

Grau en Estadística

Títol: Creació d'interfície interactiva amb shiny per la modelització de sèries temporals: estudi mediambiental de la qualitat de l'aire

Autor: Nora Amama Ben Hassun

Director: Josep Anton Sànchez Espigares

Departament: Departament d'estadística i investigació operativa

Convocatòria: Juny 2022



Resum

En aquest treball es busca trobar una solució a un problema mediambiental mitjançant la predicció de les concentracions dels contaminants a l'aire.

S'ha creat una interfície interactiva mitjançant *shiny* amb format d'un *dashboard* per modelitzar de manera àgil qualsevol tipus de sèries temporals aplicant els mètodes d'allisat exponencial, la metodologia Box-Jenkins i l'algoritme *prophet*.

Finalment, s'ha dut a terme l'anàlisi i la creació dels diferents models de les $PM_{2.5}$, PM_{10} , NO_2 , CO i O_3 registrats a Califòrnia durant el 2021. D'aquesta forma, i amb ajuda de la plataforma web, s'ha identificat que per les partícules pesants i el diòxid de nitrogen el millor model prové de la metodologia Box-Jenkins, un model SARIMA i pels altres contaminants s'ha adaptat millor per a la predicció un model creat a partir dels mètodes d'allisat exponencial.

Paraules clau: Sèries temporals, modelització, shiny, contaminació.

This paper tries to find a new way to solve an environmental problem by predicting the concentrations of air pollutants.

An interactive interface has been created using shiny with a dashboard format to easily model any type of time series by applying exponential smoothing methods, the Box-Jenkins methodology and the prophet algorithm.

Finally, the analysis and the development of the different models of the $PM_{2.5}$, PM_{10} , NO_2 , CO and O_3 recorded in California during 2021 has been carried out. In this way, and with the help of the web platform, it has been identified that for particulate matter and nitrogen dioxide the best model fitted is by Box-Jenkins methodology, a SARIMA model, and for the other pollutants a model created from exponential smoothing methods has been best suited for prediction.

Key words: Time series, modelling, shiny, pollution.

Classificació AMS:

- 37M10 Time series analysis
- 62M10 Time series, auto-correlation, regression, etc.
- 62M20 Prediction; filtering
- 91B70 Stochastic models
- 91B76 Environmental economics (natural resource models, harvesting, pollution, etc.)

Agraïments

Primerament, voldria donar les gràcies al meu tutor per accedir a portar-me el TFG. He tingut la llibertat d'experimentar amb el tema i provar diferents metodologies per encarar-ho fins a obtenir aquest resultat.

En segon lloc, m'agradaria agrair tant a familiars com als amics que han estat al meu costat durant aquesta etapa. Gràcies a aquestes persones he pogut viure al màxim l'experiència del grau.

Finalment, estic contenta i agraïda a les Universitats i sobretot, al professorat que m'han guiat fins a descobrir què és el que m'agrada i voldré dedicar-me en el futur.

Índex

Classificació AMS:	ii
Agraïments	iii
Índex de figures	x
Índex de taules	xi
Listings	xii
1 Introducció	1
2 Descripció de les dades	6
2.1 Origen, definició i simptomatologia dels contaminants	6
2.2 Depuració de les dades	8
2.3 Anàlisi descriptiva de les dades	8
2.3.1 Partícules en suspensió (<i>PM</i>)	9
2.3.2 Òxids de nitrogen (<i>NO₂</i>)	10
2.3.3 Ozó troposfèric (<i>O₃</i>)	11
2.3.4 Monòxid de carbó (<i>CO</i>)	12
2.4 Rangs de la qualitat de l'aire	12
3 Metodologia	14
3.1 Components d'una sèrie temporal	14
3.2 Mètodes d'allisat exponencial	15
3.2.1 Esquema additiu	16
3.2.2 Esquema multiplicatiu	18
3.2.3 Identificació i predicció automàtica	19
3.3 Processos estocàstics	19
3.4 Metodologia Box-Jenkins	22
3.4.1 Identificació	23
3.4.2 Estimació	25

3.4.3	Validació	26
3.4.4	Model adequat	30
3.4.5	Identificació i predicció automàtica	30
3.5	Prophet	31
3.6	Mètriques de regressió	31
3.7	Recursos informàtics	33
4	Guia d'ús de la interfície	35
4.1	Guia d'ús	35
4.2	Predicció amb filtres	37
4.3	Preparació prèvia al model	40
4.4	Models	42
4.5	Validació	42
4.6	Predicció	45
4.7	Prophet	48
4.8	Millor model	50
5	Anàlisis del contaminant NO_2	52
5.1	Predicció amb filtres	52
5.2	Preparació prèvia al model	56
5.2.1	Observar si la variància és constant	57
5.2.2	Observar si existeix patró estacional	58
5.2.3	Observar si la mitjana és 0	59
5.2.4	Observar si la sèrie transformada és estacionària	59
5.3	Identificació i estimació dels models	60
5.4	Validació del model	63
5.4.1	Normalitat	63
5.4.2	Variància	64
5.4.3	Autocorrelació	64
5.4.4	Estacionarietat i invertibilitat	66
5.4.5	Volatilitat	66
5.5	Predicció	67
5.5.1	Identificació automàtica del model	68
5.6	Prophet	69
5.7	Millor model	71
6	Resum dels anàlisis dels contaminants	72
6.1	Mètodes d'allisat exponencial	72
6.2	ETS	73
6.3	Box-Jenkins	73

6.4	Modelització automàtica	74
6.5	Prophet	75
6.6	Millor model	76
7	Conclusions	79
	Bibliografia	83
A	Anàlisi dels contaminants estudiats	86
B	Codi Shiny	101
B.1	Server	101
B.2	Ui	129

Índex de figures

1.1	Contaminació en una planta de Pensilvania al 1973.	1
1.2	Cicle de gestió de dades mitjançant el cicle investigatiu.	2
2.1	Gràfics temporals de les partícules en suspensió	9
2.2	Gràfic temporal del NO_2	10
2.3	Gràfic temporal del contaminant O_3	11
2.4	Gràfic temporal del contaminant CO	12
3.1	Esquema de les components que formen una sèrie temporal	15
3.2	Exemple de l'ACF per una sèrie no estacionària i una sèrie estacionària	21
3.3	ACF d'una sèrie que té propietats de soroll blanc.	22
3.4	Esquema de la metodologia Box-Jenkins	22
3.5	Gràfics per estudiar la constància de la variància	23
3.6	Gràfic per estudiar l'estacionalitat d'ordre m	24
3.7	Gràfics per verificar la normalitat dels residus.	27
3.8	Gràfic amb els p-valors del test Ljung-Box.	29
3.9	Gràfic amb les circumferències per verificar la invertibilitat i estacionarietat del model.	30
4.1	Primera pàgina de l'aplicació web	35
4.2	Pantalla per pujar les dades a l'aplicació	36
4.3	Càrrega d'una sèrie temporal a la web	36
4.4	Panell per fer prediccions segons la metodologia d'allisat exponencial	37
4.5	Visualització de les pestanyes Taula, Descomposició i Tipus de sèrie .	38
4.6	Pestanyes Estacionalitat i Gràfic de la predicció	39
4.7	Pestanyes Taula de la predicció i ets	39
4.8	Panell per a la preparació prèvia al model	40
4.9	Pestanyes Mitjana-Var, Monthplot i Transformacions	41
4.10	Pestanyes Mitjana i ACF/PACF	41
4.11	Panell Models per modelitzar un model SARIMA	42
4.12	Validació de la normalitat	43

4.13	Pestanya Variància dins de la validació	43
4.14	Pestanya ACF/PACF dins de la validació	44
4.15	Pestanya Ljung-Box	44
4.16	Pestanya Estacionarietat/Invertible	45
4.17	Pestanya Tests numèrics	45
4.18	Panell per la predicció de la metodologia Box-Jenkins	46
4.19	Pestanya Predicció Out-of-Sample	46
4.20	Pestanyes Prediccions i Taula de prediccions	47
4.21	Pestanya Auto.arima	47
4.22	Panell principal pel prophet	48
4.23	Predicció amb els punts de canvi	48
4.24	Pestanya Components	49
4.25	Pestanyes Cross Validation i Gràfic de Cross Validation	50
4.26	Pestanya Taula de les mètriques CV	50
4.27	Panell principal pel Millor model	51
5.1	Descomposició de les components de la sèrie NO_2	53
5.2	Càlcul de l'indicador MSE per cada tipus de sèrie en base a les dades del NO_2	53
5.3	Tipus de sèrie temporal que segueix NO_2 i l'estimació dels coeficients.	54
5.4	Estimació dels coeficients de l'estacionalitat de cada dia de la setmana pel NO_2	55
5.5	Predicció dels pròxims 212 dies per la concentració del NO_2	55
5.6	Predicció dels pròxims 30 dies del NO_2 segons l'algoritme automàtic.	56
5.7	Gràfics per estudiar la variància de la sèrie	57
5.8	Gràfics per estudiar la variància de la sèrie aplicada la transformació	57
5.9	Monthplot de la sèrie NO_2 amb mitjana constant	58
5.10	Gràfics per estudiar la mitjana de la sèrie NO_2	59
5.11	ACF i PACF de la sèrie NO_2	60
5.12	Estimació del model 1	61
5.13	Estimació del model 1 amb els paràmetres significatius	62
5.14	Estimació del model 2	62
5.15	Validació de la normalitat	63
5.16	Validació de la normalitat amb els contrastos numèrics	63
5.17	Validació de la variància	64
5.18	Validació de la variància amb el contrast Breusch-Pagan	65
5.19	Gràfics per validar la independència dels residus	65
5.20	Validació de la independència dels residus	66
5.21	Validació de l'estacionarietat del model	66

5.22	Comprovació de la volatilitat dins del model	67
5.23	Anàlisi de l'estabilitat del model	67
5.24	Predicció Out-of-Sample	68
5.25	Predicció de les 30 observacions futures	68
5.26	Identificació automàtica del model i predicció de 30 observacions futures	69
5.27	Ajust i predicció de la sèrie NO_2	69
5.28	Estimació de la tendència i punts de canvi d'aquesta component. . . .	69
5.29	Components del model no paramètric proporcionat per Prophet	70
5.30	Mètriques de regressió	70
5.31	Mètriques de regressió per comparar els diferents models	71
6.1	Estimació de les components segons Prophet	75
6.2	Millor model pel $PM_{2.5}$	76
6.3	Millor model pel PM_{10}	77
6.4	Millor model pel CO	77
6.5	Millor model per l' O_3	78
A.1	Components del $PM_{2.5}$	86
A.2	Estimació dels paràmetres $PM_{2.5}$	86
A.3	Patró estacional $PM_{2.5}$	87
A.4	Predicció $PM_{2.5}$	87
A.5	ETS $PM_{2.5}$	88
A.6	ACF i PACF $PM_{2.5}$	88
A.7	Estimació del model pel $PM_{2.5}$	89
A.8	Test numèrics per la validació del $PM_{2.5}$	89
A.9	Estabilitat del model pel $PM_{2.5}$	89
A.10	Predicció $PM_{2.5}$	89
A.11	Auto.arima $PM_{2.5}$	89
A.12	Predicció segons prophet $PM_{2.5}$	90
A.13	Components del PM_{10}	90
A.14	Estimació dels paràmetres PM_{10}	90
A.15	Predicció PM_{10}	91
A.16	ETS PM_{10}	91
A.17	ACF i PACF PM_{10}	92
A.18	Estimació del model pel PM_{10}	92
A.19	Test numèrics per la validació del PM_{10}	92
A.20	Estabilitat del model pel PM_{10}	92
A.21	Predicció PM_{10}	93
A.22	Auto.arima PM_{10}	93
A.23	Predicció segons prophet PM_{10}	93

A.24 Components del CO	93
A.25 Estimació dels paràmetres CO	94
A.26 Patró estacional CO	94
A.27 Predicció CO	95
A.28 ETS CO	95
A.29 ACF i PACF CO	96
A.30 Estimació del model pel CO	96
A.31 Test numèrics per la validació del CO	96
A.32 Estabilitat del model pel CO	96
A.33 Predicció CO	97
A.34 Auto.arima CO	97
A.35 Predicció segons prophet CO	97
A.36 Components del O_3	97
A.37 Estimació dels paràmetres O_3	97
A.38 Predicció O_3	98
A.39 ETS O_3	98
A.40 ACF i PACF O_3	98
A.41 Estimació del model pel O_3	99
A.42 Test numèrics per la validació del O_3	99
A.43 Estabilitat del model pel O_3	99
A.44 Predicció O_3	99
A.45 Auto.arima O_3	99
A.46 Predicció segons prophet O_3	100

Índex de taules

2.1	Descripció dels contaminants de l'aire que s'han estudiat.	7
2.2	Efectes en la salut dels diferents contaminants de l'aire.	7
2.3	Taules resum dels estadístics principals	9
2.4	Taula resum dels estadístics principals pel NO_2	10
2.5	Taula resum dels estadístics principals pel O_3	11
2.6	Taula resum dels estadístics principals pel CO	12
2.7	Taula resum dels rangs de concentració que ha de presentar cada contaminant per determinar el nivell de qualitat de l'aire.	13
3.1	criteri per identificar el tipus de sèrie temporal	16
3.2	Moments dels processos estacionaris.	20
3.3	Fórmules de l'autocorrelació.	20
5.1	Identificació dels models per la sèrie NO_2	61
5.2	Model seleccionat per analitzar la sèrie NO_2	62
6.1	Tipus de sèrie per cada un dels contaminants	72
6.2	Tipus de sèrie identificats per la funció ETS	73
6.3	Identificació dels models per cada un dels contaminants	74
6.4	Identificació automàtica dels models per cada un dels contaminants	75

Listings

2.1	Sintaxi aplicada en R per substituir els valors negatius per NA	8
2.2	Sintaxi aplicada en R per imputar nous valors	8
3.1	Sintaxi aplicada en R per estimar una sèrie de tipus I	16
3.2	Sintaxi aplicada en R per estimar una sèrie de tipus II	17
3.3	Sintaxi aplicada en R per estimar una sèrie de tipus III	17
3.4	Sintaxi aplicada en R per estimar una sèrie de tipus IV	18
3.5	Sintaxi aplicada en R per estimar una sèrie de tipus V	18
3.6	Sintaxi aplicada en R per estimar una sèrie de tipus VI	18
3.7	Sintaxi aplicada en R per d'identificació automàtica del tipus de sèrie	19
3.8	Sintaxi aplicada en R per estimar els coeficients d'un model SARIMA tenint en compte l'intercept	26
3.9	Sintaxi aplicada en R per estimar els coeficients d'un model SARIMA tenint en compte l'intercept	26
3.10	Sintaxi aplicada en R pel contrast Shapiro-Wilk	27
3.11	Sintaxi aplicada en R pel contrast Anderson-Darling	27
3.12	Sintaxi aplicada en R pel contrast Jarque-Bera	28
3.13	Sintaxi aplicada en R pel contrast de Breusch-Pagan	28
3.14	Sintaxi aplicada en R pel contrast de Durbin-Watson	29
3.15	Sintaxi aplicada en R pel modelatge automàtic (auto.arima)	30

Capítol 1

Introducció

Un dels problemes més antics i sense una solució definitiva és la contaminació. Des de fa milers d'anys que s'han estat emetent diferents contaminants a l'aire. Si ens remuntem a l'inici de la història, els focs de llenya emetien fum contaminat que afectava la salut enagrint els pulmons. Es creu que en l'època dels romans es va iniciar l'emissió de contaminants metàl·lics com el plom i amb la Revolució Industrial els efectes dels contaminants provinents dels cotxes, indústries i el carbó van incrementar les conseqüències. En aquesta etapa es van agreujar els efectes a la salut i augmentar els episodis de ciutats cobertes de núvols de contaminació (1.1).



Figura 1.1: Contaminació en una planta de Pensilvania al 1973.

Observant tots aquests efectes, al llarg de la història s'han intentat implementar diverses mesures per tenir un aire net sense cap millora considerable, ja que es prioritza la producció i els beneficis econòmics. Avui dia, encara estem vivint dia rere dia els efectes perjudicials dels diferents contaminants, tant a curt com a llarg termini mitjançant malalties respiratòries, cardiovasculars o càncer de pulmó mentre s'està buscant una solució.

La qüestió és, com ens podem protegir? O de manera més específica, com podem saber el futur nivell de contaminació per posar una solució preventiva?

Actualment, hi ha a disposició de tot el públic diferents eines que visualitzen la qualitat de l'aire en temps real i bases de dades de diferents estacions que recullen les dades cada cert temps dels diferents contaminants de l'aire. Però, després d'una exhaustiva recerca en els països més avançats en aquest àmbit, no s'ha trobat una eina per predir la qualitat de l'aire en un futur pròxim. A més, s'ha observat una gran quantitat de fòrums d'anàlisi de dades proposant reptes per crear un model que ajustés i, sobretot, predigues de manera precisa els diferents contaminants de l'aire. Per exemple, estava públic a DRIVENDATA un concurs organitzat per la NASA on l'objectiu principal era crear un model per poder predir la concentració de les partícules pesants que es troben a l'aire (mirar *Competition: NASA Airathon: Predict Air Quality (Particulate Track)*, [s. d.]). Agafant de referència aquesta informació, tenint en compte la situació alarmant que s'està vivint pel canvi climàtic i com la recent pandèmia va aconseguir disminuir la contaminació de manera lleugera en diverses ciutats industrialitzades, és el moment òptim per invertir i investigar solucions per reduir la contaminació.

En aquest estudi, a través de la metodologia PPDAC (1.2), es vol fer front a la problemàtica que causa la contaminació. Concretament, es vol posar a l'abast de la ciutadania els valors futurs de la qualitat de l'aire pels diferents contaminants.

The Investigative Cycle (PPDAC)

Problem Solving PPDAC Cycle

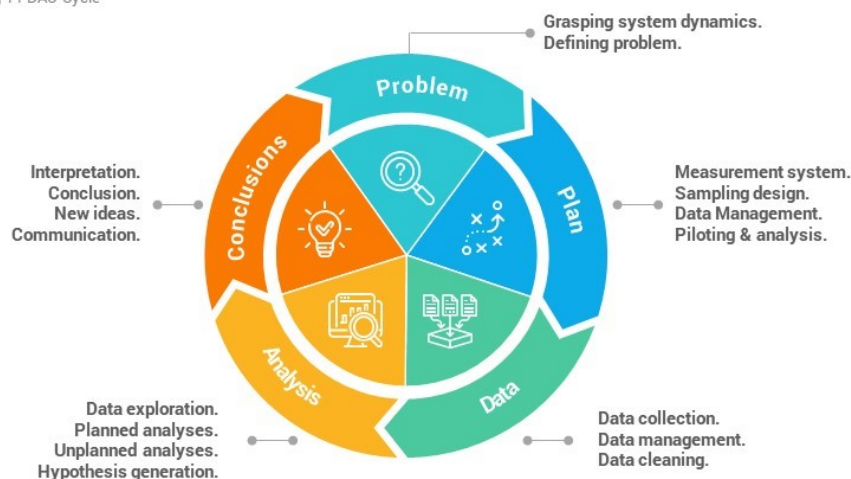


Figura 1.2: Cicle de gestió de dades mitjançant el cicle investigatiu.

Seguint la metodologia de treball, en primer lloc, es necessita identificar el problema. Aquest, com ja s'ha explicat, és la contaminació. En ser un tema tan ampli i amb diversos focus, s'ha decidit proposar una solució a la desinformació

dels futurs nivells de contaminants. És a dir, posar a l'abast de tota la ciutadania un possible valor de referència de la concentració dels diferents contaminants en un futur pròxim.

En segon lloc, s'ha plantejat un pla per fer front a la qüestió. Aquest consisteix a crear una interfície interactiva per facilitar la modelització. És a dir, s'ha creat una plataforma amb el paquet *shiny* que proporciona el llenguatge de programació R, on es pot aplicar tres metodologies de modelització diferents, tant a les dades que proporciona per defecte aquesta plataforma com a altres dades que es puguin carregar de diferents estacions. Dins d'aquesta interfície es podrà identificar quina metodologia de modelització s'adapta millor a les dades recollides del contaminant que s'estigui utilitzant i retornarà el nivell de concentració que presentarà en el futur.

Cal fer èmfasi en el fet que la creació d'aquesta plataforma web permet no només tractar amb les dades recollides pels diferents contaminants sinó que també ofereix la possibilitat d'analitzar i modelitzar qualsevol altre tipus de sèries temporals sense la necessitat de conèixer la sintaxi de programació. És a dir, com a objectiu paral·lel a l'estudi de la qualitat de l'aire, també s'ha proposat de cara a l'ensenyament i/o agilitzar el tractament de sèries temporals, fer que la plataforma disposi de diferents opcions per tractar sèries temporals amb característiques diverses i no necessàriament iguals a les que es presenten en aquest àmbit.

Definit el pla a seguir, es passa a la tercera part de l'estudi. La "D" que identifica el procés de selecció de dades. En aquesta recerca, s'ha decidit implementar les dades de cinc contaminants diferents dins de la interfície interactiva web per poder seleccionar-les i treballar amb elles per defecte. Tot i això, com s'ha esmentat anteriorment, aquesta disposarà de l'opció de càrrega d'altres sèries temporals d'altres contaminants o d'altres àmbits.

En aquest projecte, fent una breu recerca dins dels països pioners en la recollida de dades ambientals, s'ha decidit treballar amb les dades de Califòrnia per una qüestió històrica. En un dels articles publicats sobre les diferents conseqüències de la qualitat de l'aire en la salut humana, s'explica que l'estat de Califòrnia va ser el primer en regular la qualitat de l'aire. En honor a aquest fet, s'ha considerat un bon exemple d'estudi.

Seguidament, per la part d'anàlisi, es presenten dos blocs dins d'aquesta memòria. En el primer capítol es contarà cada una de les funcions que s'han implementat dins de la plataforma web. En aquest apartat es podrà veure cada una de les sortides que ofereix la web i com es relaciona amb cada una de les metodologies estudiades i més usades per l'estudi de sèries temporals. En el segon capítol es presentarà pas a

pas, l'anàlisi d'un dels contaminants que ofereix la interfície i a quines conclusions s'ha arribat.

Per acabar l'anàlisi, s'afegirà un breu resum de les conclusions a les quals s'han arribat per cada un dels contaminants i es presenta una breu discussió de les semblances i diferències entre ells.

Les hipòtesis prèvies a l'anàlisi de la qualitat de l'aire són que aquest tipus de dades que s'estudiaran presentaran poca tendència perquè la concentració de contaminant en l'aire depèn molt de factors externs i sovint succeeixen transformacions en les molècules. A més, es creu que les sèries seran volàtils, és a dir, es presentaran alts nivells de manera ràpida i tornarà a estabilitzar-se la concentració. Molt probablement això serà a causa de factors com el temps, la temperatura i altres factors meteorològics o esdeveniments puntuals. També, és molt probable que s'observi estacionalitat setmanal amb un patró d'altres emissions de contaminants durant els dies feiners.

Per acabar amb el recorregut d'anàlisis i seguint la metodologia PPDAC, es plantejaran les conclusions de l'estudi. S'identificarà quina ha estat la metodologia òptima per modelitzar els diferents contaminants i quina serà la concentració en el futur pròxim.

Per dur a terme aquest estudi per poder ajustar i predir de la forma més idònia, s'estudiaran les diferents sèries a través de les metodologies de mitjanes mòbils, és a dir, els models d'allisat exponencial, s'aplicarà la metodologia Box-Jenkins amb els models SARIMA i es treballarà amb el paquet creat per Facebook, *prophet*, on s'ajusta la sèrie de manera no paramètrica.

Aquestes metodologies es presentaran de manera més extensa dins de l'apartat de metodologies.

Finalment, per dur a terme el treball proposat, complir els objectius i estudiar les hipòtesis plantejades, es recollirà dins d'aquest informe els següents capítols:

- **Descripció de les dades:** on s'explicarà de manera més detallada les dades amb les quals s'han elaborat els models i el motiu de per què s'han escollit aquestes de manera concreta.
- **Metodologia:** es detallarà cada una de les metodologies aplicades en l'anàlisi i implementades en la interfície per poder modelitzar de manera automàtica. A més, es presentaran els recursos informàtics per dur a terme aquest treball.

- **Guia d'ús de la interfície:** en aquest capítol es presenta un manual d'ús de l'aplicació per les dades implementades en ella i les seves diferents opcions i funcionalitats.
- **Anàlisi del contaminant NO_2 :** es presentarà pas a pas el procediment que s'ha seguit per modelitzar i complir el propòsit de mostrar prediccions pel contaminant NO_2 .
- **Resum dels resultats:** per acabar l'anàlisi es presentarà els diferents resultats que s'han obtingut per la resta de contaminants.
- **Conclusions:** finalment s'ha fet una valoració del treball fent menció als resultats aconseguits, el recorregut que s'ha dut a terme per la creació d'aquesta eina interactiva, com es pot implementar en diferents àmbits i possibles actualitzacions pel futur.

Capítol 2

Descripció de les dades

En aquest apartat es definiran els contaminants que s'utilitzaran en aquest projecte, la simptomatologia que provoquen a la salut en inhalar-los i una breu descripció de les dades.

2.1 Origen, definició i simptomatologia dels contaminants

Per assegurar que les dades publicades són de qualitat, s'ha decidit treballar amb dades d'algun dels països més avançats en aquest àmbit. Els Estats Units d'Amèrica ha patit molts canvis al llarg del temps a causa de la contaminació de l'aire. En tenir tants estats i ciutats, en aquest informe s'ha decidit analitzar les dades extretes de Califòrnia, ja que històricament aquest presentava altes concentracions de contaminants de l'aire a causa de la geografia, el clima i l'alta densitat de població.

En ser un dels estats amb pitjor qualitat de l'aire en el passat, tot i que en el mateix país ja s'havia plantejat una manera per alleujar i combatre els mals resultats recollits de la contaminació, es va donar un permís especial a Califòrnia per crear unes normes més restrictives respecte a les emissions de contaminants. Això va resultar amb una comissió per treballar en aquest problema convertint-la així en l'estat líder en aquest àmbit i amb molt pocs anys aconseguir grans resultats (*History — California Air Resources Board*, [s. d.]).

Per dur a terme aquesta petita anàlisi, s'ha decidit examinar les concentracions dels cinc principals contaminants de l'aire i dels quals es coneixen quins són els efectes a la salut. En concret, s'ha implementat dins de la interfície web i analitzat dins d'aquest informe les partícules en suspensió, tant de 2.5 com 10 micres, el diòxid

de nitrogen, l'ozó troposfèric i el monòxid de carbó durant el 2021 extretes de l'EPA (*Environmental Protection Agency*), (*Download Daily Data — US EPA*, [s. d.]).

A continuació es farà una breu definició dels contaminants estudiats (2.1).

Contaminant	Descripció
<i>Partícules en suspensió (PM)</i>	Es componen per partícules microscòpiques tant sòlides com líquides provinents de substàncies orgàniques i inorgàniques que es troben a l'aire. Es divideixen entre $PM_{2.5}$ que són totes les partícules menors a 2.5 micres i incloent les partícules ultrafines i les PM_{10} que són les menors a 10 micres.
<i>Òxids de nitrogen</i>	Concretament, s'estudiarà el diòxid de nitrogen (NO_2). Aquest és un compost químic que s'utilitza principalment en la producció de fertilitzants. Es presenta com un gas marró en altes temperatures.
<i>Ozó troposfèric</i>	Es defineix com una molècula inorgànica inestable composta per tres àtoms d'oxigen (O_3). És un gas incolor i invisible que es caracteritza per la seva olor penetrant.
<i>Monòxid de carbó</i>	Es compon per un àtom de carboni i un àtom d'oxigen connectats per un triple enllaç (CO). És un gas inflamable, incolor, inodor, insípid i altament verinós.

Taula 2.1: Descripció dels contaminants de l'aire que s'han estudiat.

Ja definits els contaminants, també és necessari conèixer els diferents efectes que tenen aquests sobre la salut segons les concentracions en què es presentin (2.2).

Contaminant	Efectes en la salut
<i>Partícules en suspensió (PM)</i>	Provoquen efectes a curt i llarg termini augmentant la probabilitat d'afeccions respiratòries i cardiovasculars. Sol ajudar en l'empitjorament dels símptomes de l'asma i altres problemes respiratoris. També incrementa el nombre de defuncions a causa de malalties cardiovasculars i respiratòries i per càncer de pulmó.
<i>Òxids de nitrogen</i>	Aquest contaminant provoca inflamació a les vies respiratòries si es té una exposició de curta durada. D'altra banda, amb una exposició llarga als òxids de nitrogen poden causar més símptomes en els nens asmàtics.
<i>Ozó troposfèric</i>	Els efectes per a la salut de l'ozó generalment són tos, irritacions a la faringe, el coll i els ulls i dificultats respiratòries com la gola seca. Aquest contaminant també augmenta els símptomes de l'asma, redueix la funció pulmonar i a vegades provoca inflamació a les vies respiratòries.
<i>Monòxid de carbó</i>	És un contaminant que afecta en major proporció a les persones que pateixen de malalties del cor o pulmó. La inhalació d'aquest contaminant pot provocar dolors de cap, nàusees, vòmits, mareigs, visió borrosa, confusió, dolors de pit, debilitat, dificultats respiratòries, convulsions i comes.

Taula 2.2: Efectes en la salut dels diferents contaminants de l'aire.

2.2 Depuració de les dades

En aquest apartat s'explica quins han estat els tractaments previs de les dades amb què es treballen.

En primer lloc, en estar actuant amb diverses bases de dades alhora i amb bastants observacions cadascuna s'ha utilitzat el paquet *tydiverse* que ofereix el llenguatge R.

En segon lloc, per poder operar sense problemes assegurant que la importació de les dades ha estat satisfactòria s'ha comprovat si totes les concentracions són positives perquè d'altra manera voldria dir que no es va registrar correctament. Aplicant la sintaxi presentada a 2.1 s'ha convertit les concentracions de les diferents dades en NA.

```
1 mutate(serie, var_concentracio = replace(var_concentracio, which(
  var_concentracio < 0), NA))
```

Listing 2.1: Sintaxi aplicada en R per substituir els valors negatius per NA

Tenint en compte que amb les metodologies que s'està treballant és necessari treballar amb dades completes, s'ha aplicat un mètode d'imputació simple per estimar un valor. En concret, com que no hi ha molts valors faltans s'ha decidit aplicar la imputació d'aquests fent interpolació lineal. Aquest mètode tracta d'estimar un valor faltant que es troba entre valors observats (consultar Johannesson et al., 2009). La sintaxi 2.2 és la que s'utilitza per estimar el nou valor en la creació de la sèrie temporal.

```
1 na_interpolation(window(ts(var_concentracio, start = 1, freq = m),
  start = 1), option='linear')
```

Listing 2.2: Sintaxi aplicada en R per imputar nous valors

Abans d'iniciar amb l'anàlisi descriptiva de les dades, cal esmentar que les sèries temporals dels contaminants escollits només s'han utilitzat la meitat per dur a terme l'entrenament dels models i totes les dades registrades i importades per fer les prediccions a futur.

2.3 Anàlisi descriptiva de les dades

En aquesta secció es presentaran de manera breu les principals estadístiques descriptives per les dades. D'aquesta manera es podrà tenir un major coneixement previ de les dades.

2.3.1 Partícules en suspensió (PM)

Com s'ha comentat a l'inici d'aquest capítol, un dels contaminants que es treballaran en aquest projecte són les partícules en suspensió, particularment les $PM_{2.5}$ i les PM_{10} .

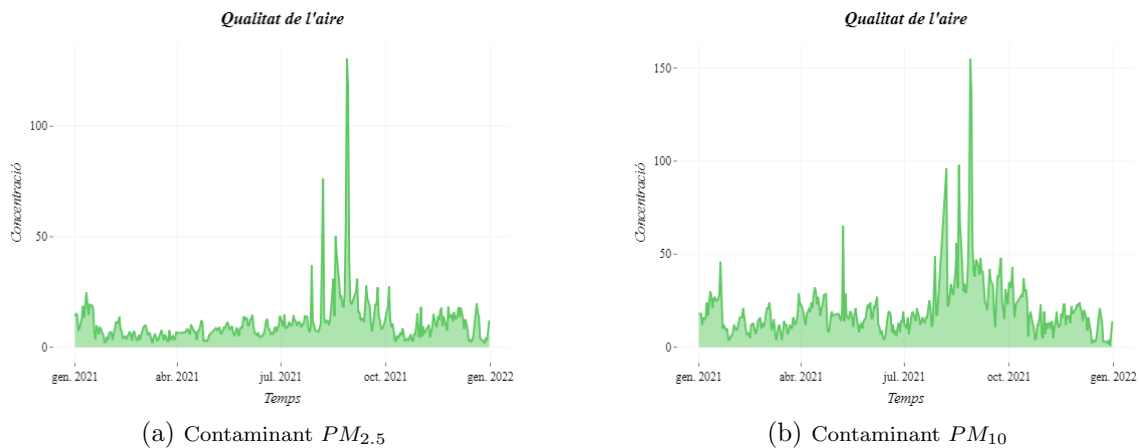


Figura 2.1: Gràfics temporals de les partícules en suspensió

Com es presenta en la figura 2.1 a l'esquerra hi ha el gràfic temporal per les $PM_{2.5}$ (2.1a) i a la dreta es pot observar com es distribueixen les PM_{10} (2.1b). Es pot veure clarament com les concentracions que es presenten pel contaminant $PM_{2.5}$ són en general menors a les del contaminant PM_{10} . Això és degut al fet que hi ha menor nombre d'elements considerats $PM_{2.5}$ que elements que es consideren PM_{10} (mirar *Principals contaminants atmosfèrics. Agència de Salut Pública de Catalunya (ASPCAT)*, [s. d.]).

Estadístic	Concentració (μ/m^3)	Estadístic	Concentració (μ/m^3)
Min	1.9	Min	1
Q1	5.9	Q1	12
Mediana	8.9	Mediana	16
Mitjana	11.28	Mitjana	19.95
Q3	13.18	Q3	23
Max	130.5	Max	155

(a) Contaminant $PM_{2.5}$

(b) Contaminant PM_{10}

Taula 2.3: Taules resum dels estadístics principals

Tant a les figures mostrades com a les taules amb els principals estadístics per cada contaminant (2.3) s'exposen resultats molt semblants entre els dos contaminants, millor dit, ambdós contaminants, al ser de la mateixa família, es distribueixen de manera molt semblant.

2.3.2 Òxids de nitrogen (NO_2)

Seguint amb l'anàlisi descriptiva de les variables amb què es treballen, en aquesta secció es presentaran les principals gràfiques i taules pel contaminant NO_2 .

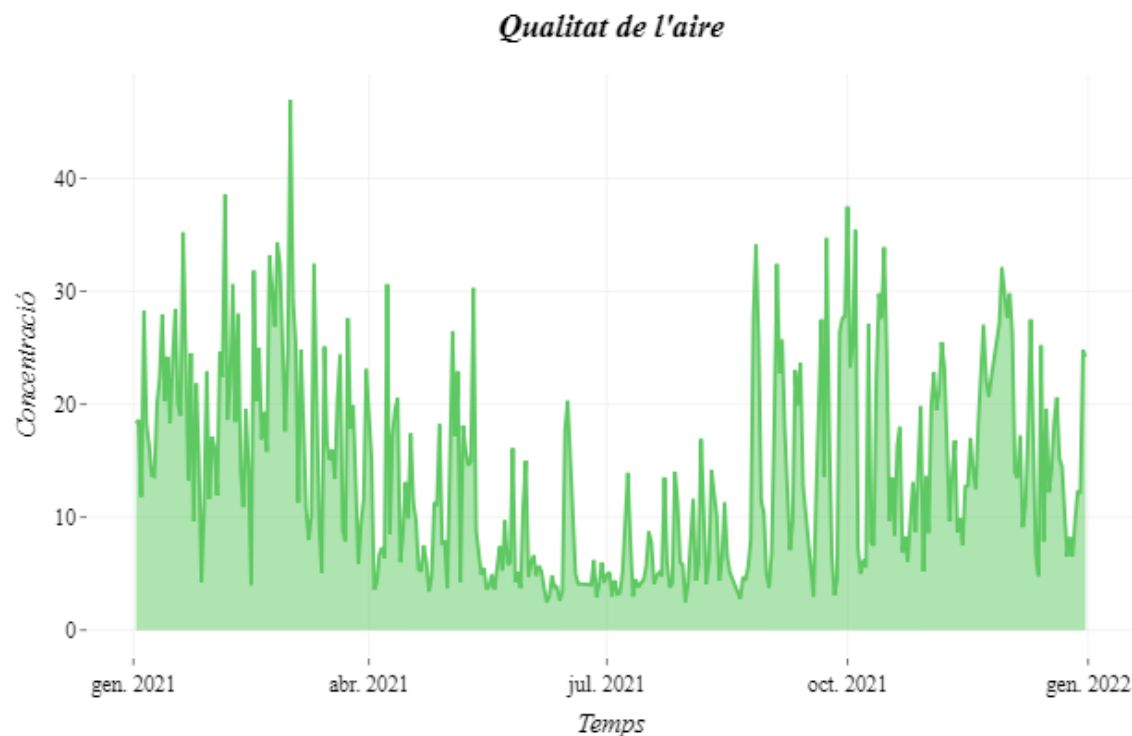


Figura 2.2: Gràfic temporal del NO_2

A la gràfica 2.2 s'observa com la concentració d'aquest contaminant és bastant volàtil al llarg del temps. A més, sembla presentar altes concentracions en els mesos més freds (majors a 20 ppb) i en els mesos d'estiu semblen presentar-se concentracions menys elevades (al voltant del 5 ppb o 10 ppb).

Estadístic	Concentració (ppb)
Min	2.5
Q1	6
Mediana	12.6
Mitjana	14.17
Q3	20.3
Max	46.9

Taula 2.4: Taula resum dels estadístics principals pel NO_2

Amb la taula 2.4 s'observa com la mitjana i la mediana presenten valors molt propers. Això és a causa que les concentracions altes no només es presenten a l'inici de l'any sinó que també al final. Com en l'anàlisi s'estarà treballant amb el màxim diari es considera important destacar que el màxim registrat el 2021 va ser de 46.9

ppb. Amb aquest valor, en els pròxims apartats es podrà interpretar si es troba, o no, dins dels límits considerats bona qualitat de l'aire.

2.3.3 Ozó troposfèric (O_3)

El punt següent s'analitzarà l'ozó troposfèric.

En el gràfic temporal (2.3) es veuen petites oscil·lacions i sembla presentar-se una tendència estable al llarg de l'any 2021. Aparentment, les concentracions que presenta aquest contaminant són molt properes al zero, tot i que en el període en què s'està treballant no s'ha registrat cap concentració de zero.

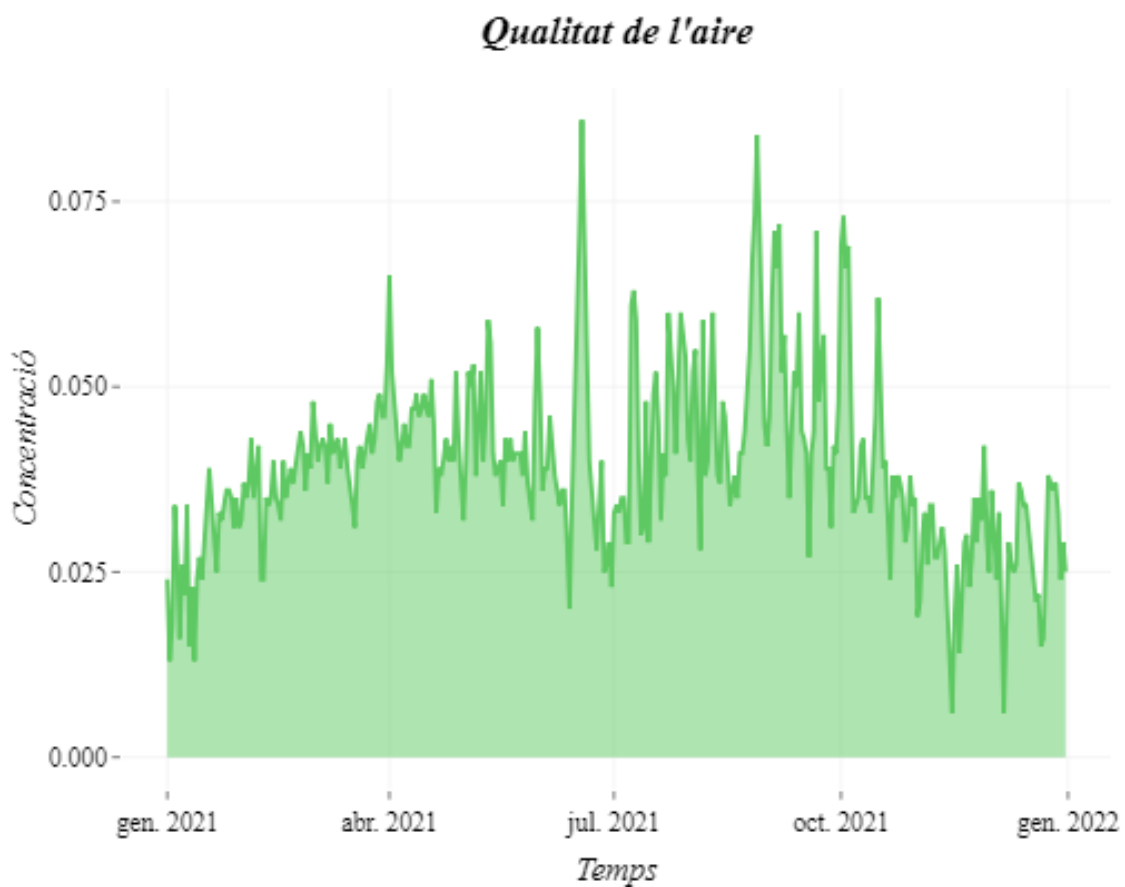


Figura 2.3: Gràfic temporal del contaminant O_3

A la taula on es presenten els principals estadístics descriptius de les dades (2.5) es pot veure com la mitjana de les concentracions registrades es troba al voltant de 0.04 ppm i s'està treballant amb concentracions d'entre 0.006 ppm a 0.086 ppm.

Estadístic	Min	Q1	Mediana	Mitjana	Q3	Max
Concentració (ppm)	0.006	0.033	0.039	0.03935	0.045	0.086

Taula 2.5: Taula resum dels estadístics principals pel O_3

2.3.4 Monòxid de carbó (CO)

Per acabar l'anàlisi descriptiva i seguint l'estructura dels anteriors apartats, s'estudiaran els gràfics i estadístics principals per conèixer com es distribueix el monòxid de carboni.

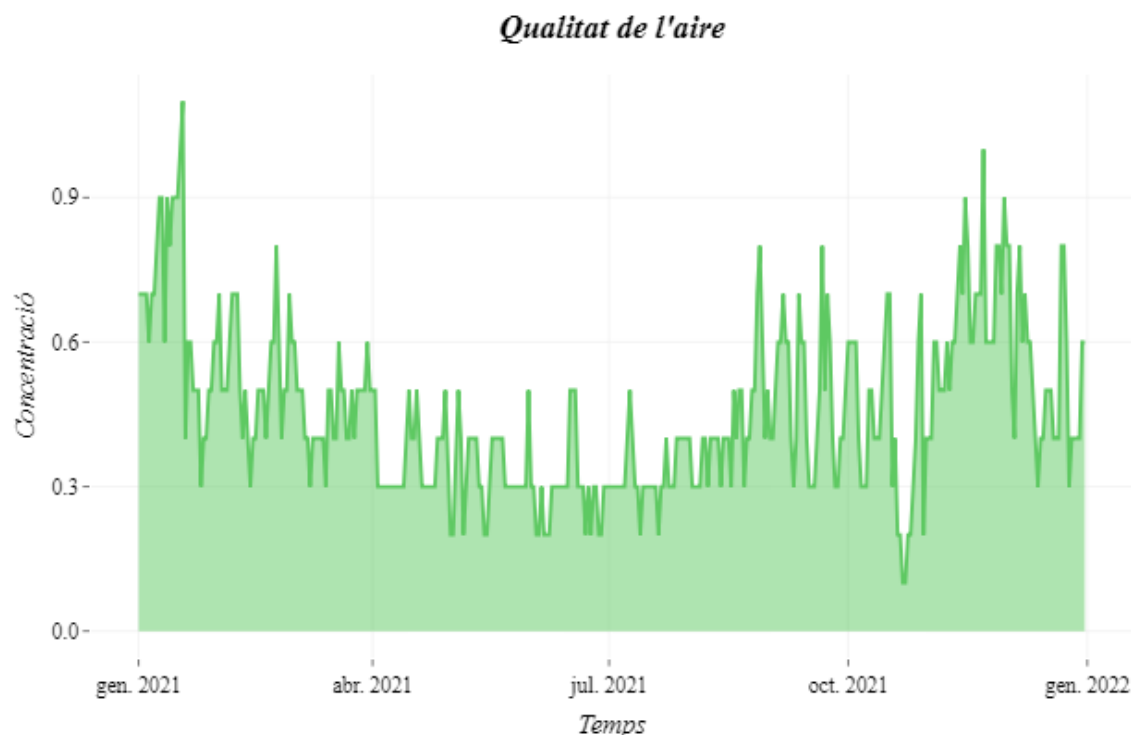


Figura 2.4: Gràfic temporal del contaminant CO

En el gràfic temporal del monòxid de carboni que es presenta a la figura 2.4 es presenta amb estabilitat al llarg del 2021 i amb diversos pics en els mesos d'hivern. En la taula 2.6 mostra de manera més detallada que la invariabilitat observada en el gràfic es troba al voltant d'una mitjana de 0.456 ppm.

Els valors registrats durant aquest any pel contaminant es distribueixen entre els 0.1 ppm i el 1.1 ppm .

Estadístic	Min	Q1	Mediana	Mitjana	Q3	Max
Concentració (ppm)	0.1	0.3	0.4	0.456	0.6	1.1

Taula 2.6: Taula resum dels estadístics principals pel CO

2.4 Rangos de la qualitat de l'aire

En aquest apartat es presenta un breu resum dels diferents rangs que ha de presentar cada un dels contaminants per determinar si es pot, o no, considerar una bona qualitat de l'aire.

Nivell de qualitat de l'aire	$PM_{2.5}$	PM_{10}	NO_2	O_3	CO
AQI bo	0 – 12.0	0 - 54	0 - 53	0 - 0.054	0 – 4.4
AQI moderat	12.1 – 35.4	55 – 154	54 - 100	0.055 - 0.070	4.5 – 9.4
AQI no saludable per grups sensibles	35.5 – 55.4	155 – 254	101 - 360	0.071 - 0.085	9.5 – 12.4
AQI no saludable	55.5 – 150.4	255 – 354	361 - 649	0.086 - 0.105	12.5 – 15.4
AQI molt poc saludable	150.5 – 250.4	355 – 424	650 - 1249	0.106 - 0.200	15.5 – 30.4
AQI perillós	250.5 – 500.4	425 – 604	1250 - 2049	*	30.5 – 50.4

Taula 2.7: Taula resum dels rangs de concentració que ha de presentar cada contaminant per determinar el nivell de qualitat de l'aire.

A la taula 2.7 es presenta un resum dels diferents rangs de concentració que s'han de presentar per cada un dels contaminants per decidir si el nivell de qualitat d'aire és bo o pot ser perjudicial d'alguna manera a la salut (mirar *How is the AQI calculated?* — *US EPA*, [s. d.]).

Capítol 3

Metodologia

En aquest capítol s'han agrupat les diferents metodologies aplicades a les diferents dades per identificar quina d'elles troba el model que millor s'adapta i les explica. Per tant, el model serà capaç de predir de manera òptima.

Cal tenir en compte que les nostres dades han estat recollides cada cert temps, en concret, s'ha treballat amb dades diàries. Amb aquest motiu, per dur a terme la seva anàlisi s'ha de tenir en compte el temps. Per fer-ho, s'aplicaran les diferents metodologies estadístiques que s'adeqüen a les sèries temporals.

3.1 Components d'una sèrie temporal

Abans d'iniciar amb l'explicació teòrica dels diferents mètodes aplicats dins l'estudi és necessari introduir les característiques de les sèries temporals.

Definició: Una sèrie temporal es defineix com una seqüència d'observacions (dades) sobre una variable aleatòria ordenada cronològicament i recollides de manera equiespaiada en el temps. L'expressió matemàtica és:

$$\{X_t\}_{t=1,\dots,T} = \{X_1, X_2, \dots, X_T\}$$

On t fa referència a la posició temporal que ocupa el valor dins de la sèrie temporal.

