



Trade-offs between predictive accuracy and distributional fidelity in surrogate modeling of extreme precipitation

Nabil El Moçayd¹ and Abdelaziz Chaqdid²

¹International Water Research Institute (IWRI), College of Agriculture and Environmental Sciences, University Mohammed VI Polytechnic (UM6P), Ben Guerir, Morocco.

¹Physical Science and Engineering Division (PSE), King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.

Correspondence: Nabil El Moçayd (nabil.elmocayd@um6p.ma)

Abstract. Extreme precipitation remains a major challenge for regional climate modeling due to strong nonlinearities, spatial heterogeneity, and sensitivity to poorly constrained parameterizations. In this study, we assess the ability of surrogate models to represent extreme precipitation under parametric uncertainty in a regional climate model. An ensemble of simulations is generated using Latin Hypercube sampling of key parameters in the SUBEX large-scale cloud and precipitation scheme, and several surrogate approaches are evaluated, including polynomial chaos expansions (PCE), Gaussian process regression (Kriging), hybrid Polynomial Chaos-Kriging (PCK), and stochastic spectral embedding (SSE). Model performance is assessed using both predictive metrics (RMSE, R^2) and distribution-based metrics (Kolmogorov–Smirnov distance, Wasserstein distance, Kullback–Leibler divergence, and Energy distance), with validation performed on an independent dataset. Results reveal a clear trade-off: PCK provides the most accurate pointwise predictions, whereas SSE more faithfully reproduces the statistical distribution of extreme precipitation, particularly in the upper tail. Notably, the observed event occupies a low-probability region of the modeled distribution, although its position varies spatially. This suggests that standard model configurations may struggle to reproduce the intensity of individual extreme events, as they are typically calibrated to capture mean climatological variability rather than rare extremes. Sensitivity analysis indicates that extreme precipitation variability is primarily controlled by parameters governing precipitation efficiency over land and ocean, while spatial analysis shows that uncertainty is concentrated in coastal and mountainous regions where nonlinear interactions are strongest. These findings demonstrate that no single surrogate model is optimal for all objectives and emphasize the need to distinguish between predictive accuracy and distributional fidelity when modeling extremes. They provide new insight into how model calibration, parameter uncertainty, and surrogate structure jointly influence the representation of extreme precipitation in climate simulations.

1 Introduction

Extreme precipitation events represent one of the most challenging aspects of regional climate modeling due to their strong spatial variability, nonlinear dynamics, and high societal impact. From a modeling perspective, these events are characterized by complex interactions between convection, moisture transport, and orographic forcing, which are difficult to represent accurately in numerical models (O’Gorman, 2015; Fowler et al., 2021). As a result, regional climate models (RCMs) exhibit



substantial uncertainties in the simulation of precipitation extremes, often manifested as systematic biases in the upper tail of the distribution (Kendon et al., 2014; Kim et al., 2020).

The uncertainty associated with extreme precipitation simulations originates from multiple sources, including model structure, boundary conditions, internal variability, and the representation of unresolved physical processes (Giorgi, 2019). Previous studies have linked these uncertainties to deficiencies in the representation of present-day precipitation climatology (Zhang et al., 2022), as well as to limitations in the simulation of the dynamical processes governing extreme events (Sothearith et al., 2026). Among these sources, parameterization uncertainty is often identified as one of the dominant contributors to uncertainty in Earth system models (Berner et al., 2017; Morales et al., 2019). In particular, convection and cloud microphysics rely on simplified representations of highly nonlinear processes and involve parameters that are often poorly constrained (Hourdin et al., 2017; Norris et al., 2021; O’Gorman and Dwyer, 2018). Consequently, relatively small perturbations in parameter values can lead to substantial variability in simulated precipitation intensity. In some cases, stochastic representations of unresolved processes have even been shown to outperform deterministic parameterizations, further emphasizing the importance of parameter uncertainty in climate simulations (Berner et al., 2011).

Uncertainty quantification (UQ) provides a systematic framework for assessing how uncertainty in model parameters propagates to simulation outputs (Forest et al., 2002). Recognizing the inherently probabilistic nature of Earth system simulations, several recent studies have advocated the use of probability-informed climate modeling frameworks (Palmer, 2001; Mansfield et al., 2023; Schneider et al., 2023). However, applying UQ to RCMs remains computationally challenging because it requires a large number of simulations to adequately explore the parameter space (Jackson et al., 2004; Bracco et al., 2025). This computational burden has motivated the development of surrogate models, also referred to as Earth system emulators, which approximate the behavior of the original model at a fraction of the computational cost (Bocquet, 2023; Schöngart et al., 2024).

Among the various surrogate modeling approaches proposed in the uncertainty quantification literature, polynomial chaos expansions (PCE), Gaussian process regression (Kriging), and hybrid formulations combining both methods have emerged as particularly attractive candidates for computationally expensive environmental models (Hou and Behdinan, 2022; Al-Ghosoun et al., 2021; Roy et al., 2018). These approaches provide a probabilistic representation of model responses while requiring relatively modest training datasets compared to many data-intensive machine-learning methods. Nevertheless, their application to extreme precipitation remains largely unexplored, particularly when the objective extends beyond pointwise prediction to the accurate representation of variability and tail behavior. Most existing studies evaluate surrogate models using predictive metrics such as root-mean-square error or coefficient of determination (Jiang et al., 2026; Schöngart et al., 2024; Bouabid et al., 2024). While these metrics are appropriate for many applications, they may be insufficient for extreme precipitation, where the statistical distribution of the response, and especially its upper tail, is often of primary interest (Pierini et al., 2026). Consequently, it remains unclear whether surrogate models that perform well in a predictive sense are also capable of faithfully reproducing the probabilistic structure of extreme events. This limitation is particularly relevant in Bayesian calibration frameworks, where accurate estimation of response distributions is essential for constraining parameter uncertainty and deriving reliable posterior distributions (Reichelt et al., 2026).



This study addresses this question by systematically evaluating the performance of several surrogate modeling approaches for extreme precipitation in a regional climate model. Using an ensemble generated through Latin Hypercube Sampling of key SUBEX scheme parameters, we compare PCE (Sudret, 2014), Kriging (Kawai and Shimoyama, 2014), hybrid PC-Kriging (PCK) (Schobi et al., 2015), and stochastic spectral embedding (SSE) (Marelli et al., 2021). Model performance is assessed using both predictive and distribution-based metrics, allowing for a comprehensive evaluation of their capabilities. Our results reveal a fundamental trade-off between predictive accuracy and distributional fidelity: while PCK provides the most accurate pointwise predictions, SSE more effectively reproduces the statistical distribution of extreme precipitation. Sensitivity analysis further identifies SUBEX scheme parameters controlling precipitation efficiency, linking surrogate model behavior to underlying physical processes. These findings provide new insight into the role of surrogate model structure in climate uncertainty quantification and highlight the importance of selecting modeling approaches consistent with the processes and objectives of interest. The remainder of this paper is organized as follows.

Section 2 describes the study area, observational dataset, regional climate model configuration, ensemble generation strategy, and surrogate modeling framework. Section 3 presents the results, including the characterization of the extreme precipitation event, the predictive and distributional performance of the surrogate models, and the associated sensitivity analysis. Section 4 discusses the observed trade-off between predictive accuracy and distributional fidelity, its implications for surrogate modeling and climate simulations, and the main limitations of the study. Finally, Section ?? summarizes the main findings and outlines perspectives for future research.

2 Data and methodology

2.1 Study area: Morocco The Atlas Mountains

The study area extends across northwestern Africa (approximately 18° W–11.5°E and 20.5°–38° N), covering Morocco, parts of Algeria, and Tunisia (see Fig. 1). This domain captures the transition from the Atlantic and Mediterranean coasts, across the Atlas Mountains, into the Sahara. The Atlas Mountains are the dominant physical feature of the domain, consisting of the Rif, Middle, High (peaking at Jebel Toubkal, 4167 m), and Anti-Atlas ranges in Morocco, continuing eastward as the Tell and Saharan Atlas. These mountains play a major role in shaping the regional climate through processes explicitly represented in our configuration: orographic uplift on windward slopes, rain-shadow and foehn effects on leeward sides, barrier jets and gap-flow channeling, coastal land–sea-breeze interactions, episodic lee cyclogenesis over the High Plateaus, and winter snow accumulation followed by spring melt in the High Atlas. Collectively, these processes generate strong spatial and seasonal gradients in temperature and precipitation that coarse global climate models cannot resolve. This study focuses on an extreme precipitation event that affected the coastal region of Morocco between November 25th and December 1st, 2014. The event is characterized by intense rainfall concentrated along coastal and mountainous regions, with significant impacts including flooding. Observations are obtained from the Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset, which integrates satellite, gauge, and reanalysis products to provide high-resolution precipitation estimates. Consistent with the evaluation framework



of Tuel and El Moçayd (2023), MSWEP is used as the reference dataset for assessing the performance of both the regional climate model and the surrogate approximations.

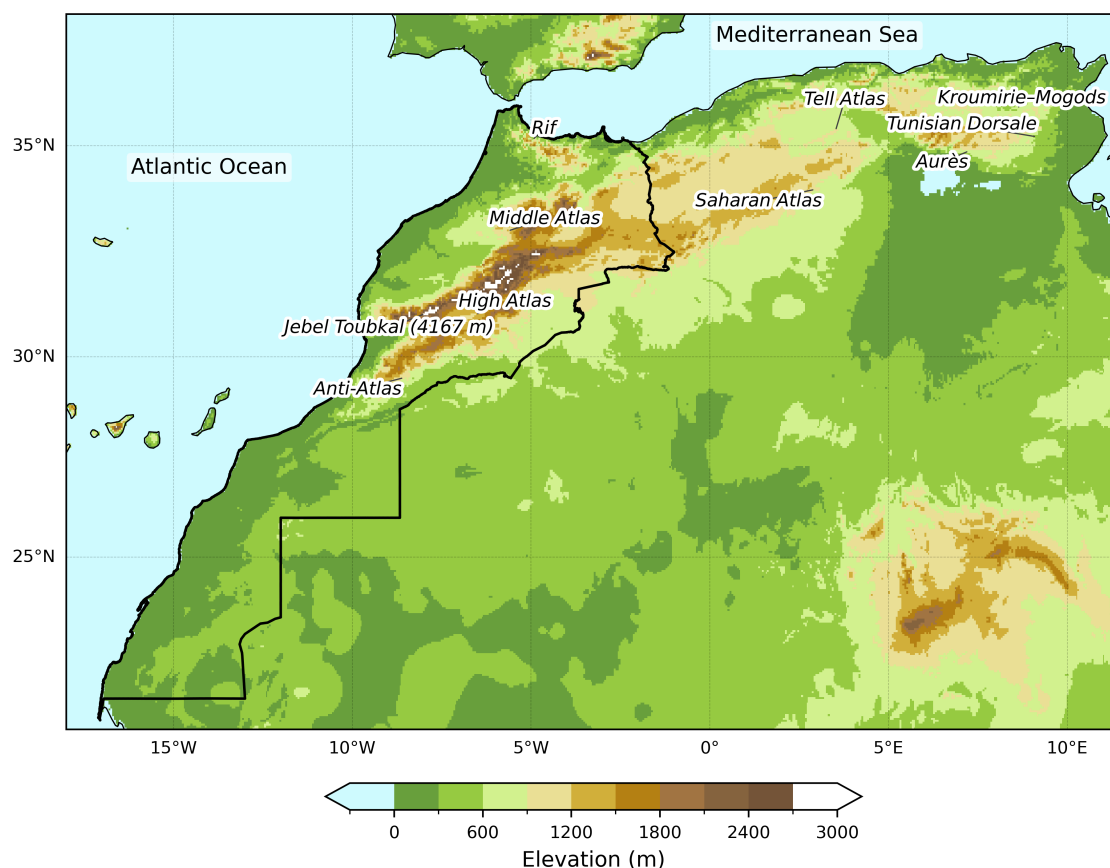


Figure 1. Topography of Northwest Africa, highlighting Morocco and the Atlas ranges that shape regional climate and elevation gradients.

2.2 Regional climate model and ensemble generation

Simulations are performed using the MRCM regional climate model, a hydrostatic, three-dimensional model derived from the ICTP RegCM3 dynamical core (Pal et al., 2007) and enhanced with targeted physics developments to improve land–atmosphere coupling, convective processes, and boundary layer representation. A key component is the coupling to the Integrated Biosphere Simulator (IBIS) land surface scheme (Foley et al., 1996; Winter et al., 2009), which replaces the legacy BATS formulation and provides improved surface energy and water budgets over vegetated and semi-arid regions (Im et al., 2014).

