WRF Surface Layer Representation from MODIS and CORINE Land Use Datasets in SE Iberian Peninsula

Author: Alejandro Pujante Pérez
Supervisor: Mireia Udina Sistach, mudina@meteo.ub.edu
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Spain.

Abstract: Land use representation influences numerical weather models, particularly near the surface, where exchanges of heat and momentum are strongly influenced by terrain parameters. This study focuses on evaluating how different land use datasets—MODIS (MLU) and CORINE (CLC)—affect surface layer representation in the WRF model through its physical schemes and parameterizations. High-resolution simulations were performed over the southeastern Iberian Peninsula, comparing key near-surface variables: land surface temperature, 2-meter air temperature, and 10-meter air wind speed. Model outputs were validated against satellite products and in-situ station observations. Results show that land use determines surface energy fluxes, with CLC generally offering a more realistic representation and better agreement with observations. During daytime 2-meter air temperature seems to be highly influenced by changes in sensible and latent heat fluxes partitioning, while land surface temperature has a greater impact during nighttime.

I. INTRODUCTION

Numerical weather prediction (NWP) models have become an important tool in recent years with several applications in sectors such as agriculture, aviation, renewable energy, and urban planning. Numerical models are influenced by many surface factors, most importantly by orography, land cover, and soil moisture. The characterization of the land use (LU) influences especially the surface layer, which is the lowest part of the planetary boundary layer (PBL) and it is a crucial input to reproduce realistic simulations (*Boisier et al.* (2012)). Furthermore, land cover is constantly changing due to industrialization, built-up areas, and perturbation by humans (*Ellis et al.* (2010)). Therefore, in recent years, significant efforts have been made to generate more accurate and higher-resolution LU information driven by the increasing availability of satellite-based remote sensing.

The representation of land use regulates the available energy and the exchanges of heat, moisture, and momentum between the land surface and the atmosphere. These processes directly influence the calculation of meteorological variables, such as land surface temperature (LST), air temperature, and wind speed. LU data are typically organized into classification schemes that distinguish whether a region is covered by urban areas, forests, wetlands, croplands, or water bodies among others. Each LU category is further defined by a set of physical static parameters, including roughness length, leaf area index, albedo, emissivity, thermal inertia, soil moisture availability, and heat capacity. The characterization of LU has been shown to play a critical role in the performance of NWP models (*Li et al.* (2018a); *Schicker et al.* (2016); *Li et al.* (2018b)) and its effects are described by the surface layer and land surface physical parameterization schemes.

LU data commonly employed in NWP models include the United States Geological Survey (USGS) dataset (*Loveland et al.* (2000)), the MODIS Land Use dataset (MLU) (*Broxton et al.* (2014)), and, specifically for Europe, the CORINE Land Cover (CLC) dataset (*EEA* (2019)). Recent studies have investigated the sensitivity of the Weather Research and Forecasting (WRF) model to different LU representations (*de Bode et al.* (2023), *Schicker et al.* (2016)) and consistently highlight that more realistic, accurate, and higher resolution LU data significantly enhance model performance.

In this work, MLU and CLC datasets in WRF simulations are compared with the aim not only of identifying their impact on the simulation of key meteorological variables, but also of providing a physical explanation for the observed differences. The study focuses on the southeastern region of the Iberian Peninsula, an area of particular interest due to its diverse land surface characteristics, including a mix of soil types, extensive agricultural zones, and land–water interface regions such as the Mar Menor. This is achieved by analysing the representation of the surface layer through land surface temperature, 2-meter air temperature, and 10-meter air wind speed compared with both in-situ observations and satellite data focusing on the physical parameters that define each LU category. The central hypothesis is that the choice of LU dataset significantly influences surface–atmosphere exchanges of energy, with higher-resolution datasets expected to yield better results, as differences can be physically interpreted through the parameterizations of each LU classification.

The structure of this study is outlined as follows: initially, the paper introduces the methodology followed during the simulation tests in the section titled "Methodology". Subsequently, the sections "Land use representation and land surface temperature" and "Validation of the air surface temperature and wind" present the outcomes of these tests along with key statistical analyses. The "Discussion" section analyses the significance of these findings, comparing them with results from other authors. Finally, the conclusion section summarizes the key findings and contributions of the study and outlines potential areas for further research.

II. METHODOLOGY

A. WRF Model setup

In this study, the WRF (Weather Research and Forecasting) model version 4.6 (*Skamarock et al.* (2019)) was used to simulate meteorological variables over the study area. WRF is a non-hydrostatic regional NWP model developed mainly by NCAR (National Center for Atmospheric Research) and its use is not only for research purposes, but is also used for operational weather forecasting. Table I shows the general model configuration used in this work. A two-way nested domain setup over the selected area of interest was employed. The outer domain (d01) covers a broader region with coarser horizontal resolution of 3 km, while the inner domain (d02) is nested within d01 and provides finer horizontal resolution of 1 km. The choice of 1 km for the innermost domain corresponds to the limit for mesoscale models before explicit turbulence-resolving approaches (such as LES) are required. The area covered by the domains d01 and d02 (which cover the southern part of the Region of Murcia) is shown in Figure 1.

For initial and boundary conditions, the NCEP FNL reanalysis product has been used. The simulated period chosen for this work involves 14 days starting on September 15, 2024, at 00:00 UTC and ending on September 29, 2024, at 00:00 UTC. This period has been selected because at the synoptic level, it presents stable conditions (weather is anticyclonic), therefore the effects of the different land use characterization are more pronounced. Regarding the physical parameterization, *CONUS* suite has been used, which automatically selects the schemes for radiation, microphysics, cumulus, planetary boundary layer (PBL), surface layer and land surface model described in table I. This model configuration was used for two simulation experiments, changing only the static land use datasets explored in this work, MLU and CLC.

Table I: Model configuration used for WRF simulations. Domain, period of simulation and physical schemes options selected in this work. Sim 1 and Sim 2 represent the different simulations using the corresponding land use data: MLU means MODIS Land Use and CLC means CORINE Land Cover.

	d01	**d02**
Horizontal resolution	3 km	1 km
Dimensions (x, y, z)	(84, 60, 45)	(70, 49, 45)
Time step	18	6
Initial and boundary conditions	NCEP FNL d083002	
Simulated period		
Start	2024-09-15 00:00 UTC	
End	2024-09-29 00:00 UTC	
Radiation (ra_physics)	RRTMG-4	
Microphysics (mp_physics)	Thompson-8	
Cumulus (cu_physics)	Tiedtke-6	
PBL (bl_pbl_physics)	MYJ-2	
Surf. layer (sf_sfclay_physics)	MYJ-2	
Land surf. (sf_surface_physics)	NOAH-MP	
LU data Sim 1	MLU	MLU
LU data Sim 2	CLC	CLC

B. Land Use data

As mentioned in section I, land use data is one of the most important static data that has an impact on numerical models, since it regulates the exchange of heat and momentum between the soil and the air. A different representation of land use can have significant impacts on meteorological simulations. In general, land use description is based on different classifications. Each classification has a list of categories such as urban, croplands, shrublands, among others, which are characterized by parameters such as roughness length, surface emissivity, albedo and leaf area index (LAI) among others.

