

The Climatic Temporal Feature Space: Continuous and Discrete

Christopher Small

*Lamont Doherty Earth Observatory,
Columbia University, Palisades, NY 10964 USA.*

CSMALL@COLUMBIA.EDU

Daniel Sousa

*Department of Geography,
San Diego State University, San Diego CA 92182 USA.*

DAN.SOUSA@SDSU.EDU

Corresponding Author: Christopher Small.

Copyright © 2021 Christopher Small and Daniel Sousa This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Climatic zones, representing seasonal variations in temperature (T) and precipitation (P), are generally mapped geographically using discrete classifications with distinct boundaries. However, it is well known that global T and P vary continuously in space and time with steep gradients occurring infrequently. The objective of this analysis is to use complementary forms of dimensionality reduction to quantify the spatiotemporal dimensionality of the climate system and to produce a continuous representation of global climate based on the temporal feature space of historical T and P alone. We characterize the continuous global feature space using principal components (PCs) to identify a parsimonious set of temporal endmember T and P patterns bounding the feature space of all observed T and P patterns. These endmember T and P patterns provide the basis for a linear temporal mixture model that can represent decadal T and P patterns of any geographic location as fractions of the endmember T and P patterns. Inverting this linear mixture model for each geographic $T+P$ time series gives an estimate of the fractional contribution of each endmember to the observed time series. The resulting temporal endmember fraction maps provide a continuous representation of the Euclidean proximity of T and P observations at every geographic location to each of the temporal endmember climates bounding the space. The spatiotemporal dimensionality implied by the variance partition of $T+P$ time series for 67,420 land-based observations suggests that the $T+P$ temporal feature space is effectively 3D, accounting for 92% of total variance. From the topology of the feature space, we identify 4 bounding temporal endmembers upon which to base the linear temporal mixture model. Inversion of the model for each normalized observed time series yields endmember fraction estimates and a model misfit distribution with 99% of misfit < 0.21 . For comparison, we also render temporal feature spaces from ensembles of 2D manifolds within the $T+P$ space derived from suites of t-distributed Stochastic Neighbor Embeddings (t-SNE) to identify discontinuities in the feature space. Comparison of spatial PC(t-SNE) across hyperparameter settings reveals consistent structure and little hyperparameter sensitivity to temporal feature spaces rendered by t-SNE. Combining the physically interpretable continuous global structure resolved by the PC feature space with the finer scale manifold structure resolved by the t-SNE feature space provides a continuous alternative to discrete classifications of climate that cannot represent the continuous character of its temporal feature space.

Keywords: Climate, Temporal Feature Space, Dimensionality reduction, PCA, t-SNE.

1. INTRODUCTION

Climatic zones, based on seasonal variations in temperature (T) and precipitation (P) are generally mapped geographically using discrete classifications with distinct boundaries separating explicitly defined zones. The most widely used, the Köppen-Geiger classification, has 30 discrete classes spanning 5 broad categories (Tropical, Dry, Temperate, Continental, Polar/Alpine) [1-3]. Class membership is determined by a decision tree in which branch thresholds are based on seasonal T and P ranges and timing. The designations and thresholds devised by Köppen, later modified by Geiger, were partially based on vegetation assemblages found in each zone – not strictly on the spatiotemporal characteristics of T and P alone. However, it is well known that T and P cycles vary continuously in space with steep gradients occurring infrequently along some coastlines and mountain range fronts. A fundamental conceptual asymmetry exists between the capability of discrete and continuous representations of the climate system: a continuous field representation could accommodate both steep and shallow gradients in T and P , but a discrete classification cannot accommodate shallow gradients. By necessity, discrete climate classifications superimpose artificial boundaries on an inherently continuous geophysical field.

The objective of this analysis is to quantify the spatiotemporal dimensionality of the global climate system and to produce a more realistic continuous representation of global climate based on the topology and dimensionality of the temporal feature space of T and P alone. We first characterize the continuous global feature space using principal components (PCs) and their corresponding Empirical Orthogonal Functions (EOFs) to identify a parsimonious set of temporal endmember $T + P$ (concatenated) patterns bounding the feature space of all observed $T + P$ patterns. These bounding temporal endmembers are used to define a temporal mixture model in which all T and P patterns in the space can be represented as linear combinations of the temporal endmembers. The mixture model is inverted to provide fraction estimates for each temporal endmember for each spatially explicit time series of $T + P$ in the set of observations comprising the feature space. The resulting temporal endmember fraction maps provide a continuous representation of the Euclidean proximity (similarity) of T and P observations at every geographic location to each of the temporal endmember climates bounding the space. For comparison, we also implement the joint characterization introduced in [4], and extended by [5], to render temporal feature spaces from ensembles of 2D manifolds within the $T + P$ space derived from suites of t-distributed Stochastic Neighbor Embeddings (t-SNE) [6], to identify any discontinuities in the feature space that may not be captured in the global structure represented by the PC-derived feature space.

The PC-derived temporal feature space renders a unique, invertible representation of uncorrelated patterns of global variance and provides a physically interpretable topology from which temporal endmember patterns of T and P cycles can be identified, as well as a statistical estimate of the spatiotemporal dimensionality of the space. In contrast, the t-SNE-derived temporal feature space attempts to preserve the local structure of manifolds embedded within the higher dimensional temporal feature space – at the expense of physically interpretable topology. Any such, local manifolds resolved as distinct clusters might be interpreted as statistically distinct climates occurring within the otherwise continuous global temporal feature space defined by the T and P time series. Together, the low dimensional parameterization of Earth's climatic space provided by the continuous temporal endmember fractions and the climatic discontinuities provided by the local embeddings

offer two complementary representations of the climate system that are both derived from, and consistent with, T and P observations alone. Because both representations are based only on the spatiotemporal characteristics of decades of T and P observations, they are free of any bias imposed by the incorporation of ancillary factors (e.g. vegetation communities) used to produce discrete climate classifications.

In brief, the temporal feature space represented by the 3 low order PCs comprising 92% of total variance of the $T + P$ product yields four well-defined temporal endmembers which bound the space. Inversion of a linear mixture model based on these four temporal endmembers for each geographically distinct $T + P$ time series yields a fraction map for each. These fraction maps represent the climatic similarity of each geographic location to each endmember $T + P$ climate. In contrast, the temporal feature spaces rendered by multiple t-SNE realizations spanning a range of hyperparameter settings show consistent global structure with distinct macroclimatic clustering in addition to a latitudinal continuum preserving global structure. Taken together, the global structure preserved by the PC temporal feature space provides a physically interpretable representation of Earth's climate system spanned by four endmember climatic extremes while the t-SNE temporal feature spaces preserve local manifold structure embedded within the global space corresponding to geographically distinct macroclimates – in addition to large scale gradients not depicted within existing discrete climate classification schemes.

Data

The T and P time series used in this analysis are decadal averages of monthly mean temperature and precipitation observations provided by the University of East Anglia (UEA) Climatic Research Unit (CRU) global gridded product (v.4.05) available from <https://crudata.uea.ac.uk/cru/data/hrg/>. The angular distance weighted interpolation used to produce the 0.5° gridded T and P products, as well as the spatial and temporal coverage of meteorological station data on which the products are based is described in detail by Harris et al. 2020, [7].

To construct the temporal feature space of the combined $T + P$ time series, we average 6 x 10 year intervals of monthly T and P observations between 1961 and 2020. We specifically use decadal intervals to allow for multi-year climatic phenomena such as El Niño-Southern Oscillation (ENSO) and North Atlantic Oscillation (NAO) that are known to influence both T and P cycles globally. In addition, decadal intervals may allow for identification of climatic manifestations of the 11 year Schwabe solar cycle. Each 6 decade average of monthly T and P is normalized separately to provide equivalent mean and standard deviation with only the uppermost tail of the P distribution exceeding the $[-1, 1]$ interval bounding the normalized T distribution. To achieve this normalization, monthly mean T values are divided by 60 and monthly mean P values are divided by 400. Normalized T and P time series are concatenated to construct the $T + P$ product, although the ordering of the temporal observations does not affect either PCs or t-SNE. Mean T and P maps, and the bivariate global distribution are shown in FIGURE 1.