Additional developments include revised surface albedo and irrigation schemes (Marcella and Eltahir, 2012, 2014), updates to convective cloud and rainfall autoconversion parameterizations (Gianotti, 2012; Gianotti and Eltahir, 2014a, b), and modifications to boundary layer height and low-cloud representation (Gianotti, 2012). These improvements have been shown to enhance model performance across a range of climatic regions, including North America, West Africa, Southwest Asia, South



Asia, and the Maritime Continent (Winter and Eltahir, 2012; Im and Eltahir, 2018; Pal and Eltahir, 2016; Choi et al., 2021; Gianotti and Eltahir, 2014a). Following Tuel and Eltahir (2020) and Pal and Eltahir (2016), the baseline configuration includes IBIS for land surface processes, the SUBEX scheme for large-scale condensation (Pal et al., 2000), the CCM3 radiation pack-
105 age (Kiehl et al., 1998), a nonlocal planetary boundary layer scheme (Holtslag and Boville, 1993), and the bulk air–sea flux formulation of Zeng et al. (1998). Deep convection is represented using the Grell scheme with Fritsch–Chappell closure. Simulations are conducted at 10 km horizontal resolution, which captures first-order orographic gradients of the Atlas Mountains and coastal land–sea contrasts that modulate extreme precipitation.

To investigate the impact of parameter uncertainty, an ensemble of simulations is generated using a Latin Hypercube Sam-
110 pling (LHS) strategy. The ensemble is constructed by perturbing a set of key parameters of the Sub-grid Explicit Moisture Scheme (SUBEX; Pal et al., 2000), which governs the formation of resolvable-scale clouds and large-scale precipitation in MRCM. The perturbed parameters are the autoconversion rates over land and ocean (*qcklland*, *qckloce*), the fractions of the Gultepe relation at which precipitation is triggered over land and ocean (*gulland*, *guloce*), the raindrop evaporation rate coefficient (*cevap*), and the raindrop accretion rate (*caccer*). These parameters control the conversion of cloud water to rain, the
115 re-evaporation of falling precipitation, and the accretion of cloud droplets, and thus directly regulate precipitation intensity. They constitute the primary sources of variability explored in this study. Throughout the paper, variations in these parameters define the ensemble used to analyze the behavior of extreme precipitation and evaluate the performance of surrogate models. Following Schneider et al. (2023), we generated an ensemble of 1,000 simulations at 10 km horizontal resolution, which was used to train the surrogate models. Their performance was then evaluated against an independent validation ensemble of
120 10,000 simulations generated using a second LHS sample.

The analysis focuses on a scalar quantity of interest defined as the maximum cumulative precipitation over the domain during the 7-day period,

$$Y_{MRCM} = \max_{s \in \Omega} \left(\sum_{t=1}^7 P(s, t) \right), \quad (1)$$

where $P(s, t)$ denotes precipitation at location s and time t . This metric is chosen to specifically target the intensity of extreme
125 events, which represents the primary source of discrepancy between simulations and observations.

2.3 Surrogate modeling approaches

To efficiently propagate parameter uncertainty and perform sensitivity analysis, surrogate models are employed to approximate the response of the regional climate model. These surrogates provide computationally efficient representations of the original model while preserving its dominant response characteristics. Four surrogate modeling approaches are considered in this
130 study: Polynomial Chaos Expansion (PCE), Gaussian Process Regression (Kriging), hybrid Polynomial Chaos Kriging (PCK), and Stochastic Spectral Embedding (SSE). Although Kriging-based methods remain the most widely used surrogate models in Earth system applications, all four approaches are well established in uncertainty quantification and sensitivity analysis, particularly in geoscience and environmental modeling. These methods represent distinct approximation philosophies, ranging from global spectral representations (PCE) and local correlation-based models (Kriging) to hybrid (PCK) and adaptive



135 partition-based approaches (SSE). The surrogate models are trained using a common experimental design composed of 1000 MRCM simulations generated through Latin Hypercube Sampling (LHS). The uncertain input vector is defined as $\mathbf{X} = [\text{cevap}, \text{gulland}, \text{guloce}, \text{qcklland}, \text{qckloce}, \text{caccr}]$, where each parameter is assumed to follow an independent uniform probability distribution. These parameters were selected because of their direct influence on large-scale (resolved-scale) microphysical precipitation processes represented by the SUBEX scheme, and therefore represent the primary sources of parameter uncertainty considered in this study. The corresponding uncertainty ranges are summarized in Table 1. By comparing these complemen-

Table 1. Uncertain parameters used in the experimental design. All variables are assumed to follow independent uniform distributions.

Parameter	Distribution	Range
<i>cevap</i>	Uniform	$[5 \times 10^{-8}, 5 \times 10^{-4}]$
<i>gulland</i>	Uniform	$[0.1, 0.7]$
<i>guloce</i>	Uniform	$[0.1, 0.7]$
<i>qcklland</i>	Uniform	$[1 \times 10^{-4}, 7 \times 10^{-4}]$
<i>qckloce</i>	Uniform	$[1 \times 10^{-4}, 7 \times 10^{-4}]$
<i>caccr</i>	Uniform	$[2, 6]$

140

tary surrogate structures within a common experimental framework, it becomes possible to assess how the choice of surrogate model influences the representation of extreme precipitation, including both predictive accuracy and distributional behavior. This comparison forms the basis of the subsequent analysis of uncertainty propagation, sensitivity estimation, and surrogate model performance.

145 2.3.1 Polynomial chaos expansion (PCE)

Polynomial Chaos Expansion (PCE) is a surrogate modeling approach widely used in uncertainty quantification and sensitivity analysis to efficiently approximate the response of computationally expensive numerical models. Considering the uncertain input vector $\mathbf{X} = (\text{cevap}, \text{gulland}, \text{guloce}, \text{qcklland}, \text{qckloce}, \text{caccr})$ the model response can be represented as a series of orthonormal polynomial basis functions,

$$150 \quad Y_{PCE}(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^d} y_{\alpha} \Psi_{\alpha}(\mathbf{X}), \quad (2)$$

where y_{α} are the spectral coefficients and Ψ_{α} are multivariate orthonormal polynomials. The polynomial basis is selected according to the probability distributions of the input variables, while multivariate basis functions are constructed through tensor products of the corresponding univariate polynomials. In practice, the expansion is truncated to a finite set of basis functions defined by a maximum polynomial degree. The unknown coefficients are then estimated from the experimental design using a regression-based approach. To improve computational efficiency and reduce the number of retained terms, sparse regression is employed using the Least Angle Regression (LAR) algorithm (Blatman and Sudret, 2011). This approach automatically identifies the most influential polynomial terms, providing an efficient representation of the model response while avoiding overfitting.



2.3.2 Gaussian process regression (Kriging)

160 Kriging models the response as a realization of a Gaussian process,

$$Y_{Krig}(\mathbf{X}) = \mu(\mathbf{X}) + Z(\mathbf{X}), \quad (3)$$

where $\mu(\mathbf{x})$ is a deterministic trend and $Z(\mathbf{x})$ is a zero-mean Gaussian process characterized by a covariance function. The covariance structure defines the correlation between model evaluations and enables interpolation of the response throughout the parameter space. Unlike PCE, which relies on a global polynomial approximation, Kriging constructs a local representation
165 of the response by exploiting the correlation between neighboring samples. This property makes it particularly suitable for capturing nonlinear behaviors and localized variations in the model response. In this study, the covariance structure is described using a Matérn (3/2) kernel, while the trend component is represented by a linear function. The hyperparameters of the Gaussian process are estimated from the experimental design through likelihood maximization. Once trained, the Kriging surrogate provides a continuous approximation of the regional climate model response and can be efficiently evaluated for uncertainty
170 quantification and sensitivity analysis.

2.3.3 Polynomial Chaos-Kriging (PCK)

Polynomial Chaos-Kriging (PCK) combines the global approximation capabilities of PCE with the local interpolation properties of Kriging. The model response is represented as

$$Y_{PCK}(\mathbf{X}) = \sum_{\alpha \in \mathbb{N}^d} y_{\alpha} \Psi_{\alpha}(\mathbf{X}) + Z(\mathbf{X}), \quad (4)$$

175 where the polynomial expansion captures the large-scale structure of the response and the Gaussian process correction accounts for local deviations not represented by the polynomial basis. As a result, PCK can be interpreted as a Kriging model whose trend component is described by a sparse polynomial chaos expansion rather than a simple low-order polynomial. This hybrid formulation was originally proposed to improve surrogate accuracy for strongly nonlinear responses by combining the complementary strengths of PCE and Kriging (Schobi et al., 2015). The polynomial component provides an efficient representation of
180 the dominant response trends, while the Gaussian process correction captures localized variability and residual nonlinearities. In this study, the polynomial basis is constructed using the same sparse LAR-PCE strategy described previously, while the residual component is modeled using a Gaussian process with a Matérn (3/2) covariance function. The resulting surrogate is particularly well suited to responses that exhibit both smooth global trends and localized nonlinear behavior.

2.3.4 Stochastic spectral embedding (SSE)

185 Stochastic Spectral Embedding (SSE) is a surrogate modeling approach designed to represent complex responses through a collection of local approximations rather than a single global surrogate. The method recursively partitions the input space into subdomains and constructs a local spectral representation within each region. The model response can therefore be expressed as



$$Y_{SSE}(\mathbf{X}) = \sum_{k=1}^{N_p} \chi_k(\mathbf{X}) Y_k(\mathbf{X}) \quad (5)$$

190 where (N_p) is the number of partitions, $(\chi_k(\mathbf{X}))$ is an indicator function defining the (k) –th subdomain, and $(Y_k(\mathbf{X}))$ denotes the local surrogate associated with that partition. In contrast to global approaches such as PCE and Kriging, SSE allows different regions of the parameter space to be represented independently. This adaptive decomposition is particularly useful when the response exhibits localized nonlinearities or distinct behavioral regimes that are difficult to capture using a single approximation. Within each partition, the local response is represented using a polynomial chaos expansion, allowing the
 195 method to retain the efficiency of spectral approximations while improving flexibility. The recursive partitioning strategy is driven by the local complexity of the response, such that additional resolution is introduced only where needed. As a result, SSE provides an efficient framework for approximating strongly nonlinear responses without requiring the construction of a globally complex surrogate. In this study, SSE is employed as an alternative local representation of the regional climate model response, complementing the global and hybrid approaches described previously. The SSE metamodel was constructed
 200 following Marelli et al. (2021).

2.4 Model evaluation and sensitivity analysis

The performance of the surrogate models is evaluated using both predictive and distribution-based metrics, reflecting the need to assess not only pointwise prediction accuracy but also the ability to reproduce the statistical structure of extreme precipitation. To ensure an unbiased evaluation, an independent validation dataset consisting of 10,000 MRCM simulations is
 205 generated using a second LHS design constructed from the same input distributions summarized in Table 1. This validation dataset is completely independent of the 1,000 simulations used to construct the surrogate models, thereby preventing any overlap between the training and testing phases. Once the surrogate models have been trained, they are evaluated against this independent dataset using complementary performance metrics. Predictive metrics quantify the accuracy of individual predictions, whereas distribution-based metrics assess the ability of each surrogate to reproduce the probability distribution
 210 of extreme precipitation. This dual evaluation framework is central to the present study, as it enables the investigation of the trade-off between predictive accuracy and distributional fidelity.

2.4.1 Predictive performance metrics

Using the independent validation dataset, the predictive performance of the surrogate models is assessed using two complementary metrics widely adopted in surrogate modeling and uncertainty quantification: the root-mean-square error (RMSE) and the coefficient of determination (R^2). Both metrics quantify the agreement between the surrogate predictions and the
 215 corresponding MRCM simulations on a point-by-point basis. The RMSE is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}, \quad (6)$$



where (Y_i) denotes the MRCM response and $\hat{Y}_i$ the corresponding prediction of the surrogate model evaluated at the same input realization. The RMSE measures the average magnitude of the prediction error, with lower values indicating better predictive performance. The coefficient of determination is given by

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2}, \quad (7)$$

where $\bar{Y}$ is the mean of the reference MRCM responses. The R^2 coefficient measures the proportion of the variance explained by the surrogate model, with values approaching unity indicating an increasingly accurate approximation. Although these metrics provide a comprehensive assessment of predictive accuracy, they evaluate the agreement between individual realizations rather than the statistical properties of the response. Consequently, they primarily characterize the ability of surrogate models to reproduce the conditional mean behavior of the regional climate model and do not necessarily reflect their capability to represent the underlying probability distribution of extreme precipitation.

2.4.2 Distributional performance metrics

While predictive metrics quantify the agreement between individual realizations, they do not assess whether a surrogate model accurately reproduces the probabilistic behavior of the regional climate model. This distinction is particularly important for extreme precipitation, where uncertainty is largely governed by the shape of the distribution and its upper tail rather than by pointwise prediction accuracy alone. To evaluate this aspect, the probability density function of the model response is first estimated using non-parametric kernel density estimation (KDE), thereby avoiding assumptions regarding the underlying distribution of extreme precipitation. The agreement between the surrogate and reference distributions is then quantified using four complementary statistical metrics.

