Università degli Studi di Trento

Doctoral School in
Civil, Environmental and Mechanical Engineering
36th cycle

Machine learning-based sensitivity analysis of surface parameters in numerical weather prediction model simulations over complex terrain

Supervisor:	**Doctoral Student:**
Prof. Lorenzo Giovannini	**Dario Di Santo**

Session X

Academic Year 2023/2024

Abstract

Land surface models (LSMs) implemented in numerical weather prediction (NWP) models use several parameters to suitably describe the surface and its interaction with the atmosphere, whose determination is often affected by many uncertainties, strongly influencing simulation results. However, the sensitivity of meteorological model results to these parameters has not yet been studied systematically, especially in complex terrain, where uncertainty is expected to be even larger. This work aims at identifying critical LSM parameters influencing the results of NWP models, focusing in particular on the simulation of thermally-driven circulations over complex terrain. While previous sensitivity analyses employed offline LSM simulations to evaluate the sensitivity to surface parameters, this study adopts an online coupled approach, utilizing the Noah-MP LSM within the Weather Research and Forecasting (WRF) model. To overcome computational constraints, a novel tool, Machine Learning-based Automated Multi-method Parameter Sensitivity and Importance analysis Tool (ML-AMPSIT), is developed and tested. This tool allows users to explore the sensitivity of the results to model parameters using supervised machine learning regression algorithms, including Random Forest, CART, XGBoost, SVM, LASSO, Gaussian Process Regression, and Bayesian Ridge Regression. These algorithms serve as fast surrogate models, greatly accelerating sensitivity analyses while maintaining a high level of accuracy. The versatility and effectiveness of ML-AMPSIT enable the fast implementation of advanced sensitivity methods, such as the Sobol method, overcoming the computational limitations encountered in expensive models like WRF. The suitability of this tool to assess model's sensitivity to the variation of specific parameters is first tested in an idealized sea breeze case study where six surface parameters are varied. Then, the analysis focuses on the evaluation of the sensitivity to surface parameters in the simulation of thermally-driven circulations in a mountain valley. Specifically, an idealized three-dimensional topography consisting of a valley-plain system is adopted, analyzing a complete diurnal cycle of valley and slope winds. The analysis focuses on all the key surface parameters governing the interactions between Noah-MP and WRF. The proposed approach, novel in the context of LSM-NWP model coupling, draws from established applications of machine learning in various Earth science disciplines, underscoring its potential to improve the estimation of parameter sensitivities in NWP models.

Contents

List of Figures

List of Tables

List of Acronyms

Introduction

Numerical Weather Prediction (NWP) models are indispensable tools in modern meteorology, providing insights into atmospheric processes and aiding in weather forecasting. A critical component of these models is the accurate representation of land surface processes. These processes, occurring at the bottom boundary of the atmospheric model, involve complex interactions between the land surface and the atmosphere, including the exchange of heat and mass fluxes. Land Surface Models (LSMs) are integrated into NWP models to simulate these interactions.

One of the most advanced LSMs frequently utilized in NWP is Noah-MP (Noah with Multi-Physics options). One of the most remarkable peculiarities of Noah-MP with respect to other LSMs is to consider the vegetation canopy as a separate layer that takes into account the leaf orientation with respect to the radiation, the space between crowns and leaves, the height and thickness of the canopies, to account for important phenomena such as radiation scattering, trapping and penetration in the canopy, and mutual shading. It also includes the treatment of all the canopy and soil processes relevant for the computation of water and energy budget, with the additional option to account for the carbon budget. Similarly to other LSMs, Noah-MP employs an ensemble of empirical parameterizations to compute these balances at the land surface. Moreover, the model uses more than 40 parameters to describe land cover and characterize its effects on surface-atmosphere exchange. Despite its comprehensive approach, Noah-MP relies on fixed values for each parameter, not accounting for the inherent variability within broad land cover categories (e.g. mixed forests). This limitation is particularly evident in complex terrain, where land cover can rapidly vary with changes in altitude, leading to a significant degree of parameter heterogeneity.

Recent studies have underscored that LSMs are a major source of uncertainty in simulating near-surface variables, especially over complex terrain (e.g., Schmidli, Billings, et al., 2011; Massey et al., 2014; Tomasi et al., 2017). This uncertainty arises from several factors, including the accuracy of the parameterizations used, their general applicability across different landforms, the representativeness of parameter values for specific regions and the grid resolution. Another possible source of errors is the lack of updated datasets.

The unique challenges posed by complex terrain areas, such as mountains, further exacerbate these uncertainties. In these regions, the topography itself acts as a physical barrier, affecting atmospheric flows and introducing additional dynamics compared to flat terrain. The geometries involved in complex terrain also give rise to thermally driven circulations, that can only be observed in these regions due to differential heating and the consequent formation of pressure gradients (Figure 1).

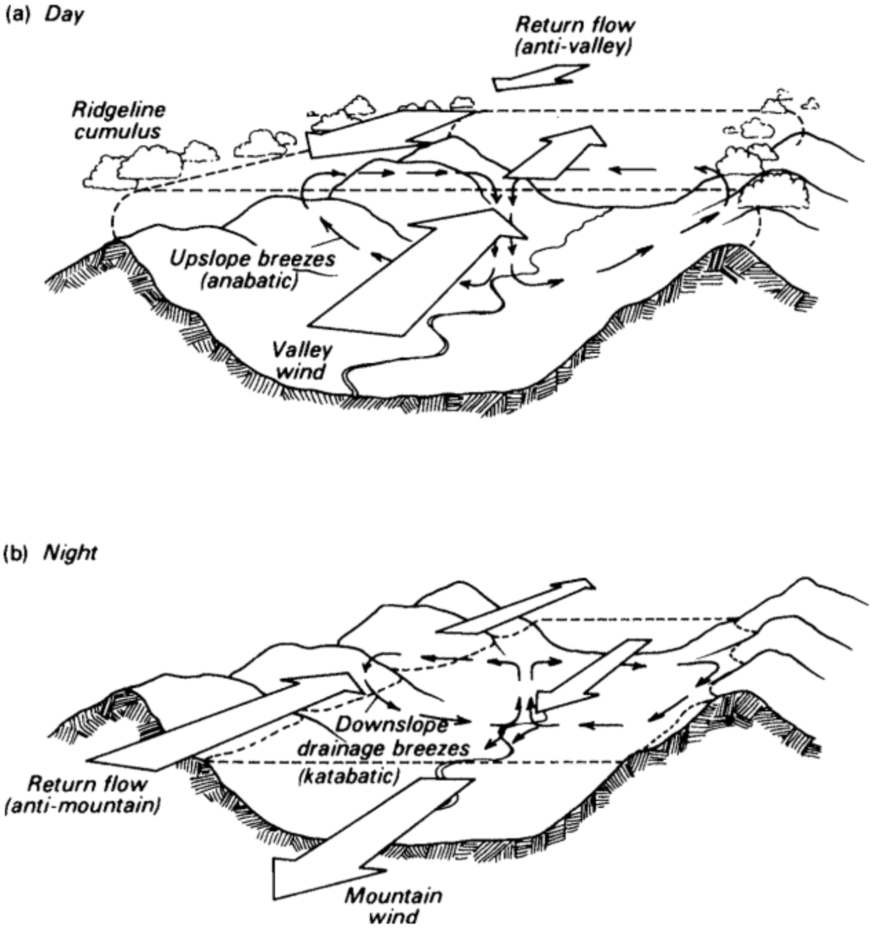


Figure 1: Diurnal (a) and nocturnal (b) atmospheric circulation in the valley, from Monteith et al., 1979.

In addition to this, LSMs implemented in meteorological models are typically single-column models, meaning they do not account for interactions between neighboring grid cells, including lateral flow. This is particularly problematic in mountainous regions, where it can significantly impact the simulation in areas characterized by strong spatial heterogeneity, for example in sparsely vegetated sites (Zhang, Chen, et al., 2020) or with spatio-temporal soil moisture variability (Krauss et al., 2010).

Moreover, the computation of surface fluxes in LSMs often relies on scaling relations based on Monin-Obukhov similarity theory. The assumptions of this theory, including quasi-stationarity and horizontal homogeneity, are frequently violated in the boundary layer of mountainous terrain, leading to discrepancies in model simulations.

Each of these sources of uncertainty can have a large impact on the final output if the model is highly sensitive to these errors, i.e., if the output is highly sensitive to parameterization-related inputs. Hence, there is a need to assess the sensitivity before interpreting the model output.

This thesis aims to investigate the sensitivity of LSM parameters in NWP models, with a specific focus on complex terrain. The objectives include quantifying the relative sensitivity of the model output to the LSM parameters that go into the forcing process calculation, evaluating their influence over complex terrain, and identifying parameters and processes with a significant impact on numerical simulations. Where possible, recommendations for optimal parameter selection will be provided.

To address these objectives, the use of idealized simulations is proposed. These simulations are preferable as they allow for the isolation of parameter-induced variations, free from confounding factors such as issues related to initial/boundary conditions, observation uncertainties, and unrelated phenomena. The idealized simulations performed for this thesis focus on synoptically undisturbed conditions in mountainous terrain. The sensitivity of the model results is evaluated with particular emphasis on thermally driven circulations, analyzing near-surface fields and vertical profiles. A first test is performed for a sea breeze system over flat terrain, then the same methodology is applied to a simplified topography inspired by other studies in the literature (Schmidli, Billings, et al., 2011), representing a valley-plain system with a horizontal inner valley floor and characteristic dimensions representative of the main Alpine valleys.

The model used for the simulations is the Weather Research and Forecasting (WRF) model. WRF is an open-source tool for both research and operational applications, widely used around the globe. The dynamical core (ARW) solves the flux-form Euler equations for a compressible non-hydrostatic fluid in terrain-following coordinates. The equations are solved by decomposing each term in its hydrostatic part plus a perturbation, with forcings coming from dedicated physics modules for radiation, microphysics, turbulent mixing, etc. It features a range of refined numerical techniques for time and space discretization, including a time-split method with an outer third-order Runge-Kutta integration scheme and an inner loop for acoustic modes. The spatial discretization is staggered with an Arakawa-C grid, and the accuracy order can be chosen by the user up to the fifth order for horizontal and vertical advection. It can employ numerical diffusion terms to smooth out turbulent eddies resulting in more stable simulations.

The sensitivity analysis methodology involves assessing the impact of small perturbations in input parameters on the model output. However, traditional one-at-a-time (OAT) sensitivity methods, which treat parameters as independent from each other, can be misleading due to potential interactions between factors.

Moreover, when dealing with one or two parameters, it might be feasible to perform an OAT sensitivity analysis where the output response (the "response curve") can be easily assessed by eye, but when dealing with multiple input parameters, the problem becomes high-dimensional, and the response curve becomes a hypersurface that is very difficult to assess with simple methods. In light of this, and since it is rare to know a priori the magnitude of parameters' interactions, there is a need for a more refined approach.

An advanced method to perform sensitivity analyses is the Sobol method, which can be categorized as a variance-based method. Its goal is to compute sensitivity indices, the Sobol indices, obtained by decomposing the variance V of the model output into multiple components accounting for the individual parameter contribute (first-order term) and its interaction with the other parameters (higher-order terms). This method allows evaluating if a parameter is involved in strong interactions, or if it is a noninfluential factor that can be fixed at any value without affecting the output variance, i.e. how much it is possible to reduce the variance of the output V if that parameter is fixed.

Although being an advanced method, the Sobol method is also known for its very high computational cost. For simple and fast models, this method has been used in the literature with more than 10000 runs (e.g. Cuntz, Mai, Zink, et al., 2015; Cuntz, Mai, Samaniego, et al., 2016). For more complex and computationally expensive models such as WRF, this may be unfeasible. The minimum number of simulations required to obtain consistent sensitivity indices also depends on the specificity of each problem and cannot be known a priori. Therefore, given the computational intensity of WRF and the need for a large number of simulations, alternative approaches are considered.

The concept of surrogate models has become very popular in fluid dynamics (Queipo et al., 2005; Fernández-Godino et al., 2017; Taira et al., 2017; Vinuesa et al., 2022; Longo et al., 2020; Lamberti et al., 2021), where data-driven models have been used in the last few years in methods such as proper orthogonal decomposition (POD), dynamical modes decomposition (DMD), and reduced order models (ROM). All these methods employ a data-driven approach to make predictions based on data coming typically from a small ensemble of high-fidelity, more computationally intensive simulations. The goal of these models is to build a faster surrogate model to reduce the computational burden.

Among the data-driven techniques used for building surrogate models, Machine Learning (ML) methods are increasingly being used in various fields where the evaluation of the sensitivity of parameters, usually referred to as "predictors" or "model features", is a crucial task for problems with high demands in terms of precision, such as system control engineering (Queipo et al., 2005; Brunton, Noack, et al., 2020; Brunton and Kutz, 2019), medical and genetic fields (Sharma et al., 2014; Cui et al., 2016; Cui et al., 2016; Antonogeorgos et al., 2009; Maroco et al., 2011; Yang, Wang, et al., 2022), but also in earth science fields such as those studying the susceptibility of landslides and fire risk (Yilmaz, 2010; Catani et al., 2013; Pradhan, 2013; Youssef et al., 2016; Kalantar et al., 2018; Lee et al., 2018; Chen, Zhu, et al., 2020; Daviran et al., 2023; Elia et al., 2023; Oliveira et al., 2012; Bar Massada et al., 2013; Arpaci et al., 2014; Pourtaghi et al., 2016; Satir et al., 2016; Gigović et al., 2019). At the cost of less interpretability compared to ROMs, machine learning is usually a faster method for model surrogation, making it attractive for complex models such as NWPs, where computational power is critical.

An example of the use of this approach in climate sciences can be found in Dagon et al., 2020, where an ML-based tool for parameter calibration developed by the Community Land Model (CLM) team from the National Center for Atmospheric Research (NCAR) is presented. The Authors found that an ensemble with perturbed land model parameters, in which each simulation has a duration of approximately 2 hours, can be replaced by the results of an artificial neural network algorithm that uses 2.6 s to generate predictions, with a correlation over 90% between original and surrogate model values.

In light of the above discussion, the present thesis proposes a sensitivity analysis tool based on ML algorithms (ML-AMPSIT) which has been developed in collaboration with the Noah-MP group at NCAR. This tool aims to guide users from start to end in a comprehensive sensitivity analysis, utilizing ML data-driven methods to obtain predictions based on simulation data, mixing the use of ML-based regression and Sobol-based sensitivity. While it currently focuses on weighting model parameters, its versatility allows for potential future applications in various other tasks such as spatial/temporal interpolation and prediction of observables.

Currently, the ML-AMPSIT implements the following regression algorithms: Random Forest (RF), Least Absolute Shrinkage and Selection Operator (LASSO), Support Vector Machine (SVM), Bayesian Ridge Regression (BRR), Gaussian Process Regression (GPR), Extreme Gradient Boosting (XGboost), and Classification And Regression Trees (CART). Most of these regression algorithms provide regression weights by default, while BRR and GPR are probabilistic methods that are used within ML-AMPSIT to compute sensitivity indices via an implementation of the Sobol method.

The rationale for adopting a multi-method approach in this analysis tool stems from the recognition that different regression algorithms, despite aiming to achieve the same goal, operate based on fundamentally distinct principles. For instance, LASSO employs a least squares approach with regularization, making it well-suited for linear problem-solving. In contrast, algorithms like Random Forest are based on decision tree ensembles and are inherently more suited at capturing non-linear relationships.

This diversity in algorithmic approaches is particularly valuable in scenarios where the underlying characteristics of the problem are not clearly understood a priori. For example, in cases dominated by linear relationships, methods like LASSO might offer more accurate predictions. Conversely, for problems with complex, non-linear dynamics, a method like Random Forest might potentially yield better results.

Given this context, ML-AMPSIT allows users to explore a variety of methodologies, providing a comprehensive perspective on the problem at hand. After evaluating the performance of different algorithms, users have the flexibility to either average the results for a more balanced view or select the best-performing method based on their specific dataset and analysis goals. This multi-faceted approach equips users with the necessary tools to effectively tackle a wide range of problems, including those that are complex or not fully understood, thereby enhancing the robustness and reliability of the analysis.

As mentioned above, the case study used to test the performance of the new tool refers to the simulation of a sea breeze over flat terrain with the systematic variation of six parameters related to the vegetation canopy. The results show that most of the algorithms are able to replicate the original model, yielding predictions with correlations above 90% with respect to the original model values. Notably, these results are obtained reaching convergence after only about 20 simulations used for training.

After validating the ML-AMPSIT tool, the work focuses on thermally driven winds in an idealized valley. All the relevant Noah-MP vegetation-related parameters for the chosen configuration are considered for the sensitivity analysis, for a total of 21 parameters.

The outline of the thesis is as follows:

Chapter 1 provides an overview of the methods used in this work. It begins with a detailed description of the Noah-MP land surface model, focusing on its treatment of canopy effects on radiation transfer, heat and momentum fluxes, and a rapid overview of the different options to parameterize physical processes. This is followed by an introduction to the Weather Research and Forecasting model. The chapter then discusses sensitivity analysis methods, including sampling design, computational cost, global sensitivity analysis, and the Sobol method. The final part of the chapter discusses data-driven modeling, the variance-bias tradeoff, and various machine learning algorithms employed in ML-AMPSIT.

Chapter 2 reviews the state of the art in idealized simulations of the valley wind system, sensitivity analysis applied to land surface models, and machine learning-based sensitivity analysis.

Chapter 3 presents the research article "ML-AMPSIT, a machine learning-based automated multi-method parameter sensitivity and importance analysis tool". The paper delves into the ML-AMPSIT workflow, the Sobol method, and the implemented ML algorithms. A case study is presented, focusing on the WRF/Noah-MP model and its setup. The sea breeze ensemble results, ML-AMPSIT output, convergence analysis, parameter importance analysis, and vertical variability are discussed.

Chapter 4 presents the research article "Machine learning-based sensitivity analysis of surface parameters in numerical simulations of thermally-driven circulations in a mountain valley". The modeling setup and the Noah-MP parameters chosen for the analysis are first described. Then, the results obtained, considering in particular the simulation of the valley circulation and the output of the sensitivity analysis, are presented.

Finally, the thesis concludes with a summary of the results, their implications, and suggestions for future research in the field of land surface modeling and numerical weather prediction, as well as further potential developments of the ML-AMPSIT tool.

CHAPTER 1

Methods

1.1 Noah-MP

The early LSMs, emerging around four decades ago, were relatively simplistic in their approach to representing land surface processes. These models primarily focused on the energy balance equation, utilizing basic assumptions like prescribed albedo, constant soil moisture, and often neglecting soil heat storage (Bonan, 2008). Their approach to simulate sensible and latent heat fluxes was guided by bulk aerodynamic transfer equations. Vegetation's impact on latent heat flux was indirectly included via a β factor, which modified potential evaporation to reflect limited soil moisture conditions (Bonan, 2008). The models' hydrology was rudimentarily represented using a simple bucket model (Manabe, 1969).

The advancement to second-generation models was led by the work of Deardorff (1978), introducing a separate vegetation canopy layer and a two-layer soil column. This development allowed for differentiated calculations of canopy and ground temperatures and distinct turbulent and radiative fluxes. These models, like BATS (Dickinson, Henderson-sellers, Kennedy, and Wilson, 1986) and SiB (Sellers et al., 1986), were more complex, incorporating detailed radiative transfer calculations and differentiating soil evapotranspiration mechanisms. They marked a significant leap in land surface modeling, integrating data from satellite products for vegetation and soil characteristics (Bonan, 2008).

Third-generation LSMs advanced further, utilizing biochemical models of photosynthesis (Farquhar et al., 1980) and semi-empirical models for stomatal conductance (Ball et al., 1987b; Leuning, 1990; Collatz et al., 1992). These models enhanced the simulation of atmospheric CO_2, integrating more detailed responses of stomata to environmental factors.

The Noah LSM was a second-generation model widely used in weather and climate predictions, representing a collaborative effort across multiple institutions (Mahrt et al., 1984; Schaake et al., 1996a; Schaake et al., 1996b; Ek et al., 2003).

This model, however, faced limitations in its structure, particularly in representing snowpack dynamics and soil moisture processes, which affected its accuracy in various environmental simulations (Barlage et al., 2010).

To address these limitations, significant augmentations were introduced in the following third-generation version, Noah-MP. These included a vegetation canopy layer, a modified two-stream radiation transfer scheme, a Ball-Berry type stomatal resistance scheme, and a dynamic vegetation model (Yang and Friedl, 2003; Niu and Yang, 2004; Ball et al., 1987a; Dickinson, Henderson-sellers, Kennedy, and Wilson, 1986). Additionally, the model was improved with a simple groundwater model, a three-layer snow model, and a frozen soil scheme, enhancing its permeability and overall accuracy (Niu, Yang, Dickinson, and Gulden, 2005; Niu, Yang, Dickinson, Gulden, and Su, 2007; Yang and Friedl, 2003).

Noah-MP is currently one of the most used LSMs and has been implemented inside many of the principal weather and hydrological models available today, such as the NCAR WRF model, the NOAA National Water Model (NWM), the NASA Land Information System (LIS), the NCAR High-Resolution Land Data Assimilation System (HRLDAS), the NOAA Unified Forecast System (UFS), as well as the new NCAR Model for Prediction Across Scales (MPAS).

In the next sections, the main components of Noah-MP are briefly presented and discussed.

1.1.1 Canopy effects on radiation transfer

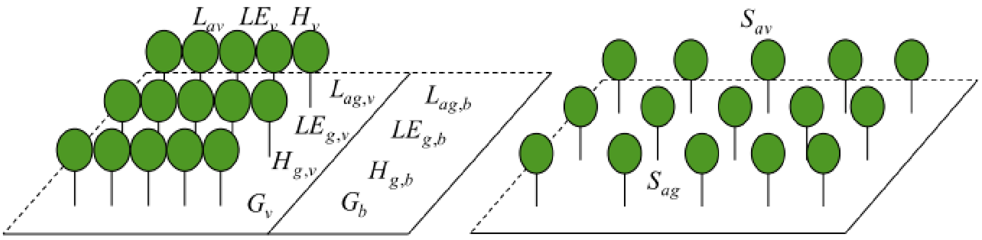


Figure 1.1: Semi-tile treatment for radiative and heat fluxes in Noah-MP, from Niu, Yang, Mitchell, et al., 2011.

Noah-MP uses a semi-tile subgrid scheme (Figure 1.1) to separate the fractional vegetated area from the fractional bare ground area in the computation of longwave radiation, latent heat, sensible heat, and ground heat fluxes. The shortwave radiation transfer is computed across the entire grid cell, taking into account probabilities of gaps between and within canopy clusters to include both scattering and multiple reflections of radiative fluxes. The arrangement and size of these gaps are dependent on factors such as the canopy's radius and density, the spacing between tree trunks, and the solar zenith angle (SZA). This approximation assumes a uniform distribution of canopies over the grid cell, calculating surface-absorbed and ground-absorbed solar radiation. The scheme allows for the utilization of different vegetation fractions in energy budget computations.

The between-crown gap probability and the within-crown gap probability are computed from the geometrical properties of the canopy summarized in Figure 1.2, defined by the crown radius R, the top and bottom of the canopy H_{top} and H_{bottom}, the SZA θ, the vegetation fraction F_{veg}, and the foliage volume density F_a.

The gaps computation contributes to the definition of the direct and diffuse radiation that interacts with the canopy, the optical depths, and the upscattering parameters, leading to the total radiation flux f_{ab} absorbed by the vegetation expressed as:

$$f_{ab} = 1 - f_r \left(1 - \alpha_d\right) - f_{td} \left(1 - \alpha_i\right) \cdot f_{ti} \tag{1.1}$$

where f_r is the total radiation flux reflected by the vegetation and the ground, f_{td} and f_{ti} are the downward direct and diffuse radiation fluxes below the vegetation, α_d and α_i are direct and diffuse albedos of the underlying soil, functions of the soil moisture.

The effects of multiple reflections and absorption inside the canopy emulate the radiation trapping effects observed in real canopies, which significantly reduces the amount of radiation reaching the ground, as shown in Figure 1.3.

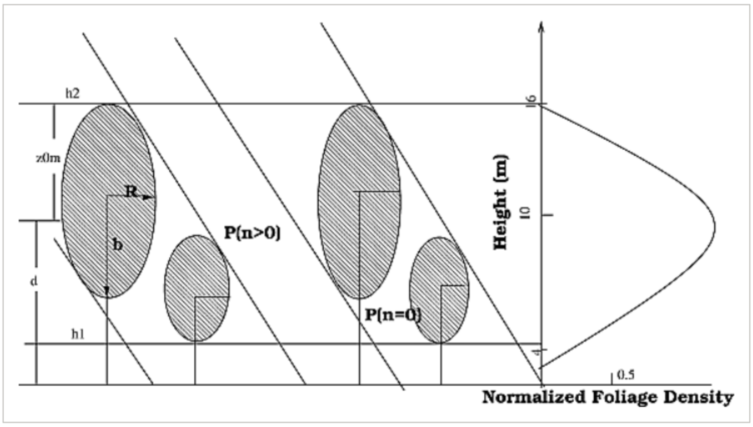


Figure 1.2: Canopy 3d structure in Noah-MP radiation treatment, from Yang and Friedl, 2003.

Energy fluxes are calculated differently for the bare ground fraction, the vegetated ground fraction, and the vegetation canopy.

Using subscripts b, g, and v for bare, vegetated ground, and canopy respectively, the total fluxes entering the energy budget are:

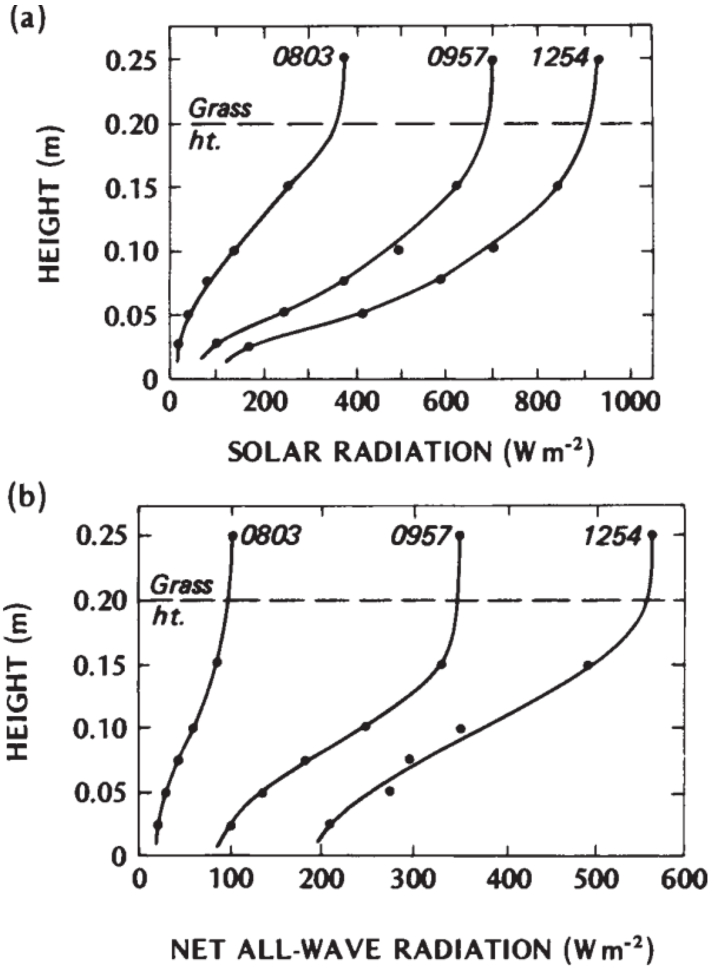


Figure 1.3: Dampening effect of vegetation over solar radiation (a) and net all-wave radiation (b) at 3 different times, from Monteith et al., 1979.

$$L_a = (1 - F_{veg})L_{a,b} + F_{veg}(L_{av} + L_{ag}) \tag{1.2}$$

$$LE = (1 - F_{veg})LE_b + F_{veg}(LE_v + LE_g) \tag{1.3}$$

$$H = (1 - F_{veg})H_b + F_{veg}(H_v + H_g) \tag{1.4}$$

$$G = (1 - F_{veg})G_b + F_{veg}G_v \tag{1.5}$$

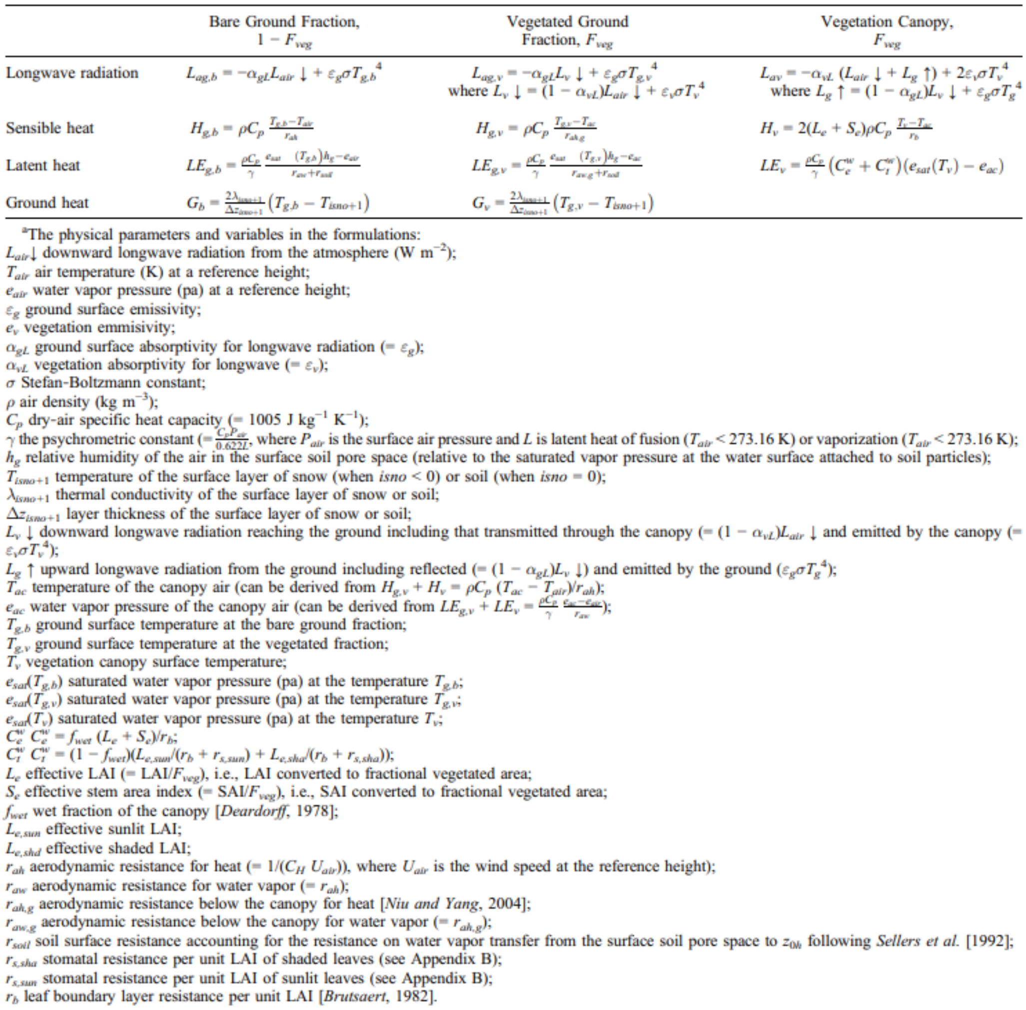
where L_a is the net longwave radiation, LE is the total latent heat flux, H is the total sensible heat flux, G is and the total ground heat flux.

The surface energy balance is:

$$S_{av} + S_{ag} = L_a + LE + H + G \tag{1.6}$$

where S_{av} and S_{ag} are the solar radiation component absorbed by vegetation and ground.

Each component in the Equations (1.2)-(1.5) is computed as summarized in Figure 1.4. Sensible and latent heat fluxes follow bulk formulae as a function of temperature and water vapor pressure, modulated by the thermal resistances of air, soil, stomata, and leaf boundary layer, in analogy with the electrical circuit representation. Ground heat is also a function of the surface layer thickness and thermal conductivity.



	Bare Ground Fraction, $1 - F_{veg}$	Vegetated Ground Fraction, F_{veg}	Vegetation Canopy, F_{veg}
Longwave radiation	$L_{ag,b} = -\alpha_{gL} L_{air}\downarrow + \varepsilon_g \sigma T_{g,b}{}^4$	$L_{ag,v} = -\alpha_{gL} L_v \downarrow + \varepsilon_g \sigma T_{g,v}{}^4$ where $L_v \downarrow = (1 - \alpha_{vL}) L_{air}\downarrow + \varepsilon_v \sigma T_v{}^4$	$L_{av} = -\alpha_{vL}\,(L_{air}\downarrow + L_g \uparrow) + 2\varepsilon_v \sigma T_v{}^4$ where $L_g \uparrow = (1 - \alpha_{gL}) L_v \downarrow + \varepsilon_g \sigma T_g{}^4$
Sensible heat	$H_{g,b} = \rho C_p \frac{T_{g,b} - T_{air}}{r_{ah}}$	$H_{g,v} = \rho C_p \frac{T_{g,v} - T_{ac}}{r_{ah,g}}$	$H_v = 2(L_e + S_e)\rho C_p \frac{T_v - T_{ac}}{r_b}$
Latent heat	$LE_{g,b} = \frac{\rho C_p}{\gamma} \frac{e_{sat}\ (T_{g,b}) h_g - e_{air}}{r_{aw} + r_{soil}}$	$LE_{g,v} = \frac{\rho C_p}{\gamma} \frac{e_{sat}\ (T_{g,v}) h_g - e_{ac}}{r_{aw,g} + r_{soil}}$	$LE_v = \frac{\rho C_p}{\gamma}\left(C_e^w + C_t^w\right)\left(e_{sat}(T_v) - e_{ac}\right)$
Ground heat	$G_b = \frac{2\lambda_{isno+1}}{\Delta z_{isno+1}}\left(T_{g,b} - T_{isno+1}\right)$	$G_v = \frac{2\lambda_{isno+1}}{\Delta z_{isno+1}}\left(T_{g,v} - T_{isno+1}\right)$	

[a]The physical parameters and variables in the formulations:
$L_{air}\downarrow$ downward longwave radiation from the atmosphere (W m^{-2});
T_{air} air temperature (K) at a reference height;
e_{air} water vapor pressure (pa) at a reference height;
ε_g ground surface emissivity;
e_v vegetation emmisivity;
α_{gL} ground surface absorptivity for longwave radiation ($= \varepsilon_g$);
α_{vL} vegetation absorptivity for longwave ($= \varepsilon_v$);
σ Stefan-Boltzmann constant;
ρ air density (kg m^{-3});
C_p dry-air specific heat capacity ($=$ 1005 J kg^{-1} K^{-1});
γ the psychrometric constant ($= \frac{C_p P_{air}}{0.622L}$, where P_{air} is the surface air pressure and L is latent heat of fusion ($T_{air} < 273.16$ K) or vaporization ($T_{air} < 273.16$ K);
h_g relative humidity of the air in the surface soil pore space (relative to the saturated vapor pressure at the water surface attached to soil particles);
T_{isno+1} temperature of the surface layer of snow (when $isno < 0$) or soil (when $isno = 0$);
λ_{isno+1} thermal conductivity of the surface layer of snow or soil;
Δz_{isno+1} layer thickness of the surface layer of snow or soil;
$L_v \downarrow$ downward longwave radiation reaching the ground including that transmitted through the canopy ($= (1 - \alpha_{vL}) L_{air}\downarrow$ and emitted by the canopy ($= \varepsilon_v \sigma T_v{}^4$);
$L_g \uparrow$ upward longwave radiation from the ground including reflected ($= (1 - \alpha_{gL}) L_v \downarrow$) and emitted by the ground ($\varepsilon_g \sigma T_g{}^4$);
T_{ac} temperature of the canopy air (can be derived from $H_{g,v} + H_v = \rho C_p\,(T_{ac} - T_{air})/r_{ah}$);
e_{ac} water vapor pressure of the canopy air (can be derived from $LE_{g,v} + LE_v = \frac{\rho C_p}{\gamma}\frac{e_{ac} - e_{air}}{r_{aw}}$);
$T_{g,b}$ ground surface temperature at the bare ground fraction;
$T_{g,v}$ ground surface temperature at the vegetated fraction;
T_v vegetation canopy surface temperature;
$e_{sat}(T_{g,b})$ saturated water vapor pressure (pa) at the temperature $T_{g,b}$;
$e_{sat}(T_{g,v})$ saturated water vapor pressure (pa) at the temperature $T_{g,v}$;
$e_{sat}(T_v)$ saturated water vapor pressure (pa) at the temperature T_v;
C_e^w $C_e^w = f_{wet}\,(L_e + S_e)/r_b$;
C_t^w $C_t^w = (1 - f_{wet})(L_{e,sun}/(r_b + r_{s,sun}) + L_{e,sha}/(r_b + r_{s,sha}))$;
L_e effective LAI ($=$ LAI/F_{veg}), i.e., LAI converted to fractional vegetated area;
S_e effective stem area index ($=$ SAI/F_{veg}), i.e., SAI converted to fractional vegetated area;
f_{wet} wet fraction of the canopy [*Deardorff*, 1978];
$L_{e,sun}$ effective sunlit LAI;
$L_{e,shd}$ effective shaded LAI;
r_{ah} aerodynamic resistance for heat ($= 1/(C_H\, U_{air})$), where U_{air} is the wind speed at the reference height);
r_{aw} aerodynamic resistance for water vapor ($= r_{ah}$);
$r_{ah,g}$ aerodynamic resistance below the canopy for heat [*Niu and Yang*, 2004];
$r_{aw,g}$ aerodynamic resistance below the canopy for water vapor ($= r_{ah,g}$);
r_{soil} soil surface resistance accounting for the resistance on water vapor transfer from the surface soil pore space to z_{0h} following *Sellers et al.* [1992];
$r_{s,sha}$ stomatal resistance per unit LAI of shaded leaves (see Appendix B);
$r_{s,sun}$ stomatal resistance per unit LAI of sunlit leaves (see Appendix B);
r_b leaf boundary layer resistance per unit LAI [*Brutsaert*, 1982].

Figure 1.4: Derivation of quantities used in fluxes calculation, from Niu, Yang, Mitchell, et al., 2011.

The radiation is computed for both the near-infrared and visible bands, accounting for the strong dependence on the wavelength of the vegetation reflectivity, transmissivity, and absorptivity observed in nature, as reported in Figure 1.5.

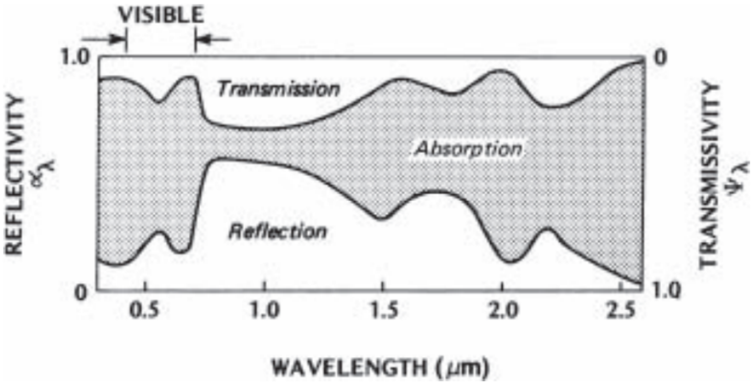


Figure 1.5: Leaf reflectivity, transmissivity and absorption, as functions of wavelength, from Monteith et al., 1979.

1.1.2 Canopy effects on heat and momentum fluxes

At the canopy top H_{can} the wind speed is derived from the logarithmic law:

$$U_c = \frac{u_*}{k} \ln \left(\frac{H_{\text{can}} - d}{z_{0M}} \right) \tag{1.7}$$

where u_* is the friction velocity:

$$u_* = \frac{U_R k}{\left[\ln \left(\frac{z_{\text{ref}} - d}{z_{0M}} \right) - \psi_m \left(\frac{z_{\text{ref}} - d}{L_{MO}} \right) \right]} \tag{1.8}$$

k is the von Kármán constant, H_{can} is the height of the canopy top, z_{0M} is the momentum roughness length, $d = 0.65 \cdot H_{can}$ is the zero-plane displacement height, U_R is the velocity at reference height $z_{ref} = d + z'_a$, where z'_a is the first level height above the surface zero plane received from the host model. ψ_m is the stability function for momentum from Monin-Obukhov Similarity Theory (MOST, Monin 1954) and L_{MO} is the Monin-Obukhov length.

Heat and momentum fluxes are modulated by resistances in a two-layer model, as shown in Figure 1.6. Between the reference height and the height $d + z_{0M}$, the aerodynamic resistances for heat and momentum are (Bonan, 1996):

$$ra_{hc} = \frac{1}{\kappa u_*} \left[\ln \left(\frac{z_{\text{ref}} - d}{z_{0M}} \right) - \psi_h \left(\frac{z_{\text{ref}} - d}{L_{MO}} \right) \right] \tag{1.9}$$

$$ra_{mc} = \frac{U_R}{u_*^2} \tag{1.10}$$

while, under the canopy, the resistances are modified to account for the canopy extinction (Brutsaert, 1982a):

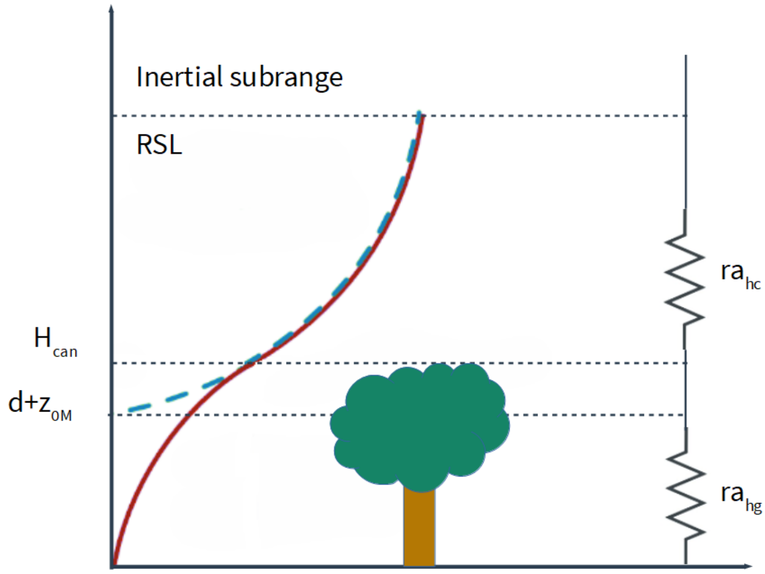


Figure 1.6: Canopy effect over the heat and momentum fluxes modeled in terms of resistances.

$$ra_{hg} = \frac{H_{can}e^{a}}{K(H_{can})a}\left[\exp\left(-a\frac{z_{0HG}}{H_{can}}\right) - \exp\left(-a\frac{z_{0H}+d}{H_{can}}\right)\right] \tag{1.11}$$

where z_{0HG} and z_{0H} are the roughness lengths for heat fluxes for ground and vegetation, which Noah-MP assumes to be equal to the momentum roughness lengths z_{0MG} and z_{0M}, a is the canopy wind extinction coefficient modulated by an empirical wind extinction parameter $CWPVT$ and the vegetation area index VAI:

$$a = (CWPVT \times VAI \times H_{can} \times \psi_m)^{0.5} \tag{1.12}$$

and K_c is the turbulent transfer coefficient for sensible heat:

$$K_c(H_{can}) = \kappa u_*(H_{can} - d) \tag{1.13}$$

The latent heat flux exchange coefficients ra_{wg} and ra_{wc} are assumed equal to the sensible heat coefficients.

The leaf boundary layer resistance is computed as

$$R_{\text{leaf,bdy}} = 50 \times \frac{a}{1-e^{-a/2}} \times \sqrt{\frac{D_{leaf}}{U_c}} \tag{1.14}$$

where D_{leaf} is a characteristic leaf dimension.

Finally, the stomatal resistance is computed based on the Ball-Berry (1987) scheme or the Jarvis (1976) scheme depending on the chosen option.

1.1.3 Options

Beyond its enhanced complexity and open-source accessibility, a pivotal attribute of the Noah-MP model lies in its remarkable flexibility through the availability of multiple options for each physical process. Users can choose between disabling dynamic vegetation and using prescribed values for Leaf Area Index (LAI) and Green Vegetation Fraction (GVF), or enabling it with predictions from the leaf dynamic, requiring the Ball-Berry implementation for stomatal resistance (Schaake et al., 1996a). Stomatal resistance itself offers two options: the Ball-Berry type, which correlates resistance with photosynthesis rates affected by Photosynthetically Active Radiation (PAR), and the Jarvis type, which accommodates variations in sunlit and shaded LAI and their PAR (Schaake et al., 1996a). Soil moisture factors influencing stomatal resistance come in three forms: Noah type based on soil moisture, and the CLM and SSiB types utilizing matric potential with different functional expressions (Xue et al., 1991; Oleson et al., 2004; Yang, Dickinson, et al., 1997). Runoff and groundwater modeling in Noah-MP include four schemes ranging from simple TOPMODEL-based approaches to the more complex BATS runoff scheme (Niu, Yang, Dickinson, and Gulden, 2005;Niu, Yang, Dickinson, Gulden, and Su, 2007; Schaake et al., 1996b; Yang, Dickinson, et al., 1997). For the surface exchange coefficient for heat, options include the standard Noah version 3.0 method and another based on MOST (Chen, Janjic, et al., 1997; Brutsaert, 1982b). Handling of supercooled liquid water in frozen soil has two alternatives, one following a general freezing-point depression equation and the other including an extra term for increased interface between soil and water (Niu and Yang, 2006; Koren et al., 1999). Frozen soil permeability can be modeled with either a linear or nonlinear effect of soil ice on infiltration (Niu and Yang, 2006; Koren et al., 1999). Radiation transfer through the vegetation canopy offers three options, differing in their assumptions about vegetation distribution (Yang and Friedl, 2003; Niu and Yang, 2004). For snow surface albedo, Noah-MP provides two choices, BATS and CLASS, each accounting differently for variables like snow age and SZA (Dickinson, Henderson-sellers, and Kennedy, 1993; Verseghy, 1991). Finally, the model offers three methods for partitioning precipitation into rainfall and snowfall, each varying in complexity and assumptions about surface air temperature (Jordan, 1991).

1.2 Weather Research and Forecasting model

The Weather Research and Forecasting (WRF) model (Skamarock, Klemp, Dudhia, Gill, Liu, Berner, Wang, et al., 2019) is a highly advanced mesoscale numerical weather prediction system designed for both atmospheric research and operational forecasting needs. Developed collaboratively by several agencies, including the

National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration (NOAA), the U.S. Air Force, the Naval Research Laboratory, the University of Oklahoma, and the Federal Aviation Administration (FAA), WRF has gained widespread usage across the globe due to its accuracy, computational efficiency, and versatility. WRF is designed to be a community model, meaning it is continually being updated and improved by contributions from users and developers around the world. The model supports parallel computation, making it suitable for running on a variety of computing platforms, from small clusters to supercomputers.

The WRF dynamical core ARW solves the compressible non-hydrostatic Euler equations in flux form (Skamarock, Klemp, Dudhia, Gill, Barker, et al., 2008):

$$\partial_t U + (\nabla \cdot Vu) + \mu_d \alpha \partial_x p + \left(\frac{\alpha}{\alpha_d}\right) \partial_\eta p \partial_x \varphi = F_U \tag{1.15}$$

$$\partial_t V + (\nabla \cdot Vv) + \mu_d \alpha \partial_y p + \left(\frac{\alpha}{\alpha_d}\right) \partial_\eta p \partial_y \varphi = F_V \tag{1.16}$$

$$\partial_t W + (\nabla \cdot Vw) - g \left[\left(\frac{\alpha}{\alpha_d}\right) \partial_\eta p - \mu_d\right] = F_W \tag{1.17}$$

$$\partial_t \Theta_m + (\nabla \cdot V\theta_m) = F_{\Theta_m} \tag{1.18}$$

$$\partial_t \mu_d + (\nabla \cdot V) = 0 \tag{1.19}$$

$$\partial_t \varphi + \frac{1}{\mu_d} \big[(V \cdot \nabla\varphi) - gW\big] = 0 \tag{1.20}$$

$$\partial_t Q_m + (\nabla \cdot Vq_m) = F_{Q_m} \tag{1.21}$$

In the above equations u, v, w are the velocity components, $\theta_m = \theta\left(1 + \frac{R_v}{R_d} q_v\right) \approx \theta\,(1 + 1.61 q_v)$ is the moist potential temperature, q_m are the mixing ratios, $\phi = gz$ the geopotential. $\mu_d = \frac{\partial p_d}{\partial \eta}$ is the vertical coordinate metric accounting for the hybrid terrain-following coordinate, which is a modified version of the standard terrain-following coordinate η, defined similarly to the one used in the Community Atmospheric Model (CAM), as described in the NCAR CAM3.0 Technical Note. μ_d enters in the flux form definitions for the prognostic variables as:

$$V = \mu_d \mathbf{v} = (U, V, W), \tag{1.22}$$

$$\Omega = \mu_d \omega, \tag{1.23}$$

$$\Theta_m = \mu_d \theta_m, \tag{1.24}$$

$$Q_m = \mu_d q_m. \tag{1.25}$$

The advantage of this new vertical coordinate is to remove the influence of the terrain more rapidly with height (Figure 1.7), preventing gravity wave amplification.

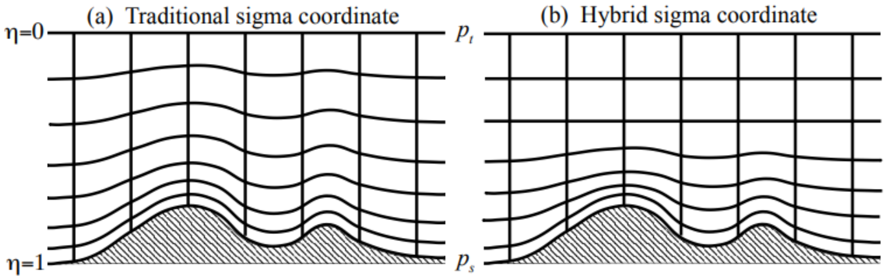


Figure 1.7: Resulting vertical grid stretching over complex terrain by using traditional sigma coordinate (a) and hybrid sigma coordinate (b), from Skamarock, Klemp, Dudhia, Gill, Barker, et al., 2008.

In addition to the prognostic equations, there are the following diagnostic equations for dry hydrostatic pressure and full pressure:

$$\partial_\eta \varphi = -\alpha_d \mu_d \tag{1.26}$$

$$p = p_0 \left(\frac{R_d \Theta_m}{p_0 \alpha_d} \right)^\gamma \tag{1.27}$$

where α_d is the inverse of the dry air density, $\gamma = cp/cv = 1.4$ is the ratio of the heat capacities for dry air, R_d is the gas constant for dry air, and p_0 is a reference surface pressure

The governing equations are re-casted in perturbation form to reduce truncation errors in the horizontal pressure gradient calculations and machine rounding errors in the vertical pressure gradient and buoyancy calculations. Each variable is decomposed in a reference state in hydrostatic balance $\bar{x}(\bar{z})$, and a perturbation part x'.

$$x = \bar{x}(\bar{z}) + x', \tag{1.28}$$

On the right-hand side of the governing equations (1.15)-(1.21) solved by ARW there are all the forcing terms F_U, F_V, F_W, and F_{Θ_m}, coming from physical processes parameterized and computed in dedicated modules: microphysics, radiation, turbulent mixing, spherical projection, Coriolis effect, etc.

The time discretization is performed in a time-split strategy, with an outer loop for the large time-step computed with a 3rd order Runge-Kutta integration, and an inner loop for the acoustic mode integration with time-step $\Delta\tau < \frac{1}{2} \cdot \frac{\Delta x}{c_s}$, where Δx is the spatial resolution and c_s is the speed of sound. This allows for separating the treatment of acoustic modes from the integration of more significant atmospherical processes.

Spatial discretization is performed as in Figure 1.8 over an Arakawa C grid (Arakawa et al., 1977), providing a better resolution of gravity waves. With this prescription, pressure gradient and divergence terms are computed over grid levels. The grids are rectangular, horizontally uniform, and stretched vertically. The ARW core provides finite difference spatial discretization from 2nd to 6th order for advection terms, with the possibility to specify the order for horizontal and vertical terms separately.