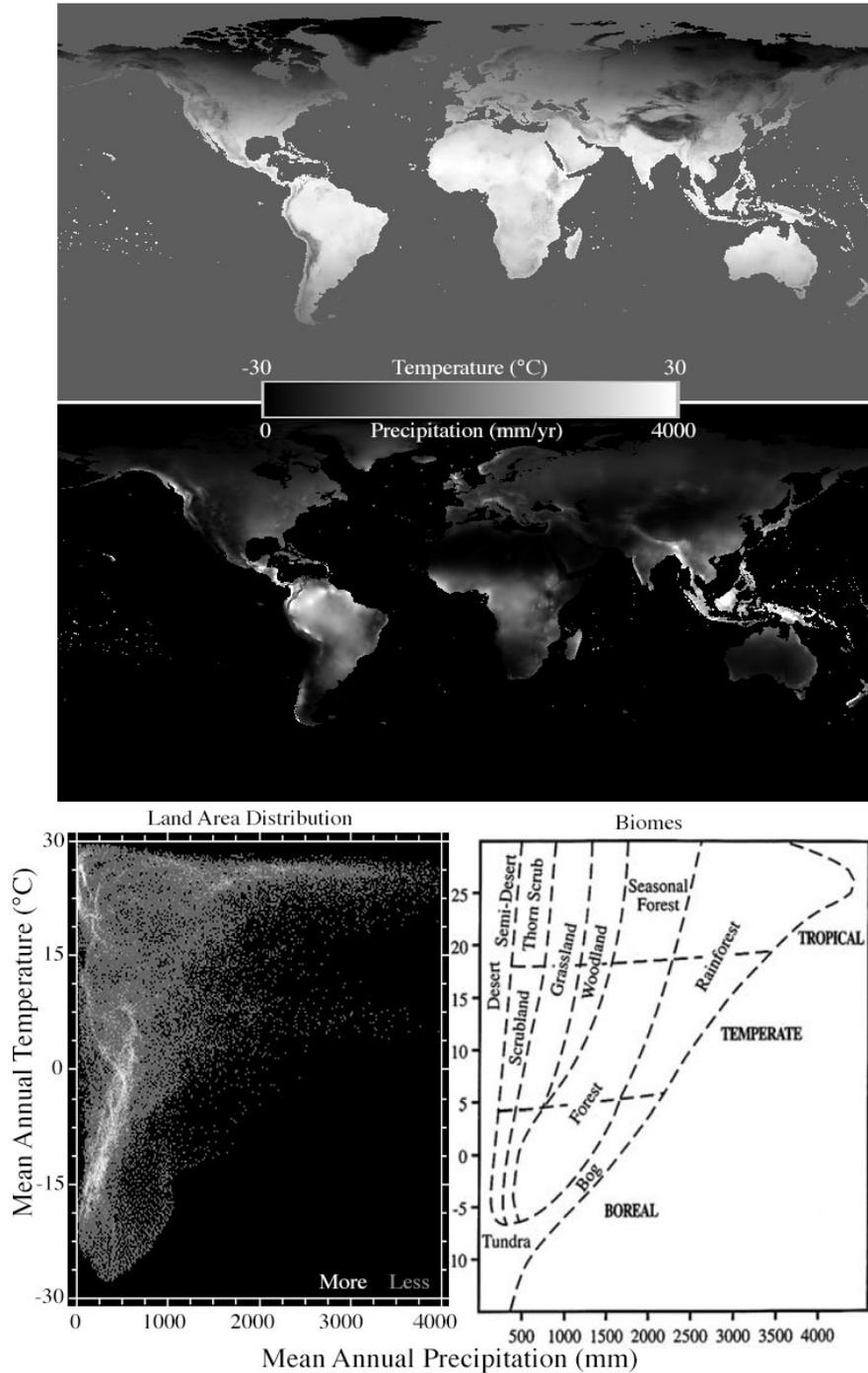


Figure 1: Mean annual temperature and precipitation maps (top) from CRU TS gridded climate products with climatic parameter space (lower left) based on 0.5° resolution monthly mean temperature and precipitation (1900-2020) from Harris et al. (2020). Biome classification (lower right) adapted from Houghton et al. 1996, [8].

2. METHODS

Linear characterization of spatiotemporal dimensionality and temporal feature space topology as a basis for construction of linear temporal mixture models is described in detail by [9]. Briefly, the approach uses classical PC/EOF analysis [10], to quantify variance partition and infer deterministic dimensionality on the basis of periodic EOFs. The persistent challenge of ascribing physical meaning to individual statistically derived EOFs is circumvented by a second change of basis in the form of a linear temporal mixture model. Linearity of mixing and identification of the temporal endmembers defining the model is established by the topology of the temporal feature space of low order PCs. Once linearity is confirmed and temporal endmembers are identified, the linear mixture model is inverted to provide estimates of fractions of each temporal endmember for each time series in the observations. Validity of the temporal mixture model is established by positivity and range, $[0,1]$, of fraction estimate distributions, unit fraction sum ($=1$), and the distribution of misfit between observed and reconstructed time series obtained from the linear combination of fraction estimates and the temporal endmembers.

With the brief description given above, the relationship between the PC/EOF characterization and the temporal mixture model can be summarized more explicitly. As explained in more detail by Small (2012), rotating the coordinate system to align with orthogonal dimensions of uncorrelated variance allows any location-specific pixel time series P_{xt} contained in an N image time series to be represented as a linear combination of temporal patterns (basis functions), F , and their location-specific components (weights), C , as:

$$P_{xt} = \sum_{i=1}^N C_{ix} F_{it} \quad (1)$$

where C_{ix} is the spatial Principal Component and F_{it} is its corresponding temporal Empirical Orthogonal Function, and i is the dimension. EOFs are the eigenvectors of the covariability (covariance or correlation) matrix, representing uncorrelated temporal patterns of variability, at each time t , within the data. PCs are the corresponding spatial weights that represent the relative contribution of each temporal EOF to the corresponding time series, P_{xt} , at each location x . The relative contribution of each EOF to the total spatiotemporal variance is given by the eigenvalues of the covariability matrix. N is the number of discrete temporal dimensions represented by the data; which may be greater, or less, than the true physical dimensionality of the process giving rise to the observations.

In systems in which deterministic processes dominate global variance, but stochastic processes are uncorrelated, the variance of the spatiotemporal structure of the deterministic processes may be represented in the low order PC/EOF dimensions while the stochastic variance may be represented by the higher order dimensions [11]. The distinction between low order deterministic and higher order stochastic dimensions may be manifest in the variance partition given by the eigenvalues – as well as physically meaningful features (e.g. periodicities) in the temporal EOFs. When a clear distinction can be made between interpretable deterministic patterns and uninterpretable, presumed stochastic, patterns, the summation of terms given above can provide a statistical basis for separation of deterministic and stochastic components of a spatial distribution of time series.

With noisy observations, an implicit assumption is that some number, $D (\leq N)$, of low order dimensions correspond to deterministic processes and that higher order dimensions represent stochastic variance. This allows an observed collection of time series to be represented as a sum of determin-

istic and stochastic components:

$$P_{xt} = \sum_{i=1}^D C_{ix} F_{it} + \varepsilon \tag{2}$$

where the weighted sum of low order dimensions, $C_{ix} F_{it}$, represents the deterministic component and the residual, ε , represents the stochastic variance.