1. The Kolmogorov-Smirnov (KS) distance (Massey Jr, 1951) is defined as

$$D_{KS} = \sup_x |F(x) - \hat{F}(x)|, \quad (8)$$

where $F(x)$ and $\hat{F}(x)$ are the empirical cumulative distribution functions of the reference and surrogate responses, respectively. This metric measures the maximum discrepancy between the two distributions.

2. The Wasserstein distance (Villani, 2009) is given by,

$$D_W = \int |F(x) - \hat{F}(x)| dx, \quad (9)$$

and measures the average transport distance between the two distributions. Unlike the KS distance, it is particularly sensitive to differences in the location and shape of the tails.

3. The Kullback-Leibler (KL) divergence (Hall, 1987) is defined as

$$D_{KL} = \int f(x) \log \left(\frac{f(x)}{\hat{f}(x)} \right) dx, \quad (10)$$



where $f(x)$ and $\hat{f}(x)$ are the probability density functions. This metric measures the information loss when approximating the reference distribution with the surrogate.

4. The Energy distance (Rizzo and Székely, 2016) is computed as

$$D_E = 2\mathbb{E}|X - Y| - \mathbb{E}|X - X'| - \mathbb{E}|Y - Y'|, \quad (11)$$

250 where X and Y denote samples from the reference and surrogate distributions, respectively, while X' and Y' are independent copies of X and Y . The Energy distance measures the discrepancy between two probability distributions by comparing both between-distribution and within-distribution distances.

These metrics provide complementary information on the agreement between distributions, with particular emphasis on variability and tail behavior, which are critical for the analysis of extreme precipitation.

255 2.4.3 Sensitivity analysis

To identify the dominant sources of uncertainty in extreme precipitation, a global variance-based sensitivity analysis is performed using Sobol indices. These indices quantify the contribution of each uncertain input parameter to the variance of the quantity of interest while accounting for the nonlinear nature of the response, see for example Elkari et al. (2023). The first-order Sobol index associated with parameter X_i is defined as

$$260 \quad S_i = \frac{\text{Var}[\mathbb{E}(Y | X_i)]}{\text{Var}(Y)}, \quad (12)$$

where Y denotes the quantity of interest. The first-order index measures the fraction of the output variance that can be explained solely by parameter X_i , while the difference between first-order and total-order indices provides insight into interaction effects between parameters. The analysis is conducted using surrogate models, enabling efficient estimation of sensitivity measures across the parameter space. Both the construction of the surrogate models and the computation of the Sobol sensitivity indices are performed using the UQLab framework (Marelli and Sudret, 2014). UQLab provides a unified environment for uncertainty quantification, surrogate modeling, and global sensitivity analysis, ensuring a consistent implementation of all surrogate models considered in this study. Once the surrogate models are constructed, Sobol indices are estimated directly from the surrogate responses, allowing an efficient and accurate approximation of the global sensitivity measures without requiring additional evaluations of the computationally expensive regional climate model.

270 3 Results

3.1 Characterization of the extreme precipitation event and model response

This study focuses on the simulation of an extreme precipitation event that affected the coastal region of Morocco between November 25th and December 1st, 2014. Observations are provided by the MSWEP dataset, which serves as a reference for evaluating model performance. As a baseline, a simulation of the event is performed with the MRCM using standard parameter



275 values obtained by calibrating the model against the observed seasonal cycle of mean precipitation; this run is hereafter referred to as the deterministic simulation. Since this calibration targets mean conditions rather than extremes, these default values are not necessarily optimal for reproducing the intensity of individual extreme events, which motivates the parameter-uncertainty analysis that follows.

Figure 2 compares the observed and deterministic simulations of cumulative precipitation for the studied event. Overall, 280 the model successfully reproduces the large-scale spatial organization of the precipitation field, including the coastal precipitation band and the first-order influence of the Atlas Mountains. However, important discrepancies remain in the simulated precipitation intensity and its spatial distribution. In particular, the simulated domain maximum reaches 420.3 mm, nearly twice the observed value of 221.5 mm, whereas the domain-averaged precipitation is much closer between the simulation and observations (19.4 and 19.1 mm, respectively). The main discrepancy therefore concerns the spatial distribution and local peak 285 intensity rather than the domain-mean event accumulation. The largest overestimation occurs along the Atlantic coastal region, where the model produces a highly intense precipitation core that is not present in the MSWEP observations. Conversely, precipitation is substantially underestimated over the Middle Atlas, where observations indicate a localized region of intense rainfall that is only weakly represented in the simulation. These contrasting biases suggest that, although the model captures the dominant synoptic-scale organization of the event, it struggles to correctly represent the regional processes controlling 290 precipitation intensity. Such behavior is consistent with the calibration strategy adopted for the deterministic simulation. The default parameter set was originally calibrated to reproduce the seasonal cycle and long-term mean precipitation, rather than the magnitude of individual extreme events. Consequently, while the model provides a realistic representation of the large-scale structure of the event, it is less successful in reproducing the localized intensity of extreme precipitation, thereby motivating the uncertainty quantification framework developed in the remainder of this study.

295 Such discrepancies are consistent with known limitations of regional climate models in representing extreme precipitation, which is governed by strongly nonlinear interactions between convection, moisture transport, and orographic forcing. In particular, the sensitivity of the SUBEX microphysics parameterization can lead to substantial variability in simulated precipitation intensity, even when the large-scale spatial patterns remain broadly consistent.

Given that the primary discrepancies between the deterministic simulation and the observations are associated with precipi- 300 tation intensity rather than the large-scale spatial organization of the event, the analysis focuses on a scalar quantity of interest defined as the maximum cumulative precipitation over the model domain during the 7-day event. This metric directly characterizes the most extreme realization simulated by the regional climate model and provides a consistent basis for uncertainty quantification, surrogate model construction, and model intercomparison. To further investigate the spatial variability of the model response, the same quantity of interest is also evaluated locally over two representative regions exhibiting contrasting 305 model behavior. The first region is centered on Sidi Ifni along the Atlantic coast (Figure 1), where the deterministic simulation substantially overestimates precipitation intensity. The second region is located around Midelt in the Middle Atlas (Figure 1), where the model underestimates the observed precipitation. These two locations provide complementary case studies for assessing whether the conclusions drawn at the domain scale remain valid in regions characterized by opposite model biases.

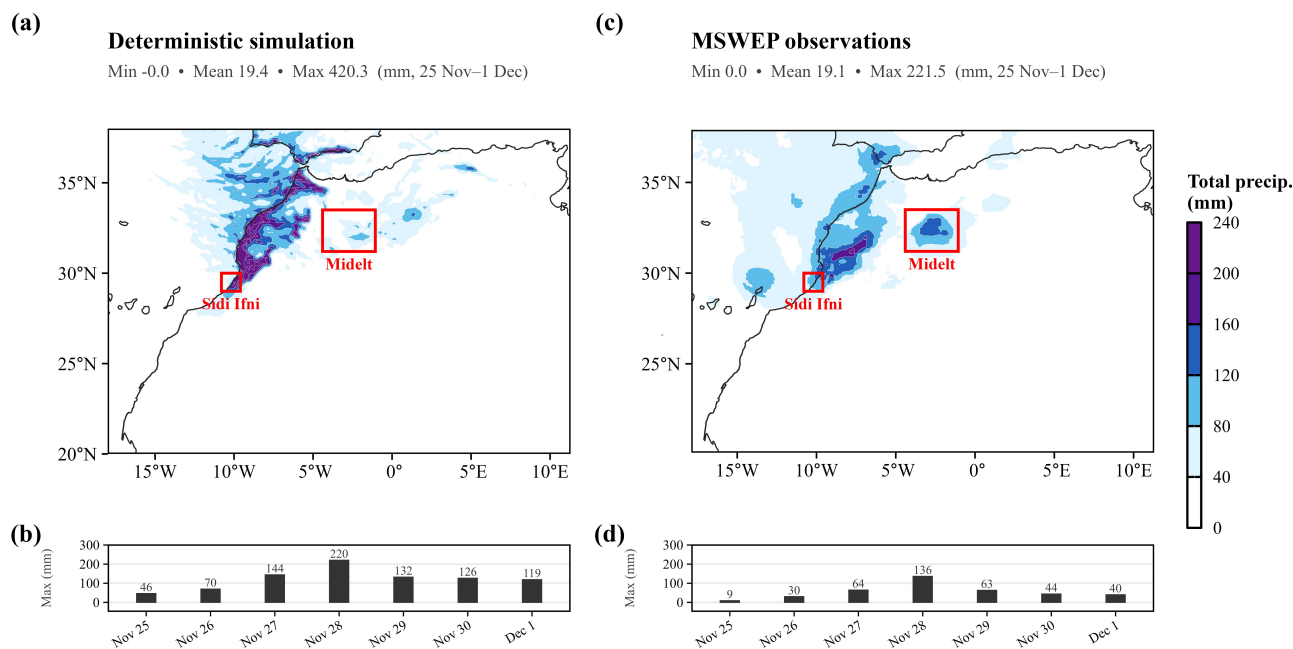


Figure 2. Cumulative precipitation (mm) for 25 November–1 December 2014 from (a) the MRCM deterministic simulation and (c) MSWEP observations, with daily spatial maxima shown in (b, d).

The complexity of the response is further illustrated in Figure 3. The relationship between the maximum cumulative precipitation and the dominant SUBEX parameters exhibits a strongly nonlinear behavior, with substantial dispersion observed for similar parameter values. This indicates that extreme precipitation cannot be represented by a simple functional relationship, highlighting the complexity of the underlying physical processes governing the model response.

These characteristics highlight the intrinsic difficulty of predicting extreme precipitation and suggest that surrogate models relying on smooth or stationary assumptions may face limitations. Motivated by these observations, an ensemble-based uncertainty quantification framework is adopted to explore the sensitivity of the model to perturbations in parameters governing precipitation processes. This provides the basis for evaluating the ability of different surrogate models to capture both the predictive behavior and the statistical structure of extreme precipitation in the following sections.

3.2 Predictive performance

Building on the complexity of the response highlighted in the previous section, the ability of surrogate models to accurately predict extreme precipitation is now evaluated. Given the strong nonlinearities and variability observed in the relationship between parameters and the maximum precipitation, this assessment focuses on the capacity of each model to reproduce unseen realizations of the quantity of interest. The predictive performance of the surrogate models is first evaluated in terms of their ability to reproduce the maximum cumulative precipitation over the domain. As introduced in the previous section, this scalar

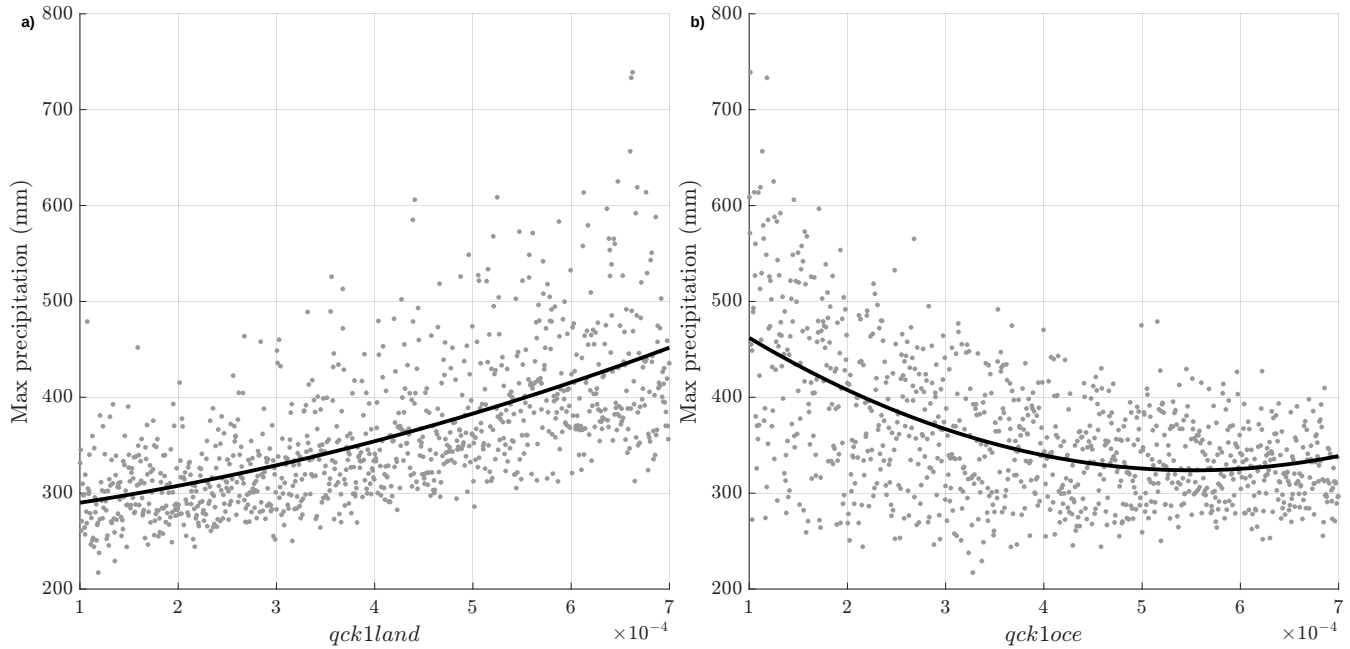


Figure 3. Relationship between maximum cumulative precipitation and key SUBEX scheme parameters. Scatter plots show the dependence of the maximum precipitation on (a) *qck1land* and (b) *qck1oce*. Grey markers represent individual ensemble simulations, and black curves correspond to second-order polynomial regressions highlighting the nonlinear trend.

quantity of interest is chosen to focus on the intensity of extreme events, which represents the primary source of discrepancy
 325 between the deterministic simulation and observations. All surrogate models are trained using the same ensemble dataset (1000
 LHS generated simulations), and their predictive skill is assessed on an independent validation set (10000 LHS generated
 simulations). Particular care has been taken to ensure the robustness of the surrogate construction. For polynomial-based
 approaches (PCE, PCK, and SSE), convergence with respect to polynomial degree has been verified, and hyperparameters
 have been selected to ensure stable approximations without overfitting.

