JGR Atmospheres

RESEARCH ARTICLE

10.1029/2025JD045540

Machine Learning-Based Analysis of Atmospheric Process Sensitivity to Land Surface Properties Over Complex Terrain

Dario Di Santo[1], **Cenlin He**[2], **Fei Chen**[3], **Andrea Zonato**[4], **and Lorenzo Giovannini**[1]

[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, [4]CIMA Research Foundation, Savona, Italy

Key Points:

- Stomatal resistance, canopy height, and near-infrared leaf reflectivity dominate sensitivity in mixed forests
- Leaf and stem area indices and near-infrared leaf reflectivity are the key parameters controlling grassland simulations
- ML-based surrogate modeling enables efficient global sensitivity analysis of coupled land-atmosphere models

Correspondence to:

D. Di Santo,
dario.disanto@unitn.it

Citation:

Di Santo, D., He, C., Chen, F., Zonato, A., & Giovannini, L. (2026). Machine learning-based analysis of atmospheric process sensitivity to land surface properties over complex terrain. *Journal of Geophysical Research: Atmospheres*, *131*, e2025JD045540. https://doi.org/10.1029/2025JD045540

Received 7 OCT 2025
Accepted 21 APR 2026

Abstract This study investigates the sensitivity of numerical simulations of atmospheric processes over complex terrain to land surface model (LSM) parameters, focusing on thermally driven circulations in an idealized valley. The sensitivity analysis is performed using the Weather Research and Forecasting model coupled with the Noah-MP LSM, for forest and grassland land cover classes. An ensemble of 400 simulations for mixed forest and grassland is emulated with a suite of machine-learning regressors, and variance-based global sensitivity indices are computed from the surrogates. Sensitivity is evaluated for near-surface wind, potential temperature, turbulent kinetic energy (TKE), and surface heat fluxes at representative valley locations. Results clearly discriminate influential and non-influential parameters, providing guidance in identifying those that most affect model results and therefore should be treated with the most care. Over forests, stomatal minimum resistance (*RS*) is the most influential parameter overall, followed by canopy height (*HVT*), with importances intensifying during the night. Near-infrared leaf reflectivity (*RHOL_NIR*), leaf area index, and optimal transpiration temperature (*TOPT*) also result as important parameters. Over grasslands, by contrast, both leaf and stem (*SAI*) area indices dominate, followed by *RHOL_NIR*. The roughness length (*Z0MVT*) is the most important parameter affecting turbulent kinetic energy during daytime, while *RS* and *HVT* are largely non-influential. The results highlight that the proposed approach, integrating variance-based global sensitivity analysis with machine learning, represents a reliable framework to perform sensitivity analyses for complex numerical models, providing valuable insights into the dynamic interactions between land surface parameters and atmospheric conditions over complex terrain.

Plain Language Summary The presence of mountains and valleys creates local winds that significantly impact the weather and climate in these regions. Numerical weather prediction models use empirical parameters to describe the exchange of energy and water between the land and the atmosphere. However, these parameters are often uncertain and can vary greatly in real landscapes, particularly in complex terrain. In this study, we analyze the results of simulations with a numerical weather prediction model with machine learning techniques to identify the most important land surface parameters controlling the development of valley winds and energy exchanges. Through an ensemble of simulations, we highlight that parameters such as stomatal resistance, canopy height, and leaf reflectivity in the near-infrared spectrum of sunlight are particularly important in forests, while leaf and stem density play the biggest role in grasslands. Identifying the most influential parameters can guide future measurement and data collection efforts to improve the representation of atmospheric processes in meteorological models over mountain regions. This approach also demonstrates how machine learning can make complex sensitivity studies significantly faster and more efficient.

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 NWP simulations (e.g., Massey et al., 2014; Schmidli 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 values are often estimated empirically. A delicate issue regarding the use of specific parameter values stems from the representativeness of the land data. In fact, land surface properties/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 class. 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 conditions. This issue tends to be even more critical over complex terrain, where land cover characteristics usually present larger temporal and spatial heterogeneity. This affects the estimation of land surface parameter values and consequently the quality of NWP simulations. 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 relevant empirical parameters, in order to concentrate the effort on the parameters that most affect model results.

One of the most popular LSMs is Noah-MP (He et al., 2023c; Niu et al., 2011; Yang et al., 2011), which is available in the widely used 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 multi-parameterization options, reaching a total number of more than 4,500 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 specific choices of parameter values representative of different types of vegetation cover. The choice of these parameters is fixed to semi-empirical values, defined in the dedicated look-up file MPTABLE.TBL. Although they are used to define bulk land cover classes, no range of uncertainty is explicitly specified. 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 uncertainties in the model results.

The importance of each parameter to NWP model outputs can be assessed through sensitivity analyses. Analyzing the impact of small input perturbations on model outputs has historically been the first type of approach for carrying out a "local sensitivity analysis" (Iooss & Lemaître, 2014), that is, referring to local variations of the input parameters. The main limitation of these methods is the assumption of a linear relationship between the parameter and the output variable. Since many land and atmospheric processes are unavoidably non-linear, the situation is usually more complex. Moreover, in a typical NWP model, a parameter not only directly affects the outputs, but also the interaction between the other parameters and the outputs. This can be defined as the "interaction problem." To overcome this problem, more refined and global-oriented methods (i.e., "global sensitivity analysis") were introduced, including the variance-based approach that decomposes the variance of model outputs into multiple terms accounting for all the parameter interactions (Saltelli et al., 2010). The most famous and well-established variance-based method is the Sobol method (Saltelli & 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 different types of environmental models (Gan et al., 2014; Gong et al., 2015; Pyo et al., 2021; Rosolem et al., 2012; van Griensven et al., 2006; Wang & 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. The authors 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. In recent years, different extensive studies have evaluated the performance of Noah-MP using global sensitivity analysis (GSA). Cuntz et al. (2016) highlighted that many hard-coded model parameters, that is, 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. Li et al. (2018), evaluating the behavior of Noah-MP over the central Tibetan Plateau, found that energy and water vapor fluxes were dominated by soil parameters. Moreover, the canopy height 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 and latent heat fluxes, and net ecosystem exchange. Some of the most relevant parameters were the Clapp-Hornberger *b* parameter, field water capacity, soil porosity, canopy height, roughness length, and soil water availability.

The above-mentioned studies, which used sensitivity analysis to evaluate the impact of model parameters on Noah-MP results, were performed with the model executed in offline mode, that is, forcing the LSM with observations without a coupled meteorological model, in order to evaluate only the LSM behavior. Therefore, the influence on atmospheric processes was not assessed. On the other hand, 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 when 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 lead to 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 resulting from the differential heating and cooling of the lower atmospheric layers during fair weather days. In particular, mountain valleys are characterized by the development of slope winds, which respond rapidly to the heating/cooling of the valley slopes, blowing upslope after sunrise and downslope after sunset, and of valley winds, which develop due to pressure differences between different valley sections or between the valley and the plain, blowing down-valley during the night and up-valley during the day (Zardi & Whiteman, 2013). The alternation of these circulations typically marks the diurnal wind cycle in mountain valleys, strongly affecting the weather and climate of these regions.

The sensitivity of these circulations to the geometric characteristics of mountain valleys 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 & Zardi, 2010, 2011). For example, Wagner et al. (2015a, 2015b) performed different LES with the WRF model, highlighting variations in wind intensity depending on the valley floor slope, the valley volume, depth, width, and length. An idealized valley geometry was also used by Schmidli et al. (2011) to compare the up-valley wind system simulated by nine 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 prediction. However, a comprehensive evaluation of the sensitivity of NWP simulations to surface parameters over complex terrain is missing in the literature. As highlighted above, the literature studies applying sensitivity analysis to evaluate the impact of model parameters on Noah-MP results used the LSM in an offline mode, not coupled with a NWP model. The low computational cost of these offline simulations allowed these studies to test model outputs 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 & Boukouvala, 2020; Vinuesa & Brunton, 2022), specifically an ensemble of machine learning (ML) algorithms that are trained over samples of the original “high-fidelity” WRF model outputs, and emulate the input-output relations through regression tasks. These algorithms are implemented into an automated software, the ML-AMPSIT tool, written specifically to perform ML-based sensitivity studies and presented in Di Santo et al. (2025). ML extraction of feature importance is widely used in different Earth science fields (Alqahtani et al., 2022; Daviran et al., 2023; Elia et al., 2023; Gigović et al., 2019; Lei et al., 2023; Rahmati et al., 2016; Sridhara et al., 2023; Trabelsi & Bel Hadj Ali, 2022; Wu et al., 2023; Yu et al., 2024), and ML-based emulators are also increasingly applied in climate model calibration and parameter estimation (Cinquegrana et al., 2023; Couvreux et al., 2021; Dagon et al., 2020; Watson-Parris et al., 2021), aiding in process-based model development and in the optimization of physical parameters in NWP models. Nonetheless, to the authors' knowledge, the present study is the first to apply ML algorithms to extract parameter importance from an LSM coupled with an NWP model and to quantify the associated NWP model sensitivity to land surface parameters.

The choice of adopting an idealized experimental design is based on the aim of isolating the influence of Noah-MP vegetation parameters on the development of thermally driven circulations. As stated above, this is a solid strategy widely used in the literature for valley flow studies. Idealized valleys provide a controlled environment in which slope and valley winds can develop without the influence of confounding factors such as heterogeneous terrain, non-uniform land cover, and complex lateral boundary conditions. This framework ensures that any variability in the simulated circulation can be attributed exclusively to the perturbed surface parameters. In a real-world scenario, land cover patterns may be spatially heterogeneous. However, introducing such heterogeneity in the present idealized setup would hinder the interpretation of parameter-induced signals. Instead, this study aims to obtain parameter rankings to provide useful guidance on identifying the vegetation parameters most likely to introduce the largest uncertainty in real-case simulations under well-developed, thermally driven conditions. The quantification of these parameters should therefore be prioritized in future observational efforts and parameter tuning.

Despite the extensive literature on Noah-MP parameter sensitivity in offline mode, Sobol-type variance-based GSA remains computationally infeasible for online-coupled NWP-LSM configurations. In such systems, each model evaluation is expensive, and two-way land-atmosphere coupling amplifies nonlinearity and higher-order interaction effects. In Sobol-Saltelli sampling (Saltelli et al., 2010), the number of required model evaluations scales linearly with parameter dimension but remains large in absolute terms. For a configuration with $p \sim 20$ uncertain parameters and a base sample size $N \sim 10^3$, that is, the number of samples used to construct the sampling matrices, a direct high-fidelity computation including second-order effects (Saltelli et al., 2008) would require $N_{\text{eval}} = N(2p + 2) \approx 4 \times 10^4$ online WRF/Noah-MP runs, which is well beyond practical computational limits. The proposed workflow makes Sobol-type variance decomposition operationally achievable for this class of problems by combining Sobol sampling with ML surrogate emulation. Once trained on $O(10^2)$ high-fidelity WRF/Noah-MP realizations, the surrogate enables $O(10^4)$ Sobol evaluations at negligible marginal cost, yielding, for the presented case-study, an effective reduction of 10^4 in the number of high-fidelity simulations relative to a direct Sobol implementation. To the best of the authors' knowledge, this is the first study that (a) makes Sobol-type variance-based GSA computationally tractable for online-coupled NWP–LSM systems at this level of acceleration, and (b) proposes a multi-emulator ML-based strategy that stabilizes sensitivity inference in strongly nonlinear, interaction-dominated regimes characteristic of land-atmosphere feedbacks in NWP applications. Therefore, leveraging this methodology, this study is also the first to assess, with such a systematic approach, the influence of surface parameters on the development of thermally driven circulations and the associated thermodynamic fields in a mountain valley environment.

The paper is organized as follows: Section 2 presents the model setup, detailing the specifics of the idealized valley geometry, the computational domain, the initial and boundary conditions, and the physics options used in the Noah-MP LSM. It also describes the methodology for the sensitivity analysis, including the ML algorithms used, the parameters tested, and the statistical methods used to evaluate parameter 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.

2. Methods

2.1. Model Setup

The sensitivity analysis was performed by means of simulations over an idealized valley geometry with no background wind. This configuration corresponds to the standard setup adopted in previous idealized studies of valley thermodynamics (e.g., Rampanelli et al., 2004; Schmidli et al., 2011; Schmidli & Rotunno, 2010; Serafin & Zardi, 2010, 2011; Wagner et al., 2015a, 2015b) aiming at analyzing the development of clean, thermally driven slope and valley circulations. These conditions are not intended to represent a specific real-life scenario, but rather to isolate the physical mechanisms linking land-surface parameter uncertainty to valley wind evolution.

The idealized valley-plain topography used in the present work is based on Schmidli and Rotunno (2010) and Schmidli 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 and interferences 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

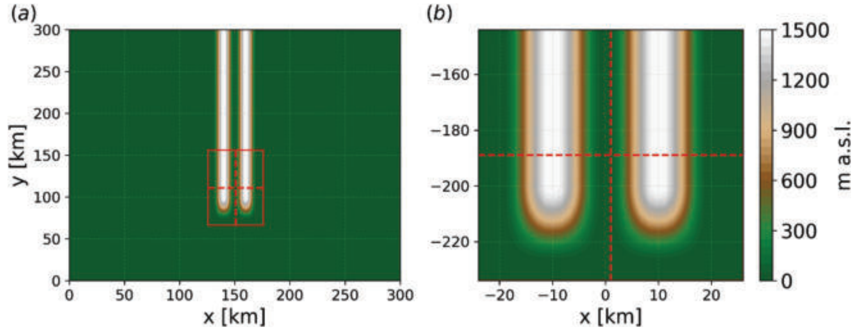


Figure 1. (a) Model topography, where dashed lines represent the position of the cross-valley and along-valley sections presented in Section 3.1, respectively in Figures 3–8. Their extension is limited by the red rectangle, which also represents the zoomed-in area shown in (b), where the four points chosen to study the different regions inside and at the exit of the valley are represented by red dots.

geometry, with the axis oriented in the North-South direction, is defined by the following analytical expression (Figure 1):

$$z = h(x,y) = h_p h_x(x) h_y(y), \tag{1}$$

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

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

$$h_y(y) = \begin{cases} 1, & \text{for } |y| \leq L_y \\ \frac{1}{2} + \frac{1}{2}\cos\left(\pi \frac{|y| - L_y}{S_y}\right), & \text{for } L_y < |y| < L_y + S_y \\ 0, & \text{for } |y| \geq L_y + S_y \end{cases} \tag{3}$$

where $h_p = 1.5$ km is the valley depth, $L_y = 200$ km is half the valley length, $S_x = S_y = 9$ km is 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 plateau width in the cross-valley direction. The analytical formulation of the valley geometry and the associated geometric parameters (X_1, X_2, X_3, X_4, L_y, S_x, S_y, and h_p) follow the configuration introduced by Schmidli et al. (2011).

The computational domain extends 300 km in both the South-North and West-East directions, while the domain top is placed at 12.2 km height. The height of the model top was selected based on previous idealized simulations of valley winds (e.g., Reif et al., 2024; Schmidli et al., 2011) to prevent the artificial reflection of gravity waves and to allow the thermally driven circulation to develop without interference from upper-boundary effects. The choice of the horizontal domain extension 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 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 13 levels below 1 km height. The vertical spacing is defined using a normalized sigmoid function:

$$\frac{s(x)}{s(0)}, \quad s(x) = \frac{1}{2}\left[1 + \tanh\left(\frac{b(x-c) - \ln(a)}{2}\right)\right]. \tag{4}$$

The parameters $a = 0.01$, $b = 0.5$, and $c = 2$ are selected empirically to obtain a dense distribution of vertical levels near the surface and a smooth transition to coarser spacing aloft. This configuration ensures adequate resolution in the boundary layer, where the slope/valley wind system develops, without introducing abrupt changes in the vertical grid spacing.

The center of the domain is set at 47°N 11°E; the simulations start at 06:00 UTC (LST-1), 19 March and run for 30 hr, with the first 6 hr used as spin-up period to allow the development of the thermally driven circulation. Therefore, a full diurnal cycle has been analyzed to evaluate the different phases of the valley circulation, with a radiation cycle representative of the equinox in an Alpine valley.

The initial state of the atmosphere has been set as stable stratified with the following potential temperature profile:

$$\theta(z) = \theta_s + \Gamma z + \Delta\theta\left(1 - e^{-\beta z}\right), \tag{5}$$

where $\theta_s = 280$ K is the reference potential temperature at the surface, $\Gamma = 3.2 \text{ K km}^{-1}$ is the lapse rate, and the term $\Delta\theta\left(1 - e^{-\beta z}\right)$ modulates the near-surface profile, with $\Delta\theta = 5$ K and $\beta = 0.002 \text{ m}^{-1}$. The values of θ_s, Γ, $\Delta\theta$, and β were chosen to represent a stably stratified pre-sunrise atmosphere. This initial condition facilitates the development of realistic slope and valley winds under clear-sky conditions.

The initial relative humidity is constant and equal to 30% along the entire vertical profile and the atmosphere is initially at rest. The soil moisture is initialized at 20%.

