog, que proporciona accés a les dades i eines de gestió espacial d'aquestes, incloent la visualització i edició de les metadades (dades sobre el mapa com ara la seva font, any, atributs, tipus de coordenades espacials, etc.), i (iii) les aplicacions de geoprocessament (incloent ArcToolbox), que proporcionen un entorn per a la implementació de l'anàlisi SIG, i que poden ser utilitzades individualment o bé combinades dins d'altres models per tal d'efectuar tasques d'anàlisi SIG, incloent la conversió de les dades a altres formats.

Per a realitzar l'anàlisi geostatística s'ha utilitzat l'extensió *Geostatistical Analyst*, integrada dins de l'aplicació ArcMap i concebuda per al modelatge avançat de superfícies mitjançant mètodes deterministes i geostatístics. Aquesta extensió ofereix eines per a realitzar una anàlisi exploratòria de les dades espacials, així com un assistent que guia a través del procés de creació de superfícies vàlides en termes estadístics. El *Geostatistical Analyst* constitueix una mena d'enllaç en la separació que existeix entre el SIG i la geostatística. Tot i que les eines geostatístiques fa temps que es troben disponibles, aquestes no acostumen a trobar-se plenament integrades dins de l'entorn de modelatge SIG. La integració entre les dues metodologies és important per tal que els usuaris dels SIGs siguin capaços de quantificar la qualitat dels seus models superficials, mesurant l'error estadístic associat a les superfícies predites.

L'ajust d'una superfície mitjançant l'eina *Geostatistical Analyst* està descrit per tres passos claus:

- Anàlisi exploratòria de les dades espacials
- Anàlisi estructural (càlcul i modelatge de les propietats de la superfície per localitzacions que es troben properes)
- Predicció de la superfície i avaluació dels resultats

En aquesta Tesi, l'operació *kriging* ha estat utilitzada amb finalitats purament funcionals, per a obtenir representacions gràfiques que facilitin la comprensió dels resultats obtinguts. En cap cas s'ha realitzat una validació del mètode ni tampoc una avaluació exhaustiva de l'error de predicció lligat al procés d'interpolació de les dades, com seria el cas d'altres estudis [64-65]. En tots els casos, la quantitat de punts disponibles ha estat inferior als necessaris per tal de poder obtenir una bona predicció. A més a més, l'adaptació del procés d'interpolació de dades encara no es troba ben desenvolupat dins del medi fluvial (en flux continu). En canvi, si que ha estat aplicat, en alguns casos, per a aigües subterrànies [66].

### 2.2. Mètodes deterministes

Per tal de realitzar la modelització determinista de la dinàmica d'una determinada conca hidrogràfica i del transport de contaminants dins d'aquesta conca, són necessaris diferents passos. En primer lloc, caldrà deduir l'estructura física de la conca, determinant-ne els seus límits, els patrons d'escolament, així com el tipus i ocupacions del sòl que s'hi troben presents. Posteriorment, caldrà simular els diferents processos hidrològics i d'erosió que hi tenen lloc. I, en darrera instància, es podrà passar a la simulació del transport de contaminants dins d'aquesta així com a l'aplicació de les diferents estratègies de gestió per tal de veure l'impacte que aquestes tenen sobre la qualitat de l'aigua de la conca.

Els primers models hidrològics informatitzats daten de fa una quarantena d'anys [66]. Es tractava de models essencialment globals, en els quals el cabal era simulat a nivell de conca, tractant-la com un tot indivisible. Ben aviat, però, comencen a aparèixer models que tenen en compte la variabilitat dels fenòmens hidrològics dins de les mateixes conques hidrogràfiques [67-70]. A més, també es proposen models de tipus físic, com per exemple el model SHE [71-72] orientat a la simulació de l'erosió, que a la pràctica requereixen una quantitat d'informació tan elevada, que de vegades es fa difícil el seu maneig i aplicació pràctica.

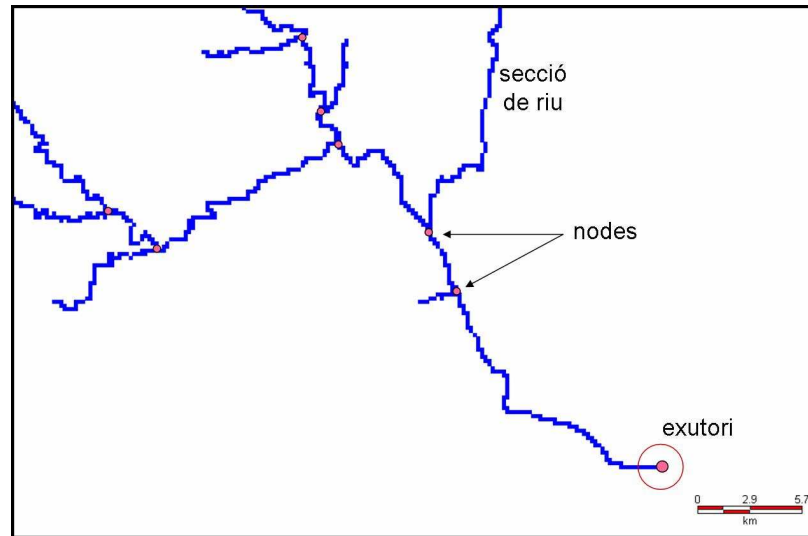
Cap d'aquests models havia estat encara concebut per tal de poder incorporar les dades procedents de la teledetecció o dels sistemes d'informació geogràfica (SIG) més recents. Amb aquesta finalitat, es crea una nova generació de models compatibles amb aquest nou tipus de dades [73]. Recentment, models hidrològics basats en el tractament de models numèrics d'elevacions (MNE) han estat desenvolupats [74], i models com l'anteriorment anomenat SHE s'han beneficiat d'aquestes millores. Paral·lelament, també s'han desenvolupat models compatibles amb la teledetecció i els SIGs [75].

#### 2.2.1. Model PHYSITEL

El model PHYSITEL (INRS-ETE, Québec) permet obtenir l'estructura de drenatge d'una conca hidrogràfica determinada a partir d'una base de dades de la conca incorporada com a informació d'entrada del model hidrològic [76]. PHYSITEL requereix un model digital d'elevacions (MDE), un mapa d'usos del sòl i un mapa de tipus (textura) de sòl. D'aquesta manera es poden determinar les unitats hidrològiques relativament homogènies (UHRH). Cadascuna d'aquestes unitats correspon a una petita subconca de la conca principal a l'escala de la qual es simularan els diferents processos hidrològics.

El MDE proporciona la representació de les elevacions de la zona d'estudi, representades mitjançant cel·les quadrades de dimensions fixades. A partir del MDE es poden obtenir inicialment la pendent i l'orientació, paràmetres que permetran determinar la direcció d'escolament de cada cel·la. Les etapes que el procediment ha de seguir per a la determinació de l'estructura de drenatge d'una conca particular, es descriuen a continuació:

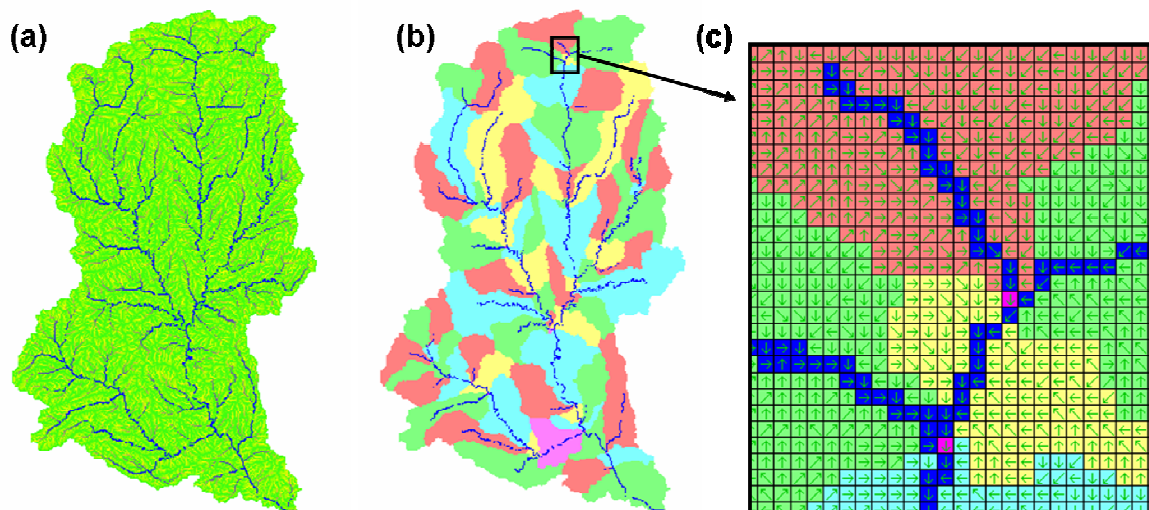
1. Importació i edició del model digital d'elevacions (MDE)
2. Determinació i edició dels pendents (s'entén com a "pendent" la diferència d'elevació entre dues cel·les veïnes en funció de la distància)
3. Importació i edició de la xarxa vectorial imposada (la xarxa hidrogràfica)
4. Determinació i edició de la xarxa matricial imposada (es tracta de convertir la representació vectorial de la xarxa hidrogràfica en una representació en forma de matriu de dades.
5. Determinació i edició del model digital d'elevacions "alterat" (es tracta d'accentuar les elevacions dins i a la perifèria de la xarxa matricial imposada per tal de reforçar el sentit de l'escolament de l'aigua). A la pràctica, el tractament del model d'elevacions pot presentar certes dificultats relacionades amb la precisió del MDE o de la naturalesa del terreny, sobretot quan els pendents són relativament febles. Per aquest motiu, generalment caldrà realitzar un filtratge previ del MDE, eliminar els valors extrems restants i construir un tipus de model d'elevacions "alterat" per tal d'assegurar la continuïtat de l'escolament.
6. Determinació i edició de les orientacions (consisteix a assignar una orientació d'escolament de l'aigua a cada cel·la del model d'elevacions "alterat")
7. Determinació de la conca hidrogràfica (veure més avall). Totes les cel·les que acaben drenant a la cel·la exutori de la conca (veure Figura 2.17) es considera que pertanyen dins de l'àrea de la conca hidrogràfica, mentre que les cel·les restants en són excloses.
8. Determinació de l'estructura d'escolament de la conca hidrogràfica (xarxa fluvial). El valor de cada cel·la es determina en funció del nombre de cel·les l'aigua de les quals acaba drenant a la cel·la en qüestió. Com més aigua va recopilant una cel·la, més elevat és el valor que li és assignat. És normal doncs, que els valors siguin baixos als extrems de la xarxa, que vagin creixent a mesura que s'apropin al centre, i que presentin el seu màxim prop de l'exutori. La xarxa creada dins d'aquesta etapa està constituïda de seccions, les quals tenen nodes com a extremitats (veure Figura 2.17). Un node neix de la intersecció entre dues seccions.



**Figura 2.17.** Exemple de representació d'una xarxa fluvial en format matricial, indicant les seccions de riu, nodes i el node exutori de la conca

9. Reorientació de les cel·les al voltant de les confluències. Es tracta d'una correcció efectuada sobre les cel·les orientades directament cap a un node determinat per tal que aquest no rebi altres aportacions d'aigua que no procedeixin directament de les seccions que el connecten.
10. Determinació de les unitats hidrològiques relativament homogènies (UHRH). Es tracta de petites subconques, cadascuna d'elles associada a una secció de la xarxa hidrogràfica modelitzada sobre la qual seran realitzats els càlculs posteriors (veure Figura 2.18b).
11. Importació i edició de les ocupacions del sòl. Mitjançant aquesta etapa es realitza un inventari de les diferents ocupacions existents sobre cada UHRH. En base a finalitats de modelització hidrològica, les classes en què es reparteix l'ocupació del sòl haurien de ser caracteritzades en funció de l'efecte produït sobre els processos hidrològics. Cadascuna de les classes, està caracteritzada per atributs quantitius (alçada, profunditat de les arrels, índex foliar, etc) que varien en funció del temps i que poden ser utilitzats en els diferents algoritmes del model. El percentatge d'ocupació de cadascuna de les classes sobre cada UHRH s'obté mitjançant la proporció entre el nombre de píxels (cel·les) corresponents a una classe determinada i el nombre total de píxels de la UHRH.
12. Importació i edició dels tipus de sòl, amb la mateixa funció que l'etapa anterior. Aquesta informació serà necessària per definir les característiques hidràuliques de cada tipus de sòl. El model proposa valors per defecte extrets de Rawls i Brakensiek (1989) [77], però l'usuari els pot substituir pels seus propis valors, en el cas que es disposi de valors més apropiats.

13. Exportació del projecte PHYSITEL a format HYDROTEL



**Figura 2.18.** Operacions efectuades mitjançant l'aplicació del software PHYSITEL a la conca del riu Llobregat. (a) determinació de la conca hidrogràfica; (b) determinació de les UHRHs; (c) direccions de l'escolament dins de les UHRH

## 2.2.2. Model HYDROTEL

El model hidrològic HYDROTEL (INRS-ETE, Québec) ha estat utilitzat per a la simulació dels processos hidrològics [78-79]. Especialment, HYDROTEL és un model hidrològic distribuït, de manera que els càlculs efectuats es realitzen sobre un conjunt d'elements de càlcul repartits a l'interior d'una conca hidrogràfica. Aquesta aproximació permet tenir en compte la variabilitat espacial de les característiques fisiogràfiques i meteorològiques que afecten a la conca en qüestió. Generalment, dins d'HYDROTEL, una conca es troba dividida en centenars o milers d'UHRHs, de manera que és impossible atribuir valors individuals de les variables fisiogràfiques o meteorològiques a cadascuna d'elles. Una de les principals raons es deu a que normalment l'usuari no disposa d'informació prou detallada sobre la conca per tal d'atribuir valors satisfactoris amb aquest nivell de precisió. Per aquest motiu, és possible treballar amb agrupacions d'UHRHs sobre la base d'ocupacions o tipus de sòl semblants.

Temporalment, HYDROTEL és un model que funciona en continu (no ha estat dissenyat per a la simulació d'esdeveniments puntuals), que disposa d'un començament i una finalització (data i hora). Per a un determinat període de simulació, el model genera resultats per cada pas de temps. Es defineix com a pas de temps l'interval de temps constant en el qual es subdivideix un període complet de simulació. Aquest ha de ser un múltiple de 24 hores. Les dades d'entrada del model que varien temporalment (dades meteorològiques i hidromètriques) han de ser

## Capítol 2

conegudes a intervals de temps que no sobrepassin la durada del pas de temps de simulació d'HYDROTEL.

Un dels principals objectius per al desenvolupament del model fou la possibilitat de poder-lo aplicar sobre diverses conques hidrogràfiques amb característiques diferents amb un esforç de calibratge mínim.

HYDROTEL ofereix una concepció modular, en la qual la simulació dels diferents processos hidrològics es divideix en 6 submodels (veure Taula 2.1). Per cadascun dels processos hidrològics, HYDROTEL ofereix la possibilitat d'elegir entre un o diversos algorismes. Aquest fet, permet la selecció d'aquell algorisme que millor s'adapta a les característiques de la conca estudiada, tenint en compte la disponibilitat de la informació d'entrada.

Cal dir que el model HYDROTEL va ser inicialment desenvolupat per conques de regions força diferents a les dels rius catalans i que, per tant, caldrà adaptar-lo per tal que pugui ser utilitzat correctament sobre les nostres conques (així, per exemple, l'algorisme que descriu l'evolució de la coberta nival tindrà una importància molt inferior a la que se li atribueix quan aquest és utilitzat sobre regions més fredes).

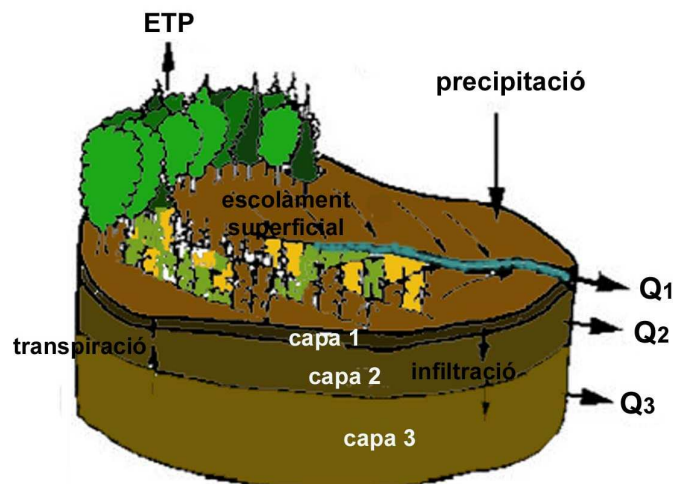
Processos hidrològics	Algorismes disponibles
1. Interpolació de dades meteorològiques	1.1. Polígons de Thiessen 1.2. Mitjana ponderada de les 3 estacions més properes
2. Evolució de la coberta nival	2.1. Mètode mixte dels (graus-dies)- (balanç energètic)
3. Evapotranspiració potencial	3.1. Thornthwaite (1948) 3.2. Hydro-Québec 3.3. Linacre (1977) 3.4. Penman-Monteith (1965) 3.5. Priestley-Taylor (1972)
4. Balanç d'aigua vertical	4.1. Balanç vertical en 3 capes (BV3C) 4.2. CEQUEAU
5. Escolament sobre la superfície de les UHRHs	5.1 Ona cinemàtica
6. Escolament per la xarxa hidrogràfica	6.1. Ona cinemàtica 6.2. Ona difusiva

**Taula 2.1.** Processos i algorismes disponibles per les simulacions hidrològiques amb HYDROTEL

Quan es porta a terme una simulació, és possible la utilització de només alguns dels submodels de la Taula 2.1 (no és necessari utilitzar-los tots). A més, HYDROTEL permet simulacions a passos de temps inferiors a 24 hores, quan les dades necessàries es troben disponibles i quan s'escullen les condicions adequades.

Els 4 primers processos de la Taula 2.1 són simulats a nivell de cada UHRH, o dit d'una altra manera, la distribució espacial dels processos a l'interior d'una conca hidrogràfica es realitza a nivell d'una unitat hidrològica, tal com es representa esquemàticament a la Figura 2.19. I el mateix pel que fa als fluxos. L'escolament sobre la part terrestre de la conca és simulat per tal

d'obtenir un hidrograma geomorfològic específic per cada UHRH (o agrupació d'UHRHs, si és el cas) i es genera per l'escolament cel·la a cel·la d'una làmina d'aigua de referència sobre l'UHRH. El submodel que efectua el càlcul de l'escolament sobre la part terrestre de la conca rep a cada pas de temps i a nivell de l'UHRH, una làmina d'aigua que es distribueix al llarg del temps en funció de l'hidrograma geomorfològic prèviament estimat a partir de l'estructura específica d'escolament intern de l'UHRH. Finalment, l'escolament per la xarxa hidrogràfica es simula a nivell de seccions de riu. Les entrades de cada secció procedeixen de la secció que queda per sobre i de les aportacions laterals de les UHRHs de la vora.



**Figura 2.19.** Diferents processos simulats sobre una UHRH

Per tal de realitzar una simulació mitjançant el software HYDROTEL es necessiten dades d'entrada de tipus (a) hidrològic, (b) meteorològic i (c) fisiogràfic.

### (a) DADES HIDROLÒGIQUES

És necessari disposar de mesures del cabal, com a mínim a nivell diari, a diferents estacions localitzades al llarg de la xarxa hidrogràfica. És imprescindible disposar de mesures del cabal a l'exutori de la conca (o en un punt proper a aquest). Les dades diàries del cabal per cada estació hidrològica s'emmagatzemen en fitxers \*.hyd, els quals presenten el format que es mostra a la Figura 2.20.



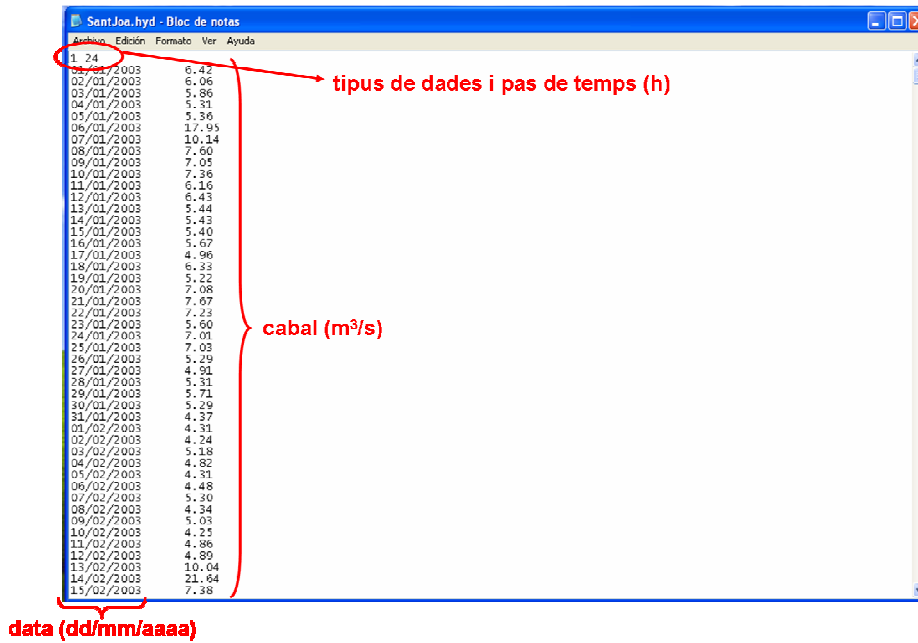


Figura 2.20. Exemple d'un fitxer \*.hyd amb les dades hidrològiques de cabal diari mesurat a l'estació de Sant Joan Despi (Barcelonès)

(b) DADES METEOROLÒGIQUES

Es necessiten dades diàries de temperatura (màxima i mínima) i precipitació. Depenent de l'algoritme que es vulgui utilitzar, caldrà disposar també de paràmetres com la humitat relativa, la radiació solar o la velocitat del vent a una alçada determinada. Les dades meteorològiques s'emmagatzemen en fitxers \*.met (veure format a la Figura 2.21). El valor -9999 correspon a NODATA (no existeix mesura).

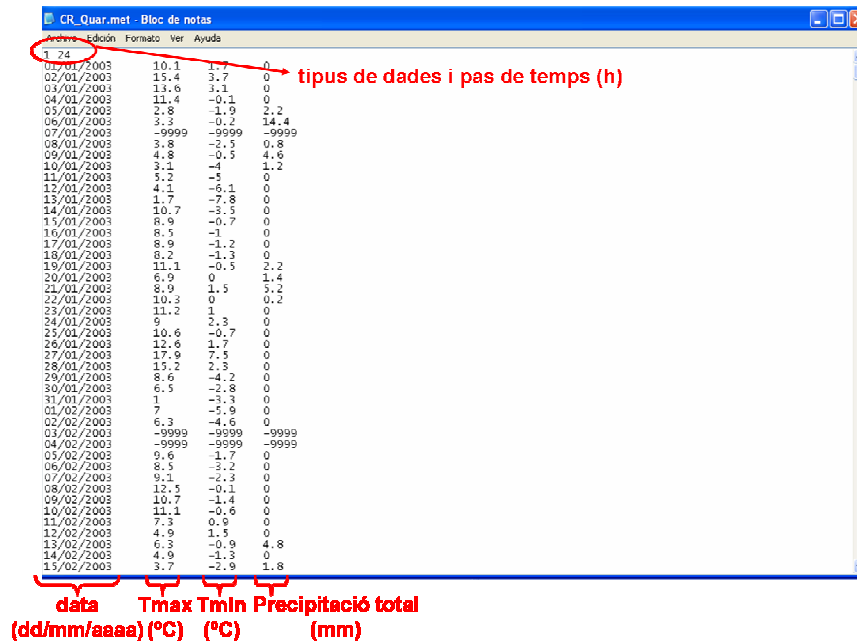


Figura 2.21. Exemple d'un fitxer \*.met amb les dades meteorològiques diàries de l'estació de La Quar (Berguedà) La precipitació es dona en mm, unitat equivalent a litres per metre quadrat

### (c) DADES FISIOLÒGUES

Caldrà que s'afegeixin al model dades sobre el valor de l'índex foliar (índex que representa els  $m^2$  de fulla /  $m^2$  de sòl) i la profunditat de les arrels (m) per les diferents classes d'ocupació del sòl i per diferents períodes de l'any. Aquestes dades s'acostumen a emmagatzemar en fitxers anuals (si es disposa de prou informació), però també existeix la possibilitat d'utilitzar el mateix fitxer per tot el període estudiat (en el cas que aquest correspongui a anys diferents). Les dades fisiològiques s'emmagatzemen en fitxers \*.def.

Per breuetat, a continuació només es descriuran els algorismes seleccionats per al cas d'estudi investigat en aquesta Tesi. La resta d'algorismes es troben àmpliament descrits i discutits a la literatura [78-80].

### 1. Interpolació de dades meteorològiques

La interpolació de dades meteorològiques (precipitació, temperatura de l'aire, humitat relativa, vent, etc.) és el procés mitjançant el qual les dades mesurades a les estacions meteorològiques són estimades sobre cadascuna de les unitats de càlcul.

En el cas dels polígons de Thiessen, s'atribueixen a cada punt de la conca les dades meteorològiques mesurades a l'estació més propera al punt. En l'aplicació presentada en aquesta Tesi, l'algorisme utilitzat ha estat la mitjana ponderada de les tres estacions més properes. Aquest algorisme, atribueix a cada UHRH d'una conca hidrogràfica les dades mesurades a les tres estacions que es troben més properes a aquesta UHRH, amb una ponderació per cadascuna de les estacions en funció de la inversa de la distància entre una estació i la UHRH. La suma dels pesos atribuïts a cada UHRH serà igual a 1.

La interpolació dels valors mesurats a les estacions es realitza amb la finalitat de tenir en compte la variació espacial de les dades meteorològiques. No obstant, aquesta variació no és la única que intervé entre una estació i una UHRH, ja que les variables com ara la precipitació i la temperatura varien també en funció de l'altitud. Aquests paràmetres poden ser corregits a l'hora d'aplicar el model mitjançant la definició de gradients verticals (veure Equacions 2.33 i 2.34 per a la temperatura i la precipitació, respectivament).

$$T_c = T_m + G_T \left( \frac{A_s - A_u}{100} \right) \quad \text{Equació 2.33}$$

On  $T_c$  és la temperatura corregida ( $^{\circ}\text{C}$ ),  $T_m$  la temperatura mesurada ( $^{\circ}\text{C}$ ),  $G_T$  el gradient vertical de temperatura ( $^{\circ}\text{C}/100\text{m}$ ),  $A_s$  l'altitud de l'estació (m) i  $A_u$  l'altitud mitjana de l'UHRH (m).

$$P_c = P_m \left[ 1 + G_p \left( \frac{A_s - A_u}{100} \right) \right] \quad \text{Equació 2.34}$$

On  $P_c$  és la precipitació corregida (mm),  $P_m$  la precipitació mesurada (mm),  $G_p$  el gradient vertical del coeficient multiplicatiu de precipitació (1/100m),  $A_s$  l'altitud de l'estació (m) i  $A_u$  l'altitud mitjana de l'UHRH (m).

Durant l'etapa d'interpolació de les dades meteorològiques, el model també efectua una repartició de la precipitació total en precipitacions líquides i sòlides sobre cada UHRH. Aquesta repartició es fa de la mateixa manera sigui quin sigui l'algoritme d'interpolació aplicat. És important remarcar que les dades poden ser entrades dins del model en forma de precipitació total o bé de pluja i neu per separat (en aquest darrer cas, el model determina la precipitació total). La repartició de la precipitació (líquida i sòlida) sobre cada unitat dependrà bàsicament de les temperatures màxima i mínima de l'aire que han estat interpolades sobre l'UHRH pel pas de temps considerat.

## 2. Evolució de la coberta nival

El model permet simular la variació diària del gruix de la coberta de neu, del seu equivalent en aigua i de la seva densitat a partir de les precipitacions sòlides, i la fusió sobre cadascuna de les unitats hidrològiques de la conca, tenint en compte l'efecte de compressió en l'estimació de l'espessor de la coberta. El dèficit calorífic i el contingut en aigua líquida de la coberta també són estimats mitjançant equacions, de manera que es simula l'evolució de la coberta de neu de manera semblant a com es faria mitjançant una aproximació de tipus energètic. Dins HYDROTEL, en aquests moments existeix un sol algoritme que permet simular l'evolució de la coberta nival, el qual està basat en un mètode mixte (graus-dia)-(balanç energètic). S'anomena grau-dia a la quantitat de calor que correspon a cada augment de la temperatura d'un grau per sobre del llindar considerat, quan el pas de temps és d'un dia.

Per al càlcul de l'algoritme, es consideren tres classes d'ocupació del sòl amb un efecte significatiu sobre la fusió de la neu: els boscos de coníferes, els boscos d'arbres de fulla caduca i les zones obertes. Diferents valors del factor de fusió ( $C_f$ ) són atribuïts a cadascuna de les classes. Aquest factor serà més elevat a mesura que el medi es vagi trobant més desforestat. Per tant, el llindar de temperatura per sota del qual hi haurà fusió disminueix a mesura que el sòl està menys cobert de vegetació (a menys vegetació, fusió més ràpida de la neu). Així, s'efectua una simulació diferent per cadascun dels medis. Les làmines d'aigua de

fusió produïdes per cada medi es ponderen en funció del percentatge de la superfície total de cada unitat hidrològica ocupada pel medi en qüestió per tal de donar com a resultat la làmina d'aigua de fusió mitjana disponible per infiltrar-se dins del sòl.

A més, existeix també la possibilitat de simular la variació de l'albedo de la coberta de neu, o sigui, la relació entre la quantitat de llum reflectida per la coberta i la quantitat de llum rebuda per aquesta (un cos totalment blanc tindrà un albedo igual a 1, ja que reflexa tota la llum rebuda). Això es pot fer utilitzant dos algorismes diferents.

L'algoritme d'evolució i fusió de la coberta nival descriu la simulació de diferents fenòmens, els quals es troben descrits en detall a Fortin i Royer (2004) [80]. L'evolució de la coberta nival és un procés amb una importància considerable en el model HYDROTEL quan aquest és aplicat a zones caracteritzades per la presència de neu bona part de l'any. De fet, el model va ser desenvolupat per la seva aplicació en una zona de clima amb característiques continentals i polars, i és per això que dóna tanta importància a la simulació d'aquest procés. En el cas d'aquesta Tesi, l'evolució i fusió de la coberta de neu té una importància força més limitada. Per aquest motiu, els paràmetres necessaris per al càlcul de l'algoritme han estat definits inicialment i no han estat modificats durant el procés de calibratge del model, ja que s'ha considerat que el seu impacte té una importància més limitada en el cas d'estudi.

### 3. Evapotranspiració potencial

L'avaluació del balanç hidrològic vertical a cada pas de temps requereix l'estimació de l'evapotranspiració real (ETR) de cada tipus de coberta vegetal a la superfície del sòl i sobre sòl nu. Aquestes estimacions depenen del perfil d'humitat del sòl i d'altres paràmetres com la resistència estomàtica, que no acostumen a ser fàcils d'obtenir. L'aproximació habitual en hidrologia consisteix en estimar l'evapotranspiració potencial (ETP).

Existeixen cinc opcions disponibles per l'estimació de l'ETP dins d'HYDROTEL: Thornthwaite (1948), Hydro-Québec, Linacre (1977), Penman-Monteith (1965) i Priestley-Taylor (1972). L'equació que requereix més informació sobre paràmetres meteorològics és la de Penman, mentre que la Hydro-Québec, juntament amb la de Thornthwaite, és la que en requereix menys (temperatures mínima i màxima diàries de l'aire). No obstant, això no significa que aquestes darreres equacions siguin menys precises que la de Penman, sinó que la precisió dependrà de les condicions meteorològiques. En el cas d'estudi inclòs en aquesta Tesi s'ha aplicat l'equació Hydro-Québec. El desavantatge que té és que no distingeix entre classes diferents d'ocupació del sòl, de manera que l'ETP obtinguda és idèntica per a cada classe. Tot i això, ha estat utilitzada en diverses ocasions i s'ha comprovat que els resultats obtinguts no difereixen massa

dels d'altres equacions considerant les diferents ocupacions del sòl. Un factor de multiplicació s'utilitza com a paràmetre de calibratge per tal d'ajustar l'ETP.

Les equacions de Thornthwaite, Hydro-Québec i Linacre no poden ser estimades directament a un pas de temps inferior a 24 hores. Per passos de temps més curts caldrà fer una transformació.

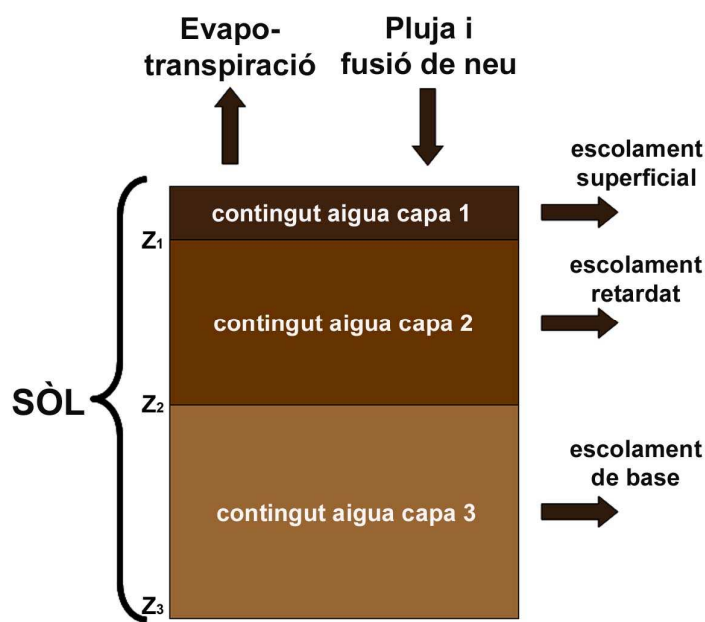
### 4. Balanç d'aigua vertical

El balanç d'aigua vertical es realitza per tal de separar, sobre cada unitat hidrològica, les làmines d'aigua que s'escolen per la superfície del terreny d'aquelles que s'hi infiltren. Un cop s'han infiltrat, aquestes làmines d'aigua modifiquen l'estat de saturació del perfil del sòl i participen en l'escolament retardat de la part superior no saturada del sòl i en els escolaments entre capes existents a la zona saturada.

Per la simulació d'aquest procés, s'ofereixen dos algoritmes diferents: un concebut específicament per als objectius d'HYDROTEL pel que fa a la reproducció dels processos físics i a la compatibilitat amb la teledetecció i els SIG [78], i l'altre extret del model conceptual CEQUEAU [70], el qual pot ser utilitzat si no es disposa de prou informació per al primer algoritme. En aquesta Tesi s'ha utilitzat el primer, l'algoritme del balanç vertical basat en processos físics en tres capes (*Bilan vertical en 3 couches*, BV3C). L'algoritme no dóna una representació exacta del moviment de l'aigua a través del sòl però presenta, en forma de promig, el que passa sobre una conca hidrogràfica.

Per la realització de les simulacions, la columna de sòl es subdivideix en 3 capes o horitzons (veure Figura 2.22). La capa superficial és relativament prima (de 10 a 20 cm) per tal de correspondre al gruix afectat per l'evaporació d'un sòl nu. Aquesta capa controla l'escolament superficial. La segona capa serveix per descriure la zona de transició entre la primera i la tercera capa. També té la funció de regular els escolaments retardats dins de la capa superior del sòl. El cabal base (cabal mínim necessari per al manteniment de les característiques essencials i l'estructura de l'ecosistema aquàtic) està controlat per la tercera capa de sòl. Aquesta capa se sol trobar relativament propera a la saturació.

El balanç vertical d'aigua es calcula sobre cada una de les UHRHs. Per tant, els límits inferior i superior de cada capa seran diferents per cada UHRH o grup d'UHRHs, i les característiques hidràuliques canviaran en funció del tipus de sòl. Finalment, a més dels límits de les capes de sòl i les característiques hidràuliques, el balanç hídic també dependrà del pendent d'escolament que existeixi sobre cada unitat hidrològica.



**Figura 2.22.** Subdivisió del balanç vertical en tres capes dins del submodel BV3C d'HYDROTEL.  $Z_1$ ,  $Z_2$ , i  $Z_3$  corresponen als límits inferiors en metres de les capes de sòl 1, 2 i 3 respectivament (Fortin i Royer, 2004)

A l'inici de cada pas de temps del model, la infiltració procedent de la pluja o bé de la fusió de la neu és controlada per l'aportació en condicions no saturades, o per la conductivitat hidràulica de la capa en condicions de saturació. L'excés d'aigua a la superfície esdevé escolament superficial. Finalment, el contingut en aigua  $\Theta_j$  de cada capa  $j$  varia en funció dels fluxos verticals  $q_{1,2}$  i  $q_{2,3}$  entre les capes, de l'escolament fora de la columna de sòl  $Q_2$  i  $Q_3$ , de la transpiració procedent de cada capa i de l'evaporació del sòl nu per la capa superficial (veure Figura 2.22).

Per tal d'estimar tots els fluxos entre les capes i els escolaments cap a l'exterior de la columna de sòl, s'estima la conductivitat hidràulica  $K$  de cada capa en funció del seu valor a la saturació, dels valors del contingut en aigua i de la distribució de les dimensions dels porus.

L'escolament retardat a la part superior del sòl pot estar lligat a diversos fenòmens. Un model simple basat en l'ona cinemàtica pot ser adequat, tenint en compte l'heterogeneïtat presentada per les conques hidrogràfiques naturals [81]. Segons l'aproximació de l'ona cinemàtica, l'escolament lateral  $Q_2$  de la segona capa només esdevé important si el contingut en aigua d'aquesta capa es troba entre la capacitat de camp i la saturació. Es defineix la capacitat de camp com la retenció màxima d'aigua al sòl, corresponent a la quantitat d'aigua retinguda per un sòl 48 hores després de la seva inundació i del posterior goteig lliure de l'aigua cap a la capa freàtica. L'escolament lateral  $Q_2$  és estimat en funció del pendent  $S_n$  de l'UHRH, del guix ( $Z_2 - Z_1$ ) de la capa i de la seva conductivitat hidràulica  $K(\Theta_2)$ , per l'equació:

$$Q_2 = K(\theta_2) \sin |\arctan(S_n)| (Z_2 - Z_1) \quad \text{Equació 2.35}$$

Finalment, l'escolament  $Q_3$  de la capa 3 és modelitzat per l'Equació 2.36 en funció del contingut en aigua  $\Theta_3$  d'aquesta capa, el seu gruix  $(Z_3 - Z_2)$  i d'un coeficient de recessió CR, el qual pot ser determinat a partir dels cabals observats:

$$Q_3 = CR (Z_3 - Z_2) \theta_3 \quad \text{Equació 2.36}$$

La distribució de l'aigua al sòl i l'escolament superficial dependran en gran mesura de la profunditat de les capes del sòl. Aquestes profunditats, juntament amb el coeficient de recessió, esdevindran paràmetres importants durant el procés de calibratge.

L'evapotranspiració real (ETR) de cada classe de sòl es deriva de l'evapotranspiració potencial (ETP) de la classe. Es calcula un valor de transpiració  $T_r$ , que es reparteix entre les capes en funció de la relació entre el contingut d'aigua de cada capa i el contingut mitjà d'aigua del perfil de sòl, per tal d'obtenir les evapotranspiracions reals corresponents a cada capa. Per la capa superficial, l'evaporació del sòl nu és afegida a la transpiració. Finalment, l'ETR sobre cada UHRH es determina calculant la mitjana ponderada de l'ETR de cada classe d'ocupació del territori en funció del percentatge de superfície atribuïda a cadascuna d'aquestes classes sobre l'UHRH. L'evaporació es considera nul·la en el cas de superfícies impermeables.

### 5. Escolament sobre la superfície de les UHRHs

L'escolament sobre la superfície d'una conca és el procés que descriu el desplaçament de l'aigua fins que aquesta arriba a la xarxa hidrogràfica. Aquest escolament integra (a) els escolaments que es realitzen directament sobre el sòl a través de la vegetació i d'altres obstacles, (b) els escolaments a l'interior de canals, ja siguin naturals o artificials, de dimensions suficients per a ser considerats una part del curs de l'aigua, i (c) els escolaments efectuats a l'interior del sòl. Així, si el submodel BV3C s'utilitzava per al càlcul del balanç vertical, el submodel d'escolament sobre la part terrestre efectuarà el transport fins al curs d'aigua més proper de la làmina d'aigua produïda tant per l'escolament superficial com pels escolaments provinents de les capes 2 i 3.

Per a la simulació d'aquest procés s'utilitza l'algoritme de l'ona cinemàtica. Aquesta formulació fa intervenir, entre d'altres, un paràmetre anomenat coeficient de rugositat de Manning, el qual serà diferent per les tres classes d'ocupació del sòl que l'algoritme és capaç de discriminar

(medi forestal, aigua i altres medis). Aquest coeficient de rugositat estableix la dificultat que troba l'aigua en desplaçar-se per cadascun dels medis.

A cada pas de temps es determina l'escolament cel·la a cel·la sobre cada unitat hidrològica i, per consegüent, sobre tota la conca hidrogràfica considerada. El càlcul de l'escolament es divideix en dues etapes: (1) el càlcul d'un hidrograma geomorfològic de referència per cada UHRH i (2) el càlcul de la quantitat d'aigua que abandona l'UHRH a cada pas de temps.

Les característiques de l'estructura d'escolament a l'interior de cada UHRH es tenen en compte amb l'ajuda d'un hidrograma geomorfològic específic per cada unitat hidrològica [82]. La forma d'aquest hidrograma es determina per cada unitat a l'inici de cada simulació, amb la propagació d'un impuls d'aigua unitari generat uniformement a l'interior de cada UHRH segons el model de l'ona cinemàtica utilitzat en aquest submodel. Posteriorment, les làmines d'aigua produïdes a cada pas de temps sobre cada unitat es distribueixen en el temps segons l'hidrograma geomorfològic definit abans, simulant així l'efecte tampó de l'UHRH sobre els cabals a la sortida. Cada vegada que es modifica la làmina de referència o els coeficients de Manning es realitza un nou càlcul de l'hidrograma. El càlcul del cabal a l'exutori d'una UHRH a cada pas de temps s'efectua fent la combinació lineal dels hidrogrames de referència. La combinació lineal dels diferents hidrogrames geomorfològics dona lloc a la corba de l'escolament acumulat.

### 6. Escolament per la xarxa hidrogràfica

Es tracta del procés mitjançant el qual l'aigua es desplaça a través de la xarxa hidrogràfica fins a l'exutori de la conca. Per a cadascuna de les seccions de riu, es simula l'escolament natural mitjançant aproximacions del sistema d'equacions de Saint-Venant. Existeixen dos algoritmes disponibles dins el model, dels quals en aquesta Tesi s'ha utilitzat el de l'ona cinemàtica.

Normalment, abans d'efectuar una simulació per al període d'estudi amb HYDROTEL, s'acostuma a realitzar una simulació inicial considerant un període de temps més curt, idealment un període immediatament anterior al considerat en l'estudi. Això permet establir els valors per a les anomenades variables d'estat, les quals descriuen l'estat de la conca en termes de quantitat d'aigua en un moment determinat. L'emmagatzemament d'aquestes variables permet l'inici de la nova simulació a partir d'uns determinats estocs d'aigua, situació més realista que el seu inici a partir de valors per al contingut d'aigua iguals a zero.

Durant l'etapa de calibratge del model, els cabals simulats es comparen amb els cabals mesurats a les diferents estacions hidrològiques, de manera que els paràmetres del model poden anar essent ajustats per tal de reproduir aquests cabals mesurats. Tot i que el model és capaç de considerar l'efecte dels embassaments, aquests no han estat considerats en aquesta



Tesi degut a la falta d'informació referent a la seva operació (la simulació dels embassaments és molt exigent en termes de quantitat d'informació).

### 2.2.3. Sistema GIBSI

El sistema integrat de modelització per a la gestió de conques hidrogràfiques GIBSI (*Gestion Intégrée des Bassins versants à l'aide d'un Système Informatisé* (INRS-ETE, Québec)) [83-85], ha estat concebut com una eina per a l'avaluació de l'impacte que els diferents usos de l'aigua i del territori poden tenir sobre la qualitat de l'aigua d'una conca hidrogràfica. Com es mostra a la Figura 2.23, GIBSI es troba constituït d'un sistema gestor de bases de dades, diferents models físics de simulació (hidrologia, erosió del sòl, transport de fertilitzants i productes fitosanitaris, i qualitat de l'aigua), diferents models de gestió (font de contaminació puntual, font de contaminació difosa i gestió d'embassaments) i un sistema d'informació geogràfica. El SIG utilitzat és GRASSLAND [86], un SIG obert construït en base a GRASS [87]. La base de dades inclou informació espacial i alfanumèrica (per exemple, el model digital d'elevacions, dades meteorològiques, característiques del tipus de sòl, indicació de la localització de les diferents estacions hidrològiques, meteorològiques i de qualitat, resultats de les simulacions, dades qualitatives i quantitatives associades a les diferents unitats administratives com són la gestió de cultius o les unitats ramaderes, etc.). Es pot trobar una descripció exhaustiva de la base de dades del sistema a Villeneuve et al. (1998) [85].

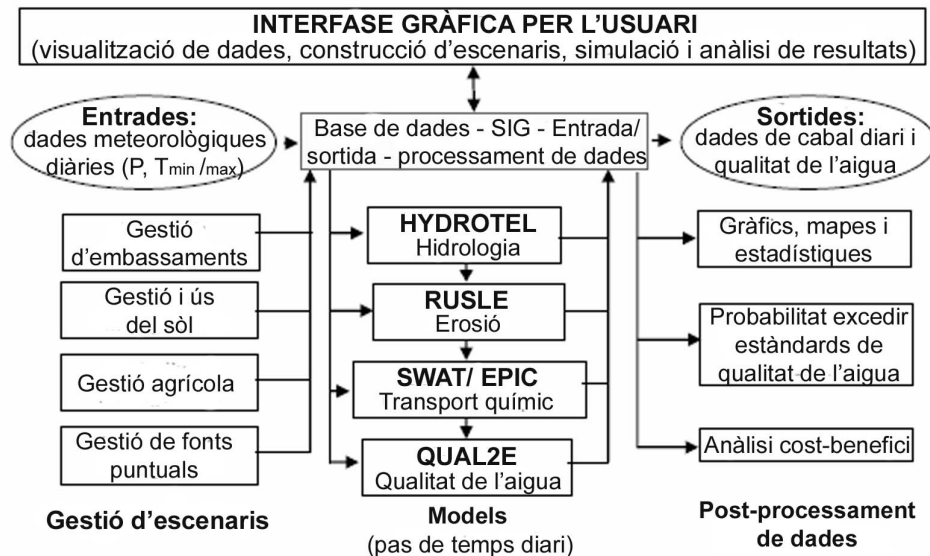


Figura 2.23. Representació dels components i de l'estructura del sistema GIBSI (Quilbé i Rousseau, 2007)

Per tal d'avaluar els impactes causats per projectes municipals, industrials, forestals i agrícoles sobre la qualitat de l'aigua, GIBSI proposa una aproximació basada en la comparació

d'escenaris. Partint d'un escenari de referència (l'estat actual de la conca), noves pràctiques o mesures de control de la contaminació poden ser introduïdes o modificades per tal de construir un nou escenari. L'impacte d'aquest nou escenari sobre la hidrologia i la qualitat de l'aigua pot ser avaluada mitjançant una comparació entre aquest i l'escenari de referència.

### Escenaris

Dins de GIBSI es poden definir quatre tipus d'escenaris diferents:

- embassaments: permet l'addició de nous embassaments a qualsevol de les seccions de riu o l'edició de les seves característiques
- ús del sòl: permet el canvi d'una determinada categoria d'ús del sòl per una altra a qualsevol escala espacial (per exemple, una o diverses UHRH, subconques o bé unitats administratives)
- agrícola: permet l'edició dels tipus de cultiu, les pràctiques de gestió de nutrients, tractaments amb productes fitosanitaris, dates en què tenen lloc les diferents pràctiques agrícoles etc., a qualsevol escala espacial
- plantes de tractament d'aigües residuals: permet l'addició de noves plantes a qualsevol de les seccions de riu o l'edició de les seves característiques (per exemple, el tipus de tractament aplicat, els efluent a la sortida, etc.)

