The observed spatio-temporal patterns of Land surface temperature over the Contiguous United States

Laura Torres-Rojas[1], Tyler Waterman[2], Jiaxuan Cai[1], Enrico Zorzetto[3], Haruko Murakami Wainwright[4], and Nathaniel W. Chaney[1]

[1]Duke University
[2]Duke University Civil and Environmental Engineering
[3]Princeton University
[4]Lawrence Berkeley National Laboratory

October 30, 2023

The observed spatio-temporal patterns of Land surface temperature over the Contiguous United States

L. Torres-Rojas[1], T. Waterman[1], J. Cai[1], E. Zorzetto[2], H.M. Wainwright[3,4] , and N. W. Chaney[1]

[1] Department of Civil and Environmental Engineering, Duke University, Durham, NC, United States of America.

[2] Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, United States of America.

[3] Earth and Environmental Sciences Area, Lawrence Berkeley National Laboratory, Berkeley, CA, United States of America.

[4] Massachusetts Institute of Technology, Cambridge, MA, United States of America.

Corresponding author: Laura Torres-Rojas (laura.torres@duke.edu)

Key Points:

- The Empirical Spatio-Temporal Covariance Function is a flexible tool to describe the spatiotemporal dependence structure of observed fields.
- A parametric covariance model can concisely describe the spatio-temporal patterns captured from empirical data.
- Multidimensional clustering can be used to identify areas with similar spatiotemporal dependence structures over continental scales.

Abstract

Surface fluxes and their related processes and states tend to recur and remain consistent across various spatial and temporal scales forming patterns. For multiple applications, identifying spatio-temporal patterns is desirable, as they provide information about the dynamics of the processes involved. This is especially true for land surface temperature, a key variable that plays a primary role in the energy and water exchange between land and atmosphere. This study introduces the Empirical Spatio-Temporal Covariance Function (ESTCF) as a tool to identify and characterize spatio-temporal patterns in remotely sensed land surface temperature fields. The method is demonstrated over the Contiguous United States by splitting the entire area into 1.0°x1.0° domains. The summer day-time surface temperature ESTCFs are derived for each domain, and a parametric covariance model is fitted. Clustering analysis is then applied to detect areas with similar spatio-temporal land surface temperature dynamics. The results are assessed to determine if particular spatio-temporal features are present in domains where landscape characteristics make interactions with the atmosphere likely. The proposed tool accurately characterizes the spatio-temporal interdependence of the fields, summarizing features such as spatio-temporal variance, spatial coherence structure, temporal persistence, and space-time interactions. The increased temporal persistence and space-time interaction drive the grouping in mountainous and coastal domains. The tools introduced here provide a pathway to formally identify and summarize the spatio-temporal patterns observed in remotely sensed fields and relate those to more complex processes within the Soil-Vegetation-Atmosphere System.

Plain Language Summary

Specific processes on Earth's surface, like heat and water fluxes and air movement, often exhibit coherent patterns across space and time. Figuring out where these patterns occur can be helpful as they can help explain how the underlying physical processes work. This research introduces a new tool to find and describe these patterns and applies it to temperature data derived from satellites. The United States is divided into smaller areas, and the developed tool is used to analyze the land temperature data within those areas; then, regions with similar temperature patterns are identified. These results help to determine if certain landscape features, like coastlines, mountains, and cities, influence how temperature patterns behave. Ultimately, the method can accurately describe how the temperature patterns vary over time, space, and both. As expected, the results show that places with similar temperature patterns are often near coasts and high mountains due to how temperature changes over time and interacts with its surroundings. The tools presented here are a step toward better understanding how air, water, and heat move near the Earth's surface, how they distribute in space, and how they change over time.

1 Introduction

Processes at the land-atmosphere interface are mainly driven by water and energy fluxes at the land surface, which are mediated by vegetation (Koch et al., 2017; Simon et al., 2021), land use, and topography. These surface fluxes have proved to control the overlying atmospheric distributions of water vapor, temperature, precipitation, and cloud properties, modulating the hydrological cycle and the surface energy budget (Dickinson, 1995; Dirmeyer et al., 2013; Y. Wang et al., 2023; Wu et al., 2015). More specifically, the multi-scale spatial heterogeneity of the physical environment (e.g., vegetation, soils, elevation, and land use) has been acknowledged to influence the spatial and temporal distribution of the fluxes in a nonlinear manner (Dickinson, 1995; R. A. Fisher & Koven, 2020; Koch et al., 2017; Nicholson, 1988; Simon et al., 2021; Tesfa et al., 2014; Torres-Rojas et al., 2022; Vergopolan et al., 2022). Besides, it is widely recognized that certain fluxes and associated processes and variables tend to recur and appear consistently across different scales in space, time, or both. These recurrent attributes are commonly referred to as patterns, and they can be a consequence of the self-organization of the systems and the organization in the systems' controlling factors (i.e., influence of the physical environment) (Koch et al., 2017; Vereecken et al., 2016).

The dynamic processes producing patterns expand over a wide range of spatial and temporal scales and encompass all compartments of the soil-vegetation-atmosphere system (SVAS). Some processes are quasi-static in nature (e.g., bedrock generation, rising and sinking motions of the Earth's mantle, and soil generation (Jenny, 1941)). On the global and synoptic scales, quasi-static surface patterns such as land-sea distribution and orography are known to control the spatial distribution of atmospheric variables such as precipitation, pressure, and temperature (Vereecken et al., 2016). From the meso- to the micro-scale, topography also induces cloud and precipitation patterns and orography-following flow patterns (e.g., Lee-wave clouds and valley-slope wind systems) (Brunsell & Gillies, 2003; Paleri et al., 2022). Furthermore, heterogeneity in vegetation and soil type distributions may cause patterns in the surface fluxes, leading to lake breeze circulation systems that can impact cloud cover and precipitation (Mahfouf et al., 1987; Nair et al., 2011). Finally, short-term atmospheric systems can directly influence the soil and vegetation states by determining the soil moisture and temperature distributions through precipitation and evapotranspiration patterns. Soil moisture patterns are constrained by hydrological processes such as infiltration and runoff, dependent on the quasi-static soil properties. Soil temperature is also directly intertwined with soil moisture patterns and quasi-static soil thermal properties. Interestingly, soil moisture and temperature patterns also play a crucial role in defining atmospheric stability and available moisture for precipitation, making them critical in the two-way interactions between the atmosphere and the land surface (Ferguson et al., 2012; Ferguson & Wood, 2011; Levine et al., 2016; Phillips et al., 2017; Taylor et al., 2013; Tuttle & Salvucci, 2016). In general, patterns can be detected in a variety of spatio-temporal fields, including hydraulic properties in soils (Chaney et al., 2016; Gueting et al., 2015; Qu et al., 2014), surface soil moisture and soil temperature (Martini et al., 2015; Poltoradnev et al., 2016; Seyfried et al., 2016; Vergopolan et al., 2021), latent and sensible heat (Jung et al., 2011; Simon et al., 2021), convection-induced atmospheric boundary layer (ABL) circulations (Taylor et al., 2007, 2011), and vegetation properties and states (Van der Putten et al., 2013).

Several approaches have been developed to identify, summarize, and extract relevant patterns from spatio-temporal geophysical datasets. The reader is referred to (Cressie & Wikle, 2015; Vereecken et al., 2016) for a detailed exploration of the approaches, their advantages, and their information content. Some techniques are based on decomposing the spatial and temporal signals according to their statistics or scales. Examples of decomposition approaches include the Empirical Orthogonal Function (EOF) method, Principal Component Analysis (PCA), Orthogonal Probability Density Function Decomposition (OPDFD), Wavelet Transform (WT), and Empirical Mode Decomposition (EMD). These methods can be applied in both space and time, and they have proved to be helpful in the simplification of complex datasets, the decomposition and identification of relevant temporal signals, and the determination of critical scales of processes within the SVAS (Biswas, 2014; Z. Fang et al., 2015; Graf et al., 2012; Katul et al., 2001; Katul & Parlange, 1995; Kim & Barros, 2002; Koch et al., 2015; Korres et al., 2010; Rudi et al., 2010; Stoy et al., 2005; Vargas et al., 2010; D. Wagner et al., 1990). However, in general, they are only meant to analyze time and frequency dimensions.

For climate, environmental, and hydrological applications, the simultaneous identification of spatially coherent persistent structures (i.e., spatio-temporal patterns) of the relevant state variables is desirable, as they provide information about the dynamics of the processes affecting them. In other words, the signatures of spatio-temporal processes in these variables are not concrete, independent objects or events but patterns appearing and evolving simultaneously over space and time (Faghmous & Kumar, 2014). For instance, let us consider the distribution of inundation in a flooding-prone area. Multiple processes determine the inundation dynamics within a watershed, including the spatio-temporal distribution of precipitation, watershed-distributed physical characteristics (i.e., soil properties, antecedent soil moisture content), and human modifications. The spatio-temporal evolution of the flooding as it moves downstream due to re-infiltration, evaporation, preferential flow, and other processes would be missed by looking only at the spatial or temporal dimensions independently (Cressie & Wikle, 2015). Another case where the simultaneous space-time evolution of variables is pivotal for process understanding is the initiation of heterogeneity-driven circulations at multiple scales. Under favorable ABL and synoptic background conditions, larger spatial scales of surface heterogeneity can generate temporally persistent structures of surface heating and moisture, initiating circulations that for large enough scales, can produce areas of shallow or even deep convection (F. Chen & Avissar, 1994; Cheng & Cotton, 2004; Courault et al., 2007; Gentine et al., 2019; Pielke, 2001; Taylor et al., 2007; Weaver, 2004; Wu et al., 2015). The two previous examples highlight the need for approaches that succinctly and effectively identify and summarize the spatio-temporal patterns observed in climatic, environmental, and hydrological datasets.

The geostatistics field has been concerned with addressing variables varying in space and time for decades, primarily due to the improvement in predictions obtained by including correlations in two dimensions instead of a single one (Cressie & Huang, 1999; Lee et al., 2010; Rodríguez-Iturbe & Mejía, 1974; Rouhani & Myers, 1990; Varouchakis, 2018). Even though geostatistics methods can make significant assumptions about the characteristics of the spatio-temporal fields and associated patterns (i.e., stationarity, spatio-temporal dependence, isotropy, and homoscedasticity, among others), they are still widely used as compact and straightforward evaluation tools of the structure of the observation fields. Moving window sampling techniques have been implemented to deal with the stationarity assumption (i.e., while environmental phenomena exhibit heterogeneity in both their mean and covariance structure, it is often possible to regard the process as approximately homogeneous within subregions) (Guttorp & Sampson, 1994; Haas, 1990a, 1990b; Risser & Turek, 2020). Among the tools developed for geostatistical analysis, the empirical spatio-temporal covariance function (ESTCF) stands out for its simplicity. Under the assumptions of second-order stationarity (i.e., the covariance between two points is the same for a given distance and direction, regardless of which two points are chosen) in space and time and isotropy in space, the ESTCF can be estimated directly from the observed data, providing a measure of the strength and structure of dependence between different locations and time points (W. Chen et al., 2021; Cressie & Huang, 1999; Faghmous & Kumar, 2014; Gneiting, 2002; Stein, 2005). The ESTCF is also able to capture various forms of dependence, such as spatial correlation, temporal correlation, and spatio-temporal interactions (Cressie & Huang, 1999; Gneiting, 2002; Guttorp & Sampson, 1994; Ma, 2003; Stein, 2005). Finally, the method can handle irregularly sampled data or missing values more effectively than other approaches (e.g., spectral analysis), making it more suitable for real-world applications where data may be sparse or irregularly collected (Demel & Du, 2015; Montero et al., 2015; Stein, 1999). Once the ESTCF is computed, a parametric class of covariance model can be selected, and the parameters estimated by fitting the model to the empirical function. This procedure allows learning about the spatio-temporal properties and interactions of the original field from the estimated parameters (W. Chen et al., 2021; Gneiting, 2002).

Traditional spatio-temporal geostatistical methods use a set of spatially-distributed in-situ measurements to model the variation of the field of values as a function of the distance between locations, reflecting Tobler's first law of geography: "Everything is related to everything else, but near things are more related than distant things." (Tobler, 1970; Vereecken et al., 2016). However, for regional and continental scales, widespread spatially distributed in situ observations of surface fluxes and states do not exist (Stisen et al., 2011, 2021; Vereecken et al., 2008; Zink et al., 2018). Hence, satellite remote sensing remains the only direct source of spatially distributed Earth surface observations. Although the quantitative precision of this data is still hard to determine, its main asset is its spatial and temporal information content over extensive domains (Crow et al., 2009; H. T. Li et al., 2009; Stisen et al., 2011). Currently, sensors onboard satellites provide spatially distributed estimates of several surface states, including land surface temperature (LST) (L. Fang et al., 2014; Shi & Bates, 2011; Wan, 1996, 2014; Yu et al., 2012), soil moisture content (Chan et al., 2018; Entekhabi et al., 2010; Kerr et al., 2012; Parinussa et al., 2015; W. Wagner et al., 2013), evapotranspiration (Boschetti et al., 2019; J. B. Fisher et al., 2020; Martens et al., 2017; Running et al., 2019; Su, 2002), snow cover fraction (Painter et al., 2009; Tsai et al., 2019), and changes in water storage (Tapley et al., 2004). Therefore, the joint use of the ever-growing available remote sensing spatio-temporal data and spatio-temporal geostatistics methods provides a promising path forward in the multi-scale characterization of processes' heterogeneity and dynamics in multiple SVAS compartments.

A promising and relatively unexplored source of remote sensing data for the analysis of spatio-temporal patterns of surface states on sub-diurnal scales is LST (Duffy et al., 2022; Koch et al., 2016; Zink et al., 2018). As a critical state variable of the land surface, LST encodes information on local energy and water fluxes, including energy partitioning into sensible and latent heat fluxes (Duffy et al., 2022; Holzman et al., 2014; K. Li et al., 2021; Sims et al., 2008; K. Wang & Dickinson, 2012). This information is vital as energy partitioning can affect the state of the atmosphere by supplying water vapor, inducing convection and lateral convergence, and growing the planetary boundary layer (Levine et al., 2016; Pielke, 2001; Tuttle & Salvucci, 2016). Recently, enhanced spatio-temporal resolution global LST products have been released, including the ECOsystem Spaceborne Thermal Radiometer on the International Space Station (ECOSTRESS) (Hook & Hulley, 2019), Landsat Provisional (Anderson et al., 2012), the National Oceanic and Atmospheric Administration (NOAA) Geostationary Operational Environmental Satellites (GOES) (L. Fang et al., 2014; Yu et al., 2012), and Sentinel-3 (Polehampton et al., 2022). However, increased spatial resolutions are often not accompanied by enhanced temporal resolutions. Products derived from sensors onboard satellites with polar orbits, such as NASA's Moderate Resolution Imaging Spectroradiometer (MODIS), are accurate and extensively validated. However, due to the nature of the satellite orbit and the intermittency of the revisit times, the temporal resolution of the resulting products is limited, and diurnal and sub-diurnal variations cannot be captured. On the other hand, sensors onboard geostationary satellites (e.g., GOES) remain in fixed positions overlooking the Earth, providing full disk observations every 10 minutes. Geostationary satellites also provide increased robustness to cloud coverage (Duffy et al., 2022; Hashimoto et al., 2021), a desirable feature for the study of atmospheric motions generated by landscape discontinuities (F. Chen & Avissar, 1994).

This study introduces the Empirical Spatio-Temporal Covariance Function (ESTCF) as a tool to assess the spatial coherence and memory of remotely sensed spatio-temporal fields and identify patterns that might be relevant in the dynamics of processes within the SVAS. Additionally, the study presents a new 4-parameter covariance model to summarize the spatio-temporal structure displayed by the ESTCF and provides physical interpretations for its parameters. These tools are applied to remotely sensed fields of LST to evaluate whether they can identify areas where landscape features (e.g., coastlines, topographic gradients, and urban areas, among others) might be responsible for triggering heterogeneity-driven atmospheric circulations. These developments are implemented and tested over CONUS by splitting the entire country into a mosaic of 1.0°x1.0° domains, deriving the summer day-time surface temperature ESTCF for each domain independently, and fitting the parametric covariance model to each specific domain. Once covariance functions for all domains are known, clustering analysis is applied to the obtained parameter maps in order to detect areas with similar spatio-temporal surface temperature dynamics. Finally, we present a combined metric for quantifying the local spatio-temporal variability based on normalized values of the covariance parameters. The key developments in this study include (a) a flexible and comprehensive tool to characterize and represent the spatio-temporal dependence structure of remotely sensed fields, (b) a parametric covariance function model to more concisely describe the spatio-temporal patterns captured with the ESTCF, and (c) a multi-dimensional clustering approach to determine areas with similar spatio-temporal dependence structures. The tools introduced here provide a pathway forward to formally identify and summarize the spatio-temporal patterns observed in remotely sensed fields and relate those to the footprint of more complex dynamic processes within the SVAS.

2 Data and Methods

2.1 GOES-16 LST and Sea Surface Temperature (SST)

The NOAA's Geostationary Operational Environmental Satellites (GOES) are the latest and main operational geostationary weather satellites in orbit over the Western hemisphere (Desai et al., 2021). Recently, the GOES-R Advanced Baseline Imagers (ABIs) on board the new generation GOES-16 and GOES-17 satellites have been generating an LST operational product based on scans at roughly 5 minutes with an approximate 2 km spatial granularity over the continental United States (CONUS). The GOES ABI LST estimates are produced using a thermal channel split-window retrieval based on the bands centered at 10.8 and 12.3 μm with high surface emission and low atmospheric absorption. Additionally, the algorithm uses a prescribed surface emissivity and an atmospheric radiative transfer model. For further details on the retrieval algorithm, the reader is referred to (Yu & Yu, 2020). The final operational product has been generated at an hourly time scale from May 2017 to the present. Evaluations have shown that the product is high quality, with validation studies indicating an approximate accuracy of 1.5K (Chang et al., 2021; Desai et al., 2021; Yu et al., 2012).

Figure 1 shows two locations over CONUS and the temporal evolution of the retrieved GOES-16 LST over several daytime hours in the summer. The figure demonstrates how GOES-

16 LST data can capture differential heating of the surface due to landscape features (e.g., mountains and lakes in domain 1 and urban areas in domain 2) with relatively high frequency. Additionally, it demonstrates how the spatial and temporal evolution of LST is heavily site-specific. Given the characteristic spatial and temporal scales reported for mesoscale heterogeneity-driven circulations, it is expected that GOES-16 provides an observational source with both sufficiently high spatial resolution (i.e., ~2km over. CONUS) and high temporal resolution (i.e., one hour) to perform the subsequent analyses.