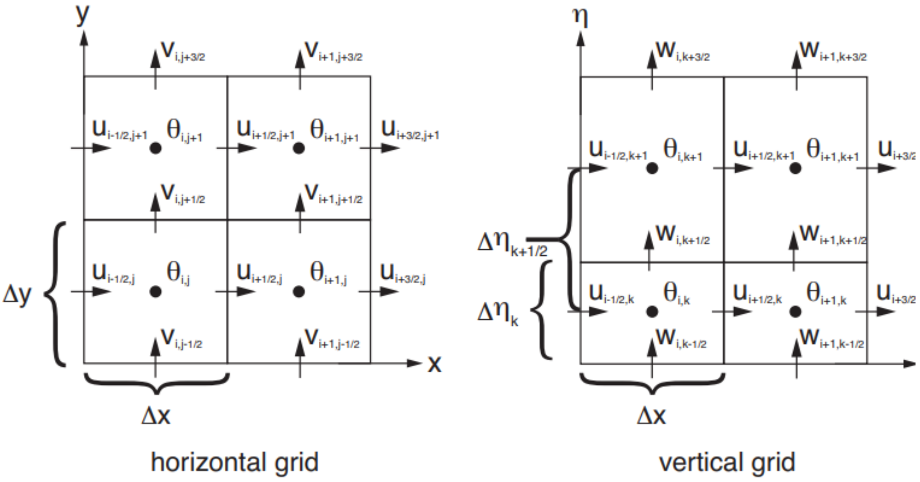


Figure 1.8: Arakawa-C staggering grid employed in WRF, from Skamarock, Klemp, Dudhia, Gill, Barker, et al., 2008.

The ARW core offers the capability of explicitly solving spatial diffusion both vertically and horizontally, or alternatively to solve only horizontal diffusion with vertical parametrized turbulence, or even using manually prescribed eddy viscosities. If the vertical turbulent mixing is computed independently by a dedicated planetary boundary layer scheme, the horizontal eddy viscosity can still be computed with a Smagorinsky closure from the deformation tensor D as:

$$K_h = C_s^2 l^2 \left[0.25(D_{11} - D_{22})^2 + D_{12_{xy}}^2\right]^{1/2} \tag{1.29}$$

with $l = \Delta x \Delta y$ and $C_s = 0.25$.

A sixth-order spatial filter can also be applied horizontally to remove noise from the kinematical fields (Knievel et al., 2007).

Different options allow the management of acoustic modes and gravity waves on the vertical, including the prescription of an absorbing layer, multiple dumping options, or a Rayleigh damping formulation for vertical velocity with the option to automatically activate it only in locations where w approaches the Courant number.

Allowed lateral boundary conditions are periodic, symmetric, open (also called "gravity-wave radiating", as in Klemp et al., 1978), or specified (mostly for real datasets).

More options are available in WRF for real simulations, but, in the present thesis, idealized simulations are performed with the microphysics scheme deactivated and without considering the Coriolis force. Among the main parameterizations available in WRF that play a role in the simulations performed for this thesis, it is worth mentioning the surface schemes, which compute the momentum and heat fluxes within the surface layer, and the planetary boundary layer schemes, which parameterize turbulence. The following describes the PBL and surface schemes available in WRF which are used in the case studies proposed in this thesis.

The Mellor-Yamada-Janic (MYJ) scheme (Janjic, 2002; Janjić, 1990) is a 1.5-order local scheme, i.e. it uses local vertical gradients to solve a prognostic equation for turbulent kinetic energy and uses it to compute the eddy diffusivity K as a function of height:

$$K_{m,h} = l_v \, (2e)^{\frac{1}{2}} \, S^* \tag{1.30}$$

where l_v is a mixing length following Prandtl theory, e is the turbulent kinetic energy and S^* is a nondimensional profile function. The subscripts m and h refer to momentum and heat diffusivities. From the eddy viscosity coefficient, the fluxes are evaluated as in the flux-gradient theory:

$$(\overline{u'w'}) = -K_m \frac{\partial \overline{u}}{\partial z} \tag{1.31}$$

The surface layer scheme for MYJ is the Eta-similarity scheme (Janjić, 1990), which incorporates the Zilitinkevich's (1995) parameterization for a viscous sub-layer with variable roughness height, relating temperature and momentum transport as:

$$z_{0T} = z_{0M} exp\left(-A_0 \sqrt{\frac{u_* z_{0M}}{\nu}}\right), \tag{1.32}$$

where u_* is the friction velocity, ν is the kinematic viscosity, and A_0 is an empirical constant. The correction term βw_*^2 formulated by Beljaars (1995) is applied for free convection, where β is the empirical free convection coefficient and w_* is the Deardorff velocity scale for convective mixed layers. The Beljaars term represents the near-surface wind induced by large eddies and avoids singularities for vanishing velocities.

The YSU scheme (Hong et al., 1996) is a non-local scheme which estimates the PBL height as:

$$h = Rib_{cr} \frac{\theta_{va} |U(h)|^2}{g(\theta_v(h) - \theta_s)} \tag{1.33}$$

where Rib_{cr} is the critical bulk Richardson number, θ_{va} is the potential temperature at the lowest model level, θ_s is a prescribed surface virtual potential temperature, $U(h)$ and $\theta_v(h)$ are the horizontal wind speed and virtual potential temperature at h.

h is used to compute the eddy viscosity for momentum (subscript m) and heat fluxes (subscript h) as a function of height as:

$$K_{m,h} = kz_s \left(1 - \frac{z}{h}\right)^2, \tag{1.34}$$

YSU also makes use of a counter-gradient term γ_h in the Reynolds averaged equation for $\overline{\theta}$, to account for the effects of deep convection typically not captured in local schemes, and a top PBL flux term $-\left(\overline{w'\theta'}\right)_h$ multiplied by $\left(\frac{z}{h}\right)^3$, inversely proportional to the surface flux:

$$\frac{\partial\overline{\theta}}{\partial t} = \ldots + \frac{\partial}{\partial z}\left[K_h\left(\frac{\partial\overline{\theta}}{\partial z} - \gamma_h\right) - \left(\overline{w'\theta'}\right)_h\left(\frac{z}{h}\right)^3\right] \tag{1.35}$$

The YSU parameterization scheme for surface fluxes is MM5, which uses Paulson (1970) stability functions to compute the scaling parameters:

$$u_* = \frac{kU}{\ln\left(\frac{z}{z_0}\right) - \psi_m} \tag{1.36}$$

$$\theta_* = -\frac{k(\theta - \theta_0)}{Pr\left[\ln\left(\frac{z}{z_0}\right) - \psi_h\right]} \tag{1.37}$$

$$q_* = -\frac{Mk}{ku_*z}\left[\left(\frac{Q - Q_s\left(\frac{\theta}{\theta_0}\right)}{\ln\left(\frac{z}{z_0}\right) - \psi_h}\right)\right] \tag{1.38}$$

where u_* is the friction velocity, q_* represents the scaling parameter for moisture or latent heat flux, θ_* is the scaling parameter for sensible heat flux, U is the mean horizontal wind speed, k is the von Kármán constant (typically around 0.4), z is the height above ground level, z_0 is the roughness length, M is a moisture availability parameter, Q and Q_s are the specific humidity at height z and at the surface, respectively, θ and θ_0 are the potential temperature at height z and at the surface, respectively, and Pr is the Prandtl number.

ψ_m and ψ_h are stability correction functions properly prescribed for unstable free convection, unstable forced convection, and stable conditions. The stability is determined from the bulk Richardson number.

The quasi-normal scale elimination (QNSE) model is a quasi-Gaussian spectral closure model from renormalization perturbation theory (Sukoriansky et al., 2005). It is defined as a self-contained algorithm of small-scale mode elimination to estimate effective viscosities and diffusivities. The complete theoretical treatment of the spectral model can be found in Sukoriansky et al. (2005), while the implementation in WRF is as follows: the definition of the eddy viscosity K_M and diffusivity K_H contains approximate stability functions $\alpha_M = K_M/K_0$ and $\alpha_H = K_H/K_0$ predicted by the spectral theory as functions of the local gradient Richardson number Ri:

$$\alpha_M = \frac{1 + 8Ri^2}{1 + 2.3Ri + 35Ri^2} \tag{1.39}$$

$$\alpha_H = \frac{1.4 - 0.01Ri + 1.29Ri^2}{1 + 2.34Ri + 19.8Ri^2} \tag{1.40}$$

The eddy viscosity for neutral flow K_0 is evaluated from the Prandtl-Kolmogorov expression:

$$K_0 = C_0 l e^{1/2}, \quad C_0 = 0.55. \tag{1.41}$$

where e is the turbulent kinetic energy solved from a prognostic equation.

The length scale l is given by

$$\frac{1}{l} = \frac{1}{l_B} + \frac{1}{l_N} \tag{1.42}$$

where l_B is the Blackadar scale (Blackadar, 1962) and l_N is a length scale limitation due to stable stratification:

$$l_N = c_N \frac{K^{1/2}}{N}, \quad c_N = 0.75 \tag{1.43}$$

The surface scheme for QNSE derives the drag coefficients for momentum and heat, C_D, C_H as:

$$C_D = \frac{\kappa^2}{\left(\ln \frac{z}{z_0} + \psi_M(\zeta) - \psi_M(\zeta_0)\right)^2} \tag{1.44}$$

$$C_H = \frac{\kappa^2}{\left(\ln\frac{z}{z_{0T}} + \psi_M(\zeta) - \psi_M(\zeta_{0T})\right)\left(\frac{1}{Pr_0}\ln\frac{z}{z_{0T}} + \psi_H(\zeta) - \psi_H(\zeta_{0T})\right)} \tag{1.45}$$

where $\zeta = \frac{z}{L}$, $\zeta_0 = \frac{z_0}{L}$, L is the Monin-Obukhov length, $Pr_0 = 0.71$ is the turbulent Prandtl number for fully developed neutral turbulence, and the stability factors ψ_M and ψ_H are:

$$\psi_M(\zeta) = 2.25\zeta - 0.2\zeta^2 \tag{1.46}$$

$$\psi_H(\zeta) = 2Pr_0\zeta + 0.1\left[(\zeta - 0.5)^2 - 0.5^2\right] \tag{1.47}$$

1.3 Sensitivity Analysis

Measuring the impact of small input perturbations on model output has historically been the first methodology to perform a sensitivity analysis (Iooss et al., 2014). Morris (1991) first introduced the concept of "elementary effect" to define the effect of the perturbation as the output obtained by that perturbation divided by the perturbation itself. This approach is based on a One-At-a-Time (OAT) methodology, i.e., changing only one parameter at a time, leaving all the others at fixed values. To translate these local perturbations into a global measure of sensitivity, Morris introduced an index defined as the average of all elementary effects for each of the parameters. This novel approach put the foundations of the so-called global sensitivity analysis (GSA). The Morris method is still used today with good results for screening procedures to discriminate the most influential parameters from the least influential ones. One of the main problems with the Morris method and, in general, with OAT approaches, is the assumption of a linear relationship between the parameter and the output variable, which does not allow to account for any interaction effect between parameters.

Figure 1.9 shows an example of a 2-parameter model where first-order and second-order effects are compared to first-order effects only. In this example, the output is related to the two parameters by the expression $z(X,Y) = 2X + 3Y + 0.5XY^2$. On the other hand, if the interaction effect is neglected, i.e. changing a single parameter leaving the other fixed, leads to $z(X) = 2X + 3c + 0.5Xc^2$, and $z(Y) = 2c + 3Y + 0.5cY^2$, with c being an arbitrary constant. It is important to note from the results of this simplified example that an OAT strategy for sensitivity analysis can be significantly misleading in understanding the system response to input perturbations.

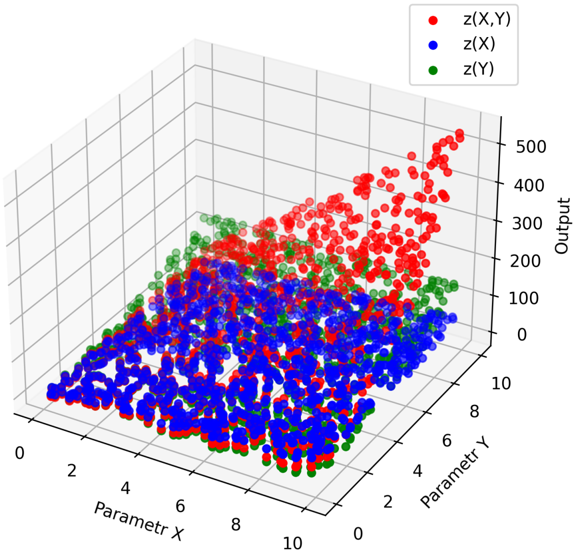


Figure 1.9: Interaction effect for a simple toy model of two parameters X and Y affecting a variable z.

A "variance-based" approach is more appropriate and leads to more robust results. The variance-based methods consist of decomposing the variance of the model output into multiple terms that account for all parameter interactions (Saltelli, Annoni, et al., 2010). With such tools, according to Saltelli et al. (2004), it is possible to correctly estimate both the overall input factors' influence on the output and their individual contributions to the total variance.

The most well-established variance-based method is the Sobol method (Saltelli and Sobol', 1995), which allows the computation, for each parameter, of the Sobol total index (which is the sum of all decomposition terms) and of multiple additional sensitivity indices, up to an arbitrary order of interactions (first-order, second-order, etc.) in the variance decomposition. The main problem with the Sobol method is its very high computational cost. For this reason, the implementation of the method is usually coupled with a screening procedure performed with the Morris method (Cuntz, Mai, Zink, et al., 2015; Cuntz, Mai, Samaniego, et al., 2016; Li, Zhang, Barlage, et al., 2018). Since the Morris method can decrease the computational costs by up to two orders of magnitude compared to the Sobol method (Herman et al., 2013), it is also often used as a proxy for the Sobol total order index.

As stated in Tarantola et al. (2024), there is a marked spread between local sensitivity analysis methods and global sensitivity analysis methods in the literature (Figure 1.10), with the majority of papers using GSA coming from engineering and environmental sciences (Figure 1.11). The Authors assert that the slower uptake of GSA methods is most probably due to divergent research cultures and traditions in the different scientific disciplines, heterogeneous computational resources in the various research fields, and varying degrees of proficiency and familiarity with sensitivity analysis methods among researchers in the different disciplines, which could explain why the majority of scientific fields favor older but more familiar methods. There is no doubt that the increasing complexity of models requires increasingly robust sensitivity analysis techniques, and a key requirement of good standard should be the ability of such techniques to provide a global perspective on the influence of uncertainties in model inputs, considering its criticality for decision making in systems with high degrees of nonlinearity and parameter interactions.

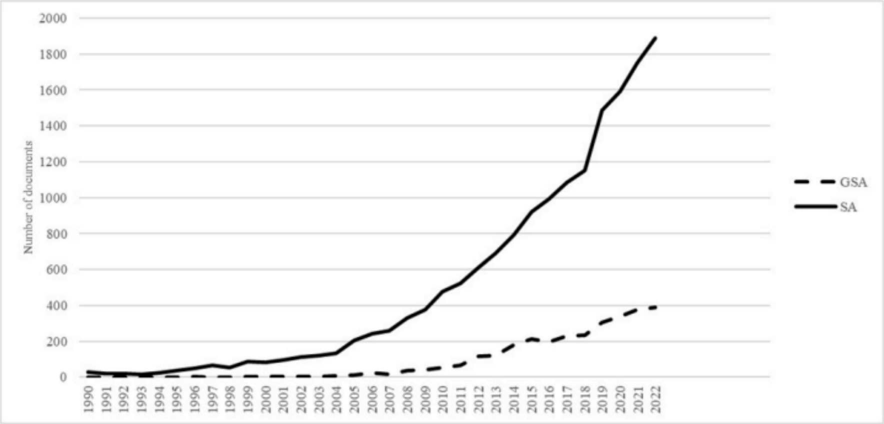


Figure 1.10: Meta-analysis of sensitivity related literature from Tarantola et al., 2024, depicting papers that utilize GSA versus papers using classical SA.

1.3.1 Sampling design and computational cost

To perform a GSA, it is necessary to set up each model run with a different set of values for the selected parameters of interest. These values must be chosen carefully, as good sampling statistics are key to obtaining meaningful results. The criteria for selecting a sampling method should take into account the degree of interaction expected for the specific problem under analysis. If there is only a single parameter to be analyzed, the algorithm may be as simple as a Monte Carlo sampling (Fishman, 1996) of values within a given range with a normal distribution, which, depending on the sample size, may have the drawback of showing unwanted clusters and gaps within the interval. A better strategy for ensuring homogeneity would be a "stratified sampling", where the domain is divided into sub-intervals, constrained by the rule to have an equal number of values in each. If the problem has multiple parameters, the simplest extension would be

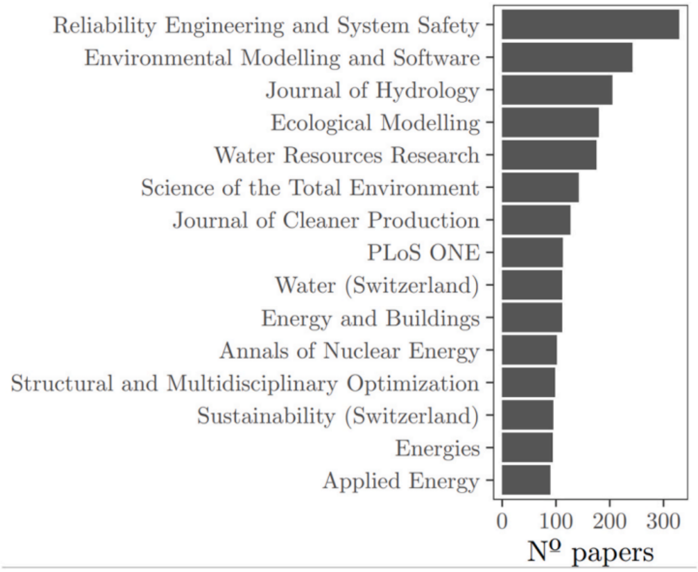


Figure 1.11: Meta-analysis of sensitivity related literature from Tarantola et al., 2024, classification by field of research.

to change one parameter at a time (OAT), leaving all the others fixed. While this is an acceptable sampling strategy for local sensitivity analysis, it does not take into account the interaction effects between the parameters involved, thus limiting the analysis to a very restricted ensemble of all possible output behaviors. Unfortunately, a full factorial sampling method, i.e. a sampling that explores all possible OAT combinations, requires 2^k simulations for k parameters, where each combination would represent a corner of a k-dimensional hypercube (Saltelli, Ratto, et al., 2008). Such sampling, called "two-level full factorial design", can be further improved in terms of exploration adding more points from inside the hypercube, thus obtaining an s-level design with s^k model runs, which increases the accuracy but inevitably also the computational cost. Selecting only a fraction of the combinations of a full factorial design leads to a fractional factorial (FF) design. Despite the convenience of reducing the sampling size, a two-level FF is often not enough to give insights into the variation of the model output (Saltelli, Ratto, et al., 2008). The Latin Hypercube (LH) sampling (Mckay et al., 1979) is considered one of the best methods to choose strategically the fraction of a full factorial design with $s > 2$. This method consists of constraining the values to ensure that each parameter is individually stratified over $s > 2$ levels, and that each level contains the same number of points.

The resulting sampling is stratified in an ensemble of non-overlapping regions ensuring that all parts of the original sample space are represented. Figure 1.13 shows an example of a 3-level FF with LH design in 2 dimensions: unlike the unstratified random sampling of Figure 1.12, for which a fractional design is visibly non-homogeneous, the 3-level FF-LH sampling ensures that each of the four

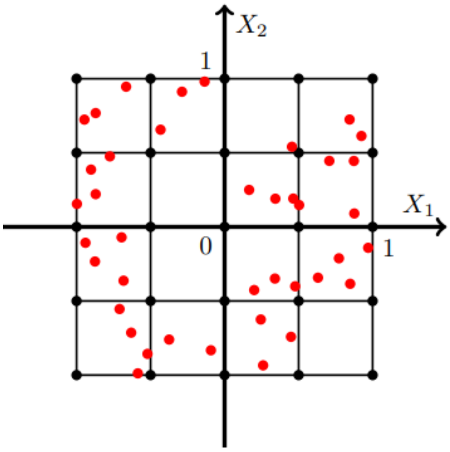


Figure 1.12: Example of random sampling in the parameter space spanned by two parameters.

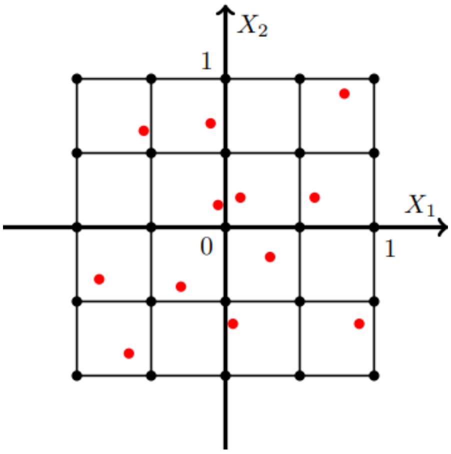


Figure 1.13: Example of FF-LH sampling in the parameter space spanned by two parameters.

quadrants contains the same number of points and each row and column contains 3 points.

Despite the advantages, LH sampling still misses important properties, such as the possibility to expand an already existing sampling, or the reproducibility offered by a deterministic procedure. Moreover, an LH sampling does not inherently incorporate any convergence criterion to ensure the best minimization of the number of simulations.

A pseudo-random number generator is a desirable alternative procedure due to its properties of straightforwardness and flexibility, allowing for easy expansion if additional parameters or simulations are needed. While classic randomly generated samples suffer from irregular distribution, pseudo-random sampling is a deterministic sequence built to increase the homogeneity as fast as possible with the number of samples, with a convergence criterion based on the discrepancy of the sequence, ensuring a superior optimization of the number of simulations compared to an LH sampling (Renardy et al., 2021). Discrepancy is a metric used to quantify the unevenness of point distribution in a multidimensional space, with lower values indicating more uniform distributions, i.e. ideal for sensitivity analysis. Low-discrepancy sequences, or pseudo-random sequences, are generated by algorithms that strategically bias the placement of new points to avoid proximity to existing ones, promoting uniformity across the entire sample space.

The benefits of low discrepancy become more apparent with larger sample sizes, even if more parameters necessitate larger quasi-random samples for the advantages of low discrepancy to manifest.

The Sobol sequence is a type of low-discrepancy sequence, also known as a quasi-random or quasi-Monte Carlo sequence, specifically designed to cover the multidimensional space more uniformly than uncorrelated random points. It is a sequence of high-dimensional points that are used to create a more uniform distribution of points within a unit hypercube. The Sobol sequence is particularly beneficial in numerical integration and in the estimate of the distribution of complex models, where traditional random sampling may fail to provide a sufficient level of accuracy.

The Sobol sequence (Sobol', 1967), and the quasi-Monte Carlo sampling methodologies in general, can be conceptualized as follows (Hung, 2023). Let D in $[0,1]^d$ be a d-dimensional unit hypercube, and f a real integrable function over D. The aim is to construct a sequence p_i in D so that:

$$\left| \frac{1}{n} \sum_{i=1}^{n} f(p_i) - \int_D f(x)\, dx \right| \leq M(P)V(f), \tag{1.48}$$

where $M(P) = \|F_n(x) - F(x)\|$ is the discrepancy defined as the difference between the artificial distribution based on the allocation $P = \{p_1, \ldots, p_n\}$ and the exact uniform distribution $F(x)$ over D, and $V(f)$ measures the fluctuation of the function f (Niederreiter, 1992). The aim is to minimize the discrepancy, hence to solve an optimization problem to find an allocation $P*$ such that:

$$P^* = \arg \min_{P\{D\}} M(P) \tag{1.49}$$

This results in more accurate estimations for integrals and model distributions in higher-dimensional spaces when compared to standard random sampling techniques.

1.3.2 Global Sensitivity Analysis

For each parameter, an elementary effect can be defined as the ratio between the difference in the output response produced by a perturbation of its original value and the perturbation itself. Using the notation of Saltelli et al. (2008), if each input parameter is perturbed once, for a given set of k input parameters $X = X_1, X_2, ..., X_i, ..., X_k$, the elementary effect (EE) of the $i-th$ input parameter is defined as:

$$EE_i = \frac{[Y(X_1, X_2, ..., X_i + \Delta, ..., X_k) - Y(X_1, X_2, ..., X_i, ..., X_k)]}{\Delta}, \qquad (1.50)$$

where Δ is one of the N random perturbations of the $i-th$ parameter. A global measure of the sensitivity is obtained by computing an average of all the N elementary effects of a single parameter:

$$\mu_i = \frac{1}{N}\sum_{i=1}^{N} EE_i. \qquad (1.51)$$

Since μ_i is sensible to effects of opposite signs, it is usually substituted by a more practical form introduced by Campolongo et al. (2007):

$$\mu_i^* = \frac{1}{N}\sum_{i=1}^{N} |EE_i|, \qquad (1.52)$$

Due to the implicit linear assumption in Equation (1.52), its standard deviation is typically used, since its introduction by Morris (1991), to assess the non-linear or interacting nature of the averaged elementary effects:

$$\sigma_i^2 = \frac{1}{N-1}\sum_{i=1}^{N} (EE_i - \mu)^2, \qquad (1.53)$$

It is important to note that, although σ_i^2 measures the difference between the direct linear effect of a single parameter and any other effect, it does not allow to distinguish between non-linear effects and interactions.

1.3.3 The Sobol Method

As previously stated, the Sobol method decomposes the variance of the model output Y into fractions that can be attributed to multiple inputs. The variance decomposition $V(Y)$ is written as:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12...k}, \qquad (1.54)$$

and dividing by $V(Y)$, we can introduce the Sobol indices $S_i = V_i/V$, $S_{ij} = V_{ij}/V$, etc., leading to:

$$1 = \sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \cdots + S_{12\ldots k}. \tag{1.55}$$

The total effect, or total order index S_{Ti}, is the total contribution to the output variation due to a specific factor X_i, i.e. a specific parameter. Hence, for instance, for a set of three parameters, X_1, X_2, X_3, the total effect index for the parameter X_1 is:

$$S_{T1} = S_1 + S_{12} + S_{123}. \tag{1.56}$$

The total order index, by definition, is in principle equivalent to the Morris μ_i^* index, and thus, for a reasonable amount of samples that ensure good convergence of the method, μ_i^* should be a good proxy for S_{Ti}.

The analysis of the first-order index S_i and the total index S_{Ti} provides important information about the parameter X_i. (Saltelli, Ratto, et al., 2008):

- if the difference between S_{Ti} and S_i is high, the parameter X_i is involved in strong interactions;
- a total index $S_{Ti} \approx 0$ means that the parameter X_i is a noninfluential factor and can be fixed at any value without affecting the output variance $V(Y)$;
- the magnitude of S_i indicates how much it is possible to reduce the variance of the output $V(Y)$ if the parameter X_i could be fixed;
- the difference $1 - \sum_i S_i$ is an overall indicator of the presence of interactions in the model.

As shown in Cuntz et al. (Cuntz, Mai, Samaniego, et al., 2016), when studying time series, i.e. when the model output under analysis is a time sequence, the Sobol indices computed for each time step can also be studied as a time series. As shown in Equation (1.55), the sum of all the indices theoretically amounts to unity (although this could not be true in case of convergence inaccuracy). This implies that ranking the k Sobol indices as $S_{max} < \ldots < S_{min}$ for a specific hour of the day, results in a ranking of the weights that each parameter has on the total variance with respect to the other parameters under analysis. In other words, the Sobol indices obtained are strictly bounded to the specific set of the chosen parameters and the specific hour over which the variance is considered, i.e. the time evolution of the Sobol indices has to be interpreted as a change in the relative weights of the parameters. Conversely, it would be an error to interpret the change of a single parameter sensitivity index without considering the other parameter sensitivity changes.

In principle, the brute-force computation of the Sobol first-order index would require N^2 model runs per input factor, with reliable results typically obtained in the range of N≈10^2-10^4. The following will describe the derivation of the algorithm proposed by Sobol and Saltelli (2010) to circumvent this prohibitive computational cost.

To be coherent with the original mathematical treatment of Saltelli, slight changes in the notation are first introduced. The variance of the model Y due to the change of all the factors except $X_i = x*$, will be written as the conditional variance $V_{X_{\sim i}}[Y|X_i = x*]$, where $x*$ is a fixed value for the factor X_i. The problem of local sensitivity states that this is not a sufficient measure of importance for the factor X_i, because it is dependent on the value $x*$. Taking the average over all the possible $x*$ will get $E_{X_i}[V_{X_{\sim i}}[Y|X_i]]$, which is now independent from $x*$. The total variance of the model Y can be written as:

$$V[Y] = E_{X_i}\left[V_{X_{\sim i}}[Y|X_i]\right] + V_{X_i}\left[E_{X_{\sim i}}[Y|X_i]\right] \tag{1.57}$$

where the term $V_{X_i}\left[E_{X_{\sim i}}[Y|X_i]\right]$ is the first-order factor from which the sensitivity index can be recovered:

$$S_i = \frac{V_{X_i}\left[E_{X_{\sim i}}[Y|X_i]\right]}{V[Y]} \tag{1.58}$$

The variance appearing in the nominator can be explicitly written in terms of its definition:

$$\begin{aligned} V_{X_i}[E_{X_{\sim i}}[Y|X_i]] &= E_{X_i}[E^2_{X_{\sim i}}[Y|X_i]] - (E_{X_i}[E_{X_{\sim i}}[Y|X_i]])^2 \\ &= \int E^2_{X_{\sim i}}[Y|X_i]\, dx_i - \left(\int E_{X_{\sim i}}[Y|X_i]\, dx_i\right)^2 \end{aligned} \tag{1.59}$$

The integrals in Eq. (1.59) can be furtherly made explicit considering that:

$$E_{X_{\sim i}}[Y|X_i] = \int f(x_{\sim i}, x_i) dx_{\sim i} \tag{1.60}$$

And by defining a dummy vector variable $\mathbf{x}'_{\sim i}$:

$$\begin{aligned} E^2_{X_{\sim i}}[Y|X_i] &= \int f(x_{\sim i}, x_i)\, dx_{\sim i} \cdot \int f(x_{\sim i}, x_i)\, dx_{\sim i} \\ &= \int\int f(x'_{\sim i}, x_i) f(x_{\sim i}, x_i)\, dx'_{\sim i}\, dx_i \end{aligned} \tag{1.61}$$

Hence the explicit form of the variance:

$$V_{X_i}[E_{X_{\sim i}}[Y|X_i]] = \int \int f(x'_{\sim i}, x_i) f(x_{\sim i}, x_i) dx'_{\sim i} dx_i - \left(\int f(x) dx \right)^2 \tag{1.62}$$

which leads to the integral form of the first-order sensitivity index:

$$\begin{aligned} S_{X_i} &= \frac{V_{X_i}[E_{X_{\sim i}}[Y|X_i]]}{V[Y]} \\ &= \frac{\int \int f(x'_{\sim i}, x_i) f(x_{\sim i}, x_i)\, dx'_{\sim i}\, dx_i - \left(\int f(x)\, dx \right)^2}{\int f(x)^2\, dx - \left(\int f(x)\, dx \right)^2} \end{aligned} \tag{1.63}$$

The same procedure applies to derive the total effect index, which yields:

$$\begin{aligned} S_{T_{X_i}} &= \frac{E_{X_{\sim i}}[V_{X_i}[Y|X_{\sim i}]]}{V[Y]} \\ &= \frac{\int f^2(x)\, dx - \int \int f(x_{\sim i}, x'_i) f(x_{\sim i}, x_i)\, dx'_i\, dx_i}{\int f(x)^2\, dx - \left(\int f(x)\, dx \right)^2} \end{aligned} \tag{1.64}$$

The discretization of the integrals leads to the algorithm found by Saltelli, which prescribes the following steps:

1. generate a $(N,\, 2k)$ matrix of random numbers (k is the number of inputs) and define two matrices of data (A and B), each containing half of the sample. N is called a base sample; to give an order of magnitude, N can vary from a few hundred to a few thousand. Sobol recommends using sequences of quasi-random numbers (Sobol', 1967);

2. define a matrix Ci formed by all columns of B except the ith column, which is taken from A;

3. compute the model output for all the input values in the sample matrices A, B, and Ci, obtaining three vectors of model outputs of dimension $N \times 1$: $y_A = f(A)$, $y_B = f(B)$, $y_C = f(C)$.

After the above steps, the following expression for the first-order index is obtained:

$$S_i = \frac{V\left[E[Y|X_i]\right]}{V[Y]} = \frac{y_A \cdot y_{C_i} - f_0^2}{y_A \cdot y_A - f_0^2} = \frac{\frac{1}{N}\sum_{j=1}^{N} y_A^{(j)} y_{C_i}^{(j)} - f_0^2}{\frac{1}{N}\sum_{j=1}^{N} \left(y_A^{(j)}\right)^2 - f_0^2} \quad (4.21) \tag{1.65}$$

where

$$f_0^2 = \left(\frac{1}{N} \sum_{j=1}^{N} y_A^{(j)} \right)^2 \tag{1.66}$$

is the mean, and the symbol $\cdot$ denotes the scalar product of two vectors.

The total effect is obtained from:

$$S_{T_i} = 1 - \frac{V\left[E[Y|X_{\sim i}]\right]}{V[Y]} = 1 - \frac{y_B \cdot y_{C_i} - f_0^2}{y_A \cdot y_A - f_0^2} = 1 - \frac{\frac{1}{N}\sum_{j=1}^{N} y_B^{(j)} y_{C_i}^{(j)} - f_0^2}{\frac{1}{N}\sum_{j=1}^{N} \left(y_A^{(j)}\right)^2 - f_0^2} \tag{1.67}$$

The matrices A and B account for $2N$ model evaluations, while the matrices Ci require $N \cdot k$ model runs for k input factors, for a total of $N \cdot (k+2)$ model runs, which is significantly lower than the N^2 required for the brute-force approach.

As proposed in the present thesis, this computational efficiency can be further enhanced by employing data-driven modeling techniques to construct a surrogate model. Such a surrogate model approximates the behavior of the original complex model, enabling the computation of the required evaluations for the Sobol method at a significantly reduced computational cost. By fitting a surrogate model to a subset of the original model evaluations, the responses for a wide range of input parameters can be estimated without the need to run the full model each time. This approach is particularly advantageous in scenarios where the original model is computationally intensive, as it substantially decreases the total number of model runs required for a comprehensive sensitivity analysis. The surrogate model, once validated for accuracy, serves as a rapid computational tool, ensuring that the Sobol method's application remains both efficient and reliable, even for models with a high computational burden.

1.4 Data-driven modeling

Data-driven modeling for regression tasks involves a sequence of key steps designed to build and evaluate predictive models based on continuous data. While the specific details of the workflow may vary across different studies, reflecting their unique requirements and objectives, the core components of this process remain consistent in every workflow. A typical workflow is shown in Figure 1.14, which is specific to weather applications using three different machine learning algorithms. Each step in the ideal workflow contributes to the overall accuracy and reliability of the model.

1. Data Collection and Preprocessing: this step includes data cleaning (handling missing values, removing duplicates), data transformation (converting data

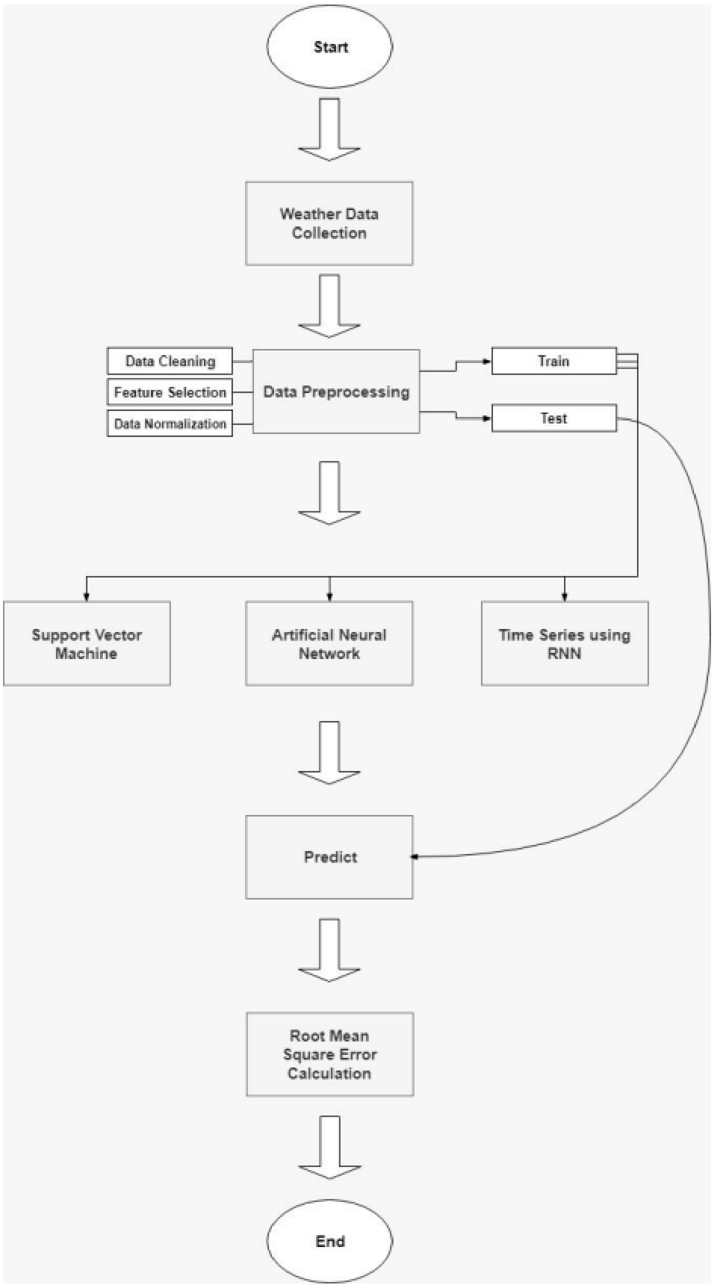


Figure 1.14: Workflow from Singh et al., 2019

into a suitable format), and feature engineering (creating new features from existing data).

2. Data Standardization/Normalization: this step prevents features with larger scales from dominating the model's learning process.

3. Data Splitting (Train-Test Split): typically, data is split into a training set (used to train the model) and a testing set (used to evaluate the model). A common split ratio is 70:30 or 80:20 (train:test). Splitting data into these two sets is crucial for assessing the model's ability to generalize beyond the data it was trained on. If the model only performs well on the training data

but not on the test data, it may be affected by overfitting.

4. Model Training: the training dataset is used to train the machine learning model. It includes both the input data and the corresponding expected output. The model learns to map inputs to outputs from this dataset. The choice of the regression algorithm (linear regression, decision trees, etc.) depends on the nature of the task and data. The model learns to map input features to the target variable during training.

5. Cross-Validation: this step provides a more robust way to assess the model's performance. It is used to prevent overfitting. The training set is divided into smaller subsets. The model is trained on some subsets and validated on others, typically using techniques like k-fold cross-validation (Figure 1.15). This process is repeated multiple times to ensure the model's stability and reliability.

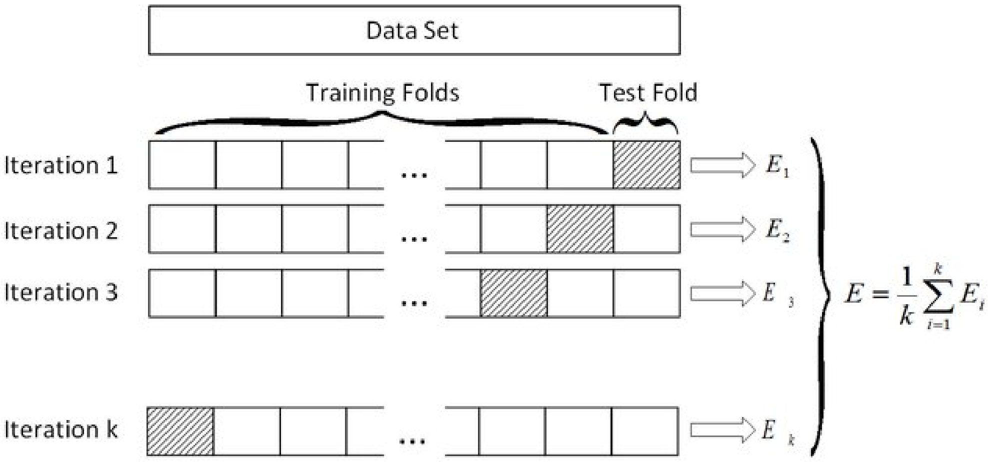


Figure 1.15: K-fold cross-validation performed by defining a training fold set and a test fold in each iteration, from De Oliveira et al., 2017.

6. Hyperparameter Tuning: hyperparameters are the parameters inherent to the machine learning algorithms. They are set prior to the training process and remain constant during it. Tuning of the hyperparameters is important to optimize the algorithms and improve their performance. Techniques like grid search or random search are used to find the optimal set of hyperparameters.

   Grid Search (Figure 1.16) is a brute-force approach to hyperparameter tuning that systematically builds and evaluates a model for each combination of algorithm parameters specified in a grid.

   Random Search (Figure 1.17) addresses some of the limitations of Grid Search by randomly selecting combinations of hyperparameters to evaluate. It can sample a broader range of values than Grid Search and it is possible to incorporate prior knowledge by specifying the distributions from which hyperparameters are sampled. It often achieves better performance with fewer evaluations compared to Grid Search, particularly when the hyperparameter space has a low intrinsic dimensionality.

Gradient descent (Figure 1.18) is a first-order iterative optimization algorithm for finding a local minimum of a differentiable function. In the context of hyperparameter tuning, gradient descent can be used to minimize a loss function $L(\theta)$, which depends on the model's hyperparameters θ. The hyperparameters are updated iteratively to minimize the loss function. When dealing with complex functions that have multiple local minima, gradient descent can end in one of these local minima rather than finding the global minimum, especially if the function is not convex. This issue can be circumvented by modifying the algorithm with more refined strategies or by using a different methodology such as Bayesian Optimization.

Bayesian Optimization is a more sophisticated approach that uses the Bayesian technique to direct the search in order to find the minimum or maximum of an objective function efficiently. It is particularly useful when the evaluation of the function is expensive, i.e. time-consuming. The key idea is to estimate the posterior probability distribution of the objective function and use it to select the most promising hyperparameters to evaluate them in the true function. In Bayesian Optimization, the choice of the next point to evaluate is based on the posterior distribution derived from the previous evaluations, typically using the Gaussian Process method to create a surrogate model.

The surrogate model in a GP framework is defined as:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')), \tag{1.68}$$

where $f(x)$ is the unknown objective function, $m(x)$ is the mean function, and $k(x, x')$ is the kernel function defining the covariance between two points x and x' in the hyperparameter space. The acquisition function $\alpha(x)$ determines the query point x_n solving an optimization problem. After evaluating the objective function at x_n, the surrogate model is updated with a new observation x_{n+1} from the dataset. This update is done using Bayes' theorem, hence the name Bayesian Optimization. The update rule for GPs, taking into account the new data point x_{n+1}, y_{n+1}, is:

$$p(f|x_{(1:n+1)}, y_{(1:n+1)}) \propto p(y_{n+1}|f, x_{n+1}) p(f|x_{1:n}, y_{1:n}), \tag{1.69}$$

where $p(f|x_{(1:n)}, y_{(1:n)})$ is the posterior probability of f given the data points observed so far, and $p(y_{n+1}|f, x_{n+1})$ is the likelihood of observing x_{n+1}, y_{n+1}.

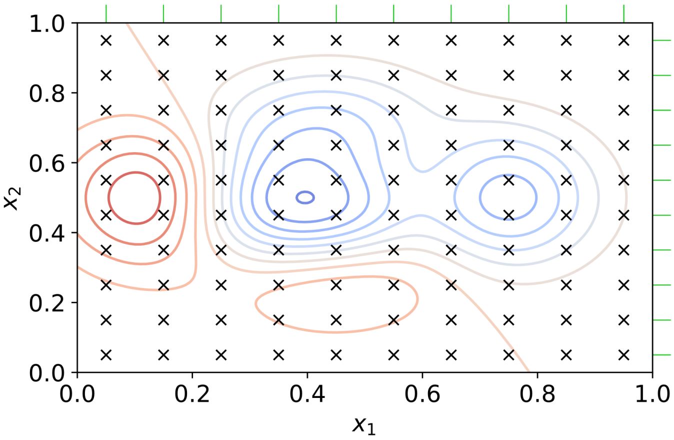


Figure 1.16: Homogeneous optimization procedure performed in a grid-search algorithm (`https://en.wikipedia.org/wiki/Hyperparameter_optimization`).

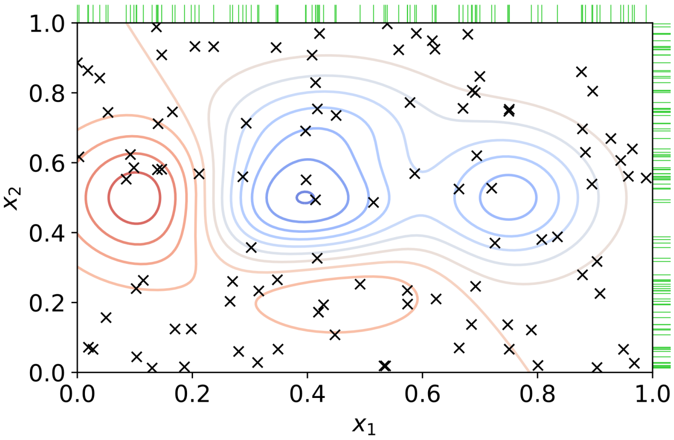


Figure 1.17: random optimization procedure performed in a random-search algorithm (`https://en.wikipedia.org/wiki/Hyperparameter_optimization`).

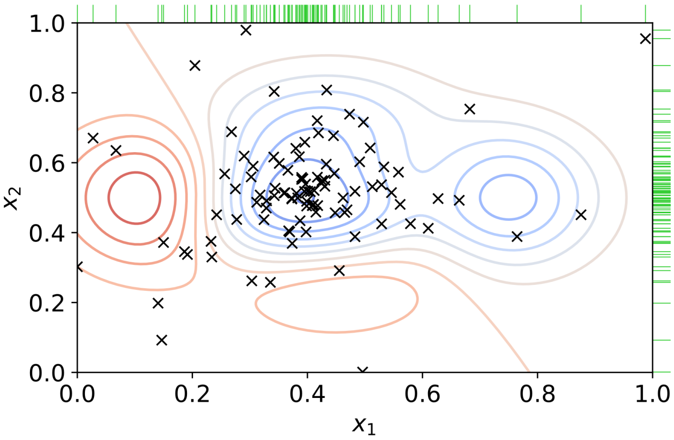


Figure 1.18: Optimization procedure based on a gradient-descent strategy (`https://en.wikipedia.org/wiki/Hyperparameter_optimization`).

7. Model Test: The test dataset is a separate set of data not used during training. It is used to evaluate the performance of the model after training. Metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), or the determination coefficient R^2 are typically used to quantify the model's accuracy. The model test provides insights into how well the model generalizes to new, unseen data.

1.4.1 Variance-bias tradeoff

The final aim of the model is to avoid overfitting and find the best compromise between the variance and the bias. Achieving this balance is crucial, as it directly influences the model's overall performance and its ability to generalize well to new, unseen data. To achieve this goal, it is necessary to consider the contributions to the prediction error of an ML-based regression algorithm.

The total expected prediction error can be written as (e.g. Hastie et al., 2004):

$$\text{Total Error} = (\text{Bias}^2) + \text{Variance} + \text{Irreducible Error} \tag{1.70}$$

where the bias refers to the error introduced by overly simplistic assumptions in the learning algorithm. High bias can cause a model to miss relevant relations between features and target outputs (underfitting). The bias is defined as the difference between the expected prediction of the model and the true output values:

$$\text{Bias}^2 = \left(E[\hat{f}(x)] - f(x)\right)^2 \tag{1.71}$$

On the other hand, the variance measures how much the predictions of a model vary for a given data point when trained on different datasets. High variance is observed when a model is overly sensitive to small fluctuations in the training set (overfitting).

$$\text{Variance} = E\left[\left(\hat{f}(x) - E[\hat{f}(x)]\right)^2\right] \tag{1.72}$$

Overfitting occurs when a statistical model or machine learning algorithm captures the noise in the data rather than the underlying pattern. It occurs when a model is overly complex, such as having too many parameters relative to the number of observations. An overfitted model performs very well on the training data, but poorly on unseen data (test data). This is because it has learned both the true underlying relationships in the training data and the random fluctuations that are not actually part of the underlying data pattern. In other words, overfitting occurs when the algorithm learns the specifics of the examples rather than the generalized input-output relationship.

Increasing model complexity typically decreases the bias but increases the variance, and vice versa. Simplifying a model can lead to higher bias (underfitting), while increasing its complexity can lead to higher variance (overfitting). This is shown in Figure 1.19, where the error from sample training and testing data is plotted against model complexity. While the error is guaranteed to decrease during training for any level of model complexity (cyan colored line), the error associated with the testing phase (red colored line) eventually reaches an inflection point, meaning that beyond that point the algorithm will overfit the training data and lose the ability to obtain good general results.

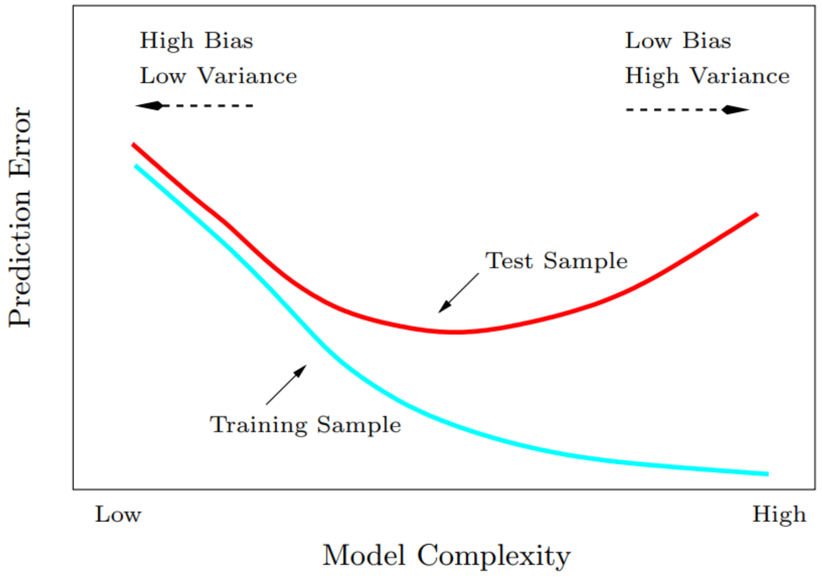


Figure 1.19: Variance-Bias tradeoff in the prediction error curve for the training sample and the test sample, from Hastie et al., 2004.

The goal in many ML tasks is to find the tradeoff between bias and variance, minimizing the total error. This optimal complexity point does not necessarily mean the lowest possible bias or variance, but the point where their combination is minimized.

Thoroughly following the good practices described in the previous section, as well as using the specific strategies discussed in the following section, which some regression methods inherently employ (e.g. regularization, bootstrapping, boosting, etc.), is typically considered the most effective empirical strategy to identify the optimal tradeoff point between bias and variance for the specific problem under analysis. The accuracy metrics used to evaluate the goodness of the test phase are the final indicators to assess whether the obtained tradeoff is sufficient for the research objectives.

1.4.2 ML algorithms

As of today, there is a great variety of ML methods for data-driven modeling.

In the field of ML, algorithms can be divided in two main categories: supervised and unsupervised learning, both of which can be applied to regression and classification tasks. Supervised learning algorithms rely on labeled data sets to train the model. This means that the input data is paired with the correct output, allowing the algorithm to learn over time how to predict the output from the input. Supervised learning works well for both classification tasks, where the output is a category or label, and regression tasks, where the output is a continuous value.

Unsupervised learning algorithms, on the other hand, do not use labeled data. They work by identifying patterns or inherent structures in the input data. These algorithms are typically used for clustering tasks, where the goal is to group similar items together, or to identify rules that describe large portions of the data.

In the specific case of this thesis, the aim is to use input features and associated outputs to train the models, hence the focus is on supervised learning methods. Furthermore, since the focus is exclusively on datasets with continuous values, the attention turns to regression tasks within the supervised learning paradigm. Regression models are trained to understand and predict continuous outcomes based on the relationships derived from the input features and their corresponding continuous outputs.

1.4.2.1 Decision Trees

A decision tree is a hierarchical method with a flowchart tree-like structure, where each internal node (the "decisional nodes") defines a test over an attribute, and has two or more branches leading to the test outcomes (the "leaf nodes"). Over a continuous variable, these tests act by splitting the data into smaller subsets based on criteria over input and output pairs, with the aim of obtaining the data segmentation that minimizes the residuals.

Using the same notation of Hastie et al. (2004), if the dataset is composed of input x_i and output y_i, and is segmented by the algorithm into M partitions $R_1, R_2, \ldots, R_M$, with the output modeled as a constant c_m in each region, the function built by the decision tree is:

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m). \tag{1.73}$$

The search of a splitting point s performed over the variable j consists of obtaining two partitions defined as:

$$R_1(j,s) = \{X|X_j \leq s\} \text{ and } R_2(j,s) = \{X|X_j > s\}. \tag{1.74}$$

and the minimization to be solved, using the sum of squares $\sum(y_i - f(x_i))^2$ as the cost function, is:

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i - c_1)^2 + \min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i - c_2)^2\right]. \tag{1.75}$$

which finds the best splitting of the data.