Les sèries temporals estan formades per diferents components (3.1), les quals ens ajudaran a identificar quin tipus de tractament és necessari segons si estan presents o no.

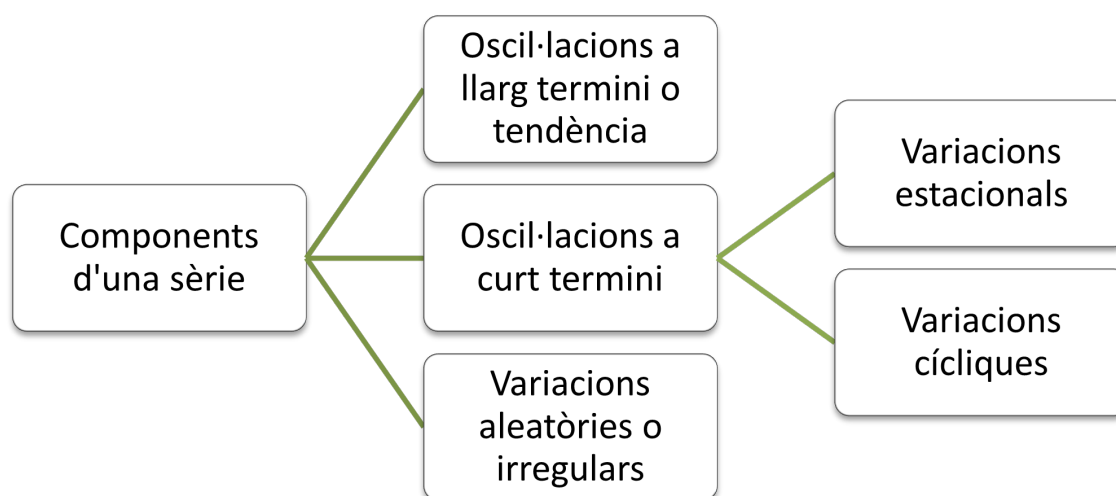


Figura 3.1: Esquema de les components que formen una sèrie temporal

- **Component tendència:** es defineix com la tendència general creixent o decreixent de les dades durant un llarg període o temps. La tendència sol ser suau i una component de llarg termini. No necessàriament ha de créixer o decreixer, pot presentar-se estable en el temps.
- **Component estacional:** aquesta recull les oscil·lacions regulars i periòdiques en un període de temps inferior a un any. Amb altres paraules, és una component que presenta el mateix patró o un patró molt semblant amb una freqüència menor a un any.
- **Component cíclica:** es presenta com les variacions de la sèrie temporal que es produeixen durant un període de temps superior a un any.
- **Component irregular:** l'última component representa les fluctuacions de la variable estudiada que no són regulars i són purament aleatòries o irregulars.

3.2 Mètodes d'allisat exponencial

Aquesta metodologia considera cada una de les observacions amb què s'està treballant donant-los diferents ponderacions, o sigui, s'apliquen diferents ponderacions al llarg del temps segons una exponencial en decaiguda. Amb altres termes, amb aquest mètode es dona major pes a les últimes observacions, és a dir, al passat més pròxim per poder predir sobretot tenint en compte els últims esdeveniments registrats.

Per iniciar amb aquesta metodologia cal identificar quines components es presenten en la sèrie temporal i amb l'ajuda de la taula 3.1 s'identificarà quin tipus de sèrie s'adapta a les dades.

		Tendència	
		No	Sí
Estacionalitat Additiva	No	Tipus I	Tipus II
	Sí	Tipus III	Tipus IV
Estacionalitat Multiplicativa	Sí	Tipus V	Tipus VI

Taula 3.1: Criteri per identificar el tipus de sèrie temporal

En cas de no identificar de manera evident si la sèrie que s'està estudiant presenta tendència o estacionalitat, un dels possibles criteris per decidir-ho objectivament és estimant per la sèrie temporal amb la qual s'està treballant cada un dels tipus de sèries i calcular el MSE (*Mean Square Error*). El MSE es calcula amb la següent fórmula:

$$MSE = \frac{1}{N} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Calculat l'indicador per cada tipus de sèrie, s'escollirà el que presenti el menor valor.

Ja sigui per la identificació de les components presents en la sèrie temporal que s'està estudiant o mitjançant el criteri d'escollir el menor MSE, per fer les prediccions s'han d'utilitzar les expressions descrites a continuació. Aquestes es basen en la mitjana ponderada definida per Holt-Winters (mirar Holt, 2004).

Seguidament, es presenten les expressions que es fan servir per al càlcul de les diferents components segons el tipus de sèrie i l'equació per dur a terme la predicció.

3.2.1 Esquema additiu

- **Sèrie tipus 1: Sense tendència ni estacionalitat (ANN)**

Equació de predicció: $\hat{X}_{T+h|T} = l_T$

Equació pel suavitzat del nivell: $l_t = \alpha X_t + (1 - \alpha)l_{t-1}$

Valors inicials: α, l_0 Paràmetre a estimar: α

L'expressió aplicada en el llenguatge R per estimar una sèrie de tipus I dins de la web es defineix en 3.1.

```
1 HoltWinters(serie, alpha = NULL, beta = F, gamma = F, seasonal
  = "additive")
```

Listing 3.1: Sintaxi aplicada en R per estimar una sèrie de tipus I

- **Sèrie tipus 2: Amb tendència i sense estacionalitat (AAN)**

Equació de predicció: $\hat{X}_{T+h|T} = l_T + hb_T$

Equació pel suavitzat del nivell: $l_t = \alpha X_t + (1 - \alpha)(l_{t-1} + b_{t-1})$

Equació pel suavitzat de la tendència: $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$

Valors inicials: α, l_0, β, b_0 Paràmetres a estimar: α, β

Dins de la web s'ha implementat la següent sintaxi per estimar les sèries de tipus II (3.2)

```
1 HoltWinters(serie, alpha = NULL, beta = NULL, gamma = F,
  seasonal = "additive")
```

Listing 3.2: Sintaxi aplicada en R per estimar una sèrie de tipus II

- **Sèrie tipus 3: Sense tendència i amb estacionalitat d'ordre m (ANA)**

Equació de predicció: $\hat{X}_{T+h|T} = l_T + S_{T+h-m}$

Equació pel suavitzat del nivell: $l_t = \alpha(X_t - S_{t-m}) + (1 - \alpha)l_{t-1}$

Equació pel suavitzat de l'estacionalitat: $S_t = \gamma(X_t - l_{t-1}) + (1 - \gamma)S_{t-m}$

Valors inicials: $\alpha, l_0, \gamma, S_{-m+1}, \dots, S_0$ Paràmetres a estimar: α, γ

Per dur a terme les estimacions de manera automàtica dins del programa aplicant les equacions descrites s'utilitza la següent funció (3.3).

```
1 HoltWinters(serie, alpha = NULL, beta = F, gamma = NULL,
  seasonal = "additive")
```

Listing 3.3: Sintaxi aplicada en R per estimar una sèrie de tipus III

- **Sèrie tipus 4: Amb tendència i estacionalitat d'ordre m (AAA)**

Equació de predicció: $\hat{X}_{T+h|T} = l_T + hb_T + S_{T+h-m}$

Equació pel suavitzat del nivell: $l_t = \alpha(X_t - S_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1})$

Equació pel suavitzat de la tendència: $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$

Equació pel suavitzat de l'estacionalitat: $S_t = \gamma(X_t - l_{t-1}) + (1 - \gamma)S_{t-m}$

Valors inicials: $\alpha, l_0, \beta, b_0, \gamma, S_{-m+1}, \dots, S_0$ Paràmetres a estimar: α, β, γ

Les estimacions de les sèries de tipus IV s'han fet amb la funció que ofereix el R, "HoltWinters" (3.4).

```
1 HoltWinters(serie, alpha = NULL, beta = NULL, gamma = NULL,
  seasonal = "additive")
```

Listing 3.4: Sintaxi aplicada en R per estimar una sèrie de tipus IV

3.2.2 Esquema multiplicatiu

- **Sèrie tipus 5: Sense tendència i amb estacionalitat multiplicativa d'ordre m (ANM)**

Equació de predicció: $\hat{X}_{T+h|T} = l_T * S_{T+h-m}$

Equació pel suavitzat del nivell: $l_t = \alpha \frac{X_t}{l_{t-1}} + (1 - \alpha)l_{t-1}$

Equació pel suavitzat de l'estacionalitat: $S_t = \gamma \frac{X_t}{l_{t-1}} + (1 - \gamma)S_{t-m}$

Valors inicials: $\alpha, l_0, \gamma, S_{-m+1}, \dots, S_0$ Paràmetres a estimar: α, γ

El codi utilitzat per estimar una sèrie de tipus V dins de la interfície es defineix en el listing 3.5.

```
1 HoltWinters(serie, alpha = NULL, beta = F, gamma = NULL,
  seasonal = "multiplicative")
```

Listing 3.5: Sintaxi aplicada en R per estimar una sèrie de tipus V

- **Sèrie tipus 6: Amb tendència i estacionalitat multiplicativa d'ordre m (AAM)**

Equació de predicció: $\hat{X}_{T+h|T} = (l_T + hb_T) * S_{T+h-m}$

Equació pel suavitzat del nivell: $l_t = \alpha \frac{X_t}{S_{t-m}} + (1 - \alpha)(l_{t-1} + b_{t-1})$

Equació pel suavitzat de la tendència: $b_t = \beta(l_t - l_{t-1}) + (1 - \beta)b_{t-1}$

Equació pel suavitzat de l'estacionalitat: $S_t = \gamma \left(\frac{X_t}{l_{t-1} + b_{t-1}} \right) + (1 - \gamma)S_{t-m}$

Valors inicials: $\alpha, l_0, \beta, b_0, \gamma, S_{-m+1}, \dots, S_0$ Paràmetres a estimar: α, β, γ

On l'expressió utilitzada es defineix com (3.6):

```
1 HoltWinters(serie, alpha = NULL, beta = NULL, gamma = NULL,
  seasonal = "multiplicative")
```

Listing 3.6: Sintaxi aplicada en R per estimar una sèrie de tipus VI

3.2.3 Identificació i predicció automàtica

Una de les opcions que ofereix el llenguatge informàtic amb què es treballa en aquest projecte és la funció “ets”. Aquesta és una funció que identifica de manera automàtica el tipus de sèrie basant-se en els mètodes descrits en l'article Hyndman; Koehler et al., 2002.

```
1 library(forecast)
2 ets(serie, allow.multiplicative.trend=T)
```

Listing 3.7: Sintaxi aplicada en R per d'identificació automàtica del tipus de sèrie

Aquesta funció només necessita les dades temporals que es volen estudiar i analitzarà si realment les components de la sèrie són, o no, prou significatives com per a determinar que formen part de la sèrie i seran necessàries per fer prediccions futures (veure 3.7).

Finalment, aquesta funció retorna el model aplicat a les dades mitjançant tres caràcters, per exemple ETS(A,N,N). La primera lletra fa referència el tipus d'error, la segona lletra identifica el tipus de tendència i l'última mostra el tipus d'estacionarietat. En tots els casos, “N” significa que no presenta la component, “A” representa la component additiva i “M” fa referència al fet que la component és multiplicativa.

3.3 Processos estocàstics

Aquesta secció introdueix la terminologia i els conceptes previs a tenir en compte per iniciar amb la metodologia de Box-Jenkins.

Principalment, per poder aplicar la metodologia de Box-Jenkins, cal treballar amb sèries temporals que siguin processos estocàstics, és a dir, serà necessari que es compleixin un seguit de propietats.

Definició: Un procés estacionari és aquell en el que la distribució conjunta de qualsevol subconjunt d'observacions $(X_{t_1}, \dots, X_{t_2})$ és la mateixa al llarg de tot el procés (més ben dit no depèn de l'origen de la sèrie):

$$F_{(X_{t_1}, \dots, X_{t_2})}(x_{t_1}, \dots, x_{t_2}) = F_{(X_{t_1+h}, \dots, X_{t_2+h})}(x_{t_1+h}, \dots, x_{t_2+h}) \quad \forall t_1, t_2, h$$

En particular, les distribucions marginals també són independents de l'origen:

$$F_{X_t}(x_t) = F_{X_{t+h}}(x_{t+h}) \quad \forall t, h$$

En sentit estricte, la sèrie temporal estudiada hauria de complir la definició

anterior, però, en ser aquesta una definició bastant restrictiva s'accepta la següent definició equivalen i més flexible.

Definició: Un procés estacionari en sentit dèbil (o de segon ordre) és aquell en què la mitjana del procés (primer moment) i les covariàncies (segon moment) no depenen de l'origen de la sèrie (mirar 3.2).

Moment	Teòric (procés estocàstic)	Mostral (sèrie temporal)
Mitjana	$E(X_t) = \mu$	$\bar{X}_t = \frac{1}{T} \sum_{t=1}^T X_t$
Variància $\sigma_X^2 = \gamma(0)$	$V(X_t) = E[(X_t - \mu)^2] = \sigma^2$	$\frac{1}{T} \sum_{t=1}^T (X_t - \bar{X})^2$
Autocovariància $\gamma(k)$	$Cov(X_t, X_{t+k}) = E((X_t - \mu)(X_{t+k} - \mu)) = \sigma_k$	$\frac{1}{T} \sum_{t=1}^{T-k} (X_{t+k} - \bar{X})(X_t - \bar{X})$

Taula 3.2: Moments dels processos estacionaris.

Dit d'una altra manera, una sèrie estacionaria en el sentit dèbil presenta la mitjana i la variància constants i les covariàncies només depenen de la distància entre les observacions (3.3).

Moment	Teòric (procés estocàstic)	Mostral (sèrie temporal)
Autocorrelació $\rho(k) = \frac{\gamma(k)}{\gamma(0)}$	$\rho(X_t, X_{t+h}) = \frac{E((X_t - \mu)(X_{t+h} - \mu))}{E[(X_t - \mu)^2]} = \frac{\sigma_k}{\sqrt{\sigma_t^2 \sigma_{t+k}^2}} = \frac{\sigma_k}{\sigma^2}$	$\frac{\sum_{t=1}^{T-k} (X_{t+k} - \bar{X})(X_t - \bar{X})}{\sum_{t=1}^T (X_t - \bar{X})^2}$

Taula 3.3: Fórmules de l'autocorrelació.

Si la sèrie compleix la definició i les propietats anteriors, només es necessitarà que el model probabilístic associat al procés segueixi una distribució normal. D'aquesta manera, el procés serà estacionari en sentit estricte i es podrà aplicar la metodologia Box-Jenkins.

Abans d'avançar amb l'explicació de les metodologies, cal mencionar les característiques i propietats que presenten els processos estacionaris, ja que aquests seran necessaris per a la identificació i validació en el procés de modelatge.

En primer lloc, la funció d'autocovariància pels processos estacionaris es defineix com:

$$\gamma(h) = Cov(X_t, X_{t+h}) = E((X_t - \mu)(X_{t+h} - \mu))$$

I presenta les següents propietats:

- $\gamma(0) = \sigma^2 > 0$
- $\gamma(k) = \gamma(-k) \forall k$
- $|\gamma(k)| \leq \gamma(0)$

En segon lloc, la funció d'autocorrelació o també denominada Funció d'autocorrelació simple (FAS), o en anglès *Auto-correlation Function (ACF)* es defineix com:

$$\rho(h) = \rho(X_t, X_{t+h}) = \frac{E((X_t - \mu)(X_{t+h} - \mu))}{\sqrt{E(X_t - \mu)^2 E(X_{t+h} - \mu)^2}} = \frac{\gamma(h)}{\gamma(0)}$$

Presentant les següents propietats:

- $\rho(0) = 1$
- $\rho(k) = \rho(-k) \forall k$
- $|\rho(k)| \leq 1 \forall k$

Per poder identificar els paràmetres que s'utilitzaran en la modelització de la sèrie, serà necessari que prèviament les diferents sèries siguin estacionàries tal com s'ha explicat anteriorment i estudiar el *Correlogram* (gràfic de l'ACF $\rho(k)$ en funció del retard k). Per verificar que la sèrie és estacionària, si s'estudia el *Correlogram* s'ha d'observar com l'ACF decau de manera ràpida cap al 0 (3.2).

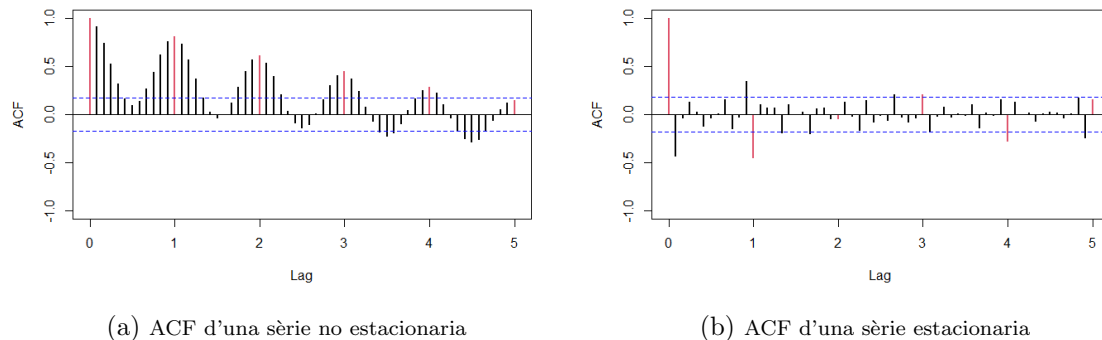


Figura 3.2: Exemple de l'ACF per una sèrie no estacionària i una sèrie estacionària

En tercer lloc, per definir la resta de paràmetres del model SARIMA caldrà estudiar la funció d'autocorrelació parcial (FAP), o en anglès *Partial Auto-correlation Function (PACF)* que correspon amb la relació entre dues variables després d'excloure l'efecte d'un o més factors independents corresponents a les observacions entre els dos retards.

$$\text{corr}(X_{t+h}, X_t | X_{t+h-1}, \dots, X_{t+1})$$

En darrer terme, per facilitar la lectura dels pròxims apartats s'han definit els termes següents:

- **Soroll blanc:** aquest és un cas particular de procés estacionari on les variables que el componen són incorrelacionades ($\rho(h) = 0 \forall h$) i està centrat en zero ($E(X_t) = 0$). Per tant, l'ACF es presenta com a la figura 3.3.

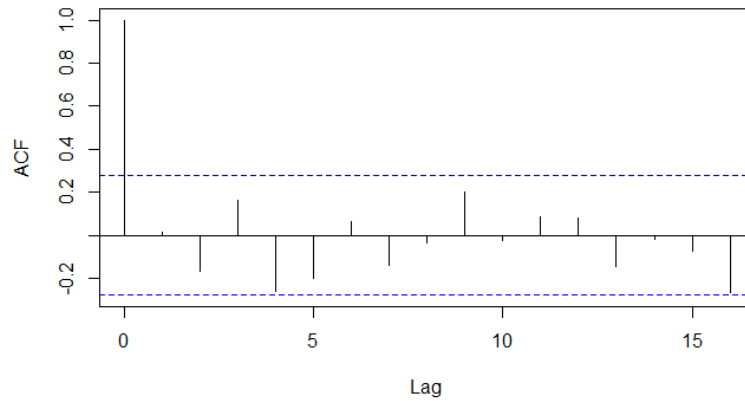


Figura 3.3: ACF d'una sèrie que té propietats de soroll blanc.

- **Operador B (*Backshift*)**: aquest operador si s'aplica a un procés (o sèrie temporal) equival a retrocedir un subíndex:

$$BX_t = X_{t-1}$$

$$B^k X_t = B \dots B^k \dots BX_t = X_{t-k}$$

3.4 Metodologia Box-Jenkins

La metodologia Box-Jenkins, anomenada així en honor als estadístics George Box i Gwilym Jenkins s'aplica als models autoregressius i mitjana mòbil, $ARMA(p, q)$, als models autoregressius integrats de mitjana mòbil, $ARIMA(p, d, q)$ o als models estacionals autoregressius integrats de mitjana mòbil, $SARIMA(p, d, q)(P, D, Q)_s$ seguint l'esquema de la figura 3.4.

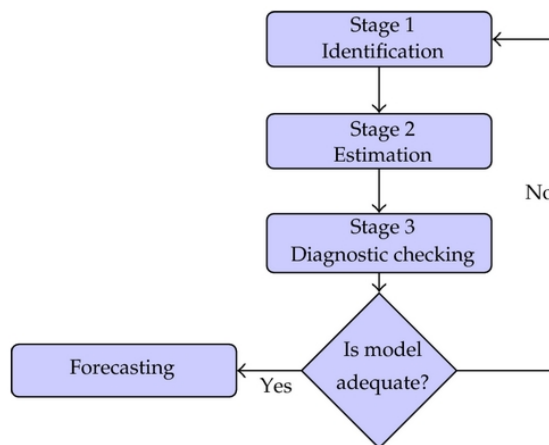


Figura 3.4: Esquema de la metodologia Box-Jenkins

La metodologia es basa en un sistema iteratiu fins a trobar el model òptim per poder acabar fent prediccions (mirar Brockwell et al., 2002). A continuació es detallarà cada una de les etapes que la formen.

3.4.1 Identificació

Per iniciar amb la identificació del model s'ha de determinar si la sèrie temporal és estacionària, és a dir, es comprovarà que la variància és constant i la mitjana és 0. A més, serà necessari observar si presenta un patró estacional per poder recollir-ho dins del model.

Per comprovar que la variància és constant s'analitzarà gràficament el diagrama de caixes per la freqüència que presenta la sèrie i es considerarà si les diferents caixes del gràfic són aproximadament de la mateixa mida. D'altra banda, també es pot comprovar si la variància és constant mitjançant un gràfic de relació entre la mitjana i la variància segons la freqüència de la sèrie. En aquest gràfic no s'ha de percebre cap tendència creixent o decreixent, sinó que les diferents observacions han de trobar-se en el mateix nivell.

En l'exemple que es presenta a la il·lustració 3.5 es pot veure clarament com la sèrie temporal estudiada no presenta variància constant.

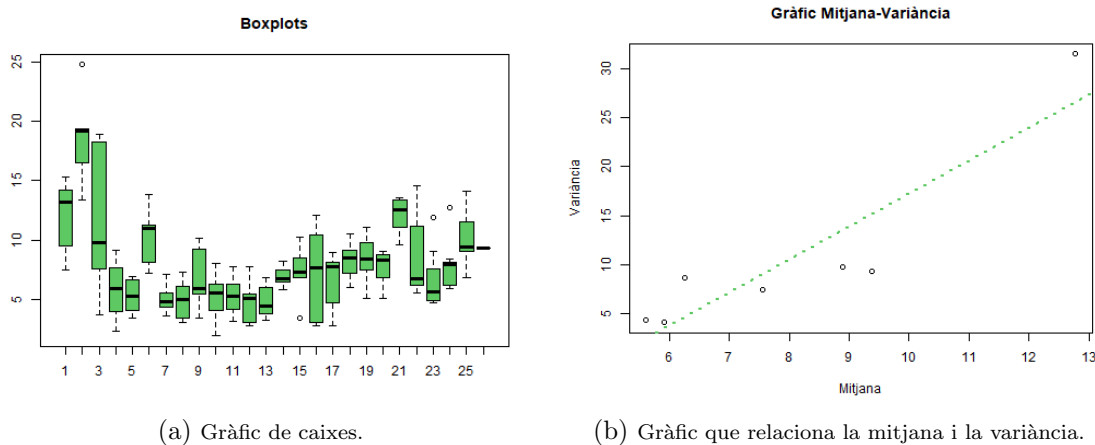


Figura 3.5: Gràfics per estudiar la constància de la variància

Si amb aquests gràfics es conclou que la variància és constant, es continuarà treballant amb la sèrie sense cap canvi, però si aquesta presenta patrons de no constància, s'ha d'aplicar una transformació logarítmica a les dades.

En finalitzar l'estudi de la variància l'altre component que cal verificar és l'existència d'un patró estacional dins de la sèrie temporal.

Per identificar si aquesta presenta estacionalitat, s'analitza el gràfic que presen-

ta l'estacionalitat per cada estació, és a dir, per l'ordre m amb què s'està treballant. Aquest ha de presentar les mitjanes en el mateix nivell en els diferents ordres. Per exemple, si tenim dades on l'estacionalitat és setmanal, el gràfic hauria de presentar l'estacionalitat de cada dia de la setmana i les seves mitjanes haurien d'estar alineades com es mostra a la figura 3.6.

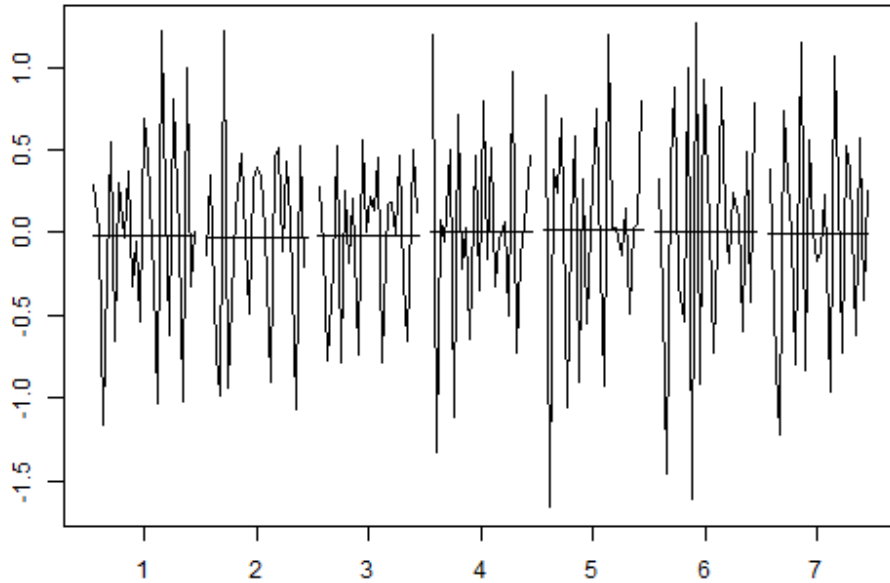


Figura 3.6: Gràfic per estudiar l'estacionalitat d'ordre m

Si s'observa que hi ha estacionalitat, és a dir, que les mitjanes dels diferents ordres són diferents caldrà aplicar una diferenciació estacional a la sèrie amb la mitjana constant. Si anomenem la sèrie temporal amb la qual s'està treballant amb mitjana constant com Z_t per aplicar la diferenciació estacional s'ha de fer:

$$W_t = Z_t - Z_{t-s}, \quad t = 2, 3, \dots, n$$

Per acabar l'anàlisi prèvia al model, cal comprovar que la mitjana de la sèrie és zero. Més ben dit, la sèrie ha de fluctuar presentant una mitjana de zero al llarg del temps.

Si es revela que la mitjana de la sèrie no és zero, s'haurà d'aplicar diferenciacions regulars a la sèrie. Les diferenciacions regulars s'aplicaran de manera reiterada a la sèrie evitant cometre sobre-diferenciació. Per comprovar que no sobre-diferenciem la sèrie s'avaluarà la variància d'aquesta. Si la sèrie amb una diferenciació més resulta amb una variància major, es treballarà amb la sèrie amb una diferenciació regular

menys. Matemàticament, les diferenciacions regulars es calculen de la següent forma:

$$W_t = Z_t - Z_{t-1}, \quad t = 2, 3, \dots, n$$

Si s'aplica una segona diferenciació regular, matemàticament es defineix com:

$$W_t = (Z_t - Z_{t-1}) - (Z_{t-1} - Z_{t-2}), \quad t = 3, \dots, n$$

I d'aquesta manera, s'aplicaran tantes com siguin necessàries.

Aplicades les transformacions que feien falta a la sèrie temporal s'avaluarà la funció ACF. Si aquesta decau de manera ràpida al 0, és una sèrie temporal estacionària tal com s'ha presentat anteriorment.

En tenir preparada la sèrie, es procedirà a la identificació d'un model SARIMA (*Seasonal Autorregresive Integrated Moving Average*). L'expressió general d'aquests models és:

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D X_t = \theta_q(B)\Theta_Q(B^s)Z_t$$

On $Z_t \sim WN$, és a dir, es distribueix com un soroll blanc.

La primera part representa la part regular del model que segueix una estructura $ARIMA(p, d, q)$ i la segona part fa referència a la part estacional del model que segueix una estructura $ARIMA(P, D, Q)$. Els paràmetres d i D defineixen l'ordre de diferenciacions regulars i estacionals respectivament. Respecte als paràmetres p i q són l'ordre d'autoregressió i mitjana mòbil regular que s'obtenen amb l'ACF i el PACF observant els primers *lags* (retards) significatius. La P i la Q són els paràmetres per la part estacional i s'identifiquen veient els *lags* múltiples a s (on s representa l'estacionalitat).

En cas de dubte o identificació d'un possible retard significatiu es recomana tenir-lo en compte en aquesta etapa perquè en l'etapa d'estimació es verificarà si realment aquest és, o no, significatiu per la creació del model.

3.4.2 Estimació

Identificat el possible model que correspon a la sèrie temporal i els paràmetres, es durà a terme l'estimació dels coeficients mitjançant mètodes no lineals per trobar la funció de màxima versemblança (Gardner et al., 1980).

En estimar els paràmetres del model caldrà verificar si aquests són significatius perquè així es pot simplificar el model final. Per avaluar si són significatius es

planteja el següent contrast d'hipòtesi:

$$H_0 : \phi_i = 0 \quad H_1 : \phi_i \neq 0$$

Utilitzant l'estadístic $\hat{\phi}_i \approx N(\phi_i, \sigma_{\phi_i})$, $\hat{t} = \frac{\hat{\phi}_i}{se(\hat{\phi}_i)} \approx t - Student_{T-k}$ on $se(\hat{\phi}_i)$ és l'estimació de σ_{ϕ_i} basada en la matriu Hessiana i k és el nombre total de paràmetres del model es conclourà que els coeficients estimats són significatius comprovant que l'estadístic (*T-ratio*) calculat és major a dos en valor absolut.

La sintaxi utilitzada per estimar els coeficients tenint en compte *l'intercept* (3.8):

```
1 arima(serie, order = c(p, 0, q), seasonal = list(order = c(P, 0, Q)
  , period = s))
```

Listing 3.8: Sintaxi aplicada en R per estimar els coeficients d'un model SARIMA tenint en compte l'intercept

I també s'ha implementat en aquest projecte l'expressió 3.9 per estimar els coeficients sense tenir en compte *l'intercept*.

```
1 arima(serie, order = c(p, d, q), seasonal = list(order = c(P, D, Q)
  , period = s))
```

Listing 3.9: Sintaxi aplicada en R per estimar els coeficients d'un model SARIMA tenint en compte l'intercept

En aquesta anàlisi, tot i identificar significatiu *l'intercept* es considerarà no significatiu per avançar amb el procés plantejat en la metodologia Box-Jenkins. Això és degut al fet que si es considera significatiu, la formulació del model i les prediccions es compliquen considerablement.

En cas que l'estimació del coeficient més gran sigui no significatiu s'aplicarà el principi de parsimònia. Això vol dir que, posant pel cas que en l'etapa d'identificació s'estava dubtant entre un model *AR*(2) i un *AR*(3) per la part regular, si en l'estimació l'últim coeficient del model *AR*(3) és no significatiu, gràcies al principi de parsimònia on es busca el model més senzill que expliqui millor, se seleccionaria el model *AR*(2).

3.4.3 Validació

Si el model de Box-Jenkins és un bon model per les dades, s'hauran de satisfer certs supòsits dels residus del model.

Els residus han de presentar-se tant com un soroll blanc o independents seguint una distribució normal amb mitjana i variància constant.

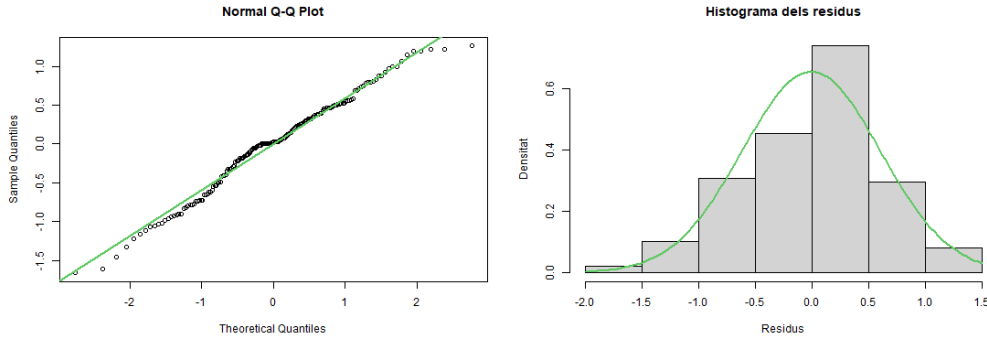


Figura 3.7: Gràfics per verificar la normalitat dels residus.

En aquest estudi s'apliquen per comprovar la normalitat de manera gràfica el *qqnorm* i un histograma amb la corba de la normal com es mostra a la imatge 3.7.

D'altra banda, es pot comprovar si els residus es distribueixen segons una distribució normal mitjançant contrastos d'hipòtesi numèrics plantejant com a hipòtesis:

H_0 : La mostra prové d'una població normalment distribuïda

H_1 : La mostra no prové d'una població normalment distribuïda

Els tests que s'ofereixen dins la plataforma per dur l'anàlisi són:

- **Shapiro-Wilk:** és considerat un dels tests més potents per comprovar la normalitat i l'estadístic que el defineix és $W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$ on $x_{(i)}$ fa referència el número que ocupa la posició i (ordenant la mostra de menor a major), \bar{x} és la mitjana mostral i $a_i = \frac{m^T V^{-1}}{(m^T V^{-1} V^{-1} m)^{1/2}}$ on m representa els valors mitjans de l'estadístic ordenat (mirar Royston, 1982).

Per aplicar-ho s'ha implementat la funció definida en el listing 3.10.

```
1 shapiro.test(model()$residuals)
```

Listing 3.10: Sintaxi aplicada en R pel contrast Shapiro-Wilk

- **Anderson-Darling:** és una prova no paramètrica que contrasta si les dades d'una mostra provenen d'una determinada distribució probabilística. La fórmula de l'estadístic del test determina quina distribució probabilística s'està estudiant. En aquest cas, l'estadístic es defineix com $A = -n - \frac{1}{n} \sum_{i=1}^n [2i - 1][\ln(p_{(i)}) + \ln(1 - p_{(n-i+1)})]$ on $p_{(i)} = \Phi([x_{(i)} - \bar{x}]/s)$ (mirar Thode, 2002).

La sintaxi utilitzada per aquest test en el llenguatge R és la que es mostra a continuació (3.11).

```
1 library(nortest)
2 ad.test(resid(model))
```

Listing 3.11: Sintaxi aplicada en R pel contrast Anderson-Darling

- **Jarque-Bera:** Aquesta prova determina si les dades segueixen la distribució normal utilitzant els coeficients d'asimetria i curtosis (llegir Cromwell et al., 1994). Per tant, l'estadístic es calcula de la següent manera:

$$JB = \frac{n}{6} \left(S^2 + \frac{1}{4} (K - 3)^2 \right)$$

on S i K es defineixen com:

$$S = \frac{\hat{\mu}_3}{\hat{\sigma}^3} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{3/2}}$$

$$K = \frac{\hat{\mu}_4}{\hat{\sigma}^4} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left(\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^2}$$

Utilitzant la següent funció en R, s'obté tant l'estadístic del test com el p-valor per acceptar o rebutjar el contrast (3.12).

```
1 library(tseries)
2 jarque.bera.test(resid(model))
```

Listing 3.12: Sintaxi aplicada en R pel contrast Jarque-Bera

Per contrastar l'homoscedasticitat es grafica els residus i l'arrel dels residus absoluts. En els gràfics s'ha de veure un patró aleatori al llarg del temps i en el mateix nivell. A més, per recolzar el que s'ha observat en el gràfic també es planteja el test de Breusch-Pagan. Aquest contrast determina si el terme de pertorbació del model plantejat és homoscedàstic o presenta heteroscedasticitat (mirar Breusch et al., 1979). Per tant, el contrast es defineix com:

H_0 : El model és homoscedàstic

H_1 : El model és heteroscedàstic

Per calcular-ho, s'ha utilitzat la següent sintaxi dins de la plataforma (3.13).

```
1 library(lmtest)
2 bptest(resid(model) ~ I(serie - resid(model)))
```

Listing 3.13: Sintaxi aplicada en R pel contrast de Breusch-Pagan

Per la validació de la independència dels residus s'estudien els gràfics de l'ACF i PACF. Dins d'aquests cal confirmar com els diferents retards es distribueixen seguint el patró del soroll blanc que s'ha explicat en la secció anterior. D'aquesta manera es podrà assegurar al 95% de probabilitat que les dades no estan autocorrelacionades.

Seguint l'esquema anterior, per comprovar aquesta propietat de manera numèrica es proposa aplicar el test de Durbin-Watson. Aquest test detecta la presència d'au-

tocorrelació, és a dir, si hi ha relació entre els valors separats per un interval de temps (llegir Durbin et al., 1950). Per tant, el contrast d'hipòtesi es pot definir com:

H_0 : Les dades són independents

H_1 : Les dades no són independents

L'estadístic que s'utilitza per al contrast es defineix com $d = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$

S'obté els resultats d'aquest contrast de manera automàtica amb la següent funció (3.14).

```
1 library(lmtest)
2 dwtest(resid(model)~I(1:length(resid(model))))
```

Listing 3.14: Sintaxi aplicada en R pel contrast de Durbin-Watson

Per acabar d'assegurar que els residus són independents entre si, s'ofereix dins de la interfície web els resultats visualment i en una taula del test de Ljung-Box (mirar Box et al., [s. d.]). Amb aquest test es pot identificar a partir de quin retard les dades comencen, o no, a estar auto-correlacionades entre elles.

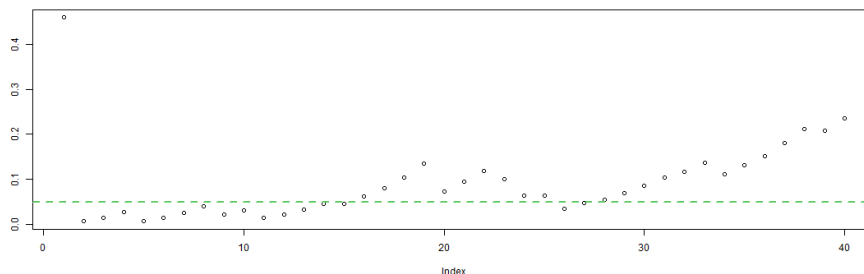


Figura 3.8: Gràfic amb els p-valors del test Ljung-Box.

Amb l'exemple que es mostra a la figura 3.8 es pot dir que a partir del segon retard les dades estan autocorrelacionades ja que el p-valor és menor al nivell de significació $\alpha = 0.05$.

Finalment, és una bona praxi verificar si el model proposat presenta volatilitat mitjançant els gràfics de l'ACF i el PACF al quadrat. Per considerar que aquests no presenten volatilitat s'ha de manifestar una distribució igual al soroll blanc.

A més, per considerar que el model plantejat per la sèrie temporal que s'està estudiant és estacionària i invertible s'ha d'examinar que les arrels del polinomi característic es trobin fora del cercle unitari (llegir Quinn, 1982). En altres paraules, s'ha donat una opció dins de la interfície web per visualitzar de manera gràfica si les

arrels inverses del polinomi característic es troben dins de les circumferències com a la imatge 3.9 per sentenciar que el model és estacionari i invertible.

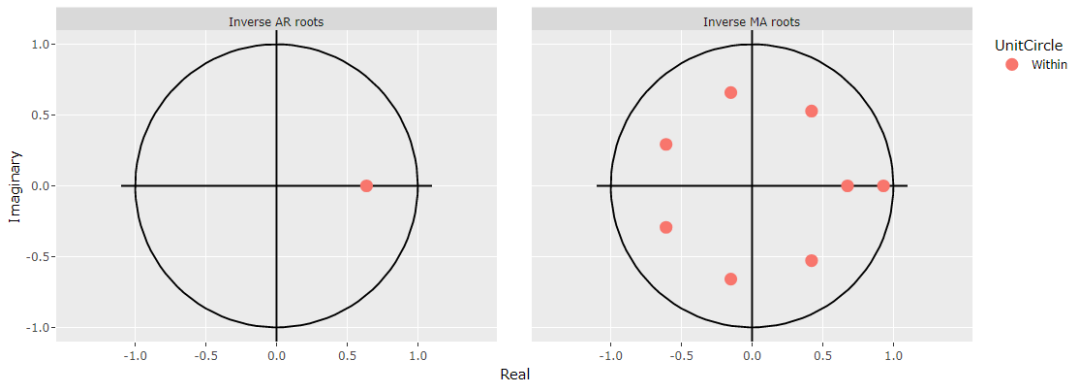


Figura 3.9: Gràfic amb les circumferències per verificar la invertibilitat i estacionarietat del model.

3.4.4 Model adequat

En aquesta última etapa, si el model que s'ha plantejat a l'inici és validat, es podrà fer les prediccions necessàries.

En cas que s'hagin plantejat més d'un model i tots ells han estat validats, s'escollirà el model òptim d'acord amb el criteri de l'AIC (criteri d'informació d'Akaike).

$$AIC = -2\log L(\phi, \theta, \sigma^2 | X) + 2p$$

On p és el nombre de paràmetres i $L(\phi, \theta, \sigma^2 | X)$ és el valor de màxima versemblança. Així doncs, el model òptim serà el que presenti un menor AIC.

3.4.5 Identificació i predicció automàtica

Com s'ha fet amb la metodologia d'allisat exponencial, seguint la metodologia Box-Jenkins també hi ha un algoritme que construeix de manera automàtica un model òptim segons el criteri AIC (3.15).

```
1 library(forecast)
2 auto.arima(serie)
```

Listing 3.15: Sintaxi aplicada en R pel modelatge automàtic (auto.arima)

Amb aquesta funció s'aplica l'algoritme per passos definit en l'article Hyndman; Khandakar, 2008 on es busca el model òptim segons les dades que es proporcionen. De manera automàtica s'estudia si són necessàries les transformacions descrites anteriorment i com es presenten les components.

Tot i seguir la mateixa estructura descrita en aquest capítol pot ser que no s'obtingui el mateix model de manera automàtica a causa que és bastant restrictiu. Si l'algoritme detecta en algun moment que la significança d'algun dels contrastos es troba en el límit, ja no proposa el model i continua buscant un model millor.

3.5 Prophet

L'any 2017, uns investigadors de Facebook van publicar un article (Taylor et al., 2018) que va presentar la llibreria oberta Facebook *Prophet* que ofereix la modelització de sèries temporals de manera ràpida, potent i accessible per tothom.

L'algoritme que recull el paquet *Prophet* és particularment bo per modelitzar sèries temporals que presenten diverses estacionalitats. El model que crea es formula com:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

On $g(t)$ és la funció que caracteritza la tendència, $s(t)$ qualifica l'estacionalitat, $h(t)$ representa les vacances i dies festius del país on s'han recollit les dades i ϵ_t engloba l'error.

La funció que estudia la tendència i determina els punts de canvi de tendència es pot definir com una tendència lineal ($y = mx + b$) o com una tendència logística ($g(t) = \frac{C(t)}{1+x^{-k(t-m)}}$).

La funció que recull l'estacionalitat es defineix com una sèrie de Fourier en funció del temps ($s(t) = \sum_{n=1}^N (a_n \cos(\frac{2\pi nt}{P}) + b_n \sin(\frac{2\pi nt}{P}))$).

Es mostra més detallada la metodologia que segueix l'algoritme *Prophet* en la pàgina web *Quick Start — Prophet*, [s. d.].

En ser una metodologia completament automàtica i relativament nova, dins del projecte s'ha adaptat per visualitzar els exemples proporcionats en la documentació del paquet *Forecasting PM2.5 with forecast and prophet — R-bloggers*, [s. d.]. D'aquesta manera s'ha fet ús de manera adequada les diferents funcions que conté aquesta llibreria i les diferents interpretacions.

3.6 Mètriques de regressió

Explicades les diferents metodologies aplicades per modelitzar les diferents sèries temporals dels contaminants de Califòrnia es decidirà quin mètode és millor segons les mètriques de regressió.

Es calcularan les següents mètriques tenint en compte que y_i fa referència a les observacions i \hat{y}_i és la predicció:

- **Mitjana de l'error absolut (MAE):** és una magnitud que determina la mitjana dels errors en fer prediccions sense tenir en compte el signe.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- **L'error quadràtic mig (RMSE):** aquest és un indicador molt freqüent on s'observa la diferència dels valors pronosticats pel model i els valors que realment s'han registrat. Aquest es pot interpretar com una mesura de capacitat de predicció.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- **Mitjana de l'error absolut en percentatge (MAPE):** aquest és un indicador de precisió percentual de l'error. En mesurar-se en percentatges facilita la interpretació, per exemple, si en calcular aquest indicador ens resulta que ens dona 10 això vol dir que la predicció té un 10% d'error.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right|$$

- **L'error quadràtic mig percentual (RMSPE):** aquesta és una mètrica que presenta les mateixes propietats que l'RMSE però és en valors percentuals. Per tant, facilitarà la interpretació dels resultats.

$$RMSPE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)^2}$$

- **Mitjana de l'error absolut normalitzat (nMAE):** amb aquesta mètrica s'aconsegueix comparar dades amb diferents escales segons el MAE.

$$nMAE = \frac{MAE}{\bar{y}} = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|}{\frac{1}{n} \sum_{i=1}^n |y_i|}$$

- **L'error quadràtic mig percentual normalitzat (nRMSE):** amb aquesta mètrica s'obté una comparació de dades amb diferents escales segons el RMSE.

$$nRMSE = \frac{RMSE}{\bar{y}} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sum_{i=1}^n |y_i|}$$

Per determinar que un model és millor a un altre segons les mètriques de regressió, s'ha d'escollir el model que presenti un menor resultat per la mètrica estudiada. En aquest treball es donarà especial importància a la mètrica RMSPE, ja que és una mètrica bastant utilitzada per determinar si un model prediu correctament.

3.7 Recursos informàtics

Per tancar el capítol de la metodologia cal anomenar els programes informàtics que s'han aplicat per dur a terme aquest petit projecte.

Primerament, el llenguatge informàtic que s'ha utilitzat per dur a terme aquest estudi ha estat RStudio. Aquest és un bon programa per dur a terme anàlisis estadístiques perquè disposa d'una alta varietat de funcions i paquets amb els quals treballar.

En segon lloc, per crear l'aplicació web s'ha emparat la llibreria *shiny* (llegir *Shiny*, [s. d.]), *shinydashboard* (mirar *Shiny Dashboard*, [s. d.]) i *shinyWidgets* (consultar *Shiny - Widget Gallery*, [s. d.]). Aquestes llibreries permeten crear la interfície on s'han desenvolupat les diferents metodologies per modelitzar sèries temporals de manera interactiva. Per poder fer-les servir, s'ha hagut de fer una recerca prèvia del funcionament d'aquestes llibreries mitjançant cursos al *LinkedIn Academy* (consultar *LinkedIn Learning: Online Courses for Creative, Technology, Business Skills*, [s. d.]), la documentació que proporciona el mateix programa i diferents fòrums de debat (per exemple, *Stack Overflow en español*, [s. d.]).

Per poder actuar amb diverses bases de dades alhora i processar-les de manera senzilla s'ha fet ús de la llibreria *tidyverse* (H. Wickham; M. H. Wickham, 2017). Aquesta llibreria inclou els paquets necessaris per fer anàlisis de dades, des dels paquets principals per importar-les fins als paquets per dissenyar els gràfics o taules idonis per l'estudi. Per aprendre a utilitzar-la i descobrir les diferents funcionalitats que proporciona es va llegir un llibre dedicat a aquest paquet que ofereix el RStudio (H. Wickham; Golemund, 2016).

La llibreria *plotly* (consultar Sievert, 2020) i *imputeTS* (llegir Moritz et al., 2017) han estat necessaris per a la visualització interactiva de les diferents sèries temporals i poder resoldre els problemes de les dades mancants respectivament. Pel

que fa a les prediccions s'han aplicat diferents funcions de la llibreria *forecast*. Durant el procés de validació dels models fonamentalment s'han necessitat les llibreries *lmtest* (Hothorn et al., 2015), *tseries* (Trapletti et al., 2022) i *nortest* (Gross et al., 2015).

