Check for updates

Joint Characterization of Spatiotemporal Data Manifolds

Daniel Sousa[1]* and Christopher Small[2]

[1]Department of Geography, San Diego State University, San Diego, CA, United States, [2]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, United States

Modeling spatiotemporal data can be a challenge due to the plethora of processes, both independent and interacting, which may or may not contribute to the measurements. Characterization can be considered a complement to modeling by helping guide assumptions about generative processes and their representation in the data. For high-D signals, Dimensionality Reduction (DR) is a frequently implemented type of characterization designed to mitigate the effects of the so-called "curse of dimensionality". For decades, Principal Component (PC) and Empirical Orthogonal Function (EOF) analysis has been used as a linear, invertible approach to dimensionality reduction and spatiotemporal analysis. Recent years have seen the additional development of a suite of nonlinear DR algorithms, frequently categorized as "manifold learning". Here, we explore the idea of joint characterization of spatiotemporal data manifolds using the PC/EOF approach alongside two nonlinear DR approaches: Laplacian Eigenmaps (LE) and t-distributed Stochastic Neighbor Embedding (t-SNE). Starting with a synthetic example and progressing to global, regional, and field scale spatiotemporal datasets spanning roughly 5 orders of spatial magnitude and 2 orders of temporal magnitude, we show these three DR approaches can yield complementary information about the topology of spatiotemporal data manifolds. Compared to the PC/EOF projections, the nonlinear DR approaches yield more compact manifolds with decreased ambiguity in temporal endmembers (LE) and/or in spatiotemporal clustering (t-SNE), compared to the relatively diffuse temporal feature space produced by the PC/EOF approach. However, these properties are compensated by the greater interpretability of PCs and EOFs than of the LE or t-SNE dimensions, as well as significantly lower computational demand and diminished sensitivity to spatial aliasing for PCs/EOFs than LE or t-SNE. Taken together, we find the joint characterization using the three complementary DR approaches capable of providing substantially greater insight about the generative processes represented in spatiotemporal datasets than is possible using any single approach alone. This parsimonious, complementary characterization of both local manifold structure and global variance can advance remote sensing time series analysis by providing important context to constrain and guide design of effective spatiotemporal models.

Keywords: spatiotemporal, characterization, dimensionality reduction (DR), temporal feature space (TFS), manifold, topology, t-SNE (t-Distributed Stochastic Neighbor Embedding), Laplacian Eigenmaps (LE)

INTRODUCTION

From agriculture to coastal erosion, and from vehicle traffic to disease transmission, many phenomena on Earth's surface are inherently *spatiotemporal*: that is, variable across both space and time. But despite the ubiquity of spatiotemporal processes, meaningful quantitative analysis has been limited for centuries by observational and computational capacity (Eshel, 2011; Christakos, 2017). Fortunately, in recent years drastic reductions in costs to sense, transmit, store, and process spatiotemporal observations have inverted this centuries-old paradigm, leading to the dawn of the era of so-called "Big Data".

The data analysis landscape has thus fundamentally shifted: today, spatiotemporal observations abound, and scientists are in need of effective and efficient tools to analyze patterns, inform modeling, and ultimately discriminate between signal and noise. This asymmetry between the volume of observations and capacity of inference tools is imperfectly represented by a comparison of Google Ngram word usage for" big data" versus "spatiotemporal" (**Figure 1**). It is perhaps an illustrative coincidence that the visual departure of the "big data" usage curve occurs in 2008, the year the Landsat satellite image archive was made freely available to the public (Woodcock et al., 2008).

Spatiotemporal processes frequently possess some (or all) of a suite of challenging properties. Such characteristics include high dimensionality, possibility of both abrupt and gradual changes (Verbesselt et al., 2010), spatial and temporal autocorrelation (Henebry, 1995), and cross-variable coupling (Lotsch et al., 2003).

Together, these factors present formidable analytic challenges which have driven over a century of development and refinement of analytic tools (Pearson, 1901; Lorenz, 1956; Ng et al., 2002; van der Maaten and Hinton, 2008).

Characterization of spatiotemporal signals complements modeling by informing presuppositions about the number, identity, and interaction of real-world generative processes which may be imperfectly represented by a set of observations (Small, 2012). Accurate characterization can be challenging for high dimensional (high-D) spatiotemporal signals, due to a range of properties of high-D spaces popularly deemed the "curse of dimensionality" (Bellman, 1957). One approach to this challenging problem of spatiotemporal characterization focuses on analysis of the temporal feature space (TFS). Here, low-D projections of high-D spatiotemporal data are visualized and analyzed to characterize their underlying geometric and topological structure. Once characterized, this structure can then be used to design parsimonious, well-posed inverse models (Small, 2012). Conceptually, geometry implies algebra.

The TFS has been used to produce accurate characterization of complex spatiotemporal processes as diverse as oak woodland drought stress (Sousa and Davis, 2020), post-cyclone mangrove recovery (Small and Sousa, 2019), dynamics of commercial (Sousa and Small, 2019) and smallholder (Small, 2012) agriculture, cloud forest phenology and disturbance response (Sousa et al., 2019), urban development and nighttime light (Small and Sousa, 2016), and pathogen transmission (Small and Sousa, 2021). Currently, TFS characterization relies on

FIGURE 1 | Historical word usage for key terms. English usage of the term "spatiotemporal" has gradually increased for decades, while usage of "big data" has sharply spiked since 2008 **(A)**. Note the timing of the publication of Lorenz's scientific report on Empirical Orthogonal Functions and Weather Predidication (1956) and the opening of the Landsat archive (2008). Similarly, the term "principal component" has steadily increased since Karl Pearson's 1901 publication, but usage of the term "manifold learning" has been a more recent phenomenon **(B)**.

estimating the global variance structure of a spatiotemporal dataset using linear, mutually orthogonal, variance-ordered basis functions, i.e., Principal Component (PC) and Empirical Orthogonal Function (EOF) analysis (Pearson, 1901; Lorenz, 1956). For the remainder of the text, we use the term "PC/EOF" to refer to this approach to both reflect the intrinsic complementarity between PCs and EOFs, and to be inclusive of the range of nomenclature used by various subfields throughout the literature.