Tree-based algorithms such as CART (Classification and Regression Trees, Breiman et al., 1984) are based on the "greedy algorithm" to explore the input space and find split points with recursive binary slitting using a cost function. A stopping criterion is chosen beforehand for the greedy algorithm, such as reaching a certain amount of samples in each leaf node. In the building of a tree structure, trees with fewer branches are favored by the use of tree pruning techniques, avoiding an over-clusterization of the data. A curve regression performed by simple decision trees would appear as in Figure 1.20, showing the effects of changing the depth of the tree, i.e. the length of the longest path from the root node to a leaf node. The discontinuous nature of the individual decision branches results in discretized steps which can be made more dense increasing the depth of the tree, but the proper size for each step must be tuned for the specific data set without ending up in an underfitting/overfitting scenario.

Despite their simplicity and intuitiveness, basic decision tree algorithms such as CART can easily suffer overfitting, high variance, and low bias. A solution to make the method more robust to outliers and more reliable is to a use multitude of decision trees and to average the results obtained from each tree separately. This approach is called "ensemble learning" and is the base of more refined methods such as Random Forests (RF, Breiman, 2001).

In the RF algorithm, each tree is obtained from a subsample of the dataset (Figure 1.21) extracted by a bootstrap procedure.

The bootstrap method is a resampling technique used to estimate statistics on a population by sampling a dataset with replacement. In the context of ML, it is often used to create multiple datasets from a single original dataset. Each element of the original dataset has the same probability of appearing in each bootstrap sample, and elements can appear multiple times in a single sample. This is different from traditional sampling methods, where each element can appear only once in each sample (sampling without replacement).

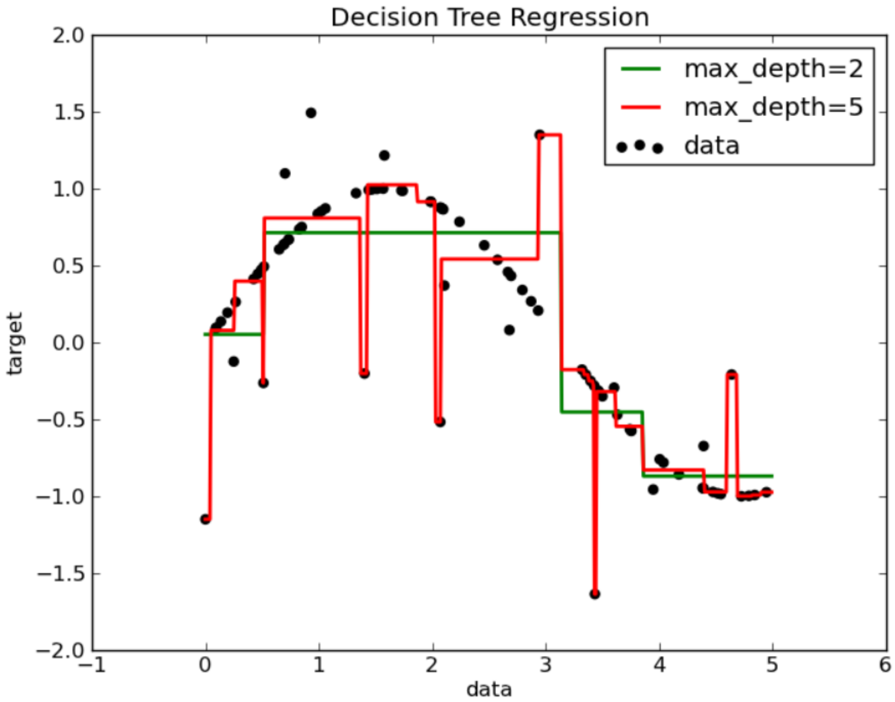


Figure 1.20: Decision tree regression for 2 values of the hyperparameter *max_depth* (`https://scikit-learn.org/stable/auto_examples/tree/plot_tree_regression.html`).

In particular, RFs utilize the Out-Of-Bag (OOB) method, which is a bootstrapping method for validating the model without needing a separate validation set. During the creation of each bootstrap sample, some instances from the original dataset are left out (i.e. not sampled). These left-out data are referred to as "out-of-bag". The OOB instances can be used to test the performance of each model in the ensemble, since they were not used during the model's training. The error rate of predictions on the OOB instances is known as the OOB error, which is a measure of prediction accuracy. Since OOB instances are a random sample of the original dataset, the OOB error can be a reliable estimate of the model's performance on unseen data, similar to cross-validation. The variance reduction comes from choosing the best split from these random subsets of features.

The tree-based algorithms can achieve even more benefits from a boosting approach (e.g. XGboost, Chen and Guestrin, 2016). Boosting is an ensemble technique where multiple models (usually weak learners) are trained sequentially. Each new model focuses on the errors of the previous models and tries to correct them. The key idea is to train subsequent models by emphasizing the data points that were incorrectly predicted by earlier models, hence each subsequent tree is built with an information function coming from the old tree, as shown in Figure 1.22. This is often achieved by assigning higher weights to these misclassified points or by focusing on the residual errors. The final model is typically a weighted

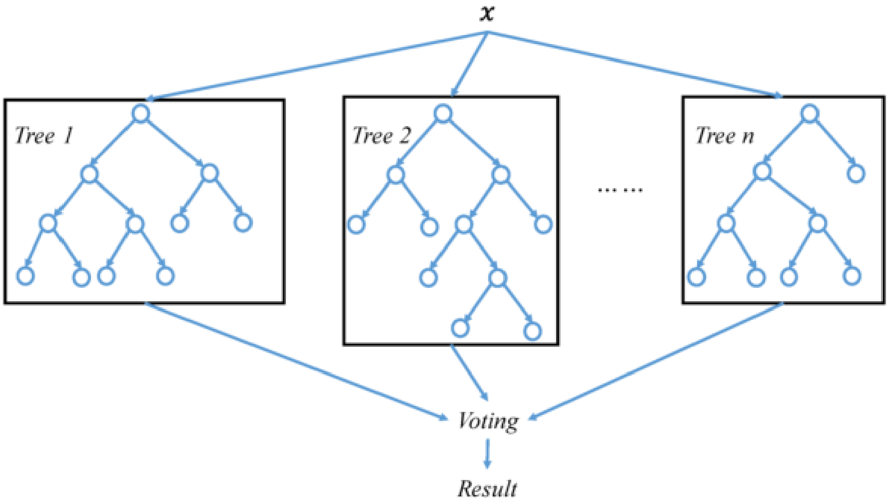


Figure 1.21: Random forest voting system depiction, from Wang, Pan, et al., 2019

average of all the "weak learners", where more accurate learners are given more weight. Specifically, gradient boosting is a type of boosting that optimizes a cost function. In gradient boosting, the decision trees are added sequentially to correct the residuals (i.e. the errors) of the combined ensemble. The gradient of the loss function is used to minimize the errors. The next model in the sequence specifically targets reducing the overall loss, making the learning process more direct and efficient. Mathematically, the procedure of gradient boosting, given a loss function $L(y, f(x))$, and a number of iterations M, is as follows (Hastie et al., 2004):

1. Initialize $f_0(x) = \arg\min_\gamma \sum_{i=1}^{N} L(y_i, \gamma)$.
2. For $m = 1$ to M:
   (a) For $i = 1, 2, \ldots, N$ compute
   $$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_{m-1}}. \tag{1.76}$$
   with r_{im} being called pseudo-residuals.
   (b) Fit a regression tree to the targets r_{im} resulting into data partitions $R_{jm}, j = 1, 2, \ldots, J_m$.
   (c) For $j = 1, 2, \ldots, J_m$ compute
   $$\gamma_{jm} = \arg\min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma). \tag{1.77}$$
   (d) Update $f_m(x) = f_{m-1}(x) + \sum_{j=1}^{J_m} \gamma_{jm} I(x \in R_{jm})$.

3. Output $\hat{f}(x) = f_M(x)$.

Figure 1.23 shows how the regression of a boosting method for decision trees changes with the number of estimators, i.e. the number of trees that are iteratively added to correct the errors of the combined ensemble.

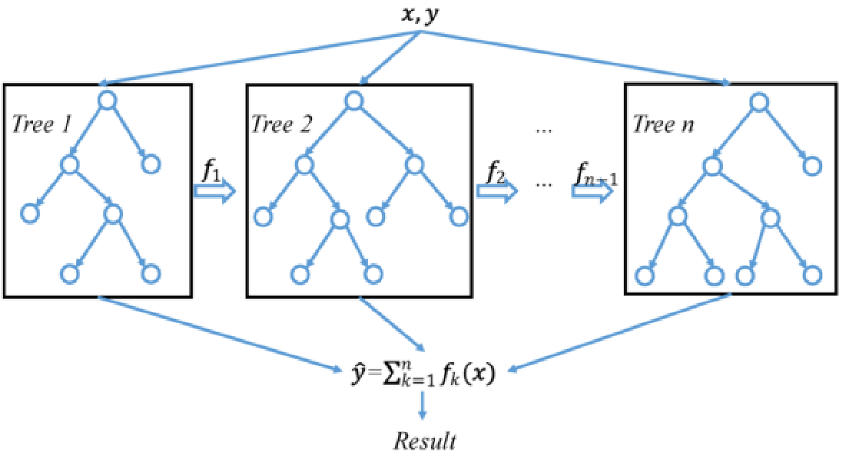


Figure 1.22: Gradient boosting procedure depiction, from Wang, Pan, et al., 2019

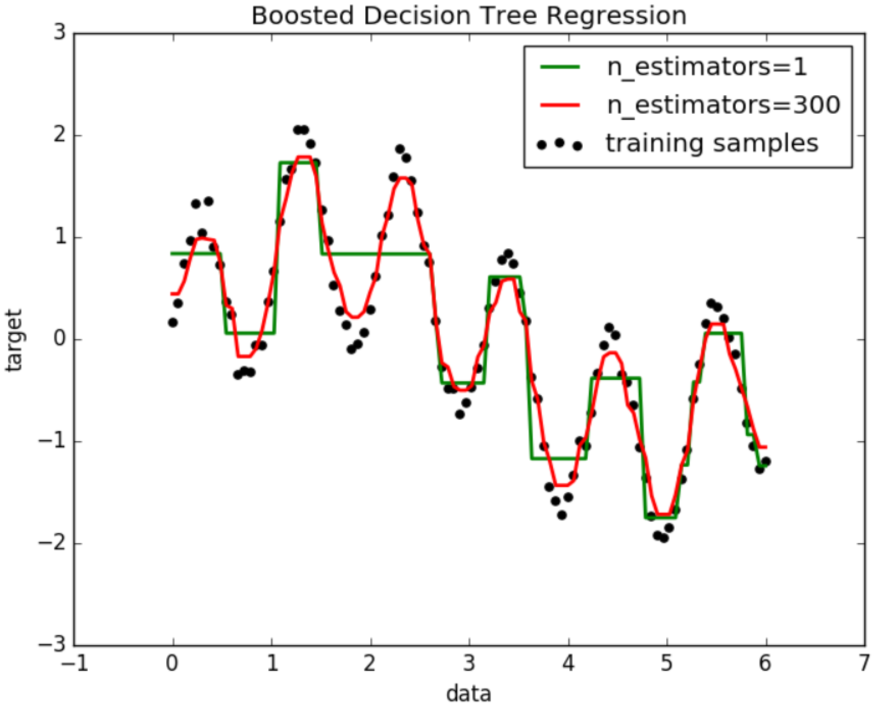


Figure 1.23: Boosting method for decision tree regression with 2 values of the hyperparameter *n_estimators* (`https://scikit-learn.org/stable/auto_examples/ensemble/plot_adaboost_regression.html`).

1.4.2.2 Regression curves

Curve-fitting-based algorithms like Support Vector Machine (SVM, Vapnik, 1963) and LASSO (Least Absolute Shrinkage and Selection Operator, Tibshirani, 1996) focus on finding a function that best fits the data. Considering a linear model:

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j \tag{1.78}$$

a curve-fitting regression task will search for the unknown coefficients β_j's from a set of training data $(x_1, y_1), \ldots, (x_N, y_N)$, constrained by the minimization of a cost function:

$$\mathrm{RSS}(\beta) = \sum_{i=1}^{N} (y_i - f(x_i))^2 = \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 . \tag{1.79}$$

In methods like LASSO, the minimization accounts for an additional constraint, called "regularization", and can be written as (Hastie et al., 2004):

$$\hat{\beta}_{\text{lasso}} = \arg\min_{\beta} \left\{ \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 \right\} \quad \text{subject to} \quad \sum_{j=1}^{p} |\beta_j| \leq t. \tag{1.80}$$

or, in the equivalent Lagrangian form:

$$\mathcal{L}(\beta, \lambda) = \arg\min_{\beta} \left(\sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij} \beta_j \right)^2 + \lambda \left(\sum_{j=1}^{p} |\beta_j| \right) \right) . \tag{1.81}$$

where $\lambda \left(\sum_{j=1}^{p} |\beta_j| \right)$ is called "penalty term", which is the key concept of regularization. Applying this term to the cost function helps the model to prevent overfitting, because it manages the bias-variance tradeoff in model training. In practical terms, while adding a penalty increases the bias, it reduces the variance, helping the model to generalize better to unseen data.

LASSO uses an "L1 Regularization", i.e. it adds a penalty equal to the absolute value of the magnitude of the coefficients. This can lead to models where some coefficients are exactly zero, which is useful for feature selection, because it leads to sparser models (i.e., models with fewer predictors), which can be advantageous for interpretation and when dealing with high-dimensional data.

Another type of regularization is the L2 regularization (Ridge). This consists of adding a penalty equal to the square of the magnitude of the coefficients. This term tends to shrink the coefficients but does not set any of them to zero, resulting in models where all features are included but with reduced magnitude. L2 regularization tends to work better when most features have a significant impact on the output variable. Unlike L1 regularization, which can completely eliminate the weight of some features, L2 regularization only shrinks the coefficients towards zero, but it does not set them to zero. This makes it a better choice when discarding features is not desirable.

SVM is a model that makes use of ridge regularization. It estimates the β coefficients as:

$$H(\beta, \beta_0) = \sum_{i=1}^{N} V(y_i - f(x_i)) + \frac{\lambda}{2}\|\beta\|^2, \tag{1.82}$$

where V is a function which introduces a tolerance ϵ, such that errors of size less than ϵ are ignored:

$$V_\epsilon(r) = \begin{cases} 0 & \text{if } |r| < \epsilon, \\ |r| - \epsilon, & \text{otherwise.} \end{cases} \tag{1.83}$$

The use of a tolerance error, often referred to as the ϵ-insensitive loss (or ϵ-tube) in SVM regression, has several implications for the model's behavior and performance. If the predicted value is within an ϵ margin from the actual value, it is considered an acceptable prediction, and no penalty is applied. This concept introduces a form of robustness in the model, as it is not heavily affected by small deviations.

Moreover, using an ϵ-insensitive loss often leads to sparsity in the solution. Since only the data points that lie outside the ϵ-tube (or on its boundary, where they are defined support vectors) contribute to the model, the model is effectively influenced only by the points that are either errors (outside the margin) or are critical to defining the margin itself.

The SVM minimization problem has solutions:

$$\hat{\beta} = \sum_{i=1}^{N} (\hat{\alpha}_i^* - \hat{\alpha}_i) x_i, \tag{1.84}$$

$$\hat{f}(x) = \sum_{i=1}^{N} (\hat{\alpha}_i^* - \hat{\alpha}_i)\langle x, x_i\rangle + \beta_0, \tag{1.85}$$

where $\hat{\alpha}_i, \hat{\alpha}_i^*$ are positive and solve the quadratic problem:

$$\min_{\alpha_i,\alpha_i^*}\left[\sum_{i=1}^{N}(\alpha_i^*+\alpha_i)-\sum_{i=1}^{N}y_i(\alpha_i^*-\alpha_i)+\frac{1}{2}\sum_{i,i'=1}^{N}(\alpha_i^*-\alpha_i)(\alpha_{i'}^*-\alpha_{i'})\langle x_i,x_{i'}\rangle\right] \tag{1.86}$$

subject to the constraints:

$$\begin{aligned} &0\le\alpha_i,\alpha_i^*\le\frac{1}{\lambda},\\ &\sum_{i=1}^{N}(\alpha_i^*-\alpha_i)=0, \quad (11.41)\\ &\alpha_i\alpha_i^*=0. \end{aligned} \tag{1.87}$$

It can be shown that approximating the solution of Eq. (1.85) using a representation with basis functions $\{h_m(x)\}, m = 1, 2, \dots, M$, leads to a "kernel formulation":

$$f(x)=\sum_{m=1}^{M}\beta_m h_m(x)+\beta_0=\sum_{i=1}^{N}\hat{a}_i K(x,x_i) \tag{1.88}$$

with $K(x, y) = \sum_{m=1}^{M} h_m(x)h_m(y)$ known as "kernel". Depending on the specificity of the SVM application, by using a particular class of functions as kernel to represent the approximate solution, the algorithm can be significantly faster and accurate.

1.4.2.3 Probabilistic approach

Probabilistic-based data-driven methods represent an inherently different approach compared with the other mentioned methods, because they integrate the quantification of uncertainty into the model predictions. This feature makes them particularly suitable for creating surrogate models, since they can effectively emulate the original model variability and generate a realistic ensemble of surrogate outputs. Since the surrogate outputs are obtained significantly faster than the original model outputs, it is possible to generate a large enough number of surrogate outputs to implement the Sobol algorithm described in Section 1.3.3. The following part of the section discusses the two probabilistic methods implemented in ML-AMPSIT: Gaussian Process Regression (GPR) and Bayesian Ridge Regression (BRR).

GPR is equivalent to the more famous spatial interpolation method known as "kriging" (Stein, 1999). It treats each individual parameter-output relationship as a member of a multivariate Gaussian distribution of unknown mean and variance:

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} \sim \mathcal{N}(\mu, \Sigma) \tag{1.89}$$

The covariance matrix $\Sigma = \mathrm{Cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)^T]$ describes the shape of the distribution around the mean vector μ. Inside Σ, the diagonals represent the variances σ_i^2 along the i_{th} variable (i.e. the i_{th} training/test point), and the off-diagonal elements describe the correlation σ_{ij} between the i_{th} and the j_{th} variable. Such a covariance matrix is determined by specifying a kernel, which defines the functional space from which a function can be extracted as a candidate for the fitting; thus, the kernel controls the possible shape that a fitted function can adopt. A common kernel used for GPR, also adopted in this study, is the Radial Basis Function (RBF) kernel, which contains all the functions of the form $\sigma^2 \exp\left(-\frac{(t-t')^2}{2l^2}\right)$ for each $t, t' \in \mathbb{R}^n$ received as input and length scale l. Each time a couple t, t' is extracted from the training data, the kernel function returns the covariance measure between the two points as:

$$k : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad \Sigma = \mathrm{Cov}(X, X') = k(t, t') \tag{1.90}$$

Initially, when no training data are available, a prior distribution is assumed. Each time new training data are considered, a new conditioning is performed, i.e. a new Bayesian inference is computed for the values of the covariance matrix. In other words, the training data constrains the set of functions to only those that pass through it. Gaussian processes also model the errors of each measurement by adding noise to the training points and then obtaining probabilistic predictions by sampling from the resulting distribution and considering a confidence interval. An example is shown in Figure 1.24 for a GPR over 10 randomly generated data points with an RBF kernel.

BRR (Box et al., 1992) is another probabilistic approach starting from the assumption that data comes from a linear model with additive Gaussian noise (Tipping, 2001). Hence the weights w of the regression follow a prior distribution whose parameters α and λ are also assumed to be probabilistic, following hyperpriors Gamma distributions. Knowing the priors and the likelihood, the parameters and hyperparameters are updated by computing the posteriors from Bayes' rule during the assimilation of the training data. It is noteworthy that in BRR the hyperparameter α, controlling the variance of the Gaussian distribution, is also implicitly applying a regularization, since a higher α, i.e., a smaller Gaussian

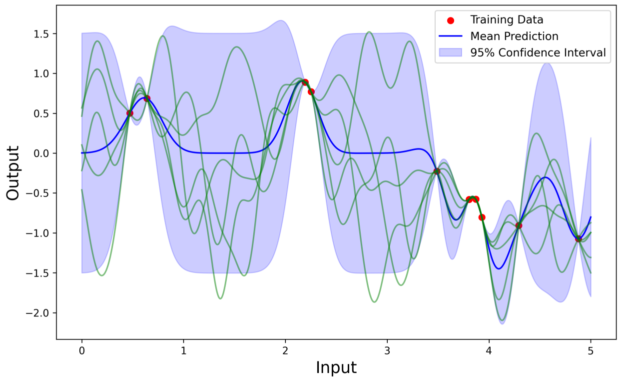


Figure 1.24: Visual example of a GPR with an RBF kernel. Ten training data are used to show the constraints they place on the probability density function.

variance, shrinks the coefficient weights asymptotically toward zero, as in the ridge-type regularization.

Chapter 2

State of the art

2.1 Idealized simulations of the valley wind system

Mountain valleys are characterized, especially during fair weather days in the warm season, by the development of thermally driven circulations due to the differential heating and cooling of the lower atmospheric layers. The daily cycle of this type of circulation for a simplified valley-plain configuration was first theorized by Wagner (e.g. Rampanelli et al., 2004) and subsequently schematized by Defant (Defant, 1951). In the Defant scheme (Figure 2.1), four main wind components alternate: upslope and upvalley winds develop during the day, while downslope and downvalley winds develop during the night. In the slope atmosphere (Whiteman et al., 1984), along-slope air currents result from pressure gradients moving air from the mountain basis to the top of the ridge (anabatic flow) during the day and from the ridge to the bottom of the valley (katabatic flow) during the night. Above the slope circulation, a counterflow compensating the thrust of the low-level air masses is generally present during the day (Monteith et al., 1979), with subsidence in the center of the valley. On the other hand, the convergence of upslope winds can generate rising plumes at the top of the ridge. The valley atmosphere is instead characterized by along-valley winds, which develop due to the combination of different contributions to the thermal gradient. First, the difference in the air volume between the valley and the plain, i.e. the "volume effect" (Rampanelli et al., 2004), causes faster heating and cooling of the former with respect to the latter (Zardi et al., 2013). Then, during daytime, the overheating of the valley atmosphere is enhanced by the subsidence in the valley center due to the presence of the upslope circulation, while the nighttime cooling of the valley is promoted by the convergence towards the valley floor of downslope winds, which accumulate cold air in the lower layers (Rampanelli et al., 2004; Serafin et al., 2011). As a consequence, atmospheric pressure tends to decrease from the plain to the inner sections of the valley during daytime, promoting the development of upvalley winds, while the opposite occurs during the night, with the development of downvalley winds. Above the valley core there is typically another counterflow closing the plain-mountain circulation (Monteith et al., 1979). Radiative processes within the valley also contribute to the strengthening of the thermal gradient with the plain region, since the valley air mass is partially enclosed by radiating surfaces (Monteith et al., 1979; Zardi et al., 2013). Slope circulations react fast to the heating and cooling of the sidewalls, so they generally develop immediately after sunrise (upslope) and sunset (downslope), with brief transition phases. On the other hand, the inertia of the along-valley circulation is greater, with a gradual weakening of the circulation before the reversal of the wind direction. For example, during the evening transition phase, when the stable layer begins to grow from the valley floor, the downvalley wind gradually develops starting from the lower layers, moving the up-valley wind to the valley core (Giovannini et al., 2017).

The sensitivity of this thermally-driven circulation to the geometric properties of an idealized valley has been extensively studied using both Reynolds-averaged Navier Stokes (RANS) simulations and Large-Eddy Simulations (LES). Wagner et al. (2015) performed LES with the WRF model adopting an idealized valley-plain configuration (Figure 2.2), showing that the increase of the valley floor inclination causes a strong intensification of the plain-valley temperature contrast and the subsequent along-valley wind penetrates much deeper into the valley. The increase in wind speed also occurs with the reduction in valley volume and especially with the narrowing of the valley walls. In all simulations the Authors observed a valley inversion layer separating two vertically stacked circulation cells, which is weaker for wider valleys (Wagner et al., 2014). With a similar configuration (Figure 2.3), Wagner et al. (2014) investigated the effects of a systematic variation of valley depth, width, and length, performing several WRF-LES simulations and found that valley deepening strengthens upvalley winds, increases the formation of vertically stacked cells, and of a higher inversion layer; on the contrary, valley widening weakens upvalley winds. The Authors also found that vertical transport is stronger for deep and narrow valleys, which then leads to a more pronounced exchange between the boundary layer and the free atmosphere over the valley compared to the plain region.

The idealized valley configuration has also been used for a model intercomparison study by Schmidli et al. (Figure 2.4), who extended the analytical expression for the topography proposed in Rampanelli et al. (2004). In particular, the Authors compared nine different non-hydrostatic mesoscale models with the same initial and boundary conditions in the simulation of the daytime valley circulation, finding a similar behavior except for a time shift of up to two hours. Their analysis suggests that these inconsistencies are primarily caused by differences in the simulation of the surface energy budget. In the same study, it was shown that different turbulence schemes have a strong influence on the vertical structure of the boundary layer and

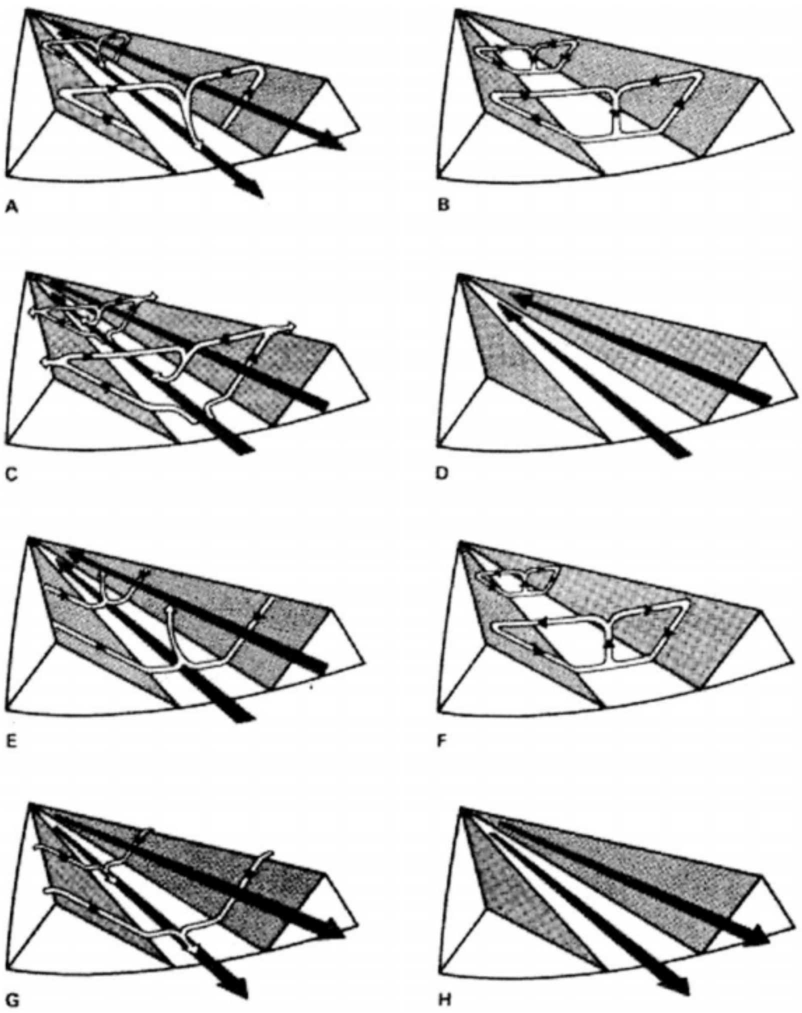


Figure 2.1: Defant schematization of the valley wind cycle, From Rampanelli et al., 2004. (a) Sunrise initiates upslope winds (white arrows) alongside persistent mountain winds (black arrows), with the valley cooler than the plain. (b) By mid-morning, strong slope winds emerge and transition from mountain to valley winds, equalizing valley and plain temperatures. (c) By noon and early afternoon, slope winds subside as valley winds fully develop, warming the valley above plain temperatures. (d) In the late afternoon, hillside winds cease while valley winds continue, maintaining a warmer valley. (e) Evening brings the onset of downslope winds and a decrease in valley winds, with the valley remaining slightly warmer than the plains. (f) Early night brings the establishment of downslope winds and a shift from valley to mountain winds, equalizing temperatures between the valley and the plains. (g) Overnight, continuous downslope and fully developed mountain winds cool the valley below plain temperatures. (h) Late night into morning, downslope winds cease and mountain winds dominate, keeping the valley colder than the plain.

the along-valley wind; in particular, the K-profile-based models tend to simulate the deepest boundary layer. The use of computational horizontal mixing, such as Smagorinsky-type mixing, also results in a wider convergence zone over the ridge. Idealized simulations have also been successfully employed to reproduce the triggering of specific phenomena to better understand the dynamics involved. For instance, Leukaf et al. (2016) performed a series of WRF-LES simulations to study

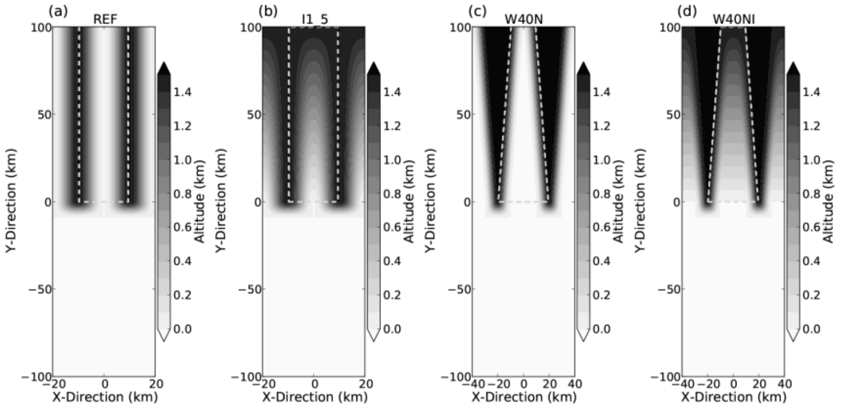


Figure 2.2: Representation of the valley topography used for simulations in Wagner et al., 2014, varying geometric configurations across (a)-(d) setups. Dashed gray boxes indicate horizontal areas at crest height.

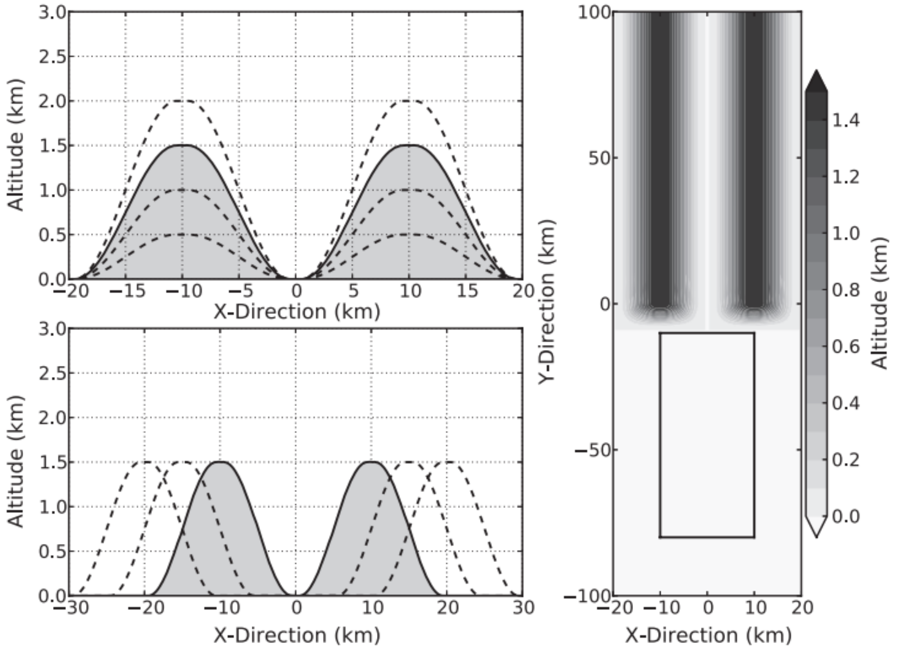


Figure 2.3: Valley topographies for various valley depths and widths used for simulations in Wagner et al., 2015.

the mechanisms leading to the breakup of the valley temperature inversion varying the surface heating and the Brunt-Vaisala frequency. The progressive heating of the valley atmosphere initially favors the development of a convective boundary layer (CBL) with a stable layer on top (the stable core) and a slope wind layer with a stability-dependent height that decreases with increasing height.

Studies such as those presented above contributed to building a solid and reliable tool to reproduce and investigate multiple aspects of valley circulations. This background forms the basis for the idealized configurations used in the present work.

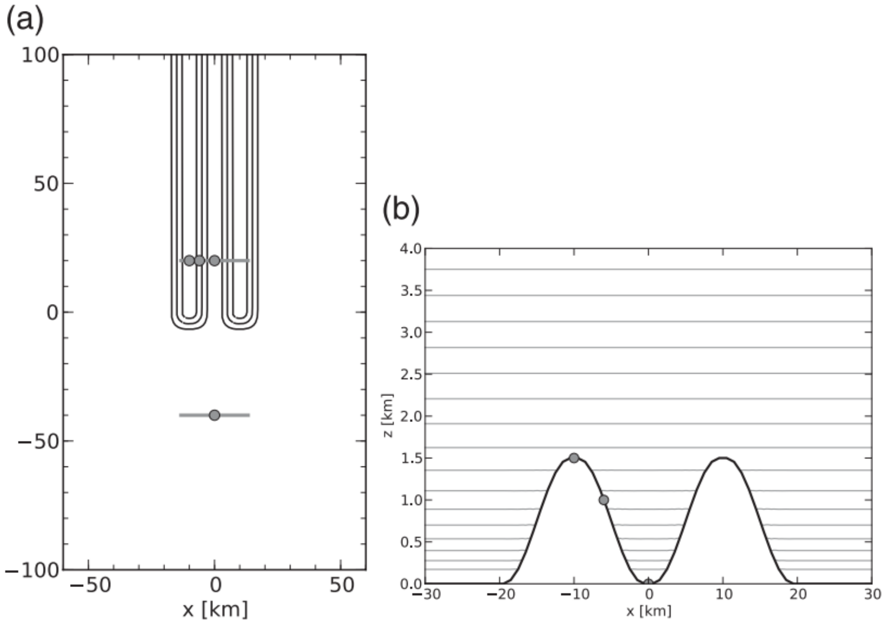


Figure 2.4: Sections of a three-dimensional valley-plain topography used in Schmidli, Billings, et al., 2011.(a) Upper section of the domain at different contour heights (b) Cross-section showing the topography and the initial profile for the potential temperature, with a contour interval of 1 K.

2.2 Sensitivity analysis applied to LSMs

The Sobol method has long been applied to hydrological and environmental models (Gan et al., 2014; Wang and Solomatine, 2019; Pyo et al., 2021; Gong et al., 2015; Rosolem et al., 2012; Griensven et al., 2006; Rosero et al., 2010). Rosero et al. (2010) used this method to evaluate the performance of the Noah land surface model by comparing three different versions, the standard version, the standard version with the adjoint of a groundwater module, and the standard version with the adjoint of a dynamic phenology module. Comparing the Sobol first-order index with the total index, the study showed that the parameter interaction becomes increasingly significant with the complexity of the model version, but also that the variance of the model response is mainly controlled by a few important parameters such as the soil porosity, the bare soil evaporation exponent fx (empirical parameter from the first-layer direct soil evaporation scheme, as reported in Ek et al. 2003) and the leaf area index. The dynamic vegetation module was found to cause the largest increase in interactions but did not improve the model's performance. In addition, the Authors highlighted the presence of interactions between parameters with no functional relation, suggesting that the model was operating differently from the originally intended purposes. The Authors argued that the problems could be related to the absence of a separate canopy layer and of a Ball-Berry stomatal conductance formulation. While both of these options are now available in Noah-MP, this work demonstrated how useful the Sobol method can be in addressing model issues.

In recent years, many extensive studies have been published focusing on testing Noah-MP through GSA. Cuntz et al. (2016) showed that the most influential parameter of an ensemble composed of 42 changeable parameters and 75 hard-coded parameters was a coefficient appearing in the formulation of the soil surface resistance to direct evaporation, which happened to be a hard-coded parameter. Furthermore, only two-thirds of the 42 modifiable parameters were meaningfully sensitive parameters. Notably, as in Rosero et al. (2010), the dynamic vegetation option was considered problematic, arguing that with this option activated the leaf area index sensitivity is spread to new vegetation parameters, increasing the number of interactions and leading to less clear results.

In evaluating Noah-MP parametric ensemble simulations over the Central Tibetan Plateau, Jianduo Li et al. (2018) found that the energy and water fluxes are dominated by soil parameters. Moreover, the canopy height was found to be the most sensitive vegetation parameter, while the Clapp-Hornberger b parameter was the most sensitive soil parameter. With the same workflow philosophy, Kristi et al. (2018) identified five soil parameters and six vegetation parameters as primary controls of soil moisture, sensible heat, latent heat, and net ecosystem exchange, among which some of the most important are the b parameter, the field water capacity, the soil porosity, the canopy height, the roughness length, and the matric potential. In two other works, Jianduo li et al. (2019; 2020) focused on the sensitivity of subprocesses, emphasizing the significant influence on the model performance of those subprocesses related to the computation of the surface exchange coefficients, the soil moisture threshold, the radiation transfer, the surface resistance to evaporation, the runoff, and the groundwater.

The studies presented in the literature review on GSA applied to Noah-MP have always been performed with the model running offline, forced by experimental datasets as initial and boundary conditions. Since the focus of the present thesis is to assess the influence of surface parameters on the local thermally driven circulation in a mountain valley, it is important to evaluate the interaction between the LSM and the coupled numerical weather prediction model, in particular the way the land surface parameters affect atmospheric variables. For this reason, in the present thesis, simulations have been performed with an online coupling between WRF and Noah-MP.

2.3 Machine learning-based sensitivity analysis

Machine learning algorithms have been increasingly employed for feature selection in the last decades, offering significant improvements in various applications. While the following literature review does not aim to be exhaustive, it highlights some notable examples that underscore the evolving role of machine learning in this field.

Regression algorithms have a long history in data science, with some of the most famous methods introduced between the 1960s-1990s: for example, the Support Vector Machine (SVM) was introduced by Vapnik et al. Vapnik, 1963 (1963), the Classification And Regression Trees (CART) by Breiman et al. (1984), the Least Absolute Shrinkage and Selection Operator (LASSO) by Tibshirani et al. (1996), the Bayesian Ridge Regression (BRR) by Box et al.(1992), the Gaussian Process Regression (GPR) by Stein et al. (1999), and the Random Forest (RF) by Breiman et al. (2001). Despite this, their applicability to sensitivity analysis has been fully formalized in later years. Oakley et al. (2004) addressed the complexity of mathematical models used to simulate real-world phenomena, particularly focusing on the challenge of understanding how changes in model inputs affect outputs. Given the uncertainty in determining accurate input values for specific applications, the Authors introduced a unified framework for probabilistic sensitivity analysis. This method quantifies uncertainty in model outputs induced by the variation in inputs framed within a joint probability distribution. The significant contribution of Oakley et al. (2004) is the introduction of a Bayesian framework to unify various tools of probabilistic sensitivity analysis. This approach is computationally efficient, requiring fewer model runs than standard Monte Carlo methods while still enabling a comprehensive analysis. This efficiency is especially crucial for models where each run demands considerable computational time. One year later, Queipo et al. (2005) introduced the definition of "Surrogate-Based Approach for Analysis and Optimization (SBAO)" to emphasize how surrogate models, constructed from data derived from high-fidelity models, offer fast approximations of objectives and constraints, making sensitivity and optimization studies feasible despite the computational expense of accurate, high-fidelity models. At the Fourth International Conference on Sensitivity Analysis Model Output (SAMO) and in continuation with the work of Oakley, O'Hagan (2006) presented a systematic approach to sensitivity analysis called Bayesian Analysis of Computer Code Outputs (BACCO), which put the foundation for the use of probabilistic methods as the Gaussian Process Regression emulation in surrogate-based sensitivity analysis. In more recent years, a systematic review performed by Brunton et al. (2020) meticulously outlines the history of ML in fluid mechanics, highlighting the successes of ML in critical tasks such as dimensionality reduction, feature extraction, Particle Image Velocimetry (PIV) processing, super-resolution, reduced-order modeling, turbulence closure, shape optimization, and flow control. Kim et al. (2020) explored the efficiency of subset selection regression techniques in developing surrogate models for simulation-based or data-driven system optimization. The Authors compared these techniques with nonlinear interpolating functions, such as radial basis functions and Kriging, finding that subset selection shows promising performance for low-dimensional problems, although showing performance degradation in higher dimensions. Antoniadis et al. (2021) focused on GSA methods in complex computational models with numerous input variables. The Authors proposed using meta-modeling as a quantitative approach for nonparametric GSA, highlighting the efficiency of RF in building meta-models for sensitivity analysis and showing that the RF methodology offers advantages in handling correlation, high-dimensional data, and variable interactions. Moroever, the Authors demonstrated the effectiveness of RF in estimating GSA measures through simulations and real dataset analysis, even in cases of dependent data and small sample sizes. Vinuesa et al. (2022) highlighted the growing importance of machine learning in computational fluid dynamics (CFD), identifying areas such as the acceleration of direct numerical simulations, improvement of turbulence closure modeling, and development of enhanced reduced-order models. The Authors discussed emerging ML areas with a high potential impact on CFD, along with potential limitations.

Compared to their proliferation in fluid dynamics, ML-based surrogate models are still quite new in the weather and climate literature, but there have been foundational works also in this field in recent years: Singh et al. (2019) explored the application of SVM, ANN, and a Time Series based Recurrent Neural Network (RNN) for weather prediction. The study emphasized the superior performance of RNN in predicting weather, highlighting the effectiveness of these models in meteorological applications. Parente et al. (2020) developed a multi-fidelity framework to predict the turbulent Schmidt number (Sc_t) with applications in atmospheric dispersion modeling. This framework utilized high-fidelity data from a delayed Detached Eddy Simulation (dDES) and employed data-driven models and Principal Component Analysis (PCA) to derive reliable information for Sc_t, demonstrating the potential of multi-fidelity approaches in enhancing the accuracy of turbulence models for atmospheric dispersion studies. Chen et al. (2020) reviewed the use of ML in tropical cyclone forecasting, covering genesis, track, intensity, and associated extreme weather forecasts. The Authors highlighted the untapped potential of ML and multi-source data in improving forecast accuracy, addressing the challenges of predictability and stability in tropical cyclone prediction.

Dagon et al. (2020) advanced the application of ML in land modeling by developing a calibration procedure for the Community Land Model version 5 (CLM5) using artificial neural networks. In particular, the Authors focused on refining the model's parameters based on carbon and water flux observations. The innovative approach involved creating emulators of CLM5 using feed-forward artificial neural networks, trained to predict annual mean spatial variability in carbon and water fluxes. Techniques such as regularization and early stopping were utilized to prevent overfitting and ensure generalizability. This study highlighted the significance of parameters like plant hydraulic conductivity and the stomatal conductance–photosynthesis relationship in influencing the predictive skill of the emulator. This approach marked a shift from traditional hand-tuning methods to a more efficient machine learning-based optimization of parameter values, considering observational uncertainty and model error.

Mansfield et al. (2020) introduced an ML approach to expedite climate change projections and decipher early indicators of long-term climate responses. Utilizing existing climate model simulations, the study aimed at individuating relationships between short-term and long-term temperature responses under various climate forcing scenarios. The study compared Ridge Regression and GPR with traditional pattern scaling, finding that the latter often underestimated the range and variability of responses, especially in short-lived forcing scenarios. Both Ridge Regression and GPR generally outperformed pattern scaling at different spatial scales, with GPR showing the lowest errors. The study demonstrated that ML models, particularly at the grid-scale level, can provide more accurate and diverse climate projections than traditional methods.

The application of ML techniques to climate modeling was also proposed by Chaure et al. (2023), who tackled the challenges in climate modeling related to the computational demands of fully coupled General Circulation Models (GCMs). Their research explored the potential of using GPR as a surrogate model. Given the high computational resources required by GCMs, surrogate models offer a more efficient alternative. The Authors emphasized the advantages of GPR over deep learning surrogates, particularly its probabilistic approach and the ability to provide nuanced modeling of uncertainty estimates.

A different application of ML algorithms was proposed by Gorle et al. (2021), who presented a multi-fidelity ML framework that combines RANS and LES to accurately predict wind loads on buildings. This framework efficiently balances the computational cost with prediction accuracy, demonstrating the utility of multi-fidelity ML approaches in complex turbulent flow simulations.

ML approaches are also used for solar radiation forecasts: for example Wu et al. (2021) assessed the transferability of RF and SVM in estimating daily global solar radiation across different climate zones. The study underscored RF's superior performance and stability over SVM, suggesting specific conditions under which RF and SVM can be effectively applied for solar radiation estimation.

Finally, ML was also recently used for identifying sensitive parameters in the WRF model by Baki et al. (2022), considering the simulation of tropical cyclones over the Bay of Bengal. Employing GSA methods, the study found that a subset of parameters significantly influenced the meteorological variables, offering insights into the optimization of tropical cyclone simulations.

Chapter 3

ML-AMPSIT: Machine Learning-based Automated Multi-method Parameter Sensitivity and Importance analysis Tool

This chapter has been submitted to GMD in 2024 and authored by: Dario Di Santo[1], Cenlin He[2], Fei Chen[3], and Lorenzo Giovannini[1].

Abstract

The accurate calibration of parameters in atmospheric and Earth system models is crucial for improving their performance, but remains a challenge due to their inherent complexity, which is reflected in input-output relationships often characterized by multiple interactions between the parameters and thus hindering the use of simple sensitivity analysis methods. This paper introduces the Machine Learning-based Automated Multi-method Parameter Sensitivity and Importance analysis Tool (ML-AMPSIT), a new tool designed to estimate the sensitivity and importance analysis of parameters in complex numerical weather prediction models. This tool leverages the strengths of multiple regression-based and probabilistic machine learning methods including LASSO, Support Vector Machine, Classification and Decision Trees, Random Forest, Extreme Gradient Boosting, Gaussian Process Regression, and Bayesian Ridge Regression. These regression algorithms are used to construct computationally inexpensive surrogate models to effectively predict model outputs from input parameters, thereby significantly reducing the computational burden of running high-fidelity models for sensitivity analysis. Through a detailed case study with the Weather Research and Forecasting (WRF) model coupled with the Noah-MP land surface model, ML-AMPSIT is demonstrated to efficiently predict the behavior of Noah-MP model parameters with a relatively small number

---

[1] Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

[2] NSF National Center for Atmospheric Research (NCAR), Boulder, CO, USA

[3] Division of Environment and Sustainability, Hong Kong University of Science and Technology, Hong Kong, China

of model runs, by simulating a sea breeze circulation over an idealized flat domain. ML-AMPSIT's ability to consider non-linearities and interactions between variables, as well as to provide a comparative analysis across various machine learning methods, offers an alternative promising new approach for sensitivity analysis in atmospheric sciences.

3.1 Introduction

One of the primary sources of error in atmospheric/Earth system models stems from inaccurate parameter values (Clark et al., 2011; Li, Chen, Zhang, et al., 2018), which can affect different physical parameterizations. Although model parameter tuning can help to alleviate this issue, determining optimal values is highly dependent on model structures and how model outputs are influenced by input parameters. Sensitivity analysis is commonly used to evaluate these input-output relationships and parameter importance, but traditional one-at-a-time (OAT) methods yield varying results depending on the interdependence of parameters, particularly within complex models, leading to issues of poor reproducibility and inability to generalize results. Consequently, more advanced variance-based techniques like the Sobol method, in the context of global sensitivity analysis (GSA, Saltelli, Ratto, et al., 2008), exhibit superior performance in such tasks, albeit being computationally intensive (Herman et al., 2013) and sometimes infeasible, especially when dealing with complex weather/climate models like the widely used Weather Research and Forecasting (WRF) model (Skamarock, Klemp, Dudhia, Gill, Liu, Berner, Wanh, et al., 2021).

An alternative approach that avoids running numerous model realizations is the utilization of surrogate models or emulators (Fernández-Godino et al., 2017; Forrester et al., 2008; Kim et al., 2020; Lamberti et al., 2021; Longo et al., 2020; O'Hagan, 2006; Queipo et al., 2005). A surrogate model/emulator is a simpler model trained using the input-output pairs of the original complex high-fidelity model, that can be used in substitution of it. The emulator makes the model process more computationally efficient in producing model realizations, while it still provides accurate predictions of the output variable. Machine Learning (ML) algorithms designed for regression tasks offer a computationally efficient means to build surrogate models to be used for sensitivity analysis (Antoniadis et al., 2021; Engelbrecht et al., 1995; Muthukrishnan et al., 2016; Shen et al., 2008; Torres, 2021; Zouhri et al., 2022). Over time, a variety of algorithms have been tested in the literature and used in different fields.

These algorithms can also be used to extract feature importance, which has become a well-established methodology widely employed in different geoscience fields, such as landslide susceptibility (Catani et al., 2013; Chen, Zhu, et al., 2020; Daviran et al., 2023; Elia et al., 2023; Kalantar et al., 2018; Lee et al., 2018; Liu et al., 2021; Pradhan, 2013; Yilmaz, 2010; Youssef et al., 2016; Zhou et al., 2018), forest fire susceptibility (Arpaci et al., 2014; Bar Massada et al., 2013; Gigović et al., 2019; Oliveira et al., 2012; Pourtaghi et al., 2016; Satir et al., 2016), water quality assessment (Alqahtani et al., 2022; Haghiabi et al., 2018; Palani et al., 2008; Rodriguez-Galiano et al., 2014; Sarkar et al., 2015; Shah et al., 2021; Trabelsi et al., 2022), hydrological modelling (Zhang, Srinivasan, et al., 2009), air quality assessment (Lei et al., 2023; Maleki et al., 2019; Sihag et al., 2019; Suárez Sánchez et al., 2011; Yu et al., 2016), groundwater mapping (Rahmati et al., 2016), agronomy (Kok et al., 2021; Sridhara et al., 2023; Wu, Cui, et al., 2023), climatological applications (Dey et al., 2022; Wu, Li, et al., 2021), renewable energy (Meenal et al., 2022; Wolff et al., 2017), earthquake detection (Murti et al., 2022), and it also has significant relevance in civil engineering (Farooq et al., 2020; Gholampour et al., 2017; Salmasi et al., 2020; Tian, 2013), genetics (Sharma et al., 2014), biology (Cui et al., 2016), and medical research (Antonogeorgos et al., 2009; Maroco et al., 2011; Yang, Wang, et al., 2022). While ML techniques have gained traction in weather and climate modeling and observations (Schneider et al., 2022; Schultz et al., 2021), particularly in parameter optimization tasks like calibration (Bocquet et al., 2020; Bonavita et al., 2020), spatial interpolation (Sekulić et al., 2020; Stein, 1999), downscaling (Fowler et al., 2007; Leinonen et al., 2021; Maraun et al., 2018), parameterization substitution (Grundner et al., 2022; Han et al., 2020; Mooers et al., 2021; Rasp et al., 2018; Ross et al., 2023; Yuval et al., 2020), and image-based classification (Chase, Harrison, Burke, et al., 2022; Chase, Harrison, Lackmann, et al., 2023), the extraction of feature importance remains relatively uncommon in the meteorological/climate modeling literature (Baki et al., 2022; Ren, Li, et al., 2020). Additionally, several studies have explored the use of emulators for tuning weather prediction and climate models, highlighting the potential of these methods for parameter space exploration and model calibration (Cinquegrana et al., 2023; Couvreux et al., 2021; Dagon et al., 2020; Watson-Parris et al., 2021; Williamson et al., 2013).

Many of the previous studies have proposed comparisons between several feature importance analysis algorithms. This is because the ability of these algorithms to best capture feature relevance is influenced by a variety of factors that can change with applications, depending on the context under analysis, such as the degree of nonlinearity of the input-output relationships, the interaction degree between features, the dimensionality of the features, the size and quality of the data used for training, the shape and smoothness of the distribution of the training data, and ultimately the validity of each algorithm's assumptions. Often, the influence of one or more of the factors on the quality of the chosen method cannot be assessed in advance, leading to a trial-and-error procedure that would benefit from a multi-algorithm approach where the results of different methods can be compared. For this reason, the present work shares the same multi-method philosophy of many of the studies mentioned above, extracting the most popular algorithms available in the literature and combining them into a single flexible efficient framework for analysis.

In light of the above considerations, this paper proposes a Machine Learning-based Automated Multi-method Parameter Sensitivity and Importance analysis Tool (ML-AMPSIT) that applies a series of ML feature importance extraction algorithms to model parameters (using the widely-used WRF/Noah-MP coupled weather model as a case study), accommodating any user-specified model configuration. The included algorithms are among the most commonly used in literature, namely: LASSO, Support Vector Machine, Classification and Decision Trees, Random Forest, Extreme Gradient Boosting, Gaussian Process Regression and Bayesian Ridge Regression. While most of them directly provide a measure of feature importance, the last two methods are probabilistic methods, specifically used in this framework for a fast implementation of the Sobol method through the SALib Python library, leading to a computationally efficient way to obtain the Sobol sensitivity indices directly from the ML-inferred relation between input and output data. Our tool's objective is overall to assist users in evaluating parameter sensitivity and importance using computationally inexpensive and nonlinear interaction-aware approaches.