Per acabar aquest capítol, convé recalcar que tota la nomenclatura presentada en aquest apartat apareixerà tant en el capítol de la guia d'ús de la interfície web com en l'exemple pràctic del procés de modelització amb un dels contaminants.

Capítol 4

Guia d'ús de la interfície

En aquest apartat s'han agrupat les diferents instruccions i opcions que s'integren dins de la interfície interactiva web que s'ha creat per agilitzar la modelització de les sèries temporals aplicant diverses metodologies.

Aquesta web ha estat dissenyada amb el llenguatge R, concretament, s'ha utilitzat l'aplicació RStudio juntament amb les seves extensions *shiny*, per fer una pàgina web interactiva i *shinydashboard* per aconseguir un disseny fàcil per la lectura dels gràfics, taules i les diverses funcions que s'ofereixen.

4.1 Guia d'ús

En entrar per primer cop en aquesta interfície web, es podrà observar la primera pantalla tal com es mostra a la figura 4.1. Aquesta plana mostra com a l'esquerra, sota el títol del *dashboard*, hi ha un menú que correspon a les diferents etapes o metodologies que es poden aplicar dins de la web. Per defecte ens apareix seleccionada la “Guia d'ús” perquè es presenta una breu introducció a l'aplicació en la primera pestanya i una opció per pujar dades.

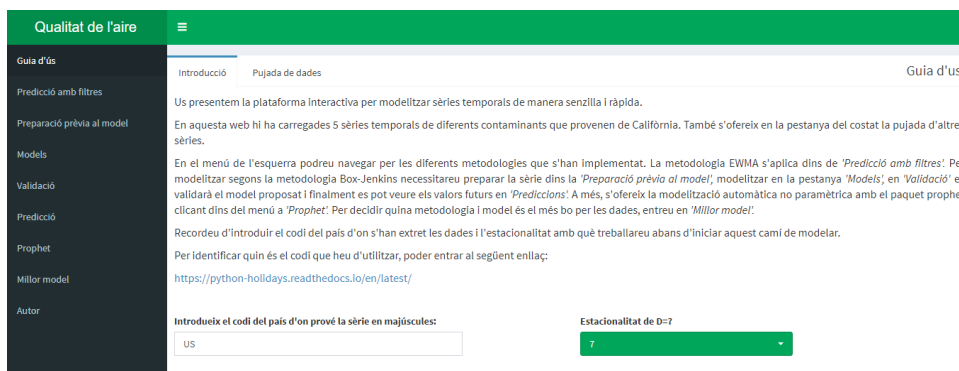


Figura 4.1: Primera pàgina de l'aplicació web

Per iniciar el modelatge, cal introduir el codi del país d'on s'han extret les dades (això és un requisit per aplicar l'algoritme *prophet*) i l'estacionalitat amb la qual es treballarà. Si es vol treballar amb les dades que ja hi ha carregades a la web, no serà necessari canviar aquestes opcions perquè en aquest projecte s'ha treballat amb dades diàries de Califòrnia on s'estudia l'estacionalitat setmanal. D'altra banda, si es vol utilitzar aquesta interfície per modelitzar unes altres dades, caldrà pujar-les amb l'ajuda que s'ofereix en la pestanya contigua.

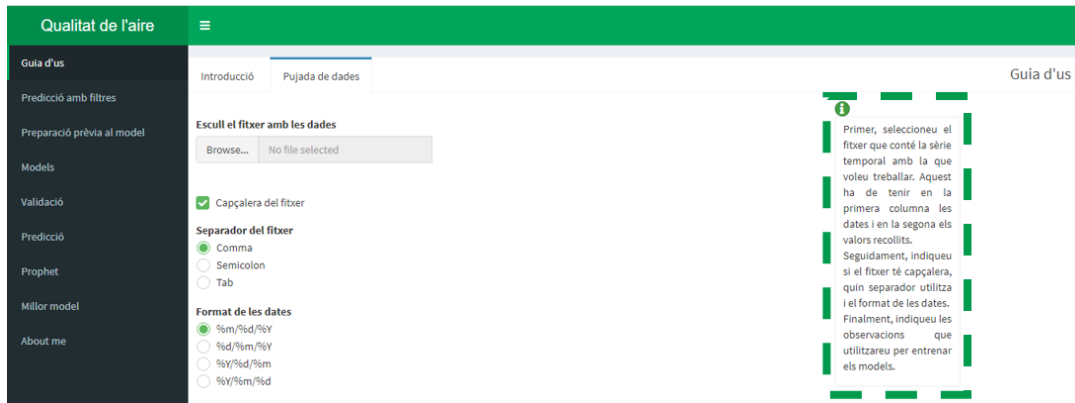


Figura 4.2: Pantalla per pujar les dades a l'aplicació

Primerament, s'ha de seleccionar el fitxer on s'ha guardat la sèrie temporal mitjançant el *Browse* (mirar 4.2). Aquest fitxer ha de tenir en la primera columna les dates i en la segona els valors recollits.

Seguidament, apareix al costat dret com s'han llegit les dades dins de la web. Si es veu la figura 4.3, ressaltat amb una línia discontinua verda, s'observen les primeres sis observacions del fitxer que s'han pujat. Com es pot veure a la imatge, en aquest cas no es detecta que hi ha més d'una columna, és a dir, no se separen les dates i els valors recollits.

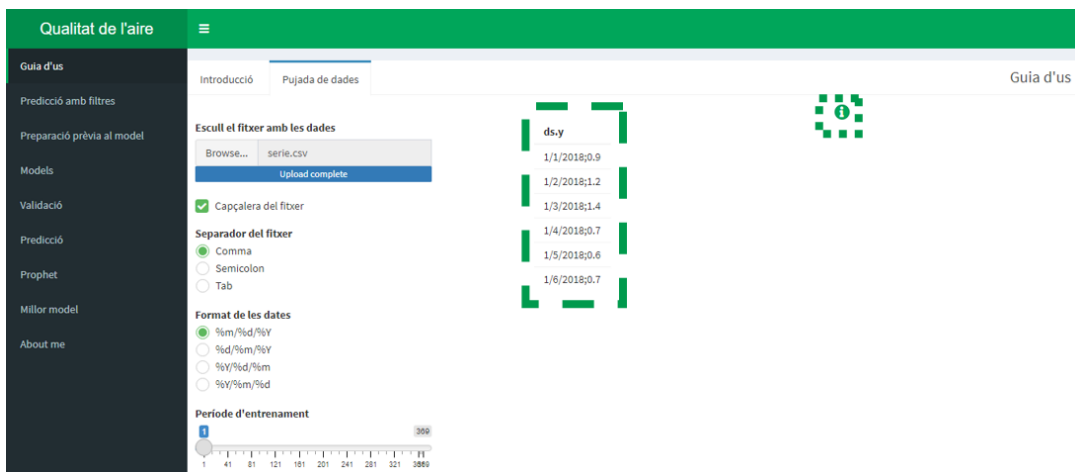


Figura 4.3: Càrrega d'una sèrie temporal a la web

Per fer una lectura correcta de les dades, hi ha un seguit d'opcions sota el *Browse* per corregir-ho. En primer lloc, cal indicar si la primera línia correspon a la capçalera del fitxer. En segon lloc, cal seleccionar quin és el separador que s'està utilitzant per separar les diferents columnes del fitxer. En l'exemple que es mostra a la il·lustració 4.3, és un punt i coma. En tercer lloc, tot i que no es vegi evident en la importació de les dades, cal que s'indiqui quin és el format de les dates. Per tant, per poder escollir, s'ha de tenir en compte que *d* fa referència al dia, *m* fa referència al mes i *Y* fa referència a l'any. Per acabar amb aquest procés, saben que les diferents metodologies de modelatge de sèries temporals necessiten un joc de dades per entrenar el model i després fer ús de la sèrie temporal completa per fer les prediccions de manera precisa, s'ha de seleccionar fins quin observació s'inclourà en el període d'entrenament.

Tenint en compte tots aquests procediments per la càrrega de les dades, ja es pot procedir a la modelització. Per ajudar als analistes a entendre cada una de les planes, s'ofereix una breu ajuda clicant al símbol *i* com es mostra a la figura 4.3 i apareix una explicació igual que es pot veure a la 4.2.

En els pròxims apartats s'explicà cada una de les opcions del menú que ofereix la web.

4.2 Predicció amb filtres

En aquesta secció, la qual també correspon amb el nom de l'opció que ofereix el menú de la web, s'engloba tot el procés per fer prediccions segons la metodologia de l'allisat exponencial.

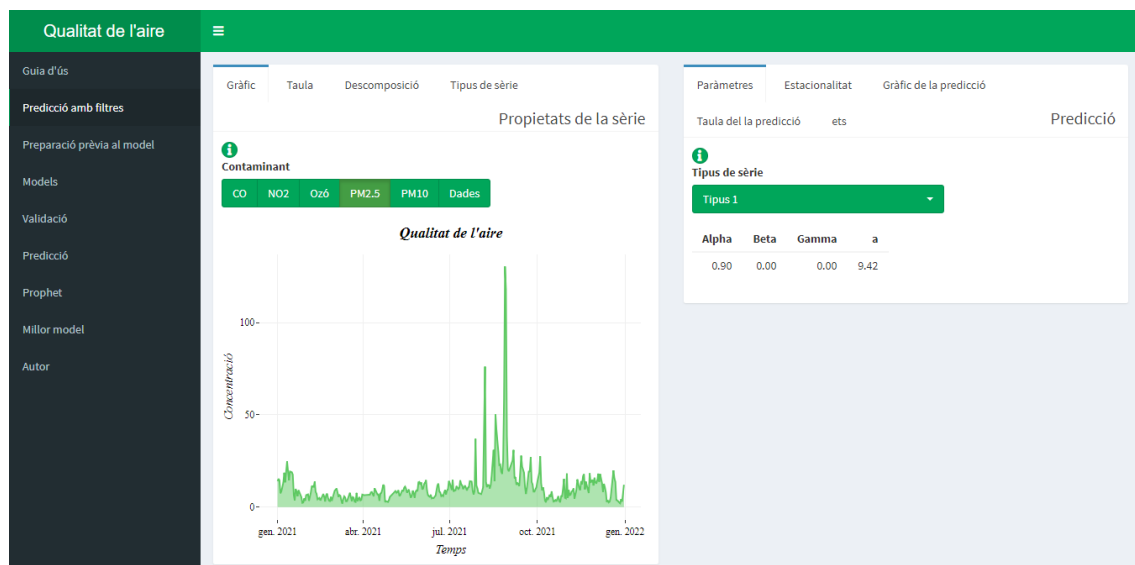


Figura 4.4: Panell per fer prediccions segons la metodologia d'allisat exponencial

Observant la figura 4.4, s'identifiquen dues planes, a l'esquerra es recullen les propietats de la sèrie i a la dreta tot el que és necessari per fer la predicció.

En el bloc de l'esquerra, dins de la pestanya de gràfic, es mostren les diferents opcions dels contaminants que s'han utilitzat en aquest projecte i també hi ha un botó anomenat *Dades* que fa referència a les dades que ha pujat a la web prèviament. Sota les diferents opcions per treballar amb les dades es presenta un gràfic interactiu temporal de les dades completes que s'han escollit. En aquest gràfic es pot fer zoom per seleccionar una part específica de la gràfica i guardar-ho com a imatge per la creació d'informes.

Avançant dins d'aquest bloc, es pot veure dins de la pestanya “Taula” (4.5) es troben els valors i dates de la sèrie temporal que s'ha escollit per analitzar. Es pot seleccionar el nombre d'observacions a mostrar a la taula, navegar per les diferents entrades que hi ha al final de la taula i buscar observacions concretes.

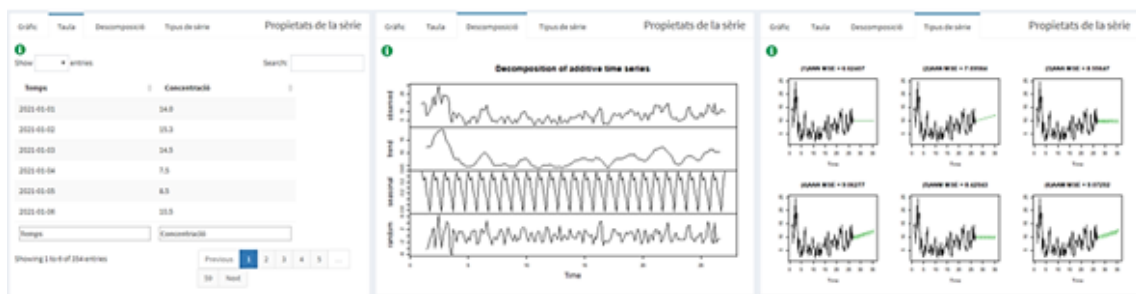


Figura 4.5: Visualització de les pestanyes Taula, Descomposició i Tipus de sèrie

Dins de la pestanya “Descomposició” es presenta en un gràfic totes les components presents en una sèrie temporal. Com ja s’ha explicat en la metodologia, per determinar quin tipus de sèrie s’està analitzant cal estudiar prèviament les components que formen la sèrie temporal. La primera imatge que mostra el gràfic són les observacions de la sèrie tenint en compte l’estacionalitat, a sota es mostra la component tendència al llarg del temps. Després, es presenta el patró de l’estacionalitat d’aquesta sèrie i al final de tot del gràfic es veu com es distribueix l’error, és a dir, les oscil·lacions aleatòries.

Per acabar amb el bloc de les propietats de la sèrie, s’ha implementat la pestanya “Tipus de sèrie”. En aquesta es presenten tots els tipus de sèries temporals i ajuda a decidir quin tipus de sèrie correspon amb les dades que s’està treballant. Per poder avançar i fer prediccions d’acord amb la metodologia d’allisat exponencial, es pot decidir el tipus de sèrie segons l’anàlisi descriptiu de les components o també pot ser de gran ajuda calcular l’indicador MSE (*Mean square error*) que fa referència a la mitjana de l’error al quadrat.

Estudiant de nou la figura 4.4, en el bloc de la dreta es recull tots els resultats després de decidir el tipus de sèrie per poder fer prediccions segons la metodologia d'allisat exponencial. Primerament, és necessari indicar el tipus de sèrie amb el desplegable que apareix en la pestanya de “Paràmetres”. D'aquesta manera s'actualitzaran els paràmetres estimats a la taula (4.6).

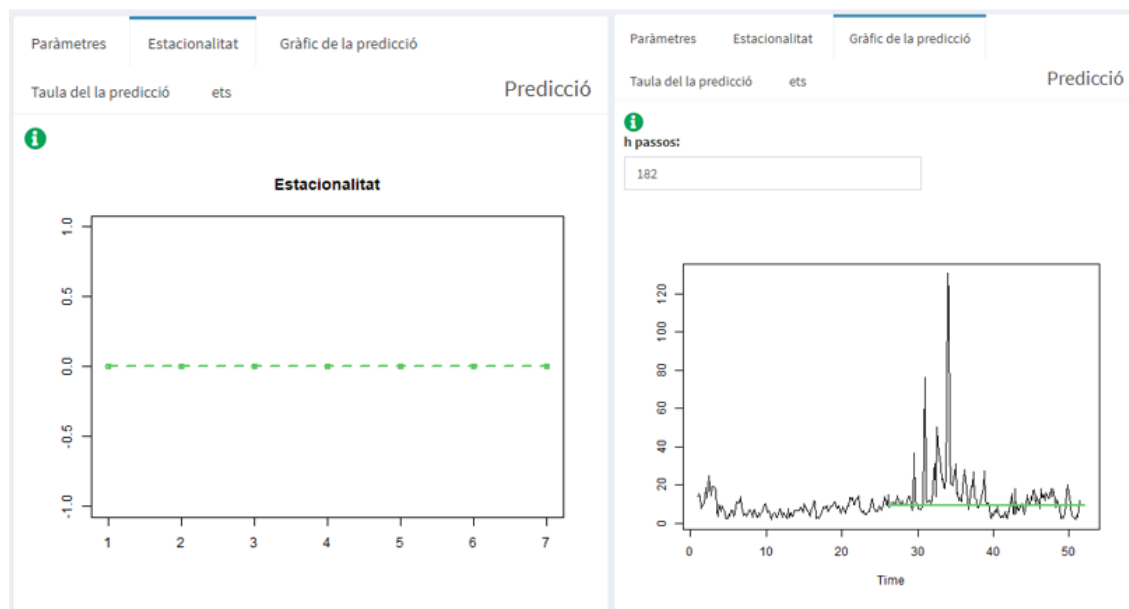


Figura 4.6: Pestanyes Estacionalitat i Gràfic de la predicció

Per predir segons la metodologia d'allisat exponencial cal entrar a la pestanya “Gràfic de predicció”. En aquesta pestanya inicialment s'ha d'introduir el nombre de dies a predir des de l'última observació utilitzada en la sèrie temporal per entrenar.

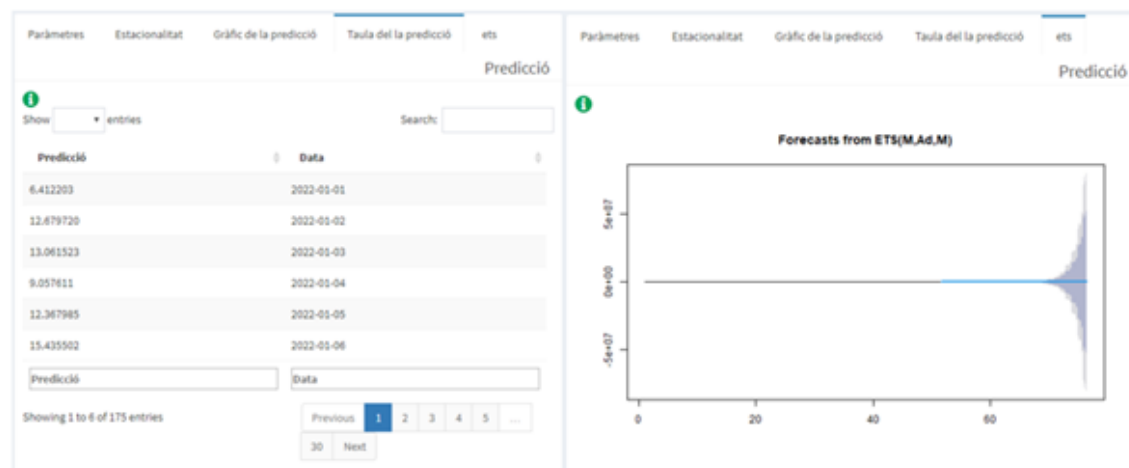


Figura 4.7: Pestanyes Taula de la predicció i ets

Per veure el resultat de la predicció numèricament es presenta l'opció en la pestanya de “Taula de la predicció” (4.7). Aquesta taula presenta les mateixes opcions que la taula vista anteriorment amb els valors de la sèrie.

Finalment, per acabar d'explicar la metodologia i funcionalitats que presenta la interfície per la metodologia d'allisat exponencial, es presenta la pestanya “ets” la qual recull gràficament el tipus de sèrie que ha estimat automàticament el programa i gràfica com es distribuïrien les prediccions al llarg dels dies escollits anteriorment.

4.3 Preparació prèvia al model

Per iniciar la metodologia Box-Jenkins és necessari que la sèrie temporal amb què s'està treballant sigui estacionària. Per aquest motiu, dins d'aquest apartat de la interfície es presenten diversos gràfics i opcions per decidir els canvis a aplicar a les dades i així, fer que sigui estacionària per poder modelitzar.

A simple vista sembla que el bloc de l'esquerra anomenat gràfics previs i el bloc de la dreta anomenat gràfics transformats es presenten quasi els mateixos gràfics. Encara que realment, el bloc de l'esquerra s'enfoca en l'anàlisi gràfica per comprovar si la sèrie temporal és estacionària i el bloc de la dreta presenta, principalment, els mateixos gràfics, però amb les transformacions que s'han considerat necessàries a la sèrie temporal.

Per iniciar l'estudi, dins del bloc de gràfics previs es mostra el gràfic de caixes perquè és adequat per decidir si la variància és constant al llarg del temps (4.8).

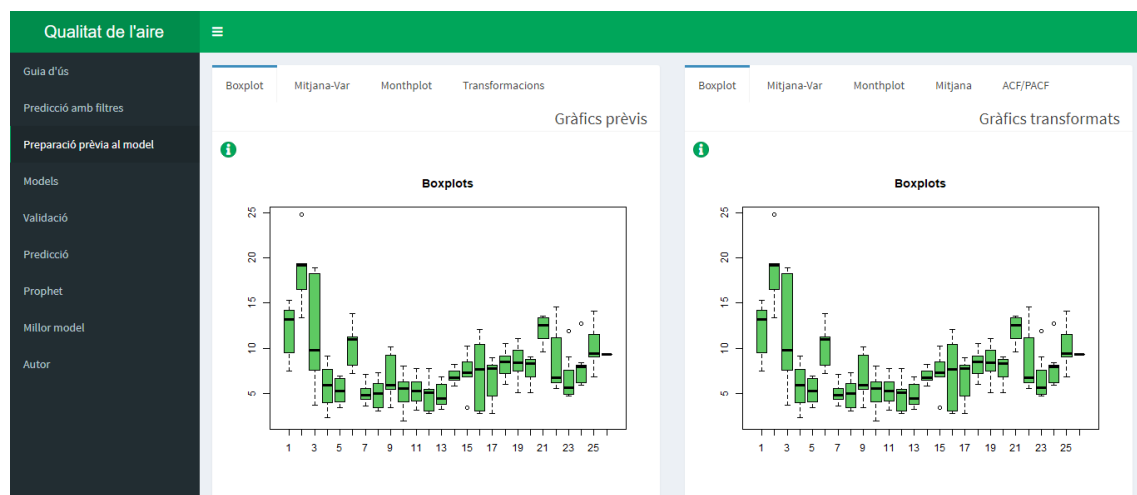


Figura 4.8: Panell per a la preparació prèvia al model

Un altra manera de comprovar que la variància és constant al llarg del temps és mitjançant el gràfic mitjana-variància que hi ha present a la pestanya “Mitjana-Var” (mirar a l'esquerra 4.9).

A la imatge del mig (4.9) es presenta el *monthplot*, gràfic necessari per determinar si la sèrie temporal presenta estacionalitat. Cal analitzar si aquesta component

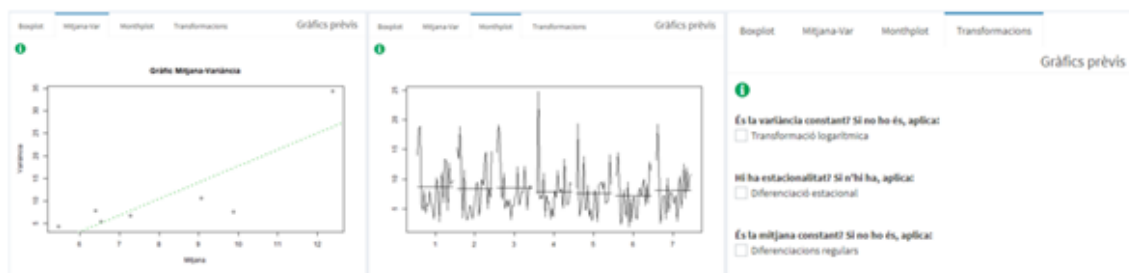


Figura 4.9: Pestanyes Mitjana-Var, Monthplot i Transformacions

existeix en la sèrie per poder seleccionar o no, en la pestanya de “Transformacions”, la transformació corresponent. En la pestanya de transformacions, ja analitzats els gràfics anteriors, s’ofereix la possibilitat d’aplicar transformacions o diferenciacions a la sèrie. Cada cop que s’apliqui alguna de les transformacions o diferenciacions proposades es pot observar com han canviat els gràfics dins del bloc de gràfics transformats. D’aquesta manera es veurà clarament si realment són necessàries les transformacions aplicades.

Una de les propietats que cal verificar abans d’iniciar la modelització és que la mitjana sigui 0. Per aquest motiu, en el bloc de gràfics transformats s’ha implementat la pestanya “Mitjana” on apareix en verd la mitjana de la sèrie temporal amb la qual s’està treballant. Si no s’està segur d’aplicar les diferenciacions regulars a les dades, aquest està programat per aplicar tantes com siguin necessàries per aconseguir que la mitjana sigui 0 sense arribar a presentar sobredispersió.

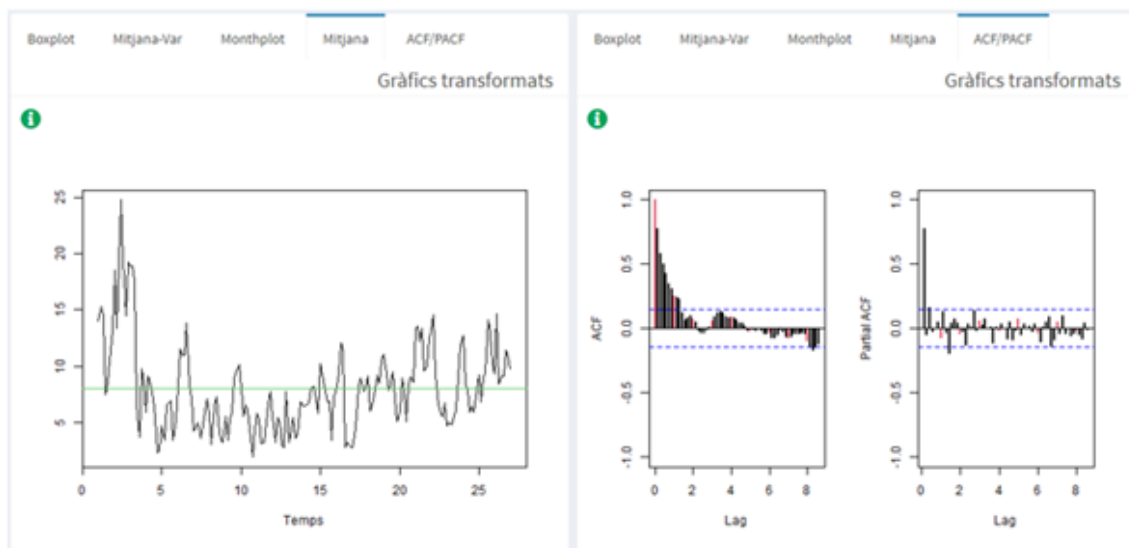


Figura 4.10: Pestanyes Mitjana i ACF/PACF

Analitzades i aplicades les transformacions necessàries a la sèrie amb l’ACF i el PACF (4.10) es pot contrastar si finalment s’ha aconseguir que la sèrie sigui estacionària, tal com requereix en la metodologia Box-Jenkins.

4.4 Models

En aquest apartat es presenten les implementacions de la web per poder iniciar amb el procés de modelització segons els models SARIMA. Amb els gràfics estudiats anteriorment es poden proposar els paràmetres del model en la caixa de l'esquerra del panell i automàticament s'actualitzaran en el bloc dret (4.11) les estimacions dels paràmetres proposats. No només es farà l'estimació tenint en compte l'*intercept* en la pestanya "Estimació A" sinó que també sense tenir-lo en compte en la pestanya "Estimació B".

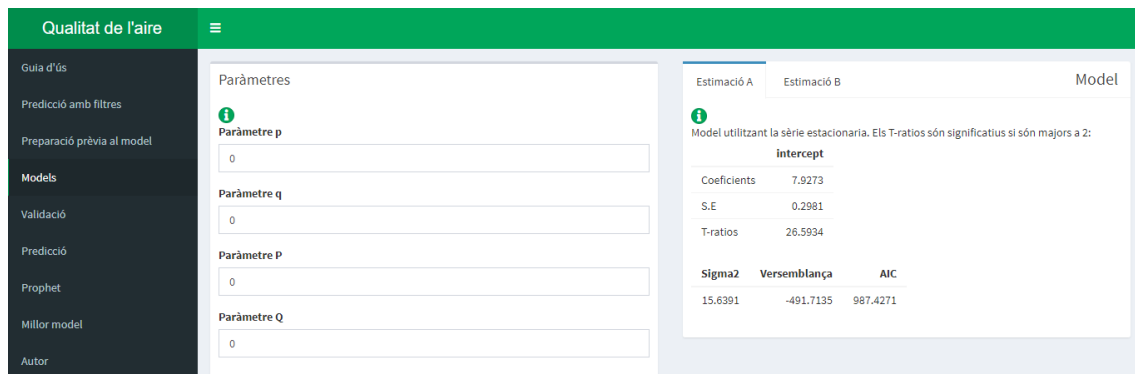


Figura 4.11: Panell Models per modelitzar un model SARIMA

A més dels coeficients estimats, es calcula l'error per cada coeficient i els *T-ratios* per cada un dels paràmetres. Probablement, amb l'anàlisi prèvia a la modelització s'han proposat més d'un model amb diferents paràmetres, per això s'ha implementat en el panell una taula resum on es recull la variància del model, el valor de màxima versemblança que s'ha utilitzat pel càlcul del model i el valor de l'AIC que ajuda a determinar si aquest model serà millor o pitjor als altres proposats.

4.5 Validació

Per aconseguir visualitzar de manera interactiva l'etapa de validació de la metodologia Box-Jenkins, es pot seleccionar en el menú de la interfície l'opció de validació. En aquest panell es recullen tant numèricament com gràficament els tests per comprovar les propietats que han de complir els residus del model. Com s'ha explicat dins del capítol de metodologia, cal comprovar la normalitat dels residus, la constància en la variància, cal que no es presentin autocorrelacions entre els diferents retards, es visualitza si el model proposat presenta volatilitat i també s'estudiarà si el model proposat és invertible i estacionari.

Per iniciar amb la validació, en el panell principal apareix seleccionada la pestanya "Normalitat" per poder comprovar-la amb els gràfics necessaris.

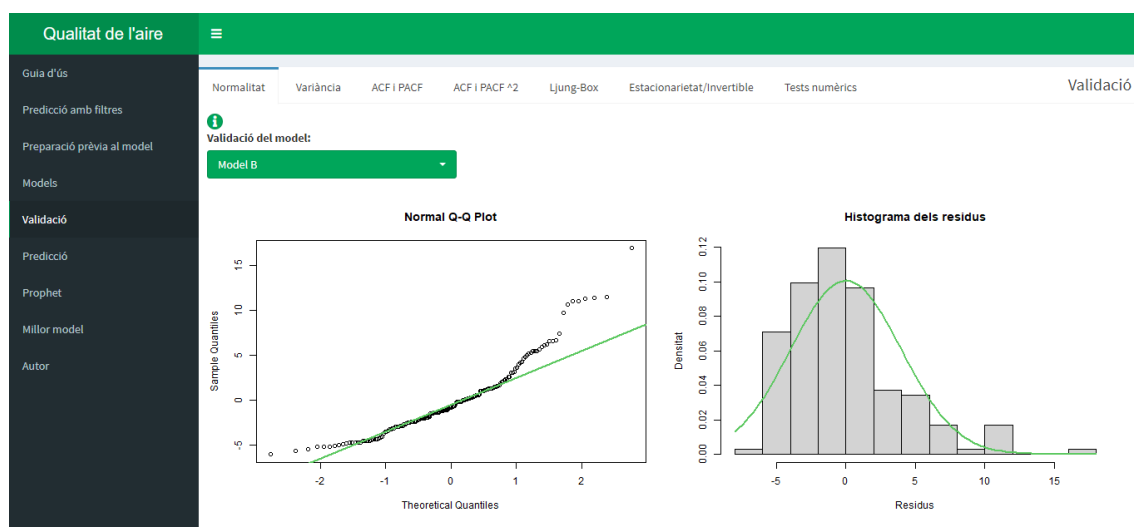


Figura 4.12: Validació de la normalitat

A la primera pantalla (4.12) es visualitza gràficament si es compleix la distribució Normal en els residus i a més, hi ha un desplegable per seleccionar el model amb l'estimació A o B que s'han pogut observar en l'apartat anterior.

Per comprovar que la variància dels residus és constant, dins de la pestanya "Variància" es troben presents tant la gràfica dels residus com la gràfica de l'arrel dels residus absoluts (mirar 4.13).

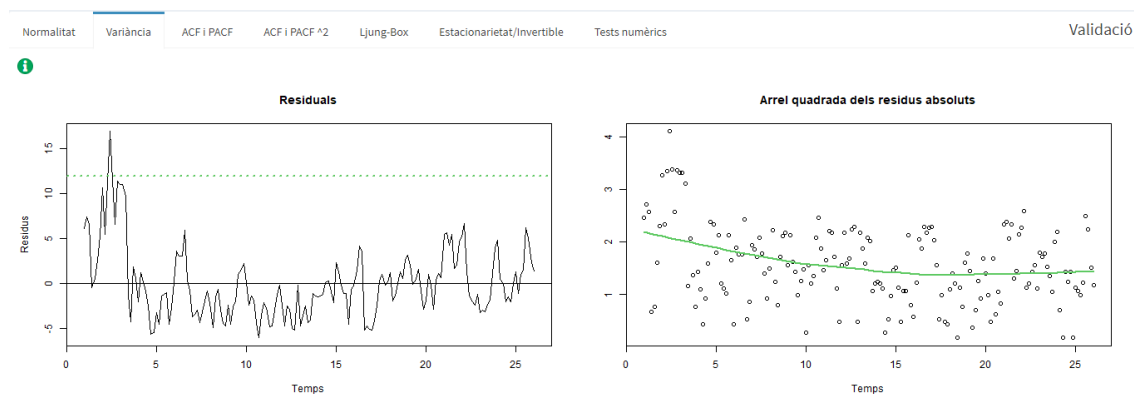


Figura 4.13: Pestanya Variància dins de la validació

Una altra de les propietats que ha de complir el terme de pertorbació és la no presència d'autocorrelació en els diferents retards. Mitjançant aquesta interfície interactiva es pot visualitzar en la pestanya "ACF/PACF" (com en 4.14) si aquests es troben dins de l'interval de confiança al 95%. Amb altres paraules, les diferents barres que apareixen en els gràfics per determinar l'autocorrelació han de presentar soroll blanc com s'ha explicat en l'apartat de la metodologia.

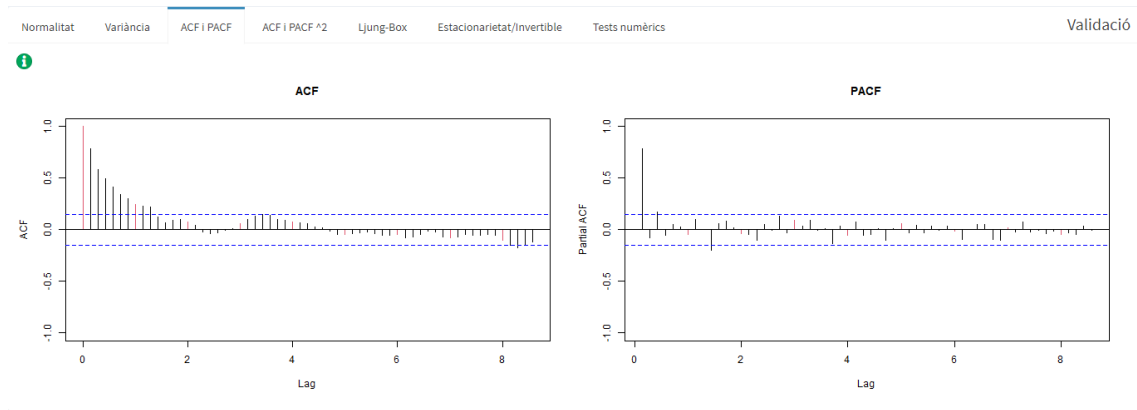


Figura 4.14: Pestanya ACF/PACF dins de la validació

Seguin el procés de validació de la metodologia Box-Jenkins, també s'ha implementat dins d'aquesta interfície la visualització dels gràfics l'ACF i el PACF al quadrat distribuïts com a la figura 4.14. Ambdós gràfics també han de presentar soroll blanc, d'aquesta manera es pot verificar que amb el model plantejat no hi haurà presència de volatilitat.

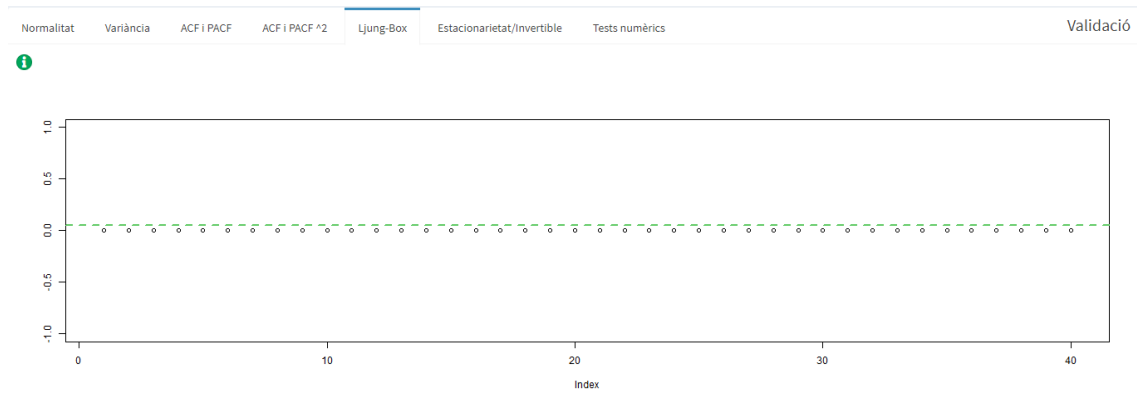


Figura 4.15: Pestanya Ljung-Box

En el gràfic que es presenta a la figura 4.15 es poden estudiar els p-valors del test Ljung-Box visualment. Aquest és un test per detectar si els residus del model estan autocorrelacionats. Tots els punts han d'estar per sobre de 0.05, visualment es detecta el límit amb l'ajuda de la línia verda present en el gràfic.

Acabant la validació gràfica del model, és una bona praxi comprovar si es compleix l'estacionarietat i si aquest model també és invertible. Per verificar-ho, utilitzant els paràmetres estimats automàticament, es grafica si les arrels dels paràmetres tant de la part autoregressiva com per la mitjana mòbil es troben dins o fora del cercle unitari (es presenten una circumferència per cada una de les parts igual que a la figura 4.16). Per assegurar que les arrels es consideren dins de la circumferència, el gràfic proporcionat és interactiu, és a dir, es pot seleccionar l'arrel i veure numèricament la posició en què es troba. A més, en plantejar el model, au-

tomàticament apareixerà una llegenda indicant el color de les arrels que es troben dins de la circumferència.

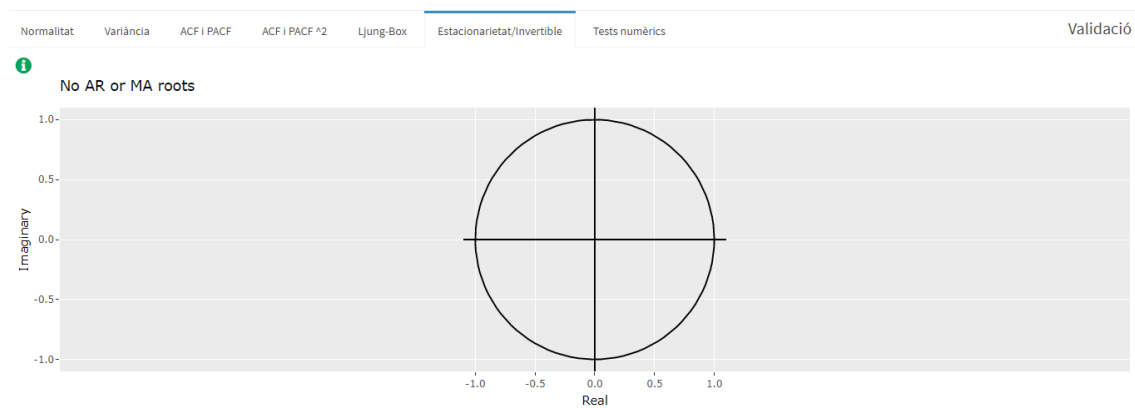


Figura 4.16: Pestanya Estacionarietat/Invertible

Per acabar amb aquesta etapa, en la darrera pestanya d'aquesta opció que ofereix la web creada, es recullen tots els tests numèrics equivalents als gràfics explicats fins ara per comprovar que es compleixen les propietats necessàries en la modelització.

Normalitat: Shapiro-Wilks		Homocedasticitat: Breusch-Pagan		Independència: Durbin-Watson		Independència: Ljung-Box		
W	p_valor	BP	p_valor	DW	p_valor	lag,df	statistic	p.value
0.90805	0.00000	0.00000	0.00000	0.42117	0.00000	1.00000	110.26164	0.00000
						2.00000	172.05100	0.00000
						3.00000	217.13785	0.00000
						4.00000	248.12938	0.00000
						7.00000	297.84695	0.00000
						14.00000	324.92481	0.00000
						21.00000	326.62131	0.00000
						28.00000	346.25273	0.00000

Normalitat: Anderson-Darling		Normalitat: Jarque-Bera	
A	p_valor	X_Squared	p_valor
3.91926	0.00000	80.91200	0.00000

Figura 4.17: Pestanya Tests numèrics

Com es mostra a la figura 4.17, es fa el contrast amb tres estadístics diferents per comprovar la normalitat, un contrast per l'homocedasticitat, un altre contrast per la independència i finalment el test d'independència que també mostra els p-valors que s'han ensenyat anteriorment en el gràfic de Ljung-Box. Als diferents contrastos s'han plantejat en l'apartat on s'expliquen les metodologies.

4.6 Predicció

Per acabar la metodologia Box-Jenkins implementada en la interfície i complint realment l'objectiu principal de predir valors futurs, s'ofereix dins de l'opció predicció diverses pestanyes per fer-ho.

Per dur a terme la predicció, equivalent a l'última etapa de la metodologia Box-Jenkins, s'escau que el model sigui estable. Dins d'aquesta interfície, ja que s'està treballant amb la sèrie temporal per entrenar i la sèrie temporal completa es presenten de nou les estimacions del model que s'han pogut veure en l'apartat d'estimació i s'estima els paràmetres del model utilitzant totes les dades.

Semblant a les sortides que s'han presentat en la part d'estimació, s'han visualitzat a l'esquerra les estimacions per la sèrie temporal incompleta i a la dreta per totes les dades que s'han carregat (4.18).

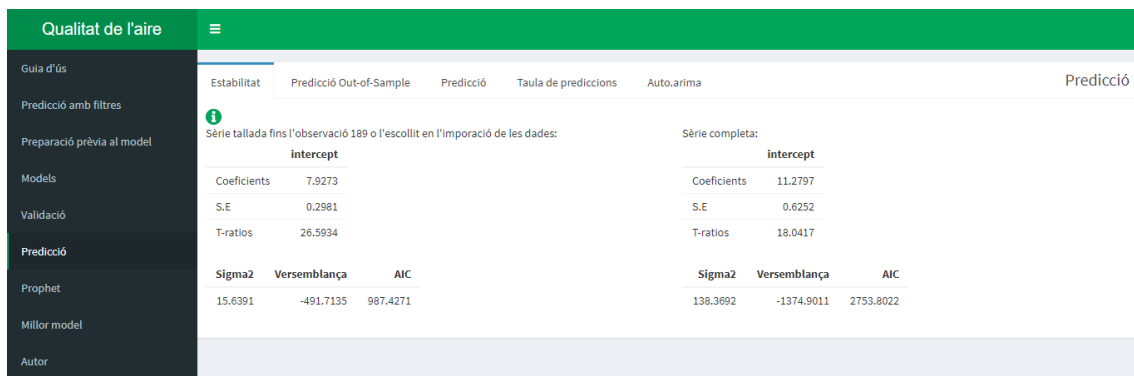


Figura 4.18: Panell per la predicció de la metodologia Box-Jenkins

Abans de fer la predicció a partir de l'última data que s'ha recollit el valor, s'ha presentat dins de la pestanya "Predicció *Out-of-Sample*" com hagués predit el model els valors observats en la sèrie temporal amb la qual s'està treballant, però no s'ha utilitzat en el procés d'entrenament del model.

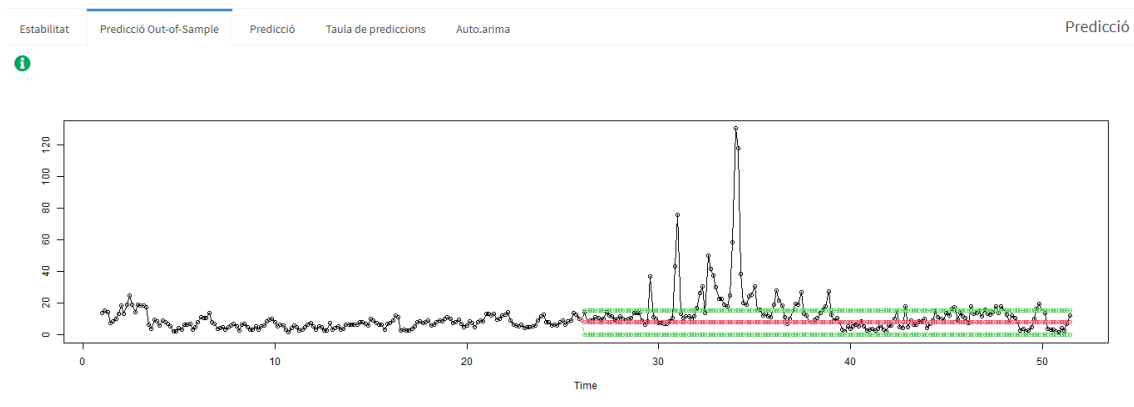


Figura 4.19: Pestanya Predicció Out-of-Sample

Si s'observa la il·lustració 4.19, s'observa com la predicció inicia en l'última data utilitzada per entrenar el model. A més de l'estimació puntual que es mostra en vermell també s'ha dibuixat els intervals de confiança en verd.

Finalment, en la pestanya "Predicció" (4.20) s'ofereix la possibilitat d'escollir el nombre de dies a predir a partir de l'últim recollit en les dades i es veu un

gràfic semblant al previ. A més, hi ha l'opció de navegar i consultar els valors numèrics exactes de les prediccions en la pestanya contigua anomenada “Taula de prediccions”.



Figura 4.20: Pestanyes Prediccions i Taula de prediccions

Per acabar, de la mateixa manera que s’ha dissenyat per la metodologia d’allisat exponencial una manera de fer prediccions automàtica en aquest apartat també s’ha afegit l’opció “Auto.arima”.

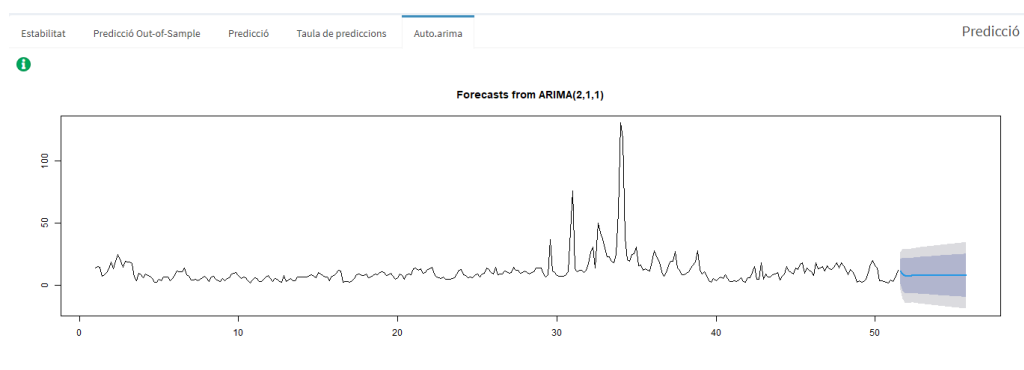


Figura 4.21: Pestanya Auto.arima

A la figura 4.21 es pot veure com en aquesta pestanya es presenta el model que ha calculat el programa automàticament, pot o no coincidir amb el que s’ha creat en

tot el procés de modelització anterior, i s'ha visualitzat les prediccions que fa segons el nombre de dies seleccionats en l'anterior pestanya.

4.7 Prophet

L'última metodologia implementada en aquesta interfície per modelitzar de manera senzilla, és l'algoritme *prophet*, on aquest model es basa en un model no paramètric i completament automàtic per dur a terme les prediccions.