Like the PC transform, a temporal mixture model represents a set of time series as a linear combination of temporal patterns (endmembers) and residual variance. Each location-specific (x) P_{xt} in an N image time series can be represented as a linear combination of D' ($\ll N$) temporal endmembers, E_{it} , and a residual component, ε' , as:

$$P_{xt} = \sum_{i=1}^{D'} f_{ix} E_{it} + \varepsilon' \tag{3}$$

where the pixel-specific fractions f_{ix} may represent either the areal fraction of the mixed observation exhibiting the temporal pattern of the corresponding endmember or, more generally, the Euclidian proximity of that pixel to the corresponding endmember in the temporal feature space. The result is a set of fraction (more generally proximity) maps representing the spatial distribution of different endmember abundances (or similarities). Dynamically, the temporal endmembers, E_{it} , can represent distinct processes and ε' represents the unmodeled residual component of the observations. The model is generally parsimonious if the modeled dimensionality, D' , is assumed less than or equal to the true dimensionality. Given an observed pixel time series P_{xt} with sufficient temporal sampling, and knowledge of a relatively small number of endmembers, E_{it} , the linear mixture model, $P = fE + \varepsilon'$, can be inverted to minimize ε' , yielding fraction estimates of each endmember for each pixel [12,13]. The linear mixture model can be represented more explicitly as $P_{xt} = f_{ix} E_{it} + \varepsilon'$. The least squares solution for the fraction estimates f_{ix} is obtained as: $(E^T E)^{-1} E^T$.

Operationally, the process of spatiotemporal characterization quantifies the spatiotemporal dimensionality of the feature space, confirms (or refutes) the linearity of mixing within the space, and reveals the identity of the temporal endmembers. The combination of temporal endmembers defines the linear mixture model to be inverted for each time series of observations. In addition to the system of mixing equations given by equation 3, a unit sum constraint can be included to urge the resulting fractions to sum to unity. In the case of the CRUTS product used in this analysis, $12 \times 10 = 120$ monthly T observations concatenated with 120 P observations gives 240 temporal dimensions normalized as described above. To balance the number of temporal mixing equations, the unit sum constraint is given a weight of 100 in the inversion presented here.

As a complement to the global feature space topology implied by the PCs, we also characterize the local structure of any manifolds embedded within the 240 dimensional temporal feature space using t-distributed Stochastic Neighbor Embedding [6]. To circumvent the stochastic nature of the t-SNE seeding algorithm, we compute multiple realizations of 2D t-SNE and then compute the PCs of each suite of individual t-SNE realizations. The temporal feature space of low order dimensions of PC(t-SNE) preserves contiguity of clusters that are persistent across multiple t-SNE realizations while achieving a greater cluster separability in the low order PC(t-SNE) feature space. In order to confirm both cluster persistence and any global structure that is captured in the 2D t-SNE feature space, we compute PC(t-SNE) for suites of 20 individual t-SNE realizations for multiple combinations of t-SNE hyperparameter settings (perplexity = 30 & 100 and learning rate 200 & 943).

Because spatiotemporal PC 1 corresponds to spatial variability of temporal means [9], the similarity (spatial correlation) of PC(t-SNE) 1 across multiple combinations of hyperparameter settings further establishes (or refutes) persistence of clusters and relative insensitivity (or sensitivity) of cluster membership for hyperparameter settings.

Computational Specifics

Computation of spatiotemporal PCs and EOFs and inversion of the temporal mixture model was performed using IDL/ENVI software. Specifics of the PC transform implementation are given by: <https://www.13harrisgeospatial.com/docs/principalcomponentanalysis.html>

The PC transform was performed using both covariance and correlation matrices with nearly identical results. Inversion of the linear mixture model used the least squares solution given above. Misfit between model-derived fractions and observations was quantified as the Root Mean Square (RMS) difference between observed and reconstructed $T + P$ time series.

Computation of t-SNE was performed in Python 3.8.8 using scikit-learn v0.24.2. Computation time was found to be sensitive to perplexity but not learning rate. On a 3.1 GHz 6-Core Intel Core i5 CPU with 64 GB of 2667 MHz DDR4 RAM and an AMD Radeon Pro 5300 4GB GPU, computation time for a single t-SNE realization with perplexity = 100 was just under 4 minutes (230 seconds for learning rate = 200; 233 seconds for learning rate = 943). Computation time decreased when the perplexity hyperparameter was lowered to its default of 30 using either the higher or lower learning rate (152 seconds at learning rate = 943; 150 seconds at learning rate = 200). Learning rate value of 943 was determined by the formula [14,15]:

$$\max(N / \text{early_exaggeration} / 4, 50)$$

where N was the total number of terrestrial pixels and `early_exaggeration` was set to the default value of 12.

3. RESULTS

The low order PC map, the temporal feature space of the decadal $T+P$ product, and the corresponding variance distribution are shown in FIGURE 2. The PC 1/2 projection bears a strong similarity to the T/P bivariate distribution shown in FIGURE 1. The variance partitions of the normalized T and P products show 3 and 4 low order dimensions (respectively) as distinct from the continuum of higher order dimensions. The 4 low order dimensions of T account for 99% of total variance while the 4 low order dimensions of P account for only 82%, with a correspondingly higher stochastic residual in the higher dimensions. The variance partition of the combined $T + P$ product shows a less distinct difference between deterministic low order dimensions and stochastic residual, but nonetheless accounts for 92% of total variance in the first 3 dimensions.

The temporal EOFs of the first 14 dimensions contain clear annual and semiannual periodicities. In the same way that any pair of periodic EOFs can define a 2D phase space in which linear combinations of periodicities of equal frequency can represent all phases of that frequency, higher

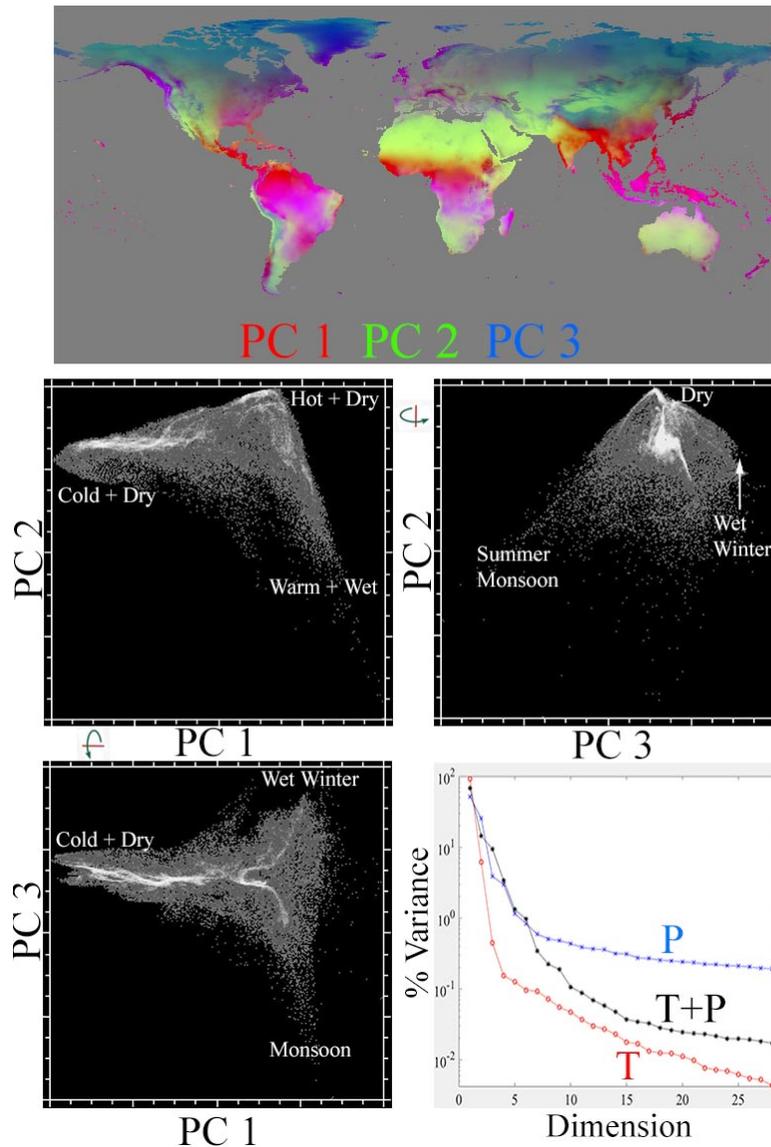


Figure 2: Principal component map and temporal feature space for the 1961-2020 mean decadal T+P monthly time series. The normalized T+P temporal feature space has 68% + 14% + 10% = 92% of total variance in the first 3 dimensions, while T and P have 99% and 81% respectively. The temporal endmembers at the apexes of the T+P feature space represent climatic extremes bounding the space.

dimensional phase spaces of annual and semiannual periodicities can represent an even greater variety of periodicities by incorporating more subtle temporal asymmetries from combinations of EOFs with slightly different phases. In this way, the quasi-sinusoidal EOFs can capture the decidedly non-sinusoidal, but still periodic, cycles of *P* commonly observed in many climates. In contrast, annual *T* cycles tend to be more sinusoidal as they are driven primarily by annual and latitudinal

variations in solar insolation. Together, these 14 low order dimensions with quasi-periodic EOFs account for 98.95% of total variance in the decadal $T + P$ product.