3.2 Methods

3.2.1 ML-AMPSIT workflow

The ML-AMPSIT workflow (Figure 3.1) can be divided into 5 main steps, each of which involves one or more Python/Bash-based scripts: the pre-processing phase, the model run phase, the post-processing phase, the sensitivity analysis phase, and an optional convergence analysis phase.

1. The selection of the input features is accomplished by specifying the parameter names within the configuration file *configAMPSIT.json*. The compiled configuration file related to the case study discussed in this paper is reported in the Appendix. There is no upper limit for the number of parameters that can be analyzed, but it is worth noting that the sensitivity analysis could converge significantly more slowly in high-dimensional (i.e., with more parameters) problems. Moreover, the scalability with the number of parameters

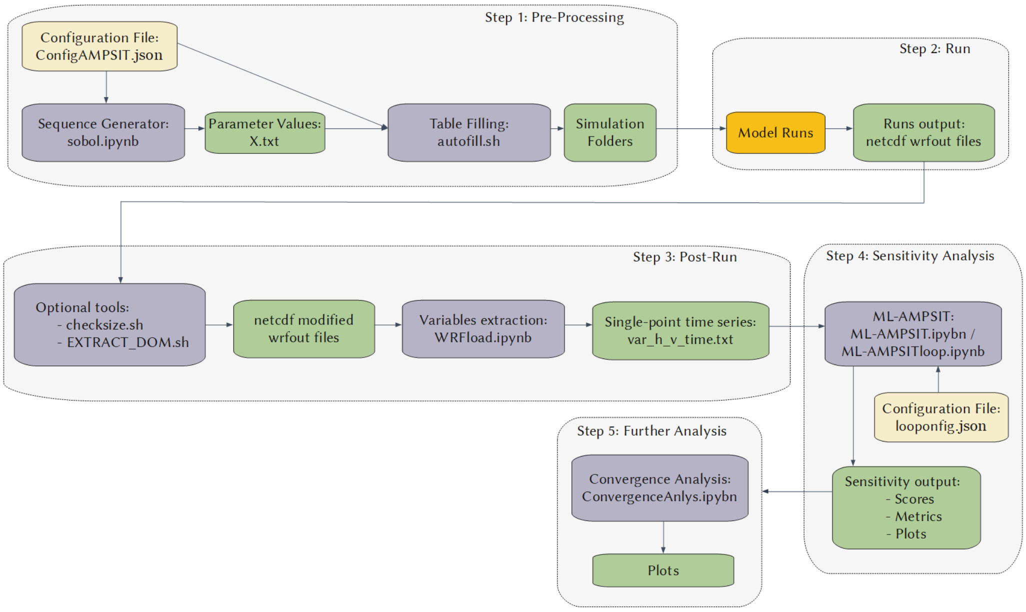


Figure 3.1: ML-AMPSIT workflow. The main code scripts are indicated with blue boxes. The yellow boxes indicate the configuration files that need to be filled by the user, the green boxes refer to all the output files that eventually become inputs for other subsequent scripts, and the orange box indicates the generic model execution that varies depending on the model involved.

can highly depend on the case study considered. The number of simulations to perform must also be specified through the configuration variable *totalsim*. To generate the values of the parameters to be tested, a Sobol sequence of the same length of *totalsim* is produced for each parameter from the preprocessing script *sobol.ipynb*. The Sobol sequence (Saltelli, Annoni, et al., 2010, Bratley et al., 1988) is a quasi-random low-discrepancy distribution designed to produce well-spaced points in the unit hypercube representing the parameter space. Unlike random sampling, each point in the sequence considers the positions of the previous points, resulting in a more uniform filling of gaps, as shown in Figure 3.2. Consequently, a robust sequence is generated more efficiently compared to random sampling, requiring fewer points.

Once the Sobol sequences are generated, a user-specified reference value and a maximum perturbation percentage need to be specified in *configAMPSIT.json*, which will be passed to the preprocessing script *autofill.sh*. These values are used to rescale the sequence values from the [0,1] range to the actual parameter range space. The output of the *sobol.ipynb* script is the file *X.txt*, containing a m x n matrix, where m is the number of simulations and n the number of parameters tested.

Therefore, each row specifies a different set of parameter values that will be

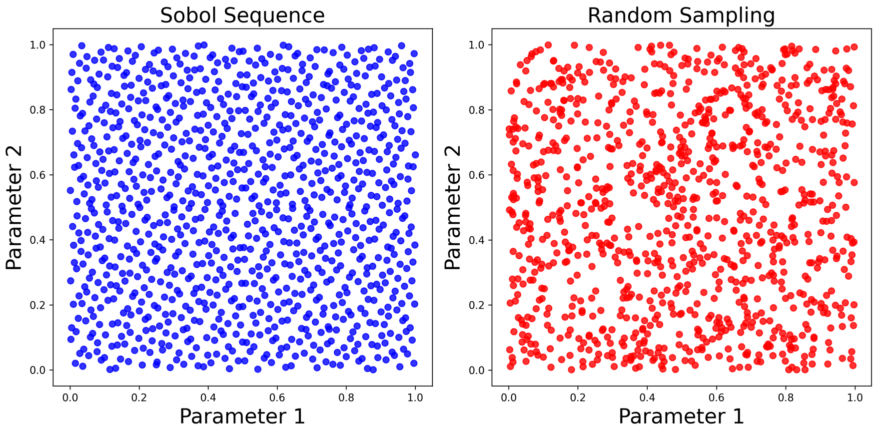


Figure 3.2: Demonstration of the differences between random sampling (red dots) and Sobol sampling (blue dots) in representing the parameter space. The Sobol sequence is able to more uniformly cover the parameter space, avoiding the presence of very close points, as it occurs in the random sequence.

used in each particular model realization. Based on these data, the *autofill.sh* script creates multiple copies of the folder in which the model is run and then searches for each parameter name within the original model parameter look-up table in each newly created folder. The values of the parameters are then changed according to the *X.txt* file for each realization. Since this script edits the original model parameter look-up table, which is MPTABLE.TBL for the WRF/Noah-MP model in the case study, it is necessarily model-dependent and thus needs to be adapted if used with other models, to suitably modify the values of the tested parameters.

2. After all the simulation folders have been created, the original high-fidelity model can be run by the user as usual. It will be necessary to collect all the output files into one single folder, whose path must be specified in the configuration file, so that the post-processing script can find it.

3. Once the user has completed all the high-fidelity model runs, a post-processing script named *WRFload.ipynb* is provided to extract single-point time series for each output variable at specific coordinates in the simulation domain, as specified by the user in the configuration file. The resulting output data, which serves as input to the sensitivity analysis tool *ML-AMPSIT.ipybn*, consists of different files with the naming convention *var_h_v_time.txt*, where *var* is the variable name, *h* and *v* indicate the labels identifying the horizontal and vertical grid cell, while *time* represents the simulated time. The script *WRFload.ipynb* specifically extracts variables from NetCDF files that follow a WRF-like format (a widely used format for weather/climate models). If the user's model output follows a different format, the script must be modified accordingly.

Extracting single-point time series means that the tool has no information about spatial patterns, nor can it capture time patterns, because the sensitivity analysis is performed on each single point and time separately. Consequently, the importance of different parameters can be directly compared only for the specific point and time being analyzed. However, the observation of consistent relative importance between parameters across different points and times can confirm the statistical robustness of the results.

4. The sensitivity phase is performed by the main script of the ML-AMPSIT tool *ML-AMPSIT.ipybn*, which accomplishes a regression task based on different ML algorithms offered to the user. As mentioned in the Introduction, this multi-method approach is useful for comparing different results, since each algorithm is structurally different and could be more or less appropriate for the problem at hand. This ML-based ensemble philosophy is similar to an ensemble learning (EL) approach (Ren, Zhang, et al., 2016), which combines the predictions of multiple base models to improve the overall performance, but the present tool does not yet include an option to integrate the methods into a stacking, bagging, or boosting procedure, allowing for the user to choose any single or multiple methods independently. For each method offered by the present tool, the input and output data are split into training and testing sets in proportions of 70% and 30%, and each set is scaled separately to have zero mean and unit variance with respect to the ensemble. The training set is used to fit the model to the data, while the testing set is used to evaluate the model's ability to reproduce new data. This strategy is used to mitigate the risk of overfitting. The coefficient of determination (R^2), the mean squared error (MSE), and the mean absolute error (MAE) have been used as measures of goodness when comparing the predicted output against the actual "truth", i.e. the results of the original physical model simulations. Since all the variables are scaled before calculating these error metrics, MSE and MAE are not affected by the different scales of the variables. This allows for a fair and meaningful comparison of the model's performance across different variables. The coefficient of determination $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$, is used as a measure of goodness of fit, where SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares. R^2 indicates how much variation in the target variable can be explained by the model's predictors. R^2 is typically a value between 0 and 1, where values closer to 1 indicate a better ability of the model to explain the variance in the data. Eventually, if the chosen model fits worse than the average value, then $\frac{SS_{res}}{SS_{tot}}$ can be greater than 1 and R^2 is negative. If the model has low values of MSE and MAE, but also low values of R^2, it might indicate that the relationship between the input data and the target variable cannot be properly explained in terms of linear weights only. This is

an indication of nonlinearity in the output response. In addition to the R^2 coefficient, the associated p-value is also computed and saved.

The script *SAML.ipybn* produces an interactive graphical user interface (GUI) built from the ipywidgets python library, which allows the user to specify which vertical level and surface point to consider in the analysis, the output variable for which to compute the sensitivity analysis, the number of simulations to consider, the algorithm to use, and the output time to plot for punctual evaluations. The flexibility of this GUI allows the user to quickly check the influence of the number of simulations on the robustness of the results and the performance of the different ML methods implemented.

Based on the selected options, unless the specified methods are Gaussian Process Regression or Bayesian Ridge Regression, the tool produces four plots: the upper two dedicated to the feature importance time series and the time-evolution of the metrics for the whole simulation duration, and the lower two showing the metrics and feature ranking specific to a particular time selected. Hence the user is provided with both the global result and the analysis related to a single output time. An example from the proposed case study will be provided in Section 3.3. If the specified method is either Gaussian Process Regression or Bayesian Ridge Regression, the features are ranked based on the total Sobol index, and the tool produces two additional plots, one showing the second-order Sobol interaction index between each couple of parameters, and the other showing the feature ranking based on the first-order Sobol index, which could potentially be different from the total order based ranking if parameters' interactions are strong enough.

An alternative version of the main script is the ML-*AMPSITloop.ipynb* loop suite, which allows the user to automate multiple analyses by using the *loopconfig.json* configuration file, specifying all the combinations of settings to be explored, without the need to manually set each combination through the graphical interface, saving time for long in-depth analyses.

5. It is possible to further investigate the convergence of the obtained feature importance and metrics with the optional tool *convergenceAnlys.ipybn.* This script manipulates all the outputs of *ML-AMPSIT.ipybn* or *ML-AMPSITloop.ipybn* to visualize the sensitivity of the parameter scores to the number of simulations, the vertical grid point, the horizontal grid point, and the algorithm used. The plots produced by this tool, related to the case study presented in this paper, will be shown and commented on in more depth in Section 4.2.

3.2.2 Sobol method

In classical sensitivity analysis, given a set of input parameters $\{X_1, ..X_k\}$, the elementary effect of a single perturbation Δ in the input parameter X_i over the output $Y(X)$ is defined as ("Elementary Effects Method" 2007):

$$EE_i = \frac{[Y(X_1, X_2, ..., X_i + \Delta, ..., X_k) - Y(X_1, X_2, ..., X_i, ..., X_k)]}{\Delta}, \tag{3.1}$$

The above definition assumes a linear relationship between the parameter and the output variable and it becomes ineffective in the presence of non-linearities or interactions between parameters. To achieve the highest level of generalization in sensitivity analysis, both the effect of the single input parameter and the additional effect of its interaction with other parameters must be evaluated.

A variance-based approach achieves this while also not relying on a linear assumption. The most well-established variance-based method is the Sobol method (Saltelli and Sobol', 1995). With this approach, the variance $V(Y)$ is decomposed as:

$$V(Y) = \sum_i V_i + \sum_i \sum_{j>i} V_{ij} + \cdots + V_{12...k}, \tag{3.2}$$

where V_i is the main effect variance, representing the contribution of the i-th input parameter to the output variance, V_{ij} is the second-order interaction effect variance, representing the combined contribution of the i-th and j-th input parameters to the output variance, and so on up to $V_{12..k}$, which represents the interaction effect variance of all k input parameters together.

Dividing by $V(Y)$, the Sobol indices are derived as $S_i = V_i/V(Y)$, $S_{ij} = V_{ij}/V(Y)...$, leading to:

$$1 = \sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \cdots + S_{12...k}. \tag{3.3}$$

The total effect, or total order index S_{Ti} is the total contribution to the output variation due to a specific factor X_i i.e. a specific parameter. Hence, for instance, for a set of three parameters X_1, X_2, X_3 the total effect index for the parameter X_1 is:

$$S_{T1} = S_1 + S_{12} + S_{123}. \tag{3.4}$$

Both first-order effects S_i and total effects S_{Ti} are important to assess the overall influence of an input parameter.

3.2.3 Implemented ML algorithms

3.2.3.1 LASSO

The Least Absolute Shrinkage and Selection Operator, (LASSO, Tibshirani, 1996) is an ML method used for feature selection and regression. It derives from the basic concept of curve fitting in the context of optimization, therefore it is one of the simplest algorithms among the most widely used. The goal of LASSO is to identify a subset of input features that are most predictive of the output variable, while also performing regularization to prevent overfitting.

In a classical regression problem, the goal is to find a function that maps the input features to the output variable. This is done by minimizing an objective function that takes into account the differences between the observations and the predictions, as a measure of how well the model fits the data. The minimization is performed during the training process to find the optimal values of model coefficients. In the LASSO algorithm, a penalty term (the "regularization") is added to the objective function, encouraging the model coefficients to be small. Specifically, the objective function for LASSO regression is the residual sum of squares (RSS), while the penalty term is based on the sum of the absolute values (L1 norm) of the coefficients, which promotes sparsity of the solution, and it can be adjusted to control the amount of shrinkage applied to the coefficients. As a result, LASSO can be particularly useful to identify the most important features by setting the coefficients of the less important features to zero.

The present tool uses LassoCV from the Python library scikit-learn, which adds a cross-validation strategy to the standard LASSO algorithm. In cross-validation, data are divided into multiple subsets or "folds", and the model is trained and evaluated multiple times, each time using a different fold as the validation set and the rest of the data as the training set (e.g. as explained in (Brunton, Noack, et al., 2020)). By averaging the results of the multiple evaluations, cross-validation provides a more accurate estimate of the model's performance than a single evaluation on a single training-validation split.

3.2.3.2 Support Vector Machine

Support vector machine (SVM) is an ML method that can be used for regression as well as classification tasks. The original algorithm was proposed in 1963 (Vapnik, 1963), was refined multiple times (Boser et al., 1992; Cortes et al., 1995; Schölkopf et al., 2002), and is considered a core method in ML.

For regression tasks, SVM aims to find a hyperplane that approximates the underlying relationship between the input variables (i.e. the model parameters) and the continuous output values. In order to do this, SVM solves a convex optimization problem by minimizing a cost function that incorporates a margin of error and an L2-norm-based regularization term (ridge-type regularization). Unlike the LASSO-type regularization, the L2 regularization, also known as ridge-type regularization, is a penalty based on the square of the coefficients. This penalty term is less strict compared to LASSO, because it helps to shrink the coefficient values towards zero without eliminating them completely. This can be preferred to a LASSO regularization in situations when there are many important parameters, since it avoids eliminating them in favor of the most important ones. The regularization term controls the trade-off between the complexity of the model and the amount of error allowed, and its strength is controlled by the hyperparameter "C". The optimization problem in the proposed implementation is solved by the default quadratic "liblinear" Python solver. SVM offers several advantages, including the ability to handle high-dimensional data and resistance to overfitting when properly regularized.

The tool presented in this paper implements a simple linear kernel, which maps the input data into a high-dimensional space using a linear function.

3.2.3.3 Classification and Regression Trees

The Classification and Regression Trees (CART, Breiman et al., 1984) algorithm is one of the most basic and straightforward decision tree methods, making it widely used and popular for various applications. The main idea behind CART is to recursively partition the input space into smaller regions based on the values of the input variables, with the goal of minimizing the impurity (or variance) within each resulting region. This partitioning process creates a binary tree structure called a "decision tree", where each internal node represents a splitting rule based on a selected input variable and a threshold value. The leaf nodes of the tree represent the final prediction or class assignment.

In the proposed tool, from the training dataset consisting of input-output pairs, the algorithm recursively selects the best splitting rule at each internal node based on the mean squared error for regression. The splitting process continues until a stopping criterion is met, such as a maximum tree depth or a minimum number of samples required to split a node. For regression tasks, the predicted output is the average value of the training samples within the leaf node.

3.2.3.4 Random Forest

Random Forest (RF, Breiman, 2001) is technically an ensemble learning method, that can be used both for classification and regression tasks. Differently from the

CART algorithm, a multitude of decision trees are constructed in an RF, and the final prediction is based on the average of the predictions of all the trees. For this reason, RF usually performs significantly better than CART.

The advantage of using RF is that it can handle high-dimensional data and complex interactions between variables, making it a useful tool for sensitivity analysis. However, it can suffer from overfitting, where the model performs well on the training data but poorly on new data. To mitigate this, both cross-validation and Bayesian optimization are used to make the model more robust.

3.2.3.5 Extreme Gradient Boosting

The Extreme Gradient Boosting (XGboost, Chen and Guestrin, 2016) builds an ensemble of decision trees, similarly to RF, but each new tree is trained to correct the errors of the previous trees. It is hence a refined version of an RF and it is usually expected to perform better.

3.2.3.6 Gaussian Process Regression

Gaussian Process Regression (GPR), better known in geosciences as kriging (Stein, 1999) when applied to spatial interpolation, is a nonparametric algorithm for computing the probability density function of the regression curve, instead of a single fitting curve and can be used as a supervised ML technique (Rasmussen et al., 2005). GPR assumes that the output values follow a Gaussian distribution with unknown mean μ and variance σ which must be predicted, given a set of input-output pairs. To achieve this, GPR models the output values as a function of the input variables, where the function is assumed to be smooth and continuous. GPR is often described as a non-parametric method because it does not assume a specific functional form for the relationship between input and output variables. Instead, it models this relationship as a distribution over possible functions, allowing for flexibility in the shape of the regression curve. However, it is important to note that there are underlying assumptions about the functional form embedded in the chosen kernel. The kernel influences the shape and properties of the functions that the Gaussian Process can learn. The kernel implemented inside ML-AMPSIT is the Radial Basis Function (RBF) kernel, which uses functions of the type $\sigma^2 \exp\left(-\frac{(t-t')^2}{2l^2}\right)$ where σ^2 is a variance, l is a length scale and t, t' represent pairs of values extracted from the training data. The choice of an RBF kernel is particularly advantageous due to its ability to model complex, non-linear relationships without imposing strong parametric constraints. The RBF kernel's smoothness assumption is well-suited for many real-world applications where the underlying function is expected to be continuous and differentiable. During the training phase, GPR estimates the parameters of this kernel function and calculates the covariance matrix between the input-output pairs. Using this covariance matrix and the training data, GPR then estimates the mean and variance of the distribution for each new input value through Bayesian inference.

3.2.3.7 Bayesian Ridge Regression

In Bayesian Ridge Regression (BRR), the hyperparameters of a classical ridge regression (i.e. a linear regression which implements a ridge regularization term) are associated with a priorly assumed probability distribution (also called hyperpriors), and tuned through training in a Bayesian inference approach (Box et al., 1992). Defining both a prior distribution $p(H)$ for the model parameters H and a likelihood function $p(E|H)$ for the observed data E, the BRR model computes the posterior distribution over functions $p(H|E)$ given the observed data through the use of Bayes' theorem $p(H|E) = \frac{p(E|H) \cdot p(H)}{p(E)}$, where $p(E) = \int p(E|H) \cdot p(H)\, dH$ is the marginal likelihood. Once the posterior distribution is obtained, the model is used to make predictions for unseen data points. These predictions come with uncertainty estimates, which are derived from the posterior distribution. BRR, as in the case of the SVM algorithm, employs an L2 regularization, hence it spreads the coefficient values more evenly, stabilizing the model and preventing overly large coefficient estimates.

3.2.4 Feature Extraction Computation

Each of the algorithms implemented in this study provides a method for calculating feature importance, albeit through different approaches. In principle, a single sensitivity method could be used to evaluate feature importance across all algorithms. However, some algorithms have built-in methods specifically designed to align with their inherent characteristics.

- Tree-based algorithms: for Random Forest, Decision Trees, and Extreme Gradient Boosting, feature importance is assessed using the Mean Decrease in Impurity (MDI) method. This method quantifies the contribution of each feature to the overall prediction accuracy by measuring how much each feature decreases the impurity of the splits in which it is involved, averaged over all the trees in the ensemble. In the scikit-learn library, these contributions are accessible through the feature_importances_ attribute. The MDI method is particularly effective because it directly measures the impact of each feature on the model's decision process, providing a clear indication of feature importance.

- Fitting Methods: LASSO and Linear Support Vector Machine derive feature importance from the model coefficients. In these linear models, the magnitude of the coefficients indicates the strength and direction of the relationship between each feature and the target variable. Specifically, in scikit-learn, this can be accessed through the best_estimator_.coef_ attribute. Larger absolute values of these coefficients indicate greater importance.

- Probabilistic methods: Gaussian Process Regression and Bayesian Ridge Regression do not have a built-in mechanism for directly assessing feature importance. Therefore, in this work, the Sobol method was used to infer feature importance. Once built and tested against the original model outputs, the GPR and BRR surrogate models can be used to perform a GSA in substitution of the original model. By using a surrogate model, the computational cost of running the original model for a large number of input combinations is avoided. Instead, the surrogate model can be used to generate a large number of input combinations with significantly less computational time and evaluate their impact on the output. Over these samples, the Sobol sensitivity indices can be feasibly computed following for instance the definition proposed by Saltelli (Saltelli, Ratto, et al., 2008). The user can then compare the Sobol indices evaluated with both GPR and BRR, providing information on their robustness and reliability. In the proposed tool, after the algorithm generates the optimal surrogate model, it uses the Python library SALib to compute the Sobol first-order index as a score for the sensitivity importance of each parameter.

Despite the differences in the feature importance calculation approaches of the different algorithms, each method is applied to standardized, non-dimensional data and each feature importance set is scaled between [0,1]. This ensures that feature importance scores are comparable across models. The primary objective of all these methods is to quantify the sensitivity of the model output to changes in the input features. Consequently, the feature importance scores obtained from these different methods provide a well-posed comparison of parameter sensitivities. By evaluating and comparing these scores, we gain a comprehensive understanding of the relative importance of each feature across different modeling approaches, which increases the robustness of the results.

3.2.5 Hyperparameter Tuning

In the proposed tool, the hyperparameters of each implemented algorithm are tuned based on a cross-validation score obtained through Bayesian optimization.

Bayesian optimization is an iterative process that seeks to explore the hyperparameter space while also exploiting regions of the space that are expected to yield good performance. At each iteration, the method proposes a new set of hyperparameters based on a probabilistic model of the function behavior and then evaluates the function at that point. The results of the evaluation are used to update the probabilistic model, which is then used to propose a new set of hyperparameters for the next iteration.

3.3 Case study

The proposed case study, adopted to highlight the functionalities of ML-AMPSIT, is based on idealized coupled WRF/Noah-MP simulations of a sea breeze circulation over a flat three-dimensional domain. The objective of this case study is to evaluate the impact of a prescribed set of Noah-MP parameters on the development of the thermally-driven wind.

3.3.1 The WRF/Noah-MP model

The Weather Research and Forecasting (WRF) model is a widely-used state-of-the-art mesoscale numerical weather prediction model for atmospheric research and operational forecasting applications which is supported by the NSF National Center for Atmospheric Research (NCAR), with more than 50000 registered users from more than 160 countries (Skamarock, Klemp, Dudhia, Gill, Liu, Berner, Wanh, et al., 2021). It offers a wide range of customization options consisting of dedicated modules and physics schemes to meet the state of the art in atmospheric science and adapt to a wide variety of scenarios. The dynamical core used for this case study is the Advanced Research WRF (ARW), which uses a third-order Runge-Kutta scheme for time integration with a time-split method for solving acoustic modes (Wicker et al., 2002) and an Arakawa-C grid staggering for spatial discretization. In the case study presented here to illustrate the functionalities of ML-AMPSIT, WRF is coupled with the Noah with Multi-Physics options (Noah-MP, **He** Niu, Yang, Mitchell, et al., 2011, Yang, Niu, et al., 2011) Land Surface Model (LSM). Noah-MP is one of the most used LSMs available in WRF to calculate surface-atmosphere exchanges and interactions. It is an augmented version of the Noah land surface model (Ek et al., 2003), that allows the usage of different physical schemes and multi-parametrization options, reaching a total number of 4,584 possible combinations (https://www.jsg.utexas.edu/noah-mp/).

3.3.2 Model setup

Simulations are performed using one domain with 201 x 201 cells in the horizontal plane with a grid spacing of 3 km. 65 vertical levels are used, transitioning from a vertical resolution of 7 m close to the surface up to 600 m at the top of the domain, which is placed at 16 km above sea level. The domain is completely flat and subdivided into two equally-sized rectangular sub-regions of land and water, with the interface line oriented along the west-east direction. The aim is to simulate the diurnal cycle of a sea/land breeze. Boundary conditions are set as open at all the boundaries.

The atmospheric initial potential temperature profile is set using the following expression, representative of a stable atmosphere:

$$\theta(z) = \theta_s + \Gamma z + \Delta\theta(1 - e^{-\beta z}), \tag{3.5}$$

with $\theta_s = 280$ K, $\Gamma = 3.2$ K Km^{-1}, $\Delta\theta = 5$ K, and $\beta = 0.002$ m^{-1}. The atmosphere is initially at rest and the relative humidity is set constant over all the domain and equal to 30%.

The idealized simulations start at 13:00 UTC, 19 March, are centered at 47°N, 11°E, and last for 35 hours, thus with a solar radiation cycle representative of the equinoxes at mid-latitudes. The first eleven simulation hours are not analyzed and considered as spin-up period. Therefore, analyses concentrate on a full diurnal cycle, from 00:00 UTC 20 March to 00:00 UTC 21 March. Model output is saved every hour.

The physical parameterization schemes selected for the present work are the Rapid Radiative Transfer Model (RRTM) for longwave radiation (Mlawer et al., 1997), the Dudhia scheme (Dudhia, 1989) for shortwave radiation and the YSU scheme (Hong et al., 1996) as planetary boundary layer (PBL) parameterization, coupled to the MM5 similarity scheme for the surface layer. Since this idealized study aims at reproducing a sea/land circulation, which best develops under completely clear sky, the microphysics parameterization is switched off, as well as the convective scheme, since convection is explicitly resolved at the resolution used.

As said above, the Noah-MP model is used to evaluate land-atmosphere exchange. In Noah-MP, the canopy radiative transfer scheme used is the "modified two-stream" (Niu and Yang, 2004), one of the most used, which aggregates cloudy leaves into evenly distributed tree crowns with gaps. The gaps are computed according to the specified vegetation fraction. The Ball-Berry scheme, the most common choice in literature, is used for the stomatal resistance computation, with the Noah-type soil moisture factor (Schaake et al., 1996a), while the default TOPMODEL with groundwater option is used for runoff and groundwater processes (Niu and

Yang, 2007). The surface resistance to evaporation and sublimation processes is set following Sakaguchi et al. (2009a). The surface-layer drag coefficient, used to compute heat and momentum exchange coefficients, is calculated with the traditional Monin-Obukhov-based parameterization. In this work, the dynamic vegetation option is not activated, with the prescription of a fixed vegetation fraction of 60% to consider a reasonably realistic percentage, while the monthly satellite-based climatological leaf area index is read from the MPTABLE.TBL file, which contains all the parameter values. Finally, crop and irrigation options are deactivated. It is important to underline that water physical properties are not varied in the sensitivity simulations, but changes in atmospheric variables are expected also over water, due to the indirect effects of the variations in the surface parameters over land.

In order to simplify this demonstrative case study, only six of the surface parameters defined in the look-up table MPTABLE.TBL, from which WRF reads and sets the surface values accordingly, are considered. In particular, the Noah-MP reference vegetation type adopted over land is grassland (vegtype=10 of the IGBP-Modified MODIS classification) and the parameters for which the sensitivity and relative importance is evaluated are the characteristic leaf dimension (DLEAF), the height of the vegetative canopy top (HVT), the momentum roughness length (Z0MVT), the near-infrared leaf reflectance (RHOL-NIR), the empirical canopy wind parameter (CWPVT), and the leaf area index referred to the month of March (hereafter LAI_MAR). The choice of these parameters is based on their importance in other sensitivity studies reported in the literature (Arsenault et al., 2018; Cuntz, Mai, Samaniego, et al., 2016; Mendoza et al., 2015). The final perturbed model parameter ensemble contains 100 samples, each with different parameter values based on the associated Sobol sequences. The input ensemble is generated by perturbing the parameters by up to 50% of their reference value in the look-up table MPTABLE.TBL. It should be clear that the results of a sensitivity analysis, regardless of the approach chosen, always depend on the range of exploration of the parameters, and that their transferability to arbitrary ranges of values is not guaranteed if the true sensitivity of the parameters in unexplored ranges is not known a priori. The perturbation percentage in this work has been chosen to avoid unphysical values, but it must be noted that the aim of the present work is to introduce and test the ML-AMPSIT functionalities in a simplified case study, while a more detailed analysis would require more attention to the choice of the parameter space.

The output variable for which sensitivity is evaluated is the south-north horizontal component of the wind v in the lowest 10 vertical levels at two different locations in the domain, one over land and one over water. For both locations, the average output from three adjacent cells in the south-north direction is analyzed, in order to increase the representativeness of the results. The central points of these two locations are x=100, y=95 and x=100, y=105 for water and land respectively, i.e., in the central cell in the west-east direction, and 5 grid points to the north and to the south of the land-water interface.

3.4 Results

3.4.1 Sea breeze ensemble

Before analyzing the ML-AMPSIT results, this section presents an overview of the output of the WRF/Noah-MP simulations, focusing on the horizontal south-north wind component v at the two locations selected for the application of ML-AMPSIT. Figure 3.3 shows the ensemble time series of v during the entire period analyzed at the first vertical level. The daily cycle of the sea breeze is evident at both locations, as the velocity changes sign according to the radiation pattern and the varying horizontal temperature and pressure gradient, i.e. v is negative (northerly, from land to sea) during the night and positive (southerly, from sea to land) during the day. The spread of the ensemble tends to be larger over water than over land, especially before sunrise, indicating that the variation of the input parameters has a larger effect on v over water. This highlights that local variations of surface parameters can significantly affect atmospheric conditions also in regions of the domain where no variations are prescribed, since temperature contrasts are affected, and, as a consequence, the development of thermally-driven circulations.

Figure 3.4 shows the ensembles of the vertical profiles of v in the lowest 200 m, containing the first 10 vertical levels, at three different times, 7:00 UTC, 13:00 UTC, and 19:00 UTC, which are representative respectively of the maximum intensity of the northerly land breeze, the morning transition between northerly and southerly wind, and the maximum intensity of the southerly sea breeze at the water point. It can be noted that the northerly land breeze (Fig. 3.4a,b) is shallower than the southerly sea breeze (Fig. 3.4e,f). Moreover, the comparison between Fig. 3.4e and Fig. 3.4f highlights the stronger effect of friction over land, with a more pronounced decrease of the wind speed close to the surface.

The ensemble variance is small near the ground over land and increases with elevation at 07:00 UTC and 19:00 UTC, while over water the spread is larger and more uniform along the entire vertical profile, especially at 13:00 UTC and 19:00 UTC. During the transition (Fig. 3.4c and Fig. 3.4d), the ensemble spread is very small over land, with all the simulations showing very small v values along the entire vertical profile, while a large spread is observed over water, suggesting that the variations in the surface parameters investigated significantly influence the timing of the transition from land to sea breeze over water, although preserving the shape of the vertical profile.

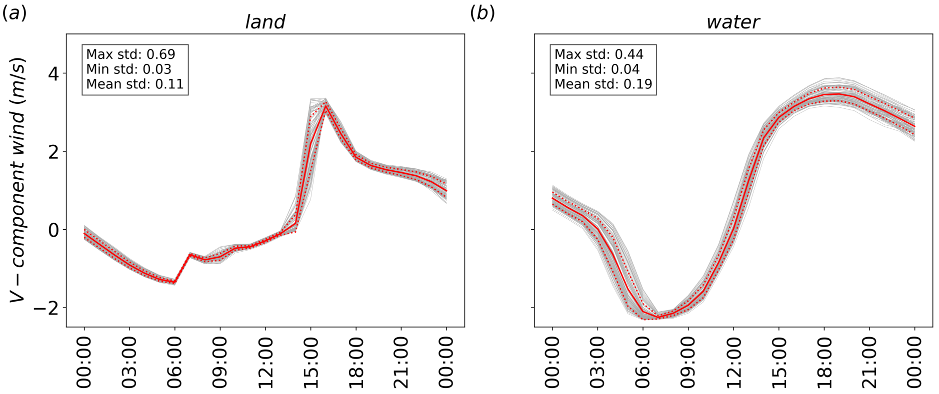


Figure 3.3: Time series of the south-north wind component at the first vertical level. The solid and dashed red lines represent the ensemble mean and standard deviation, respectively, while the gray lines represent the output of the single simulations. The values of the maximum and minimum hourly standard deviations (std) are also reported.

3.4.2 ML-AMPSIT output

After all the steps discussed in Section 2.1, the basic output that the ML-AMPSIT tool offers to the user is a composition of four plots similar to those reported in Figure 3.5, which refers to the results obtained for v at the third vertical level, over the water point (for this example, no averaging over multiple points was performed and all the simulation hours are shown, in order to present the default output of the tool) obtained with the LASSO algorithm. In particular, panel a) shows the importance time series related to each of the 6 selected parameters, panel b) shows the time evolution of the metrics, underlining how the correlation score and the errors eventually change over time, while panels c) and d) are specific to a single hour selected by the user, showing respectively the goodness of fit and the ranking of the importance of the features for that hour. In the following sections, the outputs from ML-AMPSIT will be aggregated to perform convergence analysis and methods comparison.

3.4.3 Convergence analysis

Figure 3.6 and 3.7 show the convergence of the MSE and of the feature importances computed by each method as a function of the number of simulations considered, i.e. the number of input-output relations used for training the algorithms. The analysis of the convergence is important because it indicates when the

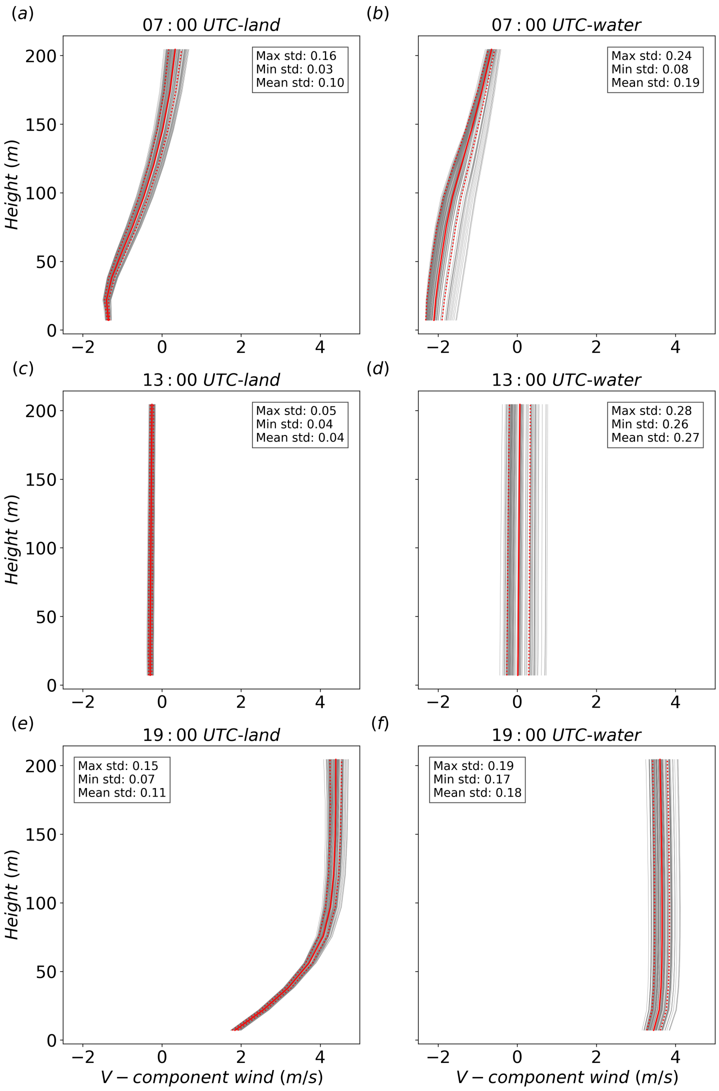


Figure 3.4: Vertical profiles of the south-north wind component in the lowest 200 m above the surface, at a) 07:00, b) 13:00, and c) 19:00 UTC. The solid and dashed red lines represent the ensemble mean and standard deviation, respectively, while the gray lines represent the output of the single simulations. The values of the maximum and minimum hourly standard deviations (std) are also reported.

regression tasks reach a stable state and additional simulations do not significantly alter the results. When convergence is reached, it can be assumed that the obtained feature importance provides a reliable representation of the underlying relationships in the system and that a sufficient number of simulations are performed to capture the essential characteristics of the system under investigation.

For the sake of brevity, only results over the land region at the first vertical level are shown here, considering results at 13:00 UTC for the metrics convergence and at 8 different times for the feature importance. However, the considerations reported below can be generalized to other points/times, since the methods maintain similar speed of convergence during the entire run in every analyzed point.

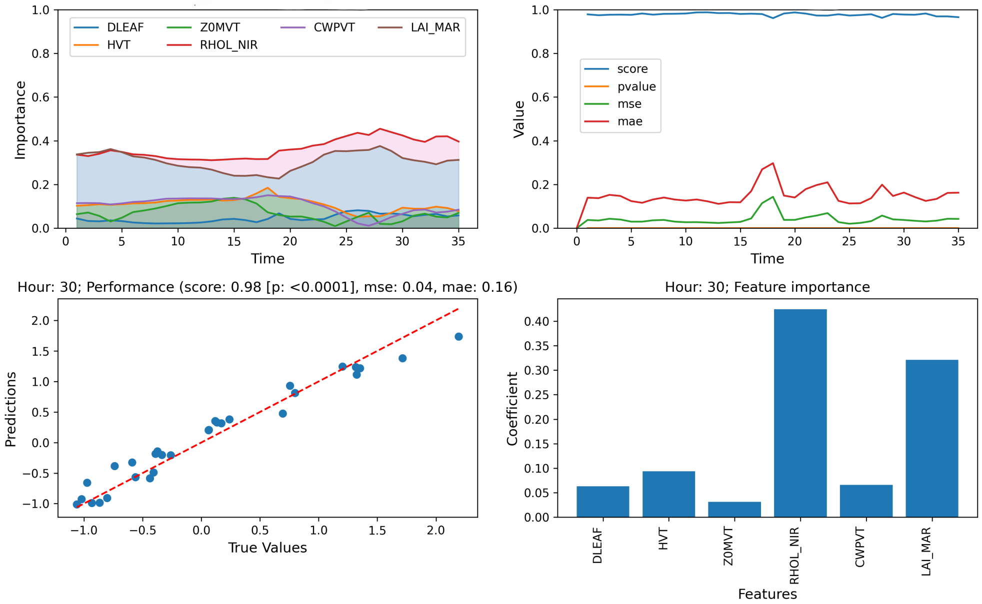


Figure 3.5: Example of the ML-AMPSIT output for a LASSO regression of 100 simulations with a duration of 36 hours (spin-up phase included), focusing on v at the third vertical level above the water point and displaying the metrics for the 30th simulated hour: a) importance time series of the 6 parameters; b) time evolution of the metrics MSE, MAE and p-value; c) quality of the test phase associated to the selected hour with the corresponding metrics; d) ranking of the importance of the features for the selected hour.

Despite each method showing some differences, especially in the oscillations around the convergence values, four out of seven of the proposed methods are able to reach a reasonably stable result after approximately 20 simulations. BRR, GPR, LASSO and SVM are the fastest and most stable algorithms to reach convergence. On the other hand, the decision-tree-based methods have significant oscillations even after 20 realizations and the metrics show that they are less consistent than the other methods. However, the results highlight that all the methods propose a stable and consistent ranking of the parameters' importance after 80 simulations and, in most cases, even with a much lower number of simulations.

3.4.4 Parameter importance analysis

Figures 3.8 and 3.9 show respectively the time series of the performance metrics for the south-north wind component at the lowest vertical level over the land and water regions, while Figures 3.10 and 3.11 show, for the same variable and regions, the time series of the feature importance for each of the proposed methods.

In this example, GPR, BRR, LASSO and SVM show the best metrics, suggesting that in the proposed case study there is no relevant difference between non-linearity-aware approaches and linear approaches, as they both correctly capture the relation between the tested parameters and the south-north wind component. These

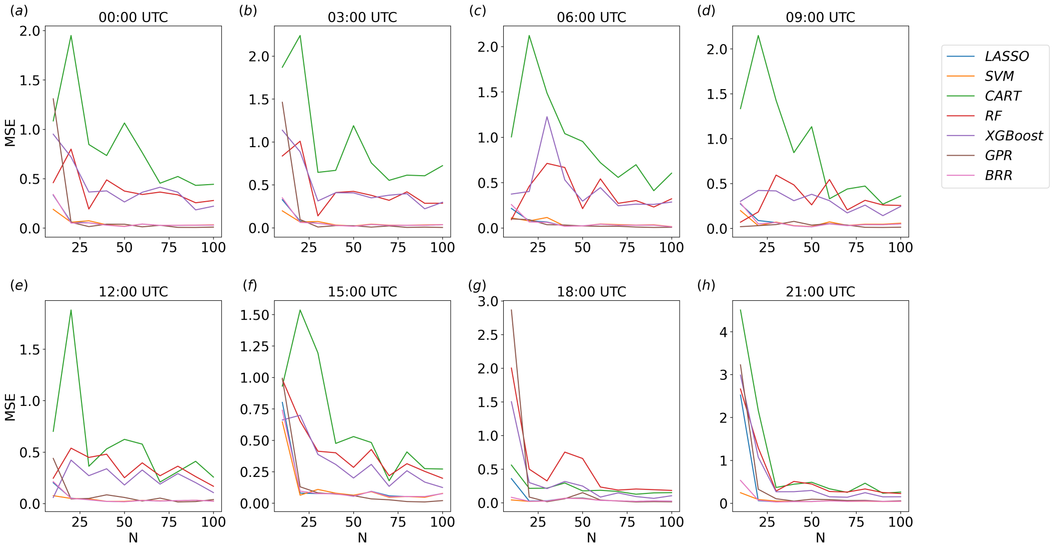


Figure 3.6: Convergence of the MSE with the number of realizations for each method implemented, considering v over land, at the first vertical level, at 8 different times.

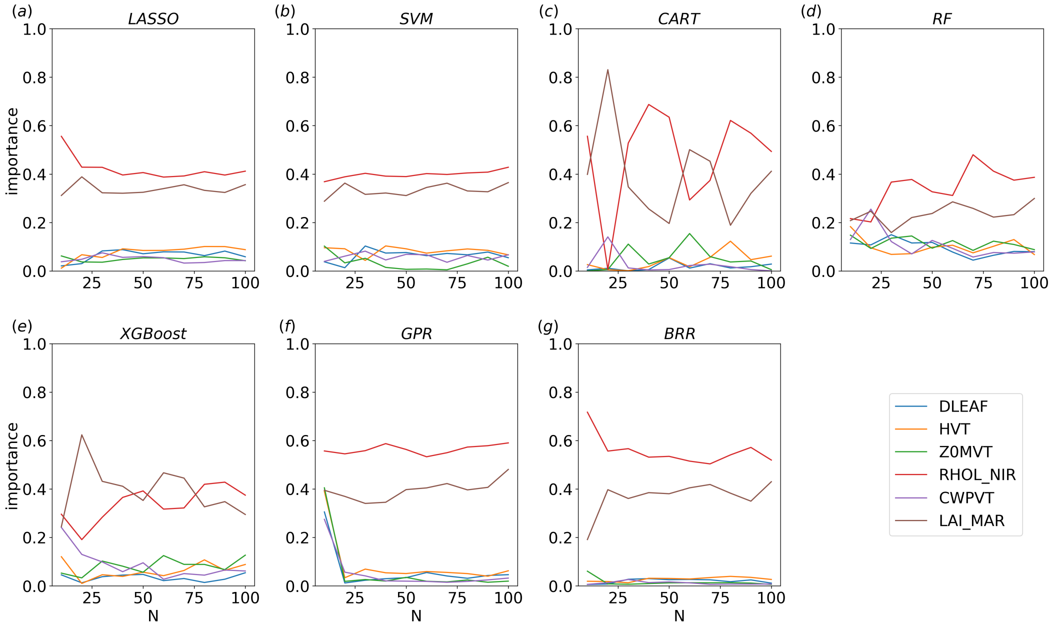


Figure 3.7: Convergence of the feature importance with the number of realizations, for each method implemented, considering v over land, at the first vertical level, at 13:00 UTC.

algorithms show very stable results, with slightly worse performance metrics only in correspondence with the peak of the sea breeze around 16:00 UTC over land and with the nocturnal peak of the northerly wind around 07:00 UTC over water. The three decision-tree-based methods present a more irregular behavior of the performance metrics, with higher errors and lower correlations. In particular, CART presents the worst performance metrics for this case study. The poorer metrics compared to more refined methods such as RF or XGboost are expected, since CART does not compute an ensemble of decision trees and does not consider the

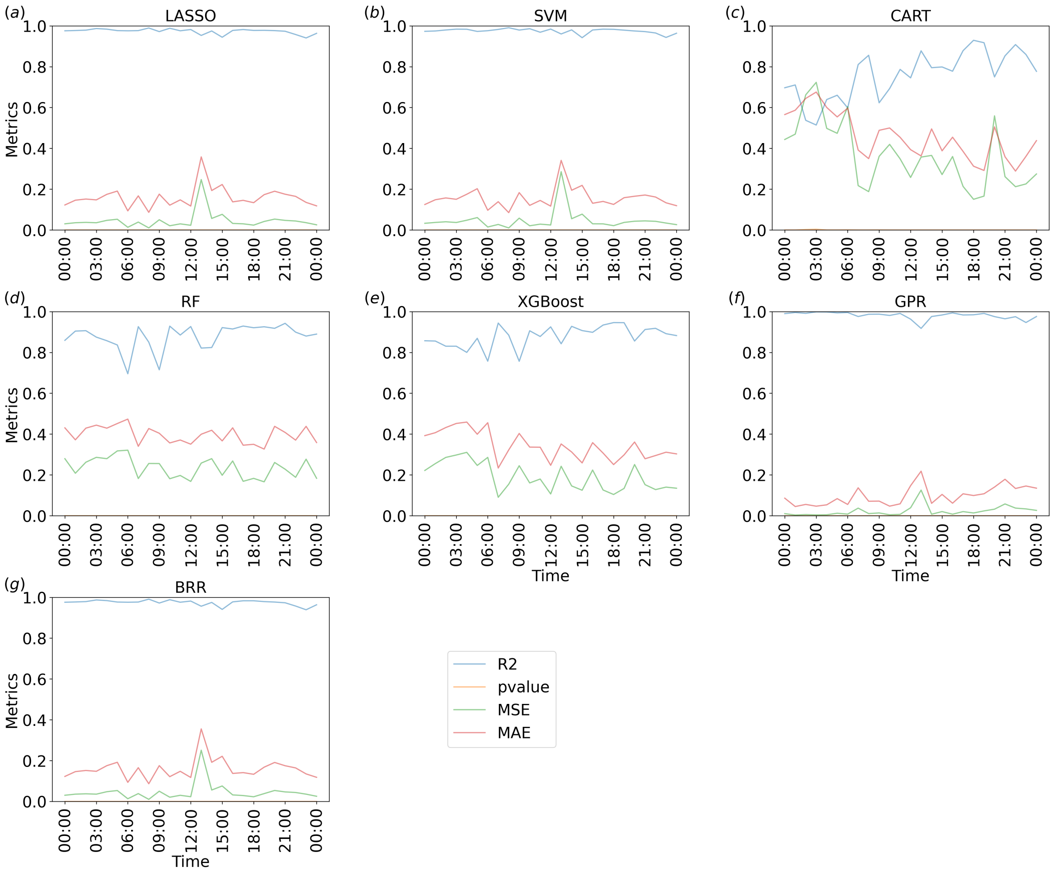


Figure 3.8: Time series of the performance metrics for each method implemented, considering v over land at the first vertical level: blue, green, red and orange lines represent respectivelyR^2, MSE, MAE and p-value.

errors of the previous branches.

For both regions, all methods agree very well on the ranking and overall magnitude ratios of the feature importance, individuating similar patterns, with only minor differences, also considering the methods showing worse performance metrics (cf. Fig. 3.8).

Figure 3.10 highlights cyclic trends of the parameters' importance over the land region, likely induced by the cycle of the diurnal thermally-driven circulation. In particular, Z0MVT and RHOL-NIR alternate as the most important parameters, with RHOL-NIR dominating during the entire day, with the exception of the hours close to sunrise and sunset, when ZOMVT becomes more important. This seems to indicate a stronger role of the surface friction when stronger winds are present (ZOMVT directly influences the surface friction), coherently with the wind speed profiles shown in Figure 3.4a,e. LAI_MAR shows an importance almost comparable to RHOL-NIR during the day, especially with LASSO, SVM and RF, while its importance is lower during the night, especially in the algorithms that tend to separate more the relevant and non-relevant parameters (i.e., GPR, BRR and decision-tree-based methods). The other parameters seem to be more relevant

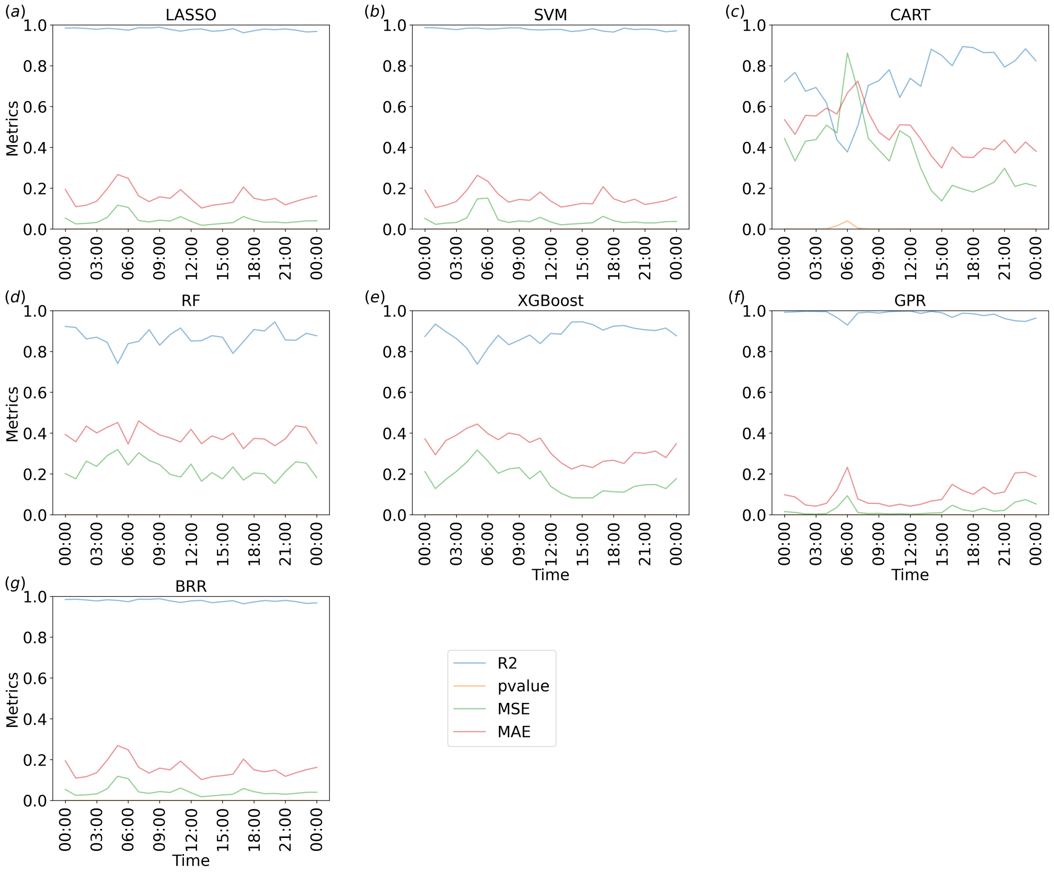


Figure 3.9: Time series of the performance metrics for each method implemented, considering v over water at the first vertical level: blue, green, red and orange lines represent respectivelyR^2, MSE, MAE and p-value.

at night, with the exception of DLEAF, which is always non-relevant for each method implemented.