The Advanced Research WRF (ARW) 4.5.1 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 & Skamarock, 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 West-East boundary conditions are set as periodic, while the South-North boundary conditions are symmetric, thus with the effect of simulating a valley that is twice as long and open on both sides (Wagner et al., 2015a). Moreover, a 2-km Rayleigh damping layer with a non-dimensional coefficient of 0.1 is set at the upper boundary. The Coriolis force is neglected.

For 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 (QNSE) scheme (Sukoriansky et al., 2005) resulted in producing the least amount of spurious convective rolls, which are typical of simulations with mesoscale meteorological models at the kilometer scale (Ching et al., 2014), hence, it is selected for this work, coupled with the corresponding surface layer scheme. The microphysics parameterization is not activated to simulate a full diurnal cycle that is not affected by clouds, to allow the full development of the valley wind circulation and simplify the interpretation of the results. Similarly, the cumulus parameterization is switched off. The effects of the topography on shading and radiation are both taken into account.

Previous sensitivity studies based on offline Noah-MP simulations have shown that increasing model complexity can obscure the attribution of output variability to individual parameters by enhancing parameter interactions (e.g., Cuntz et al., 2016; Rosero et al., 2010). In light of this, the Noah-MP configuration adopted in this study retains the key physical mechanisms governing surface energy exchange and canopy-atmosphere coupling while avoiding additional sources of variability that are not central to the research question. Accordingly, the dynamic vegetation option is not activated and a fixed vegetation fraction of 0.6 is prescribed, The value for the fixed vegetation fraction was prescribed as a reasonable intermediate value compatible with the mixed-forest and grassland Modis IGBP classes considered here. Since the focus of this work is on the sensitivity to vegetation parameters rather than land-cover fractions, prescribing a fixed value avoids introducing an additional source of variability that is not central to the study objectives. The LAI is read from the Noah-MP look-up table (MPTABLE) monthly climatology and is varied within the sensitivity analysis.

The modified two-stream canopy radiative transfer scheme (Niu & Yang, 2004) is employed to ensure a physically consistent representation of canopy radiative processes. Stomatal resistance is computed using the Jarvis scheme (Jarvis et al., 1976), in line with the prescribed-LAI benchmark adopted by Arsenault et al. (2018), and coupled with a Noah-type soil moisture stress factor (Schaake et al., 1996), in order to maintain a direct and interpretable link between vegetation parameters and surface fluxes. Runoff and groundwater processes are represented using the TOPMODEL approach with groundwater option (Niu et al., 2007). The surface resistance to evaporation and sublimation follows Sakaguchi and Zeng (2009), while heat and momentum exchange coefficients are computed using the default Monin-Obukhov-based surface layer scheme. Soil texture is prescribed from the MPTABLE, thereby deactivating pedotransfer functions. Crop, irrigation, and harvesting options are not activated. More details about Noah-MP model physics are presented in He et al. (2023a).

The sensitivity analysis presented in this work is based on 400 simulations perturbing the values of several parameters defined in the MPTABLE, as presented in Section 2.3. This number of simulations was determined through preliminary convergence tests, which showed that both parameter importance and surrogate metrics had stabilized at this sample size. In particular, the two classes “mixed-forest” and “grassland” of the MODIS IGBP classification are chosen as the reference land cover to define the unperturbed (control) values of all the tested parameters. The analysis is performed over six WRF variables: west-east (cross-valley) and south-north (along-valley) wind components, potential temperature, turbulent kinetic energy (TKE), sensible and latent heat fluxes, evaluated at four selected points, as shown in Figure 1b, representing the valley floor and slope, the mountain crest and the plain close to the valley exit.

These four locations were selected to sample the key phases and components of the thermally driven valley circulation. This includes upslope and downslope flows along the valley sidewall, features associated with the along-valley wind system, both inside the valley and over the adjacent plain, and the development of convective thermals over the ridge.

2.2. ML-AMPSIT as a Sensitivity Analysis Tool

ML-AMPSIT (Machine Learning-based Automated Multi-method Parameter Sensitivity and Importance analysis Tool) is a fully automated sensitivity analysis tool (Di Santo et al., 2025) that uses a set of ML-based regression algorithms to assess parameter importance for a given model. The algorithms currently included are LASSO, Support Vector Machine (SVM), Classification and Regression Trees (CART), Random Forest (RF), Extreme Gradient Boosting (XGB), 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 first be specified. A Sobol low-discrepancy sequence of the same length as the number of simulations is then produced for each parameter. These sequences efficiently explore the multidimensional parameter space and are consistent with variance-based GSA (more information on the Sobol sequence can be found in Saltelli et al., 2010). The resulting WRF/Noah-MP ensemble is then used to train machine learning (ML) surrogate models. The sensitivity analysis is performed at different points in the computational domain, as specified by the user, using the different ML methods mentioned above. ML-AMPSIT calculates a series of performance metrics, namely the coefficient of determination $\left(R^2\right)$, together with the p-value testing the probability of unrelated variables to produce the same 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. In this study and in the literature of surrogate models, performance metrics (R^2, MSE and MAE) are computed to quantify how well the ML surrogates reproduce the corresponding high-fidelity model simulation outputs. The “true” values for the calculation of these metrics are therefore the WRF simulations with the same parameter settings, rather than a reference (control) simulation. Metrics are calculated from standardized, non-dimensional data to ensure comparability across variables and sites.

A core feature of ML-AMPSIT is the integration of multiple regression algorithms within a unified pipeline for surrogate model training, performance evaluation, and parameter importance estimation. This approach reduces reliance on any single regressor and enables an ensemble-based assessment of the consistency of sensitivity rankings across methods. A comprehensive description of ML-AMPSIT is provided in Di Santo et al. (2025), and the open-source code is available in Di Santo (2024).

Among all the algorithms available in ML-AMPSIT, a preliminary analysis of their individual performance for the present case study led to the choice of using a method-wise average based on SVM, LASSO, RF, XGB, BRR, and GPR. CART is, therefore, not used due to its poor performance in this case study. LASSO, RF, XGB, and BRR are used with their default settings, while slight modifications of the algorithm setup are introduced to SVM and GPR to improve their accuracy and efficiency as follows.

Considering SVM, in the present work the kernel function, which determines the shape of the decision boundary, is set to polynomial, instead of linear, that is, $k(x,x') = (\gamma\langle x,x'\rangle + r)^d$, 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 perform poorly in the present case study.

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

$$k(x,x') = \left(1 + \frac{\|x - x'\|^2}{2\alpha l^2}\right)^{-\alpha} \tag{6}$$

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 speed of hyperparameter tuning compared to the default option. Differently from the non-Bayesian methods, GPR quantifies the uncertainty around the estimated regression coefficients, offering a probabilistic interpretation of the model parameters.

The tree-based algorithms used in this work (RF, XGB) inherently provide a quantification of parameter importance from the variance reduction by fixing parameters between input evaluations. The more a feature reduces the variance on average across all the trees in the forest, the more important it is considered to be. On the other hand, the non-linear SVM (as implemented here), BRR and GPR do not directly provide a direct measure of importance. The speed of these surrogate model is exploited for a fast implementation of the Sobol method to evaluate parameter importance, where the Sobol indices are obtained directly from the ML-inferred relation between the original model inputs and outputs. Lastly, LASSO assigns importance based on the magnitude of its regression coefficients, which are obtained through an L1-regularization. The importance of a feature is determined by the absolute value of its estimated coefficient. To ensure that feature importance scores are comparable across all the models, each method is applied with standardized, non-dimensional data and the results are normalized.

The input-output data used to train the algorithms are collected from the simulation ensemble generated by the quasi-random Sobol sequence sampling strategy. For each method, 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 data set.

Hyperparameter tuning is conducted separately for each regression method, as the relevant hyperparameter space and the bias-variance trade-off vary across algorithms. Specifically, for each method, ML-AMPSIT optionally performs Bayesian optimization (via BayesSearchCV) with $n_{\text{iter}} = 10$ iterations, combined with k-fold cross-validation ($k = 5$) on the training set. The optimizer explores method-specific hyperparameter ranges (e.g., tree depth and number of estimators for tree-based models, regularization strength and numerical tolerances for LASSO and BRR, kernel and margin parameters for SVM, noise level and restart settings for GPR) and selects the configuration that maximizes the cross-validated performance. The final model is subsequently refitted on the full training set using the selected hyperparameters before being used for predictions on the test set. As the analysis is repeated for multiple targets (variable-site pairs and time steps), hyperparameter tuning is performed independently for each target.

From a computational perspective, the workflow considered here consists of two distinct components: (a) the generation of the 400-member high-fidelity WRF/Noah-MP ensemble used to build the data set, and (b) the subsequent surrogate-based analysis. Let T_{WRF} denote the wall-clock time required for one high-fidelity WRF/Noah-MP realization over the simulated period. Because a single WRF/Noah-MP realization simultaneously provides the full set of model outputs, whereas ML-AMPSIT treats each variable-region-level-time combination as a separate scalar target, all surrogate run times reported below are aggregated over the complete analysis set considered in this study, namely 6 atmospheric variables, 4 valley regions, 10 vertical levels, and 24 output times. Accordingly, T_{model} denotes the total runtime required by one surrogate method over the full set of analyzed targets, rather than the cost of a single scalar prediction.

For all surrogate methods, the available 400 WRF/Noah-MP realizations are split into 70% for training and 30% for testing on unseen cases.

For methods in which Sobol indices are explicitly computed after training (SVM, BRR, and GPR in our implementation), T_{model} includes emulator training, prediction on the unseen 30% subset, and the generation and evaluation of a Sobol-Saltelli sampling design, that is, a structured set of parameter combinations used for variance-based sensitivity decomposition (Saltelli, 2002; Sobol, 1967). According to the Saltelli strategy (Saltelli et al., 2010), the number of model evaluations required to estimate first- and second-order indices is $N_{\mathrm{eval}} = N_{\mathrm{Sobol}}(2p + 2)$, where p is the number of uncertain parameters and N_{Sobol} is the base sample size. For a representative configuration with $p \approx 20$ and $N_{\mathrm{Sobol}} = 10^3$, this yields $N_{\mathrm{eval}} \approx 4 \times 10^4$. In a direct high-fidelity Sobol analysis, this would correspond to approximately 4×10^4 WRF/Noah-MP simulations. In the present workflow, by contrast, 400 high-fidelity simulations are used to train the surrogates, and the subsequent Sobol evaluations are performed with the emulators at negligible cost. Hence, the high-fidelity computational burden is reduced from approximately 4×10^4 WRF simulations to 400.

As an illustrative example, if one WRF/Noah-MP simulation requires $T_{\mathrm{WRF}} \approx 6$ h, then a direct high-fidelity Sobol analysis would take $\sim 2.4 \times 10^5$ h ($\sim$27 years) if executed sequentially. By contrast, the surrogate-based workflow requires 400 high-fidelity simulations for data set generation (i.e., $\sim 2.4 \times 10^3$ h, or $\sim$100 days sequentially), while the subsequent surrogate-based analysis is completed within 6–18 hr, depending on the emulator. These results show that surrogate-based sensitivity analysis can make variance-based GSA computationally tractable for NWP applications.

To the best of the authors' knowledge, no previous study has reported a comparable reduction in computational cost that makes variance-based Sobol analysis practical for NWP simulations.

For methods that do not rely on variance-based GSA (RF, XGB, and LASSO), T_{model} represents the time required to train the emulators, predict the unseen 30% subset, and extract feature-importance measures from the fitted models. Although these feature-importance measures are intrinsically model-based and are not obtained through an explicit variance-based decomposition, the results presented in this study show that their rankings are in close agreement with those obtained from the Sobol analysis. From a computational perspective, these methods also use the same 400-member high-fidelity ensemble and therefore avoid the additional cost of the high-fidelity simulations required for a direct Sobol analysis, as discussed above.

2.3. Noah-MP Land Surface Parameters

The Noah-MP parameters whose importance is evaluated in the present study include only those related to vegetation, defined in the look-up table (MPTABLE) and categorized on the basis of the vegetation type. Hence, the chosen parameters are strictly related to the canopy properties and have a clear role in the computation of radiation, heat and momentum fluxes. In the following two sub-sections, a concise presentation of the role played in Noah-MP by the parameters analyzed in this work on surface fluxes is provided, with the aim of aiding the interpretation of the results. For a complete overview, the reader is referred to the Noah-MP user's guide (He et al., 2023b).

2.3.1. Relevant Parameters for Radiative Fluxes

Leaves play an important role in the final canopy radiation budget. Their cellular structure tends to almost completely scatter solar radiation, resulting in relatively equal amounts of reflected and transmitted radiation (Oke, 1987). 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 LAI. Stems also play a role in the radiation budget, but, especially for forests, the main radiation exchange comes from the canopy itself, whereas trunks are less important.

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 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 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*:

$$A = 0.5 \cdot (\rho + \tau) \cdot \frac{\left[\frac{\phi_1 + \phi_2 \cdot \cos\theta'}{\phi_1 + 2 \cdot \phi_2 \cdot \cos\theta'}\right] \cdot \left[1 - \frac{\phi_1 \cdot \cos\theta'}{\phi_1 + 2 \cdot \phi_2 \times \cos\theta'} \cdot \log\left(\frac{\phi_1 \cdot (1 + \cos\theta') + 2 \cdot \phi_2 \cdot \cos\theta'}{\phi_1 \cdot \cos\theta'}\right)\right]}{(\phi_1 + 2 \cdot \phi_2 \cdot \cos\theta')} \tag{7}$$

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:

$$\theta' = \tan^{-1}\left[\left(\frac{HVT - HVB}{2RC}\right) \cdot \tan(\theta)\right] \tag{8}$$

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 (*NIR*) and the visible (*VIS*) bands for both leafs and stems, taking into account the different response of vegetation to these two wavelength ranges, that is, Noah-MP uses the following 8 parameters, *RHOL_NIR*, *RHOS_NIR*, *RHOL_VIS*, *RHOS_VIS*, *TAUL_NIR*, *TAUS_NIR*, *TAUL_VIS*, and *TAUS_VIS*, as defined in Table 1.

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 (Jarvis et al., 1976), and ultimately the transpiration contribution to latent heat. It is expected that, in the present case study, this transpiration contribution is the most significant for the total latent heat flux above the canopy, 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:

$$F_{rs,solar} = \max\left(0.0001, \frac{\frac{2 \cdot S_{par}}{RGL} + \frac{RS}{RSMAX}}{1 + \frac{2 \cdot S_{par}}{RGL}}\right) \tag{9}$$

where S_{par} is the photosynthetically active radiation, which has a different value for sunlit and shaded leaves;

$$F_{rs,temp} = \max\left(0.0001, 1 - 0.0016 \cdot (TOPT - T_v)^2\right) \tag{10}$$

where T_v is the canopy temperature;

Table 1
Parameters Analyzed in This Work

Parameters	Definition	Physical processes	Mixed-forest ref.	Grassland ref.
LAI	leaf area index, one-sided	Rad / Extinction / Stomata	2.20	0.36
SAI	stem area index, one-sided	Rad / Stomata	0.40	0.20
RHOL_NIR	near-infrared leaf reflectance	Rad / Stomata	0.45	0.4
RHOL_VIS	visible leaf reflectance	Rad / Stomata	0.10	0.09
RHOS_NIR	near-infrared stem reflectance	Rad / Stomata	0.39	0.41
RHOS_VIS	visible stem reflectance	Rad / Stomata	0.16	0.21
TAUL_NIR	near-infrared leaf transmittance	Rad / Stomata	0.25	0.19
TAUL_VIS	visible leaf transmittance	Rad / Stomata	0.05	0.05
TAUS_NIR	near-infrared stem transmittance	Rad / Stomata	–	0.13
TAUS_VIS	visible stem transmittance	Rad / Stomata	–	0.07
XL	leaf/stem orientation index	Rad / Stomata	0.25	0.15
RC	tree crown radius	Rad	1.40	0.15
HVT	height of canopy top	Rad / Extinction / Turb / Leaf BL	16.0	1.46
Z0MVT	momentum roughness length vegetated surface	Extinction / Turb / Leaf BL	0.80	0.13
DLEAF	characteristic leaf dimension	Leaf BL	0.040	0.040
CWPVT	empirical canopy wind parameter	Extinction / Leaf BL	0.29	1.49
RS	Jarvis minimum stomatal resistance	Stomata	125	232
RSMAX	Jarvis maximum stomatal resistance	Stomata	5,000	5,000
RGL	Jarvis radiation stress parameter	Stomata	30	200
TOPT	Jarvis optimum transpiration air temperature	Stomata	298	298
HS	Jarvis vapor pressure deficit parameter	Stomata	51.93	146.15

Note. The main physical processes and reference values for each parameter are given.

$$F_{rs,vpd} = \max\left(0.01, \frac{1}{1 + HS \cdot \max(0, w_{can\_air\_sat} - w_{can\_air})}\right) \tag{11}$$

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

The final stomata resistance is:

$$R_{stomata} = \frac{RS}{F_{rs,solar} \cdot F_{rs,temp} \cdot F_{rs,vpd} \cdot \beta_{tr}} \tag{12}$$

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*:

$$R_{bl} = 50 \cdot \frac{a}{1 - e^{-a/2}} \cdot \sqrt{\frac{DLEAF}{U_c}} \tag{13}$$

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

$$a = \left[CWPVT \cdot (LAI + SAI) \cdot HVT \cdot \psi_h\left(\zeta_{MO}\right)\right]^{0.5} \tag{14}$$

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

$$U_c = U_{\mathrm{ref}} \frac{\ln\left(\frac{HVT - d + Z0MVT}{Z0MVT}\right)}{\ln\left(\frac{z_{\mathrm{ref}}}{Z0MVT}\right)} \tag{15}$$

where k is the von Kármán constant, $d = 0.65 \cdot HVT$ is the zero-plane displacement height, and U_{ref} 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.