Neither ENSO nor NAO are manifest in the low order temporal EOFs, suggesting that they do not contribute sufficient coherent spatiotemporal variance in terrestrial temperature and precipitation on global scales, relative to the larger scale annual cycles of temperature and precipitation that dominate the low order temporal EOFs. This is consistent with the fact that ENSO and NAO are primarily manifest as sea surface temperature (SST) anomalies – which are not represented in the land-based CRU TS product. In an EOF analysis of >130 years of SST reconstruction, Mestas-Nuñez and Enfield in 1999 [16], find that the dominant modes of ENSO are manifest as a multidecadal signal with largest response in the North Atlantic followed by an interdecadal signal with largest response in the eastern North Pacific [16]. While terrestrial T and P anomalies are often attributed to ENSO and NAO effects, the spatial variability of these anomalies is apparently too spatially variable to be manifest as interannual variability in any specific low order temporal EOFs. In the context of the higher dimensional phase space discussed above, this highlights one of the challenges of interpreting individual temporal EOFs in the context of deterministic physical processes. Even if the process does have some degree of annual or interannual periodicity, this periodicity may be represented in 3 or more dimensions of the temporal feature space of geographically variable PCs.

The topology of the 3D $T+P$ feature space has 4 well-defined apexes defining a convex hull bounding almost all of the 67,420 land-based grid cells with T and P time series. The upper apexes in the PC 1-2 projection correspond to arid warm and cold climate temperature extremes, while the lower limb extends to two apexes corresponding to high precipitation extremes differing in seasonality of precipitation. Together, these 4 bounding temporal endmembers, shown in FIGURE 3, define the temporal mixture model spanning the temporal feature space represented by the low order PC distribution.

Inversion of the 4 endmember linear mixture model yields the fraction maps shown in FIGURE 3. As expected the hot+dry and cold+dry endmembers are inversely correlated ($r = -0.8$). As a result, the essential features of the 4 endmember model can be represented in 3 dimensions while omitting the cold+dry fraction as analogous to the origin of the endmember-based coordinate system. A continuous color composite of the three warm climate fraction maps is shown in FIGURE 3. The RMS misfit map, also shown in FIGURE 3 highlights the largest misfits associated with the highest P climates. This is primarily a result of small phase differences between the high P endmembers fractions. Overall, the RMS misfit is consistent with the $T+P$ variance partition with ~8% stochastic variance and 99% of modeled observations having misfit < 0.21. Example comparisons of observed and modeled time series, shown in FIGURE 4, show generally good agreement for seasonality of T , with some disparities in amplitude. Examples from the 2% largest misfit illustrate the small phase disparities in the P which lead to large amplitude short duration residuals. In addition, it is noteworthy that mid latitude southern hemisphere temperatures are modeled 180° phase reversed, although the seasonal amplitude is generally small because most of the mid-latitude areas in the southern hemisphere are arid environments with less pronounced T seasonality than less arid northern hemisphere temperate latitudes.

The t-SNE-derived temporal feature spaces show distinct clustering and varying degrees of contiguity. FIGURE 5 compares individual t-SNE realizations with PC(t-SNE) feature spaces for 4 combinations of hyperparameter settings. As expected, higher perplexity (P) and learning rate (LR) settings produce more distinct clustering. The PC(t-SNE) feature spaces enhance both the

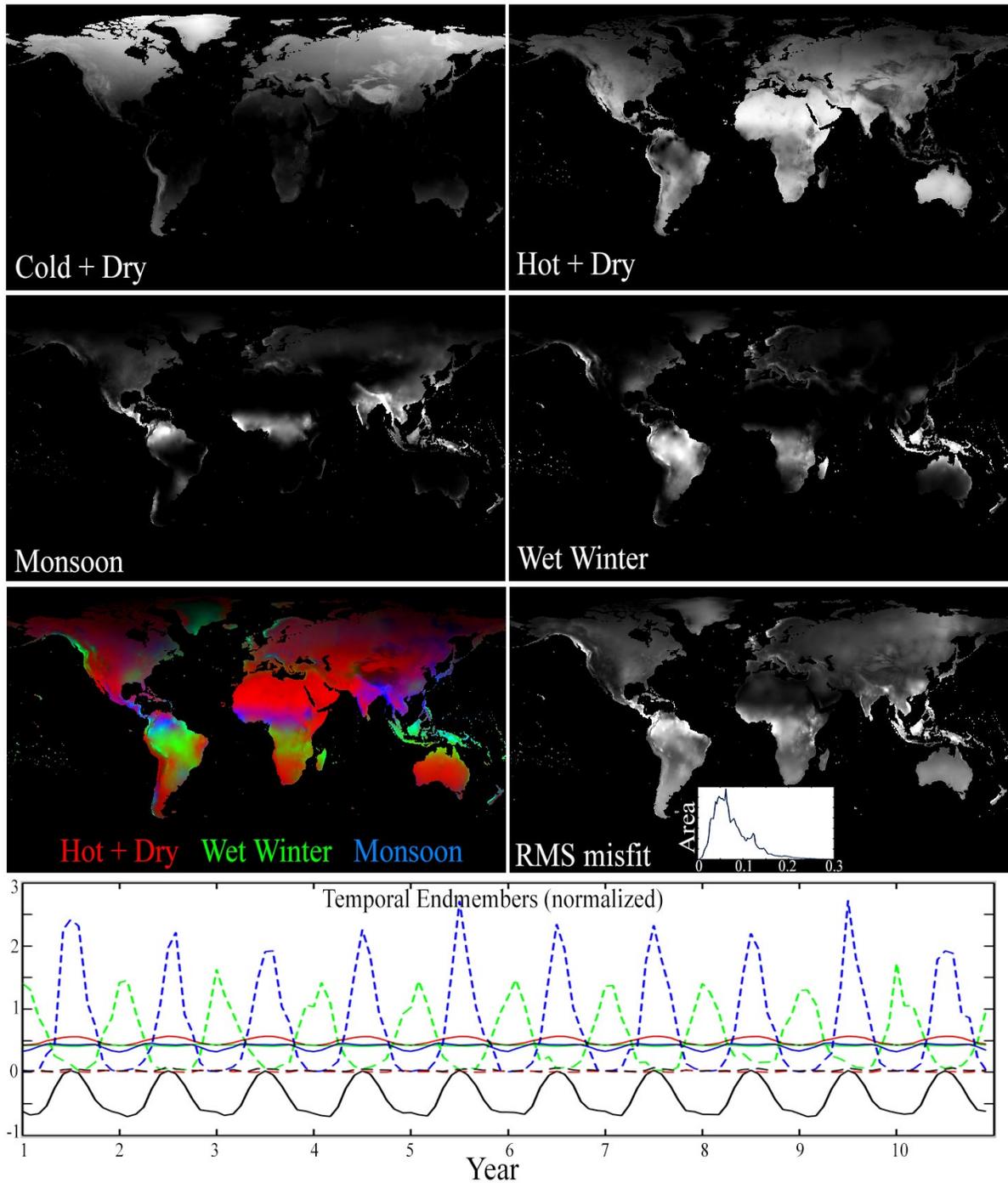


Figure 3: Temporal endmember fractions and RMS misfit obtained from inversion of the four endmember linear temporal mixture model. Shading scales linearly [0,1] for dry fractions and [0,99%] for wet fractions and RMS. Endmembers for decadal mean T(solid) and P(dashed) show wet-dry contrast among climates.