A Sea Surface Temperature (SST) product is also produced from the ABI retrievals on board the GOES satellites. The ABIs on board GOES-16 and 17 offer improved capabilities for SST retrievals, over its predecessors, including five narrow bands that can be used to estimate SST. Other advantages include accurate sensor calibration, image navigation and co-registration, spectral fidelity, and sophisticated preprocessing. Using this information, the Level 2 Preprocessed (L2P) SST product is derived at the native sensor resolution (2 km at nadir, degrading to 15 km at view zenith angle, 67°) using NOAA Advanced Clear-Sky Processor for Ocean (ACSPO) system (Ignatov et al., 2019). SST is derived from the original 10-minute full-disk brightness temperatures using the ACSPO clear-sky mask (Petrenko et al., 2010) and the Non-Linear SST algorithm (Petrenko et al., 2014). Four (4) longwave bands centered at 8.4, 10.3, 11.2, and 12.3 µm are used. The regression is calibrated against quality-controlled in situ SST observations from drifting and tropical mooring buoys in the NOAA iQuam system (Xu & Ignatov, 2014). Finally, the 10-minute full-disk data is unified in time to produce the 1-hour L2P product, with improved coverage and reduced cloud leakages and image noise, compared to each 10-minute image.

To explore the spatio-temporal patterns of remote sensing land surface temperature that can lead to the development of heterogeneity-driven atmospheric circulations, coastal regions are relevant (e.g., land-sea breezes are one of the most evident examples of an increased land-atmosphere coupling strength). In this sense, a LST product alone is insufficient to perform the analysis; an SST product must also be used. For this study, the hourly GOES-16 LST data over CONUS is superimposed to the hourly GOES-16 SST data over the Americas region from January 2018 to December 2022. The resulting 1-hour, 2-kilometer, CONUS-wide surface temperature dataset is then bounded to only consider pixels containing at least 30% of land in their area. This dataset is then used to determine the spatio-temporal dependence structure of the LST fields in different domains. The obtained structures are expected to show consistent behaviors in places where landscape features can contribute to generating heterogeneity-driven circulations. We acknowledge that differences in the algorithms used to retrieve SST and LST might generate inconsistencies in the values between water and land in the consolidated LST dataset. However, the main reason for merging the data is to analyze the contrasting temperatures in coastal patches that would be impossible to analyze using an LST product alone.

It is well established that atmospheric motions influenced by landscape discontinuities develop mainly during summer daytime hours and are optimum under clear sky conditions (F. Chen & Avissar, 1994). For this reason, analyses in this study use only warm months (i.e., June, July, August, and September), daytime hours, and clear-sky pixels of the LST dataset. Daytime hours are determined locally for each individual domain over CONUS (see Section 2.2) as the period between two (2) hours after sunrise and two (2) hours before sunset. The raw GOES-LST and GOES-SST datasets are provided in the native ABI fixed grid coordinates; therefore, reprojection to the WGS84 projection (i.e., EPSG:4326) is implemented before further analyses.

2.2 Moving Window Sampling

A sliding window approach is applied to the constructed LST dataset over CONUS to deal with the inherent limitations of the stationarity assumption of the selected geostatistics method. The approach works by first defining a domain of size 1.0°x1.0° and then moving it over the remotely sensed field by a distance of 0.25° in the vertical and horizontal directions, as shown in Figure 2b. The 1.0°x1.0° box size is determined as a typical resolution used in ESMs and General Circulation Models (GCMs). For each position of the box, the whole spatio-temporal field of observations over summer daytime is retrieved. By adopting this approach, a comprehensive analysis is performed as different combinations of landscape features are considered, and it can be assumed that the stationarity assumption holds if LST is characterized as approximately homogeneous within the subregions. Results obtained in the following steps are also expected to be smooth due to the approach. Figure 2a shows the study domain used over CONUS. The central 0.25° of each 1.0°x1.0° squared box obtained from the sliding window approach is presented as the grid. Coordinates every 5° are shown to aid in georeferencing.

2.3 Empirical Spatio-Temporal Covariance Function - ESTCF

The summer, daytime empirical spatio-temporal covariance function (ESTCF) of LST is computed for 1.0°x1.0° domains across the country (see Section 2.2 for details on the domain definition). The objective is to summarize and characterize the spatio-temporal dynamics of the LST fields in domains across the country.

The ESTCF expresses how the linear statistical dependence of two measurements in a spatio-temporal random field reduces as the distances (in space and time) between them increase, up to the lengths of statistical independence where a relation no longer exists and the covariance tends to zero (Cressie & Wikle, 2015; Mälicke et al., 2020). The spatio-temporal dependence structure displayed by the observed realizations is summarized using the ESTCF. The mathematical procedure used to compute the ESTCF for a random field is presented next and explained based on (Montero et al., 2015).

Let $Z(\cdot,\cdot)$ be an intrinsically stationary process observed on a set of n spatio-temporal locations $\{(s_1,t_1),\ldots,(s_n,t_n)\}$ where s_i is the spatial location and t_j the observation time. The classical alternative to estimate the empirical covariance function using the observed values if the process is second-order stationary is proposed by (Matheron, 1989). This classical estimation is obtained by implementing the Method-of-Moments estimator (MoM), which for the covariance function takes the form:

$$\hat{C}\big(h(l),\tau(k)\big) = \frac{1}{\#N(h(l),\tau(k))} \sum_{\substack{(s_i,t_i),(s_j,t_j) \\ \in N(h(l),\tau(k))}} \left(Z(s_i,t_i) - \bar{Z}_{t_i}\right)\left(Z(s_j,t_j) - \bar{Z}_{t_j}\right) \quad (1)$$

Where $\bar{Z}_i = \frac{1}{n}\sum_m Z(s_m,t_i)$ is an estimator of the mean μ_i of the random field for the time t_i and $N\big(h(l),\tau(k)\big) = \big\{(s_i,t_i)(s_j,t_j) : s_i - s_j \in T\big(h(l)\big),\ t_i - t_j \in T\big(\tau(k)\big)\big\}$, with $T\big(h(l)\big)$ being a tolerance region on $\mathbb{R}^d$ around $h(l)$, $T\big(\tau(k)\big)$ being a tolerance region on $\mathbb{R}$ around $\tau(k)$, and $\#N(h(l),\tau(k))$ the number of different elements in $N(h(l),\tau(k))$, with $l = 1,\dots,L$ and with $k = 1,\dots,K$.

In general, the areas $T\big(h(l)\big)$ and $T\big(\tau(k)\big)$ are chosen to yield disjoint sets with enough elements to generate stable estimates (i.e., domain sizes with enough observations to generate an adequate estimation of the ESTCF). Suppose the hypothesis of isotropy is reasonable for the spatial process under analysis. In that case, the area of spatial tolerance around each of the values $h(l)$ can be defined as $[h(l) - d_l/2,\ h(l) + d_l/2]$, with d_l being the spatial tolerance to be used. Also common is to make the temporal component take values in $\mathbb{Z}$, in which case the empirical covariance function is computed for $\tau(k) = 0, 1, \ldots$, obtained as the subsequent differences in time at which the process is observed. To better illustrate the described procedure, a simplified example is included next.

Let us suppose there is a set of spatio-temporal measurements taken in three (3) moments, t_1, t_2, and t_3, on a regular grid of size 3x3 with a spacing of 2 kilometers (Figure 3a). Assuming the resulting spatio-temporal random field is isotropic and stationary and that tolerance regions are not used, the classical ESTCF will be given by the simplified form of Equation 1:

$$\hat{C}(h,u) = \frac{1}{\#N(h,\tau)} \sum_{N(h,\tau)} \left(Z(s_i,t_i) - \bar{Z}_{t_i}\right)\left(Z(s_j,t_j) - \bar{Z}_{t_j}\right) \quad (2)$$

In this example, it is easy to show that there are $9 \times 3 = 27$ spatio-temporal points at a distance $(h,\tau) = (0km, 0)$. Therefore, for Equation 2, $\#N(h,\tau) = 27$. By definition, $\hat{C}(0km, 0)$ is the spatio-temporal variance of the random field, σ^2.

Suppose the distance is $(h,\tau) = (0km, 1)$, then $\#N(h,\tau) = 9 \times 2 = 18$. Finally, if the distance is $(h,\tau) = (0km, 2)$, $\#N(h,\tau) = 9 \times 1 = 9$. If the empirical covariance is computed for all the previously defined spatio-temporal distances, the purely temporal empirical covariance function is obtained (i.e., only time varies). It is also trivial to prove that: $\#N(2km, 0) = 12 \times 3 = 36$ and $\#N(4km, 0) = 6 \times 3 = 18$. If the empirical covariance is computed just for the distances $(h, 0)$, the purely spatial empirical covariance is obtained. Additionally, cases where both the spatial and temporal lags are different from zero, can also be considered, though the pairs of points must be determined carefully. For instance, Figure 3b shows the 12 pairs of points separated by a spatio-temporal distance $(h, u) = (4km, 2)$. It is important to mention that this procedure does not consider diagonal spatial distances for simplicity.

Since the produced LST dataset already uses a regular spatio-temporal grid, the procedure described for the example is directly applied. As mentioned before, only clear sky, summer daytime pixels are used for the analysis. If one or both points contained in any pair used to compute the ESTCF contain a missing value, the pair is ignored from the summation in Equation 2, and $\#N$ is modified accordingly by subtracting one. Additionally, to determine the spatial separation between points, h, each pixel is assigned its central coordinates in degrees..

Two conditions are implemented regarding the spatial distribution and number of missing values of the LST dataset within each domain for every time step: i) the average latitude and longitude of the LST valid pixels within the domain has to be within a range of ±0.15° of the central latitude and longitude of box; otherwise, the time step LST values for the domain are considered missing and; ii) the fraction of missing LST values within each domain per time step cannot be higher than 0.25; otherwise, the time step values are considered missing. Finally, some ESTCF results presented in the following sections are normalized by the spatio-temporal variance of the field over the local domain (σ^2), as a way to make results comparable and analyze differences beyond the magnitude of the spatio-temporal variance.

2.4 A Parametric Model for the Spatio-Temporal Covariance Function

Once the ESTCF is computed for every domain in CONUS, a parametric model is fitted to it, and the obtained parameters are used to determine the similarity in the spatio-temporal dependence structures between domains.

Typically, the end goal when estimating the ESTCF of a random field is to use the information contained within the observations to perform prediction of values for unobserved locations (i.e., spatio-temporal kriging prediction or modeling). In such cases, a positive definite covariance function is a requirement to define a valid stochastic process. However, the spatio-temporal dependence structure derived from observations (i.e., ESTCF) usually does not fulfill the condition of being positive-definite. For this reason, in practice, a parametric model of covariance that is already known to be valid is selected and fitted to the ESTCF. An extensive body of literature deals with proper covariance models for spatio-temporal prediction. For a comprehensive review of these models, the reader is referred to (Cressie & Huang, 1999; Cressie & Wikle, 2015; Montero et al., 2015; Stein, 2005). Two main types of theoretical spatio-temporal covariance models exist: separable and non-separable. Separable models are built using the sum or product of a purely spatial and purely temporal covariance, both stationary. In this sense, separable models do not consider interactions between space and time in the dependence structure of the field. Non-separable models, on the other hand, capture the space-time dependence that exists on most phenomena by including the interaction between the two dimensions.

For this study, the end goal is not predicting LST at ungauged locations but rather the characterization of the spatio-temporal dynamics of the LST fields. In this sense, the positive-definite nature of the selected parametric model is not a requirement. However, it is desirable that the chosen model uses a small set of parameters and that each is physically meaningful. For this reason, a modified form of the non-separable parametric model presented by (Cressie & Huang, 1999) is selected:

$$C(h,u) = \sigma^2 e^{-\left(\frac{\tau^a}{\gamma}\right)-\left(\frac{h^a}{\lambda}\right)} \qquad (3)$$

Where $C(h,u)$ is the parametric covariance at a spatio-temporal distance (h,u); σ^2 is the spatio-temporal variance of the random field, computed directly from the data; γ is the fitted temporal characteristic length-scale; λ is the fitted characteristic spatial length-scale; and a is the fitted spatio-temporal interaction exponent. Several other parametric models were tested on the available data, but the one selected showed improved performance with the lowest number of parameters.

2.5 Impact of the Parameters in the Spatio-Temporal Covariance Parametric Model

In the selected model, the space-time interaction exponent (a) determines the shape of the space-time interaction, while the characteristic length scales (i.e., γ and λ) modify the magnitude of the spatial and temporal distances after they have been affected by the exponent. Figure 4 displays the spatio-temporal covariance functions obtained as the values of the parameters are successively modified. The figure allows us to see how:

- Larger values of the characteristic temporal length scale, γ, result in higher memory (see Figure 4c). In other words, for the produced parametric model, high covariance values are bound to persist longer in time as γ increases. When fitted to an observational dataset, a high value of γ implies that the variable patterns tend to remain for longer for the location. The opposite behavior (i.e., shorter persistence) can also be achieved by reducing γ, as observed in Figure 4a.

- Increased spatial persistence of the modeled covariance function is achieved by raising the spatial length scale, λ (see Figure 4f). A fitted high λ suggests a significant extent of the spatial patterns in the domain. In that case, the values of the observations for two points far from each other are highly correlated. The contrasting case (i.e., smaller spatial patches of values) can be modeled by decreasing λ, as presented in Figure 4d.

- Higher values of the space-time interaction exponent, a, lead to an increased interaction between space-time in the computed covariance function. The space-time interaction is directly related to the shape of the curves in the modeled covariance function. As a rises over one (see Figure 4i), the interaction between space and time becomes stronger, and the modeled transition between the pure-spatial and pure-temporal covariance occurs by displaying a convex shape. The opposite case (i.e., $a < 1.0$) generates a concave shape in the transition between pure-spatial and pure-temporal covariance, as seen in Figure 4g. Additionally, due to the function structure, the modeled memory and spatial coherence emulated by the parametric model are also directly influenced by the magnitude of the spatio-temporal interaction exponent, a. In each dimension, the presence of the exponent determines a stretched exponential shape for the correlation function, which encompasses longer tails ($a < 1.0$) or shorter tails ($a > 1.0$) compared to a simple exponential, while retaining a characteristic scale (Laherrère & Sornette, 1998).

In general, it is expected that once applied to the LST dataset, domains with high values of the two characteristic length-scale parameters and the space-time interaction exponent (i.e., a over 1) display the coherent memory associated with the initiation of heterogeneity-driven circulation systems.

Additionally, modified forms of the spatial and temporal characteristic length scales can be derived to obtain an approximation unaffected by the spatio-temporal interaction exponent and identify the individual effects of space and time on the joint spatio-temporal dynamics. These forms are estimated as the fitted characteristic length scales operated by the fitted spatio-temporal interaction exponent, as displayed in Equations 4 and 5.

$$\gamma' = \sqrt[a]{\gamma} \qquad (4)$$

$$\lambda' = \sqrt[a]{\lambda} \qquad (5)$$

To fit the ESTCFs to the parametric model, non-linear least squares regression is used. The function is set up to use the Trust Region Reflective method, TRF, to perform the minimization. The TRF method is particularly suitable for large sparse problems with bounds, and it is generally robust. Based on an analysis of the selected function, the bounds for the parameters to be fitted are defined as $0 < \gamma \leq 100$, $0 < \lambda \leq 10$, and $0.5 < a \leq 3$.

2.6 Clustering Analysis

Once the parametric model is fitted to the ESTCFs, an unsupervised clustering algorithm determines zones with relatively homogeneous parameter values. Such spatial clustering has been used to map zones that represent co-varied features in a tractable manner (e.g., (Devadoss et al., 2020; Wainwright et al., 2022)). In this case, the identified clusters are anticipated to comprise areas with consistent potential for land-atmosphere circulations.

The commonly used k-means method is selected as a clustering algorithm. The features used to perform the clustering are the fitted spatio-temporal characteristic length scales and interaction exponent, as well as the spatio-temporal variance computed directly from the LST data. Each feature is normalized by its minimum and maximum values before performing the clustering. The dissimilarity between two data points is determined based on the Euclidean distance. To determine the appropriate number of clusters to use, the elbow method is adopted. In the elbow method, k-means clustering is performed on the dataset for a range of k values (i.e., number of clusters). Then, for each k, the method computes an average score for all clusters. By default, the distortion score is computed. The distortion scores the sum of square distances from each point to its assigned center in the clustering. Once this metric for k is plotted, it is possible to visually determine the best value for the number of clusters, as the k where the inflection point of the curve occurs. Using the obtained number of clusters, k-means is applied, and the resulting clusters are mapped out and analyzed in terms of their characteristic spatio-temporal covariance function (CSTCF), which for each cluster is computed using the parametric model with the mean value of all the domains contained within the same cluster.

2.7 Combined metric for spatio-temporal persistence

The tools developed in this study pave the way to using solely remote sensing information to detect areas characterized by homogeneous spatio-temporal dynamics. This information can, in turn, be employed to investigate which landscape features can be linked to heterogeneity-driven circulations. To this end, we construct a metric (*m*) based on normalized values of the parameters of the fitted spatio-temporal covariance function. Rescaled forms of the model parameters are used as performance metrics for each component of the spatio-temporal dependence structure. After normalization, parameter values range from zero (low spatio-temporal variability score) to one (high spatio-temporal variability score). Equations 6 to 9 present the computation of the rescaled forms for each parameter. Capital variables (i.e., Σ^2, Γ, Λ, and *A*) represent the CONUS broad fields of parameters, while lowercase variables (i.e., σ^2, γ, λ', and *a*) designate the domain or cluster of domains specific parameter values.

$$\sigma_{rs}^2 = \frac{\sigma^2 - \min(\Sigma^2)}{\max(\Sigma^2) - \min(\Sigma^2)} \qquad (6)$$

$$\gamma_{rs} = \frac{\gamma - \min(\Gamma)}{\max(\Gamma) - \min(\Gamma)} \qquad (7)$$

$$\lambda_{rs} = \frac{\lambda - \min(\Lambda)}{\max(\Lambda) - \min(\Lambda)} \qquad (8)$$

$$a_{rs} = \frac{a - \min(A)}{\max(A) - \min(A)} \qquad (9)$$

Once all the parameters are rescaled, the combined metric (*m*) for each domain or cluster of domains can be computed as shown in Equation 10. This combined metric takes values between zero (0) for locations with minimum spatio-temporal persistence to four (4) for areas with maximum spatio-temporal persistence.

$$m = \sigma_{rs}^2 + \gamma_{rs} + \lambda_{rs} + a_{rs} \qquad (10)$$

3 Results

3.1 ESTCF over CONUS

The summer daytime ESTCFs over CONUS were computed using all the available LST observations according to the procedure described in Sections 2.2 and 2.3. The obtained ESTCFs for seven (7) locations with various landscape features are presented in Figure 5. The figure shows the site-specificity of the obtained ESTCFs. As expected, the spatio-temporal variance of LST (i.e., the maximum value of the color bar) was higher for mountainous areas (i.e., Colorado and Lake Tahoe, Figures 5b and 5c), and for coastal regions, including lake coastlines (i.e., New York City and Lake Michigan, Figures 5f and 5g), due to contrasting landscape features such as topography and material thermal properties. The observed variance values for flat areas were relatively low, with values around 3K^2 for the Louisiana, Atlanta, and North Dakota domains (Figures 5d, 5e, and 5h).