While effective in many ways, the linear PC/EOF approach alone has significant limitations. First, signals that comprise real spatiotemporal data are rarely truly orthogonal, so observed signals generally must be represented by linear combinations of two or more PCs/EOFs. Second, PCs/EOFs are based solely on overall (global) variance of the entire dataset and do not leverage potentially important local scale manifold topology and connectivity structure that may be present in a high-D TFS.

These limitations suggest that the information provided by PCs/EOFs might be complemented by additional dimensionality reduction tools which do not require strict linearity and/or are able to capture local scale feature space topology. These desired properties suggest the field of *manifold learning* as a potentially useful complement to PC/EOF analysis. Manifold learning refers to a class of nonlinear dimensionality reduction (DR) techniques specifically designed to preserve the connectivity structure of the feature space via observations that are deemed proximal according to a chosen statistical distance metric. The properties of this local connectivity structure are then examined, giving information which complements the global variance information given by PCs and their corresponding EOFs.

The popularity of nonlinear DR techniques has increased in recent years with improvements in computational capacity and open-source software packages like *scikit-learn* (Pedregosa et al., 2011). Such techniques are highly general, spanning a wide range of use cases. In spatiotemporal analysis, examples of nonlinear DR include Earth system modeling and data assimilation (Safaie et al., 2017), land cover classification from satellite imagery (Yan and Roy, 2015; Zhai et al., 2018), hyperspectral imaging (Bachmann et al., 2005; Gillis et al., 2005), human mobility (Watson et al., 2020), medical imaging (Kadoury, 2018), face recognition in video data (Hadid and Pietikäinen, 2009), and more.

This analysis explores the approach of joint characterization introduced by (Sousa and Small, 2021) for the case of spatiotemporal analysis of image time series. The approach used here focuses on the complementarity of linear and nonlinear dimensionality reduction algorithms. Each algorithm emphasizes a different aspect, or variance scale, of signal within a spatiotemporal dataset. Here, we implement one linear method (PCs/EOFs) and two nonlinear methods (Laplacian Eigenmaps, LE; and t-distributed Stochastic Neighbor Embedding; t-SNE). Because t-SNE has a stochastic element that makes a single realization effectively non-repeatable, we use a Monte Carlo approach in which the low order PCs of a number of independent 2D t-SNE realizations, *PC(t-SNE)*, are used to characterize consistently recurring structure within the t-SNE

feature spaces (Small and Sousa, 2022). While we focus on these two nonlinear methods due to their popularity and complementarity, we note that DR is an active area of research—new approaches continue to be developed, and other existing approaches may be more appropriate for a given application. For these reasons, the purpose of this work is not to provide a specific methodological recipe, but rather a general conceptual framework with which to consider analysis of spatiotemporal processes. Beginning with a synthetic example and progressing to global, regional, and field scale observational datasets, we ask the following questions:

1) Which geometric and topological characteristics does each dimensionality reduction algorithm accentuate or suppress when producing a low-D representation of the topological structure of a high-D spatiotemporal dataset?
2) How do the properties of each algorithm map onto strengths and weaknesses for specific spatiotemporal analysis applications?
3) In what ways can methods be used together to yield a more effective characterization than is possible with any single method alone?

In addressing these questions, we demonstrate a generalized analytic framework for joint characterization of spatiotemporal data manifolds. We extend the concept of the temporal feature space beyond the linear domain, illustrating the potential for meaningful signal to exist across multiple variance scales within the same spatiotemporal dataset, and for the ability of complementary DR approaches to jointly capture that signal.

BACKGROUND

Manifold learning techniques cast the question of characterization in a fundamentally different light. Here, statistical similarity is evaluated for pairs (or more generally n-tuples) of observations, with "similarity" defined according to the analyst's distance metric of choice (Van Der Maaten et al., 2009). One simple metric is Euclidean distance. In the case of a higher-D dataset, Euclidean distance cannot be visualized accurately solely on the basis of a single 2-D scatterplot, and instead takes the form of a higher-D generalization of the underlying principle. While a wide range of analytic approaches can be used to characterize the pairwise statistical similarity structure of a dataset, here we focus on two nonlinear manifold learning algorithms: t-SNE and LE. These two algorithms leverage neighborhood connectivity information in fundamentally different ways.

LE uses a graph theoretic approach, considering the high-D observations as an interconnected network of nodes and edges. Connectivity among proximal data points can be considered using a distance threshold or number of nearest neighbors. Once the graph of the observations is constructed, its matrix Laplacian is computed and decomposed into eigenvectors. The contribution of each Laplacian eigenvector to each data point is then known and can be used to characterize the data. For more

FIGURE 2 | Illustration with synthetic data. Linear combinations of sinusoidal (1) and decaying exponential (2 and 3) signals **(A)** are used to generate a 100 × 100 × 100 spatiotemporal cube. Weights for each signal are randomly chosen from uniform distribution and forced to sum to unity. PCA-based linear dimensionally reduction **(B)** reveals the low order temporal feature space to be occupied by a trigonal planar mixing manifold, with >99% of variance allocated to first two dimensions. PC(t-SNE) yields a higher dimensionality (88% of variance in first two dimensions) and more clustered topology. Red, green, and cyan clusters identified from the PC(t-SNE) space effectively identify the signal with dominant weighting in each pixel - at the expense of a clear representation of the mixing continuum **(C)**. Three clusters comprised of mixtures of signals 1 and 3 form an exception (dark cyan), mapping closer to the signal 1 group PC(t-SNE) space despite slightly higher signal 3 weights. Cluster colors are applied to the other two spaces for visual comparison. In contrast, the low-order LE dimensions **(D)** yield a more continuous topology, with sharply defined edges and rounder but denser apexes than are present in the PC space.

information about LE, see (Shi and Malik, 2000; Ng et al., 2002; Von Luxburg, 2007).

In contrast, the fundamental idea underlying t-SNE is to minimize the difference between two probability distributions: one constructed over the given high-D observations and another constructed over the desired low-D map. The metric used to evaluate differences between distributions is the Kullback-Leibler divergence. The t-SNE algorithm introduced by (van der Maaten and Hinton, 2008) is a variant of the more general stochastic neighbor embedding (*SNE*) algorithm of (Hinton and Roweis, 2002). For an excellent practical introduction to t-SNE, see (Wattenberg et al., 2016).