330 Figure 4 shows the comparison between predicted and reference values for each surrogate model. Overall, all models capture
 the general trend of the response, with predictions broadly aligned along the one-to-one line. Differences in predictive skill
 are nevertheless visible, with varying degrees of dispersion around the one-to-one relationship. These differences are reflected
 in the quantitative metrics reported in Table 2. Model performance is quantified using standard pointwise metrics, namely the
 root-mean-square error (RMSE) and the coefficient of determination (R^2), computed on the independent validation dataset.
 335 PC-Kriging (PCK) achieves the best predictive performance, with an RMSE of 32.99 mm and an R^2 of 0.801. In comparison,
 PCE and Kriging exhibit similar performance levels, with RMSE values of 36.65 mm and 36.00 mm, and R^2 values of 0.754
 and 0.763, respectively. While PCK provides a noticeable improvement over these models, the differences remain moderate,
 indicating that all three approaches capture the main behavior of the response. In contrast, the stochastic spectral embedding
 (SSE) model shows significantly lower predictive accuracy, with an RMSE of 47.57 mm and an R^2 of 0.585, indicating a

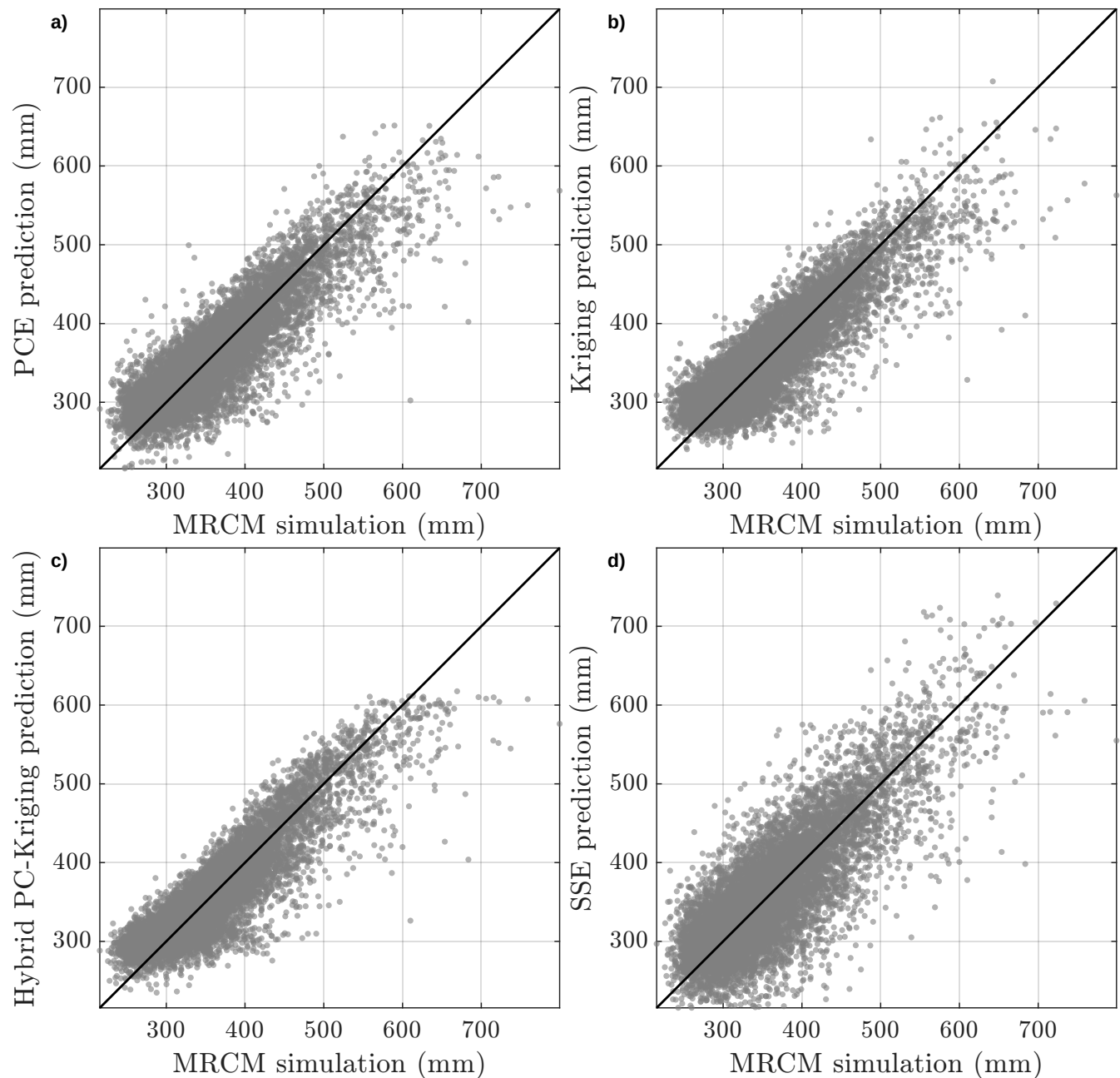


Figure 4. Scatter plots comparing surrogate model predictions and MRCM simulations for the maximum cumulative precipitation. The 1–1 line is shown for reference. All models reproduce the general trend of the response, with **(a)** PC-Kriging showing the closest agreement with the reference data, while **(d)** SSE exhibits larger dispersion. The differences illustrate the varying predictive performance of the surrogate models **(a–d)**.



Table 2. Predictive performance of surrogate models evaluated on the independent validation dataset. RMSE is expressed in mm, and R^2 denotes the coefficient of determination.

Model	RMSE (mm)	R^2
PCE	36.65	0.754
Kriging	36.00	0.763
PC-Kriging	32.99	0.801
SSE	47.57	0.585

reduced ability to reproduce individual realizations of the response. Overall, these results indicate that PC-Kriging provides the most accurate pointwise approximation of extreme precipitation among the considered models, although the improvement over PCE and Kriging remains moderate. This suggests that differences between surrogate models may become more pronounced when considering aspects of the response beyond pointwise accuracy, such as the representation of variability and extremes, which are examined in the following section.

3.3 Distributional performance

While predictive performance provides insight into the ability of surrogate models to reproduce pointwise values of extreme precipitation, it does not fully characterize their ability to represent the variability and statistical structure of the response. This is particularly important for extreme events, where the distribution, and especially its upper tail, plays a central role. To address this, the performance of the surrogate models is further evaluated in terms of their ability to reproduce the probability distribution of the maximum cumulative precipitation. Figure 5 presents kernel density estimates (KDE) of the distribution of maximum precipitation obtained from the independent validation dataset (referred to as the Reference in Figure 5) and from each surrogate model. The deterministic simulation and the observed event are also indicated for reference. The results reveal a marked contrast with the predictive performance analysis. In particular, the stochastic spectral embedding (SSE) model provides the closest match to the reference distribution, capturing both the overall shape and the heavy upper tail of the response. In contrast, PCE, Kriging, and PCK exhibit noticeable discrepancies. These differences are quantitatively confirmed by distribution-based metrics summarized in Table 3. SSE consistently yields the lowest Kolmogorov–Smirnov distance, Wasserstein distance, Kullback–Leibler divergence, and Energy distance, indicating a superior ability to reproduce the statistical distribution of extreme precipitation. Conversely, PCK, which performs best in terms of predictive accuracy, shows larger discrepancies in these distributional metrics, highlighting a divergence between pointwise accuracy and statistical fidelity. An important feature of Figure 5 is the relative position of the observed event and the deterministic simulation within the modeled probability distribution. The observed precipitation corresponds to a low-probability realization located in the lower tail of the modeled distribution, whereas the deterministic simulation is shifted toward larger precipitation values and lies on the descending branch of the distribution, closer to its mode. Consequently, the most probable precipitation intensity predicted by the model is intermediate between the observed event and the deterministic simulation. This indicates that, under the adopted

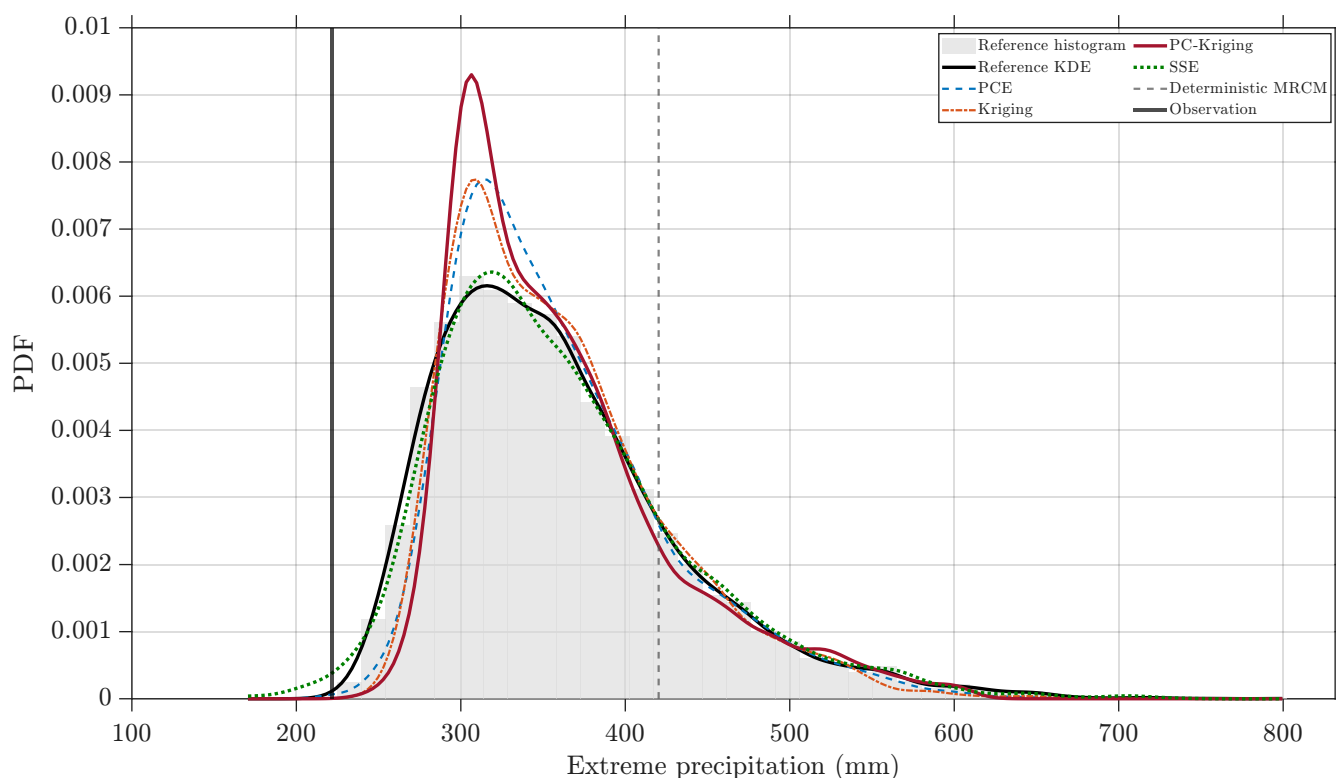


Figure 5. Kernel density estimates of the maximum cumulative precipitation for the reference ensemble and surrogate models. The deterministic simulation and observed event are indicated along the horizontal axis. SSE provides the closest match to the reference distribution, particularly in the upper tail, while other models underestimate the probability of extreme values.

