Article

An Alternative to PCA for Estimating Dominant Patterns of Climate Variability and Extremes, with Application to US Rainfall

Stephen Jewson

Risk Management Solutions Ltd., Peninsular House, 30 Monument Street, London EC3R 8NB, UK; stephen.jewson@gmail.com; Tel.: +44-(0)7858-393370

Abstract: Floods and droughts are driven, in part, by spatial patterns of extreme rainfall. Heat waves are driven by spatial patterns of extreme temperature. It is therefore of interest to design statistical methodologies that allow the identification of likely patterns of extreme rain or temperature from observed historical data. The standard work-horse for identifying patterns of climate variability in historical data is Principal Component Analysis (PCA) and its variants. But PCA optimizes for variance not spatial extremes, and so there is no particular reason why the first PCA spatial pattern should identify, or even approximate, the types of patterns that may drive these phenomena, even if the linear assumptions underlying PCA are correct. We present an alternative pattern identification algorithm that makes the same linear assumptions as PCA, but which can be used to explicitly optimize for spatial extremes. We call the method Directional Component Analysis (DCA), since it involves introducing a preferred direction, or metric, such as 'sum of all points in the spatial field'. We compare the first PCA and DCA spatial patterns for US rainfall anomalies on a 6 month timescale, using the sum metric for the definition of DCA in order to focus on total rainfall anomaly over the domain, and find that they are somewhat different. The definitions of PCA and DCA result in the first PCA spatial pattern having the larger explained variance of the two patterns, while the first DCA spatial pattern, when scaled appropriately, has a higher likelihood and greater total rainfall anomaly, and indeed is the pattern with the highest total rainfall anomaly for any given likelihood. In combination these two patterns yield more insight into rainfall variability and extremes than either pattern on its own.

Keywords: principal component analysis; PCA; directional component analysis; DCA; empirical orthogonal functions; extremes; US rainfall

1. Introduction

Principal component analysis (PCA), also known as Empirical Orthogonal Function (EOF) analysis, is often used in climate research and related fields for analysing correlated data in two or more dimensions. PCA is widely used because of its mathematical elegance, mathematical properties, and simplicity. It has various uses, such as filling gaps in historical data sets [1] and identifying patterns of variability [2]. When being used to tackle some problems, however, limitations of PCA may become apparent, and this has led to the development of various extensions of PCA, each addressing a different issue. For instance, the spatial patterns identified by PCA tend to fill the spatial domain being analysed, and in some cases it would be more appropriate to identify more localised patterns. This led to the development of rotated EOF analysis, as studied in, for instance, Mestas-Nunez [3] and Lian and Chen [4], and used in Chen and Sun [5]. In another extension, known as extended EOF analysis, PCA has been used to understand developments of patterns in time [6], and in yet another has been adapted to better handle skewed data [7]. PCA and related methods have been discussed in text books such as Wilks [8], von Storch and Zwiers [9] and Jolliffe [10] and in the review paper Hannachi et al. [11].

In a recent project to develop methods to improve resilience of financial institutions to drought shocks by identifying the patterns of rainfall that might drive the largest droughts over the domain being analysed [12] we have become aware of a property of PCA, that, in the context of the goals of this particular project, is a shortcoming. When applied to observed rainfall anomalies, the first PCA pattern maximises explained variance, by definition, but the spatial pattern does not necessarily maximise the total rainfall anomaly in any sense. The total rainfall anomaly in the first PCA spatial pattern could even be zero or very close to zero. This is simply a result of the mathematical definition of PCA and what PCA is designed to capture. As a result the first PCA spatial pattern may not be particularly relevant in terms of its impact via floods (or droughts), which are likely to be, at least to some extent, related to the size of the total anomaly over the spatial domain. One could imagine that other spatial patterns, selected based on a total rainfall anomaly criterion of some sort, may be more relevant.

Motivated by this observation, we have studied an alternative pattern identification scheme that we call DCA (Directional Component Analysis). Rather than defining the first pattern as that which maximises explained variance, DCA defines the first spatial pattern as that which has the highest likelihood for a given level of total rainfall anomaly. This spatial pattern is also, conversely, the spatial pattern with the greatest rainfall anomaly for a given likelihood of occurrence. PCA and DCA spatial patterns can be scaled to create new spatial patterns, and appropriately scaled, the first DCA spatial pattern *both* has a higher likelihood *and* contains a greater total rainfall anomaly than the first PCA spatial pattern, and as such may indeed be more relevant for understanding extremes such as floods or droughts. Similarly, if applied to temperature anomalies, the first DCA spatial pattern may have more relevance than the first PCA spatial pattern for understanding heat waves. The first DCA spatial pattern is only the same as the first PCA spatial pattern in the degenerate case in which the data is fully correlated in space, in which case the first PCA spatial pattern is uniform rainfall throughout the whole domain. In this article we will restrict most of our discussion to the properties of the spatial patterns that can be derived using PCA and DCA, as opposed to the corresponding time-series. The spatial patterns are considered as possible realisations of spatial patterns that may occur in the future.

The definitions of PCA and DCA can be contrasted as follows. Spatial patterns of variability of rainfall anomalies have various mathematical properties, including explained variance (defined below), length (based on considering patterns as vectors), total rainfall anomaly and likelihood (in the context of a statistical model such as the multivariate normal distribution). PCA is a mathematical method that considers two of these properties (explained variance and length) and finds the spatial patterns and time-series pairs with the greatest explained variance, among all spatial patterns of the same length. The standard definition of PCA quite deliberately does not take account of the total rainfall anomaly across the spatial pattern or the likelihood, although it can be reformulated in terms of likelihood (see Section 2.1 below). DCA considers a different pair of properties, and is derived by finding the spatial pattern that has the greatest total rainfall anomaly, among all patterns of the same likelihood. The definition of DCA quite deliberately does not take account of explained variance and pattern length. We see from this that PCA and DCA maximise different aspects of a pattern, with different constraints, and not surprisingly they have different uses, different possible interpretations, and give different results.