It was observed that the contrast between land cover types, particularly water vs. land, increased the space-time interaction by producing a convex transition between space and time in the ESTCFs. The same convex transition was observed for mountainous regions (Figures 5b and 5c). Additionally, domains containing features such as rivers, cities, and small lakes (Figures 5d, 5e, and 5h), which had generally homogeneous landscapes except for the small-scale features (i.e., in the order of 10-30km), displayed a relatively sharp decay in their spatial coherence. Domains displaying large-scale heterogeneity, driven by topography or contrasting land cover (especially land vs. water), showed larger spatial coherences (see Figures b, c, f, g).

Regarding the temporal persistence (i.e., memory) of the ESTCFs in the analyzed domains, it was observed that the presence of large-scale landscape features, such as bodies of water and topographic gradients, increased the temporal persistence of the ESTCF. The domains with the longer persistence of summer daytime LST were the ones located in Colorado (~5hr, Figure 5c), California (~4hr, Figure 5b), and New York City (~2.5hr, Figure 5f). Smaller persistence values, in the order of 1hr, were detected for all the other domains.

3.2 Spatio-Temporal Covariance Function: Parametric model over CONUS

Once the ESTCFs were computed for CONUS, the selected parametric model for the spatio-temporal covariance function was fitted to them, as described in Sections 2.4 and 2.5. The obtained parametric fits for the seven (7) domains analyzed in Figure 5 are presented in Figure 6. For the locations of the domains over CONUS, the reader is referred to Figure 5a.

Figure 6 shows the performance of the selected parametric model in reproducing the observed ESTCFs. The Figure displays zoomed-in satellite imagery for each domain, the ESTCFs computed from the LST observations, the obtained fit, the set of parameters corresponding to that fit, and the normalized root mean square error (nRMSE) as a performance metric. Visual inspection showed that, in general, the parametric model performed well for the selected locations, particularly for the higher covariance values on the bottom left part of the spatio-temporal domains, $0.7 \leq C/\sigma^2 \leq 1.0$. This was expected, as these values played a more critical role in the normal least squares minimization algorithm used in the fit. The obtained nRMSE values confirmed the results derived from the visual inspection. Higher nRMSE values were observed for domains with larger spatio-temporal variance values (i.e., mountainous domains and coastal domains; Figures 6a, 6b, and 6e). This behavior can be explained by the fact that the selected model was overly simplistic to represent the complex space-time interactions that could emerge in some regions.

Regarding the magnitude of the obtained parameters, it was clear that domains where interactions between land and water existed (Figures 6e and 6f) and where significant topographic gradients were present (Figures 6a and 6b) displayed higher values of the temporal characteristic length-scale, γ, and values of the spatio-temporal interaction exponent, a, over 1. For the remaining domains (Figures 6c, 6d, and 6g), the spatial characteristic length-scale, λ, seemed to play a more critical role in reproducing the ESTCF, with relatively high values.

Additionally, for these cases, the spatio-temporal interaction exponent, a, kept values slightly under or over 1.0, indicating an almost linear spatio-temporal interaction.

The fit to the parametric spatio-temporal covariance function model was performed for every 1.0°x1.0° domain under analysis with at least 2/3 of its area over land. Figure 7 presents the integrated results for this procedure as maps. Each pixel represents the central 0.25°x0.25° for each 1.0°x1.0° analyzed domain. Maps for the fitted spatio-temporal characteristic length scales, γ and λ, are presented (Figures 7b and 7c), as well as for the fitted spatio-temporal interaction exponent, a, (Figure 7d) the computed spatio-temporal variance, σ^2, (Figure 7a) and the nRMSE obtained for the fit (Figure 7e). It is worth mentioning that the ESTCFs presented in Figure 5 and Figure 6 did not share the same time lag axis limits due to the location-dependent day lengths. However, the temporal lag axis was standardized for the CONUS-wide fit, $0 \leq \tau \leq 8hr$. The results reveal:

1. Spatio-temporal variance (Figure 7a): A West to East decreasing gradient of variance was observed, showing agreement with the long-term precipitation climatology for the area. Additionally, the obtained gradient was also consistent with the Köppen-Geiger climate classification system for CONUS, with drier climates displaying a larger LST variability with a lower influence of surface soil moisture content. As expected, higher variance values were observed in regions with significant topographic gradients, coastlines, urban areas, particularly in the Midwest and the Mississippi River delta area in Louisiana. Coastal areas of the Atlantic and Pacific showed clear differences, with the Pacific coast displaying larger variance values due to a sharper contrast in temperature between land and water. The lowest values were located in flat areas in the central and eastern regions of CONUS.

2. Temporal characteristic length-scale (Figure 7b): The observations derived from Figure 6 were confirmed in this case with increased values of γ in mountain areas and coastal regions, particularly in the Sierra Nevada, Rocky Mountains, Coastal ranges, Appalachians, California Gulf, Northeast coastlines and Great Lakes shorelines. Urban areas in the Midwest, South, and Northeast also showed elevated values in comparison to their surroundings.

3. Spatial characteristic length-scale (Figure 7c): An east-to-west decreasing gradient for this parameter was observed. Higher values were identified for flat areas of the Midwest. Unlike the temporal characteristic length scale, urban sites, coastlines, and mountain ranges displayed reduced values, probably due to a heavy influence of the spatio-temporal interaction exponent.

4. Spatio-temporal characteristic length-scale (Figure 7d): The patterns observed here resemble the ones for the spatio-temporal variance (Figure 7a). Concave spatio-temporal interactions ($a < 1$) were found in flat areas of the South, non-coastal areas of the Midwest, and non-urban portions of the Northeast. Linear relationships between space and time ($a \cong 1$) were identified for urban areas of the Midwest and South, as well as in relatively homogeneous domains in the West. In general, domains containing coastlines

and significant topographic gradients consistently displayed a convex spatio-temporal interaction ($\alpha \geq 1$).

5. nRMSE (Figure 7e): The model struggled to thoroughly capture the observed dynamics in the Appalachians, Sierra Nevada, Rocky Mountains, Coastal ranges, and some urban areas of the Midwest and Northeast. However, due to the general performance, it was concluded that the selected parametric model represented the spatio-temporal dynamics of LST in a relatively accurate way, with a CONUS-wide mean nRMSE of ~3%.

Alternative forms of the length scales were derived to obtain an approximation unaffected by the spatio-temporal interaction exponent and identify the individual effects of space and time on the joint spatio-temporal dynamics. These forms were estimated as the fitted characteristic length scales operated by the fitted spatio-temporal interaction exponent (as described in Section 2.5). Another goal of this procedure was to identify locations where the spatial characteristic length-scale displayed patterns that could not be placed directly from Figure 7. Figure 8 shows the maps of the alternative forms of the spatio-temporal characteristic length scales with units. In general, the modified temporal characteristic length-scale (Figure 8a) displayed some of the same patterns identified using Figure 7b: increased values in mountainous areas and urban zones in the Midwest, South, and Northeast with elevated values compared to their surroundings. For mountainous regions, the magnitude of the modified temporal scale (i.e., memory) was in the range of 20 to 60 hours (i.e., one day to 2.5 days). For the urban areas cases, the obtained memory was on the order of 10 hours. However, unlike Figure 7b, Figure 8a showed that the contrast between land-only domains and coastal domains, particularly in the California Gulf, the Great Lakes shorelines, and the Northeast coastline, was not as high for the modified temporal characteristic length-scale, with memories slightly below 10 hours. This implies that the increased values of γ in these locations were caused by an elevated space-time interaction exponent rather than by a time-only effect. On the other hand, the modified spatial characteristic length-scale (Figure 8b) displayed increased values in domains with persistent landscape features, including coastlines and mountain ranges, a pattern expected but not observed on the original spatial characteristic length-scale map (Figure 7c). In this sense, for these locations, the spatio-temporal interaction parameter reduced the influence of the space-only characteristic length-scale, probably due to the magnitude of the distances (below 1°).

3.3 Clustering analysis

With all the parameter values from the model fit, an unsupervised clustering algorithm (i.e., k-means) was used to identify homogeneous zones. The number of clusters to be used (i.e., k=6) was determined using the elbow method based on the distortion score, as displayed in Figure 9a.

Figure 9b presents the location of the obtained clusters over CONUS. The resulting clustered regions are described below:

- **Cluster 1** occupied 41.17% of the total CONUS area, and it corresponded mainly to flat spots in the Midwest and Atlantic coastal plain regions, with some low zones on the Mountain and Pacific West.

- **Cluster 2** covered the smallest area (1.09% of the total CONUS area) and grouped domains containing the coastal regions of the Gulf and South of California and the Central Valley coastline.

- **Cluster 3** occupied 7.75% of the total area and included mostly transitional domains next to significant topographic gradients (e.g., Sierra Nevada, Rockies, Coastal ranges, Black Hills, and Appalachians), as well as some coastal or semi-coastal regions in the Northeast, the Great Lakes shorelines (except for the Lake Erie coastline, probably due to its relatively small size), the Great Salt Lake area, and the Gulf of California.

- **Cluster 4** was the second largest region, with 30.84% of the total area, and contained most of the Central region of the US, as well as most of the coastline domains in the Gulf of Mexico and Southeastern US, and domains including large urban areas of the Midwest and South (e.g., Nashville, Memphis, Saint Louis, Kansas City, Indianapolis, Chicago, Milwaukee, among others). A significant portion of the Appalachian Mountains with intermediate elevations was also included in Cluster 4.

- **Cluster 5** occupied 15.10% of the total area and primarily encompassed low regions of the Western US, including the California Central Valley, as well as flat areas of Nevada, Arizona, Utah, and Baja California in Mexico; it also included some coastal domains, mainly in the Lake Erie area, as well as Northern parts of Lake Michigan, and Eastern coasts of Lake Huron. The Chesapeake Bay area in Maryland and some domains in the Appalachian region were also included in Cluster 5.

- **Cluster 6**, occupying 4.05% of the total CONUS area, contained all the domains with the highest elevations of the Sierras, Coastal Ranges, Rockies, Appalachians, and Lake Superior shorelines.

Figure 10 presents the spatial mapping of the clusters over CONUS, as well as the parametric representation of the spatio-temporal covariance function obtained from the mean cluster value of the parameters (i.e., the mean value of the parameter values for all the domains contained within the same cluster) or characteristic spatio-temporal covariance function (CSTCF). The first observation derived from Figure 10 is the variance discrimination between groups. In general, locations with higher variances, $\overline{\sigma^2}$, (i.e., Pacific coast and Gulf of California, higher elevations of Rockies, Sierra, Coastal ranges and Appalachians, and shorelines of Lakes Superior, Michigan, Huron, and Ontario) were grouped by the clustering procedure in Cluster 2, Cluster 6, and Cluster 3, respectively (Figures 10b,10f, and 10c). Due to the topographic and material contrast within those high variance domains, the obtained shape for the CSTCF was predominately convex with mean exponent values $\bar{a}$ over one, particularly in the Cluster 2 case. That was also the case for the mean characteristic temporal length-scale, $\bar{\gamma}$, with the highest values associated with the larger variance clusters. Due to the previously discussed influence of the spatio-temporal interaction exponent, the mean characteristic spatial length-scale results were less clear. As for the lower variance clusters, Cluster 1, Cluster 4, and Cluster 5 (Figures 10a, 10d, and 10e, respectively), each exhibited distinctive characteristics. Besides presenting the lowest $\overline{\sigma^2}$ values, Cluster 1 displayed low values of $\bar{\gamma}$ and $\bar{a}$ (i.e., concave shape). Cluster 4 featured the second lowest variance with an almost linear space-time interaction exponent, $\bar{a}$, and relatively low $\bar{\gamma}$. Finally, Cluster 5 constituted a transitional group with the third lowest variance but relatively high values of both $\bar{a}$ and $\bar{\gamma}$.

The box plots in Figure 11 display the distribution of parameter values within each cluster and aid in identifying the main factors determining the grouping. Cluster 1 was mainly controlled by the lowest a and γ values and the highest λ; Cluster 2 by the highest γ, a, and σ^2 values; Cluster 3 by the second highest values of γ, and intermediate ones of a and σ^2; Cluster 4 by the second lowest γ and a and the second highest λ values; Cluster 5 by the third lowest γ, a, and σ^2 values; and Cluster 6 by the highest γ and the second highest λ, a, and σ^2 values.

3.4 Combined metric for spatio-temporal persistence

The combined metric for spatio-temporal persistence described in Section 2.7 was computed for every domain in CONUS. The combined metric values for each domain are presented in Figure 12a, and the individual contribution of the rescaled forms of the parameters to the total metric value in Figures 12b, 12c, 12d, and 12e.

In general, Figure 12a shows that domains with certain landscape features displayed increased metric values with respect to their surroundings. These features included coastlines, mountainous ranges, urban areas, and large rivers. Higher values were found in the Gulf of California coastline, Lake Superior and Michigan shorelines, Sierra Nevada, and higher elevations of the Rocky Mountains and Coastal Ranges. Lower values focused on domains within the Midwest and Atlantic coastal plain regions, excluding coastlines. Regarding the individual contributions of the parameters to the total metric value, Figure 12b showed that for most cases, the spatio-temporal variance was negligible, contributing to less than 10% of the total metric value in most locations. For the temporal characteristic length-scale displayed in Figure 12c, the contributions to the metric were higher in mountainous regions, the Southern Pacific and Northern Atlantic coastlines, and the Great Lakes shorelines, reaching values of about 40%. Figure 12d showed that the spatial characteristic length-scale contributed significatively to the metric mainly in flat domains of the Midwest and Atlantic coastal plain regions and central CONUS, with values ranging between 40% and 80%. Finally, the spatio-temporal interaction exponent (Figure 12e) was consistently high for most of the domains, with significant contributions in the Atlantic coastline, cities in the Midwest, Appalachian mountains, Great Lakes shorelines, and most of the Mountain West domains, except for the higher elevations where the temporal characteristic length-scale dominated.

The combined metric for spatio-temporal persistence was also computed for the mean value of the parameters obtained from clustering. Table 1 displays the combined metric for each cluster and the individual contributions of the rescaled forms of the mean parameters. Cluster 2, composed of the coastal domains in the Gulf of California and the South Pacific region, resulted in the highest metric value and, therefore, a presumed highest potential for the development of land-atmosphere circulations. The increased importance of the temporal characteristic length-scale and the spatio-temporal interaction exponent mainly drove this behavior. This result pointed toward the joint effects of the temporal persistence (i.e., memory) and the spatio-temporal interaction as the main factors determining the structure of the LST fields in those domains. Cluster 6 presented the second largest values of the metric, containing domains with the highest topographic gradients as well as some coastal areas in the Great Lakes region. In this case, the main driver of the metric was the temporal characteristic length-scale indicating an essential influence of the memory of the spatio-temporal structure of the fields. A moderate metric value was observed for Cluster 3 and 5, mainly driven by relatively high values of the spatio-temporal interaction exponent. Finally, Clusters 1 and 4 resulted in the lowest metric values; even though they had the highest values of the spatial characteristic length scales, low values of all the other three parameters resulted in overall reduced metric values.

Table 1. Mean parameter values per cluster, rescaled values of the parameters, and mean metric per cluster.

	σ^2		γ		λ		a		Metric (m)
	Mean value	Rescaled	Mean value	Rescaled	Mean value	Rescaled	Mean value	Rescaled	
Cluster 1	3.28	0.03	4.22	0.03	0.22	0.34	0.81	0.16	0.56
Cluster 2	60.80	0.52	86.31	0.86	0.09	0.13	2.20	0.88	2.40
Cluster 3	15.87	0.13	45.05	0.44	0.11	0.16	1.61	0.58	1.32
Cluster 4	5.51	0.05	9.74	0.09	0.17	0.26	1.07	0.30	0.69
Cluster 5	9.74	0.08	19.57	0.19	0.11	0.16	1.38	0.46	0.89
Cluster 6	18.67	0.16	89.89	0.90	0.12	0.18	1.73	0.64	1.87

4 Discussion

4.1. General implications and specific application findings

This study introduced the Empirical Spatio-Temporal Covariance Function (ESTCF) to evaluate the spatial coherence and memory of remotely sensed spatio-temporal fields. The main aim was to uncover significant spatio-temporal patterns within the observed processes in the Soil-Vegetation-Atmosphere System (SVAS). The proposed approach was applied to remotely sensed LST fields to determine whether the application could successfully pinpoint regions where landscape characteristics such as coastlines, topographic gradients, and urban areas might be influential in initiating heterogeneity-driven circulation systems. The procedure was implemented and assessed across the contiguous United States (CONUS). The country was divided into distinct 1.0°x1.0° domains. For each area, the summer day-time LST ESTCF was calculated independently, and subsequently, the parametric covariance model was fitted to the data. Following the fitting process for all domains, a clustering analysis was employed to recognize areas that share analogous spatio-temporal dynamics of LST, suggesting similar potential for heterogeneity-driven circulation generation. The key contributions of this paper encompassed a) introducing a versatile and comprehensive tool to depict and characterize the spatio-temporal interdependence structure of remotely sensed fields, b) presenting a parametric covariance function model that succinctly characterizes the spatio-temporal configurations captured by the ESTCFs, and c) proposing a multi-dimensional clustering strategy to discern regions with analogous spatio-temporal dependency structures. The methodology introduced in this study is expected to pave the way for a systematic analysis of the spatio-temporal patterns present in remotely sensed fields, as these insights can be linked to physical processes within the SVAS.

The ESTCF was easily obtained from gridded observations and proved flexible enough to deal with missing data, varying domain sizes, and differential temporal aggregations. Additionally, the ESTCF displayed the ability to characterize spatio-temporal regimes based on features of the fields such as spatio-temporal variance, spatial coherence structure, temporal persistence, and space-time interactions. Overall, the proposed parametric model of the covariance function accurately emulated the empirical data while simultaneously summarizing the dynamics within the ESTCFs. The simplified features were then used to identify areas with homogeneous spatio-temporal dynamics, successfully classifying domains based on their main spatio-temporal features.

Regarding the application of the proposed methods to the LST fields, the joint use of the clustering procedure and the proposed combined metric for spatio-temporal persistence allowed the identification of zones with higher spatio-temporal dynamics in coastal domains of the Gulf of California and the South Pacific region and domains containing the highest elevations of mountainous areas (i.e., the Sierra Nevada, Rockies, Coastal Ranges, Appalachian Mountains) as well as the coastal areas surrounding the largest lakes in the Great Lakes region. These locations coincided with those reported in the literature to have an increased likelihood of developing mesoscale heterogeneity-driven circulations. The main drivers of the increased spatio-temporal variability in these locations were the temporal characteristic length-scale and the spatio-temporal interaction component. These findings reinforce the essential influence of the memory of the spatio-temporal structure of the fields in the presumed potential of land-atmosphere coupling development. Additionally, the low individual contribution of the spatio-temporal variance to the total combined metric value underlined the necessity for Earth System Models (ESMs) to include more comprehensive metrics than spatially aggregated macroscale grid statistics (e.g., spatial mean and variance) to inform their atmospheric components of the state of their land components.