Un cop definits els escenaris, aquests poden ser integrats dins de la base de dades del sistema i posteriorment, realitzar les simulacions.

### Models de simulació

El sistema GIBSI consta de 4 models diferents, els quals són aplicats l'un darrere de l'altre (veure Figura 2.23). No és necessària l'aplicació de tots els models per tal de portar a terme una simulació. No obstant, cal que els models s'apliquin en forma de cascada, de manera que no és possible, per exemple, la simulació del transport de productes fitosanitaris sense haver simulat prèviament l'erosió a la conca hidrogràfica. Els quatre models que s'integren dins del sistema es descriuen a continuació:

#### 1. Model hidrològic

HYDROTEL [80] és el model hidrològic utilitzat per GIBSI. Aquest model ha estat descrit de manera individual en l'apartat 2.2.2 d'aquesta Tesi ja que, en aquest cas, el calibratge hidrològic de la conca ha estat realitzada fora de l'entorn GIBSI, i posteriorment els resultats han estat importats dins del sistema. La unitat de simulació espacial (USS a GIBSI, equivalent a la UHRH d'HYDROTEL) és la unitat elemental de drenatge de la conca (àrea que comprèn

## Capítol 2

tres capes de sòl de diferents profunditats, drenada per un tram o segment de riu). Per a modelitzar el moviment de l'aigua, s'han utilitzat segments unidimensionals de característiques hidràuliques i bioquímiques uniformes.

### 2. Model d'erosió del sòl

Utilitzant les dades calculades pel model hidrològic per una unitat de simulació específica, el model d'erosió, derivat dels algoritmes de l'equació revisada universal de la pèrdua de sòl (*Revised Universal Soil Loss Equation*, RUSLE) [88-89], calcula la proporció de sediment transportada al sistema fluvial corresponent a cadascun dels usos del sòl sobre la USS (veure Figura 2.24). El càlcul de la proporció de sediments transportada s'obté mitjançant la utilització d'un índex d'erosió (relacionat amb la precipitació i l'escolament superficial) i un índex de vegetació (relacionat amb l'ús del sòl i l'estat de creixement i desenvolupament de la coberta vegetal en el moment de la precipitació). El model d'erosió té en compte tres processos físics: i) despreniment de les partícules de sòl, ii) transport de les partícules erosionades i iii) sedimentació de les partícules erosionades. Com a resultat del model s'obtenen la càrrega i la concentració de sediments a cadascun dels segments de riu.

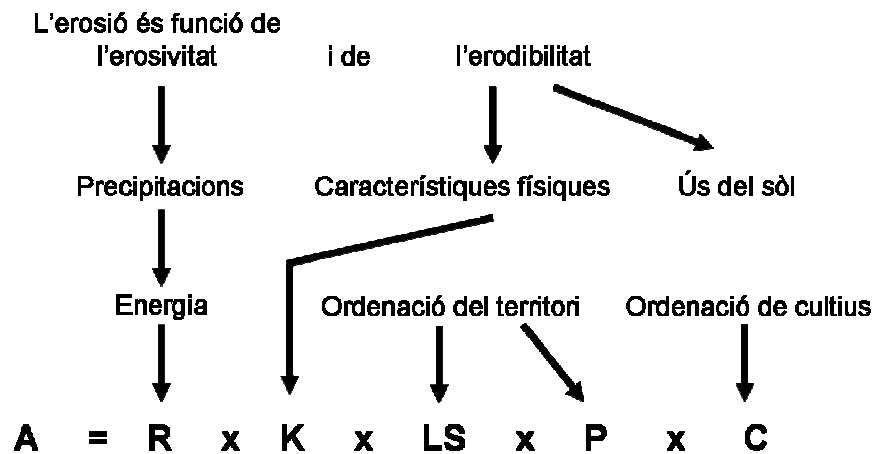


Figura 2.24. Factors que afecten l'erosió del sòl (Villeneuve et al., 1998)

### 3. Model de transport de substàncies agrícoles-químiques

El model es basa en els algoritmes de transport de nitrogen, fòsfor i productes fitosanitaris dels models SWAT (*Soil and Water Assessment Tool*) [90] i EPIC [91], i s'utilitza per calcular les càrregues de contaminants agrícoles exportades cap al flux del riu. Utilitza com a informació d'entrada les dades calculades mitjançant els models hidrològic i d'erosió. El mòdul PO, que inclou els algoritmes de transport per al nitrogen i el fòsfor, no ha estat aplicat en aquesta Tesi, ja que en aquest cas s'ha estudiat exclusivament el transport de productes fitosanitaris, inclòs en el mòdul PS. Una descripció exhaustiva dels algoritmes de simulació del mòdul PO es pot

trobar a Villeneuve et al. (1998) [85]. A continuació es descriuen breument els processos que participen en el model de transport de productes fitosanitaris. Aquests també es troben àmpliament descrits a la referència recentment citada [85].

### **Transport de productes fitosanitaris**

El mòdul PS (derivat del model SWAT) simula l'aplicació, degradació i transport de productes fitosanitaris per l'aigua i les partícules de sòl erosionades sobre cadascuna de les ocupacions agrícoles de sòl (com són, per exemple, cereals, pasturatge, fruiters, etc.) de cada unitat de simulació d'una conca. El mòdul verifica si el dia de simulació correspon a un dia d'aplicació del producte. Si és així, aquest aplica les quantitats de producte fitosanitari establertes per l'usuari en la definició de l'escenari de gestió agrícola sobre cadascuna de les ocupacions del sòl.

El mòdul estima que una fracció del producte fitosanitari passa a l'atmosfera per volatilització en el moment de la seva aplicació, que una altra fracció va a parar sobre les fulles de la vegetació (captura foliar) i que la fracció restant arriba a la superfície del sòl. Aquesta última fracció es reparteix en dues fraccions més; una en forma dissolta dins de la solució del sòl i l'altra adsorbida directament sobre les partícules de sòl. Aquesta repartició s'efectua en funció del valor del coeficient de distribució sòl/aigua o  $K_d$  (adsorció), que és un valor propi a cada producte fitosanitari. Els productes fitosanitaris que es troben a la vegetació i al sòl es degraden o es dissipen mitjançant diferents processos: fotòlisi, oxidació, hidròlisi, volatilització i transformació biològica. Tots aquests processos són simulats globalment dins del mòdul PS en funció del valor del temps de vida mitjana ( $t_{1/2}$ ) de cada producte (o sigui, el temps que necessita una determinada quantitat de producte per arribar a la meitat de la seva quantitat inicial).

Els productes fitosanitaris dissolts dins de la solució del sòl poden ser transportats a través de la xarxa hidrogràfica per escolament superficial o bé per infiltració dins del sòl, mentre que els que es troben adsorbits a les partícules de sòl poden arribar-hi a través de l'erosió del sòl (veure Figura 2.25). És per aquest motiu que els models hidrològic i d'erosió són necessaris per a la simulació del transport de productes fitosanitaris sobre les unitats de simulació. Cal remarcar, però, que el mòdul PS no simula el destí d'aquells productes volatilitzats ni dels absorbits per les plantes. Les dades d'entrada necessàries per a la utilització del mòdul PS són la identificació dels productes fitosanitaris aplicats sobre l'àrea d'estudi, les dates d'aplicació d'aquests i la quantitat en Kg/ha.

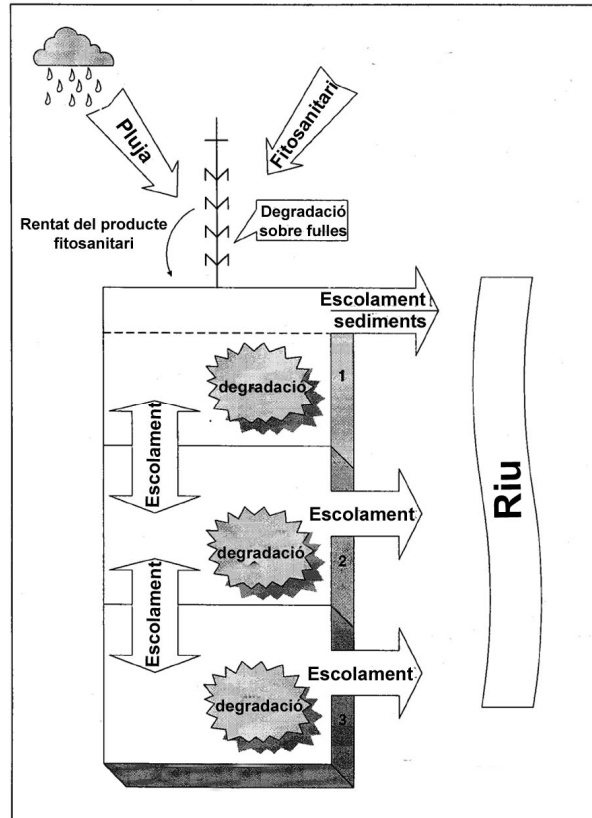


Figura 2.25. Simulació del transport de productes fitosanitaris del sòl al riu (Villeneuve et al., 1998)

#### 4. Model de qualitat de l'aigua

Es basa en el model QUAL2E [92], el qual simula els processos biològics, físics i químics que controlen el destí dels contaminants a les aigües superficials. Aquests són: i) la dispersió i advecció dels contaminants a l'aigua, ii) els processos químics i biològics locals, iii) la temperatura de l'aigua, i iv) l'erosió i el transport de sediments en el flux d'aigua. Els paràmetres de sortida obtinguts en relació a la qualitat de l'aigua són: la quantitat de coliformes, la temperatura de l'aigua i la concentració d'oxigen dissolt, clorofil-la  $\alpha$ , fòsfor orgànic i dissolt, nitrogen orgànic, nitrats, nitrats, amoni i demanda biològica d'oxigen.

Aquest model, però, no ha estat aplicat en aquesta Tesi, ja que tota la informació d'entrada requerida per a la seva aplicació no es trobava disponible per al cas d'estudi de la conca hidrogràfica del riu Llobregat (per aquest motiu, només s'ha arribat fins a la implementació de l'etapa de simulació del transport de substàncies agrícoles).

#### Procediment general d'aplicació del sistema GIBSI

Informació espacial i alfanumèrica de diferent naturalesa s'integra dins del sistema GIBSI. La integració de dades espacials es porta a terme mitjançant el SIG GRASSLAND. Normalment, es poden presentar problemes d'uniformitat de les capes d'informació gràfica, ja que sovint

aquestes són obtingudes a partir de diferents fonts i, per tant, no tenen la mateixa resolució (mida de píxel) ni es troben alineades amb les mateixes coordenades (encaixant perfectament les unes amb les altres). Les capes necessàries per a la utilització del sistema, tant en format ràster com vectorial, són: un model digital d'elevacions, els límits de la conca, els diferents municipis continguts a la conca, els usos del sòl, la pedologia, els pendents, les subconques (delimitades utilitzant HYDROTEL), les USSs (delimitades utilitzant PHYSITEL i HYDROTEL) i les seccions de riu. Normalment, es necessiten els 2 formats (vectorial i ràster) d'una mateixa capa d'informació ja que GRASSLAND necessita el format ràster per a efectuar els diferents càlculs (capa no mostrada dins de GIBSI), mentre que el format vectorial és el que s'utilitza per a la representació gràfica del mapa.

Un cop integrada la informació espacial dins del sistema GIBSI, serà necessari editar les taules amb la informació alfanumèrica relativa a la conca hidrogràfica d'estudi. Les taules que caldrà editar dependran de la finalitat de l'estudi i dels submodels que es vulguin aplicar. A l'Annex III d'aquest treball, es mostra una llista de només aquelles taules que contenen els paràmetres que s'han modificat i adaptat al cas del riu Llobregat. Per breuetat, no s'ha realitzat una descripció exhaustiva de tota la base dades.

Un cop tota la informació ha estat introduïda dins del sistema, l'aplicació del model distribuït GIBSI comprèn les etapes que es llisten a continuació:

- 1) construcció de la base de dades
- 2) aplicació del model hidrològic
- 3) calibratge i validació del model hidrològic a partir de dades de cabals mesurats
- 4) aplicació dels models d'erosió i transport de productes fitosanitaris
- 5) adaptació dels models utilitzats en funció de les condicions de la conca
- 6) calibratge i validació dels models d'erosió i transport de productes fitosanitaris a partir de dades de concentracions i cabals mesurats

Finalment, es realitza una interpretació dels resultats generats pel sistema mitjançant mapes temàtics, gràfics i taules.

# **APLICACIÓ DE DIFERENTS METODOLOGIES D'ANÀLISI I REPRESENTACIÓ DE DADES PER A L'AVALUACIÓ DE LA QUALITAT AMBIENTAL D'UNA CONCA HIDROGRÀFICA**

**Marta Terrado Casanovas  
Barcelona, 2009**



Universitat de Barcelona  
Facultat de Geologia

Departament de Geoquímica, Petrologia i  
Prospecció Geològica



Consell Superior d'Investigacions Científiques (CSIC)  
Institut de Diagnosi Ambiental i Estudis de l'Aigua  
(IDÆA)

Departament de Química Ambiental





---

# **Aplicació de diferents metodologies d'anàlisi i representació de dades per a l'avaluació de la qualitat ambiental d'una conca hidrogràfica**

Memòria realitzada per Marta Terrado Casanovas sota la direcció dels Doctors Romà Tauler Ferré i Damià Barceló Cullerés, i la tutoria del Doctor Manel Viladevall Solé, al Departament de Geoquímica, Petrologia i Prospecció Geològica de la Universitat de Barcelona, per a optar al grau de Doctor en Ciències Geològiques.

Programa de Doctorat: Ciències de la Terra (Bienni 2005-2007)

Barcelona, Setembre de 2009

Doctorand:

**Marta Terrado Casanovas**

Directors de Tesi:

**Dr. Romà Tauler Ferré**

Departament de Química Ambiental de l'Institut de Diagnosi Ambiental i Estudis de l'Aigua (IDAEA – CSIC)

**Dr. Damià Barceló Cullerés**

Departament de Química Ambiental de l'Institut de Diagnosi Ambiental i Estudis de l'Aigua (IDAEA – CSIC)

Tutor:

**Dr. Manel Viladevall Solé**

Departament de Geoquímica, Petrologia i Prospecció Geològica  
Facultat de Geologia, Universitat de Barcelona (UB)

---



# ÍNDEX

1. INTRODUCCIÓ .....	9
1.1. Caracterització de l'àmbit d'una conca hidrogràfica .....	9
1.2. Problemàtiques que afecten la qualitat ambiental d'una conca hidrogràfica ...	13
1.2.1. Legislació europea .....	15
1.2.2. La contaminació de les aigües superficials .....	16
1.2.3. La contaminació de les aigües subterrànies .....	23
1.2.4. La contaminació del sòl.....	24
1.2.5. La contaminació dels sediments .....	27
1.3. Metodologies aplicades a l'anàlisi i avaluació de la qualitat ambiental d'una conca hidrogràfica .....	29
1.3.1. Metodologies empíriques .....	29
1.3.2. Metodologies deterministes .....	35
1.3.3. Els SIG com a eina d'anàlisi ambiental .....	44
1.4. Justificació de la tesi.....	48
1.5. Objectius.....	50
2. METODOLOGIES APLICADES A L'ANÀLISI DE DADES AMBIENTALS.....	53
2.1. Mètodes estadístics i quimiomètrics .....	53
2.1.1. Mètodes univariants d'anàlisi de dades .....	54
2.1.2. Mètodes multivariants d'anàlisi de dades.....	55
2.1.2.1. Estructura de les taules de dades ambientals .....	56
2.1.2.2. Tractament preliminar de les dades .....	58
2.1.2.3. Pretractament de les dades.....	59
2.1.2.4. Anàlisi de components principals (PCA).....	64
2.1.2.5. Resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS) .....	67
Article 1: Quality assessment of the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method for the investigation of environmental pollution patterns.....	89
2.1.2.6. Resultats de l'avaluació del mètode MCR-ALS per a la investigació de pautes de contaminació en aigües superficials .....	91
2.1.2.7. Conclusions .....	99
2.1.3. Mètodes geostatístics .....	100
2.2. Mètodes deterministes.....	106
2.2.1. Model PHYSITEL .....	106
2.2.2. Model HYDROTEL.....	109
2.2.3. Sistema GIBSI .....	120

3. APLICACIÓ DE METODOLOGIES D'ANÀLISI DE DADES AMBIENTALS A L'ESTUDI DE DIFERENTS CONQUES HIDROGRÀFIQUES.....	127
3.1. Introducció .....	127
3.2. Conca hidrogràfica del riu Ebre .....	128
3.2.1. Resultats de l'aplicació de mètodes quimiomètrics a l'estudi de la contaminació a la conca del riu Ebre .....	133
Article 2: Identification and distribution of contamination sources in the Ebro River basin by chemometric modelling coupled to geographical information systems .....	141
Article 3: MCR-ALS modelling of organic pollution patterns in the Ebro River surface water-groundwater-sediment-soil system .....	143
Anàlisi de fàrmacs en aigües superficials.....	145
3.3. Delta del Riu Ebre .....	155
3.3.1. Resultats de l'aplicació de mètodes quimiomètrics i geostatístics a l'estudi de la contaminació al delta del Riu Ebre.....	158
Article 4: Use of chemometric and geostatistical methods to evaluate pesticide pollution in the irrigation and drainage channels of the Ebro River delta during the rice-growing season.....	165
3.4. Conca hidrogràfica del Riu Llobregat .....	166
3.4.1. Resultats de l'aplicació de mètodes quimiomètrics i deterministes a l'estudi de la contaminació a la conca del riu Llobregat.....	169
Article 5: Distribution and assessment of surface water contamination by application of chemometric and deterministic models .....	188
3.5. Conclusions .....	189
4. UTILITZACIÓ D'ÍNDEXS PER A L'AVALUACIÓ DE LA QUALITAT DE L'AIGUA EN UNA CONCA HIDROGRÀFICA.....	191
4.1. Introducció .....	191
4.1.1. Preselecció d'índexs per a dades de xarxes de control automàtiques....	193
4.2. Índex CMME WQI de Canadà .....	197
Article 6: Surface water quality indices for the analysis of data generated by automated sampling networks.....	199
4.2.1. Resultats de l'aplicació del CCME WQI a l'anàlisi de dades generades per xarxes de control automàtiques .....	200
4.3. Índex prioritari de risc de contaminació (basat en el procediment COMMPS) a les conques hidrogràfiques de Catalunya.....	203
4.4. Conclusions .....	214
5. CONCLUSIONS GENERALS.....	216

Referències.....	220
Annex I .....	228
Annex II .....	230
Annex III .....	232
Annex IV.....	240

# ÍNDIX DE TAULES

Taula 1.1. Models empírics de tractament de dades ambientals .....	31
Taula 1.2. Models deterministes de qualitat a escala de conca hidrogràfica.....	38
Taula 1.3. Models deterministes de qualitat per aigües subterrànies .....	41
Taula 1.4. Models deterministes de qualitat per aigües superficials .....	41
Taula 1.5. Models deterministes d'escolament urbà.....	43
Taula 2.1. Processos i algorismes disponibles per les simulacions hidrològiques amb HYDROTEL .....	110
Taula 3.1. Característiques principals de les àrees estudiades en aquesta Tesi.....	127
Taula 3.2. Resum de les pressions i els impactes ambientals existents sobre el territori de la conca del riu Ebre .....	131
Taula 3.3. Identificació dels punts de presa de mostra de la RCSP .....	134
Taula 3.4. Identificació dels punts de presa de mostra d'aigües superficials i sediments .....	137
Taula 3.5. Identificació dels punts de presa de mostra d'aigües subterrànies i sòls .....	138
Taula 3.6. Identificació dels fàrmacs analitzats en aigües superficials de la conca del riu Ebre i grup terapèutic al qual pertanyen .....	148
Taula 3.7. Identificació de les mostres recollides a la conca del Riu Ebre per a l'anàlisi de fàrmacs classificades en diferents tipus .....	148
Taula 3.8. Identificació de les estacions de presa de mostra al delta de l'Ebre .....	159
Taula 3.9. Identificació de les variables analitzades a les mostres del delta de l'Ebre.....	159
Taula 3.10. Identificació, localització i altitud de les estacions meteorològiques utilitzades en el model HYDROTEL per a la simulació de la dinàmica fluvial a la conca del riu Llobregat....	175
Taula 3.11. Identificació, localització i altitud de les estacions hidromètriques utilitzades en el model HYDROTEL per a la simulació de la dinàmica fluvial a la conca del riu Llobregat .....	176
Taula 3.12. Etapes del procés de calibratge proposades per Turcotte et al. 2003.....	179
Taula 3.13: Paràmetres utilitzats durant el procés de simulació hidrològica amb HYDROTEL la conca del riu Llobregat .....	182
Taula 3.14. Cabals observats i simulats a les diferents estacions hidromètriques mitjançant el model HYDROTEL i error del model.....	184
Taula 3.15. Valors de sorpció, solubilitat i vida mitjana dels productes fitosanitaris simulats amb el model PS.....	186
Taula 4.1. Factors de control utilitzats per al càlcul de PNECs.....	207
Taula 4.2. Puntuació assignada als efectes aquàtics indirectes .....	207
Taula 4.3. Frases R i puntuació assignada .....	208
Taula 4.4. Llista de compostos prioritari classificats a les conques catalanes .....	209

# ÍNDIX DE FIGURES

Figura 1.1. El cicle de l'aigua.....	9
Figura 1.2. Activitats de les conques hidrogràfiques que poden afectar la qualitat de les aigües superficials.....	14
Figura 1.3. Cicle d'un contaminant a la hidrosfera.....	16
Figura 1.4. Superposició de capes de diferent geometria en un SIG.....	45
Figura 1.5. Representació ràster i vectorial de la realitat.....	46
Figura 1.6. Unió de dades $\alpha$ -numèriques i cartogràfiques.....	46
Figura 2.1. Esquema d'un diagrama de caixa (a) i mapa de correlació de variables (b).....	54
Figura 2.2. Tipus d'estructures i models de dades ambientals.....	57
Figura 2.3. Diagrames de caixa d'un conjunt de dades ambientals.....	61
Figura 2.4. Nous eixos ortogonals creats a partir del conjunt de dades originals.....	64
Figura 2.5. Descomposició per PCA (model bilineal) de la matriu de dades D per un nombre de components $N = 3$ .....	65
Figura 2.6. Descomposició de valors propis d'una matriu de dades per a la determinació del nombre de components.....	66
Figura 2.7. Diagrama d'implementació del mètode MCR-ALS amb l'exemple d'aplicació de tres restriccions diferents: no negativitat, normalització dels <i>loadings</i> i trilinearitat.....	70
Figura 2.8. Diferents estructures i ordenacions de dades.....	78
Figura 2.9. Descomposició d'un conjunt de dades en tres direccions seguint un model trilineal.....	81
Figura 2.10. Descomposició de la matriu de dades augmentada $D_{aug}$ per MCR-ALS aplicant la restricció de trilinearitat.....	82
Figura 2.11. <i>Loadings</i> i <i>scores</i> (amb estructura no trilineal i estructura trilineal) de les pautes de contaminació simulades durant 4 mesos.....	92
Figura 2.12. Comparació dels <i>loadings</i> de les tres fonts de contaminació obtingudes a partir de l'anàlisi de les dades.....	94
Figura 2.13. Comparació dels perfils de <i>loadings</i> teòrics, els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades, i els perfils obtinguts desescalant els perfils escalats (dades simulades amb soroll).....	95
Figura 2.14. Comparació dels perfils de <i>loadings</i> teòrics escalats i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades amb i sense soroll.....	97
Figura 2.15. Comparació dels perfils de <i>loadings</i> teòrics i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS amb la restricció de trilinearitat per dades trilineals i dades no trilineals (dades simulades amb soroll).....	98
Figura 2.16. Paràmetres utilitzats per a l'ajust del semivariograma empíric a la funció matemàtica més idònia.....	103
Figura 2.17. Exemple de representació d'una xarxa fluvial en format matricial, indicant les seccions de riu, nodes i el node exutori de la conca.....	108
Figura 2.18. Operacions efectuades mitjançant l'aplicació del software PHYSITEL a la conca del riu Llobregat.....	109
Figura 2.19. Diferents processos simulats sobre una UHRH.....	111
Figura 2.20. Exemple d'un fitxer *.hyd amb les dades hidrològiques de cabal diari mesurat a l'estació de Sant Joan Despí (Barcelonès).....	112

Figura 2.21. Exemple d'un fitxer *.met amb les dades meteorològiques diàries de l'estació de La Quar (Berguedà) .....	112
Figura 2.22. Subdivisió del balanç vertical en tres capes dins del submodel BV3C d'HYDROTEL.....	117
Figura 2.23. Representació dels components i de l'estructura del sistema GIBSI .....	120
Figura 2.24. Factors que afecten l'erosió del sòl .....	122
Figura 2.25. Simulació del transport de productes fitosanitaris del sòl al riu.....	124
Figura 3.1. Mapa de situació de la conca del riu Ebre i comunitats autònomes dins de l'àmbit de la conca .....	129
Figura 3.2. Mapa amb la identificació del riu Ebre i els principals afluents que aporten una major quantitat d'aigua.....	129
Figura 3.3. Mapa de localització de les principals pràctiques agrícoles i dels centres industrials i urbans més importants a la conca hidrogràfica del riu Ebre.....	132
Figura 3.4. Localització dels punts de presa de mostra de la RCSP a la conca de l'Ebre.....	134
Figura 3.5. Localització dels punts de presa de mostra d'aigües superficials i sediments a la conca del riu Ebre .....	136
Figura 3.6. Localització dels punts de presa de mostra d'aigües subterrànies i sòls a la conca del riu Ebre .....	137
Figura 3.7. <i>Loadings</i> dels components principals seleccionats amb la contribució dels diferents grups de fàrmacs investigats a les aigües de riu de la conca de l'Ebre.....	149
Figura 3.8. <i>Scores</i> descrivint les diferents contribucions de les mostres d'aigua superficial de la conca de l'Ebre durant el període 2006-08.....	150
Figura 3.9. Representació dels <i>scores</i> de PC2 respecte a PC3 de les mostres de fàrmacs agrupats per grups terapèutics en aigües superficials de la conca del riu Ebre durant el període 2006-08.....	151
Figura 3.10. Representació dels <i>scores</i> de PC1 respecte a PC2 de les mostres de fàrmacs en diferents tipus d'aigües de riu i de depuradora analitzades a la conca del riu Ebre durant el període 2006-08.....	152
Figura 3.11. <i>Loadings</i> i <i>scores</i> de PC1, descrivint la composició i la contribució de la primera pauta de contaminació d'antibiòtics identificada a les mostres d'aigua superficial de la conca de l'Ebre durant el període 2006-08 .....	154
Figura 3.12. Mapa de situació del delta de l'Ebre amb els principals nuclis urbans i zones humides d'interès natural .....	156
Figura 3.13. Situació de les estacions de presa de mostra en els canals d'irrigació i drenatge del delta del riu Ebre .....	159
Figura 3.14. <i>Loadings</i> (gràfics de l'esquerra) i <i>scores</i> (gràfics de la dreta) identificats pel primer component principal (PC) mitjançant l'aplicació de PCA .....	161
Figura 3.15. Comparació de la composició de les principals pautes de contaminació de productes fitosanitaris identificades als canals del delta de l'Ebre mitjançant l'aplicació de PCA i MCR-ALS.....	162
Figura 3.16. Situació de la conca del riu Llobregat i els seus embassaments .....	167
Figura 3.17. Distribució dels diferents usos agrícoles, industrials i urbans sobre el territori de la conca del riu Llobregat .....	169
Figura 3.18. Exemple de fitxer de tipus de sòl a la conca del riu Llobregat en format ascii.....	172
Figura 3.19. Informació d'entrada per al model PHYSITEL .....	173
Figura 3.20. Representació de les UHRHs calculades dins de PHYSITEL.....	174



Figura 3.21. Localització de les estacions hidrològiques i meteorològiques sobre el territori de la conca hidrogràfica del riu Llobregat.....	176
Figura 3.22. Exemple de fitxer d'índexs foliars utilitzats en la simulació hidrològica a la conca del riu Llobregat .....	177
Figura 3.23. Divisió de la conca del riu Llobregat en 6 subconques (grups d'UHRHs) per al procediment de calibratge HYDROTEL.....	178
Figura 4.1. Exemple de càlcul de l'índex diari de qualitat d'una massa d'aigua mitjançant la funció wqindex en MATLAB .....	201
Figura 4.2. Representació gràfica de l'índex prioritari i els índexs d'exposició i d'efecte obtinguts per 50 compostos orgànics a les conques hidrogràfiques catalanes.....	211
Figura 4.3. Representació gràfica de l'índex prioritari de contaminació a les localitzacions pel territori de Catalunya.....	213

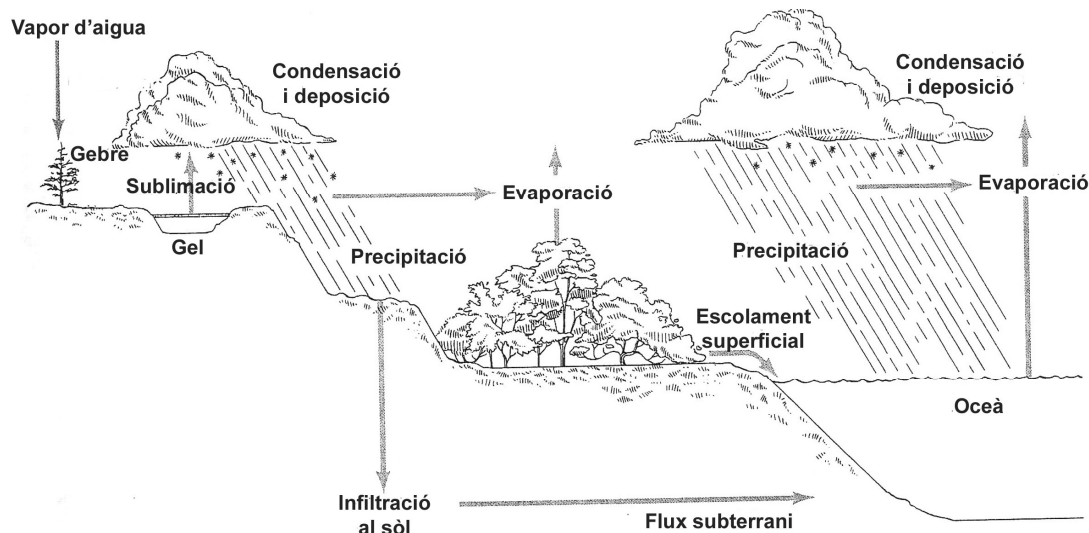


# 1. INTRODUCCIÓ

## 1.1. Caracterització de l'àmbit d'una conca hidrogràfica

Des de sempre, l'aigua ha estat sinònim de vida. Els rius són les venes de la terra; així que la xarxa fluvial constitueix la base de la vida continental de la mateixa manera que la xarxa limfàtica constitueix la base de la vida dels humans.

L'aigua al món es troba distribuïda entre els compartiments oceànic, terrestre i atmosfèric (Figura 1.1). La major part, un 97,2%, forma part dels oceans, mentre que només un 0,0001% es troba present en els corrents fluvials (*U.S. Geological Survey*). El cicle de l'aigua és un procés imprescindible d'intercanvi entre els diferents compartiments de la terra. Es tracta d'un sistema mundial enorme impulsat per l'energia solar, en el qual l'atmosfera proporciona un nexa vital entre els oceans i els continents. No obstant, al mateix temps que l'aigua és transferida d'un compartiment a l'altre, el mateix passa amb les substàncies contaminants que aquesta pugui contenir.



**Figura 1.1.** El cicle de l'aigua (Moran et al., 1986)

Aquella part de l'aigua que no s'infiltra dins de la terra o que no és evaporada, esdevé escolament superficial. Aquest escolament, es desplaça en forma de rius i torrents des de la terra fins al mar. De la mateixa manera, l'aigua subterrània, les fonts i la precipitació caiguda directament sobre els cursos fluvials, també contribueixen al seu cabal. L'àrea de territori que aboca l'aigua, sediments i substàncies dissoltes en un punt comú s'anomena conca hidrogràfica. La regió geogràfica d'una conca hidrogràfica es troba definida pel riu principal i pels afluents que desemboquen en aquest. El clima, la vegetació i la topografia d'una conca,

## Capítol 1

---

juntament amb les alteracions causades pels processos naturals i per les activitats humanes, tindran una afectació tant sobre la quantitat com sobre la qualitat de l'aigua de la conca.

En general, els rius han estat tipificats morfològicament en sentit longitudinal, dividint-los en el que s'anomena curs alt, curs mitjà i curs baix, o el que és el mateix, zona de capçalera, zona de transferència i zona de deposició. No obstant, avui en dia també es considera la seva morfologia des del punt de vista transversal, on en la majoria de casos es troben tres elements principals: el llit fluvial, per on discorre l'aigua al llarg de tot o part de l'any; la plana al·luvial o d'inundació, que en períodes de crescuda del riu i, amb freqüència irregular, pot quedar coberta per les aigües fluvials; i finalment es parla de la zona de transició, que es troba tocant a la plana d'inundació i que constitueix una connexió entre el riu i el paisatge o zona d'influència que l'envolta.

En el curs alt d'un sistema fluvial s'hi troba el naixement o zona fontinal, que acostuma a situar-se en àrees de muntanya o al marge de les valls. Quan l'aigua es mou pendent avall, la seva energia potencial es converteix en energia cinètica, i és una porció d'aquesta energia la responsable dels processos d'erosió. L'energia cinètica addicional és utilitzada per al transport de materials sòlids així com altres substàncies que entren en el riu, com poden ser residus industrials i domèstics. La zona del curs alt acostuma a estar formada per un llit de còdols o graves gruixudes i tendeix a tenir una gran concentració d'oxigen i una temperatura baixa de l'aigua. Es tracta d'aigua poc mineralitzada ja que majoritàriament rep aportacions de la precipitació atmosfèrica i de la fusió dels gels d'alta muntanya.

En el curs mitjà del riu, l'erosió disminueix a causa de la menor pendent. El llit sol estar format per graves fines o per sorres més o menys gruixudes, i comencen a assentar-s'hi plantes aquàtiques anomenades macròfits com a conseqüència de la reducció del transport de materials. En aquesta zona, la turbulència de les aigües fa que encara existeixi una incorporació força elevada d'oxigen atmosfèric. Aquí l'aigua conté una major concentració d'espècies químiques, fruit dels processos d'erosió i dissolució que tenen lloc al llarg del seu recorregut pel llit del riu. També es reben aportacions d'aigües subterrànies i aportacions antropogèniques, la qual cosa incrementa encara més la seva mineralització.

Al curs baix del riu es donen els fenòmens de deposició de sediments fins i gruixuts que han anat erosionant-se als cursos alt i mitjà, donant lloc a les grans planes al·luvials o extensions de la conca hidrogràfica inundades de manera periòdica a causa de les crescudes del riu. Al llarg del recorregut de la zona de deposició, la temperatura de l'aigua tendeix a augmentar. Aquest fet, possibilitarà l'existència d'una major concentració de nutrients la qual, al seu torn, fomentarà el desenvolupament d'algues. En aquells casos en què el creixement d'algues sigui excessiu, es poden originar fenòmens de dèficit d'oxigen.

Depenent de la dinàmica del riu i del material geològic per on discorre el seu llit, la desembocadura al mar pot donar lloc a un estuari, influït per les marees, o a un delta, formació

sovint triangular i formada per al·luvions del riu. A Catalunya hi trobem dos grans exponents d'aquesta darrera formació de la desembocadura: els deltes del riu Ebre i del riu Llobregat.

La dinàmica fluvial és una de les forces més poderoses de la natura, la qual ha anat modelant molts dels paisatges de la terra. Rius i torrents, més que cap altre element del paisatge terrestre, són vies de connexió naturals entre els sistemes més diversos des del seu naixement fins a la seva desembocadura al mar. A més, els rius també han estat artèries reguladores dels assentaments humans. La importància de l'aigua per a la vida humana ha fet que des de temps remots la humanitat hagi cercat la proximitat al medi aquàtic per a establir-se, degut a l'existència de diversos factors: sòls més productius, un subministrament garantit de l'aigua, un mitjà de transport econòmic, un potencial de pesca i recreatiu i, finalment, per raons estètiques. No obstant, convé no oblidar que les inundacions són fenòmens naturals que afecten a molts rius. Durant períodes de descàrrega anormalment llargs, el canal fluvial no és capaç d'incorporar tota l'aigua que li arriba, de manera que el seu excedent s'escampa ràpidament sobre la plana d'inundació. Lamentablement, molts assentaments humans sovint s'estenen dins el territori que constitueix aquesta zona, la qual cosa pot provocar seriosos problemes en períodes d'inundació.

En termes quantitius, es fa difícil establir recomanacions per a la predicció de la quantitat d'aigua disponible en àrees que presenten una forta dependència dels recursos aquàtics. Durant els períodes secs, en els quals l'escolament superficial cap al riu és proper a zero, la recàrrega del curs fluvial es realitza directament a partir de les aigües subterrànies o bé de les aigües que lentament s'alliberen des de les zones humides. S'anomena cabal base d'un riu a aquell subministrament d'aigua mínim necessari per al manteniment de les característiques essencials i de l'estructura de la comunitat natural aquàtica que s'hi troba present, basant-se en criteris estrictament hidrològics i propis de cada ambient fluvial considerat. Una gestió política adequada hauria de planificar l'aprofitament dels recursos aquàtics en funció d'aquest nivell de cabal base. Tot i això, convé estar preparats davant la possibilitat de llargs períodes de sequera, on fins i tot els nivells garantits de cabal base poden disminuir, la qual cosa requerirà la implementació de mesures de conservació de l'aigua.

Els materials que han estat rentats des de la superfície terrestre i incorporats al riu, juntament amb els sediments que han estat erosionats del canal, constitueixen la càrrega del riu. Aquesta càrrega és transportada en suspensió, en solució, i al llarg del llit del riu mitjançant la força del corrent. La composició química de l'aigua d'un riu pot ser molt variable, essent el resultat de la interacció entre l'aigua de la precipitació, el sòl i les roques, a la qual cosa cal afegir també les aportacions d'origen antropogènic. Aquesta composició, però, pot variar de manera estacional.

## Capítol 1

---

Així per exemple, durant la tardor i l'hivern es produeix habitualment un augment de la concentració de compostos orgànics com a resultat de la degradació de la biomassa acumulada al sòl i de l'arrossegament de material húmic. Per altra banda, a l'estiu, la disminució de les pluges sovint comporta una reducció del cabal del riu i, com a conseqüència d'això, un augment de la concentració de les espècies dissoltes.

Molts dels problemes que actualment tenen relació amb la contaminació de l'aigua tenen el seu origen en els diferents usos que existeixen dins el territori de la conca hidrogràfica (veure Figura 1.2). En termes generals, els usos presents en una conca es poden agrupar en tres categories diferents: i) àrees protegides per a la vida animal i/o recreatives, les quals es mantenen en un estat natural i relativament inalterat; ii) àrees destinades a activitats rurals o agrícoles (entre elles, les zones cultivades); iii) àrees que contenen activitats industrials i urbanes.

Cadascun d'aquests usos introduirà una barreja diferent de contaminants dins del medi d'una conca hidrogràfica, amb la conseqüent afectació sobre els seus diferents compartiments ambientals.

## 1.2. Problemàtiques que afecten la qualitat ambiental d'una conca hidrogràfica

A finals del segle XIX, quan moltes ciutats dels països desenvolupats experimentaven un fort creixement com a resposta al procés de la Revolució Industrial, les deixalles que s'acumulaven als carrers van començar a esdevenir un problema per a la població. La brutícia, juntament amb una manca de costums sanitaris que actualment es troben normalment assumits, va portar a la dispersió de malalties i epidèmies arreu del món. Molts dels recursos d'aigua utilitzats a les ciutats s'extreien directament de les aigües superficials continentals, les quals solien estar contaminades. Els esforços per tal de controlar les malalties, van començar amb la filtració de l'aigua de consum públic a través de sorres [1]. Més tard, a principis del segle XX, el clor va ser afegit a l'aigua municipal abans de ser conduïda a través de la xarxa de distribució. Aquest fet va fer que moltes de les malalties que s'havien escampat a través de l'aigua comencessin a esdevenir cada vegada menys importants. No obstant això, durant la primera meitat del segle XX, un altre tipus de contaminació, aquest cop d'origen químic, va començar a experimentar un important increment [1]. Degut a l'avenç de la tecnologia, que esdevenia cada vegada més sofisticada, les tipologies de residus generats també es van anar fent cada vegada més variats. Actualment, els recursos aquàtics són atacats per diferents productes de naturalesa química com són fertilitzants, productes fitosanitaris, detergents, quantitats traça de metalls pesants, efluents àcids procedents de mines, i substàncies radioactives, a més d'una enorme varietat de productes químics industrials. Els contaminants d'origen químic, constitueixen avui en dia una de les principals amenaces a la qualitat de les aigües (veure Figura 1.2).

Com ja anunciava l'escriptora Rachel Carson en la seva reconeguda obra *Silent Spring* (1964), *"la intoxicació que ha penetrat en les nostres aigües prové de molts orígens: residus radioactius procedents de reactors, laboratoris i hospitals; contaminacions causades per les explosions nuclears; residus domèstics en ciutats i pobles; residus químics de les fàbriques...A això cal afegir-hi una nova forma de contaminació: les polvoritzacions químiques aplicades a terrenys de cultiu, jardins, boscos i camps. Molts dels agents químics imiten i augmenten els perniciosos efectes de les radiacions i, dins dels propis grups de productes químics, existeixen interaccions sinistres i poc conegudes, transformacions i acumulacions d'efectes. Des de que els químics van començar a fabricar substàncies que la natura no va inventar mai, el problema de la purificació de l'aigua ha esdevingut complex i el perill dels usuaris de l'aigua ha augmentat"* [2].



**Figura 1.2.** Activitats de les conques hidrogràfiques que poden afectar la qualitat de les aigües superficials (Moran et al., 1986)

En les conques naturals, l'aigua (1) dissol i erosiona lentament la roca, i les zones humides (2) alliberen sediments que aniran a parar a l'aigua. En conques urbano-industrials, les plantes de tractament d'aigües residuals (3) no poden eliminar tots els residus de l'aigua; l'aigua de la pluja (4) conté residus que provenen del rentat dels carrers de les ciutats; les aigües residuals industrials (5) contenen una gran varietat de contaminants; aigües àcides s'escapen de la mineria a cel obert (6); aigua escalfada és alliberada de les plantes energètiques (7); gasos industrials són rentats de l'atmosfera constituint la pluja àcida (8); el transport de petroli (9) pot causar abocaments, incloent petits vessaments durant la seva descàrrega; i l'abocament no controlat de residus (10) pot contaminar l'aigua subterrània. En conques agrícoles, l'aplicació de productes fitosanitaris (11) incorpora aquests productes a l'escolament superficial; els fertilitzants aplicats de manera incorrecta (12) es dissolen a l'aigua d'escolament; i els residus animals (13) són rentats de les superfícies agrícoles.

Els efectes que la contaminació química comportarà a llarg termini encara no es troben prou ben entesos per la societat en el moment actual. Tot i això, els efectes a curt termini són prou visibles - la mortalitat de peixos que acompanya episodis accidentals d'abocaments de substàncies químiques, platges inutilitzables, o llacs eutrofitzats pel creixement massiu de plantes aquàtiques, entre d'altres - .



Actualment, moltes indústries i municipalitats actuen d'acord amb les directrius establertes a la legislació vigent, les quals requereixen uns mínims estàndards de qualitat per als efluent. No obstant, les directrius per al control de petites descàrregues de substàncies amb una elevada toxicitat encara no han estat totalment establertes. Tot i que actualment s'estan elaborant diverses propostes per tal de regular la presència d'aquelles substàncies considerades potencialment perilloses pel medi ambient, a la pràctica, la seva inclusió en un text legal passa per procediments que encara són massa lents.

### 1.2.1. Legislació europea

La legislació en matèria d'aigües actualment vigent dins el territori de la Unió Europea es troba regulada per la Directiva 2000/60/CE o Directiva Marc de l'Aigua (DMA) [3]. Aquesta Directiva ha estat posteriorment modificada per una altra recentment aprovada pel Consell Europeu, la Directiva 2008/105/CE, que estableix els nivells màxims per a les substàncies prioritàries per tal de complir els estàndards de qualitat en aigües de riu.

Pel que fa a les aigües subterrànies, la Directiva més recent és la 2006/118/CE [4], específicament elaborada per a aigües de subsòl. Aquesta Directiva no modifica els nivells màxims permesos de plaguicides en aigües respecte a la Directiva anteriorment vigent, la 98/83/CE [5], relativa a les aigües de consum humà. Per hidrocarburs aromàtics policíclics, continuen estant igualment en vigor els límits establerts per la Directiva de l'any 1998.

La legislació referent a productes fitosanitaris d'ús agrícola, experimenta modificacions contínuament. Sovint s'elaboren directrius específiques per a un sol compost, com és el cas de l'atrazina en la Decisió 2004/248/CE [6]. També cal destacar algunes Directives que regulen altres tipus de compostos com, per exemple, els alquilfenols, substàncies que constitueixen una amenaça per al medi ambient i que són extensament utilitzades en la indústria i en les formulacions de productes fitosanitaris. Els alquilfenols es troben regulats des del 1986 per la Directiva 86/278/CEE [7] pel que fa a la seva presència en fangs de depuradora, i des del 2003 es regula la seva comercialització i ús en els estats membres a través de la Directiva 2003/53/CE [8].

Fins al moment, els compostos estudiats en aquesta Tesi no es troben legisllats en sòls. Existeixen només alguns acords no vinculants nascuts del Conveni de Rotterdam (1998), el qual proposa la promoció de la responsabilitat compartida en el comerç internacional de certes substàncies químiques perilloses per tal de protegir la salut humana i el medi ambient; i del Conveni d'Stockholm (2001), que pretén regular la contaminació per contaminants orgànics persistents (POPs) en el medi. Aquest darrer, estableix el que s'anomena la "dotzena bruta" (*dirty dozen*). En el cas dels sediments, de la mateixa manera que ocorre amb els sòls, no

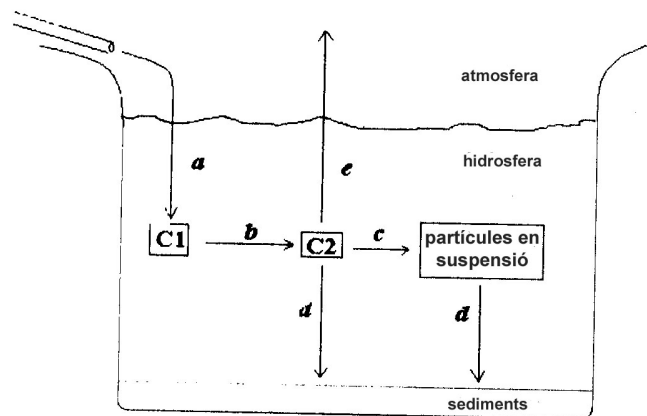
existeixen límits legiscats a Europa. Per tal de complir la legislació vigent, simplement és necessària una millora del nivell de contaminació respecte a l'any anterior.

### 1.2.2. La contaminació de les aigües superficials

Qualsevol episodi de contaminació, sigui del tipus que sigui, estarà constituït pels següents components:

- 1) una font de producció del contaminant,
- 2) un medi de transport i de transformació física i química, i
- 3) un medi receptor

La font d'origen de la contaminació determina la naturalesa química del contaminant emès, el medi on aquest s'aboca i el flux d'emissió. Un cop un contaminant (contaminant primari C1, veure Figura 1.3) és introduït en el medi, aquest es dispersa associant-se a les diferents fases que constitueixen el medi en funció de la seva naturalesa química. La solubilitat d'un contaminant serà el paràmetre que determinarà la seva dissolució en el medi hídric. Dins de l'aigua, el contaminant pot ser transportat a través de múltiples mecanismes. L'advecció, mecanisme de transport pel qual el contaminant es desplaça incorporat al moviment de la massa d'aigua, n'és un dels principals (Figura 1.3a). Altres mecanismes són la difusió molecular i la difusió turbulenta. Aquest últim, és el mecanisme més rellevant que es dona a les capçaleres dels rius, on els pendents són més acusats i, per tant, predomina el règim turbulent.