This work explores two different land use datasets, i) MODIS Land Use (MLU) which has a resolution of 1 km and classifies the categories using MODIFIED IGBP MODIS NOAH classification consisting of 21 categories and ii) CORINE Land Cover (CLC) developed by *EEA* (2019) which consists of 44 categories with 250 m of horizontal resolution. In addition, a reclassification process was applied to adapt the CLC dataset to the WRF USGS classification with its corresponding 28 categories following the methodology described by *Pineda et al.* (2004). This reclassification is essential because the WRF model requires land use data to conform to specific classification schemes, such as the USGS 28-category system.

The existing differences between MLU and CLC classifications and the spatial resolution (1 km in the case of MLU and 250 m for CLC) can have an impact on the representation of the land surface model in WRF, so we aim to compare the model performance under these two different land use datasets.

C. Land Surface Model

Changes in land use properties have several implications in the calculation of surface layer magnitudes, as represented in the physical parameterizations employed by numerical models, so we can determine what changes we can expect varying the surface parameters described in section II B. General WRF land-atmosphere interaction physics consist of surface layer (sfclay) and land surface model (LSM) schemes. The surface layer scheme determines exchange and transfer coefficients for heat and

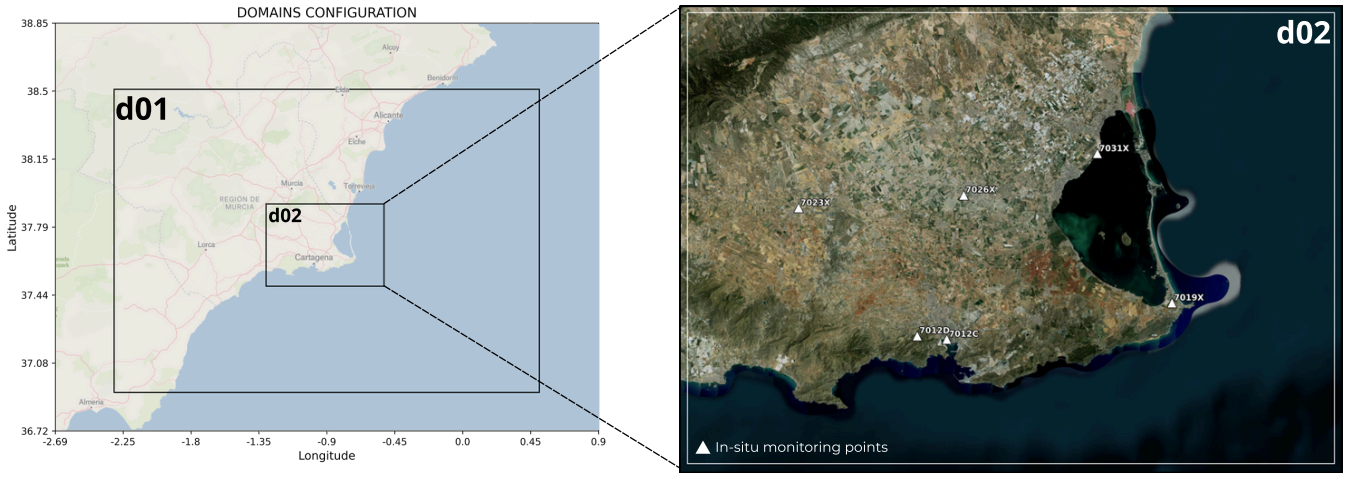

Figure 1: Domain design and spatial configuration with nested domains d01 (outer) and d02 (inner). White triangles represent observational points.

moisture to the LSM, which then provides land surface fluxes of heat and moisture to the planetary boundary layer (PBL). The surface schemes also provide friction stress and water-surface fluxes of heat and moisture to the PBL. Then, differences in land use characterization by different surface parameters could impact on magnitudes that the land surface model is calculating. In particular, the parameters we are exploring here are: albedo (α), roughness length (z_0), leaf area index (LAI) and emissivity (ϵ).

In order to understand how the surface parameters are affecting magnitudes that land surface calculates, the following key equations have been examined from *He et al.* (2023). The first one is the surface energy balance

$$S_{W\downarrow} - S_{W\uparrow} + L_{W\downarrow} - L_{W\uparrow} + H_{\text{pr,grd}} = SH + LH + GH \quad (1)$$

where S_W and L_W are the shortwave and longwave radiative fluxes, respectively, and arrows mean incoming $\downarrow$ and outgoing $\uparrow$ radiation from the surface. $H_{\text{pr,grd}}$ is the net precipitation heat flux advected to the bare ground. SH and LH are the ground sensible and latent heat fluxes, respectively, and GH is the ground heat flux. First, in equation 1 $S_{W\downarrow} = S_W(1-\alpha)$, which implies that albedo, that is the amount of solar radiation that the surface reflects, plays a significant role in the total available energy, and the terms on the right side of equation SH, LH and GH can be modified. In addition, $L_{W\uparrow}$ is described as

$$L_{W\uparrow} = \epsilon \sigma TSK^4 \quad (2)$$

where ϵ is the surface emissivity, σ is the Stefan-Boltzmann constant, and TSK is the land surface temperature. At the same time, sensible heat and latent heat fluxes are calculated based on the bulk transfer relationships as follows

$$SH = \rho_a \cdot C_h \cdot C_p \cdot U \cdot (\theta_s - \theta_a) \quad (3)$$

$$LH = \rho_a \cdot C_W \cdot C_{LH} \cdot U \cdot (q_s - q_a) \quad (4)$$

where ρ_a is the air density, C_p is the air heat capacity, C_{LH} is the specific latent heat of water vaporization, and U is the module of wind speed. θ_s and θ_a are the air potential temperatures at the surface and in the air, respectively. q_s and q_a are the specific humidity at the surface and in the air, respectively. Note that θ_s is TSK calculated from the model. C_h and C_W are the surface exchange coefficients for heat and moisture, respectively. In addition, coefficient C_h is calculated based on the Monin-Obukhov similarity theory, as:

$$C_h = \frac{\kappa^2}{\left[\ln\left(\frac{z-d_0}{z_0}\right) - \varphi_m\left(\frac{z-d_0}{L}\right)\right]\left[\ln\left(\frac{z-d_0}{z_0}\right) - \varphi_h\left(\frac{z-d_0}{L}\right)\right]}$$

where κ^2 is the von Karman constant, z is the height above the ground, d_0 is the zero displacement height, and L is the Monin-Obukhov length. φ_m and φ_h are the stability functions for momentum and heat transfer, respectively, and z_0 is the roughness length. It is important to note that z_o appears inside the logarithmic terms; therefore, an increase in the roughness length leads to a higher value of the transfer coefficient C_h. As a consequence, the sensible heat flux SH also increases following equation 3.