The momentum roughness length is also used to define the exchange coefficients for momentum:

$$C_m = \frac{k^2}{\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_m\right]^2}, \tag{16}$$

and heat:

$$C_h = \frac{k^2}{\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_m\right]\left[\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_h\right]}, \tag{17}$$

that are used to compute the surface fluxes.

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

$$C_{SH,cond,leaf} = \frac{2 \cdot EVAI}{R_{bl}} \tag{18}$$

$$C_{LH,cond,tr} = \left[\frac{EVAI_{sun}}{R_{bl} + R_{stomata,sun}} + \frac{EVAI_{shd}}{R_{bl} + R_{stomata,shd}}\right] \tag{19}$$

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 shade.

Larger LAI and SAI values imply that the leaf boundary resistance R_{bl} decreases, resulting in higher heat conductance coefficients. On the other hand, larger values of the characteristic leaf dimension $DLEAF$ increase R_{bl}, and thus decrease the heat conductance coefficients, because larger leaves produce thicker laminar layers (Oke, 1987). 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 and latent heat fluxes are also modulated by the turbulent transfer coefficient $C_{h,turb}$, entering the computation of the below-canopy aerodynamic resistances, which is directly controlled by the canopy height HVT and the roughness length $Z0MVT$ through the friction velocity u_*:

$$C_{h,turb} = \kappa u_*(HVT - d) \tag{20}$$

where the friction velocity is defined as:

$$u_* = \frac{U_{\mathrm{ref}}\kappa}{\ln\left(\frac{z_{\mathrm{ref}} - d}{Z0MVT}\right) - \psi_m}, \tag{21}$$

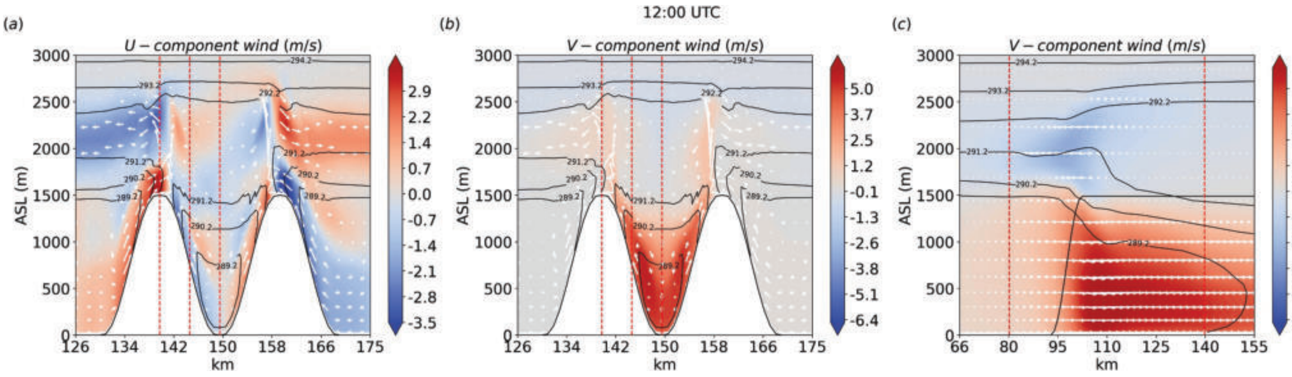


Figure 2. (a) Cross-valley section of the west-east wind component u (color shading), the wind component parallel to the section (arrows) and potential temperature (solid lines); (b) cross-valley section of the south-north wind component v (color shading), wind component parallel to the section (arrows) and potential temperature (solid lines); (c) along-valley section of the south-north wind component v (color shading), wind component parallel to the section (arrows) and potential temperature (solid lines). All variables refer to 12:00 UTC of the first simulated day. The red dotted lines indicate the position of the points considered in the following analyses (see Figure 1b). 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. See Figure 1 for the position of the sections.

and ψ_m is the Monin-Obukhov stability correction for momentum evaluated below the canopy.

Equation 20 implies that both increased roughness and increased canopy height increase the sensible heat exchange between the canopy and the atmosphere.

2.3.3. Investigated Parameters

Table 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 forests for both near-infrared and visible bands, the respective parameters of the MPTABLE *TAUS_NIR* and *TAUS_VIS* are only included in the sensitivity analysis for the grassland land cover class. Moreover, also the parameter controlling the bottom height of the canopy (*HVB*) is not taken into account, in order to avoid unphysical simulations resulting from the contemporary variation of the height of the canopy top (*HVT*).

Table 1 also shows the reference value of each parameter based on the MODIS land cover classes "mixed-forest" and "grassland." In this work, the percentage of variation adopted for the sensitivity analysis for the pseudo-random ensemble is set as 20% for all the parameters. The optimal temperature parameter *TOPT* was 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 mixed-forest- and grassland-like land cover classes. This ensures that the results are physically consistent and as realistic as possible.

3. Results

3.1. Valley Circulation

This section presents the results of the reference simulation, that is, with all the parameters set to the reference value, with the aim of describing its main features, considering in particular the development of the thermally driven circulations in the idealized valley. Results are presented at four time points representative of the diurnal cycle of both slope and valley winds, that is, 12:00, 18:00, 00:00, and 06:00 UTC and considering the cross-valley and along-valley sections shown in Figure 1b. The results shown here refer to the mixed-forest vegetation reference simulation. However, the grassland reference run shows a similar behavior and thus the description provided here applies to both vegetation classes.

At 12:00 UTC, up-slope winds are well-developed along both sidewalls, converging at the ridge crests and are topped by return branches, causing subsidence in the valley center (Figure 2a). In the along-valley direction, the

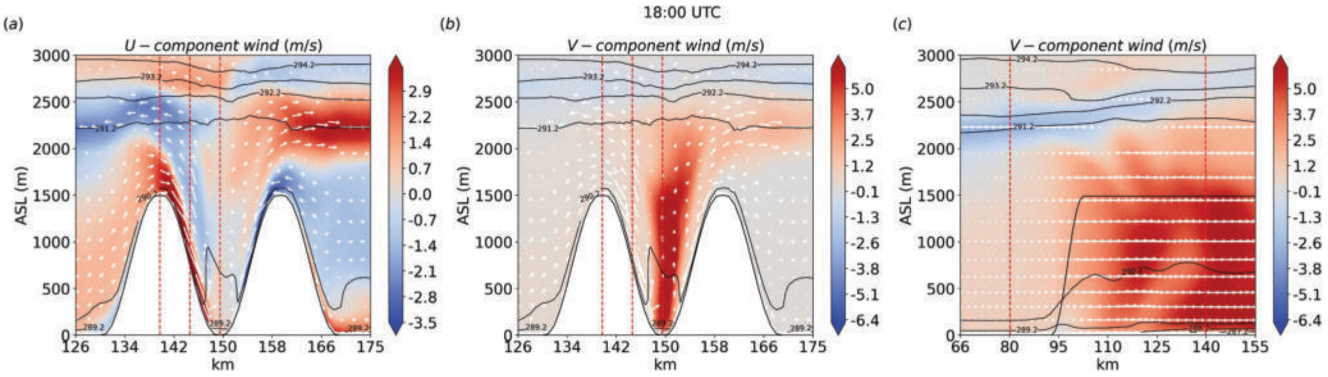


Figure 3. As Figure 2, but referring to 18:00 UTC.

up-valley wind is present up to the ridge top, with intensities around 4 m s^{-1} (Figure 2b). The along-valley section (Figure 2c) highlights higher potential temperature in the valley compared to the plain, which drives the up-valley wind. By 18:00 UTC, down-slope winds are already well-developed, especially along the western valley sidewall, due to earlier orographic shading in the afternoon (Figure 3a). Since the valley is north-south oriented, the main asymmetries in the circulation result from the differential insolation of the sidewalls, which strongly affects the development of slope winds, given their rapid response to surface heating and cooling. It is observed that a circulation converging toward the valley persists close to the external slopes in the late afternoon (Figure 4a). This feature is the remnant of the broader plain-to-mountain, thermally driven circulation induced by daytime heating. Although local downslope winds in the valley respond quickly to shading and radiative cooling, this larger-scale inflow component can decay more slowly. This leads to the temporary coexistence of downslope flow close to the surface and converging circulations near the valley slopes facing the plain, as also highlighted in the idealized numerical simulations presented in de Wekker et al. (1998) and Schmidli and Rotunno (2010). This behavior is consistent with the "Alpine pumping" mechanism, well-documented in the literature (e.g., Weissmann et al., 2005), whereby daytime heating over a mountain chain generates a mesoscale pressure deficit that induces compensating inflow from the surrounding flat area. This inflow can persist beyond the onset of the local evening slope-wind transition. At 18:00 UTC, the up-valley wind is still present, but it starts to weaken. It extends up to 2,000 m AGL, following the vertical extension of the well-mixed boundary layer (Figure 3b). In the lowest layers, a surface inversion is developing.

At 00:00 UTC, shallow down-slope winds converge toward the valley floor, contributing to the growth of the cold pool (Figure 4a). A weak down-valley wind starts to develop in the lowest atmospheric layers, topped by a residual up-valley wind above 1,000 m AGL (Figure 4b). The along-valley section (Figure 4c) highlights that the down-valley wind is mostly developed near the valley exit, driven by lower temperatures in the valley. At 06:00 UTC, the down-valley wind has strengthened, extending to higher levels and toward the inner valley sections

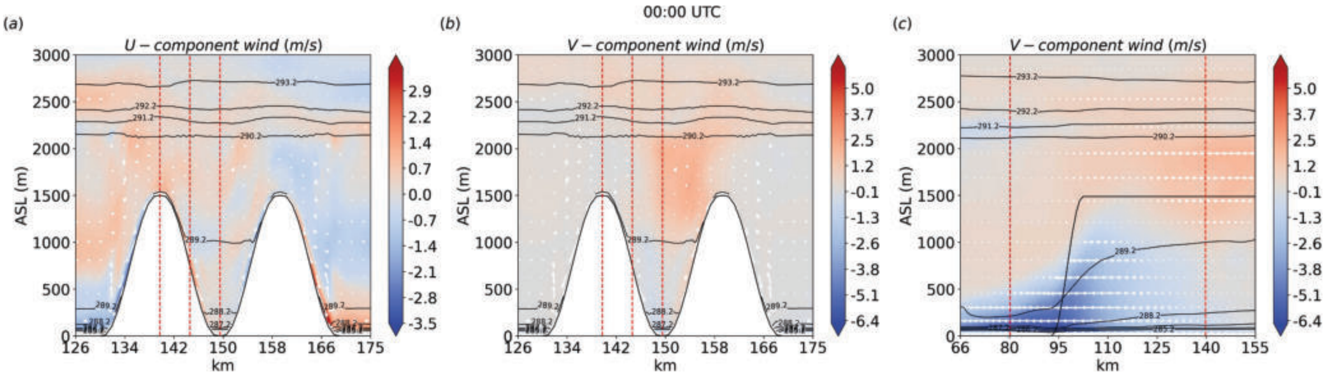


Figure 4. As Figure 2, but referring to 00:00 UTC.

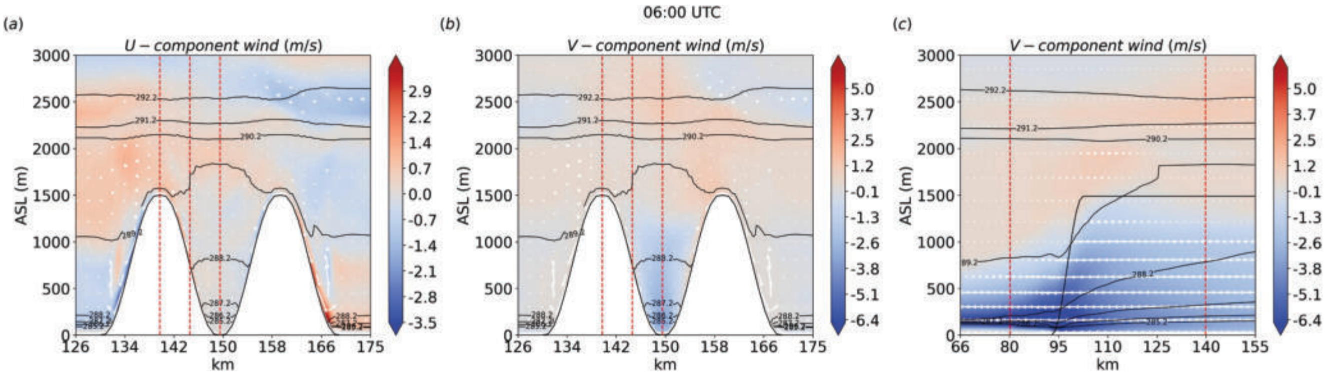


Figure 5. As Figure 2, but referring to 06:00 UTC.

(Figures 5b and 5c). Down-slope winds weaken close to the valley floor, where cold air has accumulated overnight, further stabilizing the lowest valley levels (Figure 5a).

3.2. Ensemble Variability

The variability of the simulation ensemble is evaluated by analyzing the time series of the along-valley (south-north, *v*) and cross-valley (west-east, *u*) wind components, of the TKE and the potential temperature at the lowest model level, as well as of the surface latent and sensible heat fluxes (Figures 6–11) 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 Figure 1b). Moreover, for the same four points, the ensemble results are also shown for the vertical profiles of the horizontal wind components, TKE and potential temperature in the lowest 500 m AGL at the same four representative time steps analyzed in Section 3.1 (Figures 12–14). Both the time series and the vertical profiles are spatially averaged over 5 cells along the south-north (y) direction, from 2 cells south to 2 cells north of the selected point, to obtain more robust results.

Focusing on the ensemble time series (Figures 6–11), it can be seen that the ensemble spread changes not only with time and location, but also with the variable of interest. These results refer to the simulation ensemble generated for the vegetation class "mixed-forest," which exhibits notable variability across different variables. The "grassland" ensemble (not shown) shows a very similar overall behavior but with considerably less spread, particularly in TKE and sensible heat flux during daily peaks (12:00 UTC), as well as in potential temperature throughout the day, especially on the ridge. However, given the strong similarities, the general discussion focuses on the mixed-forest ensemble, which better highlights the influence of input parameter variability.

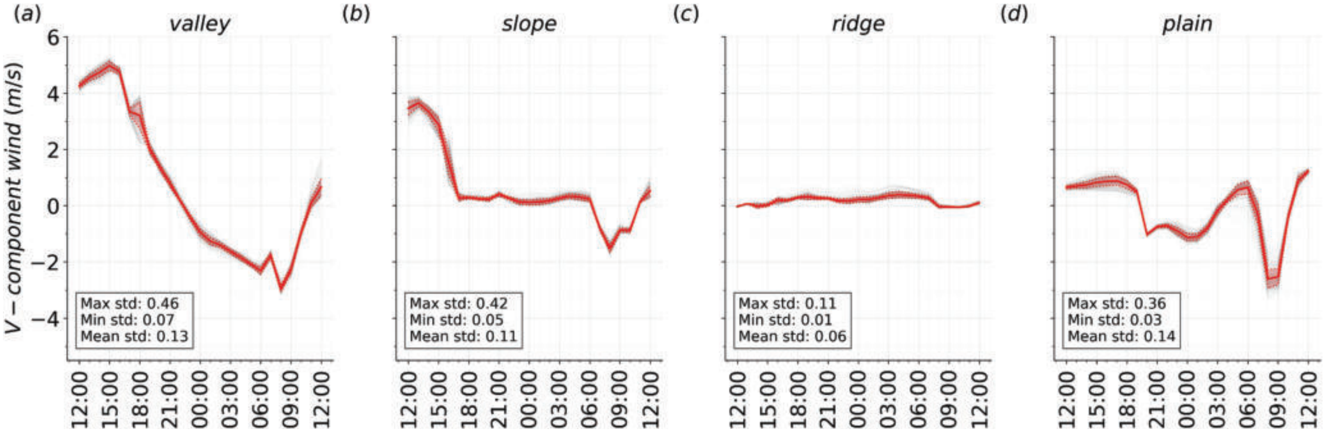


Figure 6. Ensemble time series (gray lines) for the along-valley wind component *v* at the four points shown in Figure 1b, representative of (a) the valley floor, (b) the slope, (c) the ridge and (d) the plain. The continuous and dotted red lines indicate, respectively, the ensemble mean and the standard deviation of the ensemble. The red shaded band denotes ±1σ around the mean. The box indicates the mean, minimum and maximum standard deviation of the ensemble time series.