**Figura 1.3.** Cicle d'un contaminant a la hidrosfera (Domènech, 1999). C1) contaminant primari, C2) contaminant secundari, a) transport, b) transformació química, c) sorpció, d) sedimentació i e) volatilització

Les transformacions químiques més rellevants que el contaminant pot experimentar en medi hídric són les reaccions d'hidròlisi, l'oxidació deguda a la presència d'oxigen dissolt, els processos de reducció en reservoris profunds i on la concentració d'oxigen és molt baixa i, finalment, les reaccions fotoquímiques, especialment en zones superficials on l'exposició a la

llum solar és més alta (Figura 1.3b). En el cas de transformació química, s'obté un contaminant secundari C2 (veure Figura 1.3). D'altra banda, el contaminant també pot desaparèixer del medi per volatilització (Figura 1.3e), procés que afecta a les espècies més volàtils, les quals aniran a parar a l'atmosfera, o bé pot acabar dipositant-se en els sediments en cas que es produeixin reaccions de precipitació (Figura 1.3d). A més, en el medi hídric s'hi troben partícules sòlides en suspensió amb les quals el contaminant es pot associar (Figura 1.3c) i posteriorment sedimentar. En la majoria de casos, les partícules en suspensió provenen del sòl, arrancades a través de processos d'arrossegament. Es tracta de partícules de tipus argilós, amb una elevada superfície específica i capaces d'adsorbir i retenir espècies químiques. A l'aigua s'hi troba també material orgànic particulat, procedent del sòl i generat pels propis organismes aquàtics, el qual és capaç de retenir compostos químics hidròfobs.

### **Fonts de contaminació puntual i difosa**

Les fonts de contaminació descrites per diferents aportacions procedents de punts que es troben distribuïts dins una àrea determinada, i no a partir d'una única font clara i definida com seria, per exemple, una canonada, s'anomenen fonts de contaminació difosa. Contràriament a les fonts de contaminació puntual, que consisteixen en residus concentrats, provinents per exemple, de la indústria o de plantes de tractament d'aigües residuals, la contaminació per fonts difoses és especialment difícil i cara de controlar ja que malgrat trobar-se a concentracions relativament baixes, els seus volums en el medi ambient són enormes.

Les conques industrials i urbanes contenen una gran quantitat de fonts de contaminació puntual. Tot i que la quantitat de residus no tractats que van a parar directament als cursos fluvials d'aquestes àrees ha disminuït en els últims temps, la xarxa de clavegueram encara representa una problemàtica important pel que fa a la contaminació de l'aigua superficial, ja que no totes les ciutats disposen de sistemes de tractament de les aigües residuals, i no totes les indústries netegen l'aigua que aboquen directament a la llera. Els efluents d'algunes indústries poden ser tractats per les plantes municipals, però de vegades aquests són tan tòxics que les plantes no els accepten. És per aquest motiu, que les indústries que envien els seus efluents a les plantes de tractament han de limitar-ne el nivell de toxicitat mitjançant processos de pretractament. El problema és que no totes les fàbriques estan equipades de manera adequada per tal de dur a terme aquest tipus de processos i, a banda d'això, també poden existir abocaments accidentals. Com a resultat de l'ajust dels estàndards de qualitat de l'aire i de l'aigua, moltes indústries i municipalitats s'han vist obligades a enterrar els residus més sòlids enlloc de processar-los com a efluents. Es tracta d'un procés que normalment protegeix les aigües superficials, però que constitueix una amenaça per a les aigües subterrànies quan els emplaçaments no estan ben escollits, construïts o bé gestionats.

### **Contaminants naturals**

Les àrees naturals, com els boscos, pantans, prats, etc., generalment contribueixen amb petites quantitats de materials als cursos fluvials. Les aigües superficials que s'escolen a través d'aquests sistemes contenen concentracions "normals" de substàncies dissoltes que són característiques d'aquests medis. D'altra banda, la composició dels sediments varia considerablement i, com a resultat, l'aigua que ha estat en contacte amb aquests presentarà un contingut diferent de material mineral en funció de la naturalesa del sediment. Però no tota l'aigua que prové de les àrees naturals presenta un nivell de qualitat bo. Per exemple, aquells torrents que flueixen a través de zones pantanoses presentaran una coloració més groguenca i un pH més àcid, ja que hauran incorporat àcids orgànics generats per organismes descomponedors presents en els medis pantanosos. Un altre fenomen diferent s'observa en aquells rius que s'escolen a través d'àrees constituïdes per materials rics en calci, magnesi i bicarbonats. Encara que aquests compostos no representin un perill per a la vida aquàtica, provocaran un augment de la duresa de l'aigua del riu. Així, per exemple, a més duresa de l'aigua, la quantitat de detergents que haurà de ser aplicada per a la neteja, haurà de ser també força més elevada i esdevindrà, al seu torn, un altre problema de contaminació ambiental.

### **Contaminants antropogènics**

Una àmplia varietat de contaminants generats per l'activitat humana existent dins de l'àmbit d'una conca hidrogràfica seran detectats a les aigües superficials. A continuació es descriuen aquells contaminants que apareixen amb més freqüència en les problemàtiques ambientals de les conques hidrogràfiques estudiades en aquesta Tesi.

#### **▪ Metalls pesants**

Els elements metàl·lics són utilitzats en grans quantitats en la indústria per a una variada gamma d'aplicacions (catalitzadors, pintures, detergents, materials de construcció, additius, pesticides, etc.). Bona part d'aquests elements s'incorpora al medi en forma iònica i, per tant, soluble. Encara que molts elements metàl·lics són micronutrients, és a dir, nutrients necessaris en petites quantitats pel desenvolupament dels organismes vius, quan aquests excedeixen una determinada concentració esdevenen perjudicials. Els metalls que contenen arsènic, cadmi, crom, coure, plom, níquel, platí, zinc, tal·li, vanadi i or, estan classificats com a metalls pesants i considerats perillosos. Alguns com l'arsènic, cadmi, níquel i algunes formes del crom, es consideren, a més, carcinògens. Molts metalls pesants actuen com a inhibidors d'enzims, els quals afecten als processos metabòlics dels organismes, i tenen el perill afegit que no poden ser degradats a formes no tòxiques. Així, quan els medis aquàtics són contaminats per algun d'aquests compostos, esdevenen una amenaça potencial durant anys.

### ▪ Nutrients

La presència de nutrients com el nitrogen i el fòsfor en quantitats excessives a l'aigua, provoca un ràpid creixement de les algues i altres plantes verdes, les quals recobreixen la superfície de les aigües i impedeixen que la llum solar arribi a les capes inferiors. A conseqüència de la descomposició de la biomassa generada, es consumeix l'oxigen de l'aigua i això dóna lloc al fenomen de l'eutrofització.

Existeixen diverses fonts de nutrients que contribueixen a l'eutrofització. Una d'aquestes fonts són els residus domèstics. Aproximadament el 50% del fòsfor contingut en aquest tipus de residus prové dels detergents. L'escolament urbà també constitueix una font important de nitrogen i fòsfor. Aquests materials provenen de la fertilització de la gespa, de material fecal d'origen animal, i de productes de la combustió, entre d'altres. Les entrades de nutrients procedents de la indústria són molt variables. Les indústries papereres contribueixen amb enormes quantitats de substàncies que presenten grans demandes d'oxigen però alliberen baixes concentracions de nutrients per a les plantes. D'altra banda, aquelles indústries que disposen d'extenses àrees que necessiten ser netejades, com ara els rentadors de cotxes o les fàbriques de productes làctics, utilitzen grans quantitats d'agents de neteja que contenen fòsfor. L'explotació minera de fosfats és una altra font de contaminació important. Finalment, en àrees agrícoles on es realitza una aplicació important de fertilitzants, la quantitat de nutrients que poden anar a parar a les aigües també és significativa.

### ▪ Compostos orgànics

La contaminació del medi hídric per compostos orgànics és un fenomen cada vegada més important degut a l'augment de la producció de substàncies sintètiques. Les fonts de contaminació més rellevants per compostos orgànics són:

- aigües industrials i municipals, que contenen una variada gamma de compostos orgànics sintètics entre els que destaquen els hidrocarburs aromàtics (benzè, fenol i hidrocarburs aromàtics policíclics), compostos halogenats (alifàtics i aromàtics), bifenils policlorats, detergents, etc...
- aigües d'escolament, les quals no corresponen a abocaments puntuals sinó que arrosseguen contaminants generats en grans zones, com àrees rurals (productes fitosanitaris, residus animals, fertilitzants) i àrees urbanes.
- aigües de precipitació, les quals en algunes circumstàncies poden contenir compostos orgànics volàtils (dissolvents, hidrocarburs, productes fitosanitaris, etc...) o substàncies amb un elevat pes molecular adsorbides a les partícules sòlides (bifenils policlorats, dioxines, etc...).

## Capítol 1

---

Els residus industrials i municipals acostumen a contenir concentracions elevades de substàncies orgàniques. La seva presència potencia el creixement dels organismes descomponedors, que redueixen la quantitat d'oxigen dissolt de l'aigua, i a conseqüència d'això, altres membres de la comunitat aquàtica, especialment peixos i mol·luscs, es troben mancats d'oxigen i fins i tot poden arribar a morir.

Si en algun moment s'arriba a una eliminació completa de l'oxigen del riu, ocorre un canvi en el tipus d'organismes descomponedors, i es passa dels descomponedors aerobis (que requereixen oxigen) a un tipus de descomponedors anaerobis (que requereixen absència d'oxigen). Els productes d'aquests descomponedors també difereixen. Mentre que els aerobis generen sobretot diòxid de carboni, aigua, nitrats i sulfats, els quals generalment no representen cap perill, els anaerobis es caracteritzen per la formació de metà, amoni i sulfur d'hidrogen, que són gasos perillosos. Sota condicions anaeròbies, les aigües residuals es converteixen en ambients pútrids i tèrbols, condicions sota les quals cap organisme (a excepció de les bacteries anaeròbies) és capaç de sobreviure.

La quantitat d'oxigen dissolt que necessiten els descomponedors per tal de descompondre els materials orgànics presents en un volum determinat d'aigua s'anomena Demanda Biològica d'Oxigen (DBO). Així, la DBO serà una mesura del nivell de contaminació orgànica que es troba present en aigües residuals.

### **Productes fitosanitaris**

Els productes fitosanitaris són substàncies naturals o sintètiques que serveixen per combatre els agents responsables de les plagues (fongs, insectes, males herbes, etc.). S'introdueixen al medi hídric per mitjà de l'escolament superficial, infiltracions a través del sòl cap a les aigües subterrànies, introducció directa en el moment de la seva aplicació i abocaments per part de les indústries que els produeixen, a través de les aigües residuals. En general, es tracta de substàncies que presenten una baixa solubilitat en aigua i, per tant, una elevada tendència a adsorbir-se a les partícules sòlides i a dipositar-se en els sediments. Així, la seva concentració en el medi aquós sol ser baixa. Normalment, els plaguicides organoclorats (el més conegut, el DDT) són els que presenten una major resistència a la degradació. D'entre tots els productes fitosanitaris que van a parar a les aigües superficials i que posteriorment s'acumulen a les cadenes tròfiques, els plaguicides organoclorats són els que representen un major risc per a les espècies i el medi ambient en general. L'Agència de Protecció Ambiental (EPA) ha prohibit l'ús general de molts compostos organoclorats, com per exemple, el DDT, aldrí, dieldrí, heptaclor, clordà, i toxafè, entre d'altres.

### **Bifenils policlorats (PCBs)**

Alguns compostos clorats industrials com els PCBs també es bioacumulen. Com que es tracta de compostos resistents al foc i estables a elevades temperatures, s'utilitzen com a materials aïllants en diferents tipus d'aparells i en la fabricació d'alguns plàstics. Com a resultat del seu ús extens, els PCBs poden entrar als ecosistemes aquàtics mitjançant diverses vies. Amb la combustió de paper i plàstic que conté PCBs, per exemple, aquests es vaporitzen i condensen en les partícules de pols que després es dipositaran al sòl o bé a les aigües superficials o que seran rentades de l'atmosfera amb les precipitacions. A més, aquests compostos també poden entrar als ecosistemes aquàtics per escolament superficial, evaporació a partir d'abocaments accidentals, o com a efluents de les plantes de tractament, etc.

### **Alquilfenols (APs)**

Es tracta de substàncies que presenten excel·lents propietats surfactants utilitzades en aplicacions industrials, agrícoles i domèstiques. Es troben en els productes de neteja, tèxtils, petroli, resines fenòliques, vernissos, en la indústria del paper i en la formulació de diversos pesticides. Una de les principals fonts d'alquilfenols (APs) en el medi ambient són les entrades directes d'efluents urbans o industrials, o bé a través dels efluents de les estacions depuradores d'aigües residuals (EDARs), les quals en moltes ocasions no són capaces d'eliminar completament aquestes substàncies durant els processos de tractament de les aigües, de manera que els nivells d'APs al sistema aquàtic incrementen. Són compostos que actuen com a disruptors endocrins (capaços d'alterar el funcionament normal del sistema endocrí dels animals). El nonilfenol (NP), per exemple, és el producte de degradació majoritari dels anomenats alquilfenols etoxilats, i presenta més toxicitat i més potencial estrogènic que els seus compostos parentals.

### **▪ Substàncies farmacèutiques**

Les substàncies farmacèutiques són un tipus dels anomenats nous contaminants emergents que, en els darrers anys, han anat adquirint importància degut a l'augment de la seva presència en el medi. Els productes farmacèutics en la seva forma original o bé com a metabòlits són contínuament introduïts a les aigües residuals, sobretot a través de l'excreció, l'eliminació de medicaments sense utilitzar o bé caducats, o directament a partir d'abocaments per part de les indústries farmacèutiques. Aquests compostos es troben distribuïts per tot el medi aquàtic, ja que alguns no són eliminats de manera eficient durant els processos de tractament d'aigües residuals, de manera que poden arribar a afectar les aigües superficials i subterrànies i, per tant, l'aigua de beguda. Pel fet de ser contínuament introduïts al medi ambient a través de les

aigües residuals, aquests productes poden ser classificats com a contaminants pseudopersistents.

### ▪ Altres compostos tòxics

Alguns compostos químics que s'utilitzen en la indústria comporten perills directes tant per als humans com per a la vida aquàtica. Alguns exemples en són l'amoní, cianurs, sulfurs, fluorurs, àcids forts i alcalins. Molts d'aquests compostos s'utilitzen en les seves formes concentrades, que són directament tòxiques per als organismes però indispensables per a la indústria. La descàrrega accidental d'aquests residus tòxics, generalment fruit de l'error humà, pot arribar a eliminar les comunitats aquàtiques que es troben presents riu avall de l'abocament.

L'exposició d'un organisme a substàncies tòxiques pot ser agut (com a resultat d'una sola dosi forta) o crònic (com a resultat de dosis més petites però repetides en el temps). El perill que comporta la fuga de petites quantitats de materials tòxics al medi ambient, encara que d'entrada sembli no representar una amenaça immediata per a la salut humana, és un problema tant important com podria ser-ho un gran abocament industrial. Es tracta de substàncies que tenen la capacitat de bioacumular-se en els organismes i que per tant, seran posteriorment transmeses a través de la cadena tròfica.

De la mateixa manera que els nutrients passen d'un nivell tròfic al següent, el mateix passa amb les substàncies tòxiques persistents. Si aquestes substàncies no són excretades o eliminades biològicament per l'organisme, es queden retingudes en els seus teixits. En el cas que l'organisme continuï ingerint aliment contaminat durant un cert període de temps, la concentració de l'agent contaminant va augmentant dins de l'organisme. És per això que els contaminants persistents que entren a les cadenes tròfiques es van fent cada vegada més concentrats a mesura que passen a un nivell tròfic superior. D'aquesta manera, els organismes que se situen en els nivells superiors de la cadena (entre ells els humans) es troben exposats a unes concentracions més elevades d'aquestes substàncies. Es poden identificar algunes propietats químiques que afavoreixen el procés de bioacumulació. Per exemple, aquelles substàncies més liposolubles tindran més tendència a bioacumular-se, així com els compostos que siguin més químicament estables, o sigui, capaços de persistir en el medi durant períodes de temps més llargs. Per tal de ser persistent, un compost ha de ser resistent a la biodegradació per part dels organismes, a la reacció amb aigua o amb la humitat (hidròlisi), o bé a la degradació per la llum solar (fotòlisi).



### 1.2.3. La contaminació de les aigües subterrànies

Les mateixes activitats que provoquen la contaminació de les aigües superficials d'una conca hidrogràfica, també poden arribar a contaminar les aigües subterrànies. L'abocament de residus, les activitats agrícoles, les fosses sèptiques, les llacunes que emmagatzemen residus industrials, els pous d'injecció subterrània (en els quals els residus industrials són bombejats a pressió per sota de les capes de roca permeable), i la producció de petroli i gas natural, s'identifiquen com a possibles amenaces a la qualitat de les aigües subterrànies. Cal tenir en compte que aquesta qualitat també variarà depenent de quina sigui la font d'aigua subterrània, a més de les interaccions que existeixin entre aquesta aigua i el sòl i materials rocosos a través dels quals circuli.

#### ▪ Nutrients

La infiltració de l'aigua a través del sòl és un mecanisme de purificació bastant eficient. De fet, en el cas de substàncies com els fosfats, aquests s'uneixen a les partícules de la superfície del sòl i són eliminats (per tant, no passen a les aigües subterrànies). No obstant, a diferència dels fosfats, els nitrats presenten una problemàtica major. Donat que són substàncies molt solubles en aigua, no s'adsorbeixen a les partícules del sòl, i són principalment perillosos quan van a parar a aigües dels pous. Les concentracions de nitrats més elevades en aigües subterrànies es detecten en àrees on existeix una gran aplicació de fertilitzants agrícoles.

#### ▪ Compostos orgànics

##### Hidrocarburs

Un altre problema de contaminació que amenaça freqüentment les aigües residuals és el degoteig de tancs d'emmagatzemament de combustible. Les estacions de servei, les granges i moltes indústries i institucions emmagatzemen aquestes substàncies en tancs d'acer ubicats sota terra, els quals es troben exposats a la corrosió i per tant, a fuites del seu contingut. Els sòls no retenen ni la benzina ni altres substàncies derivades del petroli, i les petites fuites no acostumen a ser detectades fins que no es descobreix algun pou contaminat. Però arribat aquest moment, el problema ja es considera seriós, pel fet que una petita quantitat de gasolina és capaç de contaminar un gran volum d'aigua.

##### Residus municipals, industrials i miners

Els residus municipals, industrials i miners també poden contaminar les aigües subterrànies. Es tracta de residus emmagatzemats normalment sota terra, on l'aigua pot infiltrar-se i dissoldre'n els compostos solubles, incorporant aquest tipus de contaminació. L'aigua que passa a través d'una zona de recàrrega contaminada pot arribar a contaminar les reserves d'aigua subterrània

durant un període de centenars i fins i tot milers d'anys, ja que dins de l'aqüífer gairebé no existeix dilució ni tampoc eliminació de contaminant. Les activitats mineres també poden contaminar les aigües subterrànies de diverses maneres. Algunes alliberen metalls pesants tòxics, altres augmenten l'acidesa de l'aigua, i n'hi ha que canvien els nivells i patrons de circulació de l'aigua subterrània. L'element clau per a la protecció de l'aigua subterrània consisteix, doncs, en evitar qualsevol opció de contaminació.

### **Productes fitosanitaris**

Els productes fitosanitaris no haurien de ser utilitzats en zones de recàrrega on els sòls no són capaços d'aturar-ne la migració. Tot i que es tracta de productes necessaris per a l'agricultura, alguns d'ells, i sobretot els més solubles, han estat freqüentment detectats a les aigües subterrànies de les àrees irrigades.

## **1.2.4. La contaminació del sòl**

El sòl ha estat, des de sempre, l'abocador per excel·lència de bona part dels residus originats per l'home, degut a la seva fàcil accessibilitat. Tan sols la pressió demogràfica ha fet que la recerca de terrenys suficientment allunyats dels nuclis urbans resulti cada cop més difícil i inciti a buscar altres medis naturals (com són la hidrosfera i l'atmosfera).

A diferència del que ocorre en el medi hídric, un contaminant té, en principi, poca mobilitat en el sòl. A més, els efectes de dilució característics d'altres medis tan sols juguen un paper marginal en aquest compartiment. És per aquest motiu que la probabilitat de que un contaminant s'incorpori a la cadena tròfica a través de la seva ingestió per part d'algun organisme que tingui el sòl com a substrat vital, és molt elevada.

Ja s'ha esmentat el paper depurador que presenten determinats components del sòl. En aquest sentit, cal destacar la funció dels materials húmics, els quals són capaços de fixar una elevada varietat de substàncies dissoltes en l'aigua edàfica, en particular, cations polivalents de metalls pesants. Per tant, la persistència dels contaminants tendeix a ser major en sòls amb nivells superiors de matèria orgànica.

### **▪ Nutrients**

Els residus de les granges contribueixen a les entrades d'ió amoni al sòl. El nitrogen orgànic, que és la forma sota la qual està present aquest element en els residus animals, es transforma en ió amoni sota l'acció catalítica de determinats bacteris anomenats amonificants. Altres aportacions de nitrogen són degudes a l'ús de fertilitzants, emissions de la indústria i abocaments d'aigües residuals domèstiques.

Fins fa poc, la generació i posterior aplicació de residus animals no suposava cap tipus de problemàtica ambiental. Històricament, els fems s'han aprofitat per a la fertilització del sòl gràcies al seu elevat valor nutritiu per les espècies vegetals, ja que no només contenen nutrients minerals sinó també una elevada proporció de matèria orgànica. No obstant, actualment es produeixen quantitats tan grans de residus animals que el sòl és incapaç d'assimilar-les, sobretot quan aquests són aplicats de manera incontrolada, sense tenir en compte ni l'època de l'any ni la quantitat, tan sols guiant-se per la necessitat de buidar les fosses. L'aplicació al sòl d'un excés de fertilitzant pot originar, a la llarga, problemes de contaminació de l'entorn.

En certes ocasions, els residus animals contenen antibiòtics o bé altres productes farmacèutics que procedeixen de la seva ingestió per part del bestiar en cas de malaltia i que no han estat assimilats per l'animal. En el cas dels antibiòtics, quan aquests passen al sòl poden alterar-ne significativament l'ecologia microbiana i, com a conseqüència, afectar els processos bioquímics que hi tenen lloc.

L'aplicació de dosis elevades de residus animals al sòl pot conduir a una salinització excessiva d'aquest, de manera que el creixement de les plantes es vegi limitat. A més, un excés de sals pot produir una desestabilització del sòl que ocasioni una disminució de la seva capacitat d'infiltració.

### ▪ **Metalls pesants**

El procés de biometilació origina la mobilització de metalls pesants tòxics, ja que produeix una reacció entre ions metàl·lics i determinades substàncies orgàniques naturals per tal de formar compostos organometàl·lics. En aquesta situació, les propietats físico-químiques del metall canvien radicalment. Els compostos resultants solen ser molt liposolubles, de manera que travessen amb facilitat les membranes biològiques i, en particular, la pell. Un dels processos de biometilació més importants és l'enllaç entre un catió metàl·lic i el grup metil (-CH<sub>3</sub>). Aquest és el principal mecanisme natural de mobilització de metalls pesants. Els metalls que ofereixen més afinitat a la biometilació són el mercuri, el plom, l'arsènic i el crom. Aquests metalls es troben presents en molts tipus de sòls a concentracions molt baixes, de manera que normalment es tracta d'un fenomen de poca transcendència. El problema esdevé més important quan aquests metalls són introduïts al sòl de manera externa, en forma d'abocaments incontrolats. En aquests casos, la probabilitat que aquests metalls s'incorporin a la cadena tròfica és bastant elevada.

### ▪ **Compostos orgànics**

#### **Productes fitosanitaris**

## Capítol 1

---

Els productes fitosanitaris poden ser aplicats directament sobre el sòl o bé en forma d'esprai, ruixant la superfície de les fulles de les plantes. En aquest darrer cas, una part del producte persisteix absorbit sobre la superfície de la planta mentre que la resta es dissol o és arrossegat per l'aigua de la pluja fins a incorporar-se al sòl.

Un cop al sòl, el producte pot migrar a través d'aquest i arribar fins a les aigües subterrànies. La facilitat amb què es porti a terme aquest desplaçament dependrà de la naturalesa i propietats químiques del producte aplicat, de la seva formulació, del mètode d'aplicació, de l'entorn on s'aplica (existència de pendents, grau de drenatge del terreny, etc...), així com del tipus d'espècie cultivada. En principi, aquells productes fitosanitaris que siguin poc solubles en aigua, poc volàtils i químicament estables, romandran al sòl durant un període de temps llarg.

Els productes fitosanitaris que més s'utilitzen en l'actualitat són els compostos organoclorats, els organofosforats i els carbamats. Els plaguicides organoclorats van ser els primers en ser comercialitzats i són, de fet, els responsables d'un impacte més negatiu sobre el medi ambient, ja que el seu efecte letal es segueix mantenint durant anys després de la seva aplicació. El DDT, que és un compost relativament fàcil de sintetitzar i econòmic en relació a altres compostos d'igual efectivitat, mata els insectes per simple contacte, per alteració del seu sistema nerviós. D'altra banda, els compostos organofosforats són menys estables que els seus anàlegs organoclorats i, per tant, menys persistents en el medi. No es bioacumulen, de manera que són menys nocius i presenten una menor probabilitat de passar a la cadena tròfica. No obstant, alguns com el metilparatió, són molt tòxics. Els carbamats també actuen sobre el sistema nerviós però són més innocus que els pesticides organofosforats.

L'aplicació indiscriminada de productes fitosanitaris al sòl, pot originar problemes greus per a l'equilibri ecològic del medi. Amb el pas del temps, si es continua aplicant el mateix producte i noves generacions de la plaga es van desenvolupant, són cada vegada més els individus que hi esdevenen resistents. Això obliga a l'agricultor a aplicar el producte amb més freqüència i en majors quantitats, de manera que l'efecte negatiu del compost es multiplica. A més, hi ha agricultors que utilitzen una bateria de productes fitosanitaris, és a dir, un conjunt de productes de naturalesa química diferent però que s'apliquen al mateix temps. En aquest cas el risc d'acumulació de residus continua essent elevat.

Un altre problema relacionat amb l'aplicació d'aquests productes és que tenen poca selectivitat, és a dir, que són poc específics respecte a una espècie determinada. Com a conseqüència d'això, quan s'aplica el producte fitosanitari existeixen moltes possibilitats de que altres espècies que cohabituen juntament amb les causants de la plaga resultin també afectades.

Tot i que existeix una necessitat real d'aplicar productes fitosanitaris per millorar el rendiment de les collites, és necessari un canvi en les pràctiques agrícoles, de manera que es vagi tendint

cap a una agricultura menys intensiva, però més selectiva. Es fa necessari el desenvolupament de noves metodologies compatibles amb el medi natural. En aquest sentit, el producte ideal seria aquell que presentés una elevada toxicitat respecte a l'espècie a eliminar, que persistís en el medi el temps just per tal d'assolir una elevada taxa d'eliminació, i que no afectés a la resta d'organismes vius de l'entorn.

### **Hidrocarburs aromàtics policíclics (PAHs)**

Els PAHs es poden formar de manera natural (formació geològica de combustibles fòssils), mitjançant processos de combustió, tant natural com antropogènica (incendis forestals, transports, incineradores, centrals elèctriques...), o per síntesi industrial. Solen acumular-se a la matèria orgànica del sòl, però depenent de la seva massa molecular, també poden tendir a evaporar-se. Degut a la seva difícil degradació (especialment els més pesants) es poden classificar com a contaminants ubics. La via més habitual d'eliminació al sòl és mitjançant la degradació microbiana. Tot i que els animals metabolitzen aquest tipus de substàncies, es pot donar bioacumulació de PAHs a les plantes.

En molts casos, els residus que es troben al sòl provenen d'abocaments indirectes, com és el cas dels òxids de sofre i de nitrogen emesos principalment a causa de la combustió de carburants fòssils. Bona part d'aquests òxids finalitzen el seu recorregut dipositant-se al sòl. El sòl rep també altres aportacions aèries com són, per exemple, les partícules sòlides que, quan provenen d'àrees urbanes i industrials, poden transportar una càrrega química complexa [9].

## **1.2.5. La contaminació dels sediments**

Quan les gotes de pluja cauen sobre el sòl desprotegit (sense vegetació) arrenquen les partícules de sòl. Aquestes partícules són transportades a través de l'escolament superficial cap a rius, llacs o oceans i són dipositades allí com a sediments. La part grollera del sediment, com són les sorres, es diposita força ràpidament. No obstant, les partícules més fines com les argiles, poden continuar en suspensió durant mesos, afavorint una major torbesa de l'aigua. Aquells sòls que es deixen sense cultivar, ja sigui per una rotació de conreus (guaret), la tala de fusta, la mineria a cel obert, la sobrepastura, la construcció de carreteres o altres activitats de construcció, estan subjectes a elevades taxes d'erosió.

Sovint, els sediments acaben emmagatzemant els contaminants hidròfobs que van a parar a les aigües superficials i que, després d'adsorbir-se a les partícules en suspensió, acaben sedimentant. Aquest seria el cas de substàncies com els productes fitosanitaris, els hidrocarburs aromàtics i els alquilfenols, entre d'altres. Els sediments constitueixen el compartiment ambiental que aporta un registre més fiable de la contaminació que afecta al riu,

## Capítol 1

---

ja que a diferència de l'aigua superficial, que flueix contínuament, el sediment és un medi bastant estàtic.

## 1.3. Metodologies aplicades a l'anàlisi i avaluació de la qualitat ambiental d'una conca hidrogràfica

L'estudi de l'àmbit d'una conca hidrogràfica inclou un ampli ventall d'aspectes, ja que aquesta es troba constituïda per diversos compartiments ambientals amb característiques molt diferents. Al llarg del temps, s'han anat desenvolupant diverses maneres d'estimar la càrrega de substàncies transportada per un riu, o bé l'escolament superficial que existeix en una determinada conca. En els estudis ambientals, és important separar aquelles contribucions que provenen de fonts diferents, a més de distingir entre la variabilitat natural i l'impacte antropogènic. Això permetrà realitzar un control ambiental més eficient i seleccionar les tècniques de gestió més adients en cada cas.

Seria difícil, a més de força car, obtenir una bona descripció dels diferents fenòmens que es donen dins del sistema d'una conca hidrogràfica basant-se només en mesures directes. És per aquest motiu que es fa necessari l'aplicació d'algun tipus de metodologia que permeti l'anàlisi i avaluació general de la qualitat ambiental d'una conca. En aquesta Tesi es presenten, a nivell general, dos tipus de metodologies per tal de dur a terme aquesta tasca d'avaluació: metodologies empíriques i metodologies deterministes.

Els models de tipus determinista [10] són mètodes numèrics basats en lleis generals o principis teòrics que permeten l'estimació de la qualitat de diferents compartiments ambientals d'una conca i la simulació del transport de substàncies a través de la xarxa hidrogràfica. Pressuposen que un sistema o procés opera de tal manera que un conjunt determinat d'esdeveniments porta a un resultat únic. D'altra banda, els models de tipus empíric [11-12] consideren la incertesa estadística tant en les condicions inicials com en les prediccions obtingudes pel model, incorporant-hi una component aleatòria.

Històricament, les comunitats científiques que han utilitzat metodologies deterministes són totalment diferents de les que han utilitzat metodologies empíriques, i existeix un gran desconeixement mutu de les possibilitats que ofereixen unes i altres. No obstant, en aquesta Tesi s'intenta apostar per la integració d'ambdós tipus de metodologies, per a la qual ja existeixen propostes incipients.

### 1.3.1. Metodologies empíriques

Les dades ambientals es caracteritzen normalment per una elevada variabilitat. Els principals orígens d'aquesta variabilitat en l'àmbit d'una conca hidrogràfica poden ser influències de tipus geogènic, hidrològic i meteorològic, però també hi tenen un paper important les fonts de tipus

## Capítol 1

---

antropogènic. Les incerteses resultants del procés de presa de mostres, dels processos de conservació i pretractament d'aquestes mostres, de les mesures analítiques i fins i tot de la fase de calibratge, es troben contingudes en les dades finals. Per aquest motiu, no és sorprenent que els resultats analítics de qualsevol estudi realitzat sobre una conca hidrogràfica continguin aquesta variabilitat produïda per les incerteses, a més de les principals característiques que influencien el medi estudiat. Per tal de resoldre el problema de les incerteses, cal que es treballi amb un elevat nombre de mostres. Això implica un gran volum de dades, les quals portaran associada molta informació redundant i latent.

Amb l'objectiu de descobrir les pautes i fonts de variació existents en les sèries i taules de dades multivariants obtingudes dels programes de seguiment ambiental sobre una conca hidrogràfica, es proposa l'aplicació de metodologies de tipus empíric. L'assumpció bàsica d'aquestes metodologies és que cadascun dels paràmetres mesurats experimentalment en una mostra particular (per exemple, la concentració d'un contaminant en un tram de riu) es troba afectat per diferents contribucions de fonts independents (que poden ser tant puntuals com difoses) d'orígens diversos.

Les metodologies empíriques es basen en la descripció i anàlisi de la variància observada en les dades experimentals. Ofereixen eines eficients per a l'extracció d'informació útil a partir de grans conjunts de dades, i són capaces d'identificar i capturar aquells paràmetres clau que condicionen el comportament dels sistemes ambientals. Ja que molts sistemes ambientals involucren processos que no es coneixen completament i per als quals no existeixen models formals que els descriguin, es presenta la necessitat d'utilitzar metodologies complementàries a les metodologies de tipus determinista. Dins del grup de metodologies empíriques existeixen, entre altres, les que es recullen a la Taula 1.1. En aquesta Tesi no es pretén fer una classificació exhaustiva de les diferents metodologies empíriques que existeixen per al tractament de les bases de dades ambientals, sinó tan sols citar-ne alguns exemples.

---

### Metodologies empíriques de tractament de dades ambientals

Mètodes d'anàlisi de dades ( <i>data mining</i> )	Mètodes de l'estadística i àlgebra multivariant	Tècniques d'anàlisi de variància (ANOVA) i anàlisi discriminant <sup>1</sup> Anàlisi de factors ( <i>factor analysis</i> ) <sup>2</sup> Anàlisi de components principals ( <i>principal component analysis, PCA</i> ) <sup>3</sup> Reconeixement de pautes ( <i>pattern recognition</i> ) <sup>4</sup> Anàlisi d'agrupacions ( <i>cluster analysis</i> ) <sup>5</sup> <i>k</i> -veí més proper ( <i>k-nearest neighbor</i> ) <sup>6</sup> SIMCA <sup>7</sup> Anàlisi de sèries temporals i d'autocorrelació ( <i>time series analysis</i> ) <sup>8</sup> ; models ARIMA <sup>9</sup>  Mètodes de regressió multivariant en les seves diferents variants, entre els que actualment destaquen
---	---	--



	els mètodes de regressió per mínims quadrats parcials ( <i>partial least squares, PLS</i> ) <sup>10</sup>
	Xarxes neuronals ( <i>artificial neural networks</i> ) <sup>11</sup>
	Xarxes de decisió basades en mètodes probabilístics o baiesians ( <i>bayesian decision networks</i> ) <sup>12</sup>
	<i>Support vector machines</i> <sup>13</sup>
	Algoritmes genètics ( <i>genetic algorithms</i> ) <sup>14</sup>
	Lògica difosa ( <i>fuzzy logic</i> ) <sup>15</sup>

Mètodes geostatístics<sup>16</sup>; desenvolupats fonamentalment des de la hidrologia i les ciències de la terra

<sup>1</sup> McLahlan 1992. Wiley, Chichester

<sup>2</sup> Malinowski 1991. Wiley, New York

<sup>3</sup> Jolliffe 2002. Springer, Berlin

<sup>4</sup> Breerton 1992. Elsevier Science, New York

<sup>5</sup> Massart and Kaufman 1983. Wiley, New York

<sup>6</sup> Fix and Hodges 1951. Report Number 4, USAF, Texas

<sup>7</sup> Wold and Sjostrom 1977. ACS Symposium Series 52

<sup>8</sup> Chatfield 1989. Chapman and Hall, London

<sup>9</sup> Abraham and Ledolter 1983. John Wiley and Sons, NJ

<sup>10</sup> Geladi and Kowalski 1986. Anal Chim Acta 185:1-17

<sup>11</sup> Zupan and Gasteiger 1991. Anal Chim Acta 248: 1-30

<sup>12</sup> Jensen 1996. Springer-Verlag, New York

<sup>13</sup> Cristianini and Shawe-Taylor 2000, Cambridge University Press, Cambridge

<sup>14</sup> Goldberg 1989. Addison-Wesley, Reading, MA

<sup>15</sup> Bandemer and Näther 1992. Kluwer, Dordrecht

<sup>16</sup> Cressie 1991. Wiley, New York

**Taula 1.1.** Models empírics de tractament de dades ambientals

## Mètodes d'anàlisi de dades

Els mètodes d'anàlisi de dades que s'han utilitzat en aquesta Tesi corresponen exclusivament al grup dels mètodes de l'estadística i àlgebra multivariant. És per aquest motiu que la resta de mètodes, tot i que es troben llistats a la Taula 1.1, no seran descrits en aquest apartat.

L'anàlisi estadística multivariant s'aplica a dades que consisteixen en múltiples mesures d'un nombre d'individus, objectes o mostres. La mesura i l'anàlisi de la dependència entre les variables és fonamental en aquest tipus de metodologia [13], ja que es pretén l'agrupació i classificació dels objectes o, en alguns casos, la modelització de les relacions que existeixen entre aquests. Dins d'aquest conjunt de mètodes hi trobem diferent tipus de tècniques com són: les tècniques d'anàlisi de la variància i anàlisi discriminant, l'anàlisi de factors, l'anàlisi de components principals (PCA), o l'anàlisi d'agrupacions, entre d'altres.

En el cas de les **tècniques d'anàlisi de la variància (ANOVA)** i de l'anàlisi discriminant, es parla de mètodes d'anàlisi supervisada de dades. En aquests, la identitat de les categories o factors és coneguda d'entrada i la finalitat és determinar si la mitjana d'una variable difereix significativament entre els diferents grups. L'extensió multivariant de l'anàlisi de la variància dona lloc a la denominada MANOVA, on l'interès rau en determinar si el conjunt de mitjanes de

les múltiples variables són diferents entre si. D'altra banda, quan es tracta de determinar si les mitjanes d'un conjunt de dades difereixen degut a diversos factors o combinació de factors (més de dos), s'utilitza la generalització d'ANOVA, anomenada N-way ANOVA.

**L'anàlisi de factors** és un mètode de l'estadística multivariant que s'utilitza per a la identificació d'un nombre reduït de factors a partir d'un conjunt de dades amb variables correlacionades. El conjunt original de dades (amb  $n$  dimensions) es projecta sobre un sistema de coordenades de dimensions més reduïdes (1, 2 o 3 dimensions). Com que cada factor pot afectar diverses variables, aquests s'anomenen factors comuns. S'assumeix que cada variable depèn de la combinació lineal dels factors comuns, i els coeficients s'anomenen *loadings*. Les variables mesurades inclouen també una component deguda a la variabilitat aleatòria, anomenada variància específica, perquè és específica a cada variable. Es tracta d'una eina útil per a extreure informació latent de les dades, com ara aquelles relacions no directament observables que existeixen entre elles.

**L'anàlisi de components principals (PCA)** és probablement el mètode més àmpliament utilitzat dins de la família de mètodes de l'anàlisi de factors. S'aplica PCA per tal de realitzar una anàlisi exploratòria de les dades, condensant el màxim d'informació sobre el sistema utilitzant el menor nombre possible de variables. El PCA aconsegueix aquesta simplificació de la informació substituint grups de variables per una nova variable. El mètode genera un nou conjunt de variables anomenades components principals, on cada component és combinació lineal de les variables originals. Tots els components principals són ortogonals i no contenen informació redundant. El primer component principal és un eix a l'espai. Quan cadascuna de les mesures originals es projecta sobre aquest eix, els valors resultants defineixen una nova variable. I la variància d'aquesta variable és la màxima d'entre totes les possibles alternatives d'aquest nou eix. El segon component principal és, de la mateixa manera, un altre eix a l'espai, el qual serà perpendicular al primer. La projecció de les mesures originals sobre aquest segon eix, defineix una nova variable, la variància de la qual també serà màxima. El conjunt total de components principals serà tant gran com el conjunt original de variables. No obstant, només els primers components principals, els quals descriuen la major part de la variància de les dades originals, seran seleccionats per explicar les dades de manera simplificada. A la secció 2.1.2.4 d'aquesta Tesi es presenta una descripció detallada del mètode PCA.

Normalment, la terminologia i els objectius per descriure l'anàlisi de factors i l'anàlisi de components principals se superposen, de manera que es fa difícil fer una distinció entre els dos mètodes. Es tracta, en tots dos casos, de tècniques de reducció de la dimensionalitat de les dades, en el sentit que poden ser utilitzades per substituir un conjunt gran de variables mesurades per un conjunt més reduït de noves variables. No obstant, els dos mètodes es

diferencien en els seus objectius i en els seus models subjacents. En general, PCA pot ser utilitzat quan es vol simplificar o realitzar una aproximació de les dades utilitzant un menor nombre de dimensions (per exemple, per tal de visualitzar-les). En canvi, l'anàlisi de factors serà utilitzat quan es faci necessària l'obtenció d'un model que expliqui les correlacions que existeixen entre les dades.

La finalitat de l'**anàlisi d'agrupacions** és la identificació de grups d'objectes relativament similars pel que fa a les seves propietats mesurades. Es tracta de crear grups d'objectes o *clusters*, de manera que els perfils dels objectes dins del mateix grup siguin molt semblants mentre que els perfils dels objectes de diferents grups siguin diferents.

Els **mètodes de reconeixement de pautes** es descriuen com a processos de formulació de regles de classificació. Partint d'un conjunt de classes donat, cadascuna de les quals conté un nombre d'observacions representat per multitud de variables, es desenvolupen pautes i regles per tal de poder classificar les noves observacions com a semblants o bé diferents als membres de les classes existents. Es parla de dos tipus generals de classificació: supervisada i no supervisada. En el reconeixement de pautes no supervisat, el nombre i tipus de classes és desconegut i caldrà que sigui determinat. En el reconeixement de pautes supervisat, per contra, l'objectiu és assignar un objecte determinat a una classe prèviament definida.

El **k-veí més proper** és un mètode de classificació no supervisada en el qual es calcula, en primer lloc, la distància euclidiana d'un objecte a la resta d'objectes. Posteriorment, la distància mínima és seleccionada per tal d'assignar l'objecte a la classe corresponent. Normalment, el nombre d'objectes veïns  $k$  es fixa, i com a resultat de l'anàlisi s'obtenen separacions força flexibles entre grups d'objectes. No obstant, la classificació obtinguda dependrà del nombre d'objectes fixats per cada classe, i en el cas de classes superposades, aquest s'assignarà a la classe amb el major nombre d'objectes.

**SIMCA** és un mètode de classificació supervisada on les classes de les mostres del conjunt de calibratge són conegudes d'antuvi. Aquestes classes o grups d'objectes són aleshores descrits mitjançant un model matemàtic independent de la resta de classes. El mètode SIMCA construeix models de components principals diferents per a cadascuna de les classes a analitzar. Cadascun dels models pot identificar un nombre diferent de components principals, de manera que poden representar línies, plans, cubs o hipercubs en més de tres dimensions. Aquests models poden ser posteriorment utilitzats per eliminar objectes aberrants, per estimar la capacitat de modelatge d'una variable particular o bé per a classificar nous objectes.

## Capítol 1

---

Una **sèrie temporal** està formada per un conjunt d'observacions mesurades de manera seqüencial en el temps. Qualsevol procés discret o bé continu que sigui mesurat regularment pot ser representat per una sèrie temporal. Les correlacions que existeixen dins de les sèries temporals són descrites en forma d'autocorrelació o autocovariància i de les seves funcions derivades. Dues sèries temporals diferents es caracteritzen per la seva correlació creuada (*cross-correlation*). La informació que s'extreu normalment d'aquest tipus de models és: i) la desviació, ii) el soroll, iii) la periodicitat dels processos, i finalment, iv) una predicció dels valors futurs en funció de la tendència presentada per la sèrie.

En els **mètodes de regressió multivariant** s'utilitzen conjunts de dades que poden ser disposades en dos blocs diferents (X i Y). Tant si aquesta relació és lineal com si no ho és, serà necessari validar adequadament els resultats obtinguts. Entre els mètodes de regressió multivariant, destaca el mètode de mínims quadrats parcials (*partial least squares*, PLS). Aquest mètode fa la regressió a partir de components que adopten un compromís entre l'explicació de la variància de X i la capacitat predictiva de Y (maximitzen la covariància entre els dos blocs de dades).

## Mètodes geostatístics

Els mètodes geostatístics consideren les variables de tipus ambiental com a funcions en les que se superposa una component aleatòria amb una component determinista de variació lenta. Com que visitar cadascuna de les localitzacions existents en una àrea d'estudi per tal de mesurar una alçada o bé la concentració d'una variable determinada és normalment difícil o bé massa car, una possible alternativa, és la mesura directa del fenomen estudiat en algunes localitzacions disperses de manera estratègica i la posterior predicció dels valors corresponents a la resta de localitzacions. Una de les principals aplicacions de la geostatística és en el camp de la cartografia automàtica, tant pel que fa a la predicció en determinades localitzacions dels valors de la variable estudiada (ja siguin puntuals o mitjans), com de la incertesa associada a aquests. S'han descrit altres aplicacions com són la simulació de processos ambientals, l'objectiu de la qual és analitzar la variabilitat espacial, o el disseny òptim de xarxes de control i presa de mostra.

Segons Matheron (1963) [14], pare de la geostatística en la seva forma actual, la geostatística és l'aplicació de la Teoria de les Variables Regionalitzades a l'estimació de processos o fenòmens geològics a l'espai. Aquesta teoria assumeix que la variació espacial d'una variable representada a localitzacions específiques, és estadísticament homogènia a través d'una superfície determinada. Així doncs, des del punt de vista matemàtic, una variable regionalitzada és simplement una funció  $z(s)$  que adopta un valor per a cada punt  $s$  a l'espai.

### 1.3.2. Metodologies deterministes

Les metodologies deterministes permeten la simulació de diferents processos basant-se en lleis físiques, químiques i biològiques. Un model determinista de la qualitat ambiental consta típicament d'una sèrie d'equacions diferencials que incorporen paràmetres o relacions matemàtiques teòriques, així com diversos coeficients mesurats, estimats o suposats, per tal de descriure el comportament de les pressions (agents que provoquen un impacte sobre el medi) i les relacions entre aquestes pressions i les característiques del medi. Un dels usos més típics de les metodologies deterministes és la quantificació de com un sistema modelat particular, per exemple una massa d'aigua, respon a pressions concretes, com ara a la presència de determinats compostos tòxics.

Les equacions utilitzades per construir el model, sovint simplifiquen la realitat de manera significativa però necessària. Aquest tipus de metodologies s'ha convertit en una eina útil per als organismes competents en matèria de gestió de l'aigua, ja que permeten la realització de simulacions mitjançant diferents escenaris. Partint d'un escenari de referència (generalment l'estat real de la conca), aquest es compara amb un escenari simulat (en unes condicions diferents, en molts casos un cop una mesura correctora ja ha estat implementada). Dita mesura pot consistir tant en una regulació introduïda per al control de la contaminació, com en el canvi d'una determinada pràctica que tingui lloc sobre l'àmbit de la conca estudiada. La comparació entre escenaris pot ser de gran ajuda en els processos de decisió. L'anàlisi d'escenaris també pot anar associada a un model de tipus predictiu que permeti la descripció del transport de substàncies a través de la xarxa hidrogràfica.

