 doi:10.20944/preprints202002.0073.v1

An alternative to PCA for estimating dominant patterns of climate variability, with application to US rainfall

Stephen Jewson

Risk Management Solutions Ltd

Keywords:

Principal Component Analysis, Directional Component Analysis, Empirical Orthogonal Functions, Extremes, US rainfall

Correspondence:

Stephen Jewson
stephen.jewson@gmail.com
Risk Management Solutions Ltd
Peninsular House
30 Monument Street
London EC3R 8NB
United Kingdom Mob: 07858 393370

 doi:10.20944/preprints202002.0073.v1

Abstract

Floods and droughts are driven, in part, by patterns of extreme rainfall. Heat waves are driven, in part, by patterns of extreme temperature. The standard work-horse for understanding patterns of climate variability is Principal Component Analysis (PCA) and its variants. But PCA does not optimize for spatial extremes, and so there is no particular reason why the first PCA 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 field'. We compare the first PCA and DCA patterns for US rainfall on a 6 month timescale, using the sum metric for the definition of DCA, and find that they are somewhat different. The definitions of PCA and DCA mean that the first PCA pattern has the larger explained variance of the two, while the first DCA pattern, when scaled appropriately, is both more likely and captures more rainfall. In combination these two patterns yield more insight into rainfall variability than either pattern on its own.

 doi:10.20944/preprints202002.0073.v1

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 (Smith and Reynolds, 2002) and identifying patterns of variability (Kurnik *et al.*, 2015). When being used to solve specific problems, however, limitations of PCA may become apparent, and this has led to the development of many extensions, each addressing a different issue. For instance, PCA patterns generally fill the 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 (2000) and Lian and Chen (2012), and used in Chen and Sun (2017). In another extension, known as extended EOF analysis, PCA has been used to understand developments of patterns in time (Fraedrich *et al.*, 1997), and in yet another has been adapted to better handle skewed data (Kim *et al.*, 2017). PCA and related methods have been discussed in text books such as Wilks (1995), von Storch and Zwiers (1999) and Jolliffe (2002) and in the review paper Hannachi *et al.* (2007).

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 (Carter and Moss, 2017) 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 does not necessarily maximise the total rainfall anomaly. This is simply a result of the mathematical definition of PCA. As a result the first PCA 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 domain. One could imagine that other 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 pattern as that which is the most likely to occur for a given level of total rainfall anomaly. This pattern is also, conversely, the pattern with the highest rainfall for a given probability density of occurrence. Appropriately scaled, the first DCA pattern is *both* more likely to occur *and* contains more rain than the first PCA pattern, and as such may be more relevant for understanding extremes such as floods or droughts. Similarly, if applied to temperature anomalies, the first DCA pattern may have more relevance than the first PCA pattern for understanding heat waves. The first DCA pattern is only the same as the first PCA pattern in the degenerate case in which the data is fully correlated and the first PCA pattern is uniform rainfall throughout the whole domain.

The definitions of PCA and DCA can be contrasted as follows. Patterns of variability of rainfall have various mathematical properties, including explained variance, length (based on considering patterns as vectors), total rainfall anomaly and probability density of occurrence. PCA is a mathematical method that considers two of these properties (explained variance and length) and finds the patterns with the greatest explained variance, among all patterns of the same length. The standard definition of PCA does not take account of the total rainfall anomaly or the probability density, although it can be reformulated in terms of probability density (see section 2.1 below). DCA considers a different pair of properties, and is derived by finding the pattern that has the greatest rainfall, among all patterns of the same probability density. The definition of DCA 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.

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 log-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 log-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 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 2x2 covariance matrix. The second (section S2) gives the general solution for PCA and DCA for any diagonal 2x2 covariance matrix. In sections S3 to S8 we prove the orthogonality of the first two DCA patterns, and provide some proofs of the main optimality properties of DCA.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

2 Principal Component Analysis

We now briefly review PCA as a basis for comparison with DCA. Consider a space-time dataset 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.* (2013)), for which the time dimension is 1363 (12 months, times 114 years, minus 5) 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

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

where $C = XX^T$ 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

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

where c is a scalar cost function, and λ is a Lagrange multiplier that multiplies $g^T g$, the variable that defines the contraint. 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.