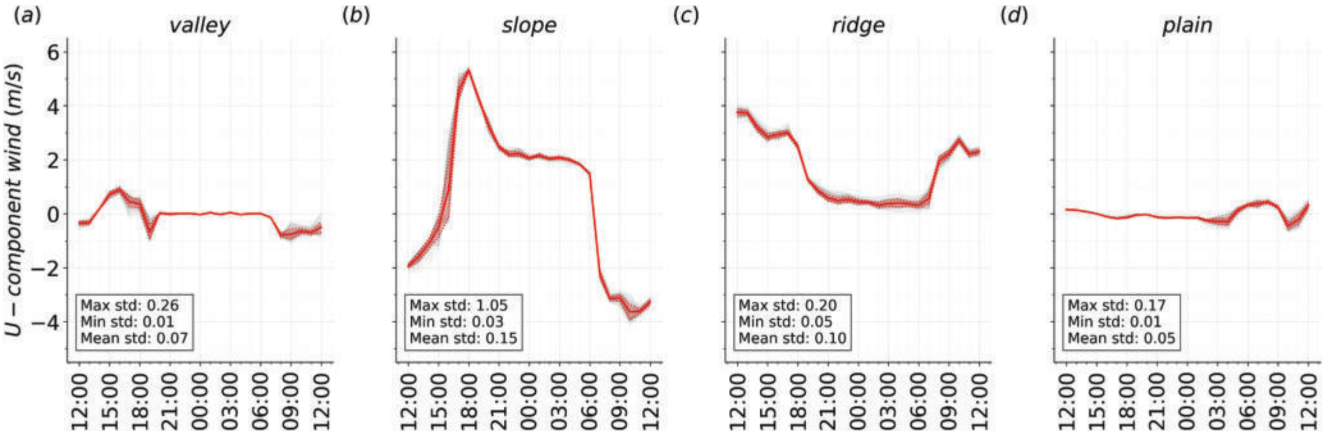


Figure 7. As Figure 6, but for the cross-valley wind component.

The variability of the along-valley wind component (Figure 6) is comparable across the regions, with a general increase of the spread in the afternoon over the valley, the plain and the slope, and during the night over the ridge. The spread over the plain is also larger in the early morning, when this location is affected by the northerly wind exiting from the valley (see Figures 4 and 5). The ensemble spread of the cross-valley wind (Figure 7) shows similar patterns, with the exception of a significantly lower spread over the valley during nighttime and the plain during daytime.

The variability of the potential temperature (Figure 8) is in general stronger; it tends to be larger during daytime on the valley floor, whereas an increase in the ensemble spread is also visible in the early morning in the plain, coherent with the increase of the variability of the south-north wind component (Figure 6). On the other hand, the ensemble spread is roughly constant and large throughout the day over the slope and the ridge. The ensemble spread of the TKE (Figure 9) and the latent and sensible heat fluxes (Figures 10 and 11) show strong differences between nighttime, when the variability is very low, and daytime, when differences between the ensemble simulations are considerably larger. This behavior reveals the significant impact of the surface parameters on daytime turbulence development and surface-atmosphere turbulent exchange, whereas the impact is lower during nighttime, when the stable stratification dampens turbulence. The TKE shows a peak, with an increase of the spread, also at 19:00 UTC over the slope, corresponding to the peak of the downslope wind, which causes a mechanical generation of turbulence.

Considering the vertical profiles of the along-valley wind component, Figure 12 shows that the ensemble spread is larger in the evening and at night on the valley floor, where the variability tends to increase at higher levels at 18:00 UTC. On the other hand, the variability of the along-valley wind component is significantly lower over the

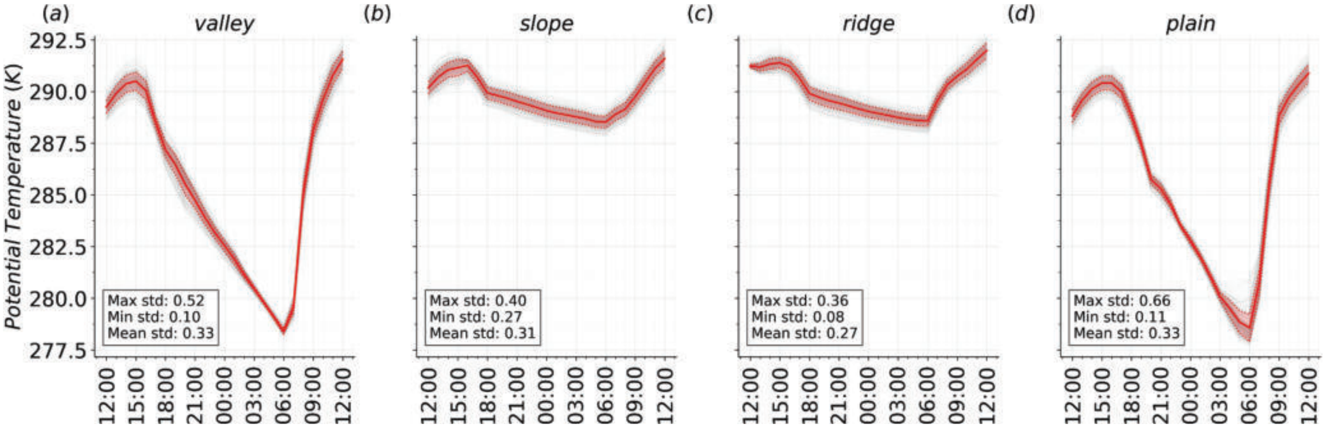


Figure 8. As Figure 6, but for the potential temperature.

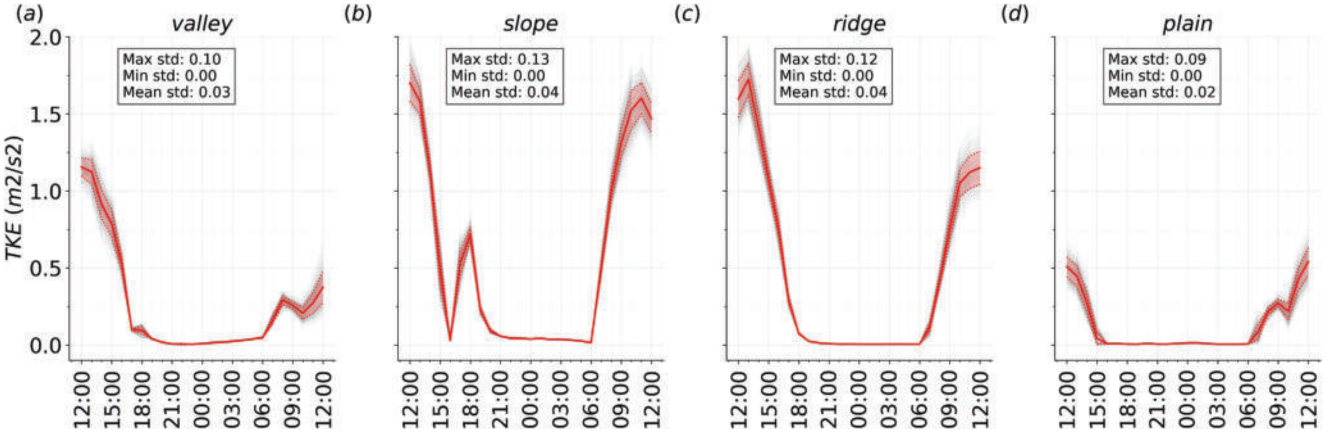


Figure 9. As Figure 6, but for the turbulent kinetic energy.

other regions. Therefore, the variability of v seems to primarily reflect differences in the timing of the evening transition of the along-valley wind and in the strength of the nocturnal down-valley wind, which result to be parameter-sensitive. The cross-valley wind component (Figure 13) shows strong variability at 18:00 UTC, that is, in the first phase of the downslope wind, at all levels over the slope, at higher levels over the ridge, and in the lowest layers over the valley floor. This likely highlights the impact of the parameters on the timing of the development of the downslope wind. The enhanced spread at 12:00 and 18:00 UTC in the slope and ridge u profiles suggests parameter-dependent variations in the elevated cross-valley component associated with the slope return flow. On the other hand, over the plain, the variability of the cross-valley wind component is particularly low.

The ensemble spread of the potential temperature is consistently large without a significant height dependence for all the locations and times considered, with the exception of a lower variability over the ridge at 12:00 UTC (Figure 14). Finally, the TKE profiles (Figure 15) show a significantly high spread at 12:00 UTC over all the points and over the entire vertical column, and a generally small spread otherwise, except over the slope, where it retains a large spread also at 18:00 UTC, again probably due to the impact of the parameter changes on the development of the downslope wind. In particular, this suggests that the perturbed parameters modulate the intensity and vertical extent of boundary-layer turbulence during the daytime convective phase, thereby feeding back on the subsequent transition of the slope cross-valley circulation.

The sensitivity in the timing of the onset and decay of slope and valley circulations observed in this ensemble is intrinsically linked to the online land–atmosphere coupling. Parameter-induced changes in surface fluxes modulate boundary-layer stability and development, thereby affecting the intensity of slope and valley winds and

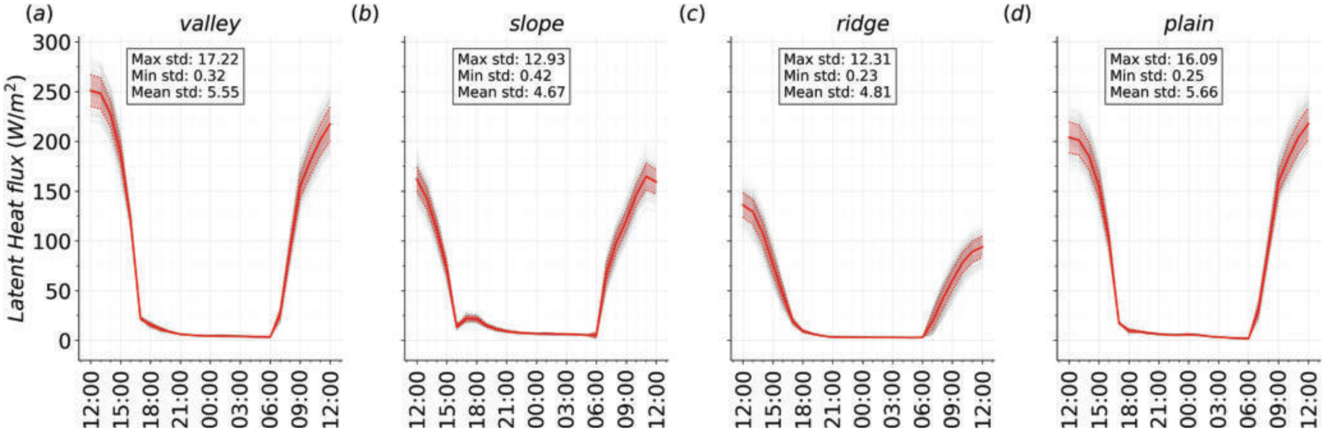


Figure 10. As Figure 6, but for the latent heat flux.

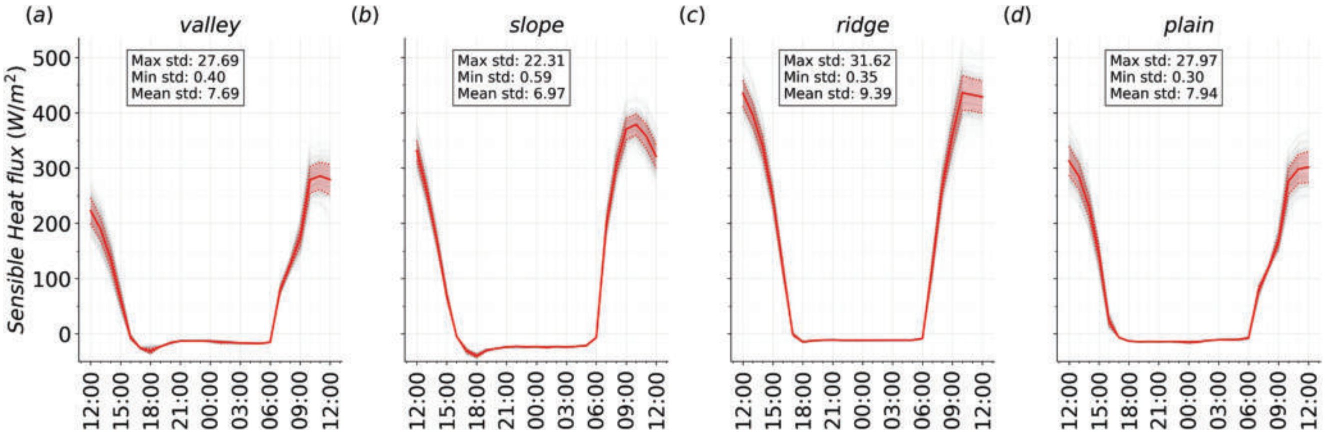


Figure 11. As Figure 6, but for the sensible heat flux.

the timing of their transitions. Such feedback mechanisms cannot be captured in offline Noah-MP sensitivity experiments, where atmospheric circulation is prescribed rather than dynamically simulated. To the best of the authors' knowledge, existing Noah-MP parameter-sensitivity studies have been conducted exclusively in offline configurations. The present results therefore provide the first systematic assessment of how land–atmosphere interactions influence the characteristics of thermally driven circulations in a mountain valley.

3.3. Sensitivity Analysis

This section presents the results of the sensitivity analysis, focusing in particular on the outcomes of the assessment of the parameter importance for both mixed forest and grassland land cover classifications. 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, as above, the analysis focuses on the four locations representative of the valley floor, the slope, the ridge and the plain (Figure 1b). To obtain robust results, also in this case, the analyses are performed by averaging model results along the south-north direction y over 5 cells. In this section, results from the analyses referring to the lowest vertical level are presented. The analysis referring to higher levels was also performed, but since the main patterns do not show a significant variation with height, the figures related to the other vertical levels analyzed are provided in the supplemental material.

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

3.3.1. Performance Metrics

The performance metrics are reported in Figure 16 for all the points and variables considered. The results refer to the average of all the surrogate models used. In particular, the metrics (R^2, MSE, and MAE) are estimates of how well the methods performed during the test phase, that is, when predicting new, unseen data. In order to ensure comparability, metrics are calculated from standardized, non-dimensional data. (Figure 17).

Cells with red borders indicate cases in which average metrics are not computed due to missing predictions in several algorithms. To avoid misinterpretation, these cases are automatically set as blank. This applies to the TKE over the plain and valley at night in both the mixed forest and the grassland ensemble, where the time series values are mostly zero (not shown), with no variability within the ensemble. Under these conditions, even minimal deviations between the predicted and observed values result in unreliable metrics (e.g., division by near-zero variance in the calculation of R^2, inflated relative errors in MAE/MSE).

For the majority of the variables and points, the results highlight that R^2 is greater than 0.85, MSE is lower than 0.3, and MAE is around 0.3–0.4. The algorithms generally display lower performance during nighttime, especially for variables showing low absolute values and variability, such as the TKE or the cross-valley wind component over the valley floor. In particular, lower R^2 values can occur for variables and periods characterized

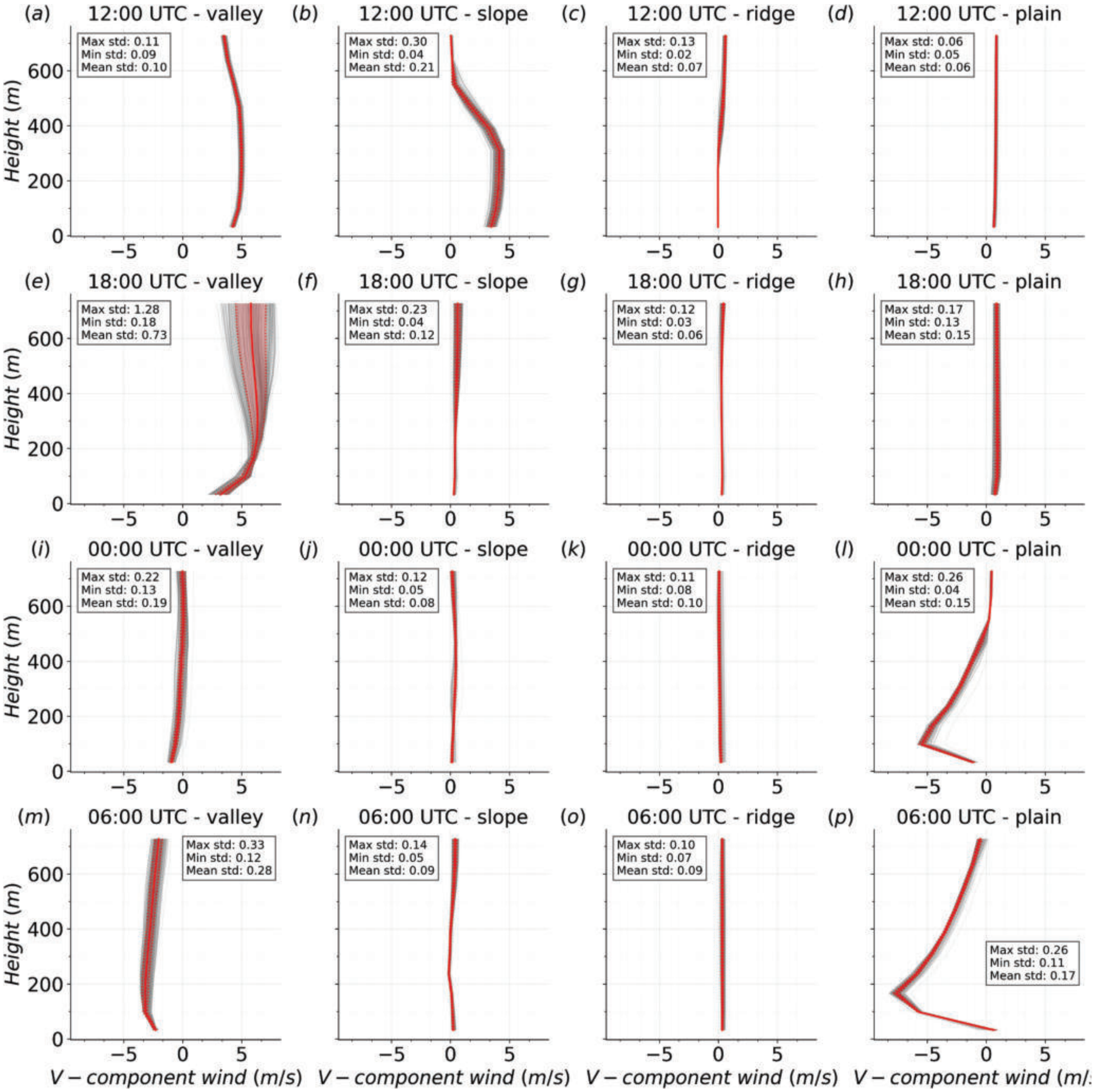


Figure 12. Ensemble vertical profiles (gray lines) in the lowest 500 m AGL for the along-valley wind component at the four points shown in Figure 1b representative of the valley floor (a-(e)-(i)-(m), the slope (b)-(f)-(j)-(n), the ridge (c)-(g)-(k)-(o) and the plain (d)-(h)-(l)-(p) and four different times representative of the daily cycle of valley and slope winds: 12:00, 18:00, 00:00, and 06:00 UTC. The continuous and dotted red lines indicate respectively the ensemble mean and the standard deviation of the ensemble. The red shaded band denotes ±1σ around the mean. The box indicates the mean, minimum and maximum standard deviation of the ensemble time series.

by very small variance (e.g., nighttime u or TKE), where minor surrogate deviations or numerical noise can lead to a substantial reduction in R^2 despite the limited physical relevance of those fluctuations.