The results are more uniform over water than over land, and the ranking of the parameters does not show significant variations during the whole day. In particular, the dominant parameters are RHOL-NIR and LAI_MAR, with Z0MVT always showing low importance values. Since the sea breeze is driven by thermal contrasts, it is expected that the parameters mainly affecting temperature, such as the reflectivity and the leaf area index, are also particularly relevant for this kind of circulation.

It is interesting to note that the decision-tree-based algorithms, CART, RF and XGboost, overall detect minor differences between the less relevant parameters, while the other methods, GPR, BRR, LASSO and SVM, enhance the differences and define a clearer ranking. The reason for these differences is reasonably due to the fact that, as mentioned in Section 2, the decision-tree-based algorithms are less strict about feature shrinkage compared to other methods containing a regularization term like LASSO, hence resulting in a less clear ranking in feature importance with respect to the other methods. However, the relative importance

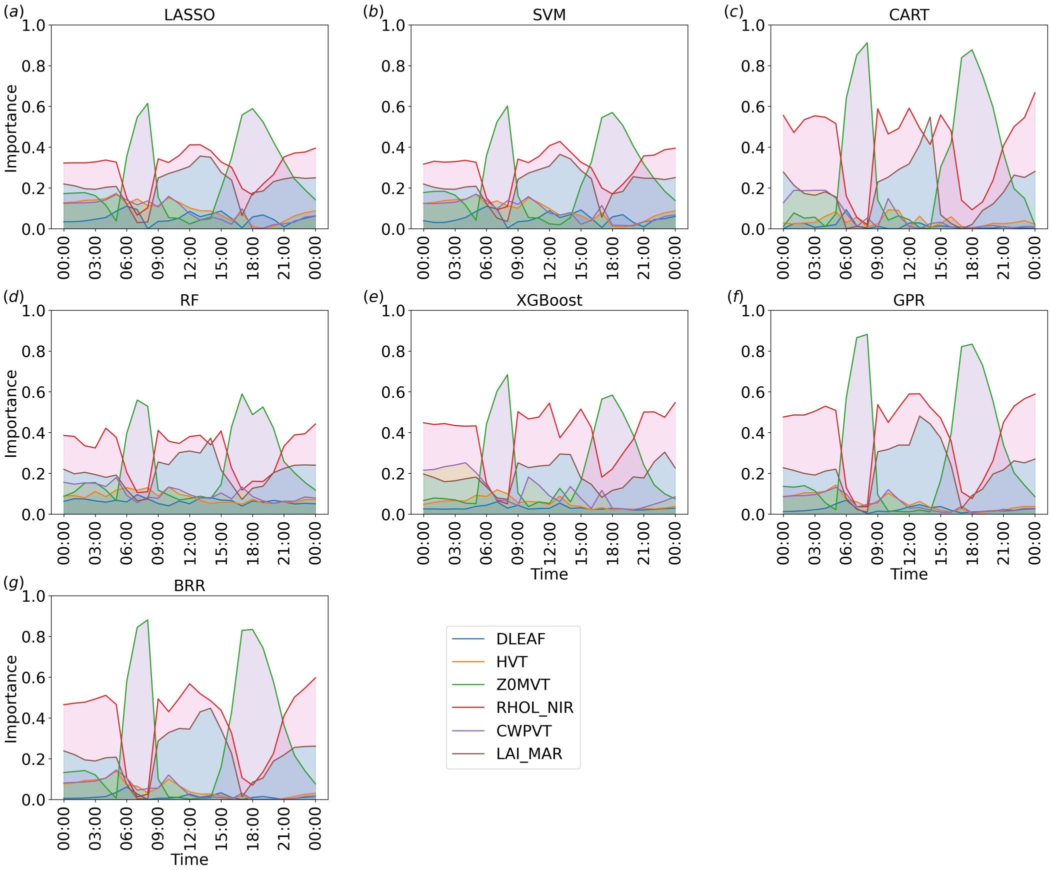


Figure 3.10: Time series of the importance of the parameters considered for each method implemented, considering v over land at the first vertical level.

between parameters is overall conserved, i.e. the feature importance ranking is mostly the same as the other methods for the entire length of the simulation.

It is also worth noting that, considering the importance time series obtained from GPR and BRR surrogate models, the surrogate Sobol first index agrees very well with the feature importance scores of the other algorithms, which indicates that the Sobol indices derived from BRR and GPR and the feature importance derived from the other methods have equivalent sensitivity estimation capability when convergence is properly achieved.

3.4.5 Vertical variability

Figure 3.12 and 3.13 show, for the land and water regions respectively, the variations in the feature importance in the lowest 10 vertical levels at different times. Since, as highlighted in the previous section, GPR is one of the methods presenting the best performance metrics, it has been chosen as the reference algorithm for this analysis. However, the results obtained with GPR are consistent with those obtained with all the other methods, particularly with LASSO, SVM and BRR.

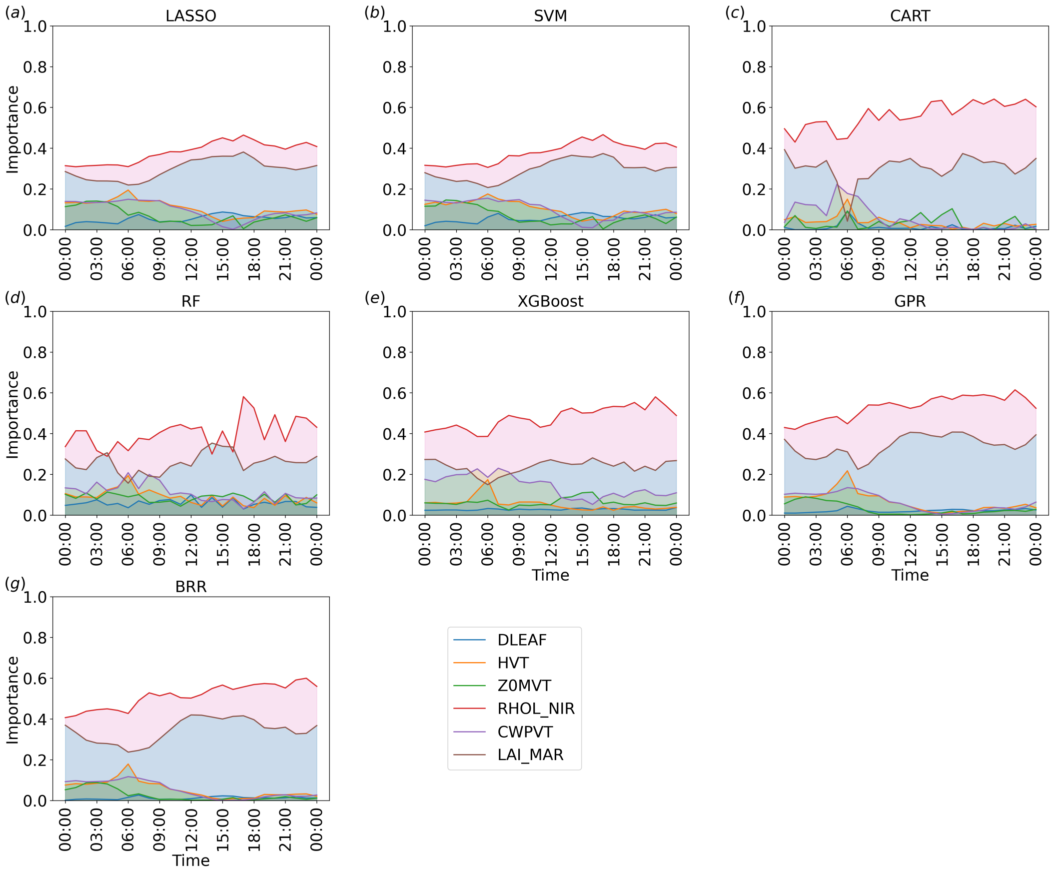


Figure 3.11: Time series of the importance of the parameters considered for each method implemented, considering v over water at the first vertical level.

Over the water region, the parameters' ranking does not show significant variations with height, except in correspondence with the nocturnal wind peak (06:00 UTC), when LAI_MAR becomes more important than RHO-NIR above the 8th model level. The situation is more complex over the land region, with more significant variations of the parameter importance with height. In particular, it can be seen that Z0MVT is more important close to the surface, especially when wind speed is stronger, consistent with the results shown in Fig. 3.10, and remarking that friction is particularly important close to the surface. The importance of LAI_MAR tends to increase with height, especially during the night, while the importance of RHOL-NIR slightly decreases with height with the northerly land breeze and strongly increases in the lowest vertical levels when the southerly sea breeze is well developed.

It is worth noting that the MSE for GPR, LASSO, BRR and SVM does not show significant variations in the lowest 10 vertical levels both over land and over water (Figure 3.14 and 3.15), meaning that the observed variations in feature importance are related to changes in the input-output relation rather than to uncertainty issues. A slightly higher variability is shown by RF and XGboost, while CART is the only method presenting a strong height dependence, in particular considering higher

MSE values close to the surface at night over land and in correspondence with the northerly land breeze peak over water. These observations strengthen the evidence that this method is not performing well in this case study.

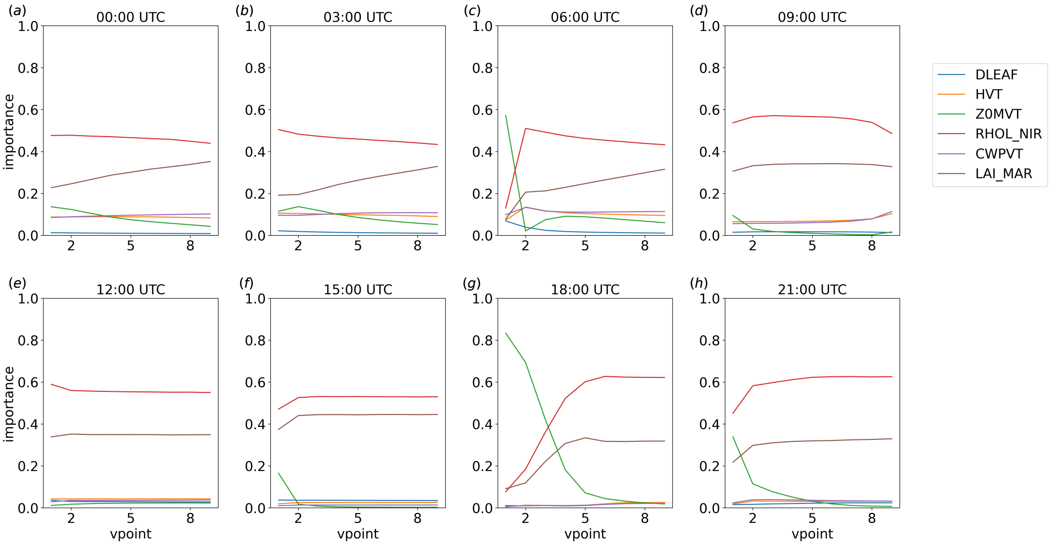


Figure 3.12: Parameter importance, considering v in the first 10 vertical levels over the land region at different times, for the GPR method.

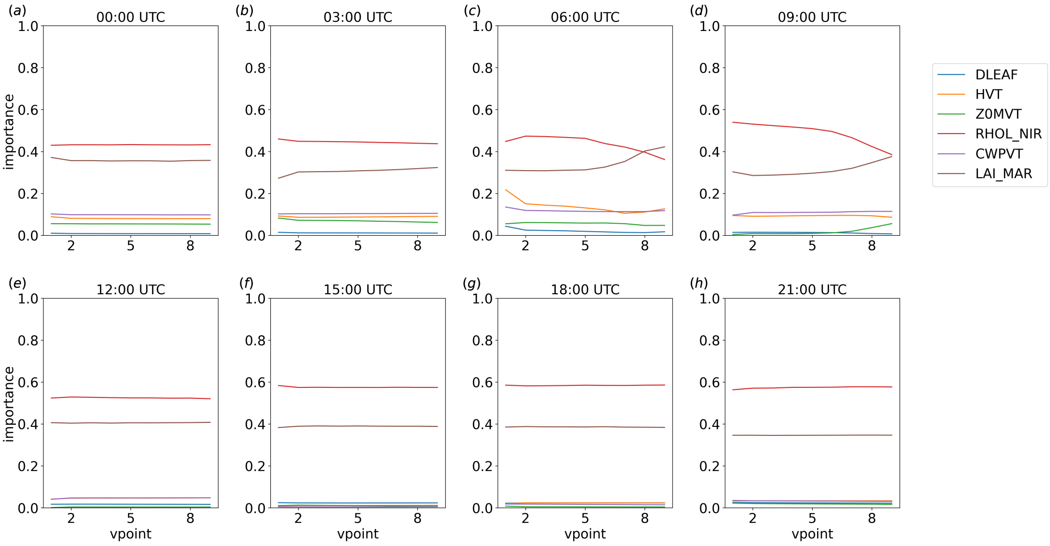


Figure 3.13: Parameter importance, considering v in the first 10 vertical levels over the water region at different times, for the GPR method.

3.5 Discussions and conclusions

This paper presented a novel automated model parameter sensitivity and importance analysis tool (ML-AMPSIT) that applies different machine learning algorithms, namely LASSO, Support Vector Machine, Classification and Decision

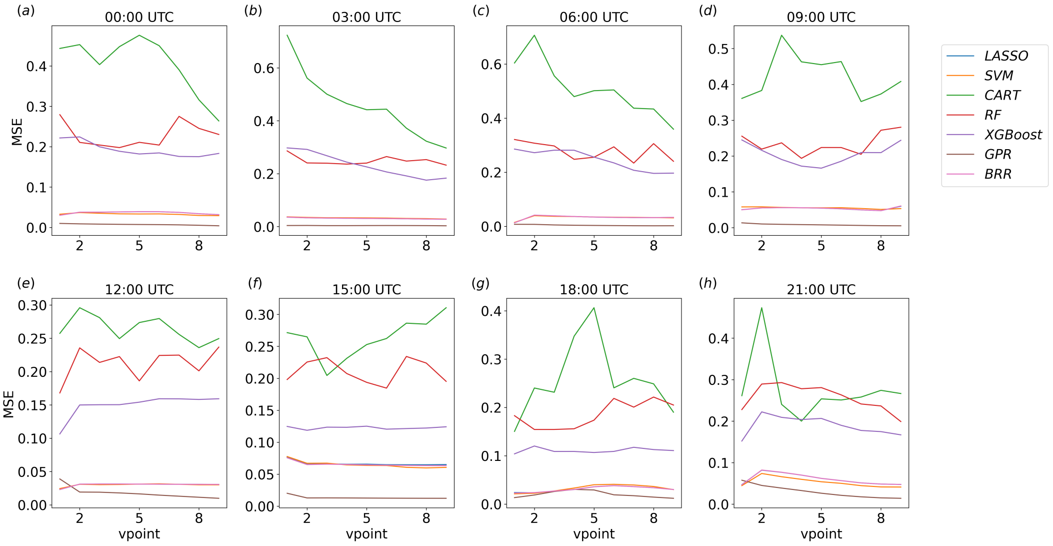


Figure 3.14: MSE for each method implemented, considering v in the first 10 vertical levels over the land region at different times.

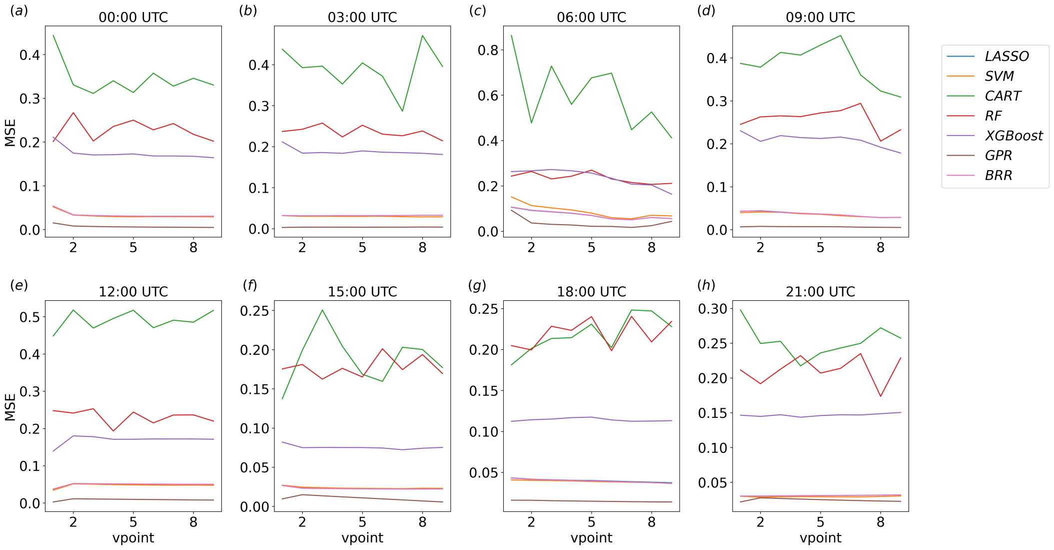


Figure 3.15: MSE for each method implemented, considering v in the first 10 vertical levels over the water region at different times.

Trees, Random Forest, Extreme Gradient Boosting, Gaussian Process Regression and Bayesian Ridge Regression, to perform sensitivity analysis and extract feature importance from input-output relationships. This tool was conceived to alleviate the computational burden usually associated with traditional global sensitivity analysis methods, which require a large number of model realizations, proposing an alternative approach using surrogate models or emulators. In fact, global sensitivity analysis methods, such as the Sobol method, demonstrate superior performance with respect to one-at-a-time approaches, which do not consider the interaction between parameters, but the large number of model realizations needed often makes their use unfeasible for complex numerical models. On the other hand, surrogate models or emulators, trained using input-output pairs of the original high-fidelity model, offer a cost-effective means of generating accurate predictions of the output variable. The utilization of machine learning techniques provides computationally efficient solutions while considering non-linearity and interactions between variables.

The implementation of different methods, also within the same family of algorithms, for example CART, Random Forest, and Extreme Gradient Boosting, which are all decision tree-based algorithms, or LASSO and SVM, which are both linear regression algorithms, or GPR and BRR, which are both probabilistic methods, is multifaceted. First, if a family of algorithms produces consistent results, this consistency increases the reliability and robustness of the outcome. Moreover, depending on the situation, it could be more convenient to rely on the fastest method instead of the most accurate, as in the case of CART with respect to XGboost. Second, the use of different families of algorithms extend the applicability and flexibility of the tool, as different families of algorithm can perform well in different scenarios.

The functionalities of the tool were tested and shown in a case study using the WRF meteorological model coupled with the Noah-MP land surface model. A sensitivity analysis applied to a set of Noah-MP parameters was presented for simulations of a sea breeze circulation over an idealized flat geometry. The different algorithms work as surrogate models of the original WRF/Noah-MP high-fidelity simulations and are able to accurately predict the original model behavior and reach robust conclusions about the parameter sensitivity given a relatively small ensemble of model runs. The efficiency of the model emulation is also tested through the computation of first-order Sobol indices from the training of Gaussian Processes Regression and Bayesian Ridge Regression, with results strongly consistent with the other proposed feature extraction methods. By integrating multiple algorithms into a flexible framework, ML-AMPSIT offers a comprehensive and reliable approach for sensitivity analysis in complex models, also allowing the evaluation of the uncertainty of the estimates by evaluating the spread between the outcomes of different algorithms.

Among different methods, Gaussian Process Regression, LASSO, Support Vector Machine, and Bayesian Ridge Regression emerged as the most reliable and robust methods. In contrast, decision-tree-based algorithms exhibited lower stability both in terms of convergence with respect to the realization number and higher uncertainty. In this case study, the linear models LASSO, Support Vector Machine, and Bayesian Ridge Regression demonstrated equal performance to the non-linearity-aware Gaussian Process Regression, suggesting the absence of strong non-linear relationships between the chosen parameters and the output variable in the analyzed domain regions.

For the best algorithms, the convergence of the feature importance was achieved with a small sample of about 20 simulations, whereas classical global sensitivity analysis approaches often require a much higher number of realizations. A qualitative comparison to evaluate the added value of ML-AMPSIT in terms of the number of simulations needed to reach robust results can be performed considering two of the most advanced methods in global sensitivity analysis, i.e. the Morris method (Morris, 1991), and the Sobol method (Saltelli and Sobol', 1995), assuming to use six parameters following a Latin Hypercube Sampling (Mckay et al., 1979) with radial design (Campolongo, Saltelli, et al., 2011). This sampling technique is one of the best trade-offs for decreasing the number of simulations needed compared to a full-factorial sampling (Saltelli, Ratto, et al., 2008). If parameter interactions are not relevant, such as for models with low complexity and low dimensionality, a viable strategy is to use the Morris method to find out the most and least relevant parameters. For Nc centroids produced with a Latin Hypercube sampling, and k perturbations produced by a radial design, the total number of sample points, i.e., the total number of model runs required, is $N = Nc(k+1)$. A sufficient number of centroids Nc for this sampling design ranges from 10 to 50 (Campolongo, Cariboni, et al., 2007), leading to 70-350 total simulations. However, even with this number of simulations, convergence is not guaranteed, as it depends on the specific case.

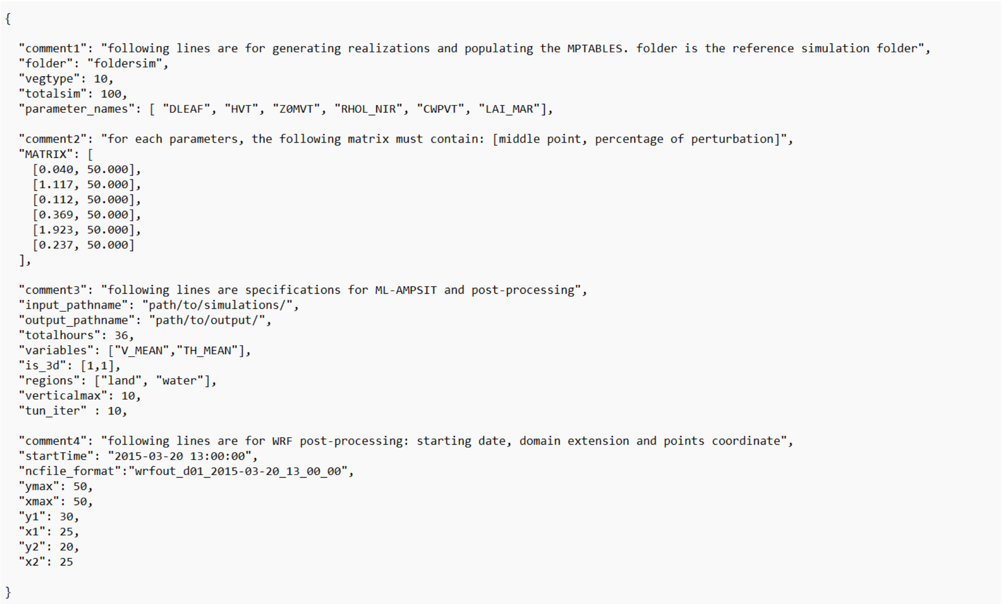
For more complex models, the Morris method can be very inefficient in stating the true parameter relevance (it is usually considered only a proxy of the true sensitivities, depending on the number of interactions and non-linearities in the model, Cuntz, Mai, Zink, et al., 2015). The Sobol method is able to weigh more accurately the interaction effects between each parameter, but it is more demanding. Following Saltelli, Annoni, et al. (2010), to circumvent some constraints over the number of model runs required, the final number would be $N_T = N_c(k+1)(k+2)$ which, using the previous assumptions for Nc, gives a minimum number of 1120-1680 runs. Aside from the minimum amount computed above, real applications of the Sobol methods can easily exceed this value to achieve robust results (Cuntz, Mai, Samaniego, et al., 2016; Cuntz, Mai, Zink, et al., 2015). However, this is not usually feasible for complex and computationally intensive models such as the WRF model.

It is then clear that ML-AMPSIT significantly reduces the number of simulations needed for sensitivity analysis and extraction of feature importance. Considering that all the proposed regression methods in ML-AMPSIT intrinsically account for interactions between parameters, this highlights its added value over classical global sensitivity analysis methods and points out its possible applications, especially in cases when the use of classical global sensitivity analysis methods is not feasible. Furthermore, the intercomparison of the results from different algorithms in ML-AMPSIT can reveal useful physical insights into model simulations.

It should be noted that the results presented in this paper are limited to the simple case study considered here to test the tool functionalities. Further investigations with a more complex scenario, where non-linearities are expected to be more strongly involved in the input-output relations, are in progress and will be presented in a future work.

Appendix

3.5.1 ML-AMPSIT configuration file



```
{

  "comment1": "following lines are for generating realizations and populating the MPTABLES. folder is the reference simulation folder",
  "folder": "foldersim",
  "vegtype": 10,
  "totalsim": 100,
  "parameter_names": [ "DLEAF", "HVT", "Z0MVT", "RHOL_NIR", "CWPVT", "LAI_MAR"],

  "comment2": "for each parameters, the following matrix must contain: [middle point, percentage of perturbation]",
  "MATRIX": [
    [0.040, 50.000],
    [1.117, 50.000],
    [0.112, 50.000],
    [0.369, 50.000],
    [1.923, 50.000],
    [0.237, 50.000]
  ],

  "comment3": "following lines are specifications for ML-AMPSIT and post-processing",
  "input_pathname": "path/to/simulations/",
  "output_pathname": "path/to/output/",
  "totalhours": 36,
  "variables": ["V_MEAN","TH_MEAN"],
  "is_3d": [1,1],
  "regions": ["land", "water"],
  "verticalmax": 10,
  "tun_iter" : 10,

  "comment4": "following lines are for WRF post-processing: starting date, domain extension and points coordinate",
  "startTime": "2015-03-20 13:00:00",
  "ncfile_format":"wrfout_d01_2015-03-20_13_00_00",
  "ymax": 50,
  "xmax": 50,
  "y1": 30,
  "x1": 25,
  "y2": 20,
  "x2": 25

}
```

Figure 3.16: An example of the configuration file for the WRF/Noah-MP model case study.

Chapter 4

Machine learning-based sensitivity analysis of surface parameters in numerical simulations of thermally-driven circulations in a mountain valley

---

This chapter has been submitted to XXX in 2024 and authored by: Dario Di Santo[1], Cenlin He[2], Fei Chen[3], Andrea Zonato[1] , and Lorenzo Giovannini[1].

Abstract

This study investigates the sensitivity of Land Surface Model (LSM) surface parameters in simulating atmospheric processes over complex terrain, with a particular focus on thermally driven circulations in an idealized valley. Using the Weather Research and Forecasting (WRF) model coupled with the Noah-MP LSM, we perform a comprehensive sensitivity analysis to evaluate the impact of empirical parameters on model accuracy and performance. A unique aspect of our methodology is the application of an ensemble of machine learning algorithms, integrated into the ML-AMPSIT tool, to emulate and evaluate the input-output relationships of the high-fidelity WRF outputs. Our results highlight the significant role of canopy height (HVT), stomatal resistance (RS), and near-infrared leaf reflectivity (RHOLNIR) in influencing the model's output surface fluxes, potential temperature, and valley wind circulations. The study highlights the importance of accurate parameterization in LSMs, especially in complex terrain, to improve the reliability of numerical weather prediction models. The integration of variance-based global sensitivity analysis with machine learning provides a novel approach to

---

[1] Department of Civil, Environmental and Mechanical Engineering, University of Trento, Trento, Italy

[2] NSF National Center for Atmospheric Research (NCAR), Boulder, CO, USA

[3] Division of Environment and Sustainability, Hong Kong University of Science and Technology, Hong Kong, China

understanding and improving model sensitivities, thus providing valuable insights into the dynamic interactions between land surface parameters and atmospheric conditions.

4.1 Introduction

In numerical weather prediction (NWP) models energy and mass exchange at the ground interface are evaluated by means of land surface models (LSMs), which, depending on their complexity, consist of a number of different interacting parameterizations to account for the relevant physical processes and to provide the atmospheric model with bottom boundary conditions. It is well known that LSMs represent one of the major sources of uncertainty in simulations with NWP models (e.g., Massey et al., 2014; Schmidli, Billings, et al., 2011; Tomasi et al., 2017), due not only to the approximations introduced by the physical parameterizations, but also to the use of a set of parameters whose value is often estimated empirically. A delicate issue regarding the use of specific parameter values stems from the representativeness of the land data. In fact, land cover parameters used in LSMs are usually defined on the basis of predefined look-up tables, which take into account representative or typical characteristics for each land cover type. However, in the real world, the actual characteristics of a specific land cover may significantly deviate from its typical features, depending for example on climatic or geographic aspects. This issue tends to be even more relevant over complex terrain, where land cover characteristics usually present larger temporal and spatial heterogeneity. This affects the estimation of the correct values for the parameters and, consequently, the quality of the simulation. The development of high-resolution and customized maps of such parameters can be extremely complex and sometimes not feasible, due to the lack of relevant information. Therefore, it is important to evaluate the sensitivity of the parameterizations to the uncertainty in the values of the empirical parameters, in order to concentrate the effort on the parameters that most affect model results.

One of the most popular and used LSMs is Noah-MP (Niu, Yang, Mitchell, et al., 2011; Yang, Niu, et al., 2011), which is available in the Weather Research and Forecasting (WRF) meteorological model. It is an augmented version of the Noah LSM (Ek et al., 2003), that allows the usage of different physical schemes and multiparametrization options reaching a total number of 4'584 possible combinations. It is an open-source model, with different contributions from the scientific community. Noah-MP uses more than 200 parameters related to land cover and soil properties, and it can employ different predefined sets of vegetation classes with particular choices of parameter values representative of different types of vegetation cover. The choice of these parameters is fixed to an empirical value, defined in the dedicated file MPTABLE.TBL, and, although they are used to define large land cover classes, no range of uncertainty is taken into account. This contrasts with a more realistic scenario in which the parameter values are not constant over the same vegetation category, hence this approach leads to potential issues in the model performance.

The importance of each parameter to the model output can be assessed through a sensitivity analysis. Analyzing the impact of small input perturbations on the model output has historically been the first type of approach for carrying out a "local sensitivity analysis" (Iooss et al., 2014), i.e. referred to local variations of the input parameters. The main limitation of these methods is the assumption of a linear relation between the parameter and the output variable. Since many simulated processes are unavoidably non-linear, the situation is usually more complex. Moreover, in a typical model, a parameter does not only affect the output, but also the behavior of the other parameter-output relations. This can be defined as the "interaction problem". To overcome this problem, more refined and global-oriented methods (hence the name "global sensitivity analysis") were introduced, among which is the variance-based approach, which decomposes the variance of the model output into multiple terms accounting for all the parameter interactions (Saltelli, Annoni, et al., 2010). The most famous and well-established variance-based method is the Sobol method (Saltelli and Sobol', 1995), which leads to multiple sensitivity indices evaluating an arbitrary number of interactions in the variance decomposition. The Sobol method has been widely applied to hydrological and environmental models (Gan et al., 2014; Gong et al., 2015; Griensven et al., 2006; Pyo et al., 2021; Rosolem et al., 2012; Wang and Solomatine, 2019). Rosero et al. (2010) used this method to assess the performance of the Noah LSM by comparing three different versions, differing in the inclusion of a groundwater module and a dynamic phenology module. Rosero et al. (2010) showed that the interaction between the model parameters is significant, increasing with the complexity of the model version, but also that the variance of the model response is mainly controlled by a few important parameters, among which is the leaf area index (LAI). This study showed that the dynamic vegetation module caused the highest increase in the interactions between the model parameters, but without a significant improvement in the model performance. In more recent years, many extensive studies have evaluated the performance of Noah-MP using global sensitivity analysis. Cuntz, Mai, Samaniego, et al. (2016) highlighted that many hard-coded parameters, i.e. parameters whose value is fixed inside the code and cannot be changed by the user, are very influential considering the sensitivity of the simulated surface runoff. Evaluating the behavior of Noah-MP over the central Tibetan Plateau, Li, Zhang, Barlage, et al. (2018) found that energy and water vapor fluxes are dominated by soil parameters. Moreover, the canopy height (HVT) was the most sensitive vegetation parameter, while the Clapp-Hornberger b parameter was the most sensitive soil parameter. With a similar approach, Arsenault et al. (2018) identified five soil parameters and six vegetation parameters acting as primary controls on soil moisture, sensible heat, latent heat, and net ecosystem exchange. Some of the most relevant parameters were the Clapp-Hornberger b parameter, the field water capacity, the soil porosity, the height of the canopy, the roughness length and the soil water availability.

The aim of the present study is to identify, through a sensitivity analysis, which LSM vegetation parameters need to be handled with the most care while simulating atmospheric processes over complex terrain, with a particular focus on the development of thermally-driven circulations and the related thermodynamic fields in an idealized valley. In such a context, it is expected that variations in relevant LSM parameters translate into different changes in the model output depending on the position in the valley and the time of the day. In fact, a significant aspect distinguishing complex terrain from simple geometry areas concerns the development of local thermally-driven circulations. In particular, mountain valleys are characterized by the development of thermally-driven circulations resulting from the differential heating and cooling of the lower atmospheric layers during fair weather days. Slope winds respond rapidly to the heating/cooling of the valley slopes, blowing upslope after sunrise and downslope after sunset. On the other hand, valley winds, which develop due to pressure differences between different valley sections or between the valley and the plain, typically blow down-valley during the night and up-valley during the day (Zardi et al., 2013). The alternation of these circulations typically marks the diurnal wind cycle in mountain valleys, strongly affecting the climatology of these regions.

The sensitivity of these circulations to the geometric characteristics of the valley has been widely studied through Reynolds-averaged Navier-Stokes (RANS) simulations and large-eddy simulations (LES), mainly using idealized valley geometries (e.g., Rampanelli et al., 2004, Serafin et al., 2010, Serafin et al., 2011). For example, Wagner et al., 2014 and Wagner et al., 2015 performed different LES with the Weather Research and Forecasting (WRF) model, highlighting variations in wind intensities depending on the valley floor slope, the valley volume, depth, width, and length. An idealized valley geometry was also used by Schmidli, Billings, et al. (2011) to compare the up-valley wind system simulated by eight different non-hydrostatic mesoscale models with the same initial and boundary conditions, finding that the main differences between the models were related to the representation of the surface energy budget and the different turbulence schemes.

Since slope and valley circulations develop as a consequence of temperature contrasts in the atmospheric boundary layer between different areas in complex terrain, it is clear that they are strongly affected by surface-atmosphere exchanges and local surface characteristics. Therefore, accurately representing these processes in NWP models is crucial for their correct simulation. However, a comprehensive evaluation of the sensitivity of their simulation to surface parameters is missing in the literature.

The studies presented above, which used sensitivity analysis to evaluate the impact of model parameters on Noah-MP results, were performed with the model executed in offline mode, i.e. forcing the LSM with observations without a coupled meteorological model, in order to evaluate only the LSM behavior. The low computational cost of these offline simulations allowed these studies to test the model output over long time periods, in order to obtain robust statistics. Since the aim of the present study is to assess the influence of surface parameters on the simulation of thermally-driven circulations in a mountain valley, it is fundamental to evaluate the interaction between the LSM and the coupled NWP model. For this reason, the analyses presented in this work are based on simulations in which the Noah-MP LSM is online coupled with the meteorological WRF model. Due to the high computational cost of the WRF simulations, this paper proposes a different approach based on surrogate models (Fernández-Godino et al., 2017; Forrester et al., 2008; Kim et al., 2020; Vinuesa et al., 2022), specifically an ensemble of machine learning (ML) algorithms that are trained over samples of the original "high-fidelty" model outputs, and emulate the input-output relations through regression tasks. These algorithms are implemented inside an automated software, the ML-AMPSIT tool, written specifically to perform ML-based sensitivity studies and presented in (Di Santo et al., 2024). ML extraction of feature importance is widely used in other Earth science fields, as stated in 3.1. Nonetheless, to the Author's knowledge, the present study is the first one that has been conducted applying ML algorithms to extract parameter importance from an LSM coupled with an NWP model in a complex terrain scenario.

The paper is organized as follows: Section 2 presents the modeling setup, detailing the specifics of the idealized valley geometry, the computational domain, the initial and boundary conditions, and the various options used in the Noah-MP land surface model. It also describes the methodology for the sensitivity analysis, including the machine learning algorithms used, the parameters tested, and the statistical methods used to evaluate parameters' importance. Section 3 reports the analysis of the simulations, discussing the variability of the simulation ensemble and the results of the sensitivity analysis, highlighting the most influential parameters affecting the valley circulation and surface fluxes. Finally, Section 4 concludes the paper with a summary of the main findings.

4.2 Methods

4.2.1 Modelling setup

The sensitivity analysis has been performed by means of simulations over an idealized valley geometry on a clear-sky day, eliminating complexities induced by real terrain and peculiar meteorological conditions, in order to facilitate the interpretation and generalization of the results.

The idealized valley-plain topography used in the present work is based on Schmidli and Rotunno, 2010 andSchmidli, Billings, et al., 2011. Following the criteria established by Rampanelli et al., 2004, the valley floor and the plain share the same elevation, avoiding contributions of up-valley anabatic winds. The valley-plain length is chosen long enough to avoid boundary effects on the along-valley flow and the valley slopes are not overly steep so that numerical inaccuracies do not arise from the terrain-following grid system. The valley geometry, with the axis oriented in the North-South direction, is defined by the following analytical expression (Fig. 4.1):

$$z = h(x, y) = h_p h_x(x) h_y(y), \tag{4.1}$$

where $h_x(x)$ and $h_y(y)$ are defined as follows:

$$h_x(x) = \begin{cases} 0, & \text{for } \ |x| \leq X_1 \\ \frac{1}{2} - \frac{1}{2} cos(\pi \frac{|x|-X_1}{S_x}), & \text{for } X_1 \leq |x| \leq X_2 \\ 1, & \text{for } X_2 < |x| < X_3 \\ \frac{1}{2} + \frac{1}{2} cos(\pi \frac{|x|-X_3}{S_x}), & \text{for } X_3 < |x| < X_4 \\ 0, & \text{for } \ |x| \geq X_4 \end{cases} \tag{4.2}$$

$$h_y(y) = \begin{cases} 1, & \text{for } \ |y| \leq L_y \\ \frac{1}{2} + \frac{1}{2} cos(\pi \frac{|y|-L_y}{S_y}), & \text{for } L_y < |y| < L_y + S_y \\ 0, & \text{for } \ |y| \geq L_y + S_y \end{cases} \tag{4.3}$$

where $h_p = 1.5$ km is the valley depth, $L_y = 200$ km half the valley length, $S_x = S_y = 9$ km the width of the slopes, $X_1 = 0.5$ km, $X_2 = 9.5$ km, $X_3 = 10.5$ km, $X_4 = 19.5$ km, and $X_3 - X_2 = 1$ km is the ridge width in the cross-valley direction.

The computational domain extends 300 km both in the North-South and West-East directions, while the domain top is placed at 16 km height. The choice of the horizontal extension of the domain was performed after a series of preliminary simulations, aiming at evaluating an optimal extension that prevents the influence of boundary effects on the simulation of the valley circulation while still keeping a

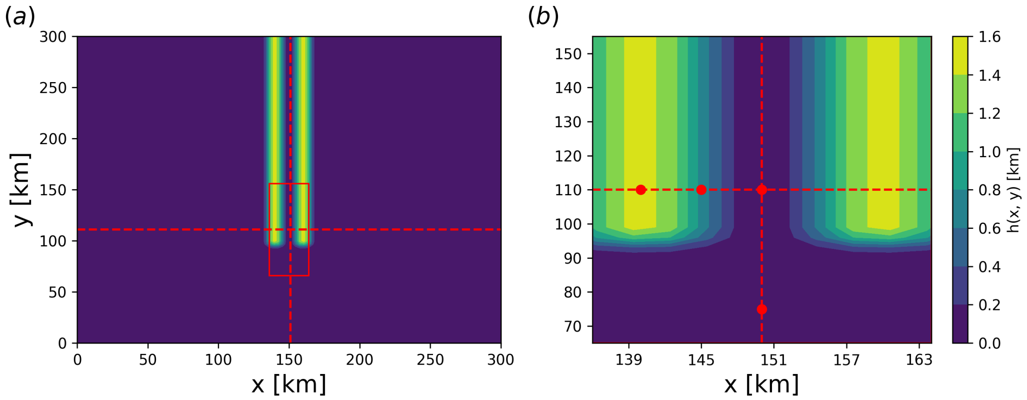


Figure 4.1: (a) Model topography, where dashed lines represent the position of the cross-valley and along-valley sections presented in section 4.3.1, respectively in Figs. 3-6 and 7-8. Their extension is limited by the red rectangle, which also represents the zoom-in area shown in (b), where the four points chosen to study the different valley regions are represented by red dots.

reasonable computational effort. The horizontal grid spacing is 1 km, representative of operational forecasts with local area meteorological models, and a total of 65 vertical eta levels are used, with 20 levels below 1 km height. The center of the domain is set at 47°N 11°E; the simulations start at 12:00 UTC (LST+1), 19 March, and run for 30 hours, with the first six hours used as spin-up period to allow the development of the thermally-driven circulation. Therefore, a full diurnal cycle has been analyzed, in order to evaluate the different phases of the the valley circulation, with a radiation cycle representative of the equinox in an Alpine valley.

The initial potential temperature vertical profile is divided into three layers, with a shallow, 500-m-deep mixed layer close to the surface topped by two stable layers, representative respectively of the troposphere and the lower stratosphere:

$$\theta(z) = \begin{cases} 280 & 0 \leq z < 500 \text{ m} \\ 280 + 0.0033 \cdot (z - 500) & 500 \leq z < 10000 \text{ m} \\ 280 + 0.0033 \cdot 9500 + 0.0098 \cdot (z - 10000) & z \geq 10000 \text{ m} \end{cases} \tag{4.4}$$

where the potential temperature is evaluated in K and the height z in m.

The initial relative humidity is constant and equal to 30% along the entire vertical profile and the atmosphere is initially at rest.

The Advanced Research WRF (ARW) dynamical core is used, which implements a third-order Runge-Kutta scheme for time integration with a time-split method to manage the acoustic modes (Wicker et al., 2002) and an Arakawa-C grid staggering for spatial discretization. In this work, 5th-order-accurate horizontal advection and 3rd-order-accurate vertical advection are used. The integration time step is 60 s. The W-E boundary conditions are set as periodic, while the N-S boundary conditions are symmetric, thus with the effect of simulating a valley that is twice as long and open on both sides (cf. Wagner et al. (2014)). Moreover, a Rayleigh sponge layer with a nondimensional damping coefficient of 0.1 has been set at the upper boundary. Coriolis force has been neglected.

As to the physics options, the Rapid Radiative Transfer Model (RRTMG) is used for shortwave and longwave radiation (Iacono et al. (2008)). After the preliminary evaluation of several planetary boundary layer parameterizations, the quasi-normal scale elimination scheme (Sukoriansky et al., 2005) resulted in producing the least amount of spurious rolls, which are typical of simulations with mesoscale meteorological models at the kilometer scale (Ching et al., 2014), hence it has been selected for this work, coupled with the corresponding surface layer scheme. The microphysics parameterization is not activated to simulate a full diurnal cycle not affected by clouds, in order to allow the full development of the valley wind circulation and simplify the interpretation of the results. Similarly, the cumulus parameterization is also switched off. The effects of the topography on shading and radiation are both taken into account.

The following options of the Noah-MP LSM have been used: the dynamic vegetation option is not activated with the prescription of a vegetation fraction, which in this work is fixed to 0.6, while the leaf area index is read from the dedicated Noah-MP table (MPTABLE), and changes in the sensitivity analysis. The radiative transfer scheme is the "modified two-stream" (Niu and Yang, 2004), which aggregates cloudy leaves into evenly distributed tree crowns with gaps. The gaps are computed according to the specified vegetation fraction and the sum of leaf and stem area indices, which in this work change in the sensitivity analysis. The Jarvis scheme for the computation of the stomatal resistance (Jarvis et al., 1976) with Noah-type soil moisture factor (Schaake et al., 1996a) and the TOPMODEL with groundwater option for runoff and groundwater processes (Niu, Yang, Dickinson, Gulden, and Su, 2007) are used. The surface resistance to evaporation and sublimation processes has been set following Sakaguchi et al., 2009b, while the surface layer drag coefficient parameterization, used to compute heat and momentum exchange coefficients, is the default Monin-Obukhov-based scheme. Finally, the soil texture definition is chosen from the MPTABLE, thus deactivating the usage of pedotransfer functions and the crop model has not been activated, as well as irrigation and harvesting options.

The sensitivity analysis presented in this work is based on 200 simulations perturbing the values of several parameters defined in the MPTABLE, as defined in Section 2.3. In particular, the class "mixed forest" of the MODIS IGBP classification has been chosen as the reference land cover to define the unperturbed values of all the tested parameters. The analysis is performed over 5 WRF variables, the west-east (cross-valley) wind component, the south-north (along-valley) wind component, the potential temperature, the sensible heat flux, and the latent heat flux.

4.2.2 ML-AMPSIT

ML-AMPSIT is a fully automated sensitivity analysis tool that uses a set of ML-based regression algorithms to assess the importance of parameters for a given model. The algorithms currently included are LASSO, Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGboost), Gaussian Process Regression (GPR), and Bayesian Ridge Regression (BRR). In the ML-AMPSIT tool, the parameters to test, their range of variability, and the number of simulations to perform with the high-fidelity model (in this case WRF/Noah-MP) must be first specified. A Sobol sequence (Saltelli, Annoni, et al., 2010) of the same length as the number of the simulations is then produced for each parameter. The sensitivity analysis is performed at different points in the computational domain, as specified by the user, using the different methods cited above. ML-AMPSIT calculates a series of performance metrics, namely the coefficient of determination (R^2), the mean squared error (MSE), and the mean absolute error (MAE), to evaluate the quality of the results. Moreover, the tool allows for the investigation of the convergence of the obtained feature importance and metrics. By integrating diverse algorithms, ML-AMPSIT offers a comprehensive toolkit for sensitivity analysis. The ensemble approach ensures that the tool can handle the problem under analysis from different perspectives balancing the outcomes of the different methods and offering robust results. A complete presentation of ML-AMPSIT can be found in Di Santo et al., 2024, while the open-source code is provided in Di Santo, 2024.

Among all the algorithms available in ML-AMPSIT, a preliminary analysis (not shown) of their individual performance for the present case study led to the choice of using here a method-wise average based on SVM, RF, XGB, and GPR. RF and XGB are used with their default settings, while slight modifications are introduced to SVM and GPR.

Considering SVM, in the present work the kernel function, which determines the shape of the decision boundary, is set to polynomial, instead of linear, i.e. for each two data points x, x' the kernel is defined as $k(x, x') = (\gamma\langle x, x'\rangle + r)^d$, where γ is a scale factor, r is the offset parameter and d is the polynomial degree. The polynomial kernel allows for non-linear relationships between features to be modeled by SVM. This choice is motivated by preliminary analyses suggesting that non-linear models are desirable over linear models, which performed poorly in the present case study.

As to GPR, compared to the default Radial Basis Function (RBF) kernel implemented in ML-AMPSIT, in the present work the more flexible rational quadratic kernel is applied, which is a mixture of multiple RBF kernels with different length scales, and is expressed as:

$$k(x, x') = \left(1 + \frac{\|x - x'\|^2}{2\alpha l^2}\right)^{-\alpha} \tag{4.5}$$

where $\|x - x'\|$ is the Euclidean distance computed between two data points (x, x'), α is the scale-mixture parameter that controls the weight of large and small scale variations across the RBF kernels mixture, and l is the length-scale parameter determining the smoothness of the function. The new kernel choice significantly improved the efficiency and fastness of hyperparameter tuning compared to the default option. GPR not only provides point estimates for the regression coefficients, but also quantifies the uncertainty around these estimates, offering a probabilistic interpretation of the model parameters.

The tree-based algorithms used in this work (RF and XGboost) inherently provide a quantification of parameter importance using the mean decrease in impurity (MDI) method. This method evaluates the importance of a feature by measuring the reduction in variance (or impurity) it provides across all the trees in the forest. The greater the average variance reduction attributed to a feature, the more important it is considered. On the other hand, GPR is a probabilistic method that does not provide a built-in method for measuring sensitivity, but it can be used to emulate the "high-fidelty" computationally expensive model (WRF/Noah-MP) and act as a surrogate, faster version of it. The speed of the surrogate model allows for a fast implementation of the Sobol method. In other words, the sensitivity indices are obtained directly from the ML-inferred relation between the original model inputs and outputs. Similarly, the Support Vector Machine (SVM) method with a non-linear kernel (as implemented here), does not inherently provide a measure of feature importance. For linear SVMs, the coefficients can indicate feature importance similarly to LASSO, where the magnitude of the regression coefficients directly indicates the importance of the corresponding features. However, with non-linear SVMs, the connection between coefficient weights and parameter importance is lost due to the complexity of the kernel transformations. Therefore, a dedicated sensitivity method must be applied after the regression task to assess feature importance. As for GPR, this modified implementation of ML-AMPSIT utilizes the Sobol method to perform the sensitivity analysis using the results of an SVM-based surrogate model.

Despite the differences in the feature importance evaluation methods, the comparison between the sensitivity indices obtained is well-posed. This is because all methods aim to quantify the relative importance of parameters and are based on standardized, non-dimensional input data.

The input-output data used to train the algorithms are collected using a sampling strategy that changes all the parameter values according to quasi-random low-discrepancy Sobol sequences to optimally explore the parameter space. For each method available in ML-AMPSIT, the ensemble of simulations generated by varying the input parameters is divided into training and test sets in a 70-30 ratio. During the training phase, the ensemble members from the training set are used by the algorithms through a cross-validation procedure to fit their hyperparameters to the provided dataset. The search for the best model parameters is aided by a Bayesian optimization algorithm for hyperparameter tuning. After the training, the algorithms use the tuning results to predict the ensemble members in the test set.

4.2.3 Noah-MP parameters

The Noah-MP parameters whose importance is evaluated in the present study include those defined in the WRF MPTABLE categorized on the basis of the vegetation type, hence excluding all the soil-related and crop-related parameters. Moreover, all the parameters associated with physical processes involving snow, ice, microphysics, runoff and groundwater have been ignored because not relevant to this case study. Hence, the chosen parameters are strictly related to the canopy properties having a clear role in the radiation and heat budget.

4.2.3.1 Relevant parameters for radiative fluxes

Leaves play an important role in the net radiation budget. Their cellular structure tends to almost completely diffuse scattering, resulting in relatively equal amounts of reflected and transmitted radiation (Monteith et al., 1979). The overall canopy albedo is usually significantly lower than the individual leaf albedo, depending on the angle of solar incidence and the stand architecture, which influence processes such as the penetration of radiation inside the canopy and radiation trapping, which is very strong in forests. These effects are mostly controlled by the scattering parameters, canopy height, and the leaf area index. Stems also play a role in the radiation budget, but, especially for forests, the main radiation exchange comes from the canopy itself, while trunks are less important. The greatest radiation trapping is present in canopies with larger leaf area index due to small leaf size, such as in conifers, because the small size increases the overall roughness and scattering.

In the modified two-stream canopy radiation scheme of Noah-MP, the leaves are treated as clouds aggregated into tree crowns with gaps, uniformly distributed in the single grid cell, and positioned over the underlying surface. Both inter-crown and intra-crown gaps are considered, and the parameters affecting these gaps are the crown radius, canopy height, and the leaf/stem area index.

In the calculation of the radiative fluxes, the vegetation absorption is regulated by the canopy leaf/stem orientation index, reflection, and transmission parameters XL, ρ, and τ respectively. The key variables entering the calculations of these fluxes are the scattered fraction of leaf intercepted radiation $\omega_L = \rho + \tau$, and the single scattering albedo A:

$$A = 0.5 \cdot (\rho + \tau) \cdot \frac{\left[\frac{\phi_1 + \phi_2 \cdot \cos\theta'}{\phi_1 + 2 \cdot \phi_2 \cdot \cos\theta'}\right] \cdot \left[1 - \frac{\phi_1 \cdot \cos\theta'}{\phi_1 + 2 \cdot \phi_2 \times \cos\theta'} \cdot \log\left(\frac{\phi_1 \cdot (1 + \cos\theta') + 2 \cdot \phi_2 \cdot \cos\theta'}{\phi_1 \cdot \cos\theta'}\right)\right]}{(\phi_1 + 2 \cdot \phi_2 \cdot \cos\theta')} \tag{4.6}$$

where both ϕ_1 and ϕ_2 are polynomials of the variable leaf/stem orientation index XL. θ is the solar zenith angle and θ' is the adjusted solar zenith angle that takes into account the geometric configuration of the canopy as:

$$\theta' = \tan^{-1}\left[\left(\frac{HVT - HVB}{2RC}\right) \times \tan(\theta)\right] \tag{4.7}$$

Where HVT and HVB are the canopy top and bottom heights, and RC is the crown radius.

The values of ρ and τ are specified for the near-infrared band and the visible band for both leaf and stem, taking into account the different response of vegetation to these two wavelength ranges, i.e. Noah-MP defines the following 8 parameters RHOL_NIR, RHOS_NIR, RHOL_VIS, RHOS_VIS, TAUL_NIR, TAUS_NIR, TAUL_VIS, TAUS_VIS.