4.2 Limitations and implications of method choices

4.2.1 Issues regarding the ESTCF

The ESTCF was selected in this study as the tool to summarize the spatio-temporal dependence structure of remotely sensed fields of LST over CONUS. The tool was chosen as it is easily attainable from the available remotely sensed data, providing a relatively dense characterization of the heterogeneity degree on different spatial and temporal scales. Although this tool provided a promising path forward for a robust evaluation and summarization of the multi-scale spatio-temporal heterogeneity in large-scale observational fields, the limitations of the methodology should be considered.

1. *Sampling issues*: The accuracy of the ESTCF is highly dependent on the number of available observations for the covariance computation in each spatio-temporal distance. As mentioned in Section 2.1, cloud cover and atmospheric aerosols directly influence the LST retrieval processes, as they can obstruct the satellite's view of the surface, leading to spatial data gaps and reduced observations over time. In general, it is well known that cloudiness leads to cool bias in satellite-derived LST, particularly within cloudy areas (e.g., mountainous areas). Additionally, developed heterogeneity-driven circulations might lead to increased cloudiness, which could negatively impact the quality of the available fields of observation in specific domains. Other factors affecting the retrieval of different variables (soil moisture content, snow coverage, vegetation fraction, and material differences, among others) will undoubtedly impact the accuracy of the computed ESTCFs if different processes are analyzed. In this sense, future work should investigate the sensitivity of the ESTCFs to the availability of observations in observation-limited domains.

2. *Selection of a spatio-temporal covariance parametric model*: The selected structure for the covariance function parametric model was chosen as a trade-off between the number of parameters (and their physical interpretability) and an accurate representation of the data-derived ESTCFs. However, this structure represents one of many possible alternatives to model the spatio-temporal dynamics of geospatial fields. Research on geostatistics has derived many forms for covariance models, and more are expected to be developed (Bolliger et al., 2007; W. Chen et al., 2021; Gneiting, 2002; de Iaco, 2010; Ma, 2003; Schepanski et al., 2015). The selected function imposes a specific a priori structure to the spatio-temporal dependence that might not be appropriate for all domains, variables, time aggregations, or applications (e.g., not all domains or variables would benefit from a non-separable, exponential parametric model). Consequently, further investigations should apply the ideas exposed throughout this study to other spatio-temporal covariance parametric models in order to determine the most appropriate version of it. Nonstationary covariance structures could be evaluated for specific processes as it is expected that, in some cases, the covariance structure may change in response to physical changes in the equilibrium state of the system under analysis.

3. *Temporal and spatial resolutions and scales:* In this study, the spatial extent of the domains was set to 1°x1° with an hourly temporal resolution over CONUS. Although these temporal and spatial resolutions were appropriate to analyze mesoscale land-atmosphere circulations with an ESM framework in mind, applications requiring finer or coarser temporal and spatial resolutions (i.e., diurnal cycle evaluations) would most likely require the definition of a different structure for the covariance function parametric model. It is possible that the goodness of fit of analytical covariance functions may exhibit some dependence on the domain size used in the analysis. Thus, further analysis of the proposed approach over varying time windows, domain sizes, and spatio-temporal scales are also a welcome follow-up contribution. Preliminary work by the coauthors has proved the utility of the ESTCF approach in summarizing the spatio-temporal information contained within a long-term (i.e., > ten years of record length), global, remotely sensed LST gridded product (Freitas et al., 2013). However, further analyses will examine the effects of the selected moving window

size, spatial offset, analyzed time period, and temporal aggregation in the obtained ESTCFs. The accuracy of the proposed parametric covariance model under these varying conditions will also be tested.

4.2.2 Remote sensing of LST and SST

In this study, a coupled LST-SST product was employed to explore the spatio-temporal patterns of remote sensing surface temperature that could lead to the development of mesoscale circulations. However, as mentioned in Section 2.1, remote sensing retrieval of LST and SST are intrinsically different, with each of them presenting particular challenges.

Surface temperature remote sensing retrieval poses inherent challenges due to multiple factors impacting measurement accuracy and precision. Among these challenges, the intrinsic diversity of Earth's surface materials stands out. Each surface type possesses distinct thermal characteristics, emissivity values, and heat exchange mechanisms, resulting in varying thermal energy emission patterns. This diversity makes the algorithm heavily dependent on the feed surface emissivity values and land-water mask. Furthermore, the presence of atmospheric water vapor significantly affects the thermal infrared signal detected by satellites, often leading to an underestimation of actual surface temperature compared to the measured brightness temperature. The relationship between radiance and temperature is also nonlinear, rendering traditional linear models, like the single and split channel methods, less precise, particularly in hot and humid atmospheric conditions (Duffy et al., 2022). This discrepancy is amplified with increasing column water vapor, making the inclusion of water vapor data crucial for enhancing LST accuracy (Sobrino et al., 1993). However, the spatial and temporal variability of the atmospheric conditions further complicates the retrieval process, as they introduce error propagation and uncertainties into the estimates.

In this sense, remote sensing-derived surface states inevitably depend on assumptions about the overlying atmosphere and landscape features, and estimations ultimately constitute a model output. This produces a relatively high uncertainty, mainly since there is no observational 'truth' at the landscape scale for comparison (Stisen et al., 2011). However, the information content present in the spatio-temporal structure of the observed satellite fields is intrinsically valuable, especially when considering the wide variety of variables of surface states and fluxes currently estimated (e.g., soil moisture content (Chan et al., 2018; Entekhabi et al., 2010; Kerr et al., 2012; Parinussa et al., 2015; W. Wagner et al., 2013), evapotranspiration (Boschetti et al., 2019; J. B. Fisher et al., 2020; Martens et al., 2017; Running et al., 2019; Su, 2002), snow cover fraction (Painter et al., 2009; Tsai et al., 2019), and changes in water storage (Tapley et al., 2004)).

4.3 ESTCF applications

4.3.1 Towards the improved representation of land-atmosphere interactions in ESMs

Research has established the significant role of landscape heterogeneities in key atmospheric processes, including atmospheric boundary layer depth determination, convection initiation, and mesoscale circulations (Bertoldi et al., 2013; Gutowski et al., 2020; Kang & Bryan, 2011; Kustas & Albertson, 2003; Ntelekos et al., 2008; Simon et al., 2021; Timmermans et al., 2010). Local studies are advancing our understanding of multi-scale landscape heterogeneity effects on micro- and mesoscale meteorological processes (H. Y. Huang et al., 2011; Senatore et al., 2015; Shrestha et al., 2014; Talbot et al., 2012). However, the extent of this effect on land-atmosphere interactions in the broader climate system remains uncertain. This uncertainty primarily stems from the limited coupling between existing sub-grid parameterizations in land surface models and the atmospheric components of ESMs. Typically, ESMs exchange spatial mean mass and energy fluxes between land and atmosphere while disregarding higher-order spatial statistics, such as spatial variance or characteristic length scales. Nevertheless, atmospheric circulation models are progressively incorporating higher-order sub-grid scale processes, as seen in examples like the Cloud Layers Unified By Binormals (CLUBB) and Eddy Diffusivity Mass Flux (EDMF) (Golaz et al., 2002; M. Huang et al., 2022; Sušelj et al., 2013). These developments provide an opportunity for potential coupling between atmospheric models and the sub-grid scale heterogeneity of the land surface. This study aims to contribute meaningfully to such efforts, ideally enhancing land surface parameterizations within the atmospheric components of ESMs with higher-order statistics.

The approach presented in this study can provide more than a tool to summarize the spatio-temporal dependence structure of remotely sensed fields; it is proved that it can also be employed to estimate the characteristic length scales of heterogeneity, providing parametrizations with useful spatio-temporal information over macroscale grid cells. The method also assesses the spatial coherence and memory of the fields and allows the identification of regions with homogeneous characteristics. By identifying these locations, the tool could help inform parametrizations schemes for ESMs by distinguishing locations and times for which the common flux averaging methods might be insufficient to represent interactions between model components, particularly the interaction between the land and atmosphere. Ultimately, the hope is that the type of approach presented through this study drives the ESM community in a direction where the representation of the subgrid-scale heterogeneity in both space and time is considered both in model development and as a model diagnostic tool.

4.3.2 Model evaluation: Spatially distributed hydrological models and Land surface models

Physically based spatially distributed hydrological models allow the simulation of the spatial distribution of hydrological and hydraulic processes within catchments while still providing discharge estimates for the river network. Their main advantage is that they emulate, to some extent, the natural spatial heterogeneity of the hydrological processes, driven by spatially distributed factors that constrain the hydrological processes, such as land use, climate, and soil properties (Koch et al., 2015). However, most spatially distributed hydrological models are still calibrated and evaluated using a goodness of fit metric describing the efficiency of the model representation of a catchment-aggregated or point-retrieved quantity, such as discharge (Zink et al., 2018). This practice generally makes the models over-parametrized relative to the data available to constrain them (Stisen et al., 2011). It is widely accepted that model calibration and validation practices for these models should take directions that agree more with the spatially distributed nature of the outputs, including continuous spatial observation data (Beven, 2001). The main issue, however, is the lack of a standard set of techniques and metrics to evaluate the goodness of fit of the models’ spatial predictions. Several spatial performance metrics have been developed (Ko et al., 2019; Koch et al., 2015, 2016, 2017; Li et al., 2009; Stisen et al., 2011, 2021; Xiao et al., 2022; Zink et al., 2018) and reviewed in their ability to constrain and evaluate models (Wealands et al., 2005). In general, simple global statistics operating locally (i.e., pixel-to-pixel comparison of the modeled and observed maps) are insufficient as they are susceptible to small-scale spatial displacement errors and do not consider information on patterns or spatial correlation of the data. More robust global statistical metrics such as mean bias, standard deviation, and variogram ranges are not entirely appropriate, as they are also pattern agnostic. An approach like the one presented in this study can provide a robust and compact tool to evaluate the performance of spatially distributed hydrological models while still being "pattern aware". The model representation of spatio-temporal variables and processes such as soil moisture content, runoff generation, infiltration, and evapotranspiration can be characterized using the ESTCF tool; then, by adding catchment aggregated observations, such as streamflow, the proposed tool would add an extra layer of constraints in the calibration stage. In addition to its flexibility in terms of spatial and temporal resolution, the proposed ESTCF method has the advantage of not being limited to square or rectangular domain shapes and being readily applicable to catchment-based hydrological models.

On the other hand, land surface models (LSMs) were initially developed to operate at continental and global scales as the land boundary condition of climate and numerical weather prediction models and ESMs (R. A. Fisher & Koven, 2020; Ko et al., 2019). Recognizing the multi-scale nature of spatial heterogeneity in land surface processes, tiling schemes were developed to represent the hierarchical structure of heterogeneity within macroscale grid cells (~100km horizontal resolution). Tiling schemes subdivide macroscale 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 (D. Li et al., 2013). Despite the significant advances regarding tiling schemes over the last decade, many issues persist, including the fact that over large-scale domains, LSM sub-grid outputs are mostly only summarized and evaluated via macroscale grid statistics: spatial mean and variance. Although informative, these statistics are insensitive to the tiles' large-scale spatial patterns (i.e., pattern-agnostic metrics) (Jupp & Twiss, 2006; Torres-Rojas et al., 2022). This issue is critical as emerging work shows the importance of correctly representing the sub-grid spatio-temporal patterns of surface states to explain the role of sub-grid heterogeneity on atmospheric response (Simon et al., 2021). An approach as the ESTCF can provide a tool to summarize the spatio-temporal dependence structure of LSM output fields, characterize it, and evaluate the accuracy of the model parametrizations of different processes by comparison to remote sensing derived ESTCF for multiple variables (e.g., soil moisture content, LST, evapotranspiration, and vegetation condition indexes, among others).

Nevertheless, when assessing either distributed hydrological models or LSMs in relation to hydrological states or fluxes derived from remote sensing, one must acknowledge that this involves comparing models to models, with considerable uncertainty inherent in both methods. This is especially true as there is no definitive observational 'truth' available for landscape-scale comparisons. Furthermore, the careful selection of suitable evaluation variables and objective functions is essential to guarantee the reliability of model assessments (Stisen et al., 2011).

4.3.3 Spatio-temporal characterization for alternative applications

This study introduces the ESTCF as a versatile and comprehensive tool to depict and characterize the spatio-temporal interdependence structure of remotely sensed fields. Even though the tool is solely applied to LST fields in this study, it is recognized that application to other spatio-temporal fields might shed light on the dynamics of processes within different compartments of the SVAS. This section explores both the remote sensing data available for other applications and systems and the processes that would benefit from the application of the ESTCF method.

- Soil moisture content (SMC): Besides the relevance of soil moisture spatio-temporal patterns in the initiation of land-atmosphere circulations (see Section 1), other essential processes such as drought onset and evolution, infiltration, surface and subsurface runoff, and inundation dynamics, all heavily depend on the spatio-temporal structure of SMC fields. However, the main limitations of the currently available SMC remote sensing products are their long revisit times and low spatial resolution. These limiting factors reduce the current applicability of the proposed methods to all the mentioned processes. However, regional flood and drought evolution analyses on longer time scales (biweekly to monthly) are still feasible using the available data.
- Fractional vegetation cover (FVC), leaf area index (LAI), normalized difference vegetation index (NDVI), and enhanced vegetation index (EVI): Spatio-temporal remotely sensed fields of vegetation-related quantities and indices contain essential information related to processes such as evapotranspiration, erosion, net primary productivity, crop productivity, agricultural droughts, and turbulent energy exchange between the land surface and the atmosphere. The temporal and spatial resolution of the currently available products would enable weekly to monthly analysis over seasonal scales and regional to continental domains.
- Evapotranspiration (ET): ET is a critical process in the hydrological cycle, linking the land surface water balance, carbon cycle, and the land surface energy balance. Remote sensing provides a method to estimate ET at regional to global scales with biweekly to weekly return rates. Spatio-temporal analysis of this variable would be primarily valuable for model evaluation and calibration purposes due to the vital role of estimated ET in model structures.
- Reanalysis of atmospheric, land, and oceanic climate variables: Global, hourly, and extended records (~1940-present) of multiple variables related to different compartments of the SVAS are included within reanalysis datasets. The wide availability of this data might allow us to analyze the impacts of climate variability and climate change on the spatio-temporal dependence structure of multiple fluxes and states in systems within the SVAS.

5 Summary and Conclusions

Several approaches have been developed to identify, summarize, and extract relevant patterns from spatio-temporal geophysical datasets. These methods can be applied in both space and time, though, in general, they are only meant to analyze independent dimensions. In climate, environmental, and hydrological applications, there is a clear advantage in concurrently detecting spatially connected and enduring structures or patterns as they offer insights into the dynamics of the processes influencing them. Among the tools developed for geostatistical analysis, the ESTCF stands out for its simplicity. Under several assumptions, the ESTCF quantifies the strength and structure of dependence between different locations and times. Once the ESTCF is computed, it becomes possible to select a parametric covariance model and estimate its parameters by fitting the model to the empirical function. This process allows us to gain insights into the spatio-temporal properties and interactions of the original field based on the estimated parameters.

This study introduced the ESTCF as a tool for evaluating the spatial consistency and temporal persistence of remotely sensed spatio-temporal fields. It was used to identify patterns that could have significance in understanding the dynamics of processes within the Soil-Vegetation-Atmosphere System (SVAS). Additionally, the study presented a parametric covariance model to summarize the spatio-temporal structure revealed by the ESTCF. These tools were then applied to remotely sensed LST fields over CONUS. The objective was to determine whether applying these tools could help pinpoint areas where landscape features played a role in initiating land-atmosphere circulation systems. Furthermore, the study proposed a metric for assessing the combined spatio-temporal persistence of the analyzed fields and a clustering approach to identify areas with homogeneous spatio-temporal dependence structures. Thus, the critical developments in this study included (a) a flexible and comprehensive tool to characterize and represent the spatio-temporal dependence structure of remotely sensed fields in the form of the ESTCF, (b) a 4-parameter covariance function model to more concisely describe the spatio-temporal patterns captured with the ESTCF, and (c) a multi-dimensional clustering approach to determine areas with similar spatio-temporal depende structures, and consequently a consistent presumed land-atmosphere circulation potential.

The ESTCF, derived from remotely sensed observations, was readily accessible and demonstrated adaptability in handling missing data, varying domain sizes, and different temporal aggregations. It showcased its capacity to characterize spatio-temporal patterns using field characteristics like spatio-temporal variance, spatial coherence structure, temporal persistence, and space-time interactions. The proposed parametric covariance function model was also reasonably accurate in emulating the empirical data while succinctly summarizing its dynamics. The simplified attributes were then utilized to pinpoint regions with consistent spatio-temporal patterns, effectively categorizing domains based on their primary spatio-temporal characteristics. The combined use of the clustering procedure and the suggested combined metric for spatio-temporal persistence facilitated the identification of zones with increased spatio-temporal dynamics. These zones included coastal areas in the Gulf of California and the South Pacific region, and regions with high elevations, such as the Sierra Nevada, Rockies, Coastal Ranges, and Appalachian Mountains. Additionally, the method identified coastal regions surrounding the largest lakes in the Great Lakes area. These findings aligned with prior literature reports suggesting an increased likelihood of mesoscale land-atmosphere circulations in locations with those landscape features. These results, however, were specific to the selected domain size, temporal aggregation, and parametric model structure. As such, it is recognized that this is just one of the many possibilities to summarize the spatio-temporal dynamics from remotely sensed fields and that more efficient and accurate strategies might exist.

The developed approach is the first attempt to objectively analyze the complex spatio-temporal dependence structure from remotely sensed fields for analysis applications. Moving forward, the transferability of the approach should be tested under various data availability scenarios, parametric model functional forms, clustering techniques, temporal windows, domain sizes, and study areas (i.e., move to global scales). Furthermore, although subject to errors and biases, using LST remotely sensed fields might help inform land-atmosphere parametrization schemes for ESMs of the real spatio-temporal distribution of the surface fluxes. The introduced approach will also be beneficial in calibrating and evaluating process-based spatially distributed hydrological models and parametrizations for LSMs. The approach can also be easily transferred to several other available remote sensing data sources, enhancing our understanding of the spatio-temporal dynamics of processes within different compartments of the SVAS. This work represents a step toward adapting model evaluation and parametrization techniques to leverage the available high-resolution data better, accounting for the dynamic nature of land surface processes. Overall, the tools introduced here provide a path forward to formally identify and summarize the spatio-temporal patterns observed in remotely sensed fields and relate those to the footprint of more complex dynamic processes within the SVAS.

Acknowledgments

This study was supported by funding from NOAA grants NA19OAR4310241 (Parameterizing the effects of sub-grid land heterogeneity on the atmospheric boundary layer and convection: Implications for surface climate, variability, and extremes) and NA22OAR4310644 (Implications of heterogeneity-aware land-atmosphere coupling in the predictability of precipitation extremes).