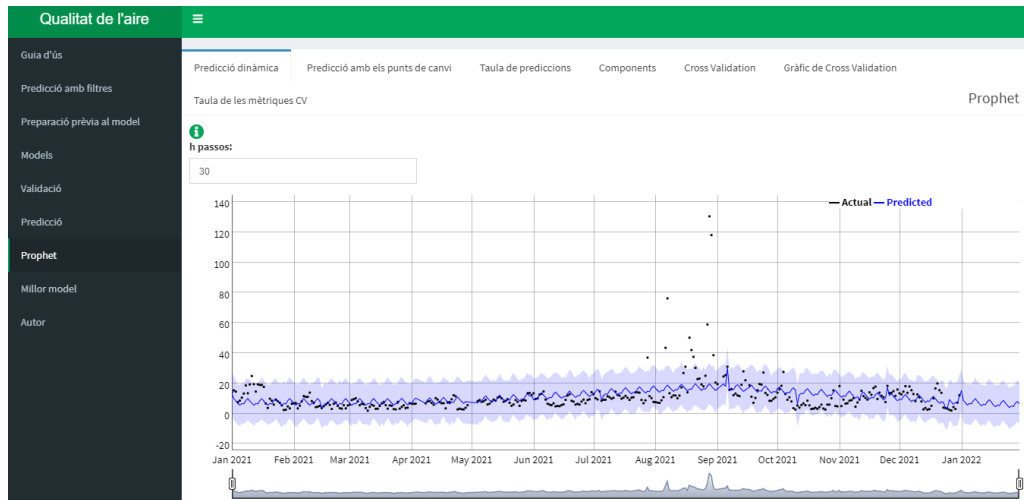


Figura 4.22: Panell principal pel prophet

En primer lloc, s'ha d'escriure el nombre de dies que es vol predir dins de la caixa que apareix sobre el gràfic de la il·lustració 4.22. Automàticament s'actualitzarà el gràfic amb les dades observades en color negre i les dades que s'han predir en blau. Amb l'interval que apareix sota el gràfic es pot acurtar el període que s'està visualitzant.

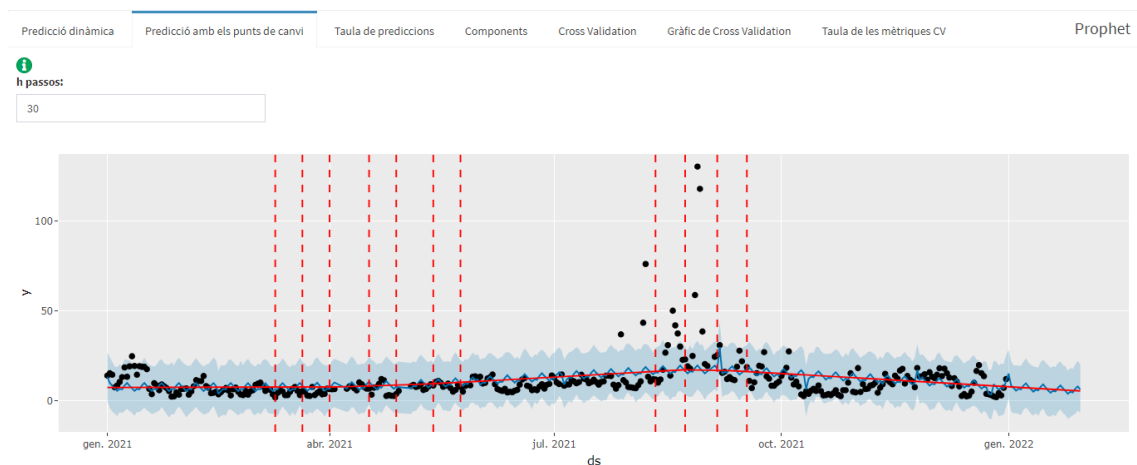


Figura 4.23: Predicció amb els punts de canvi

En la pestanya del costat es presenta el mateix gràfic que s'ha vist, però s'han

afegit en vermell l'estimació del pendent i amb línies discontinües els punts on s'han detectat de manera automàtica un canvi significatiu amb la tendència de les dades (mirar 4.23).

Dins de la pestanya “Taula de prediccions”, amb el mateix format que s’ha vist en les altres metodologies, es presenta una taula amb totes les prediccions i l’interval de confiança que s’han calculat. Com ja s’ha comentat, amb aquesta taula es pot navegar de diferents maneres.

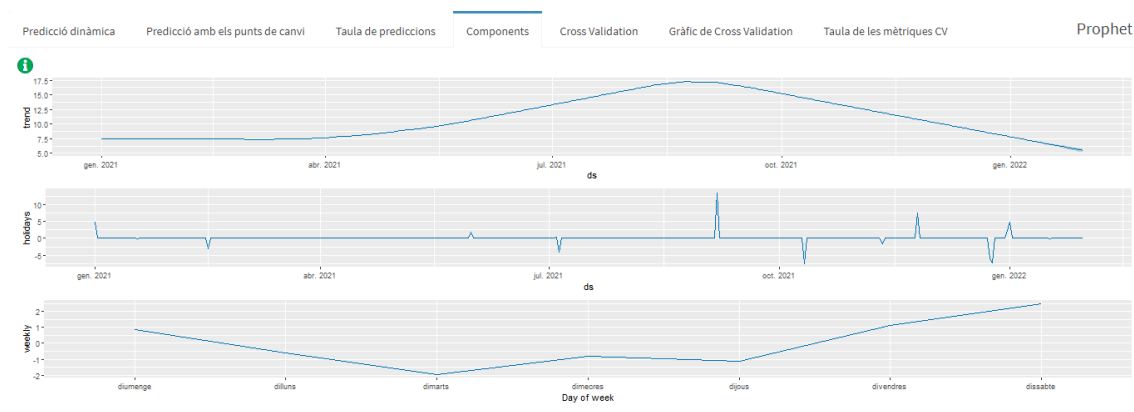


Figura 4.24: Pestanya Components

Per conèixer amb més profunditat el model que s’ha dissenyat per l’algorisme de manera automàtica, es presenten les components que ha utilitzat el model per dur a terme les posteriors prediccions. Aquest compte amb la tendència, les vacances o dies festius del país que s’ha determinat a l’inici de la web i l’estacionalitat tal com es mostra a la il·lustració 4.24. Cal esmentar que pel que fa a l’estacionalitat hi haurà un gràfic per representar l’estacionalitat setmanal i anual si hi ha suficients dades per estimar-les.

Com passa amb la majoria de models, cal validar-lo. En estar treballant amb un model no paramètric ha estat necessari fer *Cross Validation* que es presenten automàticament en els següents panells com a la figura 4.25.

En el primer resultat que es correspon a la part superior de la il·lustració es pot visualitzar cada una de les estimacions en els diferents talls que s’han fet automàticament.

Pel que fa al gràfic, es pot veure inicialment un desplegable amb les diferents mètriques per determinar si el model és prou bo per explicar i predir (a curt i llarg termini). Selecciónant les diferents mètriques s’actualitza el gràfic on mostra el resultat del càlcul de la mètrica en els diferents períodes de temps. És a dir, mostra com de precís, segons la mètrica escollida, és explicant o fent prediccions el model si es prediu un dia o més amb aquest algorisme.

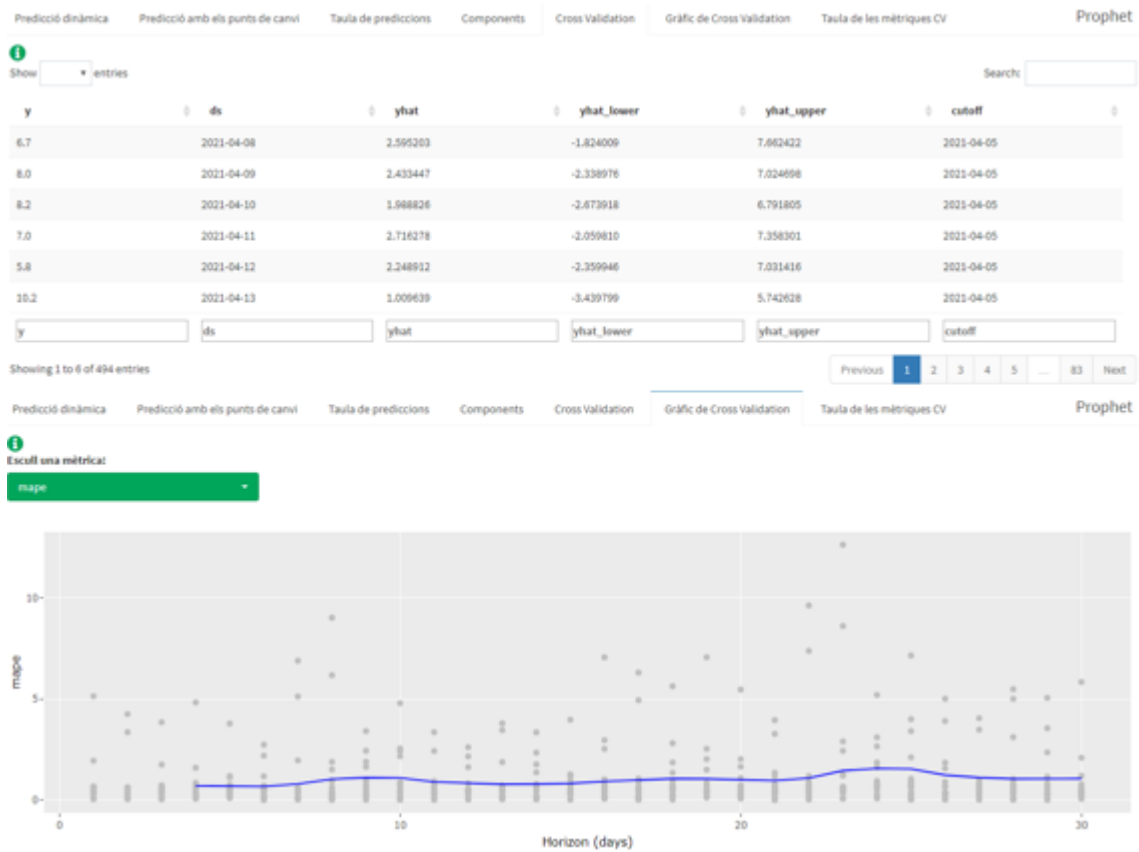


Figura 4.25: Pestanyes Cross Validation i Gràfic de Cross Validation

Per acabar amb l'explicació i visualització dels resultats d'aquest algoritme completament automàtic es visualitza en l'última pestanya una taula resum de les mètriques de validació del model (4.26).

The figure shows the 'Taula de les mètriques CV' tab in the Prophet web interface. It displays a table with the following data:

horizon	mse	rmse	mae	mape	mdape	smape	coverage
4.00	171.96	13.11	7.77	0.70	0.36	0.51	0.78
5.00	158.21	12.62	7.46	0.69	0.41	0.52	0.83
6.00	134.24	11.59	6.67	0.67	0.40	0.49	0.84
7.00	75.75	8.70	5.95	0.79	0.43	0.49	0.80
8.00	95.80	9.79	6.39	1.03	0.41	0.52	0.72
9.00	158.32	12.58	8.30	1.11	0.43	0.61	0.60

Figura 4.26: Pestanya Taula de les mètriques CV

4.8 Millor model

Aplicades totes les metodologies de modelització de sèries temporals cal recapitular als resultats per decidir quin ha estat el millor model explicant i sobretot, fent prediccions.

En aquesta opció que ofereix la web es recullen els indicadors MAPE, RMPSE, MAE, RMSE, nMAE i nRMSE per identificar quin model s'adequa millor a les dades

amb les quals s'estan treballant. Aquesta part és molt útil tant si s'està treballant amb les dades carregades en aquesta interfície o amb unes altres dades de diferent temàtica.

Però com aquesta interfície es presenta com una solució interactiva per estudiar els contaminants de l'aire, també s'han implementat unes altres sortides que ajudaran a completar els objectius de l'estudi.

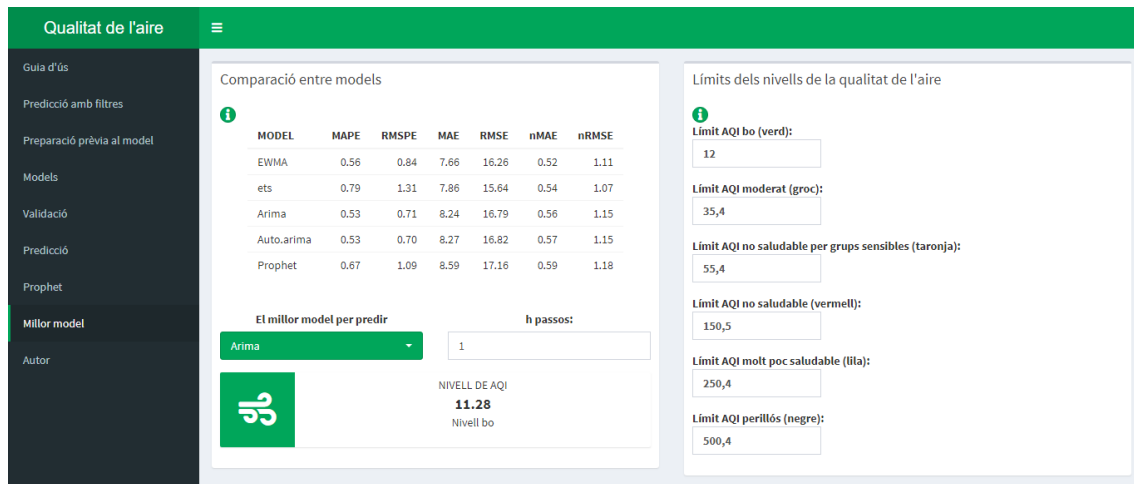


Figura 4.27: Panell principal pel Millor model

Com es pot veure a la imatge 4.27, es presenta un bloc a la dreta per escriure els límits oficials de cada contaminant per considerar-se un nivell de qualitat de l'aire bo, moderat, no saludable per grups sensibles (segons la definició oficial del país), no saludable, molt poc saludable o perillós per la salut.

En el bloc de l'esquerra apareix un desplegable per seleccionar el model òptim segons els indicadors que es mostren a la taula superior i una caixa al costat per predir el nivell de contaminant en els pròxims dies des de l'última observació guardada en la sèrie completa.

De manera dinàmica s'actualitzarà el color i la predicció de la capsa informativa del final del bloc anomenat comparació entre models. Aquesta presentarà el color corresponent al nivell de la qualitat de l'aire que es trobi el contaminant segons els límits imposats a la dreta. A més, per no induir a la confusió, apareixerà sota el valor predit si aquest és bo, o no, segons els nivells oficials (mirar *How is the AQI calculated?* — US EPA, [s. d.]).

Capítol 5

Anàlisi del contaminant NO_2

En aquest capítol s'analitzarà pas a pas, segons les metodologies explicades i implementades dins de la interfície web creada pel projecte, el contaminant diòxid de nitrogen. L'objectiu és poder predir de manera òptima el nivell de concentració del contaminant en un futur pròxim, per exemple en trenta dies.

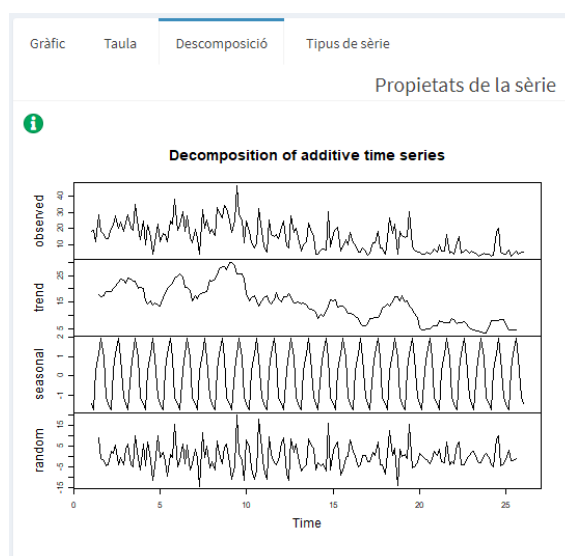
Recordant la petita descriptiva duta a terme en el capítol de la descripció de les dades, la sèrie semblava presentar volatilitat i majors concentracions en les setmanes d'hivern.

5.1 Predicció amb filtres

Utilitzant l'aplicació creada en aquest projecte per modelitzar de manera dinàmica, es començarà l'anàlisi amb el mètode d'allisat exponencial entrant dins de l'opció "Predicció amb filtres" tal com s'ha explicat dins de la guia d'ús de la interfície.

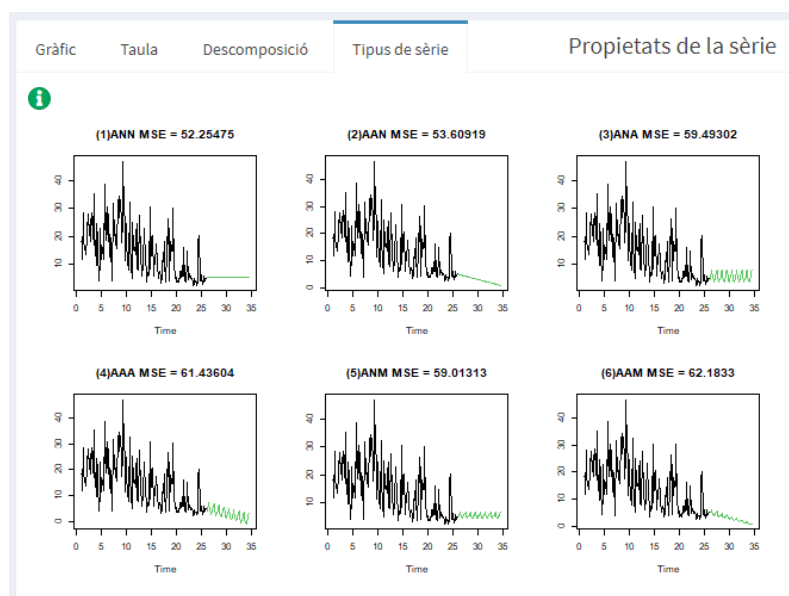
Per poder iniciar amb aquest mètode cal comprendre el comportament de les diferents components de la sèrie temporal. Amb la figura 5.1 s'estudiaran les components que formen l' NO_2 . Amb la gràfica es pot intuir certa tendència lineal decreixent de manera general. Si s'observa cada un dels intervals, també es presenten oscil·lacions mentre les dades semblen decreixer.

D'altra banda, es veu de manera clara la presència d'un patró estacional setmanal, en concret, aquest patró es presenta de manera multiplicativa. Això es pot identificar veient com en la part observada es visualitza el patró estacional de diferents mides durant el temps (en les últimes setmanes sembla presentar-se cada cop més petit).


 Figura 5.1: Descomposició de les components de la sèrie NO_2

Pel que fa al terme de pertorbació present en la sèrie temporal del registre de concentracions del diòxid de nitrogen no sembla seguir cap patró en concret, millor dit, els residus es distribueixen de manera aleatòria.

Segons l'esquema presentat en el capítol de metodologia per identificar el tipus de sèrie temporal amb què s'està treballant, s'hauria d'escollir entre una sèrie temporal de tipus V o una sèrie temporal de tipus VI, ja que són les que presenten estacionalitat multiplicativa.


 Figura 5.2: Càlcul de l'indicador MSE per cada tipus de sèrie en base a les dades del NO_2

Per acabar de decidir quin tipus de sèrie s'adapta millor amb les dades es calcularà l'indicador MSE per cada una d'elles i es compararà entre els dos tipus de sèrie.

Tenint en compte que només s'està comparant entre el tipus V i el tipus VI pel fet que s'ha considerat una estacionalitat multiplicativa, visualitzant la imatge 5.2 s'ha considerat que la sèrie temporal s'adapta a una sèrie tipus V. Observant l'indicador MSE, pel tipus V s'ha trobat un menor MSE que no pas per la sèrie tipus VI.

D'aquesta manera s'ha conclòs que les dades s'adapten a una sèrie de tipus V on la tendència no és significativa, és a dir, no s'ha considerat que les dades presenten una tendència decreixent al llarg del temps. Això és a causa que el mètode de Holt-Winters dona major pes a les últimes observacions on s'ha percebut una estabilitat durant el temps. D'altra banda, sí que s'ha identificat un clar patró estacional i a més, aquest sembla ser multiplicatiu al llarg del temps.

Decidit el tipus de sèrie temporal, cal estimar els coeficients del mètode i així, poder dur a terme les prediccions (5.3).

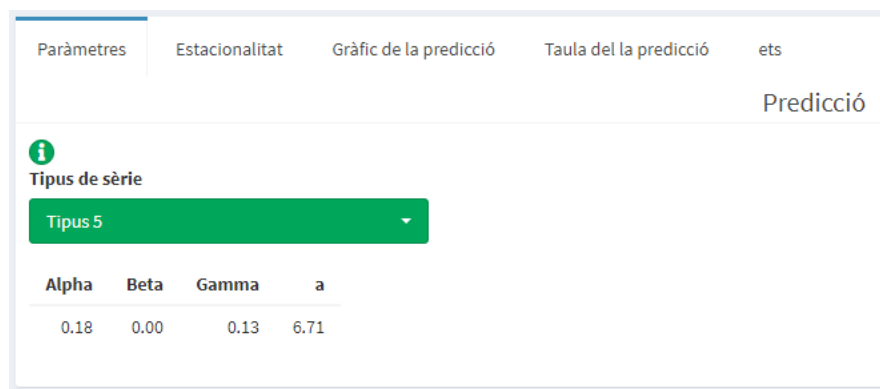


Figura 5.3: Tipus de sèrie temporal que segueix NO_2 i l'estimació dels coeficients.

Estudiant els paràmetres d'allisat sembla que α resulta bastant lluny d'1 indicant així que les últimes observacions no són les que tenen una major ponderació a l'hora d'explicar la sèrie temporal.

El paràmetre β és igual a zero, ja que no s'ha considerat la presència de la tendència significativa i el paràmetre γ ha resultat bastant lluny d'1 indicant que es dona pràcticament el mateix pes a la memòria del patró estacional que l'últim per fer les prediccions.

Per l'estimació de coeficients, a representa l'estimació del nivell de la sèrie temporal en l'últim instant observat l_t . Per tant, s'ha estimat que l'última setmana de desembre del 2021 hi va haver una concentració de 6.71 ppb de NO_2 a Califòrnia.

Per facilitar la interpretació de l'estimació de les estacionalitats de cada dia de la setmana, es visualitza el gràfic 5.4.

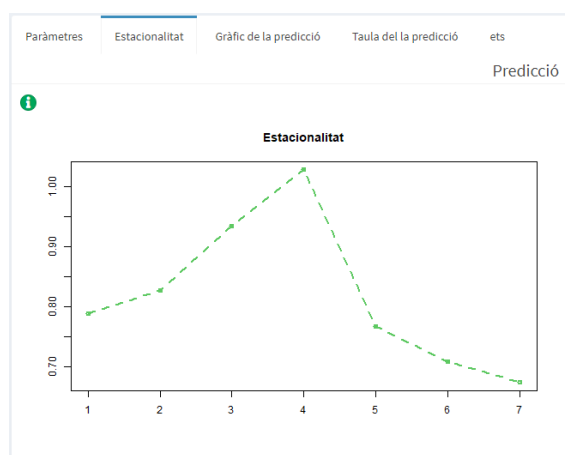


Figura 5.4: Estimació dels coeficients de l'estacionalitat de cada dia de la setmana pel NO_2 .

Cal tenir en compte que la sèrie que s'està estudiant va iniciar el 01/01/2021 el qual va ser un divendres. Per tant, el coeficient estimat en primera posició fa referència al divendres, el coeficient estimat en segona posició correspon al dissabte i així fins al final de la setmana.

Amb la figura 5.4 s'observa que tots els coeficients estimats són majors a 0 i l'estacionalitat presenta un patró en forma de pic. Amb altres paraules, inicia un creixement i després torna a decreixer a nivells més baixos.

A continuació, ja es pot començar amb el procés de predicció. Per complir l'objectiu d'observar el comportament de la predicció en el període que no s'ha utilitzat per definir el tipus de sèrie i a més, predir els pròxims trenta dies es farà la predicció de 212 observacions (trenta dies des de l'últim registre de la sèrie i 182 observacions que s'han reservat per comprovar que el mètode d'allisat exponencial s'adequa a les dades).

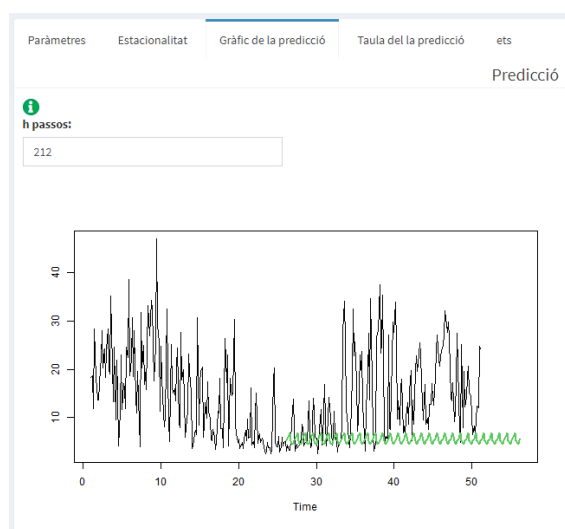


Figura 5.5: Predicció dels pròxims 212 dies per la concentració del NO_2 .

A la figura 5.5 es poden veure les prediccions de color verd. S'observa com que no presenta tendència, però en començar en un dia amb un nivell baix, la tendència es manté estable però a diferent mitjana de les concentracions observades realment. Aquest tipus de sèries temporals són difícils d'explicar segons aquest mètode perquè presenten volatilitat que no es pot recollir mitjançant les estimacions. En canvi, pel patró estacional sí que es pot veure que aquest es manté de la mateixa forma que en les observacions registrades.

Per acabar amb aquesta metodologia, s'identifica el model automàtic basat en la metodologia d'allisat exponencial.

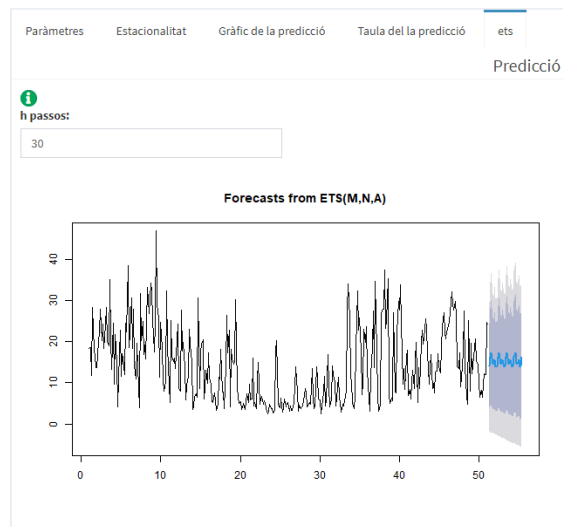


Figura 5.6: Predicció dels pròxims 30 dies del NO_2 segons l'algoritme automàtic.

Observant la figura 5.6 s'observa com l'algoritme automàtic per identificar el tipus de sèrie tampoc ha detectat la presència d'una tendència pronunciada. A diferència de les conclusions a les quals s'ha arribat amb els gràfics que descomponen la sèrie temporal, l'algoritme no ha determinat que el patró estacional sigui multiplicatiu. Per tant, ha escollit que la sèrie temporal de les concentracions del NO_2 s'expliquen segons una sèrie temporal de tipus III.

D'altra banda, en el gràfic temporal 5.6 es veu la predicció i l'interval de confiança tant al 95% (en un blau més fosc) i al 99% (en un blau més clar).

5.2 Preparació prèvia al model

Per poder iniciar amb la metodologia Box-Jenkins, és a dir, plantejar un model estocàstic per poder explicar i predir cal que la sèrie temporal amb què s'està treballant compleixi uns requisits (explicats amb més detall dins de la metodologia).

5.2.1 Observar si la variància és constant

En primer lloc, per determinar si la variància de la sèrie és constant es farà un gràfic de caixes (boxplots) per cada setmana i un gràfic de mitjana-variància. Si es considera que la variància és constant es continuarà treballant amb la sèrie temporal sense aplicar cap transformació, en canvi, si aquesta presenta patrons de no constància, s'haurà d'aplicar una transformació logarítmica.

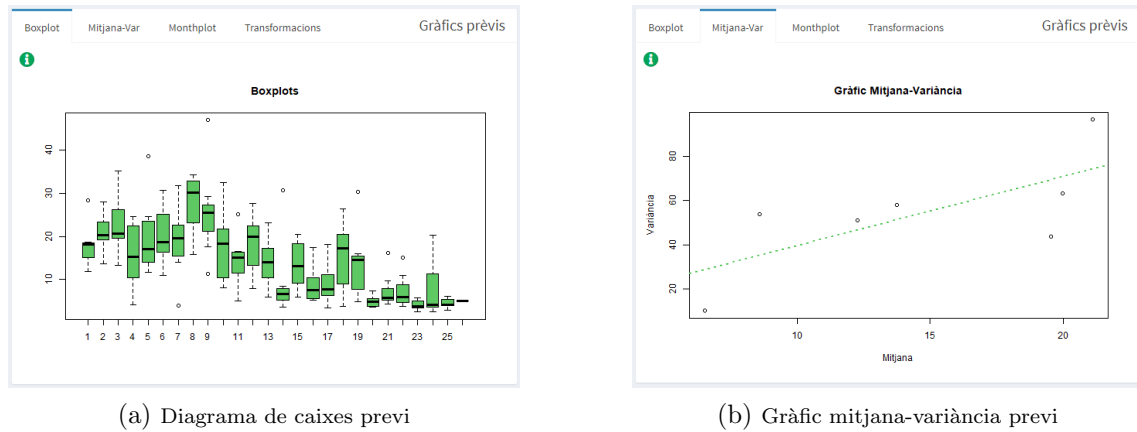


Figura 5.7: Gràfics per estudiar la variància de la sèrie

Observant els rangs interquartílics del diagrama de caixes 5.7a s'indica una possible no constància de la variància de la sèrie. A més, es pot contrastar amb el gràfic de mitjana-variància 5.7b on clarament es veu com es presenta major variància a major nivell en la sèrie.

Com s'ha explicat prèviament, al no presentar-se dins de la sèrie una variància constant al llarg del temps estudiat, s'aplicarà una transformació logarítmica a la sèrie temporal que recull les concentracions del NO_2 .

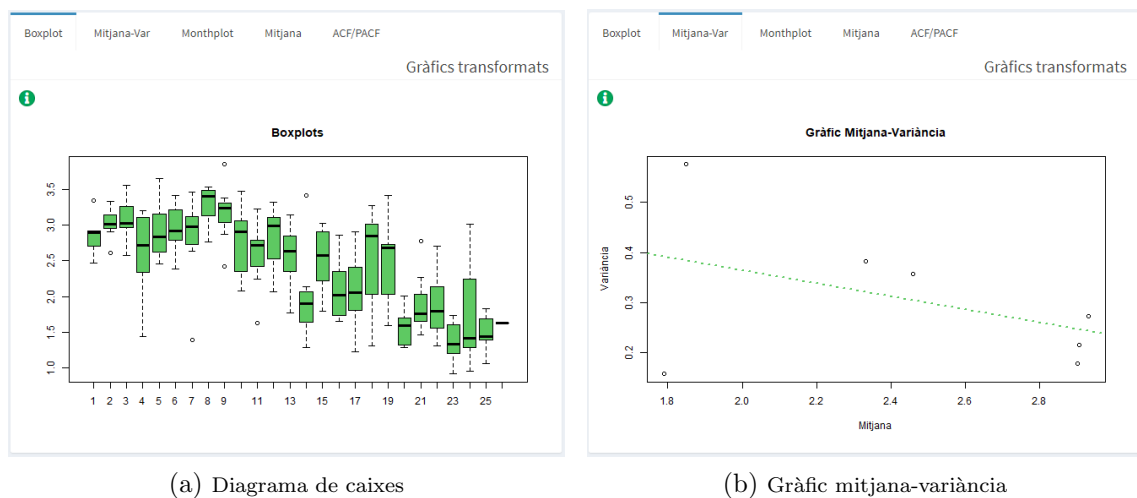


Figura 5.8: Gràfics per estudiar la variància de la sèrie aplicada la transformació

Aplicada la transformació, a la figura 5.8 es visualitzen de nou els gràfics. Es

pot considerar que la sèrie és més o menys homogènia i pel que fa a la variància aquesta ha millorat considerablement els resultats presentats anteriorment.

5.2.2 Observar si existeix patró estacional

En acabar l'estudi de la variància es procedeix amb l'anàlisi de l'estacionalitat de la sèrie temporal.

Per identificar si la sèrie temporal original de les concentracions del NO_2 a Califòrnia presenta estacionalitat, es mirarà el *monthplot* que dibuixa els termes estacionals.

Si es detecta un patró estacional, caldrà eliminar-la de la sèrie amb variància constant mitjançant una diferenciació estacional. D'altra manera es continuarà treballant amb les concentracions del NO_2 sense canvis pel que fa a l'estacionalitat.

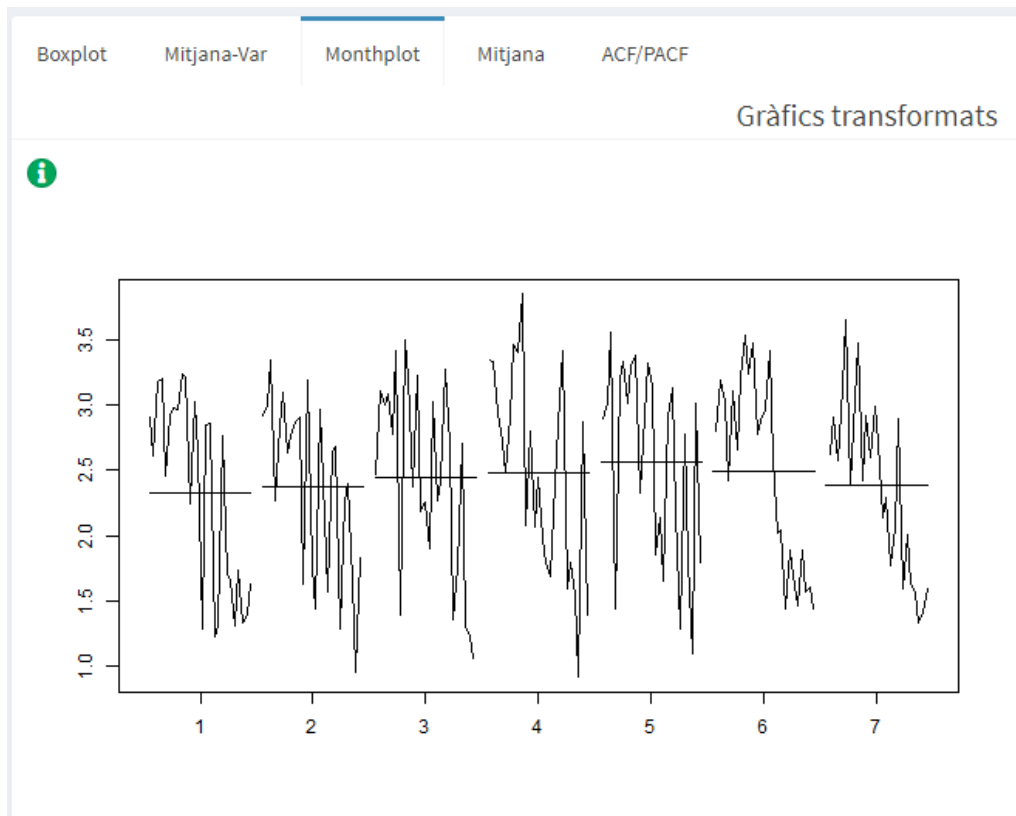


Figura 5.9: Monthplot de la sèrie NO_2 amb mitjana constant

Analitzant la figura 5.9 sembla que hi ha un lleuger patró estacional al llarg de la setmana. Tot i això, no es considera significatiu, ja que la mitjana dels diferents dies de la setmana varien del 2.4 al 2.5 aproximadament, és a dir, la diferència de les mitjanes dels termes estacionals varien relativament poc.

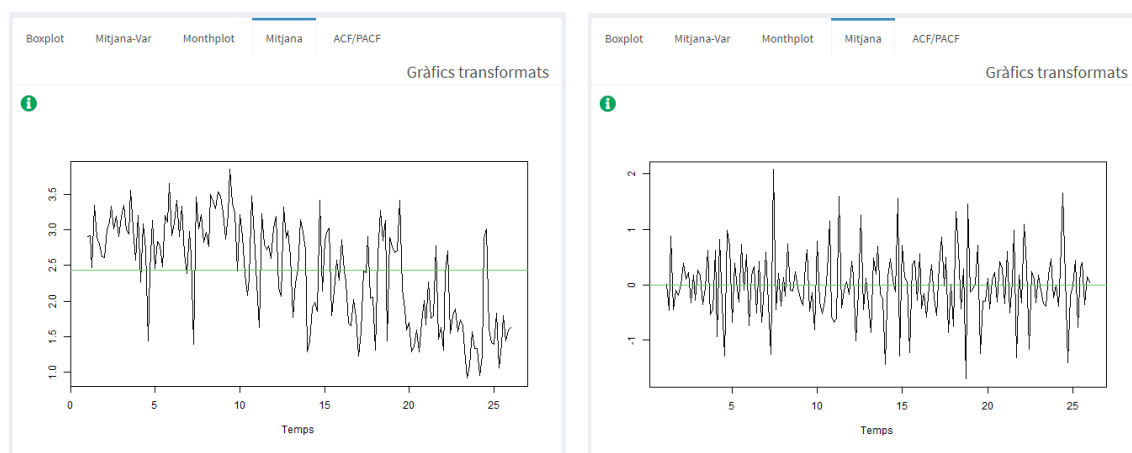
Al no concloure la presència d'un patró estacional en la sèrie temporal que

recull les concentracions del NO_2 a Califòrnia, se seguirà treballant amb la sèrie sense aplicar-hi cap canvi.

5.2.3 Observar si la mitjana és 0

En aquesta secció es durà a terme una anàlisi de la mitjana.

Com es visualitza a la figura 5.10a, no es pot considerar que la mitjana sigui constant. Per aquest motiu s'han d'aplicar tantes diferenciacions regulars com es considerin necessàries per treure les arrels unitàries. La interfície té implementada una funció per calcular les diferenciacions regulars necessàries comparant sempre amb una extra per comprovar que amb la diferenciació la variància de la sèrie no augmenta.



(a) Gràfic temporal dibuixant la mitjana abans de les diferenciacions regulars

(b) Gràfic temporal dibuixant la mitjana abans de les diferenciacions regulars

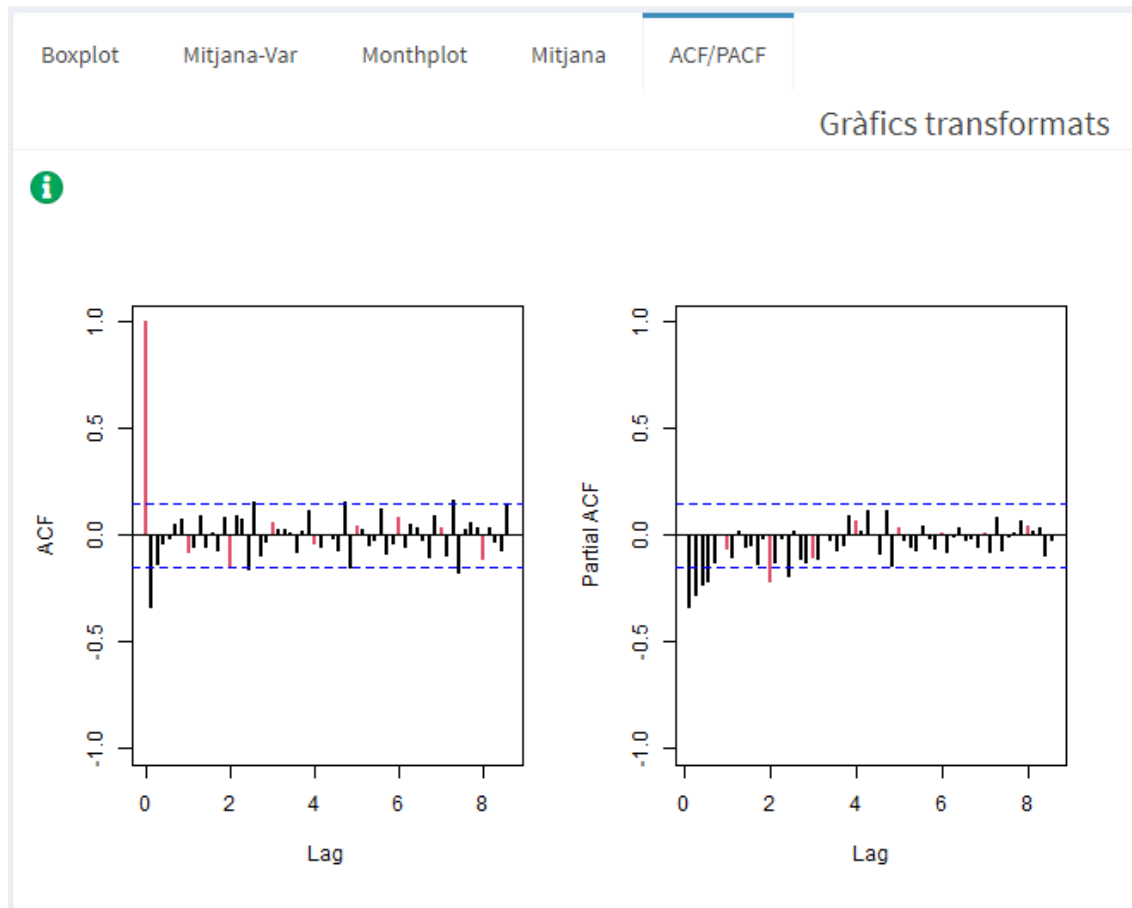
Figura 5.10: Gràfics per estudiar la mitjana de la sèrie NO_2

A la dreta de la figura 5.10b, es mostra com la mitjana ja es pot considerar constant en el 0 sense fer que hi hagi presència de sobre-dispersió.

5.2.4 Observar si la sèrie transformada és estacionària

Finalment, per determinar si la sèrie escollida en l'apartat anterior és estacionària s'analitzarà l'ACF. Per considerar que la sèrie transformada de les concentracions registrades pel diòxid de nitrogen és estacionària, s'haurà d'observar que el gràfic decau ràpidament cap a 0.

Com es pot veure a la figura 5.11, aquest mostra que la sèrie convergeix relativament ràpid decaient cap a 0. Per aquest motiu, es considera que la sèrie transformada ja és estacionària.

Figura 5.11: ACF i PACF de la sèrie NO_2

5.3 Identificació i estimació dels models

En aquest apartat es durà a terme la identificació dels possibles models d'acord amb l'ACF i el PACF tenint en compte el principi de parsimònia, és a dir, si es pot proposar un model senzill que s'adapti a la sèrie temporal de l'estudi s'evitarà complicar el model afegint paràmetres.

Amb l'ACF i PACF, s'identifiquen per la part regular del model un $AR(4)$, ja que observant el PACF es veuen quatre retards regulars fora de les bandes de confiança i un $MA(2)$ perquè en l'ACF hi ha un retard regular fora de les bandes de confiança i el segon molt pròxim. Per aquest motiu, s'estimaran aquests models i després ja es decidirà si realment són significatius mitjançant l'estimació.

Pel que fa a la part estacional es considerarà els retards de color vermell. D'aquesta manera s'ha identificat de manera clara un $SAR(2)$ mitjançant el gràfic del PACF.

Resumint les transformacions i models proposats per cada una de les parts, s'han proposat els següents models (mirar la taula 5.1):

Model 1	$AR(4)SAR(2)_7$ per W_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 1$ i $D = 0$ amb $s = 7$	$\phi_4(B)\Phi_2(B^7)(1 - B)\ln(X_t) = Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \phi_1B - \phi_2B^2 - \phi_3B^3 - \phi_4B^4)(1 - \Phi_1B - \Phi_2B^2)(1 - B)\ln(X_t) = Z_t$
Model 2	$MA(2)SAR(2)_7$ per W_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 1$ i $D = 0$ amb $s = 7$	$\Phi_2(B^7)(1 - B)\ln(X_t) = \theta_2(B)Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \Phi_1B - \Phi_2B^2)(1 - B)\ln(X_t) = (1 + \theta_1B + \theta_2B^2)Z_t$

 Taula 5.1: Identificació dels models per la sèrie NO_2

Utilitzant la interfície web creada, s'estimaran els coeficients dels dos models proposats i es comprovarà si tots els paràmetres són significatius. Si algun dels paràmetres estimats no és significatiu per explicar les dades, aplicant el principi de parsimònia se simplificarà el model.

Model utilitzant la sèrie estacionària. Els T-ratios són significatius si són majors a 2:

	ar1	ar2	ar3	ar4	sar1	sar2	intercept
Coeficients	-0.5338	-0.4478	-0.3518	-0.2126	-0.0670	-0.1433	-0.0081
S.E	0.0740	0.0806	0.0797	0.0735	0.0763	0.0779	0.0131
T-ratios	7.2139	5.5576	4.4118	2.8934	0.8775	1.8393	0.6155
Sigma2	Versemblança		AIC				
0.2758	-136.1071	288.2142					

Figura 5.12: Estimació del model 1

Com es pot veure a la figura 5.12 aplicant el contrast per determinar si els paràmetres estimats són significatius els T -Ratios per la part estacional han resultat menors a 2. Això indica que no és significativa la part estacional i només cal estimar la part regular.

Aplicant de nou l'estimació (5.13) del model es veuen els coeficients dels paràmetres estimats segons la funció de màxima versemblança, seguit dels errors estàndards de cada un dels paràmetres i els T -Ratios majors a dos en valor absolut. D'altra banda, també apareixen l'estimació de la variància de l'error del model, el logaritme de la versemblança maximitzada i l'indicador AIC. Cal recalcar que l'*intercept* no és significatiu, per tant, es treballarà amb el model sense la seva estimació. D'aquesta manera serà més senzill fer les prediccions.

Model utilitzant la sèrie estacionària. Els T-ratios són significatius si són majors a 2:

	ar1	ar2	ar3	ar4	intercept
Coefficients	-0.5455	-0.4629	-0.3374	-0.2123	-0.0075
S.E	0.0737	0.0803	0.0801	0.0735	0.0158
T-ratios	7.3970	5.7635	4.2121	2.8877	0.4768
Sigma2	Versemblança		AIC		
0.2825	-138.0311		288.0622		

Figura 5.13: Estimació del model 1 amb els paràmetres significatius

De la mateixa manera que s'ha fet pel model 1, a continuació es presentaran les estimacions pel model 2.

Estimació A	Estimació B				Model
<p>i Model utilitzant la sèrie estacionària. Els T-ratios són significatius si són majors a 2:</p>					
	ma1	ma2	sar1	sar2	intercept
Coefficients	-0.6543	-0.2598	-0.0934	-0.1634	-0.0088
S.E	0.0699	0.0712	0.0770	0.0767	0.0029
T-ratios	9.3547	3.6514	1.2125	2.1297	3.0295
Sigma2	Versemblança		AIC		
0.2546	-129.7548		271.5096		

Figura 5.14: Estimació del model 2

A la figura 5.14 es pot veure les mateixes sortides que s'han explicat anteriorment. En aquest cas, cal destacar que el model plantejat detecta com a significatiu l'*intercept*. Tot i que, segons l'AIC s'hauria de treballar amb el segon model, els següents apartats es farà l'anàlisi amb el primer model simplificat perquè aquest presenta el valor de l'*intercept* no significatiu. Cal intentar fer l'estudi amb un model sense l'*intercept* significatiu o considerar que no és significatiu per facilitar el procés de predicció.

Per tant, el model amb el qual se seguirà l'anàlisi és el que es presenta a la taula 5.2.

Model 1	$AR(4)$ per W_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 1$ i $D = 0$ amb $s = 7$	$\phi_4(B)(1 - B)\ln(X_t) = Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)(1 - B)\ln(X_t) = Z_t$

Taula 5.2: Model seleccionat per analitzar la sèrie NO_2

5.4 Validació del model

Per dur a terme la validació s'estudiaran les propietats dels termes de pertorbació i si es compleixen les característiques necessàries del model SARIMA.

5.4.1 Normalitat

En primer lloc, s'analitzarà si el terme de pertorbació del model es distribueix segons una distribució normal.