Table 3. Distributional performance metrics for surrogate models. Lower values indicate better agreement with the reference distribution.

Model	KS	Wasserstein	KL	Energy
PCE	0.067	6.12	0.045	0.59
Kriging	0.065	6.35	0.06	0.52
PCK	0.095	6.52	0.121	0.72
SSE	0.016	2.36	0.011	0.19

parameter uncertainty, the regional climate model assigns a relatively low probability to reproducing an event comparable to the observations. This behavior reflects the strongly nonlinear propagation of parameter uncertainty through the regional climate model. Although the uncertain parameters are sampled from simple uniform distributions, the resulting response is highly asymmetric, with a pronounced upper tail. The location of the observed event within the lower-probability region of this distribution suggests that the default model configuration is not well suited to reproduce this particular extreme event. More importantly, it highlights the limitations of calibration strategies based primarily on seasonal or climatological statistics.



Parameter sets that accurately reproduce long-term mean conditions do not necessarily maximize the likelihood of individual high-impact events, whose governing processes may differ substantially from those controlling the climatological average. The fact that the mode of the distribution lies between the deterministic simulation and the observation further suggests that an alternative calibration strategy, explicitly accounting for extremes, could improve the representation of rare precipitation events.

375 These characteristics also help explain the differences observed among the surrogate models. Surrogate models optimized to minimize pointwise prediction errors, such as PCK, primarily capture the central tendency of the response distribution. In contrast, the adaptive partitioning strategy employed by SSE enables a more faithful representation of the overall probability distribution, particularly in regions where the response exhibits strong nonlinearities. Overall, these results reveal a fundamental trade-off between predictive accuracy and distributional fidelity in surrogate modeling of extreme precipitation.

380 PCK provides the highest deterministic predictive accuracy according to RMSE and (R^2), SSE more accurately reproduces the statistical distribution of the regional climate model response. This distinction is particularly important for uncertainty quantification and risk-informed applications, where accurately representing the probability of rare events is often more valuable than minimizing average prediction errors.

3.4 Sensitivity analysis

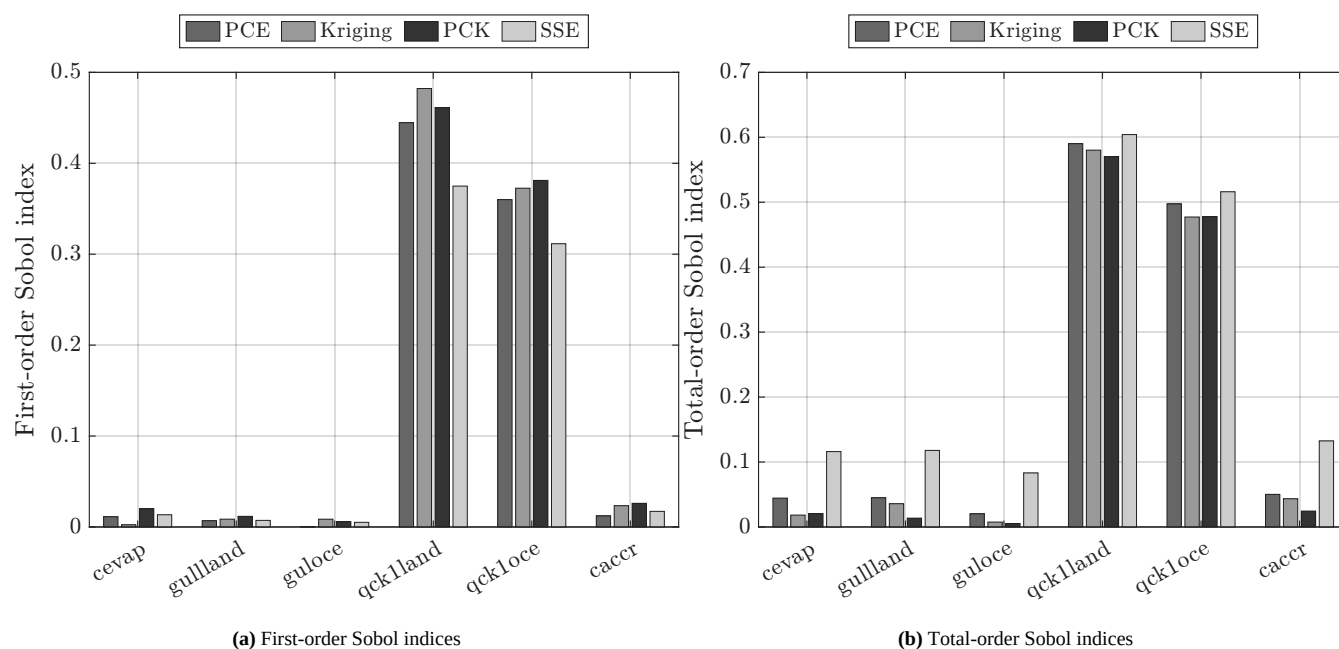


Figure 6. Sobol sensitivity indices for extreme precipitation. (a) First-order indices quantify the main effects of each parameter. (b) Total-order indices include both main effects and interactions. The dominant parameters (*qck1land* and *qck1oce*) show a clear increase from first- to total-order indices, indicating significant interaction effects in controlling extreme precipitation variability.



385 To identify the dominant drivers of extreme precipitation variability, a global sensitivity analysis is performed using Sobol indices computed from the surrogate models. These indices quantify the contribution of each input parameter to the variance of the quantity of interest, defined as the maximum cumulative precipitation over the domain. Figure 6 presents both first-order and total-order Sobol indices for all surrogate models. A consistent ranking of parameter importance is observed across all approaches, with *qcklland* and *qckloce* clearly dominating the response variability. The remaining parameters (*cevap*, *gulland*,
 390 *guloce*, and *caccr*) exhibit comparatively minor contributions. The comparison between first-order and total-order indices further reveals the presence of interaction effects, particularly among the dominant parameters. For instance, the Sobol index of *qcklland* increases from approximately 0.44 to 0.59 when interactions are included, while *qckloce* increases from about 0.36 to 0.50. This indicates that a significant fraction of the variance is explained by interactions between these parameters rather than by their individual contributions alone. Although the overall sensitivity structure is consistent across surrogate
 395 models, differences can be observed in the magnitude of the Sobol indices. In particular, the SSE model tends to assign slightly higher contributions to secondary parameters, which results in a corresponding slight underestimation of the dominant parameters compared to the other surrogate models. These differences may partly arise from the stochastic nature of Sobol index estimation, but they are also consistent with the structural characteristics of SSE. The partition-based representation of the response can redistribute variance contributions across subdomains, leading to a more diffuse attribution of sensitivity.
 400 Despite these variations in magnitude, the overall conclusions of the sensitivity analysis remain unchanged. All surrogate models consistently identify the same dominant parameters and interaction patterns, confirming that extreme precipitation variability is primarily controlled by a small subset of parameters.

From a physical perspective, the dominance of *qcklland* and *qckloce* highlights the important role of cloud-water auto-conversion in shaping extreme precipitation intensity within the SUBEX large-scale cloud and precipitation scheme. These
 405 parameters directly influence the efficiency of precipitation formation, thereby regulating rainfall intensity. The interaction between these parameters reflects the coupled influence of land–atmosphere processes and oceanic moisture supply, which is consistent with the mechanisms driving intense precipitation in this region. Overall, the sensitivity analysis provides a robust link between parameter uncertainty and the variability of extreme precipitation, and supports the interpretation of surrogate model behavior presented in the previous sections.

410 3.5 Local analysis over Sidi Ifni and Midelt

To assess the robustness of the previous findings, the analysis is repeated at the local scale by considering the same quantity of interest over two regions where the deterministic simulation exhibits contrasting biases with respect to the observations, namely Sidi Ifni along the Atlantic coast and Midelt in the Middle Atlas (Figure 1). As discussed previously, the deterministic simulation substantially overestimates precipitation over Sidi Ifni while underestimating precipitation over Midelt. These two
 415 regions therefore provide complementary test cases for evaluating whether the conclusions drawn at the domain scale remain valid under different hydrometeorological conditions. Following the same methodology as for the regional analysis, the surrogate models are evaluated using both predictive metrics (RMSE and R^2) and distributional metrics (KS, Wasserstein, KL and Energy). The corresponding results are summarized in Table 4. Overall, the local analysis confirms the conclusions obtained



at the regional scale. At both locations, PCK consistently provides the highest pointwise predictive accuracy, achieving the
 420 lowest RMSE and the highest coefficient of determination. However, this superior deterministic performance does not translate
 into the best representation of the underlying probability distribution. Conversely, SSE, despite exhibiting the lowest predictive
 accuracy, systematically provides the smallest distributional discrepancies. The contrast between predictive and distributional
 performance is especially pronounced over Midelt, where SSE reduces the Wasserstein distance by nearly a factor of five
 compared with the other surrogate models while simultaneously providing the lowest KS, KL and Energy distances. Over
 425 Sidi Ifni, the differences between surrogate models are less marked, with all approaches producing comparable KS distances.
 Nevertheless, SSE still achieves the lowest Wasserstein and KL divergence, indicating a slightly more faithful representation
 of the probability distribution despite its lower pointwise accuracy. These local results demonstrate that the trade-off identified
 at the regional scale is not an artifact of spatial aggregation but remains valid at locations characterized by markedly different
 model biases. Although the magnitude of the improvement in distributional fidelity varies between regions, the overall conclu-
 430 sion remains unchanged: surrogate models optimized for deterministic prediction are not necessarily the most appropriate for
 reproducing the probabilistic behavior of extreme precipitation. This finding further emphasizes that the choice of surrogate
 model should ultimately be guided by the intended application, whether the objective is deterministic prediction or uncertainty
 quantification.

Table 4. Predictive and distributional performance metrics at the two selected locations. RMSE and R^2 quantify predictive accuracy, whereas
 the Kolmogorov–Smirnov (KS) and Wasserstein distances evaluate the agreement between the modeled and reference distributions. Lower
 values indicate better performance for RMSE, KS, and Wasserstein distance, while higher values indicate better performance for R^2 .

Location	Model	RMSE	R^2	KS	Wasserstein	KL	Energy
Sidi Ifni	PCE	15.80	0.89	0.018	1.67	0.012	0.04
	Kriging	16.34	0.88	0.021	1.99	0.034	0.06
	PCK	14.90	0.9	0.021	1.8	0.017	0.05
	SSE	21.95	0.79	0.021	1.59	0.007	0.03
Midelt	PCE	6.48	0.57	0.056	1.38	0.065	0.09
	Kriging	6.11	0.62	0.062	1.43	0.218	0.11
	PCK	5.88	0.65	0.073	1.45	0.298	0.09
	SSE	9.02	0.17	0.012	0.32	0.007	0.01

The probability distributions of the quantity of interest at Sidi Ifni and Midelt, together with the locations of the deterministic
 435 simulation and the observations, are presented in Figure 7. Beyond evaluating the surrogate models, these distributions provide
 valuable insight into the spatial variability of the uncertainty associated with extreme precipitation. A first notable observation
 concerns the relative positions of the mode of the distribution, the deterministic simulation, and the observation. These three
 quantities are not consistently positioned across the study area, indicating that the probabilistic behavior of the regional climate
 model varies substantially from one location to another. For Sidi Ifni, the results closely resemble those obtained at the regional
 440 scale. This is expected since the regional maximum precipitation is located along the Atlantic coast, in the vicinity of Sidi Ifni,