In Section 2 we give a brief overview of PCA, as a basis for comparison with DCA. We give two derivations for PCA: the first, based on explained variance, is the more common. The second, based on maximising likelihood, is less usual, but makes a link to DCA. In Section 3 we give two derivations of DCA. The first is similar to the likelihood derivation for PCA, while the second is based on regression. In Section 4 we apply both PCA and DCA to a rainfall data-set for the United States and compare the first spatial patterns, which are somewhat different. In Section 5 we summarize and conclude. In the supporting information we give two simple examples of PCA and DCA to illustrate the DCA method.

The first (Section S1) is a numerical example with a 2 × 2 covariance matrix. The second (Section S2) gives the general solution for PCA and DCA for any diagonal 2 × 2 covariance matrix. In Sections S3–S8 we prove the orthogonality of the first two DCA patterns, and provide some proofs of the main optimality properties of DCA.

2. Principal Component Analysis

We now briefly review PCA as a basis for comparison with DCA. Consider a space-time dataset of anomalies X with spatial dimension s and temporal dimension t. In the example in Section 4 below we will use gridded maps of monthly US rainfall anomalies for 114 years (from Harris et al. [13]), for which the time dimension is 1363 (12 months, times 114 years, minus 5 because we use a 6 month running mean) and the space dimension is 3319. If the data X is projected onto an unknown $s \times 1$ spatial pattern vector g the $t \times 1$ time series p of amplitudes of the projection is given by the vector-matrix product:

$$p^T = g^T X \tag{1}$$

The variance v of this time series p is a scalar and is given by

$$v = \frac{1}{t} p^T p = \frac{1}{t} g^T X X^T g = g^T C g \tag{2}$$

where $C = \frac{1}{t} X X^T$ is the $s \times s$ empirical covariance matrix of the data X. The scalar v is known as the explained variance of the pattern g in the data set X.

We can imagine varying the vector g, subject to the constraint that g is unit length (i.e., that $g^T g = 1$), and trying to maximise the variance v. Mathematically, this can be done by maximising the Lagrange function

$$c = v - \lambda(g^T g - 1) = g^T C g - \lambda(g^T g - 1) \tag{3}$$

where c is a scalar cost function, and λ is a Lagrange multiplier that multiplies $(g^T g - 1)$, the expression that defines the constraint. In this equation $g^T C g$ is largest for long vectors that project highly onto the eigenvectors of C. The $g^T g$ term constrains length, but does not influence direction.

Differentiating with respect to the vector g gives

$$\frac{dc}{dg} = 2Cg - 2\lambda g \tag{4}$$

Setting equal to zero to find the extrema gives the equation:

$$Cg = \lambda g \tag{5}$$

the solutions of which are the eigenvectors of C, also known as the left singular vectors, or EOFs, of X. These eigenvectors can be interpreted as spatial patterns. The first eigenvector has the property that it maximises the explained variance in the data X, the second that it maximises the explained variance in what is left after the first eigenvector has been removed, and so on. The eigenvectors form an orthonormal set, and the time series associated with the eigenvectors also form an orthonormal set.

We note that there are various alternative ways to define PCA, with different terminology. For instance, finding the first PCA pattern can be defined (entirely equivalently to the above derivation) as finding a linear combination of the time series from each spatial point such that the linear combination maximises variance while the weights within the linear combination (known as the loadings) satisfy a length constraint. The spatial pattern then consists of the loadings.

2.1. PCA Alternative Derivation

An alternative derivation of PCA, which makes the link to the first of the derivations of DCA given below, is to assume that each spatial pattern in the dataset set X (that is: each column of X, each corresponding to a fixed time) is a single realisation of a multivariate random variable (also known as a vector random variable) with a multivariate normal distribution. The mean of the multivariate normal is zero (corresponding to no rain anywhere) because we are considering anomalies. The covariance matrix of the multivariate normal is estimated from data as given above and captures the variance of rainfall anomalies at each individual spatial point and the covariances of rainfall anomalies between points. Any possible spatial pattern lies within the support of the multivariate normal, and for each spatial pattern we can calculate the probability density, or likelihood, from the expression for probability density for the multivariate normal, which gives a measure of how likely that spatial pattern is to occur within the context of the distribution. Saying we are modelling the rainfall anomalies using a multivariate normal in this way is equivalent to saying that we are modelling the rainfall at each individual spatial location as normally distributed (so that the rain at each point in time is a single realisation from the normal distribution for that location), and that we model the dependencies between the rainfall anomalies at different locations using the multivariate normal structure.

Within the fitted multivariate normal, spatial patterns with very small rainfall anomalies, which are close to the mean, are more likely to occur, and have high values for the likelihood, while spatial patterns with large rainfall anomalies, which are far from the mean, are less likely to occur and have lower values for the likelihood. Also, patterns that have a spatial structure highly consistent with the covariance matrix (and hence consistent with the historical data, since the covariance matrix is derived from the historical data) are more likely to occur, and have higher likelihoods, while patterns that have spatial structure which is less consistent with the covariance matrix are less likely to occur and have lower likelihoods. To derive PCA, we can then try to find the unit vector spatial pattern with the highest likelihood in the fitted multivariate normal distribution, instead of that which explains the most variance. Likelihood itself is awkward to manipulate in this case, but instead, and equivalently, we can maximise log-likelihood, which is more convenient. The log-likelihood considered as a function of the unknown spatial pattern g for the multivariate normal is proportional to the Mahalanobis consistency, which is minus one times the Mahalanobis distance $M^2 = g^T C^{-1} g$, where C^{-1} is the inverse or pseudoinverse of the covariance matrix C (the connections between PCA, the multivariate normal distribution, and the Mahalanobis distance are discussed in detail in Wilks [8] and von Storch and Zwiers [9]). The Lagrange function for this new problem has two terms: one for the log-likelihood term $-M^2$, and one for the unit length constraint as before, and is given by:

$$c = -M^2 - \lambda(g^T g - 1) = -g^T C^{-1} g - \lambda(g^T g - 1) \tag{6}$$

In this equation the $-g^T C^{-1} g$ term is largest (most positive) for short vectors, and vectors that project highly onto the eigenvectors of C^{-1} and C. The solutions of this equation are also the eigenvectors of C, and so are also the PCA patterns.