4.2.3.2 Relevant parameters for latent heat, sensible heat and momentum fluxes

Noah-MP uses LAI, ρ, and τ to derive the fraction of vegetation under sunlight, which affects the stomatal resistance according to the Jarvis scheme, and ultimately the transpiration contribution to latent heat. It is expected that in the present case study this contribution is the most significant for the total latent heat flux, considering the absence of a wet fraction of the canopy. The stomatal resistance is composed of three different contributions referring respectively to the incoming solar radiation ($F_{rs,solar}$) affected by the maximum and minimum stomatal resistances $RSMAX$ and RS and the radiation stress factor RGL, the air temperature ($F_{rs,temp}$) affected by the optimal temperature parameter $TOPT$, and the vapor pressure deficit ($F_{rs,vpd}$) affected by the vapor pressure deficit stress factor HS, which modulates the influence of the pressure deficit. The three contributions are calculated as:

$$F_{rs,solar} = \max\left(0.0001, \frac{\frac{2\cdot S_{par}}{RGL} + \frac{RS}{RSMAX}}{1 + \frac{2\cdot S_{par}}{RGL}}\right) \tag{4.8}$$

where S_{par} is the photosynthetically active radiation, which has a different value for sunlit and shaded leaves and therefore represents a discriminant between day and night in the final stomatal resistance calculation;

$$F_{rs,temp} = \max\left(0.0001, 1 - 0.0016 \cdot (TOPT - T_v)^2\right) \tag{4.9}$$

where T_v is the canopy temperature;

$$F_{rs,vpd} = \max\left(0.01, \frac{1}{1 + HS \cdot \max(0, w_{can\_air\_sat} - w_{can\_air})}\right) \tag{4.10}$$

where $w_{can\_air\_sat}$ and w_{can_air} are the saturated mixing ratio and the mixing ratio of the canopy air.

The final stomata resistance is:

$$R_{stomata} = \frac{RS}{F_{rs,solar} \cdot F_{rs,temp} \cdot F_{rs,vpd} \cdot \beta_{tr}} \tag{4.11}$$

where β_{tr} is a soil moisture factor computed as a function of the soil wetness in each root layer, decreasing the stomatal resistance for increasingly wet soils.

Another important parameter for both latent and sensible heat fluxes is the leaf boundary layer resistance R_{bl}, which is controlled by the leaf characteristic dimension $DLEAF$, the canopy wind extinction parameter $CWPVT$, the leaf and stem area indexes LAI and SAI, the canopy height HVT, and the roughness length $Z0MVT$:

$$R_{bl} = 50 \cdot \frac{a}{1 - e^{-a/2}} \cdot \sqrt{\frac{DLEAF}{U_c}} \tag{4.12}$$

where a is the canopy wind extinction coefficient containing the Monin-Obukhov stability function (Monin et al., 1954) for heat $\psi_h(\zeta_{MO})$ evaluated below the canopy:

$$a = \left[CWPVT \cdot (LAI + SAI) \cdot HVT \cdot \psi_h(\zeta_{MO})\right]^{0.5} \tag{4.13}$$

and U_c is the wind speed evaluated at the height of the canopy top HVT:

$$U_c = \frac{u_*}{k} \ln\left(\frac{HVT - d}{Z0MVT}\right) \tag{4.14}$$

where

$$u_* = \frac{U_R k}{\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_m\right]} \tag{4.15}$$

is the friction velocity, k is the von Kármán constant, $d = 0.65 \cdot HVT$ is the zero-plane displacement height, U_R is the velocity at the reference height $z_{ref} = d + z'_a$, where z'_a is the height of the first WRF level above the zero-plane displacement height, and ψ_m is the Monin-Obukhov stability function for momentum evaluated below the canopy.

The friction velocity is also used to define the exchange coefficients for momentum:

$$C_m = \frac{k^2}{\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_h\right]^2}, \tag{4.16}$$

and heat:

$$C_h = \frac{k^2}{\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_m\right]\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_h\right]}, \tag{4.17}$$

which are used to compute the aerodynamic conductances for the surface fluxes.

R_{bl} is used to calculate the associated heat conductances for sensible and latent heat flux $C_{SH,cond,leaf}$ and $C_{LH,cond,tr}$:

$$C_{SH,cond,leaf} = \frac{2 \cdot EVAI}{R_{bl}} \tag{4.18}$$

$$C_{LH,cond,tr} = \left[\frac{EVAI_{sun}}{R_{bl} + R_{stomata,sun}} + \frac{EVAI_{shd}}{R_{bl} + R_{stomata,shd}}\right] \tag{4.19}$$

where $EVAI$ is equal to $LAI + SAI$ in the setup used in this work, while subscripts *sun* and *shaded* indicate respectively the fraction of vegetation exposed to direct solar radiation and in the shade.

Larger LAI and SAI values imply that the leaf boundary resistance R_{bl} decreases, resulting in higher heat conductance coefficients. Increasing the leaf size $DLEAF$ also increases R_{bl}, because larger leaves produce thicker laminar layers (Monteith et al., 1979). Since $CWPVT$ controls the wind extinction within the canopy, a larger value of $CWPVT$ implies lower wind speed in the canopy, reducing the heat exchange.

The sensible heat and latent heat fluxes are also modulated by a turbulent transfer coefficient, entering the computation of the below-canopy aerodynamic resistances, which is directly controlled by the canopy height HVT and the roughness length $Z0MVT$ through u_*:

$$C_{h,turb} = \kappa u_* (HVT - d) \tag{4.20}$$

Although equation (4.20) implies that both increased roughness and increased canopy height increase the heat exchange between the canopy and the atmosphere, it is worth noting that the two parameters have different weights, with $C_{h,turb}$ depending linearly on HVT, while on the roughness length as $\frac{1}{ln(1/Z0MVT)}$.

4.2.3.3 Investigated parameters

Table 4.1 shows the parameters investigated in this work, summarizing the above arguments and pointing out the main physical processes they control. Since the effect of stem transmittance is not relevant for forest vegetation for both near-infrared and visible bands, the respective parameters of the MTABLE $TAUSNIR$ and $TAUSVIS$ have been excluded from the sensitivity analysis. Moreover, also the parameter controlling the bottom height of the canopy (HVB) has not been taken into account, in order not to have unphysical simulations resulting from the contemporary variation of the top height of the canopy (HVT).

Table 4.1 also shows the reference value of each parameter based on the MODIS land cover class "mixed-forest". In this work, the percentage of variation adopted for the sensitivity analysis for the pseudorandom ensemble is set as 20% for all the parameters. The optimal temperature parameter $TOPT$ has been converted to degrees Celsius before applying the variation, to avoid unrealistic temperature settings. Since the results obtained by a regression algorithm over a specific simulation ensemble cannot be assumed to be valid for any type of ensemble, the specific variations chosen for the parameters in this study are expected to generate results that are strictly valid only for this range of variations. Nonetheless, the parameter space explored in this work is reasonably representative of the variability of forest-like land cover classes. This ensures that the results are physically consistent and as realistic as possible.

Parameters	Definition	Physical processes	Reference
LAI	leaf area index, one-sided	Rad / Ext / Stom	2.20
SAI	stem area index, one-sided	Rad / Stom	0.40
RHOLNIR	near-infrared leaf reflectance	Rad / Stom	0.45
RHOLVIS	visible leaf reflectance	Rad / Stom	0.10
RHOSNIR	near-infrared stem reflectance	Rad / Stom	0.39
RHOSVIS	visible stem reflectance	Rad / Stom	0.16
TAULNIR	near-infrared leaf transmittance	Rad / Stom	0.25
TAULVIS	visible leaf transmittance	Rad / Stom	0.05
XL	leaf/stem orientation index	Rad / Stom	0.25
RC	tree crown radius	Rad	1.40
HVT	height of canopy top	Rad / Ext / Turb / Leaf BL	16.0
Z0MVT	momentum roughness length vegetated surface	Ext / Turb / Leaf BL	0.80
DLEAF	characteristic leaf dimension	Leaf BL	0.040
CWPVT	empirical canopy wind parameter	Ext / Leaf BL	0.29
RS	Jarvis minimum Stoml resistance	Stom	125
RSMAX	Jarvis maximum Stoml resistance	Stom	5000
RGL	Jarvis radiation stress parameter	Stom	30
TOPT	Jarvis optimum transpiration air temperature	Stom	298
HS	Jarvis vapor pressure deficit parameter	Stom	51.93

Table 4.1: Parameters analyzed in this work. The main physical processes and reference values for each parameter are given.

4.3 Results

4.3.1 Valley circulation

This section presents the results of the reference simulation, i.e. with all the parameters set to the reference value, with the main aim of describing its main features, considering in particular the development of the thermally-driven circulations. Results are presented at four times representative of the diurnal cycle of both slope and valley winds, i.e., 18:00, 00:00, 06:00 and 12:00 UTC and along the cross-valley and along-valley sections shown in Fig. 4.1b.

From Fig. 4.2 it can be seen that at 18:00 UTC down-slope winds are already well-developed, especially along the western valley sidewall, due to earlier orographic shading in the afternoon. In fact, since the valley is north-south oriented, the main asymmetries in the circulation come from the changing insulation of the valley sidewalls during the day, which particularly affects the development of slope winds, which quickly react to surface heating/cooling. On the other hand, the up-valley wind is still strong in the valley core, with velocities around 5 m s^{-1} between 500 and 1000 m AGL, and extending rather homogeneously in the along-valley direction up to the ridge top. At 00:00 UTC (Fig. 4.3), shallow down-slope winds converge towards the valley floor, contributing to the development of a cold pool (Arduini et al., 2016). A weak down-valley wind is developing starting from the lowest atmospheric layers, and is topped by a residual up-valley wind extending from 1000 m AGL to the ridge top. The along-valley section (Fig. 4.3c) highlights that the down-valley wind is developed especially at the valley exit. Fig. 4.3c also shows lower temperatures in the valley with respect to the plain, driving the progressive development of the down-valley wind. Indeed, the down-valley wind becomes stronger at 06:00 UTC (Fig. 4.4), extending to higher levels and towards the inner valley sections. The down-slope winds are significantly weaker, especially close to the valley floor, where colder air has accumulated during the night. At 12:00 UTC (Fig. 4.5), up-slope winds are well-developed along both sidewalls, converging at the ridge top, and topped by return branches causing subsidence in the valley center. The along-valley section highlights higher potential temperature in the valley with respect to the plain, causing the transition from down- to up-valley wind.

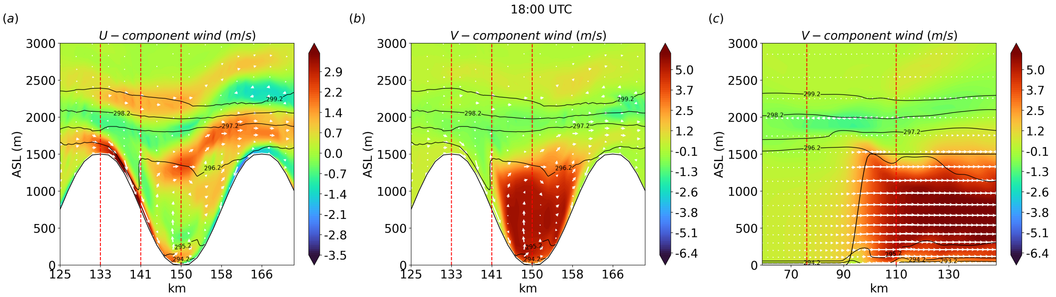


Figure 4.2: Cross-valley section (see Fig. 4.1 for the position) of (a), cross-valley wind u (color shading), wind component parallel to the section (arrows) and potential temperature (solid lines) (b) along-valley wind v (color shading), wind component parallel to the section (arrows) and potential temperature (solid lines); (c) along-valley section of along-valley wind v (color shading), wind component parallel to the section (arrows) and potential temperature (solid lines). All variables refer to 18:00 UTC of the first simulated day. Red dotted lines indicate the location of the valley floor, slope, and ridge points, which will be considered in the following analysis. For visualization purposes, the vertical wind component w in the cross-valley sections has been magnified by a factor of 10 and wind vectors are plotted every two grid points.

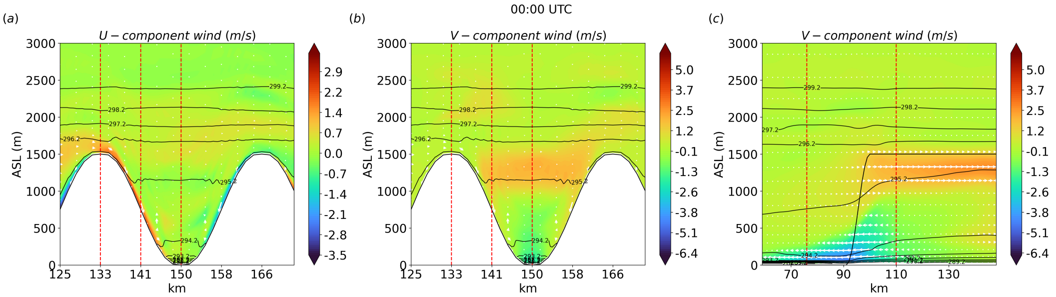


Figure 4.3: As Fig. 4.2, but referring to 00:00 UTC.

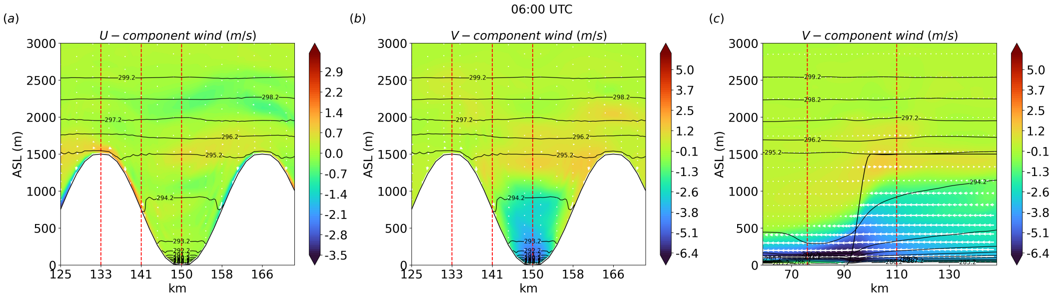


Figure 4.4: As Fig. 4.2, but referring to 06:00 UTC.

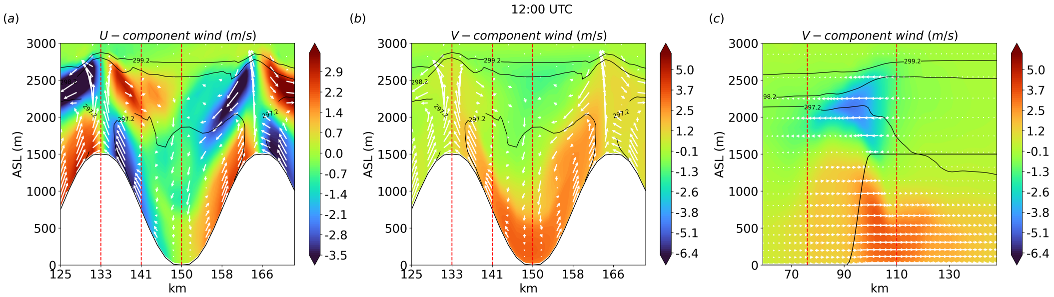


Figure 4.5: As Fig. 4.2, but referring to 12:00 UTC.

4.3.2 Simulation ensemble

The variability of the simulation ensemble is evaluated by analyzing the time series of the along-valley (south-north, v) and of the cross-valley (west-east, u) wind components and of the potential temperature at the lowest model level, as well as of the surface latent and sensible heat fluxes (Figs. 4.6-4.10) at four points of the domain representative respectively of the valley floor, the western valley slope, the western ridge and the plain close to the valley exit (see Fig. 4.1b). Moreover, for the same four points, the ensemble results are also shown for the vertical profiles of the horizontal wind components and of the potential temperature in the lowest 500 m AGL at the same four representative time steps analyzed in Section 3.1 (Figs. 4.11-4.13). Both the time series and the vertical profiles are spatially averaged over 11 cells along y, from 5 cells south to 5 cells north of the selected point, to obtain more robust results.

Focusing on the ensemble time series (Figs. 4.6-4.10), it can be seen that the ensemble spread changes not only with time and position, but also with the variable of interest. The along-valley wind component seems more sensitive to input parameter variability close to the peak of the up-valley wind in the afternoon, considering in particular the valley floor, the slope and the ridge. On the valley floor the along-valley wind component displays significant spread also during the night (Fig. 4.6). The ensemble spread of the cross-valley wind is generally larger, especially considering the slope point during the night (down-slope wind, Fig. 4.7b). The variability of the potential temperature tends to be larger during daytime on the valley floor, the ridge and the plain, while over the slope the ensemble spread is roughly constant throughout the day. The variables that present the most marked daily cycle in the ensemble spread are the latent and sensible heat fluxes, which show strong differences between nighttime, when the variability is very low, and daytime, when differences between the ensemble simulations are considerably larger.

Considering the vertical profiles of the along-valley wind component, Figure 4.11 shows that the ensemble spread is larger at night on the valley floor, where the variability tends to increase at higher levels, and over the plain, especially above the peak of the low-level jet at 100 m AGL. On the other hand, the variability of the along-valley wind component is particularly low over the ridge, except at 18:00 UTC. The cross-valley wind component (Fig. 4.12) over the slope shows rather strong variability at all levels at 18:00 UTC (first phase of the down-slope wind) and at 12:00 UTC (up-slope wind), while at 00:00 and 06:00 UTC the spread of the down-slope wind is concentrated at low levels. Therefore, the strong variability observed during nighttime in Fig. 4.7b affects only the shallow down-slope wind layer. A strong variability is observed over the ridge in the upper levels at 18:00 UTC, reducing in the following hours. The ensemble spread of the cross-valley wind component tends to be low over the valley and plain points, with slightly larger variability in both cases at 12:00 UTC and over the plain close to the surface at 06:00 UTC. The ensemble spread of the potential temperature does not present a significant height dependence for all the points and times considered, except over the plain during the night, where variability increases with height (Fig. 4.13). The variability is larger at all points at 12:00 UTC, coherently with Fig. 4.8, while it is particularly low over the ridge at night.

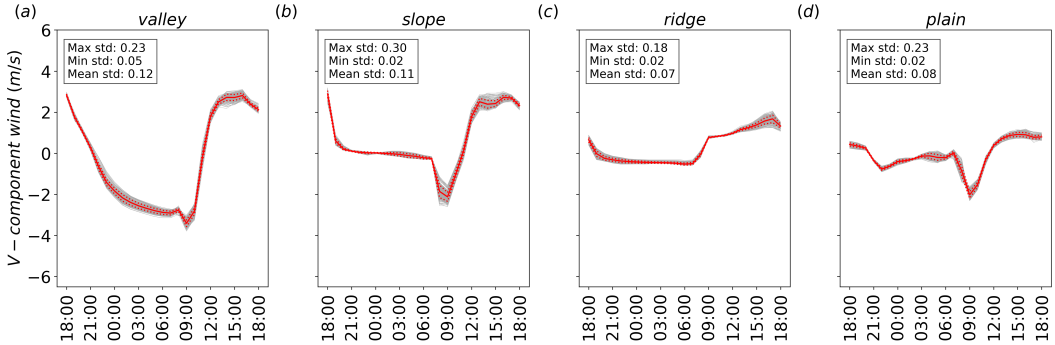


Figure 4.6: Ensemble time series for the along-valley wind component at the four points shown in Fig. 4.1b representative of the valley floor, the slope, the ridge and the plain. The continuous and dotted red lines indicate respectively the ensemble mean and the standard deviation of the ensemble. The box indicates the mean, minimum and maximum standard deviation of the ensemble time series.

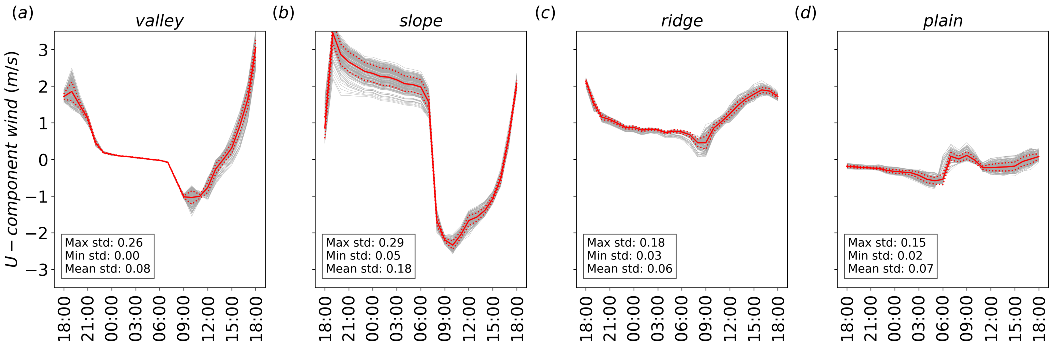


Figure 4.7: As Fig 4.6, but for the cross-valley wind component.

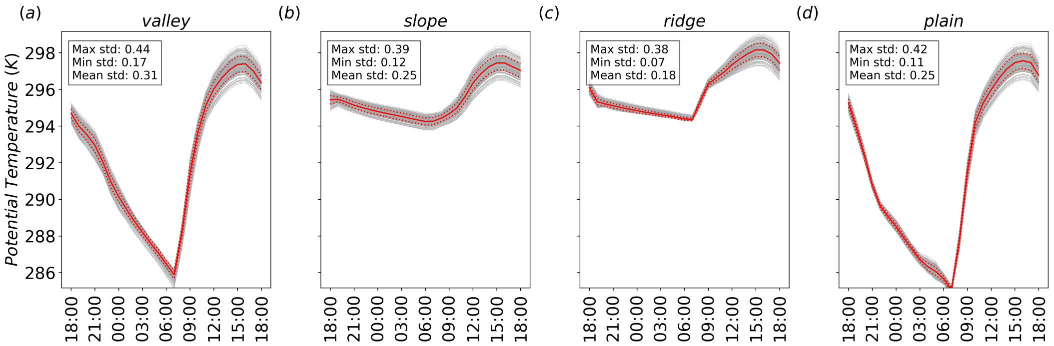


Figure 4.8: As Fig 4.6, but for the potential temperature.

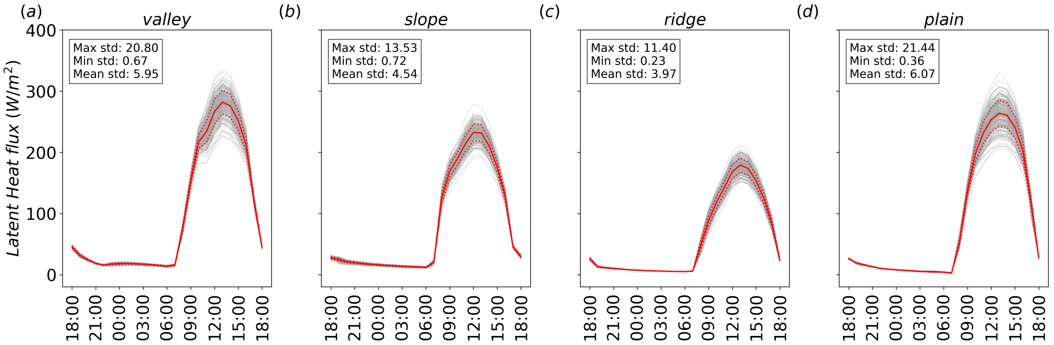


Figure 4.9: As Fig 4.6, but for the latent heat flux.

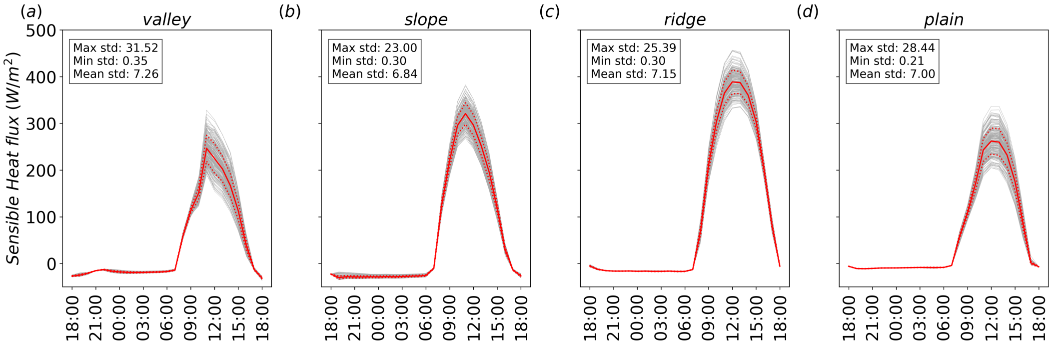


Figure 4.10: As Fig 4.6, but for the sensible heat flux.

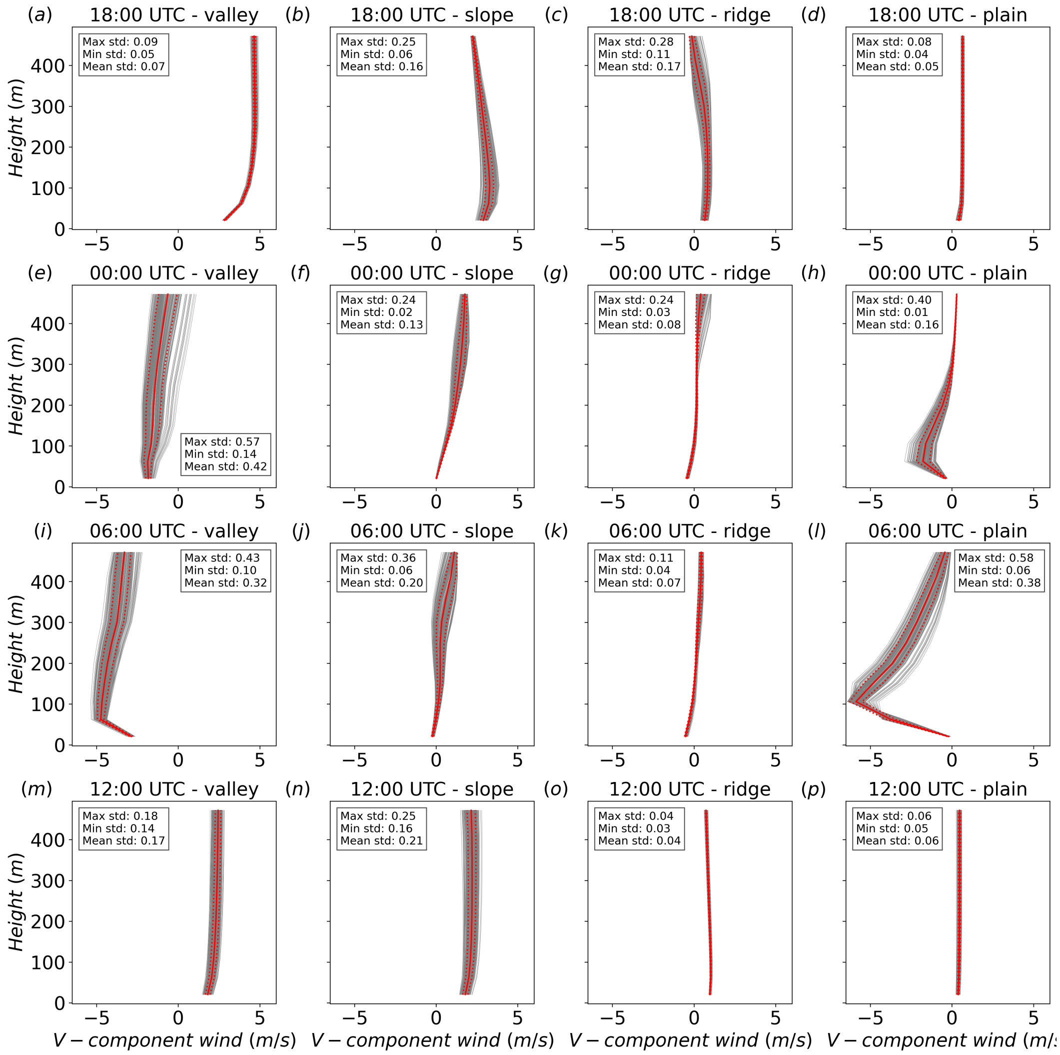


Figure 4.11: Ensemble vertical profiles in the lowest 500 m AGL for the along-valley wind component at the four points shown in Fig. 4.1b representative of the valley floor, the slope, the ridge and the plain and four different times representative of the daily cycle of valley and slope winds: 18:00, 00:00, 06:00 and 12:00 UTC. The continuous and dotted red lines indicate respectively the ensemble mean and the standard deviation of the ensemble. The box indicates the mean, minimum and maximum standard deviation of the ensemble time series.

4.3.3 Sensitivity analysis

This section presents the results of the sensitivity analysis, focusing in particular on the outcomes of the assessment of the parameters' importance. The analysis is divided between nighttime (18:00-05:00) and daytime (06:00-17:00), in order to evaluate possible differences depending on the phase of the thermally-driven circulations. Moreover, the analysis focuses on the four points representative of the valley floor, the slope, the ridge and the plain shown in Fig. 4.1b. To obtain robust results, also in this case the analyses are performed averaging model results along y over 11 cells, from 5 cells south to 5 cells north of the selected point. In

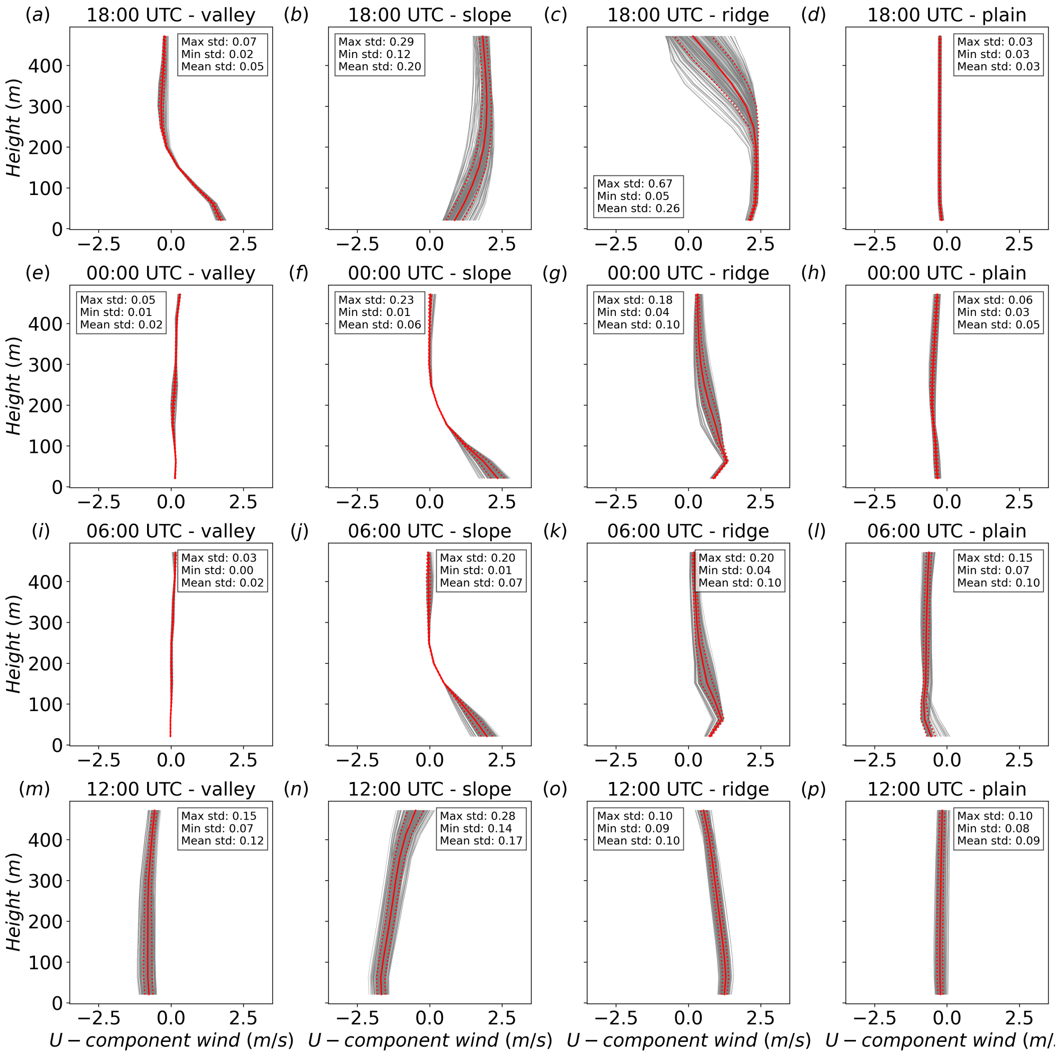


Figure 4.12: As Fig. 4.11, but for the cross-valley wind componnent.

this section, the analyses referring to the lowest vertical level are presented. The analysis referring to higher levels has also been performed, but, since the main patterns do not show a significant variation with height, the figures related to the other vertical points analyzed are left in the appendix.

The performance metrics of the surrogate models used in this work are also presented, to highlight that they are able to satisfactorily reproduce the input-output relationships of the original model.

4.3.3.1 Performance metrics

The performance metrics are reported in Figure 4.14 for all the points and variables considered. The results refer to the average of all the surrogate models used. In particular, the metrics shown, R^2, MSE, and MAE, are estimates of how well the methods performed during the test phase, i.e., when predicting new, unseen data.

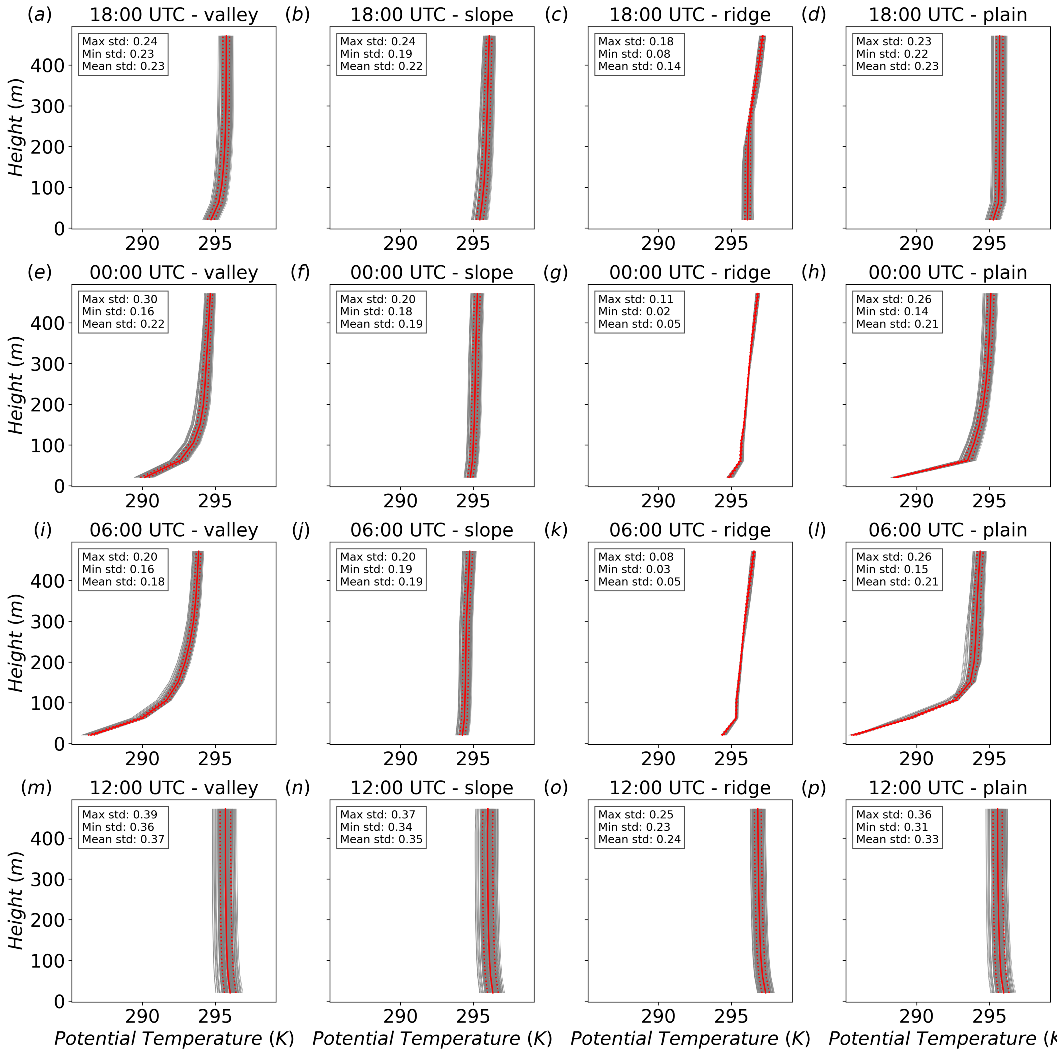
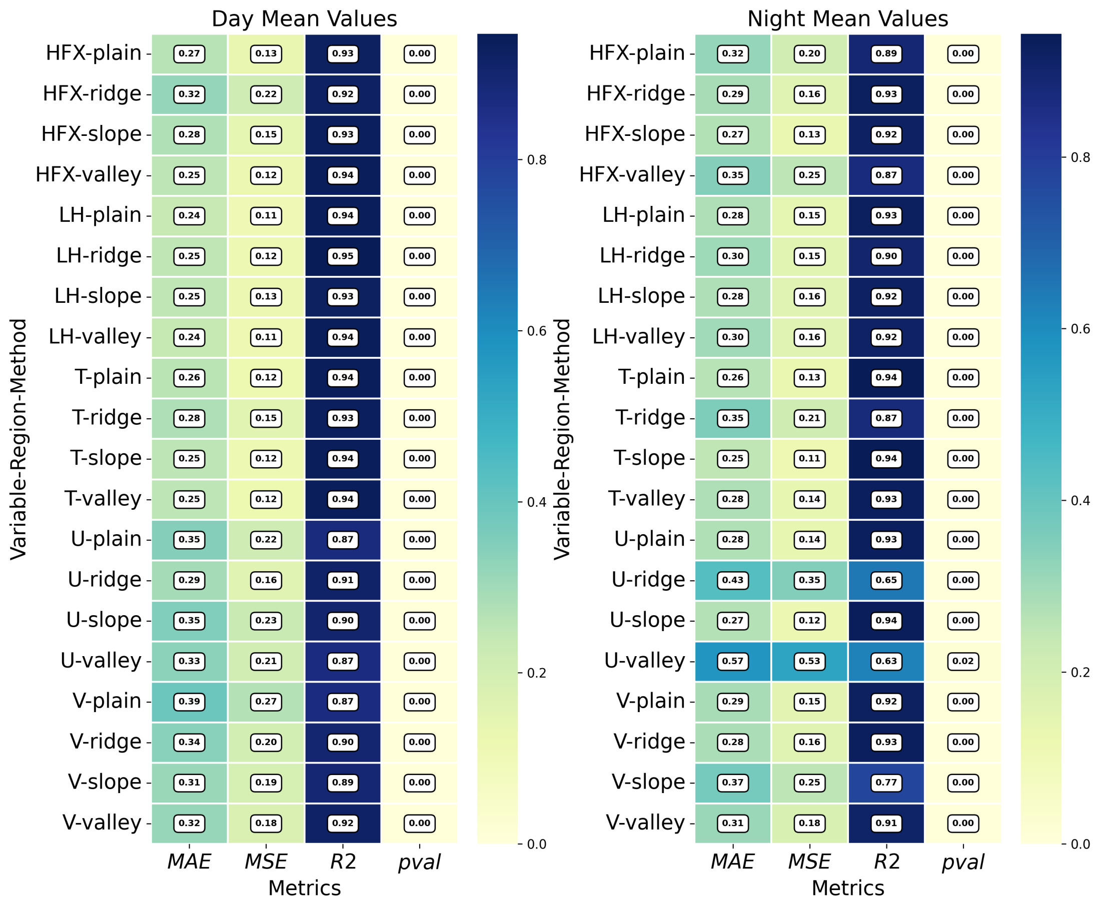


Figure 4.13: As Fig. 4.11, but for the potential temperature.

For the majority of the variables and points, the results highlight that R^2 is about 0.9, MSE is lower than 0.2, and MAE is around 0.3. Notable exceptions are the nocturnal cross-valley wind component in the valley and over the ridge, and the along-valley wind component over the slope. However, it should be noted that very good values of the metrics are obtained for the variable-point pairs most representative of the development of slope and valley winds, i.e. the along-valley wind component over the valley floor and the cross-valley wind component over the slope. This highlights the ability of the surrogate models to accurately reproduce the influence of the input parameters on the output variables, especially where a strong signal is present.

The overall good results suggest that the algorithms are correctly trained and not affected by overfitting. Even when each algorithm is evaluated separately (not shown), the metrics retain their robustness, with GPR and SVM showing the best performance.



Day Mean Values

Variable-Region-Method	MAE	MSE	R2	pval
HFX-plain	0.27	0.13	0.93	0.00
HFX-ridge	0.32	0.22	0.92	0.00
HFX-slope	0.28	0.15	0.93	0.00
HFX-valley	0.25	0.12	0.94	0.00
LH-plain	0.24	0.11	0.94	0.00
LH-ridge	0.25	0.12	0.95	0.00
LH-slope	0.25	0.13	0.93	0.00
LH-valley	0.24	0.11	0.94	0.00
T-plain	0.26	0.12	0.94	0.00
T-ridge	0.28	0.15	0.93	0.00
T-slope	0.25	0.12	0.94	0.00
T-valley	0.25	0.12	0.94	0.00
U-plain	0.35	0.22	0.87	0.00
U-ridge	0.29	0.16	0.91	0.00
U-slope	0.35	0.23	0.90	0.00
U-valley	0.33	0.21	0.87	0.00
V-plain	0.39	0.27	0.87	0.00
V-ridge	0.34	0.20	0.90	0.00
V-slope	0.31	0.19	0.89	0.00
V-valley	0.32	0.18	0.92	0.00

Night Mean Values

Variable-Region-Method	MAE	MSE	R2	pval
HFX-plain	0.32	0.20	0.89	0.00
HFX-ridge	0.29	0.16	0.93	0.00
HFX-slope	0.27	0.13	0.92	0.00
HFX-valley	0.35	0.25	0.87	0.00
LH-plain	0.28	0.15	0.93	0.00
LH-ridge	0.30	0.15	0.90	0.00
LH-slope	0.28	0.16	0.92	0.00
LH-valley	0.30	0.16	0.92	0.00
T-plain	0.26	0.13	0.94	0.00
T-ridge	0.35	0.21	0.87	0.00
T-slope	0.25	0.11	0.94	0.00
T-valley	0.28	0.14	0.93	0.00
U-plain	0.28	0.14	0.93	0.00
U-ridge	0.43	0.35	0.65	0.00
U-slope	0.27	0.12	0.94	0.00
U-valley	0.57	0.53	0.63	0.02
V-plain	0.29	0.15	0.92	0.00
V-ridge	0.28	0.16	0.93	0.00
V-slope	0.37	0.25	0.77	0.00
V-valley	0.31	0.18	0.91	0.00

Figure 4.14: Performance metrics averaged over all the algorithms used in the present study for each variable and points analyzed at the first vertical level. The heatmaps show MAE, MSE, R^2, and pvalue for daytime (left, 06:00-17:00) and nighttime (right, 18:00-05:00).

4.3.3.2 Parameters' importance

As for the performance metrics shown in the previous section, the analysis of parameter importance presented here refers to the average values of all the methods, in order to obtain an ensemble result that is robust to individual method variations. Nevertheless, a prior evaluation of the parameter importance (not shown) showed very high consistency across methods.

It is important to emphasize the relative nature of the parameters' importance analysis; each regression is run at a specific time and location with no information about other times and locations. Therefore, comparisons between different columns of the heatmap do not reflect absolute importance but should be understood in terms of the relative impact of the parameters. On the other hand, the importance of a particular parameter can be compared with the importance of the other parameters in the same column of the heatmap (same variable-location pair), to evaluate what are the parameters most affecting the variable under investigation in that particular location.

To facilitate the heatmap visibility, a filter has been applied such that any parameter importance less than 0.05 has been set to white in the colormap, highlighting the most influential parameters.

When the parameters are ranked by their importance across regions and variables, a clear hierarchy emerges. RS consistently ranks as one of the most influential parameters, especially for latent heat flux and potential temperature, with high values both during daytime and nighttime. From Eq. (4.8) and Eq. (4.11), it can be seen that that the influence of the parameter RS persists both during daytime and nighttime. Moreover, it is clear from Eq. (4.11) that RS is the most influential parameter affecting the stomatal resistance. This result is consistent with the expectation that the stomatal resistance strongly influences LH through modulating transpiration, and indirectly affecting temperature. Considering the other parameters related to the stomatal resistance, RGL and RSMAX display negligible importance. HS, which represents the weight that vapor saturation has over stomatal resistance, shows secondary but non-negligible importance across all variables. Considering that humidity plays a minor role in the present setup, this result suggests that this parameter might be much more critical in situations where mixing ratio and soil moisture are more important. Higher importance values are found for TOPT considering both wind components during daytime. TOPT is also observed to be relevant for heat fluxes, and in particular it shows a very high importance for the heat fluxes over the ridge, especially for the latent heat flux. TOPT enters the Equation (4.9) as $(TOPT - T_v)^2$, implying that, when its difference with the virtual temperature T_v is larger, this parameter can be more influential. Therefore, the higher temperatures observed over the ridge during daytime could explain a larger effect of $TOPT$ on the latent heat flux through the contribution of $F_{rs,temp}$.

HVT is also a critical parameter, showing strong effects on all variables, with the exception of the latent heat flux. HVT displays a high importance especially for the sensible heat flux and potential temperature, across all regions, with its influence often intensifying at night. Equation (4.7) shows that HVT is a crucial parameter for radiation, influencing radiation trapping, which is greater for taller canopies, and ultimately translating HVT into a temperature modulating parameter. Leaf boundary resistance is also strongly affected by HVT, as seen in Eq. (4.12), thus influencing the heat conductances for both sensible and latent heat fluxes, although a clear difference between the two is the additional presence of the stomatal resistance in the latent heat conductance, which explains the great importance of RS observed for the latent heat flux.

Considering radiation-related parameters, the only relevant parameter turns out to be the near-infrared reflectivity RHOL_NIR, which shows a significant influence during daytime on the sensible heat flux and the potential temperature, and at night for the cross-valley wind component over the slope and the ridge and the along-valley wind component over the slope. On the other hand, transmissivity, both in the near-infrared and in the visible (TAUL_NIR and TAUL_VIS) are observed to be always non-influential parameters. The presence of the near-infrared reflectivity as one of the most important parameters highlights the important role of radiation inside the canopy in modulating the surface energy budget and thus the heating-cooling of the near-surface atmospheric layers. The range chosen to vary the near-infrared leaf reflectivity and transmissivity is compatible with the empirically observed values of these two parameters, i.e. the leaf transmissivity ranges are significantly smaller compared to those of the leaf reflectivity in the near-infrared spectrum (Monteith et al., 1979). Therefore, considering their very similar role in Noah-MP, it is expected that larger variations in the reflectivity input would lead to larger effects on the analyzed variables. Similarly, since RHOL_VIS is empirically observed to be nearly 5 times smaller than RHOL_NIR, and they have been modified accordingly in this work, it is easily expected, due to their equivalence in the Noah-MP code, that RHOL_VIS is negligible compared to RHOL_NIR. Finally, TAUL_VIS can be considered negligible for the same reason as RHOL_VIS.

LAI overall displays rather low importance values, with the exception of the latent heat flux and the sensible heat flux, and a secondary role for the wind components during daytime. It is important to point out that these results are valid only for a 20% perturbation of LAI, while across the Noah-MP Modis-based vegetation classes this parameter varies more than 100%, possibly leading to a significantly different impact on the output of the simulations, as for the setup used in Di Santo et al., 2024, where LAI was varied by 50%. On the other hand, SAI is always observed as a non-influential parameter. Since the variation of SAI is generally much smaller than the variation of LAI, it is expected that SAI plays a smaller role than LAI, especially since LAI influences more processes in Noah-MP compared to SAI, e.g. the calculation of canopy radiation and transpiration.

CWPVT shows low importance values during daytime, while it becomes more important at night for the heat fluxes, especially over the ridge and slope for the latent heat flux.

Finally, some of the parameters show transient importance throughout the day, which could lead to fluctuations in feature importance that are not fully represented in the time windows chosen for the average.

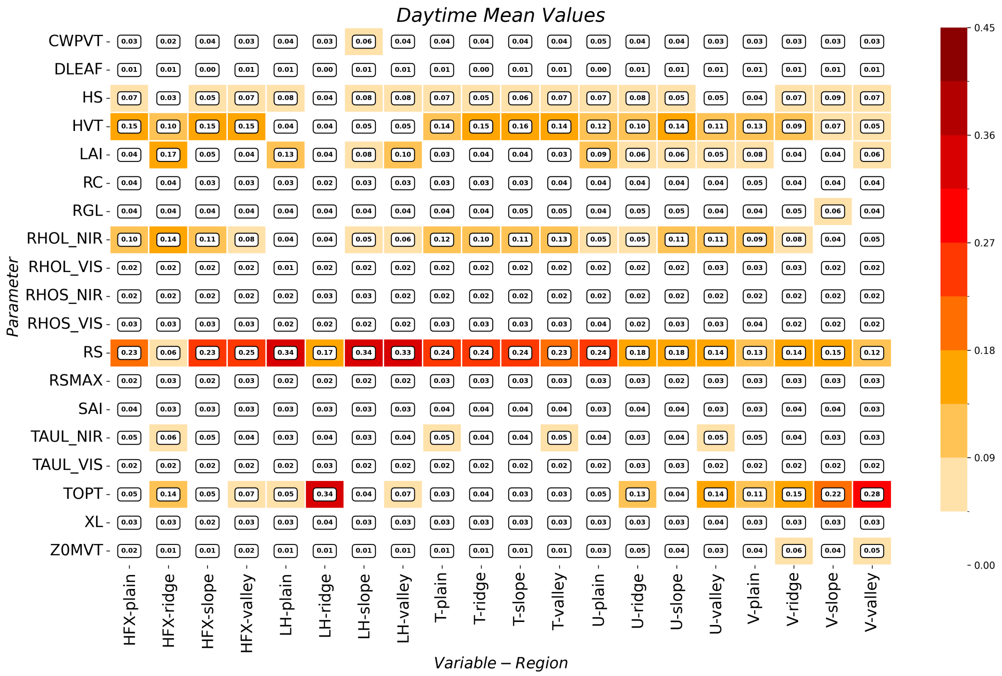


Figure 4.15: Parameters' importance during daytime (06:00 - 17:00 UTC) computed for the first model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

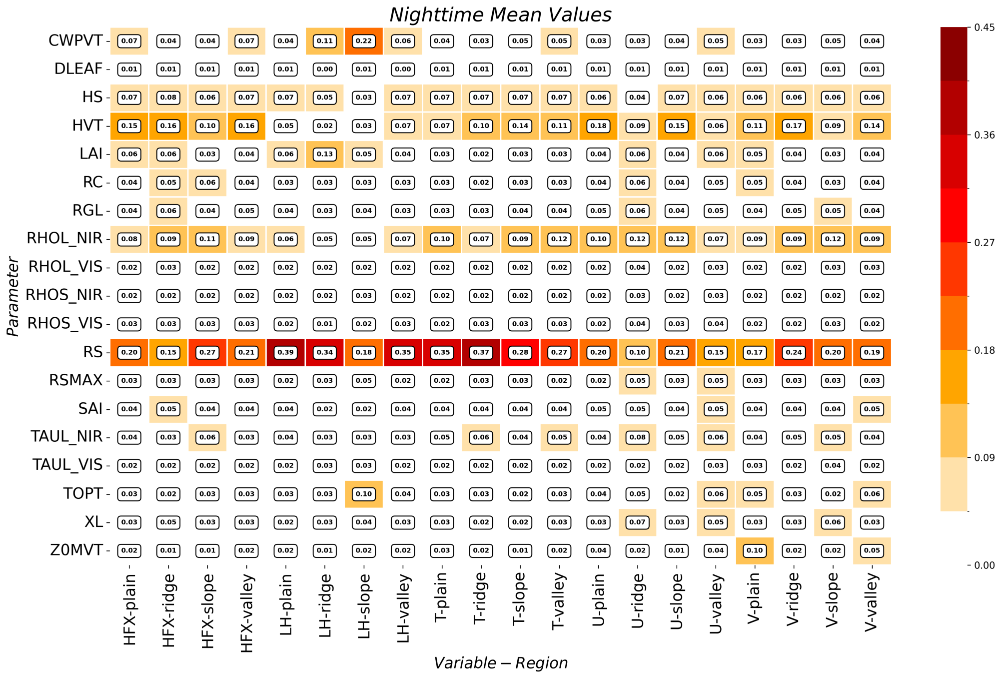


Figure 4.16: As in Fig, 4.15, but during nighttime (18:00 - 05:00 UTC).

