

Towards an Optimal Representation of Sub-Grid Heterogeneity in Land Surface Models

L. Torres-Rojas¹ , N. Vergopolan^{2,3} , J. D. Herman⁴ , and N. W. Chaney¹ 

¹Department of Civil and Environmental Engineering, Duke University, Durham, NC, USA, ²Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA, ³NOAA Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA, ⁴Department of Civil & Environmental Engineering, University of California, Davis, CA, USA

Key Points:

- A surrogate model can accurately predict the spatial structure of Land Surface Model's (LSM) hydrological output fields
- For a target level of complexity, the optimal LSM tile configuration was determined using a multi-objective Pareto efficiency analysis
- Approximately 100 tiles can effectively reproduce a quasi-fully distributed LSM setup with 1% of the computational cost

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

L. Torres-Rojas,
laura.torres@duke.edu

Citation:

Torres-Rojas, L., Vergopolan, N., Herman, J. D., & Chaney, N. W. (2022). Toward an optimal representation of sub-grid heterogeneity in land surface models. *Water Resources Research*, 58, e2022WR032233. <https://doi.org/10.1029/2022WR032233>

Received 21 FEB 2022
 Accepted 4 DEC 2022

Author Contributions:

Conceptualization: L. Torres-Rojas, N. Vergopolan, N. W. Chaney
Formal analysis: L. Torres-Rojas, N. Vergopolan, J. D. Herman, N. W. Chaney
Funding acquisition: N. W. Chaney
Methodology: L. Torres-Rojas, N. Vergopolan, J. D. Herman, N. W. Chaney
Project Administration: N. W. Chaney
Resources: N. W. Chaney
Software: N. W. Chaney
Supervision: N. W. Chaney
Writing – review & editing: L. Torres-Rojas, N. Vergopolan, J. D. Herman, N. W. Chaney

Abstract One of the persistent challenges of Land Surface Models (LSMs) is to determine a realistic yet efficient sub-grid representation of heterogeneous landscapes. This is particularly important in emulating the fine-scale and nonlinear interactions between water, energy, and biogeochemical fluxes at the land surface. In LSMs, landscape heterogeneity can be represented using sub-grid tiling techniques, which partition macroscale grid cells (e.g., 1°) into smaller units or “tiles.” However, there is currently no formal procedure to define the number of tiles required to adequately represent the heterogeneity of hydrologic processes within a macroscale grid cell, and across spatial scales. To address these challenges, a new approach is presented to diagnose sub-grid process heterogeneity formally and to infer an optimal number of tiles per macroscale grid cell. The approach is demonstrated using the HydroBlocks modeling framework coupled to Noah-MP LSM implemented over a 1.0-degree domain in Western Colorado in the United States. Our results show that (a) a surrogate model can accurately infer the spatial structure of the LSM's time-averaged hydrological fields, with over 95% overall R^2 performance in the validation stage; (b) the optimal configurations for a target level of complexity can be determined using a multi-objective Pareto efficiency analysis, which includes the simultaneous representation of the multi-scale heterogeneity of several processes; (c) the use of ~100 tiles effectively reproduces a quasi-fully distributed LSM setup (i.e., 83,000 tiles) with approximately 1% of the computational expense. This method provides a path forward to efficiently determine the optimal tile configurations for LSMs while simultaneously considering the spatial heterogeneity and spatial accuracy of hydrologic processes.

1. Introduction

Land Surface Models (LSMs) are physically based numerical models that simulate the coupled fluxes of water, energy, and carbon cycles on the land-atmosphere interphase (Fisher & Koven, 2020; Vergopolan et al., 2020). Traditionally, LSMs are used to serve as boundary conditions to large-scale models (e.g., Earth System Models (ESMs)) by representing the influence of the land on meteorological processes in terms of energy partitioning, water fluxes, surface roughness, albedo, among others (Fisher & Koven, 2020). As uncoupled models, LSMs also provide spatially distributed, physically based modeling tools, making them critical scientific assets in predicting hydrological variability at various scales. Thus, LSMs allow addressing societal needs for information about water and energy over continental and global domains (Fang et al., 2015; Koch et al., 2017; Sheffield et al., 2014; Xia et al., 2016).

Although current state-of-the-art LSMs are discretized at scales ranging from 10 to 100 km, many of the underlying processes (e.g., runoff generation) occur at much finer spatial scales (Avisar & Pielke, 1989; Beven, 2010; Blöschl & Sivapalan, 1995; Famiglietti & Wood, 1994; Vergopolan et al., 2022). Assuming homogeneity of these processes within LSM grid cells neglects the nonlinear nature of the system; which can ultimately lead to various estimation issues (e.g., errors in estimation for development of the planetary boundary layer, initiation of shallow and deep convection, and cloud formation and precipitation) (Fisher & Koven, 2020; Simon et al., 2021; Tesfa et al., 2014; Vergopolan et al., 2022). Hence, correctly representing the effects of this physical heterogeneity in the LSM macroscale grid cells is vital to accurately represent weather and climate dynamics, as well as the hydrologic cycle (Chen et al., 2020; Li et al., 2013; Salmun and Molod, 2006).

In recent decades, the land surface modeling community has seen an increase in computational power and availability of spatially distributed environmental data sets at high resolutions. This has continued the early 2010 “hyper-resolution modeling” debate on how to effectively represent sub-grid heterogeneity in LSMs (Beven

& Cloke, 2012; Beven et al., 2014; Bierkens, 2015; Wood et al., 2011). Generally speaking, two options are considered: (a) splitting the landscape (i.e., modeling domain) down to a meter-resolution with the hope that heterogeneity stops being an issue, and (b) aggregating the landscape to some scale at which the effects of the sub-grid heterogeneity are contained (Sivapalan, 2018). Significant uncertainties in model structure, parameters, and forcing make the deterministic meter-resolution fully distributed model unfeasible and impractical in most cases. On the other hand, aggregating the landscape using simplified representations of heterogeneity is a generally more feasible solution. Therefore, there is a persistent need for LSMs to derive effective yet efficient aggregation methods for spatially heterogeneous landscapes. To address this issue, most LSMs use sub-grid “tiling” techniques to divide grid cells into smaller units (i.e., tiles). Within this semi-distributed framework, each tile’s water, energy, and carbon cycles are resolved independently, assuming intra-tile homogeneity (Li et al., 2013). As the number of tiles increases, the average tile-size decreases, and the simulation outputs approach those for a fully distributed configuration (i.e., convergence of the spatial pattern (Chaney, Metcalfe, & Wood, 2016; Chaney et al., 2018, 2021)). The convergence pattern is a function of various factors, including the domain size, the implemented tiling scheme, the analyzed hydrological process, the adopted temporal aggregation, and the metric used to define agreement with the fully distributed simulation; this variety of factors makes the selection of a single optimal tiling configuration a challenging task.

Recognizing the multi-scale nature of spatial heterogeneity in land surface processes, the land modeling community is constantly rethinking how to represent the hierarchical structure of heterogeneity within tiling schemes. Tiling techniques have evolved from user-defined equally sized tiles (Chen et al., 1997) to structural grid partition techniques based on the spatial distribution of vegetation, soil properties (Melton and Arora, 2014), and topography (Hao et al., 2022), to more formal clustering techniques (Chaney, Metcalfe, & Wood, 2016; Chaney et al., 2018, 2021; Newman et al., 2014). Along with these advances, some modern tiling schemes also introduced direct inter-tile interactions at subsurface- and surface-level (Ajami et al., 2016; Chaney, Metcalfe, & Wood, 2016; Chaney et al., 2018, 2021; Clark et al., 2015; Subin et al., 2014; Swenson et al., 2019). Adding these interactions has been shown to improve the modeling of baseflow production, the representation of riparian ecosystems, the accuracy of the energy balance partitioning, and the effectiveness of the sub-grid runoff routing schemes (Chaney et al., 2018, 2021).

One flexible land surface modeling framework accounting for multidimensional clustering, subbasin structure, and inter-tile interactions is HydroBlocks (Chaney, Metcalfe, & Wood, 2016; Chaney et al., 2018, 2021). HydroBlocks uses a Hierarchical Multivariate Clustering (HMC) scheme to determine the tiles’ configuration within the modeling domain coupled to the traditionally used NOAA-MP vertical LSM (Niu et al., 2011). The tiling scheme simultaneously accounts for three levels of heterogeneity, from coarse to fine: (a) large-scale patterns over large extensions of the landscape (i.e., watershed-scale heterogeneity), (b) differences in processes due to elevation (i.e., elevation heterogeneity), and (c) small-scale heterogeneity produced by soil features and land cover (i.e., small-scale heterogeneity). The primary goal of HMC is to provide an approach that robustly accounts for the different sources of multi-scale heterogeneity while still ensuring simplicity and computational efficiency. This tiling scheme structure is flexible enough to account for a single or several proxies of physical heterogeneity, converging to more traditional structural grid partition techniques.

Despite the significant advances in tiling schemes over the last decade, many issues persist. First, defining the tile configuration per macroscale grid cell remains mostly an arbitrary decision. Besides, commonly used tiling schemes are usually over-simplistic, relying solely on one or two proxies of physical heterogeneity to determine the tiling structure and disregarding the computational expenses derived from redundant tiling. To the best of our knowledge, there has not been any attempt to objectively determine the optimal resolution of tiles per modeling domain for a given tiling scheme. Second, over large-scale domains, LSM sub-grid outputs are mostly only summarized and evaluated via macroscale grid statistics: spatial mean and variance. Equations 1 and 2 present the equations used to compute these statistics for any domain containing gridded information:

$$\mu = \frac{\sum_{i=0}^N n_i}{N} \quad (1)$$

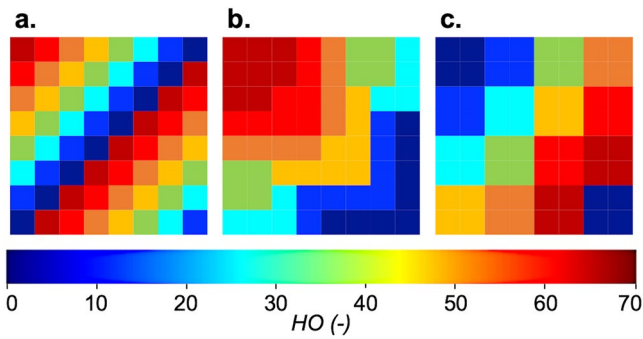


Figure 1. Three simplified land surface representations (a, b, and c) with the same macroscale spatial mean and spatial variance but different spatial patterns. The fields correspond to any hydrological output (HO).

$$\sigma = \sqrt{\frac{\sum_{i=0}^N (n_i - \mu)^2}{N}} \quad (2)$$

where μ represents the spatial mean, σ the spatial standard deviation, N the number of pixels or grid cells within the domain, and n_i the value of each pixel. Although informative, these statistics are insensitive to the tiles' large-scale spatial patterns (i.e., pattern-agnostic metrics) (Jupp and Twiss, 2006). Figure 1 shows three simplified land surface representations with identical macroscale grid statistics but different spatial patterns. By using spatial metrics that are insensitive to spatial structure (e.g., spatial mean and variance), the three cases are indistinguishable from each other. This issue is especially important as emerging work shows the importance of correctly representing the sub-grid spatial coherence to explain the role of sub-grid heterogeneity on atmospheric response (Simon et al., 2021).

Finally, as mentioned previously, the selection of a single optimal tiling configuration is a challenging task. The problem becomes more complex as the multi-scale heterogeneity of the physical environment can, at times, have different importance for different processes. In other words, a tiling structure that accurately represents one process may not be the most appropriate for a different one. Besides, the number of tiles must also be considered as it is a decisive factor in ensuring computational tractability. Therefore, to achieve accurate representation of various processes as well as computational efficiency, informed decisions about the heterogeneity representation in the models must be taken in the presence of at least two conflicting objectives.

This study presents a path forward in determining the optimal tiling configurations in LSMs using the HMC tiling framework. To this aim, we introduce a metric to assess modeled sub-grid heterogeneity along with a multi-objective approach to optimize computational expenses based on the selection of the appropriate tiling configuration. These developments are implemented and tested over a 1.0° bounding box centered at the East River Watershed Function Scientific Focus Area in Western Colorado in the United States. The high landscape heterogeneity of this domain in terms of its topographic, vegetative, and climatic features provides an ideal case for quantifying optimal sub-grid heterogeneity. As such, the key developments in this paper include (a) a pattern-aware metric to measure hydrologic heterogeneity of LSM-derived fields; (b) a sensitivity analysis of the parameters of the LSM tiling scheme to relate the convergent behavior to the source of heterogeneity driving it, and (c) a multi-objective optimization approach to determine the tile configuration that for the desired tolerance and particular set of outputs, more closely approximates a quasi-fully distributed configuration. The methodology introduced here provides a pathway forward to diagnose sub-grid process heterogeneity formally and to infer an optimal number of tiles per macroscale grid cell.

2. Data and Methods

2.1. Study Domain

A 1.0° box in central-western Colorado in the United States is used in this study (Figure 2). The domain falls within the Upper Colorado River Basin and includes catchments such as the East River, Washington Gulch, Slate River, and Coal Creek. The bounding box is part of the Elk Range in the Rocky Mountains. The domain is selected due to the observable role of the physical environment in the sub-grid heterogeneity. With elevations ranging between 2,000 and 4,000 m (Figure 2b), the area has primarily sandy soils with some clay deposits on the southwest hills (Figure 2c). The major land-cover types include deciduous forest (27%), evergreen forest (26%), shrubs and scrubs (23%), and grassland and herbaceous (13%) (Figure 2d). The region exhibits strong North-South gradients in precipitation and temperature with precipitation increasing toward the north (and vice versa with temperature).

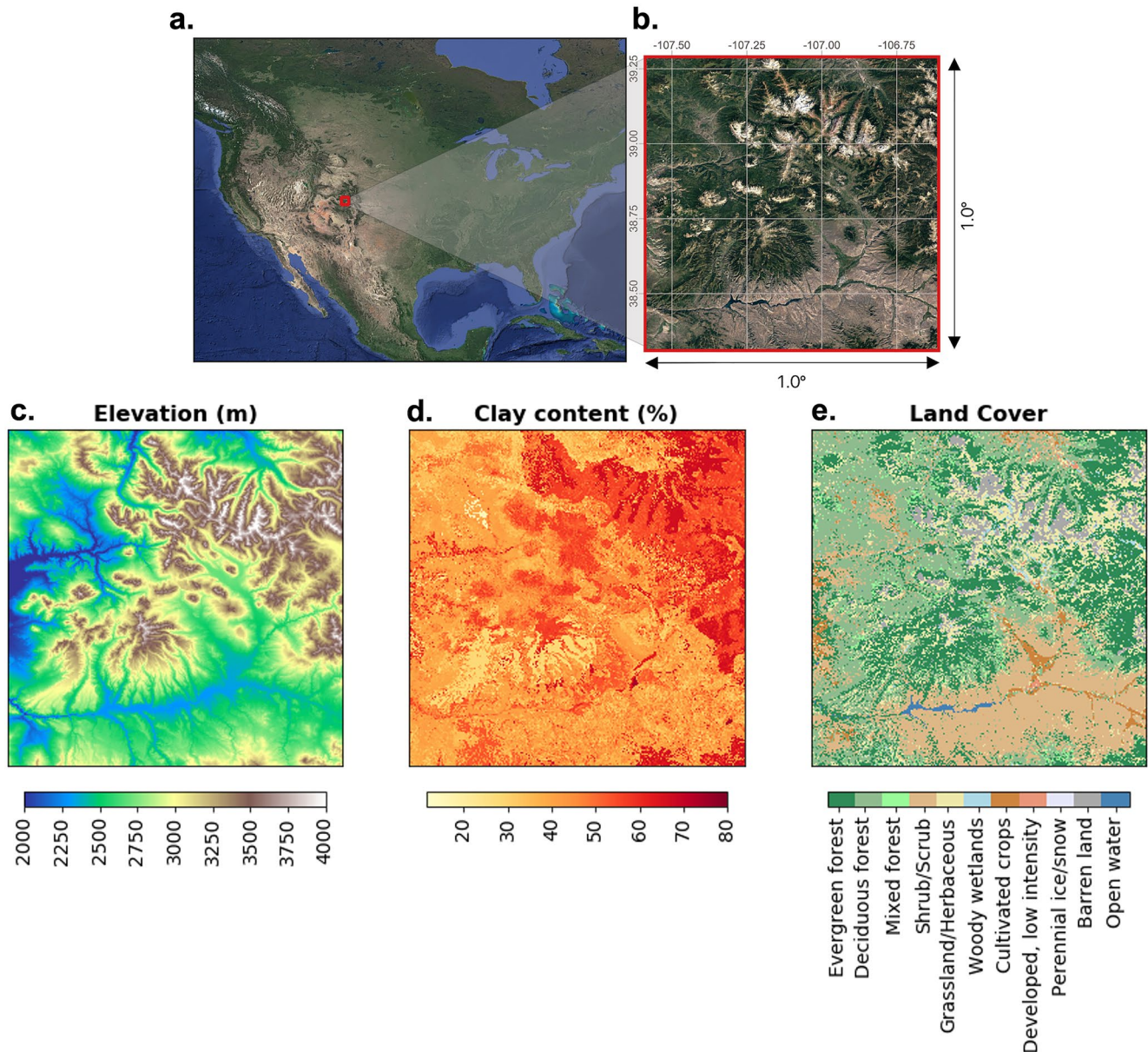


Figure 2. The study area consists of a 1.0° box in central-western Colorado. The domain contains the East River Watershed Function Scientific Focus Area; (a) location within CONUS, (b) zoomed-in satellite visible imagery and coordinates, (c) USGS 3DEP elevation, (d) POLARIS clay content, and (e) NLCD land cover classification (Chaney et al., 2019; Gesch et al., 2010; Homer et al., 2015).