The data does not in fact have to be multivariate normal for this derivation (and the subsequent derivation of DCA) to make sense. Mahalanobis consistency is a reasonable general measure for consistency of a vector g with a covariance matrix C, and can be considered a multivariate generalisation of z values, the number of standard deviations from the mean [8]. In non-normal cases, PCA, by this derivation, finds the unit vector that maximises this Mahalanobis consistency, $-M^2$.

2.2. Two Dimensional Example

PCA is illustrated in Figure 1 for a simple case. The two axes represent rainfall anomaly amounts in two locations. The ellipse represents a contour of constant likelihood (which is also a contour of constant log-likelihood, and a contour of constant Mahalanobis consistency) from the joint probability distribution of rainfall at these locations (with higher likelihoods inside the ellipse). The principal axis of the ellipse, illustrated by the double arrow, is tilted slightly to the left of vertical, indicating a negative correlation between precipitation at these two locations. The first PCA spatial pattern is a scaled version of this principal axis vector (scaled to be a unit vector), and consists of negative rainfall anomalies (or loadings) at location one (horizontal axis), and positive rainfall anomalies (or loadings) at location two (vertical axis), reflecting the negative correlation. In this example we will assume that the likelihood value that defines the ellipse has been chosen so that the PCA arrow shown is the exact unit-scaled first PCA spatial pattern. The two diagonal lines represent contour lines of total rainfall anomaly, summed across the two locations, and the highest total rainfall anomaly amounts are in the top right hand corner of the figure. The total rainfall anomaly of the first PCA spatial pattern is not particularly large in this case since there is a cancellation of rainfall anomalies to some extent between the two locations, because of the negative correlation.

The dotted arrow is the first DCA spatial pattern and is explained below.

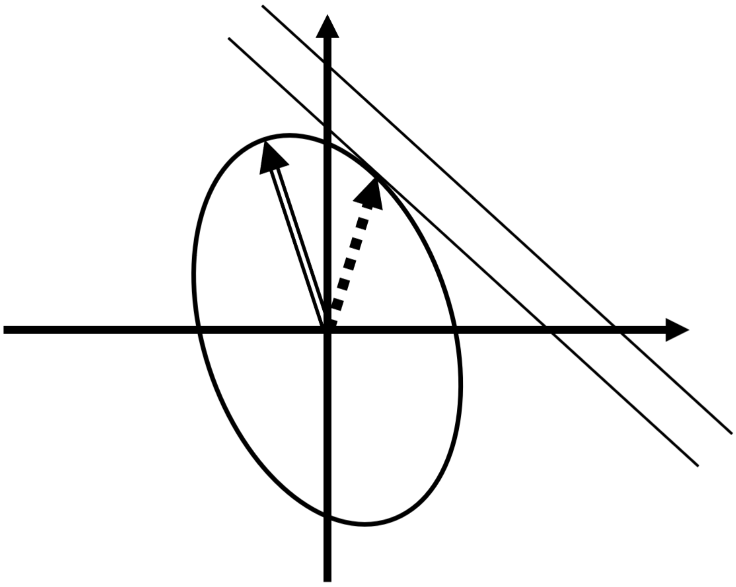

Figure 1. PCA and DCA spatial patterns in a space with two dimensions. The axes are the two dimensions, which might be, for instance, rainfall anomaly amounts at two locations. The diagonal lines then show lines of constant total rainfall anomaly. Assuming that the two variables are bivariate normal distributed in space the ellipse shows a contour of constant likelihood (probability density) or constant Mahalanobis distance, with higher likelihoods (lower Mahalanobis distances) inside the ellipse. Each point in this two dimensional space represents a spatial pattern, consisting of a single realisation from the bivariate normal, made up of rainfall anomaly values at the two locations. The tip of the double arrow gives the rainfall anomaly values for the first PCA spatial pattern, while the tip of the dotted arrow shows rainfall anomaly values for the first DCA spatial pattern. In this case the two patterns are normalized to have the same likelihood. The PCA spatial pattern is longer and has larger explained variance, while the DCA spatial pattern (which is the point on the ellipse with the greatest total rainfall anomaly, by definition) is shorter but captures more total rainfall anomaly.

3. Directional Component Analysis

Having reviewed PCA, we now describe DCA. For the same space-time data set X, again considered to be multivariate normal over the spatial dimension, and a new unknown spatial pattern g, we derive DCA by solving a different constrained maximisation problem, which is to look for the highest likelihood spatial pattern given a certain level of total rainfall anomaly. Explained another way: many different rainfall anomaly spatial patterns can give the same total rainfall anomaly. Which of them is the most likely to occur? (and from the point of view of understanding possible future extremes, presumably the most likely one is the most interesting one to look at first). Alternatively, but equivalently, we could say we are looking for the spatial pattern with the greatest total rainfall anomaly given a certain value of the likelihood. Explained another way: many different rainfall anomaly spatial patterns are equally likely. Which of them has the highest total rainfall anomaly? (again, this is presumably the most interesting spatial pattern to look at first from the point of view of understanding extremes). Essentially we are trying to find the pattern with the greatest total rainfall anomaly, while factoring in the requirement that the pattern should have a reasonably high probability of occurring in reality, so that it is relevant. Since the actual probability of any individual pattern occurring in a continuous distribution is zero, we use probability density (likelihood) instead of probability itself. Another way to see why we include probability in the derivation is that the pattern with the greatest rainfall anomaly, not factoring in probability, is simply uniform (and infinite) rainfall everywhere, which is not an interesting result. If PCA is an attempt to find the single pattern that can tell us the most about correlated variability in the data set, then DCA is an attempt to find the single pattern that can tell us the most about the extreme high or low values of the total of the anomalies in the spatial pattern. The DCA method generalizes to non-normal data in the same way that the second derivation of PCA given above in Section 2.1 does: for non-normal data we can restate the problem using Mahalanobis consistency rather than likelihood by saying we are looking for the pattern that shows the highest Mahalanobis consistency with the covariance matrix, for a given level of rainfall, or, alternatively, the pattern with the greatest total rainfall anomaly, for a given level of Mahalanobis consistency.