Elements of the t-SNE output are stochastic. Specifically, the absolute location of any given observational cluster in the low-D map produced by t-SNE is essentially random, and potentially important differences in boundaries among clusters can exist among different realizations of the algorithm. One commonly used approach to resolve this limitation is to run multiple realizations of the algorithm and choose the result with the minimum divergence (van der Maaten, 2021). Here, following (Small and Sousa, 2021), we take a different tack. Rather than choosing the single run with minimum divergence, we stack the output of multiple t-SNE realizations and compute the PCs of this stack, to which we refer as PC(t-SNE). This has the effect of capturing cluster consistency, as observations which routinely cluster together in t-SNE space plot together in PC(t-SNE) space. The eigenvalue distribution of the PC(t-SNE) result also provides information about the global dimensionality of the data manifold(s) resolved by t-SNE.

TOY EXAMPLE

Methods
We motivate the concept using a highly simplified synthetic dataset (**Figure 2**). Here, a 100 × 100 × 100 (*x, y, t*) spatiotemporal data cube is created using random linear

combinations of three signals: one sinusoid and two decaying exponentials with different amplitudes and decay constants.

The generating signals are given by the following functions (**Figure 2A**):

$$s_1 = \cos\left(\frac{\pi}{25}t\right)$$

$$s_2 = e^{\frac{-t}{5}}$$

$$s_3 = \begin{cases} 0 & for\ x = [1, 49] \\ 3e^{\frac{-(t-50)}{10}} & for\ x = [50, 100] \end{cases}$$

Within the spatiotemporal data cube of the toy example T, each synthetic "observation" $T_{x,y}(t)$ is computed as a linear combination of the above three signals, with weights chosen from a uniform random distribution and forced to sum to unity.

The forward model can be expressed as a weighted linear combination as:

$$T_{x,y}(t) = w_1 s_1 + w_2 s_2 + w_3 s_3, \text{ such that: } w_1 + w_2 + w_3 = 1$$

The inverse problem involves the identification of the temporal endmembers (s_i) and the estimation of their corresponding weights (w_i) for each observed $T_{x,y}(t)$.

Because the spatiotemporal cube has 100 time steps, each $T_{(x,y)}$ can be considered a vector residing in 100-D space. However, the simplicity of the generating functions suggests that the true dimensionality of the underlying signals is far lower than 100-D. Geometrically, the data lie on a low-D manifold within the full high-D space. Topologically, strictly imposed linear mixing dictates that this triangular manifold is fully interconnected with sharp corners and straight edges. Such properties of the low-D manifold imply fundamental constraints on the design of a well-posed inverse problem: geometric and topologic structure imply algebraic structure. In this case, the inverse problem amounts to identification of the temporal endmembers representing the generative processes and estimation of the

relative contribution of each temporal endmember to a general observation potentially containing contributions from one, two or three of the signals.

The overall joint characterization methodology can be summarized by the following steps:

1. PC transform $T_{xy}(t)$
2. Render TFS as orthogonal 2D projections of PCs 1–3
3. Identify tEMs
4. Characterize linearity of binary mixtures
5. Apply LE to $T_{xy}(t)$
6. Render TFS of LE 1-3
7. Identify LE tEMs and compare to PC tEMs
8. Iteratively apply t-SNE to $T_{xy}(t)$
9. PC transform compilation of multiple t-SNE realizations
10. Identify clusters from PC(t-SNE) 1-3
11. Back propagate clusters to spatial domain

Results

Through the lens of PC/EOF analysis, the data are decomposed onto a set of mutually orthogonal eigenvectors which are sorted on the basis of variance. The power of this approach is illustrated in Panel B of **Figure 2**. Here, two dimensions represent *T*, together comprising >99% of overall variance in the data. The algebraic structure imposed by the linear mixing equations above maps cleanly onto trigonal planar geometric structure reminiscent of a ternary diagram. Pixels with (nearly) pure contributions from each one of the three temporal endmember (tEM) signals occupy one of the three corners of this triangle. Pixels occupying the sharp linear edges represent binary mixtures of two EMs; and more generally, pixels in the body of the triangle have a contribution from each of the three EMs linearly weighted by Euclidean distance to each corner. The variance partition and topology of the temporal feature space immediately characterize the 2D manifold and three temporal endmembers corresponding to the three input signals.

In contrast, the connectivity structure of LE (**Figure 2C**) accurately reflects the generative spatial mixing process. Both the connectivity structure and bounding endmembers from Panel B are clearly preserved in LE (**Figure 2D**). However, the linear edges and sharp corners are rounded, a characteristic of LE which can be conceptually considered to reflect the (roughly) analogous nature of the Laplacian to geometric curvature.

On the other hand, PC(t-SNE) gives a fundamentally different output (**Figure 2D**) that is more clustered than PCs/EOFs or LE. On the one hand, this clustering could be considered a strength: the relatively well-separated red, green, and blue clusters accurately identify the dominant input signal contributing to each observation. On the other hand, the same property could be considered a weakness, introducing granularity to the low-D map which is not necessarily representative of any underlying generative process in the spatiotemporal dataset. One question we seek to address here is the consistency with which t-SNE clusters resolve known differences among subsets of observations. While the separability of the t-SNE clusters shown in **Figure 2** generally corresponds to the largest endmember fraction in the mixture, we highlight the exception (in dark cyan).

GLOBAL SCALE EXAMPLE—TERRESTRIAL CLIMATE

Methods

We now illustrate a joint characterization of a commonly used set of spatiotemporal observations: gridded terrestrial climate data produced by the Climatic Research Unit (CRU) at the University of East Anglia (UEA) (Mitchell and Jones, 2005). Here, we apply these DR algorithms to 103 years (1900–2002) of monthly estimates of mean temperature and total precipitation (T + P). Each pixel time series represents 103×12 monthly mean temperatures and total monthly precipitation estimated for each $1° \times 1°$ grid cell on land, excluding Antarctica.