Roughness length, that is a measure of small-scale irregularities of a surface, has not only an impact on sensible heat flux, but also on wind speed. Equation 5 describes the vertical wind profile $u(z)$ under neutral conditions at the surface layer through a logarithmic semi-empirical law:

$$u(z) = \frac{u_*}{\kappa}\left[ln\left(\frac{z}{z_o}\right) + \psi\left(\frac{z}{L}\right)\right] \quad (5)$$

where u_* is the friction velocity which is related to the turbulent transport of horizontal momentum at the surface. Note that z_0 parameter will modify the surface layer wind profile. Thus, if z_0 increases (urban land use or mountains), wind will decrease more near the surface. This analysis will allow us to justify the differences in the simulation results using different land use data.

D. In-situ data

Observational in-situ data were obtained from the operational meteorological network of AEMET (Agencia Estatal de Meteorología), the national meteorological service of Spain. The data were collected via its public API service called AEMET OPEN DATA. Six stations were selected within the higher resolution domain d02 and named by the original station code from AEMET (see Table II). The dataset spans 14 days from the simulation period defined in section II A (2024-09-15 00:00 UTC to 2024-09-29 00:00 UTC). The stations temporal reporting resolution is hourly, and although multiple variables are recorded, only 2-meter air temperature and 10-meter wind speed were used for model evaluation, as they are directly influenced by land surface characteristics and measured according to WMO standards.

Table II: Geographic coordinates and altitude of meteorological stations used in this study.

Station code	Latitude	Longitude	Altitude (m)
7012C	37.60	-0.99	17
7031X	37.78	-0.81	4
7026X	37.74	-0.97	50
7023X	37.73	-1.17	140
7019X	37.64	-0.72	2.5
7012D	37.60	-1.02	55

E. Satellite data

Given the significant impact of land surface temperature LST represented with θ_s in Eq. 3 on the computation of SH as shown in section II C, it is important to assess the land surface temperature magnitude. Therefore, data from satellite MODIS have been employed in this work. MODIS incorporates a spectroradiometer that infers the surface temperature by assuming the Earth's surface behaves similarly to a blackbody emitter and using blackbody radiation theory. In this work, MODIS LST MOD11A1 product data acquired from the MODIS Terra sensor were employed to evaluate the land surface temperature. Data are available via the NASA AppEEARS portal. Although this measurement is not in situ, according to some recent research from *Wang et al.* (2019), the temperature bias between these data and the test station is approximately 1 K, which is accurate enough.

Since we aim to compare satellite data with the two simulation experiments using WRF, MLU and CLC land use datasets, some considerations have been made. During the study period, the Terra satellite crossed the equator in descending orbit at 10:30 local time due to its sun-synchronous orbit (which means that we have one value per day). This was taken into account in order to extract the same time stamp from the WRF model. The MODIS LST product has a 1 km resolution, which fits with the domain d02 horizontal resolution, but the points of MODIS LST do not correspond exactly with the WRF model grid. For this reason, the MODIS LST data were re-sampled to match the WRF d02 grid points in order to have the same shape in dimensions (time, x-direction, y-direction) in both MODIS LST and WRF datasets. In addition, data with cloud cover have been excluded. Once this process is done, the analysis was carried out for a 1-month simulation, starting 2024-09-15 00:00 until 2024-10-15 00:00, and the comparison was made by comparing each WRF TSK value with the corresponding MODIS LST measurement at each grid point within the d02 domain.

III. LAND USE REPRESENTATION AND LAND SURFACE TEMPERATURE

This section is structured into two parts. Firstly, there is the analysis of the land use representation over the study area for the two land use datasets employed in this work: MLU and CLC, with the aim of identifying the primary differences and patterns among them. The second part refers to a comparative study between the land surface temperature outputs from the WRF simulations (MLU and CLC) against the estimated LST from satellite MODIS.

A. Land use representation in the study area

In order to compare the differences between the two land use datasets, it is necessary first to analyse what kind of land use category the WRF model is using over the study area, because this category will determine the land surface parameters to run the simulations.

Figure 2 shows the spatial distribution of the land use categories for each dataset. The following similarities and differences can be observed: Domain d02 is in both cases divided by 11 land use categories. CLC seems to be more realistic than MLU since it represents better urban areas like the city of Cartagena (on the south), the shrublands area around the coast (lower left side) and also the needleleaf forest in the upper right side. Both detect the cultivation area in the centre of the domain in a similar way. In addition, CLC seems to better represent the continent-water interface zones. These results agree

with the spatial resolution of each land use dataset, since the CLC has higher resolution than MLU as mentioned in section II.

In addition, to carry out a more detailed analysis, Figure 3 presents the differences between the two datasets (CLC minus MLU) for several surface parameters: roughness length (Fig. 3a), albedo (Fig. 3b), emissivity (Fig. 3c), and leaf area index (LAI; Fig. 3d). The most notable differences in z_0 appear in urban and forested areas, where CLC exhibits higher values. For albedo, the largest discrepancies are found in the northwestern forested region of the domain and in urban areas. Regarding emissivity, CLC shows lower values along the coast, while presenting higher values over areas classified as cropland. The LAI parameter shows only minor differences, with isolated peaks of up to 0.3 in some locations.

Regarding the land use classification at the observational stations, Table III shows the land use categories assigned for each dataset: MLU and CLC. Additionally, since each category is characterized by physical parameters such as z_0, ϵ, LAI and α, Table III also includes the corresponding values for each dataset and their differences. Significant differences in roughness length (z_0) between datasets are observed, particularly in station 7023X, where CLC classifies the site as urban ($z_0 = 0.8$ m), while MLU assigns shrublands ($z_0 = 0.06$ m), resulting in a difference of 0.74 m. Similarly, in 7012D, CLC assigns a Shrubland/Grassland category with $z_0 = 0.06$ m, contrasting with the 0.8 m of the MLU Urban class. Differences in LAI and albedo are also notable. In station 7019X, CLC identifies a Wetland area with LAI = 0.053 and albedo = 0.14, whereas MLU classifies it as Barren, assigning LAI = 0 and albedo of 0.25. Finally, differences in emissivity are also remarkable. At station 7026X, CLC classifies the area as urban while MLU assigns croplands, leading to a difference in emissivity of 0.105, with higher values associated with MLU.