It is important to stress again that these metrics are not being used here as a physical measure of the model's ability to predict observations, but rather as an indicator of the surrogate's ability to reproduce the WRF ensemble. 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.

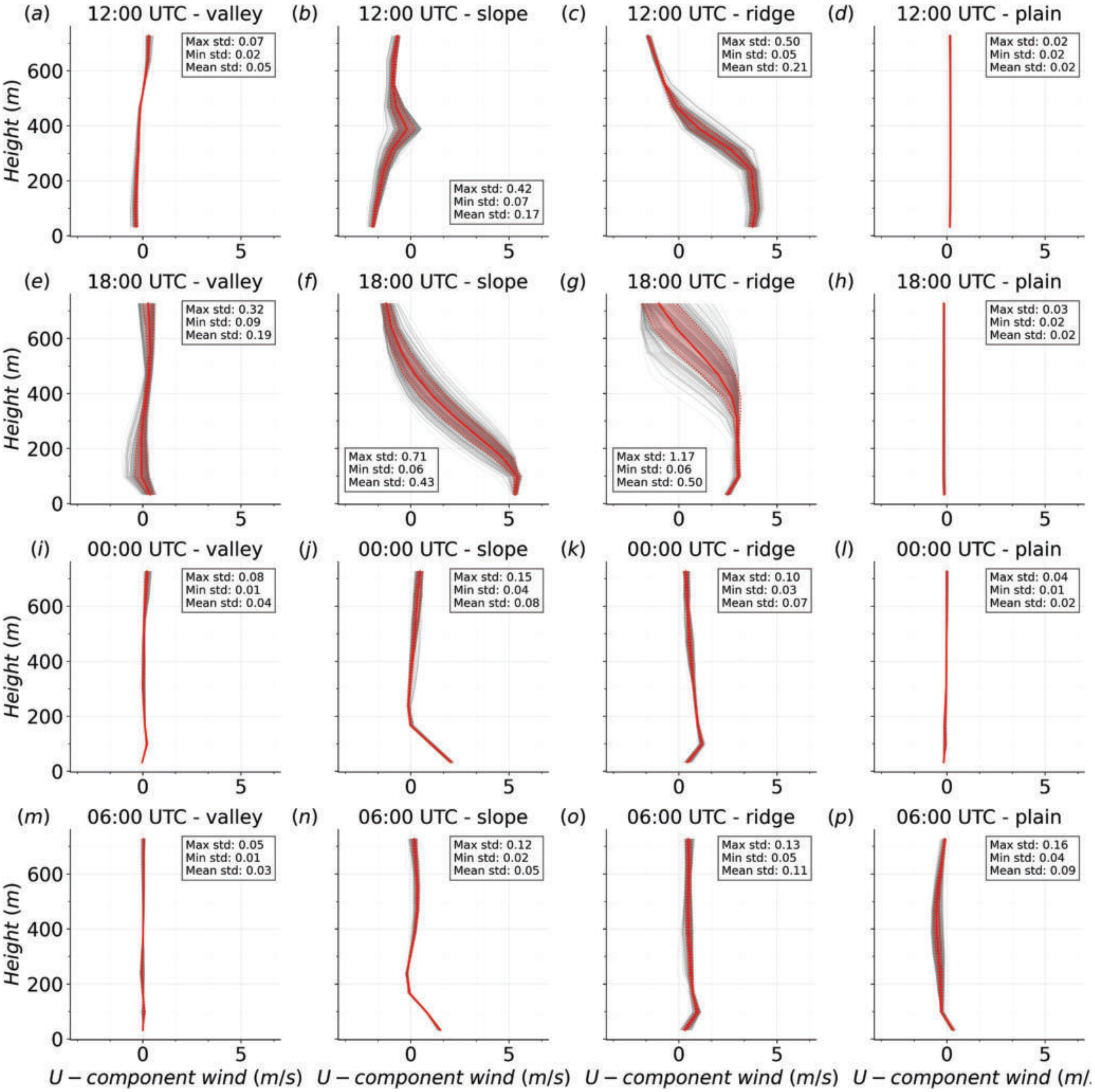


Figure 13. As Figure 12, but for the cross-valley wind component.

3.3.2. Parameter Importance Analysis

Similar to the performance metrics shown in the previous section, the analysis of the 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) highlighted very high consistency across methods.

It is important to emphasize the relative nature of the parameter 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 heatmaps presented in Figures 18–21 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

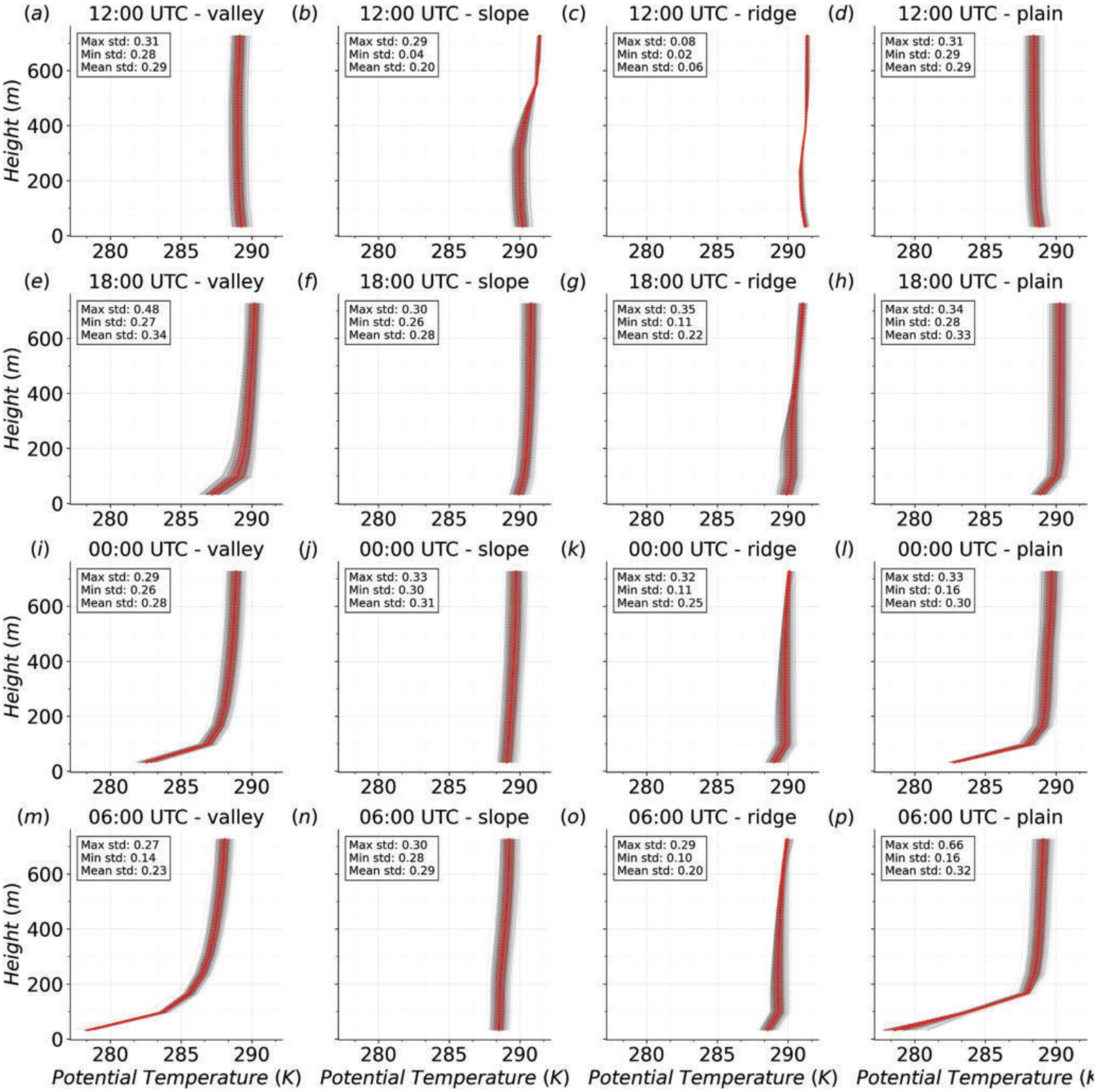


Figure 14. As Figure 12, but for the potential temperature.

parameter can be compared with the importance of the other parameters in the same column of the heatmap (same variable-location pair), to evaluate which parameters mostly affect 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. Moreover, values with a red font indicate the importance estimates associated with $R^2 < 0.6$, emphasizing values for which the model is less reliable. Variables falling below this threshold are u over the plain for grassland during daytime, u over the valley floor for both grassland and mixed forest during nighttime, and TKE over both the plain and the valley floor for grassland during nighttime. These variables present very low values with nearly absent variability within the

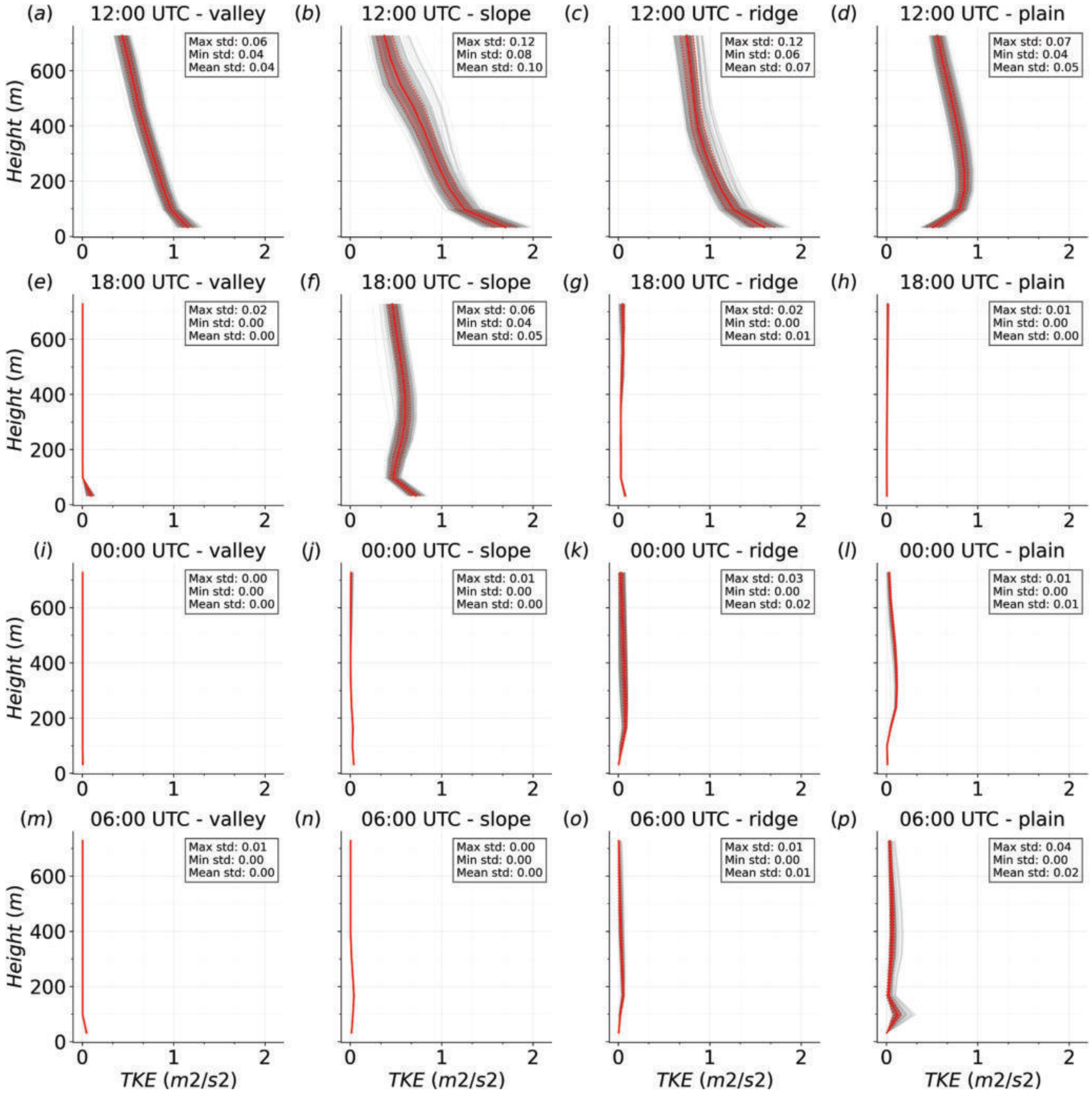


Figure 15. As Figure 12, but for the turbulent kinetic energy.

ensemble (not shown for grassland; see Figures 7 and 9 for mixed forest), suggesting that the underlying patterns are more easily covered by statistical noise, which limits the interpretability of the models. This behavior is consistent with the fact that these variable-region pairs in the above-mentioned time windows are only weakly involved in the valley circulation dynamics, leading to weakly correlated patterns in the time series.