Els models deterministes es poden dividir en 1) estacionaris i 2) dinàmics, depenent de si incorporen o no una dimensió temporal [10]. Els models estacionaris no inclouen una component temporal variable sinó que descriuen les condicions temporals mitjanes que es donen en el període d'estudi. D'altra banda, els models dinàmics presenten una dimensió de temps amb velocitats específiques pels diferents processos, donant lloc a sèries temporals. Els models estacionaris es poden dividir, al seu torn, segons quin sigui l'inici espacial pels seus càlculs, en: 1.1) models d'immissió, en els que el transport o concentració estimats d'una substància a la sortida de la conca es troben relacionats amb les característiques que es donen corrent amunt; i 1.2) els models d'emissió, els quals resumeixen els coeficients d'escolament i/o dades d'emissió empíriques per diferents classes de contribucions que es troben presents en una conca, per tal d'establir les condicions a la sortida. Els models estacionaris poden o no estar basats en els resultats d'un model hidrològic. D'altra banda, els models dinàmics de transport es basen generalment en models hidrològics, ja que el flux d'aigua constitueix el medi de transport, de manera que molta de la variabilitat en el transport de substàncies és un efecte

## Capítol 1

---

de la variabilitat hidrològica. Temporalment, els càlculs acostumen a ser repetits a intervals de temps horaris o bé diaris. Existeixen models basats en fenòmens puntuals que són capaços de simular, per exemple, el transport de substàncies durant un episodi aïllat de tempesta, mentre que n'existeixen d'altres de tipus continu que poden realitzar simulacions per determinades estacions, anys o fins i tot dècades. Espacialment, les simulacions poden anar des d'una fins a tres dimensions.

Degut al balanç que existeix entre la complexitat d'un model i el seu poder predictiu, existeixen dues categories de models dinàmics: 2.1) mecanicistes, o sigui, basats en lleis físiques, químiques i biològiques, i 2.2) conceptuals, en els quals només aquells processos més importants són descrits pel model o en els quals diversos processos s'agrupen sota una sola equació. El principal avantatge d'aquests darrers, és que requereixen un volum menor d'informació d'entrada així com un esforç computacional inferior, la qual cosa els fa adequats per a estudis a gran escala o per conques per les quals la informació de base disponible és limitada.

L'Agència de Protecció Ambiental dels Estats Units (*Environmental Protection Agency, EPA* [www.epa.gov](http://www.epa.gov)) treballa, conjuntament amb les autoritats estatals i locals, per tal de portar a terme projectes de seguiment i control de la qualitat del medi fluvial. El seu estudi es centra cada vegada més en la interacció que existeix entre els diferents compartiments ambientals dins d'una mateixa conca hidrogràfica. Aquest enfocament, permet una cada vegada major identificació dels canvis en la qualitat de l'aigua, el descobriment de nous problemes emergents i l'avaluació de diferents mesures de control de la contaminació, l'eficiència de la seva aplicació i la seva resposta davant de situacions d'emergència. Existeixen diversos models deterministes proposats per la EPA. Alguns d'aquests models, i altres, han estat utilitzats dins de l'àmbit de la Unió Europea.

Recentment, s'ha portat a terme el projecte *Benchmark Models for the Water Framework Directive (BWW)* per tal d'oferir consell als professionals del sector de l'aigua en la selecció dels models més adients per tractar una problemàtica ambiental determinada. A la literatura hi apareixen una gran quantitat de models deterministes amb noms diferents però que sovint utilitzen equacions bàsiques semblants, tot i que la seva interfase així com les rutines d'entrada i sortida, poden diferir considerablement. Existeixen molts models que no poden ser classificats específicament dins d'una categoria concreta. Sovint, alguns d'ells poden acoblar-se per a la seva implementació conjunta, de vegades sota un mateix nom i de vegades sota un nom diferent, la qual cosa pot donar lloc a confusió.

La major part d'inconvenients relacionats amb els models de tipus determinista recauen en la gran quantitat de dades d'entrada que es necessiten per al seu funcionament. Aquesta entrada de dades inicial requereix, per norma general, la inversió d'una quantitat molt important de

temps. A més, cal que l'usuari tingui un coneixement exhaustiu de l'àrea d'estudi i també molta experiència en l'aplicació del model per tal que aquest sigui utilitzat de manera eficient. Moltes vegades el model requereix l'entrada de valors de diversos paràmetres com són constants, coeficients o índexs. L'elecció dels valors correctes per aquests paràmetres requereix un coneixement i una experiència en el modelat de la qualitat de l'aigua que de vegades l'usuari no té i a més, desafortunadament, moltes vegades no es troben recomanacions per aquests paràmetres en els manuals del model. Quan un model ha estat dissenyat per ser aplicat en unes condicions determinades, el fet d'aplicar-lo en condicions diferents requerirà una modificació d'aquest, adaptant-lo a les noves característiques de l'àrea d'estudi. Finalment, el fet que les descripcions dels processos en un model determinista hagin de ser simplificades, és la causa d'incerteses en les prediccions que posteriorment donarà com a resultat el model.

Pel que fa a la modelització determinista de la qualitat de les aigües superficials dins de l'àmbit europeu, l'informe elaborat per Arheimer i Olsson (2003) [10] recull alguns dels models que s'han anat utilitzant, classificats segons el medi físic al qual està dirigida la seva aplicació. A continuació es presenta un recull no exhaustiu de diversos models deterministes que han estat aplicats amb la finalitat de modelitzar la dinàmica de diferents conques hidrogràfiques o dels compartiments ambientals que les constitueixen.

### 1. Models de conca hidrogràfica

<b>Model</b>	<b>Nom complet</b>	<b>Origen</b>	<b>Propòsit/substàncies modelitzades</b>
BASINS <sup>1</sup>	<i>Better Assessment Science Integrating point and Nonpoint Sources</i>	<i>European Protection Agency (EPA), 1996</i>	Sistema de control ambiental de les fonts de contaminació puntual i difosa a l'aigua
WCS <sup>2</sup>	<i>Watershed Characterisation System</i>	EPA, 2001	Variables físiques, variables de qualitat de l'aigua i càrregues de diferents fonts (nutrients, productes fitosanitaris, sediments, etc)
LSPC <sup>3</sup>	<i>Loading Simulation Program C++</i>	EPA, 2001	Cabal, transport de sediments i qualitat de l'aigua
AGNPS <sup>4</sup>	<i>AGricultural Non-Point Source pollution model</i>	<i>United States Department of Agriculture (USDA), 1987</i>	Nutrients i productes fitosanitaris
HBV-N <sup>5</sup>		<i>Swedish Meteorological and Hydrological Institute</i>	Control de l'eutrofització i transport de nitrogen



		(SMHI), 1994	
INCA <sup>6</sup>	<i>Integrated Nitrogen in Catchments</i>	Universitat de Reading, UK, 1998	Control de l'eutrofització i transport de nitrogen
MIKE <sup>7</sup> BASIN <sup>8</sup>		<i>Danish Hydraulic Institute</i> (DHI), anys 90	Control de l'eutrofització, transport de nitrogen i contaminants
SHETRAN <sup>9</sup>		Universitat de Newcastle, 1996	Control de contaminants i transport de nitrogen i sediments
SWAT <sup>10</sup>	<i>Soil and Water Assessment Tool</i>	USDA, 1993	Control de l'eutrofització, productes fitosanitaris, sediments i nutrients
GIBSI <sup>11</sup>	<i>Gestion Intégrée des Bassins versants à l'aide d'un Système Informatisé</i>	<i>Institut National de la Recherche Scientifique</i> (INRS), 1998	Nutrients i productes fitosanitaris, contaminació urbana (EDAR) <sup>a</sup>

<sup>1</sup>Whittemore 1998. Water Environ Technol 10:57-61

<sup>2</sup>USEPA 2001. Watershed Characterization System data, UST Tech, Region IV, Georgia

<sup>3</sup>Shen et al. 2005. J Environ Sci Health A Tox Hazard Subst Environ Eng 40:1791-1807

<sup>4</sup>Young et al. 1989. J Soil Water Conserv 44(2):168-173

<sup>5</sup>Arheimer and Wittgren 1994. Ambio 23(6):378-386

<sup>6</sup>Whitehead et al. 1998. Sci Total Environ 210/211:547-583

<sup>7</sup>Xevi et al. 1997. Water Resour Mahage 11:219-242

<sup>8</sup>Jha and Das Gupta 2003. Water Int 28:27-35

<sup>9</sup>Lunn et al. 1996. J Hydrol 174:285-304

<sup>10</sup>Arnold et al. 1998. J Am Water Res Assoc 34:73-89

<sup>11</sup>Villeneuve et al. 1998. Rapport du Projet GIBSI, Tome I R-462, INRS-Eau

<sup>a</sup>estació depuradora d'aigües residuals

**Taula 1.2.** Models deterministes de qualitat a escala de conca hidrogràfica

### **BASINS (Better Assessment Science Integrating Point and Nonpoint Sources)**

Model que té com a objectius (i) caracteritzar la qualitat de les aigües, (ii) identificar les fonts de contaminació existents en el medi i (iii) avaluar i localitzar aquelles masses d'aigua que no compleixen els estàndards de qualitat definits, per a una posterior discussió de les alternatives de gestió a aplicar. Combina un sistema d'informació geogràfica (SIG), informació diversa sobre les conques hidrogràfiques i eines d'avaluació i modelització ambiental ([www.epa.gov/athens/wwqtsc/html/basins.html](http://www.epa.gov/athens/wwqtsc/html/basins.html)).

### **WCS (Watershed Characterization System)**

Eina que permet realitzar una caracterització de les conques hidrogràfiques. Efectua una avaluació de la qualitat de l'aigua i de les fonts de contaminació potencials per les quals pot ser afectada. Com a resultat de l'aplicació del model, es generen 3 tipus d'informes: un informe físic de la conca (amb la seva delimitació, usos del sòl, clima, cabal, fonts puntuals, recursos, etc.), un informe de la qualitat de l'aigua, i un informe de les càrregues de fonts puntuals,



nutrients, productes fitosanitaris, sediments, etc. ([www.epa.gov/athens/wwqtsc/html/WCS-toolbox.pdf](http://www.epa.gov/athens/wwqtsc/html/WCS-toolbox.pdf))

### **LSPC (Loading Simulation Program C<sup>++</sup>)**

Eina que permet la simulació de la hidrologia i del transport de contaminants en una conca hidrogràfica. És capaç de representar càrregues de contaminants provinents de fonts tant puntuals com difoses. Es tracta d'un dels pocs models a nivell de conca que simula processos que afecten al sòl i al flux d'aigua simultàniament ([www.epa.gov/athens/wwqtsc/html.lspc.html](http://www.epa.gov/athens/wwqtsc/html.lspc.html))

### **AGNPS (AGricultural Non-Point Source pollution model)**

Model desenvolupat per avaluar la qualitat de l'aigua quan aquesta es troba afectada per l'erosió del sòl a causa de les pràctiques agrícoles. El model consta de 3 components principals: (i) component hidrològic, que proporciona una predicció del volum d'escolament i del cabal màxim, (ii) component d'erosió del sòl, que inclou tant l'erosió com la sedimentació, i (iii) component de contaminació per nutrients, que analitza les concentracions de nitrogen, fòsfor i la demanda química d'oxigen, tant en el material erosionat com en el propi sediment. Model recomanat principalment per a la descripció d'episodis puntuals ([www.wsi.nrcs.usda.gov/products/w2q/h&h/tools\\_models/agnps/index.html](http://www.wsi.nrcs.usda.gov/products/w2q/h&h/tools_models/agnps/index.html))

### **HBV-N**

Model que simula la pèrdua de nitrogen i el transport d'aquest cap als sistemes d'aigües subterrànies, rius i llacs. Es tracta d'un model conceptual que descriu de manera simplificada la hidrologia de la conca, sense incloure una descripció física exacta dels diferents processos. Utilitza rutines separades per la simulació del nitrogen inorgànic i orgànic ([www.smhi.se/sgn0106/if/hydrologi/hbv\\_np.htm](http://www.smhi.se/sgn0106/if/hydrologi/hbv_np.htm))

### **INCA (Integrated Nitrogen in CAatchments)**

Model que considera les diferents fonts de nitrogen que existeixen a la conca hidrogràfica estudiada, i que processa aquesta informació, després de la interacció d'aquestes fonts amb els cultius, la vegetació semi-natural, els microbis i el sòl, per tal de predir la quantitat de nitrogen que anirà a parar al riu o corrent d'aigua que constitueixi el seu destí final. Basat en balanços de massa i reaccions cinètiques, el model simula els principals mecanismes que operen en el cicle del nitrogen, incloent la mineralització, la immobilització, la nitrificació i la desnitrificació. Si el nivell de nitrogen que arriba finalment al flux d'aigua és massa elevat (superior als 50 mg/L de nitrats establerts per la Unió Europea), diferents mesures poden ser adoptades per tal de reduir-lo ([www.rdg.ac.uk/INCA](http://www.rdg.ac.uk/INCA))

## Capítol 1

---

### **MIKE BASIN**

Model utilitzat per identificar masses d'aigua en risc, les fonts de contaminació més importants que les afecten, i el càlcul de les reduccions necessàries per tal d'acomplir els objectius de qualitat definits a la normativa vigent per aquestes masses d'aigua. El model ha estat utilitzat en diversos estudis i aplicat a una gran varietat de condicions climàtiques i humanes. ([www.dhigroup.com/Software/WaterResources/MIKEBASIN.aspx](http://www.dhigroup.com/Software/WaterResources/MIKEBASIN.aspx))

### **SHETRAN**

Eina útil per a l'estudi de l'impacte ambiental de l'erosió del sòl, la contaminació i els efectes dels canvis d'usos del sòl i climàtics. També utilitzat per a l'estudi dels recursos d'aigua superficial i subterrània i la seva gestió, així com per la modelització del transport de nitrogen i sediments ([www.ceg.ncl.ac.uk/shetran](http://www.ceg.ncl.ac.uk/shetran))

### **SWAT (Soil and Water Assessment Tool)**

Model capaç de predir l'impacte de diverses pràctiques de gestió de l'aigua, sediments, nutrients i pesticides en conques grans i complexes, amb una gran varietat d'usos del sòl, textures de sòl i diferents condicions de gestió al llarg d'extensos períodes de temps. Ideal per a la predicció de canvis en la gestió dels usos del sòl (clima i canvis en la vegetació, pràctiques agrícoles, gestió d'embassaments, transferència d'aigua, extracció d'aigües subterrànies, etc.) ([www.brc.tamus.edu/swat](http://www.brc.tamus.edu/swat))

### **GIBSI (Gestion Intégrée des Bassins versants à l'aide d'un Système Informatisé)**

Sistema dissenyat com a eina d'avaluació de l'impacte que els diferents usos del sòl i dels recursos hídrics tenen sobre la qualitat de l'aigua d'una conca hidrogràfica. Inclou diferents models físics de simulació (hidrològic, erosió del sòl, transport de nutrients i productes fitosanitaris, i qualitat de l'aigua), així com diferents models de gestió (fonts de contaminació puntual, fonts de contaminació difosa i gestió d'embassaments) ([www.inrs-ete.quebec.ca/activites/modeles/gibsi](http://www.inrs-ete.quebec.ca/activites/modeles/gibsi)). Per una descripció més detallada del sistema GIBSI, veure secció 2.2.3. En aquesta Tesi, el sistema és aplicat per a la modelització de la hidrologia i el transport de productes fitosanitaris a la conca hidrogràfica del riu Llobregat.

2. Models per aigües subterrànies

Model	Nom complet	Origen	Propòsit/substàncies modelitzades
ASM <sup>1</sup>	<i>Aquifer Simulation Model</i>	<i>Swiss Federal Institute of Technology (ETH), 1986</i>	Dispersió de la contaminació
MODFLOW <sup>2</sup>		<i>US Geological Survey (USGS), 1988</i>	Flux d'aigua subterrània i transport de soluts

<sup>1</sup> Kinzelbach 1986. Elsevier, Amsterdam

<sup>2</sup> Ashley 1994. J Inst Water Environ Manage 8(6):635-645

Taula 1.3. Models deterministes de qualitat per aigües subterrànies

**ASM (Aquatic Simulation Model)**

Modelització del transport de contaminants i del flux d'aigua subterrània en dues dimensions (horitzontal i vertical) ([www.baug.ethz.ch](http://www.baug.ethz.ch))

**MODFLOW**

Model utilitzat per a la simulació del flux d'aigua subterrània i dels contaminants dins de l'aquífer (aquífer confinat, no confinat o bé una combinació d'ambdues opcions) ([www.modflow.com](http://www.modflow.com))

3. Models per aigües superficials

Model	Nom complet	Origen	Propòsit/substàncies modelitzades
WASP <sup>1</sup>	<i>Water quality Analysis Simulation Program</i>	EPA, anys 70	Paràmetres d'eutrofització, substàncies químiques orgàniques, mercuri
MIKE11 <sup>2</sup>		DHI, 1999	Qualitat de l'aigua i transport de sediments
PC-QUASAR <sup>3</sup>		<i>Center for Ecology and Hydrology, UK, 1997</i>	Cabal i qualitat de l'aigua
QUAL2E QUAL2K <sup>4</sup>	<i>River and Stream Water Quality Model</i>	EPA, 1987	Cabal i qualitat de l'aigua

<sup>1</sup> Nikolaidis et al. 2006. J Mar Sys 60:51-62

<sup>2</sup> Hanley et al. 1998. J Environ Manage 52(3):211-225

<sup>3</sup> Lewis et al 1997. Sci Total Environ 194/195:399-418

<sup>4</sup> Azzellino et al. 2006. Sci Total Environ 371:214-222

Taula 1.4. Models deterministes de qualitat per aigües superficials

### **WASP (Water quality Analysis Simulation Program)**

Programa per a la modelització dels diferents compartiments ambientals que conformen els ecosistemes aquàtics, incloent la columna d'aigua i el bènctic subjacent. D'una banda, el model simula els paràmetres convencionals que causen l'eutrofització, mentre que de l'altra, simula també la dinàmica dels contaminants tòxics (substàncies químiques orgàniques i mercuri). El principal avantatge és la flexibilitat de la seva aplicació a gairebé qualsevol tipus de massa d'aigua, així com a molts dels problemes més convencionals pels quals pot estar afectada la qualitat d'aquesta aigua ([www.epa.gov/athens/wwqtsc/html/wasp.html](http://www.epa.gov/athens/wwqtsc/html/wasp.html))

### **MIKE11**

Es tracta d'un paquet de software dirigit principalment al camp de l'enginyeria, per a la simulació del cabal, la qualitat de l'aigua i el transport de sediments en estuaris, rius, sistemes d'irrigació, canals i altres masses d'aigua. Conté un mòdul hidrodinàmic que s'utilitza per al càlcul del cabal en rius i estuaris; un mòdul d'advecció-dispersió per simular el transport dels materials dissolts i, finalment, un mòdul de qualitat de l'aigua que, juntament amb el mòdul d'advecció-dispersió, simula els processos de reacció que existeixen en els sistemes multicomponent, incloent la degradació de la matèria orgànica, la fotosíntesi i respiració de les plantes, la nitrificació i l'intercanvi d'oxigen amb l'atmosfera. Addicionalment, el software consta d'un mòdul dedicat a l'estudi del transport de sediments i de les condicions morfològiques dels rius ([www.dhigroup.com/Software/WaterResources/MIKE11.aspx](http://www.dhigroup.com/Software/WaterResources/MIKE11.aspx))

### **PC-QUASAR**

Model que descriu els canvis en la qualitat de l'aigua i el cabal a la xarxa fluvial al llarg d'un període de temps determinat. Proporciona un mètode de comparació entre l'estat actual del riu i el que existiria després d'un canvi, planificat o no, a la xarxa fluvial. Quan s'utilitza amb finalitats de planificació, el model és capaç de proporcionar distribucions del cabal i dades de la qualitat de l'aigua en punts determinats (punts clau), permetent d'aquesta manera, fixar els nivells màxims d'efluent per tal de mantenir una bona qualitat ecològica del riu. Poden ser modelats els següents paràmetres de la massa d'aigua: cabal, amoníac, pH, nitrats, temperatura, E.coli, DBO, oxigen dissolt i múltiples contaminants ([www.ceh.ac.uk/products/software/CEHSoftware-PC-QUASAR\\_000.htm](http://www.ceh.ac.uk/products/software/CEHSoftware-PC-QUASAR_000.htm))

### **QUAL2E/QUAL2K (River and Stream Water Quality Model)**

QUAL2K és la versió moderna del conegut model QUAL2E. Es tracta d'un model continu aplicat a contaminants convencionals en els cursos fluvials. Pot ser utilitzat per avaluar l'impacte provocat per abocaments residuals puntuals sobre la qualitat de l'aigua així com per a la identificació de les característiques quantitatives i qualitatives dels abocaments de tipus

difós. El model ha estat àmpliament utilitzat per a la simulació de la concentració d'oxigen dissolt en els rius. A més, també pot ser aplicat per descriure les principals reaccions que existeixen en els cicles de nutrients, la generació d'algues, la demanda de carboni, i el reaireig atmosfèric, així com els efectes d'aquestes reaccions sobre el balanç d'oxigen dissolt ([www.epa.gov/athens/wwqtsc/html/qual2k.html](http://www.epa.gov/athens/wwqtsc/html/qual2k.html))

#### 4. Models per a l'escolament urbà

Model	Nom complet	Origen	Propòsit/substàncies modelitzades
MOUSE <sup>1</sup>	<i>Modelling Of Urban SEwers</i>	DHI, anys 80	Qualitat de l'aigua i transport de sediments per sistemes de drenatge urbà, desaignes urbans i sanitaris
SWMM <sup>2</sup>	<i>Storm Water Management Model</i>	EPA, anys 70	Tots els aspectes del cicle hidrològic urbà, incloent les precipitacions, l'escolament superficial i subsuperficial, recorregut de l'aigua a través de la xarxa de drenatge, emmagatzemament i tractament

<sup>1</sup> Jensen and Linde-Jensen 1992. Water Sci Technol 25(9):101-113

<sup>2</sup> Peterson and Wicks 2006. J Hydrol 329:294-305

Taula 1.5. Models deterministes d'escolament urbà

#### **MOUSE (Modelling Of Urban SEwers)**

Sistema per a la modelització global de l'escolament superficial en àrees urbanes, el cabal en els canals oberts, el cabal dins de les canonades, la qualitat de l'aigua i el transport de sediments en els sistemes de drenatge urbà, els sistemes receptors de l'aigua de pluja i els receptors dels efluents municipals ([www.dhi.es/Software/Urbano/MOUSE.aspx](http://www.dhi.es/Software/Urbano/MOUSE.aspx))

#### **SWMM (Storm Water Management Model)**

Model global per a l'anàlisi de problemes relacionats amb la quantitat i la qualitat de l'escolament superficial. És capaç de simular tots els aspectes dels cicles urbans hidrològics i de qualitat de l'aigua, com són les pluges, el desgel, l'escolament superficial i subsuperficial, el flux d'aigua a través de la xarxa de drenatge, l'emmagatzemament i el tractament. Permet avaluar el problema de l'escolament superficial i buscar possibles solucions per a la seva reducció. D'altra banda, conté també un bloc estadístic que pot ser utilitzat per a l'anàlisi de freqüències d'hidrogrames i pol-lutogrames de llargues sèries temporals, i per a la identificació d'episodis hidrològics individuals que puguin ser d'especial interès. S'ha observat un bon funcionament del model quan aquest s'aplica a àrees urbanitzades sense drenatge, tot i que aquest també ha estat àmpliament aplicat sobre altres zones amb característiques diferents (<http://cce.oregonstate.edu/water/swmm>)

### 1.3.3. Els SIG com a eina d'anàlisi ambiental

En els darrers anys, l'activitat en el camp dels sistemes d'informació geogràfica [15-17] (SIG, o en anglès GIS, per *Geographical Information Systems*), ha evolucionat de la pura recerca universitària a una gran popularització, fins al punt que SIGs senzills ja han aparegut sota la forma de guies urbanes de diferents ciutats i municipis o d'instruments d'ajuda a la circulació automobilística.

Segons la definició de Burrough i McDonnell (1998) [15], un SIG és un conjunt d'eines per a la captació, emmagatzematge, recuperació, transformació i visualització de dades espacials del món real per a unes finalitats determinades. La representació del món real es fa en base a les coordenades geogràfiques, els atributs i la topologia dels elements representats.

Un SIG està compost per tres elements bàsics. El primer és el *hardware* sobre el qual descansa el sistema i que determinarà les capacitats de processament de què es disposa. El segon és el *software* amb què es manegen les dades, que determinarà les capacitats funcionals amb què l'usuari es pot adreçar a aquestes dades. I finalment, hi ha les dades en si, tant si són de tipus purament espacial (per exemple, els límits d'una conca hidrogràfica) com si són dades alfanumèriques de naturalesa geogràfica (per exemple, el cabal mesurat a cada tram d'una xarxa fluvial).

Més enllà de la definició relativa a l'estructura i funcionalitat d'un SIG, la característica més distintiva d'aquest és la seva capacitat d'interrelacionar o integrar diferents conjunts d'informació sobre fenòmens o localitzacions en virtut de la seva localització i relacions en l'espai. És habitual que un SIG contingui diferents capes d'informació relatives a un determinat paràmetre (altitud, usos del sòl, concentració de contaminants, tipus de textura del sòl, direcció del flux d'aigua, etc.), i que es consulti aquesta informació ja sigui de manera individualitzada, ja sigui per encreuament o superposició entre capes (Figura 1.4). Una possible consulta podria ser "quines textures del sòl emmagatzemen generalment una concentració més elevada de contaminants?".

Els principals tipus de consultes que es poden realitzar en un SIG corresponen a les anomenades "consulta per localització" i "consulta per atributs". A mode d'exemple, una consulta per localització podria ser "a quina alçada es troba una coordenada concreta?". En canvi, com a exemples d'una consulta per atributs es podria indicar "mostrar totes les zones on l'ús del sòl sigui industrial". No obstant, el tipus de qüestions plantejables en un SIG permet altres possibilitats, com ara la consulta de relacions espacials de proximitat, "a quina distància dels sòls de tipus industrial han estat detectades les concentracions més elevades d'un determinat contaminant?", la consulta de relacions espacials de veïnatge, "es troben en contacte les estacions depuradores d'aigües residuals i les concentracions més elevades de

carboni orgànic dissolt al riu?”, o la consulta de relacions espacials de contenció, “les mostres més contaminades es troben dins de l'àmbit d'una determinada comarca?”.

També es poden demanar les característiques espacials (àrea, perímetre, forma, etc.) d'un fenomen, o bé fer consultes sobre quina és la ruta més curta sobre una xarxa de corredors biològics. A més, si es disposa d'informació relativa a dos moments històrics diferents, es poden abordar qüestions sobre l'evolució que han patit diversos atributs en el temps: “en quina proporció ha disminuït la superfície d'irrigació durant els darrers 10 anys?”.

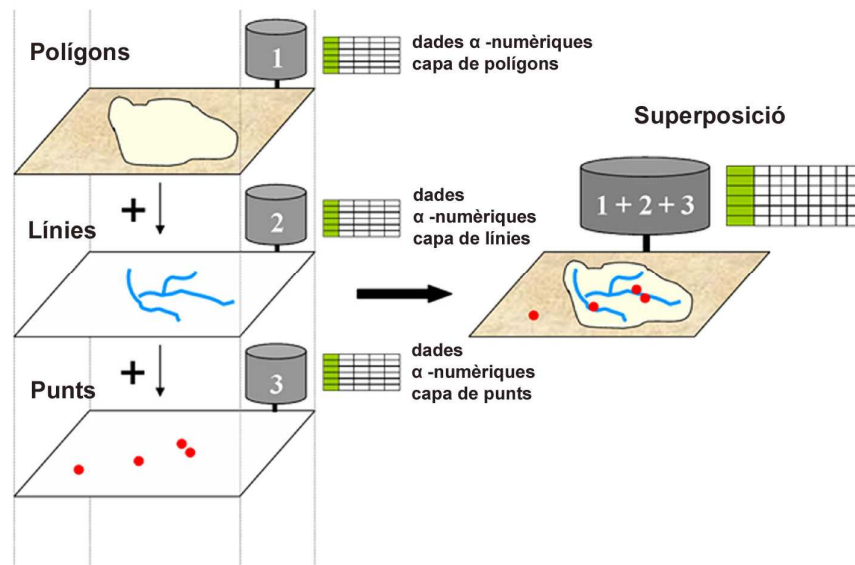


Figura 1.4. Superposició de capes de diferent geometria en un SIG

Fonamentalment, existeixen dues maneres de representar la informació digital en un SIG: els sistemes vectorials i els sistemes ràster o matricials (Figura 1.5). Un sistema vectorial permet representar les entitats espacials mitjançant punts, línies o polígons a base d'emmagatzemar les coordenades dels punts o vèrtexs. Aquests objectes gràfics poden portar associada informació diversa. Per exemple, a la Figura 1.4, els vectors tipus polígon representen el contorn de les diferents conques hidrogràfiques, els vectors tipus línia, el recorregut de la xarxa fluvial, i els vectors tipus punt la localització de les diferents mostres analitzades. Cadascun d'aquests objectes gràfics està associat amb una base de dades alfanumèrica que conté informació addicional (Figura 1.6). La utilitat d'aquest format es basa en la possibilitat de realitzar consultes de la base de dades, però amb l'interès afegit d'obtenir els resultats representats a l'espai.

El sistema matricial o ràster representa la realitat mitjançant cel·les o píxels (unitats elementals d'informació del territori). Aquestes cel·les contenen valors numèrics però, a diferència d'una matriu numèrica en el sentit matemàtic, un ràster està georreferenciat (té coordenades que tenen equivalència dins un determinat sistema de projecció).

## Capítol 1

Pel fet de ser matrius numèriques, els ràsters també permeten que es puguin fer operacions matemàtiques i lògiques entre diverses capes d'informació. Des de sempre, hi ha hagut una gran preocupació sobre quin model de dades és el millor per representar l'espai geogràfic. Tanmateix, actualment se sol acceptar la necessitat que els programes donin suport a les dues aproximacions, ja que depenent del tipus de dades o aplicacions, serà més convenient adoptar-ne una o altra.

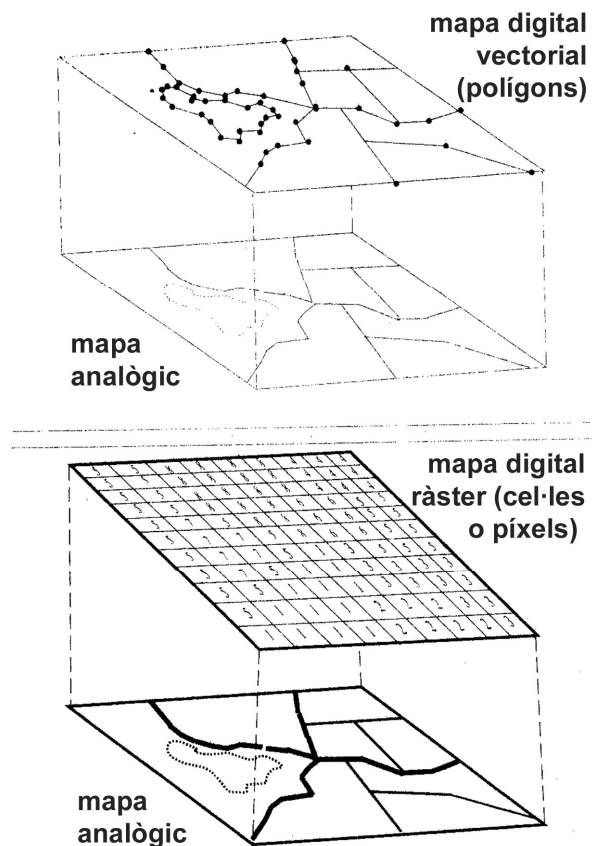


Figura 1.5. Representació ràster i vectorial de la realitat (Comas i Ruiz, 1993)

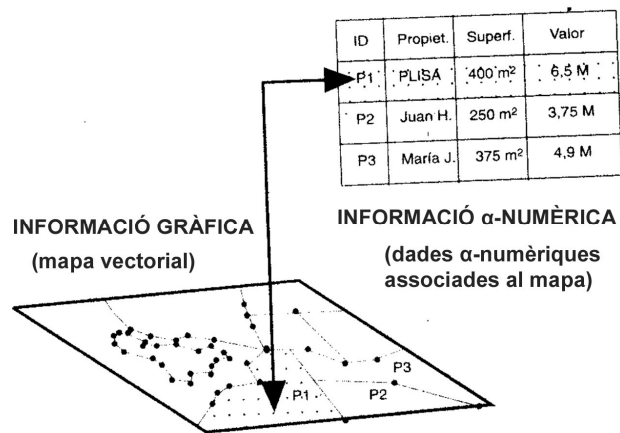


Figura 1.6. Unió de dades α-numèriques i cartogràfiques (Comas i Ruiz, 1993)



Tot i que la majoria de SIGs són bidimensionals, molts d'ells permeten el tractament del relleu a través d'aproximacions espacials. La més utilitzada és el model digital d'elevacions (MDE). Un MDE sol ser un fitxer ràster que recull, per al centre de cada cel·la, la seva altitud. Donada la naturalesa de la distribució de les dades altimètriques, l'altitud en qualsevol altre punt que no correspongui al centre d'una cel·la es pot obtenir utilitzant tècniques espacials d'interpolació. Els MDE són de gran utilitat ja que, a més de facilitar la informació altimètrica, permeten obtenir informació derivada, així com superposar-hi altres capes d'informació per tal de generar representacions tridimensionals. Els sistemes estrictament vectorials solen adoptar un model de representació que simplifica el terreny en zones triangulars irregulars de pendent i exposició similars, on la densitat de triangles és més alta en les zones de relleu complex. Aquest model s'anomena TIN (*Triangulated Irregular Network*).

La utilització dels SIGs no suposa cap disjunció respecte de l'ús dels mapes convencionals (en paper), sinó més aviat un enriquiment. Algunes de les limitacions del mapa clàssic com a sistema d'informació, com són la reducció del volum d'informació per imperatius de representabilitat i legibilitat (amb enorme dependència de l'escala escollida), la dificultat d'extreure i combinar informació entre mapes de diferent temàtica, de diferent moment temporal i de diferent full, i la dificultat d'actualització de la informació continguda en el mapa, es veuen superades gràcies als avenços de la cartografia digital. El suport informàtic dels SIGs permet una major facilitat de manteniment i recuperació de dades, l'anàlisi de canvis temporals, així com una llarga sèrie de funcionalitats que faciliten tant l'anàlisi científica i tècnica com la representació cartogràfica tradicional. A més, cal destacar que la presa de decisions utilitzant els SIGs facilita la feina i sol fer-la molt menys propensa a errors.

El punt fort dels sistemes SIG es que són capaços de tractar el que en principi semblen tipus molt diferents de dades, integrant informació que pertany a diferents àmbits mitjançant una característica comuna: la component geogràfica. La seva utilització implica la necessitat d'abstraccions o assumpcions per tal de representar la realitat geogràfica, ja que la informació ha de ser susceptible de ser processada per un ordinador.

Tot i les potencialitats dels SIGs, existeixen també alguns inconvenients que caldrà tenir en compte de cara a la seva utilització. I es que pot arribar a ser una perillosa forma de treballar allunyada de la realitat, en casos en què no es recorri al treball de camp perquè es té una inconscient fe cega en les dades que subministra l'ordinador. Així, el professional del medi natural ha de veure la informació que proporciona un SIG com una eina complementària al seu treball, potencialment tant subjecta a errors com qualsevol altre producte humà [16].

### 1.4. Justificació de la tesi

En els darrers anys, han anat proliferant un important nombre de treballs científics dedicats a la descripció de les problemàtiques que afecten la qualitat ambiental de les conques hidrogràfiques. Els entorns fluvials han patit una important degradació en tots els nivells, anant des de la contaminació química d'aigües superficials i subterrànies, sediments i sòls, a la modificació de les condicions hidromorfològiques, detracció de cabals, etc., amb la consegüent afectació que aquestes alteracions tenen sobre l'ecosistema de la conca.

Aquesta Tesi ha estat elaborada amb una intenció principalment metodològica. No té la finalitat d'abordar l'estudi específic d'una problemàtica concreta, sinó de proporcionar un conjunt d'eines adequades per a l'estudi de diverses problemàtiques de caràcter ambiental, sovint enteses des de diferents disciplines científiques. Aquestes eines són:

- ❖ Els mètodes estadístics i quimiomètrics per a l'extracció d'informació rellevant a partir de l'anàlisi de taules de dades ambientals. Aquests mètodes permeten la identificació de les principals fonts de contaminació que afecten una determinada regió, tant pel que fa a la composició química d'aquestes fonts com a la seva distribució espacial i temporal. Els mètodes estadístics s'utilitzen en totes les ciències experimentals. En particular, han adquirit força importància en l'àmbit de la química analítica.
- ❖ Els mètodes deterministes per a la modelització de les dinàmiques existents dins d'una determinada conca hidrogràfica. Aquests han estat utilitzats per a modelar la hidrologia de la conca, així com el transport de contaminants químics provinents de fonts de contaminació a través de la xarxa fluvial. Els mètodes deterministes s'utilitzen en diversos camps de les ciències ambientals i de les ciències de la terra, de la geologia, la biologia, l'ecologia, la química, la física, etc.
- ❖ Els sistemes d'informació geogràfica (SIGs) es proposen com a eines de representació de dades i de resultats de la seva anàlisi, ja que a més de la informació pròpiament continguda a les taules de dades, permeten la incorporació de la seva component espacial mitjançant l'addició de coordenades geogràfiques. Dins d'aquest apartat de representació espacial de dades, se situen també els mètodes d'interpolació geostatística, que permeten l'estimació de determinats atributs en localitzacions on aquests no han estat mesurats. L'aproximació geostatística s'acostuma a associar als camps de la geologia i la mineria.

- ❖ El càlcul d'índexs de qualitat per a la caracterització d'una determinada massa d'aigua, es proposa com un mètode per tal de resumir, de manera concisa, la informació més rellevant provinent dels programes científics de seguiment ambiental i transmetre-la de manera intel·ligible tant als organismes gestors com al públic en general. En aquest sentit, s'ha posat especial atenció en l'aprofitament d'aquelles dades procedents de les xarxes automàtiques de presa de mostra.

### 1.5. Objectius

Dins de cadascun dels tipus de metodologies aplicades en aquesta Tesi, s'han desenvolupat objectius concrets:

#### ❖ MÈTODES ESTADÍSTICS I QUIMIOMÈTRICS

1- Aplicació de mètodes estadístics i quimiomètrics a l'anàlisi de taules de dades ambientals per a la identificació de les principals fonts de contaminació així com de les seves tendències espacials i temporals:

- i) Estudi de la contaminació històrica i actual a la conca del riu Ebre a partir de dades de la *Confederación Hidrográfica del Ebro* (CHE, dades històriques) i del Projecte Europeu AquaTerra (dades recents)
- ii) Estudi de les fonts de contaminació presents a la conca del riu Llobregat
- iii) Estudi de l'impacte del cultiu de l'arròs sobre la qualitat de l'aigua dels canals d'irrigació i drenatge del delta del riu Ebre

2- Validació del mètode de resolució multivariant de corbes (MCR-ALS) per a la investigació de fonts de contaminació en aigües superficials

3- Estudi de l'efecte que diversos factors relacionats amb el procediment d'anàlisi de dades (com el mètode de pretractament de les dades, el soroll, l'ambigüitat rotacional, l'estructura de les dades o les restriccions aplicades durant el procés d'anàlisi) tenen sobre la recuperació de les fonts de contaminació

#### ❖ MÈTODES DETERMINÍSTICS

1- Calibratge del model GIBSI (hidrologia, erosió i transport de contaminants) per a la seva aplicació a la conca hidrogràfica del riu Llobregat

2- Modelització de la hidrologia de la conca del riu Llobregat

3- Simulació del transport i distribució de contaminants procedents de fonts agrícoles (aplicats sobre els sòls de l'àmbit de la conca hidrogràfica) a través de la xarxa fluvial del riu Llobregat

#### ❖ SIG I GEOSTADÍSTICA

1- Representació de dades espacials (àmbit de la conca d'estudi, estacions de presa de mostra, xarxa hidrogràfica, usos del sòl, etc.), mitjançant la cartografia digital

- 2- Representació contínua de la distribució de la contaminació sobre l'àrea d'estudi a partir de l'interpolació geostatística utilitzant el mètode *kriging*
- 3- Representació gràfica de resultats obtinguts a partir de l'anàlisi multivariant de dades

### ❖ INDEXS DE QUALITAT DE L'AIGUA

- 1- Estudi dels principals tipus d'índexs existents a la literatura utilitzats per a l'avaluació de la qualitat de les masses d'aigua. Establiment dels requisits per tal d'obtenir una bona categorització de la qualitat. Comparació de diferents índexs.
- 2- Validació de l'índex de Canadà (CCME WQI) per a l'anàlisi de dades fisicoquímiques procedents de xarxes automàtiques de presa de mostra. Aplicació de l'índex sobre diferents escenaris ambientals
- 3- Adaptació del procediment *combined monitoring-based and modelling-based priority setting scheme* (COMMPS) a les conques hidrogràfiques de Catalunya. Càlcul de l'índex prioritari per les diferents substàncies que presenten un risc per als ecosistemes aquàtics i per l'home via els ecosistemes aquàtics. Càlcul d'un índex prioritari de contaminació de localitzacions



## 2. METODOLOGIES APLICADES A L'ANÀLISI DE DADES AMBIENTALS

Les taules de dades ambientals s'obtenen generalment a partir d'estudis o programes de seguiment ambiental (monitoratge). Poden contenir paràmetres de diferent naturalesa (físics, químics, biològics, toxicològics, etc.), i les seves dades es caracteritzen per una gran variabilitat que prové de fonts molt diverses i, normalment, desconegudes. La variabilitat existent en les dades ambientals es pot assignar a tres tipus diferents de fonts o orígens [18]:

- (a) fonts naturals sense fortes influències de contaminació, les quals varien geogràficament i temporalment d'una manera estocàstica com a conseqüència dels fenòmens naturals;
- (b) fonts antropogèniques, que són causades per activitats humanes i interfereixen de manera significativa amb la variabilitat de tipus natural;
- (c) fonts de variabilitat produïdes per errors de tipus experimental i de mesura, causades a la vegada per les diferents etapes que conformen el procés analític: presa de mostra, pretractament i mesura d'aquesta i, finalment, pel procés d'anàlisi de les dades. La contribució de l'error experimental a la variabilitat de les dades augmentarà quan les concentracions determinades siguin baixes i també quan la magnitud de la mostra disminueixi.

### 2.1. Mètodes estadístics i quimiomètrics

Normalment, l'obtenció de mesures selectives de determinats compostos o de les seves fonts resulta difícil. Apart del soroll de fons, les dades experimentals es veuen afectades per interferències químiques i/o físiques procedents de la mostra o bé del procés de mesura utilitzat. En general, l'aplicació de tècniques d'anàlisi univariant de dades acostuma a no ser suficient i cada vegada existeix una major utilització de les tècniques d'anàlisi multivariant. Les dades de tipus univariant s'obtenen a partir de senyals analítiques mesurades com a funció d'una única variable de control (per exemple, la longitud d'ona, el temps, etc.). Òbviament, la quantitat d'informació que pot ser extreta a partir de senyals univariants és limitada. D'altra banda, els objectius de l'anàlisi ambiental acostumen a ser de tipus multivariant. Diversos patrons de contaminació, procedents de fonts diferents i sovint desconegudes, poden existir en el medi. Gran quantitat d'informació ambiental pot ser extreta d'un conjunt de dades multivariant si s'apliquen les eines d'anàlisi multivariant adients.

## 2.1.1. Mètodes univariants d'anàlisi de dades

Un primer anàlisi estadístic de les dades es pot dur a terme a partir de la utilització de:

- (a) Les eines de l'**estadística descriptiva** univariant que proporcionen diversos fulls de càlcul. En el cas del software EXCEL (Microsoft, Redmond, WA, USA), existeix l'opció *Anàlisi de dades* → *Estadística descriptiva*, mitjançant la qual es poden obtenir els següents paràmetres estadístics, agrupant les dades per variables i/o mostres: mínim, màxim, mitjana, mediana (valor central de la distribució), moda (valor que es repeteix més), variància (mesura de la dispersió de valors), desviació estàndard (arrel quadrada de la variància), suma i recompte del nombre total de mostres.
- (b) Els **diagrames de caixa**, una representació gràfica de les característiques principals d'una distribució de dades. En aquest tipus de diagrama, cada variable està representada per una caixa, la qual conté el 50% de les dades centrals de la distribució (veure Figura 2.1a). Els extrems inferior i superior de la caixa indiquen el 1er i 3er quartils, o el que és el mateix, el percentils 25% i 75%. Aquests indiquen les posicions a partir de les quals hi ha el 25% i 75% dels valors mesurats en la distribució després de ser ordenats de menor a major. El rang entre aquests dos quartils és el que es coneix com a rang interquartil. La línia a l'interior de la caixa indica el valor de la mediana, o sigui, el valor central de la distribució (valors ordenats de menor a major). El fet que la mediana no estigui localitzada en el centre de la caixa, indica que es tracta d'una distribució esbiaixada. Els límits superior i inferior de les línies de fora de la caixa (*whiskers*), serveixen per a identificar els valors atípics. Generalment se situen a 1.5 vegades la distància interquartil del 1er i 3er quartils (valor que se suma al 3er quartil i que es resta al 1er quartil). Els valors fora d'aquest rang i representats per signes positius (+) es consideren valors atípics o *outlayers*.

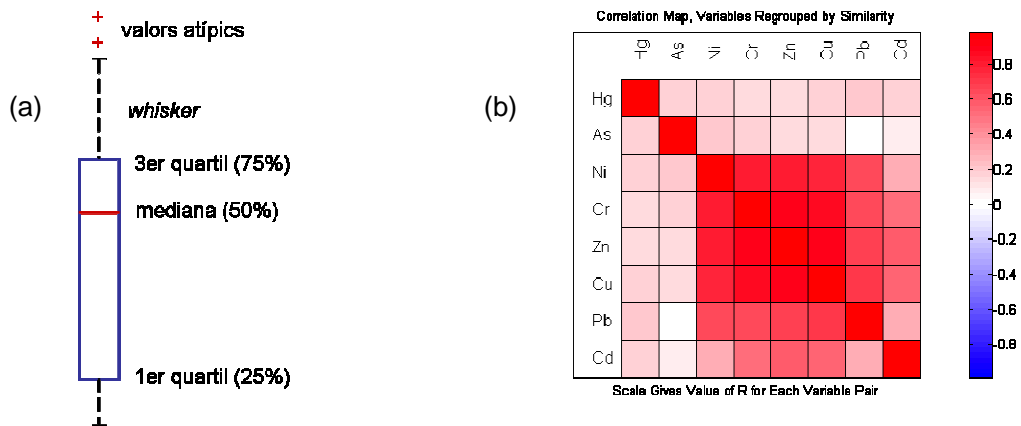


Figura 2.1. Esquema d'un diagrama de caixa (a) i mapa de correlació de variables (b)



- (c) **L'anàlisi de correlació** per parelles de variables a partir del càlcul dels coeficients de correlació entre les variables obtinguts de les dades experimentals. Aquesta anàlisi permet una primera investigació sobre la possible presència i distribució de fonts de contaminació comunes. En l'exemple de la Figura 2.1b, les variables Ni, Cr, Zn i Cu són les que presenten una correlació més elevada entre elles (una major similitud). En canvi, es veu com l'Hg, As i Cd varien de forma més independent i no es troben correlacionades amb la resta de variables analitzades.

### 2.1.2. Mètodes multivariants d'anàlisi de dades

Un dels objectius de la recerca ambiental actual és obtenir informació el més objectiva possible sobre els múltiples i complexos processos que tenen lloc en el medi ambient. El desenvolupament de mètodes quimiomètrics que ha tingut lloc durant les darreres dècades, ha aportat una eina poderosa per a la resolució d'aquest tipus de problemes.