B. Comparison with satellite data

This section analyses the land surface temperature derived from WRF for each corresponding land use dataset (CLC and MLU) and how it compares with satellite data from MODIS. To assess the agreement between WRF surface temperature and MODIS LST observations, Table IV shows the statistical metrics obtained, including MAE, RMSE, and Pearson correlation coefficient r. These metrics were calculated separately according to the land use classification as defined by each dataset (CLC and MLU) and mentioned in section II. In both the MLU and CLC simulations, the urban and built-up category shows the best results in terms of MAE, achieving values of 2.03 K for MLU and 2.56 K for CLC, with the MLU simulation exhibiting superior performance. Regarding RMSE, the MLU exhibited 2.67 K and the CLC simulation registered a value of 3.33 K. On the other hand, shrublands category shows more advantageous results concerning the Pearson correlation coefficient r, with values of 0.57 for MLU and 0.59 for CLC. Finally, overall statistical results indicated a global MAE of 2.899 °C for the CLC simulation and 3.027 °C for the MLU simulation, showing that CLC has slightly better agreement with the MODIS LST product but not by a large difference.

In order to determine the agreement and linearity of land surface temperature, figure A1 shows different scatter-plots between WRF and MODIS LST surface temperature divided by main categories of each simulation dataset, such as cropland (Fig A1 a,b), urban areas (Fig A1 c,d), shrubland (Fig A1 e,f) and grassland (Fig A1 g,h). Overall, points lie under the line $y = x$, which means that MODIS LST shows lower values than WRF TSK in both MLU and CLC simulations. The effect is most pronounced in croplands for temperature ranges between 30 °C and 32.5 °C. The grassland category exhibits the highest concordance between MODIS and WRF temperatures with r = 0.56, indicating a relatively strong linear relationship in both simulations. Similarly, croplands also show relatively high correlations with r = 0.52 in the case of MLU and r = 0.54 for CLC, while urban areas present comparable correlation coefficients with r = 0.52 in the case of MLU and 0.50 in the case of CLC. Regarding the regression slopes, in both cases, it is close to one, indicating a good representation of land surface temperatures in these areas. Conversely, the shrubland category exhibits the weakest correlation with r = 0.32 in the case of MLU and r = 0.36 for CLC, which means a larger dispersion in the data, potentially indicating greater uncertainty in land use temperature estimates. Generally, the CLC simulation tends to yield higher correlations and regression slopes approaching one, in contrast to the MLU simulation, across most land use categories.

IV. VALIDATION OF THE AIR SURFACE TEMPERATURE AND WIND

In this section, the evaluation of the WRF model performance is done through comparison with observational data from six in-situ meteorological stations distributed across the area of study. The comparison was conducted over a 14-day simulation period, from 15 to 29 September 2024, as defined in Section II. As shown before, the two WRF simulations differ only in the land use dataset employed: one using MLU and the other CLC.

The analysis focuses on two key near-surface meteorological variables: air temperature at 2 metres and wind speed module at 10 meters. To better understand the origin of the differences observed between simulations using different land use datasets, diagnostic variables related to temperature and wind speed from some WRF surface fluxes were examined. These include the TSK variable and sensible and latent heat fluxes (SH and LH, respectively), which mainly influence surface layer temperature, as well as the friction velocity (u_*), which plays a role in

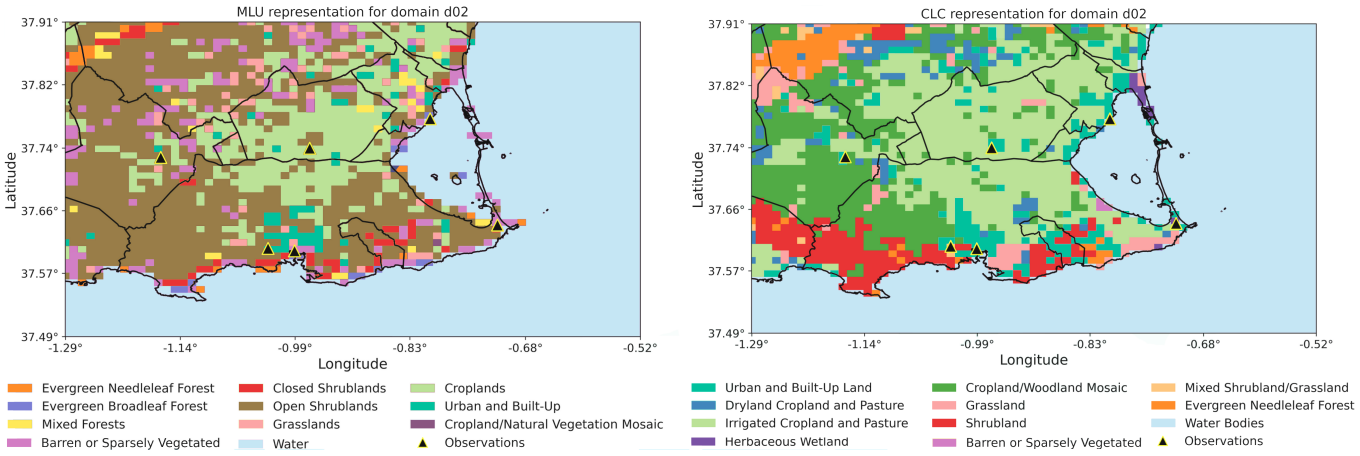

Figure 2: Land Use categories distribution over the domain d02 from the WRF model for the two datasets: MLU and CLC. Black triangles correspond with stations listed in Table II. Horizontal resolution of d02 is 1x1 km.