2.2. Modeling Framework: HydroBlocks

HydroBlocks is a modeling framework that aims to resolve the surface water and energy cycles at field scales (Chaney, Metcalfe, & Wood, 2016). The main feature of HydroBlocks is the use of the HMC tiling scheme (explained in detail in Section 2.3). By applying HMC, HydroBlocks leverages the repeating patterns over the landscape (i.e., the spatial organization) by clustering areas of assumed similar hydrologic behavior into hydrologic response units (HRUs) or “tiles” (Vergopolan et al., 2020). Note that in HydroBlocks, the regions of the landscape that belong to a given HRU or tile can be spatially separated, a departure from the original concept of an HRU (Flügel, 1995). Although the terms “HRU” and “tile” in HydroBlocks are interchangeable, this study will use “tile” for consistency.

In HydroBlocks, the macroscale grid cell is partitioned into multiple tiles using HMC, and the NOAA-MP vertical LSM (Niu et al., 2011) is then used to resolve the vertical land surface processes at each time step. The version

Table 1
Description of the Selected Schemes Within NOAH-MP LSM

Parametrization	Selected scheme
Dynamic vegetation	Off. It uses the table-based leaf area index
Canopy stomatal resistance	It parameterizes stomatal conductance as a function of soil moisture, atmospheric temperature, radiation availability, and vapor pressure deficit (Jarvis, 1976).
Soil moisture factor for stomatal resistance	Noah type: The soil moisture factor controlling stomatal resistance is parameterized as a function of soil moisture (Chen & Dudhia, 2001).
Runoff and groundwater	TOPMODEL-based scheme with equilibrium water table (Niu et al., 2005).
Surface layer drag coefficient	Monin-Obukhov-based (Brutsaert, 1982)
Supercooled liquid water in frozen soil	Variant of the freezing-point depression equation (Koren et al., 1999)
Frozen soil permeability	Inherits Koren99 scheme in Noah V3 (soil ice has a nonlinear effect on soil permeability)
Radiation transfer	Modified two-stream (Niu and Yang, 2004; Yang and Friedl, 2003)
Ground snow surface albedo	CLASS (Verseghy, 1991)
Partitioning precipitation into rainfall & snowfall	Precipitation as snowfall when $T_{air} < T_{frc}$ and rainfall otherwise
Snow/soil temperature time scheme (layer 1)	Fully implicit, as defined in the original NOAH model
Glacier treatment	Include phase change of ice
Surface resistance to evaporation/sublimation	Represents the effect of plant litter cover on water vapor transfer and considers the effect of under canopy atmospheric stability on the under canopy turbulent resistance (Sakaguchi and Zeng, 2009)
Crop model option	None
Urban physics option	Off

of NOAH-MP currently implemented in HydroBlocks is the one included within version 3.7 of the NCAR's High-Resolution Land Data Assimilation System (Chen et al., 2007). The selected schemes for the NOAH-MP LSM in this study are presented in Table 1. For further details, the reader is referred to Niu et al. (2011) and Li et al. (2022).

For each tile, the LSM parameters are computed as the mean of the parameters of all the grid cells that belong to the tile. The meteorological forcing is also aggregated from the individual values of grid cells without downscaling. For each time step, the height bands, as defined by the tiling scheme, dynamically interact laterally via subsurface flow computed as a Darcy flux between adjacent bands at each subsurface level (Chaney et al., 2018). These terms are included as divergences within the one-dimensional solution of the Richards' equation in the Noah-MP implementation for each tile belonging to the height band (Chaney et al., 2021). In this study, HydroBlocks is run at an effective 30-m spatial resolution and 1-hr temporal resolution from 2012 to 2018; the first 2 years are used as a model spin-up period. HydroBlocks is used to simulate time averaged soil moisture content (SMC), sensible heat flux (SH), latent heat flux (LH), and runoff (RO).

2.3. Hierarchical Multivariate Clustering (HMC) Tiling Scheme

HydroBlocks uses a HMC scheme to determine the tiles' configuration within the modeling domain, as displayed in Figure 3 (Chaney et al., 2018, 2021). The scheme starts by delineating the river network and watersheds from the entire area's sink-filled 30-m Digital Elevation Model (DEM). The domain is then partitioned into smaller k subdomains (or clusters of watersheds) determined by using K-means over assumed proxies of large-scale physical heterogeneity. The proxies used in the clustering for this study include average values for the geographical location (latitude and longitude) and elevation. Next, the height above the nearest drainage (HAND) values for each cluster of watersheds are discretized into height bands given n , which indicates how many times larger the areal coverage of a height band is compared to its adjacent lower one. This stage attempts to capture the subbasin structure by differentiating areas in high and lowlands. Finally, to represent fine-scale heterogeneity of land cover and soils, each height band is split into multiple clusters using K-means. The third HMC parameter, p , describes the average number of clusters per height band.

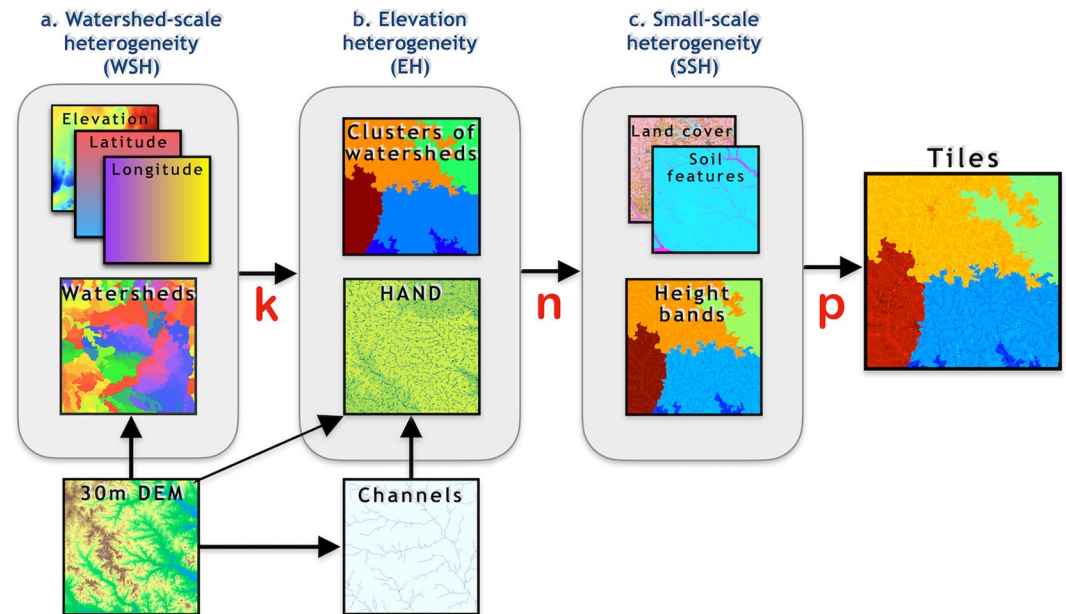


Figure 3. Process used to define the tile structure using the Hierarchical Multivariate Clustering scheme. The three main scales of heterogeneity represented are highlighted: (a) watershed-scale heterogeneity, (b) elevation heterogeneity, and (c) small-scale heterogeneity.

To summarize, the first stage (associated with k) aims to account for large-scale patterns (i.e., watershed-scale heterogeneity (WSH)) that can occur over the landscape and their implications on the land surface response (e.g., large-scale differences in climatology and energy balance partitioning between watersheds). The second stage, (related to n) attempts to capture the dynamics in riparian zones and, therefore, better reproduce the connected processes (e.g., runoff) and explicitly account for elevation-related processes (i.e., elevation heterogeneity (EH)). Finally, the last stage (associated with p) represents the intra-band heterogeneity in soil and land cover, leading to localized and small-scale differences in hydrologic outputs such as soil moisture (i.e., small-scale heterogeneity (SSH)). For a comprehensive description of the most recent HMC implementation within HydroBlocks, the reader is referred to (Chaney et al., 2021).

2.4. High-Resolution Land Surface Data Sets

In this study, a stack of high-resolution open-access environmental data sets for topography, land cover, soil features, and meteorology are used to drive HydroBlocks. The specifications of the data sets used in this study are presented in Table 2. The products include the one arcsec (~ 30 m) USGS National elevation data set (NED), the one arcsec (~ 30 m) National Land Cover Database (NLCD), the Probabilistic Remapping of SSURGO (POLARIS) (~ 30 m), and the $1/32^\circ$ (~ 3 km) Princeton CONUS Forcing (PCF) data set that provides meteorological forcing at 1-hr temporal resolution (Chaney et al., 2019; Gesch et al., 2010; Homer et al., 2015; Pan et al., 2016).

2.5. Empirical Spatial Covariance Function (ESCF)

To assess the simulated sub-grid hydrologic heterogeneity, the empirical spatial covariance function (ESCF) was implemented as a pattern-aware metric. The ESCF expresses how the linear statistical dependence of two measurements in space reduces as the distance between them increases, up to the distance of statistical independence, that is, correlation length, where a relation no longer exists (Mälicke et al., 2020). Using the ESCF to quantify the spatial heterogeneity of the fields accounts not only for the spatial variance but also for the spatial autocorrelation of the hydrologic processes.

In this study, the ESCF is calculated on the HydroBlocks simulated annual mean spatial fields of SMC, SH, LH, and RO, as displayed in Figure 4 for any hydrological variable, HO. First, the annual mean tile-level simulation of each variable is mapped out on to the 30-m tile map (computed via HMC); this enables a 30-m fully distributed

Table 2
Specifications of Environmental Data Sets Used in the Study

Group	Parameter	Data source	Period	Spatial resolution	Temporal resolution	References ¹
<i>Meteorology</i>	Precipitation	PCF	2012–2018	~3 km	Hourly	Pan et al., 2016
	Short wave radiation					
	Longwave radiation					
	2 m air temperature					
	Specific humidity					
	Surface pressure					
	Wind speed					
<i>Land cover</i>	Land cover classification	NLCD	2016	30 m	Static	Homer et al., 2015
<i>Soil properties</i>	Soil texture	POLARIS	–	30 m	Static	Chaney et al., 2019
	Organic matter content					
	Saturated hydraulic conductivity					
	Brooks-Corey water retention curve parameters					
	Saturated water content					
<i>Topography</i>	Elevation	NED	–	30 m	Static	Gesch et al., 2010

¹All the data sets are open access and available at the provided references.

representation of the simulated variables (Figure 4a). A sample of 30,000 fixed spatial points (Figure 4b) is then used to compute the covariance cloud (Figure 4c) and sample covariance for each scale bin (i.e., distance bin) on the annual mean fields of each variable. The number of points is selected by considering the available memory to store the resulting arrays (900,000,000 of elements). The covariance cloud is computed by iterating through all the positions in the sample points set, calculating the pairwise distances between them, and computing the product of the anomalies from the spatial mean between the 30-m annual mean value for two points, i and j , as indicated in Figure 4b. The covariance cloud is then binned based on distances, and the average covariance per bin is approximated as the mean value of all the anomalies belonging to the bin.

2.6. Random Forest Model for Predicting ESCFs

To understand how the different sources of heterogeneity represented by the HMC parameters influence the spatial organization features of the selected hydrological processes, a large sample of HydroBlocks simulations using a wide range of HMC parameter configurations is required. Even with HydroBlocks' computational advantages, this remains computationally expensive due to the need to sample a very large number of tiling configurations (e.g., 100,000 simulations). To overcome this limitation, a random forest model is trained to predict the resulting HydroBlocks ESCF for the mean annual fields of the selected hydrological variables. More specifically, a multi-output Random Forest (RF) machine learning approach is trained on a set of HydroBlocks simulated ESCF for the annual means of SMC, SH, LH, and RO. Each HydroBlocks simulated ESCF comes from a random choice of HMC parameters.

The HydroBlocks simulations used to train the RF model are determined from a convergence analysis of the model performance. The procedure starts with a predefined set of 100,000 configurations. The first 100 configurations are used to run HydroBlocks, derive the ESCFs, and train and test the model (70% of the HydroBlocks simulations for training and 30% for testing). The coefficient of determination (R^2) is computed between the HydroBlocks derived values and those generated by the RF model. More HydroBlocks simulations are successively added to the RF model's training and validation stages, and its performance is compared to the one from the previously fitted model. It is assumed the RF converges once the relative difference (RD) between successive R^2 is below 2% for all the hydrological variables and for the training and validation stages. For a detailed description of the RF model's convergence analysis, the reader is referred to Supporting Information S1 (Text S1).

Once trained, the RF model is used to assemble the ESCFs for 100,000 synthetic HMC configurations. The random sampling procedure for the HMC configurations for this stage is performed in logarithmic space. The

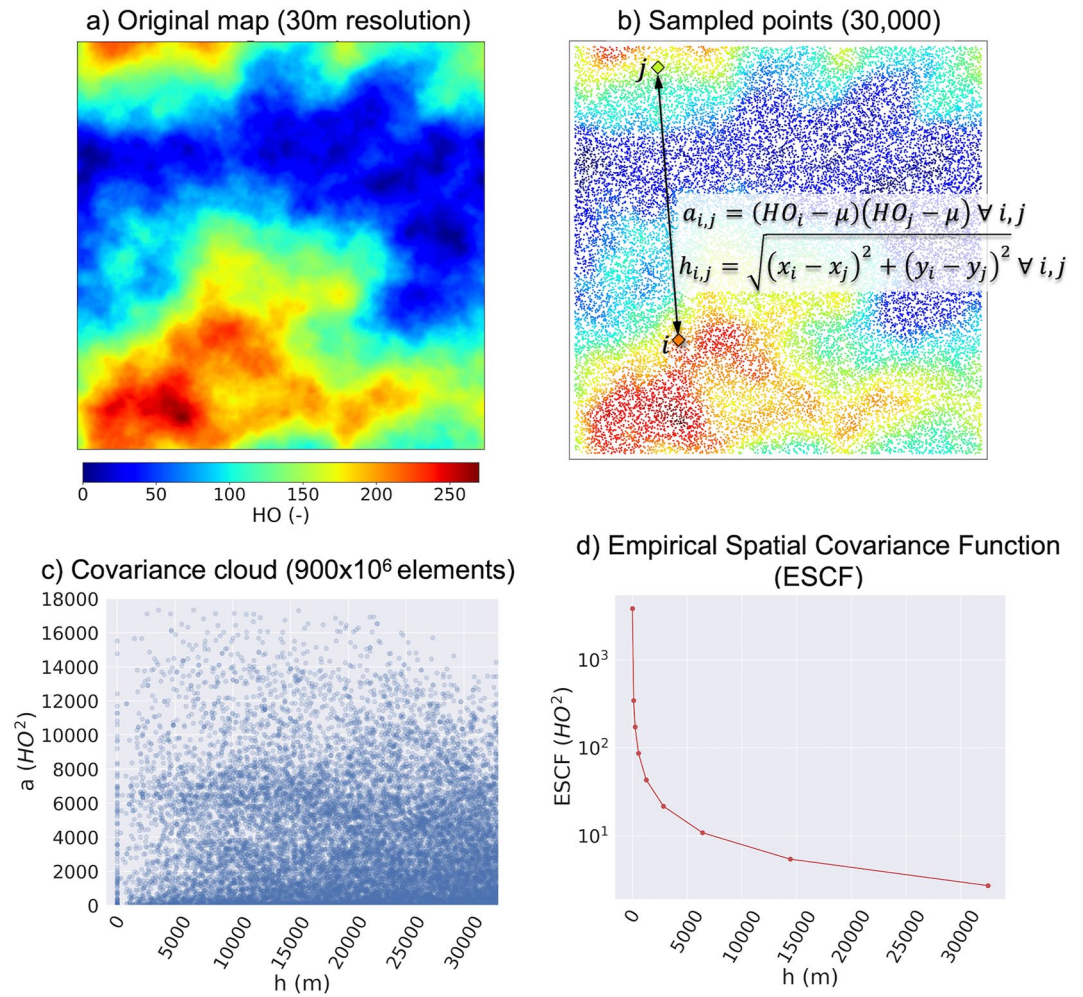


Figure 4. Process used to compute the empirical spatial covariance function over any gridded field of a hydrological variable HO. (a) 30-m original map of the variable; (b) a sample of 30,000 fixed spatial points are extracted from the original field and the anomalies and distances between them are computed; (c) the covariance cloud is built by plotting the anomalies versus the distances for all the points; (d) the covariance cloud is binned based on distances. Finally, the average behavior for each bin is approximated as the mean value of all the anomalies that belong to it.

selection of logarithmic sampling in this stage responds to the need for a more robust sample for a lower number of tiles since it has been proved by previous studies (Chaney et al., 2018, 2021) that the output states and fluxes are more sensitive to the HMC selected parameters in the lower tile range. The lower and upper limits used for sampling each parameter are presented in Table 3. These simulations are then analyzed to determine the convergence patterns of HydroBlocks over the study domain.