La quimiometria ambiental és una disciplina relativament recent. El seu propòsit general consisteix en l'aplicació de mètodes estadístics i matemàtics amb l'objectiu de trobar una manera òptima de resoldre problemes químics ambientals i extreure la màxima informació possible de les dades experimentals. Així, l'aplicació de mètodes quimiomètrics en el camp de la química ambiental és necessari i útil degut a les següents raons:

- (a) Els processos naturals i antropogènics que ocorren en el medi ambient es troben principalment caracteritzats per canvis multivariable en les concentracions de múltiples compostos i/o contaminants químics en els diferents compartiments ambientals i en les seves propietats i paràmetres fisicoquímics i biològics,
- (b) Les emissions de tipus antropogènic s'acostumen a caracteritzar pel seu caràcter multicomponent o multielement. A més, aquestes substàncies poden reaccionar de manera sinèrgica o competitiva, és a dir, que les interaccions entre contaminants i entre aquests i els components naturals del medi seran complexes i també hauran de ser considerades,
- (c) Les dades experimentals obtingudes en estudis ambientals contenen informació sobre els diferents processos ambientals i també sobre la variància causada per l'error analític i altres incerteses experimentals. Un dels propòsits d'aplicació de les metodologies de tipus quimiomètric en l'àmbit ambiental serà la minimització o bé, en cas que sigui possible, l'eliminació d'aquestes incerteses experimentals i de l'error analític associat a les mesures ambientals,

(d) Els darrers anys han estat caracteritzats per un gran desenvolupament dels instruments de química analítica i del processament de la informació obtinguda a partir d'ells. Aquests canvis han permès un seguiment més detallat i precís del medi ambient i de les reaccions de substàncies químiques en diversos medis. La necessitat de gestionar l'enorme quantitat de dades obtingudes a partir dels estudis de seguiment i monitoratge ambiental és una altra de les principals raons de l'aplicació dels mètodes quimiomètrics. El principal objectiu es centra en l'extracció i simplificació de la informació rellevant pel que fa a l'estat de la contaminació, els canvis espacials i temporals d'aquesta i la identificació de les principals fonts. La modelització del comportament dels contaminants dins i entre els diferents compartiments ambientals, el seu impacte sobre l'ésser humà i l'avaluació dels riscos ambientals i toxicològics, també requereix l'aplicació de mètodes quimiomètrics.

En resum, la quimiometria aplicada als estudis mediambientals (quimiometria ambiental) és una eina que permet realitzar una investigació més rigorosa i objectiva dels diferents processos que ocorren en el medi ambient.

### 2.1.2.1. Estructura de les taules de dades ambientals

Les dades obtingudes a partir dels estudis de seguiment i monitoratge ambiental es poden classificar, segons la seva complexitat, en dades ordenades en una direcció (*one-way data*), en dues direccions (*two-way data*), en tres direccions (*three-way data*) i en múltiples direccions (*multiway data*) [19-20]. Tal com es mostra en la Figura 2.2, els valors numèrics escalars (per exemple, la mesura d'una única variable en una mostra) correspondrien a dades no direccionades o ordenades en zero direccions (*zero-way*), mentre que les dades vectorials (per exemple, mesures de diverses variables en una mostra) es troben ordenades en una direcció. Quan es mesuren diverses variables en diferents mostres, s'obtenen dades que es poden ordenar en dues direccions, és a dir, en una taula o matriu de dades. Finalment, la recopilació de varies matrius de dades o bé de dades ordenades en tres direccions, dóna lloc al que s'anomena un cub de dades.

Una matriu de dades constitueix l'estructura més comunament trobada en els estudis de seguiment ambiental. En aquestes taules o matrius de dades, les diferents mostres analitzades es col·loquen en les diferents files de la matriu, i les diferents variables mesurades (concentracions de compostos químics, paràmetres fisicoquímics, etc.) es col·loquen en les diferents columnes de la matriu. Les tècniques estadístiques necessàries per al processament

multivariant d'aquestes dades agrupades en forma de taula o matriu utilitzen les eines, formulacions i notacions de l'àlgebra lineal. Aquestes taules o matrius de dades ambientals seran representades per una lletra majúscula en negreta. A mode d'exemple,  $\mathbf{D}(i,j)$  seria la notació utilitzada per indicar una matriu de dades ambientals constituïda d'  $i$  files (per exemple, el nombre de mostres ambientals analitzades) i  $j$  columnes (per exemple, el nombre de paràmetres ambientals analitzats). Quan es passa de dades en una direcció a dades ordenades en tres direccions, la quantitat d'informació augmenta, però augmentarà també la complexitat de la seva anàlisi.

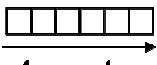
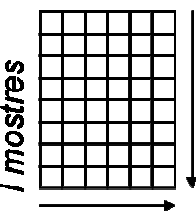
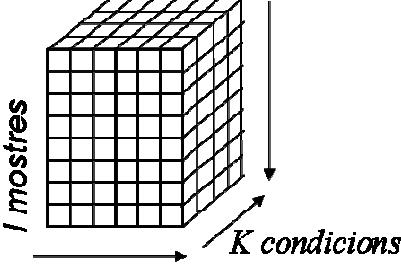
	<b>0 direccions</b>	<b>2 direccions</b>	<b>3 direccions</b>
<b>ESTRUCTURA</b>	<p><b>1 variable</b></p> <p>□</p> <p><b>1 mostra</b></p> <p><b>1 direcció</b></p> <p><b>J variables</b></p>  <p><b>1 mostra</b></p>	<p><b>J variables</b></p>  <p><b>i mostres</b></p>	<p><b>J variables</b></p>  <p><b>i mostres</b></p> <p><b>K condicions</b></p>
<b>MODEL</b>	<p><b>-Models lineals</b></p> <p><b>-Models no lineals</b></p>	<p><b>-Models bilineals</b></p> <p><b>-Models no bilineals</b></p>	<p><b>-Models trilineals</b></p> <p><b>-Models no trilineals</b></p>

Figura 2.2. Tipus d'estructures i models de dades ambientals

Depenent de l'estructura que presentin les dades a analitzar, es podran aplicar diferents tipus de models. Així, quan es tracta de dades en una direcció, els tipus de models utilitzats podran ser els models lineals (veure Equació 2.1) i no lineals (veure Equació 2.2) univariants. Per dades ordenades en dues direccions, es podran aplicar models bilineals (veure Equació 2.3) o models no bilineals. Finalment, per les dades ordenades en tres direccions, es podran aplicar models trilineals (si les dades presenten una estructura trilineal) (veure Equació 2.4) o bé models no trilineals.

$$y_i = b_0 + x_i \quad \text{Equació 2.1}$$

$$y_i = f_{\text{no lineal}}(x_i) \quad \text{Equació 2.2}$$

$$x_{ij} = \sum_{n=1}^N g_{in} f_{nj} + e_{ij} \quad \text{Equació 2.3}$$

$$x_{ijk} = \sum_{n=1}^N g_{in} f_{jn} z_{kn} + e_{ijk}$$

Equació 2.4

La descomposició de valors singulars (*Singular Value Decomposition*, SVD [21]) pot ajudar a esbrinar quina és l'estructura de les dades i el model més adequat per analitzar-les. Es dona més informació sobre els diferents tipus de models aplicats a les matrius de dades ambientals i sobre la SVD a l'apartat 2.1.2.5 i a l'article 1 de la present Tesi.

### 2.1.2.2. Tractament preliminar de les dades

Abans de poder aplicar l'anàlisi quimiomètrica, les taules de dades ambientals necessiten una preparació prèvia, que sovint acostuma a ser la part més laboriosa de tota l'anàlisi. Aquesta preparació comportarà una feina d'estructuració, homogeneïtzació i transformació de les dades, que dependrà, en cada cas, del format en què aquestes dades hagin estat inicialment obtingudes, del seu volum, etc.

Entre els problemes més comuns a resoldre, es troben:

(a) La presència de **valors buits** (*missing values*). Es tracta de valors que no es coneixen o que estan en blanc i on els compostos considerats no han estat o no han pogut ser analitzats. En el cas que el nombre de valors buits sigui petit i es trobin distribuïts d'una manera aleatòria dins de la taula de dades, aquests poden ser estimats numèricament tenint en compte les variacions observades a les variables (columnes de la matriu) i a les mostres (files de la matriu). L'estimació d'aquests valors es pot fer mitjançant mètodes quimiomètrics, com per exemple, l'anàlisi de components principals [22] o altres algorismes iteratius [23-25]. Un cop realitzada l'estimació dels valors buits, és aconsellable comprovar que les estimacions realitzades siguin plausibles i no discordants. S'ha de comprovar que els valors estimats per aquests procediments tinguin sentit des d'un punt de vista ambiental (per exemple, que no s'obtinguin valors negatius en el cas de variables que només poden ser positives) o bé que la magnitud no s'aparti excessivament d'allò que seria d'esperar tenint en compte la variació observada per a la resta de mesures. En el cas que el valor que s'ha de predir fós en realitat un valor atípic, és evident que la seva estimació a partir d'aquests procediments no seria possible a no ser que correspongués a una variació sistemàtica prèviament detectada per a la resta de dades. En el cas d'aquesta Tesi, l'estimació dels valors buits s'ha fet mitjançant les funcions d'imputació per PCA de la PLS Toolbox (Eigenvector Research Ltd., Manson, WA, USA), dins l'entorn de programació i visualització del software MATLAB (The Mathworks, MA, USA). Aquest tipus d'estimació es realitzarà substituint inicialment els valors buits de la matriu de

dades per zeros i fent la seva anàlisi per PCA. Posteriorment, els zeros són substituïts pels nous valors predits pel model PCA per un nombre de components determinat i, amb aquests nous valors predits, es torna a calcular un nou model PCA. Aquest procés es va repetint fins que els valors buits estimats per PCA convergeixen (s'acaben obtenint sempre els mateixos valors). La selecció del nombre de components principals es fa de la forma usual, intentant explicar una proporció considerable de la variància de les dades, sense incloure el soroll experimental. Quan es fa una sobreestimació del nombre de components, s'acostumen a predir valors no raonables, com ara valors negatius en el cas de concentracions de compostos químics.

(b) Valors **sota el límit de detecció** ( $<LD$ ). Aquests valors se substitueixen per la meitat del límit de detecció ( $LD/2$ ) considerat, ja que s'ha comprovat que aquesta és una bona estimació, millor que la simple substitució per zeros o pel mateix valor del límit de detecció, a més d'implacar poc esforç computacional [26]. Altres propostes com la substitució d'aquests valors pels obtinguts directament a partir de les mesures instrumentals es consideren encara millors (dades no censurades). Tot i això, no sempre es disposa dels valors dels senyals mesurats experimentalment a partir dels instruments emprats en les anàlisis [12], ja sigui perquè el senyal instrumental de la substància no ha estat detectat, o bé pel fet que tot i detectar el senyal, aquest ha estat censurat per l'analista en considerar-lo poc fiable estadísticament.

### 2.1.2.3. Pretractament de les dades

Per tal de facilitar l'extracció d'informació ambiental útil de les dades experimentals, sovint es fa necessària l'aplicació de tècniques de preprocessament o pretractament de dades. La selecció de les tècniques de pretractament més adients dependrà de diversos factors inherents a cada conjunt de dades, com per exemple, de la manera com aquestes han estat obtingudes, del tipus d'informació ambiental que es busca, o del coneixement previ que es tingui sobre la seva naturalesa. No existeix un únic mètode de pretractament de dades que pugui ser generalitzat a totes les matrius de dades ambientals. L'elecció del mètode més adequat dependrà principalment del tipus de fenomen o problema ambiental que es pretén posar de manifest.

El fet que generalment les dades de tipus ambiental no portin associades les incerteses en la seva determinació, en dificulta un pretractament òptim. Aquest pretractament implicaria la utilització de mètodes de càlcul de sospesament de dades que tinguessin en compte aquests errors o incerteses, com es fa en el mètode anomenat factorització positiva de la matriu (*Positive Matrix Factorization*, PMF [27]) o bé, el desenvolupat més recentment, mètode de mínims quadrats alternats sospesats (*Weighted Alternating Least Squares*, WALs [28]).

## Capítol 2

---

Els mètodes de pretractament de dades que més s'han aplicat al llarg d'aquesta Tesi són l'escalat i l'autoescalat. Aquests mètodes, juntament amb altres possibles tipus de pretractament, es descriuen a continuació:

### Centrat amb la mitjana de les dades

El centrat consisteix en restar a cada valor de les variables originals el valor de la seva mitjana. D'aquesta manera, s'aconsegueix que cada variable de la nova matriu de dades (matriu centrada) quedi amb una mitjana igual a zero.

$$z_{ij} = x_{ij} - \bar{x}_j \quad i = 1, \dots, I \quad j = 1, \dots, J \quad \text{Equació 2.5}$$

$$\bar{x}_j = \frac{\sum_{i=1}^I x_{ij}}{I} \quad \text{Equació 2.6}$$

on  $z_{ij}$  és el nou valor centrat,  $x_{ij}$  és el valor original de la variable, i  $\bar{x}_j$  és la mitjana dels  $x_i$  valors corresponents a la variable  $j$ . El centrat amb la mitjana de les dades realitza una translació de l'origen de coordenades de les dades des de zero fins al valor de la mitjana de les dades. Així, es fa més fàcil visualitzar les variacions respecte al valor mitjà (veure Figura 2.3b). Es descarta, per tant, aquella informació de les variables que no varia i que es pot considerar constant a la matriu (*offsets*) ja que interessa emfatitzar les variacions. No obstant, cal remarcar que el centrat farà més difícil distingir l'aportació (*apportionment*) o assignació quantitativa proporcional de les diferents fonts de contaminació sobre les mostres. El centrat de les dades és un procediment recomanat per a mètodes quimiomètrics com l'anàlisi de components principals (PCA), però es desaconsella per als mètodes de resolució multivariant [29], el PMF [27] o el PARAFAC (*Parallel Factor Analysis* [30]), quan s'apliquen restriccions de no negativitat.

### Escalat

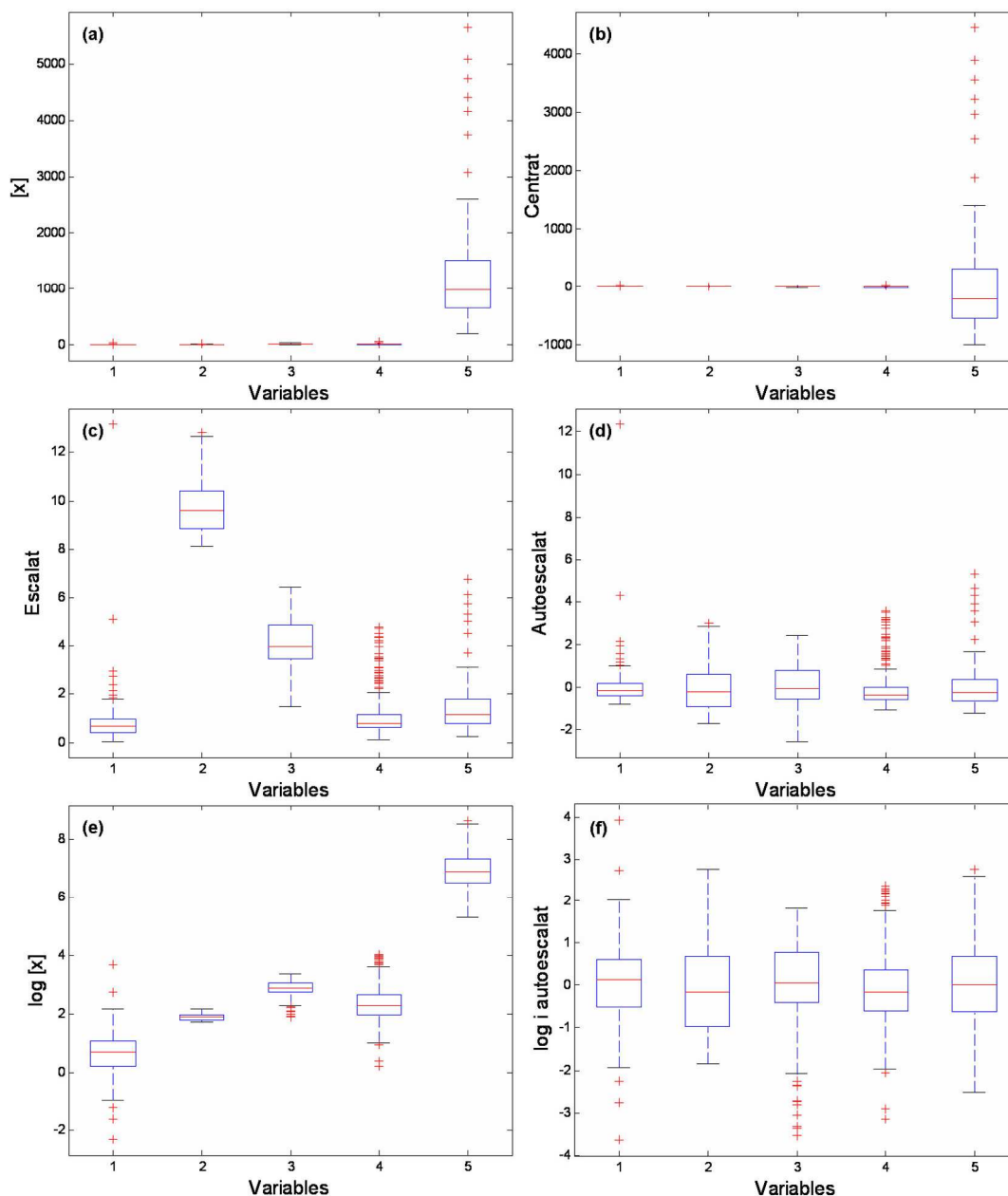
L'escalat consisteix en dividir cada valor de les variables per la seva desviació estàndard.

$$z_{ij} = \frac{x_{ij}}{s_j} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad \text{Equació 2.7}$$

on  $z_{ij}$  és el nou valor escalat,  $x_{ij}$  el valor original i  $s_j$  el valor de la desviació estàndard de la variable:

$$s_j = \sqrt{\frac{\sum_{i=1}^I (x_{ij} - \bar{x}_j)^2}{I-1}} \quad \text{Equació 2.8}$$

On  $\bar{x}_j$  és la mitjana dels valors de la variable  $j$ . L'escalat és una forma de pretractament de dades força utilitzada en el cas de conjunts de dades ambientals, sobretot quan aquestes han estat mesurades per variables, l'escala i les unitats de les quals varien. Les variables escalades que s'obtenen, presenten una distribució més homogènia, cosa que les fa més fàcilment comparables entre si (veure Figura 2.3c).



**Figura 2.3.** Diagrames de caixa d'un conjunt de dades ambientals (a) originals, (b) centrades, (c) escalades, (d) autoescalades, (e) amb transformació logarítmica i (f) amb transformació logarítmica i escalades

A més, la desviació estàndard de les variables un cop escalades és, per totes, igual a 1, de manera que s'iguali la seva importància relativa. No obstant, cal anar en compte en el cas de

## Capítol 2

---

tenir variables en les que tots els seus valors siguin molt baixos i constants en gairebé totes les mesures, ja que aleshores el seu escalat pot augmentar de manera artificial la seva importància relativa (perquè es divideixen per una desviació estàndard molt petita) quan en realitat no la tenen i, per tant, això pot portar a resultats i conclusions totalment errònies. En aquests casos, és millor eliminar directament aquestes variables o, en el cas que es consideri necessària la seva presència en l'anàlisi, donar-les-hi un pes petit (per exemple, dividint tots els valors de la variable per un valor gran, com ara 20, enlloc de dividir-los per la desviació estàndard de la variable).

L'escalat de les dades és un pretractament especialment útil quan les dades de variables diferents es troben expressades en unitats i magnituds diferents. En aquests casos, les variables amb valors més grans dominen de forma no desitjable la variància observada, i per tant, les variables amb els valors més baixos es troben emmascarades. En aquestes situacions és preferible escalar les dades per tal de disminuir l'efecte exagerat d'aquelles variables que presenten magnituds majors. En canvi, quan les variables analitzades tenen les mateixes unitats de mesura, no es veu tan clara la conveniència del seu escalat. Com ja s'ha esmentat, quan s'escala es dona la mateixa importància a totes els variables. Per exemple, existeixen casos en què certs compostos químics minoritaris tenen tanta o més importància que els majoritaris, de manera que amb concentracions més baixes, els seus efectes poden ser molt més greus, donant lloc a una contaminació més important. Totes aquestes consideracions caldrà tenir-les en compte a l'hora de decidir si s'aplica l'escalat o no de les dades, així com en decidir si aquest escalat s'aplica a totes o a només algunes de les variables.

En els casos en què es disposi de diverses matrius de dades on les variables analitzades coincideixen (per exemple, matrius d'anys diferents o bé corresponents a diferents compartiments ambientals), l'escalat pot ser aplicat sobre cadascuna de les matrius de dades individuals o conjuntament sobre la matriu formada a partir de totes les matrius de dades individuals (matriu augmentada, veure més endavant, secció 2.1.2.5) al mateix temps. L'elecció d'una opció o altra caldrà que es faci en funció de la finalitat de l'estudi. Per exemple, en aquells casos en què s'analitzin matrius obtingudes en diferents anys, caldria escalar-les conjuntament si el que interessa observar són les diferències temporals (entre anys diferents) dels components identificats. Si, en canvi, el que es pretén posar de manifest són les diferències que existeixen entre les localitzacions geogràfiques, atribuint una importància menor a les diferències entre anys, serà millor optar per l'escalat individual i per separat de cadascuna de les matrius de dades abans de la seva anàlisi conjunta. Tot això es discuteix amb més detall a l'article 4 d'aquesta Tesi.



Com ja s'ha esmentat anteriorment, les dades ambientals porten associades incerteses que no han estat determinades. Caldrà també anar en compte amb aquestes incerteses, ja que es poden veure amplificades de manera considerable a l'hora d'aplicar l'escalat.

### Autoescalat

L'autoescalat consisteix en l'aplicació simultània del centrat i de l'escalat:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad i = 1, \dots, I \quad j = 1, \dots, J \quad \text{Equació 2.9}$$

on  $z_{ij}$  és el nou valor autoescalat,  $x_{ij}$  el valor original,  $\bar{x}_j$  la mitjana dels valors de la variable (descrita a l'Equació 2.6), i  $s_j$  la desviació estàndard dels valors corresponents a la variable  $j$  (descrita a l'Equació 2.8). La distribució dels valors de les variables obtingudes en aplicar el tractament d'autoescalat és similar al cas de l'escalat, però al mateix temps les dades experimenten una translació del seu origen degut al centrat amb la mitjana de les dades (veure Figura 2.3d). L'autoescalat és el pretractament preferit en l'aplicació de mètodes com PCA, on l'interès fonamental recau en la descripció i investigació de la variància de les dades. No obstant, tal com ocorre en el cas del centrat, amb l'autoescalat apareixeran valors negatius a la nova matriu de dades autoescalada, la qual cosa impedirà l'aplicació dels procediments de resolució multivariant o d'altres quan s'apliquen restriccions de no negativitat (veure més endavant, secció 2.1.2.5).

### Transformació logarítmica

La transformació logarítmica de les dades actua, com l'escalat, augmentant el pes d'aquelles variables que presenten uns valors més baixos respecte a aquelles que presenten els valors més alts. És útil en aquells casos en què existeixen variables amb diferents magnituds, escales o unitats de mesura. Es tracta d'un procediment que sovint es troba recomanat en la literatura per conjunts esbiaixats de dades, com ara les dades d'estudis ambientals, on la major part dels valors de les variables són baixos i existeix un menor nombre de valors alts. Amb l'aplicació del logaritme, s'obté una distribució més simètrica de les dades (veure Figura 2.3e). No obstant, la transformació logarítmica també pot donar lloc a efectes indesitjats, ja que l'estructura interna de les dades pot quedar afectada i augmentar la seva possible no linealitat i el nombre de components necessari per explicar la seva variància. Es per això que es considera convenient la comparació del nombre de components obtinguts per PCA o SVD (descrits més avall) que es necessiten per explicar la mateixa quantitat de variància, abans i després de la seva transformació logarítmica. Posteriorment a la transformació logarítmica de les dades, és també possible aplicar addicionalment qualsevol dels altres pretractaments descrits anteriorment

(centrat, escalat o autoescalat (Figura 2.3f), obtenint diferents distribucions dels valors de les variables.

### 2.1.2.4. Anàlisi de components principals (PCA)

L'anàlisi de components principals (*Principal Component Analysis*, PCA) [22, 31] és un mètode d'estadística multivariant freqüentment utilitzat en l'anàlisi de taules de dades obtingudes en estudis de seguiment i monitoratge ambiental. Parteix de la hipòtesi que en el grup de dades original existeixen un conjunt reduït de factors o components dominants (fonts de variació) que influeixen de manera important sobre la variància observada en les dades, i que aquests factors o components no són mesurats directament (es tracta de factors no explícits o amagats, *hidden factors*), ja que no existeixen sensors específics per a ells o no són observables com a tals experimentalment.

Generalment, una part important de la informació continguda en les variables mesurades experimentalment és redundant (correlacionada amb les altres variables) o bé irrelevant. El PCA permet transformar aquest conjunt de variables en un nou conjunt de variables no correlacionades entre si, que és més senzill d'interpretar, fent més evident aquella informació subjacent que probablement és més rellevant. Aquestes noves variables, factors o components principals són doncs ortogonals. Això significa que el que explica un component, no ho explica l'altre. A més, en PCA, el primer component es troba en la direcció que expliqui el màxim de variància, i el mateix pel segon component per la variància residual, el tercer, i així successivament. La finalitat del PCA és trobar, a partir de la combinació lineal de les variables originals, un nou conjunt d'eixos ortogonals de coordenades sobre els que es poden projectar les dades originals i proporcionar una millor i més fàcil interpretació dels fenòmens o fonts que estan causant la variància observada en les dades ambientals mesurades (veure Figura 2.4).

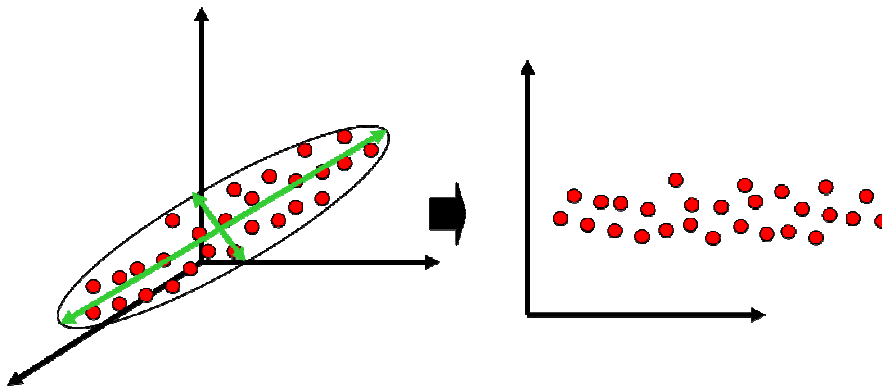


Figura 2.4. Nous eixos ortogonals creats a partir del conjunt de dades originals

El mètode PCA pretén resumir aquella informació rellevant que es troba continguda a la matriu de dades analitzada (separant-la de la part que correspon al soroll o error). Matemàticament, la matriu original de dades  $\mathbf{D}$  es descompon, seguint un model bilineal, en el producte de dues matrius ortogonals,  $\mathbf{X}$  i  $\mathbf{Y}^T$  (veure Figura 2.5)

$$\mathbf{D} = \mathbf{X} \mathbf{Y}^T + \mathbf{E} = \mathbf{D}^* + \mathbf{E} \quad \text{Equació 2.10}$$

On  $\mathbf{D}$  és la matriu de dades experimentals,  $\mathbf{X}$  la matriu d'*scores* (mapa de mostres),  $\mathbf{Y}^T$  la matriu de *loadings* (mapa de variables) i  $\mathbf{E}$  la matriu d'error (Equació 2.10). El producte de les matrius d'*scores* i de *loadings* reproduirà la matriu original de dades per un nombre determinat  $N$  de components (nombre de columnes en  $\mathbf{X}$  i de files en  $\mathbf{Y}^T$ ). Degut a què els components en PCA s'obtenen en ordre decreixent de la variància explicada, aquella informació més rellevant del conjunt de dades es troba sempre concentrada en els primers components principals generats. És per aquest motiu que els components posteriors podran ser omesos sense una pèrdua significativa de la variància o d'informació rellevant, ja que només explicaran una petita part de la variància la qual estarà relacionada amb el soroll i l'error experimental.

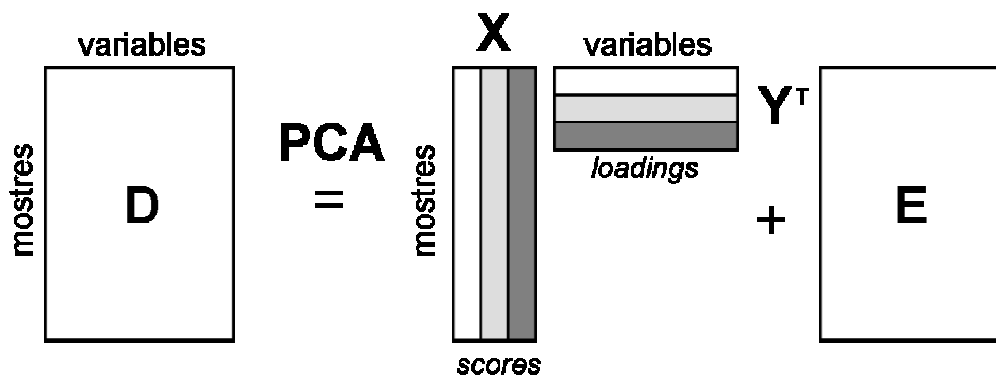
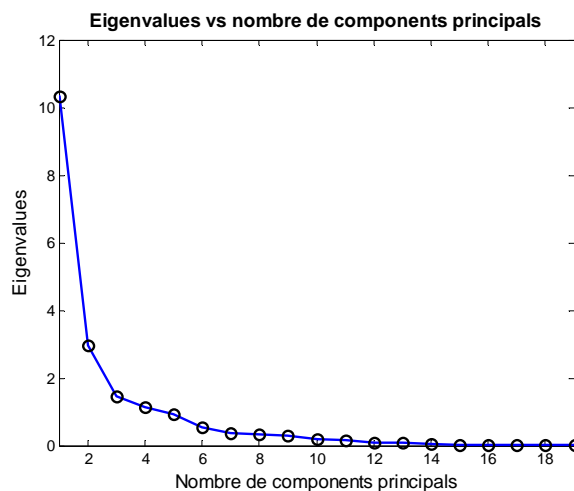


Figura 2.5. Descomposició per PCA (model bilineal) de la matriu de dades  $\mathbf{D}$  per un nombre de components  $N = 3$

El mètode utilitzat freqüentment per realitzar la descomposició PCA de la matriu  $\mathbf{D}$  és el de la descomposició en valors singulars (SVD [21]). Una primera manera de veure quants components cal incloure a l'anàlisi és mitjançant l'observació dels valors propis (*eigenvalues*) associats a aquests components, ja que aquests indiquen la quantitat de variància retinguda o explicada per cada component considerat. Els valors propis corresponen al quadrat dels valors singulars (*singular values*) trobats per SVD, i la quantitat de variància retinguda per a cada component, va disminuint successivament fins a un nivell petit i que ja varia poc per a nous components. Generalment, el nombre de components  $N$  s'escull fins que l'addició d'un nou component ja no proporciona informació addicional rellevant dins del context del problema estudiat o, dit d'una altra manera, quan ja només explica el soroll experimental. Aquells components que expliquen proporcions de variància petites no són investigats, i s'assumeix

que es troben principalment associats a petites contribucions de fons o al soroll i error experimental. El nombre escollit de components principals es relacionarà amb el nombre de fonts o pautes de variació (contaminació) independents que es troben presents en les dades analitzades. Per exemple, en el cas de la descomposició de la matriu de la Figura 2.6, el nombre de components aconsellat en funció de la representació dels valors propis seria de fins a tres (tot i que també caldria mirar el % de variància observada, que podria ser que amb dos components ja fos força elevada). A partir del tercer component, els valors propis disminueixen ja molt poc de mida i tendeixen a fer-se constants. En funció de la magnitud de la variància retinguda i de la contribució que les variables originals tenen a cada component, es pot deduir el significat ambiental dels diferents components trobats i pot determinar-se també el nivell de soroll aproximat de les dades experimentals. En aquest context, són extremadament útils les representacions de les matrius d'*scores* (matriu  $\mathbf{X}$ ) i de *loadings* (matriu  $\mathbf{Y}^T$ ) obtingudes directament de la descomposició PCA o SVD de la matriu de dades original  $\mathbf{D}$ .



**Figura 2.6.** Descomposició de valors propis d'una matriu de dades per a la determinació del nombre de components

Els valors dels *loadings* indiquen quina és la magnitud de la contribució de cadascuna de les variables originals a cada component o factor (nous eixos de representació). Quan es tracta de dades ambientals, els *loadings* poden indicar quina és la composició química de les fonts de contaminació identificades. Les variables amb valors grans de *loadings* sobre el mateix component, covarien. Si tenen el mateix signe, covarien positivament mentre que si tenen signes oposats, ho fan negativament (covarien inversament). La representació de les mostres en el nou espai definit pels nous components o factors són els *scores*. Aquestes representacions d'*scores* permeten agrupar les mostres segons la seva similitud i detectar la presència de mostres que presenten valors extrems. Des d'un punt de vista ambiental, els *scores* contenen informació sobre la distribució de les fonts de contaminació a les mostres (ja sigui informació geogràfica, temporal, per tipus de compartiment ambiental, etc.). La matriu

d'error conté aquella part de la variació de les dades que no es troba explicada pel model de PCA i que, per tant, es deu al soroll experimental. Generalment, la interpretació correcta de les diferents agrupacions de mostres obtingudes en els gràfics d'*scores* requereix que aquests siguin analitzats conjuntament amb els gràfics corresponents de *loadings*.

En determinats casos, per a una millor interpretació dels resultats obtinguts per PCA, és recomanable l'aplicació d'algun tipus de rotació dels components. Les tècniques d'anàlisi de factors utilitzen diferents tipus de rotacions. En el cas de les rotacions ortogonals, els nous eixos són també ortogonals entre ells, però els *loadings* de les variables que contribueixen a cada component es veuen modificats per tal de facilitar la interpretació de les dades. A més, la quantitat de variància explicada per cada component individual també variarà. La rotació Varimax [22, 32] intenta maximitzar la variància dels *loadings* de les variables dominants per rotació seqüencial dels *loadings* originals (obtinguts per PCA). El resultat són uns perfils de *loadings* simplificats i una redistribució de la variància, això sí, conservant-se la mateixa quantitat de variància total explicada pels factors o components identificats.

### 2.1.2.5. Resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS)

La resolució multivariant de corbes (*Multivariate Curve Resolution*, MCR) descriu una família de procediments quimiomètrics utilitzats amb la finalitat d'identificar i resoldre les contribucions existents en un conjunt de dades. Aquests procediments, sovint han estat aplicats per a la resolució de múltiples components químics en mescles investigades mitjançant tècniques d'anàlisi espectroscòpica (per exemple, per a la resolució d'espectres i perfils de concentració en mescles de compostos químics de composició desconeguda, sota diverses assumpcions) [33].

Darrerament, els procediments MCR han vist múltiples aplicacions a la literatura [34]. Alguns exemples de processos i sistemes analitzats mitjançant procediments MCR són l'anàlisi de mescles de compostos químics a partir de mesures espectrals o voltamperomètriques, l'estudi de processos i reaccions químiques (de cinètica i equilibri químic), les anàlisis cromatogràfiques complexes de mostres naturals amb detecció multicanal espectral (LD-DAD, LC-MS i GC-MS, etc.), l'anàlisi espectroscòpica de processos industrials i l'anàlisi d'imatges espectroscòpiques, entre d'altres. A més a més, aquests procediments també han estat aplicats per a la resolució de problemes de naturalesa ambiental. Aquest últim cas és el que es presenta en aquesta Tesi.

En primer lloc, per tal d'utilitzar els procediments de resolució multivariant de corbes, es requereix disposar, com a mínim, d'un conjunt de dades ordenades en dues direccions en una taula o matriu de dades **D**, per exemple un conjunt de *J* variables o concentracions de compostos químics (columnes de la matriu de dades) mesurades en *I* mostres (files de la matriu de dades) diferents. El mètode MCR es basa en la descomposició de la informació continguda en la matriu de dades **D**, seguint un model bilineal similar a l'emprat en el cas de PCA, format també per la suma de les contribucions de cadascun dels components (model additiu). Així, el model bilineal de resolució de corbes també s'expressarà com:

$$\mathbf{D} = \mathbf{X} \mathbf{Y}^T + \mathbf{E} \quad \text{Equació 2.11}$$

On  $\mathbf{Y}^T$  és la matriu de factors (anomenada de *loadings* en la notació de PCA), amb la contribució de cadascuna de les variables a les diferents fonts de contaminació (que expressarà la composició de les fonts de contaminació identificades), **X** és la matriu de factors (anomenada d'*scores* en la notació de PCA), amb la contribució de les fonts de contaminació identificades en cadascuna de les mostres analitzades (que expressarà la distribució de les fonts de contaminació sobre les mostres), i **E** conté l'error de les dades (part de la variància de les dades no explicada pel model). En aquesta Tesi, les matrius de factors **X** i  $\mathbf{Y}^T$  també s'anomenaran matrius d'*scores* i de *loadings*, conservant doncs la mateixa notació utilitzada en PCA.

Tot i que el mètode MCR utilitzi el mateix model bilineal additiu que el mètode PCA descrit anteriorment, la descomposició de la matriu de dades es fa de forma totalment diferent i donant lloc a resultats també diferents. Mentre que en PCA les matrius de factors, matriu de *loadings* i d'*scores*, són ortogonals i es troben en la direcció de màxima variància explicada (a més de que els *loadings* es troben normalitzats), en el cas dels procediments MCR, les matrius de factors es busquen de manera que aquests tinguin unes propietats similars a les que realment causen la variació experimental observada. Per exemple, en el cas ambiental, la composició química de les fonts de contaminació i la seva distribució geogràfica a les mostres només pot presentar valors positius o iguals a zero, però no negatius. Aquesta propietat no s'assolirà, en general, en el cas dels factors ortogonals trobats per PCA.

Existeixen diferents algorismes MCR de descomposició bilineal, entre els quals destaquen els procediments iteratius de resolució. Aquests mètodes de resolució iteratius recuperen el model bilineal a través d'una optimització iterativa de **X** i/o  $\mathbf{Y}^T$  o bé a través de l'optimització d'una matriu de transformació **R**. Així, el model bilineal bàsic de l'Equació 2.11, també es pot escriure com:

$$\mathbf{D} = (\mathbf{X}_0 \mathbf{R}) (\mathbf{R}^{-1} \mathbf{Y}_0^T) + \mathbf{E} \quad \text{Equació 2.12}$$

$$\text{On } \mathbf{X} = \mathbf{X}_0 \mathbf{R} \quad \text{Equació 2.13}$$

i

$$\mathbf{Y}^T = \mathbf{R}^{-1} \mathbf{Y}_0^T \quad \text{Equació 2.14}$$

On  $\mathbf{X}_0$  i  $\mathbf{Y}_0^T$  són les estimacions inicials de  $\mathbf{X}$  i  $\mathbf{Y}^T$  i poden ser factors abstractes o bé factors reals, depenent del mètode de resolució utilitzat. Tots els mètodes de resolució per iteracions comencen el procés d'optimització amb les estimacions inicials d'algunes de les matrius del model bilineal ( $\mathbf{X}_0$ ,  $\mathbf{Y}_0^T$  o bé  $\mathbf{R}_0$ ) i aquestes estimacions inicials són posteriorment modificades de manera iterativa sota l'acció de diferents restriccions. Les restriccions forcen als factors  $\mathbf{X}$  i /o  $\mathbf{Y}^T$  a obeir unes propietats químiques o matemàtiques prèviament definides, i són les característiques principals dels mètodes iteratius. Les restriccions constitueixen, al mateix temps, una manera d'introduir tota la informació matemàtica i química que es troba disponible d'entrada, per tal d'assegurar l'obtenció de solucions interpretables. A més, la correcta selecció i aplicació de les restriccions, minimitza o fins i tot pot arribar a eliminar l'ambigüitat relacionada amb els mètodes de resolució multivariant (descrita més endavant).

La diferència entre els diversos mètodes iteratius de resolució que existeixen a l'actualitat es pot relacionar amb el tipus de factors que acostumen a ser optimitzats de manera iterativa, les estimacions inicials utilitzades, la naturalesa i aplicació de restriccions, o bé el tipus d'estructura de dades necessària per tal de que l'anàlisi sigui possible.

Entre els mètodes iteratius de resolució multivariant de corbes destaca el mètode de resolució multivariant de corbes per mínims quadrats alternats (MCR-ALS). Aquest mètode utilitza una implementació específica dels mínims quadrats alternats (ALS) que ha estat proposada i desenvolupada pel grup de recerca dins del qual s'ha realitzat aquesta Tesi doctoral [29, 35-36]. El mètode bàsic del procediment MCR-ALS es basa també en el model bilineal que s'ha descrit a l'Equació 2.11. El mètode troba, de manera iterativa, les matrius dels factors  $\mathbf{X}$  i  $\mathbf{Y}^T$  a través de l'optimització d'estimacions inicials d'aquestes per un procediment de mínims quadrats alternats, que fa mínima una funció minimoquadràtica definida a partir dels residuals entre els valors de les dades experimentals i els valors corresponents estimats com a resultat de l'optimització de les matrius de factors  $\mathbf{X}$  i  $\mathbf{Y}^T$  i, cosa molt important, utilitzant determinades restriccions durant el procés d'optimització. En l'MCR-ALS, cap de les dues matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$  té prioritats respecte a l'altra, i les dues seran utilitzades en cada cicle d'iteració. El procediment

general utilitzat per MCR-ALS inclou els passos que es descriuen a continuació (veure Figura 2.7):

1. Determinació del nombre de components de la matriu de dades experimentals  $\mathbf{D}$
2. Generació de les estimacions inicials (per exemple,  $\mathbf{X}_0$ , per la matriu  $\mathbf{X}$ )
3. Càlcul de la matriu  $\mathbf{Y}^T$  sota l'aplicació de determinades restriccions
4. Càlcul de la matriu  $\mathbf{X}$  sota l'aplicació de restriccions
5. Obtenció de la matriu  $\mathbf{D}$  mitjançant el producte de  $\mathbf{X}$  i  $\mathbf{Y}^T$
6. Càlcul de la funció minimoquadràtica dels residuals definits a partir de la diferència entre la matriu experimental  $\mathbf{D}$  i la seva estimació iterativa  $\mathbf{X}\mathbf{Y}^T$ ,  $\|\mathbf{D} - \mathbf{X}\mathbf{Y}^T\|$ , o arrel quadrada de la suma dels residuals al quadrat, definits com a diferència entre els valors experimentals i els estimats per l'aplicació del model bilineal (Equació 2.11) amb les estimacions de  $\mathbf{X}$  i  $\mathbf{Y}^T$  a la iteració considerada
7. Retorn al pas número 3 fins que s'assoleix la convergència, és a dir, fins que els residuals definits en el pas 6, ja pràcticament no canviïn

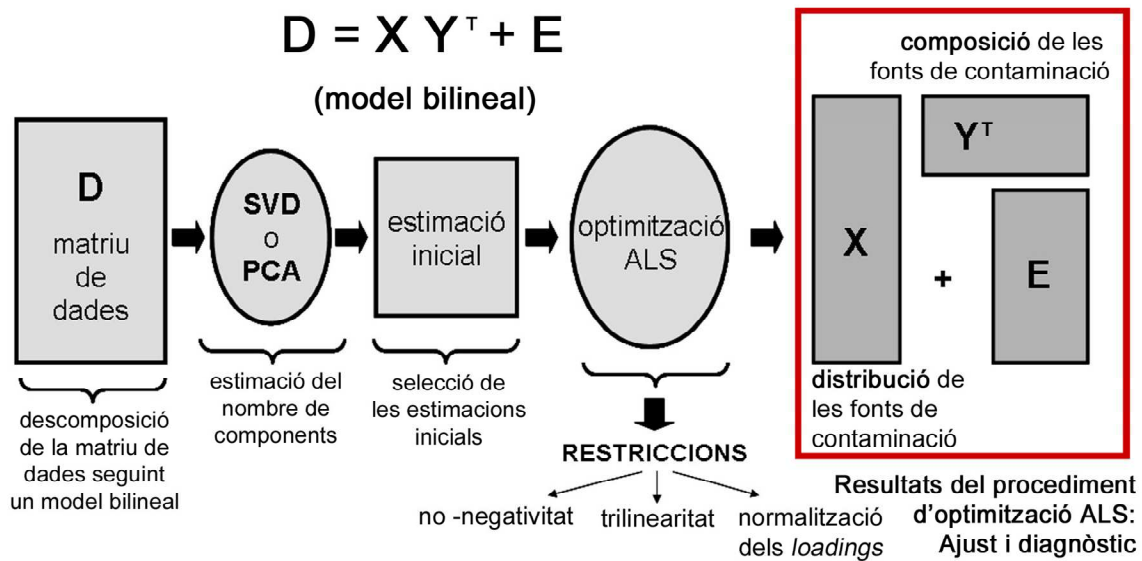


Figura 2.7. Diagrama d'implementació del mètode MCR-ALS amb l'exemple d'aplicació de tres restriccions diferents: no negativitat, normalització dels *loadings* i trilinearitat

### Determinació del nombre de components

El nombre de components ( $M$ ) de la matriu  $\mathbf{D}$  pot ser inicialment estimat utilitzant PCA o bé pot ser ja conegut d'entrada. En qualsevol cas, es recomana resoldre el sistema utilitzant diferent nombre de components, per tal de comparar els resultats i escollir aquell nombre de components que millor descriu les dades. Dit d'una altra manera, aquell nombre de components que proporcioni un error minimoquadràtic mínim i que doni solucions que compleixin les restriccions imposades i siguin plausibles dins del context del problema que es



vol resoldre. En el cas de les dades analitzades en aquesta Tesi (dades ambientals),  $N$  es referirà a les fonts o pautes de variació (contaminació ambiental) més importants i diferenciades. Un cop determinat el nombre de components, existirà un percentatge significatiu de variància no explicada provinent del soroll experimental i possiblement també de múltiples fonts de contaminació naturals i no naturals desconegudes, que tenen cadascuna d'elles una contribució a la variància petita, i que es trobaran agrupades dins de la matriu de residuals  $\mathbf{E}$ .

En els models bilineals, els percentatges de variància explicada per cada component poden ser calculats a partir de l'expressió següent:

$$R^2 = 1 - \frac{\sum_{i=1}^I \sum_{j=1}^J e_{i,j}^2}{\sum_{i=1}^I \sum_{j=1}^J d_{i,j}^2} \quad e_{i,j} = d_{i,j} - \hat{d}_{i,j} \quad \text{Equació 2.15}$$

On  $d_{i,j}$  és el valor de la variable  $j$  a la mostra  $i$  de la matriu de dades experimentals i  $\hat{d}_{i,j}$  és el valor corresponent calculat utilitzant MCR.  $R^2$  pot ser calculada per tot el model (pels  $N$  components considerats conjuntament) així com també per a cadascun dels components individuals d'un model determinat per separat.

### Selecció de les estimacions inicials

Una vegada realitzada l'estimació del nombre de components, qualsevol model de resolució iterativa basat en el model bilineal descrit a l'Equació 2.11, requereix l'estimació inicial d'una de les dues matrius,  $\mathbf{X}$  o bé  $\mathbf{Y}^T$  (veure Figura 2.7), per tal de començar el procés iteratiu d'optimització. Posteriorment, les matrius dels dos factors seran modificades pas a pas durant els cicles iteratius consecutius fins que s'assoleixi la convergència final.

Existeixen múltiples possibilitats per tal de generar les estimacions inicials, no obstant, en el context de MCR, és millor evitar les estimacions aleatòries. Enlloc d'aquestes, l'ús d'estimacions inicials que siguin plausibles (que compleixin les restriccions imposades) i tinguin un significat ambiental disminueix el temps computacional i minimitza els problemes de convergència i de mínims locals, ja que el conjunt de factors inicial obeeix les mateixes propietats que tenen els factors reals i, per tant, aquests factors inicials es troben més propers de la solució final real. Apart de la presència de mínims locals diferents, estimacions inicials diferents poden donar lloc a solucions diferents ambientalment però equivalents matemàticament degut a l'existència dels problemes d'ambigüitat rotacional (descrits més endavant). No obstant, totes aquestes solucions obtingudes presentaran un ajust òptim de la

## Capítol 2

---

funció minimoquadràtica i seran ambientalment possibles segons el coneixement previ que es tingui del sistema i les restriccions imposades.