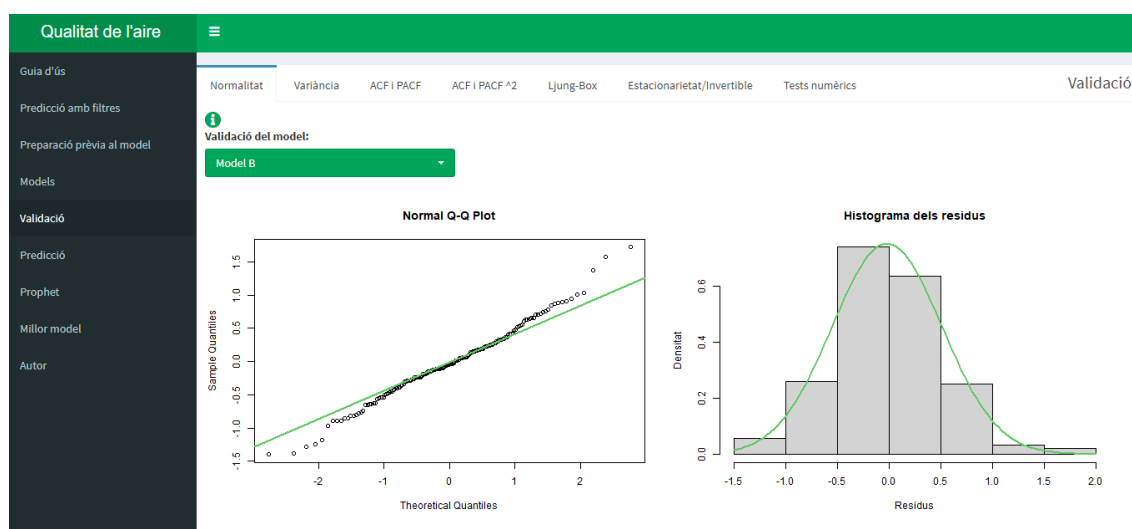


Figura 5.15: Validació de la normalitat

Segons les gràfiques presentades a la figura 5.15 el terme de pertorbació sembla tenir un bon ajust amb el “Q-Q Plot”, és a dir, els residus semblen ajustar-se bé a la distribució Normal. Amb l’histograma es verifica el que s’ha vist amb l’anterior gràfic.

Normalitat: Shapiro-Wilks

W	p_valor
0.98898	0.18881

Normalitat: Anderson-Darling

A	p_valor
0.55833	0.14724

Normalitat: Jarque-Bera

X_Squared	p_valor
4.03725	0.13284

Figura 5.16: Validació de la normalitat amb els contrastos numèrics

Fent costat a les conclusions que s'han obtingut amb els gràfics, els contrastos numèrics plantejats presenten un p-valor major al nivell de significança $\alpha = 0.05$ (5.16). Per tant, es conclou que les dades, en aquest cas els termes de perturbació es distribueixen segons una distribució Normal.

Cal esmenar, que ambdós gràfics presenten una mica de cues pesants que probablement serà a causa d'una possible presència de volatilitat en la sèrie que no s'acaba de recollir dins del model proposat. Tot i això, no sembla ser gaire significatiu, per tant, se seguirà amb el procés de validació del model.

5.4.2 Variància

Seguint amb el procés de validació del model plantejat, en aquesta secció es comprovarà si el terme de perturbació presenta variància constant.

Com s'ha definit en la metodologia Box-Jenkins, per verificar si els residus presenten una variància constant al llarg del temps s'estudiarà els gràfics dels residus i es farà el càlcul del contrast d'hipòtesis per l'homogeneïtat.

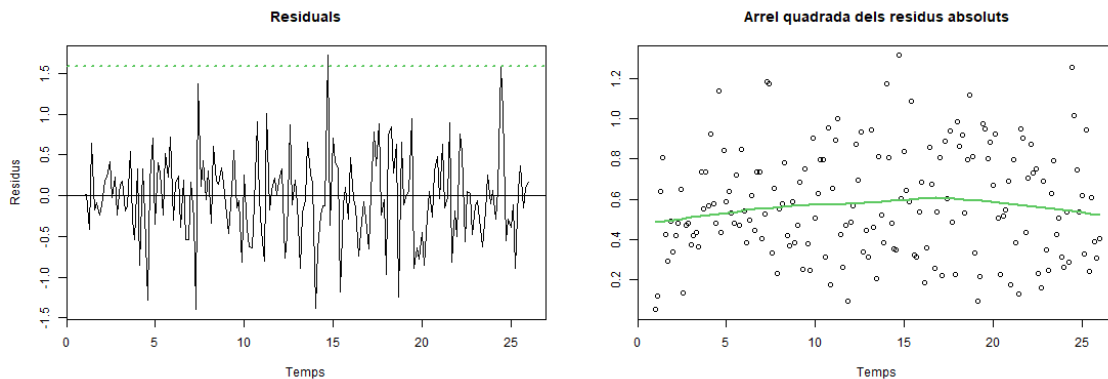


Figura 5.17: Validació de la variància

Segons es mostra a la figura 5.17 en el gràfic de l'esquerra només hi ha un residu que surt de les bandes de confiança al 95% per tant, es considera que la variància que presenta el terme de perturbació del model és constant al llarg del temps.

En el gràfic de la dreta, es recolza l'anterior conclusió, ja que si s'observa la línia verda, sembla presentar-se bastant anivellada. També s'ha comprovat numèricament mitjançant el contrast de Breusch-Pagan resultant amb un p-valor major a $\alpha = 0.05$ conclouent així que els residus són homoscedàstics (5.18).

5.4.3 Autocorrelació

Una de les propietats que ha de complir el terme de perturbació és la independència entre els diferents retards. A continuació es presentaran els diferents gràfics i con-

Homocedasticitat: Breusch-Pagan

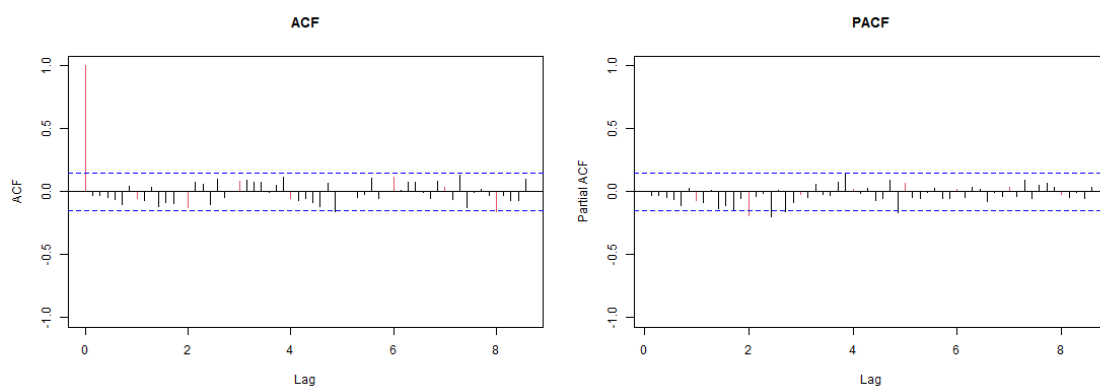
BP	p_valor
0.26020	0.60998

Figura 5.18: Validació de la variància amb el contrast Breusch-Pagan

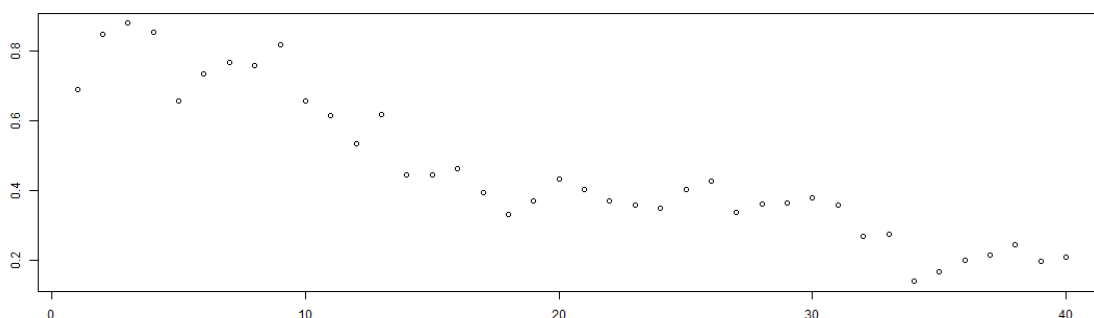
trastos per validar si es compleix aquesta propietat.

Amb l'ACF i el PACF que es mostren a la figura 5.19a no sembla que els diferents retards estiguin autocorrelacionats perquè es presenta en els gràfics la mateixa distribució que el soroll blanc. Si s'observa amb més detall, es pot considerar que en els retards més llunyans hi ha certa correlació, ja que es troben molt propers a les bandes de confiança.

Visualitzant els p-valors del contrast Ljung-Box (5.19b), sembla que tots es troben per sobre del nivell de significació de $\alpha = 0.05$. Per tant, es pot validar i concloure que els residus són independents entre ells.



(a) ACF i PACF del terme de pertorbació



(b) P-valors del contrast Ljung-Box

Figura 5.19: Gràfics per validar la independència dels residus

En quant als contrastos d'hipòtesi numèrics, s'ha aplicat els tests de Durbin-Watson i s'han ensenyat els resultats numèrics del contrast Ljung-Box.

Com es veu a la figura 5.20 els p-valors són majors al nivell de significança

$\alpha = 0.05$ corroborant així que els residus són independents.

Independència: Durbin-Watson		Independència: Ljung-Box		
DW	p_valor	lag.df	statistic	p.value
2.06015	0.62728	1.00000	0.16107	0.68818
		2.00000	0.33430	0.84607
		3.00000	0.68050	0.87778
		4.00000	1.35253	0.85240
		7.00000	4.12684	0.76506
		14.00000	14.06921	0.44457
		21.00000	21.94336	0.40278
		28.00000	30.01613	0.36245

Figura 5.20: Validació de la independència dels residus

5.4.4 Estacionarietat i invertibilitat

Com s'explica dins de la metodologia és una bona praxi comprovar si el model proposat és estacionari i invertible. Per fer-ho, s'analitzaran les arrels del polinomi característic del model.

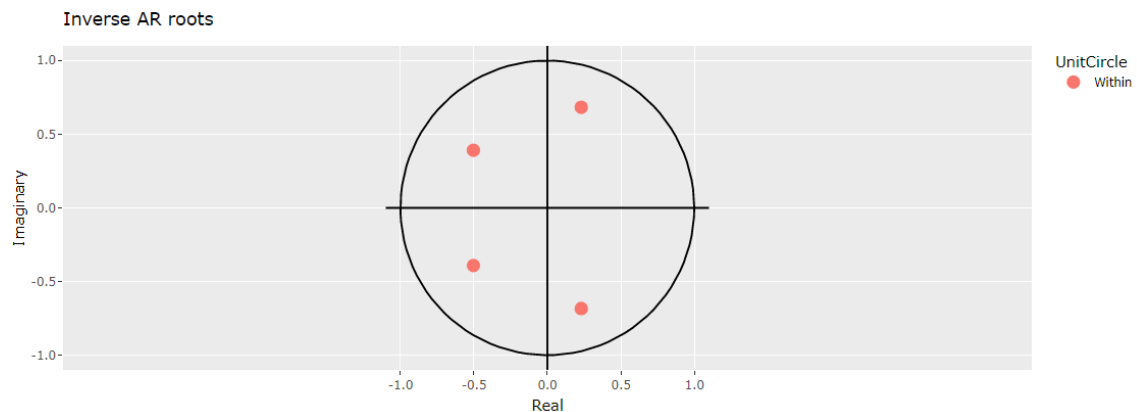


Figura 5.21: Validació de l'estacionarietat del model

Recordant que el model que s'està utilitzant per a l'anàlisi és un model auto-regressiu, una de les seves propietats és que sempre serà invertible. Per tant, cal estudiar si les arrels del polinomi característic són estacionàries. Mitjançant la figura 5.21 es verifica que totes les arrels inverses es troben dins de la circumferència unitària, com a resultat, es conclou que el model presentat és estacionari i invertible.

5.4.5 Volatilitat

En estar treballant amb una sèrie temporal que històricament ja s'han fet investigacions i se sap que presenta bastant volatilitat, s'ha cregut important comprovar si el

model recull bé la volatilitat, o no. Mitjançant l'ACF i el PACF al quadrat es pot analitzar la presència de la volatilitat. Per determinar que no n'hi ha, cal observar soroll blanc en els gràfics.

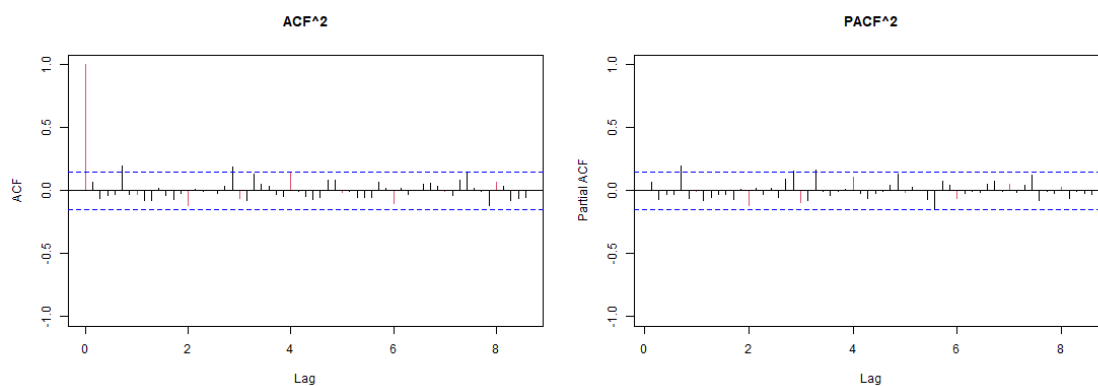


Figura 5.22: Comprovació de la volatilitat dins del model

Analitzant la figura 5.22 es pot considerar que els gràfics presenten soroll blanc perquè només hi ha dos o tres retards que surten de les bandes de confiança. Això pot ser per al 95% de significació de les bandes o simplement per l'aleatorietat. Sigui com sigui, aquesta no és una característica que sembli preocupant pel model.

5.5 Predicció

Per dur a terme la predicció i d'aquesta manera complir un dels objectius principals de l'estudi és necessari comprovar l'estabilitat del model.

Per fer-ho, s'ajusta el model per les dades d'entrenament i per totes les dades registrades l'any 2021. Per considerar el model estable, s'ha d'observar aproximadament les mateixes magnituds pel que fa als coeficients estimats i els mateixos signes.

Sèrie tallada fins l'observació 189 o l'escollit en l'imporació de les dades:					Sèrie completa:				
	ar1	ar2	ar3	ar4		ar1	ar2	ar3	ar4
Coefficients	-0.5444	-0.4617	-0.3360	-0.2112	Coefficients	-0.3748	-0.3357	-0.2669	-0.1918
S.E	0.0738	0.0803	0.0801	0.0735	S.E	0.0524	0.0543	0.0542	0.0524
T-ratios	7.3812	5.7470	4.1948	2.8721	T-ratios	7.1582	6.1806	4.9206	3.6582
Sigma2	Versemblança		AIC		Sigma2	Versemblança		AIC	
0.2829	-138.1445	286.2889		0.2974	-285.4575	580.9151			

Figura 5.23: Anàlisi de l'estabilitat del model

Com es pot veure a la figura 5.23, tots els paràmetres semblen complir les condicions d'estabilitat concloent així que el model és estable durant el temps.

En segon lloc, es realitzarà la predicció *Out-of-Sample* per observar de manera gràfica si el model realment a predit correctament els valors registrats durant la segona meitat de l'any 2021.

El gràfic que mostra la figura 5.24 mostra com el model prediu molt bé les observacions que es van registrar en l'última meitat de l'any i totes es troben dins de l'interval de confiança.

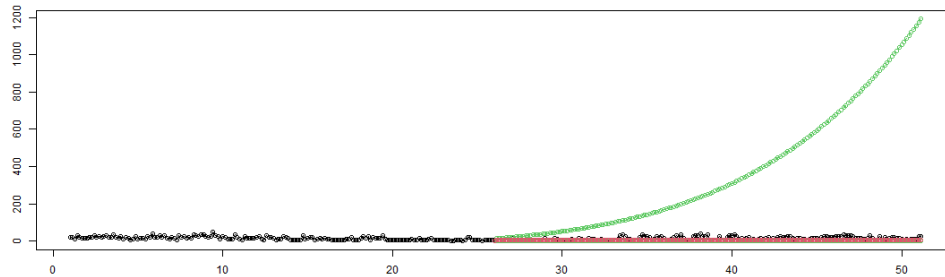


Figura 5.24: Predicció Out-of-Sample

Per acabar, es farà la predicció dels pròxims trenta dies tenint en compte totes les dades registrades resultant un gràfic com el que es mostra a la figura 5.25.

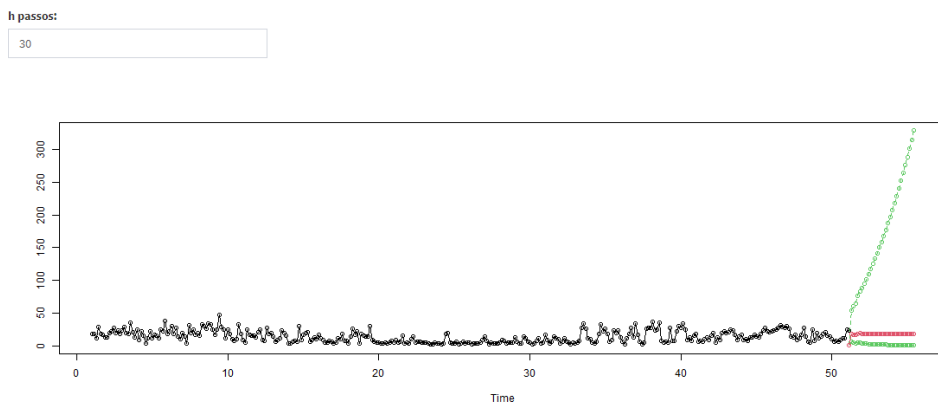


Figura 5.25: Predicció de les 30 observacions futures

5.5.1 Identificació automàtica del model

En aquesta subsecció s'aplicarà l'algoritme automàtic explicat a la metodologia per determinar si aquest ha detectat, o no, el mateix model SARIMA amb el que s'ha treballat durant aquest capítol.

A la imatge 5.26 es pot veure que l'algoritme ha detectat un model diferent per explicar les concentracions registrades de NO_2 .

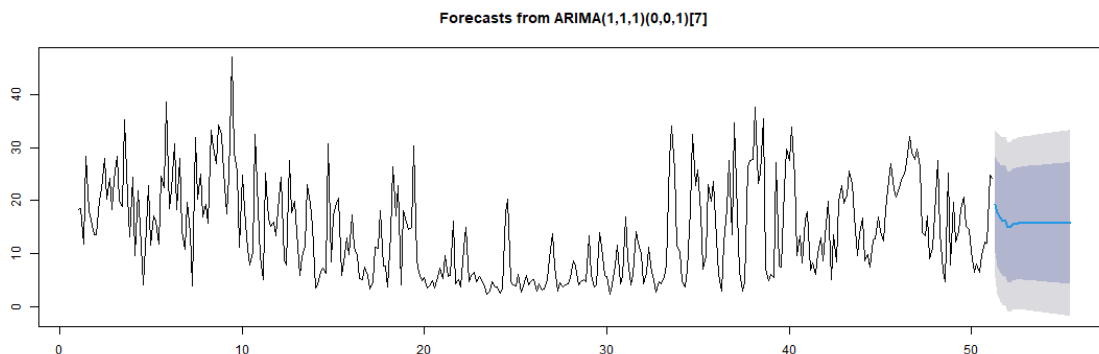


Figura 5.26: Identificació automàtica del model i predicció de 30 observacions futures

5.6 Prophet

En l'última etapa de l'anàlisi d'aquest contaminant s'aplicarà una metodologia no paramètrica per modelitzar i predir de manera automàtica.

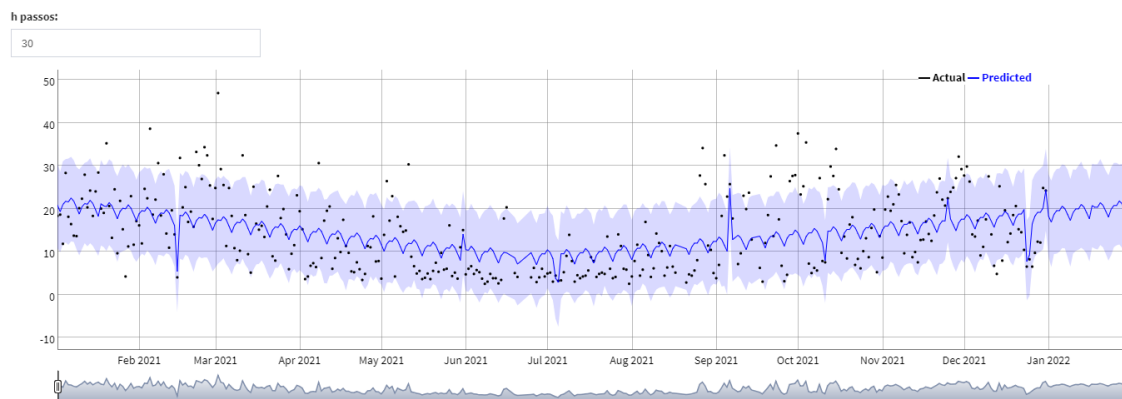


Figura 5.27: Ajust i predicció de la sèrie NO_2

De manera automàtica, l'algorisme implementat per Facebook ha dut a terme l'ajust d'un model no paramètric a la sèrie temporal i la predicció de les 30 observacions futures tal com es mostra a la figura 5.27.

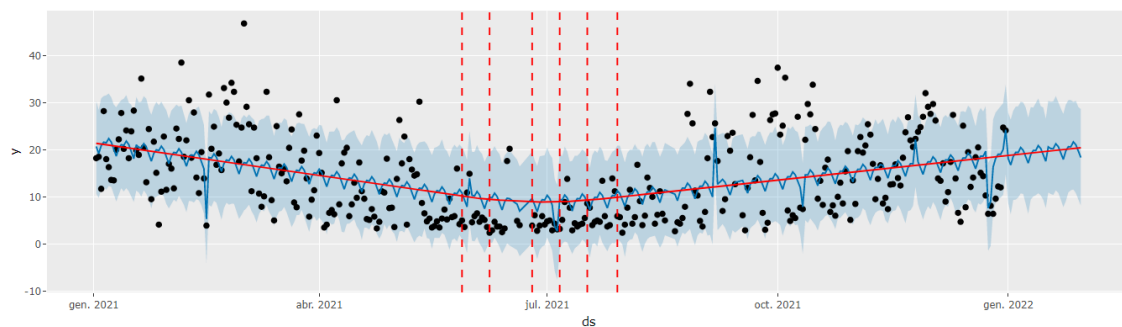


Figura 5.28: Estimació de la tendència i punts de canvi d'aquesta component.

L'algorisme plantejat també identifica de manera automàtica la tendència i els

punts de canvi d'aquesta component. Dins de la web creada s'ofereix una opció per visualitzar de manera ràpida aquests punts i l'estimació de la tendència (figura 5.28).

De la mateixa forma que s'han descrit les components dins de la metodologia d'allisat exponencial, aquest algoritme necessita identificar les components que formen el model no paramètric. Aquest ofereix la visualització de les seves components tal com es mostra a la següent il·lustració 5.29.

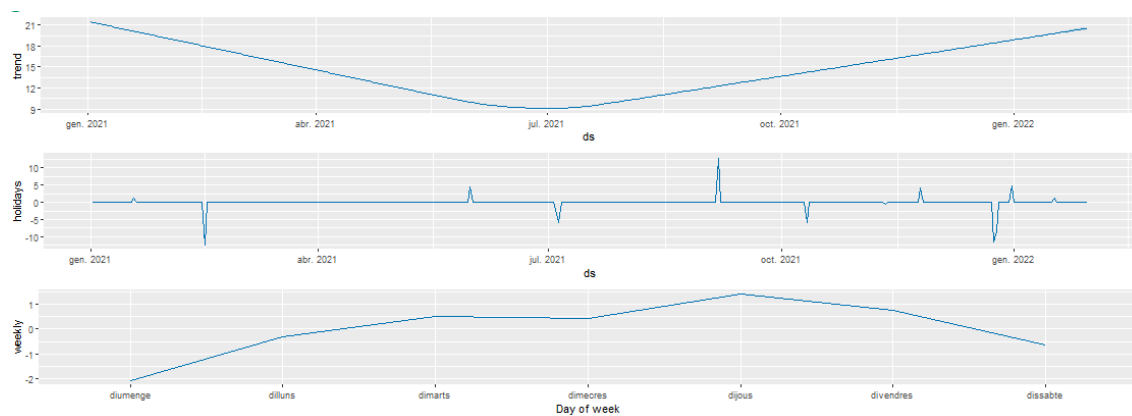


Figura 5.29: Components del model no paramètric proporcionat per Prophet

Finalment, com s'ha fet en totes les metodologies analitzades, cal validar si aquest model és bo explicant i sobretot, fent prediccions. Per fer-ho, s'ha de fer *Cross-Validation* i seguidament, el càlcul de les diferents mètriques de regressió.

horizon	mse	rmse	mae	mape	mdape	smape	coverage
4.00	84.16	9.17	6.96	0.66	0.42	0.60	0.74
5.00	64.90	8.06	6.53	0.71	0.53	0.65	0.75
6.00	75.82	8.71	7.02	0.75	0.63	0.69	0.73
7.00	61.01	7.81	6.18	0.63	0.47	0.59	0.76
8.00	65.42	8.09	6.16	0.59	0.51	0.56	0.78
9.00	73.63	8.58	6.31	0.63	0.55	0.58	0.84

Figura 5.30: Mètriques de regressió

A la figura 5.30 es mostra com el model creat per l'algoritme explica i fa bones prediccions a curt termini.

5.7 Millor model

Aplicades les diferents metodologies per modelitzar una sèrie temporal i fetes les prediccions per cada una d'elles sorgeix el dubte de quin de tots els mètodes analitzats és l'òptim per fer prediccions de la concentració de NO_2 .

Per decidir quin és el millor model es calcularan diferents mètriques de regressió que indiquen com de bo, o no, és cada model per predir. Amb aquestes mètriques es pot comparar quina metodologia ha estat millor per explicar i predir les concentracions.



Figura 5.31: Mètriques de regressió per comparar els diferents models

Segons el que es mostra a la figura 5.31 es considera que el millor model de tots els estudiats és l'ARIMA provinent de la metodologia Box-Jenkins. Tot i això, cal mencionar que els valors dels indicadors són bastant pròxims, per tant, les altres metodologies també són bones per fer prediccions.

Com que els objectius de l'estudi eren identificar quin és el model òptim per fer prediccions i estimar els valors futurs de la concentració de contaminant NO_2 en un futur pròxim, per exemple, en trenta dies, s'han afegit els límits per identificar el nivell de qualitat i s'ha introduït 30 passos per obtenir la predicció puntual segons el model seleccionat. D'aquesta manera s'observa que en trenta dies, és a dir, el 30/01/2022 s'estimava una concentració de 18.53 ppb de contaminant NO_2 en l'aire de Califòrnia. A més, es mostra que aquesta estimació es troba dins del rang de bona qualitat de l'aire segons les fonts oficials de l'estat encarregades de l'àmbit de la salut.

Capítol 6

Resum dels anàlisis dels contaminants

En aquest capítol es presenten els resultats de l'anàlisi duts a terme pels contaminants $PM_{2.5}$, PM_{10} , CO i O_3 .

De la mateixa manera que s'ha presentat dins del capítol 5, en aquesta secció s'han estudiat els diferents contaminants aplicant les diferents metodologies amb ajuda de la plataforma web creada per aquest projecte i es presenten quins han estat els principals resultats.

Per examinar els gràfics i fer un estudi profund s'ha annexat les principals sortides de l'estudi a l'Appendix A.

6.1 Mètodes d'allisat exponencial

Per iniciar l'explicació dels resultats de l'anàlisi, es presentaran quines han estat les conclusions per cada un dels contaminants pel mètode d'allisat exponencial.

Contaminant	$PM_{2.5}$	PM_{10}	CO	O_3
Tipus de sèrie	Tipus V (ANM)	Tipus I (ANN)	Tipus V (ANM)	Tipus I (ANN)

Taula 6.1: Tipus de sèrie per cada un dels contaminants

Amb ajuda de la petita estadística descriptiva presentada en el capítol 2 i l'estudi de cada una de les components de les diferents sèries que recullen les concentracions de cada contaminant, s'ha arribat a les conclusions que es presenten a la taula 6.1. Per cada un dels contaminants es presenten de manera significativa diferents característiques.

Fent l'estudi de les components com s'ha presentat en el capítol anterior, s'ha observat que el $PM_{2.5}$ tenia un patró estacional clar. A més, segons les observacions, aquesta estacionalitat que presenta ha de ser multiplicativa. Pel que fa a la tendència, no s'ha vist de manera evident la seva existència en la sèrie temporal.

Per les partícules en suspensió PM_{10} no s'ha aconseguit identificar cap de les components de la sèrie. Per aquest motiu s'ha aplicat el mètode de decisió del tipus de sèrie d'acord amb el càlcul del MSE. L'òxid de carboni, de la mateixa manera que ha passat amb les $PM_{2.5}$, s'ha considerat evident que la sèrie temporal presenta un patró estacional multiplicatiu sense tendència.

Per acabar, l' O_3 s'ha presentat molt estable durant el període estudiat i sense una estacionalitat clara.

6.2 ETS

Com s'ha fet en l'anterior capítol i s'ha explicat dins de la metodologia, l'aplicació web creada ofereix l'opció d'aplicar un algoritme automàtic per identificar el tipus de sèrie. Aquesta funció s'ha aplicat per cada una de les sèries temporals dels contaminants estudiats obtenint els resultats que es resumeixen a la taula 6.2.

Contaminant	$PM_{2.5}$	PM_{10}	CO	O_3
Resultat de l'ETS	ETS(MAM)	ETS(MAM)	ETS(MAN)	Tipus I (ANN)

Taula 6.2: Tipus de sèrie identificats per la funció ETS

Si es comparen els resultats obtinguts per l'algoritme automàtic d'identificació del tipus de sèrie i les conclusions a les quals s'ha arribat seguint els passos de la metodologia d'allisat exponencial i el criteri analític es veu com no s'ha arribat a les mateixes conclusions excepte per l'ozó.

6.3 Box-Jenkins

Aplicant la metodologia Box-Jenkins com en l'exemple presentat dins de l'anàlisi del NO_2 , s'ha dut a terme l'estudi de cada un dels contaminants identificant diversos models i escollint el més bo tenint en compte l'AIC, la significació de l'estimació de cada paràmetre i la validació dels models proposats.

Tenint en compte totes aquestes característiques, s'han definit els models que es presenten a la taula 6.3 per cada una de les sèries temporals.

Presentats els models que s'han considerat millors segons els gràfics i contrastos

$PM_{2.5}$	$ARMA(1,1)SMA(1)_7$ per W_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 1$ i $D = 1$ amb $s = 7$	$\phi(B)(1 - B)(1 - B^7)\ln(X_t) = \theta(B)\Theta(B^7)Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \phi B)(1 - B)(1 - B^7)\ln(X_t) = (1 + \theta B)(1 + \Theta B)Z_t$
PM_{10}	$ARMA(1,1)SMA(1)_7$ per W_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 1$ i $D = 1$ amb $s = 7$	$\phi(B)(1 - B)(1 - B^7)\ln(X_t) = \theta(B)\Theta(B^7)Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \phi B)(1 - B)(1 - B^7)\ln(X_t) = (1 + \theta B)(1 + \Theta B)Z_t$
CO	$AR(3)SAR(3)_7$ per W_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 1$ i $D = 1$ amb $s = 7$	$\phi_3(B)\Phi_3(B^7)(1 - B)(1 - B^7)\ln(X_t) = Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3)(1 - \Phi_1 B^7 - \Phi_2 B^{14} - \Phi_3 B^{21})(1 - B)(1 - B^7)\ln(X_t) = Z_t$
O_3	$ARMA(1,1)$ per X_t
Forma compacte del model ARIMA per $\ln(X_t)$ amb $d = 0$ i $D = 0$ amb $s = 7$	$\phi(B)X_t = \theta(B)Z_t$
Substituint cada polinomi característic s'obté:	$(1 - \phi B)X_t = (1 + \theta B)Z_t$

Taula 6.3: Identificació dels models per cada un dels contaminants

presentats en l'Appendix A, el contaminant $PM_{2.5}$ s'han validat totes les seves premisses.

Però pel contaminant de la mateixa família, el PM_{10} s'ha identificat un valor extrem que ha provocat la invalidació dels tests numèrics per la normalitat i variància constant del terme de pertorbació. Tot i això, s'ha considerat un bon model per estudiar, ja que només necessita un tractament del valor extrem.

Estudiant l'òxid de carboni s'obté la verificació de totes les propietats necessàries per continuar amb el model excepte l'homogeneïtat de la variància. Realment, la invalidació d'aquesta propietat dels residus és habitual en aquest tipus de sèries perquè solen presentar bastant volatilitat a causa dels factors externs.

Finalment, l'ozó és una de les sèries temporals estudiades més estables i sense estacionalitat de manera clara. Amb el plantejament del model pateix la invalidació de la variància constant pels residus de la mateixa forma i causes que el CO .

6.4 Modelització automàtica

En aquest apartat es presenten els resultats obtinguts per la modelització dels contaminants de l'aire mitjançant l'algoritme automàtic que proporciona el R dins de la funció "auto.arima". Aquest algoritme també segueix la metodologia de Box-

Jenkins, però aquest és molt restrictiu a l'hora de decidir quin és el millor model per explicar les dades proporcionades. Per tant, a la taula 6.4 es pot veure si aquest programa ha assolit els mateixos resultats que en l'apartat anterior.

$PM_{2.5}$	$ARIMA(2, 1, 1)$
PM_{10}	$SARIMA(2, 1, 2)$
CO	$SARIMA(0, 1, 4)(2, 0, 1)_7$
O_3	$ARIMA(2, 1, 1)$

Taula 6.4: Identificació automàtica dels models per cada un dels contaminants

Comparant els resultats obtinguts en aquest apartat i l'anterior secció es pot veure com els models plantejats fent l'estudi previ de les sèries i analitzant les diferents gràfiques sense automatitzar s'han aconseguit plantejar bons models per fer prediccions i sense utilitzar tants paràmetres. Cal recordar que l'estimació dels diferents paràmetres tenen un cost computacional, per tant, si s'assoleix ser una mica més flexibles en alguna de les premisses a validar i estalviant l'estimació d'algun paràmetre es pot obtenir un altre tipus de millora del model.

6.5 Prophet

Aplicant l'última metodologia implementada en aquest treball, l'algoritme *prophet* ha estimat un model no paramètric de manera automàtica per cada un dels contaminants.

Seguidament, es presentaran les components estimades pel model no paramètric (6.1).

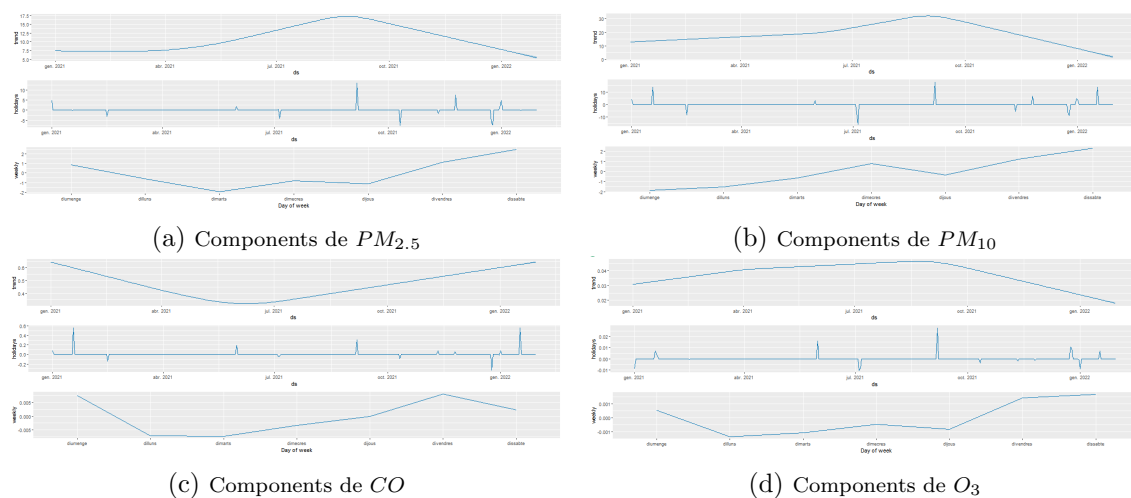


Figura 6.1: Estimació de les components segons Prophet

Tant les prediccions com les mètriques obtingudes per validar el model es troben en l'Appendix A.

6.6 Millor model

Per acabar l'anàlisi i complir l'objectiu principal es determinarà quin ha estat el millor mètode per fer prediccions segons les mètriques de regressió i es farà una predicció a un futur pròxim, en concret a trenta dies.

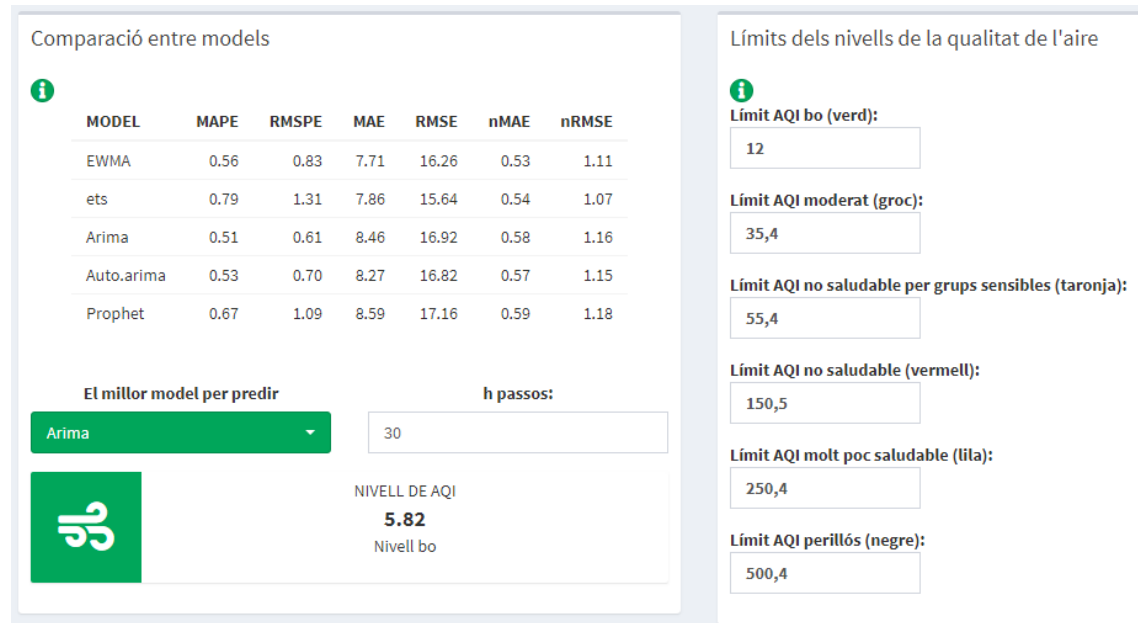
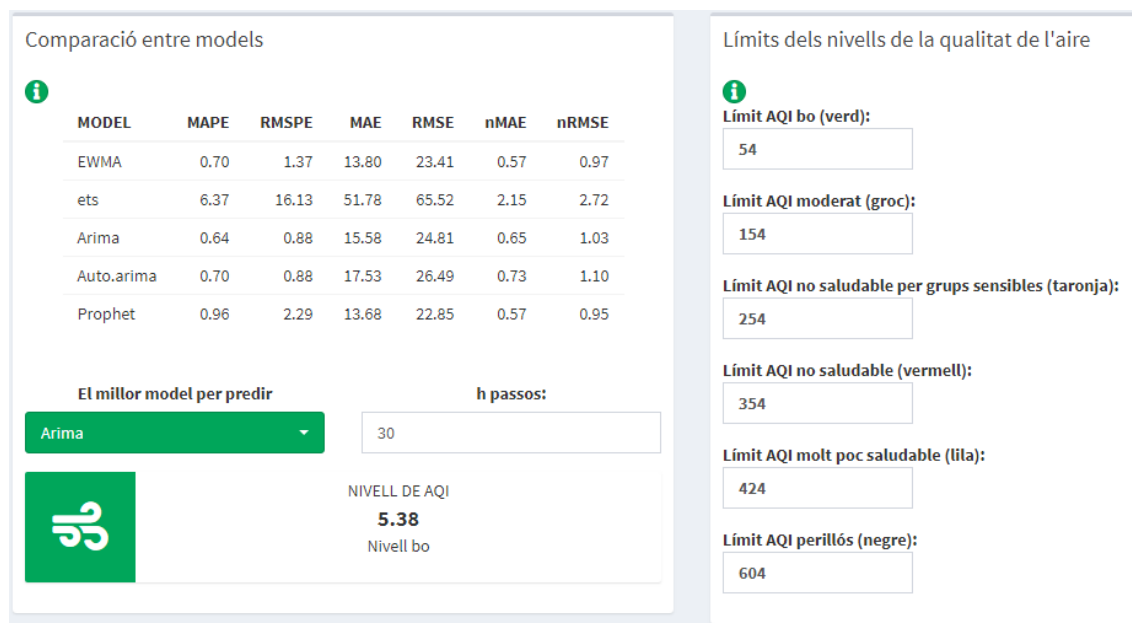


Figura 6.2: Millor model pel $PM_{2,5}$

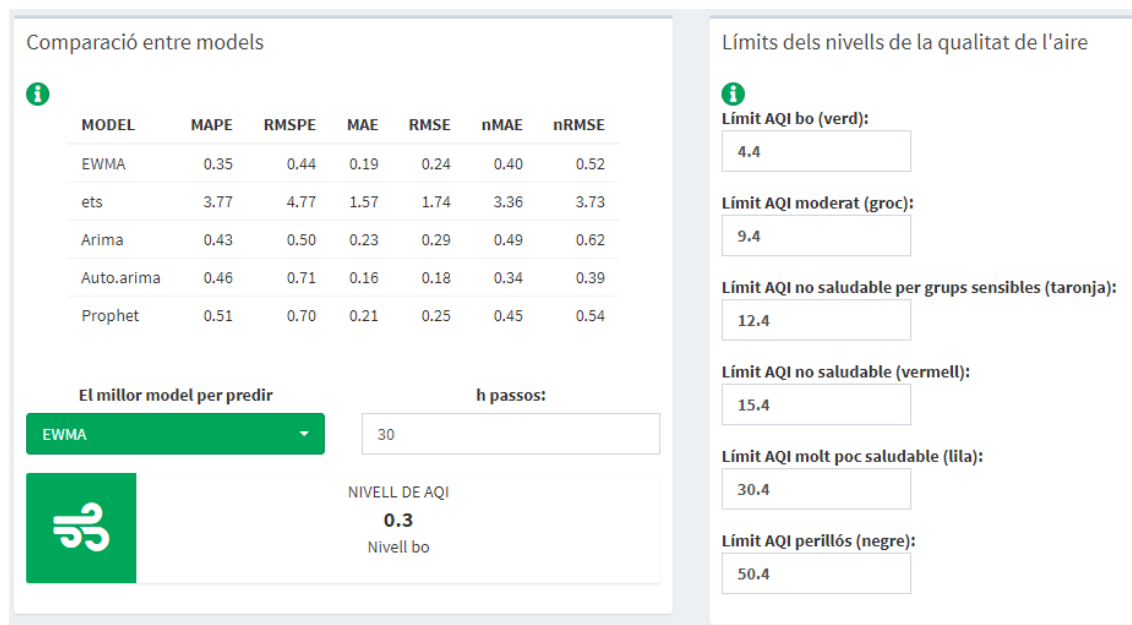
La figura 6.2 mostra com el menor resultat per la mètrica RMPSE es presenta pel model plantejat per la metodologia Box-Jenkins, el model SARIMA. Aquesta indicador el sosté el MAPE.

Escollint aquest model com l'ideal per la sèrie temporal que recull les concentracions de les partícules en suspensió menors a 2.5 micres, es fa una predicció als 30 dies.

A la figura 6.3 apareix per l'indicador RMSPE amb els menors valors els models Arima i el creat automàticament per l'Auto.arima. Per decidir quin dels dos mètodes és millor s'analitzarà la mètrica MAPE. D'aquesta manera, amb una petita diferència es determina que el model proposat manualment és millor al plantejat de manera automàtica pel contaminant PM_{10} .


 Figura 6.3: Millor model pel PM_{10}

Per l'òxid de carboni (6.4), es presenta amb menor valor en les mètriques principals el model estimat segons la metodologia d'allisat exponencial plantejat manualment. Aquest tipus de metodologia s'adapta molt bé a les sèries que no presenten estacionalitats i tendències.


 Figura 6.4: Millor model pel CO

Finalment, per l'ozó torna a ocórrer un empat dins de la mètrica RMPSE (6.5). De la mateixa manera que s'ha fet anteriorment s'estudiarà també l'indicador MAPE per decidir quin model és millor per fer les prediccions. Per una diferència mínima, el model proporcionat amb la identificació del tipus de sèrie temporal segons la

metodologia d'allisat exponencial supera el proposat automàticament pel programa.

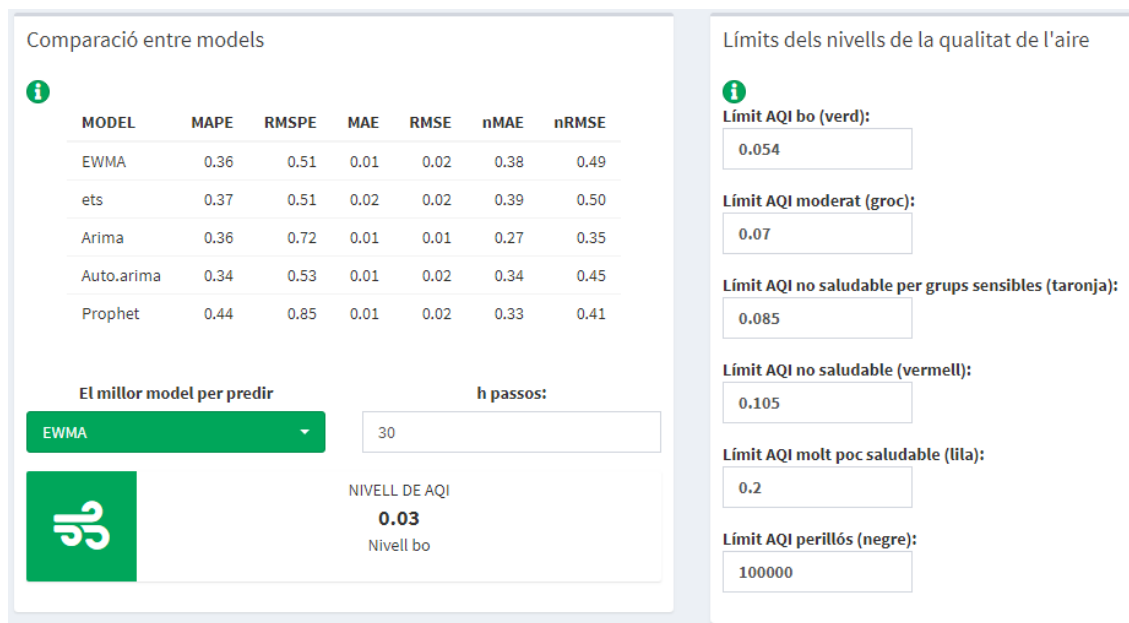


Figura 6.5: Millor model per l'O₃

Capítol 7

Conclusions

Tancant aquest cicle d'investigació s'ha pogut complir els objectius principals de l'estudi. Per una part, s'ha completat el disseny d'una plataforma interactiva en *shiny* per facilitar i agilitzar la modelització de qualsevol tipus de sèrie temporal aplicant algunes de les principals metodologies. I, per altra banda, s'ha pogut contribuir en la investigació de models explicatius i amb un percentatge de precisió satisfactori per a les prediccions de les concentracions de contaminants de l'aire a l'estat de Califòrnia.

Complint el pla definit a l'inici del projecte, finalment s'ha dissenyat una interfície web que facilita el procés de modelització de dades temporals segons les metodologies d'allisat exponencial, Box-Jenkins i *prophet* amb un disseny de *dashboard*. Dins d'aquesta aplicació s'incorpora totes les funcions, gràfics i taules necessaris per a la modelització de les sèries temporals. L'usuari pot escollir les opcions de manera interactiva. A més, facilita diferents indicadors per determinar quin de tots els models que es poden plantejar és el millor per explicar les dades amb què es treballen.