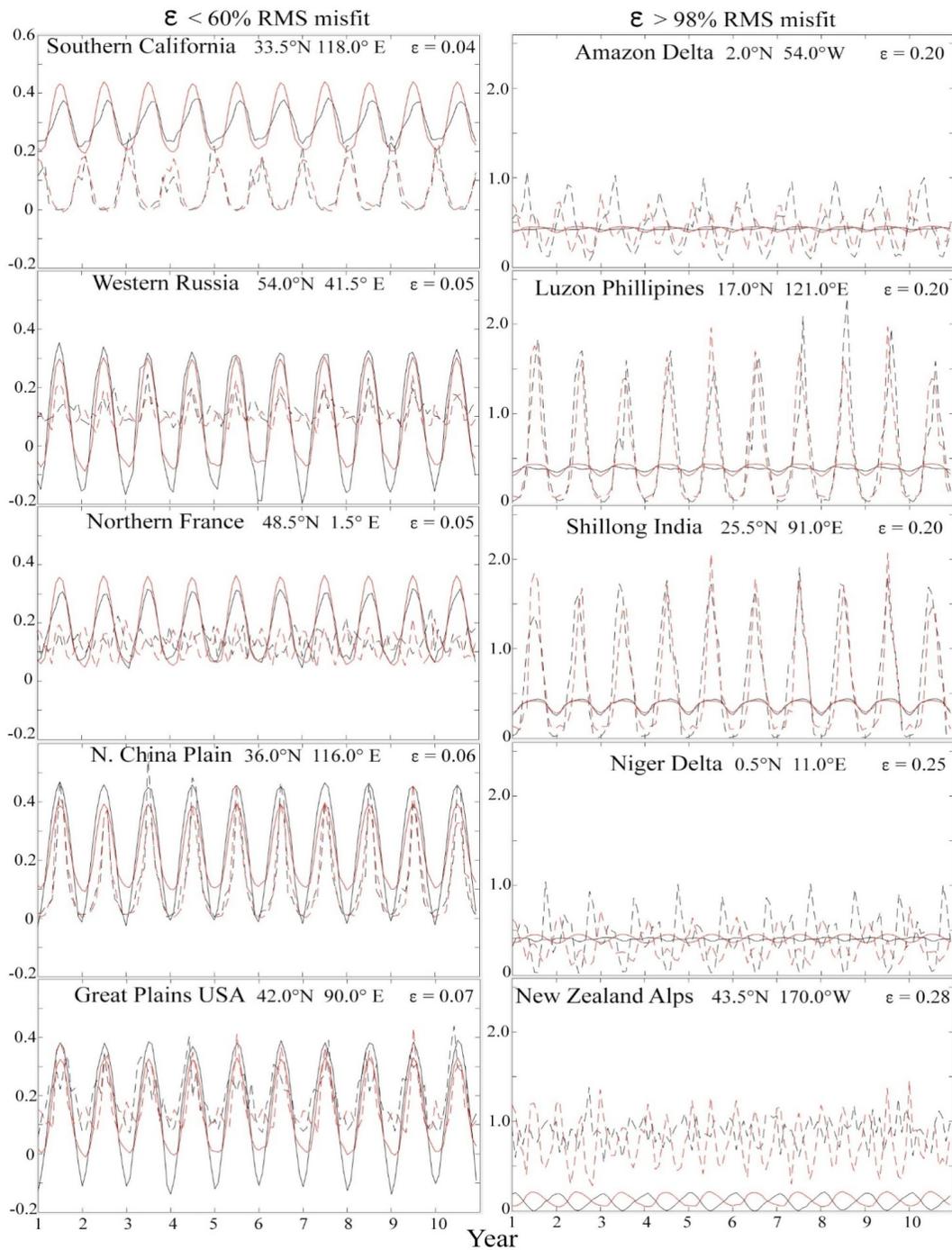


Figure 4: Comparison of observed (black) and modeled (red) temperature (solid) and precipitation (dash) for typical (<60%) RMS misfits (left) and 2% largest misfits (right). Largest misfits are associated with small phase differences in high amplitude precipitation cycles and temperate southern hemisphere phase reversals.

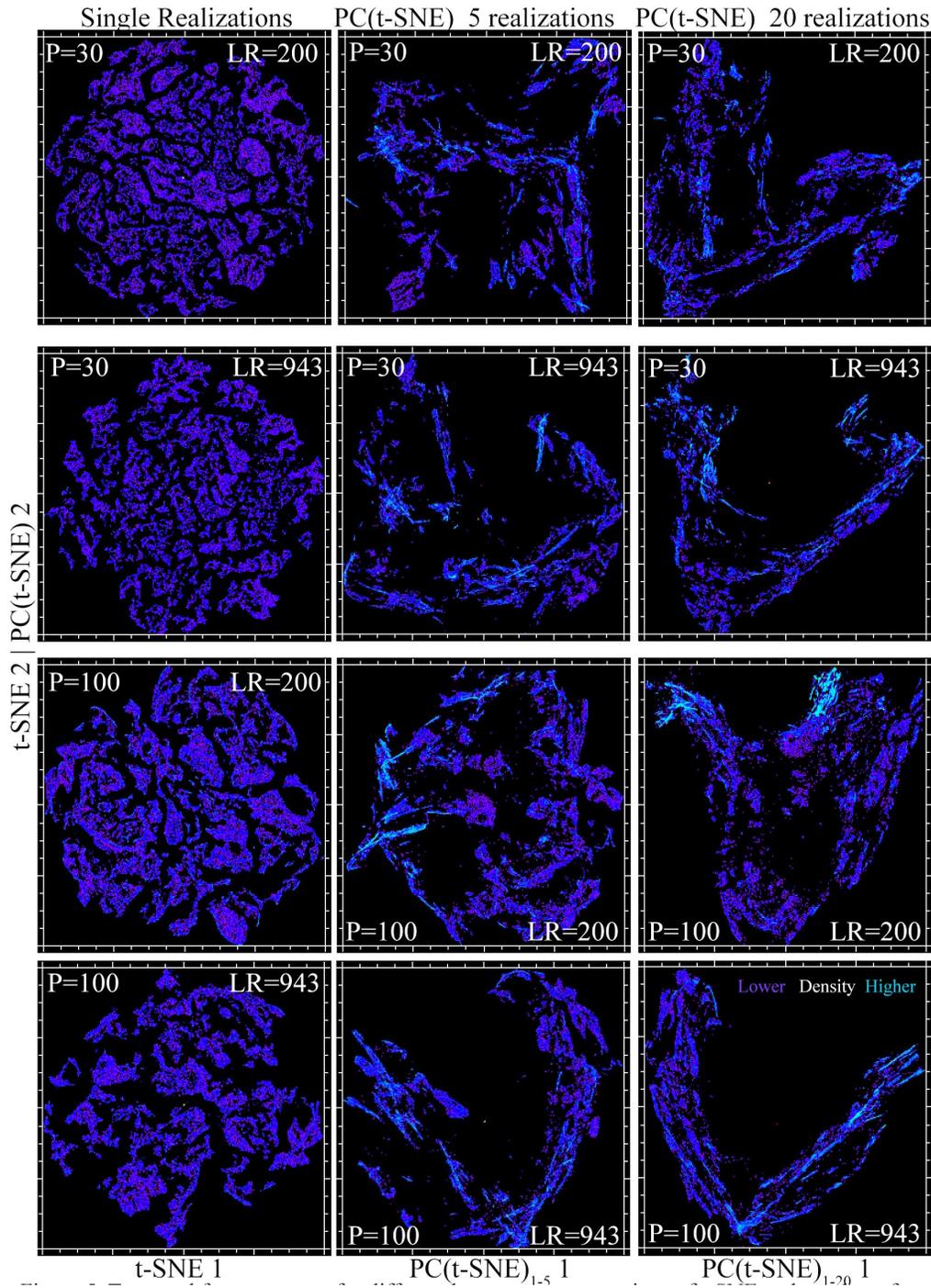


Figure 5: Temporal feature spaces for different hyperparameter settings of t-SNE and number of realizations of PC(t-SNE). Increasing perplexity (P) and learning rate (LR) both tend to increase size and connectivity of clusters while reducing number. Increasing number of t-SNE realizations increases continuity of larger clusters. All clusters are geographically contiguous.