Les estimacions inicials no aleatòries poden ser obtingudes directament a partir del conjunt de les dades sense processar, o bé poden ser el resultat de l'aplicació de mètodes quimiomètrics auxiliars. Així, la selecció d'un conjunt de columnes o d'un conjunt de files a partir de la matriu de dades original  $\mathbf{D}$ , igual en nombre als components estimats en el sistema, pot proporcionar una estimació inicial de les matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$  respectivament. Aquesta estimació es pot obtenir seleccionant directament aquelles files (per estimar  $\mathbf{Y}^T$ ) o bé aquelles columnes (per estimar  $\mathbf{X}$ ) de la matriu  $\mathbf{D}$  que siguin més diferents entre elles i, per tant, més pures o selectives. Aleshores, si el nombre de components escollit és correcte, qualsevol fila o columna de la matriu de dades serà combinació lineal d'aquestes files o columnes més pures.

En alguns casos, el simple coneixement previ del problema ambiental a analitzar pot ser suficient per tal de realitzar una selecció encertada de les estimacions inicials. D'altra banda, també existeixen mètodes quimiomètrics auxiliars quan es disposa de poca informació sobre el conjunt de dades, o bé simplement quan es prefereix obtenir estimacions inicials basades en procediments matemàtics.

Un dels mètodes quimiomètrics que permet obtenir estimacions inicials per qualsevol conjunt de dades és el mètode SIMPLISMA (*SIMPL*e-to-use *I*terative *S*elf-*M*odeling *A*nalysis [37-38]). Aquest mètode selecciona, de manera seqüencial, aquelles variables, en la direcció de les columnes o bé en la direcció de les files, que menys informació tenen en comú amb les seleccionades prèviament. El mètode SIMPLISMA és el que s'ha utilitzat per realitzar les estimacions inicials en els casos estudiats en aquesta Tesi. Altres mètodes utilitzats per aconseguir estimacions inicials es troben descrits en diversos treballs [39-40].

### Optimització ALS

Després de l'obtenció de les estimacions inicials, s'inicia el procediment de mínims quadrats alternats, el qual consisteix en resoldre les dues equacions següents sota les restriccions adjents:

$$\min_{\hat{\mathbf{X}}} \left\| \mathbf{D}_{\text{PCA}} - \hat{\mathbf{X}}\hat{\mathbf{Y}}^T \right\| \quad \text{Equació 2.16}$$

$$\min_{\hat{\mathbf{Y}}^T} \left\| \mathbf{D}_{\text{PCA}} - \hat{\mathbf{X}}\hat{\mathbf{Y}}^T \right\| \quad \text{Equació 2.17}$$

En les Equacions 2.16 i 2.17, la norma dels residuals (arrel quadrada de la suma dels residuals al quadrat) entre les dades reproduïdes per PCA,  $\mathbf{D}_{PCA}$ , utilitzant el nombre seleccionat de components, i les dades reproduïdes per l'ALS utilitzant les estimacions per mínims quadrats de les matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$ ,  $\hat{\mathbf{X}}$  i  $\hat{\mathbf{Y}}^T$ , (sota determinades restriccions) és alternativament minimitzada mantenint constant  $\hat{\mathbf{X}}$  o bé  $\hat{\mathbf{Y}}^T$ .

Es comença calculant la solució mínimoquadràtica inicial per a la matriu per la qual no s'ha realitzat l'estimació inicial. Així, si per exemple es disposa d'una estimació inicial dels factors per  $\mathbf{X}$ , la millor solució per mínims quadrats (en aquest cas, sense restriccions) de la matriu  $\mathbf{Y}^T$  serà calculada tal com s'indica a continuació:

$$\hat{\mathbf{Y}}^T = \hat{\mathbf{X}}^+ \mathbf{D}_{PCA} \quad \text{Equació 2.18}$$

On  $\hat{\mathbf{X}}^+$  és la matriu pseudoinversa de  $\mathbf{X}$  [21], definida com aquella matriu que multiplicada per  $\mathbf{X}$  dona la matriu identitat. Si, en canvi, es disposa d'una estimació inicial dels factors de concentració  $\mathbf{Y}^T$ , la millor solució es calcularà de la següent forma:

$$\hat{\mathbf{X}} = \mathbf{D}_{PCA} (\hat{\mathbf{Y}}^T)^+ \quad \text{Equació 2.19}$$

Els dos problemes de mínims quadrats de les Equacions 2.16 i 2.17, es resolen de manera seqüencial en cadascun dels cicles iteratius, de manera que la matriu  $\mathbf{Y}^T$  és calculada i posteriorment utilitzada per obtenir la matriu  $\mathbf{X}$ . Cal remarcar que la matriu emprada durant la minimització mínimoquadràtica del procediment d'optimització és, segons les Equacions 2.16 a 2.19, la matriu calculada a partir d'un model de PCA amb un nombre determinat de components,  $\mathbf{D}_{PCA}$ . Es tracta, per tant, d'una matriu a la qual ja s'ha eliminat una part important del soroll experimental i que per tant, tot i mantenir pràcticament tota la informació química rellevant continguda en les dades originals, facilita el procés d'optimització i l'assoliment més fiable de la convergència dels factors optimitzats i de l'obtenció de les solucions finals.

El criteri de convergència en l'optimització ALS es basa en la comparació de l'ajust obtingut en dues iteracions consecutives. Quan la diferència d'ajust relativa mesurada a partir de la desviació estàndard dels residuals es troba per sota d'un valor lliandar establert (normalment fixat al 0,1%), el procés d'optimització es dona per finalitzat. Altres possibilitats consisteixen en definir un nombre màxim de cicles iteratius com a criteri de finalització, o bé en comparar la forma dels factors resolts en iteracions consecutives. Sempre que s'utilitzin les restriccions adequades durant els processos d'optimització, MCR-ALS proporcionarà solucions possibles,

encara que les ambigüitats rotacionals i/o d'intensitat poden continuar-hi presents depenent de l'estructura, la complexitat i el coneixement previ del sistema sota estudi. Una versió MATLAB de l'algoritme corresponent al procediment MCR-ALS que inclou una interfase gràfica per a facilitar la seva implementació, es pot trobar disponible i pot ser descarregada lliurement des de la pàgina web [www.ub.edu/gesq/mcr/mcr.htm](http://www.ub.edu/gesq/mcr/mcr.htm) [41].

### Resultats de l'ALS. Ajust i diagnòstic

Com a resultat del procés d'optimització ALS s'obté (veure Figura 2.7) una matriu  $\mathbf{Y}^T$  o matriu de *loadings*, amb la descripció de les variables que formen part dels components que, en el cas ambiental, correspondrà a la composició de les fonts de contaminació identificades, una matriu  $\mathbf{X}$  o matriu d'*scores*, amb la distribució d'aquestes fonts a les mostres, i finalment una matriu  $\mathbf{E}$ , amb l'error o part de la variància no explicada pel model.

Ja que el nombre de components no és un paràmetre conegut d'entrada, és important comparar l'ajust obtingut per diferent nombre de components així com investigar el sentit de les solucions obtingudes en cada cas. De la mateixa manera, quan s'analitzen simultàniament múltiples conjunts de dades ordenades en matrius de mides similars, caldrà comparar els resultats obtinguts amb l'aplicació de diferents tipus de models (bilineal o trilineal) quan no s'estigui del tot segur del model que segueixen les dades.

### Incerteses inherents al mètode MCR-ALS

Les principals fonts d'incertesa associades als resultats obtinguts amb els mètodes de resolució multivariant són les ambigüitats (rotacionals i d'escala o d'intensitat) dels components recuperats i la propagació del soroll experimental [36, 42]. Per tant, l'estudi d'aquests dos efectes és necessari quan es vol avaluar la qualitat dels resultats obtinguts i es vol comparar aquests resultats amb els obtinguts mitjançant altres procediments quimiomètrics. El desenvolupament de metodologies per quantificar les incerteses no és només un tema d'interès creixent en la literatura quimiomètrica actual sinó també una necessitat per tal d'estendre la utilització d'aquests mètodes de resolució de forma més eficient.

El problema a resoldre en la resolució multivariant de corbes, pot ser descrit matemàticament de la següent manera:

Donada una matriu de dades  $\mathbf{D}$ :

- (1) caldrà trobar  $N$ , el nombre de components causant de la variància observada a les dades,

- (2) trobar la contribució dels components a les diferents mostres analitzades (perfils d'scores) o matriu  $\mathbf{X}$
- (3) trobar la contribució de les variables als components identificats (perfils de *loadings*) o matriu  $\mathbf{Y}^T$

No obstant, les solucions obtingudes per  $\mathbf{X}$  i  $\mathbf{Y}^T$  a partir de l'Equació 2.11 (que correspon a la descomposició de la matriu de dades  $\mathbf{D}$  seguint un model bilineal), serà ambigua si no es disposa d'informació addicional, o dit d'una altra manera, existirà, en les solucions obtingudes de l'Equació 2.11, una llibertat rotacional i d'escala. Aquest problema sovint rep el nom de problema d'ambigüitat de l'anàlisi de factors. Les ambigüitats rotacionals i d'escala, es mostren de manera simplificada a l'Equació 2.20:

$$\mathbf{D} = \mathbf{X}_o \mathbf{Y}_o^T = (\mathbf{X}_o \mathbf{T})(\mathbf{T}^T \mathbf{Y}_o^T) = \mathbf{X}_k \mathbf{Y}_k^T \quad \text{Equació 2.20}$$

On  $\mathbf{X}_o$  i  $\mathbf{Y}_o^T$  corresponen a matrius inicials d'scores i *loadings* obtingudes de la descomposició bilineal de la matriu  $\mathbf{D}$ , i  $\mathbf{X}_k$  i  $\mathbf{Y}_k^T$  són matrius d'scores i *loadings* obtingudes a partir del producte de les matrius inicials d'scores i *loadings* per una matriu  $\mathbf{T}$  (anomenada matriu de rotació) [43]. D'aquí se'n dedueix que, si no s'imposa cap tipus de restricció, existiran un nombre infinit de solucions a l'Equació 2.20 per qualsevol matriu invertible  $\mathbf{T}$ . Diverses combinacions lineals de les solucions poden ser obtingudes a partir de matrius de rotació apropiades  $\mathbf{T}$ .

L'existència de les ambigüitats d'intensitat o d'escala, comporta que els components estimats per MCR-ALS (o per qualsevol altre mètode de descomposició bilineal) estiguin escalats per un factor desconegut, el qual haurà de ser fixat ja sigui en base al coneixement extern o bé per algun tipus de normalització (en molts casos, els perfils de *loadings* es normalitzen entre 0 i 1), per tal d'evitar que surtin fora d'escala durant l'optimització per ALS. D'altra banda, existeix un altre tipus d'ambigüitat, anomenada ambigüitat rotacional, que és un problema persistent i més seriós que l'ambigüitat d'escala. Aquesta apareixerà quan hi ha dos o més components que se superposen i, a conseqüència d'això, és possible obtenir un conjunt de solucions diferents que ajusten igualment bé les dades i que compleixen correctament les restriccions posades al sistema (la solució no és única). És a dir, les solucions obtingudes són equivalents des d'un punt de vista d'optimització matemàtica, però són diferents des d'un punt de vista fisicoquímic. Quan això passa, no hi ha manera d'estar segur de si la solució obtinguda és la solució veritable. No obstant, normalment és possible reduir considerablement aquest nombre infinit de solucions mitjançant l'aplicació de restriccions derivades de la naturalesa i del coneixement previ del problema que s'estudia. Una discussió més detallada d'aquest tipus d'ambigüitats cau fora de l'abast d'aquesta Tesi i constitueix un dels temes actuals de recerca (veure, per exemple, Abdollahi et al. (2009) [44]).

## Capítol 2

---

A banda de les ambigüitats d'intensitat i rotacionals, existeix un altre problema que sempre es troba present en l'anàlisi de dades experimentals, que és el problema de la propagació de soroll. Fins i tot en absència d'ambigüitats rotacionals, l'error, inevitablement present en les dades reals, es propaga a les solucions obtingudes. Aquesta font d'incertesa afecta els resultats de tot tipus de mètodes d'anàlisi de dades. En els casos senzills (com el calibratge univariant o multivariant lineal), la determinació de l'efecte del soroll es pot estimar a partir de fórmules matemàtiques analítiques senzilles. Però l'error associat als mètodes de resolució multivariant és un problema complex, degut a l'elevat nombre de paràmetres no lineals que han de ser calculats i a l'aplicació de les restriccions. S'han proposat procediments d'estimació dels errors associats a les incerteses a partir de l'aplicació de procediments de simulació dels efectes del soroll utilitzant tècniques de Monte Carlo i *boot-strapping* [42, 45]. Per altra banda, la disminució i eventual eliminació dels efectes del soroll o error experimental sobre els paràmetres es pot aconseguir a partir de l'aplicació de mètodes basats en aproximacions de màxima versemblança (*maximum likelihood*) com ara els de sospesament de mínims quadrats alternats (*Weighted Alternating Least-Squares*, WALs [28, 46]) proposats recentment. Una discussió de tots aquests aspectes cau també fora de l'àmbit d'aquesta Tesi, però poden consultar-se en els diversos treballs citats en aquest apartat i d'altres [47-48].

### Restriccions

Les restriccions són propietats químiques o matemàtiques generals que les matrius de factors  $X$  i /o  $Y^T$  han de complir, les quals actuen dirigint el procediment d'optimització iteratiu cap a una solució amb significat físic i interpretable en termes ambientals. La utilització adequada de restriccions és una condició imprescindible en tots els procediments de resolució multivariant i permet disminuir l'ambigüitat de les solucions finals obtingudes. Les restriccions poden ser classificades d'acord amb la seva naturalesa o bé d'acord amb la seva utilització. Quan s'aplica una determinada restricció, el coneixement matemàtic i ambiental sobre els factors ha de ser traduït a condicions matemàtiques.

Una restricció o condició pot ser imposada de dues maneres: forçant el factor (o alguns dels seus elements) a ser iguals a determinats valors preestablerts (restriccions d'igualtat, *equality constraints*), o bé forçar-lo a ser superior o inferior a ells (restriccions de desigualtat, *inequality constraints*). D'aquesta manera, si a un factor se li aplica la restricció de no negativitat, els elements negatius d'aquest factor són forçats a ser positius (o zero). En el cas de dades ambientals, la restricció de no negativitat pot ser aplicada, per exemple, sobre concentracions químiques (valors de les variables en els *loadings*) o bé sobre les contribucions de les fonts de contaminació sobre les mostres (valors dels *scores*), ja que en els dos casos només tenen sentit si són positives [49]. Hi ha diferents formes d'aplicar la restricció de no negativitat, com la

substitució directa dels valors negatius per zero durant els diferents passos iteratius o els mètodes més rigorosos, basats en procediments per mínims quadrats no negatius (*non-negative least squares*, nnls, [49]) i les seves variants més ràpides, com per exemple la *fast non-negative least squares*, fnnls [50]. En aquesta Tesi s'ha utilitzat la basada en aquest darrer mètode de mínims quadrats no negatius ràpids (fnnls). A més, la matriu de *loadings*  $\mathbf{Y}^T$ , es normalitza per tal d'evitar problemes d'ambigüitats d'escala [36] (veure apartat de les incerteses inherents al mètode MCR-ALS).

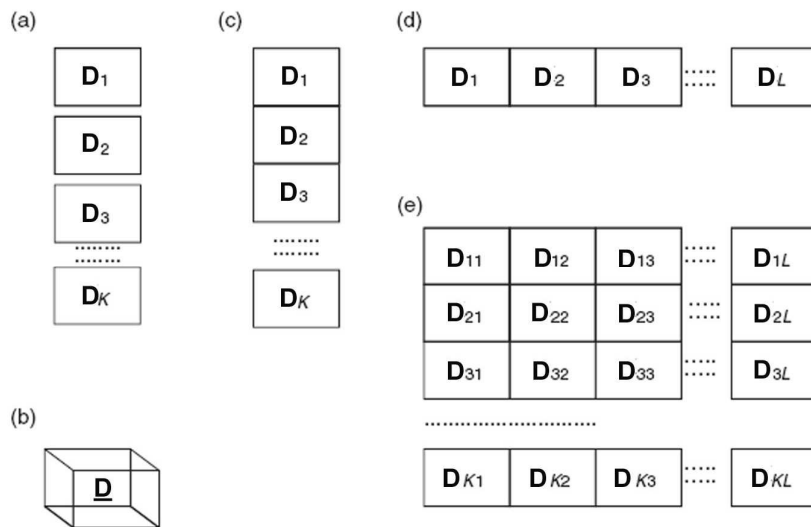
A banda de les restriccions de no negativitat i de normalització (sovint aplicades per a dades ambientals), el mètode MCR-ALS també permet l'aplicació d'altres restriccions majorment utilitzades en altres camps de la química analítica, com són l'unimodalitat (permet als perfils tenir només un màxim), la de sistema tancat (degut, per exemple, al compliment de balanços de massa), les de selectivitat o rang local, o les de modelatge rígid (*hard-modeling*), sovint utilitzades en el cas de sistemes químics que segueixen parcial o totalment models cinètics o d'equilibri. La descripció de totes aquestes possibilitats cau fora de l'àmbit d'aquesta Tesi, i es troben recollides a la bibliografia d'aplicació del procediment MCR-ALS [36, 47, 51-54].

Finalment queda un tercer tipus de restricció, la restricció de trilinearitat, que ha estat emprada en l'àmbit d'aquesta Tesi. Aquesta restricció és aplicada en l'anàlisi simultània de diverses matrius de dades (ambientals i no ambientals), i es troba descrita en el següent apartat, que descriu l'anàlisi simultània de dades ordenades en tres o més direccions (*multiway data*).

És important, a l'hora d'aplicar qualsevol restricció, considerar que aquesta estigui plenament justificada des d'un punt de vista matemàtic o bé ambiental. El fet d'utilitzar una restricció inadequada, pot donar lloc a resultats erronis i pitjors que si aquesta no s'apliqués. De vegades, poden aparèixer dubtes sobre l'aplicació o no d'una determinada restricció. En aquests casos, és millor anar introduint les restriccions de manera gradual durant el procés de resolució, i anar comprovant l'efecte que té la seva introducció a partir del nivell d'ajust obtingut en la reproducció del conjunt de dades original i l'efecte sobre les solucions obtingudes. El decreixement significatiu de l'ajust i l'aparició de residuals amb tendències no aleatòries lligats a la introducció d'una restricció determinada, és una indicació de que probablement les dades no la compleixen. Quan l'aplicació de les restriccions porta a solucions que no són raonables des d'un punt de vista físic (ambiental), caldrà comprovar si la seva aplicació ha estat massa estricta, i ampliar els nivells de tolerància per al seu incompliment. Finalment, cal remarcar que existeixen restriccions que, un cop aplicades sobre les dades, permeten eliminar l'ambigüitat rotacional. Aquest és el cas de les restriccions de trilinearitat, de les de model rígid o de les de selectivitat i rang local.

**Extensió del mètode MCR-ALS a l'anàlisi simultània de múltiples conjunts de dades**

Un dels majors avantatges del mètode MCR-ALS és que pot ser fàcilment estès a l'anàlisi simultània de diversos conjunts de dades obtinguts en experiments, situacions o condicions diferents. Això és possible a partir de l'augmentació o concatenació de les matrius de dades individuals corresponents als experiments individuals. Una matriu de dades pot ser augmentada per files (*row-wise*), per columnes (*column-wise*) o bé per files i columnes (*row-and column-wise*) per tal de formar una nova estructura multimatriu on les matrius amb el mateix nombre de files, columnes o ambdues a la vegada, són concatenades en la direcció apropiada [48]. El primer avantatge que cal destacar d'aquesta augmentació de matrius de dades és que aquelles característiques de resolució millors per una o algunes de les matrius analitzades, sempre tindran un efecte positiu sobre la resolució d'aquelles altres matrius més complexes. Tal com es mostra a la Figura 2.8, hi ha tres tipus possibles d'augmentació o concatenació de les matrius de dades, les quals poden ser indicades de manera concisa utilitzant la notació de l'entorn de programació MATLAB. Les matrius de dades augmentades per columnes,  $D_{aug}$  (veure Figura 2.8c), s'escriuen comon el símbol “;” s'utilitza per indicar que les diferents matrius de dades  $D_k$ ,  $k=1,\dots,K$ , (on  $k$  indica les diferents experiments, condicions o situacions, les quals poden correspondre per exemple, a diferents temps, a diferents compartiments ambientals, etc.) es troben concatenades per columnes, una sobre l'altra, mantenint el mateix nombre de columnes que les matrius individuals. En aquesta estructura de dades, se suposa que les diferents matrius comparteixen el seu espai vectorial per columnes.



**Figura 2.8.** Diferents estructures i ordenacions de dades (Tauler et al., 2009); (a) matrius individuals, (b) conjunt de dades multiway (cub de dades), i (c)-(e) disposició de múltiples matrius. (c) matrius augmentades per columnes, (d) matrius augmentades per files, i (e) matrius augmentades per files i per columnes



Per aquest tipus d'augmentació de les dades, el model bilineal es pot escriure mitjançant la següent equació:

$$\begin{pmatrix} D_1 \\ D_2 \\ D_3 \\ \vdots \\ D_k \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \\ X_3 \\ \vdots \\ X_k \end{pmatrix} Y^T + \begin{pmatrix} E_2 \\ E_2 \\ E_3 \\ \vdots \\ E_k \end{pmatrix} = X_{aug} Y^T + E_{aug} \quad \text{Equació 2.21}$$

O, d'una forma més compacta (notació MATLAB):

$$[D_1; D_2; D_3; \dots; D_k] = [X_1; X_2; X_3; \dots; X_k] Y^T + [E_1; E_2; E_3; \dots; E_k] = X_{aug} Y^T + E_{aug} \quad \text{Equació 2.22}$$

En el model bilineal descrit a les Equacions 2.21 i 2.22 només es formula una sola matriu  $Y^T$ , equivalent a la matriu de *loadings*, que en el cas de les taules de dades ambientals descriurà la composició de les diferents fonts de contaminació identificades, comuna a totes les matrius  $D_k$  considerades. D'altra banda, la matriu augmentada d'*scores*  $[X_1; X_2; X_3; \dots; X_k]$ , descriurà (també pel cas de taules ambientals) els canvis en la distribució de cadascun dels components o fonts de contaminació identificades en cadascuna de les mostres contingudes a les diferents matrius  $D_k$ . Així, el model bilineal de les Equacions 2.21 i 2.22 assumeix que els components identificats a les matrius de dades  $D_k$  inclosos a la matriu augmentada per columnes, comparteixen els mateixos perfils de composició, mentre que la seva distribució a les mostres és diferent en cadascuna de les  $k$  condicions o situacions.

Les matrius de dades augmentades per files (veure Figura 2.8d) es col·loquen una al costat de l'altra per a la seva concatenació. En aquest cas, es considera que les diferents matrius comparteixen l'espai vectorial de les seves files. Les matrius de dades també poden ser augmentades per columnes i per files al mateix temps (veure Figura 2.8e). Això significa que algunes de les matrius comparteixen l'espai vectorial de les seves columnes mentre que d'altres el comparteixen de les files. L'augmentació de matrius per files i en ambdues direccions (files i columnes) no es descriuen amb més detall en aquest treball pel fet que no han estat utilitzades en aquesta Tesi.

El model bilineal descrit a l'Equació 2.22 pot ser aplicat de diferents maneres partint de la informació continguda a les matrius de dades originals. En MCR, només es consideraran aquelles solucions del model que tinguin un significat físic i interpretable en termes ambientals. En aquest punt, MCR es diferencia d'altres aproximacions basades en models bilineals, com

## Capítol 2

---

ara PCA, que també pot ser estesa a l'anàlisi de matrius de dades augmentades utilitzant equacions similars. El mètode PCA busca solucions que siguin ortogonals i que expliquin la màxima variància de les dades (sense superposició, no compartida). És per aquest motiu que les solucions obtingudes de PCA no tenen un significat físic directe. En MCR, s'apliquen diferents tipus de restriccions per tal d'obtenir uns perfils dels components que tinguin un sentit físic i que, per tant, puguin ser fàcilment interpretables en termes ambientals. La implementació de restriccions naturals (com la no negativitat i la normalització, en el cas de les dades ambientals), seguirà la mateixa estratègia adoptada per a l'anàlisi de les matrius individuals. Tot i això, cal tenir en compte que les restriccions seran aplicades individualment i per separat per a cadascuna de les submatrius en les matrius augmentades  $\mathbf{X}_{\text{aug}}$  (si es realitza l'augmentació per columnes).

El problema de mínims quadrats resolt per MCR-ALS quan aquest s'aplica a conjunts de dades ordenats en múltiples direccions, és el mateix que quan aquest és aplicat a l'anàlisi d'una matriu de dades individual; la única diferència és que ara  $\mathbf{D}$  i  $\mathbf{X}$  i/o  $\mathbf{Y}^T$  són matrius augmentades. No obstant, l'aplicació d'MCR-ALS a aquest conjunt més complex de dades no hauria de ser considerada equivalent a la descomposició d'una matriu de dades individual, degut a la possible aplicació de les restriccions a nivell de submatriu. El primer avantatge clar de l'anàlisi simultània de múltiples matrius de dades és que, en general, aquesta genera estimacions per mínims quadrats que són més robustes i precises.

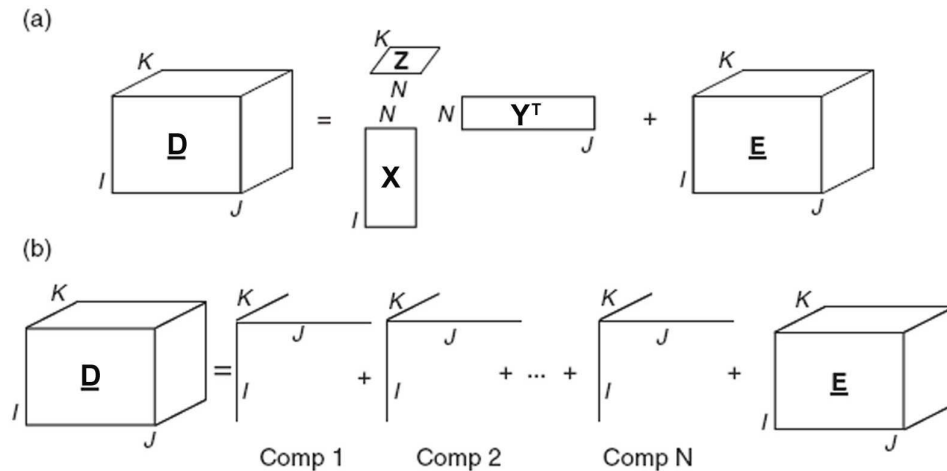
Una altra característica interessant de l'anàlisi de múltiples matrius de dades és la possibilitat d'aplicar noves restriccions relacionades amb les estructures de dades anomenades *multiway*. Aquest és el cas quan, per exemple, les diferents matrius de dades analitzades simultàniament tenen les mateixes mides (files i columnes), i es pot construir amb elles un cub de dades. En aquests casos, el model bilineal es pot utilitzar també sobre aquestes noves estructures de dades, però aquest model bilineal es pot estendre igualment a un model trilineal. Aquí, la restricció aplicada en el mètode MCR-ALS s'anomena de trilinearitat, i implica que els perfils dels factors d'un mateix component en les diferents matrius de dades analitzades simultàniament tinguin la mateixa forma i que només variïn entre si en un factor d'escala.

L'expressió del model trilineal (també anomenat anàlisi de factors paral·lels o *parallel factor analysis*, PARAFAC [30, 55]) utilitzada en aquesta Tesi per descriure la descomposició d'un conjunt de dades ordenat en tres direccions (*three-way data set*) es dona a l'Equació 2.23.

$$\mathbf{D}_k = \mathbf{XZ}_k\mathbf{Y}^T + \mathbf{E}_k$$

Equació 2.23

On  $\mathbf{X}$ ,  $\mathbf{Z}$  i  $\mathbf{Y}^T$  són les matrius de factors (perfils dels components en les tres direccions, modes o ways) i  $\mathbf{E}$  la matriu de residuals (part de la variància de les dades no explicada pel model). Aquesta descomposició trilineal d'un conjunt de dades en tres direccions  $\mathbf{D}$  es troba també representada gràficament a la Figura 2.9.



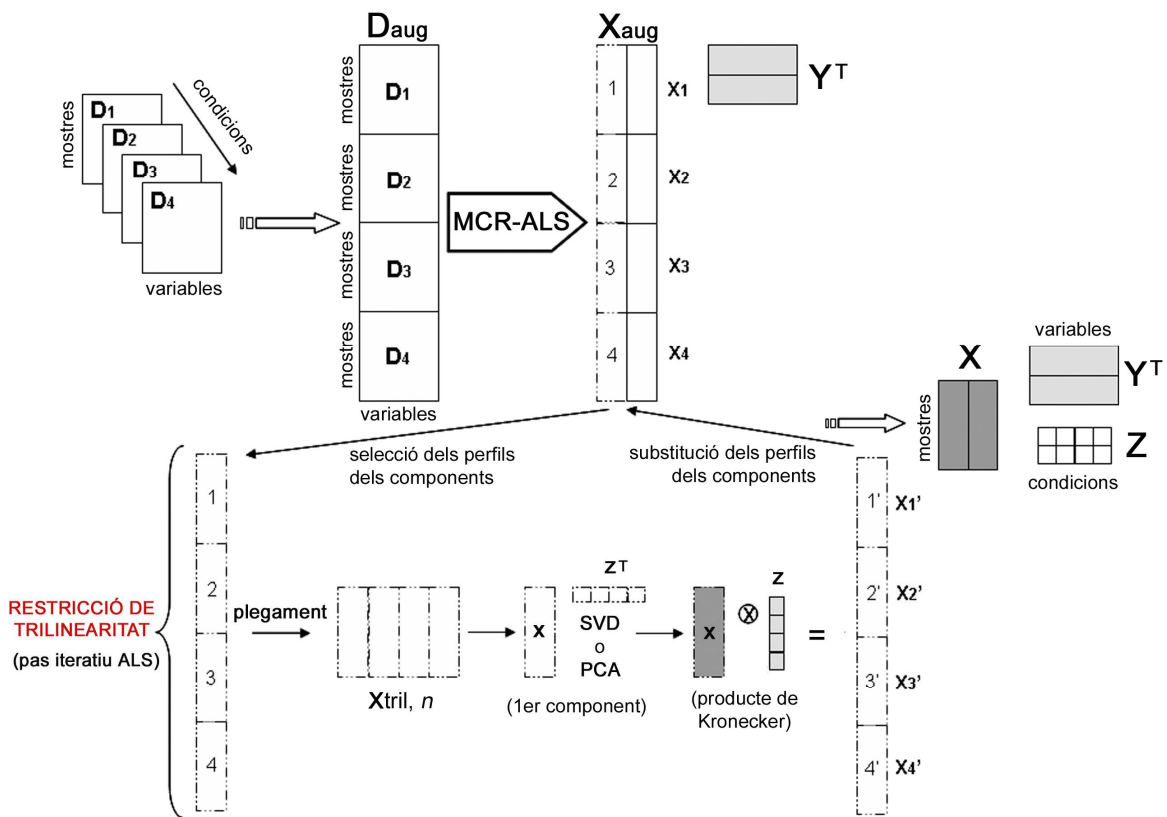
**Figura 2.9.** Descomposició d'un conjunt de dades en tres direccions seguint un model trilineal (Tauler et al., 2009) (a) expressat mitjançant la descomposició en matrius de *loadings* i (b) expressat com la suma dels components purs

L'anàlisi d'un conjunt de dades en tres direccions mitjançant un model trilineal genera directament tres conjunts de matrius de factors amb els perfils dels components (*loadings*), amb el mateix nombre de components en els tres modes o direccions. A la Figura 2.9a, la matriu  $\mathbf{X}$  correspon als *loadings* comuns en el primer mode (també anomenada *scores* en el cas dels models bilineals de PCA quan les dades s'ordenen en dues direccions) de totes les matrius analitzades, i conté la contribució dels diferents components sobre les mostres, que en termes ambientals correspon a la distribució geogràfica o temporal de les fonts de contaminació identificades. La matriu  $\mathbf{Y}^T$  són els *loadings* comuns en el segon mode de totes les matrius analitzades, i conté la contribució de les diferents variables en els components, que en el cas ambiental correspon a la composició química de les fonts de contaminació identificades. La nova matriu de dades  $\mathbf{Z}_k$  de *loadings* en el tercer mode o direcció, és una matriu diagonal, amb la contribució relativa de cada component en cadascuna de les matrius de dades considerades  $\mathbf{D}_k$ . Tal com es descriu en l'Equació 2.23, els elements de la matriu  $\mathbf{Z}_k$  seran els únics valors en el model de descomposició trilineal que canvien per a les diferents matrius  $\mathbf{D}_k$ . Les matrius  $\mathbf{X}$  i  $\mathbf{Y}^T$  són úniques (comunes) i invariants (les fonts o components identificats tenen la mateixa composició i la mateixa distribució a les mostres) per a cada component individual en totes les matrius  $\mathbf{D}_k$  analitzades simultàniament. La trilinearitat és doncs una restricció forta, que només es compleix en determinades ocasions, però per altra banda i com a conseqüència de la seva

## Capítol 2

aplicació, permet l'obtenció de solucions úniques per a les tres matrius de factors i evita la presència de les ambigüitats rotacionals associades a models bilineals.

El mètode MCR-ALS és adequat tant per a l'anàlisi de dades ordenades en tres direccions, formades per l'augmentació de matrius individuals de diferents mides i que no presenten una estructura trilineal, com per a l'anàlisi de conjunts de dades trilineals, constituïts per diferents matrius de dades amb el mateix nombre de files i columnes (cubs de dades). En aquest darrer cas, MCR-ALS aprofita els avantatges d'aquesta estructura i pot forçar les solucions obtingudes (les matrius de factors en els diferents modes) a que aquestes compleixin les premisses del model trilineal. L'estructura trilineal s'incorpora com una restricció opcional durant l'optimització ALS dels perfils de la matriu  $\mathbf{X}_{aug}$ . A més de ser opcional, aquesta restricció pot ser aplicada per separat sobre cadascun dels components. A continuació es mostra un esquema de la seva implementació en el mètode MCR-ALS [56]:



**Figura 2.10.** Descomposició de la matriu de dades augmentada  $\mathbf{D}_{aug}$  per MCR-ALS aplicant la restricció de trilinearitat.  $\mathbf{D}_{aug}$  formada per la concatenació de les matrius individuals ( $\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3, \mathbf{D}_4$ ).  $\mathbf{X}_{aug}$  és la matriu augmentada d'scores (mestres en les diferents matrius).  $\mathbf{Y}^T$  és la matriu de *loadings* en el segon mode (variables).  $\mathbf{X}$  és la matriu de *loadings* en el primer mode (mestres).  $\mathbf{Z}$  és la matriu de *loadings* en el tercer mode (condicions o situacions, equivalents a períodes de temps o compartiments en el cas de dades ambientals).

Tal com es mostra a la Figura 2.10, quan s'aplica la restricció de trilinearitat, els perfils de cada component en el primer mode (*scores* o bé *loadings* en el primer mode) de les diferents matrius  $\mathbf{X}_k$  es veuen forçats a adoptar exactament la mateixa forma (només variarà la seva escala per les diferents matrius de dades) durant cada iteració de l'optimització ALS. Per tal de fer això, en primer lloc, els perfils d'*scores* corresponents a un component en les diferents matrius es col·loquen un al costat de l'altre (plegament) per tal de generar una nova matriu,  $\mathbf{X}_{\text{tril},n}$ , amb  $I$  files (igual al nombre de files de cada matriu de dades) i  $K$  columnes (igual al nombre de matrius analitzades simultàniament),  $n$  indicant el component sobre el qual s'aplica la restricció. En el cas de l'aplicació d'un model totalment trilineal, cada component donarà una matriu  $\mathbf{X}_{\text{tril},n}$  ( $n = 1, \dots, N$ ), que es descomposarà bilinealment emprant només un component (mitjançant, per exemple, la descomposició de valors singulars (SVD)), tal com s'indica a continuació:

$$\mathbf{X}_{\text{tril},n}(I,K) = \mathbf{x}(I,1) \mathbf{z}^T(1,K) \quad \text{Equació 2.24}$$

On  $\mathbf{x}$  és el nou vector columna que conté el perfil de *loadings* comú (mitjà) en el primer mode (*scores*) del component  $n$  en les diferents matrius  $K$ , i  $\mathbf{z}^T$  és un nou vector fila amb el perfil de *loadings* en el tercer mode, que expressa les quantitats relatives d'aquest component en les diferents matrius  $K$ . El producte de Kroneker [57] d'aquests dos nous vectors de *loadings* dona lloc a un nou vector d'*scores* augmentats, el qual substitueix la columna corresponent de la matriu augmentada  $\mathbf{X}_{\text{aug}}$ . Els perfils  $\mathbf{x}$  i  $\mathbf{z}$  en cada iteració ALS donen la darrera estimació dels perfils de *loadings* en el primer i tercer mode per al component considerat. Així, quan la restricció de trilinearitat s'imposa durant el procediment d'optimització iterativa ALS, es força a que la forma dels vectors de *loadings* en el primer mode (distribució de les fonts de contaminació sobre les mostres en el cas de taules de dades ambientals) sigui la mateixa en totes les condicions (períodes de temps o compartiments ambientals). A més a més, la variació relativa de la contribució del component considerat a les diferents matrius, es captura en els *loadings* del tercer mode, mostrant les diferències d'aquest component entre les diferents taules de dades ambientals analitzades simultàniament, per exemple, en períodes de temps o compartiments ambientals diferents.

Les solucions obtingudes mitjançant l'aplicació del model trilineal són úniques i, a diferència del que ocorre en el cas dels models bilineals, no presenten ambigüitats rotacionals [36]. L'assumpció d'un model trilineal simplifica considerablement la interpretació dels resultats obtinguts ja que, en termes ambientals, això significa que no només la composició (matriu  $\mathbf{Y}^T$ ) de les fonts de contaminació identificades per MCR-ALS amb l'aplicació de trilinearitat serà comuna a totes les matrius analitzades simultàniament sinó que la seva distribució a les

## Capítol 2

---

mostres (distribució geogràfica) també serà la mateixa per a una mateixa font (component) en les diferents condicions.

Els resultats obtinguts mitjançant MCR-ALS amb la restricció de trilinearitat aplicada sobre tots els components del sistema, serien pràcticament els mateixos que s'obtidrien amb l'aplicació d'altres mètodes basats en models trilineals (com per exemple, pel mètode PARAFAC [30]) sobre el mateix sistema. No obstant, el principal avantatge del mètode MCR-ALS és que permet l'aplicació de la restricció de trilinearitat de manera independent i opcional sobre cadascun dels components del conjunt de dades, donant més flexibilitat a l'anàlisi i permetent la utilització de models trilineals totals o parcials, segons sigui convenient. Això fa que aquest mètode es distingeixi de PARAFAC, on tots els components obtinguts han de complir la restricció de trilinearitat. Aquesta flexibilitat permet un modelatge més representatiu d'algunes situacions reals, com per exemple, les d'aquells sistemes (com ocorre en determinats problemes ambientals) en què els perfils d'alguns components es poden comportar de manera trilineal mentre que els perfils d'altres components poden no comportar-s'hi.

L'estructura trilineal d'un conjunt de dades pot ser comprovada aplicant i comparant l'anàlisi SVD de les matrius de dades augmentades en els diferents modes o direccions (augmentació per files i per columnes) i mitjançant la comparació de l'ajust obtingut a partir de la utilització de models bilineals i trilineals, tal com s'explica amb més detall a continuació.

### **Rang i estructura de les dades**

Per un sistema de dades determinat ordenat en una taula o matriu de dades, s'anomena rang (representat per  $N$ ) al nombre de components linealment independents en absència de soroll que són necessaris per tal de representar la variació observada a les dades. Aquesta definició no coincidiria amb la de rang matemàtic d'una matriu de dades, la qual es referiria al nombre de files o columnes linealment independents de la matriu i tindria en compte també les variacions degudes al soroll i error experimental. El rang matemàtic serà igual al nombre de files o columnes (el que sigui inferior dels dos) de la matriu de dades, ja que sempre existirà una component aleatòria associada a l'error experimental. No obstant, en aquest cas no interessarà el rang matemàtic, sinó el nombre de causes de variació sistemàtica no associades amb el soroll ni l'error experimental. Normalment, aquest nombre de components  $N$ , no és conegut d'entrada. Tot i que existeixen diverses aproximacions basades en criteris estadístics per tal de determinar el rang d'un conjunt de dades, a la pràctica s'acostumen a escollir nombres diferents de components per tal de comparar els resultats obtinguts per MCR. En general, l'ajust millora a mesura que s'incrementa el rang, però si s'escull un rang massa elevat, els

resultats obtinguts no són raonables i s'obté un sobreajust de les dades, o en alguns casos fins i tot un empitjorament de les dades (degut a la dificultat de complir les restriccions imposades).

En el cas d'estudis de sistemes de reacció química i d'altres, és freqüent que es presenti la situació de deficiència de rang [54]. Aquest problema es presenta quan en un sistema tancat es detecta un nombre de components linealment independents inferior al nombre vertader de components existents en el sistema, és a dir, que per exemple es detectarien menys fonts de contaminació de les que realment es troben presents en el medi estudiat. Aquest problema és conseqüència de la presència de dependències lineals en els perfils dels components en un dels modes. Existeixen diferents maneres d'eliminar els problemes de deficiència de rang, una de les més importants consisteix en emprar estratègies d'augmentació de matrius de dades com les abans descrites. Quan les diferents matrius simultàniament analitzades aporten informació independent dels diferents components presents, aleshores molt probablement serà possible eliminar aquest tipus de deficiència [54, 58-59].

En el cas de l'anàlisi de matrius augmentades, el càlcul del rang també és útil a l'hora de determinar el model que segueixen les dades (només bilineal o també trilineal). El rang de les matrius augmentades es pot calcular en les seves diferents direccions o modes possibles d'augmentació, per files, per columnes, i en profunditat (en cas de tenir-ne suficients, és a dir, que el nombre de matrius simultàniament analitzades sigui més gran que el nombre de components del sistema), segons es disposin una matriu al costat de l'altra (augmentació o desplegament en la primera direcció), una a sobre de l'altra (augmentació o desplegament en la segona direcció) o es desplegui cada matriu en un vector llarg i es posin una al costat de l'altra (augmentació o desplegament en la tercera direcció). Quan en les tres direccions de desplegament o augmentació s'obté el mateix nombre de components o rang, el conjunt de dades en les tres direccions podrà descomposar-se adequadament mitjançant un model trilineal i, en canvi, quan els seus rangs en les tres direccions de desplegament siguin diferents, el conjunt de dades no s'explicarà bé segons un model trilineal, però molt probablement encara podrà ser descrit correctament amb un model bilineal. Aquest tipus de comprovació és molt senzilla i útil per tal de comprovar l'estructura del conjunt de dades. En el cas que el nombre de matrius analitzades simultàniament sigui petit (igual o inferior al nombre de components del sistema), només caldrà comprovar el rang de les matrius augmentades per files i per columnes (en els dos primers modes).

### PCA respecte a MCR-ALS

Els algorismes o mètodes de descomposició bilineals utilitzats en MCR i PCA són totalment diferents, de manera que les matrius resultants d'*scores* i *loadings* també ho seran. De fet,

## Capítol 2

---

degut a la presència de les anomenades ambigüitats rotacionals (a més de les d'escala) associades a l'anàlisi de factors [36], existeixen un nombre indeterminat de possibles descomposicions matemàtiques d'una matriu de dades que reproduïxen igualment bé la matriu de dades original. Dit d'una altra manera, mentre que en PCA, les solucions obtingudes són úniques degut a les restriccions d'ortogonalitat, màxima variància i normalització, en el cas de MCR no ho són sempre, ja que les restriccions que s'imposen són d'una altra mena. Les solucions obtingudes per PCA són solucions úniques que tenen unes propietats matemàtiques òptimes (ortogonalitat i màxima variància explicada en ordre decreixent), però generalment, no tenen un significat físic i per tant, no són directament interpretables. No obstant, la informació continguda en les dades continua essent la mateixa que en les solucions obtingudes per MCR, amb la diferència que aquesta informació està expressada de manera diferent. Les solucions buscades per MCR es troben sotmeses a una sèrie de restriccions que els proporcionen significat físic. En canvi, aquestes es trobaran correlacionades entre si (no són linealment independents). A partir del coneixement físic del sistema, es podran restringir les solucions de manera que compleixin un determinat nombre de condicions conegudes. Així es podrà limitar de forma considerable el nombre de solucions possible i obtenir, eventualment, les solucions "vertaderes". Recentment, s'han proposat diversos mètodes alternatius per tal d'avaluar el grau d'ambigüitat associat a unes determinades solucions MCR [43].

Quan s'apliquen restriccions d'ortogonalitat, com en el cas del PCA, les variàncies  $R^2$  pels components individuals d'un model particular són additives i la seva suma és igual a la  $R^2$  total explicada per tot el model. Això és totalment diferent en el cas de models on no s'utilitzen restriccions d'ortogonalitat, com ara MCR-ALS, en els quals la suma de les variàncies  $R^2$  calculades individualment per cada component excedeix la variància explicada total del model, calculada incloent simultàniament tots els components. La raó d'aquesta diferència és que les variàncies explicades dels components individuals no ortogonals se superposa, i inclou les covariàncies existents entre aquests components, les quals són diferents de zero. Això és també una conseqüència de la correlació que pot existir entre els perfils dels components purs reals en els sistemes ambientals i, per tant, de la correlació existent entre els perfils dels components trobats per MCR. En aquest cas, la superposició de perfils porta associada també una superposició de variància explicada, que es pot mesurar, per exemple, a partir de la diferència entre la suma de les variàncies explicades per cada component individualment en el model de la variància explicada i la variància dels components considerats conjuntament en el model.

En aquesta Tesi, l'anàlisi quimiomètrica de les dades per PCA s'ha realitzat utilitzant la PLS Toolbox (Eigenvector Research Ltd., Manson, WA, USA), i per MCR-ALS a partir d'una



## **Metodologies aplicades a l'anàlisi de dades ambientals**

col·lecció de rutines quimiomètriques essencials i avançades [34, 41] que treballen també dins de l'entorn de programació, anàlisi numèrica i visualització de dades del software MATLAB (The Mathworks, MA, USA).