Open Research

The GOES-16 LST and SST products used in this study are freely available from NOAA's Comprehensive Large Array-Data Stewardship System (CLASS). The data that support the findings of this study, including the scripts to reproject the original data to a WSG84 projection, combine LST and SST products for the CONUS region, merge individual hourly files into weekly netCDF4 files, extract the data for 1°x1° domains over CONUS, compute the daytime summer ESTCF for those domains, and analyze the results (i.e., mapping and clustering) are preserved at https://doi.org/10.5281/zenodo.8428629 (Torres-Rojas & Chaney, 2023).

References

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/http://dx.doi.org/10.1016/j.advwatres.2012.06.005

Bertoldi, G., Kustas, W. P., & Albertson, J. D. (2013). 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. (2001). How far can we go in distributed hydrological modelling? *Hydrology and Earth System Sciences*, *5*(1), 1–12. https://doi.org/10.5194/hess-5-1-2001

Biswas, A. (2014). Scaling analysis of soil water storage with missing measurements using the second-generation continuous wavelet transform. *European Journal of Soil Science*, *65*(4), 594–604. https://doi.org/10.1111/ejss.12145

Bolliger, J., Wagner, H. H., & Turner, M. G. (2007). Identifying and Quantifying Landscape Patterns in Space and Time, 177–194. https://doi.org/10.1007/978-1-4020-4436-6_12

Boschetti, L., Roy, D. P., Giglio, L., Huang, H., Zubkova, M., & Humber, M. L. (2019). Global validation of the collection 6 MODIS burned area product. *Remote Sensing of Environment*, *235*, 111490. https://doi.org/10.1016/J.RSE.2019.111490

Brunsell, N., & Gillies, R. R. (2003). Scale issues in land - Atmosphere interactions: Implications for remote sensing of the surface energy balance. *Agricultural and Forest Meteorology*, *117*(3–4), 203–221. https://doi.org/10.1016/S0168-1923(03)00064-9

Chan, S. K., Bindlish, R., O'Neill, P., Jackson, T., Njoku, E., Dunbar, S., et al. (2018). Development and assessment of the SMAP enhanced passive soil moisture product. *Remote Sensing of Environment*, *204*, 931–941. https://doi.org/10.1016/J.RSE.2017.08.025

Chaney, N. W., Wood, E. F., Mcbratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., & Odgers, N. P. (2016). POLARIS: A 30-meter probabilistic soil series map of the contiguous United States POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma*, *274*, 54–67. https://doi.org/10.1016/j.geoderma.2016.03.025

Chang, Y., Xiao, J., Li, X., Frolking, S., Zhou, D., Schneider, A., et al. (2021). Exploring diurnal cycles of surface urban heat island intensity in Boston with land surface temperature data derived from GOES-R geostationary satellites. *Science of the Total Environment*, *763*, 144224. https://doi.org/10.1016/j.scitotenv.2020.144224

Chen, F., & Avissar, R. (1994). The Impact of Land-Surface Wetness Heterogeneity on Mesoscale Heat Fluxes. *Journal of Applied Meteorology*, *33*(11), 1323–1340. https://doi.org/10.1175/1520-0450(1994)033<1323:tiolsw>2.0.co;2

Chen, W., Genton, M. G., & Sun, Y. (2021). Space-time covariance structures and models. *Annual Review of Statistics and Its Application*, *8*, 191–215. https://doi.org/10.1146/annurev-statistics-042720-115603

Cheng, W. Y. Y., & Cotton, W. R. (2004). Sensitivity of a cloud-resolving simulation of the genesis of a mesoscale convective system to horizontal heterogeneities in soil moisture initialization. *Journal of Hydrometeorology*, *5*(5), 934–958. https://doi.org/10.1175/1525-7541(2004)005<0934:SOACSO>2.0.CO;2

Courault, D., Drobinski, P., Brunet, Y., Lacarrere, P., & Talbot, C. (2007). Impact of surface heterogeneity on a buoyancy-driven convective boundary layer in light winds. *Boundary-Layer Meteorology*, *124*(3), 383–403. https://doi.org/10.1007/s10546-007-9172-y

Cressie, N., & Huang, H. C. (1999). Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. *Journal of the American Statistical Association*, *94*(448), 1330–1339. https://doi.org/10.1080/01621459.1999.10473885

Cressie, N., & Wikle, C. (2015). *Statistics for Spatio-Temporal Data*. Hoboken, New Jersey: John Wiley & Sons, Ltd.

Crow, W. T., Huffman, G. J., Bindlish, R., & Jackson, T. J. (2009). Improving satellite-based rainfall accumulation estimates using spaceborne surface soil moisture retrievals. *Journal of Hydrometeorology*, *10*(1), 199–212. https://doi.org/10.1175/2008JHM986.1

Demel, S. S., & Du, J. (2015). Spatio-temporal models for some data sets in continuous space and discrete time. *Statistica Sinica*, *25*(1), 81–98. https://doi.org/10.5705/ss.2013.223w

Desai, A. R., Khan, A. M., Zheng, T., Paleri, S., Butterworth, B., Lee, T. R., et al. (2021). Multi-Sensor Approach for High Space and Time Resolution Land Surface Temperature. *Earth and Space Science*, *8*(10), 1–18. https://doi.org/10.1029/2021EA001842

Devadoss, J., Falco, N., Dafflon, B., Wu, Y., Franklin, M., Hermes, A., et al. (2020). Remote sensing-informed zonation for understanding snow, plant and soil moisture dynamics within a mountain ecosystem. *Remote Sensing*, *12*(17), 2733. https://doi.org/10.3390/RS12172733

Dickinson, R. E. (1995). Land-atmosphere interaction. *Reviews of Geophysics*, *33*(2 S), 917–922. https://doi.org/10.1029/95RG00284

Dirmeyer, P. A., Jin, Y., Singh, B., & Yan, X. (2013). Trends in land-atmosphere interactions from CMIP5 simulations. *Journal of Hydrometeorology*, *14*(3), 829–849. https://doi.org/10.1175/JHM-D-12-0107.1

Duffy, K., Vandal, T. J., & Nemani, R. R. (2022). Multisensor Machine Learning to Retrieve High Spatiotemporal Resolution Land Surface Temperature. *IEEE Access*, *10*(August), 89221–89231. https://doi.org/10.1109/ACCESS.2022.3198673

Entekhabi, D., Njoku, E. G., O'Neill, P. E., Kellogg, K. H., Crow, W. T., Edelstein, W. N., et al. (2010). The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, *98*(5), 704–716. https://doi.org/10.1109/JPROC.2010.2043918

Faghmous, J. H., & Kumar, V. (2014). Spatio-temporal Data Mining for Climate Data: Advances, Challenges, and Opportunities (pp. 83–116). https://doi.org/10.1007/978-3-642-40837-3_3

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

Ferguson, C. R., & Wood, E. F. (2011). Observed land-atmosphere coupling from satellite remote sensing and reanalysis. *Journal of Hydrometeorology*, *12*(6), 1221–1254. https://doi.org/10.1175/2011JHM1380.1

Ferguson, C. R., Wood, E. F., & Vinukollu, R. K. (2012). A Global intercomparison of modeled and observed land-atmosphere coupling. *Journal of Hydrometeorology*, *13*(3), 749–784. https://doi.org/10.1175/JHM-D-11-0119.1

Fisher, J. B., Lee, B., Purdy, A. J., Halverson, G. H., Dohlen, M. B., Cawse-Nicholson, K., et al. (2020). ECOSTRESS: NASA's Next Generation Mission to Measure Evapotranspiration

From the International Space Station. *Water Resources Research*, *56*(4), e2019WR026058. https://doi.org/10.1029/2019WR026058

Fisher, R. A., & Koven, C. D. (2020, April 1). Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems. *Journal of Advances in Modeling Earth Systems*. Blackwell Publishing Ltd. https://doi.org/10.1029/2018MS001453

Freitas, S. C., Trigo, I. F., Macedo, J., Barroso, C., Silva, R., & Perdigão, R. (2013). Land surface temperature from multiple geostationary satellites. *International Journal of Remote Sensing*, *34*(9–10), 3051–3068. https://doi.org/10.1080/01431161.2012.716925

Gentine, P., Massmann, A., Lintner, B. R., Hamed Alemohammad, S., Fu, R., Green, J. K., et al. (2019). Land-atmosphere interactions in the tropics - A review. *Hydrology and Earth System Sciences*, *23*(10), 4171–4197. https://doi.org/10.5194/hess-23-4171-2019

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, *97*(458), 590–600. https://doi.org/10.1198/016214502760047113

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

Graf, A., Herbst, M., Weihermüller, L., Huisman, J. A., Prolingheuer, N., Bornemann, L., & Vereecken, H. (2012). Analyzing spatiotemporal variability of heterotrophic soil respiration at the field scale using orthogonal functions. *Geoderma*, *181–182*, 91–101. https://doi.org/10.1016/j.geoderma.2012.02.016

Gueting, N., Klotzsche, A., van der Kruk, J., Vanderborght, J., Vereecken, H., & Englert, A. (2015). Imaging and characterization of facies heterogeneity in an alluvial aquifer using GPR full-waveform inversion and cone penetration tests. *Journal of Hydrology*, *524*, 680–695. https://doi.org/10.1016/j.jhydrol.2015.03.030

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

Guttorp, P., & Sampson, P. D. (1994). 20 Methods for estimating heterogeneous spatial covariance functions with environmental applications. *Handbook of Statistics*, *12*, 661–689. https://doi.org/10.1016/S0169-7161(05)80022-7

Haas, T. C. (1990a). Kriging and automated variogram modeling within a moving window. *Atmospheric Environment Part A, General Topics*, *24*(7), 1759–1769. https://doi.org/10.1016/0960-1686(90)90508-K

Haas, T. C. (1990b). Lognormal and moving window methods of estimating acid deposition. *Journal of the American Statistical Association*, *85*(412), 950–963. https://doi.org/10.1080/01621459.1990.10474966

Hashimoto, H., Wang, W., Dungan, J. L., Li, S., Michaelis, A. R., Takenaka, H., et al. (2021). New generation geostationary satellite observations support seasonality in greenness of the

Amazon evergreen forests. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-20994-y

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

Hook, S., & Hulley, G. (2019). ECOSTRESS Land Surface Temperature and Emissivity Daily L2 Global 70 m V001. https://doi.org/https://doi.org/10.5067/ECOSTRESS/ECO2LSTE.001

Huang, H. Y., Margulis, S. A., Chu, C. R., & Tsai, H. C. (2011). Investigation of the impacts of vegetation distribution and evaporative cooling on synthetic urban daytime climate using a coupled LES-LSM model. *Hydrological Processes*, *25*(10), 1574–1586. https://doi.org/10.1002/hyp.7919

Huang, M., Ma, P. L., Chaney, N. W., Hao, D., Bisht, G., Fowler, M. D., et al. (2022). Representing surface heterogeneity in land–atmosphere coupling in E3SMv1 single-column model over ARM SGP during summertime. *Geoscientific Model Development*, *15*(16), 6371–6384. https://doi.org/10.5194/gmd-15-6371-2022

de Iaco, S. (2010). Space-time correlation analysis: A comparative study. *Journal of Applied Statistics*, *37*(6), 1027–1041. https://doi.org/10.1080/02664760903019422

Ignatov, A., Pennybacker, M., Gladkova, I., Petrenko, B., Jonasson, O., & Kihai, Y. (2019). GHRSST NOAA/STAR GOES-16 ABI L2P America Region SST v2.70 dataset (GDS version 2). https://doi.org/https://doi.org/10.25921/ayf6-c438

Jenny, H. (1941). *Factors of Soil Formation, a System of Quantitative Pedology*. New York, NY: McGraw-Hill. https://doi.org/10.2134/agronj1941.00021962003300090016x

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., et al. (2011). Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *Journal of Geophysical Research: Biogeosciences*, *116*(3), 1–16. https://doi.org/10.1029/2010JG001566

Jupp, T. E., & Twiss, S. D. (2006). A physically motivated index of subgrid-scale pattern. *Journal of Geophysical Research Atmospheres*, *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

Katul, G., & Parlange, M. B. (1995). Analysis of Land Surface Heat Fluxes Using the Orthonormal Wavelet Approach. *Water Resources Research*, *31*(11), 2743–2749. https://doi.org/10.1029/95WR00003

Katul, G., Lai, C. T., Schäfer, K., Vidakovic, B., Albertson, J., Ellsworth, D., & Oren, R. (2001). Multiscale analysis of vegetation surface fluxes: From seconds to years. *Advances in Water Resources*, *24*(9–10), 1119–1132. https://doi.org/10.1016/S0309-1708(01)00029-X

Kerr, Y., Waldteufel, P., Richaume, P., Wigneron, J. P., Ferrazzoli, P., Mahmoodi, A., et al.

(2012). The SMOS soil moisture retrieval algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(5 PART 1), 1384–1403. https://doi.org/10.1109/TGRS.2012.2184548

Kim, G., & Barros, A. P. (2002). Space-time characterization of soil moisture from passive microwave remotely sensed imagery and ancillary data. *Remote Sensing of Environment*, *81*(2–3), 393–403. https://doi.org/10.1016/S0034-4257(02)00014-7

Ko, A., Mascaro, G., & Vivoni, E. R. (2019). Strategies to Improve and Evaluate Physics-Based Hyperresolution Hydrologic Simulations at Regional Basin Scales. *Water Resources Research*, (iii), 1129–1152. https://doi.org/10.1029/2018WR023521

Koch, J., Jensen, K. H., & Stisen, S. (2015). Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study. *Water Resources Research*, *51*(2), 1225–1246. https://doi.org/10.1002/2014WR016607

Koch, J., Siemann, A., Stisen, S., & Sheffield, J. (2016). Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics. *Journal of Geophysical Research*, *121*(10), 5430–5452. https://doi.org/10.1002/2015JD024482

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

Korres, W., Koyama, C. N., Fiener, P., & Schneider, K. (2010). Analysis of surface soil moisture patterns in agricultural landscapes using Empirical Orthogonal Functions. *Hydrology and Earth System Sciences*, *14*(5), 751–764. https://doi.org/10.5194/hess-14-751-2010

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

Laherrère, J., & Sornette, D. (1998). Stretched exponential distributions in nature and economy: " fat tails " with characteristic scales. *Eur. Phys. J. B*, *2*, 525–539. Retrieved from https://link.springer.com/content/pdf/10.1007/s100510050276.pdf

Lee, S. J., Wentz, E. A., & Gober, P. (2010). Space-time forecasting using soft geostatistics: A case study in forecasting municipal water demand for Phoenix, Arizona. *Stochastic Environmental Research and Risk Assessment*, *24*(2), 283–295. https://doi.org/10.1007/s00477-009-0317-z

Levine, P. A., Randerson, J. T., Swenson, S. C., & Lawrence, D. M. (2016). Evaluating the strength of the land-atmosphere moisture feedback in Earth system models using satellite observations. *Hydrology and Earth System Sciences*, *20*(12), 4837–4856. https://doi.org/10.5194/hess-20-4837-2016

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), 11,918-11,935. https://doi.org/10.1002/2013JD020657

Li, H. T., Brunner, P., Kinzelbach, W., Li, W. P., & Dong, X. G. (2009). Calibration of a groundwater model using pattern information from remote sensing data. *Journal of Hydrology*, *377*(1–2), 120–130. https://doi.org/10.1016/j.jhydrol.2009.08.012

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

Ma, C. (2003). Spatio-temporal stationary covariance models. *Journal of Multivariate Analysis*, *86*(1), 97–107. https://doi.org/10.1016/S0047-259X(02)00014-3

Mahfouf, J.-F., Richard, E., & Mascart, P. (1987). The Influence of Soil and Vegetation on the Development of Mesoscale Circulations. *Journal of Climate and Applied Meteorology*, *26*(11), 1483–1495. https://doi.org/10.1175/1520-0450(1987)026<1483:TIOSAV>2.0.CO;2

Mälicke, M., Hassler, S. K., Blume, T., Weiler, M., & Zehe, E. (2020). Soil moisture: variable in space but redundant in time. *Hydrol. Earth Syst. Sci*, *24*, 2633–2653. https://doi.org/10.5194/hess-24-2633-2020

Martens, B., Miralles, D. G., Lievens, H., Van Der Schalie, R., De Jeu, R. A. M., Fernández-Prieto, D., et al. (2017). GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geoscientific Model Development*, *10*(5), 1903–1925. https://doi.org/10.5194/gmd-10-1903-2017

Martini, E., Wollschläger, U., Kögler, S., Behrens, T., Dietrich, P., Reinstorf, F., et al. (2015). Spatial and Temporal Dynamics of Hillslope-Scale Soil Moisture Patterns: Characteristic States and Transition Mechanisms. *Vadose Zone Journal*, *14*(4), 1–16. https://doi.org/10.2136/VZJ2014.10.0150

Matheron, G. (1989). Estimating and Choosing. *Estimating and Choosing*. https://doi.org/10.1007/978-3-642-48817-7

Montero, J.-M., Fernández-Avilés, G., & Mateu, J. (2015). *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging* (1st ed.). New York, NY: Wiley.