Both PCA and DCA spatial patterns can be scaled to create new spatial patterns with the same spatial structure but different amplitude. In the context of the multivariate normal, these new spatial patterns are also possible values for the random variable (i.e., are possible realisations from the distribution) but will have different likelihoods and different total rainfall anomaly amounts. Scaling by a factor greater than 1 increases the total rainfall anomaly, but moves the spatial pattern further from the mean of zero, and hence decreases the likelihood. Scaling with a very large factor would lead to such large rainfall anomalies that the pattern could never realistically occur: this would be reflected in the likelihood values, which become very low for very large anomalies. Scaling with different factors occurs when PCA and DCA spatial patterns are combined with their time series to reconstruct the original data.

A scaled version of the first DCA spatial pattern is illustrated in Figure 1 by the dotted arrow. The scaling in the diagram has been chosen so that the dotted arrow hits the same contour of likelihood as the first PCA spatial pattern, and so occurs with the same likelihood. The first DCA spatial pattern points to a greater extent towards the region of greatest total rainfall anomaly in the top right hand corner of the diagram, even though it is shorter than the first PCA spatial pattern. It thus achieves a greater total rainfall anomaly amount than the first PCA spatial pattern for the same likelihood, and hence we would argue is more useful for understanding extremes of total rainfall anomaly.

One can imagine scaling the lengths of the two spatial patterns in Figure 1 in different ways to create patterns with different levels of total rainfall anomaly and different likelihoods. We use this to illustrate various properties of PCA and DCA. For instance, if we were to scale (and lengthen) the first PCA spatial pattern so that it would hit the same total rainfall anomaly line as the first DCA spatial pattern it would contain more rainfall than before but would extend outside the elliptical contour and would hence occur with a lower likelihood than the scaled DCA pattern shown. Conversely if we were to scale the DCA spatial pattern to be slightly shorter, then it can be seen that it would still achieve higher rainfall than the first PCA spatial pattern, but would have a higher likelihood. This latter case is the most interesting since it creates a pattern which is both more likely and has a greater rainfall anomaly than the first PCA spatial pattern and is hence more relevant to understanding spatial extremes from both the magnitude and the likelihood perspectives. These illustrations show general properties of PCA and DCA that are discussed further below, proven in the supplementary materials and illustrated in the example in Section 4. Geometrically, we can summarize DCA using Figure 1 very simply: DCA allows us to find the point on the ellipse that has the greatest rainfall anomaly (i.e., is the furthest to the top right of the diagram).

We can derive an expression for the first DCA pattern as follows. The total rainfall anomaly in the pattern g is given by the sum of the individual components in g. It can be written in a general form as a linear function of the components in g as the vector dot product $g^T r$, where r is the vector (1, 1, 1 ..., 1), a pattern of uniform rainfall anomaly. The analysis below also applies to any other value for the vector r, other than uniform rainfall anomaly, hence the name 'directional' component analysis: the use of r introduces a preferred direction, or metric, in addition to the directions defined by the eigenvectors of the covariance matrix C. The inclusion of a preferred direction distinguishes this method from PCA and related methods. It therefore makes most sense to consider DCA not as a variant of PCA, but as a different approach.

We can maximise log-likelihood, for a given level of total rainfall anomaly , by combining $-M^2$ and $g^T r$ in the Lagrange function:

$$c = -M^2 + 2\lambda(g^T r - 1) = -g^T C^{-1} g + 2\lambda(g^T r - 1) \tag{7}$$

We have added an arbitrary factor of 2 in the definition of λ to simplify the algebra later. In this equation both terms are influenced by both the length and direction of g. Compared with the Lagrange function used to derive PCA in Section 2.1, only the second term is different.

Differentiating with respect to the vector g gives

$$\frac{dc}{dg} = -2C^{-1}g + 2\lambda r \tag{8}$$

and setting equal to zero gives:

$$C^{-1}g = \lambda r \tag{9}$$

or

$$g = \lambda C r \tag{10}$$

To make the solution unique g can be be normalized to unit length, giving the definition of the first DCA pattern g_1, as:

$$g_1 = \frac{Cr}{|Cr|} = \frac{Cr}{\sqrt{r^T C^2 r}} \tag{11}$$

The second derivative of c is:

$$\frac{d^2c}{dg^2} = -2C^{-1} \tag{12}$$

and so we see that the solution is a maximum.

The solution for the first DCA spatial pattern is very simple. When r is a vector of 1's, the first DCA spatial pattern is simply proportional to the sums of rows in the covariance matrix.

3.1. Derivation of the Second DCA Pattern

We can derive a time-series corresponding to the first DCA spatial pattern by projecting the original data X onto the spatial pattern, giving $g_1^T X$. An approximate reconstruction of X can then be created by combining the first DCA spatial pattern with this time series, giving $g_1(g_1^T X)$, and the residuals X_2 from this approximation can be derived as:

$$X_2 = X - g_1(g_1^T X) \tag{13}$$

The second DCA spatial pattern can be derived by repeating the entire DCA analysis given above on this second data-set X_2, using $C_2 = \frac{1}{t} X_2 X_2^T$, and giving $g_2 = \frac{C_2 r}{|C_2 r|}$. This process can be continued to derive a series of spatial patterns, with corresponding time series, and as with PCA the number of spatial pattern time series pairs is equal to the rank of X. This series of spatial patterns will be mutually

Peer-reviewed version available at *Atmosphere* **2020**, *11*; doi:10.3390/atmos11040354

orthonormal, by construction, and so together will form an orthonormal set. The orthogonality of the first two patterns is proved in the supporting information (Section S3).

There are in fact an infinite number of possible orthonormal sets of spatial patterns, of which the PCA and DCA spatial patterns are both examples. However, there is only one orthonormal set for which the time-series of different patterns are uncorrelated, which is PCA. The time-series for different DCA patterns are correlated, except in the degenerate case when the DCA patterns are the same as the PCA patterns.

The set of DCA spatial patterns are empirical, and orthogonal, and so one could refer to them as empirical orthogonal functions (EOFs). However, the term EOF analysis is currently used synonymously with PCA, and so should perhaps reserved for that usage.