4.4 Conclusions

This study presented a detailed investigation of the importance of LSM parameters in atmospheric simulations over complex terrain, with a focus on thermally driven circulations in an idealized valley. Recognizing that LSMs are a significant source of uncertainty in NWP models, the study investigated how the uncertainty in the representation of LSM parameters can affect atmospheric processes in the planetary boundary layer, highlighting the challenges posed by spatial and temporal heterogeneity of surface conditions in mountainous areas. The methodology integrated a variance-based approach within a data-driven modeling framework, using the Sobol method enhanced by machine learning algorithms to determine parameter significance. This combination allowed a fast evaluation of parameter effects on model outputs, taking into account non-linear relationships and interactions.

The analysis revealed that stomatal resistance (RS), canopy height (HVT), and near-infrared leaf reflectivity (RHOL_NIR) are critical in modulating atmospheric boundary layer processes both during daytime and nighttime, with RS being the most influential parameter across all regions and variables. Leaf area index (LAI) and optimal temperature (TOPT) were also identified as influential. The canopy wind extinction parameter (CWPVT) was observed to be influential mostly during the night for the latent heat flux.

The results obtained in this work highlight what are the LSM parameters that mostly affect model results, and, as a consequence, those parameters whose quantification should be treated with most care. It is expected that the optimization of the quantification of these parameters can lead to significant benefits to the overall quality of the simulation of atmospheric boundary layer processes over complex terrain.

Appendix

4.4.1 vertical sensitivity

4.4.1.1 second level

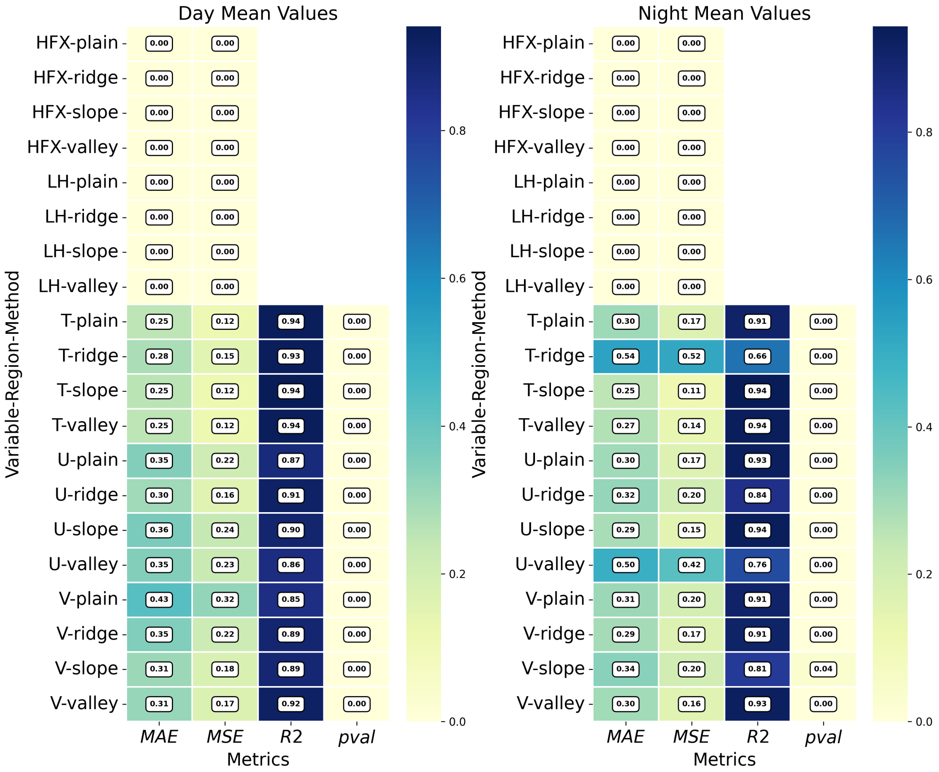


Figure 4.17: Performance metrics averaged over all the algorithms used in the present study for each variable and points analyzed at the second vertical level. The heatmaps show MAE, MSE, R^2, and pvalue for daytime (left, 06:00-17:00) and nighttime (right, 18:00-05:00).

4.4.1.2 third level

4.4.1.3 fourth level

4.4.1.4 fifth level

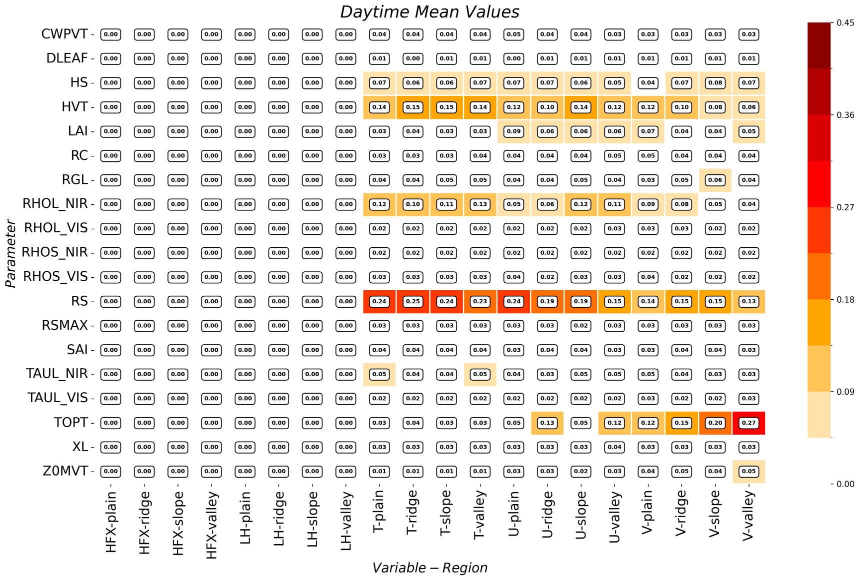


Figure 4.18: Parameters' importance during daytime (06:00 - 17:00 UTC) computed for the second model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

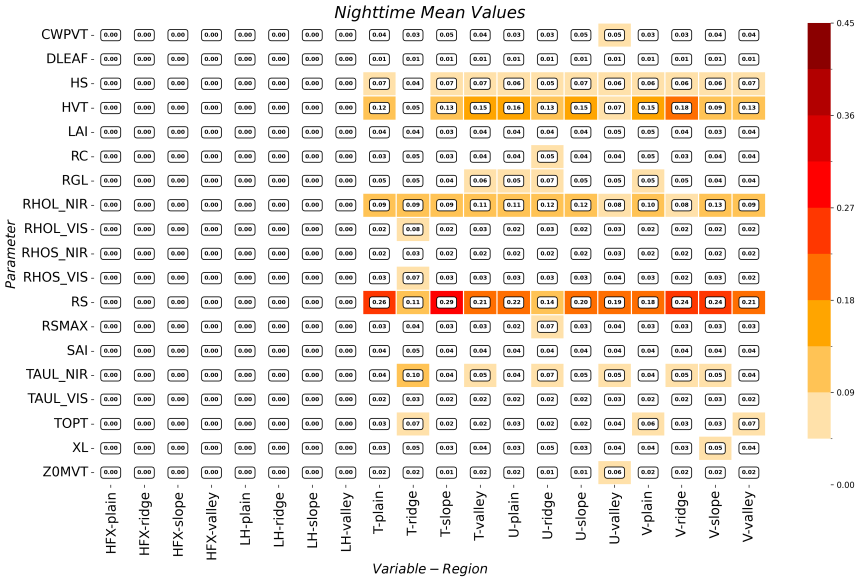


Figure 4.19: Parameters' importance during nighttime (18:00 - 05:00 UTC) computed for the second model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

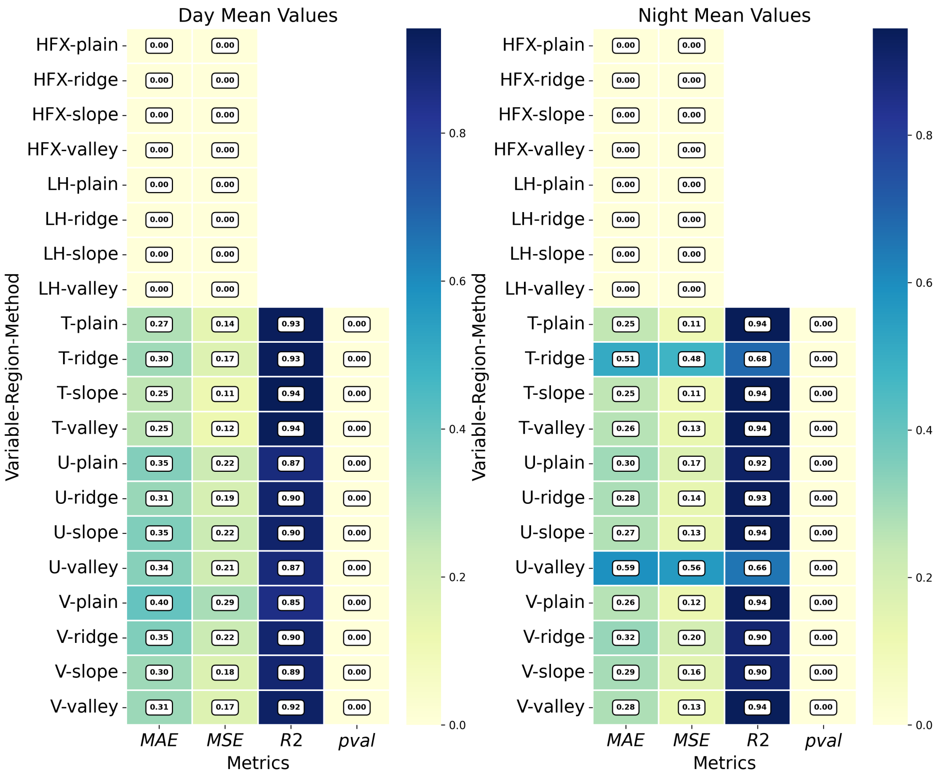


Figure 4.20: Performance metrics averaged over all the algorithms used in the present study for each variable and points analyzed at the third vertical level. The heatmaps show MAE, MSE, R^2, and pvalue for daytime (left, 06:00-17:00) and nighttime (right, 18:00-05:00).

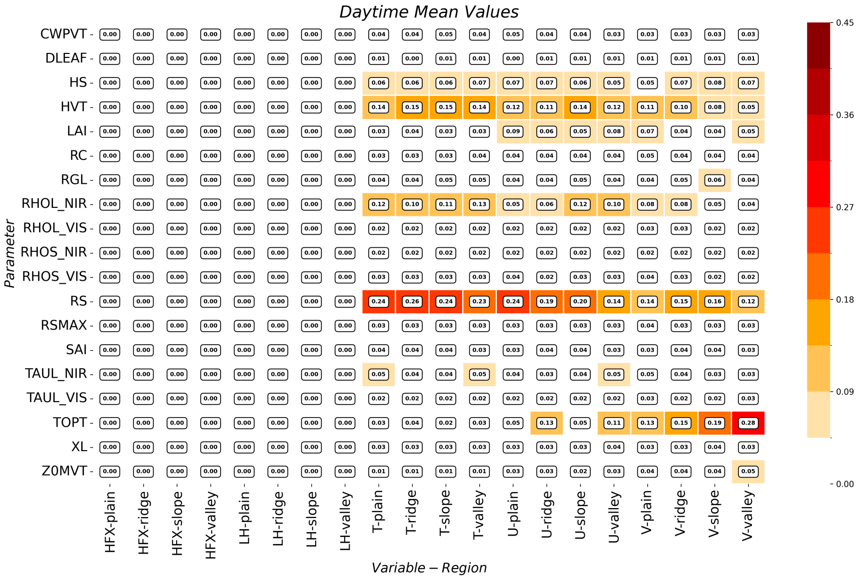


Figure 4.21: Parameters' importance during daytime (06:00 - 17:00 UTC) computed for the third model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

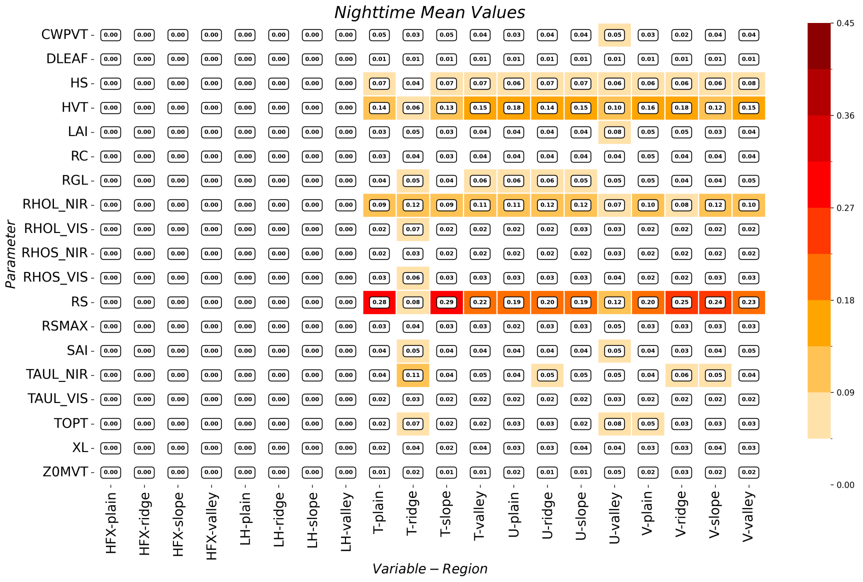


Figure 4.22: Parameters' importance during nighttime (18:00 - 05:00 UTC) computed for the third model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

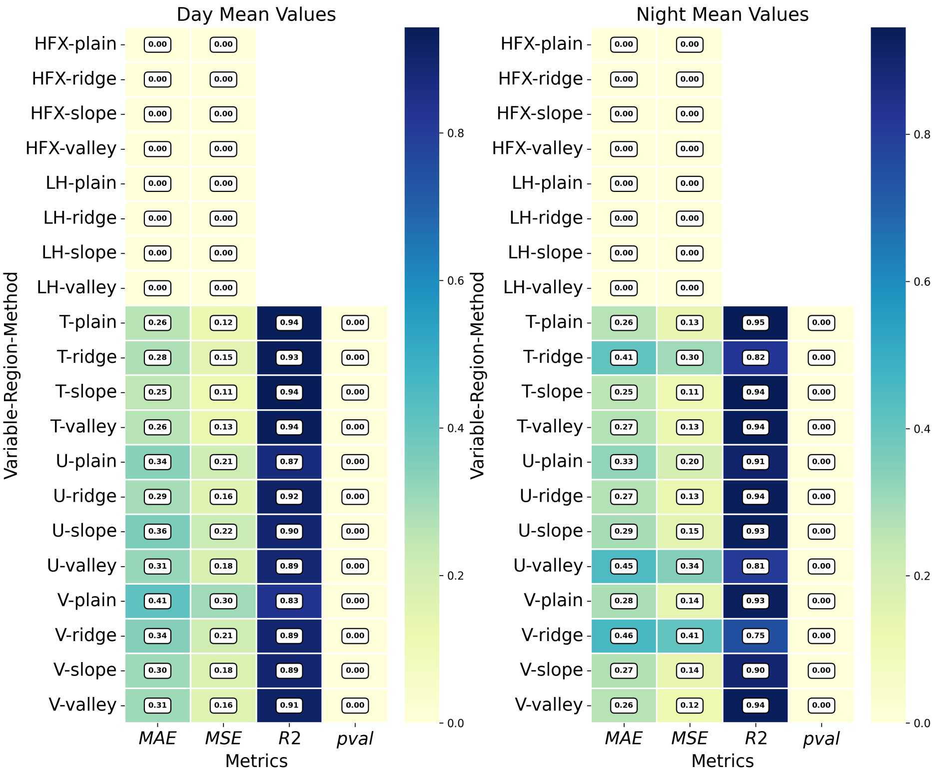


Figure 4.23: Performance metrics averaged over all the algorithms used in the present study for each variable and points analyzed at the fourth vertical level. The heatmaps show MAE, MSE, R^2, and pvalue for daytime (left, 06:00-17:00) and nighttime (right, 18:00-05:00).

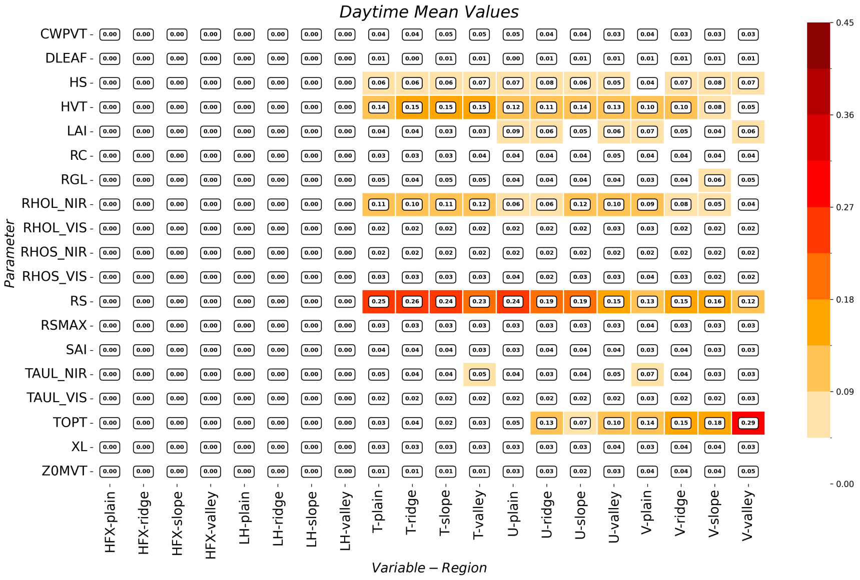


Figure 4.24: Parameters' importance during daytime (06:00 - 17:00 UTC) computed for the fourth model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

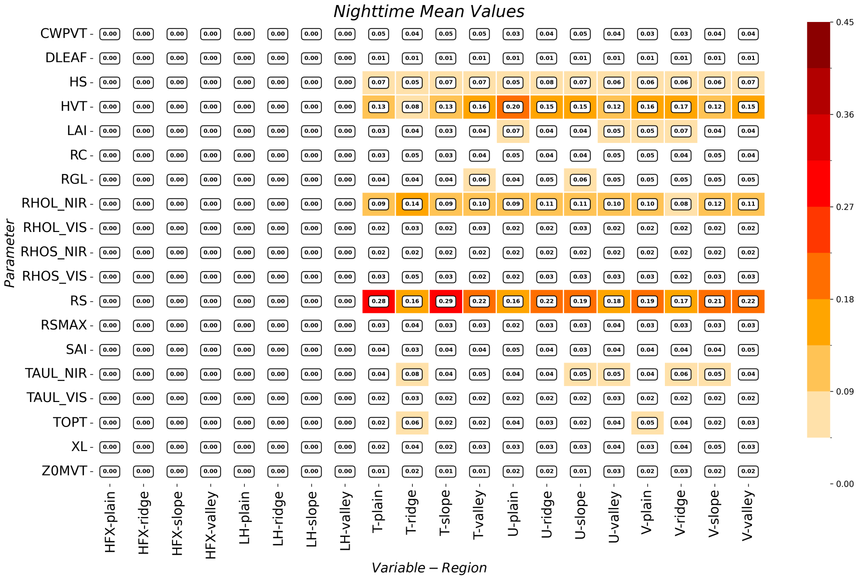


Figure 4.25: Parameters' importance during nighttime (18:00 - 05:00 UTC) computed for the fourth model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

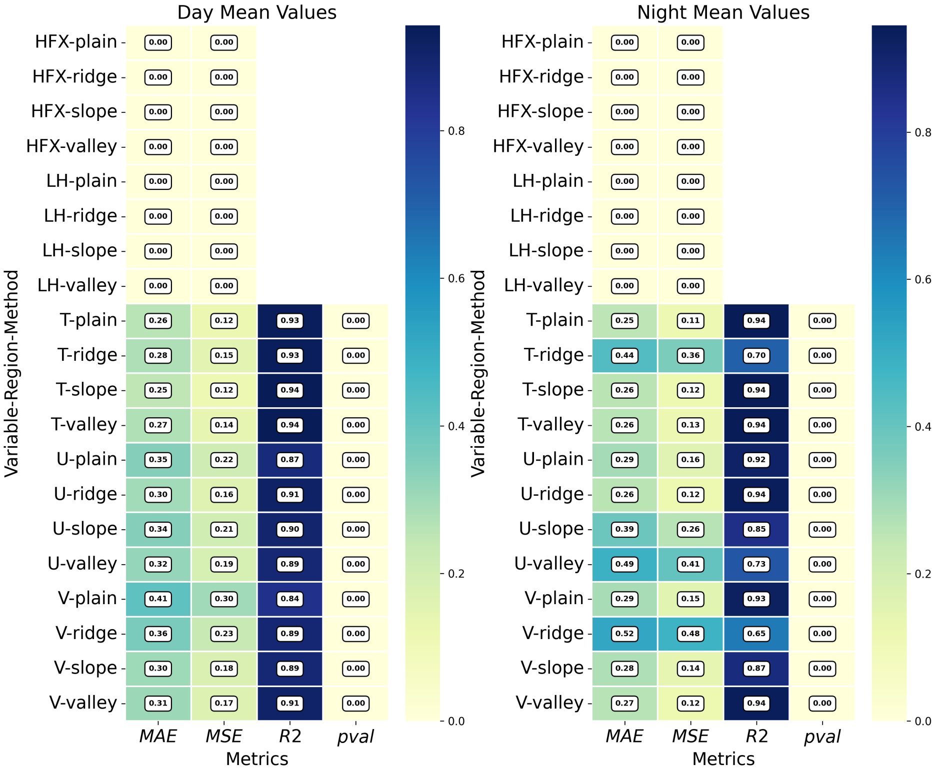


Figure 4.26: Performance metrics averaged over all the algorithms used in the present study for each variable and points analyzed at the fifth vertical level. The heatmaps show MAE, MSE, R^2, and pvalue for daytime (left, 06:00-17:00) and nighttime (right, 18:00-05:00).

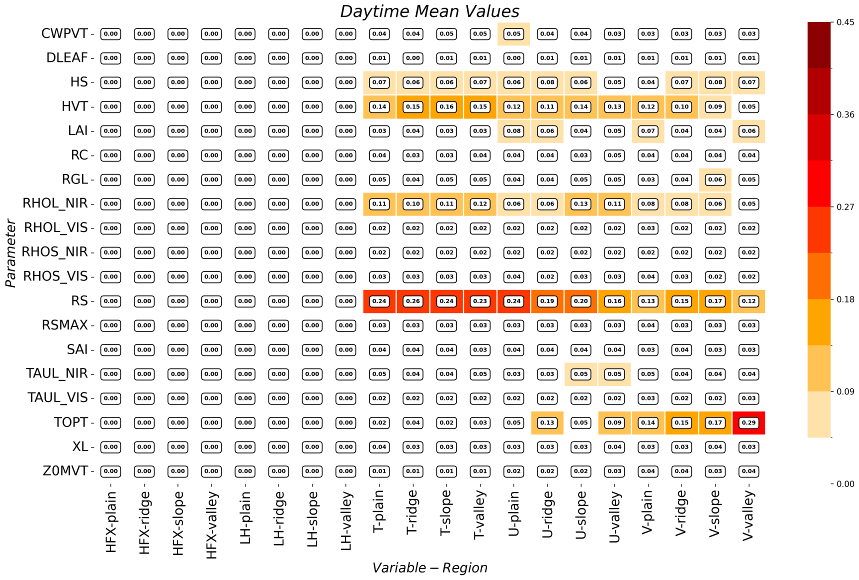


Figure 4.27: Parameters' importance during daytime (06:00 - 17:00 UTC) computed for the fifth model level as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

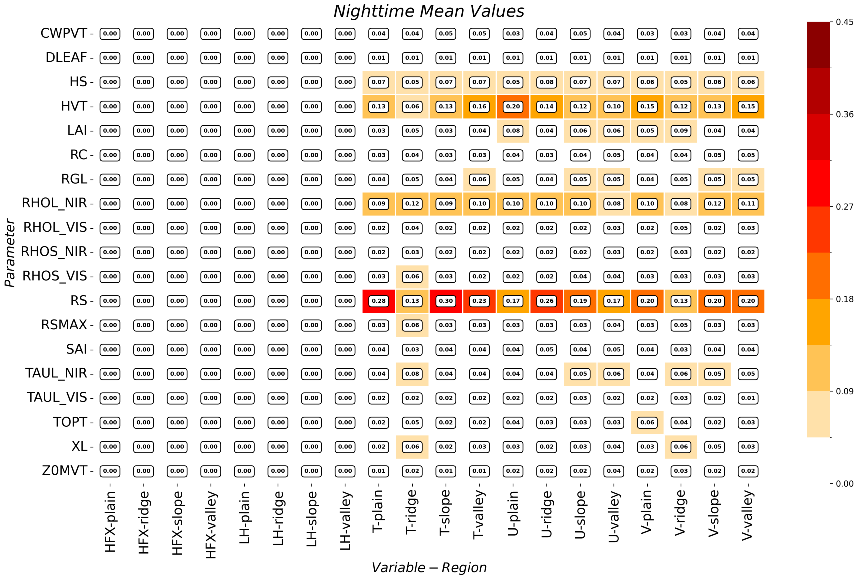


Figure 4.28: Parameters' importance during nighttime (18:00 - 05:00 UTC) computed for the fifth modellevel as an average of the five methods chosen for this case study: SVM, RF, XGB, and GPR.

Conclusions

This thesis proposed and tested a comprehensive methodology to perform sensitivity analyses of the results of simulations with a numerical weather prediction (NWP) model to land surface model (LSM) parameters, with a particular emphasis on atmospheric processes over complex terrain.

The advanced capabilities of Noah-MP, with its unique treatment of canopy processes and its extensive parametric ensemble, were used to analyze the variability in the output of an atmospheric host model to specific land surface input parameters. Such variability can be a problematic source of uncertainty when, for instance, parameters do not accurately reflect the heterogeneity of natural landscapes, especially in the highly variable environments typical of complex terrain regions.

The problem of quantifying and classifying model sensitivity to a set of input parameters has been addressed through the development of a machine learning-based automated multi-method parameter sensitivity and importance analysis tool (ML-AMPSIT). The new tool has been thoroughly leveraged to deal with the problem of high dimensionality, which often makes sensitivity analysis infeasible. ML-AMPSIT utilizes an array of machine learning algorithms, including LASSO, Support Vector Machine, Classification and Regression Trees, Random Forest, Extreme Gradient Boosting, Gaussian Process Regression, and Bayesian Ridge Regression, to create surrogate models. It employs multiple strategies to prevent overfitting, including cross-validation and Bayesian optimization for hyperparameter tuning.

ML-AMPSIT addresses the computational challenges associated with traditional global sensitivity analysis methods, like the Sobol method, which requires numerous model realizations to perform effectively. By leveraging machine learning algorithms to construct surrogate models, ML-AMPSIT significantly reduces the number of simulations needed for robust sensitivity analysis.

ML-AMPSIT performance is first evaluated in a case study involving simulations of a sea breeze circulation over an idealized flat domain with the Weather Research and Forecasting (WRF) model coupled with the Noah-MP land surface model. It was shown that ML-AMPSIT was able to efficiently predict the sensitivity of selected Noah-MP model parameters with a relatively small number of model runs. Convergence analysis revealed that the system response stabilizes after about 20 simulations for several methods, while the ensemble of methods generally agrees on the importance rankings of the parameters after 80 simulations. This stabilization suggests that the obtained feature importance reliably reflects the underlying dynamics of the system, capturing essential features with fewer simulations than typically expected. For the specified setup, ML-AMPSIT identified near-infrared leaf reflectivity and leaf area index as the most consistently influential parameters among the 6 parameters analyzed, while momentum roughness length was found to be critical in the transition time windows of sunrise and sunset.

Then, ML-AMPSIT was applied to evaluate the sensitivity to surface parameters of WRF/Noah-MP simulations of thermally-driven circulations in a mountain valley, characterized by the development of slope and valley winds due to differential heating. The study provided insights into the response of the WRF model to an extensive set of Noah-MP input parameters. Across 19 parameters analyzed, the results highlighted that the stomatal resistance, the canopy height, and the near-infrared leaf reflectivity are critical in modulating both diurnal and nocturnal atmospheric boundary layer processes. In particular, the stomata resistance emerged as the most influential parameter across all scenarios. The leaf area index, the optimal transpiration air temperature, the vapor pressure deficit parameter, the canopy wind extinction parameter, and the momentum roughness length also showed significant effects on the model outputs. The study demonstrates that parameter sensitivity is highly variable depending on the time of the day and location, emphasizing the complexity of accurately modeling land-atmosphere interactions. These findings suggest that special attention should be paid to these parameters when configuring LSMs for NWP models, especially over mountainous terrain.

The importance of sensitivity analysis in meteorological models cannot be overstated. It contributes to quantifying uncertainty in models that potentially impact decision-making processes related to public safety and resource management. Currently, there is a noticeable lack in the application of modern sensitivity methods in the field of meteorology, although it is well known that the chaotic nature of the atmospheric dynamics as well as the interactions between a large number of parameterizations lead to models that are highly sensitive to input variations, which are also characterized by a very high dimensionality. Therefore, uncertainty quantification still represents a crucial issue in improving atmospheric model reliability. ML-AMPSIT has been developed to fill this gap by providing a new toolkit for sensitivity analysis.

ML-AMPSIT offers several avenues for expansion and improvement. For example, the implementation of a neural network architecture, due to its high scalability potential, could extend the tool's ability to provide deeper insights and more robust feature recognition capabilities. Moreover, expanding the array of available methods to the class of physics-informed machine learning algorithms could significantly improve the interpretability of results. Furthermore, the integration of non-ML data-driven methods, such as data assimilation techniques or reduced-order models (ROMs), could greatly expand the capabilities of the tool. While ML-based algorithms are designed to emulate the data set without considering the underlying physics, ROM methods would be able to synthesize the principal components of the dynamics inherent in the input-output relationships. A mixture of these two approaches could offer the benefits of both, resulting in speed, accuracy, and physical interpretability.

The inclusion of additional sensitivity analysis methods beyond the Sobol method would also be valuable. The presence of multiple techniques such as the Morris method, FAST (Fourier Amplitude Sensitivity Testing), and other variance-based methods could provide a more comprehensive toolkit, consistent with the multi-method purpose of ML-AMPSIT, for understanding parameter influences and interactions.

In the context of sensitivity analysis, ML-AMPSIT could be used in the future to address other parameterization-induced problems, i.e. when the uncertainty comes from the parameterization employed; within atmospheric models these include land-surface interactions, turbulence closure, microphysics, convection, and more. ML-AMPSIT could be employed in the future not only to quantify the importance across input parameters, but also to evaluate the impact of multiple simulation setups, e.g. physical schemes, sub-processes, land covers, numerical strategies, and geometrical configurations.

In addition, the methodologies employed in ML-AMPSIT are not exclusively confined to applications within LSMs but are, in principle, adaptable to any dataset of input-output pairs, regardless of the nature of these data. Therefore there is potential for ML-AMPSIT to extend its application domain to any task that benefits from regression, such as spatial and temporal interpolation, which are critical in operational meteorology. By using data-driven modeling, these tasks can be performed more quickly and with potentially less data. Similarly, applying the tool to the prediction of observables could streamline the forecasting process, making forecasts more efficient and accessible. Lastly, since understanding relationships and patterns within data sets can be meaningful in any scientific or statistical field, the applicability of a data-driven tool like ML-AMPSIT potentially extends far beyond the specific applications that could be enumerated here.

Bibliography

Alqahtani, Abdulaziz, Muhammad Izhar Shah, Ali Aldrees, and Muhammad Faisal Javed (2022)
"Comparative Assessment of Individual and Ensemble Machine Learning Models for Efficient Analysis of River Water Quality". In: *Sustainability* 14.3. ISSN: 2071-1050. DOI: `10.3390/su14031183`. URL: `https://www.mdpi.com/2071-1050/14/3/1183`.

Antoniadis, Anestis, Sophie Lambert-Lacroix, and Jean-Michel Poggi (2021)
"Random forests for global sensitivity analysis: A selective review". In: *Reliability Engineering System Safety* 206, p. 107312. ISSN: 0951-8320. DOI: `https://doi.org/10.1016/j.ress.2020.107312`. URL: `https://www.sciencedirect.com/science/article/pii/S0951832020308073`.

Antonogeorgos, George, Demosthenes B. Panagiotakos, Kostas N. Priftis, and Anastasia Tzonou (Mar. 2009)
"Logistic Regression and Linear Discriminant Analyses in Evaluating Factors Associated with Asthma Prevalence among 10- to 12-Years-Old Children: Divergence and Similarity of the Two Statistical Methods". In: *International Journal of Pediatrics* 2009, p. 952042. ISSN: 1687-9740. DOI: `10.1155/2009/952042`. URL: `https://doi.org/10.1155/2009/952042`.

Arakawa, Akio and Vivian R. Lamb (1977)
"Computational Design of the Basic Dynamical Processes of the UCLA General Circulation Model". In: URL: `https://api.semanticscholar.org/CorpusID:118807126`.

Arduini, Gabriele, Chantal Staquet, and Charles Chemel (Oct. 2016)
"Interactions Between the Nighttime Valley-Wind System and a Developing Cold-Air Pool". In: *Boundary-Layer Meteorology* 161, pp. 49–72. DOI: `10.1007/s10546-016-0155-8`.

Arpaci, A., B. Malowerschnig, O. Sass, and H. Vacik (2014)
"Using multi variate data mining techniques for estimating fire susceptibility of Tyrolean forests". In: *Applied Geography* 53, pp. 258–270. ISSN: 0143-6228. DOI: `https://doi.org/10.1016/j.apgeog.2014.05.015`. URL: `https://www.sciencedirect.com/science/article/pii/S0143622814001106`.

Arsenault, Kristi R., Grey S. Nearing, Shugong Wang, Soni Yatheendradas, and Christa D. Peters-Lidard (2018)
"Parameter Sensitivity of the Noah-MP Land Surface Model with Dynamic Vegetation". In: *Journal of Hydrometeorology* 19.5, pp. 815–830. DOI: `10.1175/`

jhm-d-17-0205.1. URL: https://journals.ametsoc.org/view/journals/hydr/19/5/jhm-d-17-0205_1.xml.

Baki, H., S. Chinta, C Balaji, and B. Srinivasan (2022)
"Determining the sensitive parameters of the Weather Research and Forecasting (WRF) model for the simulation of tropical cyclones in the Bay of Bengal using global sensitivity analysis and machine learning". In: *Geoscientific Model Development* 15.5, pp. 2133–2155. DOI: 10.5194/gmd-15-2133-2022. URL: https://gmd.copernicus.org/articles/15/2133/2022/.

Ball, J. Timothy, Ian E. Woodrow, and Joseph A. Berry (1987a)
"A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions". In: *Progress in Photosynthesis Research: Volume 4 Proceedings of the VIIth International Congress on Photosynthesis Providence, Rhode Island, USA, August 10–15, 1986*. Ed. by J. Biggins. Dordrecht: Springer Netherlands, pp. 221–224. ISBN: 978-94-017-0519-6. DOI: 10.1007/978-94-017-0519-6_48. URL: https://doi.org/10.1007/978-94-017-0519-6_48.

Ball, John, Ian Woodrow, and Joseph Berry (Jan. 1987b)
"A Model Predicting Stomatal Conductance and Its Contribution to the Control of Photosynthesis Under Different Environmental Conditions". In: *Progress in Photosynthesis Research* 4, pp. 221–224. DOI: 10.1007/978-94-017-0519-6_48.

Bar Massada, Avi, Alexandra D. Syphard, Susan I. Stewart, and Volker C. Radeloff (2013)
"Wildfire ignition-distribution modelling: a comparative study in the Huron–Manistee National Forest, Michigan, USA". In: *International Journal of Wildland Fire* 22.2, pp. 174–183. DOI: 10.1071/WF11178. URL: https://doi.org/10.1071/WF11178.

Barlage, Michael, Fei Chen, Mukul Tewari, Kyoko Ikeda, David Gochis, Jimy Dudhia, Roy Rasmussen, Ben Livneh, Mike Ek, and Ken Mitchell (2010)
"Noah land surface model modifications to improve snowpack prediction in the Colorado Rocky Mountains". In: *Journal of Geophysical Research: Atmospheres* 115.D22. DOI: https://doi.org/10.1029/2009JD013470. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD013470. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD013470.

Beljaars, Anton C. M. (1995)
"The parametrization of surface fluxes in large-scale models under free convection". In: *Quarterly Journal of the Royal Meteorological Society* 121.522, pp. 255–270. DOI: https://doi.org/10.1002/qj.49712152203. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49712152203.

URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49712152203`.

Blackadar, Alfred K. (1962)
"The vertical distribution of wind and turbulent exchange in a neutral atmosphere". In: *Journal of Geophysical Research (1896-1977)* 67.8, pp. 3095–3102. DOI: `https://doi.org/10.1029/JZ067i008p03095`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/JZ067i008p03095`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JZ067i008p03095`.

Bocquet, Marc, Julien Brajard, Alberto Carrassi, and Laurent Bertino (2020)
"Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization". In: *Foundations of Data Science* 2.1, pp. 55–80. DOI: `10.3934/fods.2020004`. URL: `/article/id/eb1e67dd-ccb9-4ab2-8526-15c23feae415`.

Bonan, Bertrand (1996)
"A Land Surface Model (LSM Version 1.0) for Ecological, Hydrological, and Atmospheric Studies: Technical Description and User's Guide". In: URL: `https://api.semanticscholar.org/CorpusID:130358464`.

Bonan, Gordon (Jan. 2008)
Ecological Climatology: Concepts and Applications, p. 550. ISBN: 9780521872218. DOI: `10.1017/CBO9780511805530`.

Bonavita, Massimo and Patrick Laloyaux (2020)
"Machine Learning for Model Error Inference and Correction". In: *Journal of Advances in Modeling Earth Systems* 12.12. e2020MS002232 10.1029/2020MS002232, e2020MS002232. DOI: `https://doi.org/10.1029/2020MS002232`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002232`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002232`.

Boser, Bernhard E., Isabelle M. Guyon, and Vladimir N. Vapnik (1992)
"A training algorithm for optimal margin classifiers". In: COLT '92, pp. 144–152. DOI: `10.1145/130385.130401`. URL: `https://doi.org/10.1145/130385.130401`.

Box, George E.P. and George C. Tiao (1992). In: *Bayesian Inference in Statistical Analysis*. John Wiley Sons, Ltd. ISBN: 9781118033197. DOI: `https://doi.org/10.1002/9781118033197`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118033197`. URL: `https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118033197`.

Bratley, Paul and Bennett L. Fox (Mar. 1988)
"Algorithm 659: Implementing Sobol's quasirandom sequence generator". In:

ACM Trans. Math. Softw. 14.1, pp. 88–100. ISSN: 0098-3500. DOI: 10.1145/42288.214372. URL: https://doi.org/10.1145/42288.214372.

Breiman, L. (2001)
“Random Forests”. In: *Machine Learning* 45, pp. 5–32. URL: https://api.semanticscholar.org/CorpusID:89141.

Breiman, L., Jerome H. Friedman, Richard A. Olshen, and C. J. Stone (1984)
“Classification and Regression Trees”. In: *Biometrics* 40, p. 874. URL: https://api.semanticscholar.org/CorpusID:29458883.

Brunton, Steven, Bernd Noack, and Petros Koumoutsakos (Jan. 2020)
“Machine Learning for Fluid Mechanics”. In: *Annual Review of Fluid Mechanics* 52. DOI: 10.1146/annurev-fluid-010719-060214.

Brunton, Steven L. and J. Nathan Kutz (2019)
Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control. Cambridge University Press. DOI: 10.1017/9781108380690.

Brutsaert, Wilfried (1982a)
“Evaporation into the atmosphere”. In: URL: https://api.semanticscholar.org/CorpusID:92106959.

— (1982b)
“Evaporation into the atmosphere”. In: URL: https://api.semanticscholar.org/CorpusID:92106959.

Campolongo, Francesca, Jessica Cariboni, and Andrea Saltelli (2007)
“An effective screening design for sensitivity analysis of large models”. In: *Environmental Modelling Software* 22.10. Modelling, computer-assisted simulations, and mapping of dangerous phenomena for hazard assessment, pp. 1509–1518. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2006.10.004. URL: https://www.sciencedirect.com/science/article/pii/S1364815206002805.

Campolongo, Francesca, Andrea Saltelli, and Jessica Cariboni (2011)
“From screening to quantitative sensitivity analysis. A unified approach”. In: *Computer Physics Communications* 182.4, pp. 978–988. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2010.12.039. URL: https://www.sciencedirect.com/science/article/pii/S0010465510005321.

Catani, F., D. Lagomarsino, S. Segoni, and V. Tofani (2013)
“Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues”. In: *Natural Hazards and Earth System Sciences* 13.11, pp. 2815–2831. DOI: 10.5194/nhess-13-2815-2013. URL: https://nhess.copernicus.org/articles/13/2815/2013/.

Chase, Randy J., David R. Harrison, Amanda Burke, Gary M. Lackmann, and Amy McGovern (2022)
“A Machine Learning Tutorial for Operational Meteorology. Part I: Traditional

Machine Learning". In: *Weather and Forecasting* 37.8, pp. 1509–1529. DOI: `https://doi.org/10.1175/WAF-D-22-0070.1`. URL: `https://journals.ametsoc.org/view/journals/wefo/37/8/WAF-D-22-0070.1.xml`.

Chase, Randy J., David R. Harrison, Gary M. Lackmann, and Amy McGovern (2023)
"A Machine Learning Tutorial for Operational Meteorology, Part II: Neural Networks and Deep Learning". In: *Weather and Forecasting*. DOI: `https://doi.org/10.1175/WAF-D-22-0187.1`. URL: `https://journals.ametsoc.org/view/journals/wefo/aop/WAF-D-22-0187.1/WAF-D-22-0187.1.xml`.

Chaure, Anmol and Sudip Bhattacharya (May 2023)
"Gaussian Process Regression for Climate Modeling: Potentials, Limitations, and Advances in Emulation Techniques". In: DOI: `10.1729/Journal.34007`.

Chen, Fei, Zavisa I. Janjic, and Kenneth E. Mitchell (1997)
"Impact of Atmospheric Surface-layer Parameterizations in the new Land-surface Scheme of the NCEP Mesoscale Eta Model". In: *Boundary-Layer Meteorology* 85, pp. 391–421. URL: `https://api.semanticscholar.org/CorpusID:122919976`.

Chen, Rui, Weimin Zhang, and Xiang Wang (2020)
"Machine Learning in Tropical Cyclone Forecast Modeling: A Review". In: *Atmosphere* 11.7. ISSN: 2073-4433. DOI: `10.3390/atmos11070676`. URL: `https://www.mdpi.com/2073-4433/11/7/676`.

Chen, Tao, Li Zhu, Rui-qing Niu, C. John Trinder, Ling Peng, and Tao Lei (Mar. 2020)
"Mapping landslide susceptibility at the Three Gorges Reservoir, China, using gradient boosting decision tree, random forest and information value models". In: *Journal of Mountain Science* 17.3, pp. 670–685. ISSN: 1993-0321. DOI: `10.1007/s11629-019-5839-3`. URL: `https://doi.org/10.1007/s11629-019-5839-3`.

Chen, Tianqi and Carlos Guestrin (Aug. 2016)
"XGBoost: A Scalable Tree Boosting System". In: KDD '16. DOI: `10.1145/2939672.2939785`. URL: `http://dx.doi.org/10.1145/2939672.2939785`.

Ching, J., R. Rotunno, M. LeMone, A. Martilli, B. Kosovic, P. A. Jimenez, and J. Dudhia (2014)
"Convectively Induced Secondary Circulations in Fine-Grid Mesoscale Numerical Weather Prediction Models". In: *Monthly Weather Review* 142.9, pp. 3284–3302. DOI: `10.1175/MWR-D-13-00318.1`. URL: `https://journals.ametsoc.org/view/journals/mwre/142/9/mwr-d-13-00318.1.xml`.

Cinquegrana, Davide, Alessandra Lucia Zollo, Myriam Montesarchio, and Edoardo Bucchignani (2023)
"A Metamodel-Based Optimization of Physical Parameters of High Resolution

NWP ICON-LAM over Southern Italy". In: *Atmosphere* 14.5. ISSN: 2073-4433. DOI: `10.3390/atmos14050788`. URL: `https://www.mdpi.com/2073-4433/14/5/788`.

Clark, Martyn P., Dmitri Kavetski, and Fabrizio Fenicia (2011)
"Pursuing the method of multiple working hypotheses for hydrological modeling". In: *Water Resources Research* 47.9. DOI: `https://doi.org/10.1029/2010WR009827`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010WR009827`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR009827`.

Collatz, Gj, Miquel Ribas-Carbo, and Joseph Berry (Jan. 1992)
"Coupled Photosynthesis-Stomatal Conductance Model for Leaves of C 4 Plants". In: *Functional Plant Biology* 19. DOI: `10.1071/PP9920519`.

Cortes, Corinna and Vladimir Naumovich Vapnik (1995)
"Support-Vector Networks". In: *Machine Learning* 20, pp. 273–297. URL: `https://api.semanticscholar.org/CorpusID:52874011`.

Couvreux, Fleur, Frédéric Hourdin, Daniel Williamson, Romain Roehrig, Victoria Volodina, Najda Villefranque, Catherine Rio, Olivier Audouin, James Salter, Eric Bazile, Florent Brient, Florence Favot, Rachel Honnert, Marie-Pierre Lefebvre, Jean-Baptiste Madeleine, Quentin Rodier, and Wenzhe Xu (2021)
"Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement". In: *Journal of Advances in Modeling Earth Systems* 13.3. e2020MS002217 2020MS002217, e2020MS002217. DOI: `https://doi.org/10.1029/2020MS002217`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002217`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002217`.

Cui, Caihao and Dianhui Wang (2016)
"High dimensional data regression using Lasso model and neural networks with random weights". In: *Information Sciences* 372, pp. 505–517. ISSN: 0020-0255. DOI: `https://doi.org/10.1016/j.ins.2016.08.060`. URL: `https://www.sciencedirect.com/science/article/pii/S0020025516306314`.

Cuntz, Matthias, Juliane Mai, Luis Samaniego, Martyn Clark, Volker Wulfmeyer, Oliver Branch, Sabine Attinger, and Stephan Thober (2016)
"The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model". In: *Journal of Geophysical Research: Atmospheres* 121.18, pp. 10, 676–10, 700. DOI: `https://doi.org/10.1002/2016JD025097`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016JD025097`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JD025097`.

Cuntz, Matthias, Juliane Mai, Matthias Zink, Stephan Thober, Rohini Kumar, David Schäfer, Martin Schrön, John Craven, Oldrich Rakovec, Diana Spieler,

Vladyslav Prykhodko, Giovanni Dalmasso, Jude Musuuza, Ben Langenberg, Sabine Attinger, and Luis Samaniego (2015)
"Computationally inexpensive identification of noninformative model parameters by sequential screening". In: *Water Resources Research* 51.8, pp. 6417–6441. DOI: `https://doi.org/10.1002/2015WR016907`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2015WR016907`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015WR016907`.

Dagon, Katherine, Benjamin Sanderson, Fisher Rosie, and David Lawrence (Dec. 2020)
"A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5". In: *Advances in Statistical Climatology Meteorology and Oceanography* 6, pp. 223–244. DOI: `10.5194/ascmo-6-223-2020`.

Daviran, M., M. Shamekhi, R. Ghezelbash, and A. Maghsoudi (Jan. 2023)
"Landslide susceptibility prediction using artificial neural networks, SVMs and random forest: hyperparameters tuning by genetic optimization algorithm". In: *International Journal of Environmental Science and Technology* 20.1, pp. 259–276. ISSN: 1735-2630. DOI: `10.1007/s13762-022-04491-3`. URL: `https://doi.org/10.1007/s13762-022-04491-3`.

De Oliveira, Rodrigo, Ramon Cristian Fernandes Araújo, Fabrício Barros, Adriano Segundo, Ronaldo Zampolo, Wellington Fonseca, Victor Dmitriev, and Fernando Brasil (Sept. 2017)
"A System Based on Artificial Neural Networks for Automatic Classification of Hydro-generator Stator Windings Partial Discharges". In: *Journal of Microwaves, Optoelectronics and Electromagnetic Applications* 16, pp. 628–645. DOI: `10.1590/2179-10742017v16i3854`.

Deardorff, J. W. (1978)
"Efficient prediction of ground surface temperature and moisture, with inclusion of a layer of vegetation". In: *Journal of Geophysical Research: Oceans* 83.C4, pp. 1889–1903. DOI: `https://doi.org/10.1029/JC083iC04p01889`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/JC083iC04p01889`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JC083iC04p01889`.

Defant, Friedrich (1951)
"Local Winds". In: *Compendium of Meteorology: Prepared under the Direction of the Committee on the Compendium of Meteorology*. Ed. by Thomas F. Malone. Boston, MA: American Meteorological Society, pp. 655–672. ISBN: 978-1-940033-70-9. DOI: `10.1007/978-1-940033-70-9_54`. URL: `https://doi.org/10.1007/978-1-940033-70-9_54`.

Dey, Aiendrila, Debi Prasad Sahoo, Rohini Kumar, and Renji Remesan (2022) “A multimodel ensemble machine learning approach for CMIP6 climate model projections in an Indian River basin”. In: *International Journal of Climatology* 42.16, pp. 9215–9236. DOI: `https://doi.org/10.1002/joc.7813`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.7813`. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7813`.

Di Santo, D., C. He, F. Chen, and L. Giovannini (2024) “ML-AMPSIT: Machine Learning-based Automated Multi-method Parameter Sensitivity and Importance analysis Tool”. In: *Geoscientific Model Development Discussions* 2024, pp. 1–37. DOI: `10.5194/gmd-2024-56`. URL: `https://gmd.copernicus.org/preprints/gmd-2024-56/`.

Di Santo, Dario (Mar. 2024) *ML-AMPSIT*. Version v1.0.0. DOI: `10.5281/zenodo.10789930`. URL: `https://doi.org/10.5281/zenodo.10789930`.

Dickinson, E., A. Henderson-sellers, and John J. Kennedy (1993) “Biosphere-atmosphere transfer scheme(BATS) version 1e as coupled to the NCAR community climate model”. In: URL: `https://api.semanticscholar.org/CorpusID:128818661`.

Dickinson, E., A. Henderson-sellers, John J. Kennedy, and Fred Wilson (1986) “Biosphere-atmosphere Transfer Scheme (BATS) for the NCAR Community Climate Model”. In: *CTIT technical reports series*. URL: `https://api.semanticscholar.org/CorpusID:129114679`.

Dudhia, Jimy (1989) “Numerical Study of Convection Observed during the Winter Monsoon Experiment Using a Mesoscale Two-Dimensional Model”. In: *Journal of Atmospheric Sciences* 46.20, pp. 3077–3107. DOI: `10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/atsc/46/20/1520-0469_1989_046_3077_nsocod_2_0_co_2.xml`.

Ek, M. B., K. E. Mitchell, Y. Lin, E. Rogers, P. Grunmann, V. Koren, G. Gayno, and J. D. Tarpley (2003) “Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model”. In: *Journal of Geophysical Research: Atmospheres* 108.D22. DOI: `https://doi.org/10.1029/2002JD003296`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD003296`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD003296`.

“Elementary Effects Method” (2007). In: *Global Sensitivity Analysis. The Primer*. John Wiley Sons, Ltd. Chap. 3, pp. 109–154. ISBN: 9780470725184. DOI: `https://doi.org/10.1002/9780470725184.ch3`. eprint: `https://onlinelibrary.`

wiley.com/doi/pdf/10.1002/9780470725184.ch3. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470725184.ch3.

Elia, Letizia, Silvia Castellaro, Ashok Dahal, and Luigi Lombardo (July 2023) "Assessing multi-hazard susceptibility to cryospheric hazards: Lesson learnt from an Alaskan example". In: *Science of The Total Environment* 898, p. 165289. DOI: 10.1016/j.scitotenv.2023.165289.

Engelbrecht, A. P., I. Cloete, and J. M. Zurada (1995) *Determining the significance of input parameters using sensitivity analysis*. Undetermined. DOI: 10.1007/3-540-59497-3_199.

Farooq, Furqan, Muhammad Nasir Amin, Kaffayatullah Khan, Muhammad Rehan Sadiq, Muhammad Faisal Javed, Fahid Aslam, and Rayed Alyousef (2020) "A Comparative Study of Random Forest and Genetic Engineering Programming for the Prediction of Compressive Strength of High Strength Concrete (HSC)". In: *Applied Sciences* 10.20. ISSN: 2076-3417. DOI: 10.3390/app10207330. URL: https://www.mdpi.com/2076-3417/10/20/7330.

Farquhar, Graham, Susanne von Caemmerer, and Joseph Berry (June 1980) "A biochemical model of photosynthetic CO". In: *Planta* 149, pp. 78–90. DOI: 10.1007/BF00386231.

Fernández-Godino, M. Giselle, Chanyoung Park, Nam Kim, and Raphael Haftka (Mar. 2017) "Review of multi-fidelity models". In: *Advances in Computational Science and Engineering* 1, pp. 351–400. DOI: 10.3934/acse.2023015.

Fishman, G. (1996) *Monte Carlo: Concepts, Algorithms, and Applications*. Springer Series in Operations Research and Financial Engineering. Springer New York. ISBN: 9781475725537. URL: https://books.google.it/books?id=dogHCAAAQBAJ.