When the parameters are ranked by their importance across regions and variables, a clear hierarchy emerges. For the mixed forest ensemble, the highest importance values are displayed by *RS* for most of the variables and locations, with consistently high importance both during daytime and nighttime. The importance is particularly high for heat fluxes and potential temperature. Given the parameterization of $R_{stomata}$ reported in Equation 12, it was expected that *RS* would have been the most important parameter within the stomata-related group, as it

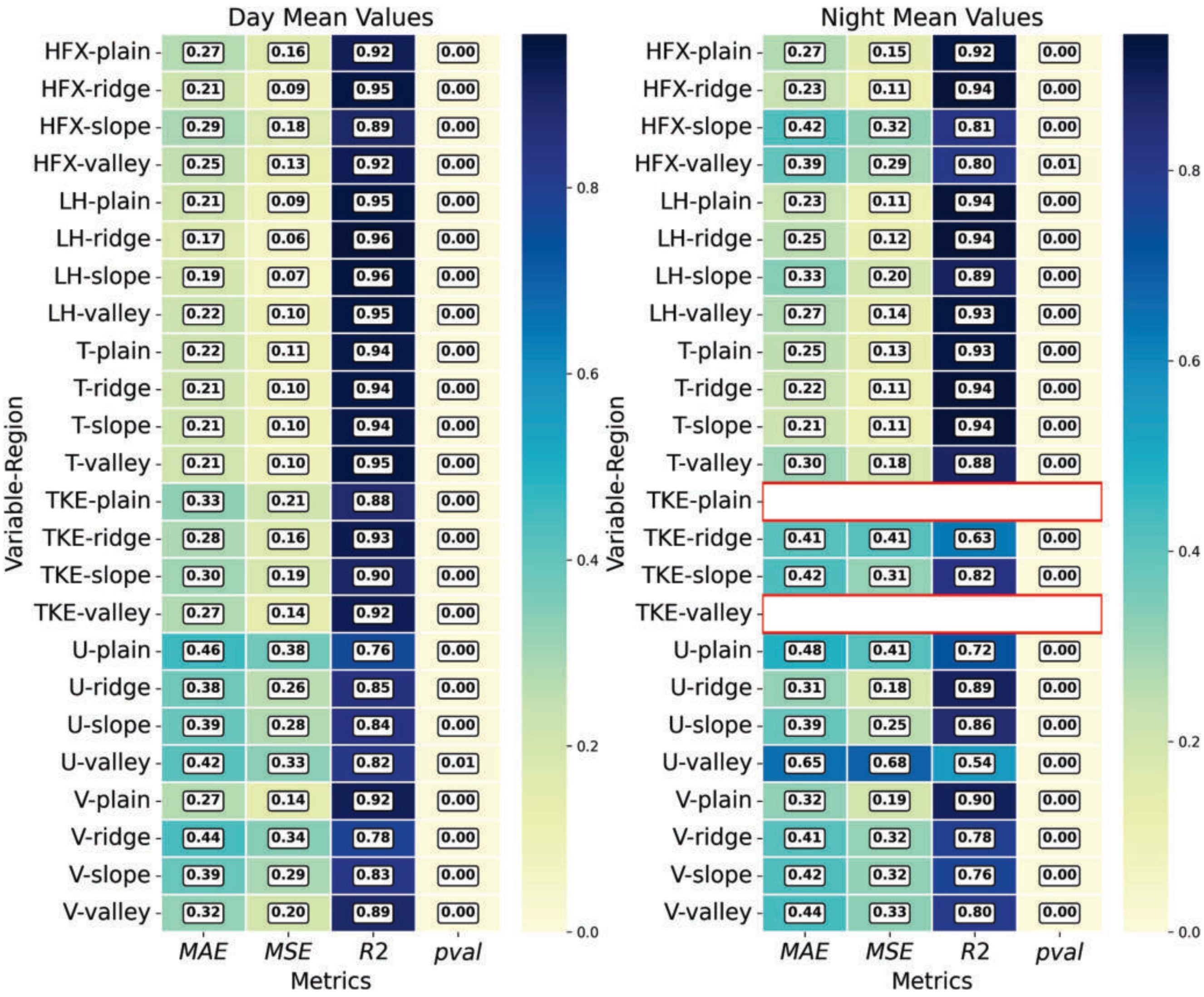


Figure 16. Performance metrics for mixed forest land cover averaged over all the algorithms used in the present study for each variable and point analyzed at the first vertical level. The heatmaps show MAE, MSE, R^2, and p-value for daytime (left, 06:00–17:00) and nighttime (right, 18:00-05:00). The red boxes highlight values that were not computed due to missing predictions.

directly controls transpiration rates. Ultimately, *RS* strongly influences latent heat flux by modulating transpiration, which in turn indirectly affects temperature. In contrast, RS shows weak or negligible importance for the same variable-location pairs in the grassland ensemble, as indicated by the corresponding heatmaps (Figures 20 and 21). This suggests that stomatal control contributes less to the variance of the latent heat flux over grassland than over mixed forest under the present setup and within the tested perturbation ranges.

HVT consistently ranks as one of the most influential parameters for the mixed forest ensemble, showing strong effects on all variables and regions. This result aligns with expectations, as *HVT* is a crucial parameter for canopy forests, affecting radiation trapping and heat retention, which are more pronounced for taller canopies. Additionally, *HVT* strongly affects leaf boundary resistance, making it a key factor in both sensible and latent heat flux conductances (Equations 18 and 19). However, an important distinction between sensible and latent heat conductances is the presence of $R_{stomata}$ in the latent heat conductance, which explains the greater importance of *RS* for the latent heat flux. Unlike the mixed forest case, *HVT* is found to be not influential for the grassland ensemble, where the absence of tall, structured canopies reduces its role in governing heat and radiation exchange processes.

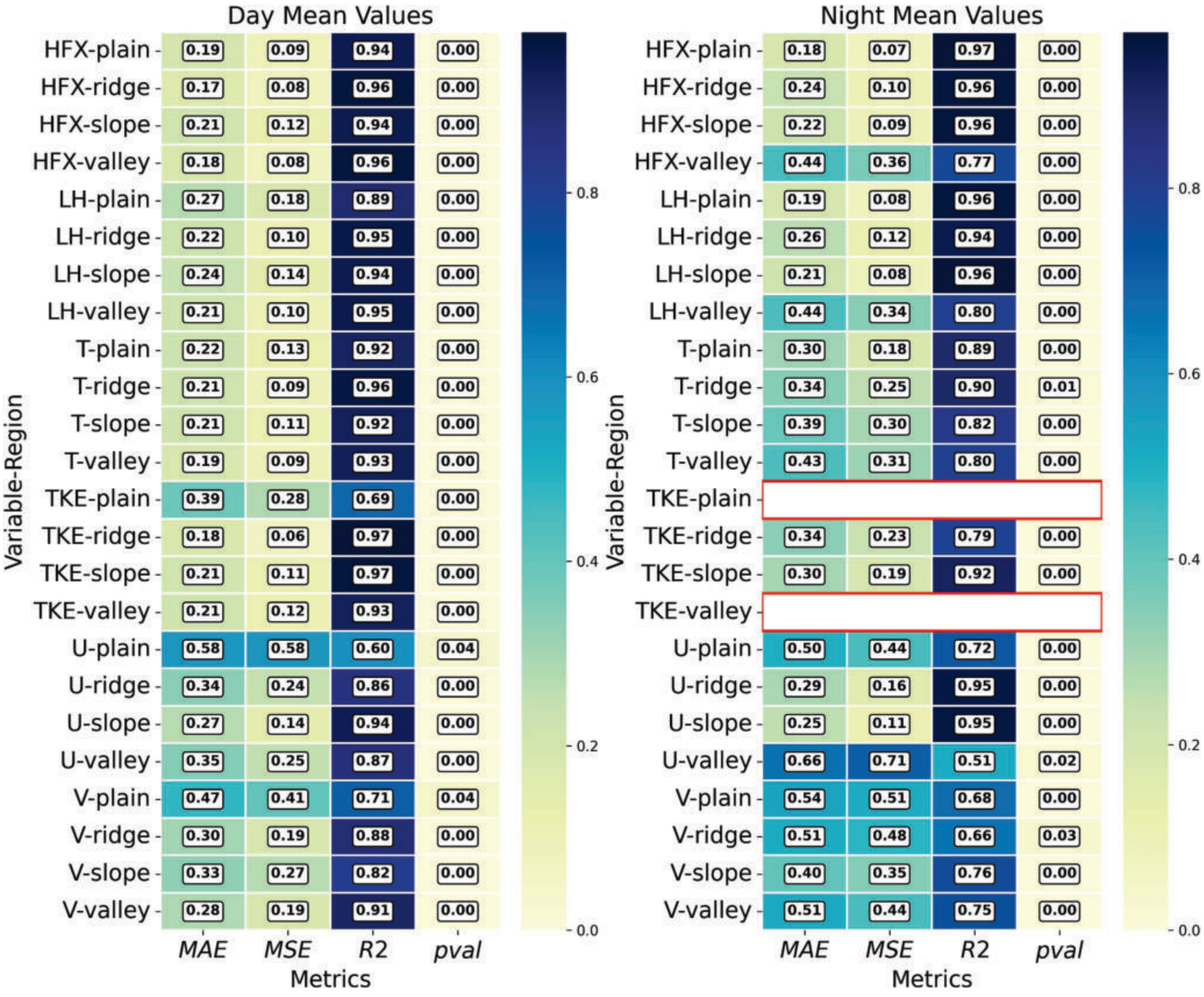


Figure 17. As Figure 16, but for grassland land cover.

Considering the other parameters related to stomatal resistance (Equations 9–12), *RGL* and *RSMAX* display negligible importance for both mixed forest and grassland. *HS*, which represents the weight that vapor saturation has over stomatal resistance, also exhibits negligible importance for both mixed forest and grassland ensembles. *TOPT* shows very strong importance for mixed forest, especially for heat fluxes. On the other hand, it is observed to be negligible for grassland.

For the mixed forest ensemble, when considering radiation-related processes, the only influential parameter turns out to be the near-infrared reflectivity *RHOL_NIR*, which shows a significant influence both during daytime and nighttime, especially on potential temperature. This parameter is also important for the grassland ensemble, further highlighting its role in modulating energy exchanges through radiative effects. The presence of the near-infrared reflectivity as one of the most influential 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.

On the other hand, leaf transmissivities, both in the near-infrared and in the visible (*TAUL_NIR* and *TAUL_VIS*) are observed to be always non-influential parameters for both mixed forest and grassland, as well as stem transmissivities *TAUS_NIR* and *TAUS_VIS* for grassland.

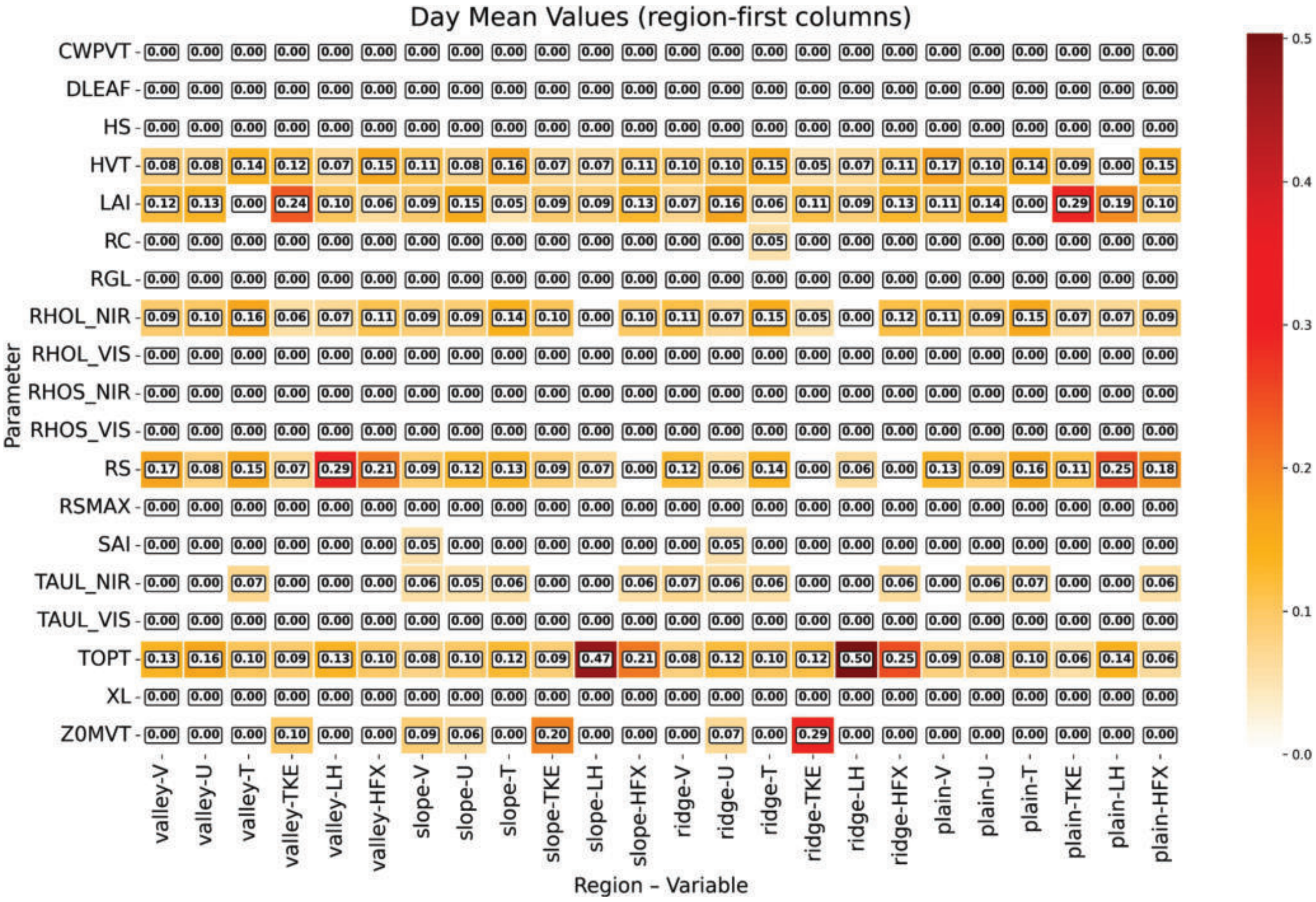
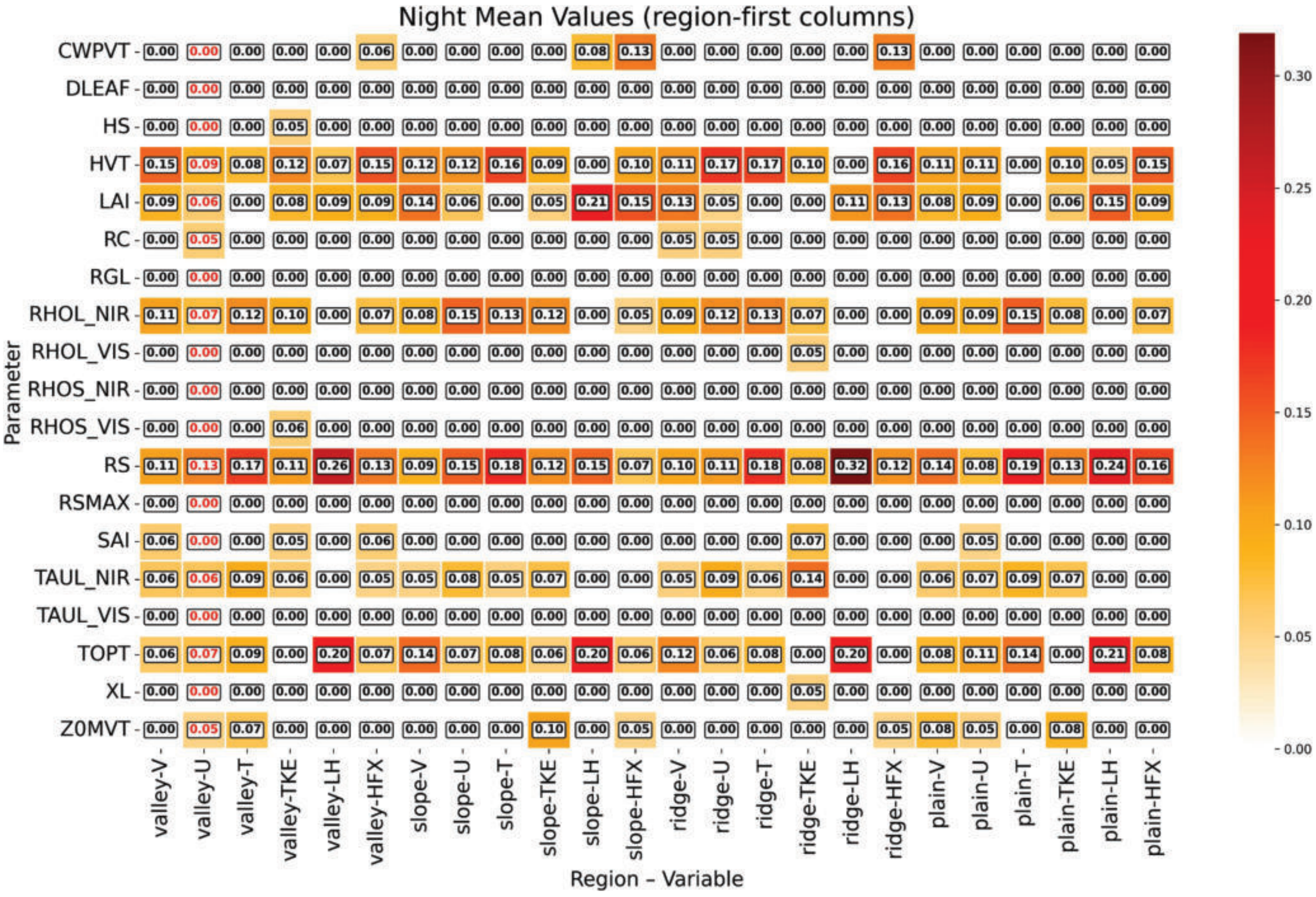


Figure 18. Parameter importance during daytime (06:00–17:00 UTC) for mixed forest land cover computed for the first model level as an average of all the algorithms used in the present study. Values with a red font indicate importance estimates associated with $R^2 < 0.6$.

For the mixed forest ensemble, *LAI* overall displays significant importance for most variables, particularly TKE during daytime and latent and sensible heat fluxes at night. It is important to note that this result applies to a 20% perturbation of *LAI*, whereas across the Noah-MP Modis-based vegetation classes, this parameter varies by more than 100%, potentially leading to a significantly greater impact on simulation outputs. On the other hand, *SAI* is consistently observed as a non-influential parameter in the mixed forest case. Given that the range of variation of *SAI* is significantly lower than that of *LAI*, it was expected that, despite playing a similar role in the Noah-MP code, it would have had a much smaller influence. In contrast, for the grassland ensemble, *LAI* emerges as the most important parameter overall, with *SAI* ranking as the second most influential. This difference can be attributed to the fact that *SAI* values in grassland are more comparable to *LAI*, unlike in forests, where the disparity between the two parameters is much larger. *Z0MVT* appears as an important parameter for TKE, especially during daytime. It also displays significant importance for the wind components.

Finally, some parameters exhibit transient importance throughout the day in some of the ML methods used. This does not appear in this analysis due to the time windows adopted and the averages across different methods. However, here we decided to focus on the parameters that demonstrated persistently higher importance, as these represent the main results that are both consistent across methods and stable over time.