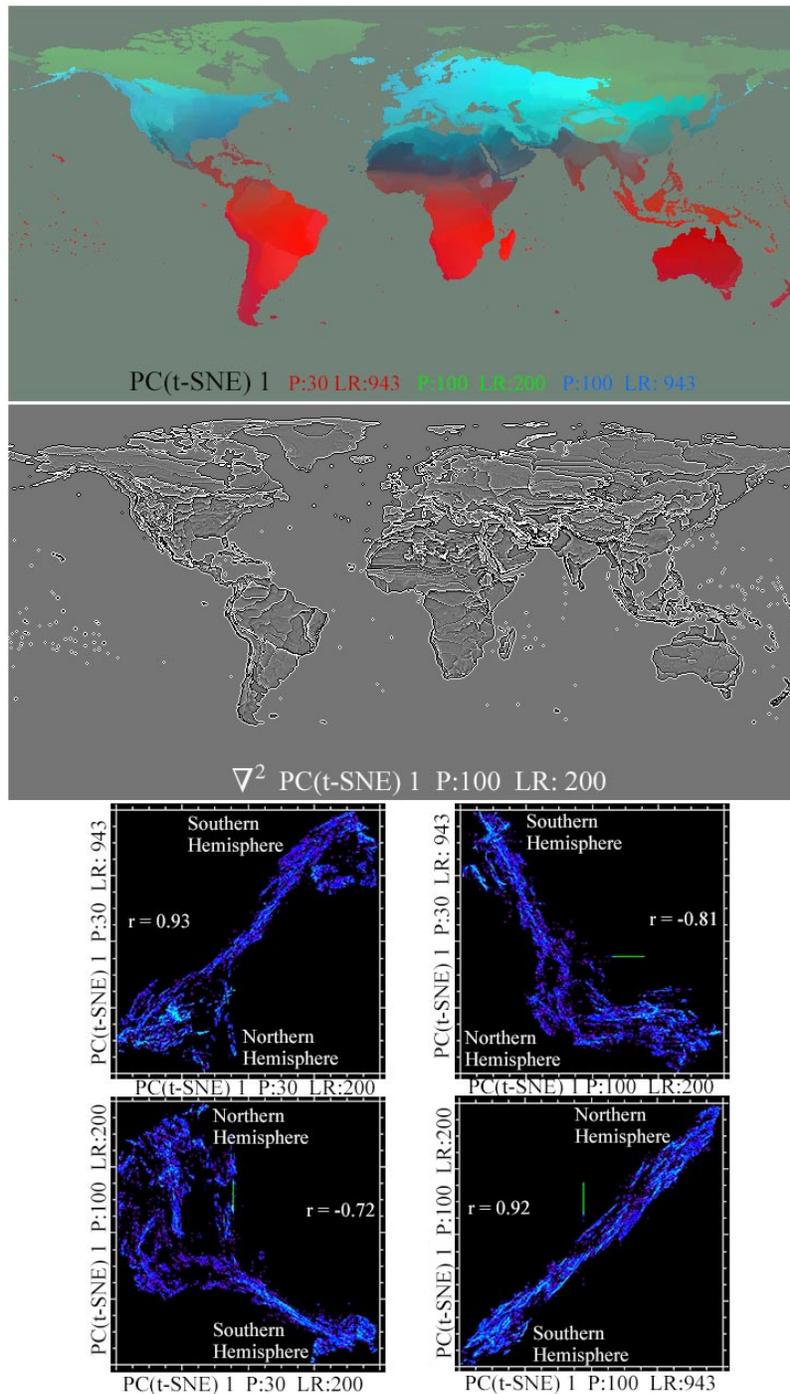


Figure 6: PC(t-SNE) 1 maps and temporal feature spaces for spatial PCs of multiple t-SNE hyperparameter settings. Color composite of 3 correlated ($r > 0.7$) PC(t-SNE) 1 (top) shows global structure preserves geography. The Laplacian cluster boundaries correspond to local manifold structures embedded within the more continuous global feature space.

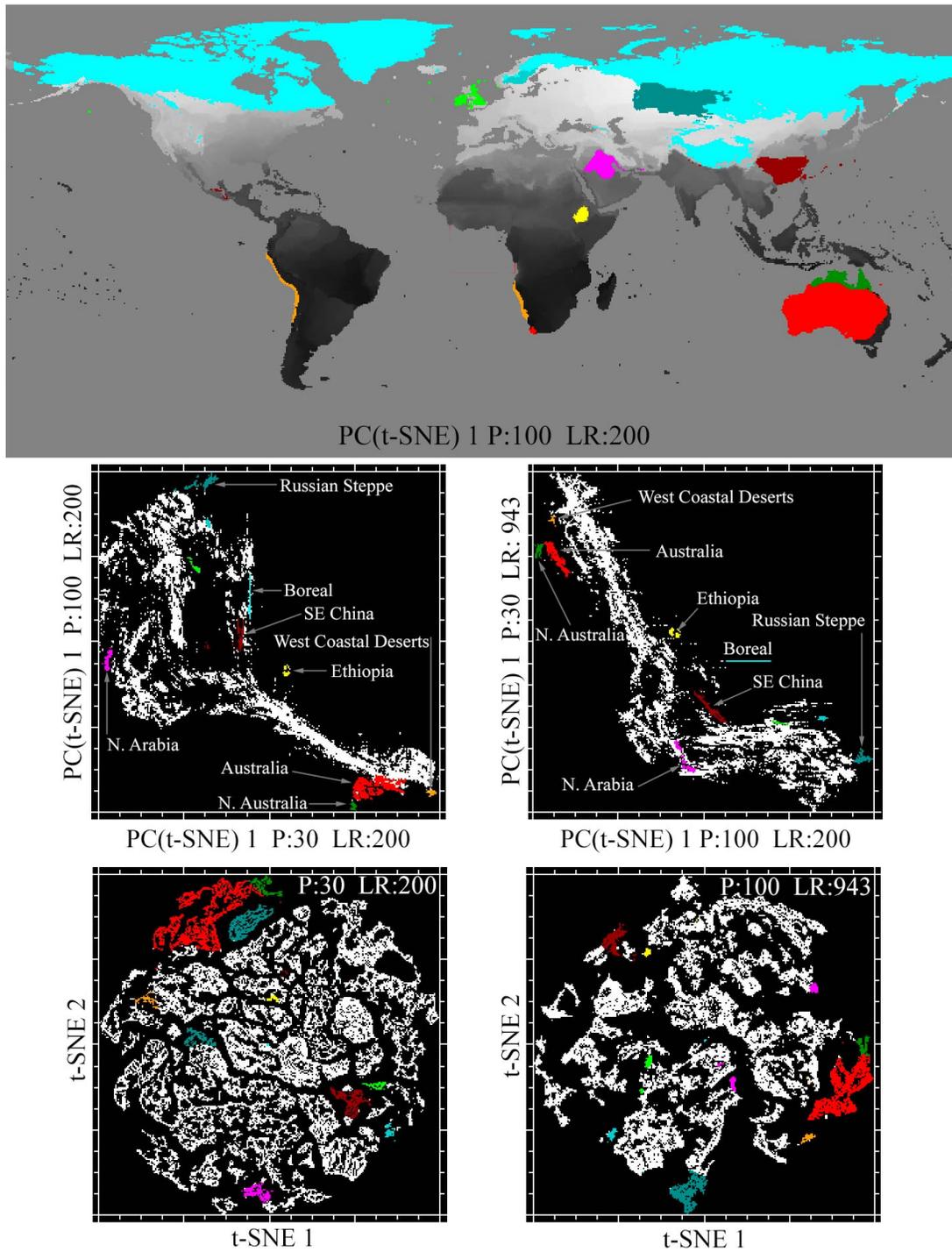


Figure 7: Clusters labeled in PC(t-SNE) feature space map onto geographically contiguous regions with distinct macroclimates. Clusters labeled in PC(t-SNE) space also map onto individual clusters when back-propagated to realizations of t-SNE feature spaces with different hyperparameters.