2.7. Approximating a Quasi-Fully Distributed Configuration: Spatial Standard Deviation Versus the ESCF-KGE Metric

A quasi-fully distributed HMC configuration (QFD), simulated with HydroBlocks, is used as the benchmark. This benchmark allows for a straightforward assessment of the synthetic configurations' performance in representing the spatial heterogeneity of a fully distributed solution. By comparing the synthetic configurations against this benchmark, the goal is to find the optimal (i.e., minimal number of tiles) HMC configuration that mimics the hydrologic spatial representation of the QFD simulation. In other words, the QFD simulation is an upper limit for computational burden. By

Table 3
Sampling Space for 100,000 Hierarchical Multivariate Clustering (HMC) Synthetic HMC Configurations Evaluated Using the RF Model

Parameter	Lower limit	Upper limit
k	1	100
p	1	50
n	1.1	101

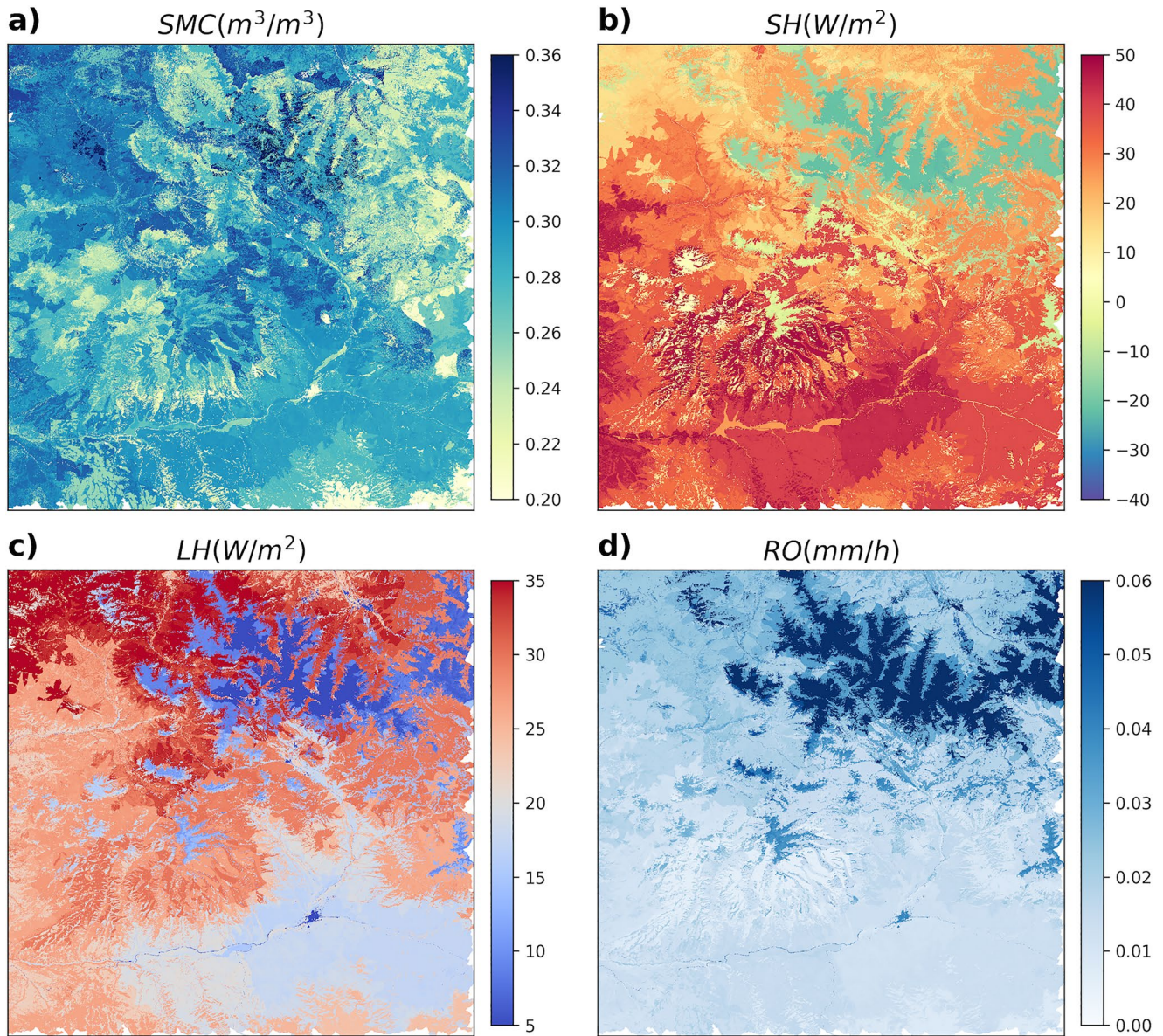


Figure 5. HydroBlocks modeling system simulations using the quasi-fully distributed configuration. Each panel shows the long-term temporal mean of (a) soil moisture content, (b) sensible heat flux with a positive sign indicating upward direction of the flow and negative, downward direction, and (c) latent heat flux, and (d) runoff.

comparing the synthetic configurations to it, the expectation is to find the configuration that approximately mimics the fully distributed behavior while minimizing the required number of tiles.

Ideally, the benchmark should correspond to a 30-m fully distributed HydroBlocks simulation containing $\sim 18,000,000$ of tiles. However, the computational resources for this task are limited. The size of the study domain, as well as the simulation period and time step, increase computational requirements to the extent that cannot be met by the available infrastructure. Therefore, the configuration with the highest order (i.e., number of tiles) that can be handled with the available computational resources (i.e., QFD) is selected as the benchmark.

In this study, the QFD contains 83,000 tiles, achieved by setting the HMC parameters to $k = 60$, $n = 1.2$, and $p = 60$. Figure 5 shows the temporal mean of SMC, SH, LH, and RO between 2014 and 2018 of the QFD configuration.

Chaney, Metcalfe, and Wood (2016); Chaney et al. (2018) evaluated the simulated sub-grid representation of the land surface states and fluxes of HydroBlocks via the macroscale grid spatial variance and mean (i.e., pattern-agnostic metric). However, to determine the advantages of a pattern-aware metric over a pattern-agnostic one, two approaches are compared. For the baseline case, the spatial variances of the synthetic configurations are compared to the same quantity for the QFD simulation. For the pattern-aware method, the Kling–Gupta efficiency metric (KGE; Gupta et al., 2009; Kling et al., 2012) is computed between the ESCFs for the QFD configuration and the ESCFs for each synthetic configuration. This metric is given the name ESCF-KGE. The proposed procedure allows comprehensively assessing the representation's performance on several scales.

KGE is selected as a performance indicator as it accounts for several levels of agreement between values by combining terms for linear Pearson correlation (ρ), standard deviation bias ($\alpha = \frac{\sigma}{\sigma_{QFD}}$), and mean bias ($\beta = \frac{\mu}{\mu_{QFD}}$), as displayed in Equation 3. An optimal fit occurs for KGE equal to 1, while negative values represent a very poor fit.

$$KGE = 1 - \sqrt{(\rho - 1)^2 + (\beta - 1)^2 + (\alpha - 1)^2} \quad (3)$$

2.8. Sobol Sensitivity Analysis

The hydrologic heterogeneity simulated by HydroBlocks is largely impacted by the chosen HMC parameter values. However, there is limited knowledge on how these parameters, and the sources of heterogeneity they represent influence the hydrologic convergence. A Sobol sensitivity analysis is thus performed to understand the sensitivity of convergence to the heterogeneity types and to discern their role in the quality of hydrologic heterogeneity in comparison to the QFD configuration. More specifically, we evaluate this sensitivity for the case of the spatial variance and the ESCF-KGE method.

The Sobol sensitivity analysis (Sobol, 1993) is a global method that decomposes the variance of the model output, Y , into contributions from each parameter, X_i , and its interactions with other parameters: Chaney et al., 2016.

$$\text{Var}(Y) = \sum_i V_i + \sum_{i < j} V_{ij} + \sum_{i < j < k} V_{ijk} + V_{12, \dots, n} \quad (4)$$

where V_i represents the first-order variance contribution of the parameter X_i , V_{ij} is the interaction between parameters X_i and X_j and $V_{12, \dots, n}$ includes all the interactions higher than third order. It is possible to define the first order (S_i) and total-order sensitivity (S_{Ti}) indices as:

$$S_i = \frac{\text{Var}[E(Y|X_i)]}{\text{Var}(Y)} = \frac{V_i}{\text{Var}(Y)} \quad (5)$$

$$S_{Ti} = \frac{\text{Var}[E(Y|X_{\sim i})]}{\text{Var}(Y)} = \frac{V_{\sim i}}{\text{Var}(Y)} \quad (6)$$

S_i represents the expected reduction in variance if the parameter X_i was fixed without accounting for interactions with other parameters. On the other hand, S_{Ti} represents the variance reduction ($V_{\sim i}$) associated with all the parameters being fixed except for X_i , accounting for all the interactions with other parameters. The implementation of the Sobol sensitivity analysis used in this study comes from SALib (Herman & Usher, 2017).

2.9. Multi-Objective Optimization of the Tiling Configuration

One of the main goals of this study is to optimize the heterogeneous representation of multiple hydrologic states and fluxes simultaneously while minimizing the number of required tiles. This goal implies a multi-objective optimization approach as optimal decisions must be taken in the presence of two or more conflicting objectives. In general, there is no unique solution that simultaneously optimizes all objectives. Therefore, the purpose of the multi-objective optimization problem is to find the best trade-off solutions among all the conflicting objectives (i.e., Pareto optimal solutions or non-dominated solutions) (Deshpande et al., 2013). The Pareto optimal solutions set is known as the Pareto front. In this study, a multi-objective approach is needed to approximate the Pareto front of optimal configurations from the set of synthetic configurations at the site. This Pareto front simultaneously represents the independent spatial quality of several hydrological processes when optimizing the number

of tiles (i.e., single-process Pareto Front analysis). In this study, the obtained Pareto fronts can be considered near-optimal because the sampling provides good coverage of the three-parameter HMC space.

As a complement to the single-process Pareto front analysis, the multiple objectives for the hydrological variables are summarized into a single value using a weighted convergence metric (m). To this end, a re-scaled form of the ESCF-KGE ($rsESCF_KGE$) is used as a performance metric for each hydrological variable. The $rsESCF_KGE$ is calculated by comparing the synthetic HMC configurations to the QFD one. For each hydrological variable hv ($rsESCF_KGE_{hv}$) the $rsESCF_KGE$ is computed as shown below:

$$rsESCF_KGE_{hv} = 1 - \frac{1 - ESCF_KGE}{\max(\{1 - ESCF_KGE\}_{hv})} \quad (7)$$

where $\{1 - ESCF_KGE\}_{hv}$ represents the set of all the analyzed configurations. The scaling makes the values to range from zero (poor performance) to one (optimal performance). Once the $rsESCF_KGE$ is computed for all the hydrological variables, the combined metric (m) for each synthetic configuration is computed as:

$$m = \sum_{hv=1}^{HV} w_{hv} rsESCF_KGE_{hv} \quad (8)$$

where HV represents the total number of variables to be simultaneously optimized and w_{hv} the weight assigned to it (ranging between zero for no influence and one to exclusive influence). The combined metric takes values between one for a perfect representation of heterogeneity of all the processes to zero for an incomplete representation of it. For this study, all the analyzed hydrological variables are given equal weight.

2.10. A Path Forward in the Determination of Optimal Tiling Configurations in LSMs

The approach introduced in this study consist of two main interacting stages, as displayed in Figure 6: (a) surrogate model training, and (b) HMC convergence and sensitivity analyses.

2.10.1. Surrogate Model Training

This first stage outputs the surrogate random forest model predicting the ESCF (see Section 2.5) of the mean annual fields of SMC, SH, LH, and RO based on user-input HMC tiling parameters (described in Section 2.6 and in more detail in Text S1 in Supporting Information S1).

2.10.2. HMC Convergence and Sensitivity Analyses

Stage b), presented in Sections 2.8 and 2.9, involves: (a) the use of the generated surrogate model to evaluate the convergent behavior of the four output variables as they approach a quasi-fully distributed HydroBlocks simulation (see Section 2.7); (b) the derivation of a Pareto front for the convergence patterns identified in (a); (c) a Sobol sensitivity analysis to relate convergence to the source of heterogeneity driving it (see Section 2.8); and (d) a multi-objective optimization approach to determine the tile configuration that more closely approximates the quasi-fully distributed configuration for the four selected HOs while minimizing the required number of tiles (see Section 2.9). All the processes in Stage b) are performed for pattern-agnostic and pattern-aware heterogeneity metrics, and the two approaches are extensively compared. The comparison encompasses one to one relations of values between metrics for the four selected hydrological variables and the obtained Pareto fronts for each case.

3. Results

3.1. Performance of the Surrogate Model

The ESCF-RF model was trained using 490 HydroBlocks simulations and tested using 210. The one-to-one comparison of the ESCFs obtained using the surrogate model (ESCF-RF) and the real simulated HydroBlocks fields is shown in Figure 7. The model performed well, with minimal accuracy difference between the hydrological variables. Text S1 in Supporting Information S1 presents more details of the convergence analysis performed to determine the appropriate sampling for the ESCF-RF model training and a comprehensive evaluation of its performance. This analysis showed that training the RF model with 490 different HydroBlocks simulations and their associated HMC configurations was sufficient. As described in Section 2.6 and Figure 6, once trained, this

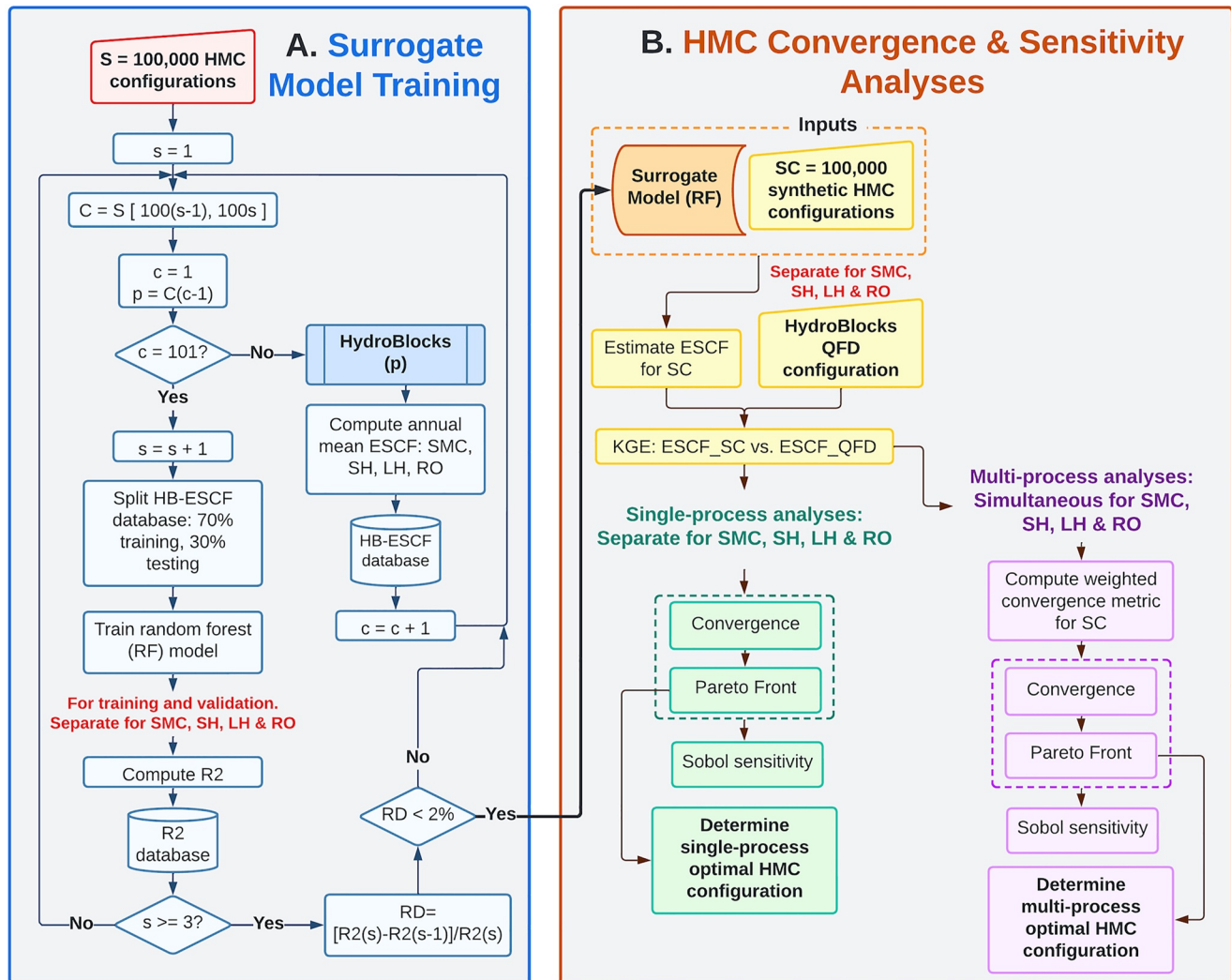


Figure 6. Flowchart summarizing the method for determining the optimal tile configuration in Hierarchical Multivariate Clustering (HMC). The approach relies in two main stages: (a) surrogate model training and (b) HMC convergence and sensitivity analysis.