El resultat d'aquest aplicatiu té un doble interès: per una banda de tipus pedagògic ja que ajuda a l'usuari a implementar les metodologies de forma interactiva i d'altra banda, de tipus pràctic, ja que permet treballar sobre qualsevol sèrie obtenint resultats d'utilitat aplicada.

Tenint en compte que l'objectiu d'anàlisi era oferir els valors futurs de la qualitat de l'aire per poder ajudar tant a la ciutadania com als centres d'investigació a prendre mesures preventives davant d'altres concentracions de mala qualitat de l'aire, amb ajuda de la plataforma web per modelitzar sèries temporals, s'ha pogut fer un estudi de com funciona i com s'ha pogut determinar quin és el millor model per cada contaminant de l'aire de Califòrnia.

Recordant les hipòtesis definides a l'inici de l'estudi, amb les diferents anàlisis s'ha comprovat que efectivament, les sèries temporals que recullen el registre de concentració dels diferents contaminants a Califòrnia, no presentaven una tendència significativa, però sí que es percebien oscil·lacions causades per la volatilitat. Pels contaminants $PM_{2.5}$, CO i NO_2 s'han vist de manera clara la presència d'un patró estacional setmanal on es presenta major concentració de contaminació durant els dies feiners. En canvi, pels contaminants PM_{10} i O_3 no s'ha detectat cap mena d'estacionalitat significativa amb les dades registrades durant l'any 2021.

Dut a terme la identificació, estimació i predicció de cada un dels models de les diferents metodologies per cada contaminant, s'ha estudiat el grau de precisió de cada un dels models a l'hora de fer les prediccions mitjançant les mètriques MAPE i RMSPE. D'aquesta manera, s'ha determinat que les partícules pesants s'han ajustat millor amb un model SARIMA plantejat segons la metodologia Box-Jenkins i el monòxid de carboni i l'ozó, s'expliquen i prediuen millor amb el model proporcionat per la metodologia d'allisat exponencial.

Per acabar amb la demostració de l'ús de la plataforma en dades mediambientals, concretament per determinar la qualitat de l'aire, s'ha afegit cada un dels límits oficials que determinen si la qualitat de l'aire és bona o no i s'ha fet una predicció de trenta dies. D'aquesta manera s'ha vist que fent les prediccions amb els models òptims per cada contaminant han resultat amb una concentració de contaminació futura dins de l'interval considerat bona qualitat. D'aquesta manera, aquestes prediccions poden ser la base d'un mecanisme d'alarma davant d'un increment de la pol·lució que serveixi per avançar mesures preventives.

Veient aquest exemple detallat dins de l'informe, es pot percebre el potencial d'aquesta aplicació pel món acadèmic (aquesta interfície es pot proporcionar per dur a terme de manera àgil la modelització de diferents sèries temporals), pel món de la investigació (proporcionant aquesta plataforma es pot adjuntar a diferents recerques per modelitzar de manera ràpida les sèries temporals amb diferents característiques i obtenir així prediccions quasi a l'instant) i en diferents empreses (perquè es pot implementar en processos interns i fer prediccions de les vendes, els ingressos, etc.).

Finalment, les projeccions pel futur d'aquest treball de recerca és la creació de diverses actualitzacions on s'implementaran dins de la plataforma el tractament de valors atípics, l'addició de metodologies pel tractament de sèries temporals com *Random Forest Regression* i la implementació de la lectura de dades des de la web (*webscraping*).

Cal dir que, a través d'aquest treball, durant aquests mesos, no només s'ha

investigat sobre un dels temes més importants que afecten cada dia com és la contaminació sinó que també s'ha millorat les habilitats de programació creant una pàgina web interactiva des de zero. A més, tot aquest informe ha estat redactat amb llenguatge *LaTeX* per adquirir noves habilitats i destreses. Ha estat un camí d'aprenentatge constant i amb moltes novetats però sempre amb la base teòrica obtinguda en el grau d'estadística.

En particular, el desenvolupament d'aquest treball m'ha permès formar-me en metodologies que no he vist anteriorment en el grau, com és el cas de la tècnica de predicció *Prophet* i en entorns de programació també innovadors com el *Shiny*.

Bibliografia

- BOX, GEP; PIERCE, D, [s. d.]. DA 1970. Distribution 01 residual autocorrelations in autoregressive integrated moving average time series models. *Journal*. Vol. 1, pàg. 1509-1524.
- BREUSCH, Trevor S; PAGAN, Adrian R, 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the econometric society*, pàg. 1287-1294.
- BROCKWELL, Peter J; DAVIS, Richard A, 2002. *Introduction to time series and forecasting*. Springer.
- Competition: NASA Airathon: Predict Air Quality (Particulate Track)*, [s. d.] **online**[cons. 2022-06-20]. Disp. a: <https://www.drivendata.org/competitions/88/competition-air-quality-pm/>.
- CROMWELL, Jeff B; TERRAZA, Michel, 1994. *Multivariate tests for time series models*. Sage. Núm. 100.
- Download Daily Data — US EPA*, [s. d.] **online**[cons. 2022-06-21]. Disp. a: <https://www.epa.gov/outdoor-air-quality-data/download-daily-data>.
- DURBIN, James; WATSON, Geoffrey S, 1950. Testing for serial correlation in least squares regression: I. *Biometrika*. Vol. 37, núm. 3/4, pàg. 409-428.
- Forecasting PM2.5 with forecast and prophet — R-bloggers*, [s. d.] **online**[cons. 2022-06-23]. Disp. a: <https://www.r-bloggers.com/2018/02/forecasting-pm2-5-with-forecast-and-prophet/>.
- GARDNER, G; HANEY, A; PHILIPS, G, 1980. Exact Maximum Likelihood Estimation of Autoregressive-Moving Average Models Average Models by Means of Kalman Filtering. *Applied Statistics*. Vol. 29, pàg. 311-322.
- GROSS, Juergen; LIGGES, Uwe; LIGGES, Maintainer Uwe, 2015. Package ‘nortest’. *Five omnibus tests for testing the composite hypothesis of normality*.
- History — California Air Resources Board*, [s. d.] **online**[cons. 2022-06-20]. Disp. a: <https://ww2.arb.ca.gov/es/about/history>.
- HOLT, Charles C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*. Vol. 20, núm. 1, pàg. 5-10.

- ISSN 0169-2070. Disp. a DOI: <https://doi.org/10.1016/j.ijforecast.2003.09.015>.
- HOTHORN, Torsten; ZEILEIS, Achim; FAREBROTHER, Richard W; CUMMINS, Clint; MILLO, Giovanni; MITCHELL, David; ZEILEIS, Maintainer Achim, 2015. Package 'lmtest'. *Testing linear regression models*. <https://cran.r-project.org/web/packages/lmtest/lmtest.pdf>. Accessed. Vol. 6.
- How is the AQI calculated? — US EPA*, [s. d.] **online**[cons. 2022-06-21]. Disp. a: <https://www.epa.gov/outdoor-air-quality-data/how-aqi-calculated>.
- HYNDMAN, Rob J; KHANDAKAR, Yeasmin, 2008. Automatic time series forecasting: the forecast package for R. *Journal of statistical software*. Vol. 27, pàg. 1-22.
- HYNDMAN, Rob J; KOEHLER, Anne B; SNYDER, Ralph D; GROSE, Simone, 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*. Vol. 18, núm. 3, pàg. 439-454. ISSN 0169-2070. Disp. a DOI: [https://doi.org/10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8).
- JOHANNESSON, Tomas; BJORNSSON, Halldor; GROTHENDIECK, Gabor; JOHANNESSON, Maintainer Tomas, 2009. Package 'stinepack'. *Icelandic Meteorological Office*.
- LinkedIn Learning: Online Courses for Creative, Technology, Business Skills*, [s. d.] **online**[cons. 2022-06-23]. Disp. a: <https://www.linkedin.com/learning/?u=85379114>.
- MORITZ, Steffen; BARTZ-BEIELSTEIN, Thomas, 2017. imputeTS: time series missing value imputation in R. *R J*. Vol. 9, núm. 1, pàg. 207.
- Principals contaminants atmosfèrics. Agència de Salut Pública de Catalunya (ASP-CAT)*, [s. d.] **online**[cons. 2022-06-21]. Disp. a: https://salutpublica.gencat.cat/ca/ambits/proteccio_salut/aire/Principals-contaminants/.
- Quick Start — Prophet*, [s. d.] **online**[cons. 2022-06-20]. Disp. a: https://facebook.github.io/prophet/docs/quick_start.html.
- QUINN, BG, 1982. Stationarity and invertibility of simple bilinear models. *Stochastic Processes and their Applications*. Vol. 12, núm. 2, pàg. 225-230.
- ROYSTON, J Patrick, 1982. An extension of Shapiro and Wilk's W test for normality to large samples. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. Vol. 31, núm. 2, pàg. 115-124.
- Shiny*, [s. d.] **online**[cons. 2022-06-23]. Disp. a: <https://shiny.rstudio.com/>.
- Shiny - Widget Gallery*, [s. d.] **online**[cons. 2022-06-23]. Disp. a: <https://shiny.rstudio.com/gallery/widget-gallery.html>.
- Shiny Dashboard*, [s. d.] **online**[cons. 2022-06-23]. Disp. a: <https://rstudio.github.io/shinydashboard/index.html>.

- SIEVERT, Carson, 2020. *Interactive web-based data visualization with R, plotly, and shiny*. CRC Press.
- Stack Overflow en español*, [s. d.] **online**[cons. 2022-06-23]. Disp. a: <https://es.stackoverflow.com/>.
- TAYLOR, Sean J; LETHAM, Benjamin, 2018. Forecasting at scale. *The American Statistician*. Vol. 72, núm. 1, pàg. 37-45.
- THODE, H, 2002. Testing for normality marcel dekker. *Inc. New York*, pàg. 99-123.
- TRAPLETTI, Adrian; HORNIK, Kurt; LEBARON, Blake; HORNIK, Maintainer Kurt, 2022. *Package 'tseries'*.
- WICKHAM, Hadley; GROLEMUND, Garrett, 2016. *R for data science: import, tidy, transform, visualize, and model data.* ” O'Reilly Media, Inc.”.
- WICKHAM, Hadley; WICKHAM, Maintainer Hadley, 2017. Package tidyverse. *Easily Install and Load the 'Tidyverse'*.

Apèndix A

Anàlisi dels contaminants estudiats

En aquest apartat s'han afegit totes les sortides utilitzades per decidir els models utilitzats dins del resum d'anàlisi dels contaminants.

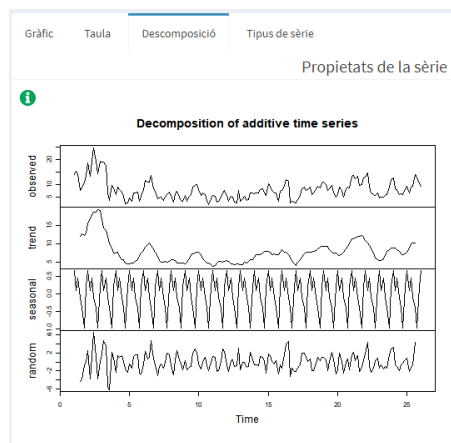


Figura A.1: Components del $PM_{2.5}$

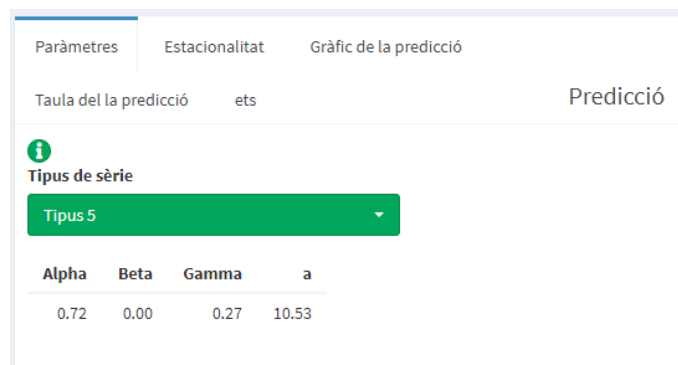


Figura A.2: Estimació dels paràmetres $PM_{2.5}$

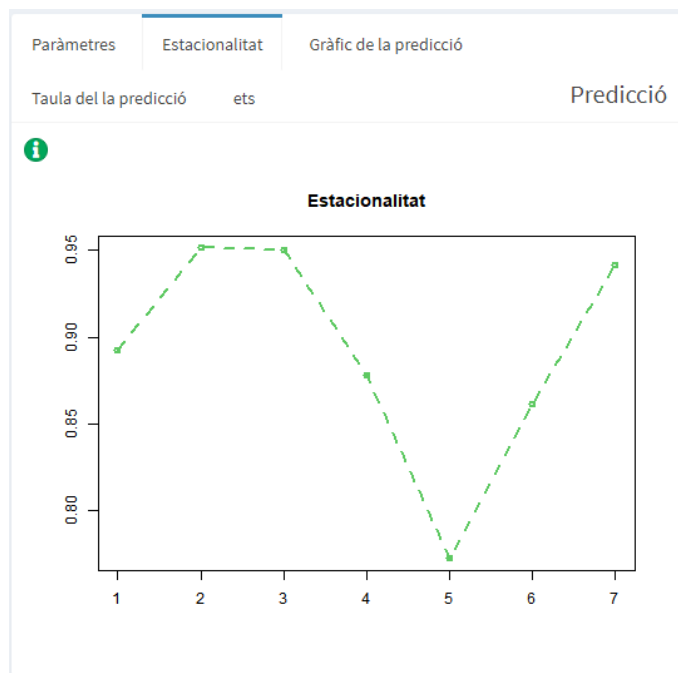


Figura A.3: Patró estacional $PM_{2.5}$

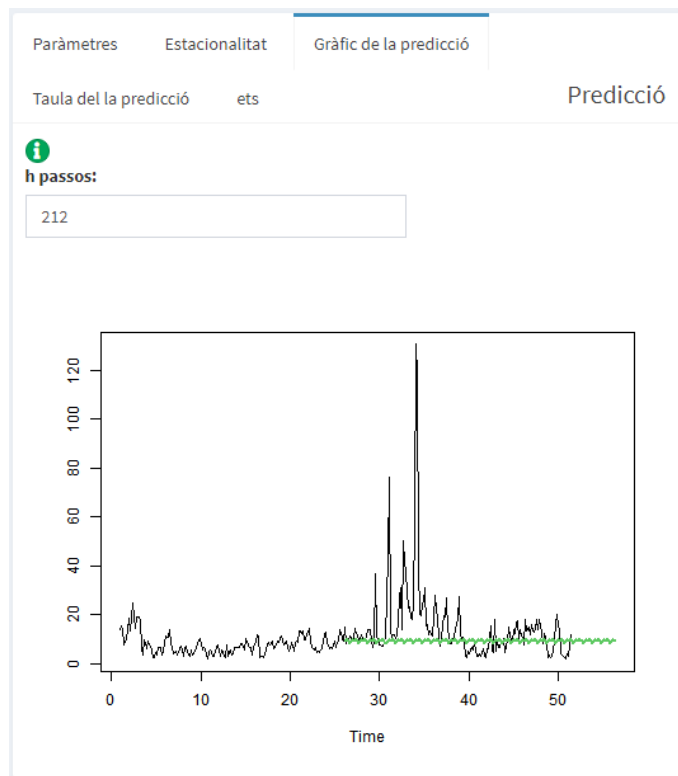


Figura A.4: Predicció $PM_{2.5}$

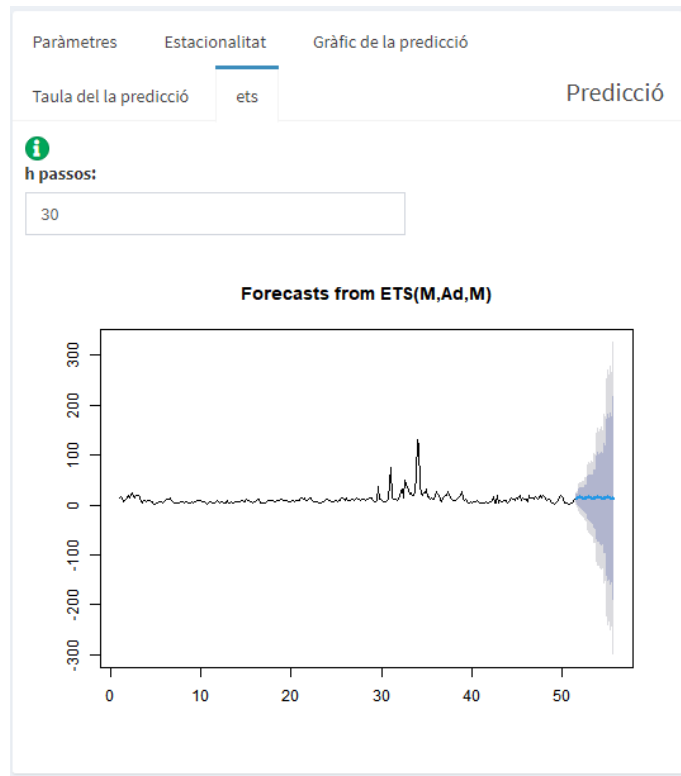


Figura A.5: ETS $PM_{2.5}$

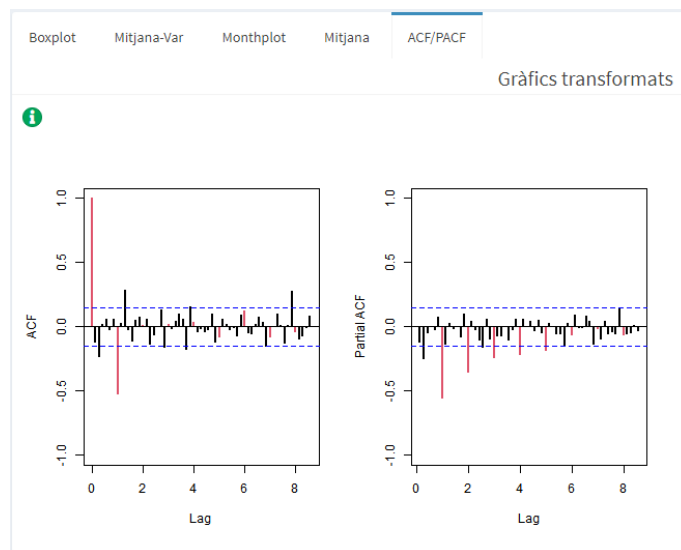


Figura A.6: ACF i PACF $PM_{2.5}$

Estimació A	Estimació B	Model	
<p>i Model utilitzant la sèrie no estacionària. Els T-ratios són significatius si són majors a 2:</p>			
	ar1	ma1	sma1
Coefficients	0.6438	-0.9266	-1.0000
S.E	0.0766	0.0390	0.0865
T-ratios	8.4092	23.7305	11.5667
Sigma2	Versemblança		AIC
0.1123	-67.0500		142.1000

Figura A.7: Estimació del model pel $PM_{2.5}$

Normalitat: Shapiro-Wilks		Homocedasticitat: Breusch-Pagan		Independència: Durbin-Watson		Independència: Ljung-Box		
W	p_valor	BP	p_valor	DW	p_valor	lag.df	statistic	p.value
0.98257	0.02678	3.25201	0.07134	1.93197	0.29786	1.00000	0.26108	0.60938
Normalitat: Anderson-Darling						2.00000	2.96981	0.22652
A	p_valor					3.00000	3.37849	0.33687
0.58319	0.12720					4.00000	3.54083	0.47170
Normalitat: Jarque-Bera						7.00000	7.59402	0.36975
X_Squared	p_valor					14.00000	16.03279	0.31138
10.70987	0.00472					21.00000	21.93569	0.40322
						28.00000	30.60759	0.33478

Figura A.8: Test numèrics per la validació del $PM_{2.5}$

Sèrie tallada fins l'observació 189 o l'escollit en l'imporació de les dades:				Sèrie completa:			
ar1	ma1	sma1		ar1	ma1	sma1	
Coefficients	0.6438	-0.9266	-1.0000	Coefficients	0.6424	-0.9142	-1.0000
S.E	0.0766	0.0390	0.0865	S.E	0.0612	0.0330	0.0386
T-ratios	8.4092	23.7305	11.5667	T-ratios	10.4944	27.6719	25.9392
Sigma2	Versemblança		AIC	Sigma2	Versemblança		AIC
0.1123	-67.0500		142.1000	0.1564	-184.6305		377.2611

Figura A.9: Estabilitat del model pel $PM_{2.5}$

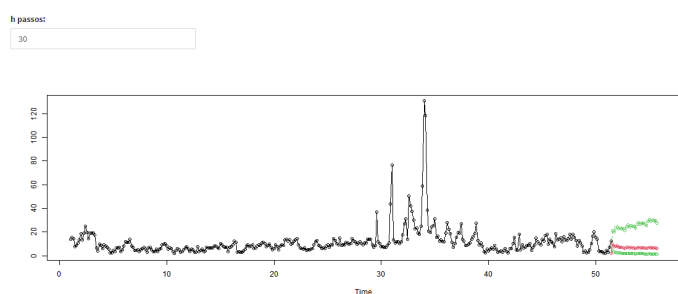


Figura A.10: Predicció $PM_{2.5}$

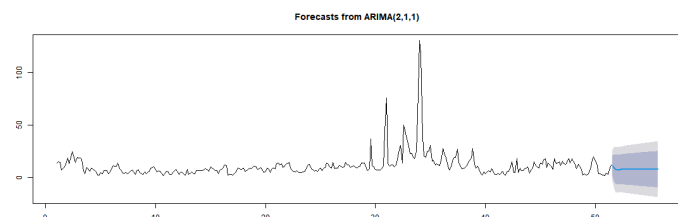


Figura A.11: Auto.arima $PM_{2.5}$

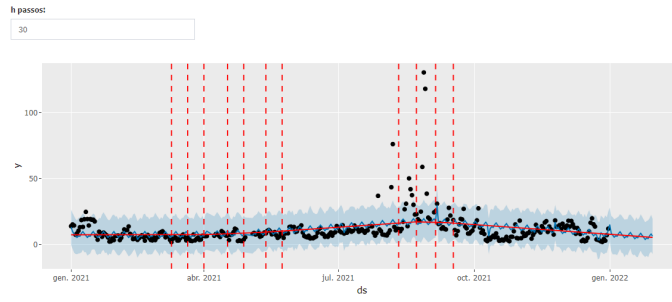


Figura A.12: Predicció segons prophet $PM_{2.5}$

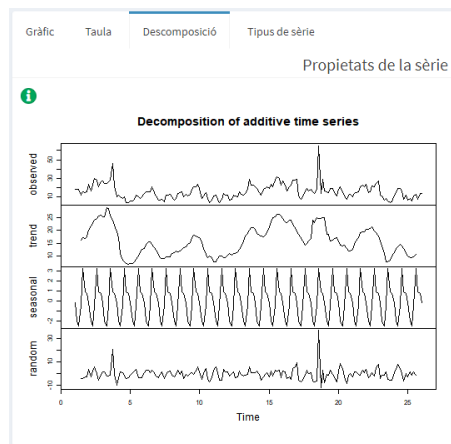


Figura A.13: Components del PM_{10}

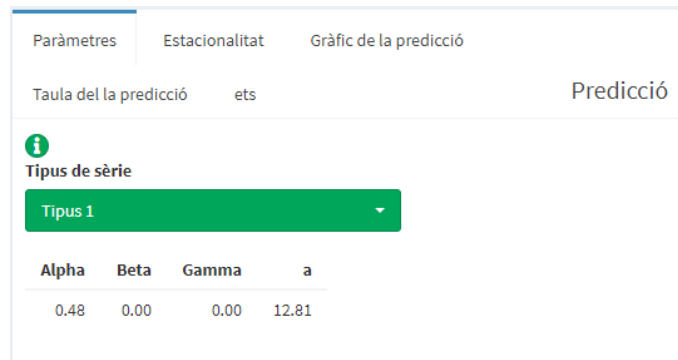


Figura A.14: Estimació dels paràmetres PM_{10}

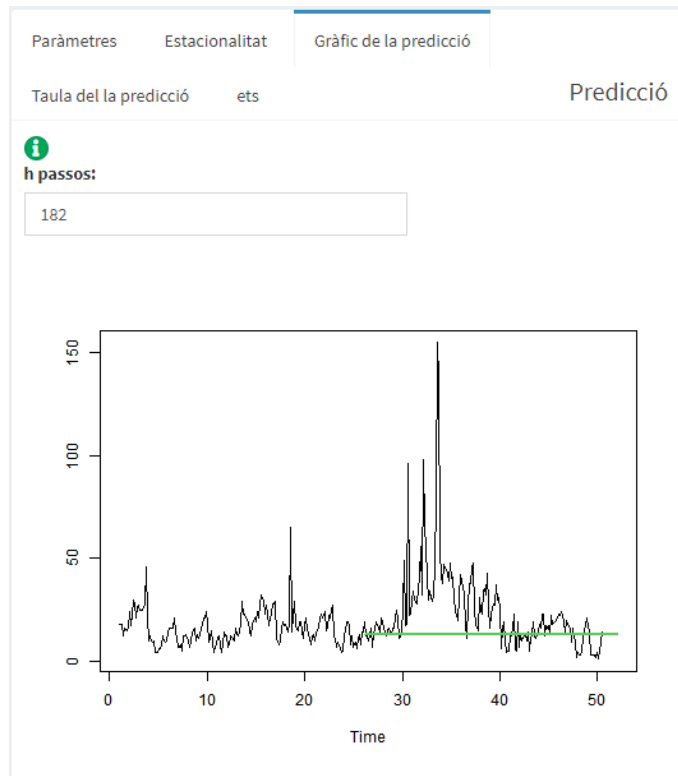


Figura A.15: Predicció PM_{10}

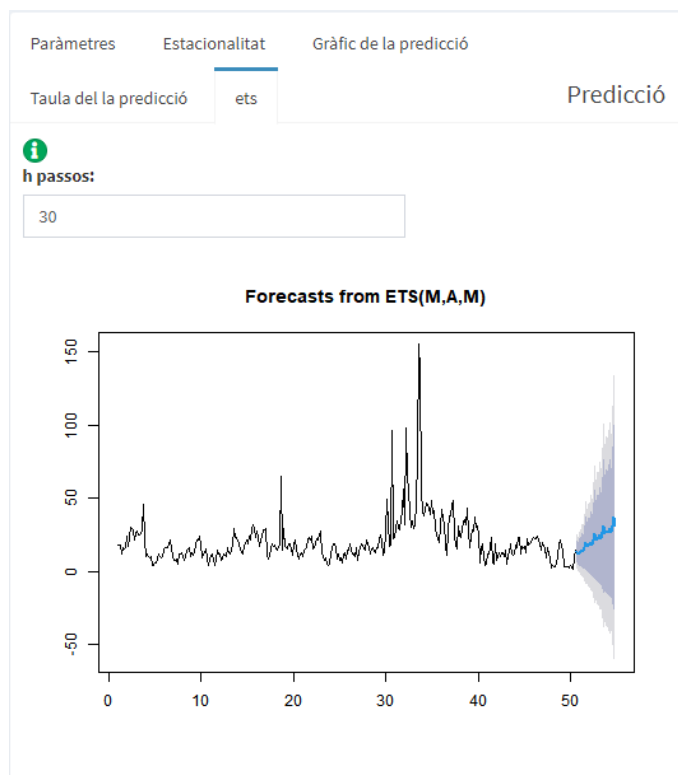


Figura A.16: ETS PM_{10}

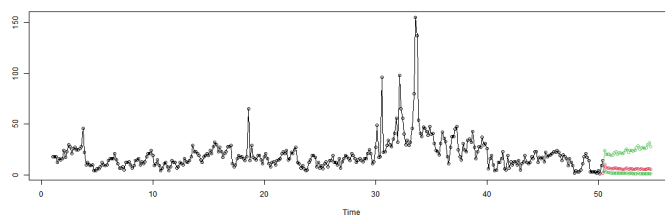


Figura A.21: Predicció PM_{10}

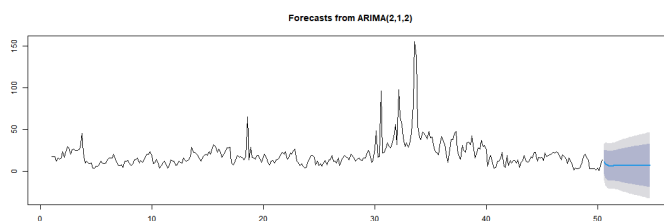


Figura A.22: Auto.arima PM_{10}

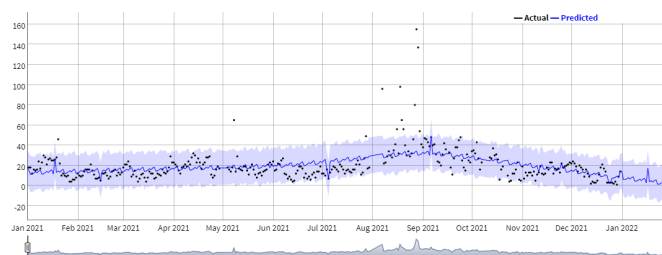


Figura A.23: Predicció segons prophet PM_{10}

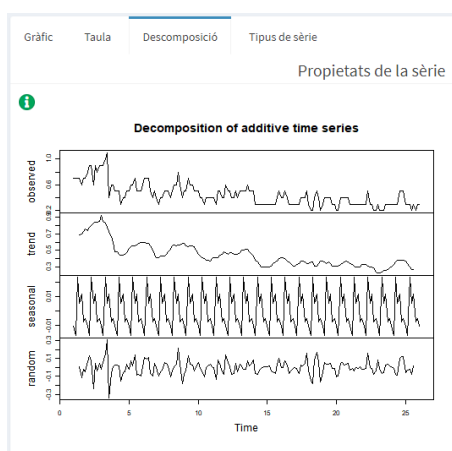


Figura A.24: Components del CO



Figura A.25: Estimació dels paràmetres CO

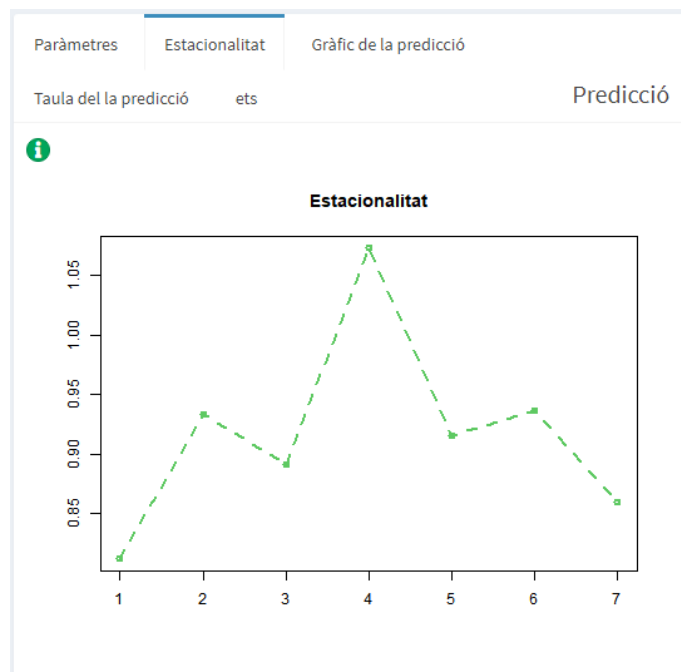


Figura A.26: Patró estacional CO

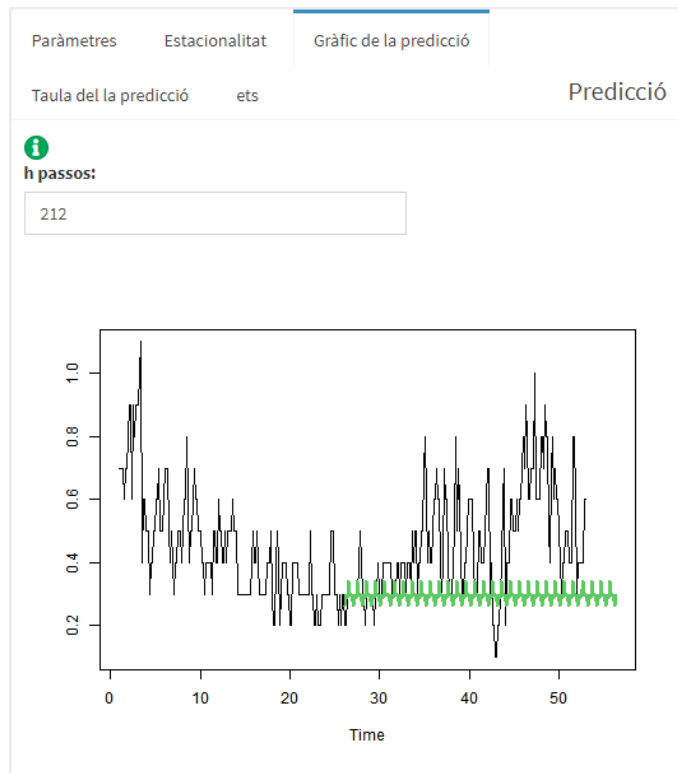


Figura A.27: Predicció CO

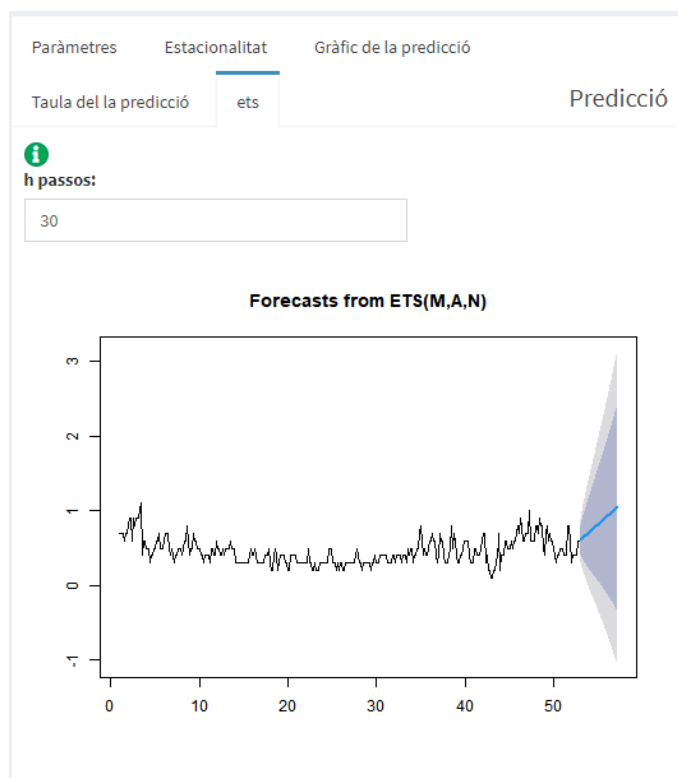


Figura A.28: ETS CO

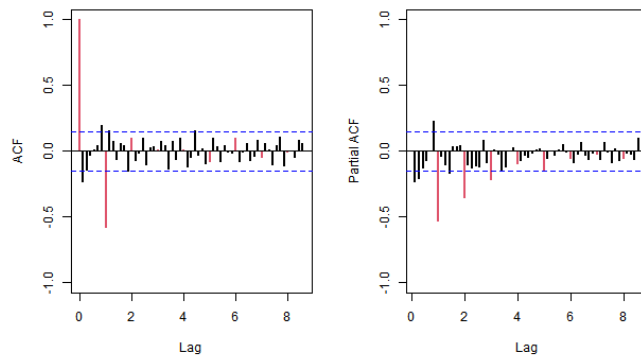


Figura A.29: ACF i PACF CO

Sèrie tallada fins l'observació 189 o l'escollit en l'importació de les dades:							Sèrie completa:						
	ar1	ar2	ar3	sma1	sma2	sma3		ar1	ar2	ar3	sma1	sma2	sma3
Coefficients	-0.2460	-0.2329	-0.1859	-1.0624	0.2397	-0.0980	Coefficients	-0.2074	-0.1315	-0.1451	-1.0205	0.0325	-0.0120
S.E	0.0762	0.0766	0.0756	0.0906	0.1278	0.0895	S.E	0.0525	0.0531	0.0525	0.0742	0.0704	0.0535
T-ratios	3.2265	3.0426	2.4577	11.7038	1.8756	1.0944	T-ratios	3.9474	2.4740	2.7664	13.7524	0.4612	0.2245
Sigma2	Versemblança		AIC				Sigma2	Versemblança		AIC			
0.0538	0.0471		13.9058				0.0641	-29.9932		73.9863			

Figura A.30: Estimació del model pel CO

Normalitat: Shapiro-Wilks		Homocedasticitat: Breusch-Pagan		Independència: Durbin-Watson		Independència: Ljung-Box			
W	p_valor	BP	p_valor	DW	p_valor	lag,df	statistic	p.value	
0.97550	0.00337	32.95800	0.00000	2.05256	0.60796	1.00000	0.11583	0.73360	
Normalitat: Anderson-Darling									
A	p_valor								
1.70699	0.00022								
Normalitat: Jarque-Bera									
X_Squared	p_valor								
9.67583	0.00792								
							2.00000	0.55983	0.75585
							3.00000	1.08354	0.78105
							4.00000	4.94461	0.29303
							7.00000	6.00389	0.53930
							14.00000	10.94101	0.69066
							21.00000	14.96793	0.82456
							28.00000	16.43835	0.95887

Figura A.31: Test numèrics per la validació del CO

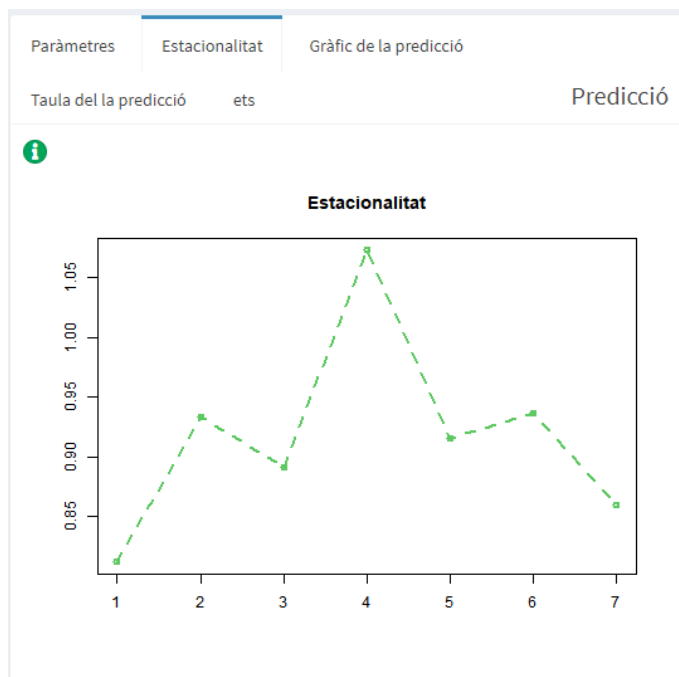


Figura A.32: Estabilitat del model pel CO

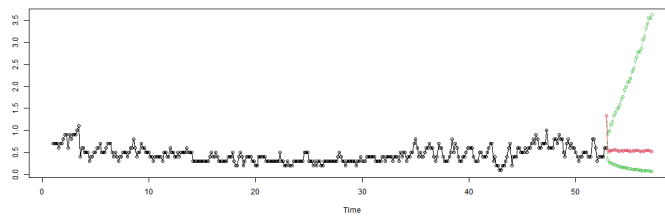


Figura A.33: Predicció CO

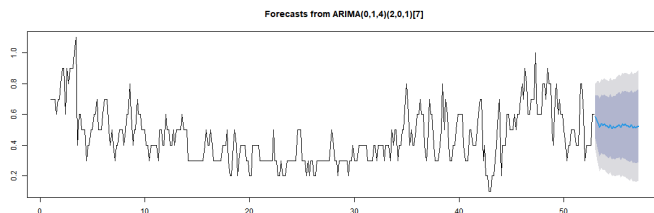


Figura A.34: Auto.arima CO

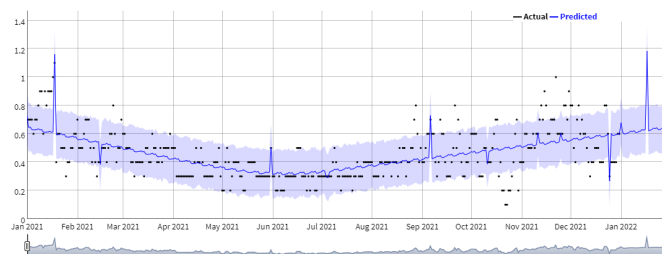


Figura A.35: Predicció segons prophet CO

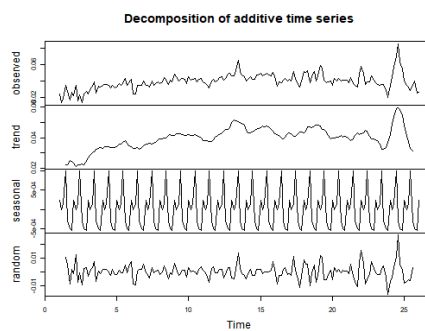


Figura A.36: Components del O_3

Paràmetres Estacionalitat Gràfic de la predicció

Taula del la predicció ets Predicció

i
Tipus de sèrie

Tipus 1

Alpha	Beta	Gamma	a
0.81	0.00	0.00	0.03

Figura A.37: Estimació dels paràmetres O_3

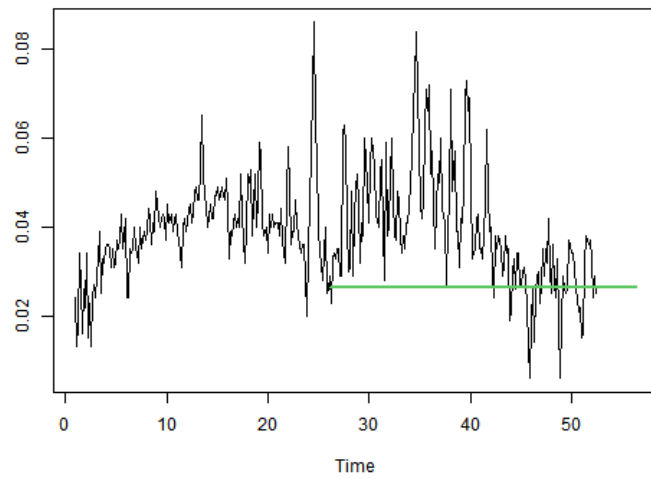


Figura A.38: Predicció O_3

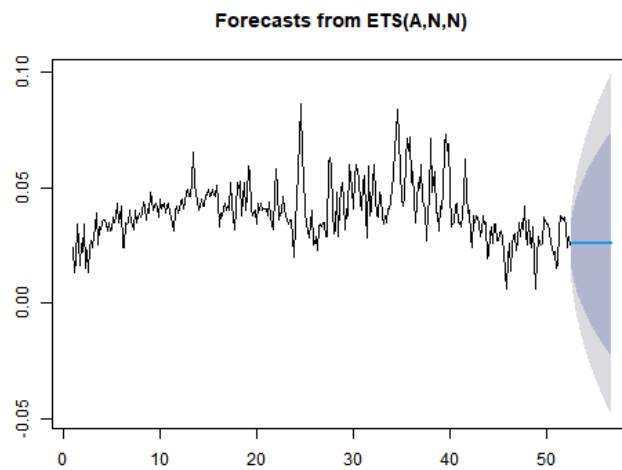


Figura A.39: ETS O_3

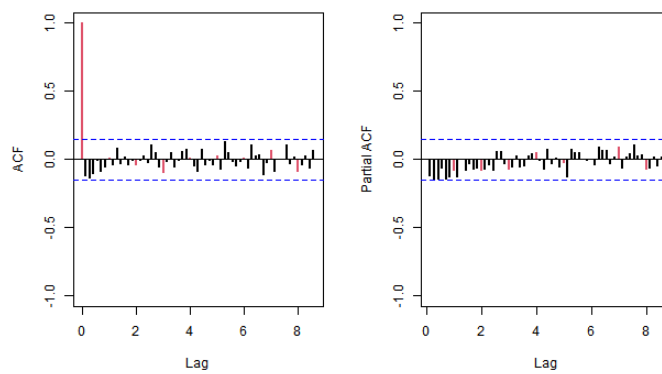


Figura A.40: ACF i PACF O_3

Estimació A	Estimació B	Model	
<p>i Model utilitzant la sèrie no estacionària. Els T-ratios són significatius si són majors a 2:</p>			
	ar1	ma1	
Coefficients	0.6336	-0.9433	
S.E	0.0701	0.0270	
T-ratios	9.0400	34.9620	
Sigma2	Versemblança	AIC	
0.0000	632.6373	-1259.2746	

Figura A.41: Estimació del model pel O_3

Normalitat: Shapiro-Wilks		Homocedasticitat: Breusch-Pagan		Independència: Durbin-Watson		Independència: Ljung-Box		
W	p_valor	BP	p_valor	DW	p_valor	lag,df	statistic	p.value
0.98511	0.05810	4.34255	0.03717	1.94725	0.33411	1.00000	0.29264	0.58853
Normalitat: Anderson-Darling						2.00000	0.61544	0.73512
A	p_valor					3.00000	1.09371	0.77859
0.69031	0.07027					4.00000	1.10633	0.89327
Normalitat: Jarque-Bera						7.00000	3.13327	0.87241
X_Squared	p_valor					14.00000	5.11051	0.98421
8.54552	0.01394					21.00000	11.03192	0.96216
						28.00000	13.93230	0.98766

Figura A.42: Test numèrics per la validació del O_3

Sèrie tallada fins l'observació 189 o l'escollit en l'imporació de les dades:				Sèrie completa:			
	ar1	ma1			ar1	ma1	
Coefficients	0.6336	-0.9433		Coefficients	0.6129	-0.9390	
S.E	0.0701	0.0270		S.E	0.0499	0.0191	
T-ratios	9.0400	34.9620		T-ratios	12.2842	49.1031	
Sigma2	Versemblança	AIC		Sigma2	Versemblança	AIC	
0.0000	632.6373	-1259.2746		0.0001	1237.4163	-2468.8326	

Figura A.43: Estabilitat del model pel O_3

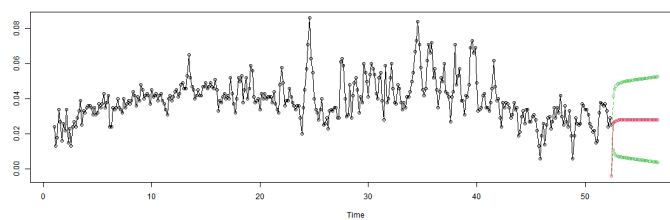


Figura A.44: Predicció O_3

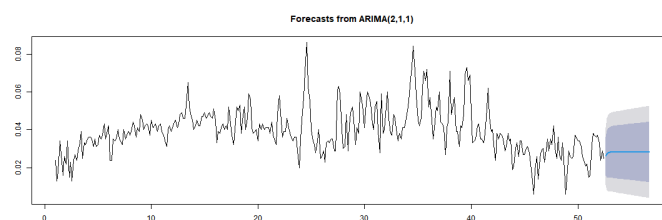


Figura A.45: Auto.arima O_3

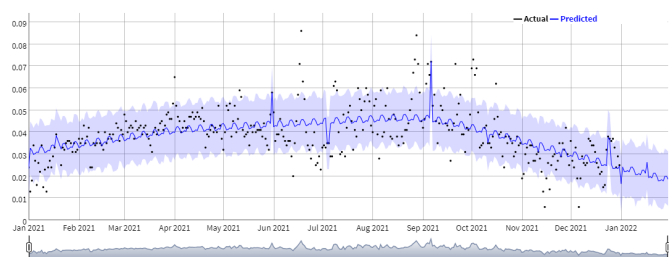


Figura A.46: Predicció segons prophet O_3

Apèndix B

Codi Shiny

B.1 Server

```
1 #
2 # This is the server logic of a Shiny web application. You can run
  the
3 # application by clicking 'Run App' above.
4 #
5 # Find out more about building applications with Shiny here:
6 #
7 #   http://shiny.rstudio.com/
8 #
9
10 library(shiny)
11 library(shinydashboard)
12 library(plotly)
13
14 function(input, output, session) {
15   library(tidyverse)
16
17   # Importaci de dades
18   dades <- reactive({
19     inFile <- input$file1
20     if (is.null(inFile)){
21       return(NULL)
22     }else{
23       read.csv(inFile$datapath, header=input$Cap alera, sep=input$
24       Separador)
25     }
26   })
27   dadescomp <- eventReactive(dades(), {
```

```

28   dades() %>% mutate(ds=as.Date(ds, format = input$formatdata))
    %>%
29     mutate(y=as.numeric(y))
30   })
31
32
33   output$dadestaula <- renderTable({
34     head(dades())
35   })
36
37
38   output$slider <- renderUI({
39     if(is.null(dades())){
40       cat("")
41     }else{
42       sliderInput("slider", "Per ode d'entrenament", min = 1, max =
43         nrow(dades()),
44         value = 1)
45     }
46   })
47
48   # Funci per crear la s rie completa
49   crearserie <- function(serie){
50     serie = serie[order(serie$Date, serie[,2]),]
51     serie = serie[!duplicated(serie$Date),]
52     serie <- serie %>% mutate(serie, Date=as.Date(Date, format = "%m
53     /%d/%Y"))
54     return(serie)
55   }
56
57   # California PM2.5
58   #PM25_2018 <- read_csv("D:/TFG/dades/2018 California PM25.csv")
59   #PM25_2019 <- read_csv("D:/TFG/dades/2019 California PM25.csv")
60   #PM25_2020 <- read_csv("D:/TFG/dades/2020 California PM25.csv")
61   #PM25_2022 <- read_csv("D:/TFG/dades/2022 California PM25.csv")
62   PM25serie <- bind_rows(PM25_2021)
63   PM25serie <- PM25serie %>% select(Date, 'Daily Mean PM2.5
64     Concentration')
65   PM25serie <- crearserie(PM25serie)
66   PM25serie <- rename(PM25serie, Temps = Date,
67     Concentraci = 'Daily Mean PM2.5
68     Concentration')
69   PM25serie <- mutate(PM25serie, Concentraci =
70     replace(Concentraci, which(Concentraci
71     <0), NA))