Forrester, Alexander, Andras Sobester, and Andy Keane (July 2008) *Engineering Design Via Surrogate Modelling: A Practical Guide*. ISBN: 978-0-470-06068-1. DOI: 10.1002/9780470770801.

Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007) "Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling". In: *International Journal of Climatology* 27.12, pp. 1547–1578. DOI: https://doi.org/10.1002/joc.1556. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1556. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1556.

Gan, Yanjun, Qingyun Duan, Wei Gong, Charles Tong, Yunwei Sun, Wei Chu, Aizhong Ye, Chiyuan Miao, and Zhenhua Di (2014) "A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model". In: *Environmental Modelling Software* 51, pp. 269–285. ISSN: 1364-8152. DOI: `https://doi.org/10.1016/j.envsoft.2013.09.031`. URL: `https://www.sciencedirect.com/science/article/pii/S1364815213002338`.

Gholampour, Aliakbar, Amir H. Gandomi, and Togay Ozbakkaloglu (2017) "New formulations for mechanical properties of recycled aggregate concrete using gene expression programming". In: *Construction and Building Materials* 130, pp. 122–145. ISSN: 0950-0618. DOI: `https://doi.org/10.1016/j.conbuildmat.2016.10.114`. URL: `https://www.sciencedirect.com/science/article/pii/S0950061816317408`.

Gigović, Ljubomir, Hamid Reza Pourghasemi, Siniša Drobnjak, and Shibiao Bai (2019) "Testing a New Ensemble Model Based on SVM and Random Forest in Forest Fire Susceptibility Assessment and Its Mapping in Serbia's Tara National Park". In: *Forests* 10.5. ISSN: 1999-4907. DOI: `10.3390/f10050408`. URL: `https://www.mdpi.com/1999-4907/10/5/408`.

Giovannini, Lorenzo, Lavinia Laiti, Stefano Serafin, and Dino Zardi (2017) "The thermally driven diurnal wind system of the Adige Valley in the Italian Alps". In: *Quarterly Journal of the Royal Meteorological Society* 143.707, pp. 2389–2402. DOI: `https://doi.org/10.1002/qj.3092`. eprint: `https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3092`. URL: `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3092`.

Gong, Wei, Qingyun Duan, Jianduo Li, Chen Wang, Zhenhua di, Aizhong ye, Chiyuan Miao, and Y Dai (July 2015) "An Intercomparison of Sampling Methods for Uncertainty Quantification of Environmental Dynamic Models". In: *Journal of Environmental Informatics* 28. DOI: `10.3808/jei.201500310`.

Griensven, Ann van, Thomas Meixner, Sabine Grunwald, Thomas Bishop, M. Diluzio, and R. Srinivasan (June 2006) "A Global Sensitivity Analysis Tool for the Parameters of Multi-Variable Catchment Models". In: *Journal of Hydrology* 324, pp. 10–23. DOI: `10.1016/j.jhydrol.2005.09.008`.

Grundner, Arthur, Tom Beucler, Pierre Gentine, Fernando Iglesias-Suarez, Marco A. Giorgetta, and Veronika Eyring (2022) "Deep Learning Based Cloud Cover Parameterization for ICON". In: *Journal of Advances in Modeling Earth Systems* 14.12. e2021MS002959 2021MS002959, e2021MS002959. DOI: `https://doi.org/10.1029/2021MS002959`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002959`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002959`.

Haghiabi, Amir Hamzeh, Ali Heidar Nasrolahi, and Abbas Parsaie (Jan. 2018)
"Water quality prediction using machine learning methods". In: *Water Quality Research Journal* 53.1, pp. 3–13. ISSN: 1201-3080. DOI: `10.2166/wqrj.2018.025`. eprint: `https://iwaponline.com/wqrj/article-pdf/53/1/3/224144/wqrjc0530003.pdf`. URL: `https://doi.org/10.2166/wqrj.2018.025`.

Han, Yilun, Guang J. Zhang, Xiaomeng Huang, and Yong Wang (2020)
"A Moist Physics Parameterization Based on Deep Learning". In: *Journal of Advances in Modeling Earth Systems* 12.9. e2020MS002076 2020MS002076, e2020MS002076. DOI: `https://doi.org/10.1029/2020MS002076`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002076`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002076`.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, and James Franklin (Nov. 2004)
"The Elements of Statistical Learning: Data Mining, Inference, and Prediction". In: *Math. Intell.* 27, pp. 83–85. DOI: `10.1007/BF02985802`.

Herman, J. D., J. B. Kollat, P. M. Reed, and T. Wagener (2013)
"Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models". In: *Hydrology and Earth System Sciences* 17.7, pp. 2893–2903. DOI: `10.5194/hess-17-2893-2013`. URL: `https://hess.copernicus.org/articles/17/2893/2013/`.

Hong, Song-You and Hua-Lu Pan (1996)
"Nonlocal Boundary Layer Vertical Diffusion in a Medium-Range Forecast Model". In: *Monthly Weather Review* 124.10, pp. 2322–2339. DOI: `10.1175/1520-0493(1996)124<2322:NBLVDI>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/mwre/124/10/1520-0493_1996_124_2322_nblvdi_2_0_co_2.xml`.

Hung, Ying-Chao (2023)
"A review of Monte Carlo and quasi-Monte Carlo sampling techniques". In: *WIREs Computational Statistics*. URL: `https://api.semanticscholar.org/CorpusID:265182838`.

Iacono, Michael J., Jennifer S. Delamere, Eli J. Mlawer, Mark W. Shephard, Shepard A. Clough, and William D. Collins (2008)
"Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models". In: *Journal of Geophysical Research: Atmospheres* 113.D13. DOI: `https://doi.org/10.1029/2008JD009944`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD009944`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD009944`.

Iooss, Bertrand and Paul Lemaître (Apr. 2014)
“A Review on Global Sensitivity Analysis Methods”. In: *Operations Research/Computer Science Interfaces Series* 59. DOI: `10.1007/978-1-4899-7547-8_5`.

Janjic, Zavisa (Jan. 2002)
“Nonsingular Implementation of the Mellor–Yamada Level 2.5 Scheme in the NCEP Meso Model”. In: *NCEP Office Note* 436.

Janjić, Zaviša I. (1990)
“The Step-Mountain Coordinate: Physical Package”. In: *Monthly Weather Review* 118.7, pp. 1429–1443. DOI: `10.1175/1520-0493(1990)118<1429:TSMCPP>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/mwre/118/7/1520-0493_1990_118_1429_tsmcpp_2_0_co_2.xml`.

Jarvis, P. G., John Lennox Monteith, and Paul Egerton Weatherley (1976)
“The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field”. In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 273.927, pp. 593–610. DOI: `10.1098/rstb.1976.0035`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rstb.1976.0035`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rstb.1976.0035`.

Kalantar, Bahareh, Biswajeet Pradhan, Seyed Amir Naghibi, Alireza Motevalli, and Shattri Mansor (2018)
“Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN)”. In: *Geomatics, Natural Hazards and Risk* 9.1, pp. 49–69. DOI: `10.1080/19475705.2017.1407368`. eprint: `https://doi.org/10.1080/19475705.2017.1407368`. URL: `https://doi.org/10.1080/19475705.2017.1407368`.

Kim, Sun and Fani Boukouvala (June 2020)
“Machine learning-based surrogate modeling for data-driven optimization: a comparison of subset selection for regression techniques”. In: *Optimization Letters* 14. DOI: `10.1007/s11590-019-01428-7`.

Klemp, J. B. and D. K. Lilly (1978)
“Numerical Simulation of Hydrostatic Mountain Waves”. In: *Journal of Atmospheric Sciences* 35.1, pp. 78–107. DOI: `10.1175/1520-0469(1978)035<0078:NSOHMW>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/atsc/35/1/1520-0469_1978_035_0078_nsohmw_2_0_co_2.xml`.

Knievel, Jason C., George H. Bryan, and Joshua P. Hacker (2007)
“Explicit Numerical Diffusion in the WRF Model”. In: *Monthly Weather Review* 135.11, pp. 3808–3824. DOI: `10.1175/2007MWR2100.1`. URL: `https://journals.ametsoc.org/view/journals/mwre/135/11/2007mwr2100.1.xml`.

Kok, Zhi Hong, Abdul Rashid Mohamed Shariff, Meftah Salem M. Alfatni, and Siti Khairunniza-Bejo (2021)
"Support Vector Machine in Precision Agriculture: A review". In: *Computers and Electronics in Agriculture* 191, p. 106546. ISSN: 0168-1699. DOI: `https://doi.org/10.1016/j.compag.2021.106546`. URL: `https://www.sciencedirect.com/science/article/pii/S0168169921005639`.

Krauss, Liane, Christian Hauck, and Christoph Kottmeier (Dec. 2010)
"Spatio-temporal soil moisture variability in Southwest Germany observed with a new monitoring network within the COPS domain". In: *Meteorologische Zeitschrift* 19, pp. 523–537. DOI: `10.1127/0941-2948/2010/0486`.

Lamberti, Giacomo and Catherine Gorlé (July 2021)
"A multi-fidelity machine learning framework to predict wind loads on buildings". In: *Journal of Wind Engineering and Industrial Aerodynamics* 214, p. 104647. DOI: `10.1016/j.jweia.2021.104647`.

Lee, Jung-Hyun, Maher Ibrahim Sameen, Biswajeet Pradhan, and Hyuck-Jin Park (2018)
"Modeling landslide susceptibility in data-scarce environments using optimized data mining and statistical methods". In: *Geomorphology* 303, pp. 284–298. ISSN: 0169-555X. DOI: `https://doi.org/10.1016/j.geomorph.2017.12.007`. URL: `https://www.sciencedirect.com/science/article/pii/S0169555X16311060`.

Lei, Thomas, Stanley Ng, and Shirley Siu (Mar. 2023)
"Application of ANN, XGBoost, and Other ML Methods to Forecast Air Quality in Macau". In: *Sustainability* 15, p. 5341. DOI: `10.3390/su15065341`.

Leinonen, Jussi, Daniele Nerini, and Alexis Berne (2021)
"Stochastic Super-Resolution for Downscaling Time-Evolving Atmospheric Fields With a Generative Adversarial Network". In: *IEEE Transactions on Geoscience and Remote Sensing* 59.9, pp. 7211–7223. DOI: `10.1109/TGRS.2020.3032790`.

Leukauf, Daniel, Alexander Gohm, and Mathias Rotach (Oct. 2016)
"Quantifying horizontal and vertical tracer mass fluxes in an idealized valley during daytime". In: *ATMOSPHERIC CHEMISTRY AND PHYSICS* 16, pp. 13049–13066. DOI: `10.5194/acp-16-13049-2016`.

Leuning, Ray (1990)
"Modelling Stomatal Behaviour and and Photosynthesis of Eucalyptus grandis". In: *Australian Journal of Plant Physiology* 17, pp. 159–175. URL: `https://api.semanticscholar.org/CorpusID:84236513`.

Li, Jianduo, Fei Chen, Xingjie Lu, Wei Gong, Guo Zhang, and Yanjun Gan (2020)
"Quantifying Contributions of Uncertainties in Physical Parameterization Schemes and Model Parameters to Overall Errors in Noah-MP Dynamic Vegetation Mod-

eling". In: *Journal of Advances in Modeling Earth Systems* 12.7. e2019MS001914 2019MS001914, e2019MS001914. DOI: `https://doi.org/10.1029/2019MS001914`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2019MS001914`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001914`.

Li, Jianduo, Fei Chen, Guo Zhang, Michael Barlage, Yanjun Gan, Yufei Xin, and Chen Wang (2018)
"Impacts of Land Cover and Soil Texture Uncertainty on Land Model Simulations Over the Central Tibetan Plateau". In: *Journal of Advances in Modeling Earth Systems* 10.9, pp. 2121–2146. DOI: `https://doi.org/10.1029/2018MS001377`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001377`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001377`.

Li, Jianduo, Guo Zhang, Michael Barlage, Yanjun Gan, Yufei Xin, and Chen Wang (Aug. 2018)
"Impacts of Land Cover and Soil Texture Uncertainty on Land Model Simulations Over the Central Tibetan Plateau". In: *Journal of Advances in Modeling Earth Systems* 10. DOI: `10.1029/2018MS001377`.

Li, Jianduo, Guo Zhang, Fei Chen, Xindong Peng, and Yanjun Gan (2019)
"Evaluation of Land Surface Subprocesses and Their Impacts on Model Performance With Global Flux Data". In: *Journal of Advances in Modeling Earth Systems* 11.5, pp. 1329–1348. DOI: `https://doi.org/10.1029/2018MS001606`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018MS001606`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001606`.

Liu, Yue, Peihua Xu, Chen Cao, Bo Shan, Kuanxing Zhu, Qiuyang Ma, Zongshuo Zhang, and Han Yin (2021)
"A comparative evaluation of machine learning algorithms and an improved optimal model for landslide susceptibility: a case study". In: *Geomatics, Natural Hazards and Risk* 12.1, pp. 1973–2001. DOI: `10.1080/19475705.2021.1955018`. eprint: `https://doi.org/10.1080/19475705.2021.1955018`. URL: `https://doi.org/10.1080/19475705.2021.1955018`.

Longo, Riccardo, Aurelie Bellemans, Marco Derudi, and Alessandro Parente (Nov. 2020)
"A multi-fidelity framework for the estimation of the turbulent Schmidt number in the simulation of atmospheric dispersion". In: *Building and Environment* 185, p. 107066. DOI: `10.1016/j.buildenv.2020.107066`.

Mahrt, Larry and Pan HL (Jan. 1984)
"A 2-layer model of soil hydrology". In: *Boundary-Layer Meteorology* 29, pp. 1–20. DOI: `10.1007/BF00119116`.

Maleki, Heidar, Armin Sorooshian, Gholamreza Goudarzi, Zeynab Baboli, Yaser Tahmasebi Birgani, and Mojtaba Rahmati (Aug. 2019)
"Air pollution prediction by using an artificial neural network model". In: *Clean Technologies and Environmental Policy* 21.6, pp. 1341–1352. ISSN: 1618-9558. DOI: `10.1007/s10098-019-01709-w`. URL: `https://doi.org/10.1007/s10098-019-01709-w`.

Manabe, Syukuro (1969)
"Climate and the ocean circulation: I. the atmospheric circulation and the hydrology of the earth's surface". In: *Monthly Weather Review* 97.11, pp. 739–774. DOI: `10.1175/1520-0493(1969)097<0739:CATOC>2.3.CO;2`. URL: `https://journals.ametsoc.org/view/journals/mwre/97/11/1520-0493_1969_097_0739_catoc_2_3_co_2.xml`.

Mansfield, L. A., Peer Johannes Nowack, Matthew R. Kasoar, Richard G. Everitt, William J. Collins, and Apostolos Voulgarakis (2020)
"Predicting global patterns of long-term climate change from short-term simulations using machine learning". In: *npj Climate and Atmospheric Science* 3, pp. 1–9. URL: `https://api.semanticscholar.org/CorpusID:227033420`.

Maraun, Douglas and Martin Widmann (2018)
Statistical Downscaling and Bias Correction for Climate Research. Cambridge University Press. DOI: `10.1017/9781107588783`.

Maroco, João, Dina Silva, Ana Rodrigues, Manuela Guerreiro, Isabel Santana, and Alexandre de Mendonça (Aug. 2011)
"Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests". In: *BMC Research Notes* 4.1, p. 299. ISSN: 1756-0500. DOI: `10.1186/1756-0500-4-299`. URL: `https://doi.org/10.1186/1756-0500-4-299`.

Massey, Jeffrey D., W. James Steenburgh, Sebastian W. Hoch, and Jason C. Knievel (2014)
"Sensitivity of Near-Surface Temperature Forecasts to Soil Properties over a Sparsely Vegetated Dryland Region". In: *Journal of Applied Meteorology and Climatology* 53.8, pp. 1976–1995. DOI: `10.1175/JAMC-D-13-0362.1`. URL: `https://journals.ametsoc.org/view/journals/apme/53/8/jamc-d-13-0362.1.xml`.

Mckay, M., Richard Beckman, and William Conover (May 1979)
"A Comparison of Three Methods for Selecting Vales of Input Variables in the Analysis of Output From a Computer Code". In: *Technometrics* 21, pp. 239–245. DOI: `10.1080/00401706.1979.10489755`.

Meenal, R., D. Binu, K. C. Ramya, Prawin Angel Michael, K. Vinoth Kumar, E. Rajasekaran, and B. Sangeetha (Aug. 2022)
"Weather Forecasting for Renewable Energy System: A Review". In: *Archives of Computational Methods in Engineering* 29.5, pp. 2875–2891. ISSN: 1886-1784. DOI: `10.1007/s11831-021-09695-3`. URL: `https://doi.org/10.1007/s11831-021-09695-3`.

Mendoza, Pablo A., Martyn P. Clark, Michael Barlage, Balaji Rajagopalan, Luis Samaniego, Gab Abramowitz, and Hoshin Gupta (2015)
"Are we unnecessarily constraining the agility of complex process-based models?" In: *Water Resources Research* 51.1, pp. 716–728. DOI: `https://doi.org/10.1002/2014WR015820`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014WR015820`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014WR015820`.

Mlawer, Eli J., Steven J. Taubman, Patrick D. Brown, Michael J. Iacono, and Shepard A. Clough (1997)
"Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave". In: *Journal of Geophysical Research: Atmospheres* 102.D14, pp. 16663–16682. DOI: `https://doi.org/10.1029/97JD00237`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/97JD00237`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/97JD00237`.

Monin, A. and A.M. Obukhov (1954)
"Basic laws of turbulent mixing in the surface layer of the atmosphere". In:

Monteith, John Lennox and Timothy R. Oke (1979)
"Boundary Layer Climates." In: *Journal of Applied Ecology* 17, p. 517. URL: `https://api.semanticscholar.org/CorpusID:87534854`.

Mooers, Griffin, Michael Pritchard, Tom Beucler, Jordan Ott, Galen Yacalis, Pierre Baldi, and Pierre Gentine (2021)
"Assessing the Potential of Deep Learning for Emulating Cloud Superparameterization in Climate Models With Real-Geography Boundary Conditions". In: *Journal of Advances in Modeling Earth Systems* 13.5. e2020MS002385 2020MS002385, e2020MS002385. DOI: `https://doi.org/10.1029/2020MS002385`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020MS002385`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002385`.

Morris, Max D. (1991)
"Factorial sampling plans for preliminary computational experiments". In: *Quality Engineering* 37, pp. 307–310.

Murti, Muhammad Ary, Rio Junior, Ali Najah Ahmed, and Ahmed Elshafie (Dec. 2022)

"Earthquake multi-classification detection based velocity and displacement data filtering using machine learning algorithms". In: *Scientific Reports* 12.1, p. 21200. ISSN: 2045-2322. DOI: `10.1038/s41598-022-25098-1`. URL: `https://doi.org/10.1038/s41598-022-25098-1`.

Muthukrishnan, R and R Rohini (2016)
"LASSO: A feature selection technique in predictive modeling for machine learning". In: pp. 18–20. DOI: `10.1109/ICACA.2016.7887916`.

Niederreiter, Harald (1992)
"Random number generation and Quasi-Monte Carlo methods". In: URL: `https://api.semanticscholar.org/CorpusID:121347446`.

Niu, Guo-Yue and Zong-Liang Yang (2004)
"Effects of vegetation canopy processes on snow surface energy and mass balances". In: *Journal of Geophysical Research: Atmospheres* 109.D23. DOI: `https://doi.org/10.1029/2004JD004884`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2004JD004884`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004JD004884`.

— (2006)
"Effects of Frozen Soil on Snowmelt Runoff and Soil Water Storage at a Continental Scale". In: *Journal of Hydrometeorology* 7.5, pp. 937–952. DOI: `10.1175/JHM538.1`. URL: `https://journals.ametsoc.org/view/journals/hydr/7/5/jhm538_1.xml`.

— (2007)
"An observation-based formulation of snow cover fraction and its evaluation over large North American river basins". In: *Journal of Geophysical Research: Atmospheres* 112.D21. DOI: `https://doi.org/10.1029/2007JD008674`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007JD008674`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007JD008674`.

Niu, Guo-Yue, Zong-Liang Yang, Robert E. Dickinson, and Lindsey E. Gulden (2005)
"A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models". In: *Journal of Geophysical Research: Atmospheres* 110.D21. DOI: `https://doi.org/10.1029/2005JD006111`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006111`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006111`.

Niu, Guo-Yue, Zong-Liang Yang, Robert E. Dickinson, Lindsey E. Gulden, and Hua Su (2007)
"Development of a simple groundwater model for use in climate models and evaluation with Gravity Recovery and Climate Experiment data". In: *Journal of*

Geophysical Research: Atmospheres 112.D7. DOI: `https://doi.org/10.1029/2006JD007522`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006JD007522`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JD007522`.

Niu, Guo-Yue, Zong-Liang Yang, Kenneth E. Mitchell, Fei Chen, Michael B. Ek, Michael Barlage, Anil Kumar, Kevin Manning, Dev Niyogi, Enrique Rosero, Mukul Tewari, and Youlong Xia (2011)
"The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements". In: *Journal of Geophysical Research: Atmospheres* 116.D12. DOI: `https://doi.org/10.1029/2010JD015139`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010JD015139`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015139`.

O'Hagan, A. (2006)
"Bayesian analysis of computer code outputs: A tutorial". In: *Reliability Engineering System Safety* 91.10. The Fourth International Conference on Sensitivity Analysis of Model Output (SAMO 2004), pp. 1290–1300. ISSN: 0951-8320. DOI: `https://doi.org/10.1016/j.ress.2005.11.025`. URL: `https://www.sciencedirect.com/science/article/pii/S0951832005002383`.

Oakley, Jeremy E. and Anthony O'Hagan (2004)
"Probabilistic sensitivity analysis of complex models: a Bayesian approach". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 66. URL: `https://api.semanticscholar.org/CorpusID:6130150`.

Oleson, Keith W., Yongjiu Dai, Bertrand Bonan, Mike Bosilovichm, Robert E. Dickinson, Paul A. Dirmeyer, Forrest M. Hoffman, Paul R. Houser, Samuel Levis, Guo-yue Niu, Peter E. Thornton, Mariana Vertenstein, Zong-Liang Yang, and Xubin Zeng (2004)
"Technical Description of the Community Land Model (CLM)". In: URL: `https://api.semanticscholar.org/CorpusID:129695923`.

Oliveira, Sandra, Friderike Oehler, Jesús San-Miguel-Ayanz, Andrea Camia, and José M.C. Pereira (2012)
"Modeling spatial patterns of fire occurrence in Mediterranean Europe using Multiple Regression and Random Forest". In: *Forest Ecology and Management* 275, pp. 117–129. ISSN: 0378-1127. DOI: `https://doi.org/10.1016/j.foreco.2012.03.003`. URL: `https://www.sciencedirect.com/science/article/pii/S0378112712001272`.

Palani, Sundarambal, Shie-Yui Liong, and Pavel Tkalich (2008)
"An ANN application for water quality forecasting". In: *Marine Pollution Bulletin* 56.9, pp. 1586–1597. ISSN: 0025-326X. DOI: `https://doi.org/10.`

1016/j.marpolbul.2008.05.021. URL: https://www.sciencedirect.com/science/article/pii/S0025326X08003196.

Paulson, C. A. (1970)
"The Mathematical Representation of Wind Speed and Temperature Profiles in the Unstable Atmospheric Surface Layer". In: *Journal of Applied Meteorology and Climatology* 9.6, pp. 857–861. DOI: 10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/apme/9/6/1520-0450_1970_009_0857_tmrows_2_0_co_2.xml.

Pourtaghi, Zohre Sadat, Hamid Reza Pourghasemi, Roberta Aretano, and Teodoro Semeraro (2016)
"Investigation of general indicators influencing on forest fire and its susceptibility modeling using different data mining techniques". In: *Ecological Indicators* 64, pp. 72–84. ISSN: 1470-160X. DOI: https://doi.org/10.1016/j.ecolind.2015.12.030. URL: https://www.sciencedirect.com/science/article/pii/S1470160X15007359.

Pradhan, Biswajeet (2013)
"A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS". In: *Computers Geosciences* 51, pp. 350–365. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2012.08.023. URL: https://www.sciencedirect.com/science/article/pii/S0098300412003093.

Pyo, JongCheol, Yong Kwon, Jae-Hyun Ahn, Sangsoo Baek, Yong-Hwan Kwon, and Kyung Cho (Feb. 2021)
"Sensitivity Analysis and Optimization of a Radiative Transfer Numerical Model for Turbid Lake Water". In: *Remote Sensing* 13, p. 709. DOI: 10.3390/rs13040709.

Queipo, Nestor V., Raphael T. Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P. Kevin Tucker (2005)
"Surrogate-based analysis and optimization". In: *Progress in Aerospace Sciences* 41.1, pp. 1–28. ISSN: 0376-0421. DOI: https://doi.org/10.1016/j.paerosci.2005.02.001. URL: https://www.sciencedirect.com/science/article/pii/S0376042105000102.

Rahmati, Omid, Hamid Reza Pourghasemi, and Assefa M. Melesse (2016)
"Application of GIS-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at Mehran Region, Iran". In: *CATENA* 137, pp. 360–372. ISSN: 0341-8162. DOI: https://doi.org/10.1016/j.catena.2015.10.010. URL: https://www.sciencedirect.com/science/article/pii/S0341816215301326.

Rampanelli, Gabriele, Dino Zardi, and Richard Rotunno (Dec. 2004) "Mechanisms of Up-Valley Winds". In: *Journal of The Atmospheric Sciences - J ATMOS SCI* 61, pp. 3097–3111. DOI: `10.1175/JAS-3354.1`.

Rasmussen, Carl Edward and Christopher K. I. Williams (Nov. 2005) *Gaussian Processes for Machine Learning.* The MIT Press. ISBN: 9780262256834. DOI: `10.7551/mitpress/3206.001.0001`. URL: `https://doi.org/10.7551/mitpress/3206.001.0001`.

Rasp, Stephan, Michael S. Pritchard, and Pierre Gentine (2018) "Deep learning to represent subgrid processes in climate models". In: *Proceedings of the National Academy of Sciences* 115.39, pp. 9684–9689. DOI: `10.1073/pnas.1810286115`. eprint: `https://www.pnas.org/doi/pdf/10.1073/pnas.1810286115`. URL: `https://www.pnas.org/doi/abs/10.1073/pnas.1810286115`.

Ren, Xiaoli, Xiaoyong Li, Kaijun Ren, Junqiang Song, Zichen Xu, Kefeng Deng, and Xiang Wang (Dec. 2020) "Deep Learning-Based Weather Prediction: A Survey". In: *Big Data Research* 23, p. 100178. DOI: `10.1016/j.bdr.2020.100178`.

Ren, Ye, Le Zhang, and P.N. Suganthan (2016) "Ensemble Classification and Regression-Recent Developments, Applications and Future Directions [Review Article]". In: *IEEE Computational Intelligence Magazine* 11.1, pp. 41–53. DOI: `10.1109/MCI.2015.2471235`.

Renardy, Marissa, Louis R. Joslyn, Jess A. Millar, and Denise E. Kirschner (2021) "To Sobol or not to Sobol? The effects of sampling schemes in systems biology applications". In: *Mathematical Biosciences* 337, p. 108593. ISSN: 0025-5564. DOI: `https://doi.org/10.1016/j.mbs.2021.108593`. URL: `https://www.sciencedirect.com/science/article/pii/S0025556421000419`.

Rodriguez-Galiano, Victor, Maria Paula Mendes, Maria Jose Garcia-Soldado, Mario Chica-Olmo, and Luis Ribeiro (2014) "Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain)". In: *Science of The Total Environment* 476-477, pp. 189–206. ISSN: 0048-9697. DOI: `https://doi.org/10.1016/j.scitotenv.2014.01.001`. URL: `https://www.sciencedirect.com/science/article/pii/S0048969714000102`.

Rosero, Enrique, Zong-Liang Yang, Thorsten Wagener, Lindsey E. Gulden, Soni Yatheendradas, and Guo-Yue Niu (2010) "Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the Noah land surface model over transition zones during the warm season". In: *Journal of Geophysical Research: Atmospheres* 115.D3. DOI: `https://doi.org/10.1029/2009JD012035`. eprint: `https:`

//agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009JD012035. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD012035`.

Rosolem, Rafael, Hoshin Gupta, W. Shuttleworth, and Xubin Zeng (Apr. 2012) "A fully multiple-criteria implementation of the Sobol method for parameter sensitivity analysis". In: *Journal of Geophysical Research D: Atmospheres* 117. DOI: `10.1029/2011JD016355`.

Ross, Andrew, Ziwei Li, Pavel Perezhogin, Carlos Fernandez-Granda, and Laure Zanna (2023) "Benchmarking of Machine Learning Ocean Subgrid Parameterizations in an Idealized Model". In: *Journal of Advances in Modeling Earth Systems* 15.1. e2022MS003258 2022MS003258, e2022MS003258. DOI: `https://doi.org/10.1029/2022MS003258`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2022MS003258`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2022MS003258`.

Sakaguchi, Koichi and Xubin Zeng (2009a) "Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5)". In: *Journal of Geophysical Research: Atmospheres* 114.D1. DOI: `https://doi.org/10.1029/2008JD010834`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD010834`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010834`.

— (2009b) "Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the Community Land Model (CLM3.5)". In: *Journal of Geophysical Research: Atmospheres* 114.D1. DOI: `https://doi.org/10.1029/2008JD010834`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2008JD010834`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008JD010834`.

Salmasi, Farzin, Meysam Nouri, Parveen Sihag, and John Abraham (Sept. 2020) "Application of SVM, ANN, GRNN, RF, GP and RT models for predicting discharge coefficients of oblique sluice gates using experimental data". In: *Water Supply* 21.1, pp. 232–248. ISSN: 1606-9749. DOI: `10.2166/ws.2020.226`. eprint: `https://iwaponline.com/ws/article-pdf/21/1/232/839979/ws021010232.pdf`. URL: `https://doi.org/10.2166/ws.2020.226`.

Saltelli, Andrea (Jan. 2004) "Global sensitivity analysis: An introduction". In: *Proceedings of the 4th international conference on sensitivity analysis of model output (SAMO 2004)*.

Saltelli, Andrea, Paola Annoni, Ivano Azzini, Francesca Campolongo, Marco Ratto, and Stefano Tarantola (2010)

"Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index". In: *Computer Physics Communications* 181.2, pp. 259–270. ISSN: 0010-4655. DOI: `https://doi.org/10.1016/j.cpc.2009.09.018`. URL: `https://www.sciencedirect.com/science/article/pii/S0010465509003087`.

Saltelli, Andrea, Marco Ratto, Andres Terry, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola (2008) *Global Sensitivity Analysis: The Primer*. URL: `https://api.semanticscholar.org/CorpusID:115957810`.

Saltelli, Andrea and I.M. Sobol' (Jan. 1995) "Sensitivity analysis for nonlinear mathematical models: Numerical experience". In: *Matematicheskoe Modelirovanie* 7.

Sarkar, Archana and Prashant Pandey (2015) "River Water Quality Modelling Using Artificial Neural Network Technique". In: *Aquatic Procedia* 4. INTERNATIONAL CONFERENCE ON WATER RESOURCES, COASTAL AND OCEAN ENGINEERING (ICWRCOE'15), pp. 1070–1077. ISSN: 2214-241X. DOI: `https://doi.org/10.1016/j.aqpro.2015.02.135`. URL: `https://www.sciencedirect.com/science/article/pii/S2214241X15001364`.

Satir, Onur, Suha Berberoglu, and Cenk Donmez (2016) "Mapping regional forest fire probability using artificial neural network model in a Mediterranean forest ecosystem". In: *Geomatics, Natural Hazards and Risk* 7.5, pp. 1645–1658. DOI: `10.1080/19475705.2015.1084541`. eprint: `https://doi.org/10.1080/19475705.2015.1084541`. URL: `https://doi.org/10.1080/19475705.2015.1084541`.

Schaake, John C., Victor I. Koren, Qing-Yun Duan, Kenneth Mitchell, and Fei Chen (1996a) "Simple water balance model for estimating runoff at different spatial and temporal scales". In: *Journal of Geophysical Research: Atmospheres* 101.D3, pp. 7461–7475. DOI: `https://doi.org/10.1029/95JD02892`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/95JD02892`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95JD02892`.

— (1996b) "Simple water balance model for estimating runoff at different spatial and temporal scales". In: *Journal of Geophysical Research: Atmospheres* 101.D3, pp. 7461–7475. DOI: `https://doi.org/10.1029/95JD02892`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/95JD02892`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/95JD02892`.

Schmidli, J., B. Billings, F. K. Chow, S. F. J. de Wekker, J. Doyle, V. Grubišić, T. Holt, Q. Jiang, K. A. Lundquist, P. Sheridan, S. Vosper, C. D. Whiteman, A. A. Wyszogrodzki, and G. Zängl (2011)
"Intercomparison of Mesoscale Model Simulations of the Daytime Valley Wind System". In: *Monthly Weather Review* 139.5, pp. 1389–1409. DOI: `10.1175/2010MWR3523.1`. URL: `https://journals.ametsoc.org/view/journals/mwre/139/5/2010mwr3523.1.xml`.

Schmidli, Juerg and Richard Rotunno (Sept. 2010)
"Mechanisms of Along-Valley Winds and Heat Exchange over Mountainous Terrain". In: *Journal of The Atmospheric Sciences - J ATMOS SCI* 67, pp. 3033–3047. DOI: `10.1175/2010JAS3473.1`.

Schneider, Rochelle, Massimo Bonavita, Alan Geer, Rossella Arcucci, Peter Dueben, Claudia Vitolo, Bertrand Le Saux, Begüm Demir, and Pierre-Philippe Mathieu (June 2022)
"ESA-ECMWF Report on recent progress and research directions in machine learning for Earth System observation and prediction". In: *npj Climate and Atmospheric Science* 5.1, p. 51. ISSN: 2397-3722. DOI: `10.1038/s41612-022-00269-z`. URL: `https://doi.org/10.1038/s41612-022-00269-z`.

Schölkopf, Bernhard and Alex Smola (Apr. 2002)
"Support Vector Machines and Kernel Algorithms". In: *Encyclopedia of Biostatistics, 5328-5335 (2005).*

Schultz, M. G., C. Betancourt, B. Gong, F. Kleinert, M. Langguth, L. H. Leufen, A. Mozaffari, and S. Stadtler (2021)
"Can deep learning beat numerical weather prediction?" In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194, p. 20200097. DOI: `10.1098/rsta.2020.0097`. eprint: `https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2020.0097`. URL: `https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2020.0097`.

Sekulić, Aleksandar, Milan Kilibarda, Gerard B.M. Heuvelink, Mladen Nikolić, and Branislav Bajat (2020)
"Random Forest Spatial Interpolation". In: *Remote Sensing* 12.10. ISSN: 2072-4292. DOI: `10.3390/rs12101687`. URL: `https://www.mdpi.com/2072-4292/12/10/1687`.

Sellers, P. J., Y. Mintz, Y. C. Sud, and A. Dalcher (1986)
"A Simple Biosphere Model (SIB) for Use within General Circulation Models". In: *Journal of Atmospheric Sciences* 43.6, pp. 505–531. DOI: `10.1175/1520-0469(1986)043<0505:ASBMFU>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/atsc/43/6/1520-0469_1986_043_0505_asbmfu_2_0_co_2.xml`.

Serafin, Stefano and Dino Zardi (Apr. 2010)
"Structure of the Atmospheric Boundary Layer in the Vicinity of a Developing Upslope Flow System: A Numerical Model Study". In: *Journal of The Atmospheric Sciences - J ATMOS SCI* 67, pp. 1171–1185. DOI: `10.1175/2009JAS3231.1`.

— (Sept. 2011)
"Daytime Development of the Boundary Layer over a Plain and in a Valley under Fair Weather Conditions: A Comparison by Means of Idealized Numerical Simulations". In: *Journal of The Atmospheric Sciences - J ATMOS SCI* 68, pp. 2128–2141. DOI: `10.1175/2011JAS3610.1`.

Shah, Muhammad Izhar, Wesam Salah Alaloul, Abdulaziz Alqahtani, Ali Aldrees, Muhammad Ali Musarat, and Muhammad Faisal Javed (2021)
"Predictive Modeling Approach for Surface Water Quality: Development and Comparison of Machine Learning Models". In: *Sustainability* 13.14. ISSN: 2071-1050. DOI: `10.3390/su13147515`. URL: `https://www.mdpi.com/2071-1050/13/14/7515`.

Sharma, Alok, Kuldip K. Paliwal, Seiya Imoto, and Satoru Miyano (Apr. 2014)
"A feature selection method using improved regularized linear discriminant analysis". In: *Machine Vision and Applications* 25.3, pp. 775–786. ISSN: 1432-1769. DOI: `10.1007/s00138-013-0577-y`. URL: `https://doi.org/10.1007/s00138-013-0577-y`.

Shen, Kai-Quan, Chong-Jin Ong, Xiao-Ping Li, and Einar P. V. Wilder-Smith (Jan. 2008)
"Feature selection via sensitivity analysis of SVM probabilistic outputs". In: *Machine Learning* 70.1, pp. 1–20. ISSN: 1573-0565. DOI: `10.1007/s10994-007-5025-7`. URL: `https://doi.org/10.1007/s10994-007-5025-7`.

Sihag, Parveen, Vinod Kumar, Fazal Rahman Afghan, Siraj Muhammed Pandhiani, and Ali Keshavarzi (Dec. 2019)
"Predictive modeling of PM2.5 using soft computing techniques: case study—Faridabad, Haryana, India". In: *Air Quality, Atmosphere & Health* 12.12, pp. 1511–1520. ISSN: 1873-9326. DOI: `10.1007/s11869-019-00755-z`. URL: `https://doi.org/10.1007/s11869-019-00755-z`.

Singh, Siddharth, Mayank Kaushik, Ambuj Gupta, and Anil Malviya (Jan. 2019)
"Weather Forecasting using Machine Learning Techniques". In: *SSRN Electronic Journal*. DOI: `10.2139/ssrn.3350281`.

Skamarock, C., Bogumiła Klemp, Jimy Dudhia, O. Gill, Zhiquan Liu, Judith Berner, Wei Wang, G. Powers, Greg Duda, Dale Melvyn Barker, and Xiangyu Huang (2019)
"A Description of the Advanced Research WRF Model Version 4". In: URL: `https://api.semanticscholar.org/CorpusID:196211930`.

Skamarock, William C, J.B. Klemp, J. Dudhia, D.O. Gill, Z. Liu, J. Berner, W. Wanh, J.G. Powers, M.G Duda, D. Barker, and X.-Y. Huang (2021)
"A Description of the Advanced Research WRF Model Version 4.3". In: *No. NCAR/TN-556+STR*. DOI: `10.5065/1dfh-6p97`.

Skamarock, William C., Joseph B. Klemp, Jimy Dudhia, Dave Gill, Dale Melvyn Barker, Michael G. Duda, Xiang-Yu Huang, Wei Wang, and Jordan G Powers (2008)
"A Description of the Advanced Research WRF Version 3". In: URL: `https://api.semanticscholar.org/CorpusID:6114213`.

Sobol', I.M (1967)
"On the distribution of points in a cube and the approximate evaluation of integrals". In: *USSR Computational Mathematics and Mathematical Physics* 7.4, pp. 86–112. ISSN: 0041-5553. DOI: `https://doi.org/10.1016/0041-5553(67)90144-9`. URL: `https://www.sciencedirect.com/science/article/pii/0041555367901449`.

Sridhara, Shankarappa, Konapura Nagaraja Manoj, Pradeep Gopakkali, Girish R. Kashyap, Bappa Das, Kamlesh Kumar Singh, and Amit Kumar Srivastava (Jan. 2023)
"Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India". In: *International Journal of Biometeorology* 67.1, pp. 165–180. ISSN: 1432-1254. DOI: `10.1007/s00484-022-02396-x`. URL: `https://doi.org/10.1007/s00484-022-02396-x`.

Stein, Michael L (1999)
Interpolation of spatial data: some theory for kriging. Springer Science & Business Media.

Suárez Sánchez, A., P.J. García Nieto, P. Riesgo Fernández, J.J. del Coz Díaz, and F.J. Iglesias-Rodríguez (2011)
"Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain)". In: *Mathematical and Computer Modelling* 54.5, pp. 1453–1466. ISSN: 0895-7177. DOI: `https://doi.org/10.1016/j.mcm.2011.04.017`. URL: `https://www.sciencedirect.com/science/article/pii/S0895717711002196`.

Sukoriansky, Semion, Boris Galperin, and Ilya Staroselsky (Aug. 2005)
"A quasinormal scale elimination model of turbulent flows with stable stratification". In: *Physics of Fluids* 17.8, p. 085107. ISSN: 1070-6631. DOI: `10.1063/1.2009010`. eprint: `https://pubs.aip.org/aip/pof/article-pdf/doi/10.1063/1.2009010/16050470/085107\_1\_online.pdf`. URL: `https://doi.org/10.1063/1.2009010`.

Taira, Kunihiko, Steven Brunton, Scott Dawson, Clarence Rowley, Tim Colonius, Beverley McKeon, Oliver Schmidt, Stanislav Gordeyev, V. Theofilis, and

Lawrence Ukeiley (Feb. 2017)
"Modal Analysis of Fluid Flows: An Overview". In: *AIAA Journal* 55. DOI: `10.2514/1.J056060`.

Tarantola, Stefano, Federico Ferretti, Samuele Lo Piano, Mariia Kozlova, Alessio Lachi, Rossana Rosati, Arnald Puy, Pamphile Roy, Giulia Vannucci, Marta Kuc-Czarnecka, and Andrea Saltelli (2024)
"An annotated timeline of sensitivity analysis". In: *Environmental Modelling Software* 174, p. 105977. ISSN: 1364-8152. DOI: `https://doi.org/10.1016/j.envsoft.2024.105977`. URL: `https://www.sciencedirect.com/science/article/pii/S1364815224000380`.

Tian, Wei (2013)
"A review of sensitivity analysis methods in building energy analysis". In: *Renewable and Sustainable Energy Reviews* 20, pp. 411–419. ISSN: 1364-0321. DOI: `https://doi.org/10.1016/j.rser.2012.12.014`. URL: `https://www.sciencedirect.com/science/article/pii/S1364032112007101`.

Tibshirani, Robert (1996)
"Regression Shrinkage and Selection Via the Lasso". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, pp. 267–288. DOI: `https://doi.org/10.1111/j.2517-6161.1996.tb02080.x`. eprint: `https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1996.tb02080.x`. URL: `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.2517-6161.1996.tb02080.x`.

Tipping, Michael (Jan. 2001)
"Sparse Bayesian Learning and the Relevance Vector Machine". In: *Journal of Machine Learning Research* 1, pp. 211–244.

Tomasi, Elena, Lorenzo Giovannini, Dino Zardi, and Massimiliano De Franceschi (Oct. 2017)
"Optimization of Noah and Noah$_MPWRFLandSurfaceSchemesinSnow-MeltingConditionsoverComplexTerrain$". In: *Monthly Weather Review* 145. DOI: `10.1175/MWR-D-16-0408.1`.

Torres, Marcella (June 2021)
"A Machine Learning Method for Parameter Estimation and Sensitivity Analysis". In: pp. 330–343. ISBN: 978-3-030-77976-4. DOI: `10.1007/978-3-030-77977-1_26`.

Trabelsi, Fatma and Salsebil Bel Hadj Ali (2022)
"Exploring Machine Learning Models in Predicting Irrigation Groundwater Quality Indices for Effective Decision Making in Medjerda River Basin, Tunisia". In: *Sustainability* 14.4. ISSN: 2071-1050. DOI: `10.3390/su14042341`. URL: `https://www.mdpi.com/2071-1050/14/4/2341`.

Vapnik V.N. Chervonenkis, A.Y. (1963)
"On a class of algorithms of learning pattern recognition". In: 25, pp. 112–120.

Vinuesa, Ricardo and Steven Brunton (June 2022)
"Enhancing computational fluid dynamics with machine learning". In: *Nature Computational Science* 2, pp. 358–366. DOI: `10.1038/s43588-022-00264-7`.

Wagner, J., Alexander Gohm, and Mathias Rotach (Oct. 2014)
"The impact of valley geometry on daytime thermally driven flows and vertical transport processes". In: *Quarterly Journal of the Royal Meteorological Society* 141. DOI: `10.1002/qj.2481`.

— (June 2015)
"Influence of along-valley terrain heterogeneity on exchange processes over idealized valleys". In: *Atmospheric Chemistry and Physics Discussions* 15, pp. 415–451. DOI: `10.5194/acpd-15-415-2015`.

Wang, Anqi and Dimitri Solomatine (May 2019)
"Practical Experience of Sensitivity Analysis: Comparing Six Methods, on Three Hydrological Models, with Three Performance Criteria". In: *Water* 11, p. 1062. DOI: `10.3390/w11051062`.

Wang, Yu, Z. Pan, J. Zheng, Lewei Qian, and M. Li (2019)
"A hybrid ensemble method for pulsar candidate classification". In: *Astrophysics and Space Science* 364. URL: `https://api.semanticscholar.org/CorpusID:201663653`.

Watson-Parris, D., A. Williams, L. Deaconu, and P. Stier (2021)
"Model calibration using ESEm v1.1.0 – an open, scalable Earth system emulator". In: *Geoscientific Model Development* 14.12, pp. 7659–7672. DOI: `10.5194/gmd-14-7659-2021`. URL: `https://gmd.copernicus.org/articles/14/7659/2021/`.

Whiteman, Charles D. and E. Dreiseitl (1984)
"Alpine meteorology: translations of classic contributions by A. Wagner, E. Ekhart, and F. Defant". In:

Wicker, Louis and William Skamarock (Aug. 2002)
"Time-Splitting Methods for Elastic Models Using Forward Time Schemes". In: *Monthly Weather Review - MON WEATHER REV* 130. DOI: `10.1175/1520-0493(2002)130<2088:TSMFEM>2.0.CO;2`.

Williamson, Daniel, Michael Goldstein, Lesley Allison, Adam Blaker, Peter Challenor, Laura Jackson, and Kuniko Yamazaki (Oct. 2013)
"History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble". In: *Climate Dynamics* 41, pp. 1703–1729. DOI: `10.1007/s00382-013-1896-4`.

Wolff, Bjorn, Oliver Kramer, and Detlev Heinemann (2017)
"Selection of Numerical Weather Forecast Features for PV Power Predictions

with Random Forests". In: ed. by Wei Lee Woon, Zeyar Aung, Oliver Kramer, and Stuart Madnick, pp. 78–91.

Wu, Wei, Mao-Fen Li, Xia Xu, Xiao-Ping Tang, Chao Yang, and Hong-Bin Liu (Oct. 2021)
"The transferability of random forest and support vector machine for estimating daily global solar radiation using sunshine duration over different climate zones". In: *Theoretical and Applied Climatology* 146.1, pp. 45–55. ISSN: 1434-4483. DOI: `10.1007/s00704-021-03726-6`. URL: `https://doi.org/10.1007/s00704-021-03726-6`.

Wu, Zongjun, Ningbo Cui, Daozhi Gong, Feiyu Zhu, Liwen Xing, Bin Zhu, Xi Chen, Shengling Wen, and Quanshan Liu (2023)
"Simulation of daily maize evapotranspiration at different growth stages using four machine learning models in semi-humid regions of northwest China". In: *Journal of Hydrology* 617, p. 128947. ISSN: 0022-1694. DOI: `https://doi.org/10.1016/j.jhydrol.2022.128947`. URL: `https://www.sciencedirect.com/science/article/pii/S0022169422015177`.

Yang, Rongqian and Mark A. Friedl (2003)
"Modeling the effects of three-dimensional vegetation structure on surface radiation and energy balance in boreal forests". In: *Journal of Geophysical Research: Atmospheres* 108.D16. DOI: `https://doi.org/10.1029/2002JD003109`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2002JD003109`. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2002JD003109`.

Yang, Wan-Xia, Fang-Fang Wang, Yun-Yan Pan, Jian-Qin Xie, Ming-Hua Lu, and Chong-Ge You (Dec. 2022)
"Comparison of ischemic stroke diagnosis models based on machine learning". In: *Frontiers in Neurology* 13. DOI: `10.3389/fneur.2022.1014346`.

Yang, Zong-Liang, Robert E. Dickinson, Alan Robock, and K. Ya Vinnikov (1997)
"Validation of the Snow Submodel of the Biosphere–Atmosphere Transfer Scheme with Russian Snow Cover and Meteorological Observational Data". In: *Journal of Climate* 10.2, pp. 353–373. DOI: `10.1175/1520-0442(1997)010<0353:VOTSSO>2.0.CO;2`. URL: `https://journals.ametsoc.org/view/journals/clim/10/2/1520-0442_1997_010_0353_votsso_2.0.co_2.xml`.

Yang, Zong-Liang, Guo-Yue Niu, Kenneth E. Mitchell, Fei Chen, Michael B. Ek, Michael Barlage, Laurent Longuevergne, Kevin Manning, Dev Niyogi, Mukul Tewari, and Youlong Xia (2011)
"The community Noah land surface model with multiparameterization options (Noah-MP): 2. Evaluation over global river basins". In: *Journal of Geophysical Research: Atmospheres* 116.D12. DOI: `https://doi.org/10.1029/2010JD015140`. eprint: `https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/`

2010JD015140. URL: `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JD015140`.

Yilmaz, Işık (Aug. 2010)
"Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine". In: *Environmental Earth Sciences* 61.4, pp. 821–836. ISSN: 1866-6299. DOI: `10.1007/s12665-009-0394-9`. URL: `https://doi.org/10.1007/s12665-009-0394-9`.

Youssef, Ahmed Mohamed, Hamid Reza Pourghasemi, Zohre Sadat Pourtaghi, and Mohamed M. Al-Katheeri (Oct. 2016)
"Landslide susceptibility mapping using random forest, boosted regression tree, classification and regression tree, and general linear models and comparison of their performance at Wadi Tayyah Basin, Asir Region, Saudi Arabia". In: *Landslides* 13.5, pp. 839–856. ISSN: 1612-5118. DOI: `10.1007/s10346-015-0614-1`. URL: `https://doi.org/10.1007/s10346-015-0614-1`.

Yu, Ruiyun, Yu Yang, Leyou Yang, Guangjie Han, and Oguti Ann Move (2016)
"RAQ–A Random Forest Approach for Predicting Air Quality in Urban Sensing Systems". In: *Sensors* 16.1. ISSN: 1424-8220. DOI: `10.3390/s16010086`. URL: `https://www.mdpi.com/1424-8220/16/1/86`.

Yuval, Janni and Paul A. O'Gorman (July 2020)
"Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions". In: *Nature Communications* 11.1, p. 3295. ISSN: 2041-1723. DOI: `10.1038/s41467-020-17142-3`. URL: `https://doi.org/10.1038/s41467-020-17142-3`.

Zardi, Dino and C. David Whiteman (2013)
"Diurnal Mountain Wind Systems". In: *Mountain Weather Research and Forecasting: Recent Progress and Current Challenges*. Ed. by Fotini K. Chow, Stephan F.J. De Wekker, and Bradley J. Snyder. Dordrecht: Springer Netherlands, pp. 35–119. ISBN: 978-94-007-4098-3. DOI: `10.1007/978-94-007-4098-3_2`. URL: `https://doi.org/10.1007/978-94-007-4098-3_2`.

Zhang, Guo, Fei Chen, Yueli Chen, Jianduo Li, and Xindong Peng (2020)
"Evaluation of Noah-MP Land-Model Uncertainties over Sparsely Vegetated Sites on the Tibet Plateau". In: *Atmosphere* 11.5. ISSN: 2073-4433. DOI: `10.3390/atmos11050458`. URL: `https://www.mdpi.com/2073-4433/11/5/458`.

Zhang, Xuesong, Raghavan Srinivasan, and Michael Van Liew (2009)
"Approximating SWAT Model Using Artificial Neural Network and Support Vector Machine1". In: *JAWRA Journal of the American Water Resources Association* 45.2, pp. 460–474. DOI: `https://doi.org/10.1111/j.1752-1688.2009.00302.x`. eprint: `https://onlinelibrary.wiley.com/doi/pdf/`

10.1111/j.1752-1688.2009.00302.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1752-1688.2009.00302.x.

Zhou, Chao, Kunlong Yin, Ying Cao, Bayes Ahmed, Yuanyao Li, Filippo Catani, and Hamid Reza Pourghasemi (2018)
"Landslide susceptibility modeling applying machine learning methods: A case study from Longju in the Three Gorges Reservoir area, China". In: *Computers Geosciences* 112, pp. 23–37. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2017.11.019. URL: https://www.sciencedirect.com/science/article/pii/S0098300417304843.

Zilitinkevich, Sergej (1995)
"Non-local turbulent transport pollution dispersion aspects of coherent structure of convective flows". In: URL: https://api.semanticscholar.org/CorpusID:119885197.

Zouhri, Wahb, Lazhar Homri, and Jean-Yves Dantan (2022)
"Handling the impact of feature uncertainties on SVM: A robust approach based on Sobol sensitivity analysis". In: *Expert Systems with Applications* 189, p. 115691. ISSN: 0957-4174. DOI: https://doi.org/10.1016/j.eswa.2021.115691. URL: https://www.sciencedirect.com/science/article/pii/S0957417421010769.