ESCF-RF model was used to assess a set of 100,000 synthetic HMC configurations (i.e., the combination of HMC parameters).

3.2. Moving Towards a Pattern-Aware Metric of Agreement Between Heterogeneity Measures: ESCF-KGE

First, the relation between the spatial variance and the ESCF-KGE metric was investigated. To demonstrate the disagreement between the two approaches when approximating the QFD configuration, Figure 8 shows the one-to-one comparison of the ESCF-KGE values and the difference between the spatial variance for each experiment and the QFD solution ($\Delta\sigma^2$). The “converged” solutions are in the upper left quadrant of each plot, where the ESCF-KGE values are close to one, and the differences between the spatial variances to the QFD values are minimized. If using spatial variance as a measure of heterogeneity is equivalent to the ESCF-KGE method, a linear behavior for the entire domain should be observed. In fact, some linear relationship holds on the right side of the plots (i.e., configurations with a low number of tiles). However, as the number of tiles increases and the experiments move toward convergence, the relationship becomes noisier, implying that the ESCF-KGE and the spatial variance cease to be equivalent. This implies that pattern-agnostic and pattern-aware metrics were

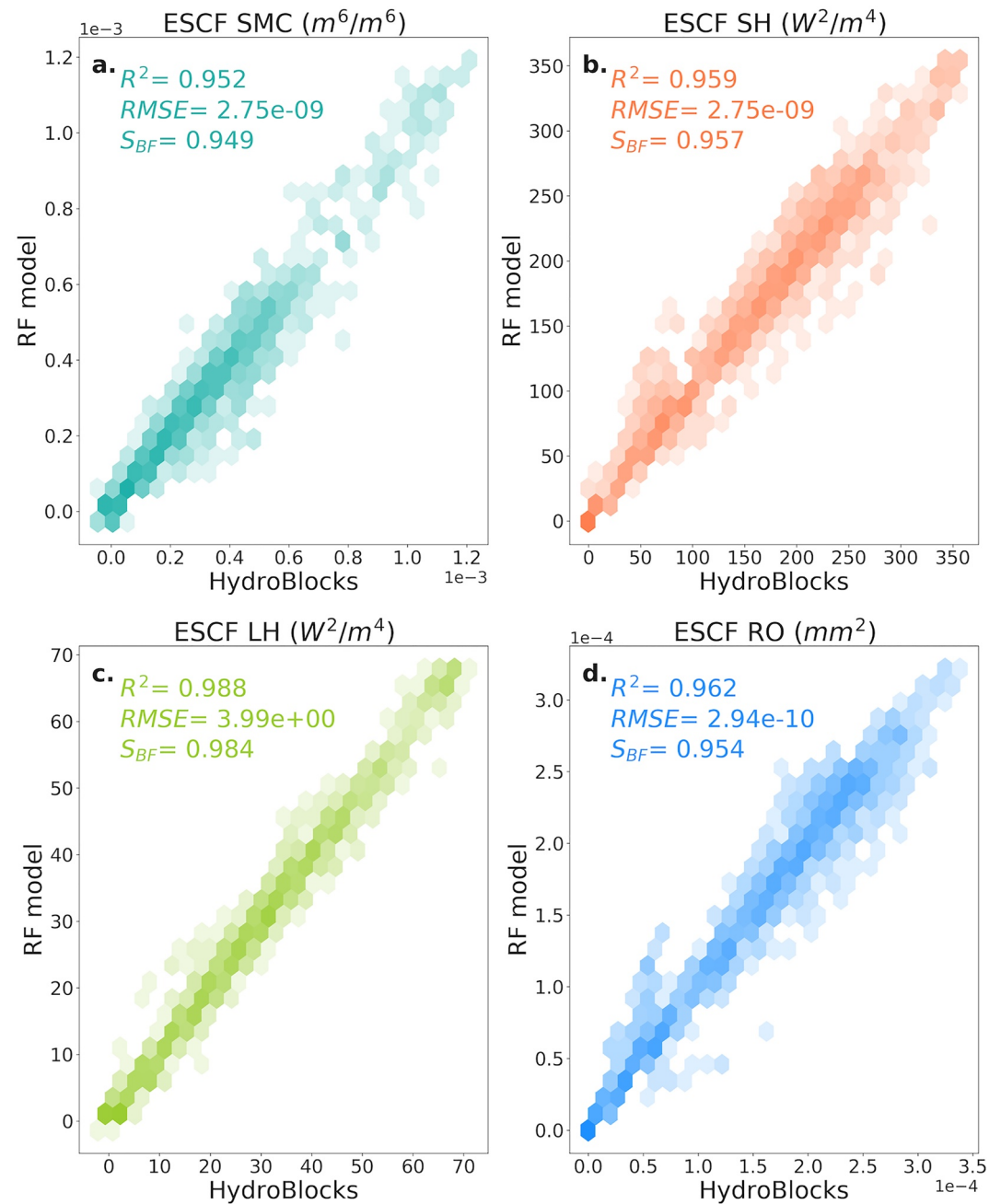


Figure 7. Validation of the empirical spatial covariance functions of the HydroBlocks simulations and its emulator model (random forest regressor) for (a) soil moisture, (b) sensible heat, (c) latent heat, and (d) runoff. R^2 score (R^2), root mean square error, and the slope of the best fitting line are presented as statistical metrics of the quality of the model.

comparable in the early stages of the convergence path (i.e., configurations far away from a converged value). Nevertheless, they diverged as the configurations approach the convergence value.

Another observation derived from Figure 8 is the process specificity of the relationships. For SMC, the relationship between spatial variance and ESCF-KGE was essentially linear. However, the relationships broke down for the other variables (SH, LH, RO), especially for configurations with a higher number of tiles. This behavior is most likely related to the long-term spatial patterns of the analyzed variables over the domain. The nonlinear behavior might be related to the more decisive influence of finer scales of heterogeneity on the obtained values of the ESCF-KGE metric closer to the QFD configuration compared to the convergence pattern of the variance.

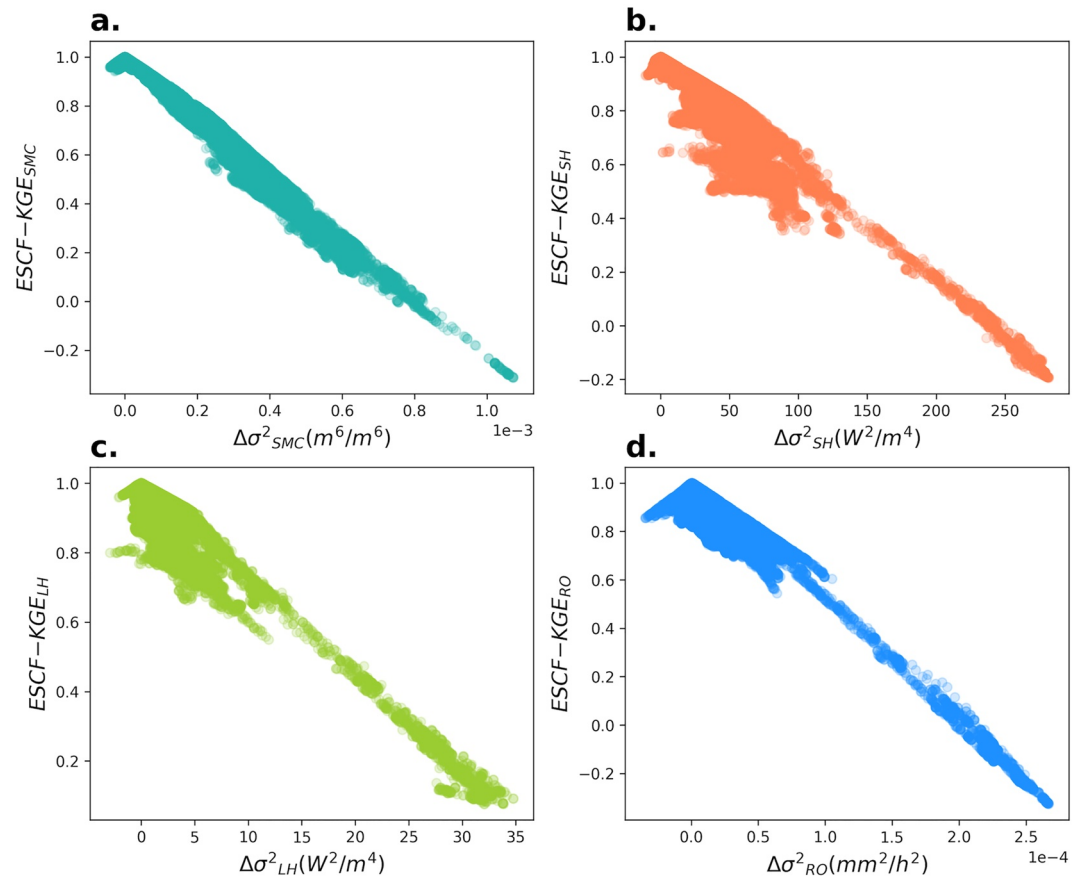


Figure 8. One to one comparison between using empirical spatial covariance function-Kling-Gupta efficiency metric and spatial variance to approximate the quasi-fully distributed solution. The comparison is performed for (a) soil moisture, (b) sensible heat, (c) latent heat, and (d) runoff.

To further illustrate the differences between the two approaches, Figure 9 presents the relationship between the spatial variance of SMC for the 100,000 synthetic configurations versus the number of tiles associated with each (Figure 9a). Ten configurations were selected for detailed analysis (b–k points in Figure 9a). Panels b–k display the ESCFs of the chosen configurations (colored lines) compared to the ESCF for the QFD (dotted gray line). The colors in the figure indicate the ESCF-KGE value for each configuration. Results illustrate how better performance was generally obtained as the number of tiles increases toward the QFD configuration for both spatial variance and ESCF-KGE. However, the figure also highlights the differences between the two metrics when comparing specific points, specifically Figure 9e versus Figures 9f and Figure 9g versus Figure 9h. For these configurations, higher spatial variances, and smaller differences from the QFD variance, did not translate into ESCF-KGE values closer to one. This behavior was consistent with Figure 8 results and highlights the potential differences that can arise in the optimization results due to the selected convergence metric. These differences are further analyzed in Figure 11.

To determine the most influential sources of heterogeneity and understand their role in the convergence of each variable, the steepest part of the Pareto front for the ESCF-KGE metric (i.e., the largest increase in ESCF-KGE per increase in tiles) was analyzed. The same analysis for the spatial variance case can be found in Text S2 in Supporting Information S1. As indicated before, the sources of heterogeneity considered by HMC in this study included: watershed-scale heterogeneity (WSH, associated with k), elevation-related heterogeneity (EH, associated with n), and small-scale heterogeneity (SSH, associated with p). The Pareto fronts for each hydro-

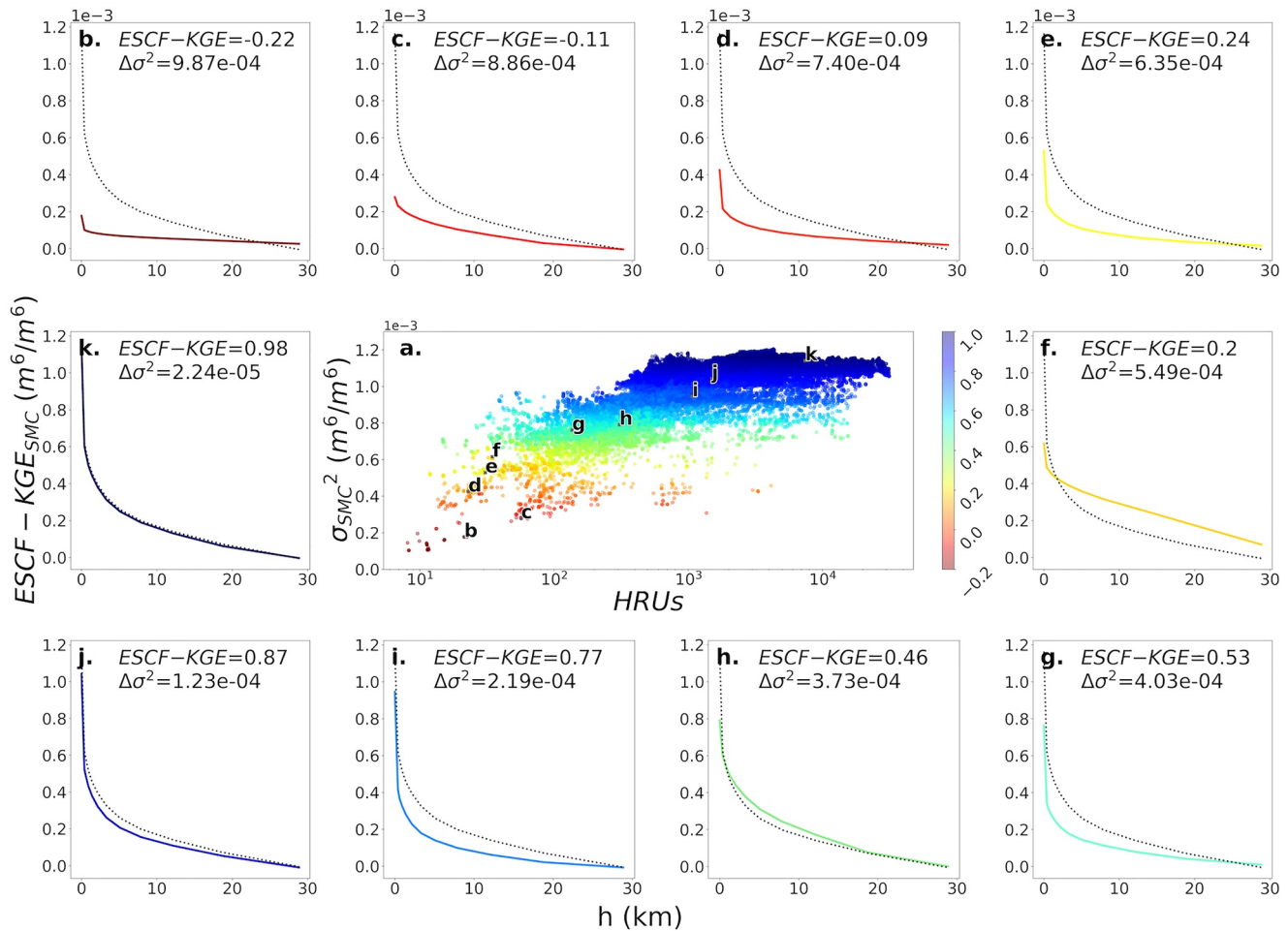


Figure 9. The central panel (a) shows the convergence for the spatial variance of soil moisture content of the 100,000 synthetic Hierarchical Multivariate Clustering (HMC) configurations. The colors indicate the empirical spatial covariance function-Kling-Gupta efficiency metric (ESCF-KGE) values obtained from comparing the ESCF for each configuration and the ESCF for the quasi-fully distributed (QFD) solution. Ten different HMC configurations (b–k) are selected for further illustration. The external panels (b–k) compare the ESCFs between the selected configurations (colored lines) and the ESCF for the QFD (gray dotted lines). The x-axis for the outer panels is simply the distance associated with each ESCF value. For each specific case, the ESCF-KGE value and the difference between the spatial variance for each configuration and the QFD one ($\Delta\sigma^2$) are presented.

logical variable were computed from the sample of 100,000 synthetic HMC configurations and are presented in Figure 10. Our results reveal:

1. *SMC*: The two well-defined steep zones were identified. For the first one, between 7 and 150 tiles, the combined effect of EH, SSH, and WSH drove the convergence of SMC. This indicates the relevance of all the scales of heterogeneity and the hydrological connectivity in representing SMC in this domain. For the second zone, between 400 and 550 tiles, WSH was the only factor influencing the climb, with marginal contributions of SSH and EH. Therefore, for the SMC case, the large-scale variability of watersheds ultimately controlled the observed heterogeneity in the QFD configuration. This behavior is related to a greater influence of climatological features determining the SMC spatial patterns on annual scales over the selected study site.
2. *SH*: Only one steep zone was identified for this variable (between 7 and 1,000 tiles). Initially, the rise was driven by EH. The second increase stage was determined by a significant surge of SSH, accompanied by a marginal, not continuous, influence of EH. Then, an isolated contribution of WSH led the curve up to half of the slope. Finally, for the last part of the slope, SSH regained importance, being the final factor taking the curve to a flatter zone. Therefore, for the SH case in the study domain, the convergence depended heavily on finer scales, with the variance between small patches (i.e., SSH and EH) being more relevant than the one between watershed-sized clusters. This result revealed the influence of differences in land cover, vegetation

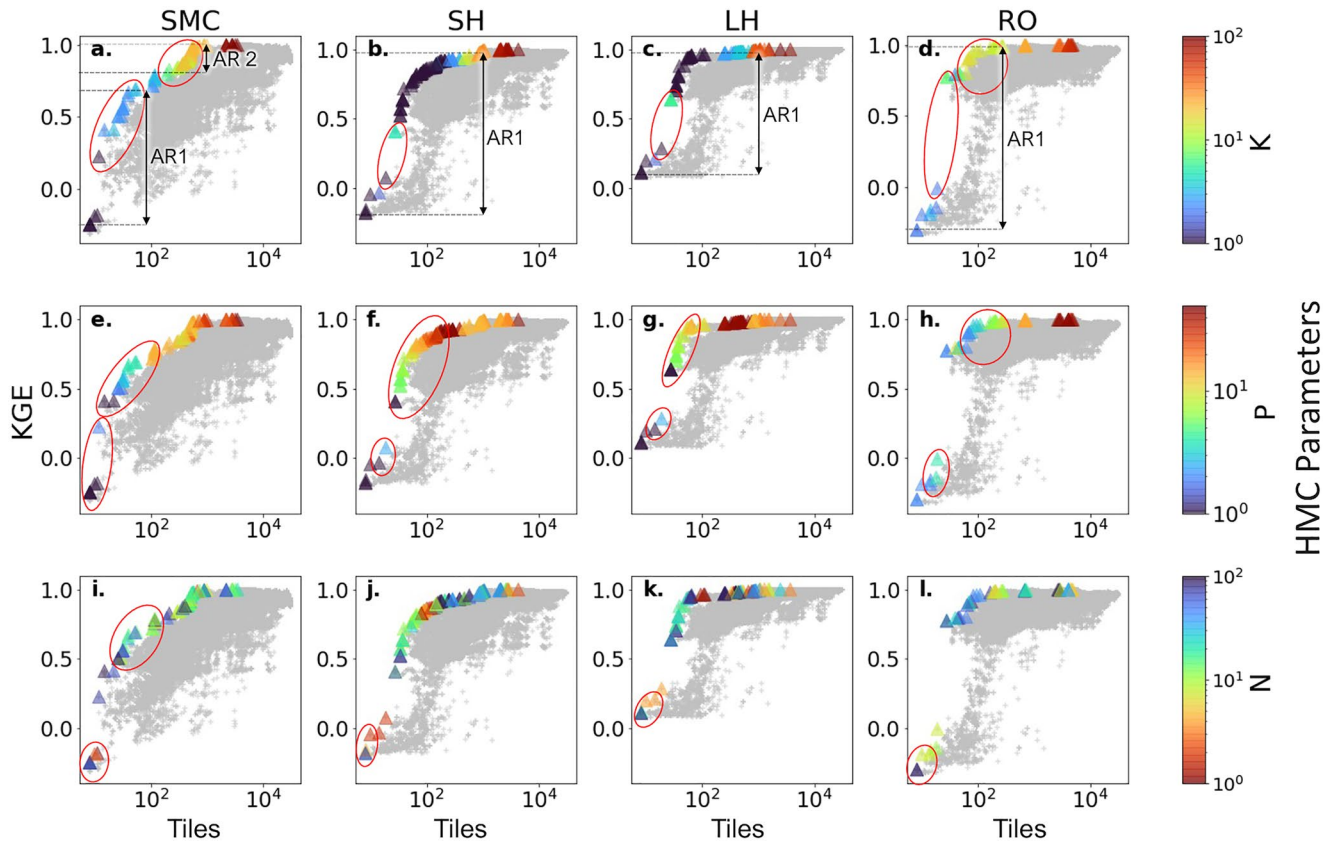


Figure 10. Pareto front for the empirical spatial covariance function-Kling-Gupta efficiency metric (ESCF-KGE) metric; a separate Pareto front is calculated for each variable. Colored data points indicate the front members and gray dots the remaining configurations. The colormaps indicate the values of the parameters k , p , and n . The analyzed regions (AR) for each hydrological variable are presented in the first row. Red circles indicate variations in the Hierarchical Multivariate Clustering parameters leading to increasing ESCF-KGE values. Results are presented for soil moisture content (a, e, and i), SH (b, f, and j), LH (c, g, and k), and RO (d, h, and l).

type, and small-scale topographical features (i.e., slope and aspect) in determining the long-term spatial patterns of sensible heat in the selected domain.

3. *LH*: As for the SH case, only one steep zone was observed in the Pareto Front, between 10 and 60 tiles. EH and SSH initially drove the significant convergence jump. The last two zones were determined mainly by the large-scale heterogeneity, WSH, followed by a contribution of smaller scales, SSH, for the final stage. Therefore, all the partition scheme steps were relevant for representing the heterogeneity related to latent heat flux in the Colorado domain. However, it appeared that SSH and EH play a more meaningful role than WSH, probably due to the same reasons exposed for the SH case: the convergence of latent heat flux in the domain depended on scales smaller than the watershed ones (i.e., EH and SSH) determined by small-scale landscape features (i.e., land cover, vegetation cover, slope, and aspect).
4. *RO*: A single steep zone was observed in the convergence for runoff, between 10 and 80 tiles. Three zones were identified in the rise: In the first stage, SSH and EH were the determining factors, followed by WSH in the second zone. Eventually, WSH and SSH took the path to the convergence zone. These results showed how the representation of the large-scale patterns, WSH, played a fundamental role in reaching convergence for long-term mean runoff over the domain. However, contrary to the SMC case, for the runoff case the importance was shared with a detailed definition of streams and riparian zones, EH, and a marginal contribution of small-scale patterns representation, SSH.

Finally, to further evaluate the differences between the pattern-agnostic and pattern-aware metrics of convergence (i.e., spatial variance and ESCF-KGE), the obtained Pareto fronts for each metric and process were compared (Figure 11). First, it could be noted how the shape of the Pareto front differs between metrics, mainly for the SH and LH cases. This behavior was consistent with the one-to-one relationships between metrics analyzed in Figure 8 (i.e., a more direct linear relationship between metrics for the SMC and RO cases and nonlinear relations

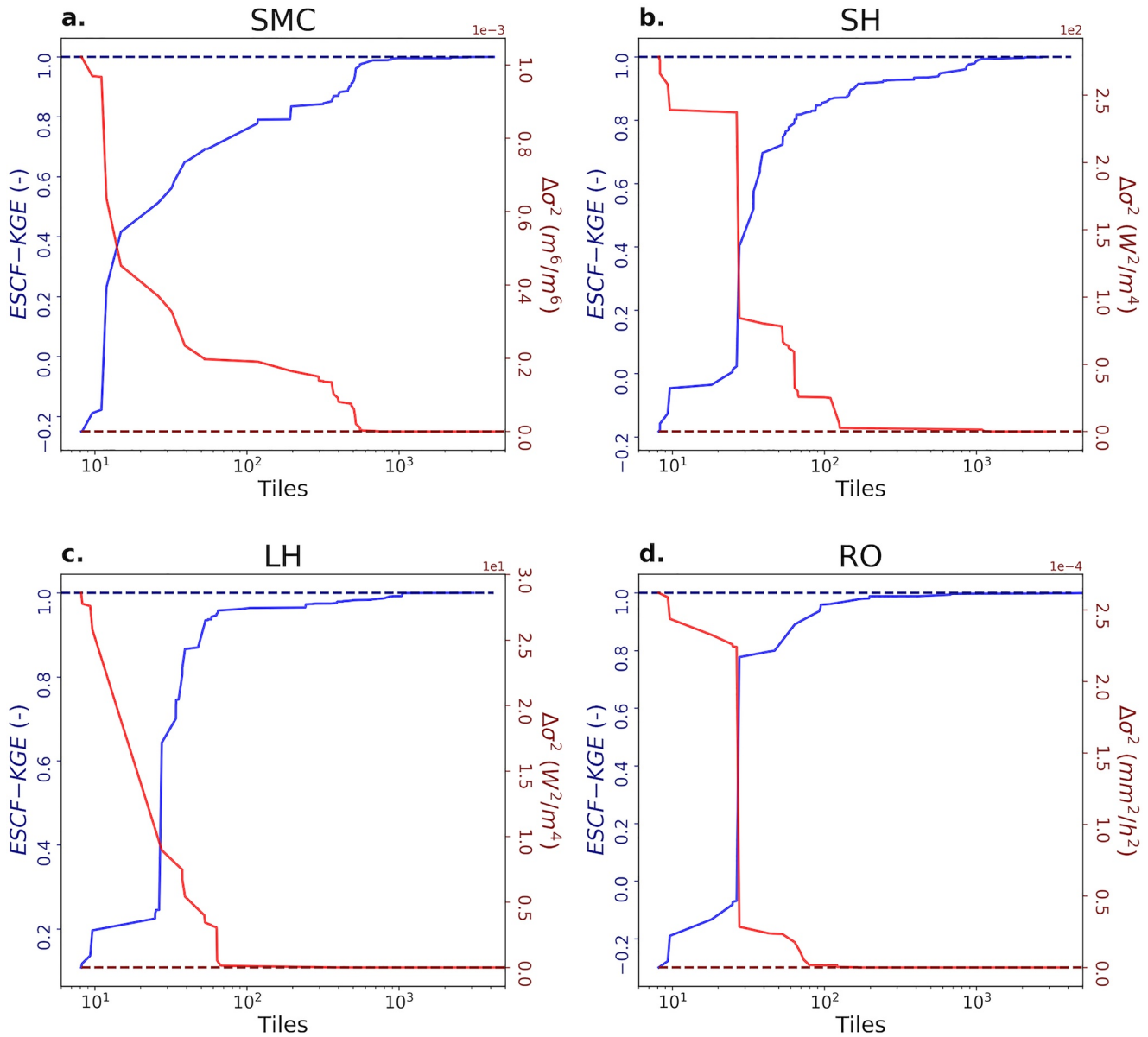


Figure 11. A comparison of the Pareto fronts obtained using the difference of spatial variance to quasi-fully distributed (red, right axes) and empirical spatial covariance function-Kling-Gupta efficiency metric (ESCF-KGE) (blue, left axes). Solid lines represent the Pareto fronts obtained for each metric, and the dotted lines the convergence value for each of them (zero for $\Delta\sigma^2$ and one for ESCF-KGE). (a) SMC, (b) SH, (c) LH, and (d) RO.

for LH and SH). Second, this comparison allowed to assess how the number of tiles needed to converge to the QFD configuration was indeed influenced by the selected convergence metric. In general, it could be observed how the number of tiles required to converge to the QFD was higher for the ESCF-KGE case than for the spatial variance case (i.e., the number of tiles needed for the solid and dotted lines to intersect). Again, these differences were evident for the SH and LH cases and significant for the RO case. These differences might be related to the more decisive influence of finer scales of heterogeneity (SSH and EH) on the obtained Pareto fronts for the ESCF-KGE metric for the three processes, which ultimately led to increases in the required number of tiles.

3.3. Sobol Sensitivity Analysis

The Pareto front evaluation was complemented by performing a Sobol sensitivity analysis on the HMC parametric space for the spatial variance and the ESCF-KGE cases (Figure 12). For latent heat flux and runoff,

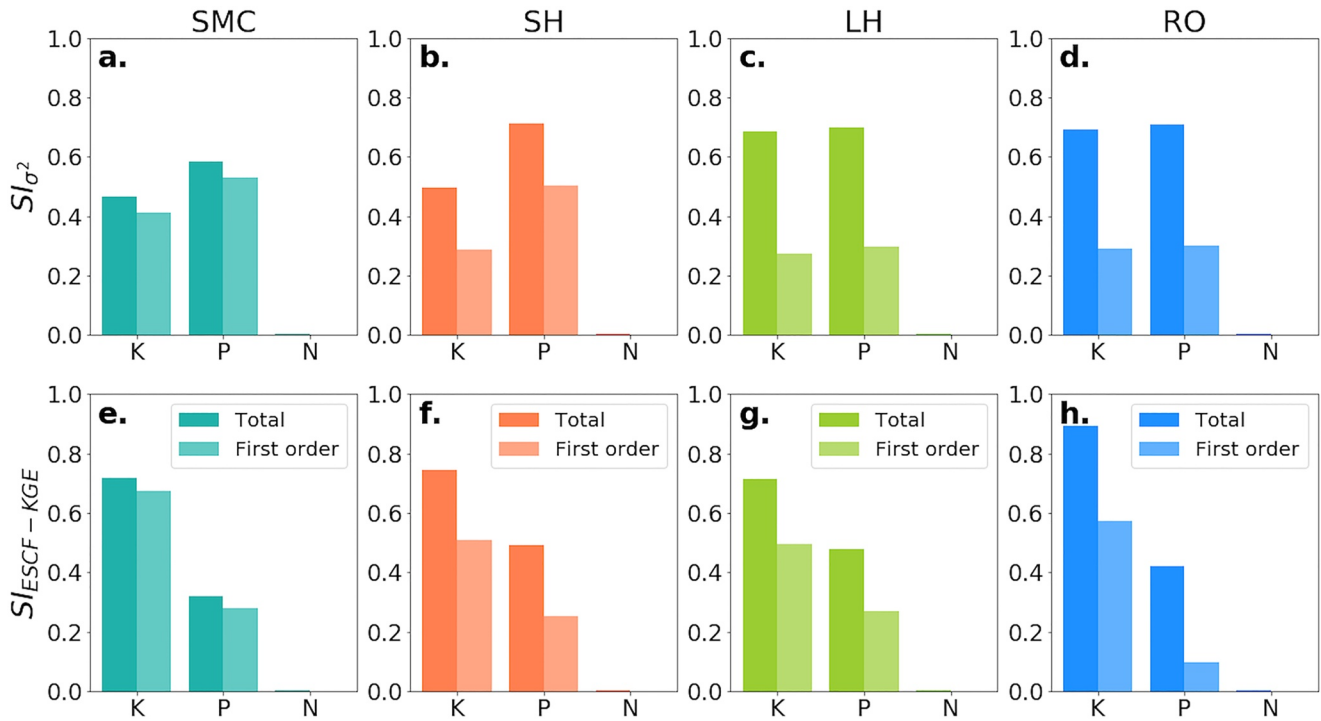


Figure 12. Sobol sensitivity analysis for spatial variance (first row) and soil moisture content (ESCF-KGE) (second row). Lighter and darker colors for the bars are used to display the first-order and total-effect Sobol indices (SI), respectively. The results are displayed for soil moisture content (a and e), SH (b and f), LH (c and g), and RO (d and h).

WSH and SSH (represented by k and p parameters respectively) had comparable importance for spatial variance metric while WSH dominated the behavior for the ESCF-KGE metric results. This implies that the large-scale features represented by the clusters of watersheds strongly influenced the pattern-aware method. The results for SMC and SH cases were similar: the small-scale features (SSH) dominated the spatial variance, while the watershed-scale ones (WSH) were the factors that primarily contributed to the variability of the ESCF-KGE approach. This behavior can be explained by the fact that WSH is more likely to affect the ESCF-KGE values than the small-scale, local-scale variability (i.e., EH and SSH) for the processes being analyzed.

Finally, contrary to WSH and SSH, EH did not seem to play a significant role in the sensitivity analysis. This result occurred because decreasing the n parameter (elevation heterogeneity) did not lead to significant variations, neither on the spatial variance values nor in the ESCF-KGE ones, for any analyzed variable. Therefore, the variance associated with this parameter was low compared to the other stages. As the HMC was structured for this study, the channel network was always represented as an independent tile. Besides, a significant part of the elevation-induced heterogeneity was already considered in the first stage of the tiling scheme (i.e., WSH). For the processes being analyzed, these two issues implied that an important fraction of the variability related to gradients between highlands and lowlands was already intrinsically included in the configurations, without a direct influence of the height discretization parameter, n . However, when assessing the convergence paths presented in Figure 10, n showed a more substantial effect because that analysis focused on the simultaneous reduction of units and improvement in the performance metric, and this parameter had a significant impact on the resulting number of tiles.

3.4. Combined Metric

The combined metric described in Section 2.9 was computed for the synthetic configurations, and the convergence patterns obtained were analyzed in terms of the HMC parameters' values. Figures 13a–13c shows the simultaneous hydrologic convergence of the combined metric, accompanied by the values of k , p , and n that produced the results. As expected, by simultaneously increasing the effects of WSH, SSH, and EH (k , p , and n , respectively) the number of tiles increased, and the metric value approached the QFD solution at $m = 1$.