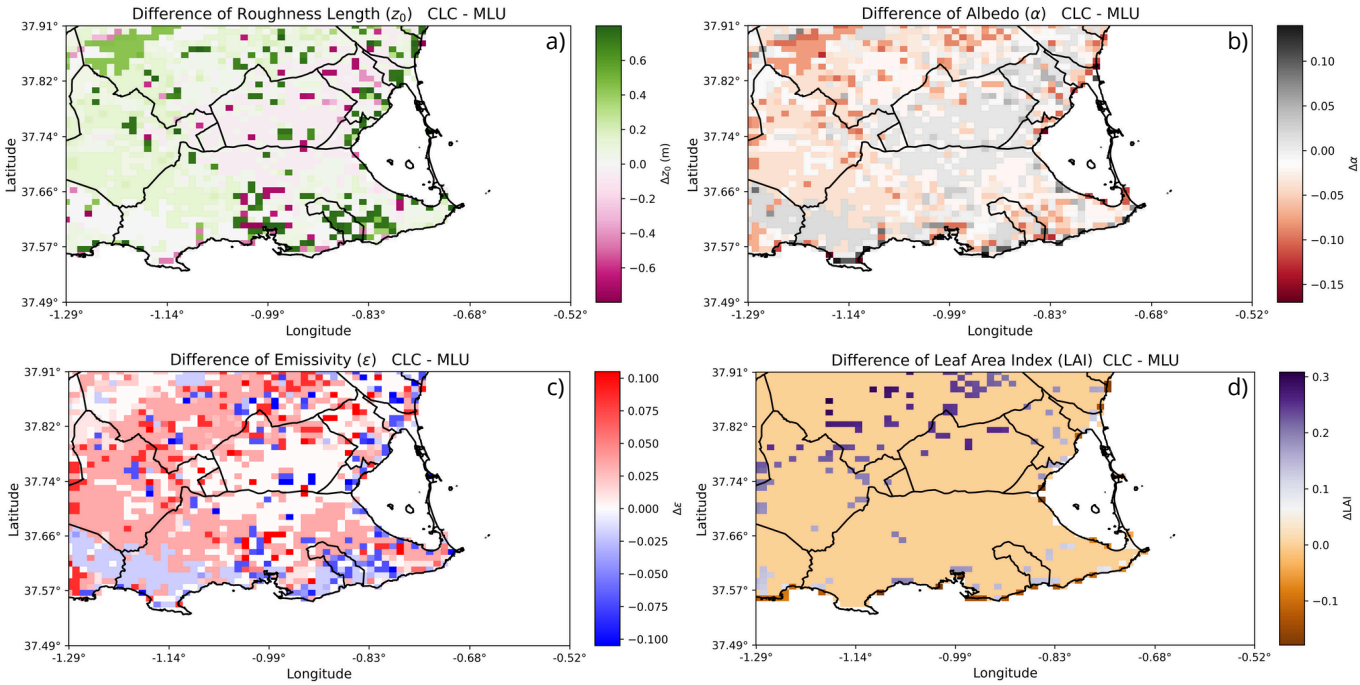

Figure 3: Spatial distribution of differences in land surface parameters between the CLC and MLU datasets over the study region (CLC - MLU). Positive values indicate higher parameter values in CLC compared to MLU.

modulating near-surface wind. The differences can also be explained in terms of the land use parameters shown in Table III such as z_0, LAI fraction, and albedo.

A. 2-m temperature and surface parameters

Figure 4 shows the time evolution of 2-meter air temperature, land surface temperature (TSK in the WRF model) sensible heat flux, and latent heat flux at station 7012C using the CLC and MLU land use datasets. CLC classifies the station point as an urban category, while MLU establishes barren land use. Results for daytime show that, in general, the CLC simulation reproduces air temperatures closer to observational data. In particular, MLU simulation tends to underestimate daytime temperatures. This fact can be explained by looking at surface heat fluxes. The sensible heat flux of the CLC simulation presents higher values than the MLU, which means that a greater portion of the available energy is being used to warm the near-surface air. Meanwhile, the latent heat flux in the case of the CLC simulation is zero, there is not any latent flux at this point, while it is up to 200 W/m^2 in the case of the MLU simulation, indicating that more

Table III: Land use assignation in each in-situ location for the two datasets MLU and CLC. ZNT-C and ZNT-M is the roughness length in each point for the CLC and MLU datasets and ΔZ is the difference. The same in the case of variables such emissivity, LAI fraction and albedo.

Station	LU-CLC	LU-MLU	Z-C	Z-M	ΔZ	EM-C	EM-M	ΔEM	LAI-C	LAI-M	ΔLAI	ALB-C	ALB-M	ΔALB
7023X	Urban	Shrublands	0.8	0.06	0.74	0.88	0.95	-0.07	0.19	0.19	0	0.15	0.2	-0.05
7012D	Shrub/Grass	Urban	0.06	0.8	-0.74	0.95	0.88	0.07	0.188	0.188	0	0.2	0.15	0.05
7012C	Urban	Barren	0.8	0.01	0.79	0.88	0.9	-0.02	0.181	0	0.181	0.15	0.25	-0.1
7019X	Wetland	Barren	0.2	0.01	0.19	0.95	0.9	0.05	0.053	0	0.053	0.14	0.25	-0.11
7031X	Urban	Evergreen	0.8	0.5	0.3	0.88	0.95	-0.07	0.12	0.12	0	0.15	0.12	0.03
7026X	Urban	Croplands	0.8	0.15	0.65	0.88	0.985	-0.105	0.26	0.26	0	0.15	0.17	-0.02

Table IV: Evaluation results for each land use category comparing WRF land surface temperature with MLU and CLC datasets against the satellite MODIS LST product.

MLU				CLC			
Land Use Category	**MAE**	**RMSE**	**r**	**Land Use Category**	**MAE**	**RMSE**	**r**
Evergreen Needleleaf Forest	6.93 K	7.50 K	0.50	Urban and Built-Up Land	**2.56 K**	**3.33 K**	0.50
Evergreen Broadleaf Forest	5.47 K	6.05 K	0.43	Dryland Cropland and Pasture	2.64 K	3.43 K	0.52
Mixed Forests	3.50 K	4.51 K	0.48	Irrigated Cropland and Pasture	2.94 K	3.59 K	0.55
Closed Shrublands	5.17 K	6.11 K	0.32	Cropland/Woodland Mosaic	2.67 K	3.50 K	0.52
Open Shrublands	3.34 K	4.19 K	**0.57**	Grassland	3.39 K	4.18 K	0.56
Grasslands	2.51 K	3.24 K	0.56	Shrubland	4.35 K	5.20 K	0.36
Croplands	2.45 K	3.15 K	0.52	Mixed Shrubland/Grassland	5.14 K	5.79 K	**0.59**
Urban and Built-Up	**2.03 K**	**2.67 K**	0.52	Evergreen Needleleaf Forest	6.83 K	7.37 K	0.59
Cropland/Natural Veg. Mosaic	3.61 K	4.39 K	0.40	Water Bodies	4.45 K	4.97 K	0.20
Barren or Sparsely Vegetated	4.19 K	4.77 K	0.42	Herbaceous Wetland	6.45 K	6.93 K	0.32
Water	6.47 K	6.90 K	0.24	Barren or Sparsely Vegetated	2.18 K	2.62 K	0.48

energy is being invested in phase changes rather than directly heating the air. Due to energy conservation as exposed in equation 1, these fluxes are balanced, so an increase in one flux necessarily implies a reduction in the other one (excluding the small fraction of ground flux). Thus, the higher latent heat flux in the MLU simulation results in less energy available for sensible heating, which explains the lower near-surface air temperatures observed in that case. This fact can also be explained by examining the surface parameters shown in Table III for station 7012C and the equations described in Section II C. For MLU, the albedo is higher than in CLC, which implies lower $S_{W\downarrow}$ and therefore less available energy according to the equation 1. The roughness length (z_0) also contributes to this effect, as MLU has a lower z_0, according to the equation for the coefficient C_h described in section II C, this clearly results in a lower SH. During nighttime in this station, the MLU simulation reproduces lower temperatures, while the CLC matches the observed values more closely. Since sensible and latent heat fluxes are near zero at night in both MLU and CLC, this difference must be attributed to land surface temperature. In the case of MLU, the skin temperature is lower than in CLC, which leads to a lower 2-m air temperature. Additionally, this behaviour can be physically explained by differences in emissivity: for MLU, $\epsilon = 0.9$, while for CLC, $\epsilon = 0.88$. According to Equation 2, higher emissivity values lead to lower land surface temperatures.