The results obtained in this work share some similarities with findings from other literature studies in which the sensitivity to Noah-MP parameters was tested with offline simulations. In accordance with our results, Cuntz et al. (2016) also showed that *HVT* displayed consistently high importance for the latent heat flux in mixed forest due to its impact on canopy interception and sub-canopy radiation, while almost no sensitivity was found for grassland. Their results also confirm *LAI* as a dominant driver for transpiration across both vegetation types. *RHOL_NIR* was found to be relevant mainly for transpiration in grassland, but not in forests, mirroring our results



Parameter	valley-V	valley-U	valley-T	valley-TKE	valley-LH	valley-HFX	slope-V	slope-U	slope-T	slope-TKE	slope-LH	slope-HFX	ridge-V	ridge-U	ridge-T	ridge-TKE	ridge-LH	ridge-HFX	plain-V	plain-U	plain-T	plain-TKE	plain-LH	plain-HFX
CWPVT	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.08	0.13	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.00	0.00	0.00	0.00
DLEAF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HS	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HVT	0.15	0.09	0.08	0.12	0.07	0.15	0.12	0.12	0.16	0.09	0.00	0.10	0.11	0.17	0.17	0.10	0.00	0.16	0.11	0.11	0.00	0.10	0.05	0.15
LAI	0.09	0.06	0.00	0.08	0.09	0.09	0.14	0.06	0.00	0.05	0.21	0.15	0.13	0.05	0.00	0.00	0.11	0.13	0.08	0.09	0.00	0.06	0.15	0.09
RC	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RGL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHOL_NIR	0.11	0.07	0.12	0.10	0.00	0.07	0.08	0.15	0.13	0.12	0.00	0.05	0.09	0.12	0.13	0.07	0.00	0.00	0.09	0.09	0.15	0.08	0.00	0.07
RHOL_VIS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHOS_NIR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHOS_VIS	0.00	0.00	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RS	0.11	0.13	0.17	0.11	0.26	0.13	0.09	0.15	0.18	0.12	0.15	0.07	0.10	0.11	0.18	0.08	0.32	0.12	0.14	0.08	0.19	0.13	0.24	0.16
RSMAX	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SAI	0.06	0.00	0.00	0.05	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00
TAUL_NIR	0.06	0.06	0.09	0.06	0.00	0.05	0.05	0.08	0.05	0.07	0.00	0.00	0.05	0.09	0.06	0.14	0.00	0.00	0.06	0.07	0.09	0.07	0.00	0.00
TAUL_VIS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOPT	0.06	0.07	0.09	0.00	0.20	0.07	0.14	0.07	0.08	0.06	0.20	0.06	0.12	0.06	0.08	0.00	0.20	0.00	0.08	0.11	0.14	0.00	0.21	0.08
XL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z0MVT	0.00	0.05	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.05	0.08	0.05	0.00	0.08	0.00	0.00

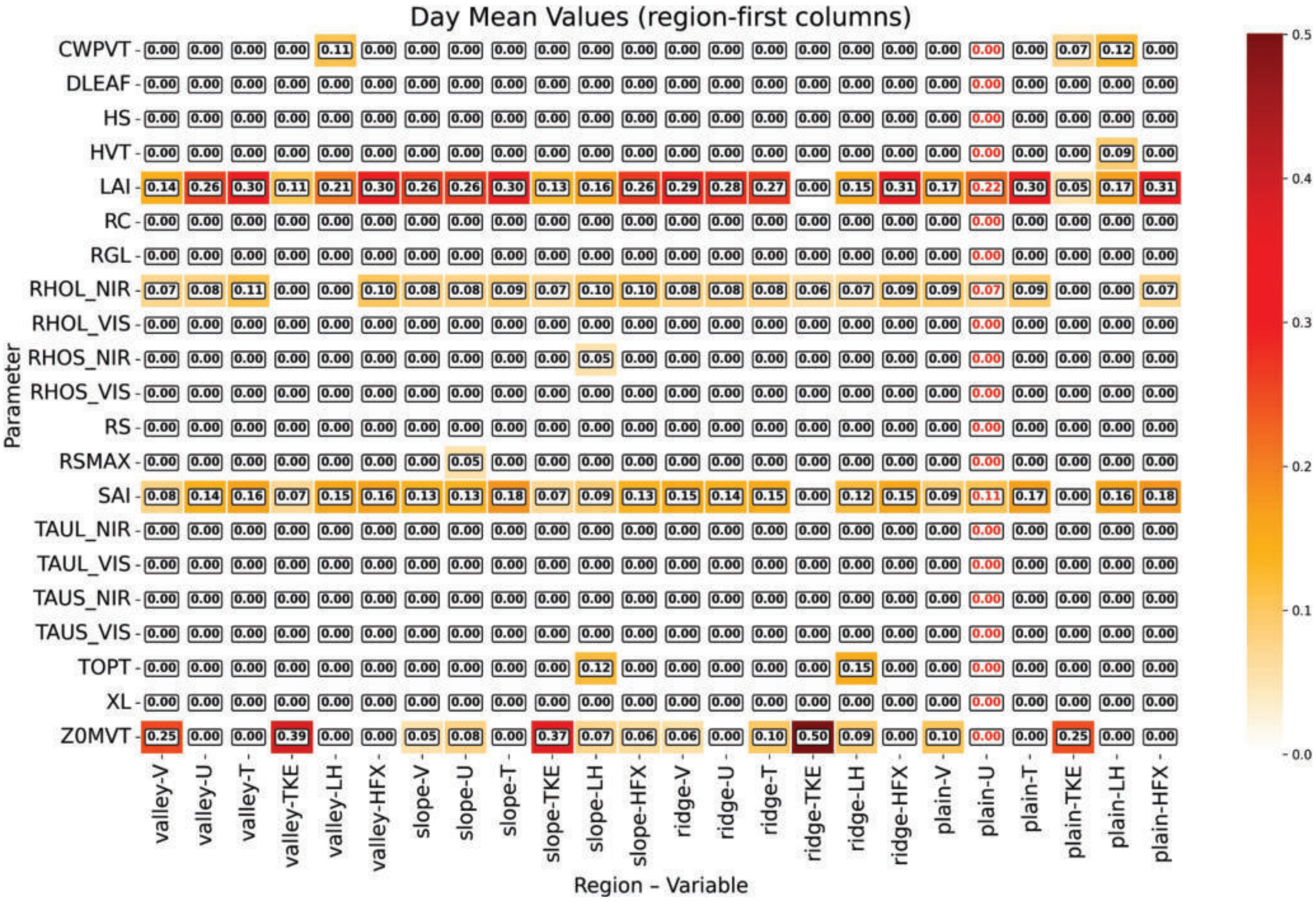
Figure 19. As in Figure 18, but during nighttime (18:00–05:00 UTC).

that show *RHOL_NIR* as significant in some locations and for some variables. Arsenault et al. (2018) also identified *LAI* and *SAI* as substantially affecting the surface energy balance across multiple sites due to their role in determining the total available energy at the surface. They also highlighted the significant importance of the roughness length *Z0MVT* in both grassland and forested sites considering the surface energy fluxes, and in particular that *Z0MVT* and *HVT* both displayed a relevant influence over a mixed deciduous forest site. Furthermore, they observed the relevant sensitivity of the sensible heat flux to *RHOL_NIR* and of the latent heat flux to *TOPT*.

4. Conclusions

This study presented a detailed investigation of the importance of LSM parameters in coupled land-atmosphere simulations over complex terrain, with a focus on thermally driven circulations in an idealized valley. Recognizing that LSMs are a significant source of uncertainties in NWP modeling, 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. We 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 approach allowed a fast evaluation of parameter effects on model outputs, taking into account non-linear relationships and interactions.

To the best of the authors' knowledge, this is the first study to combine ML-based surrogate modeling and variance-based GSA to rank Noah-MP vegetation parameters within a fully coupled WRF configuration targeting thermally driven valley circulations. In contrast to previous sensitivity studies based on offline Noah-MP experiments, the present framework explicitly represents land-atmosphere feedbacks governing the development of



Parameter	valley-V	valley-U	valley-T	valley-TKE	valley-LH	valley-HFX	slope-V	slope-U	slope-T	slope-TKE	slope-LH	slope-HFX	ridge-V	ridge-U	ridge-T	ridge-TKE	ridge-LH	ridge-HFX	plain-V	plain-U	plain-T	plain-TKE	plain-LH	plain-HFX
CWPVT	0.00	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.07	0.12	0.00
DLEAF	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
HVT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.09	0.00
LAI	0.14	0.26	0.30	0.11	0.21	0.30	0.26	0.26	0.30	0.13	0.16	0.26	0.29	0.28	0.27	0.00	0.15	0.31	0.17	0.22	0.30	0.05	0.17	0.31
RC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RGL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHOL_NIR	0.07	0.08	0.11	0.00	0.00	0.10	0.08	0.08	0.09	0.07	0.10	0.10	0.08	0.08	0.08	0.06	0.07	0.09	0.09	0.07	0.09	0.00	0.00	0.07
RHOL_VIS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHOS_NIR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RHOS_VIS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RSMAX	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SAI	0.08	0.14	0.16	0.07	0.15	0.16	0.13	0.13	0.18	0.07	0.09	0.13	0.15	0.14	0.15	0.00	0.12	0.15	0.09	0.11	0.17	0.00	0.16	0.18
TAUL_NIR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TAUL_VIS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TAUS_NIR	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TAUS_VIS	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TOPT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.12	0.00	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00	0.00
XL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Z0MVT	0.25	0.00	0.00	0.39	0.00	0.00	0.05	0.08	0.00	0.37	0.07	0.06	0.06	0.00	0.10	0.50	0.09	0.00	0.10	0.00	0.00	0.25	0.00	0.00

Figure 20. Parameter importance during daytime (06:00–17:00 UTC) for grassland land cover computed for the first model level as an average of all the algorithms used in the present study. Values with a red font indicate importance estimates associated with $R^2 < 0.6$.

along- and cross-valley circulations. The coupled configuration highlights that parameter uncertainty propagates beyond surface fluxes, also influencing the intensity of slope and valley circulations and the timing of their transitions. These effects cannot be represented in offline Noah-MP sensitivity experiments.

The analysis revealed that canopy height (*HVT*), LAI, optimal temperature (*TOPT*), stomatal resistance (*RS*), and near-infrared leaf reflectivity (*RHOL_NIR*) are critical parameters for atmospheric processes in the mountain valley with mixed forests, with *RS* being the most influential overall. Momentum roughness length (*Z0MVT*) is shown to be strongly influential for wind and TKE. For atmospheric processes in the mountain valley with grassland, *LAI* and *SAI* are the dominant parameters, with *RHOL_NIR* ranking third overall. *Z0MVT* has a much stronger influence during daytime compared to nighttime.

The results obtained in this work highlighted which LSM parameters have the greatest impact on model results, identifying those whose quantification should be treated with the most care for correctly representing surface-atmosphere exchange processes in mesoscale meteorological models. The comparative analysis between mixed forest and grassland ensembles revealed that results can be different depending on surface characteristics, suggesting that the analysis should be enlarged to other land cover types to obtain more general results. Moreover, it should be remembered that parameter importance is not absolute but rather context-dependent, potentially being influenced by several factors, and in particular by the physical processes that mostly affect surface-atmosphere exchange in different conditions. It is also important to highlight that these results are also constrained by how processes are parameterized within Noah-MP and thus findings might be different when using different LSMs.

Moreover, the sensitivity rankings reported here are expected to be most directly applicable to thermally driven valley circulations under clear-sky conditions and weak synoptic forcing, consistent with the idealized design

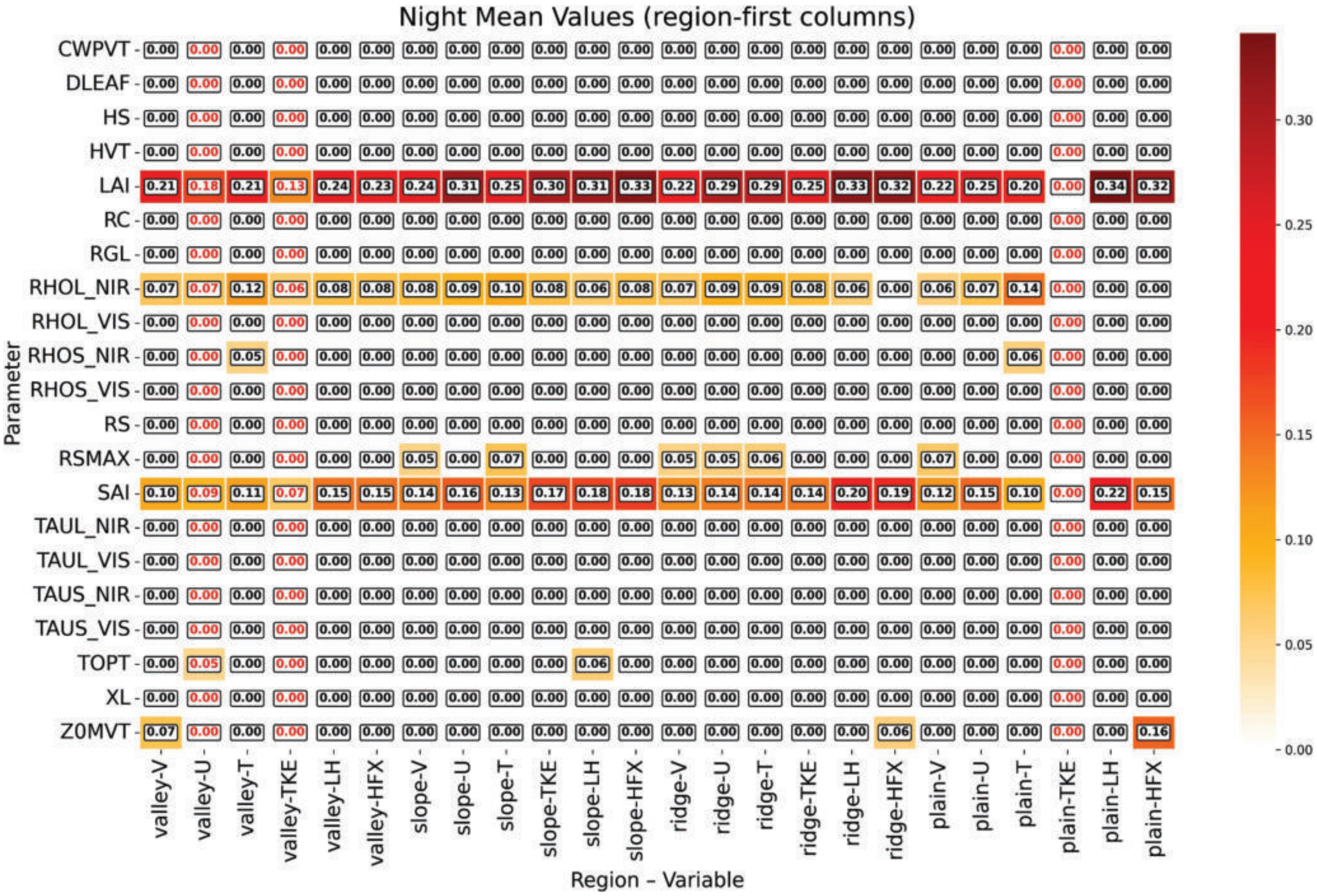


Figure 21. As in Figure 20, but during nighttime (18:00–05:00 UTC).

adopted in this study. While this controlled configuration allowed a clear identification of parameter effects, it also constrains the range of interacting processes represented. Under conditions characterized by cloud cover or stronger background winds, additional physical mechanisms and nonlinear feedbacks may become important, potentially altering the relative importance of individual parameters.

This complexity highlights the need for tailored investigations and reinforces the value of using GSA as a diagnostic tool for each specific case study. The proposed sensitivity analysis framework based on fast surrogate models provides an effective tool for this task, allowing for the application of GSA to the output of complex and computationally demanding meteorological models. Similarly, the same approach could also be used for the optimization of parameter values, potentially leading to significant benefits to the overall quality of meteorological simulations, considering in particular atmospheric boundary layer processes.

Beyond its physical findings, this study provides a methodological contribution aimed at making variance-based GSA computationally feasible for highly nonlinear models. Here the approach is demonstrated for an online-coupled NWP-LSM system in an idealized complex terrain configuration. By combining ML-based surrogate emulation with Sobol variance decomposition, the framework enables the large number of model evaluations required by Sobol analysis without incurring the prohibitive computational cost of executing the full high-fidelity model for each realization. The use of a multi-algorithm surrogate ensemble further increases the robustness of the inferred sensitivity rankings by reducing reliance on any single regression model and by identifying importance patterns that are consistent across methods and regimes. To the best of the authors' knowledge, this represents the first application in which Sobol-type variance decomposition is rendered computationally practical for a sensitivity analysis with a NWP model, avoiding approximately 4×10^4 high-fidelity simulations compared to a direct high-fidelity Sobol implementation.

The objective of this work was not to optimize parameters, but to quantify the relative influence of uncertain vegetation parameters on simulated atmospheric processes. As the present experiment is idealized and does not rely on an observational benchmark, it is not possible to evaluate the accuracy improvement associated with parameter tuning. Instead, the sensitivity values reported in the heatmaps quantify each parameter's contribution to the output variance within the tested uncertainty ranges. This provides guidance on which parameters to prioritize for future measurements and calibration efforts in real-world applications.