3.2. Regression-Based Derivation

An alternative regression-based derivation of the DCA spatial patterns proceeds as follows.

First, we construct a time series T of the total rainfall anomaly at each point in time, by projecting the data X onto the uniform rainfall vector r:

$$T = X^T r \tag{14}$$

We then regress the data X onto this time series, to give regression slopes b:

$$X = bT^T + \epsilon \tag{15}$$

The standard estimator for b is given by:

$$\begin{aligned} b &= XT(T^T T)^{-1} && (16)\\ &= XX^T r(T^T T)^{-1} && (17)\\ &= tCr(T^T T)^{-1} && (18)\\ &\propto Cr && (19) \end{aligned}$$

and we see that b is parallel to the first DCA spatial pattern g.

Once again, the second pattern can be produced by removing the data explained by the pattern b, and repeating the process, and the series of patterns thus obtained will be the same as those derived in the previous section.

3.3. Properties of DCA

We now summarise some of the mathematical properties of the first DCA spatial pattern, with the assumption that X represents rainfall anomalies as illustration. From the derivations of the first PCA and DCA spatial patterns we can say that:

- Since PCA is designed to maximise explained variance, the explained variance of the first DCA pattern will be less than or equal to the explained variance of the first PCA pattern. The explained variance will only be equal to that of the first PCA pattern in the degenerate case that the first PCA spatial pattern equals the vector r, in which case the first DCA spatial pattern will also equal r.
- Since the first DCA spatial pattern is designed to maximise total rainfall anomaly (for a given value of likelihood) the total rainfall anomaly $g_1^T r$ for the first DCA spatial pattern will be greater than or equal to that of the first PCA spatial pattern. Once again it will only be equal in the degenerate case. This property can be shown by comparing the definition of DCA with the second defintion of PCA given above. It is also is proven more carefully in the supporting information, sections S4 and S5.

- If the first DCA spatial pattern is scaled to have the same amount of total rainfall anomaly as the first PCA spatial pattern, it will have a higher or equal likelihood, equal only in the degenerate case. This property can be shown from the definitions of PCA and DCA, and is also proven in the supporting information, section S6.
- If the first DCA spatial pattern is scaled to have the same likelihood as the first PCA spatial pattern (which is how the arrows are scaled in Figure 1) it will have a greater or equal value for the total rainfall anomaly, equal only in the degenerate case. This property follows from the definitions, but is also proven carefully in the supporting information, section S7.
- In the non-degenerate case, the first DCA spatial pattern can be scaled to the in-between case where it has *both* more rainfall *and* a higher likelihood than the first PCA spatial pattern. This is the most interesting property of DCA in comparison with PCA, and is the property which suggests that DCA is the better method for identifying spatial extremes (defined here as extremes in the total anomaly summed across the pattern). It is proven in the supporting information, section S8.

We now illustrate these properties with an example.

4. Application of DCA to Observed US Rainfall

We now derive the first PCA and DCA spatial patterns for an example dataset. We use the CRU data for US rainfall from 1901 to 2014 [13] from which we calculate anomalies from the 114 year time mean. We then smooth the data with a 6 month running mean, to emphasize a 6 month timescale. We do not separate the seasons, since our goal is to identify the pattern of greatest total rainfall anomaly for a given likelihood irrespective of when it occurs during the year, although one could apply a similar analysis to rainfall for shorter periods such as certain seasons or months.

We calculate the first PCA and DCA spatial patterns on the anomalies, and we assume the data is multivariate normal in space. The multivariate normal assumption is not essential, but allows us to relate the values of the Mahalanobis consistency to the likelihood, which is easier to understand.

4.1. Unit Vector Scaling

The first PCA and DCA spatial patterns for the 6 month rainfall data are shown in Figure 2 and Figure 3. The patterns are shown as unit vectors, corresponding to the definitions. As possible realisations of the random variable of the fitted multivariate normal distribution, these are both patterns that could occur in reality, if the multivariate normal assumption is correct. These patterns could also occur in reality but with different scalings, which would lead to different total rainfall anomaly amounts and different likelihoods. The explained variances of the two patterns are 43% and 35% respectively: as would be expected from the definition of PCA, the first PCA pattern has higher explained variance. We can calculate the total rainfall anomaly in each of these spatial patterns by summing across the points within the spatial pattern. This is the total rainfall anomaly that would occur due to this pattern in a reconstruction of the data if the corresponding time series had an amplitude of 1. The PCA spatial pattern has a total rainfall anomaly of 16.6, and the DCA spatial pattern has a total rainfall anomaly of 38.1: as would be expected from the definition of DCA the DCA pattern has higher total rainfall anomaly, which in this case is higher by a factor of 2.3. The Mahalanobis distance values are $M^2_{PCA} = 1.32 \times 10^{-6}$ and $M^2_{DCA} = 2.60 \times 10^{-6}$, respectively. In summary the DCA spatial pattern has a greater total rainfall anomaly, but a lower likelihood (since it has a higher Mahalanobis distance). This is not particularly remarkable since many other patterns would have this combination of properties. The interesting properties of the first DCA spatial pattern only emerge after appropriate scaling (see below).

The ratio of likelihoods of the two patterns can be calculated from the Mahalanobis distance values as $pd_{DCA}/pd_{PCA} = \exp\left(M^2_{PCA} - M^2_{DCA}\right)$, and is extremely close to 1 i.e., the patterns have almost the same likelihood. This is because both patterns are scaled to represent very small anomalies by the requirement that they are unit vectors, and because the density of a multivariate normal distribution is almost flat near the mean. If the patterns were both scaled to represent larger anomalies then the difference in likelihoods would be larger: this is explored below in Section 4.5.