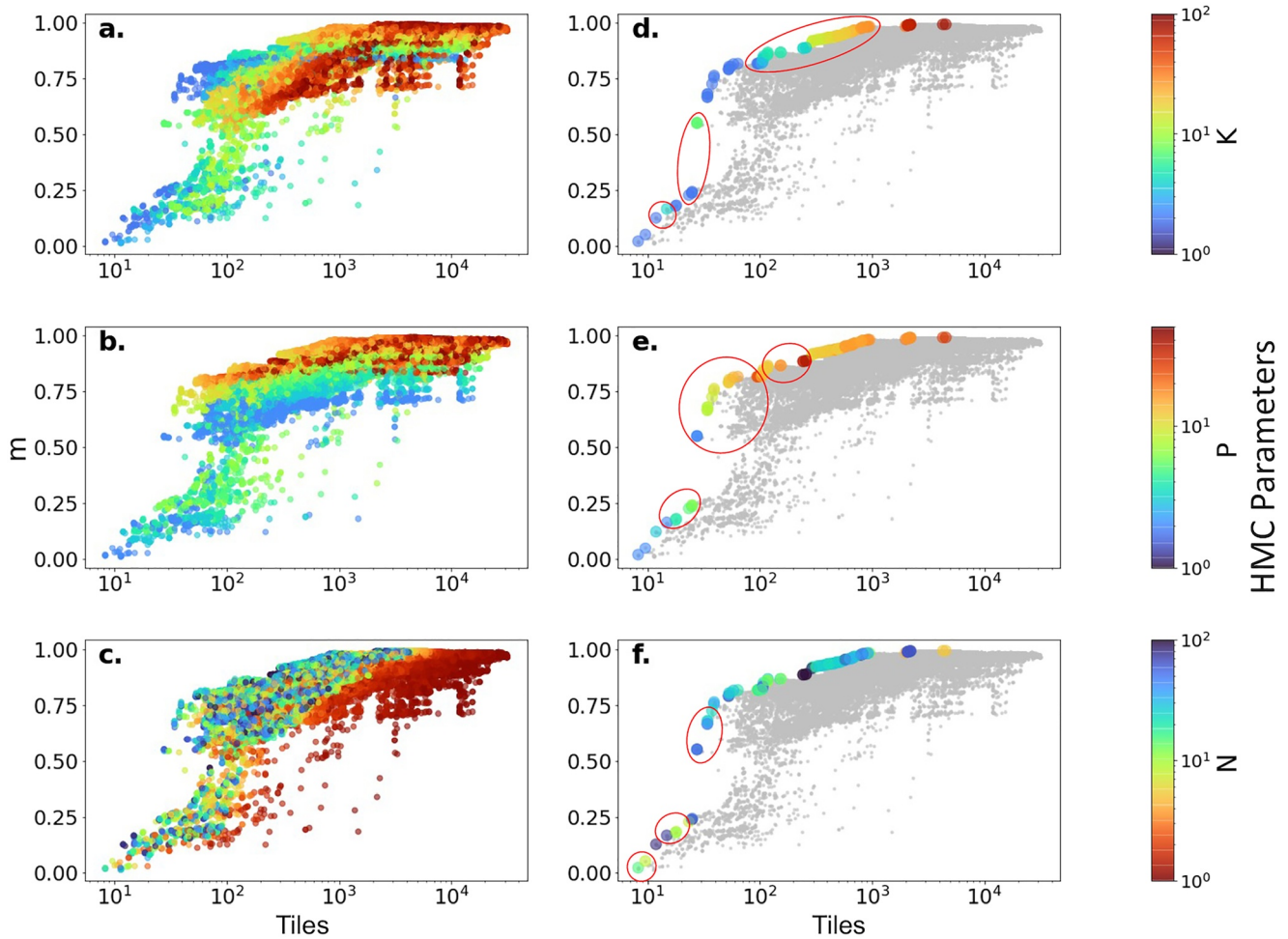


Figure 13. Convergence of the number of tiles, as computed by the combined metric described in Equation 8. The colormaps indicate the values of the parameters k , p , and n used. Plots a, b and c show the results for all the simulations; d, e, and f show the obtained Pareto front as colored data points and the remaining configurations as gray dots. Red circles indicate variations in the Hierarchical Multivariate Clustering parameters leading to increasing metric values.

The plots presented in Figures 13a–13c were modified to highlight the points belonging to the Pareto front (Figures 13d–13f). These results allowed identifying the importance of each one of the tiling scheme stages in the convergence of the metric. Initially, from 9 to 10 tiles, convergence was driven by EH, followed by a considerable contribution of WSH from 10 to 20 tiles. Then, SSH and EH gained importance (from 20 to 25 tiles), followed by WSH. The rise driven by WSH was relevant since it takes the metric from 0.25 to over 0.5 using fewer than 10 extra tiles. In this sense, the importance of the large-scale watershed patterns found independently for all the processes in previous analyses reappeared during the simultaneous evaluation. Joint SSH-EH and WSH-SSH mainly drove the last two increases in performance. This result implied that all the HMC tiling stages had some influence in reaching convergence for the higher number of tiles. Hence, when looking at all the processes simultaneously, initial convergence was determined by riparian zones, followed by large-scale heterogeneity of watersheds and small-scale heterogeneity representation (i.e., land cover and soil features). Finally, all the effects together contributed to fully emulating the QFD solution over the study domain: meteorology and elevation (WSH), riparian zones and channels representation (EH), and small-scale features (SSH).

Regarding the number of units required to achieve acceptable performance on the representation of the four processes in this domain, from the Pareto front results presented in Figure 13, a configuration with about 800 tiles ensured a combined metric value converging to 1. The approximate values of the HMC parameters generating this configuration were 30 for k , 20 for p , and 30 for n . The obtained converged configuration was most likely related to the constraint imposed by the units required to converge on the QFD for the sensible heat case (around 1,000

units). However, a word of caution is needed: the previous results are site-, time aggregation-, parameter-set- and tiling scheme-specific.

4. Discussion

4.1. Implications to the LSM Community and Case Study Findings

For the first time, this study provided the LSM community with an approach that allows to objectively determine the optimal tile configuration per macroscale grid cell and apply it to a comprehensive tiling scheme (i.e., HMC). To that end, the introduced approach used a pattern-aware metric of heterogeneity (i.e., ESCF-KGE) to summarize and evaluate the LSM aggregated sub-grid outputs. Additionally, the approach adopted a multi-objective Pareto front analysis to determine the optimal tiling structure that accurately represented multiple processes simultaneously while minimizing the number of tiles. The results derived from the application of the approach also aided in performing detailed analyses on the tiling scheme parametric space and assessing the sensitivity of the hydrological spatial patterns to these parameters, allowing evaluation of how different paths of heterogeneity representation led to a similar performance concerning a QFD solution. Additionally, the training and application of a surrogate model to predict the spatial structure of the LSM HO fields showed promise in reducing the computational burden derived from the need for extensive LSM simulations to perform sensitivity analyses. Finally, the performed analyses highlighted the differences between pattern-agnostic and pattern-aware heterogeneity metrics, providing a new perspective on the traditionally used methodologies to evaluate LSM macroscale grid cell outputs.

The main result of this study for the selected domain was that about 100 tiles are enough to converge on the multi-scale response of a quasi-fully distributed model when considering single processes. The multi-process case was more complex, requiring around 800 tiles to reach convergence. Still, the ratio between the number of units used for the QFD solution and the selected converged configuration showed that a reduced-order configuration effectively reproduced a highly heterogeneous model setup with a significantly lower computational expense (0.12% and 1% for the single- and multi-process cases, respectively). However, these findings are specific to the selected domain and study design. Further analysis would be required to generalize the results to other locations, temporal aggregations, tiling structures, and parameterizations.

The approach enabled the optimization of the number of tiles so that more tiles could be used in locations where more detail was needed. In contrast, fewer tiles could be used in places with smaller variability (e.g., flat, homogeneous domains). Since, for the current study, a site with large topographic gradients and significant physical heterogeneity was chosen, it is not surprising that the required number of tiles is large. However, the overall idea that tiles can be computed a priori could help address the challenge of using a limited number of tiles in LSM.

4.2. General Implications for the ESM Field

It has been proved that heterogeneities emerging over the landscape can have a crucial role in many critical atmospheric processes, such as determining the atmospheric boundary layer depth, triggering convection, and initiating mesoscale circulations (Bertoldi et al., 2012; Gutowski et al., 2020; Kang & Bryan, 2011; Kustas & Albertson, 2003; Ntelekos et al., 2008; Simon et al., 2021; Timmermans et al., 2008). Progress is being made in regional and local studies to understand the role of the multi-scale landscape heterogeneity on micro- and meso-scale meteorological processes (Huang & Margulis, 2013; Senatore et al., 2015; Shrestha et al., 2014; Talbot et al., 2012; Vergopalan et al., 2022). However, the effect of heterogeneity remains unknown, primarily when it comes to land-atmosphere interactions in the climate system as a whole. This uncertainty mainly arises from the simplistic coupling between existing sub-grid parameterizations in land-surface models (i.e., the tiling schemes described extensively in this study) and the atmospheric components of ESMs. Typically, ESMs only exchange spatial mean fluxes of mass and energy between the land and atmosphere and ignore high-order spatial statistics, such as spatial variance or correlation lengths.

Nevertheless, convection and turbulence parameterizations in atmospheric circulation models are starting to include higher-order sub-grid scale processes. Some examples of these efforts include the Cloud Layers Unified By Binormals (CLUBB) and Eddy Diffusivity Mass Flux (EDMF) (Golaz et al., 2002; Sušelj et al., 2013). These developments can be seen as opportunities for the potential coupling of atmospheric models with the sub-grid

scale heterogeneity of the land surface. This study is framed to be meaningful for such efforts. Ideally, the methods exposed here can help inform land surface parametrizations within the atmospheric components of ESMs with higher-order statistics. A pattern-aware metric such as the ESCF-KGE metric can provide more than information about the spatial fitness of the obtained fields concerning the QFD configuration; it can also be used to estimate the characteristic scales of heterogeneity of such fields (i.e., characteristic lengths), providing parametrizations with useful spatial information over macroscale grid cells. Besides, ESMs could benefit from approaches that are aware of the information lost related to certain arbitrary decisions (i.e., number of subgrid tiles, proxies of physical heterogeneity and tiling schemes, HOs considered in models' evaluation). Ultimately, the hope is that the type of approach presented through this study drives the broader community in a direction where the representation of the subgrid-scale heterogeneity is tackled more strictly.

4.3. Limitations and Impacts of Method Choices

4.3.1. Using a QFD Configuration as a Benchmark

As stated in Section 2.7, a QFD HMC configuration (~83,000 tiles), simulated with HydroBlocks, was used as the evaluation benchmark. This benchmark allowed for a straightforward assessment of the synthetic configurations' performance in representing the spatial heterogeneity of a fully distributed solution. However, it is clear that a 30-m fully distributed HydroBlocks simulation would have been the ideal benchmark. However, computational limitations hindered the use of the fully distributed fields. In this sense, the possibility of the results of this study being influenced by this simplification must be acknowledged. This fact is especially true in the selected study domain, given that its landscape heterogeneity is significantly high, and more tiles than the ones contained within the QFD solution might be necessary to emulate it fully. However, even the fully distributed configuration cannot be considered the most realistic representation of reality. In fact, the significant uncertainty in model structure, parameters, and forcing minimizes the prospective advantages of modeling at higher spatial scales, as pointed out by numerous literature sources (Beven & Cloke, 2012; Beven et al., 2014; Bierkens, 2015). In this context, the only way to ensure a realistic representation of the output fields is by setting an observational data source as the evaluation threshold; in the future, using remotely sensed fields appears as a promising alternative.

4.3.2. Transferability of the Optimal Tiling Configurations

Although the results presented here provide a promising path forward for a robust and efficient representation of multi-scale heterogeneity in large-scale models, important limitations should be considered.

1. *LSM Model structure*: The obtained set of optimal tile configurations is tied to the process representation within the adopted LSM. Adding more processes and modifying their representation will immediately influence the spatial patterns and the required number of units to emulate them (i.e., as more complexity is added to the model's structure, more units would be required to represent the spatial fields).
2. *Tiling scheme*: The adopted tiling scheme, HMC, tightly constrains the results presented here. This scheme represents just one of many possible alternatives to attack the challenge of determining the tiling structure to use within LSM macroscale grid cells, imposing a specific a priori structure to the tiling that might not be appropriate for all domains or applications. Consequently, further work must look into applying the ideas exposed throughout this study to other tiling schemes/techniques to determine the most appropriate tiling structure.
3. *Clustering*: The results suggest that around 800 tiles are necessary to successfully approximate the QFD spatial structure of annual mean fields of latent heat flux, soil moisture content, surface runoff, and sensible heat flux for the Upper Colorado domain. However, this result is dependent on both the (a) chosen proxies of spatial heterogeneity and (b) the clustering method adopted by the tiling scheme.
 - (a) *Selecting proxies of spatial heterogeneity*: A series of factors contributes to generating surface heterogeneity: the variability associated with vegetation cover, or in a more general fashion, with surface type and land use; the variability in morphology and soil characteristics; the spatial variability in climatic forcing, among others. By design, this approach assumes that the observed physical features are tightly coupled to the small-scale processes involved in the water, energy, and biogeochemical cycles. Future work should investigate the influence of covariates informing HMC in improving the representation of spatial patterns of hydrological variables.

- (b) Clustering mechanism: Several clustering applications characterizing the sub-grid heterogeneity within ESMs and LSMs can be found in the literature (Loritz et al., 2018; Mälicke et al., 2020; Melton et al., 2017; Newman et al., 2014; Wainwright et al., 2022). However, a comprehensive evaluation of the impact of the selected clustering algorithm on the accuracy of the representation of different processes has not been performed. This advance will require careful analysis of the additional computational costs and accuracy gains of different clustering methods in various tiling schemes.
4. *Temporal and spatial scales*: In this study, the sets of optimal configurations are determined as a function of the time-averaged behavior of a QFD configuration. The spatial extent of the domain is set to one arc degree by one arc degree area. Although these temporal and spatial scales may be appropriate for climate modeling, applications requiring finer temporal and spatial resolutions would most likely need different sets of optimal configurations. As for the increasing complexity in the processes, an accurate representation for increasing resolutions will presumably demand more units in addition to different parameter combinations. For instance, for smaller timescales, it is hypothesized that the height discretization stage of HMC would gain importance as the hillslope processes become relevant. Therefore, the transferability of the optimal set of configurations between scales and applications is restricted. Hence, further analysis of the proposed scheme over smaller time windows and different domain sizes must be performed.
 5. *Selected pattern-aware metric of heterogeneity*: This study's selected measure of pattern heterogeneity was the Empirical Spatial Covariance Function. This metric was chosen as it was easily attainable from the HydroBlocks output fields, providing a relatively dense characterization of the heterogeneity degree on different scales. However, other geostatistical tools can be used to obtain such scale-comprehensive descriptions. The variogram and correlogram provide measures of the spatial correlation structure of continuous quantitative fields. Additionally, if the typical spatial scales of variation are the goal, other tools such as the Fourier and wavelet analysis may be appropriate. Even more, if the dynamic nature of the problem is to be considered, the temporal dimension of the fields can be added, leading to spatiotemporal variogram or covariance function models.

4.4. Future Developments

4.4.1. Assembling Optimal Tiling Configurations Over Continental Scales

This study shows how a relatively low number of tiles (~800 tiles) in comparison to a QFD setup (~83,000 tiles) are enough to converge on the behavior of a highly complex configuration for the case study. Therefore, it provides a middle ground for large-scale models: the role of the sub-grid fine-scale features is captured while computational efficiency is preserved. However, further work is needed to extend this approach to enable determining the optimal HMC parameters to use over different domains without performing intensive LSM simulations. A computationally feasible option to achieve this goal is to sample a comprehensive subset of domains (i.e., macroscale grid cells) that compose a geographically distributed set with varied topographic, climatological, and physical characteristics. For each domain, the HMC parameters would be independently optimized, as described in this study. Finally, a machine learning scheme (e.g., random forest or neural networks) would be trained on the subset and used to regionalize the behavior of the optimal tile configurations. The model would be trained on sets of environmental covariates, their spatial organization features over the modeling domain (i.e., correlation lengths), and the required error in the hydrological target variables. This procedure would allow the assembly of the optimal configuration for any domain, avoiding excessive computational expenses.