The ML-AMPSIT framework, originally introduced in Di Santo et al. (2025), and applied here for the first time in a complex terrain configuration, thus provides a transferable diagnostic workflow that can be readily extendable to other land-cover types, meteorological regimes, and progressively more realistic setups. By enabling computationally feasible sensitivity assessments with complex models, ML-AMPSIT offers a flexible tool for systematically investigating how parameter uncertainty affects simulation results across different configurations.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

The ML-AMPSIT tool, complete with source code and detailed usage instructions, is publicly available (Di Santo, 2024). This study employed version 4.5.1 of the Weather Research and Forecasting model (Skamarock et al., 2019), which incorporates version 4.4 of the Noah-MP land surface model. The ensemble dataset used in the ML-processing of this study and the namelist file used for simulations are publicly available in Zenodo (Di Santo, 2025).

Acknowledgments
This research was funded by Euregio Science Fund (3rd Call, IPN101) of the Europaregion Euregio. Open access publishing facilitated by Universita degli Studi di Trento, as part of the Wiley - CRUI-CARE agreement.

References

Alqahtani, A., Shah, M. I., Aldrees, A., & Javed, M. F. (2022). Comparative assessment of individual and ensemble machine learning models for efficient analysis of river water quality. *Sustainability*, *14*(3), 1183. https://doi.org/10.3390/su14031183

Arsenault, K. R., Nearing, G. S., Wang, S., Yatheendradas, S., & Peters-Lidard, C. D. (2018). Parameter sensitivity of the noah-mp land surface model with dynamic vegetation. *Journal of Hydrometeorology*, *19*(5), 815–830. https://doi.org/10.1175/jhm-d-17-0205.1

Ching, J., Rotunno, R., LeMone, M., Martilli, A., Kosovic, B., Jimenez, P. A., & Dudhia, J. (2014). Convectively induced secondary circulations in fine-grid mesoscale numerical weather prediction models. *Monthly Weather Review*, *142*(9), 3284–3302. https://doi.org/10.1175/MWR-D-13-00318.1

Cinquegrana, D., Zollo, A. L., Montesarchio, M., & Bucchignani, E. (2023). A metamodel-based optimization of physical parameters of high resolution nwp icon-lam over southern Italy. *Atmosphere*, *14*(5), 788. https://doi.org/10.3390/atmos14050788

Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., et al. (2021). Process-based climate model development harnessing machine learning: I. a calibration tool for parameterization improvement. *Journal of Advances in Modeling Earth Systems*, *13*(3), e2020MS002217. https://doi.org/10.1029/2020MS002217

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., et al. (2016). The impact of standard and hard-coded parameters on the hydrologic fluxes in the noah-mp land surface model. *Journal of Geophysical Research: Atmospheres*, *121*(18), 10676–10700. https://doi.org/10.1002/2016JD025097

Dagon, K., Sanderson, B., Rosie, F., & Lawrence, D. (2020). A machine learning approach to emulation and biophysical parameter estimation with the community land model, version 5. *Advances in Statistical Climatology, Meteorology and Oceanography*, *6*(2), 223–244. https://doi.org/10.5194/ascmo-6-223-2020

Daviran, M., Shamekhi, M., Ghezelbash, R., & Maghsoudi, A. (2023). Landslide susceptibility prediction using artificial neural networks, svms and random forest: Hyperparameters tuning by genetic optimization algorithm. *International Journal of Environmental Science and Technology*, *20*(1), 259–276. https://doi.org/10.1007/s13762-022-04491-3

de Wekker, S. F. J., Zhong, S., Fast, J. D., & Whiteman, C. D. (1998). A numerical study of the thermally driven plain-to-basin wind over idealized basin topographies. *Journal of Applied Meteorology*, *37*(6), 606–622. https://doi.org/10.1175/1520-0450(1998)037⟨0606:ANSOTT⟩2.0.CO;2

Di Santo, D. (2024). Ml-ampsit. *Zenodo*. https://doi.org/10.5281/zenodo.10789930

Di Santo, D. (2025). Datasetvalleycirculation v1.0.1 [Dataset]. *Zenodo*. https://doi.org/10.5281/zenodo.17804228

Di Santo, D., He, C., Chen, F., & Giovannini, L. (2025). Ml-ampsit: Machine learning-based automated multi-method parameter sensitivity and importance analysis tool. *Geoscientific Model Development*, *18*(2), 433–459. https://doi.org/10.5194/gmd-18-433-2025

Ek, M. B., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., et al. (2003). Implementation of noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model. *Journal of Geophysical Research*, *108*(D22). https://doi.org/10.1029/2002JD003296

Elia, L., Castellaro, S., Dahal, A., & Lombardo, L. (2023). Assessing multi-hazard susceptibility to cryospheric hazards: Lesson learnt from an Alaskan example. *Science of The Total Environment*, *898*, 165289. https://doi.org/10.1016/j.scitotenv.2023.165289

Fernández-Godino, M. G., Park, C., Kim, N., & Haftka, R. (2017). Review of multi-fidelity models. *Advances in Computational Science and Engineering*, *1*, 351–400. https://doi.org/10.3934/acse.2023015

Forrester, A., Sobester, A., & Keane, A. (2008). Engineering design via surrogate modelling: A practical guide. https://doi.org/10.1002/9780470770801

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., et al. (2014). A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. *Environmental Modelling Software*, *51*, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Gigović, L., Pourghasemi, H. R., Drobnjak, S., & Bai, S. (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. *Forests*, *10*(5), 408. https://doi.org/10.3390/f10050408

Gong, W., Duan, Q., Li, J., Wang, C., di, Z., ye, A., et al. (2015). An intercomparison of sampling methods for uncertainty quantification of environmental dynamic models. *Journal of Environmental Informatics*, *28*. https://doi.org/10.3808/jei.201500310

He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R., et al. (2023a). The community noah-mp land surface modeling system technical description version 5.0. *No. NCAR/TN-575+STR*. https://doi.org/10.5065/ew8g-yr95

He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R., et al. (2023c). Modernizing the open-source community noah with multi-parameterization options (Noah-mp) land surface model (version 5.0) with enhanced modularity, interoperability, and applicability. *Geoscientific Model Development*, *16*(17), 5131–5151. https://doi.org/10.5194/gmd-16-5131-2023

He, C., Valayamkunnath, P., Barlage, M., Chen, F., Gochis, D., Cabell, R. S., et al. (2023b). The community Noah-MP land surface modeling system technical description version 5.0. [Tech. Rep.]. https://doi.org/10.5065/ew8g-yr95

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *Journal of Geophysical Research*, *113*(D13). https://doi.org/10.1029/2008JD009944

Iooss, B., & Lemaître, P. (2014). A review on global sensitivity analysis methods. *Operations Research/ Computer Science and Information Systems*, *59*, 101–122. https://doi.org/10.1007/978-1-4899-7547-8_5

Jarvis, P. G., Monteith, J. L., & Weatherley, P. E. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, *273*(927), 593–610. https://doi.org/10.1098/rstb.1976.0035

Kim, S., & Boukouvala, F. (2020). Machine learning-based surrogate modeling for data-driven optimization: A comparison of subset selection for regression techniques. *Optimization Letters*, *14*(4), 989–1010. https://doi.org/10.1007/s11590-019-01428-7

Lei, T., Ng, S., & Siu, S. (2023). Application of ann, xgboost, and other ml methods to forecast air quality in Macau. *Sustainability*, *15*(6), 5341. https://doi.org/10.3390/su15065341

Li, J., Zhang, G., Barlage, M., Gan, Y., Xin, Y., & Wang, C. (2018). Impacts of land cover and soil texture uncertainty on land model simulations over the central Tibetan Plateau. *Journal of Advances in Modeling Earth Systems*, *10*. https://doi.org/10.1029/2018MS001377

Massey, J. D., Steenburgh, W. J., Hoch, S. W., & Knievel, J. C. (2014). Sensitivity of near-surface temperature forecasts to soil properties over a sparsely vegetated dryland region. *Journal of Applied Meteorology and Climatology*, *53*(8), 1976–1995. https://doi.org/10.1175/JAMC-D-13-0362.1

Monin, A., & Obukhov, A. (1954). Basic laws of turbulent mixing in the surface layer of the atmosphere.

Niu, G.-Y., & Yang, Z.-L. (2004). Effects of vegetation canopy processes on snow surface energy and mass balances. *Journal of Geophysical Research*, *109*(D23). https://doi.org/10.1029/2004JD004884

Niu, G.-Y., Yang, Z.-L., Dickinson, R. E., Gulden, L. E., & Su, H. (2007). Development of a simple groundwater model for use in climate models and evaluation with gravity recovery and climate experiment data. *Journal of Geophysical Research*, *112*(D7). https://doi.org/10.1029/2006JD007522

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community noah land surface model with multi-parameterization options (Noah-mp): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, *116*(D12), D12109. https://doi.org/10.1029/2010JD015139

Oke, T. (1987). *Boundary layer climates*. Routledge. Retrieved from https://books.google.it/books?id=K_2dW7crfVIC

Pyo, J., Kwon, Y., Ahn, J.-H., Baek, S., Kwon, Y.-H., & Cho, K. (2021). Sensitivity analysis and optimization of a radiative transfer numerical model for turbid lake water. *Remote Sensing*, *13*(4), 709. https://doi.org/10.3390/rs13040709

Rahmati, O., Pourghasemi, H. R., & Melesse, A. M. (2016). Application of gis-based data driven random forest and maximum entropy models for groundwater potential mapping: A case study at mehran region, Iran. *Catena*, *137*, 360–372. https://doi.org/10.1016/j.catena.2015.10.010

Rampanelli, G., Zardi, D., & Rotunno, R. (2004). Mechanisms of up-valley winds. *Journal of The Atmospheric Sciences - J ATMOS SCI*, *61*(24), 3097–3111. https://doi.org/10.1175/JAS-3354.1

Reif, M., Rotach, M. W., Gohm, A., & Wohlfahrt, G. (2024). Carbon dioxide exchange in an idealized valley. *Environmental Modelling Software*, *171*, 105887. https://doi.org/10.1016/j.envsoft.2023.105887

Rosero, E., Yang, Z.-L., Wagener, T., Gulden, L. E., Yatheendradas, S., & Niu, G.-Y. (2010). Quantifying parameter sensitivity, interaction, and transferability in hydrologically enhanced versions of the noah land surface model over transition zones during the warm season. *Journal of Geophysical Research*, *115*(D3). https://doi.org/10.1029/2009JD012035

Rosolem, R., Gupta, H., Shuttleworth, W., Zeng, X., & de Gonçalves, L. G. G. (2012). A fully multiple-criteria implementation of the sobol method for parameter sensitivity analysis. *Journal of Geophysical ResearchD: Atmosphere*, *117*(D7). https://doi.org/10.1029/2011JD016355

Sakaguchi, K., & Zeng, X. (2009). Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the community land model (clm3.5). *Journal of Geophysical Research*, *114*(D1). https://doi.org/10.1029/2008JD010834

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, *145*(2), 280–297. https://doi.org/10.1016/S0010-4655(02)00280-1

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., et al. (2008). *Global sensitivity analysis: The primer*. Wiley.

Saltelli, A., & Sobol', I. (1995). *Sensitivity analysis for nonlinear mathematical models: Numerical experience* (Vol. 7). Matematicheskoe Modelirovanie.

Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., & Chen, F. (1996). Simple water balance model for estimating runoff at different spatial and temporal scales. *Journal of Geophysical Research*, *101*(D3), 7461–7475. https://doi.org/10.1029/95JD02892

Schmidli, J., Billings, B., Chow, F., De Wekker, S., Doyle, J., Grubišić, V., et al. (2011). Intercomparison of mesoscale model simulations of the daytime valley wind system. *Monthly Weather Review - MON WEATHER REV*, *139*(5), 1389–1409. https://doi.org/10.1175/2010MWR3523.1

Schmidli, J., & Rotunno, R. (2010). Mechanisms of along-valley winds and heat exchange over mountainous terrain. *Journal of the Atmospheric Sciences*, *67*(9), 3033–3047. https://doi.org/10.1175/2010JAS3473.1

Serafin, S., & Zardi, D. (2010). Structure of the atmospheric boundary layer in the vicinity of a developing upslope flow system: A numerical model study. *Journal of The Atmospheric Sciences - J ATMOS SCI*, *67*(4), 1171–1185. https://doi.org/10.1175/2009JAS3231.1

Serafin, S., & Zardi, D. (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. *Journal of The Atmospheric Sciences - J ATMOS SCI*, *68*(9), 2128–2141. https://doi.org/10.1175/2011JAS3610.1

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). *A description of the advanced research wrf version 4* (Tech. Rep.). National Center for Atmospheric Research. https://doi.org/10.5065/1dfh-6p97

Sobol, I. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, *7*(4), 86–112. https://doi.org/10.1016/0041-5553(67)90144-9

Sridhara, S., Manoj, K. N., Gopakkali, P., Kashyap, G. R., Das, B., Singh, K. K., & Srivastava, A. K. (2023). Evaluation of machine learning approaches for prediction of pigeon pea yield based on weather parameters in India. *International Journal of Biometeorology*, *67*(1), 165–180. https://doi.org/10.1007/s00484-022-02396-x

Sukoriansky, S., Galperin, B., & Staroselsky, I. (2005). A quasinormal scale elimination model of turbulent flows with stable stratification. *Physics of Fluids*, *17*(8), 085107. https://doi.org/10.1063/1.2009010

Tomasi, E., Giovannini, L., Zardi, D., & De Franceschi, M. (2017). Optimization of noah and noah_mp wrf land surface schemes in snow-melting conditions over complex terrain. *Monthly Weather Review*, *145*(12), 4727–4745. https://doi.org/10.1175/MWR-D-16-0408.1

Trabelsi, F., & Bel Hadj Ali, S. (2022). Exploring machine learning models in predicting irrigation groundwater quality indices for effective decision making in medjerda river basin, Tunisia. *Sustainability*, *14*(4), 2341. https://doi.org/10.3390/su14042341

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. *Journal of Hydrology*, *324*(1–4), 10–23. https://doi.org/10.1016/j.jhydrol.2005.09.008

Vinuesa, R., & Brunton, S. (2022). Enhancing computational fluid dynamics with machine learning. *Nature Computational Science*, *2*(6), 358–366. https://doi.org/10.1038/s43588-022-00264-7

Wagner, J., Gohm, A., & Rotach, M. (2015a). The impact of valley geometry on daytime thermally driven flows and vertical transport processes. *Quarterly Journal of the Royal Meteorological Society*, *141*(690), 1780–1794. https://doi.org/10.1002/qj.2481

Wagner, J., Gohm, A., & Rotach, M. (2015b). Influence of along-valley terrain heterogeneity on exchange processes over idealized valleys. *Atmospheric Chemistry and Physics Discussions*, *15*, 415–451. https://doi.org/10.5194/acpd-15-415-2015

Wang, A., & Solomatine, D. (2019). Practical experience of sensitivity analysis: Comparing six methods, on three hydrological models, with three performance criteria. *Water*, *11*(5), 1062. https://doi.org/10.3390/w11051062

Watson-Parris, D., Williams, A., Deaconu, L., & Stier, P. (2021). Model calibration using esem v1.1.0 – An open, scalable earth system emulator. *Geoscientific Model Development*, *14*(12), 7659–7672. https://doi.org/10.5194/gmd-14-7659-2021

Weissmann, M., Braun, F. J., Gantner, L., Mayr, G. J., Rahm, S., & Reitebuch, O. (2005). The alpine mountain–plain circulation: Airborne doppler LiDAR measurements and numerical simulations. *Monthly Weather Review*, *133*(11), 3095–3109. https://doi.org/10.1175/MWR3012.1

Wicker, L., & Skamarock, W. (2002). Time-splitting methods for elastic models using forward time schemes. *Monthly Weather Review - MON WEATHER REV*, *130*. https://doi.org/10.1175/1520-0493(2002)130⟨2088:TSMFEM⟩2.0.CO;2

Wu, Z., Cui, N., Gong, D., Zhu, F., Xing, L., Zhu, B., et al. (2023). Simulation of daily maize evapotranspiration at different growth stages using four machine learning models in semi-humid regions of northwest China. *Journal of Hydrology*, *617*, 128947. https://doi.org/10.1016/j.jhydrol.2022.128947

Yang, Z.-L., Niu, G.-Y., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community noah land surface model with multiparameterization options (Noah-mp): 2. evaluation over global river basins. *Journal of Geophysical Research*, *116*(D12), D12110. https://doi.org/10.1029/2010JD015140

Yu, Z., Dai, H., Yang, J., Zhu, Y., & Yuan, S. (2024). Global sensitivity analysis with deep learning-based surrogate models for unraveling key parameters and processes governing redox zonation in riparian zone. *Journal of Hydrology*, *638*, 131442. https://doi.org/10.1016/j.jhydrol.2024.131442

Zardi, D., & Whiteman, C. D. (2013). Diurnal mountain wind systems. In F. K. Chow, S. F. De Wekker, & B. J. Snyder (Eds.), *Mountain weather research and forecasting: Recent progress and current challenges* (pp. 35–119). Springer Netherlands. https://doi.org/10.1007/978-94-007-4098-3_2