B. 10-m wind and momentum fluxes

Figure 5 shows the time evolution of 10-m wind speed at station 7026X and the friction velocity for simulations using the CLC and MLU land use datasets during the simulation period. The MLU simulation overestimates the wind speed compared to both the CLC simulation and observations from the meteorological station, especially in the peaks of wind speed. On the other hand, CLC simulation performs better compared with observational data. As well as in the case of temperature, these differences can be explained based on the surface properties assigned in each land use dataset, and particularly in the roughness length z_0. As shown in Table III, MLU dataset assigns the point located at station 7026X to the croplands category , meanwhile in the case of CLC this point is classified as urban. The z_0 parameter is higher in the case of urban terrain (CLC) with 0.5 m than the croplands category (MLU) with z_0 value of 0.0692, this difference in land use categories gives a $\Delta z_0 = 0.43$. The result of this discrepancy can be noted in the time evolution of friction velocity, where CLC simulation shows higher values of u_* over the entire period. This means that the lower roughness in the MLU simulation results in reduced surface drag, allowing the wind to retain more of its momentum and giving higher wind speeds at 10 meters.

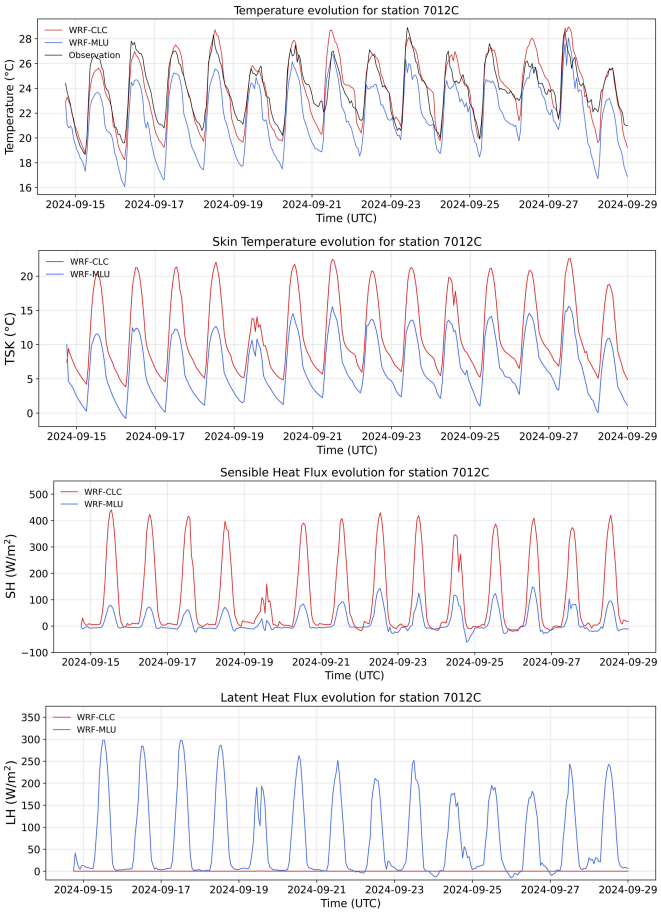

Figure 4: 2-meter air temperature, sensible heat flux (SH) and latent heat flux (LH) evolution during the simulation period for station 7012C.

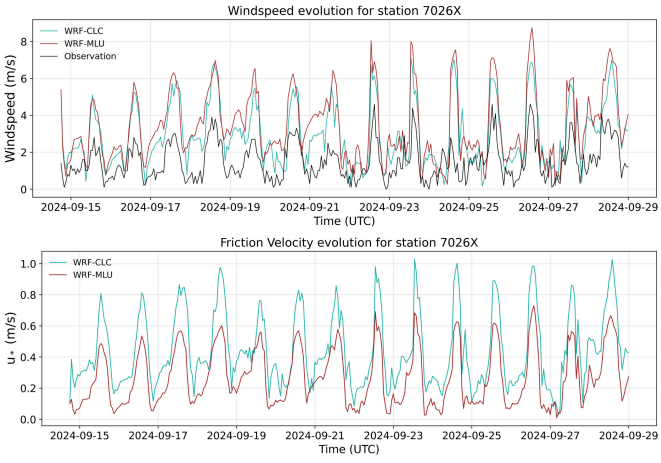

Figure 5: Module of wind and friction velocity (u_*) evolution during the simulation period for station 7026X.

To evaluate the global model performance under different land use scenarios, statistical analysis was carried out comparing observational data with both MLU and CLC simulations in each one of the in-situ locations. Statistical parameters such as MAE, RMSE, Pearson r, Mean Bias, and RBias (%) were obtained and are shown in Table V.

In general, results show that the WRF model presents a better agreement with observational data when using the CLC dataset compared to MLU, particularly for temperature simulations, where the CLC configuration yields lower MAE and RMSE values across most stations, and a higher correlation coefficient r Pearson. For instance, at station 7012C, the temperature MAE and RMSE under CLC are significantly lower (0.831°C and 1.064°C, respectively) than with MLU (2.182°C and 2.428°C).

For wind speed, the CLC simulation presents better results in terms of MAE, RMSE and MBias, indicating that the difference between observation and observational values in case of CLC is lower than MLU simulation. In contrast, MLU showed better results in terms of r Pearson correlation in 4 of the 6 stations.

Table V: Evaluation results for in-situ data. Statistic parameters MAE, RMSE, MBias and r comparison of temperature and wind speed for each in-situ location. Results marked in bold mean the best one.

Stat.	Var	LU	MAE	RMSE	MBias	r
7012C	Temp	CLC	**0.831 K**	**1.064 K**	**-0.060 K**	**0.921**
		MLU	2.182 K	2.428 K	-2.130 K	0.896
	Wind	CLC	**1.487 m/s**	**1.779 m/s**	**1.372 m/s**	0.716
		MLU	1.879 m/s	2.209 m/s	1.816 m/s	**0.746**
7012D	Temp	CLC	**1.058 K**	**1.364 K**	**-0.113 K**	**0.897**
		MLU	1.127 K	1.504 K	-0.117 K	0.872
	Wind	MLU	**1.515 m/s**	**1.987 m/s**	**1.354 m/s**	0.715
		CLC	2.017 m/s	2.505 m/s	1.887 m/s	**0.766**
7019X	Temp	CLC	**0.814 K**	**1.123 K**	0.151 K	**0.882**
		MLU	2.154 K	2.463 K	**-2.068 K**	0.809
	Wind	CLC	**1.182 m/s**	1.608 m/s	0.109 m/s	0.706
		MLU	1.188 m/s	**1.564 m/s**	**-0.105 m/s**	**0.737**
7026X	Temp	CLC	1.236 K	1.595 K	0.892 K	**0.951**
		MLU	**1.167 K**	**1.411 K**	**-0.414 K**	0.936
	Wind	CLC	**1.509 m/s**	**1.785 m/s**	**1.411 m/s**	**0.771**
		MLU	1.939 m/s	2.276 m/s	1.882 m/s	0.766
7031X	Temp	CLC	**1.058 K**	**1.356 K**	0.385 K	**0.897**
		MLU	1.436 K	1.742 K	**-1.049 K**	0.878
	Wind	CLC	**1.187 m/s**	**1.588 m/s**	**-0.698 m/s**	0.657
		MLU	1.313 m/s	1.704 m/s	-1.052 m/s	**0.677**