```

```

69 PM25serie <- PM25serie[order(PM25serie$Temps),]
70
71 # California CO
72 #CO_2018 <- read_csv("D:/TFG/dades/2018 California CO.csv")
73 #CO_2019 <- read_csv("D:/TFG/dades/2019 California CO.csv")
74 #CO_2020 <- read_csv("D:/TFG/dades/2020 California CO.csv")
75 CO_2021 <- read_csv("D:/TFG/dades/2021 California CO.csv")
76 COserie <- bind_rows(CO_2021)
77 COserie <- COserie %>% select(Date, 'Daily Max 8-hour CO
    Concentration')
78 COserie <- crearserie(COserie)
79 COserie <- rename(COserie, Temps = Date,
80                   Concentraci = 'Daily Max 8-hour CO
    Concentration')
81 COserie <- mutate(COserie, Concentraci =
82                   replace(Concentraci , which(Concentraci
    <0), NA))
83 COserie <- COserie[order(COserie$Temps),]
84
85 # California NO2
86 #NO2_2018 <- read_csv("D:/TFG/dades/2018 California NO2.csv")
87 #NO2_2019 <- read_csv("D:/TFG/dades/2019 California NO2.csv")
88 #NO2_2020 <- read_csv("D:/TFG/dades/2020 California NO2.csv")
89 NO2_2021 <- read_csv("D:/TFG/dades/2021 California NO2.csv")
90 NO2serie <- bind_rows(NO2_2021)
91 NO2serie <- NO2serie %>% select(Date, 'Daily Max 1-hour NO2
    Concentration')
92 NO2serie <- crearserie(NO2serie)
93 NO2serie <- rename(NO2serie, Temps = Date,
94                   Concentraci = 'Daily Max 1-hour NO2
    Concentration')
95 NO2serie <- mutate(NO2serie, Concentraci =
96                   replace(Concentraci , which(Concentraci <0)
    , NA))
97 NO2serie <- NO2serie[order(NO2serie$Temps),]
98
99 # California ozone
100 #ozone_2018 <- read_csv("D:/TFG/dades/2018 California Ozone.csv")
101 #ozone_2019 <- read_csv("D:/TFG/dades/2019 California Ozone.csv")
102 #ozone_2020 <- read_csv("D:/TFG/dades/2020 California Ozone.csv")
103 ozone_2021 <- read_csv("D:/TFG/dades/2021 California Ozone.csv")
104 #ozone_2022 <- read_csv("D:/TFG/dades/2022 California Ozone.csv")
105 ozoneserie <- bind_rows(ozone_2021)
106 ozoneserie <- ozoneserie %>% select(Date,
107                                     'Daily Max 8-hour Ozone
    Concentration')
108 ozoneserie <- crearserie(ozoneserie)

```

```

109 ozoneserie <- rename(ozoneserie, Temps = Date,
110                      Concentraci = 'Daily Max 8-hour Ozone
      Concentration')
111 ozoneserie <- mutate(ozoneserie, Concentraci =
112                      replace(Concentraci, which(Concentraci
      <0), NA))
113 ozoneserie <- ozoneserie[order(ozoneserie$Temps),]
114
115 # California Pb
116 # Pb_2018 <- read_csv("D:/TFG/dades/2018 California Pb.csv")
117 # Pb_2019 <- read_csv("D:/TFG/dades/2019 California Pb.csv")
118 # Pb_2020 <- read_csv("D:/TFG/dades/2020 California Pb.csv")
119 # Pb_2021 <- read_csv("D:/TFG/dades/2021 California Pb.csv")
120 # Pb_2022 <- read_csv("D:/TFG/dades/2022 California Pb.csv")
121
122 # California PM10
123 #PM10_2018 <- read_csv("D:/TFG/dades/2018 California PM10.csv")
124 #PM10_2019 <- read_csv("D:/TFG/dades/2019 California PM10.csv")
125 #PM10_2020 <- read_csv("D:/TFG/dades/2020 California PM10.csv")
126 PM10_2021 <- read_csv("D:/TFG/dades/2021 California PM10.csv")
127 PM10serie <- bind_rows(PM10_2021)
128 PM10serie <- PM10serie %>% select(Date, 'Daily Mean PM10
      Concentration')
129 PM10serie <- crearserie(PM10serie)
130 PM10serie <- rename(PM10serie, Temps = Date,
131                      Concentraci = 'Daily Mean PM10
      Concentration')
132 PM10serie <- mutate(PM10serie, Concentraci =
133                      replace(Concentraci, which(Concentraci
      <0), NA))
134 PM10serie <- PM10serie[order(PM10serie$Temps),]
135
136
137
138 interserie <- reactive({
139   if(input$contaminant=="CO"){
140     as.data.frame(COserie)
141   }else if(input$contaminant=="NO2"){
142     as.data.frame(NO2serie)
143   }else if(input$contaminant=="Oz "){
144     as.data.frame(ozoneserie)
145   }else if(input$contaminant=="PM2.5"){
146     as.data.frame(PM25serie)
147   }else if(input$contaminant=="PM10"){
148     as.data.frame(PM10serie)
149   }else if(input$contaminant=="Dades"){
150     as.data.frame(dadescomp())

```

```

151   }
152   })
153
154   #Escala de colors
155   a <-c("#fde725", "#5ec962", "#21918c", "#3b528b", "#440154")
156
157   # Plot descriptiu
158   output$descriptiu <- renderPlotly({
159     if((input$contaminant=="CO")||(input$contaminant=="NO2")||
160       (input$contaminant=="Oz ")||(input$contaminant=="PM2.5")||
161       (input$contaminant=="PM10")){
162       interserie() %>%
163       ggplot(aes(x=Temps, y=Concentraci)) +
164       geom_area(fill=a[2], alpha=0.5) +
165       geom_line(color=a[2]) + ggtitle("Qualitat de l'aire") +
166       theme(panel.grid.major = element_line(colour = "gray94"),
167             panel.grid.minor = element_line(colour = NA),
168             axis.title = element_text(family = "serif", face = "
169 italic"),
170             axis.text = element_text(family = "serif", colour = "
171 black"),
172             plot.title = element_text(family = "serif", face = "
173 bold.italic",
174                                     hjust = 0.5),
175             panel.background = element_rect(fill = "white"))
176     }else if(input$contaminant=="Dades"){
177       interserie() %>%
178       ggplot(aes(x=ds, y=y)) +
179       geom_area(fill=a[2], alpha=0.5) + xlab("Temps") + ylab("")
180     +
181     geom_line(color=a[2]) +
182     theme(panel.grid.major = element_line(colour = "gray94"),
183           panel.grid.minor = element_line(colour = NA),
184           axis.title = element_text(family = "serif", face = "
185 italic"),
186           axis.text = element_text(family = "serif", colour = "
187 black"),
188           plot.title = element_text(family = "serif", face = "
189 bold.italic",
190                                     hjust = 0.5),
191           panel.background = element_rect(fill = "white"))
192     }
193   })
194
195   # Taula primeresobs

```

```

189 output$obs <- renderDataTable(interserie(), options = list(
    pageLength=6))
190
191 # Creaci ts s rie
192 serie <- reactive({
193   library(imputeTS)
194   if(input$contaminant=="CO"){
195     na_interpolation(window(ts(COserie$Concentraci , start = 1,
196       freq = as.numeric(as.numeric(input
197 $Diffest))),
198       end = 52/2), start = 1), option='
199 linear')
200   }else if(input$contaminant=="NO2"){
201     na_interpolation(window(ts(NO2serie$Concentraci , start = 1,
202       freq = as.numeric(as.numeric(input
203 $Diffest))),
204       end = 52/2), start = 1), option='
205 linear')
206   }else if(input$contaminant=="Oz "){
207     na_interpolation(window(ts(ozoneserie$Concentraci , start =
208 1,
209       freq = as.numeric(as.numeric(input
210 $Diffest))),
211       end = 52/2), start = 1), option='
212 linear')
213   }else if(input$contaminant=="PM2.5"){
214     na_interpolation(window(ts(PM25serie$Concentraci , start =
215 1,
216       freq = as.numeric(as.numeric(input
217 $Diffest))),
218       end = 52/2), start = 1), option='
219 linear')
220   }else if(input$contaminant=="PM10"){
221     na_interpolation(window(ts(PM10serie$Concentraci , start =
222 1,
223       freq = as.numeric(as.numeric(input
224 $Diffest))),
225       end = 52/2), start = 1), option='
226 linear')
227   }else if(input$contaminant=="Dades"){
228     na_interpolation(window(ts(dadescomp()$y, start = 1,
229       freq = as.numeric(as.numeric(input
230 $Diffest))),
231       end = c(input$slider%%as.numeric(
232 input$Diffest),
233       input$slider%%as.numeric(
234 input$Diffest))),

```

```

219         start = 1),
220         option='linear')
221     }
222 })
223
224 totalserie <- reactive({
225     library(imputeTS)
226     if(input$contaminant=="CO"){
227         na_interpolation(window(ts(COserie$Concentraci , start = 1,
228             freq = as.numeric(as.numeric(input
229                 $Diffest))),
230                 start = 1),option='linear')
231     }else if(input$contaminant=="NO2"){
232         na_interpolation(window(ts(NO2serie$Concentraci , start = 1,
233             freq = as.numeric(input$Diffest)),
234                 start = 1),option='linear')
235     }else if(input$contaminant=="Oz "){
236         na_interpolation(window(ts(ozoneserie$Concentraci , start =
237             1,
238             freq = as.numeric(input$Diffest)),
239                 start = 1), option='linear')
240     }else if(input$contaminant=="PM2.5"){
241         na_interpolation(window(ts(PM25serie$Concentraci , start =
242             1,
243             freq = as.numeric(input$Diffest)),
244                 start = 1), option='linear')
245     }else if(input$contaminant=="PM10"){
246         na_interpolation(window(ts(PM10serie$Concentraci , start =
247             1,
248             freq = as.numeric(input$Diffest)),
249                 start = 1), option='linear')
250     }
251 })
252
253 # Plot decompose
254 output$decompose <- renderPlot({
255     plot(decompose(serie()))
256 })
257
258 # Tipus de s ries
259 output$tipusdeserie <- renderPlot({
260     par(mfrow = c(2, 3))
261

```



```

262 #####
263 ##### Serie Tipo 1 #####
264 #####
265
266 tuneS1 <- function(par){
267   hw1 <- HoltWinters(serie(), alpha = par, beta = F, gamma = F)
268   mean((serie()-fitted(hw1)[, 1])^2)
269 }
270 par = 0.5
271 sol1 <- optim(par, tuneS1, method = "L-BFGS-B", lower = 1e-4,
272             upper = 0.9999)
273 hw1 <- HoltWinters(serie(), alpha = sol1$par, beta = F, gamma = F
274 )
275 phw1 <- predict(hw1, 60)
276 ts.plot(serie(), phw1, col = c(1, a[2]), main = paste0("(1) ANN
277 MSE = ",
278                                                     round(sol1
279 $value, 5)))
280 #####
281 ##### Serie Tipo 2 #####
282 #####
283 tuneS2 <- function(par){
284   hw2 <- HoltWinters(serie(), alpha = par[1], beta = par[2],
285                     gamma = F)
286   mean((serie()-fitted(hw2)[, 1])^2)
287 }
288 par = c(0.5, 0.5)
289 sol2 <- optim(par, tuneS2, method = "L-BFGS-B", lower = c(1e-4, 1
290 e-4),
291             upper = c(0.9999, 0.9999))
292 hw2 <- HoltWinters(serie(), alpha = sol2$par[1], beta = sol2$par
293 [2],
294                 gamma = F)
295 phw2 <- predict(hw2, 60)
296 ts.plot(serie(), phw2, col = c(1, a[2]), main = paste0("(2) AAN
297 MSE = ",
298                                                     round(sol2
299 $value, 5)))
300 #####
301 ##### Serie Tipo 3 #####
302 #####
303 tuneS3 <- function(par){
304   hw3 <- HoltWinters(serie(), alpha = par[1], beta = FALSE, gamma
305 = par [2])
306   mean((serie()-fitted(hw3)[, 1])^2)

```

```

300 }
301 par = c(0.5, 0.5)
302 sol3 <- optim(par, tuneS3, method = "L-BFGS-B", lower = c(1e-4, 1
    e-4),
303             upper = c(0.9999, 0.9999))
304 hw3 <- HoltWinters(serie(), alpha = sol3$par[1], beta = FALSE,
305                  gamma = sol3$par[2])
306 phw3 <- predict(hw3, 60)
307 ts.plot(serie(), phw3, col = c(1, a[2]),
308         main = paste0("(3)ANA MSE = ", round(sol3$value, 5)))
309
310 #####
311 ##### Serie Tipo 4 #####
312 #####
313 tuneS4 <- function(par){
314     hw4 <- HoltWinters(serie(), alpha = par[1], beta = par[2],
315                      gamma = par[3])
316     mean((serie()-fitted(hw4)[, 1])^2)
317 }
318 par = c(0.5, 0.5, 0.5)
319 sol4 <- optim(par, tuneS4, method = "L-BFGS-B", lower = c(1e-4, 1
    e-4, 1e-4),
320             upper = c(0.9999, 0.9999, 0.9999))
321 hw4 <- HoltWinters(serie(), alpha = sol4$par[1], beta = sol4$par
    [2],
322                  gamma = sol4$par[3])
323 phw4 <- predict(hw4, 60)
324 ts.plot(serie(), phw4, col = c(1, a[2]), main = paste0("(4)AAA
    MSE = ",
325                                                         round(sol4
    $value, 5)))
326 #####
327 ##### Serie Tipo 5 #####
328 #####
329 tuneS5 <- function(par){
330     hw5 <- HoltWinters(serie(), alpha = par[1], beta = FALSE, gamma
    = par[2],
331                      seasonal = "multip")
332     mean((serie()-fitted(hw5)[, 1])^2)
333 }
334 par = c(0.5, 0.5)
335 sol5 <- optim(par, tuneS5, method = "L-BFGS-B", lower = c(1e-4, 1
    e-4),
336             upper = c(0.9999, 0.9999))
337 hw5 <- HoltWinters(serie(), alpha = sol5$par[1], beta = FALSE,
338                  gamma = sol5$par[2], seasonal = "m")

```

```

339 phw5 <- predict(hw5, 60)
340 ts.plot(serie(), phw5, col = c(1, a[2]), main = paste0("(5) ANM
    MSE = ",
341
                                                    round(sol5
    $value, 5)))
342
343 #####
344 ##### Serie Tipo 6 #####
345 #####
346 tuneS6 <- function(par){
347   hw6 <- HoltWinters(serie(), alpha = par[1], beta = par[2],
    gamma = par[3],
348
                                                    seasonal = "multip")
349   mean((serie()-fitted(hw6)[, 1])^2)
350 }
351 par = c(0.3, 0.1, 0.1)
352 sol6 <- optim(par, tuneS6, method = "L-BFGS-B", lower = c(1e-4, 1
    e-4, 1e-4),
353
                                                    upper = c(0.9999, 0.9999, 0.9999))
354 hw6 <- HoltWinters(serie(), alpha = sol6$par[1], beta = sol6$par
    [2],
355
                                                    gamma = sol6$par[3], seasonal = "m")
356 phw6 <- predict(hw6, 60)
357 ts.plot(serie(), phw6, col = c(1, a[2]), main = paste0("(6) AAM
    MSE = ",
358
                                                    round(sol6
    $value, 5)))
359 })
360
361 # Parametres
362
363 hw <- reactive({
364   if(input$serie == "Tipus 1"){
365     HoltWinters(serie(), alpha=NULL, beta=F, gamma=F, seasonal="
    additive")
366   }else if(input$serie == "Tipus 2"){
367     HoltWinters(serie(), alpha=NULL, beta=NULL, gamma=F, seasonal="
    additive")
368   }else if(input$serie == "Tipus 3"){
369     HoltWinters(serie(), alpha=NULL, beta=F, gamma=NULL, seasonal="
    additive")
370   }else if(input$serie == "Tipus 4"){
371     HoltWinters(serie(), alpha=NULL, beta=NULL, gamma=NULL, seasonal="
    additive")
372   }else if(input$serie == "Tipus 5"){
373     HoltWinters(serie(), alpha=NULL, beta=F, gamma=NULL, seasonal="
    multiplicative")

```

```

374 }else if(input$serie == "Tipus 6"){
375   HoltWinters(serie(),alpha=NULL,beta=NULL,gamma=NULL,
376             seasonal="multiplicative")
377 }
378 })
379
380 output$parametres <- renderTable({
381   if(input$serie == "Tipus 1"||input$serie == "Tipus 3"||input$
382     serie == "Tipus 5"){
383     data.frame(Alpha = round(hw()$alpha,2), Beta = round(hw()$beta
384     ,2),
385               Gamma = round(hw()$gamma,2),
386               a = round(hw()[["coefficients"]][["a"]],2))
387   }else if(input$serie == "Tipus 2"||input$serie == "Tipus 4"||
388   input$serie == "Tipus 6"){
389     data.frame(Alpha = round(hw()$alpha,2), Beta = round(hw()$beta
390     ,2),
391               Gamma = round(hw()$gamma,2),
392               a = round(hw()[["coefficients"]][["a"]],2),
393               b = round(hw()[["coefficients"]][["b"]],2))
394   }
395 },options = list(paging = F, searching = F, info = F))
396
397 # Estacionalitat
398 output$parmseason <- renderPlot({
399   if(input$serie == "Tipus 4"||input$serie == "Tipus 6"){
400     plot(hw()[["coefficients"]][-c(1,2)], type="o",
401           col=a[2], lty=2, lwd=2, main = "Estacionalitat", xlab=" ",
402           ylab=" ")
403     abline(h=0)
404   }else if(input$serie == "Tipus 3"||input$serie == "Tipus 5"){
405     plot(hw()[["coefficients"]][-c(1)], type="o",
406           col=a[2], lty=2, lwd=2, main = "Estacionalitat", xlab=" ",
407           ylab=" ")
408     abline(h=0)
409   }else{
410     plot(rep(0,input$Diffest), type="o",
411           col=a[2], lty=2, lwd=2, main = "Estacionalitat", xlab=" "
412           , ylab=" ")
413   }
414 })
415
416 # Predicci
417 output$hpasos <- renderPlot({
418   pred=predict(hw(),input$num)
419   ts.plot(totalserie(),pred,col=c(1,a[2]),lwd=c(1,2))

```

```

414 })
415
416
417 output$hpasostaula <- renderDataTable({
418   d <- data.frame(Y = as.matrix(predict(hw(),input$num)),
419                 date = seq.Date(interserie()[length(totalserie())
420   )-length(serie()),1] + 1,
421                 by = "day", length.out = input$num))
422   names(d) <- c("Predicci ", "Data")
423   d
424   }, options = list(pageLength=6))
425
426 output$etsplot <- renderPlot({
427   library(forecast)
428   fit <- ets(totalserie(), allow.multiplicative.trend=T)
429   plot(predict(fit,h=input$etsnum),ylab="")
430 })
431 #####*
432 ##### ARIMA #####*
433 #####*
434
435 # Observar si la vari ncia s constant
436 output$boxplot <- renderPlot({
437   boxplot(serie()~floor(time(serie()))), xlab = "", ylab = "", col
438   = a[2],
439   main = "Boxplots")
440 })
441
442 output$mediavar <- renderPlot({
443   m=apply(matrix(serie(),ncol=as.numeric(input$Diffest)),2,mean)
444   v=apply(matrix(serie(),ncol=as.numeric(input$Diffest)),2,var)
445   plot(v~m,main="Gr fic Mitjana-Variancia", ylab = "Variancia",
446   xlab = "Mitjana")
447   abline(lm(v~m),col=a[2],lty=3,lwd=2)
448 })
449
450 # Observar si hi ha estacionalitat
451 output$monthplot <- renderPlot(monthplot(serie(), ylab = ""))
452
453 # Transformacions de la s rie
454
455 #Funci per transformar la s rie
456 transformadaserie <- function(s,v,e,m,dif){
457   tserie <- s
458   lnserie <- s

```

```

458 SD <- 0
459 Rd <- 0
460 if(v){
461   tserie <- log(tserie)
462   lnserie <- log(lnserie)
463 }
464 if(e){
465   tserie <- diff(tserie,lag=as.numeric(dif))
466   SD <- 1
467 }
468 if(m){
469   Rd <- 0
470   x <- diff(tserie)
471   while (var(x) < var(tserie)) {
472     tserie <- x
473     x <- diff(tserie)
474     Rd <- Rd + 1
475     print(Rd)
476   }
477 }
478 return(list(Tserie = tserie, D = SD, d = Rd, NE = lnserie))
479 }
480
481 transerie <- reactive({
482   transformadaserie(serie(),input$Var,input$Est,input$Mitjana,
483     as.numeric(input$Diffest))
484 })
485
486 transerietotal <- reactive({
487   transformadaserie(totalserie(),input$Var,input$Est,input$Mitjana
488     ,
489     as.numeric(input$Diffest))
490 })
491
492 # Observar si la variància és constant
493 output$boxplott <- renderPlot({
494   boxplot(transerie()$Tserie~floor(time(transerie()$Tserie)),
495     xlab = "", ylab = "", col = a[2], main = "Boxplots")
496 })
497
498 output$mediavart <- renderPlot({
499   m=apply(matrix(transerie()$Tserie,ncol=as.numeric(input$Diffest)
500     ),2,mean)
501   v=apply(matrix(transerie()$Tserie,ncol=as.numeric(input$Diffest)
502     ),2,var)
503   plot(v~m,main="Gràfic Mitjana-Variància", ylab = "Variància",

```

```

502     xlab = "Mitjana")
503     abline(lm(v~m),col=a[2],lty=3,lwd=2)
504 })
505
506 # Observar si hi ha estacionalitat
507 output$monthplott <- renderPlot(monthplot(transerie())$Tserie, ylab
    = ""))
508
509 # Observar si la mitjana s constant
510 output$ctemean <- renderPlot({
511 plot(transerie())$Tserie,main="", ylab = "", xlab = "Temps")
512 abline(h=0)
513 abline(h=mean(transerie())$Tserie), col=a[2])
514 })
515
516 # Observar si la s rie transformada s estacionaria
517 output$Pacf <- renderPlot({
518     par(mfrow=c(1,2))
519     acf(transerie())$Tserie,ylim=c(-1,1),lag.max=60,
520         col=c(2,rep(1,as.numeric(input$Diffest)-1)),lwd=2, main = ""
521     )
522     pacf(transerie())$Tserie,ylim=c(-1,1),
523         lag.max=60,col=c(rep(1,as.numeric(input$Diffest)-1),2),lwd
524         =2, main = "")
525 })
526
527 # Creaci del model A
528 output$model1A_1 <- renderTable({
529     mod1A=arima(transerie())$Tserie,order=c(input$p,0,input$q),
530     seasonal=list(order=c(input$P,0,input$Q),
531     period=as.numeric(input$Diffest)))
532     t <- as.data.frame(matrix(c(mod1A$coef,sqrt(diag(mod1A$var.coef)
533     ),
534     abs(t(mod1A$coef/sqrt(diag(
535     mod1A$var.coef))))),
536     ncol = length(mod1A$coef), byrow = T))
537     names(t) <- names(mod1A$coef)
538     row.names(t) <- c("Coefficients", "S.E", "T-ratios")
539     t
540 }, rownames=TRUE, digits = 4)
541
542 output$model1A_2 <- renderTable({
543     mod1A=arima(transerie())$Tserie,order=c(input$p,0,input$q),
544     seasonal=list(order=c(input$P,0,input$Q),
545     period=as.numeric(input$Diffest)))
546     data.frame(Sigma2 = mod1A$sigma2, Versemblan a = mod1A$loglik,
547     AIC = mod1A$aic)

```

```

544 }, digits = 4)
545
546
547 # Creaci del model B
548 output$model1B_1 <- renderTable({
549   mod1B=arima(transerie()$NE, order=c(input$p,transerie()$d,input$
550     q),
551     seasonal=list(order=c(input$P,transerie()$D,input$Q
552       ),
553         period=as.numeric(input$Diffest))
554   t <- as.data.frame(matrix(c(mod1B$coef,sqrt(diag(mod1B$var.coef)
555     ),
556     abs(t(mod1B$coef/sqrt(diag(
557     mod1B$var.coef))))),
558     ncol = length(mod1B$coef), byrow = T))
559   names(t) <- names(mod1B$coef)
560   row.names(t) <- c("Coeficients", "S.E", "T-ratios")
561   t
562 }, rownames=TRUE, digits = 4)
563
564 output$model1B_2 <- renderTable({
565   mod1B=arima(transerie()$NE, order=c(input$p,transerie()$d,input$
566     q),
567     seasonal=list(order=c(input$P,transerie()$D,input$Q)
568     ,
569     period=as.numeric(input$Diffest))
570   data.frame(Sigma2 = mod1B$sigma2, Versemblan a = mod1B$loglik,
571     AIC = mod1B$aic)
572 }, digits = 4)
573
574
575 # Funci per la validaci
576
577 # Definici del model
578 modelo <- reactive({
579   if(input$choosemod=="Model A"){
580     arima(transerie()$Tserie, order=c(input$p,0,input$q),
581       seasonal=list(order=c(input$P,0,input$Q),
582         period=as.numeric(input$Diffest)))
583   }else if(input$choosemod=="Model B"){
584     arima(transerie()$NE, order=c(input$p,transerie()$d,input$q),
585       seasonal=list(order=c(input$P,transerie()$D,input$Q),
586         period=as.numeric(input$Diffest)))
587   }
588 })
589
590 modelototal <- reactive({

```



```

585   if(input$choosemod=="Model A"){
586     arima(transerietotal()$Tserie, order=c(input$p,0,input$q),
587         seasonal=list(order=c(input$P,0,input$Q),
588             period=as.numeric(input$Diffest)))
589   }else if(input$choosemod=="Model B"){
590     arima(transerietotal()$NE, order=c(input$p,transerietotal()$d,
591         input$q),
592         seasonal=list(order=c(input$P,transerietotal()$D,input$Q
593         ),
594             period=as.numeric(input$Diffest)))
595   }
596   })
597
598   #validation=function(inpmodel){
599
600     output$normalitat <- renderPlot({
601       s=as.numeric(input$Diffest)
602       resid=modelo()$residuals
603       par(mfrow=c(1,2))
604
605       #Normal plot of residuals
606       qqnorm(resid)
607       qqline(resid,col=a[2],lwd=2)
608
609       ##Histogram of residuals with normal curve
610       hist(resid,freq=FALSE, main = "Histograma dels residus", xlab =
611         "Residus",
612           ylab = "Densitat")
613       curve(dnorm(x,mean=mean(resid),sd=sd(resid)),col=a[2],lwd=2,add=
614         T)
615     })
616
617     output$variancia <- renderPlot({
618       s=as.numeric(input$Diffest)
619       resid=modelo()$residuals
620       par(mfrow=c(1,2))
621       #Residuals plot
622       plot(resid,main="Residuals", xlab = "Temps", ylab = "Residus")
623       abline(h=0)
624       abline(h=c(-3*sd(resid),3*sd(resid)),lty=3,col=a[2],lwd=2)
625       #Square Root of absolute values of residuals (Homocedasticity)
626       scatter.smooth(sqrt(abs(resid)),main="Arrel quadrada dels
627         residus absoluts",
628           ylab = "", xlab = "Temps", lpars=list(col=a[2],
629           lwd=2))
630     })

```

```

626 #ACF & PACF of residuals
627 output$acfpacf1 <- renderPlot({
628   resid=modelo()$residuals
629   par(mfrow=c(1,2))
630   acf(resid,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,as.numeric(input
631     $Diffest)-1)),
632     lwd=1, main = "ACF")
633   pacf(resid,ylim=c(-1,1),lag.max=60,col=c(rep(1,as.numeric(input$
634     Diffest)-1),
635     2),lwd=1, main = "PACF"
636   })
637
638 #ACF & PACF of square residuals
639 output$acfpacf2 <- renderPlot({
640   resid=modelo()$residuals
641   par(mfrow=c(1,2))
642   acf(resid^2,ylim=c(-1,1),lag.max=60,col=c(2,rep(1,as.numeric(
643     input$Diffest)-1)
644     ),lwd=1, main = "ACF
645     ^2")
646   pacf(resid^2,ylim=c(-1,1),lag.max=60,col=c(rep(1,as.numeric(
647     input$Diffest)-1),
648     2),lwd=1, main = "
649     PACF^2")
650   })
651
652 #Ljung-Box p-values
653 output$pval <- renderPlot({
654   par(mfrow=c(1,1))
655   plot(apply(matrix(1:40),1, function(e1)
656     Box.test(resid(modelo()),lag=e1,type="Lj")$p.value),type="p"
657     , ylab = "")
658   abline(h=0.05,lwd=2,lty=2,col=a[2])
659   })
660
661 #Stationary and Invertible
662 output$stainv1 <- renderPlotly({
663   autoplot(modelo())
664   })
665
666 ##Shapiro-Wilks Normality test
667 output$shapiro <- renderTable({
668   data.frame(W=as.numeric(shapiro.test(modelo())$residuals)$
669     statistic),
670     p_valor=shapiro.test(modelo())$p.value)
671   }, digits = 5)

```

```

664
665 ##Anderson-Darling test
666 output$andar <- renderTable({
667   library(nortest)
668   data.frame(A=as.numeric(ad.test(resid(modelo()))$statistic),
669             p_valor=ad.test(resid(modelo()))$p.value)
670 }, digits = 5)
671
672 ##Jarque-Bera test
673 output$jarbe <- renderTable({
674   library(tseries)
675   data.frame(X_Squared=as.numeric(jarque.bera.test(resid(modelo
676   ()))$statistic),
677             p_valor=jarque.bera.test(resid(modelo()))$p.value)
678 }, digits = 5)
679
680 ##Breusch-Pagan test
681 output$brpa <- renderTable({
682   library(lmtest)
683   data.frame(BP=as.numeric(bptest(
684     resid(modelo())~I(serie()-resid(modelo()))$statistic),
685     p_valor=bptest(
686     resid(modelo())~I(serie()-resid(modelo()))$p.
687     value)
688   }, digits = 5)
689
690 ##Durbin-Watson test
691 output$durwat <- renderTable({
692   library(lmtest)
693   dwtest(resid(modelo())~I(1:length(resid(modelo()))))
694   data.frame(DW=as.numeric(
695     dwtest(resid(modelo())~I(1:length(resid(modelo()))))$
696     statistic),
697     p_valor=dwtest(resid(modelo())~I(1:length(resid(modelo()))))
698     $p.value)
699 }, digits = 5)
700
701 ##Ljung-Box test
702 output$ljbox <- renderTable({
703   s=as.numeric(input$Diffest)
704   l <- t(apply(matrix(c(1:4,(1:4)*s)),1,function(e1) {
705     te=Box.test(resid(modelo()),type="Ljung-Box",lag=e1)
706     c(lag=(te$parameter),statistic=te$statistic[[1]],p.value=te$
707     p.value)}))
708   data.frame(l)
709 }, digits = 5)
710

```

```

706 # Predicci #
707 # Estabilitat
708 output$modelserie <- renderTable({
709   #mod1B=arima(transerie()$NE, order=c(input$p,transerie()$d,
input$q),
710   #           seasonal=list(order=c(input$P,transerie()$D,input
711   #           period=as.numeric(input$Diffest)))
712   t <- as.data.frame(matrix(c(modelo()$coef,sqrt(diag(modelo()$
var.coef))),
713   #           abs(t(modelo()$coef/sqrt(diag(
modelo()$var.coef))))),
714   #           ncol = length(modelo()$coef), byrow
= T))
715   names(t) <- names(modelo()$coef)
716   row.names(t) <- c("Coeficients", "S.E", "T-ratios")
717   t
718 }, rownames=TRUE, digits = 4)
719
720 output$modelserie1 <- renderTable({
721   #mod1B=arima(transerie()$NE, order=c(input$p,transerie()$d,
input$q),
722   #           seasonal=list(order=c(input$P,transerie()$D,input
723   #           period=as.numeric(input$Diffest)))
724   data.frame(Sigma2 = modelo()$sigma2, Versemblan a = modelo()$
loglik,
725   #           AIC = modelo()$aic)
726 }, digits = 4)
727
728 output$modelserietotal <- renderTable({
729   #mod1B=arima(transerie()$NE, order=c(input$p,transerie()$d,
input$q),
730   #           seasonal=list(order=c(input$P,transerie()$D,input
731   #           period=as.numeric(input$Diffest)))
732   t <- as.data.frame(matrix(c(modelototal()$coef,sqrt(diag(
modelototal()$var.coef))),
733   #           abs(t(modelototal()$coef/sqrt(diag
734   #           (modelototal()$var.coef))))),
#           ncol = length(modelototal()$coef),
735   #           byrow = T))
736   names(t) <- names(modelototal()$coef)
737   row.names(t) <- c("Coeficients", "S.E", "T-ratios")
738   t
739 }, rownames=TRUE, digits = 4)

```

```

740 output$modelserietotal1 <- renderTable({
741   #mod1B=arima(transerie()$NE, order=c(input$p,transerie()$d,
input$q),
742   #           seasonal=list(order=c(input$P,transerie()$D,input
$Q),
743   #           period=as.numeric(input$Diffest)))
744   data.frame(Sigma2 = modelototal()$sigma2, Versemblan a =
modelototal()$loglik,
745             AIC = modelototal()$aic)
746 }, digits = 4)
747
748 # Predicci Out-of-Sample
749 pred <- reactive({
750   predict(modelo(),n.ahead=length(totalserie())-length(serie()))
751 })
752
753 pr <- reactive({
754   ts(c(tail(transerie()$Tserie,1),pred())$pred),
755     start=end(serie()),
756     freq=as.numeric(input$Diffest))
757 })
758
759 se <- reactive({
760   ts(c(0,pred())$se),
761     start=end(serie()),
762     freq=as.numeric(input$Diffest))
763 })
764
765 output$outofsample <- renderPlot({
766
767   if(input$Var){
768     tl=ts(exp(pr()-1.96*se()),
769         start=end(serie()),
770         freq=as.numeric(input$Diffest))
771     tu=ts(exp(pr()+1.96*se()),
772         start=end(serie()),
773         freq=as.numeric(input$Diffest))
774     pr=ts(exp(pr()),
775         start=end(serie()),
776         freq=as.numeric(input$Diffest))
777   }else{
778     tl=ts((pr()-1.96*se()),
779         start=end(serie()),
780         freq=as.numeric(input$Diffest))
781     tu=ts((pr()+1.96*se()),
782         start=end(serie()),
783         freq=as.numeric(input$Diffest))

```

```

784     pr=ts((pr()),
785           start=end(serie()),
786           freq=as.numeric(input$Diffest))
787   }
788
789   ts.plot(totalserie(),t1,tu,pr,lty=c(1,2,2,1),col=c(1,a[2],a
790 [2],2),
791           type="o",main="")
792   #abline(v=(ultim[1]-3):(ultim[1]+2),lty=3,col=4)
793 })
794
795 # Predicci a h passos
796 pred1 <- reactive({
797   predict(modelototal(),n.ahead=input$hpasosarima)
798 })
799
800 pr1 <- reactive({
801   ts(c(tail(transerietotal())$Tserie,1),pred1()$pred),
802       start=end(totalserie()),
803       freq=as.numeric(input$Diffest))
804 })
805
806 se1 <- reactive({
807   ts(c(0,pred1()$se),
808       start=end(totalserie()),
809       freq=as.numeric(input$Diffest))
810 })
811
812 output$prediccioarima <- renderPlot({
813   #Intervals
814   if(input$Var){
815     t11=ts(exp(pr1()-1.96*se1()),
816            start=end(totalserie()),
817            freq=as.numeric(input$Diffest))
818     tu1=ts(exp(pr1()+1.96*se1()),
819            start=end(totalserie()),
820            freq=as.numeric(input$Diffest))
821     pr1=ts(exp(pr1()),
822            start=end(totalserie()),
823            freq=as.numeric(input$Diffest))
824   }else{
825     t11=ts((pr1()-1.96*se1()),
826            start=end(totalserie()),
827            freq=as.numeric(input$Diffest))
828     tu1=ts((pr1()+1.96*se1()),
829            start=end(totalserie()),

```

```

830         freq=as.numeric(input$Diffest))
831     pr1=ts((pr1()),
832           start=end(totalserie()),
833           freq=as.numeric(input$Diffest))
834 }
835
836     ts.plot(totalserie(),t11,tu1,pr1,lty=c(1,2,2,1),col=c(1,a[2],a
[2],2),
837           type="o",main="")
838     #abline(v = 2009:2022, col = "blue", lty = 3)
839 })
840
841 output$taulahpasosarima <- renderDataTable({
842     #Intervals
843     if(input$Var){
844         t11=ts(exp(pr1()-1.96*se1()),
845               start=end(totalserie()),
846               freq=as.numeric(input$Diffest))
847         tu1=ts(exp(pr1()+1.96*se1()),
848               start=end(totalserie()),
849               freq=as.numeric(input$Diffest))
850         pr1=ts(exp(pr1()),
851               start=end(totalserie()),
852               freq=as.numeric(input$Diffest))
853     }else{
854         t11=ts((pr1()-1.96*se1()),
855               start=end(totalserie()),
856               freq=as.numeric(input$Diffest))
857         tu1=ts((pr1()+1.96*se1()),
858               start=end(totalserie()),
859               freq=as.numeric(input$Diffest))
860         pr1=ts((pr1()),
861               start=end(totalserie()),
862               freq=as.numeric(input$Diffest))
863     }
864     data.frame(Data = seq.Date(interserie()[nrow(interserie())
,1]+1,
865                               by = "day", length.out = (input$
hpasosarima)+1),
866               ICI=t11,ICS=tu1,Predicci =pr1)
867 },options = list(pageLength=6))
868
869
870 # Autoarima
871 output$autoarima <- renderPlot({
872     library(forecast)
873     fit <- auto.arima(totalserie())

```

```

874     plot(forecast::forecast(fit,h=input$hpasosarima))
875   })
876
877   #####*
878   ##### PROPHECT #####*
879   #####*
880
881   library(prophet)
882   dfprophet <- reactive({
883     if((input$contaminant=="CO")||(input$contaminant=="NO2")||
884       (input$contaminant=="Oz ")||(input$contaminant=="PM2.5")||
885       (input$contaminant=="PM10")){
886     df <- mutate(
887       interserie(),
888       ds = Temps,
889       y = Concentraci
890     )
891     df <- column_to_rownames(df, var = "Temps")
892     m <- prophet()
893     m <- add_country_holidays(m,input$pais)
894     fit.prophet(m,df)
895   }else if(input$contaminant=="Dades"){
896     #df <- mutate(
897     # interserie(),
898     # ds = ds,
899     # y = y
900     #)
901     #df <- column_to_rownames(df, var = "ds")
902     m <- prophet()
903     m <- add_country_holidays(m,input$pais)
904     fit.prophet(m,interserie())
905   }
906   })
907
908   forecast <- reactive({
909     future <- make_future_dataframe(dfprophet(), periods = input$
910     prophetnum)
911     predict(dfprophet(), future)
912   })
913
914   output$predproph <- renderDygraph({
915     dyplot.prophet(dfprophet(),forecast())
916   })
917
918   output$plotpredcanvi <- renderPlotly({
919     plot(dfprophet(), forecast()+add_changepoints_to_plot(
920     dfprophet())

```



```
919   })
920
921   output$taulaprophet <- renderDataTable(
922     forecast()[c("ds", "yhat", "yhat_lower", "yhat_upper")],
923     options = list(pageLength=6)
924   )
925
926   output$components <- renderPlot(
927     prophet_plot_components(dfprophet(), forecast())
928   )
929
930   cvtable <- reactive({
931     cross_validation(dfprophet(), horizon = input$prophetnum,
932       units = 'days')
933   })
934
935   output$cv <- renderDataTable(
936     cvtable(),
937     options = list(pageLength=6)
938   )
939
940   output$metplot <- renderPlotly(
941     plot_cross_validation_metric(cvtable(), input$metrica,
942       rolling_window = 0.1)
943   )
944
945   output$mettable <- renderTable(
946     head(performance_metrics(cvtable()))
947   )
948
949   ##### Final
950
951   RMSPE <- function(obs,pr){
952     r <- sqrt(sum(((obs-pr)/obs)^2)/(length(totalserie())-length(
953     serie()))))
954     return(r)
955   }
956
957   MAPE <- function(obs,pr){
958     r <- sum(abs(obs-pr)/obs)/(length(totalserie())-length(serie(
959     )))
960     return(r)
961   }
962
963   MAE <- function(obs,pr){
964     r <- sum(abs(obs-pr))/(length(totalserie())-length(serie()))
```

```

963     return(r)
964 }
965
966 RMSE <- function(obs,pr){
967   r <- sqrt(sum(((obs-pr))^2)/(length(totalserie())-length(serie
968   ())))
969   return(r)
970 }
971
972 nRMSE <- function(obs,pr){
973   r <- RMSE(obs,pr)/abs(mean(obs))
974   return(r)
975 }
976
977 nMAE <- function(obs,pr){
978   r <- MAE(obs,pr)/abs(mean(obs))
979 }
980
981 pr_filtres <- reactive(as.matrix(predict(hw(),length(totalserie
982   ())))
983   length(serie()))))
984
985 pr_ets <- reactive({
986   library(forecast)
987   fitets <- forecast::ets(totalserie(), allow.multiplicative.
988   trend = T)
989   as.matrix(forecast::forecast(fitets, h=length(totalserie())-
990   length(serie()))$mean)
991 })
992
993 pr_arima <- reactive({
994   #Arima
995   if(input$Var){
996     as.matrix(ts(exp(pr()), start=end(serie()), freq=as.numeric(
997     input$Diffest)))[-1,]
998   }else{
999     as.matrix(ts((pr()), start=end(serie()), freq=as.numeric(
1000     input$Diffest)))[-1,]
1001   }
1002 })
1003
1004 pr_autoarima <- reactive({
1005   library(forecast)
1006   fitauto <- auto.arima(totalserie())
1007   as.matrix(forecast::forecast(fitauto, h=length(totalserie())-
1008   length(serie()))$mean)
1009 })

```

```

1005
1006 pr_prophet <- reactive({
1007   future <- make_future_dataframe(dfprophet(), periods = 1)
1008   forecast <- predict(dfprophet(), future)$yhat
1009   as.matrix(forecast)[1:(length(totalserie())-length(serie()))
1010   ,1]
1011 })
1012
1013 obs <- reactive({
1014   as.matrix(tail(interserie(),length(totalserie())-length(serie
1015   ()))[,2])
1016 })
1017
1018 output$metricasperf <- renderTable({
1019   c_mape <- c(MAPE(obs(),pr_filtres()), MAPE(obs(),pr_ets()),
1020   MAPE(obs(),pr_arima()),
1021   MAPE(obs(),pr_autoarima()), MAPE(obs(),pr_prophet
1022   ()))
1023   c_rmspe <- c(RMSPE(obs(),pr_filtres()), RMSPE(obs(),pr_ets()),
1024   RMSPE(obs(),pr_arima()),
1025   RMSPE(obs(),pr_autoarima()), RMSPE(obs(),pr_
1026   prophet()))
1027   c_mae <- c(MAE(obs(),pr_filtres()), MAE(obs(),pr_ets()), MAE(
1028   obs(),pr_arima()),
1029   MAE(obs(),pr_autoarima()), MAE(obs(),pr_prophet()))
1030   c_rmse <- c(RMSE(obs(),pr_filtres()), RMSE(obs(),pr_ets()),
1031   RMSE(obs(),pr_arima()),
1032   RMSE(obs(),pr_autoarima()), RMSE(obs(),pr_prophet
1033   ()))
1034   c_nmae <- c(nMAE(obs(),pr_filtres()), nMAE(obs(),pr_ets()),
1035   nMAE(obs(),pr_arima()),
1036   nMAE(obs(),pr_autoarima()), nMAE(obs(),pr_prophet
1037   ()))
1038   c_nrmse <- c(nRMSE(obs(),pr_filtres()), nRMSE(obs(),pr_ets()),
1039   nRMSE(obs(),pr_arima()),
1040   nRMSE(obs(),pr_autoarima()), nRMSE(obs(),pr_
1041   prophet()))
1042
1043   r <- data.frame(MODEL = c("EWMA", "ets", "Arima", "Auto.arima"
1044   , "Prophet"),
1045   MAPE = c_mape, RMSPE = c_rmspe, MAE = c_mae,
1046   RMSE = c_rmse,
1047   nMAE = c_nmae, nRMSE = c_nrmse)
1048 })
1049
1050 #####

```

```

1037   f_pr_filtres <- reactive(as.matrix(predict(hw(),input$
      predicciofinal)))
1038
1039   f_pr_ets <- reactive({
1040     library(forecast)
1041     fitets <- forecast::ets(totalserie(), allow.multiplicative.
      trend = T)
1042     as.matrix(forecast::forecast(fitets, h=input$predicciofinal)$
      mean)
1043
1044   })
1045
1046   predf <- reactive({
1047     predict(modelototal(),n.ahead=input$predicciofinal)
1048   })
1049
1050   prf <- reactive({
1051     ts(predf()$pred,
1052        start=end(totalserie()),
1053        freq=as.numeric(input$Diffest))
1054   })
1055
1056   f_pr_arima <- reactive({
1057     #Arima
1058     if(input$Var){
1059       as.matrix(ts(exp(prf()), start=end(totalserie()), freq=as.
      numeric(input$Diffest)))
1060     }else{
1061       as.matrix(ts((prf()), start=end(totalserie()), freq=as.
      numeric(input$Diffest)))
1062     }
1063   })
1064
1065   f_pr_autoarima <- reactive({
1066     library(forecast)
1067     fitauto <- auto.arima(totalserie())
1068     as.matrix(forecast::forecast(fitauto,h=input$predicciofinal)$
      mean)
1069   })
1070
1071   f_pr_prophet <- reactive({
1072     future <- make_future_dataframe(dfprophet(), periods = input$
      predicciofinal)
1073     forecast <- predict(dfprophet(), future)$yhat
1074     as.matrix(forecast)
1075   })
1076