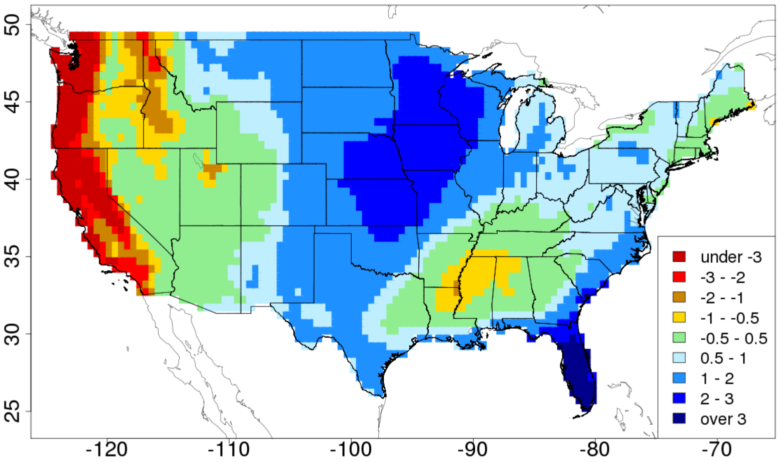

Figure 2. The first PCA spatial pattern for 6 month US rainfall anomalies. This pattern maximises explained variance.

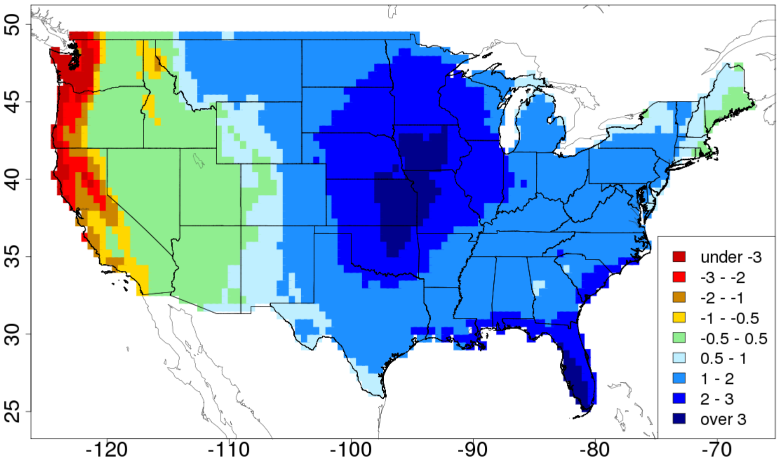

Figure 3. The first DCA spatial pattern for 6 month US rainfall anomalies. With appropriate scaling this pattern is both more likely and represents a greater total rainfall anomaly than any given scaling of the first PCA spatial pattern. As a result it is more appropriate for understanding possible extremes in total rainfall anomalies than the first PCA spatial pattern.

4.2. Equal Total Rainfall Anomaly Scaling

The interesting properties of the first DCA spatial pattern, as compared to the first PCA spatial pattern, become apparent when we adjust the scaling of one or both of the patterns. The properties of the scaled patterns illustrate that the first DCA spatial pattern is more suitable than the first PCA spatial pattern for understanding possible future extremes of total rainfall anomaly. In practice, one would likely not calculate these alternative scalings, but would simply consider the pattern structure, which is discussed in Section 4.6 below.

One possible scaling is equal total rainfall anomaly scaling i.e., scaling the two patterns to have equal total rainfall anomaly amounts. This does, of course, change the likelihoods of the patterns. Since the ratio of the spatial pattern total rainfall anomalies for DCA to PCA (given above in Section 4.1) is 2.3, if we multiply the DCA spatial pattern by the inverse of 2.3, which is 0.43, then the two patterns will have the same total rainfall anomaly. The Mahalanobis distance scales by the scaling factor squared, and so the Mahalanobis distance for this new scaled DCA pattern is 4.91×10^{-7}, which is 0.43^2 times less than before. This is now lower than the Mahalanobis distance for the unit-scaled PCA spatial pattern and hence the likelihood for the scaled DCA spatial pattern is higher than that of the PCA spatial pattern. In summary, we have derived a pattern which can be considered as a realisation from the modelled rainfall anomaly distribution, and which could occur in reality, if the multivariate normal distribution is correct. The pattern has the same total rainfall anomaly as the first PCA spatial pattern, but a higher likelihood. It is a better indicator of how a large total rainfall anomaly might occur in the future than the first PCA spatial pattern, because of the higher likelihood.

4.3. Equal Likelihood Scaling

Another possibility is equal likelihood scaling, in which we scale the DCA spatial pattern so that it has the same likelihood as the PCA spatial pattern. Since the ratio of the Mahalanobis distances for the unit vector patterns is 1.97, if we scale the DCA spatial pattern by inverse of the square root of 1.97, which is 0.71, then the Mahalanobis distances of the PCA spatial pattern and the scaled DCA spatial pattern will be the same. This also makes the likelihoods the same. The rainfall of the DCA spatial pattern will scale to 0.71 times its original value, and becomes 27.2, which is still 1.64 times the rainfall of the PCA spatial pattern. In summary we have derived a spatial pattern which can be considered as a realisation from the modelled rainfall anomaly distribution, and has the same likelihood as the PCA spatial pattern, but has a greater total rainfall anomaly. It is therefore a better indicator than the first PCA spatial pattern of how large total rainfall anomalies might materialize in the future, because of the higher total rainfall anomaly at the same level of likelihood.

4.4. Intermediate Scaling

We can also scale the DCA spatial pattern to an intermediate case in between equal total rainfall anomaly scaling and equal likelihood scaling, and the results are then more interesting than the previous two scalings in terms of resulting mathematical properties. For instance, if we apply a scaling of 0.57 to the DCA spatial pattern, which is half way between the scalings used above, then we achieve a total rainfall anomaly of 21.9 (1.31 times the rainfall of the PCA spatial pattern) and a Mahalanobis distance of 8.55×10^{-7} (0.65 times the Mahalanobis distance of the PCA spatial pattern). We see that we have created a spatial pattern which *both* has a greater total rainfall anomaly *and* has a higher likelihood (has a lower Mahalanobis distance) than the first PCA spatial pattern. In fact any scaling between the two scalings used above, and so any scaling in the range (0.43, 0.71), would have this property. Since the pattern we have created has a higher likelihood and a greater total rainfall anomaly than the first PCA spatial pattern it is presumably more relevant for understanding extreme floods and droughts: it is both more likely and worse.