V. DISCUSSION

Our initial hypothesis has been confirmed, as the results show that CLC (the higher-resolution dataset) provides a more realistic representation of land use categories than MLU (Fig 2), consistent with the findings of *Bachantourian et al.* (2022) and *Schicker et al.* (2016). In MLU, some categories are missing, while CLC captures them well, particularly in urban areas. These differences significantly affect roughness length, albedo, and emissivity, although not much influence on LAI. The TSK simulated by the WRF model generally agrees with the MODIS LST values. Compared to *Kadaverugu* (2023), our results show consistently higher or comparable values. For urban areas, we obtained r = 0.52 (CLC), improving upon their January values (r = –0.07 to 0.48). In croplands, our correlations (r = 0.52 for MLU and 0.55 for CLC) exceed those reported by the same authors (r = 0.11–0.44). For forests, we reached up to r = 0.59 (CLC), compared to 0.23–0.58 reported in *Kadaverugu* (2023).

Regarding the 2-meter air temperature, it was found that SH and LH differences significantly impact daily values, while TSK is the most relevant variable for the nightime period. CLC tends to produce better results than MLU in terms of r, MAE, RMSE, and MBias across most in situ stations, which aligns with the findings of *Golzio et al.* (2021), where lower MAE and RMSE values were reported for CLC compared to MLU. For 10-meter wind speed, changes in the z_0 parameter impacted the friction velocity u_*, and consequently, on wind speed. This leads to a reduction in the systematic overestimation of wind speed peaks also identified by *Wang et al.* (2025) for all PBL schemes at velocities above 5 m/s. Similarly to the 2-meter air temperature results, CLC outperformed MLU when compared with observations in terms of r, MAE, RMSE, and MBias in most stations, consistent with the results of *Golzio et al.* (2021).

However, this study has some limitations to be considered. First, the simulation period spans only 14 days within a single season, limiting the assessment of the model's sensitivity to LU changes across different seasonal conditions. Second, the geographical scope is confined to the southeastern Iberian Peninsula, reducing the generalizability of the results to other regions with distinct surface characteristics. Third, the use of only six in-situ meteorological stations could be improved by incorporating additional sites. Finally, the comparison with satellite data was limited to a single daily value. Future improvements could involve the integration of in situ LST observations, which would allow a more accurate representation of the daily evolution.

VI. CONCLUSIONS AND FUTURE WORK

This work assesses the impact of land use input for the surface layer representation in the mesoscale meteorological model WRF from different land use data: i) MODIS Land Use (MLU) and ii) CORINE Land Cover (CLC). The methodology uses both in-situ data from meteorological stations and satellite observations to validate different surface layer magnitudes such as land surface temperature, 2-meter air temperature, and 10-meter air wind speed over a 14 days simulation in the southeastern Iberian Peninsula. We can highlight some conclusions and point out directions for future research.

- Accurate higher-resolution land use representation increases NWP performance. In particular, CLC shows improved representation of urban areas, shrublands and evergreen needleleaf forest categories, and yields better agreement with both in-situ and satellite observations for land surface temperature, 2-meter air temperature and 10-meter wind speed when compared to MLU in terms of statistical parameters MAE, RMSE, MBias and r.
- WRF TSK generally overestimates MODIS LST, with best agreement found in urban and cropland areas; overall, CLC simulations present slightly higher correlation and lower errors than MLU.
- For 2-meter air temperature, daily values are mainly influenced by sensible and latent heat fluxes partitioning, while land surface temperature has a greater impact on nightime periods. CLC fits better with observations due to a changes in surface parameters as albedo and emissivity.
- A more accurate representation of the surface roughness length leads to a reduction in the overestimation of 10-meter wind speed due to a differences in friction velocities.
- Higher-resolution and more up-to-date land use datasets, such as CLC, improve the model's ability to represent transition zones between land cover types, especially in heterogeneous areas like urban, where multiple land uses coexist at small scales, often in close proximity to other categories and subject to rapid temporal changes. Finer spatial resolution helps the model assign more representative surface categories in these complex environments.

One potential extension of this study would be to explore the impact of land use characterization under explicit large-scale turbulence-resolving using WRF-LES and improving surface layer representation. Finally, although some authors have already applied this approach to air quality studies *Tao et al.* (2018), there is limited literature in the state of the art. Applying this methodology to chemical transport models such as CHIMERE-WRF (*Menut et al.* (2024)) represents a promising research direction, since meteorological magnitudes have an influence on chemical reactions and pollutant dispersion.

ACKNOWLEDGMENTS

I would like to sincerely thank my supervisor Mireia Udina, for her continuous support and involvement during this work, for sharing with me the knowledge in micrometeorology and for the many productive discussions we had while analysing the results. I am also grateful to my parents and my friends for their constant support and encouragement along this challenging year.

REFERENCES

Agencia Estatal de Meteorología (AEMET), Aemet OpenData portal, `https://opendata.aemet.es`, accessed: 2025-02, 2024.

Bachantourian, M., K. Chaleplis, A. Gemitzi, K. Kalabokidis, P. Palaiologou, and C. Vasilakos, Evaluation of MODIS, climate change initiative, and CORINE land cover products based on a ground truth dataset in a mediterranean landscape, *Land*, *11*(9), 1453, 2022.

Boisier, J., N. de Noblet-Ducoudré, A. Pitman, F. Cruz, C. Delire, B. Van den Hurk, M. Van Der Molen, C. Müller, and A. Voldoire, Attributing the impacts of land-cover changes in temperate regions on surface temperature and heat fluxes to specific causes: Results from the first LUCID set of simulations, *Journal of Geophysical Research: Atmospheres*, *117*(D12), 2012.

Broxton, P. D., X. Zeng, D. Sulla-Menashe, and P. A. Troch, A global land cover climatology using MODIS data, *Journal of Applied Meteorology and Climatology*, *53*(6), 1593–1605, 2014.

de Bode, M., T. Hedde, P. Roubin, and P. Durand, A method to improve land use representation for weather simulations based on high-resolution data sets—application to corine land cover data in the WRF model, *Earth and Space Science*, *10*(2), e2021EA002,123, 2023.