Results

The resulting maps (**Figure 3A**) of the low-D spaces illustrate complementarities among the DR approaches. Traditional PC/EOF DR (top) captures broad zonal and longitudinal climatic patterns, representing the vast majority of points as occupying a continuum spanning differing endmember climate zones. In contrast, t-SNE (lower right) represents the global terrestrial T + P TFS as an amalgamation of highly discretized subregions, conceptually resembling a choropleth map. Some features, like the longitudinal Eurasian dipole, are accentuated in t-SNE that are either relegated to higher dimensions or represented as a linear combination of two or more PC/EOF dimensions. LE (lower left) complements both of these two approaches, representing the global T + P space as fundamentally comprised of continuous nonlinear gradients which are notably sharper than present in the PC/EOF space, but lack the discrete clustering of the t-SNE space. Discrete physiographic features like the Rocky, Himalayan, and Andean mountain ranges emerge clearly using PCs/EOFs and LE, but not in t-SNE. Comparison of PC/EOF, LE and t-SNE also reveals that even the low order dimensions of PCs/EOFs resolve finer scale, but physically meaningful, distinctions such as the orographic high precipitation regions of the Western Ghat in India and the Chittagong Hill Tracts in Bangladesh. These features are resolved as distinct clusters in individual t-SNE realizations, but not in the low order dimensions of PC(t-SNE). They appear within the continuum of the LE TFS.

Temporal feature spaces derived from each approach (**Figure 3B**) illustrate key differences in geometry and topology. PCs/EOFs (top) represents the majority of grid cells as occupying a position within a continuum bounded by endmember climates. The PC1 vs PC2 feature space effectively discriminates between hot and dry, hot and wet, and cold climatic endmembers, bearing a strong resemblance to the mean temperature vs precipitation space on which terrestrial biomes are defined, e.g., in (Small and Sousa, 2016). Density shading reveals prevalence of linear and nonlinear mixing relationships, as well as sparse data representation of high variance patterns with low PC 2 and low PC 3 values. Points with maximum PC 1 values differentiate along the PC 3 axis between Polar Ice Cap (Köppen Climate Type EF, Greenland) and Polar—Tundra (Köppen Climate Type EF, Siberia and northern Canada).

In contrast, the clustering evident in the t-SNE map is clearly represented in the corresponding low-order TFS (bottom row). Here, the spatiotemporal data manifold is represented as a
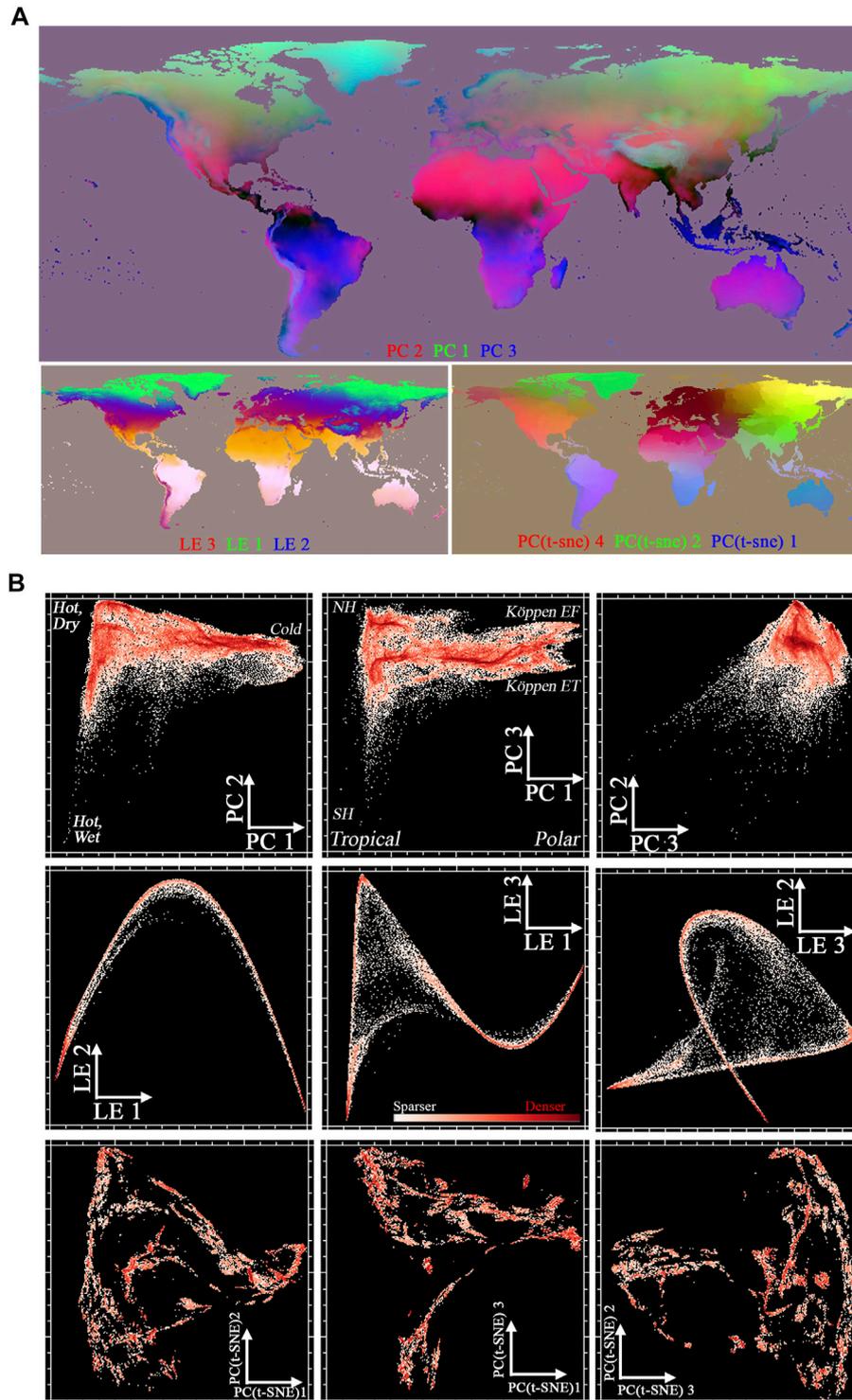
FIGURE 3 | (A) Joint characterization yields complementary spatiotemporal patterns when applied to a century of gridded global climate data. Characterization of monthly mean air temperature and total precipitation estimates for 100 years (1900-2000) using linear and nonlinear dimensionality reduction reveals both continuous gradation and discrete clustering. Linear variance-based decomposition (PCA, top) captures broad bioclimatic patterns. Non-linear approaches produce complementary maps: Laplacian Eigenmaps (bottom left) accentuates gradients, with clear separation among tropical, arid, temperate, and boreal climates in the three lowest dimensions. In contrast, a compilation of 100 t-SNE runs captures both connectivity relationships and clearly identifiable geospatial clusters. **(B)** Joint characterization yields complementary temporal feature spaces when applied to a century of global climate reanalysis data. The continuous gradation and discrete clustering observed in geographic space in Fgure Xa maps onto temporal feature space structure. Liner variance-based decomposition (PCA, top row) captures