## **Article científic 1**

**Article 1: Quality assessment of the Multivariate Curve Resolution Alternating Least Squares (MCR-ALS) method for the investigation of environmental pollution patterns**

Marta Terrado, Damià Barceló and Romà Tauler

*Environmental Science and Technology* 43 (2009) 5321-5326



### 2.1.2.6. Resultats de l'avaluació del mètode MCR-ALS per a la investigació de pautes de contaminació en aigües superficials

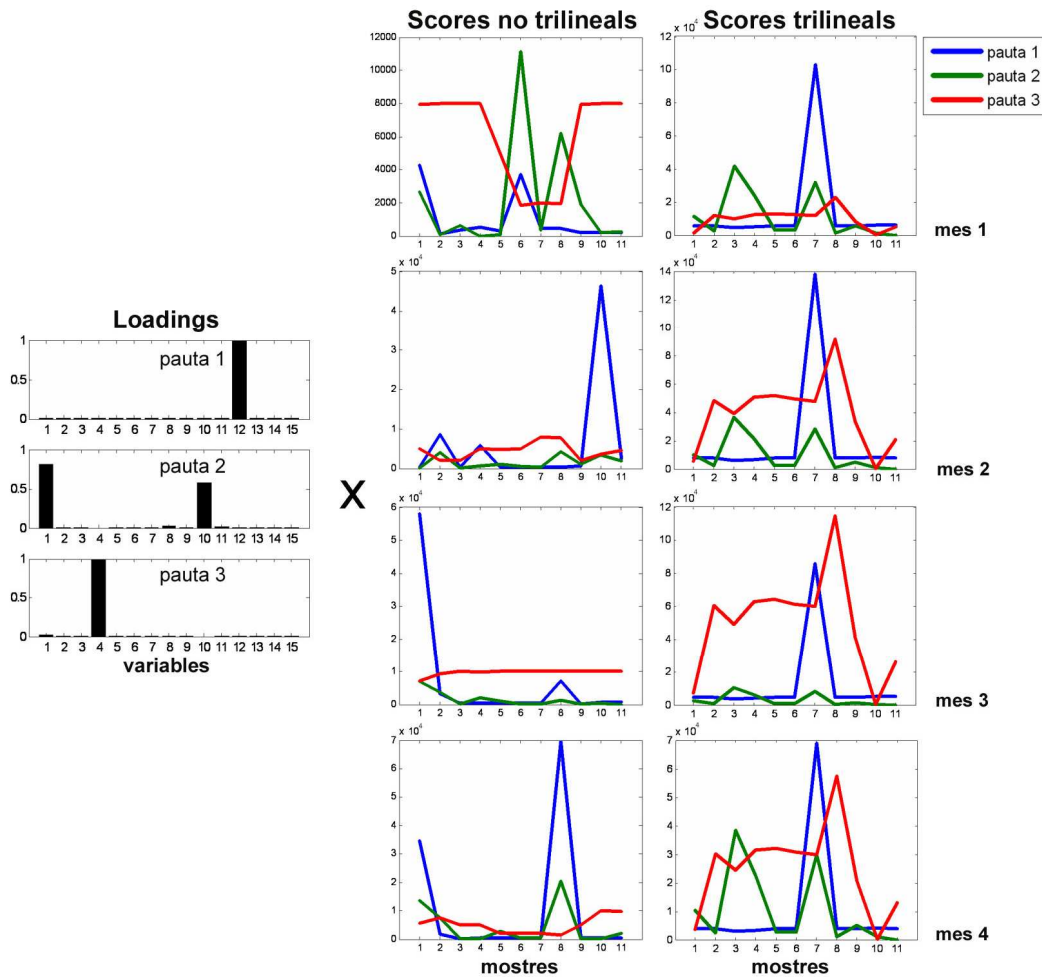
A l'article científic número 1, s'ha presentat l'avaluació de la qualitat dels resultats obtinguts mitjançant el mètode MCR-ALS en el cas de la investigació de fonts o pautes de contaminació en aigües superficials a partir de dades de monitoratge ambiental. A més, diversos conjunts de dades simulades han estat dissenyats per tal de poder realitzar la validació del mètode per a les diverses possibles estructures de dades obtingudes (simplement bilineal o també trilineal) a partir de diferents estudis de seguiment ambiental. A partir de les simulacions utilitzades, també ha estat possible contrastar l'efecte de l'anàlisi directa de les dades originals (sense cap pretractament) respecte a l'anàlisi de dades que han estat prèviament pretractades, per exemple per procediments d'escalat. Finalment, també s'ha estudiat l'efecte de la propagació del soroll sobre les fonts de contaminació resoltes, per tal d'establir diferències entre aquesta propagació i l'efecte de les possibles ambigüitats rotacionals inherents als mètodes de resolució multivariant de corbes.

Per a la preparació de les matrius de dades utilitzades en la present avaluació, s'han pres com a referència les dades de l'estudi experimental sobre la contaminació de productes fitosanitaris als canals del delta del riu Ebre durant la principal estació de creixement de l'arròs (veure article científic número 4). S'han simulat, per tant, 4 matrius de dades individuals (corresponents als 4 mesos de l'aplicació d'aquests productes). Aquestes matrius han estat creades assumint el model bilineal descrit a l'Equació 2.11, descrita a l'apartat anterior. Per a cada matriu individual  $\mathbf{D}_k$ ,  $k=1,2,3,4$ , s'ha suposat una composició determinada de les fonts de contaminació de productes fitosanitaris (files de la matriu  $\mathbf{F}^T$  o matriu de *loadings*) i una distribució d'aquestes fonts a les mostres (columnes de les matrius  $\mathbf{G}_k$  o matriu d'*scores*), diferent per a cada matriu de dades. D'aquesta manera, cada matriu  $\mathbf{D}_k$  expressarà les concentracions dels diferents productes fitosanitaris a cada una de les mostres analitzades durant les 4 campanyes. Tal com s'observa en el gràfic de *loadings* de la Figura 2.11, en aquest cas les pautes de contaminació són bastant ortogonals (no existeix gaire superposició de variables en les 3 fonts de contaminació simulades). En aquesta figura, la contribució de les variables minoritàries als gràfics de *loadings* gairebé no es percep, però aquestes sí que s'hi troben presents, amb contribucions que tot i ser baixes, són diferents de zero. Aquesta és una situació freqüentment trobada a la pràctica, i per aquest motiu ha estat utilitzada en la preparació de les dades d'aquest estudi.

S'han simulat dos tipus diferents de matrius d'*scores*  $\mathbf{G}_k$ , una d'elles seguint una estructura no trilineal i l'altra amb una estructura trilineal. En el cas de la matriu d'*scores* no trilineal de la

Figura 2.11, la distribució de les tres pautes de contaminació té unes característiques diferents per a cada matriu individual. En canvi, quan la matriu d'scores presenta una estructura trilineal, la distribució de les pautes de contaminació a les mostres és invariant en el temps, canviant només en la seva escala relativa però no en la seva forma. Com ja s'ha explicat en la descripció del mètode MCR-ALS, per tal de comprovar si les dades presenten una estructura trilineal, només cal aplicar la descomposició de valors singulars (SVD) a les matrius de dades augmentades per files i per columnes, i comprovar si el nombre de components principals coincideix. Si és així, les dades segueixen un model trilineal.

Del producte de la matriu de *loadings* amb les diferents matrius d'scores, se n'obtenen les diferents matrius que contenen la concentració de productes fitosanitaris a les diferents mostres en quatre condicions de temps diferents (mes 1 a 4), les quals seran posteriorment analitzades mitjançant el mètode MCR-ALS.



**Figura 2.11.** *Loadings* i *scores* (amb estructura no trilineal i estructura trilineal) de les pautes de contaminació simulades durant 4 mesos. Variables: 1, MCPA; 2, mecoprop; 3, 2,4-D; 4, bentazona; 5, simazina; 6, isoproturon; 7, clortoluron; 8, atrazina; 9, diuron; 10, propanil; 11, molinat; 12, alaclor; 13, metolaclor; 14, fenitrotion; 15, diazinon. Mostres: 1, canal d'entrada 1; 2-9, canals d'irrigació 1 al 8; 10, llac de l'Encanyissada; 11, Badia dels Alfacs

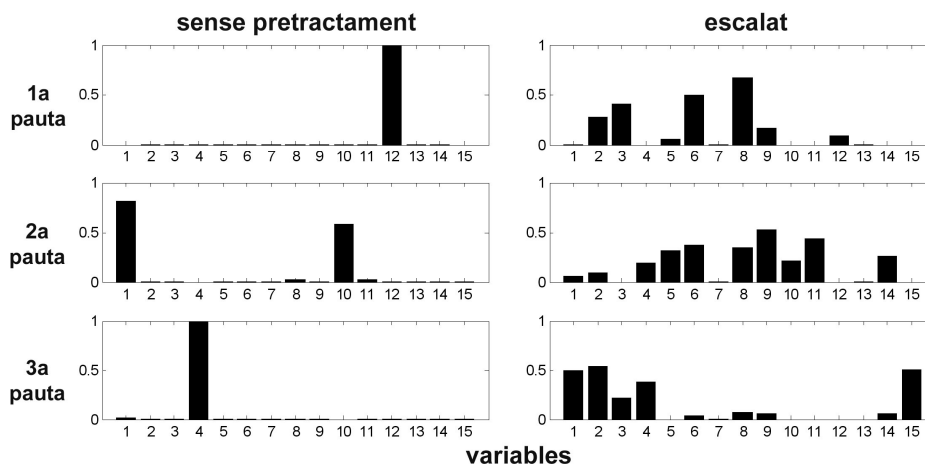
Ambdós tipus de conjunts de dades (trilineal i no trilineal) han estat preparats per duplicat. En un d'ells s'ha addicionat soroll mentre que a l'altre no. Tal com es descriu a l'article científic número 1, la matriu de soroll  $\mathbf{E}_k$ ,  $k=1,2,3,4$ , presenta la contribució de dos tipus diferents d'error per cadascuna de les matrius de dades: un tipus d'error que és constant per cada variable (en aquest cas del 10% de la màxima concentració de la variable corresponent) i un altre que és proporcional al valor mesurat (també del 10%). El valor final de l'error segueix una distribució aleatòria Gaussiana amb un valor mitjà igual a zero i desviació estàndard igual a l'arrel quadrada de la suma dels dos tipus d'error (constant i proporcional) calculats per cada valor particular de la matriu de dades  $\mathbf{D}_k$ . Ja que l'addició de l'error estimat d'aquesta manera pot donar lloc a valors negatius, els quals no tenen sentit per aquest tipus de dades ambientals, en aquest cas els valors negatius han estat substituïts per un valor positiu molt baix i proper a zero. Aquest valor presenta un efecte negligible sobre els resultats finals obtinguts.

Les 4 matrius  $\mathbf{D}_k$  finalment generades (amb estructura trilineal i no trilineal, amb soroll i sense soroll) han estat analitzades mitjançant el mètode MCR-ALS, i els seus resultats han estat posteriorment comparats per tal de veure quin és l'efecte que tenen sobre el resultat final (perfils d'*scores* i *loadings* obtinguts), els diferents tipus de restriccions i de pretractaments aplicats a les dades. Els principals pretractaments investigats han estat: (1) no pretractament, (2) escalat de les dades, (3) escalat de les dades i posterior desescalat dels perfils obtinguts (*loadings*) per als factors. El desescalat dels *loadings* consisteix en aplicar l'operació inversa de l'escalat sobre les dades (és a dir, es multipliquen cadascun dels valors de les variables en els *loadings* per la desviació estàndard prèviament estimada pels valors de la mateixa variable a les dades originals), i s'aplica per tal que els perfils obtinguts quan s'escalen les dades siguin comparables amb els perfils obtinguts quan no s'escalen les dades. L'escalat sobre les variables de les dades originals només hauria d'afectar als *loadings* i no als *scores*. Pel que fa a la comprovació de l'efecte de les restriccions, l'efecte principalment estudiat ha estat el de la restricció de trilinearitat durant l'anàlisi MCR-ALS: (4) aplicació de la restricció de trilinearitat utilitzant dades brutes o sense pretractament i (5) aplicació de la restricció de trilinearitat utilitzant dades escalades. L'efecte dels diferents factors s'ha investigat a partir de la comparació entre els perfils d'*scores* i *loadings* simulats (perfils teòrics) i els obtinguts després de l'aplicació d'MCR-ALS per cadascuna de les opcions 1 a 5 descrites en aquest paràgraf.

### Efecte de l'escalat de les dades

En aquells casos en què les variables presenten diferències en les seves magnituds, és aconsellable realitzar algun tipus d'escalat abans d'aplicar l'anàlisi quimiomètrica. En cas contrari, només aquelles variables que dominen la variància de les dades presentaran una

elevada contribució significativa en els perfils de *loadings* obtinguts. En el cas de la Figura 2.12, per exemple, es mostren els resultats obtinguts de l'aplicació de l'anàlisi MCR-ALS sobre la matriu augmentada de dades experimentals (no simulades) amb les concentracions de productes fitosanitaris al delta del riu Ebre durant l'any 2005, en el cas en què les dades no han estat pretractades i en el cas en què les dades han estat escalades. Quan no es realitza cap pretractament a les dades, només les variables alaclor (1a pauta o font de contaminació), MCPA i propanil (2a font de contaminació) i bentazona (3a font de contaminació) presenten una contribució significativa en els perfils de *loadings*, mentre que la resta queden amagades. És doncs en casos com aquest, que l'escalat de les variables és particularment important, ja que facilita la identificació de les contribucions de les diferents variables menors (amb concentracions baixes dels compostos químics) en els perfils de *loadings* i, conseqüentment, contribuirà a millorar la interpretació ambiental de les diferents fonts de contaminació identificades.



**Figura 2.12.** Comparació dels *loadings* de les tres fonts de contaminació obtingudes a partir de l'anàlisi de les dades experimentals de productes fitosanitaris al delta del riu Ebre, dades sense pretractament i dades escalades. Variables: 1, MCPA; 2, mecoprop; 3, 2,4-D; 4, bentazona; 5, simazina; 6, isoproturon; 7, clortoluron; 8, atrazina; 9, diuron; 10, propanil; 11, molinat; 12, alaclor; 13, metolaclor; 14, fenitrotion; 15, diazinon

No obstant això, a l'hora d'escalar les variables, cal anar en compte amb aquelles que presenten valors massa petits i que a més gairebé no varien, ja que en escalar-los, es corre el risc d'augmentar innecessàriament la seva importància relativa. Això té lloc quan s'inclouen mesures de variables amb valors per sota del límit de detecció en gairebé tots els casos. Aquest efecte es presenta en el cas d'algunes de les variables de la segona font de contaminació obtinguda a la Figura 2.12 després de l'aplicació de l'anàlisi quimiomètrica sobre les dades prèviament escalades. A l'apartat 2.1.2.3 d'aquesta Tesi, s'han exposat diferents alternatives que permeten evitar aquest efecte negatiu, per exemple, a partir de la divisió dels valors petits i pràcticament constants d'aquestes variables per un valor numèric relativament gran (com ara 20) que fa que finalment la contribució d'aquestes variables continuï sent petita a



les dades un cop escalades. És important també tenir en compte que l'escalat és una operació matemàtica que, en principi, només afectarà a les variables (ja que només s'aplica sobre les columnes de la matriu de dades) mentre que aquest pretractament no ha de tenir cap efecte sobre les mostres (sobre les files de la matriu de dades). Tot i això, algunes de les diferències observades en la recuperació dels perfils d'*scores* seran degudes als efectes del soroll, inherent a les dades de tipus ambiental. L'escalat també presentarà altres efectes, com per exemple un repartiment més uniforme de la variància de les dades en les fonts de contaminació identificades quan aquestes són prèviament escalades (veure Figura 2.13).

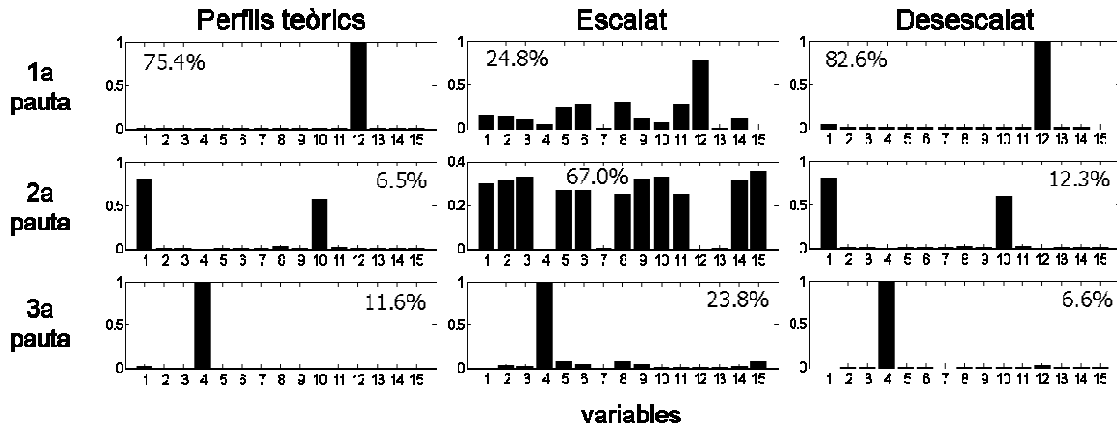


Figura 2.13. Comparació dels perfils de *loadings* teòrics, els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades, i els perfils obtinguts desescalant els perfils escalats (dades simulades amb soroll)

A la Figura 2.13 es poden observar els tres casos estudiats. En primer lloc, es mostren els perfils de *loadings* simulats o teòrics (bastant ortogonals i amb una contribució petita de la majoria de les variables), en segon lloc, es mostren els perfils de *loadings* obtinguts aplicant el mètode MCR-ALS sobre la matriu de dades augmentada (pels 4 mesos) prèviament escalada. En aquest cas, la importància d'algunes de les variables que presentaven poca contribució en els perfils de *loadings* teòrics, s'ha vist incrementada per l'escalat. També s'observa clarament l'efecte de la redistribució de la variància, la qual en els perfils teòrics es trobava majoritàriament concentrada en la primera font de contaminació i, en les dades escalades, es troba més repartida entre les tres fonts identificades. Finalment, es mostren els perfils recuperats un cop els perfils de *loadings* obtinguts a partir de les dades escalades han estat posteriorment desescalats. L'operació de desescalat, es mostra en forma matemàtica a les següents equacions:

$$\mathbf{X} = \{x_{i,j}\} \tag{Equació 2.25}$$

$$s_j = \sqrt{\frac{\sum_{i=1}^I (x_{i,j} - \bar{x}_j)^2}{I-1}} \tag{Equació 2.26}$$

$$sX_{ij} = \frac{x_{ij}}{s_j} \quad \text{Equació 2.27}$$

$$\mathbf{SX} = \{sx_{ij}\} \quad \text{Equació 2.28}$$

$$lx_{n,j} = lSX_{n,j}s_j \quad \text{Equació 2.29}$$

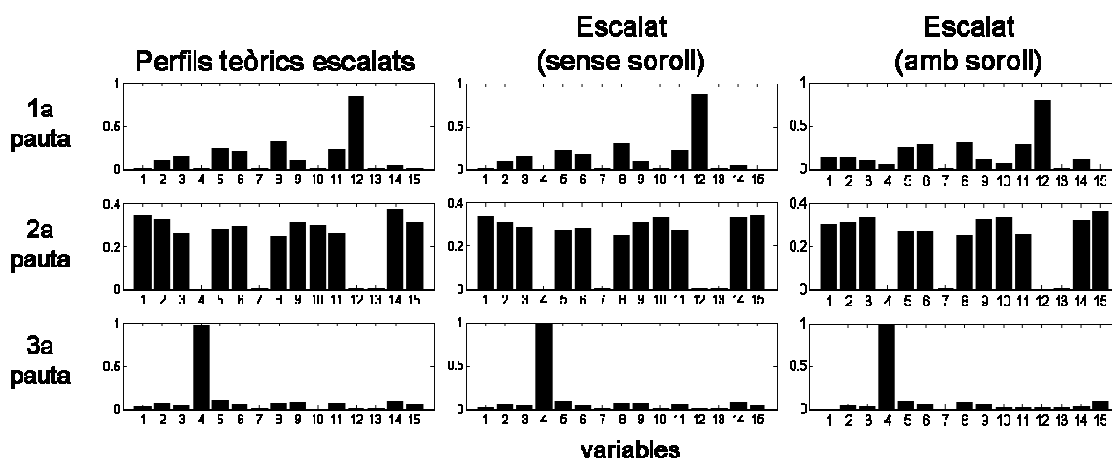
A les Equacions 2.25 a 2.29,  $x_{ij}$  correspon a la mesura individual de la variable  $j$  en una mostra  $i$ . El conjunt de les mesures de les diferents variables a les diferents mostres donen lloc a la matriu de dades  $\mathbf{X}$  (Equació 2.25). El paràmetre  $s_j$  (Equació 2.26) correspon a la desviació estàndard (descrita a l'apartat 2.1.2.3 d'aquesta Tesi) de la variable  $j$  (per a totes les mostres), i  $sx_{ij}$  són els valors de les mesures originals escalades, els quals donen lloc a la matriu  $\mathbf{SX}$  (Equació 2.28). El valor dels *loadings* desescalats per la variable  $j$  en el component  $n$ , ve donat per  $lx_{n,j}$ , un cop els *loadings* obtinguts per les dades escalades,  $lSX_{n,j}$ , s'han multiplicat per  $s_j$ , la desviació estàndard obtinguda a partir dels valors originals de les mesures de la variable (Equació 2.29).

És interessant comprovar que l'operació de l'escalat no comporta cap pèrdua d'informació significativa, ja que quan els perfils de contaminació obtinguts a partir de l'anàlisi de les dades escalades (perfils escalats) es desescalen, s'obtenen altra vegada uns perfils molt similars als teòrics (veure Figura 2.13 i els valors de recuperació dels perfils de *loadings* escalats i desescalats de la Taula 2 en l'article científic número 1). Finalment, com a resultat de l'estudi de les simulacions realitzades mitjançant MCR-ALS, s'observa que sovint l'escalat té tendència a disminuir el sobreajust inherent al model MCR-ALS. A la Taula 2 de l'article científic número 1, es mostra com en la majoria dels casos de dades simulades amb soroll, el sobreajust obtingut analitzant dades sense pretractar (diferència entre la variància explicada,  $R^2$ , amb l'aplicació del model a dades sense pretractar i la  $R^2$  teòrica), és superior al sobreajust que s'obté mitjançant l'anàlisi de dades escalades (diferència entre la  $R^2$  obtinguda amb l'aplicació del model a dades escalades i la  $R^2$  teòrica).

### Efecte del soroll de les dades

Els efectes del soroll existent en els conjunts de dades ambientals poden ser avaluats comparant els resultats obtinguts en l'aplicació del mètode MCR-ALS sobre dades simulades en presència i en absència de soroll. En absència de soroll, les petites diferències detectades en la recuperació dels perfils seran degudes fonamentalment a les ambigüitats rotacionals inherents als mètodes de resolució de corbes. Com ja s'ha explicat anteriorment en l'apartat de descripció del mètode MCR-ALS, aquest tipus d'ambigüitat ocorre quan dos o més components se superposen, i una combinació lineal d'aquests ajusta igualment les dades i compleix

igualment bé les restriccions del sistema, de manera que les solucions no són úniques. A més, en presència de soroll, és més complicat distingir aquells components que expliquen la variància estructural de les dades d'aquells que només expliquen soroll.



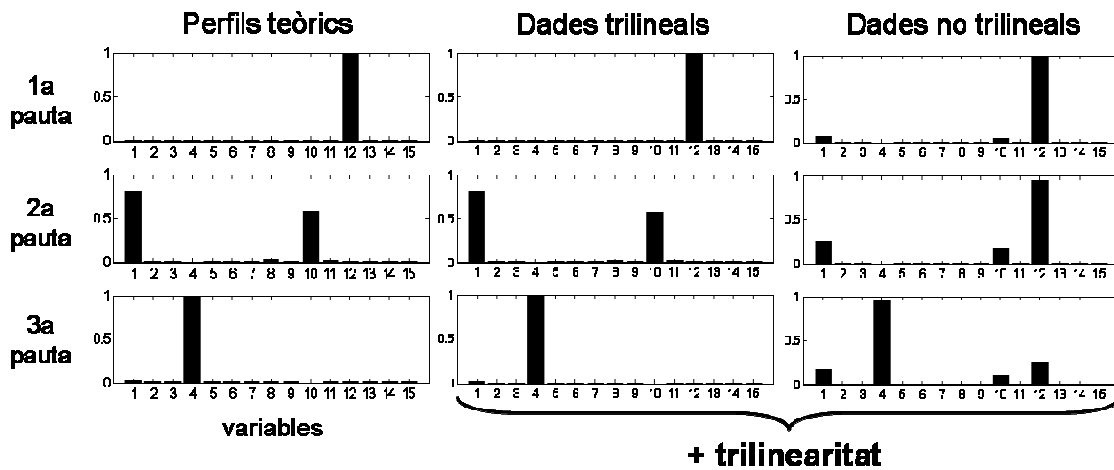
**Figura 2.14.** Comparació dels perfils de *loadings* teòrics escalats i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS sobre les dades escalades amb i sense soroll

A la Figura 2.14, s'observa com en el cas de les dades amb soroll, els resultats obtinguts són pitjors i apareixen més dificultats en la correcta recuperació dels perfils originals o teòrics. El soroll és propagat i amplificat des de les dades experimentals originals fins als resultats de l'MCR-ALS com a conseqüència de les diferents operacions matemàtiques realitzades (començant, per exemple, amb l'escalat de les dades), la qual cosa fa més difícil la correcta identificació especialment d'aquells components més minoritaris, que presenten una relació senyal/soroll més petita. A l'article científic número 1, es dona una taula (Taula 2) amb els valors de la recuperació dels perfils teòrics de *loadings* ( $r^2$ ) un cop les dades amb i sense soroll han estat analitzades aplicant l'escalat. Les recuperacions dels perfils *d'scores* es poden consultar a la informació suplementària d'aquest article. S'observa que les recuperacions sempre són lleugerament més bones en els casos sense soroll per als tres perfils de *loadings*. En aquesta mateixa taula també es pot observar com afecta el soroll en els casos en què s'aplica el desescalat als perfils de *loadings* obtinguts de l'anàlisi de la matriu escalada. Les recuperacions sempre són molt bones en el cas del desescalat dels perfils en absència de soroll ( $r^2 = 1$ ), mentre que per les dades amb soroll, aquest entra a dins dels perfils obtinguts mitjançant l'anàlisi de les dades escalades i els seus efectes es mantenen també durant el desescalat d'aquests perfils de *loadings*.

**Efecte de la restricció de trilinearitat**

L'efecte que té l'aplicació de la restricció de trilinearitat sobre els resultats de l'MCR-ALS s'ha avaluat a partir de la comparació d'aquests resultats quan la restricció s'aplica sobre matrius de dades amb estructura trilineal així com quan s'aplica sobre matrius de dades amb estructura no trilineal.

L'assumpció d'un model trilineal simplifica considerablement la interpretació ambiental de les pautes de contaminació identificades, ja que no només la composició de les fonts de contaminació (*loadings* en el segon mode) és la mateixa per totes les situacions o condicions (per exemple, períodes de temps o compartiments ambientals), sinó que la seva distribució espacial també es manté constant (*loadings* en el primer mode o *scores*), variant només en un factor d'escala (*loadings* en el tercer mode).



**Figura 2.15.** Comparació dels perfils de *loadings* teòrics i els perfils obtinguts mitjançant l'aplicació del mètode MCR-ALS amb la restricció de trilinearitat per dades trilineals i dades no trilineals (dades simulades amb soroll)

A la Figura 2.15, es mostren, en primer lloc, els perfils de *loadings* teòrics simulats, i a continuació els perfils de *loadings* obtinguts un cop s'ha aplicat l'anàlisi MCR-ALS sobre les dades amb soroll, aplicant la restricció de trilinearitat: (i) per a dades que segueixen un model trilineal i (ii) per a dades que no segueixen un model trilineal (veure els perfils d'*scores* trilineals i no trilineals representats a la Figura 2.11).

Com a resultat de l'aplicació de l'MCR-ALS amb la restricció de trilinearitat sobre el conjunt de dades amb estructura trilineal, els perfils de *loadings* obtinguts coincideixen força bé amb els perfils teòrics (algunes petites diferències que poden existir, seran degudes principalment a l'efecte del soroll). En canvi, quan l'anàlisi MCR-ALS amb la restricció de trilinearitat és aplicat sobre dades que no segueixen el model trilineal, els resultats obtinguts són erronis, ja que les dades es veuen forçades a seguir una estructura que en realitat no posseeixen. En el cas mostrat a la Figura 2.15, per exemple, les tres fonts de contaminació simulades inicialment

(perfils teòrics), no es recuperen per separat sinó de manera barrejada en els tres perfils de *loadings* resultants de l'anàlisi de les dades no trilineals. Per tant, els resultats obtinguts de l'assumpció d'un model trilineal per aquest tipus de dades no trilineals no seran correctes, de manera que les fonts o pautes de contaminació identificades a arrel de l'anàlisi quimiomètrica seran errònies (o no coincidents amb els perfils teòrics). Aquestes recuperacions es troben quantificades a l'article científic número 1, dins de la Taula 2 pel que fa a les recuperacions dels perfils de *loadings* i a la informació suplementària d'aquest mateix article pel que fa a les recuperacions dels perfils d'*scores*.

Com ja s'ha explicat dins de l'apartat de descripció del mètode MCR-ALS, l'ambigüitat rotacional pot ser completament eliminada mitjançant l'aplicació de la restricció de trilinearitat durant l'anàlisi MCR, de manera que les solucions obtingudes en aquests casos siguin úniques. És per això que és recomanable aplicar aquesta restricció en tots aquells casos en què sigui possible i adequat fer-ho. No obstant, en els estudis de resolució amb MCR-ALS, cal tenir en compte que aquesta restricció només té sentit d'ésser imposada quan les dades presenten una estructura trilineal. En aquest últim cas, a més d'eliminar l'ambigüitat rotacional, la resolució de les fonts o pautes de contaminació s'aconsegueix amb un menor sobreajust del model (la diferència entre la variància explicada  $R^2$  mitjançant l'aplicació de l'MCR-ALS i la  $R^2$  teòrica disminueix quan s'aplica la trilinearitat). I aquest possible sobreajust encara disminueix més quan les dades es troben prèviament escalades (veure resultats obtinguts per  $R^2$  a la Taula 2 de l'article científic número 1).

### 2.1.2.7. Conclusions

Les conclusions principals extretes de la secció de mètodes d'anàlisi multivariant, es recullen a continuació:

- Es posa de manifest el potencial dels mètodes quimiomètrics PCA i MCR-ALS per a l'anàlisi de taules de dades ambientals
- Es valida la utilització del mètode MCR-ALS per a l'anàlisi de dades i per a la identificació de les principals fonts o pautes de contaminació que afecten les aigües superficials (i, per extensió, altres compartiments ambientals)

## Capítol 2

---

- Es comprova que l'aplicació de qualsevol restricció durant l'anàlisi de dades multivariant ha d'estar d'acord amb les característiques i l'estructura interna de les dades (sinó, els resultats obtinguts poden ser erronis)
- S'identifica que l'escalat de les dades (i l'autoescalat, en el cas de PCA) és un pretractament adient i recomanable per tal de facilitar la interpretació ambiental dels resultats obtinguts, especialment quan les magnituds o les unitats de les variables difereixen. No obstant, cal anar en compte amb aquelles variables que presenten valors petits que gairebé no varien, ja que es poden veure exagerades per aquesta operació
- S'aconsella l'aplicació de la restricció de trilinearitat per a l'anàlisi de dades en tres direccions o modes sempre que sigui possible, ja que així les solucions obtingudes són úniques (s'elimina l'ambigüitat rotacional) i són més fàcilment interpretables des del punt de vista ambiental. A més, s'observa que l'aplicació de la trilinearitat també permet disminuir el possible sobreajust del model, associat als models bilineals (sobretot quan els dades es troben prèviament escalades)
- Es confirma que per a la identificació d'una determinada pauta o font de contaminació, el senyal o contribució de les variables que descriuen aquesta font ha de ser superior al nivell de soroll contingut a les dades (sinó, el senyal quedarà tapat pel soroll)

### 2.1.3. Mètodes geostatístics

Com ja s'ha comentat, les dades ambientals es troben típicament distribuïdes en l'espai i en el temps. Per tant, el coneixement del valor d'un atribut, com per exemple la qualitat d'un mineral o bé la concentració d'un determinat contaminant, presenta molt poc interès quan no es coneix la seva localització o bé el moment de la seva mesura, i això no és tingut en compte durant el procediment d'anàlisi de les dades. És en aquest sentit que la geostatística [11, 14, 60-61] proporciona un conjunt d'eines estadístiques que permeten incorporar les coordenades espacials i temporals de les observacions durant el processament de les dades. D'aquesta manera, es pot descriure la possible continuïtat espacial de qualsevol fenomen natural. A més d'interpol·lar el valor d'un determinat atribut en aquelles localitzacions on aquest no ha estat o no ha pogut ser mesurat, els mètodes geostatístics també permeten obtenir una modelització de les incerteses que existeixen en la predicció dels valors d'aquest atribut.

La geostatística està basada en el concepte de funcions estocàstiques, en el qual el conjunt de valors mesurats es considera com un conjunt de variables aleatòries espacialment dependents. Una variable aleatòria és una variable que pot prendre una sèrie de valors diferents en funció d'una distribució de probabilitats. En les funcions aleatòries, cada realització de la funció pot ser concebuda com la suma d'una component estructurada i una altra aparentment erràtica o estocàstica. La component estructurada és la que permet assegurar que els valors observats tinguin una variació sistemàtica, és a dir, que si ens trobem, per exemple, en una zona on s'han realitzat diverses mesures per sobre del que és normal, el més probable és que les mesures addicionals també siguin elevades. D'altra banda, la component aleatòria és la que dificulta predir amb exactitud el valor d'aquestes mesures hipotètiques, ja que aquestes presenten fluctuacions impredecibles.

Amb la finalitat de predir el valor de determinats atributs en localitzacions no mesurades, existeixen diversos algorismes, com és el cas del mètode anomenat *kriging* [62].

### **Kriging**

El mètode *kriging* està basat en la teoria de les variables regionalitzades [63], la qual assumeix que la variació espacial d'una variable presentada a localitzacions específiques és estadísticament homogènia a través d'una superfície definida. Es tracta d'una aproximació geostatística que quantifica l'estructura de correlació espacial entre les diferents localitzacions en funció de la seva distància de separació. Tot i que existeixen diferents tipus de *kriging*, el *kriging* ordinari és el que acostuma a utilitzar-se més, i el que serà descrit en aquest apartat. La predicció del valor d'un atribut en una localització determinada es pot descriure de la següent manera:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad \text{Equació 2.30}$$

On  $\hat{Z}(s_0)$  és el valor de l'atribut a la localització predita  $s_0$ ,  $Z(s_i)$  són els valors de l'atribut a les localitzacions mesurades  $s_i$ ,  $i=1, \dots, n$ , on  $n$  és el nombre de valors per als quals es disposa d'una mesura, i  $\lambda_i$  correspon als pesos que es donen a cada valor mesurat, i que dependran de:

- (i) el semivariograma empíric (explicat a sota)
- (ii) les distàncies a la localització predita
- (iii) les relacions espacials entre els valors mesurats al voltant de la localització que es vol predir

Les equacions utilitzades durant el procediment de predicció de valors en el mètode *kriging* no dependran dels valors mesurats de les variables sinó només de la seva posició i del semivariograma. Els passos que es segueixen per portar a terme la tasca de predicció es descriuen a continuació:

- 1. Estimació del semivariograma empíric.** La funció bàsica que descriu la variabilitat espacial del fenomen d'interès es coneix com el semivariograma empíric (inversa de l'anomenat diagrama de correlació experimental). El semivariograma  $\hat{\gamma}(h)$  s'utilitza per quantificar l'estructura de correlació espacial o covariància, o sigui per a explorar les relacions que existeixen entre les dades. Es tracta doncs d'una mesura de la dissimilaritat mitjana existent entre parells de dades separades per una distància o vector  $h$ , i es calcula de la següent manera:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(s_i + h) - z(s_i)]^2 \quad \text{Equació 2.31}$$

On  $N(h)$  és el nombre de parells de mesures separades per una distància  $h$  que presenten un valor mesurat de  $z(s_i)$ .

- 2. Determinació del model que segueixen les dades.** Tot i que el semivariograma empíric proporciona força informació (discreta) sobre el comportament espacial d'una variable, és convenient ajustar una funció contínua que faciliti la quantificació del grau i l'escala de la variació espacial i que permeti realitzar, posteriorment, estimacions acurades en les localitzacions no mesurades. En geostatística existeixen diversos models d'ajust de la funció, entre els quals els més utilitzats són l'esfèric, l'exponencial, el gaussià i el lineal. L'ajust de la funció permet obtenir una sèrie de paràmetres que són els que seran utilitzats per a realitzar la interpolació geostatística (*kriging*) i que defineixen el grau i l'escala de la variació espacial. Aquests paràmetres són el rang ( $A_0$ ), el *nugget* ( $C_0$ ), el *sill* ( $C_0+C$ ) i la proporció de la variància explicada ( $C/C_0+C$ ), sovint expressada en tant per cent (veure Figura 2.16). El rang ( $A_0$ ) és el valor de  $h$  (distància entre parells de dades) dins del qual té lloc l'estabilització, i per sobre del qual la semivariància ja deixa d'augmentar. El *nugget* ( $C_0$ ) és la variància no explicada pel model, calculada com la intercepció amb l'eix Y quan es fa la representació dels valors del semivariograma  $\hat{\gamma}(h)$  respecte a la separació entre parells de dades. També es



coneix com a variància error. La màxima semivariància identificada entre parells de punts s'anomena *sill* i descriu la variància que existeix entre el conjunt de mostres, que és el resultat de la suma de la variància error ( $C_0$ ) i la variància estructural explicada de les dades ( $C$ ). Per valors elevats d' $h$ , el semivariograma empíric s'estabilitza, indicant que ja no existeix cap correlació entre les localitzacions.

En general, al voltant d'uns 30 punts són necessaris per tal de generar un semivariograma raonable. La part més important del semivariograma és la seva forma al voltant de l'origen, ja que els punts més propers són els que reben més pes en el procés d'interpolació.

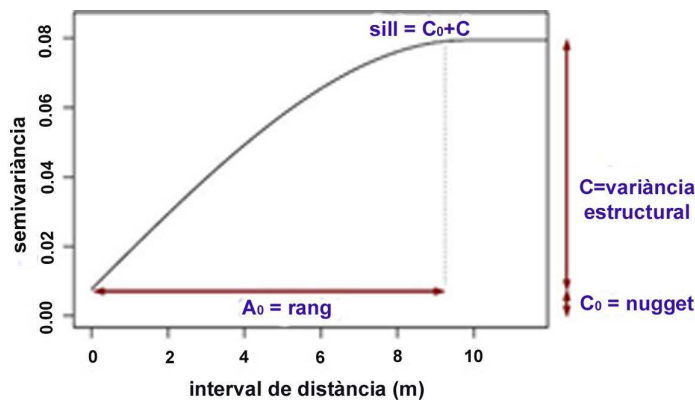


Figura 2.16. Paràmetres utilitzats per a l'ajust del semivariograma empíric a la funció matemàtica més idònia

3. **Creació de les matrius de pesos per al *kriging*.** Les equacions per al *kriging* ordinari es troben contingudes en matrius i vectors que depenen de l'autocorrelació espacial entre els valors a les localitzacions mesurades i predites. Caldrà calcular, en primer lloc, la matriu de covariància **C** entre els diferents parells de localitzacions  $i$ , posteriorment, una matriu **D** amb les distàncies entre les localitzacions mesurades i aquelles que es volen predir. D'aquesta manera, la matriu de pesos  $\lambda$  es calcula mitjançant el producte entre la matriu de distàncies i la matriu de covariàncies transposada:

$$\lambda = C^{-1}D \quad \text{Equació 2.32}$$

4. **Predicció.** A partir dels pesos assignats a les diferents localitzacions mesurades, es calcula una predicció per aquelles localitzacions per les quals no es coneixen els valors. Per al càlcul del valor de l'atribut en un punt en què aquest no ha estat mesurat, es realitza el producte de la matriu de pesos obtinguda i dels diferents valors a les localitzacions mesurades (veure Equació 2.30).

## Capítol 2

---

Abans de generar la superfície final, caldrà tenir una idea de la qualitat de la predicció obtinguda (com de bé el model prediu el valor de l'atribut a les localitzacions no mesurades). El mètode de la validació creuada (*cross-validation leave one out*) utilitza totes les dades menys una per tal d'estimar el model d'autocorrelació que s'utilitza per predir el valor prèviament no considerat. Després torna a posar la dada extreta i en treu una altra de nova per tal de tornar a predir el seu valor. I així successivament per a totes les dades del conjunt. Cada vegada es calcula la diferència entre el valor predit i el valor mesurat, i d'aquesta manera s'obté un resum estadístic amb els errors de la predicció, el qual pot ser utilitzat a mode de diagnòstic.

L'anàlisi SIG així com l'anàlisi geostatística de les dades s'ha dut a terme utilitzant el software ArcGis (ESRI, Redlands, CA, USA), que conté les següents aplicacions: (i) ArcMap, que permet visualitzar, crear i consultar mapes així com editar-ne la seva base de dades, (ii) ArcCatalog, que proporciona accés a les dades i eines de gestió espacial d'aquestes, incloent la visualització i edició de les metadades (dades sobre el mapa com ara la seva font, any, atributs, tipus de coordenades espacials, etc.), i (iii) les aplicacions de geoprocessament (incloent ArcToolbox), que proporcionen un entorn per a la implementació de l'anàlisi SIG, i que poden ser utilitzades individualment o bé combinades dins d'altres models per tal d'efectuar tasques d'anàlisi SIG, incloent la conversió de les dades a altres formats.

Per a realitzar l'anàlisi geostatística s'ha utilitzat l'extensió *Geostatistical Analyst*, integrada dins de l'aplicació ArcMap i concebuda per al modelatge avançat de superfícies mitjançant mètodes deterministes i geostatístics. Aquesta extensió ofereix eines per a realitzar una anàlisi exploratòria de les dades espacials, així com un assistent que guia a través del procés de creació de superfícies vàlides en termes estadístics. El *Geostatistical Analyst* constitueix una mena d'enllaç en la separació que existeix entre el SIG i la geostatística. Tot i que les eines geostatístiques fa temps que es troben disponibles, aquestes no acostumen a trobar-se plenament integrades dins de l'entorn de modelatge SIG. La integració entre les dues metodologies és important per tal que els usuaris dels SIGs siguin capaços de quantificar la qualitat dels seus models superficials, mesurant l'error estadístic associat a les superfícies predites.

L'ajust d'una superfície mitjançant l'eina *Geostatistical Analyst* està descrit per tres passos claus:

- Anàlisi exploratòria de les dades espacials
- Anàlisi estructural (càlcul i modelatge de les propietats de la superfície per localitzacions que es troben properes)
- Predicció de la superfície i avaluació dels resultats

En aquesta Tesi, l'operació *kriging* ha estat utilitzada amb finalitats purament funcionals, per a obtenir representacions gràfiques que facilitin la comprensió dels resultats obtinguts. En cap cas s'ha realitzat una validació del mètode ni tampoc una avaluació exhaustiva de l'error de predicció lligat al procés d'interpolació de les dades, com seria el cas d'altres estudis [64-65]. En tots els casos, la quantitat de punts disponibles ha estat inferior als necessaris per tal de poder obtenir una bona predicció. A més a més, l'adaptació del procés d'interpolació de dades encara no es troba ben desenvolupat dins del medi fluvial (en flux continu). En canvi, si que ha estat aplicat, en alguns casos, per a aigües subterrànies [66].

### 2.2. Mètodes deterministes

Per tal de realitzar la modelització determinista de la dinàmica d'una determinada conca hidrogràfica i del transport de contaminants dins d'aquesta conca, són necessaris diferents passos. En primer lloc, caldrà deduir l'estructura física de la conca, determinant-ne els seus límits, els patrons d'escolament, així com el tipus i ocupacions del sòl que s'hi troben presents. Posteriorment, caldrà simular els diferents processos hidrològics i d'erosió que hi tenen lloc. I, en darrera instància, es podrà passar a la simulació del transport de contaminants dins d'aquesta així com a l'aplicació de les diferents estratègies de gestió per tal de veure l'impacte que aquestes tenen sobre la qualitat de l'aigua de la conca.

Els primers models hidrològics informatitzats daten de fa una quarantena d'anys [66]. Es tractava de models essencialment globals, en els quals el cabal era simulat a nivell de conca, tractant-la com un tot indivisible. Ben aviat, però, comencen a aparèixer models que tenen en compte la variabilitat dels fenòmens hidrològics dins de les mateixes conques hidrogràfiques [67-70]. A més, també es proposen models de tipus físic, com per exemple el model SHE [71-72] orientat a la simulació de l'erosió, que a la pràctica requereixen una quantitat d'informació tan elevada, que de vegades es fa difícil el seu maneig i aplicació pràctica.

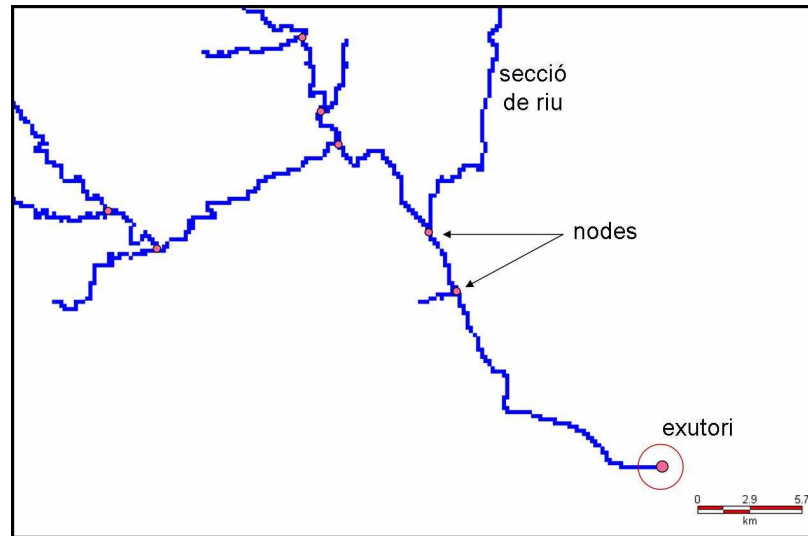
Cap d'aquests models havia estat encara concebut per tal de poder incorporar les dades procedents de la teledetecció o dels sistemes d'informació geogràfica (SIG) més recents. Amb aquesta finalitat, es crea una nova generació de models compatibles amb aquest nou tipus de dades [73]. Recentment, models hidrològics basats en el tractament de models numèrics d'elevacions (MNE) han estat desenvolupats [74], i models com l'anteriorment anomenat SHE s'han beneficiat d'aquestes millores. Paral·lelament, també s'han desenvolupat models compatibles amb la teledetecció i els SIGs [75].

#### 2.2.1. Model PHYSITEL

El model PHYSITEL (INRS-ETE, Québec) permet obtenir l'estructura de drenatge d'una conca hidrogràfica determinada a partir d'una base de dades de la conca incorporada com a informació d'entrada del model hidrològic [76]. PHYSITEL requereix un model digital d'elevacions (MDE), un mapa d'usos del sòl i un mapa de tipus (textura) de sòl. D'aquesta manera es poden determinar les unitats hidrològiques relativament homogènies (UHRH). Cadascuna d'aquestes unitats correspon a una petita subconca de la conca principal a l'escala de la qual es simularan els diferents processos hidrològics.

El MDE proporciona la representació de les elevacions de la zona d'estudi, representades mitjançant cel·les quadrades de dimensions fixades. A partir del MDE es poden obtenir inicialment la pendent i l'orientació, paràmetres que permetran determinar la direcció d'escolament de cada cel·la. Les etapes que el procediment ha de seguir per a la determinació de l'estructura de drenatge d'una conca particular, es descriuen a continuació:

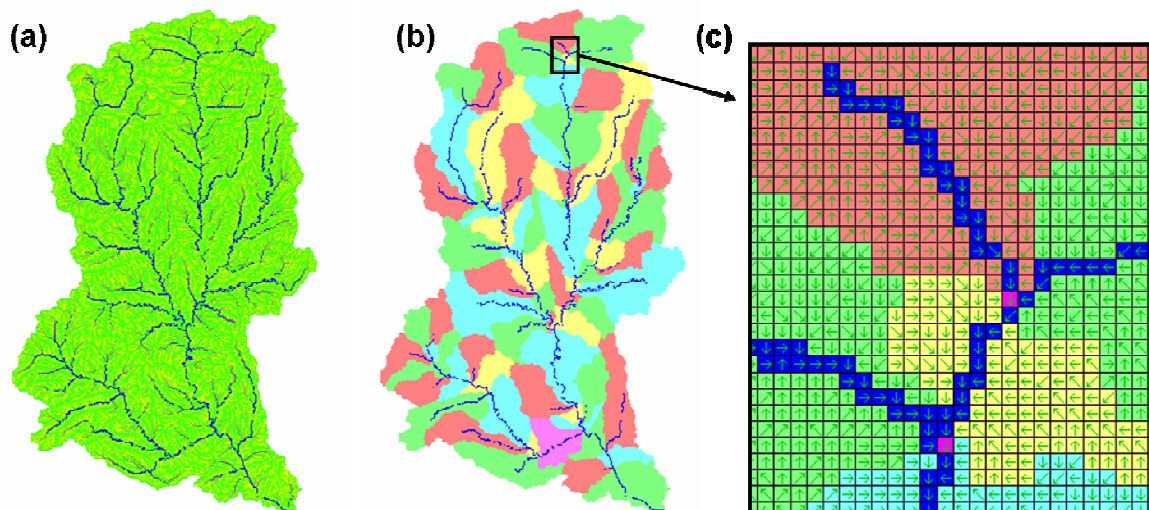
1. Importació i edició del model digital d'elevacions (MDE)
2. Determinació i edició dels pendents (s'entén com a "pendent" la diferència d'elevació entre dues cel·les veïnes en funció de la distància)
3. Importació i edició de la xarxa vectorial imposada (la xarxa hidrogràfica)
4. Determinació i edició de la xarxa matricial imposada (es tracta de convertir la representació vectorial de la xarxa hidrogràfica en una representació en forma de matriu de dades.
5. Determinació i edició del model digital d'elevacions "alterat" (es tracta d'accentuar les elevacions dins i a la perifèria de la xarxa matricial imposada per tal de reforçar el sentit de l'escolament de l'aigua). A la pràctica, el tractament del model d'elevacions pot presentar certes dificultats relacionades amb la precisió del MDE o de la naturalesa del terreny, sobretot quan els pendents són relativament febles. Per aquest motiu, generalment caldrà realitzar un filtratge previ del MDE, eliminar els valors extrems restants i construir un tipus de model d'elevacions "alterat" per tal d'assegurar la continuïtat de l'escolament.
6. Determinació i edició de les orientacions (consisteix a assignar una orientació d'escolament de l'aigua a cada cel·la del model d'elevacions "alterat")
7. Determinació de la conca hidrogràfica (veure més avall). Totes les cel·les que acaben drenant a la cel·la exutori de la conca (veure Figura 2.17) es considera que pertanyen dins de l'àrea de la conca hidrogràfica, mentre que les cel·les restants en són excloses.
8. Determinació de l'estructura d'escolament de la conca hidrogràfica (xarxa fluvial). El valor de cada cel·la es determina en funció del nombre de cel·les l'aigua de les quals acaba drenant a la cel·la en qüestió. Com més aigua va recopilant una cel·la, més elevat és el valor que li és assignat. És normal doncs, que els valors siguin baixos als extrems de la xarxa, que vagin creixent a mesura que s'apropin al centre, i que presentin el seu màxim prop de l'exutori. La xarxa creada dins d'aquesta etapa està constituïda de seccions, les quals tenen nodes com a extremitats (veure Figura 2.17). Un node neix de la intersecció entre dues seccions.