4.4.2. Dynamic Tiling Configurations

In this study, some of the features used to define heterogeneity (e.g., elevation) are fixed on general timescales of ESMs (i.e., tens to hundreds of years), while others should be dynamic in time—for example, changes in vegetation and land use by natural or anthropogenic causes (Landry et al., 2016). Hence, the degree of clustering required for specific features can vary in time; in other words, the sub-grid heterogeneous representation may not be static. Such representation is already used within ESMs to account for land use change due to harvesting, deforestation, and fires, among others (e.g., Milly et al., 2014). However, this scheme has not been extended to account for the time-varying soil moisture and floodplain dynamics, which will play an important role in determining the level of detail required. As HMC is implemented in this study, the a priori computation of the covariance between environmental properties simplifies the problem by assuming the drivers for the tiling are fixed in both space and time. However, updating the tiling configuration should be a continuous activity during

the model runtime rather than a one-time a priori exercise. The concept of continuously checking the clustering structure and “re-clustering” as needed is called an adaptive clustering scheme and has been applied before to several simple models (Ehret et al., 2020; Loritz et al., 2021). A step forward in this respect would include the development and application of an adaptive clustering scheme into the HMC structure.

However, the simultaneous operation of an HMC-adaptive clustering scheme within the modeling framework represents a substantial increase in the complexity and cost of representation of the land surface. For this reason, the marginal efficiency of the dynamic approach against a static tiling structure must be carefully considered before implementing these approaches. If found useful, these dynamic/adaptive schemes must be implemented in a flexible way such as they can either be turned on or off. Besides, the degree of temporal disaggregation must be modified following the nature of the research question. For instance, a recent Mälicke et al. (2020) study looked into the spatiotemporal variability of soil moisture from in-situ observations. This work revealed that while the spatial organization of soil moisture can be highly variable in space, it is also persistent over time, with the temporal variability of its patterns being strongly determined by meteorology (i.e., the forcing data). Other authors such as Mittelbach & Seneviratne (2012) and Teuling et al. (2006) also found characteristic spatial patterns persisting over time. In this sense, the implemented schemes should be able to consider the temporal persistence of the processes of interest and include this information in their dynamic components. Finally, the applicability of optimization techniques presented in this study to these adaptive structures would be limited, especially if using a fully distributed simulation as a benchmark. If using a dynamic framework, tiling evaluations must certainly be performed against observed, spatially distributed information.

4.4.3. Moving From Landscape Proxies of Hydrologic Heterogeneity to Hydrologic Processes Heterogeneity

The implementation of HMC throughout this study assumes that the observed characteristics of the physical environment (i.e., proxies of physical heterogeneity) provide a robust representation of the heterogeneity of the water and energy cycles. Although this assumption is generally appropriate, it only indirectly addresses the overarching goal, which is to characterize the processes' real, observed natural multi-scale heterogeneity. Besides, the approach taken in this study set the QFD simulation as the benchmark for the optimization process. By taking this route, the computational efficiency of the model is prioritized over the realism of the generated fields.

Moving forward, future optimization attempts should move beyond approaching fully distributed simulations and use satellite remote sensing products that directly measure the states and fluxes of the cycles at high spatial and temporal resolutions as benchmarks for the evaluations. For example, land surface temperature (LST) retrievals from GOES-16 (Fang et al., 2014; Yu et al., 2010) and ECOSTRESS (Hook & Hulley, 2019) would provide key observations of the spatial and temporal distributions of surface fluxes within modeling domains. As a critical state variable of the land surface, LST encodes information of local energy and water fluxes. Accordingly, several relevant hydrological quantities such as SMC, vegetation water stress, gross primary production, and crop yield correlate strongly with LST, making it a key variable in understanding the physics of multiple land surface processes (Holzman et al., 2014; Li et al., 2021; Sims et al., 2008; Vergopolan et al., 2021; Wang and Dickinson, 2012). Other variables that would be useful include Leaf Area Index and the Normalized Vegetation Difference Index from MODIS (Spruce et al., 2016; Yuan et al., 2011), and evapotranspiration from Landsat (Anderson et al., 2012), among others. Although biased, these data would provide a more formal connection between the model and the observed heterogeneity, particularly regarding the spatial organization of the fields and signal persistence. Besides, using these high-resolution data sets would also open a novel path to assimilating them into field-scale land models.

5. Summary and Conclusions

One persistent challenge for LSMs and ESMs is to derive an effective yet efficient representation of the spatially heterogeneous physical landscape and resolve the nonlinear hydrological processes on scales smaller than those resolved by the numerical models. Most modern LSMs use sub-grid “tiling” techniques to divide their grid cells into smaller units and support applications across various scales. However, tiling techniques face several challenges. First, defining the number of tiles remains mostly an arbitrary decision. Second, grid-box summary statistics are usually employed to summarize the land surface fluxes and storage. Despite being informative, these statistics are insensitive to the large-scale spatial patterns within the macro-scale grid cell. Third, the number

of tiles must be considered as an influencing factor within the tiling structure determination as it increases the computational burden. Finally, the multi-objective nature of the problem must be considered as usually the accurate representation of a single state/flux is not the end goal.

This study introduced a pattern-aware metric of heterogeneity based on the concept of the empirical covariance function to measure the degree of heterogeneity representation of several tiling configurations. In addition, we introduced a multi-objective optimization approach using a large sample of tiling scheme parameter combinations. Ultimately, we converged toward the multi-process response of a quasi-fully distributed configuration while minimizing the associated computational cost. The selected tiling configuration reduced the computational burden to 1% of the quasi-fully distributed structure for a case study centered on a $1 \times 1^\circ$ region of complex terrain (in mountainous Colorado, USA). An additional feature of the approach included the ability to perform detailed analyses on the tiling scheme parametric space.

The results showed how (a) a predictive random forest model can be used to predict the spatial structure of the LSM's time-averaged hydrological fields; (b) an LSM with a reduced-order tile configuration reproduced a highly heterogeneous model setup with a significantly lower computational expense; (c) hydrological convergence is highly process-, time aggregation-, site-, parameter set, and tiling scheme-dependent; (d) the optimal hydrological convergence routes can be identified using a multi-objective Pareto efficiency analysis, and (e) a single metric can be used to determine the optimum set of configurations to be used in the simultaneous representation of the multi-scale heterogeneity of several processes. These results, however, were specific to the selected tiling scheme (HMC). As such, it is recognized that this is just one of the many possibilities for representant heterogeneity and that more efficient and accurate tiling strategies might exist.

The developed approach is only the first attempt to objectively select the parameters to use within LSM's and ESM's tiling schemes and simultaneously evaluate the quality of the LSM's modeled hydrological processes. Moving forward, the transferability of the approach should be tested under various tiling schemes, hydrological model structures, clustering techniques, proxies of physical heterogeneity, temporal windows, and domain sizes. Furthermore, the community will benefit from a generalized method to precompute the tiling structure on any domain over continental scales (e.g., CONUS), as well as from a method to objectively quantify the heterogeneity and information losses derived from arbitrary selections within models. This work represents a step toward adapting the current tiling schemes to better leverage the available high-resolution data to account for the dynamic nature of land surface processes. Overall, this novel approach provides a path forward to precomputing robust tile configuration for ESMs and LSMs while considering the spatial heterogeneity and accuracy of hydrologic processes.

Data Availability Statement

The HydroBlocks model code used in this study is preserved at <https://doi.org/10.5281/zenodo.4071692> (Chaney & Vergopolan, 2020). The data that support the findings of this study, including the HydroBlocks inputs, the ESCFs resulting from HydroBlocks simulations used to train and test the Random Forest Models (RFM), the converged RFM, and the ESCFs for the QFD solution are preserved at <https://doi.org/10.5281/zenodo.7051439> (Torres-Rojas and Chaney, 2022). Table 2 references the open access data sets used to force HydroBlocks.

Acknowledgments

This study was supported by funding from NOAA Grant NA19OAR4310241 (Parameterizing the effects of sub-grid land heterogeneity on the atmospheric boundary layer and convection: Implications for surface climate, variability, and extremes) and NOAA Grant NA19OAR4310244 (3D-Land Energy Exchanges: Harnessing High Resolution Terrestrial Information to Refine Atmosphere-to-Land interactions in Earth System Models). We also want to thank the reviewers for their informative and constructive comments and suggestions.

References

- Ajami, H., Khan, U., Tuteja, N. K., & Sharma, A. (2016). Development of a computationally efficient semi-distributed hydrologic modeling application for soil moisture, lateral flow, and runoff simulation. *Environmental Modelling & Software*, 85, 319–331. <https://doi.org/10.1016/j.envsoft.2016.09.002>
- Anderson, M. C., Kustas, W. P., Alfieri, J. G., Gao, F., Hain, C., Prueger, J. H., et al. (2012). Mapping daily evapotranspiration at Landsat spatial scales during the BEAREX'08 field campaign. *Advances in Water Resources*, 50, 162–177. <https://doi.org/10.1016/j.advwatres.2012.06.005>
- Avissar, R., & Pielke, R. A. (1989). A parameterization of heterogeneous land surfaces for atmospheric numerical models and its impact on regional meteorology. *Monthly Weather Review*, 117(10), 2113–2136. [https://doi.org/10.1175/1520-0493\(1989\)117<2113:APOHLS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<2113:APOHLS>2.0.CO;2)
- Bertoldi, G., Kustas, W. P., & Albertson, J. D. (2012). Evaluating source area contributions from aircraft flux Measurements over heterogeneous land using large-eddy simulation. *Boundary-Layer Meteorology*, 147(2), 261–279. <https://doi.org/10.1007/S10546-012-9781-Y>
- Beven, K. (2010). *Rainfall-runoff modelling: The primer/Keith Beven* (2nd ed., pp. 157–175). John Wiley & Sons, Ltd.
- Beven, K., & Cloke, H. (2012). Comment on “hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water” by Eric F. Wood et al. In *Water resources research* (Vol. 48). John Wiley & Sons, Ltd. <https://doi.org/10.1029/2011WR010982>

- Beven, K., Cloke, H., Pappenberger, F., Lamb, R., & Hunter, N. (2014). Hyperresolution information and hyperresolution ignorance in modelling the hydrology of the land surface. *Science China Earth Sciences*, 58(1), 25–35. <https://doi.org/10.1007/S11430-014-5003-4>
- Bierkens, M. F. P. (2015). Global hydrology 2015: State, trends, and directions. *Water Resources Research*, 51(7), 4923–4947. <https://doi.org/10.1002/2015WR017173>
- Blöschl, G., & Sivapalan, M. (1995). Scale issues in hydrological modelling: A review. *Hydrological Processes*, 9(3–4), 251–290. <https://doi.org/10.1002/HYP.3360090305>
- Brutsaert, W. A. (1982). In D. Reidel (Eds.), *Evaporation into the atmosphere* (p. 299).
- Chaney, N., & Vergopolan, N. (2020). Chaneyn/HydroBlocks: HBrouting_Oct2020, Zenodo. [code]. <https://doi.org/10.5281/zenodo.4071692>
- Chaney, N. W., Herman, J. D., Ek, M. B., & Wood, E. F. (2016). Deriving global parameter estimates for the Noah land surface model using FLUXNET and machine learning. *Journal of Geophysical Research: Atmospheres*, 121(22), 13218–13235. <https://doi.org/10.1002/2016JD024821>
- Chaney, N. W., Metcalfe, P., & Wood, E. F. (2016). HydroBlocks: A field-scale resolving land surface model for application over continental extents. *Hydrological Processes*, 30(20), 3543–3559. <https://doi.org/10.1002/hyp.10891>
- Chaney, N. W., Minasny, B., Herman, J. D., Nauman, T. W., Brungard, C. W., Morgan, C. L. S., et al. (2019). POLARIS soil properties: 30-m probabilistic maps of soil properties over the contiguous United States. *Water Resources Research*, 55(4), 2916–2938. <https://doi.org/10.1029/2018WR022797>
- Chaney, N. W., Torres-Rojas, L., Vergopolan, N., & Fisher, C. K. (2021). HydroBlocks v0.2: Enabling a field-scale two-way coupling between the land surface and river networks in Earth system models. *Geoscientific Model Development*, 14(11), 6813–6832. <https://doi.org/10.5194/gmd-14-6813-2021>
- Chaney, N. W., Van Huijgevoort, M. H. J., Shevliakova, E., Malyshev, S., Milly, P. C. D., Gauthier, P. P. G., & Sulman, B. N. (2018). Harnessing big data to rethink land heterogeneity in Earth system models. *Hydrology and Earth System Sciences*, 22(6), 3311–3330. <https://doi.org/10.5194/hess-22-3311-2018>
- Chen, F., & Dudhia, J. (2001). Coupling an Advanced land surface–hydrology model with the Penn state–NCAR MM5 modeling system. Part I: Model implementation and sensitivity. *Monthly Weather Review*, 129(4), 569–585. [https://doi.org/10.1175/1520-0493\(2001\)129<0569:caalsh>2.0.co;2](https://doi.org/10.1175/1520-0493(2001)129<0569:caalsh>2.0.co;2)
- Chen, F., Manning, K. W., Lemone, M. A., Trier, S. B., Alfieri, J. G., Roberts, R., et al. (2007). Description and evaluation of the characteristics of the NCAR high-resolution land data assimilation system. *Journal of Applied Meteorology and Climatology*, 46(6), 694–713. <https://doi.org/10.1175/JAM2463.1>
- Chen, G., Wei, Z., Jin, X., Dong, W., Zhu, X., Liu, Y., et al. (2020). The effects of the modified mosaic approach method on regional simulations of surface meteorological variables in Western China. *International Journal of Climatology*, 40(9), 4053–4066. <https://doi.org/10.1002/joc.6440>
- Chen, T. H., Henderson-Sellers, A., Milly, P. C. D., Pitman, A. J., Beljaars, A. C. M., Polcher, J., et al. (1997). Cabauw experimental results from the project for intercomparison of land-surface parameterization schemes. *Journal of Climate*, 10(6), 1194–1215. [https://doi.org/10.1175/1520-0442\(1997\)010<1194:CERFTP>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1194:CERFTP>2.0.CO;2)
- Clark, M. P., Fan, Y., Lawrence, D. M., Adam, J. C., Bolster, D., Gochis, D. J., et al. (2015). Improving the representation of hydrologic processes in Earth System Models. *Water Resources Research*, 51(8), 5929–5956. <https://doi.org/10.1002/2015WR017096>
- Deshpande, S., Watson, L. T., & Canfield, R. A. (2013). Pareto front approximation using a hybrid approach. *Procedia Computer Science*, 18, 521–530. <https://doi.org/10.1016/J.PROCS.2013.05.216>
- Ehret, U., van Puijssjen, R., Bortoli, M., Loritz, R., Azmi, E., & Zehe, E. (2020). Adaptive clustering: Reducing the computational costs of distributed (hydrological) modelling by exploiting time-variable similarity among model elements. *Hydrology and Earth System Sciences*, 24(9), 4389–4411. <https://doi.org/10.5194/hess-24-4389-2020>
- Famiglietti, J. S., & Wood, E. F. (1994). Multi-scale modeling of spatially variable water and energy balance processes. *Water Resources Research*, 30(11), 3061–3078. <https://doi.org/10.1029/94WR01498>
- Fang, L., Yu, Y., Xu, H., & Sun, D. (2014). New retrieval algorithm for deriving land surface temperature from geostationary orbiting satellite observations. *IEEE Transactions on Geoscience and Remote Sensing*, 52(2), 819–828. <https://doi.org/10.1109/TGRS.2013.2244213>
- Fang, Z., Bogena, H., Kollet, S., Koch, J., & Vereecken, H. (2015). Spatio-temporal validation of long-term 3D hydrological simulations of a forested catchment using empirical orthogonal functions and wavelet coherence analysis. *Journal of Hydrology*, 529, 1754–1767. <https://doi.org/10.1016/J.JHYDROL.2015.08.011>
- Fisher, R. A., & Koven, C. D. (2020). Perspectives on the future of land surface models and the challenges of representing complex terrestrial systems. In *Journal of advances in modeling Earth systems* (Vol. 12, p. e2018MS001453). Blackwell Publishing Ltd. <https://doi.org/10.1029/2018MS001453>
- Flügel, W.-A. (1995). Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the River Bröl, Germany. *Hydrological Processes*, 9(3–4), 423–436. <https://doi.org/10.1002/HYP.3360090313>
- Gesch, D., Evans, G., Mauck, J., Hutchinson, J., & Carswell, W. J., Jr. (2010). The National Map-Elevation National elevation data set: U.S. Geological Survey fact sheet. Tech. Rep. 2009–3053. Retrieved from http://www.usgs.gov/ngpo/ngp_liaisons.html
- Golaz, J. C., Larson, V. E., & Cotton, W. R. (2002). A PDF-based model for boundary layer clouds. Part I: Method and model description. *Journal of the Atmospheric Sciences*, 59(24), 3540–3551. [https://doi.org/10.1175/1520-0469\(2002\)059<3540:APBMFB>2.0.CO;2](https://doi.org/10.1175/1520-0469(2002)059<3540:APBMFB>2.0.CO;2)
- Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of Hydrology*, 377(1–2), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>
- Gutowski, W. J., Ullrich, P. A., Hall, A., Leung, L. R., O'Brien, T. A., Patricola, C. M., et al. (2020). The ongoing need for high-resolution regional climate models: Process understanding and stakeholder information. *Bulletin of the American Meteorological Society*, 101(5), E664–E683. <https://doi.org/10.1175/BAMS-D-19-0113.1>
- Hao, D., Bisht, G., Huang, M., Ma, P. L., Tesfa, T., Lee, W. L., et al. (2022). Impacts of sub-grid topographic representations on surface energy balance and boundary conditions in the E3SM land model: A case study in Sierra Nevada. *Journal of Advances in Modeling Earth Systems*, 14(4), e2021MS002862. <https://doi.org/10.1029/2021MS002862>
- Herman, J., & Usher, W. (2017). SALib: An open-source Python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9), 97. <https://doi.org/10.21105/joss.00097>
- Holzman, M. E., Rivas, R., & Bayala, M. (2014). Subsurface soil moisture estimation by VI-LST method. *IEEE Geoscience and Remote Sensing Letters*, 11(11), 1951–1955. <https://doi.org/10.1109/LGRS.2014.2314617>
- Homer, C., Dewitz, J., Yang, L., Jin, S., Danielson, P., Xian, G., et al. (2015). Completion of the 2011 national land cover database for the conterminous United States – Representing a decade of land cover change information. *Photogrammetric Engineering & Remote Sensing*, 81(5), 2971. [https://doi.org/10.1016/S0099-1112\(15\)30100-2](https://doi.org/10.1016/S0099-1112(15)30100-2)