(*Continued*)

FIGURE 3 | dominant patterns of overall variance across the entire domain, yielding a diffuse point cloud with substructure indicative of regional-scale gradients. Non-linear approaches produce complementary results: Laplacian Eigenmaps (center row) captures a continuous helicoid manifold that clearly expresses endmember climate and mixing relations, PC(t-SNE) captures both connectivity relationships and clearly identifiable clusters. Eacg temporal feature space emphasizes different aspects of the geometric and topological structure of the spatiotemporal data manifolf.

discretized nonlinear continuum, with point density spread in a relatively uniform manner among clusters. Mixing relationships are similar to those found with PCs/EOFs, but effectively collapsed onto a narrow manifold rather than diffusely spread throughout the low-order TFS. It is noteworthy that the small isolated clusters in the t-SNE space correspond to geographically isolated locations (e.g., Madagascar, Australia, Greenland) as well as topographically isolated regions (e.g., Altiplano, Anatolia, Tibetan Plateau), providing a clear physical basis for distinction from the more continuous climatic gradients. The LE TFS (center row) possesses a fundamentally distinct geometric structure. Here, the global terrestrial T + P space is represented as a tight manifold with a nonlinear continuum spanning a low dimensional surface bounded by three unambiguous endmembers corresponding to those bounding the PC1/PC2 feature space. For this popular spatiotemporal dataset, each of the three DR approaches clearly provides information that the other two do not, illustrating the potential complementarity benefits of joint characterization of the global climate space.

REGIONAL SCALE EXAMPLE—SAHEL VEGETATION PHENOLOGY

Methods

Next, we investigate the ability of joint characterization to provide useful information across spatiotemporal scales and generative processes. Specifically, we explore rainfall-driven vegetation phenology across the African Sahel. This spatiotemporal cube is a time series of MODIS Enhanced Vegetation Index (EVI) vegetation abundance maps in which each pixel time series represents the aggregate vegetation phenology within a 250 × 250 m footprint. Relative to the global climate illustration above, these data are ~2x more frequent (16-day composite) and ~400x spatially finer (250 m).

Results

The TFS resulting from each DR approach is shown in **Figure 4A**. PCs/EOFs identify clear temporal endmembers corresponding to barren (EM1) and evergreen (EM2) phenologies, connected by a diffuse helical manifold of seasonal grass/shrub vegetation driven by latitudinal migration of the InterTropical Convergence Zone (ITCZ) and associated rainfall. EOF 1 controls the amplitude of the seasonal vegetation abundance. EOFs 2 and 3 reveal complementary out-of-phase sinusoids which together combine to form a phase plane. An additional double monsoon signal is seen in EM3, distinct from the single annual signal.

Nonlinear dimensionality reduction methods identify a much more compact low-D manifold, emphasizing some parts of the structure of the linear space in each case. LE collapses the

EM1—EM3—EM2 continuum into a single, well-defined curvilinear helicoid, with double monsoon EM3 diverging from the primary manifold near its midpoint. PC(t-SNE) identifies a similar overall manifold structure, but with much more defined clustering. Unvegetated pixels form a large, diffuse cloud in PC(t-SNE) space. PC(t-SNE) also separates coherent clusters which are not clearly defined using either PC/EOF or LE methods alone. The stochastic nature of individual t-SNE realizations is illustrated through the example realizations shown in the bottom right quadrant of the figure.

The geometry and topology from **Figure 4A** can then be leveraged to design a temporal mixture model of the vegetation phenology. Here, we illustrate a simple linear model constructed using four temporal endmember phenologies identified from the TFS characterization (**Figure 4B** inset). The resulting phenology map, shown in **Figure 4B**, is effective at capturing a continuum of distinct phenological zones, as well as variability within and among those zones.

FIELD SCALE EXAMPLE—CALIFORNIA AGRICULTURE

Methods

We next extend the investigation of scale dependence to a study area relevant to practical land management: field-level agricultural mapping in California (**Figure 5**). This extends our investigation of scale dependence by covering a ~20x shorter record (1 year) with ~3x more frequent revisit (3–5 days) and 1,000x finer spatial scale (9 km^2) than was present for the MODIS example above. The study area here is an orchard-dominated 3 km × 3 km subset of Kern County, the highest-value crop producing county within California's agriculturally diverse Central Valley (Kern County, 2019). Here, we use subpixel vegetation fraction estimated from 30 m Harmonized Landsat-Sentinel (HLS) multispectral imagery by inversion of a linear spectral mixture model using generalized global spectral endmembers from (Small, 2018).

The area shown here is dominated by two high-value orchard crops: almonds and pistachios. A publicly available county-level crop map (Kern County, 2019) provides ground truth of crop type (far left, top; superimposed on false color reflectance image). Average vegetation fraction time series of each crop (far left, bottom) show the almond crop is characterized by earlier, higher amplitude green-up and more rapid senescence than the pistachio crop. Two pistachio fields (white in reflectance image) behave differently, with lower overall vegetation fraction presumably due to age of orchard.

Results

PCs/EOFs (center left) generally capture these differences, with almonds and pistachios forming diffuse clusters largely separable

FIGURE 4 | (A) Complementary temporal feature spaces for Sahel MODIS EVI vegetation phenology. Principal components yield a diffuse continuous helical feature space in which PC1 controls amplitude while PC2 and PC3 form a phase plane to represent the seasonal shift in greening and senescence following the latitudinal shift of precipitation. Temporal endmembers (inset) from apexes of the PC cloud form a convex hull bounding the full range of phenologies. Temporal EOFs show the latitudinal phase shift and isolated double monsoon seen in EM3 AND EM4. Laplacian Eigenmaps yield a much less diffuse, more continuous feature space with similar structure and EMs. Individual t-SNE spaces vary in shape and orientation but with similar topology, also showing an amplitude continuum with distinct clusters for non-vegetated areas and closed shrublands fed by the Somali double monsoon (EM4). PC(t-SNE) of 10 2D t-SNE realizations produces a more compressed helical

(Continued)

FIGURE 4 | (Continued).