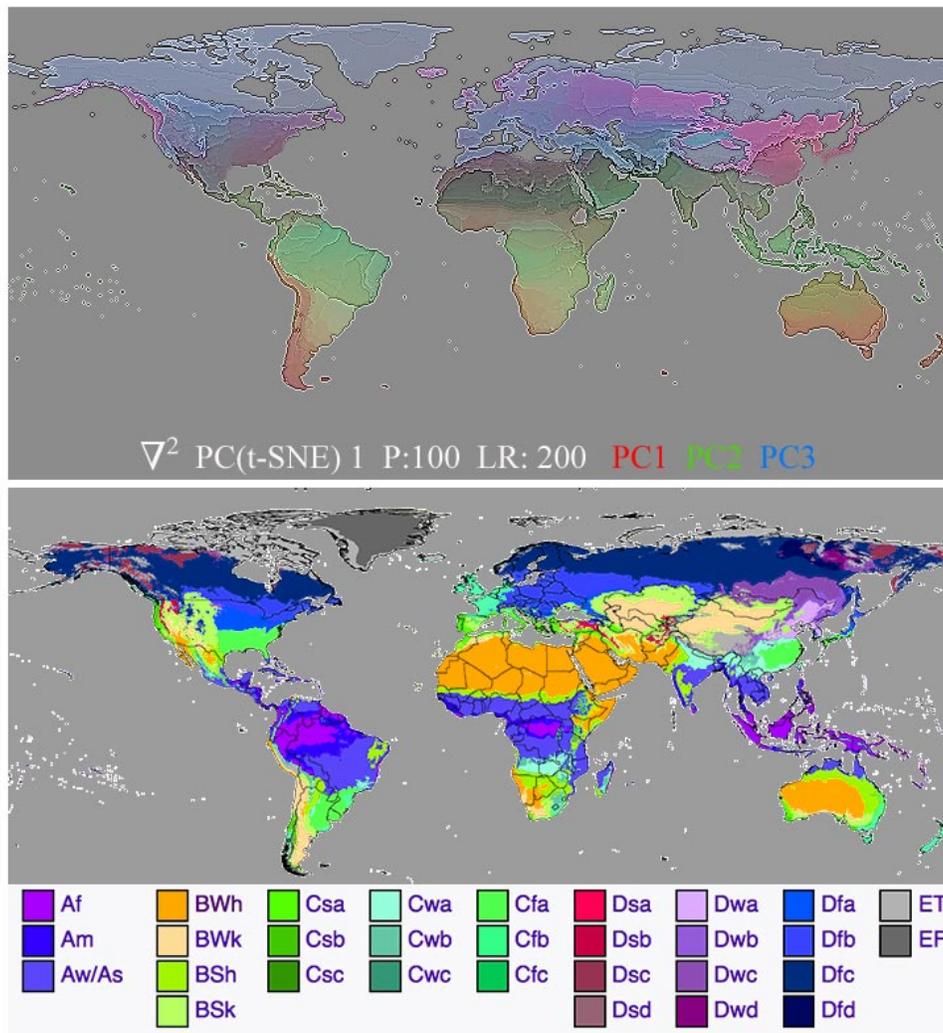


Figure 8: Comparison of PC(t-SNE) continuous+discrete climate map with updated Köppen-Geiger climatic zones from Beck et al. (2018). Hues on PC(t-SNE) map show global feature space characteristics while superimposed boundaries show persistent t-SNE manifold clusters over a range of hyperparameter settings. Broad scale features and transitions between climate groups (A-E) generally agree, finer scale features differ considerably. For detailed descriptions of Köppen-Geiger climate class codes see: https://en.wikipedia.org/wiki/Köppen_climate_classification

smaller more isolated clusters as well as the larger more contiguous clusters. Larger numbers of t-SNE realizations produce more distinct clusters, suggesting that the membership of the clusters is consistent across t-SNE realizations. This is confirmed by correlations > 0.7 for all 4 pairs of PC(t-SNE) dimension 1 for different hyperparameter combinations. Pairwise projections of these PC(t-SNE) 1, shown in FIGURE 6, have a consistent latitude-dependent global structure, as well as consistent distinction between the latitudinal continuum and the smaller, more isolated clusters.

Mapping triplets of PC(t-SNE) 1 as color composites shows both geographic contiguity of global structure, and varying degrees of spatial discontinuity, suggesting climatic discontinuities. Larger discontinuities are clearly identified by steep gradients in hue, while smaller discontinuities appear as subtle differences of similar hue. Computing the spatial Laplacian (2D curvature) highlights the discontinuities more clearly (FIG. 6). All pairwise PC(t-SNE) 1 feature spaces show the same isolated clusters, all corresponding to geographically contiguous areas with distinct climates. Labeling these isolated clusters in the PC(t-SNE) 1 feature space allows them to be mapped geographically, and back propagated to the individual 2D t-SNE feature spaces, confirming their contiguity in both geography and feature space, as shown in FIGURE 7.

Fusing any of the PC(t-SNE) 1 color composites with any of the individual PC(t-SNE) 1 Laplacian maps shows both the global structure (hue continuity) and gradients (hue discontinuity with gradient enhancement), as illustrated in FIGURE 8. Comparing this map with an updated Köppen-Geiger map reveals both similarities and differences. While both maps agree in the steep transitions between climate groups (hue gradients on both maps), there are considerable differences in the discontinuities within climate groups between the two maps.

4. DISCUSSION

The Continuous Linear Model: Why It Works

The defining characteristics of the global climate system, amplitude and timing of T and P seasonality, emerge naturally in the low order PCs and EOFs of the normalized T+P product. Linearity (or convexity) of gradients is indicated by linearity (or convexity) of binary mixing lines between the Hot+Dry temporal endmember and the other 3 endmembers, combined with the near-orthogonality of all 4 endmembers (Fig. 3) are consistent with the possibility of a low dimensional approximation of the salient features of Earth's climatic feature space. Indeed, the intrinsic spatiotemporal dimensionality suggested by the variance partition (Fig. 2) implies that the global feature space is effectively 3D. While EOFs 1-6 all show annual periodicity of both T and P , EOFs 4-6 account for < 3% of variance each.

The 3D topology of the PC feature space is also consistent with a linear mixture model. Concavity of the feature space between the Cold+Dry endmember and the 2 high P endmembers is accommodated by the linear mixture model, whereas convexity of the feature space between the two high P endmembers is not. However, the relatively small number of observations falling outside the binary mixing line spanning the high P endmembers project onto the mixing line with misfits proportional to distance. The linear model reproduces 99% of observations with < 0.21 misfit. The relatively low RMS misfit distribution is a result of the fact that mean T and P are accurately represented by the model and that relatively small amplitude and phase disparities account for the largest misfits. The most significant shortcoming of the model is the failure to accommodate the phase reversal of the southern hemisphere mid-latitude temperatures - although the fact that most are arid environments with little seasonal T variation results in relatively small RMS misfits. The New Zealand Alps example shown in FIGURE 4 is among the most egregious of misfits due to its more pronounced temperature seasonality.

Different T and P normalizations with equivalent means and standard deviations accommodate the long upper tail of the P distribution and suppresses the amplitude of the T sufficiently that the relatively low RMS conceals phase disparities in both T and P . The much longer tail of the tropical P distributions results in large excursions - even for small phase disparities. These drive the upper tail of the RMS distribution and suppress the mode. Concealed within the mode of the RMS distribution are some substantial phase disparities in T that produce relatively small misfits compared to P . However, these disparities are small in both number and amplitude due to the relatively small land area of temperate mid-latitude in the southern hemisphere. These disparities could be considered a small price to pay for the simplicity, stability, and physical interpretability of the 4 endmember model. Particularly since the phase disparities in T are primarily in the southern hemisphere deserts, where there is very low P . The phase disparities in P are also relatively small, but produce large RMS misfit because the upper tail of P amplitude is large. The normalization that gives T and P distributions with similar moments produces a temporal feature space that captures the latitudinal T gradient in both hemispheres, as well as the two most distinctive seasonal P extremes (earlier vs later year). The resulting model has effectively no overfitting (endmember fractions outside $[0,1]$ or $\text{Sum} > 1$), but substantial relative misfit in a few places. However, the misfit in those places could be considered tolerable since it does not affect which climatic temporal endmember they are most similar to. In other words, the climatic temporal endmember fractions are still reasonable everywhere because the model is driven by annual mean T , annual mean P , and seasonal timing of P . Seasonal timing of T is effectively redundant because phase polarity is redundant and latitudinal amplitude is the dominant feature of the T dimension.