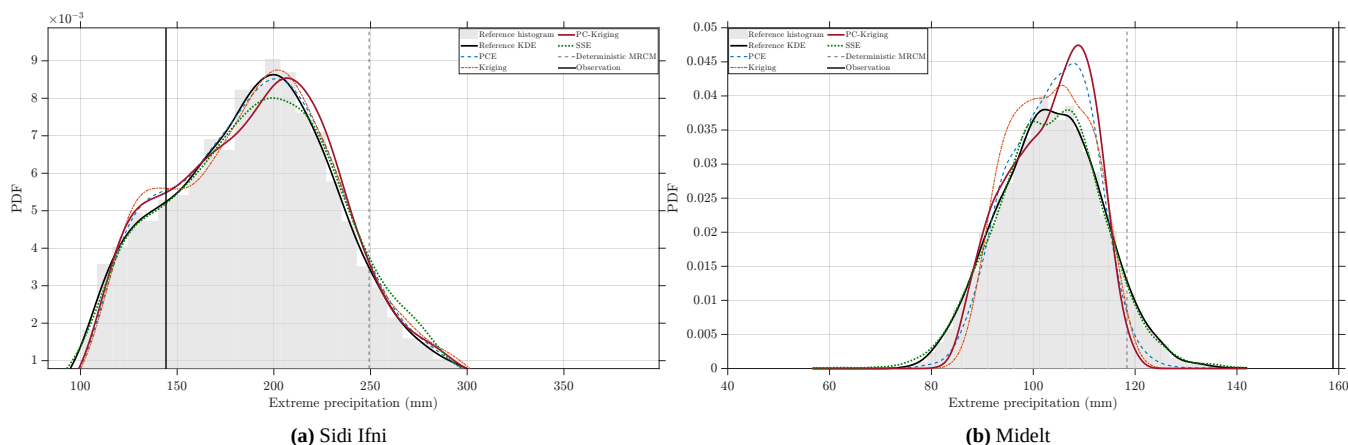


Figure 7. Kernel density estimates of the maximum cumulative precipitation for the reference ensemble and surrogate models at Sidi Ifni (a) and Midelt (b)

making the regional quantity of interest largely representative of the local behavior. Consequently, the observation lies in the lower-probability region of the distribution, while the deterministic simulation is shifted toward higher precipitation values, with the mode located between the two. In contrast, a markedly different behavior is observed over Midelt. Here, the observed precipitation is located far into the upper tail of the modeled distribution, indicating that the regional climate model assigns a very low probability to reproducing an event of comparable magnitude. The deterministic simulation is also located within the upper tail but remains substantially closer to the mode than to the observation. This suggests that the model is able to generate intense precipitation over the Middle Atlas, yet still struggles to reproduce the most extreme observed events. Although the relative positions of the deterministic simulation and the observations over Sidi Ifni are consistent with the regional analysis, the shape of the probability distribution differs considerably. In particular, the local distribution exhibits negative skewness, in contrast to the positively skewed distribution obtained at the regional scale, and is characterized by a larger kurtosis than both the regional distribution and that obtained over Midelt. These differences demonstrate that the statistical characteristics of the model response are themselves spatially variable and cannot be inferred solely from the regional analysis. Overall, these findings reinforce the conclusion that uncertainty associated with extreme precipitation exhibits pronounced spatial heterogeneity. They also suggest that calibration strategies based exclusively on regional or climatological statistics may fail to adequately represent the diversity of local precipitation regimes. Consequently, adaptive calibration approaches capable of accounting for spatially varying uncertainty may be required to improve the reliability of regional climate model simulations for extreme precipitation and to increase confidence in their use for impact assessment and risk-informed decision making.



4 Discussion

4.1 Trade-off between predictive accuracy and distributional fidelity

460 The results presented in sections 3.2 and 3.3 reveal a clear trade-off between predictive accuracy and distributional fidelity in surrogate modeling of extreme precipitation. This behavior is further confirmed by the local analyses presented in section 3.5, demonstrating that the trade-off is robust across regions exhibiting contrasting precipitation characteristics. While PCK consistently provides the highest pointwise predictive accuracy, as indicated by lower RMSE and higher R^2 , the SSE surrogate more faithfully reproduces the probability distribution of the regional climate model response, particularly in the low-probability

465 regions of the distribution. This apparent contradiction originates from the fundamentally different objectives optimized by the two surrogate modeling strategies. PCK combines a global polynomial approximation with a local Gaussian process correction, resulting in a surrogate specifically designed to minimize prediction errors over the entire parameter space (Schöbi et al., 2015). Since the model is trained using error-based criteria such as mean squared error or likelihood maximization, it primarily approximates the conditional mean response (Gneiting and Raftery, 2007). Consequently, PCK accurately repro-

470 duces the central tendency of the response but does not explicitly seek to preserve the underlying probability distribution or the occurrence of rare events. Conversely, SSE relies on an adaptive partitioning of the input space in which local surrogate models are constructed (Marelli et al., 2021). Instead of representing the response through a single global approximation, SSE decomposes the problem into multiple locally homogeneous regions, allowing the surrogate to adapt to different response regimes. This local representation naturally provides greater flexibility for capturing abrupt nonlinear transitions and hetero-

475 geneous response behavior, leading to a more accurate reconstruction of the overall probability distribution, including its tails. The different behaviors of PCK and SSE are consistent with previous studies on surrogate modeling of rare events. Schöbi and Sudret (2017) showed that classical PCK requires adaptive enrichment of the experimental design to accurately approximate low-probability regions of the response. Similarly, Singh et al. (2022) demonstrated that globally trained Gaussian-process surrogates may fail to reproduce rare extreme responses, motivating the development of locally adaptive surrogate

480 formulations. In this context, SSE can be viewed as an alternative approach that improves the representation of rare events through structural adaptability rather than adaptive sampling. More fundamentally, this trade-off reflects the intrinsic complexity of the regional climate model response. As shown in Figure 3, the relationship between the uncertain SUBEX parameters and extreme precipitation is strongly nonlinear, with substantial dispersion indicating that similar parameter values may lead to markedly different precipitation intensities. Furthermore, the nonlinear propagation of parameter uncertainty transforms simple

485 uniform input distributions into highly asymmetric and heavy-tailed response distributions. Under these conditions, minimizing pointwise prediction errors does not necessarily imply an accurate representation of the statistical properties of the response. Consequently, surrogate models optimized for deterministic prediction naturally favor the central tendency of the response, whereas models with greater structural flexibility are better suited to preserving the variability and probabilistic characteristics of extreme precipitation. These findings demonstrate that no single surrogate model is universally optimal. Instead, the choice

490 of surrogate model should be guided by the intended application. For deterministic prediction and parameter estimation, PCK remains an excellent choice owing to its superior predictive accuracy. In contrast, for uncertainty quantification, probabilistic



hazard assessment, and risk-informed applications, accurately reproducing the probability distribution of the response becomes equally important, making adaptive surrogate models such as SSE particularly attractive. More broadly, this study highlights that surrogate evaluation should not rely solely on deterministic performance metrics but should also incorporate measures of distributional fidelity when the objective is to characterize uncertainty associated with extreme events.

4.2 Implications for regional climate model behavior and calibration

An important finding of this study is that the observed precipitation consistently occupies low-probability regions of the modeled distributions, whereas the deterministic simulations remain closer to the central tendency of the response. The exact position of the observation varies spatially: it lies in the lower tail for the regional analysis and Sidi Ifni, but in the upper tail over Midelt. Despite these differences, the common feature is that the regional climate model assigns a relatively low probability to reproducing the observed event. This behavior is consistent with the mismatch between modeled and observed precipitation extremes documented across multiple regions and climate models (Bador et al., 2020). Our results therefore suggest that, although the default model configuration can represent the more probable responses arising from parameter uncertainty, it has greater difficulty representing the range of responses associated with individual extreme events. The difficulty of reproducing the magnitude and frequency of extreme precipitation with regional climate models is well documented (Kendon et al., 2014; O’Gorman, 2015). Such discrepancies are commonly attributed to insufficient spatial resolution and to limitations in the representation of processes such as convection, cloud microphysics, and land–atmosphere interactions (Ivušić et al., 2021; Kalmár et al., 2024). However, increasing model resolution or improving individual physical parameterizations does not necessarily eliminate biases in precipitation extremes. Substantial differences in extreme precipitation have been reported even among high-resolution climate simulations (Wehner et al., 2014; Bador et al., 2020). The present results suggest that model calibration provides an additional perspective from which these discrepancies should be considered. Climate models are generally calibrated against multiple aspects of the observed climate, including long-term means, seasonal cycles, large-scale circulation, temperature, precipitation, and cloud properties (Hourdin et al., 2017; Mauritsen et al., 2012). Such multi-objective tuning is necessary to obtain a physically consistent representation of the mean climate, but it does not explicitly guarantee an accurate representation of rare events. Parameter combinations that perform well against climatological targets may therefore differ from those required to reproduce individual precipitation extremes. Consequently, calibration can preferentially constrain the more frequently sampled states of the climate system, while low-probability regions of the response remain comparatively weakly constrained (Gagne et al., 2020; Fischer et al., 2024; Zhang et al., 2022; Ho et al., 2012). The uncertainty analysis presented here is consistent with this interpretation. Despite exploring a broad range of physically plausible parameter combinations, the observations remain located in low-probability regions of the modeled distributions at both regional and local scales. Reproducing these events therefore requires model responses that occur relatively infrequently within the explored parameter space. Similar difficulties have been reported when optimizing different parameter combinations for climate-model representation (Fischer et al., 2025). Moreover, these sparsely represented portions of the response are intrinsically difficult to characterize from finite ensembles (Fischer et al., 2024, 2025), particularly when parameter perturbations propagate nonlinearly through the climate model (Zhang et al., 2022). The resulting uncertainty is therefore not solely a question of whether an individual



parameterization is physically adequate, but also of whether the calibration and sampling framework sufficiently constrains the model behavior relevant to extremes. The physical characteristics of the study region further reinforce this interpretation. Extreme precipitation over southern Morocco arises from interactions among Atlantic moisture transport, coastal processes, and orographic lifting over the Atlas Mountains (Chaqdid et al., 2023). Because these processes interact nonlinearly, relatively small changes in their parameterized representation can produce substantial differences in localized precipitation intensity. Parameter combinations that reproduce average climatic conditions are therefore not necessarily those that best represent the coupled processes governing individual extreme events. The contrasting probability distributions obtained over Sidi Ifni and Midelt further indicate that this limitation can vary spatially, emphasizing the difficulty of defining a single calibration strategy that performs equally well across climatological regimes and extreme-event conditions. Overall, these findings suggest that improving the simulation of extreme precipitation may require calibration strategies that explicitly incorporate information about extremes alongside conventional climatological targets. This does not imply replacing climatological calibration, which remains essential for maintaining a physically consistent mean climate, but rather extending it toward multi-objective frameworks that also constrain the tails of relevant response distributions. More generally, evaluating regional climate models solely through deterministic simulations or mean-state performance may overlook important information contained in their probabilistic response. Integrating uncertainty quantification and extreme-sensitive objectives into the calibration process therefore offers a promising pathway toward regional climate simulations that remain climatologically credible while providing a more reliable representation of high-impact precipitation events.

4.3 Implications for surrogate modeling and uncertainty quantification

The results of this study have important implications for the use of surrogate models in uncertainty quantification of climate simulations. In particular, they demonstrate that surrogate models should not be evaluated solely based on predictive accuracy, as this criterion does not necessarily reflect their ability to represent the statistical structure of extreme events. Although this has not been fully established, similar arguments have been advanced in previous studies. For example, Weber et al. (2020) shows that the use of a neural-network-based surrogate whose structure is assessed using predictive accuracy metrics fails to adequately represent precipitation in regions where it is mainly driven by intense events. In addition, many of the surrogate models developed in the literature use relatively small numbers of samples (usually fewer than one hundred simulations; Couvreur et al. (2021); Weber et al. (2020); Fischer et al. (2024, 2025)) compared with the current study (over one thousand simulations), which favors the characterization of the first and second statistical moments rather than higher-order moments associated with the tail of the distribution. This generally leads to the well-known overfitting problem, which is also relevant for statistically based emulators that can be trained using small training samples (Van Katwyk et al., 2026). The observed trade-off between predictive accuracy and distributional fidelity indicates that different surrogate models are suited to different objectives. Gaussian-based models such as PC-Kriging, which are optimized to minimize prediction error, are well adapted to tasks requiring accurate emulation of individual realizations, such as parameter estimation. This was the case, for example, in the work of Fischer et al. (2025). However, they can show some limitations regarding the estimation of extreme values. In the work of Schöbi et al. (2016), the authors demonstrate that an adaptive experimental design is necessary for PCK to estimate rare



560 events. In contrast, approaches such as SSE, which rely on more flexible representations of the response, are better suited to capturing the statistical distribution and tail behavior, making them more appropriate for applications involving risk assessment or extreme-event analysis. This has been further demonstrated in the study of Wagner et al. (2022), where the use of SSE has been shown to be effective in tracking rare events. These findings suggest that surrogate model selection should be guided by the intended application rather than by a single performance metric. In particular, the use of distribution-based metrics, 565 such as those employed in this study, provides a more comprehensive evaluation of model performance in the context of extremes. This is especially relevant for climate applications, where the accurate representation of variability and tail behavior is often more critical than minimizing average prediction error. More broadly, this work highlights a limitation of current uncertainty quantification practices, which often rely on surrogate models calibrated and evaluated using global error measures. Such approaches may overlook important features of the response, particularly in low-probability regions associated with 570 extreme events. Addressing this limitation may require the development of hybrid strategies that combine accurate predictive performance with improved representation of variability, for example, through adaptive sampling, local surrogate modeling, or distribution-aware training criteria. Overall, the results emphasize the need to explicitly account for the probabilistic structure of model responses when applying surrogate models to climate simulations. By distinguishing between predictive accuracy and distributional fidelity, this study provides a framework for more robust and application-specific surrogate modeling in the 575 context of extreme precipitation.