FIGURE 5 | Dimensionality reduction techinques capture complementary aspects of crop phenology. Other than roads and structure, a 3 km × 3 km subset of Kern County, CA is entirely comprised of almond and pistachio orchards (upper left; image date 23 March 2020). Mean time series of all pixels from each crop (lower left) reveals similar phenology, but with earlier and greener leaf-on maximum, and more rapid senescence, for almonds (A) than for pistachios (P). Each technique resolves spatially coherent (top) patterns which resolve variance both within and among classes. The locus of convergence on PC(t-SNE) space corresponds to the top apex in LE space, which features s lower amplitude phenology indicative of a younger orchard with smaller, less full canopies.

on the basis of PC/EOF 2. The two distinct pistachio fields plot as a smaller cluster closer to the center of the linear feature space. However, the diffuse topology of the data in linear feature space yields ambiguity in the optimal location of decision boundaries.

In effect, within-class variance is accentuated at the expense of between-class variance.

Nonlinear dimensionality reduction algorithms capture different aspects of the high-D data structure. Almond and

pistachio fields are sharply distinct as well-separated (Transformed Divergence = 2.0) apexes in LE space, with the two distinct pistachio fields forming a third well-separated apex at high LE 2. In contrast, PC(t-SNE) captures a continuously varying nonlinear data manifold. Almonds and pistachios form opposite ends of this manifold, with the two indistinct fields plotting near the lengthwise center of manifold.

DISCUSSION

Strengths and Limitations

Any low-dimensional representation of a high-dimensional space will generally result in retention of some subset, and loss of some other subset, of the total high-D information. The analysis presented here illustrates complementarity among three philosophically differing approaches to achieving such low-D representations, focusing on the special case of spatiotemporal characterization. Depending on the application, the subset of information retained by each approach may be more or less useful. In many cases, comparative analysis facilitated by joint characterization using multiple DR approaches may prove to be more useful than the sum of its parts.

Specifically, important differences among algorithms manifest as differential sensitivities to scales, geometries, and topologies of variance. Perhaps the broadest difference is between the global and local scales of operation between the linear and nonlinear DR approaches. The linear (PC/EOF) approach implemented here operates on the overall global variance structure of the spatiotemporal dataset, allowing each observation explicit contextual representation in terms of the global space, but sacrificing potentially important local structure for generality. In contrast, both of the nonlinear approaches (LE and t-SNE) focus on local statistical similarity at the expense of downweighing global structure. This difference in formulation of the DR problem maps directly onto geometry and topology of each low-order Temporal Feature Space. The linear, global TFS yields relatively diffuse point clouds due to its strict orthogonality requirement and usage of a metric of global dispersion (variance). In contrast, the low-order TFS for each nonlinear algorithm yields a visibly tighter manifold, with decreased ambiguity in temporal endmembers (LE) and/or in spatiotemporal clustering (t-SNE)—a direct result of optimizing metrics focused on local structure at the expense of global structure.

However, neither algorithmic strengths nor limitations occur in isolation. For instance, the diffuse nature of the PC/EOF TFS poses a challenge, but is compensated by the explainable connection between spatial and temporal dimensions provided by explicit EOF time series associated with each PC spatial pattern. This relative ease of interpretability is not a feature of LE or t-SNE. Similarly, the unambiguous clustering relations identified by t-SNE are in some sense a feature, for instance allowing clear determination of decision boundaries for discrete classification; but can also be a limitation by artificially introducing clustering that may not be present in the generative processes, as observed in the synthetic example of **Figure 2**. As the degree of clustering can be influenced by the

perplexity parameter chosen, in some cases a sensitivity analysis may be warranted to quantify the persistence of specific clusters. Similarly, LE is effective at removing ambiguity in endmember spatiotemporal patterns, but is limited in explainability. Further, because both t-SNE and LE operate on statistical similarity relations among observations, both are considerably more sensitive to spatial aliasing than PCs/EOFs, which treat each observation as independent of all others. Notably, this is not necessarily true for temporal aliasing, as the symmetry between spatial and temporal dimensions in broken in LE and t-SNE.

Broad Efficacy

The global, regional, and field scale examples shown above span roughly 5 orders of magnitude spatially and 2 orders of magnitude temporally, illustrating the ability of joint characterization to provide useful information at global, local, and regional scales. In particular, the approach is shown to be effective in both characterizing processes occurring at scales coarser than pixel resolution (field-level agriculture, broad climate gradients) and their associated subpixel spatial mixing processes (smallholder Sahel agriculture). Further, joint characterization is shown to yield useful results for disparate biogeophysical spatiotemporal processes including climate as well as natural and human-managed vegetation. This broad efficacy suggests considerable potential for relevance to other non-spatiotemporal but high-D data types, for instance to the field of imaging spectroscopy as recently suggested in (Small and Sousa, 2022).