The primary limitation of the continuous linear model is its inability to represent the more subtle discontinuities preserved by the t-SNE feature spaces. Because these discontinuities generally contribute little variance compared to the differences in endmember climates, they have relatively little influence on the PC rotation and end up relegated to the higher order dimensions where they may be dispersed over multiple dimensions and remain effectively unresolved as stochastic residual. However, the utility of the 4 endmember linear model, which can be reduced to 3 distinct warm climates with the Cold+Dry endmember residing at the origin (low proximity to the other 3) of a 3D space, is that it allows the essential features of the climate system to be represented by 3 numbers globally while preserving both steep and shallow climatic gradients.

The conceptual analog revealed by the topology of the feature space is straightforward. Seasonality (range + phase) of temperature (cold vs. warm) is the primary modulator of Earth's climatic system, while annual amount of precipitation (wet vs. dry) is the secondary modulator. Timing of precipitation relative to the temperature cycle is the tertiary modulator. Qualitatively, these are completely consistent with the parameters implicit in the Köppen-Geiger classification – but they are derived quantitatively and repeatably from T and P observations alone. Any bias related to plant community adaptation to regional geologic influences on soil properties, ecosystem feedbacks, or evolutionary lag with respect to changing climate are eliminated by using T and P alone.

Future Work: Limitations and Potential of the t-SNE Temporal Feature Space

As with many non-linear manifold learning algorithms, t-SNE is subject to non-uniqueness of result. Specifically, the relative position of clusters and precise location of cluster boundaries can vary considerably across multiple t-SNE realizations. Here, we address this limitation using a Monte

Carlo approach in which stability of cluster membership across multiple t-SNE realizations is taken as suggestive of physical significance. This interpretation is supported by the well-known climatic distinction and geographic contiguity of clusters like the Ethiopian highlands, British Isles and southern hemisphere west coastal deserts identified in FIGURE 7 – but further work is needed to determine the generality with which this approach can be used for other physical systems. One fundamental question yet to be resolved is the degree to which the more subtle climatic gradients implied by the clusters less clearly resolved by t-SNE correspond to physically meaningful differences in actual climate. Idiosyncracies in the changing spatiotemporal distribution of the meteorological station observations upon which the CRU TS product is based could certainly manifest as subtle discontinuities within the spatiotemporal feature space. Independent validation of the diversity of gradients implied by the t-SNE feature spaces is clearly beyond the scope of this analysis, but must eventually be addressed. This limitation has analogs with other applications of t-SNE and other dimensionality reduction approaches. In some applications, including many used for t-SNE examples, the significance of clusters is based on self-evident distinctions in discrete classes of inputs (e.g. images of alphanumeric). While these examples are illustrative of the power of proximity preservation in manifold retrieval algorithms, they say little about the potential of these algorithms to provide insight or understanding in applications where the answer is not self-evident. Independent confirmation of the physical meaning of manifold embeddings must be established for the value of manifold retrieval algorithms to achieve their full potential.

Discretizing any continuous field requires the imposition of hard decision boundaries, and the global climate space is no exception. Algorithms such as t-SNE which are inherently designed to preserve statistically distinct manifolds run the risk of introducing arbitrary cutoffs among discrete classes which may or may not reflect real geophysical properties. Approaches like Köppen-Geiger attempt to develop optimal decision boundaries on the basis of expert knowledge, merging climate information with ancillary biogeographic information and incorporating both the strengths and weaknesses of that ancillary information. In one sense, the approach presented here provides a complementary cognitive model that is rooted in statistical characterization of physical climate observations alone – leveraging information from both high-variance global feature space topology and low-variance local manifolds to characterize multiscale spatiotemporal patterns of climate variability.

5. CONCLUSION

We present a continuous representation of the global climate space. While traditional discrete models are only capable of representing abrupt transitions in temperature and/or precipitation, the continuous model is capable of representing both abrupt transitions and shallow gradients. The continuous model we present is parsimonious, using a linear combination of only 4 $T + P$ patterns, in contrast to the 30 discrete classes used by the popular Köppen-Geiger classification. We find the continuous model capable of representing 99% of historical terrestrial time series observations with root-mean-square misfit < 0.21 . Model performance is consistent with effectively 3D (92% variance) global climate spatiotemporal dimensionality. Further information is obtained by identifying stable local manifolds using Monte Carlo t-distributed stochastic neighbor embedding. Taken together, the continuous representation presented here is found to reflect the continuous character of the global climatic temporal feature space more faithfully than could be possible using a discrete classification alone.

6. ACKNOWLEDGEMENTS

CS was supported by the endowment of the Lamont Doherty Earth Observatory of Columbia University. DS was funded by the taxpayers of the state of California. DS thanks Dave Osleger and Isabel Montañez for years of illuminating conversations about the Earth system. The authors are grateful to the anonymous reviewers for helpful comments and suggestions.

7. CONFLICT OF INTEREST

The authors declare no financial or ethical conflicts of interest.

8. DATA AVAILABILITY

The T and P time series used in this analysis are provided by the University of East Anglia (UEA) Climatic Research Unit (CRU) global gridded product (v.4.05) available from <https://crudata.uea.ac.uk/cru/data/hrg/>. The results of our analysis, including the gridded endmember fraction maps and t-SNE embeddings are available from the authors upon request.

References

- [1] Köppen W. Klassifikation der Klimate nach Temperatur, Niederschlag and Jahreslauf, Petermanns Geographische Mitteilungen. 1918;64:193–203.
- [2] W. Köppen and R. Geiger.. 'Das geographische System der Klimate [The geographic system of climates]' in (eds.), Handbuch der Klimatologie [Handbook of Climatology]. Borntraeger: Berlin.1936.
- [3] R. Geiger. Überarbeitete Neuauflage von Geiger, R.: Köppen-Geiger/Klima der Erde. (Wandkarte 1:16 Mill.), Klett-Perthes, Gotha. 1961.
- [4] <https://arxiv.org/ftp/arxiv/papers/2108/2108.09545.pdf>
- [5] <https://arxiv.org/ftp/arxiv/papers/2112/2112.01416.pdf>
- [6] Maaten LVD, Hinton G. Visualizing Data Using t-SNE', Journal of Machine Learning Research.2008;9:2579–605.
- [7] Harris I, Osborn TJ, Jones P, Lister D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset, Scientific Data. 2020;7:1-18.
- [8] HoughtonJT, Filho LGM, Callander BA, Harris N, Kattenberg A, et al. Climate Change 1995: The Science of Climate Change. Cambridge University Press: Cambridge UK. 1996.
- [9] Small C. Spatiotemporal Dimensionality and Time-Space Characterization of Multitemporal Imagery. Remote Sensing of Environment.2012;124:793-809.

- [10] Lorenz EN. Empirical orthogonal functions and statistical weather prediction. 1956.
- [11] Preisendorfer RW, Preisendorfer R, Mobley C, Preisendorfer R, Mobley W. Principal Component Analysis In Meteorology And Oceanography. (Elsevier: Amsterdam). 1988.
- [12] Boardman JW. Inversion of Imaging Spectrometry Data Using Singular Value Decomposition. In IGARSS'89 12th Canadian Symposium on Remote Sensing. Vancouver, B. C. 1989;2069-2072.
- [13] Settle JJ, Drake NA. Linear Mixing and the Estimation of Ground Cover Proportions. International Journal of Remote Sensing. 1993;14:1159-1177.
- [14] Tabula Muris Consortium. Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris. Nature. 2018;562:367-372.
- [15] Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, et al. Molecular Architecture of the Mouse Nervous System. Cell. 2018;174:999–1014.
- [16] Mestas-Nuñez AM, Enfield DB. Rotated Global Modes of Non-enso Sea Surface Temperature Variability. Journal of Climate. 1999;12: 2734-2746.