4.4 Limitations and perspectives

Several limitations of the present study should be acknowledged. First, the analysis is based on a single extreme precipitation event, which limits the generality of the conclusions. While the results provide clear evidence of a trade-off between predictive accuracy and distributional fidelity, further investigation across multiple events, climatic regimes, and model configurations would be necessary to assess the robustness of this behavior in a broader context. Second, the study focuses on a 580 scalar quantity of interest defined as the maximum cumulative precipitation over the domain. This choice is well suited to the analysis of extreme intensity but does not capture the full spatial structure of precipitation fields. Extending the analysis to spatially distributed quantities would provide additional insight into how surrogate models represent localized processes and spatial variability, although this would introduce additional methodological challenges related to high-dimensional outputs. 585 Third, the surrogate models are constructed within a predefined parameter space, and their performance is inherently conditioned on the sampling strategy used to generate the ensemble. Although Latin Hypercube Sampling provides a space-filling design, extreme events remain associated with low-probability regions that are only sparsely sampled. This limitation is intrinsic to ensemble-based approaches and highlights the need for strategies that better target tail behavior, such as adaptive sampling or importance sampling techniques. From a methodological perspective, the comparison performed in this study 590 is limited to a set of surrogate models representative of current practices in uncertainty quantification. While these models capture a wide range of approximation strategies, alternative approaches—such as deep-learning-based emulators or hybrid probabilistic frameworks—have recently been proposed to address the challenges of modeling complex and nonlinear climate responses. However, these approaches typically require substantial computational resources for training and hyperparameter



tuning, particularly when applied to high-dimensional inputs or large ensembles. In the context of regional climate modeling, where each simulation is already computationally expensive, the additional cost associated with such methods can represent a significant limitation. As a result, the use of relatively efficient surrogate models, such as those considered in this study, remains highly relevant for practical uncertainty quantification workflows. Investigating how to balance model complexity, computational cost, and the representation of extreme events therefore constitutes an important direction for future research. This is particularly critical in regional climate simulations, where ensemble generation remains the dominant computational cost. Finally, the present results highlight the importance of calibration strategies in shaping model behavior. The fact that extreme events are poorly represented within the modeled distribution suggests that current calibration practices, which prioritize mean climatological performance, may not be sufficient for applications focused on extremes. Future work could explore calibration approaches that explicitly incorporate distributional or tail-based objectives, thereby improving the representation of rare but high-impact events. Overall, these limitations point toward several avenues for future work, including the extension of the analysis to multiple events and regions, the development of surrogate models tailored to extreme-value behavior, and the integration of distribution-aware calibration strategies into regional climate modeling frameworks.

5 Conclusions

This study investigated the performance of surrogate models for representing extreme precipitation under parametric uncertainty in a regional climate model. By combining ensemble simulations with both predictive and distribution-based evaluation metrics, we provided a comprehensive assessment of how different surrogate modeling approaches capture the behavior of extreme events. The results reveal a fundamental trade-off between predictive accuracy and distributional fidelity. While PC-Kriging provides the most accurate pointwise predictions of the maximum cumulative precipitation, it tends to reproduce the central tendency of the response and underrepresent variability in the upper tail. In contrast, the stochastic spectral embedding approach offers a more faithful representation of the statistical distribution, particularly for extreme values, but at the cost of reduced predictive accuracy for individual realizations. Beyond surrogate model comparison, the analysis highlights an important limitation of regional climate modeling. The observed event occupies a low-probability region of the modeled response, but its position varies spatially: it lies toward the lower tail at the domain scale and over Sidi Ifni and toward the upper tail over Midelt. Thus, the model does not exhibit a uniform underestimation of extreme precipitation; rather, the magnitude and sign of the bias depend strongly on location. This spatially heterogeneous behavior is consistent with a model configuration calibrated primarily against climatological rather than event-specific characteristics. Sensitivity analysis further shows that extreme precipitation is primarily controlled by parameters governing precipitation efficiency over land and ocean, thereby linking the statistical behavior of the response to underlying physical processes. These findings emphasize the importance of accounting for parameter uncertainty when analyzing extreme events. Overall, this study demonstrates that surrogate model evaluation for climate applications should extend beyond predictive accuracy and explicitly consider distributional properties, particularly when targeting extremes. The results provide practical guidance for selecting surrogate models depending on the



intended application and highlight the need for uncertainty quantification frameworks that better capture rare but high-impact events in regional climate simulations.

Code and data availability. The complete model source code, processing scripts, configuration files, and analysis datasets used in this study are permanently archived and publicly available via the Zenodo repository at <https://doi.org/10.5281/zenodo.22051098> (El Mocayd, 2026).

630 *Author contributions.* NEM conceptualized the study, supervised the research, secured the project funding, performed the data analysis and developed the surrogate modeling software. AC conducted the model simulations and contributed to the data interpretation. Both authors actively contributed to the writing and editing of the manuscript.

Competing interests. The authors declare that they have no competing interests.

Financial support. This work was supported by the OCP Foundation within the framework of the UM6P-MIT Research Program (UMRP).

635 *Acknowledgements.* The authors sincerely thank Prof. Elfatih Eltahir and Dr. Yeonwoo Choi from Parsons Laboratory of the Civil and Environmental Department at MIT for providing the MRCM source code and for their valuable support and guidance during the setup and execution of the climate model simulations. The authors also gratefully acknowledge the University Mohammed VI Polytechnic (UM6P) for providing the high-performance computing resources on the Toubkal supercomputer used to perform the simulations for this study (Kissami et al., 2025).



640 References

- Al-Ghosoun, A., El Moçayd, N., and Seaid, M.: A surrogate model for efficient quantification of uncertainties in multilayer shallow water flows, *Environmental Modelling & Software*, 144, 105 176, 2021.
- Bador, M., Boé, J., Terray, L., Alexander, L. V., Baker, A., Bellucci, A., et al.: Impact of Higher Spatial Atmospheric Resolution on Precipitation Extremes Over Land in Global Climate Models, *Journal of Geophysical Research: Atmospheres*, 125,
 645 <https://doi.org/10.1029/2019JD032184>, 2020.
- Berner, J., Ha, S.-Y., Hacker, J., Fournier, A., and Snyder, C.: Model uncertainty in a mesoscale ensemble prediction system: Stochastic versus multiphysics representations, *Monthly Weather Review*, 139, 1972–1995, 2011.
- Berner, J., Achatz, U., Batte, L., Bengtsson, L., Cámara, A. d. l., Christensen, H. M., Colangeli, M., Coleman, D. R., Crommelin, D., Dolaptchiev, S. I., et al.: Stochastic parameterization: Toward a new view of weather and climate models, *Bulletin of the American*
 650 *Meteorological Society*, 98, 565–588, 2017.
- Blatman, G. and Sudret, B.: Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of computational Physics*, 230, 2345–2367, 2011.
- Bocquet, M.: Surrogate modeling for the climate sciences dynamics with machine learning and data assimilation, *Frontiers in Applied Mathematics and Statistics*, 9, 1133 226, 2023.
- 655 Bouabid, S., Sejdinovic, D., and Watson-Parris, D.: FaIRGP: A Bayesian energy balance model for surface temperatures emulation, *Journal of Advances in Modeling Earth Systems*, 16, e2023MS003 926, 2024.
- Bracco, A., Brajard, J., Dijkstra, H. A., Hassanzadeh, P., Lessig, C., and Monteleoni, C.: Machine learning for the physics of climate, *Nature Reviews Physics*, 7, 6–20, 2025.
- Chaqqid, A., Tuel, A., El Fatimy, A., and El Moçayd, N.: Extreme rainfall events in Morocco: Spatial dependence and climate drivers,
 660 *Weather and Climate Extremes*, 40, 100 556, 2023.
- Choi, Y.-W., Campbell, D. J., Aldridge, J. C., and Eltahir, E. A.: Near-term regional climate change over Bangladesh, *Climate Dynamics*, 57, 3055–3073, <https://doi.org/10.1007/s00382-021-05856-z>, 2021.
- Couvreux, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., et al.: Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement, *Journal of Advances in Modeling Earth Systems*,
 665 13, <https://doi.org/10.1029/2020MS002217>, 2021.
- El Moçayd, N.: Code and Data for MRCM and Surrogate Model Extreme Precipitation Analysis, <https://doi.org/10.5281/zenodo.22051098>, 2026.
- Elkarii, M., Boukharfane, R., Benjelloun, S., Bouallou, C., and El Moçayd, N.: Global sensitivity analysis for phosphate slurry flow in pipelines using generalized polynomial chaos, *Physics of Fluids*, 35, 2023.
- 670 Fischer, M., Knippertz, P., van der Linden, R., Lemburg, A., Pante, G., Proppe, C., and Marsham, J. H.: Quantifying uncertainty in simulations of the West African monsoon with the use of surrogate models, *Weather and Climate Dynamics*, 5, 511–536, <https://doi.org/10.5194/wcd-5-511-2024>, 2024.
- Fischer, M., Knippertz, P., and Proppe, C.: Surrogate-based model parameter optimization in simulations of the West African monsoon, *Weather and Climate Dynamics*, 6, 113–130, <https://doi.org/10.5194/wcd-6-113-2025>, 2025.



- 675 Foley, J. A., Prentice, I. C., Ramankutty, N., Levis, S., Pollard, D., Sitch, S., and Haxeltine, A.: An Integrated Biosphere Model of Land Surface Processes, Terrestrial Carbon Balance, and Vegetation Dynamics, *Global Biogeochemical Cycles*, 10, 603–628, <https://doi.org/10.1029/96GB02692>, 1996.
- Forest, C. E., Stone, P. H., Sokolov, A. P., Allen, M. R., and Webster, M. D.: Quantifying Uncertainties in Climate System Properties with the Use of Recent Climate Observations, *Science*, 295, 113–117, <https://doi.org/10.1126/science.1064419>, 2002.
- 680 Fowler, H. J., Lenderink, G., Prein, A. F., Westra, S., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Do, H. X., et al.: Anthropogenic intensification of short-duration rainfall extremes, *Nature Reviews Earth & Environment*, 2, 107–122, 2021.
- Gagne, D. J., Christensen, H. M., Subramanian, A. C., and Monahan, A. H.: Machine Learning for Stochastic Parameterization: Generative Adversarial Networks in the Lorenz '96 Model, *Journal of Advances in Modeling Earth Systems*, 12, <https://doi.org/10.1029/2019MS001896>, 2020.
- 685 Gianotti, R. L.: Convective Organization and Precipitation over the Maritime Continent: New Insights from Observation and Modeling, Ph.D. thesis, Massachusetts Institute of Technology, <https://dspace.mit.edu/handle/1721.1/71107>, 2012.
- Gianotti, R. L. and Eltahir, E. A. B.: Regional Climate Modeling over the Maritime Continent. Part I: New Parameterization for Convective Cloud Fraction, *Journal of Climate*, 27, 1488–1503, <https://doi.org/10.1175/JCLI-D-13-00127.1>, 2014a.
- Gianotti, R. L. and Eltahir, E. A. B.: Regional Climate Modeling over the Maritime Continent. Part II: New Parameterization for Autocon-
- 690 version of Convective Rainfall, *Journal of Climate*, 27, 1504–1523, <https://doi.org/10.1175/JCLI-D-13-00171.1>, 2014b.
- Giorgi, F.: Thirty years of regional climate modeling: where are we and where are we going next?, *Journal of Geophysical Research: Atmospheres*, 124, 5696–5723, 2019.
- Gneiting, T. and Raftery, A. E.: Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association*, 102, 359–378, 2007.
- 695 Hall, P.: On Kullback-Leibler loss and density estimation, *The Annals of Statistics*, pp. 1491–1519, 1987.
- Ho, C. K., Stephenson, C. A. T., Collins, M., Ferro, C. A., and Brown, S. J.: Calibration Strategies: A Fundamental Indeterminacy in Postprocessing 21st Century Climate Projections, *Bulletin of the American Meteorological Society*, 93, <https://doi.org/10.1175/2011BAMS3110.1>, 2012.
- Holtzlag, A. A. M. and Boville, B. A.: Local versus Nonlocal Boundary-Layer Diffusion in a Global Climate Model, *Journal of Climate*, 6, 1825–1842, [https://doi.org/10.1175/1520-0442\(1993\)006<1825:LVNBLD>2.0.CO;2](https://doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2), 1993.
- 700 Hou, C. K. J. and Behdinan, K.: Dimensionality reduction in surrogate modeling: A review of combined methods, *Data science and engineering*, 7, 402, 2022.
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J.-C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., et al.: The art and science of climate model tuning, *Bulletin of the American Meteorological Society*, 98, 589–602, 2017.
- 705 Im, E.-S. and Eltahir, E. A. B.: Simulations of the Observed ‘Jump’ in the West African Monsoon and Its Underlying Dynamics Using the MIT Regional Climate Model, *International Journal of Climatology*, 38, 841–852, <https://doi.org/10.1002/joc.5214>, 2018.
- Im, E.-S., Gianotti, R. L., and Eltahir, E. A. B.: Improving the Simulation of the West African Monsoon Using the MIT Regional Climate Model, *Journal of Climate*, 27, 2209–2229, <https://doi.org/10.1175/JCLI-D-13-00188.1>, 2014.
- Ivušić, S., Güttler, I., Somot, S., Guérémy, J.-F., Horvath, K., and Alias, A.: Modelling extreme precipitation over the Dinaric Alps: An
- 710 evaluation of the CNRM-ALADIN regional climate model, *Quarterly Journal of the Royal Meteorological Society*, 147, 4425–4453, 2021.