Synthesis

The concept of the temporal feature space was initially introduced as the basis for characterization of high dimensional spatiotemporal data to inform the feasibility and design of temporal mixture models. By providing a model-agnostic projection of the temporal feature space, the combination of eigenvalue variance partition, temporal EOFs and spatial PCs, the Singular Value Decomposition of spatiotemporal data informs the 1) spatiotemporal dimensionality, 2) linearity (or lack thereof) and 3) temporal endmember selection (Small, 2012). While individual temporal EOFs can be informative, it is their linear combinations that represent the actual data. Individual EOFs almost never act alone. As such, the temporal endmembers identified from the TFS that provide a more physically meaningful set of basis functions on which to project the higher dimensional spatiotemporal data. If a TFS structure can be accurately represented (as quantified by model misfit to observation and positivity of EM fractions) with a relatively small number of temporal EMs, a temporal mixture model can provide a useful abstraction of high dimensional spatiotemporal data. In addition, the variance partition given by the eigenvalues of the covariability matrix can provide a basis for distinguishing between deterministic and stochastic components of the spatiotemporal data (Lorenz, 1956; Small, 2012). The only assumptions implicit in the use of Principal Components and EOFs are that global variance corresponds to information content and that correlation corresponds to redundancy. While these are often valid assumptions, they also bias the projection of the temporal feature space accordingly. The

potential value of supplementing the global variance structure given by the PCs/EOFs with nonlinear manifold learning tools lies in the preservation of low variance information content that may be lost in the higher order dimensions of the PCs/EOFs. In other words, the assumptions implicit in some nonlinear manifold learning approaches are complementary to the assumptions on which PC/EOF analysis is based. As shown in the examples presented here, LE and t-SNE are complementary in that the LE spaces are simpler, less diffuse and more continuous than the PC-based TFS, while the t-SNE (and PC(t-SNE)) spaces can be both continuous and clustered with verifiably meaningful clusters not resolved by the PC-based TFS (Small and Sousa, 2021; Sousa and Sousa, 2021).

Practical Considerations

Even if an approach is highly effective, practical limitations can severely limit its widespread adoption. Fortunately, code facilitating each of the DR algorithms used in this study is freely available from numerous open source web repositories, most notably Python's scikit-learn. But despite code availability, computational limitations remain a key chokepoint for easy adoption, especially for the nonlinear algorithms. Both mathematical elegance and decades of computer science have facilitated algorithms enabling efficient PC/EOF analysis of datasets with millions of spatial observations, each with hundreds of time steps, on typical personal computing hardware. The same is unfortunately not (yet) true of t-SNE and LE. Because of the nature of such manifold-based approaches, computational resources scale roughly with the square of the number of spatial—but not temporal—samples. The key limitation is generally sufficient RAM to store large matrices of pairwise (or n tuple-wise) metrics. For reference, a Lenovo laptop running x64 Linux with 32 GB of RAM and scikit-learn version 0.24.2 was able to compute both LE on datasets with a maximum of roughly 25,000 spatial samples, and t-SNE on datasets with a maximum of roughly 250,000 spatial samples. Temporal depth was not found to be a limitation for either algorithm for the datasets examined here.

Notably, other linear and non-linear DR approaches are also worthy of consideration: e.g., ISOMAP, Locally Linear Embedding, Hessian Eigenmapping, Local Tangent Space Alignment, metric and non-metric multidimensional scaling, UMAP, and more. Because the purpose of this work is to introduce the idea of the joint characterization for spatiotemporal data manifolds, rather than present a thorough intercomparison of all available methods, we defer detailed examination of these other approaches to future work. However, we do note that in some sense PCs/EOFs, Laplacian Eigenmaps, and t-SNE can be argued to represent endmember DR philosophies and thus span a substantial subset of the current DR landscape.

For some applications, algorithm automation is a primary concern. We note here that in a fundamental sense, the joint characterization approach is inherently non-automated. This is because joint characterization requires the scientist to utilize visual perception and critical thinking in exploratory data analysis and interpretation of geometric and topological TFS structure. That being said, the information provided by joint characterization can then be used to design parsimonious inverse models which could potentially be fully automated.

CONCLUSION

This analysis explores the utility of joint characterization of spatiotemporal data using synthetic, reanalysis, and satellite datasets. Joint characterization is implemented using a set of three linear and nonlinear approaches to dimensionality reduction: Empirical Orthogonal Functions, Laplacian Eigenmaps, and t-distributed Stochastic Neighbor Embedding. The three approaches are shown to yield complementary characterizations for a variety of generative processes at global, regional, and field scales. The results of this analysis suggest that joint characterization can provide a useful framework to visualize the geometric and topologic structure of high dimensional spatiotemporal data manifolds to assist with the design of effective and parsimonious inverse models, as well as improved class separability for discrete thematic classifications.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

Study conception: DS and CS. Computation: DS. Analysis and Interpretation: DS and CS. Draft manuscript preparation: DS. Manuscript revision: DS and CS. All authors reviewed the results and approved the final version of the manuscript.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Bachmann, C. M., Ainsworth, T. L., and Fusina, R. A. (2005). Exploiting Manifold Geometry in Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sensing* 43, 441–454. doi:10.1109/tgrs.2004.842292

Bellman, R. (1957). *Dynamic Programming*. Courier Dover Publications.

Christakos, G. (2017). *Spatiotemporal Random fields: Theory and Applications*. 2nd Edn. Elsevier.

Eshel, G. (2011). *Spatiotemporal Data Analysis*. Princeton University Press. doi:10.1515/9781400840632

Gillis, D., Bowles, J., Lamela, G. M., Rhea, W. J., Bachmann, C. M., Montes, M., et al. (2005). "Manifold Learning Techniques for the Analysis of Hyperspectral Ocean Data," in *Presented at the Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XI, International Society for Optics and Photonics*.

Hadid, A., and Pietikäinen, M. (2009). "Manifold Learning for Video-To-Video Face Recognition," in *Biometric ID Management and Multimodal Communication. Lecture Notes in Computer Science*. Editors J. Fierrez, J. Ortega-Garcia, A. Esposito, A. Drygajlo, and M. Faundez-Zanuy (Berlin: Springer), Vol. 5707.

Henebry, G. M. (1995). Spatial Model Error Analysis Using Autocorrelation Indices. *Ecol. Model.* 82, 75–91. doi:10.1016/0304-3800(94)00074-R

Hinton, G., and Roweis, S. T. (2002). "Stochastic Neighbor Embedding Presented at the NIPS Citeseer," in *Advances in Neural Information Processing Systems*. (MIT Press), Vol. 15.

Jianbo Shi, J., and Malik, J. (2000). Normalized Cuts and Image Segmentation. *IEEE Trans. Pattern Anal. Machine Intell.* 22, 888–905. doi:10.1109/34.868688

Kadoury, S. (2018). Manifold Learning in Medical Imaging," in *Manifolds II-Theory and Applications*. IntechOpen. doi:10.5772/intechopen.79989

Kern County (2019). *2018 Kern County Agricultural Crop Report*. Bakersfield, CA: Department of Agriculture and Measurement Standards. Available at: http://www.kernag.com/caap/crop-reports/crop10_19/crop2018.pdf.