- Hook, S., & Hulley, G. (2019). ECOSTRESS land surface temperature and Emissivity Daily L2 global 70 m V001. <https://doi.org/10.5067/ECOSTRESS/ECO2LSTE.001>
- Huang, H.-Y., & Margulis, S. A. (2013). Impact of soil moisture heterogeneity length scale and gradients on daytime coupled land-cloudy boundary layer interactions. *Hydrological Processes*, 27(14), 1988–2003. <https://doi.org/10.1002/hyp.9351>
- Jarvis, P. G. (1976). The interpretation of the variations in leaf water potential and stomatal conductance found in canopies in the field. *Philosophical Transactions of the Royal Society of London B Biological Sciences*, 273(927), 593–610. <https://doi.org/10.1098/rstb.1976.0035>
- Jupp, T. E., & Twiss, S. D. (2006). A physically motivated index of subgrid-scale pattern. *Journal of Geophysical Research*, 111(19), 19112. <https://doi.org/10.1029/2006JD007343>
- Kang, S. L., & Bryan, G. H. (2011). A large-eddy simulation study of moist convection initiation over heterogeneous surface fluxes. *Monthly Weather Review*, 139(9), 2901–2917. <https://doi.org/10.1175/MWR-D-10-05037.1>
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 424–425. <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Koch, J., Mendiguren, G., Mariethoz, G., & Stisen, S. (2017). Spatial sensitivity analysis of simulated land surface patterns in a catchment model using a set of innovative spatial performance metrics. *Journal of Hydrometeorology*, 18(4), 1121–1142. <https://doi.org/10.1175/JHM-D-16-0148.1>
- Koren, V., Schaake, J., Mitchell, K., Duan, Q. Y., Chen, F., & Baker, J. M. (1999). A parameterization of snowpack and frozen ground intended for NCEP weather and climate models. *Journal of Geophysical Research*, 104(D16), 19569–19585. <https://doi.org/10.1029/1999JD900232>
- Kustas, W. P., & Albertson, J. D. (2003). Effects of surface temperature contrast on land-atmosphere exchange: A case study from monsoon 90. *Water Resources Research*, 39(6), 1159. <https://doi.org/10.1029/2001WR001226>
- Landry, J. S., Ramankutty, N., & Parrott, L. (2016). Investigating the effects of subgrid cell dynamic heterogeneity on the large-scale modeling of albedo in boreal forests. *Earth Interactions*, 20(5), 1–23. <https://doi.org/10.1175/EI-D-15-0022.1>
- Li, D., Bou-Zeid, E., Barlage, M., Chen, F., & Smith, J. A. (2013). Development and evaluation of a mosaic approach in the WRF-Noah framework. *Journal of Geophysical Research: Atmospheres*, 118(21), 11918–11935. <https://doi.org/10.1002/2013JD020657>
- Li, J., Miao, C., Zhang, G., Fang, Y. H., Shanguan, W., & Niu, G. Y. (2022). Global evaluation of the Noah-MP land surface model and suggestions for selecting parameterization schemes. *Journal of Geophysical Research: Atmospheres*, 127(5), 1–33. <https://doi.org/10.1029/2021JD035753>
- Li, K., Guan, K., Jiang, C., Wang, S., Peng, B., & Cai, Y. (2021). Evaluation of four new land surface temperature (LST) products in the U.S. Corn belt: ECOSTRESS, GOES-R, Landsat, and Sentinel-3. *Ieee Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14, 9931–9945. <https://doi.org/10.1109/JSTARS.2021.3114613>
- Loritz, R., Gupta, H., Jackisch, C., Westhoff, M., Kleidon, A., Ehret, U., & Zehe, E. (2018). On the dynamic nature of hydrological similarity. *Hydrology and Earth System Sciences*, 22(7), 3663–3684. <https://doi.org/10.5194/hess-22-3663-2018>
- Loritz, R., Hrachowitz, M., Neuper, M., & Zehe, E. (2021). The role and value of distributed precipitation data in hydrological models. *Hydrology and Earth System Sciences*, 25(1), 147–167. <https://doi.org/10.5194/hess-25-147-2021>
- Mälicke, M., Hassler, S. K., Blume, T., Weiler, M., & Zehe, E. (2020). Soil moisture: Variable in space but redundant in time. *Hydrology and Earth System Sciences*, 24(5), 2633–2653. <https://doi.org/10.5194/hess-24-2633-2020>
- Melton, J. R., & Arora, V. K. (2014). Sub-grid scale representation of vegetation in global land surface schemes: Implications for estimation of the terrestrial carbon sink. *Biogeosciences*, 11(4), 1021–1036. <https://doi.org/10.5194/bg-11-1021-2014>
- Melton, J. R., Sospedra-Alfonso, R., & McCusker, K. E. (2017). Tiling soil textures for terrestrial ecosystem modelling via clustering analysis: A case study with CLASS-CTEM (version 2.1). *Geoscientific Model Development*, 10(7), 2761–2783. <https://doi.org/10.5194/gmd-10-2761-2017>
- Milly, P. C. D., Malyshev, S. L., Shevliakova, E., Dunne, K. A., Findell, K. L., Gleeson, T., et al. (2014). An enhanced model of land water and energy for global hydrologic and Earth-system studies. *Journal of Hydrometeorology*, 15(5), 1739–1761. <https://doi.org/10.1175/JHM-D-13-0162.1>
- Mittelbach, H., & Seneviratne, S. I. (2012). A new perspective on the spatio-temporal variability of soil moisture: Temporal dynamics versus time-invariant contributions. *Hydrology and Earth System Sciences*, 16(7), 2169–2179. <https://doi.org/10.5194/hess-16-2169-2012>
- Newman, A. J., Clark, M. P., Winstral, A., Marks, D., & Seyfried, M. (2014). The use of similarity concepts to represent subgrid variability in land surface models: Case study in a snowmelt-dominated watershed. *Journal of Hydrometeorology*, 15(5), 1717–1738. <https://doi.org/10.1175/JHM-D-13-038.1>
- Niu, G. Y., & Yang, Z. L. (2004). Effects of vegetation canopy processes on snow surface energy and mass balances. *Journal of Geophysical Research*, 109(D23), 1–15. <https://doi.org/10.1029/2004JD004884>
- Niu, G. Y., Yang, Z. L., Dickinson, R. E., & Gulden, L. E. (2005). A simple TOPMODEL-based runoff parameterization (SIMTOP) for use in global climate models. *Journal of Geophysical Research*, 110(D21), 1–15. <https://doi.org/10.1029/2005JD006111>
- Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., et al. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *Journal of Geophysical Research*, 116(12), D12109. <https://doi.org/10.1029/2010JD015139>
- Ntelekos, A. A., Smith, J. A., Baeck, M. L., Krajewski, W. F., Miller, A. J., & Goska, R. (2008). Extreme hydrometeorological events and the urban environment: Dissecting the 7 July 2004 thunderstorm over the Baltimore MD metropolitan region. *Water Resources Research*, 44(8), 8446. <https://doi.org/10.1029/2007WR006346>
- Pan, M., Cai, X., Chaney, N. W., Entekhabi, D., & Wood, E. F. (2016). An initial assessment of SMAP soil moisture retrievals using high-resolution model simulations and in situ observations. *Geophysical Research Letters*, 43(18), 9662–9668. <https://doi.org/10.1002/2016GL069964>
- Sakaguchi, K., & Zeng, X. (2009). Effects of soil wetness, plant litter, and under-canopy atmospheric stability on ground evaporation in the community land model (CLM3.5). *Journal of Geophysical Research*, 114(D1), D01107. <https://doi.org/10.1029/2008JD010834>
- Salmun, H., & Molod, A. (2006). Progress in modeling the impact of land cover change on the global climate. *Progress in Physical Geography: Earth and Environment*, 30(6), 737–749. <https://doi.org/10.1177/0309133306071956>
- Senatore, A., Mendicino, G., Gochis, D. J., Yu, W., Yates, D. N., & Kunstmann, H. (2015). Fully coupled atmosphere-hydrology simulations for the central Mediterranean: Impact of enhanced hydrological parameterization for short and long time scales. *Journal of Advances in Modeling Earth Systems*, 7(4), 1693–1715. <https://doi.org/10.1002/2015MS000510>
- Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S., Yuan, X., et al. (2014). A drought monitoring and forecasting system for Sub-Saharan African water resources and food security. *Bulletin of the American Meteorological Society*, 95(6), 861–882. <https://doi.org/10.1175/BAMS-D-12-00124.1>
- Shrestha, P., Sulis, M., Masbou, M., Kollet, S., & Simmer, C. (2014). A scale-consistent terrestrial systems modeling platform based on COSMO, CLM, and ParFlow. *Monthly Weather Review*, 142(9), 3466–3483. <https://doi.org/10.1175/MWR-D-14-00029.1>

- Simon, J. S., Bragg, A. D., Dirmeyer, P. A., & Chaney, N. W. (2021). Semi-coupling of a field-scale resolving land-surface model and WRF-LES to investigate the influence of land-surface heterogeneity on cloud development. *Journal of Advances in Modeling Earth Systems*, 13(10), e2021MS002602. <https://doi.org/10.1029/2021MS002602>
- Sims, D. A., Rahman, A. F., Cordova, V. D., El-Masri, B. Z., Baldocchi, D. D., Bolstad, P. V., et al. (2008). A new model of gross primary productivity for North American ecosystems based solely on the enhanced vegetation index and land surface temperature from MODIS. *Remote Sensing of Environment*, 112(4), 1633–1646. <https://doi.org/10.1016/J.RSE.2007.08.004>
- Sivapalan, M. (2018). From engineering hydrology to Earth system science: Milestones in the transformation of hydrologic science. *Hydrology and Earth System Sciences*, 22(3), 1665–1693. <https://doi.org/10.5194/hess-22-1665-2018>
- Sobol, M. (1993). Sensitivity analysis for non-linear mathematical models. *Mathematical Modelling in Civil Engineering*, 1, 407–414. <https://ci.nii.ac.jp/naid/10027137978>
- Spruce, J. P., Gasser, G. E., & Hargrove, W. W. (2016). MODIS NDVI data, Smoothed and Gap-filled, for the Conterminous US: 2000-2015. ORNL DAAC. <https://doi.org/10.3334/ORNLDAAC/1299>
- Subin, Z. M., Milly, P. C. D., Sulman, B. N., Malyshev, S., & Shevliakova, E. (2014). Resolving terrestrial ecosystem processes along a subgrid topographic gradient for an Earth-system model. *Hydrology and Earth System Sciences Discussions*, 11(7), 8443–8492. <https://doi.org/10.5194/hessd-11-8443-2014>
- Sušelj, K., Teixeira, J., & Chung, D. (2013). A unified model for moist convective boundary layers based on a stochastic eddy-diffusivity/mass-flux parameterization. *Journal of the Atmospheric Sciences*, 70(7), 1929–1953. <https://doi.org/10.1175/JAS-D-12-0106.1>
- Swenson, S. C., Clark, M., Fan, Y., Lawrence, D. M., & Perket, J. (2019). Representing intrahillslope lateral subsurface flow in the community land model. *Journal of Advances in Modeling Earth Systems*, 11(12), 4044–4065. <https://doi.org/10.1029/2019MS001833>
- Talbot, C., Bou-Zeid, E., & Smith, J. (2012). Nested mesoscale large-eddy simulations with WRF: Performance in real test cases. *Journal of Hydrometeorology*, 13(5), 1421–1441. <https://doi.org/10.1175/JHM-D-11-048.1>
- Tesfa, T. K., Li, H.-Y., Leung, L. R., Huang, M., Ke, Y., Sun, Y., & Liu, Y. (2014). A subbasin-based framework to represent land surface processes in an Earth system model. *Geoscientific Model Development*, 7(3), 947–963. <https://doi.org/10.5194/gmd-7-947-2014>
- Teuling, A. J., Uijlenhoet, R., Hupet, F., van Loon, E. E., & Troch, P. A. (2006). Estimating spatial mean root-zone soil moisture from point-scale observations. *Hydrology and Earth System Sciences*, 10(5), 755–767. <https://doi.org/10.5194/hess-10-755-2006>
- Timmermans, W., Bertoldi, G., Albertson, J., Olioso, A., Su, Z., & Gieske, A. (2008). Accounting for atmospheric boundary layer variability on flux estimation from RS observations. *International Journal of Remote Sensing*, 29(17–18), 5275–5290. <https://doi.org/10.1080/01431160802036383>
- Torres-Rojas, L., & Chaney, N. W. (2022). Supporting data set: Towards the optimal representation of sub-grid heterogeneity in land surface models—Upper Colorado River Basin study case [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.7051439>
- Vergopolan, N., Chaney, N. W., Beck, H. E., Pan, M., Sheffield, J., Chan, S., & Wood, E. F. (2020). Combining hyper-resolution land surface modeling with SMAP brightness temperatures to obtain 30-m soil moisture estimates. *Remote Sensing of Environment*, 242, 111740. <https://doi.org/10.1016/J.RSE.2020.111740>
- Vergopolan, N., Sheffield, J., Chaney, N. W., Pan, M., Beck, H. E., Ferguson, C. R., et al. (2022). High-resolution soil moisture data reveal complex multi-scale spatial variability across the United States. *Geophysical Research Letters*, 49(15), e2022GL098586. <https://doi.org/10.1029/2022GL098586>
- Vergopolan, N., Xiong, S., Estes, L., Wanders, N., Chaney, N. W., Wood, E. F., et al. (2021). Field-scale soil moisture bridges the spatial-scale gap between drought monitoring and agricultural yields. *Hydrology and Earth System Sciences*, 25(4), 1827–1847. <https://doi.org/10.5194/HESS-25-1827-2021>
- Verseghy, D. L. (1991). Class—A Canadian land surface scheme for GCMS. I. Soil model. *International Journal of Climatology*, 11(2), 111–133. <https://doi.org/10.1002/joc.3370110202>
- Wainwright, H. M., Uhlemann, S., Franklin, M., Falco, N., Bouskill, N. J., Newcomer, M. E., et al. (2022). Watershed zonation through hillslope clustering for tractably quantifying above-and below-ground watershed heterogeneity and functions. *Hydrology and Earth System Sciences*, 26(2), 429–444. <https://doi.org/10.5194/hess-26-429-2022>
- Wang, K., & Dickinson, R. E. (2012). A review of global terrestrial evapotranspiration: Observation, modeling, climatology, and climatic variability. *Reviews of Geophysics*, 50(2), 2005. <https://doi.org/10.1029/2011RG000373>
- Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., et al. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, 47(5), W05301. <https://doi.org/10.1029/2010WR010090>
- Xia, Y., Cosgrove, B. A., Mitchell, K. E., Peters-Lidard, C. D., Ek, M. B., Brewer, M., et al. (2016). Basin-scale assessment of the land surface water budget in the National Centers for Environmental Prediction operational and research NLDAS-2 systems. *Journal of Geophysical Research: Atmospheres*, 121(6), 2750–2779. <https://doi.org/10.1002/2015JD023733>
- Yang, R., & Friedl, M. A. (2003). Modeling the effects of three-dimensional vegetation structure on surface radiation and energy balance in boreal forests. *Journal of Geophysical Research*, 108(D16), 8615. <https://doi.org/10.1029/2002JD003109>
- Yu, Y., Tarpley, D., Hui, X., & Chen, M. (2010). GOES-R advanced baseline imager (ABI) algorithm theoretical basis document for land surface temperature. *September*, 1–60.
- Yuan, H., Dai, Y., Xiao, Z., Ji, D., & Shangquan, W. (2011). Reprocessing the MODIS Leaf Area Index products for land surface and climate modelling. *Remote Sensing of Environment*, 115(5), 1171–1187. <https://doi.org/10.1016/J.RSE.2011.01.001>