- Jackson, C., Sen, M. K., and Stoffa, P. L.: An Efficient Stochastic Bayesian Approach to Optimal Parameter and Uncertainty Estimation for Climate Model Predictions, *Journal of Climate*, 17, 2828–2841, [https://doi.org/10.1175/1520-0442\(2004\)017<2828:AESBAT>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2828:AESBAT>2.0.CO;2), 2004.
- 715 Jiang, Z., Isenberg, N. M., Subba, T., Woo, H.-M., Serbin, S. P., Urban, N. M., and Kuang, C.: A framework for parametric and predictive uncertainty quantification in the E3SM land model: Assessing site and observable generalizability, *Journal of Advances in Modeling Earth Systems*, 18, e2025MS005 562, 2026.
- Kalmár, T., Pongrácz, R., Pieczka, I., and Hollós, R.: Evaluation of RegCM simulation ensemble using different parameterization scheme combinations: a case study for an extremely wet year in the Carpathian region, *Climate Dynamics*, 62, 8201–8225, 2024.
- 720 Kawai, S. and Shimoyama, K.: Kriging-model-based uncertainty quantification in computational fluid dynamics, in: 32nd AIAA Applied Aerodynamics Conference, p. 2737, 2014.
- Kendon, E. J., Roberts, N. M., Senior, C. A., and Roberts, M. J.: Heavier summer downpours with climate change revealed by weather forecast resolution model, *Nature Climate Change*, 4, 570–576, 2014.
- Kiehl, J. T., Hack, J. J., Bonan, G. B., Boville, B. A., Williamson, D. L., and Rasch, P. J.: The National Center for Atmospheric Research Community Climate Model: CCM3, *Journal of Climate*, 11, 1131–1149, [https://doi.org/10.1175/1520-0442\(1998\)011<1131:TNCFAR>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<1131:TNCFAR>2.0.CO;2), 1998.
- 725 Kim, Y.-H., Min, S.-K., Zhang, X., Sillmann, J., and Sandstad, M.: Evaluation of the CMIP6 multi-model ensemble for climate extreme indices, *Weather and Climate Extremes*, 29, 100 269, 2020.
- Kissami, I., Basmadjian, R., Chakir, O., and Abid, M. R.: TOUBKAL: a high-performance supercomputer powering scientific research in Africa: I. Kissami et al., *The Journal of Supercomputing*, 81, 1401, 2025.
- 730 Mansfield, L. A., Gupta, A., Burnett, A. C., Green, B., Wilka, C., and Sheshadri, A.: Updates on model hierarchies for understanding and simulating the climate system: A focus on data-informed methods and climate change impacts, *Journal of Advances in Modeling Earth Systems*, 15, e2023MS003 715, 2023.
- Marcella, M. P. and Eltahir, E. A. B.: Modeling the Summertime Climate of Southwest Asia: The Role of Land Surface Processes in Shaping the Climate of Semiarid Regions, *Journal of Climate*, 25, 704–719, <https://doi.org/10.1175/2011JCLI4080.1>, 2012.
- 735 Marcella, M. P. and Eltahir, E. A. B.: Introducing an Irrigation Scheme to a Regional Climate Model: A Case Study over West Africa, *Journal of Climate*, 27, 5708–5723, <https://doi.org/10.1175/JCLI-D-13-00116.1>, 2014.
- Marelli, S. and Sudret, B.: UQLab: A framework for uncertainty quantification in Matlab, in: *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, pp. 2554–2563, 2014.
- 740 Marelli, S., Wagner, P.-R., Lataniotis, C., and Sudret, B.: Stochastic spectral embedding, *International Journal for Uncertainty Quantification*, 11, 2021.
- Massey Jr, F. J.: The Kolmogorov-Smirnov test for goodness of fit, *Journal of the American statistical Association*, 46, 68–78, 1951.
- Mauritsen, T., Stevens, B., Roeckner, E., Crueger, T., Esch, M., Giorgetta, M., et al.: Tuning the climate of a global model, *Journal of Advances in Modeling Earth Systems*, 4, <https://doi.org/10.1029/2012MS000154>, 2012.
- 745 Morales, A., Posselt, D. J., Morrison, H., and He, F.: Assessing the influence of microphysical and environmental parameter perturbations on orographic precipitation, *Journal of the Atmospheric Sciences*, 76, 1373–1395, 2019.
- Norris, J., Hall, A., Neelin, J. D., Thackeray, C. W., and Chen, D.: Evaluation of the tail of the probability distribution of daily and subdaily precipitation in CMIP6 models, *Journal of Climate*, 34, 2701–2721, 2021.
- O’Gorman, P. A.: Precipitation extremes under climate change, *Nature Climate Change*, 2015.



- 750 O’Gorman, P. A. and Dwyer, J. G.: Using Machine Learning to Parameterize Moist Convection: Potential for Modeling of Climate, Climate Change, and Extreme Events, *Journal of Advances in Modeling Earth Systems*, 10, 2548–2563, <https://doi.org/10.1029/2018MS001351>, 2018.
- Pal, J. S. and Eltahir, E. A. B.: Future Temperature in Southwest Asia Projected To Exceed a Threshold for Human Adaptability, *Nature Climate Change*, 6, 197–200, <https://doi.org/10.1038/nclimate2833>, 2016.
- 755 Pal, J. S., Small, E. E., and Eltahir, E. A. B.: Simulation of Regional-Scale Water and Energy Budgets: Representation of Subgrid Cloud and Precipitation Processes within RegCM, *Journal of Geophysical Research: Atmospheres*, 105, 29 579–29 594, <https://doi.org/10.1029/2000JD900415>, 2000.
- Pal, J. S., Giorgi, F., Bi, X., Elguindi, N., Solmon, F., Gao, X., et al.: Regional Climate Modeling for the Developing World: The ICTP RegCM3 and RegCM3, *Bulletin of the American Meteorological Society*, 88, 1337–1349, <https://doi.org/10.1175/BAMS-88-9-1395>, 2007.
- 760 Palmer, T. N.: A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models, *Quarterly Journal of the Royal Meteorological Society*, 127, 279–304, 2001.
- Pierini, L., Gudmundsson, L., Quilcaille, Y., and Seneviratne, S. I.: Framework for global emulation of extreme precipitation along global warming trajectories, *Environmental Research Letters*, 21, 084 017, 2026.
- 765 Reichelt, T., Rainforth, T., and Watson-Parris, D.: Calibration of climate model parameterizations using Bayesian experimental design, *Machine Learning: Earth*, 2, 015 003, 2026.
- Rizzo, M. L. and Székely, G. J.: Energy distance, *wiley interdisciplinary reviews: Computational statistics*, 8, 27–38, 2016.
- Roy, P. T., El Moçayd, N., Ricci, S., Jouhaud, J.-C., Goutal, N., De Lozzo, M., and Rochoux, M. C.: Comparison of polynomial chaos and Gaussian process surrogates for uncertainty quantification and correlation estimation of spatially distributed open-channel steady flows, *Stochastic environmental research and risk assessment*, 32, 1723–1741, 2018.
- 770 Schneider, T., Behera, S., Boccaletti, G., Deser, C., Emanuel, K., Ferrari, R., Leung, L. R., Lin, N., Müller, T., Navarra, A., et al.: Harnessing AI and computing to advance climate modelling and prediction, *Nature Climate Change*, 13, 887–889, 2023.
- Schöbi, R. and Sudret, B.: Rare event estimation using polynomial-chaos-based Kriging, *Structural Safety*, 2017.
- Schöbi, R., Sudret, B., and Wiart, J.: Polynomial-chaos-based Kriging, *International Journal for Uncertainty Quantification*, 5, 2015.
- 775 Schöbi, R., Sudret, B., and Marelli, S.: Rare event estimation using Polynomial-Chaos Kriging, *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, 3, <https://doi.org/10.1061/AJRUA6.0000870>, 2016.
- Schöngart, S., Gudmundsson, L., Hauser, M., Pfeleiderer, P., Lejeune, Q., Nath, S., Seneviratne, S. I., and Schleussner, C.-F.: Introducing the MESMER-M-TPv0. 1.0 module: spatially explicit Earth system model emulation for monthly precipitation and temperature, *Geoscientific Model Development*, 17, 8283–8320, 2024.
- 780 Singh, D., Dwight, R. P., Laugesen, K., Beaudet, L., and Viré, A.: Probabilistic surrogate modeling of offshore wind-turbine loads with chained Gaussian processes, *arXiv preprint arXiv:2204.02258*, 2022.
- Sothearith, M., Park, D., and Ahn, K.-H.: Dynamic internal variability dominates uncertainty in modeling future extreme precipitation, *npj Climate and Atmospheric Science*, 2026.
- Sudret, B.: Polynomial chaos expansions and stochastic finite element methods, *Risk and reliability in geotechnical engineering*, pp. 265–300, 2014.
- 785 Tuel, A. and El Moçayd, N.: Evaluating extreme precipitation in gridded datasets with a novel station database in Morocco, *Stochastic Environmental Research and Risk Assessment*, 37, 3085–3097, 2023.



- Tuel, A. and Eltahir, E. A. B.: Why Is the Mediterranean a Climate Change Hot Spot?, *Journal of Climate*, 33, 5829–5843, <https://doi.org/10.1175/JCLI-D-19-0910.1>, 2020.
- 790 Van Katwyk, P., Fox-Kemper, B., Hewitt, H. T., and Bergen, K. J.: Rewiring climate modeling with machine learning emulators, *Communications Earth & Environment*, <https://doi.org/10.1038/s43247-026-03238-z>, 2026.
- Villani, C.: The wasserstein distances, in: *Optimal transport: old and new*, pp. 93–111, Springer, 2009.
- Wagner, P.-R., Marelli, S., Papaioannou, I., Straub, D., and Sudret, B.: Rare event estimation using stochastic spectral embedding, *Structural Safety*, 96, 102 179, <https://doi.org/10.1016/j.strusafe.2021.102179>, 2022.
- 795 Weber, T., Corotan, A., Hutchinson, B., Kravitz, B., and Link, R.: Technical note: Deep learning for creating surrogate models of precipitation in Earth system models, *Atmospheric Chemistry and Physics*, 20, 2303–2317, <https://doi.org/10.5194/acp-20-2303-2020>, 2020.
- Wehner, M. F., Reed, K. A., Li, F., Prabhat, Bacmeister, J., Cheng-Ta, C., Paciorek, C., Gleckler, P. J., Sperber, K. R., Collins, W. D., Gettelman, A., and Jablonowski, C.: The effect of horizontal resolution on simulation quality in the Community Atmospheric Model, CAM5.1, *Journal of Advances in Modeling Earth Systems*, 6, 980–997, 2014.
- 800 Winter, J. M. and Eltahir, E. A. B.: Modeling the Hydroclimatology of the Midwestern United States. Part 1: Current Climate, *Climate Dynamics*, 38, 573–593, <https://doi.org/10.1007/s00382-011-1182-2>, 2012.
- Winter, J. M., Pal, J. S., and Eltahir, E. A. B.: Coupling of Integrated Biosphere Simulator to Regional Climate Model Version 3, *Journal of Climate*, 22, 2743–2757, <https://doi.org/10.1175/2008JCLI2541.1>, 2009.
- Zeng, X., Zhao, M., and Dickinson, R. E.: Intercomparison of Bulk Aerodynamic Algorithms for the Computation of Sea Surface Fluxes Using TOGA COARE and TAO Data, *Journal of Climate*, 11, 2628–2644, [https://doi.org/10.1175/1520-0442\(1998\)011<2628:IOBAAF>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<2628:IOBAAF>2.0.CO;2), 1998.
- 805 Zhang, W., Furtado, K., Zhou, T., Wu, P., and Chen, X.: Constraining extreme precipitation projections using past precipitation variability, *Nature Communications*, 13, 6612, <https://doi.org/10.1038/s41467-022-34006-0>, 2022.