```

```

1077   inpara <- reactive(as.character(input$ara))
1078
1079   prfinal <- reactive({
1080     if(inpara()=="EWMA"){
1081       as.numeric(f_pr_filtres()[nrow(f_pr_filtres()),])
1082     }else if(inpara()=="ets"){
1083       as.numeric(f_pr_ets()[nrow(f_pr_ets()),])
1084     }else if(inpara()=="Arima"){
1085       as.numeric(f_pr_arima()[nrow(f_pr_arima()),])
1086     }else if(inpara()=="Auto.arima"){
1087       as.numeric(f_pr_autoarima()[nrow(f_pr_autoarima()),])
1088     }else if(inpara()=="Prophet"){
1089       as.numeric(f_pr_prophet()[nrow(f_pr_prophet()),1])
1090     }else{
1091       0
1092     }
1093   })
1094
1095   #####
1096
1097   output$predicciofinal <- renderInfoBox({
1098     a <- round(prfinal(),2)
1099     if(a<input$green){
1100       infoBox(title = "Nivell de AQI", value = a,
1101               subtitle = "Nivell bo", color = "green",
1102               icon = icon("fa-duotone fa-wind"))
1103     }else if(a<input$yellow){
1104       infoBox(title = "Nivell de AQI", value = a,
1105               subtitle = "Nivell moderat", color = "yellow",
1106               icon = icon("fa-duotone fa-wind"))
1107     }else if(a<input$orange){
1108       infoBox(title = "Nivell de AQI", a,
1109               subtitle = "No saludable per grups sensibles", color
1110 = "orange",
1111               icon = icon("fa-duotone fa-wind"))
1112     }else if(a<input$red){
1113       infoBox(title = "Nivell de AQI", value = a,
1114               subtitle = "No saludable", color = "red",
1115               icon = icon("fa-duotone fa-wind"))
1116     }else if(a<input$purple){
1117       infoBox(title = "Nivell de AQI", value = a,
1118               subtitle = "Molt poc saludable", color = "purple",
1119               icon = icon("fa-duotone fa-wind"))
1120     }else if(a<input$black){
1121       infoBox(title = "Nivell de AQI", value = a,
1122               subtitle = "Perill s", color = "black",
1123               icon = icon("fa-duotone fa-wind"))

```

```

1123     }
1124   })
1125
1126
1127 #}
1128
1129 }
```

B.2 Ui

```

1 #
2 # This is the user-interface definition of a Shiny web application.
  # You can
3 # run the application by clicking 'Run App' above.
4 #
5 # Find out more about building applications with Shiny here:
6 #
7 #   http://shiny.rstudio.com/
8 #
9
10 library(shiny)
11 library(shinydashboard)
12 library(plotly)
13 library(shinyWidgets)
14 library(dygraphs)
15 library(datamods)
16
17 dashboardPage(
18   skin = "green",
19   dashboardHeader(title = "Qualitat de l'aire"),
20   dashboardSidebar(
21     sidebarMenu(
22       menuItem("Guia d' s ", tabName = "home"),
23       menuItem("Predicci amb filtres", tabName = "filtres"),
24       menuItem("Preparaci pr via al model", tabName = "model"),
25       menuItem("Models", tabName = "Arima"),
26       menuItem("Validaci ", tabName = "Validaci "),
27       menuItem("Predicci ", tabName = "Predicci "),
28       menuItem("Prophet", tabName = "Prophet"),
29       menuItem("Millor model", tabName = "bestmodel"),
30       menuItem("Autora", tabName = "finaljo")
31     )
32   ),
33   dashboardBody(
34     tabItems(
35
```

```

36 ##### TABO #####
37   tabItem(tabName = "home",
38           fluidRow(
39             tabBox(title = "Guia d'us", id = "home",
40                   width = "100%",
41                   height = "100%",
42                   tabPanel("Introducci ",
43                             fluidRow(
44                               column(12,
45                                     p("Us presentem la plataforma
46                                     interactiva
47                                     per modelitzar s ries
48                                     temporals de manera
49                                     senzilla i r pida.",
50                                     style="font-size:16px; text
51                                     -align: justify"),
52                                     p("En aquesta web hi ha
53                                     carregades 5 s ries
54                                     temporals de diferents
55                                     contaminants que
56                                     provenen de Calif rnia.
57                                     Tamb s'ofereix
58                                     en la pestanya del costat
59                                     la pujada
60                                     d'altres s ries.",
61                                     style="font-size:16px; text
62                                     -align: justify"),
63                                     p("En el men de l'esquerra
64                                     podreu navegar
65                                     per les diferents
66                                     metodologies que s'han
67                                     implementat. La metodologia
68                                     EWMA s'aplica
69                                     dins de", em("'Predicci
70                                     amb filtres'."),
71                                     "Per modelitzar segons la
72                                     metodologia
73                                     Box-Jenkins necessitareu
74                                     preparar la s rie
75                                     pr via al model',""),
76                                     , em("'Models',""),
77                                     "en", em("'Validaci '"), "
78                                     es validar
79                                     el model proposat i
80                                     finalment es pot veure

```

```

65     els valors futurs en", em("
    'Prediccions'."),
66     "A m s , s'ofereix la
    modelitzaci
67     autom tica no param trica
    amb el paquet
68     prophet clicant dins del
    men a",
69     em("'Prophet'."), "Per
    decidir quina
70     metodologia i model s el
    m s bo per
71     les dades , entreu en",
72     em("'Millor model'."),
73     style="font-size:16px; text
    -align: justify"
74     ),
75     p("Recordeu d'introduir el
    codi del pa s
76     d'on s'han extret les
    dades i
77     l'estacionalitat amb qu
    treballareu
78     abans d'iniciar aquest
    cam de modelar.",
79     style="font-size:16px;
    text-align: justify"),
80     p("Per identificar quin s
    el codi que heu
81     d'utilitzar , poder entrar
    al seg ent enlla :",
82     style="font-size:16px; text
    -align: justify"),
83     a("https://python-holidays.
    readthedocs.io/en/latest/",
84     style="font-size:16px; text
    -align: justify"),
85     br(), br(), br(),
86     fluidRow(
87         column(6,
88             textInput(inputId = "pais", label = "Introdueix
89                 el codi del pa s d'on
    prov la s rie
90                 en maj scules:",
91                 value = "US", width = "75%",
92                 placeholder = NULL)
93             ), column(6,

```



```

94         pickerInput(
95             inputId = "Diffest", label = "Estacionalitat
de D=?",
96             choices = c("5","7","12"),
97             options = list(
98                 style = "btn-success"),
99             selected = "7")
100         )
101     )
102 )
103 )
104 ),
105     tabPanel("Pujada de dades",
106         fluidRow(
107             column(9, align="right",
108                 dropdownButton(
109                     p("Primer, seleccioneu el
110 fitxer que cont la
111 s rie temporal amb la qual voleu
112 treballar. Aquest
113 ha de tenir en la primera columna
114 les dates i
115 en la segona els valors recollits.
116 ",br(),
117 "Seguidament, indiqueu si
118 el fitxer t cap alera ,
119 quin separador utilitza i el
120 format de les
121 dates.", br(), "Finalment ,
122 indiqueu les
123 observacions que emprareu per
124 entrenar
125 els models.",style="text-align:
126 justify"),
127             circle = TRUE,
128             status = "success",
129             size = "xs",
130             icon = icon("fa-duotone fa-
info"),
131             label = NULL,
132             tooltip = FALSE,
133             right = F,
134             up = FALSE,
135             width = NULL,
136             margin = "10px",
137             inline = TRUE,
138             inputId = NULL

```

```

130         ),
131         fluidRow(
132             column(6, align="left",
133                 fileInput("file1", "Escull el fitxer amb les
dades",
134                     multiple = FALSE,
135                     accept=c('text/csv',
136                             'text/comma-separated-values
,text/plain',
137                             '.csv')
138             ),
139             awesomeCheckbox('Cap alera', 'Cap alera del
fitxer', TRUE,
140                 status = "success"),
141             awesomeRadio('Separador', 'Separador del fitxer
',
142                 c(Coma=',', Punt_i_coma=';', Tab='
\t'),',',',
143                 status = "success"),
144             awesomeRadio(
145                 inputId = "formatdata",
146                 label = "Format de les dates",
147                 choices = c("%m/%d/%Y", "%d/%m/%Y", "%Y/%d/%m
", "%Y/%m/%d"),
148                 selected = "%m/%d/%Y",
149                 status = "success"
150             ),
151             uiOutput("slider")
152         ),
153         column(6, align="left",
154             tableOutput("dadestaula")
155         )
156     )
157 )
158 )
159 )
160 )
161 )
162 )
163 )
164 ),
165 #####
166
167 ##### TAB1 #####
168 tabItem(tabName = "filtres",
169     fluidRow(
170         tabBox(title = "Propietats de la s rie", id = "

```

```

tabsetpropserie",
171     tabPanel("Gràfic",
172             dropdownButton(
173                 p("Podeu seleccionar qualsevol de les
sries temporals
174                 que corresponen als diferents
contaminants o l'opció ",
175                 em("Dades"),"que correspon al fitxer
que heu carregat
176                 anteriorment.",br(),"Automàticament
veureu el gràfic
177                 de la srie temporal escollida.",
178                 style="text-align: left"),
179                 circle = TRUE,
180                 status = "success",
181                 size = "xs",
182                 icon = icon("fa-duotone fa-info"),
183                 label = NULL,
184                 tooltip = FALSE,
185                 right = F,
186                 up = FALSE,
187                 width = NULL,
188                 margin = "10px",
189                 inline = TRUE,
190                 inputId = NULL
191             ),
192             radioGroupButtons(
193                 inputId = "contaminant",
194                 label = "Contaminant",
195                 choices = c("CO", "NO2", "Oz", "PM2
.5", "PM10",
196                 "Dades"),
197                 selected = "PM2.5",
198                 status = "success"
199             ),
200             plotlyOutput("descriptiu")),
201     tabPanel("Taula",
202             dropdownButton(
203                 p("En aquesta pestanya veureu la taula
amb les
204                 observacions de la srie temporal
total.",
205                 style="text-align: left"),
206                 circle = TRUE,
207                 status = "success",
208                 size = "xs",
209                 icon = icon("fa-duotone fa-info"),

```

```

210         label = NULL,
211         tooltip = FALSE,
212         right = F,
213         up = FALSE,
214         width = NULL,
215         margin = "10px",
216         inline = TRUE,
217         inputId = NULL
218     ),
219     dataTableOutput("obs")),
220     tabPanel("Descomposici ",
221             dropdownButton(
222                 p("Aqu  podeu estudiar gr ficament
les components de
223                 la s rie.",
224                 style="text-align: left"),
225                 circle = TRUE,
226                 status = "success",
227                 size = "xs",
228                 icon = icon("fa-duotone fa-info"),
229                 label = NULL,
230                 tooltip = FALSE,
231                 right = F,
232                 up = FALSE,
233                 width = NULL,
234                 margin = "10px",
235                 inline = TRUE,
236                 inputId = NULL
237             ),
238             plotOutput("decompose")),
239     tabPanel("Tipus de s rie",
240             dropdownButton(
241                 p("En aquesta pestanya es presenten els
diferents tipus
242                 de s rie que defineix la metodologia
d'allisat
243                 exponencial. Segons les components
que s'han vist
244                 presents anteriorment, haureu d'
aplicar un tipus
245                 de s rie diferent.",br(), "Si no heu
vist clar quin
246                 tipus de components presenta la
s rie que esteu
247                 estudiant, podeu ajudar-vos amb el
valor MSE.
248                 Escolliu el que presenti menor MSE.",

```

```

249         style="text-align: left"),
250         circle = TRUE,
251         status = "success",
252         size = "xs",
253         icon = icon("fa-duotone fa-info"),
254         label = NULL,
255         tooltip = FALSE,
256         right = F,
257         up = FALSE,
258         width = NULL,
259         margin = "10px",
260         inline = TRUE,
261         inputId = NULL
262     ),
263     plotOutput("tipusdeserie")
264 ),
265     tabBox(title = "Predicci ", id = "tabsetpred",
266           tabPanel("Par metres",
267                   dropdownButton(
268                       p("En aquesta pestanya heu de
seleccionar el tipus de
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
s rie que heu determinat
"Seguidament, us apareixeran els
par metres del
model segons la metodologia d'allisat
exponencial.",
                style="text-align: left"),
                circle = TRUE,
                status = "success",
                size = "xs",
                icon = icon("fa-duotone fa-info"),
                label = NULL,
                tooltip = FALSE,
                right = F,
                up = FALSE,
                width = NULL,
                margin = "10px",
                inline = TRUE,
                inputId = NULL
            ),
            pickerInput(
                inputId = "serie", label = "Tipus de
s rie",
                choices = c("Tipus 1","Tipus 2","Tipus
3","Tipus 4",
                "Tipus 5", "Tipus 6"),

```

```

290         options = list(
291             style = "btn-success"),
292         selected = "Tipus 1"),
293         tableOutput("parametres")
294     ),
295     tabPanel("Estacionalitat",
296         dropdownButton(
297             p("Observeu de manera gràfica els
par metres estimats
298             per l'estacionalitat.",
299             style="text-align: left"),
300             circle = TRUE,
301             status = "success",
302             size = "xs",
303             icon = icon("fa-duotone fa-info"),
304             label = NULL,
305             tooltip = FALSE,
306             right = F,
307             up = FALSE,
308             width = NULL,
309             margin = "10px",
310             inline = TRUE,
311             inputId = NULL
312         ),
313         plotOutput("parmseason")),
314     tabPanel("Gràfic de la predicció",
315         dropdownButton(
316             p("Seleccioneu el nombre de dies que
voleu predir.
317             En el gràfic veureu la predicció
des de l'últim
318             dia de la sèrie temporal utilitzada
per crear el
319             model.",
320             style="text-align: left"),
321             circle = TRUE,
322             status = "success",
323             size = "xs",
324             icon = icon("fa-duotone fa-info"),
325             label = NULL,
326             tooltip = FALSE,
327             right = F,
328             up = FALSE,
329             width = NULL,
330             margin = "10px",
331             inline = TRUE,
332             inputId = NULL

```

```

333         ),
334         numericInput(inputId = "num", label = "h
passos:",
335                             value = 182, min = 1, max =
NA),
336         plotOutput("hpassos")),
337     tabPanel("Taula del la predicci ",
338         dropdownButton(
339             p("Observeu els valors predits
num ricament en la taula.",
340                 style="text-align: left"),
341             circle = TRUE,
342             status = "success",
343             size = "xs",
344             icon = icon("fa-duotone fa-info"),
345             label = NULL,
346             tooltip = FALSE,
347             right = F,
348             up = FALSE,
349             width = NULL,
350             margin = "10px",
351             inline = TRUE,
352             inputId = NULL
353         ),
354         dataTableOutput("hpassostaula")),
355     tabPanel("ets",
356         dropdownButton(
357             p("El gr fic mostra el model proposat
autom ticament
358             prediccions del
359                 nombre de dies que heu indicat.",
360                 style="text-align: left"),
361             circle = TRUE,
362             status = "success",
363             size = "xs",
364             icon = icon("fa-duotone fa-info"),
365             label = NULL,
366             tooltip = FALSE,
367             right = F,
368             up = FALSE,
369             width = NULL,
370             margin = "10px",
371             inline = TRUE,
372             inputId = NULL
373         ),
374         numericInput(inputId = "etsnum", label =

```

```

    "h passos:",
375                                     value = 30, min = 1, max =
NA),
376                                     plotOutput("etsplot"))
377     )
378   )
379 ),
380 #####
381
382 ##### TAB2 #####
383   tabItem(tabName = "model",
384     fluidRow(
385       tabBox(title = "Gr fics pr vis", id = "previ",
386         tabPanel("Boxplot",
387           dropdownButton(
388             p("Per aplicar la metodologia Box
-Jenkins cal
389               tenir una s rie amb vari ncia
constant.",br(),
390               "Observa el gr fic de box plot
per decidir si
391               la vari ncia s constant.",
392               style="text-align: left"),
393             circle = TRUE,
394             status = "success",
395             size = "xs",
396             icon = icon("fa-duotone fa-info")
,
397             label = NULL,
398             tooltip = FALSE,
399             right = F,
400             up = FALSE,
401             width = NULL,
402             margin = "10px",
403             inline = TRUE,
404             inputId = NULL
),
405         plotOutput("boxplot")),
406     tabPanel("Mitjana-Var",
407       dropdownButton(
408         p("El gr fic mitjana-vari ncia
s un altre
410           m tode visual que ajuda a
identificar si la
411           s rie presenta vari ncia
constant.",br(),
412         "Per identificar-ho, cal que la

```



```

    l n i a  v e r d a
413         sigui el m s horitzontal
possible.",
414         style="text-align: left"),
415         circle = TRUE,
416         status = "success",
417         size = "xs",
418         icon = icon("fa-duotone fa-info")
,
419         label = NULL,
420         tooltip = FALSE,
421         right = F,
422         up = FALSE,
423         width = NULL,
424         margin = "10px",
425         inline = TRUE,
426         inputId = NULL
427     ),
428     plotOutput("mediavar")),
429     tabPanel("Monthplot",
430         dropdownButton(
431             p("Podeu analitzar si la s rie
presenta
432             estacionalitat amb aquest
gr fic.",br()),
433             "Podeu considerar que hi ha un
patr estacional
434             si les mitjanes no estan
alineades de manera
435             significativa.",
436             style="text-align: left"),
437             circle = TRUE,
438             status = "success",
439             size = "xs",
440             icon = icon("fa-duotone fa-info")
,
441             label = NULL,
442             tooltip = FALSE,
443             right = F,
444             up = FALSE,
445             width = NULL,
446             margin = "10px",
447             inline = TRUE,
448             inputId = NULL
449         ),
450     plotOutput("monthplot")),
451     tabPanel("Transformacions",

```

```

452     dropdownButton(
453         p("Segons les conclusions a les
quals heu arribat
454             anteriorment necessitareu, o no
, aplicar les
455             diferents transformacions que
es suggereixen
456             en el seg ent llistat.",br(),"
Podeu veure
457             com canvien els gr fics en la
pantalla de la
458             dreta per confirmar que s
necessaria la
459             transforamci a la s rie.",
460             style="text-align: left"),
461         circle = TRUE,
462         status = "success",
463         size = "xs",
464         icon = icon("fa-duotone fa-info")
,
465         label = NULL,
466         tooltip = FALSE,
467         right = F,
468         up = FALSE,
469         width = NULL,
470         margin = "10px",
471         inline = TRUE,
472         inputId = NULL
473     ),
474     br(),br(),
475     strong(" s la vari ncia constant?
Si no ho s ,
476     aplica:"),
477     prettyCheckbox(inputId = "Var",
478         label = "
Transformaci
479             logar tmica",
480             value = F,
481             status = "success",
482             fill = F),
483     br(), strong("Hi ha estacionalitat?
Si n'hi ha,
484     aplica:"),
485     prettyCheckbox(inputId = "Est",
486         label = "
Diferenciaci
487             estacional",

```

```

488                                     value = F,
489                                     status = "success",
490                                     fill = F),
491     br(), strong(" s la mitjana
constant? Si no ho s ,
492     aplica:"),
493     prettyCheckbox(inputId = "Mitjana",
494                   label = "
Diferenciacions
495                                     regulars",
496                                     value = F,
497                                     status = "success",
498                                     fill = F)
499     )
500   ),
501   tabBox(title = "Gr fics transformats", id = "post"
,
502         tabPanel("Boxplot",
503                 dropdownButton(
504                   p("Observa el gr fic de box plot
amb la
505                   transformaci aplicada a la
s rie per
506                   estabilitzar la vari ncia.
Creieu que necessiteu
507                   aplicar la transformaci ?",
508                     style="text-align: left"),
509                   circle = TRUE,
510                   status = "success",
511                   size = "xs",
512                   icon = icon("fa-duotone fa-info")
,
513                   label = NULL,
514                   tooltip = FALSE,
515                   right = F,
516                   up = FALSE,
517                   width = NULL,
518                   margin = "10px",
519                   inline = TRUE,
520                   inputId = NULL
521                 ),
522         plotOutput("boxplott")),
523   tabPanel("Mitjana-Var",
524           dropdownButton(
525             p("Observa el gr fic de mitjana-
vari ncia amb la
526             transformaci aplicada a la

```

```

s rie per
527     establitzar la vari ncia .

Creieu que necessiteu
528     aplicar la transformaci ?",
529     style="text-align: left"),
530     circle = TRUE,
531     status = "success",
532     size = "xs",
533     icon = icon("fa-duotone fa-info")
,
534     label = NULL,
535     tooltip = FALSE,
536     right = F,
537     up = FALSE,
538     width = NULL,
539     margin = "10px",
540     inline = TRUE,
541     inputId = NULL
542 ),
543     plotOutput("mediavart")),
544     tabPanel("Monthplot",
545     dropdownButton(
546     p("Observa el gr fic de l'
estacionalitat amb la
547     transformaci aplicada a la
s rie per
548     eliminar la component estacional.
Creieu que necessiteu
549     aplicar la transformaci ?",
550     style="text-align: left"),
551     circle = TRUE,
552     status = "success",
553     size = "xs",
554     icon = icon("fa-duotone fa-info")
,
555     label = NULL,
556     tooltip = FALSE,
557     right = F,
558     up = FALSE,
559     width = NULL,
560     margin = "10px",
561     inline = TRUE,
562     inputId = NULL
563 ),
564     plotOutput("monthplott")),
565     tabPanel("Mitjana",
566     dropdownButton(

```

```

567     p("Per acabar de preparar la
s rie temporal
568     abans d'iniciar amb la
modelitzaci necessiteu
569     que la mitjana sigui 0.",br(),
570     "Si la mitjana no s 0,
apliqueu en el panell
571     de l'esquerra les
diferenciacions regulars.
572     Autom ticament se'n faran
tantes com siguin
573     necess ries sense fer que la
s rie sigui
574     sobre dispersa.",
575     style="text-align: left"),
576     circle = TRUE,
577     status = "success",
578     size = "xs",
579     icon = icon("fa-duotone fa-info")
,
580     label = NULL,
581     tooltip = FALSE,
582     right = F,
583     up = FALSE,
584     width = NULL,
585     margin = "10px",
586     inline = TRUE,
587     inputId = NULL
588 ),
589     plotOutput("ctemean")),
590     tabPanel("ACF/PACF",
591     dropdownButton(
592     p("Comproveu que la s rie est
preparada per
593     modelitzar amb els seg ents
gr fics. Per
594     determinar-ho, cal observar que
els retards
595     decauen r pidament a 0.",br(),
"Les barres
596     negres representen la component
regular del
597     model i les vermelles la
component estacional.
598     Amb aquests gr fics podeu
determinar els possibles
599     par metres per crear un model

```

```

SARIMA.",
600         style="text-align: left"),
601         circle = TRUE,
602         status = "success",
603         size = "xs",
604         icon = icon("fa-duotone fa-info")
,
605         label = NULL,
606         tooltip = FALSE,
607         right = F,
608         up = FALSE,
609         width = NULL,
610         margin = "10px",
611         inline = TRUE,
612         inputId = NULL
613     ),
614     plotOutput("Pacf"))
615 )
616 )
617 )
618 ),
619 #####
620
621 ##### TAB3 #####
622 tabItem(tabName = "Arima",
623         fluidRow(
624             box(title = "Par metres", id = "tabsetcreaciomodel
",
625                 dropdownButton(
626                     p("Inseriu els par metres del model SARIMA
que voleu provar
627                         en les seg ents caselles. Els par metres
que determinen
628                         tant la diferenciaci estacional com les
regulars
629                         s'inclouran en el model autom ticament
segons el que heu
630                         decidit en la preparaci previa al model."
,
631                         style="text-align: left"),
632                     circle = TRUE,
633                     status = "success",
634                     size = "xs",
635                     icon = icon("fa-duotone fa-info"),
636                     label = NULL,
637                     tooltip = FALSE,
638                     right = F,

```

```

639         up = FALSE,
640         width = NULL,
641         margin = "10px",
642         inline = TRUE,
643         inputId = NULL
644     ),br(),
645     numericInput(inputId = "p",label = "Par metre
p",
646                 value = 0, min = 0, max = NA),
647     numericInput(inputId = "q", label = "Par metre
q",
648                 value = 0, min = 0, max = NA),
649     numericInput(inputId = "P", label = "Par metre
P",
650                 value = 0, min = 0, max = NA),
651     numericInput(inputId = "Q", label = "Par metre
Q",
652                 value = 0, min = 0, max = NA)
653 ),
654     tabBox(title = "Model", id = "tabsetmodel",
655            tabPanel("Estimaci A",
656                    dropdownButton(
657                        p("Observeu els par metres que s
'han estimat i
658                        si s n significatiu. L'
estimaci A
659                        considera", em("l'intercept"),
"tot i que pels
660                        pr xims passos i facilitar la
modelitzaci es
661                        considerar que no s
significatiu."),
662                        style="text-align: left"),
663                        circle = TRUE,
664                        status = "success",
665                        size = "xs",
666                        icon = icon("fa-duotone fa-info")
,
667                        label = NULL,
668                        tooltip = FALSE,
669                        right = F,
670                        up = FALSE,
671                        width = NULL,
672                        margin = "10px",
673                        inline = TRUE,
674                        inputId = NULL
675                    ),br(),

```

```

676         "Model utilitzant la s rie
677         estacionaria. Els T-ratios s n
significatiu si
678         s n majors a 2:",
679         tableOutput("model1A_1"),
680         tableOutput("model1A_2")
681     ),
682     tabPanel("Estimaci B",
683         dropdownButton(
684             p("Observeu els parmetres que s
'han estimat i
685             si s n significatiu. L'
estimaci B no
686             considera", em("l'intercept."),
687             style="text-align: left"),
688             circle = TRUE,
689             status = "success",
690             size = "xs",
691             icon = icon("fa-duotone fa-info")
,
692             label = NULL,
693             tooltip = FALSE,
694             right = F,
695             up = FALSE,
696             width = NULL,
697             margin = "10px",
698             inline = TRUE,
699             inputId = NULL
700         ),br(),
701         "Model utilitzant la s rie no
702         estacionaria. Els T-ratios s n
significatiu si
703         s n majors a 2:",
704         tableOutput("model1B_1"),
705         tableOutput("model1B_2")
706     )
707 )
708 )
709 ),
710 #####
711
712 ##### TAB4 #####
713     tabItem(tabName = "Validaci ",
714             fluidRow(
715                 tabBox(title = "Validaci ", id = "valid",
716                       width = "100%",
717                       height = "100%",

```



```

718         tabPanel("Normalitat",
719                 dropdownButton(
720                     p("Despr s de definir el model
necessiteu validar-ho.
721                     Una de les condicions s que els
residus segueixin
722                     la distribuci normal.",br(),"
Podeu observar tant
723                     l'histograma com el qqnorm per
decidir-ho
724                     gr ficament.",
725                     style="text-align: left"),
726                     circle = TRUE,
727                     status = "success",
728                     size = "xs",
729                     icon = icon("fa-duotone fa-info"),
730                     label = NULL,
731                     tooltip = FALSE,
732                     right = F,
733                     up = FALSE,
734                     width = NULL,
735                     margin = "10px",
736                     inline = TRUE,
737                     inputId = NULL
738                 ),
739                 pickerInput(
740                     inputId = "choosemod", label = "
Validaci del
741                     model:",
742                     choices = c("Model A","Model B"),
743                     options = list(
744                         style = "btn-success"),
745                     selected = "Model B"),
746                     plotOutput("normalitat")
747                 ),
748         tabPanel("Vari ncia",
749                 dropdownButton(
750                     p("Els residus tamb han de
751                     presentar vari ncia
constant. Per decidir-ho de manera
752                     gr fica podeu
observar les gr fiques dels
753                     residus i de l'arrel
dels residus absoluts.",
754                     style="text-align: left"),
755                     circle = TRUE,
756                     status = "success",

```

```

757         size = "xs",
758         icon = icon("fa-duotone fa-info"),
759         label = NULL,
760         tooltip = FALSE,
761         right = F,
762         up = FALSE,
763         width = NULL,
764         margin = "10px",
765         inline = TRUE,
766         inputId = NULL
767     ),
768     plotOutput("variancia")),
769     tabPanel("ACF i PACF",
770         dropdownButton(
771             p("Per determinar si els residus no
772 estan
773         autocorrelacionats s'observa tant
774 l'ACF com el
775         PACF. Ambd s gr fics han de
776 presentar els retards
777         dins l'interval de confian a al
778 95% seguint el
779         patr del soroll blanc.",
780         style="text-align: left"),
781         circle = TRUE,
782         status = "success",
783         size = "xs",
784         icon = icon("fa-duotone fa-info"),
785         label = NULL,
786         tooltip = FALSE,
787         right = F,
788         up = FALSE,
789         width = NULL,
790         margin = "10px",
791         inline = TRUE,
792         inputId = NULL
793     ),
794     plotOutput("acfpac1")),
795     tabPanel("ACF i PACF ^2",
796         dropdownButton(
797             p("Per garantir que no est present
798 volatilitat
799         cal que en els gr fics de l'ACF i
800 PACF al quadrat
801         presentin els diferents retards
802 dins de l'interval
803         de confian a al 95% seguint el

```

```

    patr del soroll blanc.",
797         style="text-align: left"),
798         circle = TRUE,
799         status = "success",
800         size = "xs",
801         icon = icon("fa-duotone fa-info"),
802         label = NULL,
803         tooltip = FALSE,
804         right = F,
805         up = FALSE,
806         width = NULL,
807         margin = "10px",
808         inline = TRUE,
809         inputId = NULL
810     ),
811     plotOutput("acfpacf2")),
812     tabPanel("Ljung-Box",
813         dropdownButton(
814             p("En aquest gràfic s'observen els
p-valors del
815             test Ljung-Box. Aquest és un test
per detectar
816             si els residus del model estan
autocorrelacionats.
817             Tots els punts han d'estar per
sobre de 0.05,
818             visualment veureu una línia verda
que determina
819             aquest límit."),
820             style="text-align: left"),
821             circle = TRUE,
822             status = "success",
823             size = "xs",
824             icon = icon("fa-duotone fa-info"),
825             label = NULL,
826             tooltip = FALSE,
827             right = F,
828             up = FALSE,
829             width = NULL,
830             margin = "10px",
831             inline = TRUE,
832             inputId = NULL
833         ),
834         plotOutput("pval")),
835     tabPanel("Estacionarietat/Invertible",
836         dropdownButton(
837             p("és una bona pràctica comprovar que

```

```

    el model proposat
838         s invertible i estacionari. Per
fer-ho cal que
839         les arrels inverses del polinomi
caracter stic
840         estiguin dins del cercle unitari."
,
841         style="text-align: left"),
842     circle = TRUE,
843     status = "success",
844     size = "xs",
845     icon = icon("fa-duotone fa-info"),
846     label = NULL,
847     tooltip = FALSE,
848     right = F,
849     up = FALSE,
850     width = NULL,
851     margin = "10px",
852     inline = TRUE,
853     inputId = NULL
854 ),
855     plotlyOutput("stainv1")),
856     #tableOutput("stainv2")),
857     tabPanel("Tests num rics",
858         dropdownButton(
859             p("Tal com s'ha fet gr ficamente, en
aquesta
860             pantalla es presenten els tests
num rics per
861             validar el model. Els tests
num rics
862             poden no ser fiables com els
gr fics a causa de
863             la sensibilitat als outliers.",
864             style="text-align: left"),
865             circle = TRUE,
866             status = "success",
867             size = "xs",
868             icon = icon("fa-duotone fa-info"),
869             label = NULL,
870             tooltip = FALSE,
871             right = F,
872             up = FALSE,
873             width = NULL,
874             margin = "10px",
875             inline = TRUE,
876             inputId = NULL

```

```

877         ),
878         fluidRow(column(3, align = "center",
879         strong("Normalitat: Shapiro-Wilks"),
880         tableOutput("shapiro"),
881         strong("Normalitat: Anderson-Darling")
882     ,
883         tableOutput("andar"),
884         strong("Normalitat: Jarque-Bera"),
885         tableOutput("jarbe")),
886         column(3, align = "center",
887         strong("Homocedasticitat: Breusch-
Pagan")),
888         tableOutput("brpa")),
889         column(3, align = "center",
900         strong("Independència: Durbin-Watson"
901     ),
902         tableOutput("durwat")),
903         column(3, align = "center",
904         strong("Independència: Ljung-Box"),
905         tableOutput("ljbox"))
906     ),
907     )
908 ),
909 #####
910 ##### TAB5 #####
911 tabItem(tabName = "Predicci ",
912         fluidRow(
913             tabBox(title = "Predicci ", id = "predar",
914                 width = "100%",
915                 height = "100%",
916                 tabPanel("Estabilitat",
917                     dropdownButton(
918                         p("Per poder fer prediccions amb
919 el model ARIMA ,
920 primer cal comprovar que s un
921 model estable.
922 Es pot comprovar observant les
923 estimacions
924 dels par metres del model amb
925 la s rie que
926 s'ha utilitzat per modelar i la
927 s rie completa.
928 Aquests han de presentar valors
929 semblants i

```

```

915         els mateixos signes.",
916         style="text-align: left"),
917         circle = TRUE,
918         status = "success",
919         size = "xs",
920         icon = icon("fa-duotone fa-info")
921     ,
922     label = NULL,
923     tooltip = FALSE,
924     right = F,
925     up = FALSE,
926     width = NULL,
927     margin = "10px",
928     inline = TRUE,
929     inputId = NULL
930 ),
931 fluidRow(
932     column(6,
933         "S rie tallada fins l'
observaci 189 o
de les dades:",
934         tableOutput("modelserie"),
935         tableOutput("modelserie1")
936     ),
937     column(6,
938         "S rie completa:",
939         tableOutput("
modelserietotal"),
940         tableOutput("
modelserietotal1")
941     )
942 )
943 ),
944 tabPanel("Predicci Out-of-Sample",
945     dropdownButton(
946         p("El gr fic mostra les
prediccions que faria
947         el model a partir de la s rie
amb qu s'ha
948         treballat i el que realment s'
ha observat durant
949         aquell per ode. En vermell es
mostra la
950         predicci puntual i en verd l'
interval de confian a
951         al 95%.",

```

```

952         style="text-align: left"),
953         circle = TRUE,
954         status = "success",
955         size = "xs",
956         icon = icon("fa-duotone fa-info")
957     ,
958     label = NULL,
959     tooltip = FALSE,
960     right = F,
961     up = FALSE,
962     width = NULL,
963     margin = "10px",
964     inline = TRUE,
965     inputId = NULL
966 ),
967     plotOutput("outofsample")
968 ),
969     tabPanel("Predicci ",
970         dropdownButton(
971             p("Selecioneu el nombre de dies
972             que voleu predir.
973             En el gr fic veureu la
974             predicci puntual de color
975             vermell i l'interval de
976             confian a al 95% de
977             color verd.",
978             style="text-align: left"),
979             circle = TRUE,
980             status = "success",
981             size = "xs",
982             icon = icon("fa-duotone fa-info")
983         ,
984         label = NULL,
985         tooltip = FALSE,
986         right = F,
987         up = FALSE,
988         width = NULL,
989         margin = "10px",
990         inline = TRUE,
991         inputId = NULL
992     ),
993     numericInput(inputId = "hpassosarima",
994         label = "h passos:",
995         value = 30, min = 1,
996         max = NA),
997     plotOutput("prediccioarima")

```

```

992         ),
993         tabPanel("Taula de prediccions",
994             dropdownButton(
995                 p("Observeu els valors predits
num ricament en
996                 la taula.",
997                 style="text-align: left"),
998                 circle = TRUE,
999                 status = "success",
1000                 size = "xs",
1001                 icon = icon("fa-duotone fa-info")
,
1002                 label = NULL,
1003                 tooltip = FALSE,
1004                 right = F,
1005                 up = FALSE,
1006                 width = NULL,
1007                 margin = "10px",
1008                 inline = TRUE,
1009                 inputId = NULL
1010             ),
1011             dataTableOutput("taulahpasosarima")
),
1012         tabPanel("Auto.arima",
1013             dropdownButton(
1014                 p("El gr fic mostra el model
proposat
1015                 autom ticament amb la funci ",
em("auto.arima()"),
1016                 "i les prediccions del nombre de
dies que heu
1017                 indicat anteriorment.",
1018                 style="text-align: left"),
1019                 circle = TRUE,
1020                 status = "success",
1021                 size = "xs",
1022                 icon = icon("fa-duotone fa-info")
,
1023                 label = NULL,
1024                 tooltip = FALSE,
1025                 right = F,
1026                 up = FALSE,
1027                 width = NULL,
1028                 margin = "10px",
1029                 inline = TRUE,
1030                 inputId = NULL
1031             ),

```



```

1032         plotOutput("autoarima")
1033     )
1034 )
1035 )
1036 ),
1037 #####
1038
1039 ##### TAB6 #####
1040 tabItem(tabName = "Prophet",
1041     fluidRow(
1042         tabBox(title = "Prophet", id = "prophetli",
1043             width = "100%",
1044             height = "100%",
1045             tabPanel("Predicci din mica",
1046                 dropdownButton(
1047                     p("Cal que inseriu el nombre de
1048 dies que voleu
1049 presentar el model
1050 no param tric creat segons l'
1051 algoritme
1052 prophet.",
1053                 style="text-align: left"),
1054                 circle = TRUE,
1055                 status = "success",
1056                 size = "xs",
1057                 icon = icon("fa-duotone fa-info")
1058             ,
1059                 label = NULL,
1060                 tooltip = FALSE,
1061                 right = F,
1062                 up = FALSE,
1063                 width = NULL,
1064                 margin = "10px",
1065                 inline = TRUE,
1066                 inputId = NULL
1067             ),
1068             numericInput(inputId = "prophetnum"
1069             ,
1070                 label = "h passos:",
1071                 value = 30, min = 1,
1072                 max = NA),
1073             dygraphOutput("predproph")
1074         ),
1075         tabPanel("Predicci amb els punts de canvi"
1076         ,
1077             dropdownButton(

```

```

1072         p("Aquest gràfic mostra els
mateixos resultats
1073         presentats en la pestanya
anterior, per
1074         indicant els punts que s'han
detectat
1075         automàticament un canvi
significatiu de la
1076         tendència.",
1077         style="text-align: left"),
1078         circle = TRUE,
1079         status = "success",
1080         size = "xs",
1081         icon = icon("fa-duotone fa-info")
,
1082         label = NULL,
1083         tooltip = FALSE,
1084         right = F,
1085         up = FALSE,
1086         width = NULL,
1087         margin = "10px",
1088         inline = TRUE,
1089         inputId = NULL
1090     ),
1091     numericInput(inputId = "
prophetnumpred",
1092                 label = "h passos:",
1093                 value = 30, min = 1,
max = NA),
1094     plotlyOutput("plotpredcanvi")
1095 ),
1096     tabPanel("Taula de prediccions",
1097             dropdownButton(
1098                 p("Aquesta taula presenta els
valors predits.",
1099                 style="text-align: left"),
1100                 circle = TRUE,
1101                 status = "success",
1102                 size = "xs",
1103                 icon = icon("fa-duotone fa-info")
,
1104                 label = NULL,
1105                 tooltip = FALSE,
1106                 right = F,
1107                 up = FALSE,
1108                 width = NULL,
1109                 margin = "10px",

```

```

1110         inline = TRUE,
1111         inputId = NULL
1112     ),
1113     dataTableOutput("taulaprophet")),
1114     tabPanel("Components",
1115         dropdownButton(
1116             p("Podeu veure quina distribuci
1117
1118                 de les
1119                 algoritme. A
1120                 diferència de la metodologia d
1121                 'allisat exponencial,
1122                 aquest t en consideraci les
1123                 vacances o
1124                 dies festius del pa s que heu
1125                 indicat en la
1126                 guia d' s.",
1127                 style="text-align: left"),
1128             circle = TRUE,
1129             status = "success",
1130             size = "xs",
1131             icon = icon("fa-duotone fa-info")
1132         ),
1133         label = NULL,
1134         tooltip = FALSE,
1135         right = F,
1136         up = FALSE,
1137         width = NULL,
1138         margin = "10px",
1139         inline = TRUE,
1140         inputId = NULL
1141     ),
1142     plotOutput("components")),
1143     tabPanel("Cross Validation",
1144         dropdownButton(
1145             p("En estar treballant amb un
1146
1147                 model no
1148                 paramtric cal validar-lo amb
1149                 cross-validation.
1150                 En la taula es mostra els
1151                 resultats que s'han
1152                 obtingut d'aquest proc s.",br
1153                 (),
1154                 "Recordeu que no podreu validar
1155                 -lo si el
1156                 nombre de prediccions s major
1157                 al nombre

```

```

1145         d'observacions que presenta la
s rie temporal
1146         amb qu entreneu",
1147         style="text-align: left"),
1148         circle = TRUE,
1149         status = "success",
1150         size = "xs",
1151         icon = icon("fa-duotone fa-info")
,
1152         label = NULL,
1153         tooltip = FALSE,
1154         right = F,
1155         up = FALSE,
1156         width = NULL,
1157         margin = "10px",
1158         inline = TRUE,
1159         inputId = NULL
1160     ),
1161     dataTableOutput("cv")),
1162     tabPanel("Gr fic de Cross Validation",
1163             dropdownButton(
1164                 p("En aquesta pestanya es
visualitzen diferents
1165                 m triques per determinar si el
model creat
1166                 s prec s i prediu b .",
1167                 style="text-align: left"),
1168                 circle = TRUE,
1169                 status = "success",
1170                 size = "xs",
1171                 icon = icon("fa-duotone fa-info")
1172             ),
1173             label = NULL,
1174             tooltip = FALSE,
1175             right = F,
1176             up = FALSE,
1177             width = NULL,
1178             margin = "10px",
1179             inline = TRUE,
1180             inputId = NULL
1181         ),
1182         pickerInput(
1183             inputId = "metrica", label = "
Escull una
1184             m trica:",
1185             choices = c("mse","rmse","mae","
mape",

```

```

1185         "coverage"),
1186         options = list(
1187           style = "btn-success"),
1188         selected = "mape"),
1189         plotlyOutput("metplot")
1190       ),
1191       tabPanel("Taula de les mètriques CV",
1192         fluidRow(
1193           column(12, align = "center",
1194             tableOutput("mettable")
1195           )))
1196     )
1197   )
1198 )
1199 ),
1200 #####
1201
1202 ##### TAB7 #####
1203 tabItem(tabName = "bestmodel",
1204         fluidRow(
1205           box(title = "Comparaci entre models", id = "
1206 bmodels",
1207             fluidRow(
1208               column(12,
1209                 dropdownButton(
1210                   p("En primer lloc, observeu quin
1211 model s el millor
1212 per fer les prediccions amb els
1213 resultats de la
1214 taula. El millor model ser el
1215 que presenti un
1216 MAPE i RMPSE menor.",br(),"En
1217 segon lloc, indica
1218 el nmero del dia que vols predir
1219 comen ant des
1220 del final de la s rie. s a dir,
1221 si la s rie
1222 acaba el 31/12/2021 i vols la
1223 predicci del
1224 07/01/2022 cal que escrigueu en la
1225 casella 7.",
1226 br(),"Finalment, us apareixer el
1227 valor predit
1228 i el nivell d'AQI en qu us
1229 trobareu.",
1230           style="text-align: left"),
1231           circle = TRUE,

```

```

1221     status = "success",
1222     size = "xs",
1223     icon = icon("fa-duotone fa-info"),
1224     label = NULL,
1225     tooltip = FALSE,
1226     right = F,
1227     up = FALSE,
1228     width = NULL,
1229     margin = "10px",
1230     inline = TRUE,
1231     inputId = NULL
1232   ),
1233   fluidRow(
1234     column(12, align="center",
1235       tableOutput("metricasperf"),
1236       br(),
1237       fluidRow(
1238         column(6,
1239           pickerInput(
1240             inputId = "ara", label = "El millor
model
1241             per predir",
1242             choices = c("EWMA","ets","Arima","Auto.
arima",
1243                       "Prophet"),
1244             options = list(
1245               style = "btn-success"),
1246               selected = "Arima")
1247         ),
1248         column(6, align = "center",
1249           numericInput(inputId = "predicciofinal"
,
1250                       label = "h passos:",
1251                       value = 1, min = 1, max = NA)
1252         )
1253       ),
1254       uiOutput("predicciofinal")
1255     )
1256   )
1257 )
1258 )
1259 ),
1260   box(title = "L mits dels nivells de la qualitat de
l'aire", id = "bmodels",
1261     dropdownButton(
1262       p("Tingueu en compte que cada contaminant
presenta uns

```

```

1263         l mits diferents de concentraci per
determinar que
1264         pot ser perill s per la salut. Per defecte
hi ha els
1265         l mits de PM2.5. Podeu extreure els
l mits del seg ent
1266         enlla :",br(),a("https://www.epa.gov/
outdoor-air-quality-data/how-aqi-calculated"),
1267         style="text-align: left"),
1268         circle = TRUE,
1269         status = "success",
1270         size = "xs",
1271         icon = icon("fa-duotone fa-info"),
1272         label = NULL,
1273         tooltip = FALSE,
1274         right = F,
1275         up = FALSE,
1276         width = NULL,
1277         margin = "10px",
1278         inline = TRUE,
1279         inputId = NULL
1280     ),
1281     fluidRow(column(12,
1282         strong("L mit AQI bo (verd):", style="text-
align: left",
1283         numericInput(inputId = "green",
1284             label = NULL,
1285             value = 12, min = 0, max = NA,
1286             width = "30%")),
1287         #br(),
1288         strong("L mit AQI moderat (groc):",
1289             style="text-align: left",
1290         numericInput(inputId = "yellow",
1291             label = NULL,
1292             value = 35.4, min = 0, max = NA,
1293             width = "30%")),
1294         #br(),
1295         strong("L mit AQI no saludable per grups
sensibles (taronja):",
1296             style="text-align: left",
1297         numericInput(inputId = "orange",
1298             label = NULL,
1299             value = 55.4, min = 0, max = NA,
1300             width = "30%")),
1301         #br(),
1302         strong("L mit AQI no saludable (vermell):",
1303             style="text-align: left",

```

```

1304         numericInput(inputId = "red",
1305                       label = NULL,
1306                       value = 150.5, min = 0, max = NA
1307     ,
1308                       width = "30%")),
1309     #br(),
1310     strong("L mit AQI molt poc saludable (lila):",
1311           style="text-align: left",
1312           numericInput(inputId = "purple",
1313                         label = NULL,
1314                         value = 250.4, min = 0, max = NA
1315     ,
1316                         width = "30%")),
1317     #br(),
1318     strong("L mit AQI perill s (negre):",
1319           style="text-align: left",
1320           numericInput(inputId = "black",
1321                         label = NULL,
1322                         value = 500.4, min = 0, max = NA
1323     ,
1324                         width = "30%"))
1325   )
1326 )
1327 )
1328 ),
1329 #####
1330
1331 ##### TAB8 #####
1332 tabItem(tabName = "finaljo",
1333         fluidRow(
1334           tabBox(title = "Estudia el futur per solucionar-ho
1335 en el present", id = "final",
1336                 width = "100%",
1337                 height = "100%",
1338                 p(strong("Nora Amama Ben Hassun"), style="
1339 font-size:30px"),
1340                 p(em("Estudiant del grau d'estad stica UB-
1341 UPC",
1342                     style="font-size:18px;")),br(),
1343                 p("El meu nom s Nora Amama Ben Hassun i
1344                 s c estudiant
1345                 d'estad stica.",
1346                 style="font-size:16px; text-align: justify

```



```

"),
1343         p("La meva missi      s ajudar a les empreses
           i entitats p bliques
1344         a tenir al seu abast an lisis predictius
per anticipar-se
1345         als futurs canvis.",
1346         style="font-size:16px; text-align: justify
"),
1347         p(" S c encara estudiant i potser aix crea
           una mica de
1348         desconfian a pel que fa a les meves
capacitats. Pensareu...",
1349         em("Aquesta noia acaba de comen ar , no
           t experincia ,
1350         encara ha de treballar molt per saber
com funciona la
1351         vida..."),"I la veritat s que teniu
tota la ra , hi
1352         ha infinitats de coses que encara he d'
aprendre, per aix
1353         mai ha estat un impediment per descobrir,
cr ixer i
1354         aconseguir els nostres prop sits.",
1355         style="font-size:16px; text-align: justify
"),
1356         p("Sempre m'han agradat les matem tiques ,
jo he estat de
1357         les poques persones que esperava les
classes de matem tiques ,
1358         ja que em semblaven divertides. Els
problemes eren com un
1359         trencaclosques i obtenir la soluci
sempre m'ha fet feli .",
1360         style="font-size:16px; text-align: justify
"),
1361         p("En cr ixer i descobrir que una de les
mat ries que m s
1362         m'agradaven podia proporcionar-me
solucions pel futur,
1363         entendre i explicar que s el que est
passant en
1364         diferents mbits em va emocionar. Tinc
moltes ganes de
1365         participar i crear diferents projectes per
analitzar i
1366         d'aquesta manera aportar el meu granet de
sorra.",

```

```

1367         style="font-size:16px; text-align: justify
1368     " )
1369         )
1370     )
1371
1372     )#tabItems
1373 )#dashboardBody
1374 )

```