Nair, U. S., Wu, Y., Kala, J., Lyons, T. J., Pielke, R. A., & Hacker, J. M. (2011). The role of land use change on the development and evolution of the west coast trough, convective clouds, and precipitation in southwest Australia. *Journal of Geophysical Research Atmospheres*, *116*(7), 1–12. https://doi.org/10.1029/2010JD014950

Nicholson, S. E. (1988). Land surface atmosphere interaction: Physical processes and surface changes and their impact. *Progress in Physical Geography*, *12*(1), 36–65. https://doi.org/10.1177/030913338801200102

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

Painter, T. H., Rittger, K., McKenzie, C., Slaughter, P., Davis, R. E., & Dozier, J. (2009). Retrieval of subpixel snow covered area, grain size, and albedo from MODIS. *Remote Sensing of Environment*, *113*(4), 868–879. https://doi.org/10.1016/J.RSE.2009.01.001

Paleri, S., Butterworth, B., & Desai, A. R. (2022). *Here, there, and everywhere: Spatial patterns* *and scales*. *Conceptual Boundary Layer Meteorology: The Air Near Here*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-817092-2.00009-6

Parinussa, R. M., Holmes, T. R. H., Wanders, N., Dorigo, W. A., & De Jeu, R. A. M. (2015). A Preliminary Study toward Consistent Soil Moisture from AMSR2. *Journal of Hydrometeorology*, *16*(2), 932–947. https://doi.org/10.1175/JHM-D-13-0200.1

Petrenko, B., Ignatov, A., Kihai, Y., & Heidinger, A. (2010). Clear-sky mask for the advanced clear-sky processor for oceans. *Journal of Atmospheric and Oceanic Technology*, *27*(10), 1609–1623. https://doi.org/10.1175/2010JTECHA1413.1

Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J., & Dash, P. (2014). Evaluation and selection of SST regression algorithms for JPSS VIIRS. *Journal of Geophysical Research: Atmospheres*, *119*(8), 4580–4599. https://doi.org/10.1002/2013JD020637

Phillips, T. J., Klein, S. A., Ma, H. Y., Tang, Q., Xie, S., Williams, I. N., et al. (2017). Using ARM Observations to Evaluate Climate Model Simulations of Land-Atmosphere Coupling on the U.S. Southern Great Plains. *Journal of Geophysical Research: Atmospheres*, *122*(21), 11,524-11,548. https://doi.org/10.1002/2017JD027141

Pielke, R. A. (2001). Influence of the spatial distribution of vegetation and soils on the prediction of cumulus Convective rainfall. *Reviews of Geophysics*, *39*(2), 151–177. https://doi.org/10.1029/1999RG000072

Polehampton, E., Cox, C., Smith, D., Ghent, D., Wooster, M., Xu, W., et al. (2022). *Copernicus Sentinel-3 SLSTR Land User Handbook*. Retrieved from https://sentinel.esa.int/documents/247904/4598082/Sentinel-3-SLSTR-Land-Handbook.pdf/

Poltoradnev, M., Ingwersen, J., & Streck, T. (2016). Spatial and Temporal Variability of Soil Water Content in Two Regions of Southwest Germany during a Three-Year Observation Period. *Vadose Zone Journal*, *15*(6), vzj2015.11.0143. https://doi.org/10.2136/vzj2015.11.0143

Van der Putten, W. H., Bardgett, R. D., Bever, J. D., Bezemer, T. M., Casper, B. B., Fukami, T., et al. (2013). Plant-soil feedbacks: The past, the present and future challenges. *Journal of Ecology*, *101*(2), 265–276. https://doi.org/10.1111/1365-2745.12054

Qu, W., Bogena, H. R., Huisman, J. A., Martinez, G., Pachepsky, Y. A., & Vereecken, H. (2014). Effects of Soil Hydraulic Properties on the Spatial Variability of Soil Water Content: Evidence from Sensor Network Data and Inverse Modeling. *Vadose Zone Journal*, *13*(12), vzj2014.07.0099. https://doi.org/10.2136/vzj2014.07.0099

Risser, M. D., & Turek, D. (2020). Bayesian inference for high-dimensional nonstationary Gaussian processes. *Journal of Statistical Computation and Simulation*, 2902–2928. https://doi.org/10.1080/00949655.2020.1792472

Rodríguez-Iturbe, I., & Mejía, J. M. (1974). The design of rainfall networks in time and space. *Water Resources Research*, *10*(4), 713–728. https://doi.org/10.1029/WR010i004p00713

Rouhani, S., & Myers, D. E. (1990). Problems in space-time kriging of geohydrological data. *Mathematical Geology*, *22*(5), 611–623. https://doi.org/10.1007/BF00890508

Rudi, J., Pabel, R., Jager, G., Koch, R., Kunoth, A., & Bogena, H. (2010). Multiscale Analysis of Hydrologic Time Series Data using the Hilbert-Huang Transform. *Vadose Zone Journal*, *9*(4), 925–942. https://doi.org/10.2136/vzj2009.0163

Running, S., Mu, Q., Zhao, M., & Moreno, A. (2019). MOD16A2GF MODIS/Terra Net Evapotranspiration Gap-Filled 8-Day L4 Global 500 m SIN Grid V006 [Data set]. https://doi.org/https://doi.org/10.5067/MODIS/MOD16A2GF.006

Schepanski, K., Klüser, L., Heinold, B., & Tegen, I. (2015). Spatial and temporal correlation length as a measure for the stationarity of atmospheric dust aerosol distribution. *Atmospheric Environment*, *122*, 10–21. https://doi.org/10.1016/j.atmosenv.2015.09.034

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

Seyfried, M., Link, T., Marks, D., & Murdock, M. (2016). Soil Temperature Variability in Complex Terrain Measured Using Fiber-Optic Distributed Temperature Sensing. *Vadose Zone Journal*, *15*(6), vzj2015.09.0128. https://doi.org/10.2136/vzj2015.09.0128

Shi, L., & Bates, J. J. (2011). Three decades of intersatellite-calibrated High-Resolution Infrared Radiation Sounder upper tropospheric water vapor. *Journal of Geophysical Research: Atmospheres*, *116*(D4), 4108. https://doi.org/10.1029/2010JD014847

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

Sobrino, J. A., Li, Z. L., & Stoll, M. P. (1993). Impact of the Atmospheric Transmittance and Total Water Vapor Content in the Algorithms for Estimating Satellite Sea Surface Temperatures. *IEEE Transactions on Geoscience and Remote Sensing*, *31*(5), 946–952. https://doi.org/10.1109/36.263765

Stein, M. L. (1999). *Interpolation of Spatial Data : Some Theory for Kriging*. New York, NY: Springer.

Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, *100*(469), 310–321. https://doi.org/10.1198/016214504000000854

Stisen, S., McCabe, M. F., Refsgaard, J. C., Lerer, S., & Butts, M. B. (2011). Model parameter analysis using remotely sensed pattern information in a multi-constraint framework. *Journal of Hydrology*, *409*(1–2), 337–349. https://doi.org/10.1016/j.jhydrol.2011.08.030

Stisen, S., Soltani, M., Mendiguren, G., Langkilde, H., Garcia, M., & Koch, J. (2021). Spatial patterns in actual evapotranspiration climatologies for europe. *Remote Sensing*, *13*(12). https://doi.org/10.3390/rs13122410

Stoy, P. C., Katul, G., Siqueira, M. B. S., Juang, J. Y., McCarthy, H. R., Kim, H. S., et al. (2005). Variability in net ecosystem exchange from hourly to inter-annual time scales at adjacent pine and hardwood forests: A wavelet analysis. *Tree Physiology*, *25*(7), 887–902. https://doi.org/10.1093/treephys/25.7.887

Su, Z. (2002). The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes. *Hydrology and Earth System Sciences*, *6*(1), 85–99. https://doi.org/10.5194/hess-6-85-2002

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

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

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505. https://doi.org/10.1126/SCIENCE.1099192/SUPPL_FILE/TAPLEY.SOM.PDF

Taylor, C. M., Parker, D. J., & Harris, P. P. (2007). An observational case study of mesoscale atmospheric circulations induced by soil moisture. *Geophysical Research Letters*, *34*(15). https://doi.org/10.1029/2007GL030572

Taylor, C. M., Parker, D. J., Kalthoff, N., Gaertner, M. A., Philippon, N., Bastin, S., et al. (2011). New perspectives on land-atmosphere feedbacks from the African Monsoon Multidisciplinary Analysis. *Atmospheric Science Letters*, *12*(1), 38–44. https://doi.org/10.1002/asl.336

Taylor, C. M., Birch, C. E., Parker, D. J., Dixon, N., Guichard, F., Nikulin, G., & Lister, G. M. S. (2013). Modeling soil moisture-precipitation feedback in the Sahel: Importance of spatial scale versus convective parameterization. *Geophysical Research Letters*, *40*(23), 6213–6218. https://doi.org/10.1002/2013GL058511

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. *Geosci. Model Dev*, *7*, 947–963. https://doi.org/10.5194/gmd-7-947-2014

Timmermans, W. J., Bertoldi, G., Albertson, J. D., Olioso, A., Su, Z., & Gieske, A. S. M. (2010). Accounting for atmospheric boundary layer variability on flux estimation from RS observations. *Https://Doi-Org.Proxy.Lib.Duke.Edu/10.1080/01431160802036383*, *29*(17–18), 5275–5290. https://doi.org/10.1080/01431160802036383

Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, *46*, 234. https://doi.org/10.2307/143141

Torres-Rojas, L., & Chaney, N. W. (2023). Supporting Dataset: The observed spatio-temporal patterns of Land surface temperature over the Contiguous United States. https://doi.org/10.5281/ZENODO.8428629

Torres-Rojas, L., Vergopolan, N., Herman, J. D., & Chaney, N. W. (2022). Towards an Optimal Representation of Sub-Grid Heterogeneity in Land Surface Models. *Water Resources Research*, *58*(12), e2022WR032233. https://doi.org/10.1029/2022WR032233

Tsai, Y.-L. S., Dietz, A., Oppelt, N., Kuenzer, C., De (A. D., Kuenzer@dlr, C., & De, C. K. (2019). Remote Sensing of Snow Cover Using Spaceborne SAR: A Review. *Remote Sensing 2019, Vol. 11, Page 1456*, *11*(12), 1456. https://doi.org/10.3390/RS11121456

Tuttle, S., & Salvucci, G. (2016). Empirical evidence of contrasting soil moisture–precipitation feedbacks across the United States. *Science*, *352*(6287), 825–828. https://doi.org/10.1126/science.aaa7185

Vargas, R., Detto, M., Baldocchi, D. D., & Allen, M. F. (2010). Multiscale analysis of temporal variability of soil CO2 production as influenced by weather and vegetation. *Global Change Biology*, *16*(5), 1589–1605. https://doi.org/10.1111/j.1365-2486.2009.02111.x

Varouchakis, E. A. (2018). *Background of spatiotemporal geostatistical analysis: Application to aquifer level mapping*. *Spatiotemporal Analysis of Extreme Hydrological Events*. Elsevier Inc. https://doi.org/10.1016/B978-0-12-811689-0.00002-1

Vereecken, H., Huisman, J. A., Bogena, H., Vanderborght, J., Vrugt, J. A., & Hopmans, J. W. (2008). On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resources Research*, *46*(4). https://doi.org/10.1029/2008WR006829

Vereecken, H., Pachepsky, Y., Simmer, C., Rihani, J., Kunoth, A., Korres, W., et al. (2016). On the role of patterns in understanding the functioning of soil-vegetation-atmosphere systems. *Journal of Hydrology*, *542*, 63–86. https://doi.org/10.1016/j.jhydrol.2016.08.053

Vergopolan, N., Chaney, N. W., Pan, M., Sheffield, J., Beck, H. E., Ferguson, C. R., et al. (2021). SMAP-HydroBlocks, a 30-m satellite-based soil moisture dataset for the conterminous US. *Scientific Data 2021 8:1*, *8*(1), 1–11. https://doi.org/10.1038/s41597-021-01050-2

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

Wagner, D., Ruprecht, E., & Simmer, C. (1990). A Combination of Microwave Observations from Satellites and an EOF Analysis to Retrieve Vertical Humidity Profiles over the Ocean. *Journal of Applied Meteorology*, *29*(11), 1142–1157. https://doi.org/10.1175/1520-0450(1990)029<1142:ACOMOF>2.0.CO;2

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., et al. (2013). The ASCAT Soil Moisture Product: A Review of its Specifications, Validation Results, and Emerging Applications. *Meteorologische Zeitschrift*, *22*(1), 5–33. https://doi.org/10.1127/0941-2948/2013/0399

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

Wan, Z. (1996). A generalized split-window algorithm for retrieving land-surface temperature from space. *IEEE Transactions on Geoscience and Remote Sensing*, *34*(4), 892–905. https://doi.org/10.1109/36.508406

Wan, Z. (2014). New refinements and validation of the collection-6 MODIS land-surface temperature/emissivity product. *Remote Sensing of Environment*, *140*, 36–45. https://doi.org/10.1016/J.RSE.2013.08.027

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

Wang, Y., Liu, H., Liu, Y., Wang, S., Wang, L., & Li, X. (2023). Effect of land–atmosphere process parameterizations on the PM simulation of a river valley city with complex topography. *Atmospheric Research*, *281*(October 2022), 106505. https://doi.org/10.1016/j.atmosres.2022.106505

Wealands, S. R., Grayson, R. B., & Walker, J. P. (2005). Quantitative comparison of spatial fields for hydrological model assessment - Some promising approaches. *Advances in Water Resources*, *28*(1), 15–32. https://doi.org/10.1016/j.advwatres.2004.10.001

Weaver, C. P. (2004). Coupling between large-scale atmospheric processes and mesoscale land-atmosphere interactions in the U.S. Southern Great Plains during summer. Part I: Case studies. *Journal of Hydrometeorology*, *5*(6), 1223–1246. https://doi.org/10.1175/JHM-396.1

Wu, C., Lo, M., Chen, W., & Lu, C. (2015). The impacts of heterogeneous land surface fluxes on the diurnal cycle precipitation: A framework for improving the GCM representation of land-atmosphere interactions. *Journal of Geophysical Research: Atmospheres*, *120*(9), 3714–3727. https://doi.org/10.1002/2014JD023030

Xiao, M., Mascaro, G., Wang, Z., Whitney, K. M., & Vivoni, E. R. (2022). On the value of satellite remote sensing to reduce uncertainties of regional simulations of the Colorado River. *Hydrology and Earth System Sciences*, *26*(21), 5627–5646. https://doi.org/10.5194/hess-26-5627-2022

Xu, F., & Ignatov, A. (2014). In situ SST quality monitor (iQuam). *Journal of Atmospheric and Oceanic Technology*, *31*(1), 164–180. https://doi.org/10.1175/JTECH-D-13-00121.1

Yu, Y., & Yu, P. (2020). Land Surface Temperature Product from the GOES-R Series. In *The GOES-R Series* (pp. 133–144). Elsevier. https://doi.org/10.1016/B978-0-12-814327-8.00012-3

Yu, Y., Tarpley, D., Hui, X., & Chen, M. (2012). *GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document For Land Surface Temperature* (Version 2.). NOAA NESDIS, Center for Satellite Applications and Research.

Zink, M., Mai, J., Cuntz, M., & Samaniego, L. (2018). Conditioning a Hydrologic Model Using Patterns of Remotely Sensed Land Surface Temperature. *Water Resources Research*, *54*(4), 2976–2998. https://doi.org/10.1002/2017WR021346

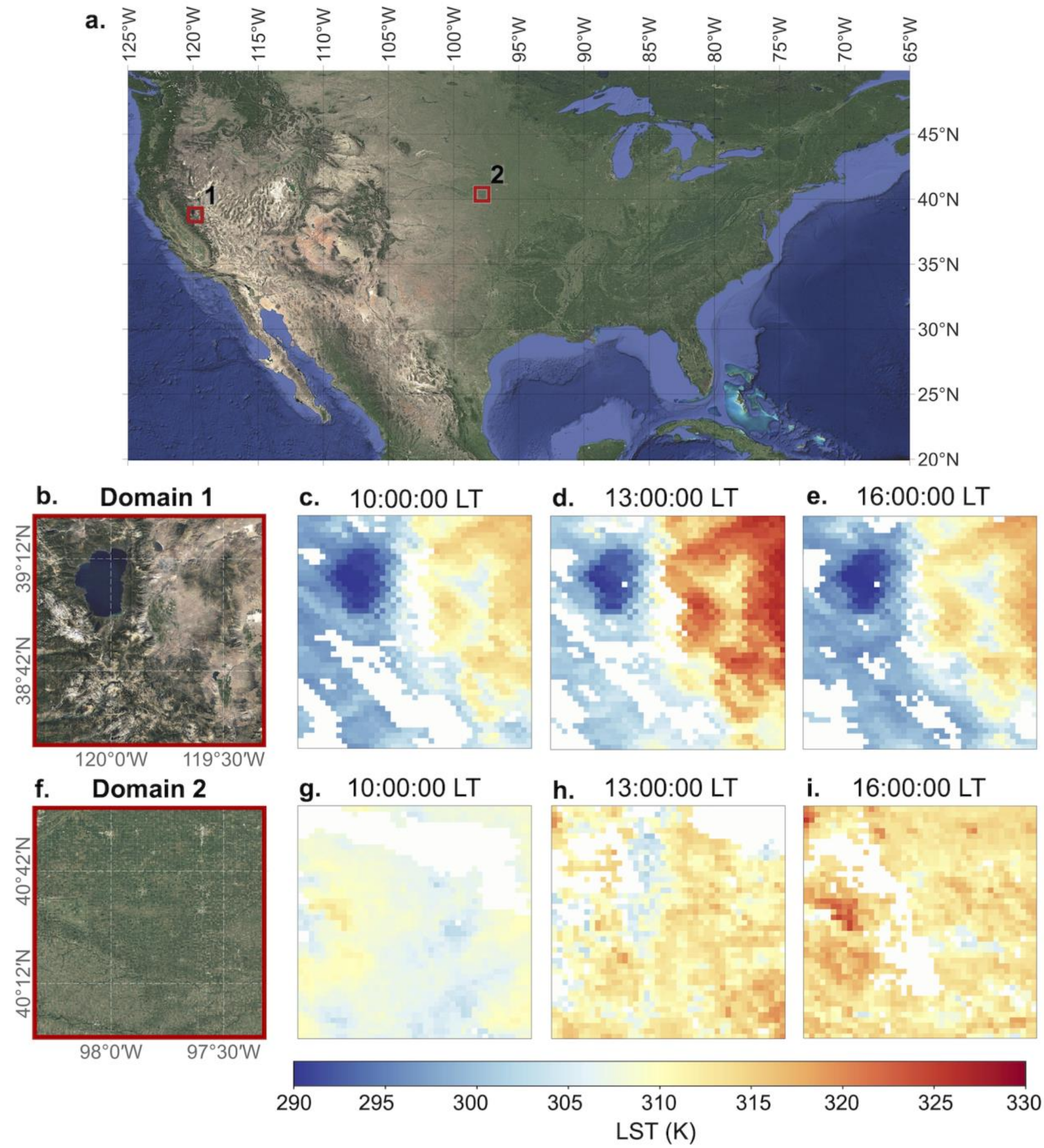

Figure 1. Temporal evolution of LST over two 1.0°x1.0° domains over CONUS on 2020-07-04. a) Location of the domains within CONUS; b and f) zoomed-in satellite visible imagery with coordinates, c and g) GOES-16 derived LST for 10:00:00 local time (LT); d and h) 13:00:00 LT, and e and i) 16:00:00 LT.