**Figura 2.17.** Exemple de representació d'una xarxa fluvial en format matricial, indicant les seccions de riu, nodes i el node exutori de la conca

9. Reorientació de les cel·les al voltant de les confluències. Es tracta d'una correcció efectuada sobre les cel·les orientades directament cap a un node determinat per tal que aquest no rebi altres aportacions d'aigua que no procedeixin directament de les seccions que el connecten.
10. Determinació de les unitats hidrològiques relativament homogènies (UHRH). Es tracta de petites subconques, cadascuna d'elles associada a una secció de la xarxa hidrogràfica modelitzada sobre la qual seran realitzats els càlculs posteriors (veure Figura 2.18b).
11. Importació i edició de les ocupacions del sòl. Mitjançant aquesta etapa es realitza un inventari de les diferents ocupacions existents sobre cada UHRH. En base a finalitats de modelització hidrològica, les classes en què es reparteix l'ocupació del sòl haurien de ser caracteritzades en funció de l'efecte produït sobre els processos hidrològics. Cadascuna de les classes, està caracteritzada per atributs quantitius (alçada, profunditat de les arrels, índex foliar, etc) que varien en funció del temps i que poden ser utilitzats en els diferents algorismes del model. El percentatge d'ocupació de cadascuna de les classes sobre cada UHRH s'obté mitjançant la proporció entre el nombre de píxels (cel·les) corresponents a una classe determinada i el nombre total de píxels de la UHRH.
12. Importació i edició dels tipus de sòl, amb la mateixa funció que l'etapa anterior. Aquesta informació serà necessària per definir les característiques hidràuliques de cada tipus de sòl. El model proposa valors per defecte extrets de Rawls i Brakensiek (1989) [77], però l'usuari els pot substituir pels seus propis valors, en el cas que es disposi de valors més apropiats.

13. Exportació del projecte PHYSITEL a format HYDROTEL



**Figura 2.18.** Operacions efectuades mitjançant l'aplicació del software PHYSITEL a la conca del riu Llobregat. (a) determinació de la conca hidrogràfica; (b) determinació de les UHRHs; (c) direccions de l'escolament dins de les UHRH

## 2.2.2. Model HYDROTEL

El model hidrològic HYDROTEL (INRS-ETE, Québec) ha estat utilitzat per a la simulació dels processos hidrològics [78-79]. Especialment, HYDROTEL és un model hidrològic distribuït, de manera que els càlculs efectuats es realitzen sobre un conjunt d'elements de càlcul repartits a l'interior d'una conca hidrogràfica. Aquesta aproximació permet tenir en compte la variabilitat espacial de les característiques fisiogràfiques i meteorològiques que afecten a la conca en qüestió. Generalment, dins d'HYDROTEL, una conca es troba dividida en centenars o milers d'UHRHs, de manera que és impossible atribuir valors individuals de les variables fisiogràfiques o meteorològiques a cadascuna d'elles. Una de les principals raons es deu a que normalment l'usuari no disposa d'informació prou detallada sobre la conca per tal d'atribuir valors satisfactoris amb aquest nivell de precisió. Per aquest motiu, és possible treballar amb agrupacions d'UHRHs sobre la base d'ocupacions o tipus de sòl semblants.

Temporalment, HYDROTEL és un model que funciona en continu (no ha estat dissenyat per a la simulació d'esdeveniments puntuals), que disposa d'un començament i una finalització (data i hora). Per a un determinat període de simulació, el model genera resultats per cada pas de temps. Es defineix com a pas de temps l'interval de temps constant en el qual es subdivideix un període complet de simulació. Aquest ha de ser un múltiple de 24 hores. Les dades d'entrada del model que varien temporalment (dades meteorològiques i hidromètriques) han de ser

## Capítol 2

conegudes a intervals de temps que no sobrepassin la durada del pas de temps de simulació d'HYDROTEL.

Un dels principals objectius per al desenvolupament del model fou la possibilitat de poder-lo aplicar sobre diverses conques hidrogràfiques amb característiques diferents amb un esforç de calibratge mínim.

HYDROTEL ofereix una concepció modular, en la qual la simulació dels diferents processos hidrològics es divideix en 6 submodels (veure Taula 2.1). Per cadascun dels processos hidrològics, HYDROTEL ofereix la possibilitat d'elegir entre un o diversos algorismes. Aquest fet, permet la selecció d'aquell algorisme que millor s'adapta a les característiques de la conca estudiada, tenint en compte la disponibilitat de la informació d'entrada.

Cal dir que el model HYDROTEL va ser inicialment desenvolupat per conques de regions força diferents a les dels rius catalans i que, per tant, caldrà adaptar-lo per tal que pugui ser utilitzat correctament sobre les nostres conques (així, per exemple, l'algorisme que descriu l'evolució de la coberta nival tindrà una importància molt inferior a la que se li atribueix quan aquest és utilitzat sobre regions més fredes).

Processos hidrològics	Algorismes disponibles
1. Interpolació de dades meteorològiques	1.1. Polígons de Thiessen 1.2. Mitjana ponderada de les 3 estacions més properes
2. Evolució de la coberta nival	2.1. Mètode mixte dels (graus-dies)- (balanç energètic)
3. Evapotranspiració potencial	3.1. Thornthwaite (1948) 3.2. Hydro-Québec 3.3. Linacre (1977) 3.4. Penman-Monteith (1965) 3.5. Priestley-Taylor (1972)
4. Balanç d'aigua vertical	4.1. Balanç vertical en 3 capes (BV3C) 4.2. CEQUEAU
5. Escolament sobre la superfície de les UHRHs	5.1 Ona cinemàtica
6. Escolament per la xarxa hidrogràfica	6.1. Ona cinemàtica 6.2. Ona difusiva

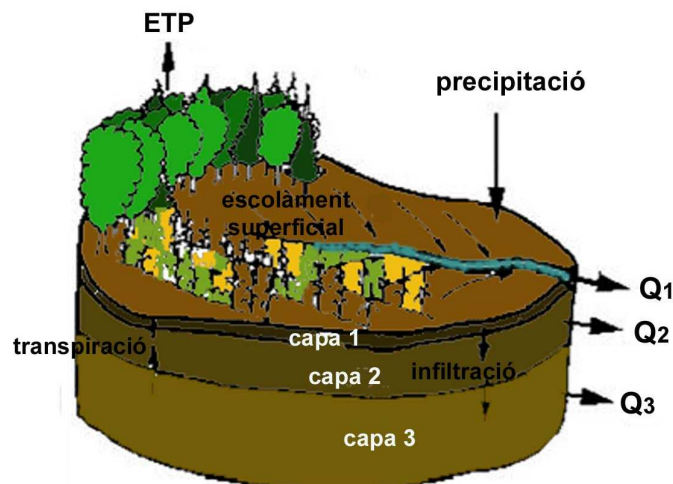
**Taula 2.1.** Processos i algorismes disponibles per les simulacions hidrològiques amb HYDROTEL

Quan es porta a terme una simulació, és possible la utilització de només alguns dels submodels de la Taula 2.1 (no és necessari utilitzar-los tots). A més, HYDROTEL permet simulacions a passos de temps inferiors a 24 hores, quan les dades necessàries es troben disponibles i quan s'escullen les condicions adequades.

Els 4 primers processos de la Taula 2.1 són simulats a nivell de cada UHRH, o dit d'una altra manera, la distribució espacial dels processos a l'interior d'una conca hidrogràfica es realitza a nivell d'una unitat hidrològica, tal com es representa esquemàticament a la Figura 2.19. I el mateix pel que fa als fluxos. L'escolament sobre la part terrestre de la conca és simulat per tal



d'obtenir un hidrograma geomorfològic específic per cada UHRH (o agrupació d'UHRHs, si és el cas) i es genera per l'escolament cel·la a cel·la d'una làmina d'aigua de referència sobre l'UHRH. El submodel que efectua el càlcul de l'escolament sobre la part terrestre de la conca rep a cada pas de temps i a nivell de l'UHRH, una làmina d'aigua que es distribueix al llarg del temps en funció de l'hidrograma geomorfològic prèviament estimat a partir de l'estructura específica d'escolament intern de l'UHRH. Finalment, l'escolament per la xarxa hidrogràfica es simula a nivell de seccions de riu. Les entrades de cada secció procedeixen de la secció que queda per sobre i de les aportacions laterals de les UHRHs de la vora.



**Figura 2.19.** Diferents processos simulats sobre una UHRH

Per tal de realitzar una simulació mitjançant el software HYDROTEL es necessiten dades d'entrada de tipus (a) hidrològic, (b) meteorològic i (c) fisiogràfic.

### (a) DADES HIDROLÒGIQUES

És necessari disposar de mesures del cabal, com a mínim a nivell diari, a diferents estacions localitzades al llarg de la xarxa hidrogràfica. És imprescindible disposar de mesures del cabal a l'exutori de la conca (o en un punt proper a aquest). Les dades diàries del cabal per cada estació hidrològica s'emmagatzemen en fitxers \*.hyd, els quals presenten el format que es mostra a la Figura 2.20.

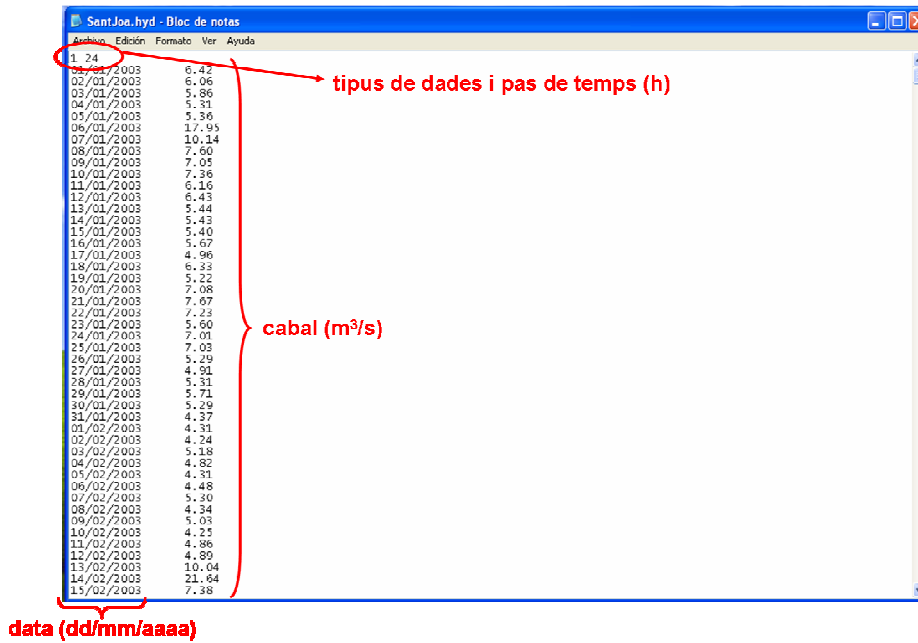


Figura 2.20. Exemple d'un fitxer \*.hyd amb les dades hidrològiques de cabal diari mesurat a l'estació de Sant Joan Despi (Barcelonès)

(b) DADES METEOROLÒGIQUES

Es necessiten dades diàries de temperatura (màxima i mínima) i precipitació. Depenent de l'algoritme que es vulgui utilitzar, caldrà disposar també de paràmetres com la humitat relativa, la radiació solar o la velocitat del vent a una alçada determinada. Les dades meteorològiques s'emmagatzemen en fitxers \*.met (veure format a la Figura 2.21). El valor -9999 correspon a NODATA (no existeix mesura).

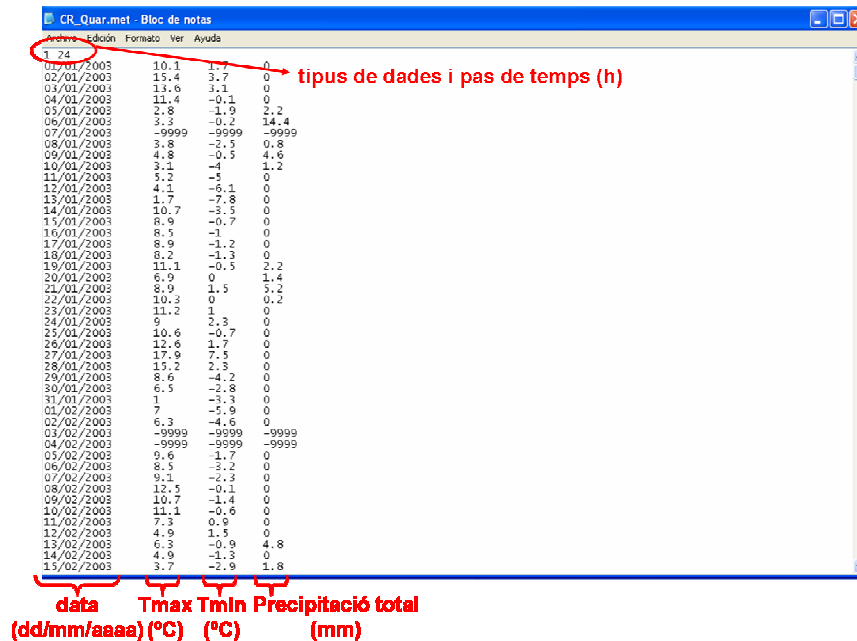


Figura 2.21. Exemple d'un fitxer \*.met amb les dades meteorològiques diàries de l'estació de La Quar (Berguedà) La precipitació es dona en mm, unitat equivalent a litres per metre quadrat

### (c) DADES FISIOLÒGUES

Caldrà que s'afegeixin al model dades sobre el valor de l'índex foliar (índex que representa els m<sup>2</sup> de fulla / m<sup>2</sup> de sòl) i la profunditat de les arrels (m) per les diferents classes d'ocupació del sòl i per diferents períodes de l'any. Aquestes dades s'acostumen a emmagatzemar en fitxers anuals (si es disposa de prou informació), però també existeix la possibilitat d'utilitzar el mateix fitxer per tot el període estudiat (en el cas que aquest correspongui a anys diferents). Les dades fisiològiques s'emmagatzemen en fitxers \*.def.

Per breuetat, a continuació només es descriuran els algorismes seleccionats per al cas d'estudi investigat en aquesta Tesi. La resta d'algorismes es troben àmpliament descrits i discutits a la literatura [78-80].

### 1. Interpolació de dades meteorològiques

La interpolació de dades meteorològiques (precipitació, temperatura de l'aire, humitat relativa, vent, etc.) és el procés mitjançant el qual les dades mesurades a les estacions meteorològiques són estimades sobre cadascuna de les unitats de càlcul.

En el cas dels polígons de Thiessen, s'atribueixen a cada punt de la conca les dades meteorològiques mesurades a l'estació més propera al punt. En l'aplicació presentada en aquesta Tesi, l'algorisme utilitzat ha estat la mitjana ponderada de les tres estacions més properes. Aquest algorisme, atribueix a cada UHRH d'una conca hidrogràfica les dades mesurades a les tres estacions que es troben més properes a aquesta UHRH, amb una ponderació per cadascuna de les estacions en funció de la inversa de la distància entre una estació i la UHRH. La suma dels pesos atribuïts a cada UHRH serà igual a 1.

La interpolació dels valors mesurats a les estacions es realitza amb la finalitat de tenir en compte la variació espacial de les dades meteorològiques. No obstant, aquesta variació no és la única que intervé entre una estació i una UHRH, ja que les variables com ara la precipitació i la temperatura varien també en funció de l'altitud. Aquests paràmetres poden ser corregits a l'hora d'aplicar el model mitjançant la definició de gradients verticals (veure Equacions 2.33 i 2.34 per a la temperatura i la precipitació, respectivament).

$$T_c = T_m + G_T \left( \frac{A_s - A_u}{100} \right) \quad \text{Equació 2.33}$$

On  $T_c$  és la temperatura corregida (°C),  $T_m$  la temperatura mesurada (°C),  $G_T$  el gradient vertical de temperatura (°C/100m),  $A_s$  l'altitud de l'estació (m) i  $A_u$  l'altitud mitjana de l'UHRH (m).

$$P_c = P_m \left[ 1 + G_p \left( \frac{A_s - A_u}{100} \right) \right] \quad \text{Equació 2.34}$$

On  $P_c$  és la precipitació corregida (mm),  $P_m$  la precipitació mesurada (mm),  $G_p$  el gradient vertical del coeficient multiplicatiu de precipitació (1/100m),  $A_s$  l'altitud de l'estació (m) i  $A_u$  l'altitud mitjana de l'UHRH (m).

Durant l'etapa d'interpolació de les dades meteorològiques, el model també efectua una repartició de la precipitació total en precipitacions líquides i sòlides sobre cada UHRH. Aquesta repartició es fa de la mateixa manera sigui quin sigui l'algoritme d'interpolació aplicat. És important remarcar que les dades poden ser entrades dins del model en forma de precipitació total o bé de pluja i neu per separat (en aquest darrer cas, el model determina la precipitació total). La repartició de la precipitació (líquida i sòlida) sobre cada unitat dependrà bàsicament de les temperatures màxima i mínima de l'aire que han estat interpolades sobre l'UHRH pel pas de temps considerat.

## 2. Evolució de la coberta nival

El model permet simular la variació diària del gruix de la coberta de neu, del seu equivalent en aigua i de la seva densitat a partir de les precipitacions sòlides, i la fusió sobre cadascuna de les unitats hidrològiques de la conca, tenint en compte l'efecte de compressió en l'estimació de l'espessor de la coberta. El dèficit calorífic i el contingut en aigua líquida de la coberta també són estimats mitjançant equacions, de manera que es simula l'evolució de la coberta de neu de manera semblant a com es faria mitjançant una aproximació de tipus energètic. Dins HYDROTEL, en aquests moments existeix un sol algoritme que permet simular l'evolució de la coberta nival, el qual està basat en un mètode mixte (graus-dia)-(balanç energètic). S'anomena grau-dia a la quantitat de calor que correspon a cada augment de la temperatura d'un grau per sobre del llindar considerat, quan el pas de temps és d'un dia.

Per al càlcul de l'algoritme, es consideren tres classes d'ocupació del sòl amb un efecte significatiu sobre la fusió de la neu: els boscos de coníferes, els boscos d'arbres de fulla caduca i les zones obertes. Diferents valors del factor de fusió ( $C_f$ ) són atribuïts a cadascuna de les classes. Aquest factor serà més elevat a mesura que el medi es vagi trobant més desforestat. Per tant, el llindar de temperatura per sota del qual hi haurà fusió disminueix a mesura que el sòl està menys cobert de vegetació (a menys vegetació, fusió més ràpida de la neu). Així, s'efectua una simulació diferent per cadascun dels medis. Les làmines d'aigua de

fusió produïdes per cada medi es ponderen en funció del percentatge de la superfície total de cada unitat hidrològica ocupada pel medi en qüestió per tal de donar com a resultat la làmina d'aigua de fusió mitjana disponible per infiltrar-se dins del sòl.

A més, existeix també la possibilitat de simular la variació de l'albedo de la coberta de neu, o sigui, la relació entre la quantitat de llum reflectida per la coberta i la quantitat de llum rebuda per aquesta (un cos totalment blanc tindrà un albedo igual a 1, ja que reflexa tota la llum rebuda). Això es pot fer utilitzant dos algorismes diferents.

L'algoritme d'evolució i fusió de la coberta nival descriu la simulació de diferents fenòmens, els quals es troben descrits en detall a Fortin i Royer (2004) [80]. L'evolució de la coberta nival és un procés amb una importància considerable en el model HYDROTEL quan aquest és aplicat a zones caracteritzades per la presència de neu bona part de l'any. De fet, el model va ser desenvolupat per la seva aplicació en una zona de clima amb característiques continentals i polars, i és per això que dóna tanta importància a la simulació d'aquest procés. En el cas d'aquesta Tesi, l'evolució i fusió de la coberta de neu té una importància força més limitada. Per aquest motiu, els paràmetres necessaris per al càlcul de l'algoritme han estat definits inicialment i no han estat modificats durant el procés de calibratge del model, ja que s'ha considerat que el seu impacte té una importància més limitada en el cas d'estudi.

### 3. Evapotranspiració potencial

L'avaluació del balanç hidrològic vertical a cada pas de temps requereix l'estimació de l'evapotranspiració real (ETR) de cada tipus de coberta vegetal a la superfície del sòl i sobre sòl nu. Aquestes estimacions depenen del perfil d'humitat del sòl i d'altres paràmetres com la resistència estomàtica, que no acostumen a ser fàcils d'obtenir. L'aproximació habitual en hidrologia consisteix en estimar l'evapotranspiració potencial (ETP).

Existeixen cinc opcions disponibles per l'estimació de l'ETP dins d'HYDROTEL: Thornthwaite (1948), Hydro-Québec, Linacre (1977), Penman-Monteith (1965) i Priestley-Taylor (1972). L'equació que requereix més informació sobre paràmetres meteorològics és la de Penman, mentre que la Hydro-Québec, juntament amb la de Thornthwaite, és la que en requereix menys (temperatures mínima i màxima diàries de l'aire). No obstant, això no significa que aquestes darreres equacions siguin menys precises que la de Penman, sinó que la precisió dependrà de les condicions meteorològiques. En el cas d'estudi inclòs en aquesta Tesi s'ha aplicat l'equació Hydro-Québec. El desavantatge que té és que no distingeix entre classes diferents d'ocupació del sòl, de manera que l'ETP obtinguda és idèntica per a cada classe. Tot i això, ha estat utilitzada en diverses ocasions i s'ha comprovat que els resultats obtinguts no difereixen massa

dels d'altres equacions considerant les diferents ocupacions del sòl. Un factor de multiplicació s'utilitza com a paràmetre de calibratge per tal d'ajustar l'ETP.

Les equacions de Thornthwaite, Hydro-Québec i Linacre no poden ser estimades directament a un pas de temps inferior a 24 hores. Per passos de temps més curts caldrà fer una transformació.

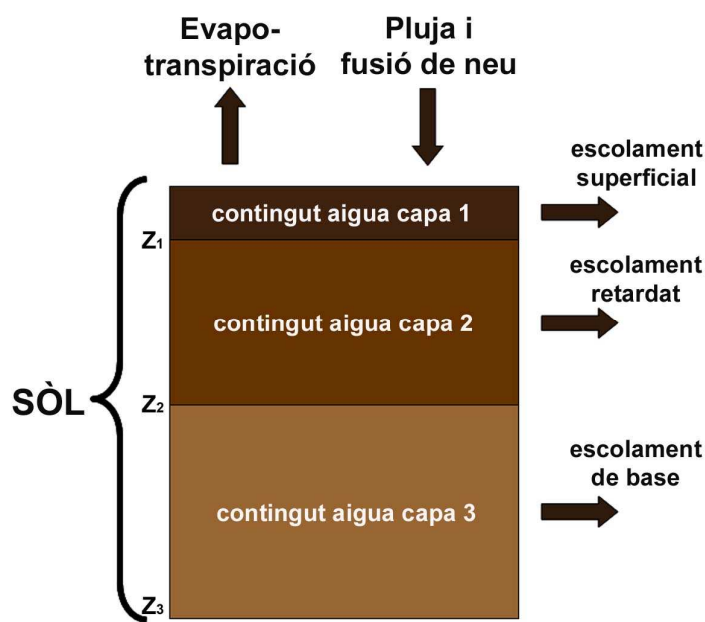
### 4. Balanç d'aigua vertical

El balanç d'aigua vertical es realitza per tal de separar, sobre cada unitat hidrològica, les làmines d'aigua que s'escolen per la superfície del terreny d'aquelles que s'hi infiltren. Un cop s'han infiltrat, aquestes làmines d'aigua modifiquen l'estat de saturació del perfil del sòl i participen en l'escolament retardat de la part superior no saturada del sòl i en els escolaments entre capes existents a la zona saturada.

Per la simulació d'aquest procés, s'ofereixen dos algoritmes diferents: un concebut específicament per als objectius d'HYDROTEL pel que fa a la reproducció dels processos físics i a la compatibilitat amb la teledetecció i els SIG [78], i l'altre extret del model conceptual CEQUEAU [70], el qual pot ser utilitzat si no es disposa de prou informació per al primer algoritme. En aquesta Tesi s'ha utilitzat el primer, l'algoritme del balanç vertical basat en processos físics en tres capes (*Bilan vertical en 3 couches*, BV3C). L'algoritme no dóna una representació exacta del moviment de l'aigua a través del sòl però presenta, en forma de promig, el que passa sobre una conca hidrogràfica.

Per la realització de les simulacions, la columna de sòl es subdivideix en 3 capes o horitzons (veure Figura 2.22). La capa superficial és relativament prima (de 10 a 20 cm) per tal de correspondre al gruix afectat per l'evaporació d'un sòl nu. Aquesta capa controla l'escolament superficial. La segona capa serveix per descriure la zona de transició entre la primera i la tercera capa. També té la funció de regular els escolaments retardats dins de la capa superior del sòl. El cabal base (cabal mínim necessari per al manteniment de les característiques essencials i l'estructura de l'ecosistema aquàtic) està controlat per la tercera capa de sòl. Aquesta capa se sol trobar relativament propera a la saturació.

El balanç vertical d'aigua es calcula sobre cada una de les UHRHs. Per tant, els límits inferior i superior de cada capa seran diferents per cada UHRH o grup d'UHRHs, i les característiques hidràuliques canviaran en funció del tipus de sòl. Finalment, a més dels límits de les capes de sòl i les característiques hidràuliques, el balanç hídic també dependrà del pendent d'escolament que existeixi sobre cada unitat hidrològica.



**Figura 2.22.** Subdivisió del balanç vertical en tres capes dins del submodel BV3C d'HYDROTEL.  $Z_1$ ,  $Z_2$ , i  $Z_3$  corresponen als límits inferiors en metres de les capes de sòl 1, 2 i 3 respectivament (Fortin i Royer, 2004)

A l'inici de cada pas de temps del model, la infiltració procedent de la pluja o bé de la fusió de la neu és controlada per l'aportació en condicions no saturades, o per la conductivitat hidràulica de la capa en condicions de saturació. L'excés d'aigua a la superfície esdevé escolament superficial. Finalment, el contingut en aigua  $\Theta_j$  de cada capa  $j$  varia en funció dels fluxos verticals  $q_{1,2}$  i  $q_{2,3}$  entre les capes, de l'escolament fora de la columna de sòl  $Q_2$  i  $Q_3$ , de la transpiració procedent de cada capa i de l'evaporació del sòl nu per la capa superficial (veure Figura 2.22).

Per tal d'estimar tots els fluxos entre les capes i els escolaments cap a l'exterior de la columna de sòl, s'estima la conductivitat hidràulica  $K$  de cada capa en funció del seu valor a la saturació, dels valors del contingut en aigua i de la distribució de les dimensions dels porus.

L'escolament retardat a la part superior del sòl pot estar lligat a diversos fenòmens. Un model simple basat en l'ona cinemàtica pot ser adequat, tenint en compte l'heterogeneïtat presentada per les conques hidrogràfiques naturals [81]. Segons l'aproximació de l'ona cinemàtica, l'escolament lateral  $Q_2$  de la segona capa només esdevé important si el contingut en aigua d'aquesta capa es troba entre la capacitat de camp i la saturació. Es defineix la capacitat de camp com la retenció màxima d'aigua al sòl, corresponent a la quantitat d'aigua retinguda per un sòl 48 hores després de la seva inundació i del posterior goteig lliure de l'aigua cap a la capa freàtica. L'escolament lateral  $Q_2$  és estimat en funció del pendent  $S_n$  de l'UHRH, del guix ( $Z_2 - Z_1$ ) de la capa i de la seva conductivitat hidràulica  $K(\Theta_2)$ , per l'equació:

$$Q_2 = K(\theta_2) \sin |\arctan(S_n)| (Z_2 - Z_1) \quad \text{Equació 2.35}$$

Finalment, l'escolament  $Q_3$  de la capa 3 és modelitzat per l'Equació 2.36 en funció del contingut en aigua  $\Theta_3$  d'aquesta capa, el seu gruix  $(Z_3 - Z_2)$  i d'un coeficient de recessió CR, el qual pot ser determinat a partir dels cabals observats:

$$Q_3 = CR (Z_3 - Z_2) \theta_3 \quad \text{Equació 2.36}$$

La distribució de l'aigua al sòl i l'escolament superficial dependran en gran mesura de la profunditat de les capes del sòl. Aquestes profunditats, juntament amb el coeficient de recessió, esdevindran paràmetres importants durant el procés de calibratge.

L'evapotranspiració real (ETR) de cada classe de sòl es deriva de l'evapotranspiració potencial (ETP) de la classe. Es calcula un valor de transpiració  $T_r$ , que es reparteix entre les capes en funció de la relació entre el contingut d'aigua de cada capa i el contingut mitjà d'aigua del perfil de sòl, per tal d'obtenir les evapotranspiracions reals corresponents a cada capa. Per la capa superficial, l'evaporació del sòl nu és afegida a la transpiració. Finalment, l'ETR sobre cada UHRH es determina calculant la mitjana ponderada de l'ETR de cada classe d'ocupació del territori en funció del percentatge de superfície atribuïda a cadascuna d'aquestes classes sobre l'UHRH. L'evaporació es considera nul·la en el cas de superfícies impermeables.

### 5. Escolament sobre la superfície de les UHRHs

L'escolament sobre la superfície d'una conca és el procés que descriu el desplaçament de l'aigua fins que aquesta arriba a la xarxa hidrogràfica. Aquest escolament integra (a) els escolaments que es realitzen directament sobre el sòl a través de la vegetació i d'altres obstacles, (b) els escolaments a l'interior de canals, ja siguin naturals o artificials, de dimensions suficients per a ser considerats una part del curs de l'aigua, i (c) els escolaments efectuats a l'interior del sòl. Així, si el submodel BV3C s'utilitzava per al càlcul del balanç vertical, el submodel d'escolament sobre la part terrestre efectuarà el transport fins al curs d'aigua més proper de la làmina d'aigua produïda tant per l'escolament superficial com pels escolaments provinents de les capes 2 i 3.

Per a la simulació d'aquest procés s'utilitza l'algorisme de l'ona cinemàtica. Aquesta formulació fa intervenir, entre d'altres, un paràmetre anomenat coeficient de rugositat de Manning, el qual serà diferent per les tres classes d'ocupació del sòl que l'algorisme és capaç de discriminar



(medi forestal, aigua i altres medis). Aquest coeficient de rugositat estableix la dificultat que troba l'aigua en desplaçar-se per cadascun dels medis.

A cada pas de temps es determina l'escolament cel·la a cel·la sobre cada unitat hidrològica i, per consegüent, sobre tota la conca hidrogràfica considerada. El càlcul de l'escolament es divideix en dues etapes: (1) el càlcul d'un hidrograma geomorfològic de referència per cada UHRH i (2) el càlcul de la quantitat d'aigua que abandona l'UHRH a cada pas de temps.

Les característiques de l'estructura d'escolament a l'interior de cada UHRH es tenen en compte amb l'ajuda d'un hidrograma geomorfològic específic per cada unitat hidrològica [82]. La forma d'aquest hidrograma es determina per cada unitat a l'inici de cada simulació, amb la propagació d'un impuls d'aigua unitari generat uniformement a l'interior de cada UHRH segons el model de l'ona cinemàtica utilitzat en aquest submodel. Posteriorment, les làmines d'aigua produïdes a cada pas de temps sobre cada unitat es distribueixen en el temps segons l'hidrograma geomorfològic definit abans, simulant així l'efecte tampó de l'UHRH sobre els cabals a la sortida. Cada vegada que es modifica la làmina de referència o els coeficients de Manning es realitza un nou càlcul de l'hidrograma. El càlcul del cabal a l'exutori d'una UHRH a cada pas de temps s'efectua fent la combinació lineal dels hidrogrames de referència. La combinació lineal dels diferents hidrogrames geomorfològics dóna lloc a la corba de l'escolament acumulat.

### 6. Escolament per la xarxa hidrogràfica

Es tracta del procés mitjançant el qual l'aigua es desplaça a través de la xarxa hidrogràfica fins a l'exutori de la conca. Per a cadascuna de les seccions de riu, es simula l'escolament natural mitjançant aproximacions del sistema d'equacions de Saint-Venant. Existeixen dos algoritmes disponibles dins el model, dels quals en aquesta Tesi s'ha utilitzat el de l'ona cinemàtica.

Normalment, abans d'efectuar una simulació per al període d'estudi amb HYDROTEL, s'acostuma a realitzar una simulació inicial considerant un període de temps més curt, idealment un període immediatament anterior al considerat en l'estudi. Això permet establir els valors per a les anomenades variables d'estat, les quals descriuen l'estat de la conca en termes de quantitat d'aigua en un moment determinat. L'emmagatzemament d'aquestes variables permet l'inici de la nova simulació a partir d'uns determinats estocs d'aigua, situació més realista que el seu inici a partir de valors per al contingut d'aigua iguals a zero.

Durant l'etapa de calibratge del model, els cabals simulats es comparen amb els cabals mesurats a les diferents estacions hidrològiques, de manera que els paràmetres del model poden anar essent ajustats per tal de reproduir aquests cabals mesurats. Tot i que el model és capaç de considerar l'efecte dels embassaments, aquests no han estat considerats en aquesta

Tesi degut a la falta d'informació referent a la seva operació (la simulació dels embassaments és molt exigent en termes de quantitat d'informació).

### 2.2.3. Sistema GIBSI

El sistema integrat de modelització per a la gestió de conques hidrogràfiques GIBSI (*Gestion Intégrée des Bassins versants à l'aide d'un Système Informatisé* (INRS-ETE, Québec)) [83-85], ha estat concebut com una eina per a l'avaluació de l'impacte que els diferents usos de l'aigua i del territori poden tenir sobre la qualitat de l'aigua d'una conca hidrogràfica. Com es mostra a la Figura 2.23, GIBSI es troba constituït d'un sistema gestor de bases de dades, diferents models físics de simulació (hidrologia, erosió del sòl, transport de fertilitzants i productes fitosanitaris, i qualitat de l'aigua), diferents models de gestió (font de contaminació puntual, font de contaminació difosa i gestió d'embassaments) i un sistema d'informació geogràfica. El SIG utilitzat és GRASSLAND [86], un SIG obert construït en base a GRASS [87]. La base de dades inclou informació espacial i alfanumèrica (per exemple, el model digital d'elevacions, dades meteorològiques, característiques del tipus de sòl, indicació de la localització de les diferents estacions hidrològiques, meteorològiques i de qualitat, resultats de les simulacions, dades qualitatives i quantitatives associades a les diferents unitats administratives com són la gestió de cultius o les unitats ramaderes, etc.). Es pot trobar una descripció exhaustiva de la base de dades del sistema a Villeneuve et al. (1998) [85].

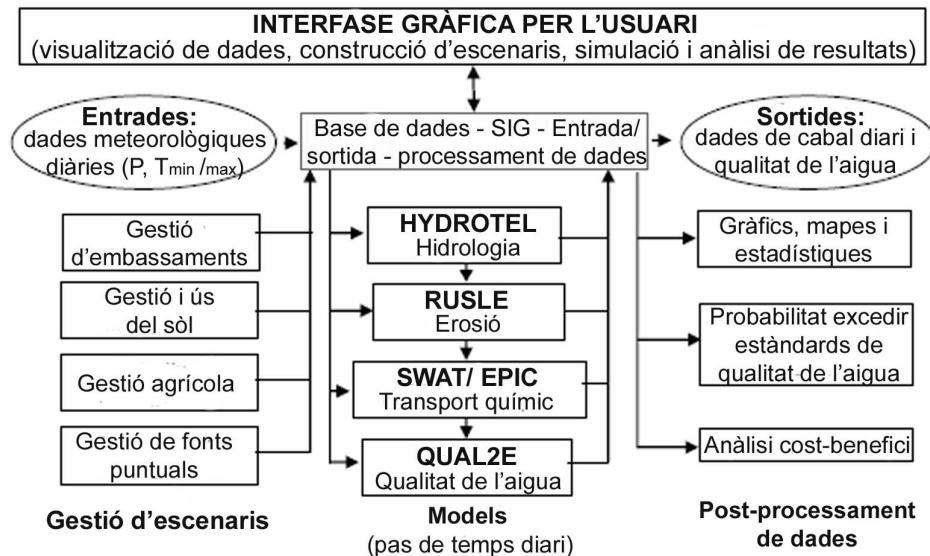


Figura 2.23. Representació dels components i de l'estructura del sistema GIBSI (Quilbé i Rousseau, 2007)

Per tal d'avaluar els impactes causats per projectes municipals, industrials, forestals i agrícoles sobre la qualitat de l'aigua, GIBSI proposa una aproximació basada en la comparació

d'escenaris. Partint d'un escenari de referència (l'estat actual de la conca), noves pràctiques o mesures de control de la contaminació poden ser introduïdes o modificades per tal de construir un nou escenari. L'impacte d'aquest nou escenari sobre la hidrologia i la qualitat de l'aigua pot ser avaluada mitjançant una comparació entre aquest i l'escenari de referència.

### Escenaris

Dins de GIBSI es poden definir quatre tipus d'escenaris diferents:

- embassaments: permet l'addició de nous embassaments a qualsevol de les seccions de riu o l'edició de les seves característiques
- ús del sòl: permet el canvi d'una determinada categoria d'ús del sòl per una altra a qualsevol escala espacial (per exemple, una o diverses UHRH, subconques o bé unitats administratives)
- agrícola: permet l'edició dels tipus de cultiu, les pràctiques de gestió de nutrients, tractaments amb productes fitosanitaris, dates en què tenen lloc les diferents pràctiques agrícoles etc., a qualsevol escala espacial
- plantes de tractament d'aigües residuals: permet l'addició de noves plantes a qualsevol de les seccions de riu o l'edició de les seves característiques (per exemple, el tipus de tractament aplicat, els efluent a la sortida, etc.)

Un cop definits els escenaris, aquests poden ser integrats dins de la base de dades del sistema i posteriorment, realitzar les simulacions.

### Models de simulació

El sistema GIBSI consta de 4 models diferents, els quals són aplicats l'un darrere de l'altre (veure Figura 2.23). No és necessària l'aplicació de tots els models per tal de portar a terme una simulació. No obstant, cal que els models s'apliquin en forma de cascada, de manera que no és possible, per exemple, la simulació del transport de productes fitosanitaris sense haver simulat prèviament l'erosió a la conca hidrogràfica. Els quatre models que s'integren dins del sistema es descriuen a continuació:

#### 1. Model hidrològic

HYDROTEL [80] és el model hidrològic utilitzat per GIBSI. Aquest model ha estat descrit de manera individual en l'apartat 2.2.2 d'aquesta Tesi ja que, en aquest cas, el calibratge hidrològic de la conca ha estat realitzada fora de l'entorn GIBSI, i posteriorment els resultats han estat importats dins del sistema. La unitat de simulació espacial (USS a GIBSI, equivalent a la UHRH d'HYDROTEL) és la unitat elemental de drenatge de la conca (àrea que comprèn

## Capítol 2

tres capes de sòl de diferents profunditats, drenada per un tram o segment de riu). Per a modelitzar el moviment de l'aigua, s'han utilitzat segments unidimensionals de característiques hidràuliques i bioquímiques uniformes.

### 2. Model d'erosió del sòl

Utilitzant les dades calculades pel model hidrològic per una unitat de simulació específica, el model d'erosió, derivat dels algoritmes de l'equació revisada universal de la pèrdua de sòl (*Revised Universal Soil Loss Equation*, RUSLE) [88-89], calcula la proporció de sediment transportada al sistema fluvial corresponent a cadascun dels usos del sòl sobre la USS (veure Figura 2.24). El càlcul de la proporció de sediments transportada s'obté mitjançant la utilització d'un índex d'erosió (relacionat amb la precipitació i l'escolament superficial) i un índex de vegetació (relacionat amb l'ús del sòl i l'estat de creixement i desenvolupament de la coberta vegetal en el moment de la precipitació). El model d'erosió té en compte tres processos físics: i) despreniment de les partícules de sòl, ii) transport de les partícules erosionades i iii) sedimentació de les partícules erosionades. Com a resultat del model s'obtenen la càrrega i la concentració de sediments a cadascun dels segments de riu.

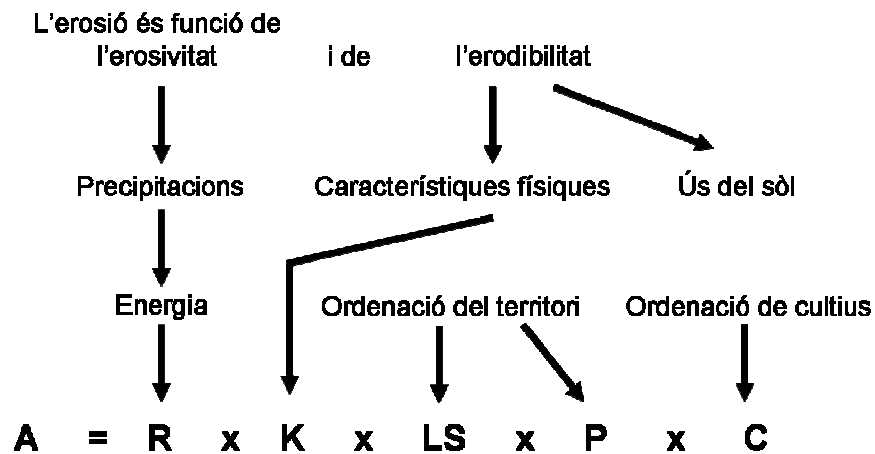


Figura 2.24. Factors que afecten l'erosió del sòl (Villeneuve et al., 1998)

### 3. Model de transport de substàncies agrícoles-químiques

El model es basa en els algoritmes de transport de nitrogen, fòsfor i productes fitosanitaris dels models SWAT (*Soil and Water Assessment Tool*) [90] i EPIC [91], i s'utilitza per calcular les càrregues de contaminants agrícoles exportades cap al flux del riu. Utilitza com a informació d'entrada les dades calculades mitjançant els models hidrològic i d'erosió. El mòdul PO, que inclou els algoritmes de transport per al nitrogen i el fòsfor, no ha estat aplicat en aquesta Tesi, ja que en aquest cas s'ha estudiat exclusivament el transport de productes fitosanitaris, inclòs en el mòdul PS. Una descripció exhaustiva dels algoritmes de simulació del mòdul PO es pot

trobar a Villeneuve et al. (1998) [85]. A continuació es descriuen breument els processos que participen en el model de transport de productes fitosanitaris. Aquests també es troben àmpliament descrits a la referència recentment citada [85].

### **Transport de productes fitosanitaris**

El mòdul PS (derivat del model SWAT) simula l'aplicació, degradació i transport de productes fitosanitaris per l'aigua i les partícules de sòl erosionades sobre cadascuna de les ocupacions agrícoles de sòl (com són, per exemple, cereals, pasturatge, fruiters, etc.) de cada unitat de simulació d'una conca. El mòdul verifica si el dia de simulació correspon a un dia d'aplicació del producte. Si és així, aquest aplica les quantitats de producte fitosanitari establertes per l'usuari en la definició de l'escenari de gestió agrícola sobre cadascuna de les ocupacions del sòl.

El mòdul estima que una fracció del producte fitosanitari passa a l'atmosfera per volatilització en el moment de la seva aplicació, que una altra fracció va a parar sobre les fulles de la vegetació (captura foliar) i que la fracció restant arriba a la superfície del sòl. Aquesta última fracció es reparteix en dues fraccions més; una en forma dissolta dins de la solució del sòl i l'altra adsorbida directament sobre les partícules de sòl. Aquesta repartició s'efectua en funció del valor del coeficient de distribució sòl/aigua o  $K_d$  (adsorció), que és un valor propi a cada producte fitosanitari. Els productes fitosanitaris que es troben a la vegetació i al sòl es degraden o es dissipen mitjançant diferents processos: fotòlisi, oxidació, hidròlisi, volatilització i transformació biològica. Tots aquests processos són simulats globalment dins del mòdul PS en funció del valor del temps de vida mitjana ( $t_{1/2}$ ) de cada producte (o sigui, el temps que necessita una determinada quantitat de producte per arribar a la meitat de la seva quantitat inicial).

Els productes fitosanitaris dissolts dins de la solució del sòl poden ser transportats a través de la xarxa hidrogràfica per escolament superficial o bé per infiltració dins del sòl, mentre que els que es troben adsorbits a les partícules de sòl poden arribar-hi a través de l'erosió del sòl (veure Figura 2.25). És per aquest motiu que els models hidrològic i d'erosió són necessaris per a la simulació del transport de productes fitosanitaris sobre les unitats de simulació. Cal remarcar, però, que el mòdul PS no simula el destí d'aquells productes volatilitzats ni dels absorbits per les plantes. Les dades d'entrada necessàries per a la utilització del mòdul PS són la identificació dels productes fitosanitaris aplicats sobre l'àrea d'estudi, les dates d'aplicació d'aquests i la quantitat en Kg/ha.



aquestes són obtingudes a partir de diferents fonts i, per tant, no tenen la mateixa resolució (mida de píxel) ni es troben alineades amb les mateixes coordenades (encaixant perfectament les unes amb les altres). Les capes necessàries per a la utilització del sistema, tant en format ràster com vectorial, són: un model digital d'elevacions, els límits de la conca, els diferents municipis continguts a la conca, els usos del sòl, la pedologia, els pendents, les subconques (delimitades utilitzant HYDROTEL), les USSs (delimitades utilitzant PHYSITEL i HYDROTEL) i les seccions de riu. Normalment, es necessiten els 2 formats (vectorial i ràster) d'una mateixa capa d'informació ja que GRASSLAND necessita el format ràster per a efectuar els diferents càlculs (capa no mostrada dins de GIBSI), mentre que el format vectorial és el que s'utilitza per a la representació gràfica del mapa.

Un cop integrada la informació espacial dins del sistema GIBSI, serà necessari editar les taules amb la informació alfanumèrica relativa a la conca hidrogràfica d'estudi. Les taules que caldrà editar dependran de la finalitat de l'estudi i dels submodels que es vulguin aplicar. A l'Annex III d'aquest treball, es mostra una llista de només aquelles taules que contenen els paràmetres que s'han modificat i adaptat al cas del riu Llobregat. Per breuetat, no s'ha realitzat una descripció exhaustiva de tota la base dades.

Un cop tota la informació ha estat introduïda dins del sistema, l'aplicació del model distribuït GIBSI comprèn les etapes que es llisten a continuació:

- 1) construcció de la base de dades
- 2) aplicació del model hidrològic
- 3) calibratge i validació del model hidrològic a partir de dades de cabals mesurats
- 4) aplicació dels models d'erosió i transport de productes fitosanitaris
- 5) adaptació dels models utilitzats en funció de les condicions de la conca
- 6) calibratge i validació dels models d'erosió i transport de productes fitosanitaris a partir de dades de concentracions i cabals mesurats

Finalment, es realitza una interpretació dels resultats generats pel sistema mitjançant mapes temàtics, gràfics i taules.