In general it is not necessary to actually calculate these scalings, as we have done here. It is just sufficient to know that they exist, and the existence of such an intermediate scaling with the these properties in all non-degenerate cases is proven in S8. The existence of these scalings then provides justification for considering the spatial structure in the DCA pattern, not that in the PCA pattern, as the first spatial structure to consider when studying extremes of total rainfall anomaly.

4.5. Equal Total Rainfall Scaling at Larger Amplitude

For the scalings discussed above in Sections 4.1, 4.2 and 4.4 the ratio of actual likelihoods is very close to one, because the amplitudes of the patterns are very small. Both patterns are very close to the mean of the distribution. To create patterns with larger differences in likelihood, we have to scale both patterns to larger amplitudes. For instance, if we start with the patterns that resulted from the intermediate scaling described in Section 4.4 above we can then apply an additional scaling to both patterns so that the Mahalanobis distance of the PCA pattern is 1. This is a way to achieve a reasonable amplitude for the patterns. A Mahalanobis distance of 1 is the multivariate equivalent of being one standard deviation from zero (and hence of typical amplitude). This scaling can be accomplished by scaling both intermediate patterns derived in Section 4.4 by the inverse of the square root of the Mahalanobis distance of the PCA pattern, giving a scaling of 870. The rainfall anomaly totals for the PCA and DCA spatial patterns increase to 14,417 and 19,029. The ratio of rainfall for the two patterns stays the same, at 1.31. The Mahalanobis distances are 1 and 0.648, and the ratio of likelihoods for the two patterns is then 1.42. In this case we have again created scaled PCA and DCA spatial patterns that can both be considered as possible samples from the modelled rainfall anomaly distribution. In addition, in this case the scaled PCA pattern now has an amplitude typical of real variability. The scaled DCA pattern has a greater rainfall anomaly than the scaled PCA pattern, and has a greater likelihood by a clear margin, and hence is more relevant for understanding possible future extremes of total rainfall anomaly.

4.6. Discussion of Pattern Structure

The scalings described above illustrate the properties of spatial patterns from PCA and DCA. However, what is typically more relevant to interpretation are the different spatial structures, rather than the amplitudes of the patterns. The PCA spatial pattern in Figure 2 shows large regions of positive and negative anomalies. Some of the same regions appear in the DCA spatial pattern in Figure 3, but the large-scale eastern dipole has disappeared, and the DCA spatial pattern shows more uniform rainfall anomalies (or more uniform rainfall loadings) over a larger area. We can measure this within-pattern variability using the spatial standard deviation of each spatial pattern. The PCA spatial pattern has a spatial standard deviation of 1.3×10^{-4} while the DCA spatial pattern has a smaller spatial standard deviation of 7.0×10^{-5}. Combined with the fact that the DCA spatial pattern for equal level of likelihood has 1.64 times the total rainfall anomaly, we would argue that the DCA spatial pattern is the more relevant indicator of possible driving spatial patterns for flood or drought on the spatial scale of this domain. The PCA spatial pattern is, however, likely to be a better indication of the overall spatial patterns of variability in the rainfall anomalies and how they correlate in space (which is the answer to a different question). Taken together, the two spatial patterns give more insight into the variability and extremes in the rainfall data than either pattern taken on its own.

5. Discussion

We have described a method for pattern identification in spatially correlated multivariate space-time datasets, that we call DCA. For spatially multivariate normal data the method finds patterns with the highest likelihood subject to a linear constraint. For non-normal data the method can be described as finding the patterns with the highest Mahalanobis consistency subject to the same constraint. In the example we have presented we used rainfall anomaly data, and a sum-of-all-data-points constraint that represents total rainfall anomaly, in order to find the rainfall anomaly pattern with the highest likelihood for any given level of total rainfall anomaly. This is a reasonable answer to the question: what single spatial pattern is most likely to drive large total rainfall anomalies in the future? Applying the method to US rainfall we were able to derive a pattern of rainfall that has both a higher likelihood and a greater total rainfall anomaly than the first PCA spatial pattern. The first PCA spatial pattern is dominated by regions of positive and negative rainfall, while the DCA spatial pattern is more uniform. This first DCA spatial pattern is arguably the single pattern which has the greatest relevance for understanding future floods and droughts at this spatial and temporal scale (at least within the limitations of linear analysis).

The original motivating project for this work involved designing a simple methodology to identify representative extreme flood and drought scenarios in various parts of the world, for use in risk management, and the DCA spatial patterns seem like they may form part of a solution to this problem. For instance, the first DCA spatial pattern could be used as part of a methodology for simple quantification of extreme flood and drought scenarios via the following steps:

- A target spatial domain and timescale needs to be identified (in our example: the continential US for a 6 month timescale)
- A target return period would be identified (such as 200 yr return period)

- Standard methodologies from extreme value theory could be used to estimate the total rainfall anomaly or total drought index over the domain at that return period.
- Given this total rainfall anomaly amount the first DCA spatial pattern could be scaled to give exactly that rainfall amount. It is an appropriate pattern to represent possible rainfall extremes at that return period, since it has a higher likelihood than any other pattern with that total rainfall anomaly (by definition of DCA)
- The DCA spatial pattern so derived could then be used to drive impact models

One could imagine similar applications for deriving patterns of extreme temperature for understanding heatwaves. In both cases (rainfall or temperature) we emphasize that the use of single forcing patterns is an extreme simplification, relative to more complex models of flood, drought or heatwaves that are based on the full distribution of possible outcomes, not just a single pattern. However, simple models can play a useful role when time and resources are limited, or when only very approximate answers are required.

One limitation of basic DCA patterns in this context is that, like PCA patterns, they are linear in that they are based entirely on the covariance matrix of the data, and do not account for correlations that change with the intensity of rainfall. To account for that effect, a more detailed analysis would be necessary. That might consist of deriving PCA or DCA patterns based on data censored to only include more extreme values, for instance.