Lorenz, E. (1956). *Empirical Orthogonal Functions and Statistical Weather Prediction (No. 1), Statistical Forecasting Project. Massachusetts Institute of Technology*. Cambridge: MA.

Lotsch, A., Friedl, M. A., Anderson, B. T., and Tucker, C. J. (2003). Coupled Vegetation-Precipitation Variability Observed from Satellite and Climate Records. *Geophys. Res. Lett.* 30, 506. doi:10.1029/2003GL017506

Mitchell, T. D., and Jones, P. D. (2005). An Improved Method of Constructing a Database of Monthly Climate Observations and Associated High-Resolution Grids. *Int. J. Climatol.* 25, 693–712. doi:10.1002/joc.1181

Ng, A. Y., Jordan, M. I., and Weiss, Y. (2002). *On Spectral Clustering: Analysis and an Algorithm Presented at the Advances in neural information processing systems,* 849

Pearson, K. (1901). LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *Lond. Edinb. Dublin Phil. Mag. J. Sci.* 2, 559–572. doi:10.1080/14786440109462720

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine Learning in Python. *J. machine Learn. Res.* 12, 2825–2830.

Safaie, A., Dang, C., Qiu, H., Radha, H., and Phanikumar, M. S. (2017). Manifold Methods for Assimilating Geophysical and Meteorological Data in Earth System Models and Their Components. *J. Hydrol.* 544, 383–396. doi:10.1016/j.jhydrol.2016.11.009

Small, C. (2018). "Multisource Imaging of Urban Growth and Infrastructure Using Landsat, Sentinel and SRTM," in *NASA Landsat-Sentinel Science Team Meeting* Rockville.MD

Small, C., and Elvidge, C. D. (2013). Night on Earth: Mapping Decadal Changes of Anthropogenic Night Light in Asia. *Int. J. Appl. Earth Observation Geoinformation* 22 (1), 40–52. doi:10.1016/j.jag.2012.02.009

Small, C., and Sousa, D. (2016). Humans on Earth: Global Extents of Anthropogenic Land Cover from Remote Sensing. *Anthropocene* 14, 1–33. doi:10.1016/j.ancene.2016.04.003

Small, C., and Sousa, D. (2022). Joint Characterization of the Cryospheric Spectral Feature Space. *Front. Remote Sens.* 2, 793228. doi:10.3389/frsen.2021.793228

Small, C., and Sousa, D. (2019). Spatiotemporal Characterization of Mangrove Phenology and Disturbance Response: The Bangladesh Sundarban. *Remote Sensing* 11, 2063. doi:10.3390/rs11172063

Small, C., and Sousa, D. (2021). Spatiotemporal Evolution of COVID-19 Infection and Detection within Night Light Networks: Comparative Analysis of USA and China. *Appl. Netw. Sci.* 6, 10. doi:10.1007/s41109-020-00345-4

Small, C., Sousa, D., Yetman, G., Elvidge, C., and MacManus, K. (2018). Decades of Urban Growth and Development on the Asian Megadeltas. *Glob. Planet. Change* 165, 62–89. doi:10.1016/j.gloplacha.2018.03.005

Small, C. (2012). Spatiotemporal Dimensionality and Time-Space Characterization of Multitemporal Imagery. *Remote Sensing Environ.* 124, 793–809. doi:10.1016/j.rse.2012.05.031

Sousa, D., and Davis, F. W. (2020). Scalable Mapping and Monitoring of Mediterranean-Climate Oak Landscapes with Temporal Mixture Models. *Remote Sensing Environ.* 247, 111937. doi:10.1016/j.rse.2020.111937

Sousa, D., and Small, C. (2021). Joint Characterization of Multiscale Information in High Dimensional Data. *Adv. Artif. Intell. Mach. Learn.* 1 (3), 13. doi:10.54364/aaiml.2021.1113

Sousa, D., and Small, C. (2019). Mapping and Monitoring Rice Agriculture with Multisensor Temporal Mixture Models. *Remote Sensing* 11, 181. doi:10.3390/rs11020181

Sousa, D., Small, C., Spalton, A., and Kwarteng, A. (2019). Coupled Spatiotemporal Characterization of Monsoon Cloud Cover and Vegetation Phenology. *Remote Sensing* 11, 1203. doi:10.3390/rs11101203

van der Maaten, L. (2021). t-SNE: FAQ. t-SNE. https://lvdmaaten.github.io/tsne/.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data Using T-SNE. *J. machine Learn. Res.* 9, 2579

Van Der Maaten, L., Postma, E., and Van den Herik, J. (2009). *Dimensionality Reduction: a Comparative Review*.

Verbesselt, J., Hyndman, R., Newnham, G., and Culvenor, D. (2010). Detecting Trend and Seasonal Changes in Satellite Image Time Series. *Remote Sensing Environ.* 114, 106–115. doi:10.1016/j.rse.2009.08.014

Von Luxburg, U. (2007). A Tutorial on Spectral Clustering. *Stat. Comput.* 17, 395–416. doi:10.1007/s11222-007-9033-z

Watson, J. R., Gelbaum, Z., Titus, M., Zoch, G., and Wrathall, D. (2020). Identifying Multiscale Spatio-Temporal Patterns in Human Mobility Using Manifold Learning. *PeerJ Comp. Sci.* 6, e276. doi:10.7717/peerj-cs.276

Wattenberg, M., Viégas, F., and Johnson, I. (2016). *How to Use T-SNE Effectively*. Distill. doi:10.23915/distill.00002

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., et al. (2008). Free Access to Landsat Imagery. *Science* 320, 1011. doi:10.1126/science.320.5879.1011a

Yan, L., and Roy, D. P. (2015). Improved Time Series Land Cover Classification by Missing-Observation-Adaptive Nonlinear Dimensionality Reduction. *Remote Sensing Environ.* 158, 478–491. doi:10.1016/j.rse.2014.11.024

Zhai, Y., Qu, Z., and Hao, L. (2018). Land Cover Classification Using Integrated Spectral, Temporal, and Spatial Features Derived from Remotely Sensed Images. *Remote Sens.*, 10 (3), 383. doi:10.3390/rs10030383