EEA, E., CORINE Land Cover 2018 (Raster 100 m), Europe, 6-Yearly—Version 2020_20u1, May 2020, 2019.

Ellis, E. C., K. Klein Goldewijk, S. Siebert, D. Lightman, and N. Ramankutty, Anthropogenic transformation of the biomes, 1700 to 2000, *Global ecology and biogeography*, *19*(5), 589–606, 2010.

Golzio, A., S. Ferrarese, C. Cassardo, G. A. Diolaiuti, and M. Pelfini, Land-use improvements in the weather research and forecasting model over complex mountainous terrain and comparison of different grid sizes, *Boundary-Layer Meteorology*, *180*(2), 319–351, 2021.

He, C., et al., The community Noah-MP land surface modeling system technical description version 5.0, *Tech. rep.*, NCAR Technical Note NCAR/TN-575+ STR, doi: 10.5065/ew8g-yr95, 2023.

Jiménez-Esteve, B., M. Udina, M. R. Soler, N. Pepin, and J. R. Miró, Land use and topography influence in a complex terrain area: A high resolution mesoscale modelling study over the Eastern Pyrenees using the WRF model, *Atmospheric Research*, *202*, 49–62, 2018.

Kadaverugu, R., A comparison between WRF-simulated and observed surface meteorological variables across varying land cover and urbanization in south-central India, *Earth Science Informatics*, *16*(1), 147–163, 2023.

Li, H., M. Wolter, X. Wang, and S. Sodoudi, Impact of land cover data on the simulation of urban heat island for Berlin using WRF coupled with bulk approach of Noah-LSM, *Theoretical and applied climatology*, *134*, 67–81, 2018a.

Li, Y., C. Zhao, T. Zhang, W. Wang, H. Duan, Y. Liu, Y. Ren, and Z. Pu, Impacts of land-use data on the simulation of surface air temperature in Northwest China, *Journal of Meteorological Research*, *32*(6), 896–908, 2018b.

Loveland, T. R., B. C. Reed, J. F. Brown, D. O. Ohlen, Z. Zhu, L. Yang, and J. W. Merchant, Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, *International journal of remote sensing*, *21*(6-7), 1303–1330, 2000.

Menut, L., A. Cholakian, R. Pennel, G. Siour, S. Mailler, M. Valari, L. Lugon, and Y. Meurdesoif, The CHIMERE chemistry-transport model v2023r1, *Geoscientific Model Development Discussions*, *2024*, 1–44, 2024.

NASA Land Processes Distributed Active Archive Center (LP DAAC), Application for extracting and exploring analysis ready samples (appeears), `https://appeears.earthdatacloud.nasa.gov/`, accessed: 2025-04, 2025.

Pineda, N., O. Jorba, J. Jorge, and J. Baldasano, Using NOAA AVHRR and SPOT VGT data to estimate surface parameters: application to a mesoscale meteorological model, *International journal of remote sensing*, *25*(1), 129–143, 2004.

Schicker, I., D. Arnold Arias, and P. Seibert, Influences of updated land-use datasets on WRF simulations for two Austrian regions, *Meteorology and Atmospheric Physics*, *128*(3), 279–301, 2016.

Skamarock, W. C., et al., A description of the advanced research WRF model version 4, *National Center for Atmospheric Research: Boulder, CO, USA*, *145*(145), 550, 2019.

Tao, H., J. Xing, H. Zhou, X. Chang, G. Li, L. Chen, and J. Li, Impacts of land use and land cover change on regional meteorology and air quality over the Beijing-Tianjin-Hebei region, China, *Atmospheric environment*, *189*, 9–21, 2018.

Wang, Q., B. Zeng, G. Chen, and Y. Li, Simulation performance of planetary boundary layer schemes in WRF v4. 3.1 for near-surface wind over the western Sichuan Basin: a single-site assessment, *Geoscientific Model Development*, *18*(5), 1769–1784, 2025.

Wang, T., J. Shi, Y. Ma, L. Husi, E. Comyn-Platt, D. Ji, T. Zhao, and C. Xiong, Recovering land surface temperature under cloudy skies considering the solar-cloud-satellite geometry: Application to MODIS and Landsat-8 data, *Journal of Geophysical Research: Atmospheres*, *124*(6), 3401–3416, 2019.

VII. APPENDIX

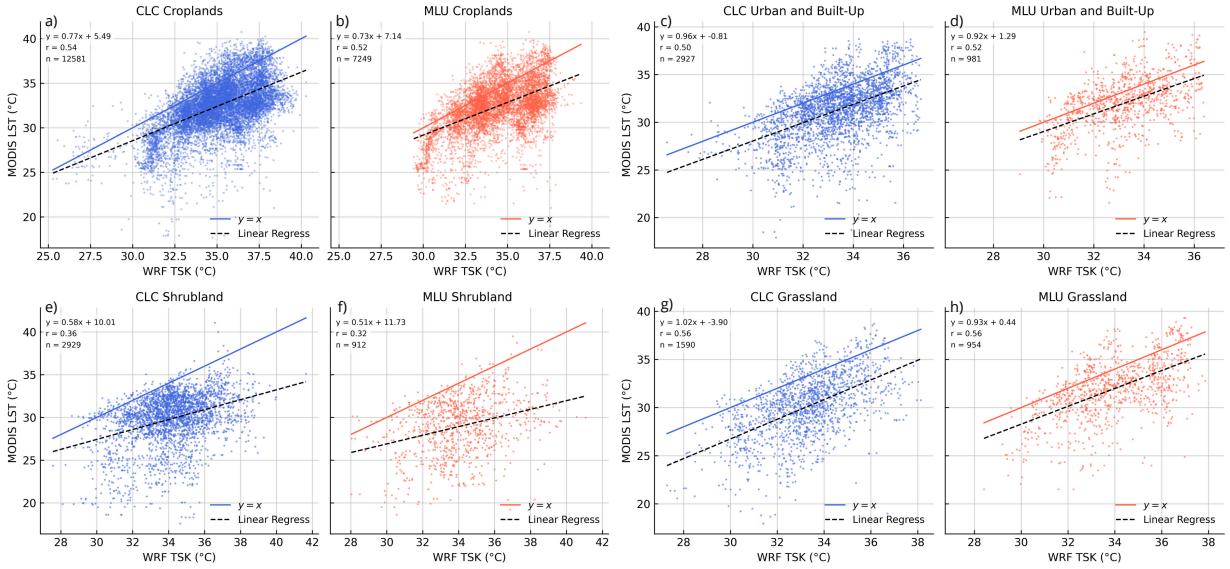

Figure A1: Scatter plot comparing WRF model surface skin temperature (TSK) and MODIS LST for different land use categories (croplands, built-up, shrublands, and grassland areas) and land use dataset (CLC and MLU). Linear regression and $y = x$ lines are included for each category.