DCA patterns also shed interesting light on the use of truncated PCA for simulating surrogate data, in which the first n PCA patterns are retained, and the remaining patterns are discarded and replaced by a simple noise model such as white or red noise [8]. If the first DCA pattern projects onto the discarded PCA patterns, then simulated data from truncated PCA will fail to capture that pattern. In other words, in the context of rainfall simulation, the simulated data may not capture the pattern that maximises total rainfall anomaly at a given likelihood. This may be unfortunate if the simulation of extreme scenarios with large total rainfall anomaly is important. To avoid this one could consider basing simulation on a truncation of the series of DCA patterns instead, as follows. First, the data X would be decomposed using DCA into:

$$X = GLQ^T \tag{20}$$

where the matrix G contains the DCA spatial patterns, the matrix L is diagonal, and the matrix Q contains time series for each pattern. The time series in Q may be correlated. To model Q one might therefore use PCA, giving:

$$Q^T = E\Lambda P^T \tag{21}$$

where the new time series P are now uncorrelated and easy to replace with simulated values. Combining these expressions gives:

$$X = GLE\Lambda P^T \tag{22}$$

Truncation can then be applied via the DCA patterns, by retaining just the first m columns of G. Truncating using DCA in this way will ensure better retention of patterns with large rainfall anomaly amounts than truncation using PCA. On the other hand, it will lead to the retention of less total explained variance as compared to truncated PCA with the same level of truncation. Which is to be preferred depends on the application.

There are various possible extensions of this research. One would be to consider other directions for the direction vector r than uniform rainfall anomaly. An obvious choice would be to use r to weight the different grid points so as to reflect different levels of possible impacts at different locations. For instance, when considering extreme wind one might want to use r to weight populated areas more heavily than unpopulated areas.

One could also mix concepts from PCA and DCA. Using two Lagrange multipliers, it is possible to derive patterns that maximise likelihood subject to both a linear and a normalisation constraint, using the Lagrange function:

$$c = -g^T C^{-1} g + \lambda_1 (g^T r - 1) - \lambda_2 (g^T g - 1) \tag{23}$$

The solutions to this equation lie in-between the PCA and DCA patterns (in some sense), depending on the values of the Lagrange multipliers.

It is also possible to consider DCA but with non-linear constraints on the unknown pattern. For instance, with a quadratic impact function of the form $g^T M g$, where M is a matrix, we have:

$$c = -g^T C^{-1} g + \lambda (g^T M g - 1) \tag{24}$$

the solutions of which are the eigenvectors of the matrix product CM.

One could also consider using a cost function of the general form:

$$c = -M^2 + \lambda (f(g) - 1) \tag{25}$$

Non-linear constraints for $f(g)$ may make sense in applications where impact is a nonlinear function of the variable, as it often is. If $f(g)$ is then approximated using $f(g) \approx r^T g + g^T M g$ we have

$$c = -M^2 + \lambda (r^T g + g^T M g - 1) \tag{26}$$

which has the solutions $g \propto (I - \lambda C M)^{-1} \lambda C r$.

There are also various other potential extensions and applications of DCA. It would be possible to consider applying some of the variations and extensions used for PCA, such as application to correlation matrices rather than covariance matrices, to DCA patterns. One could investigate the weather and climate patterns associated with the first DCA pattern by regression of other weather and climate fields, such as mean sea level pressure, onto the time series T. In addition, one could compare the first DCA spatial pattern between observations and numerical model output as a way of evaluating how well the numerical model captures spatial extremes.

Finally we note that physical interpretation of both PCA and DCA patterns is sometimes difficult. Both are defined purely in terms of the observed statistics, and do not take any physics into account directly. If the statistical assumptions are incorrect then they may represent patterns that could not occur in nature.

Funding: The author was self-funded.

Conflicts of Interest: The author works as a consultant to RMS Ltd, a company that builds models of extreme weather events and their impacts. The results of this particular research do not currently form part of any RMS product, nor are there currently any plans to use them in any RMS product.

References

1. Smith, T.; Reynolds, R. Bias corrections for historical sea surface temperature based on marine air temperature. *J. Clim.* **2002**, *15*, 73–87.
2. Kurnik, B.; Kajfez-Bogataj, L.; Horion, S. An assessment of actual evapotranspiration and soil water deficit in agricultural regions in Europe. *Int. J. Climatol.* **2015**, *35*, 2451–2471.
3. Mestas-Nunez, A.M. Orthogonality properties of rotated empirical modes. *Int. J. Climatol.* **2000**, *20*, 1509–1516.
4. Lian, T.; Chen, D. An Evaluation of Rotated EOF Analysis and Its Application to Tropical Pacific SST Variability. *J. Clim.* **2012**, *25*, 5361–5373.
5. Chen, H.; Sun, J. Characterizing present and future drought changes over eastern China. *Int. J. Climatol.* **2017**, *37*, 138–156.

6. Fraedrich, K.; McBride, J.; Frank, W.; Wang, R. Extended EOF Analysis of Tropical Disturbances: TOGA COARE. *J. Atmos. Sci.* **1997**, *41*, 2363–2372.
7. Kim, J.; Oh, H.; Lim, Y.; Kang, H. Seasonal precipitation prediction via data-adaptive principal component regression. *Int. J. Climatol.* **2017**, *37*, 75–86.
8. Wilks, D. *Statistical Methods in the Atmospheric Sciences*; Academic Press: Cambridge, MA, USA, 1995.
9. von Storch, H.; Zwiers, F.W. *Statistical Analysis in Climate Research*; Cambridge University Press: Cambridge, UK, 1999.
10. Jolliffe, I. *Principal Component Analysis*; Springer: Berlin, Germany, 2002.
11. Hannachi, A.; Jolliffe, I.; Stephenson, D. Empirical othogonal functions and related techniques in atmospheric science: A review. *Int. J. Climatol.* **2007**, *27*, 1119–1152.
12. Carter, L.; Moss, S. Drought Stress Testing: Making Financial Institutions More Resilient to Environmental Risks. 2017. Available online: https://naturalcapital.finance/wp-content/uploads/2018/11/Drought-Stress-Testing-Tool-FULL-REPORT.pdf (accessed on 25 June 2019).
13. Harris, I.; Jones, P.; Osborn, T.; Lister, D. Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *Int. J. Climatol.* **2013**, *34*, 623–642.