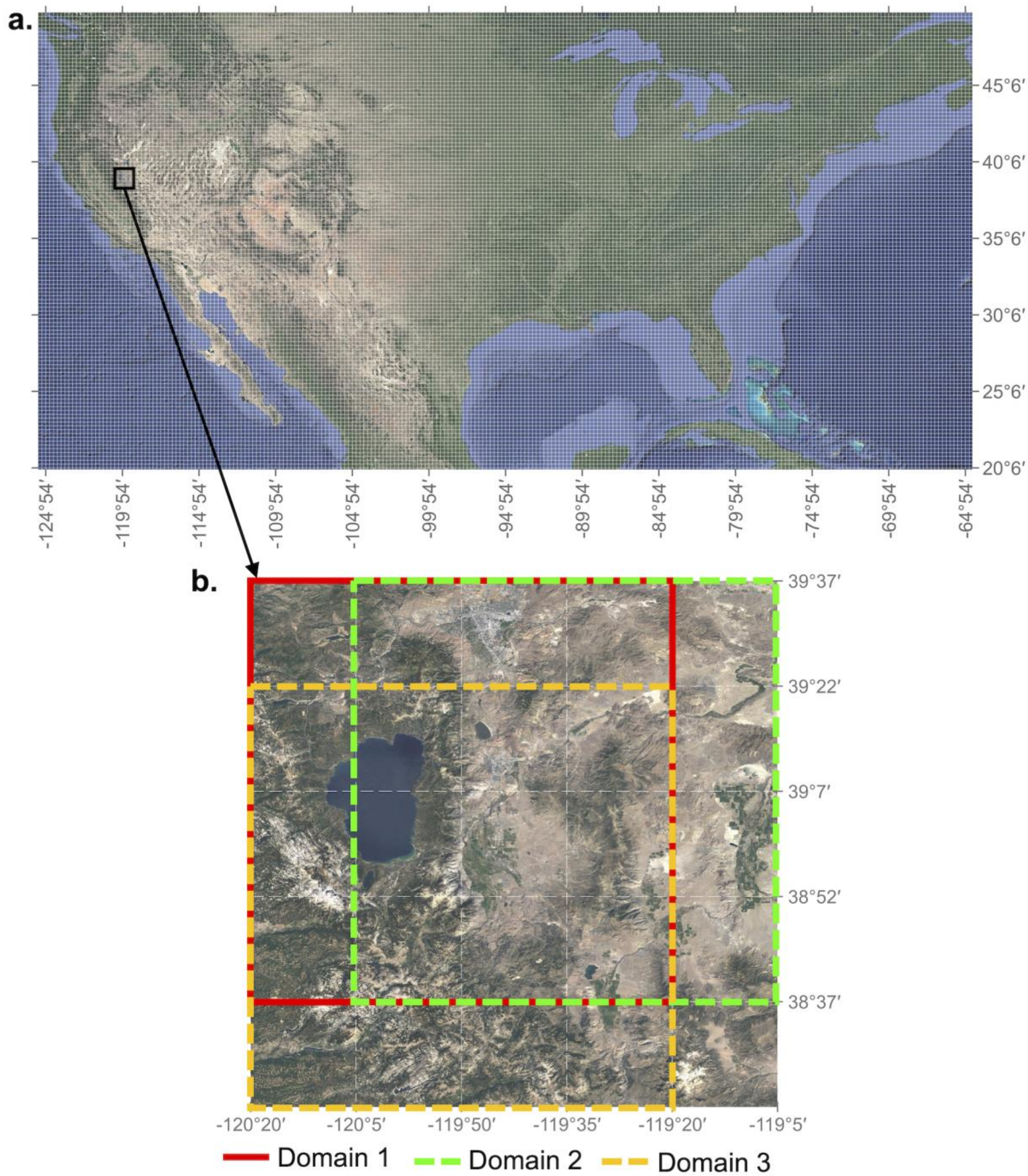

Figure 2. Study domain over CONUS and detail of the sliding window approach used. a) Study domain over CONUS; the central 0.25° of each 1.0°x1.0° squared box obtained from the sliding window approach is presented as the grid; coordinates every 5° are presented to aid in georeferencing. b) Detail of the sliding window approach used for sampling over a domain in the California-Nevada border. Three (3) 1.0°x1.0° domains separated by 0.25° from domain 1 in the horizontal (domain 2) and vertical (domain 3) directions are displayed.

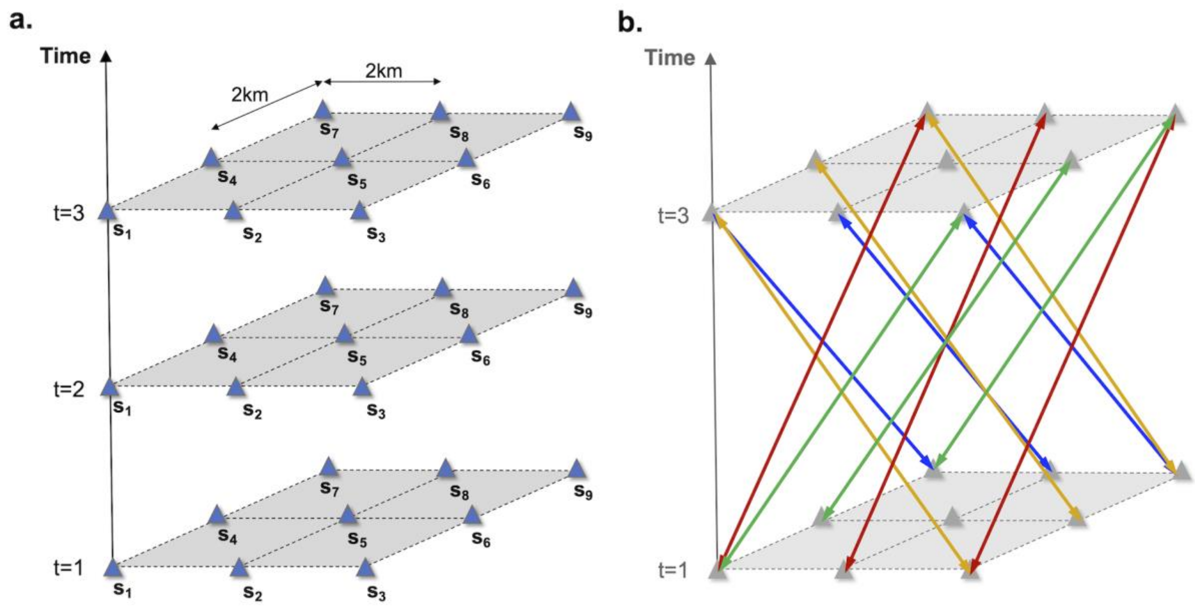

Figure 3. a) Schematic view of a regularly-spaced 2-kilometer grid over a time axis with 3 compartments; b) 12 pairs of points separated by a spatial distance of 4km and a temporal lag of 2 ($(h, u) = (4km, 2)$), colors represent sides of the spatial grid where the origin point is located: yellow for left, red for back, green for right, and blue for front.

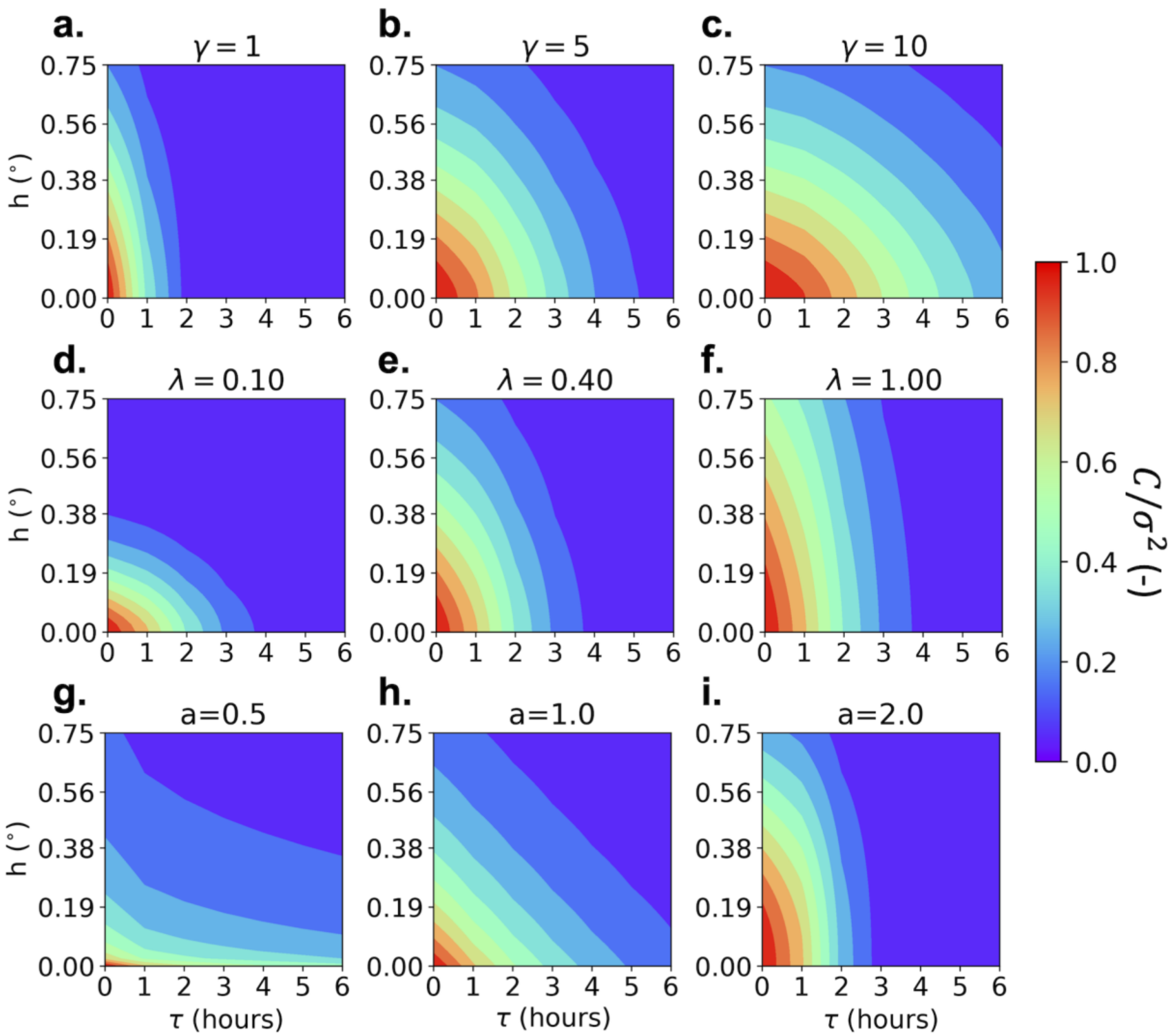

Figure 4. Parametric spatio-temporal covariance function obtained by individually increasing the values of the three main model parameters: a, b, and c) varying γ with $\lambda = 0.4$ and $a = 1.5$; d, e, and f) varying λ with $\gamma = 3$ and $a = 1.5$; and g, h, and i) varying a with $\lambda = 0.4$ and $\gamma = 3$.

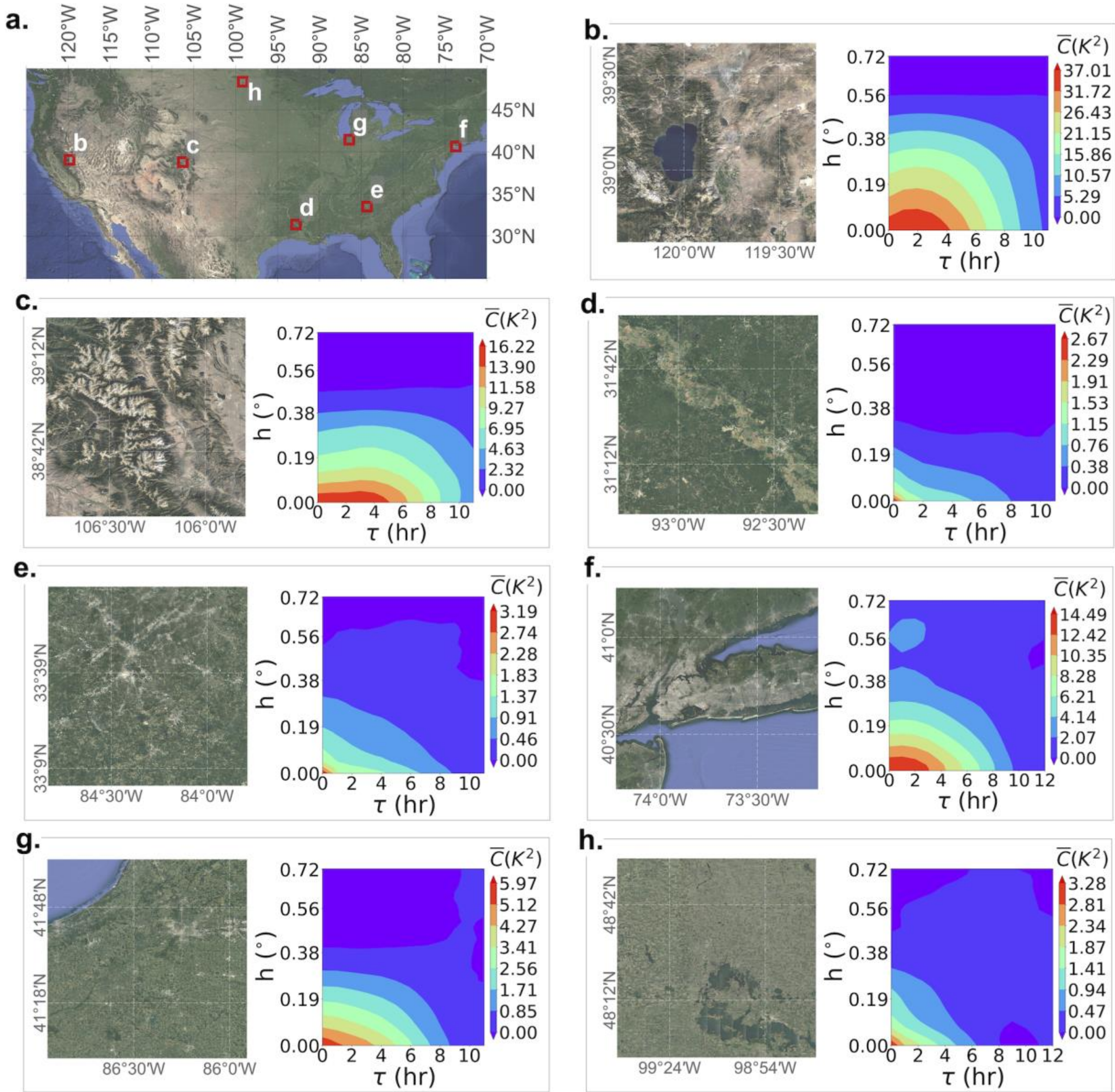

Figure 5. Zoomed-in satellite visible imagery with coordinates for seven 1.0°x1.0° domains over CONUS. The obtained summer daytime LST ESTCFs for each domain are also presented. a) Location of the seven domains within CONUS; visible satellite imagery of the landscape and computed ESTCF for b) the Lake Tahoe area, California-Nevada border; c) the Mount Mitchell area, Colorado; d) Mississippi River, Lousiana; e) Atlanta, Georgia; f) New York City; g) Lake Michigan shore, Indiana-Michigan border; h) Leeds county, North Dakota.

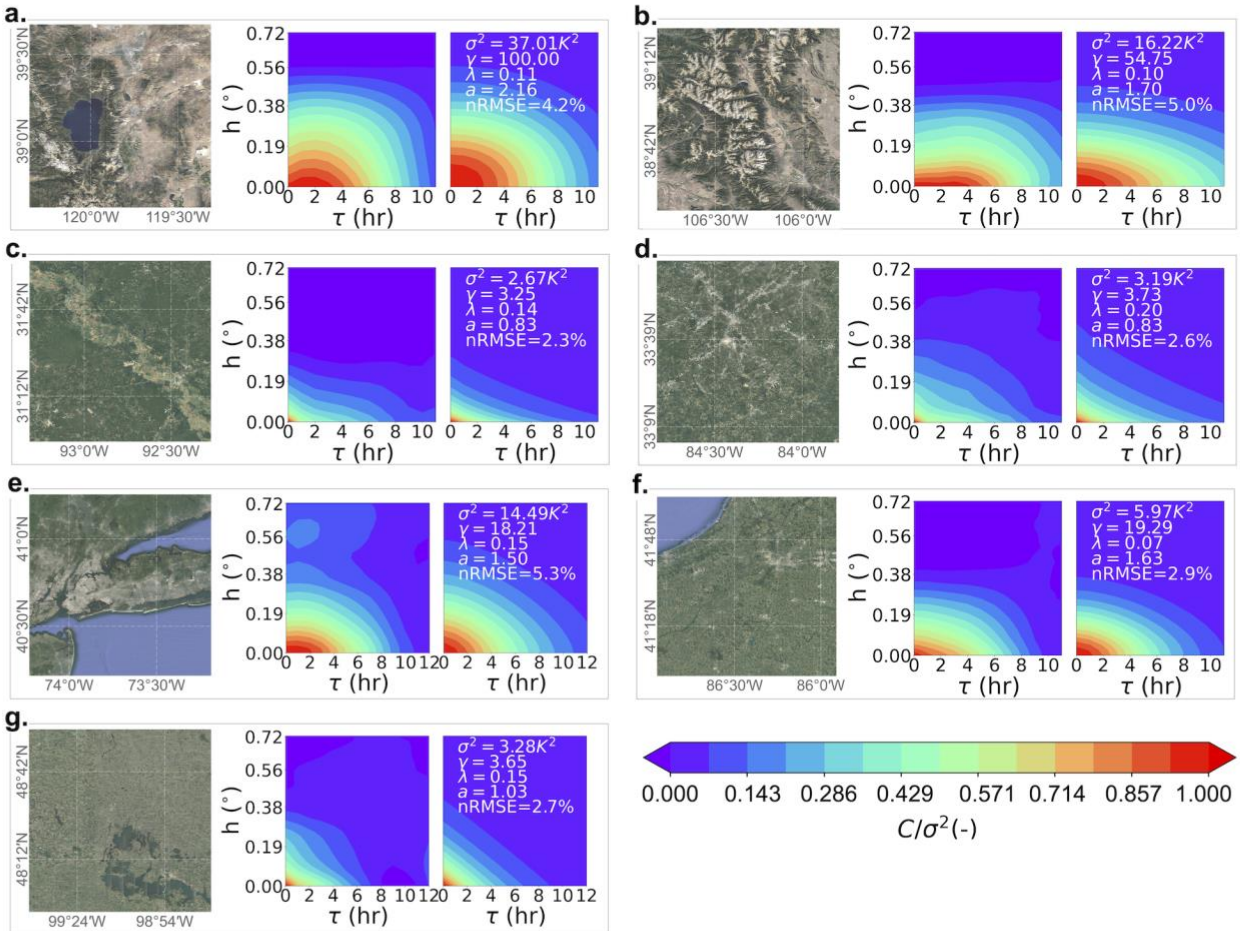

Figure 6. Zoomed-in satellite visible imagery with coordinates for the seven 1.0°x1.0° domains over CONUS. The obtained summer daytime LST ESTCFs for each domain and the fitted spatio-temporal covariance function parametric model are presented. The plots for the fitted cases include the obtained set of parameters and the normalized root mean square error (nRMSE) for the fit. Visible satellite imagery of the landscape, computed ESTCF, and fitted spatio-temporal covariance function parametric model for a) the Lake Tahoe area, California-Nevada border; b) the Mount Mitchell area, Colorado; c) Mississippi River, Lousiana; d) Atlanta, Georgia; e) New York City; f) Lake Michigan shore, Indiana-Michigan border; g) Leeds county, North Dakota.

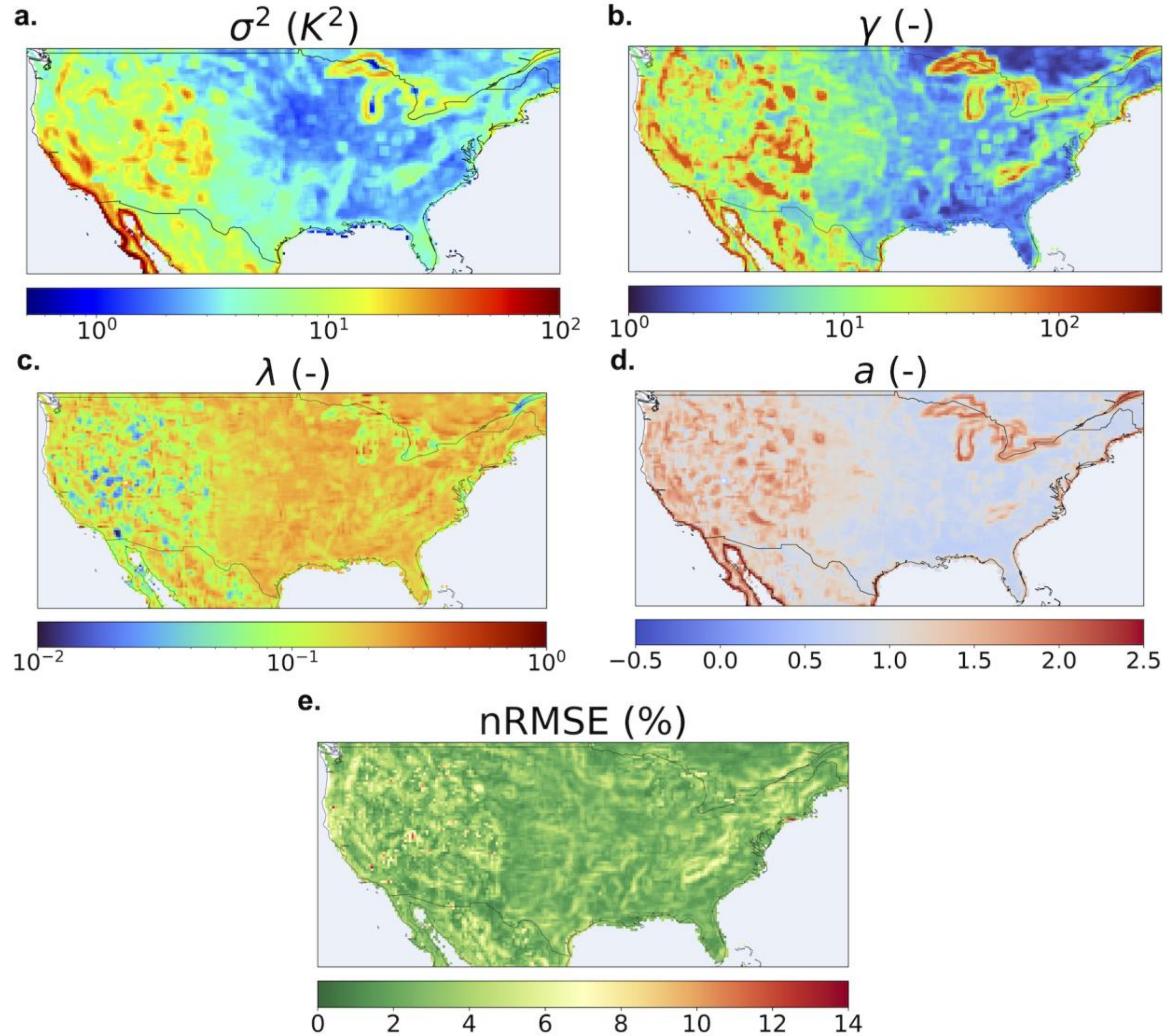

Figure 7. Maps of results for the parametric spatio-temporal covariance function fit for land-containing domains. Each pixel represents the central 0.25°x0.25° for each 1.0°x1.0° analyzed domain. a) Computed spatio-temporal variance, b) fitted temporal characteristic length-scale, c) fitted spatial characteristic length-scale, d) fitted spatio-temporal interaction exponent, and e) nRMSE for the parametric fit.

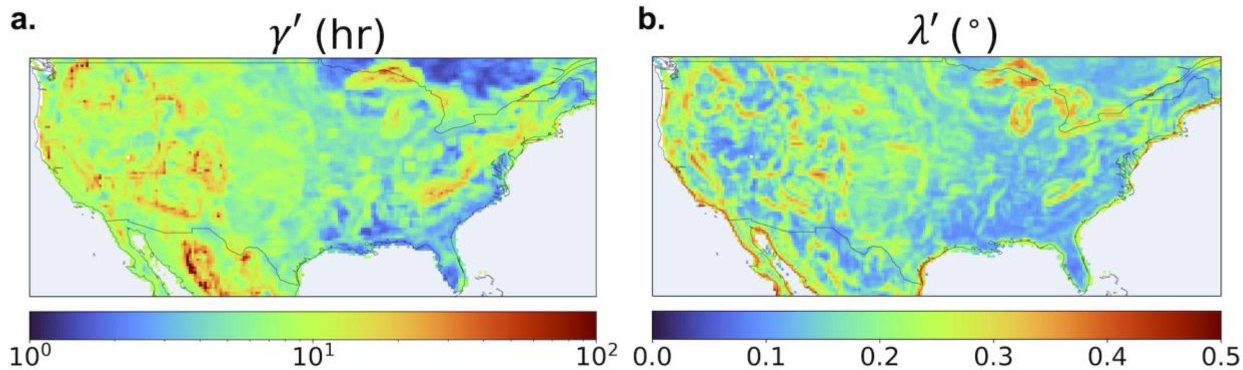

Figure 8. Maps of results for the modified forms of the spatio-temporal characteristic length scales for land-containing domains. Each pixel represents the central 0.25°x0.25° for each 1.0°x1.0° analyzed domain. a) Modified temporal characteristic length-scale, b) modified spatial characteristic length-scale.

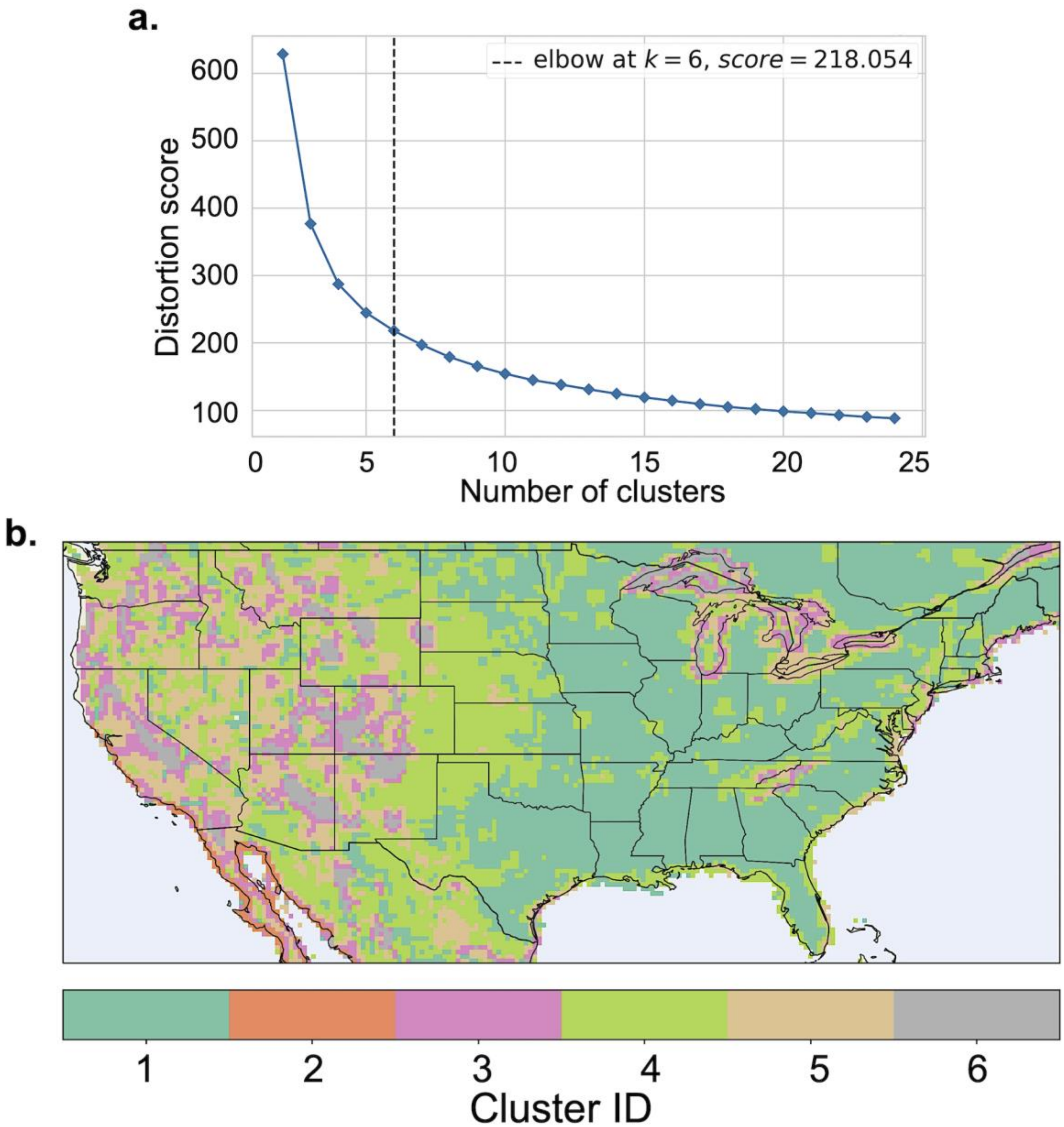

Figure 9. a) Elbow diagram for cluster number determination and, b) spatial maps of obtained clusters over CONUS.

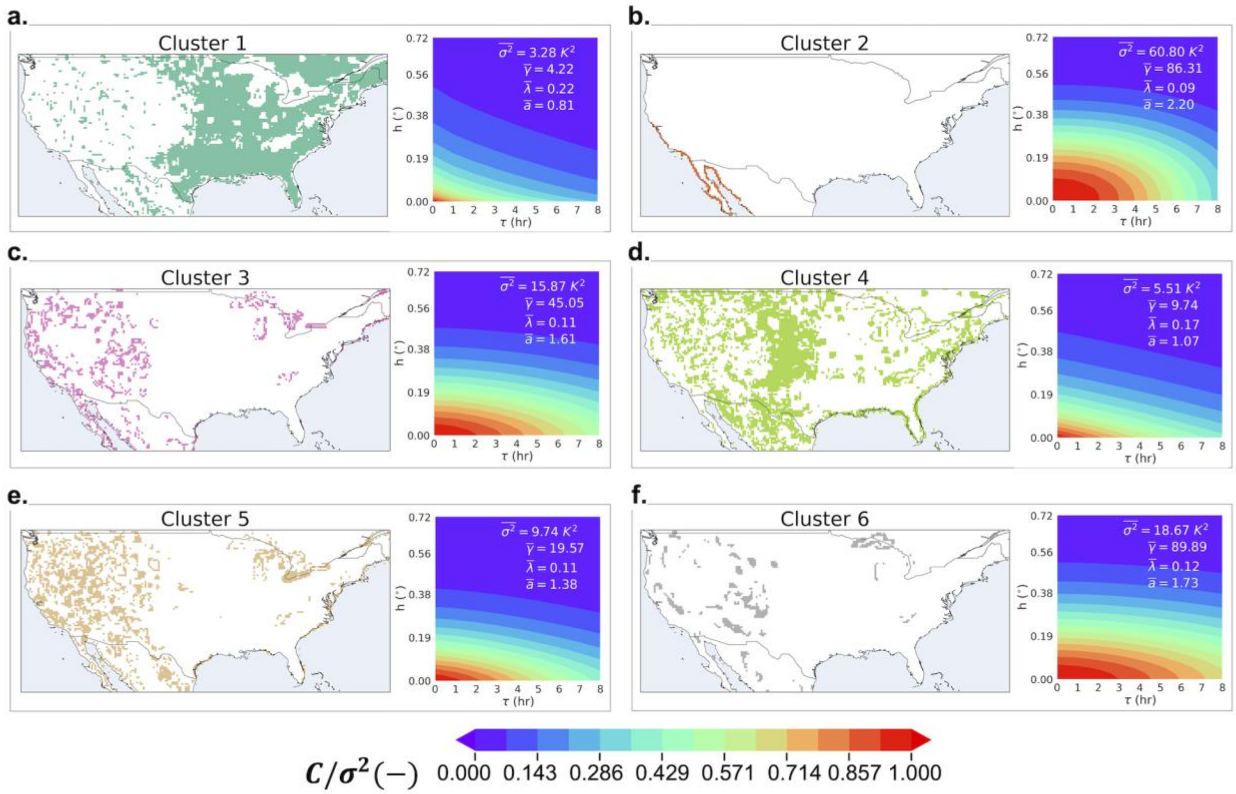

Figure 10. Individual spatial mapping of the clusters over CONUS, next to the corresponding characteristic spatio-temporal covariance function (CSTCF) obtained as the mean cluster value of the parameters for a) cluster 1, b) cluster 2, c) cluster 3, d) cluster 4, e) cluster 5, and f) cluster 6.

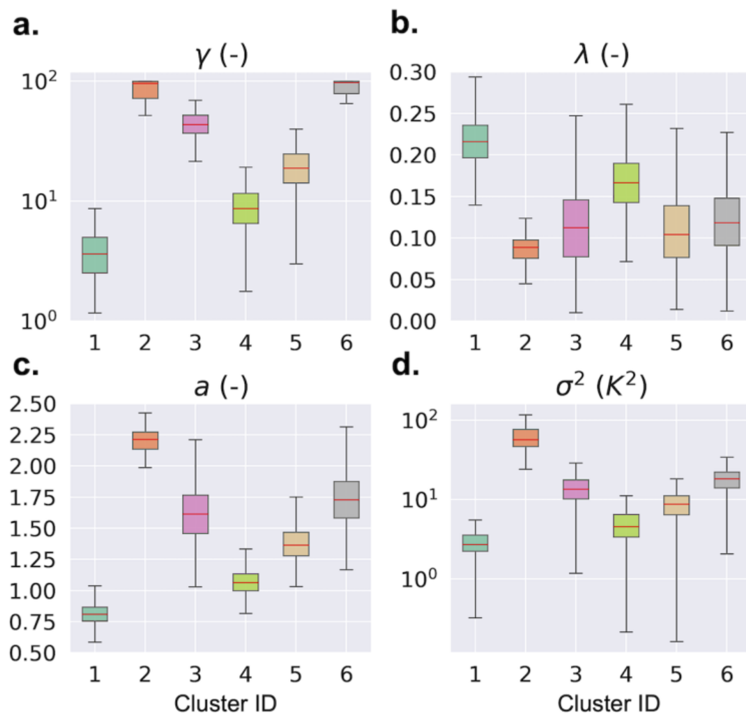

Figure 11. Box plots of parameters distributions among clusters: a) characteristic temporal length-scale, b) characteristic spatial length-scale, c) spatio-temporal interaction exponent, and d) spatio-temporal variance.

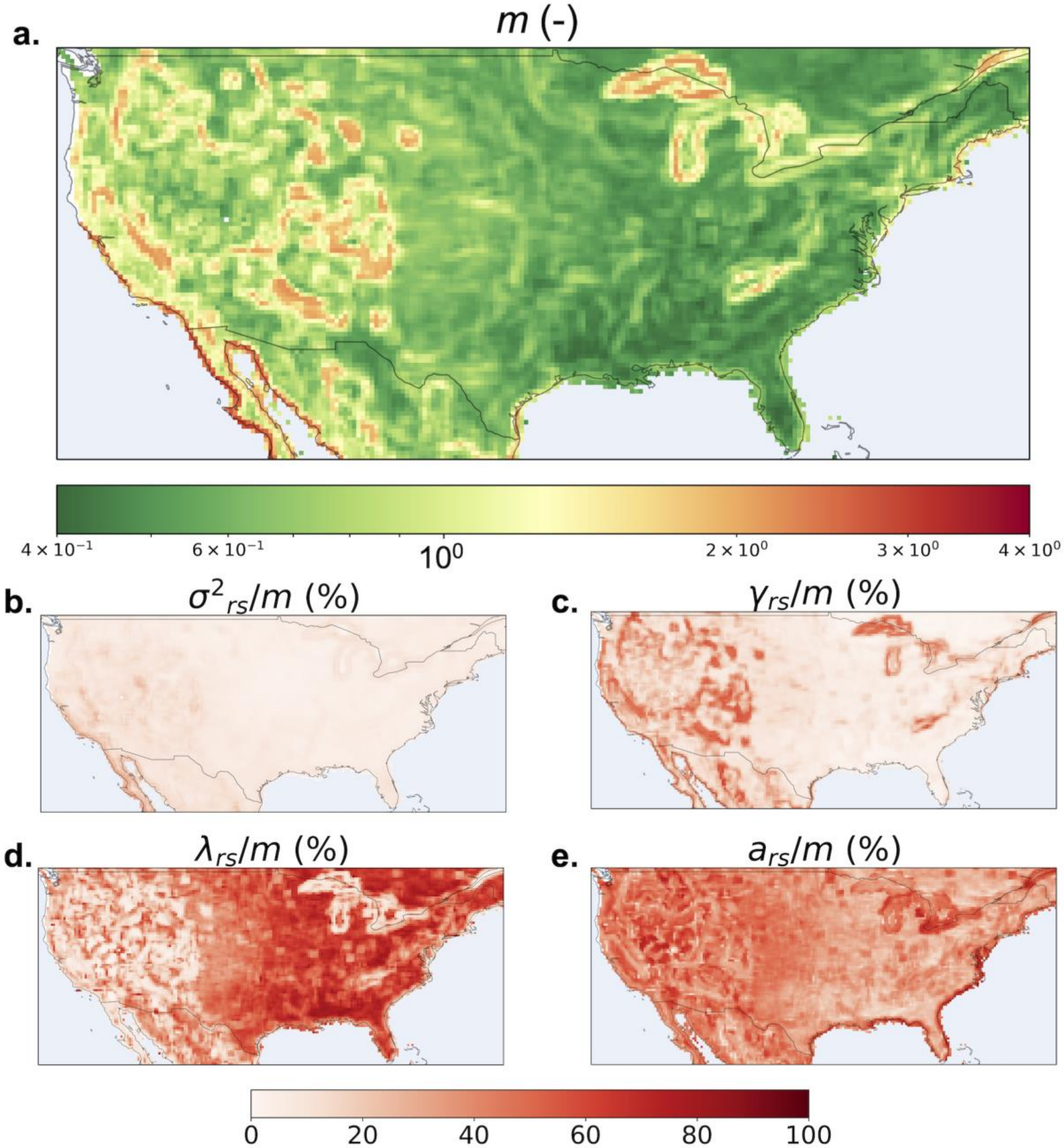

Figure 12. a) Combined metric for spatio-temporal persistence for every domain in CONUS along with individual contributions of rescaled forms of model parameters to the total metric value: b) spatio-temporal variance, c) temporal characteristic length-scale, d) spatial characteristic length-scale, and e) spatio-temporal interaction exponent.