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 the dataset X is distributed with a multivariate normal distribution, and then to try to find the unit vector that is most likely to occur, instead of that which explains the most variance. We can measure 'most likely to occur' using log-likelihood, and considered as a function of the unknown pattern g the log-likelihood for the multivariate normal is proportional to minus one times the Mahalanobis distance $M^2 = g^T C^{-1} g$, where C^{-1} is the inverse or pseudoinverse of the matrix C (the connections between PCA, the multivariate normal distribution, and the Mahalanobis distance are discussed in detail in Wilks (1995) and von Storch and Zwiers (1999)). The Lagrange function for this new problem has two terms: one for the log-likelihood $-M^2$, and one for the total rainfall anomaly constraint as before, and is given by:

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

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 to make sense. Minus one times the Mahalanobis distance $g^T C^{-1} g$ 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 (Wilks, 1995). In non-normal cases, PCA, by this derivation, finds the unit vector that maximises this Mahalanobis consistency, $-M^2$.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

2.2 2 dimensional example

PCA is illustrated in figure 1 for a simple case. The two axes represent rainfall anomalies in two locations. The ellipse represents a contour of constant probability density from the joint probability distribution of rainfall at these locations. 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 pattern is a scaled version of this principal axis vector (scaled to be a unit vector), and consists of negative rainfall anomalies at location one, and positive rainfall anomalies at location two, reflecting the negative correlation. In this example we will assume that the ellipse has been chosen so that the PCA arrow shown is the exact unit-scaled PCA pattern. The two diagonal lines represent contour lines of total rainfall anomaly, summed across the two locations, and the highest total rainfall amounts are in the top right hand corner of the figure. The total rainfall anomaly of the first PCA pattern is not particularly large since there is a cancellation of rainfall anomalies to some extent between the two locations.

3 Directional Component Analysis

Having reviewed PCA, we now describe DCA. Considering the same space-time data set X, again considered to be multivariate normal, and a new unknown pattern g, we derive DCA by solving a different maximisation problem, which is to look for the most likely pattern given a certain level of total rainfall anomaly. Alternatively, but equivalently, we could say we are looking for the pattern with the highest rainfall given a certain value of the probability density of occurrence. Essentially we are trying to find the pattern with the most rainfall, while factoring in the requirement that the pattern has a high probability density of occurring. We include the probability density because the pattern with the most rainfall, not factoring in probability density, is simply uniform rainfall everywhere, which is not an interesting result. 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 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 most rainfall, for a given level of Mahalanobis consistency.

A scaled version of the first DCA pattern is illustrated in figure 1 by the dotted arrow. The scaling has been chosen so that the dotted arrow hits the same contour of probability density as the first PCA pattern, and so occurs with the same probability density. The first DCA pattern points to a greater extent towards the region of highest total rainfall anomaly in the top right hand corner of the diagram, and even though it is shorter, it thus achieves a higher total rainfall anomaly amount than the first PCA pattern.

One can imagine scaling the lengths of the two patterns in figure 1 and we now use this to illustrate various properties of PCA and DCA. For instance, if we were to scale the length of the first PCA pattern so that it would hit the same total rainfall line as the first DCA pattern, it would extend outside the elliptical contour, and would hence occur with a lower probability density than the scaled DCA pattern shown. Conversely if we were to scale the DCA pattern to be slightly shorter, then it can be seen that it would still achieve higher rainfall than the first PCA pattern, but would be more likely. 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.

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 a normalised version of the vector $(1, 1, 1..., 1)$, a pattern of uniform rainfall. The analysis below also applies to any other value for the vector r, other than uniform rainfall, 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 and its inverse C^{-1}. The inclusion of a preferred direction distinguishes this method from EOF and rotated EOF methods.

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:

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

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

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.

3.1 Derivation of the second DCA pattern

We can derive a time-series corresponding to the first DCA pattern by projecting the original data X onto it, giving $g_1^T X$. An approximate reconstruction of X can then be created by combining the first DCA 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 pattern can be derived by repeating the entire DCA analysis given above on this second data-set X_2, using $C_2 = X_2 X_2^T$, and giving $g_2 = \frac{C_2 r}{|C_2 r|}$. This process can be continued to derive a series of patterns, and as with PCA the number of patterns is equal to the rank of X. This series of patterns will be mutually 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 patterns, of which PCA and DCA 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 patterns are empirical, and orthogonal, and so one could refer to them as empirical orthogonal functions (EOFs). However, the term EOF 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 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:

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

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

3.3 Properties of DCA

We now summarise some of the mathematical properties of the first DCA pattern, with the assumption that X represents rainfall anomalies as illustration. From the derivations of the first PCA and DCA 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 variance explained by 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 pattern equals the vector r, in which case the first DCA pattern will also equal r.

- Since DCA is designed to maximise total rainfall anomaly (for a given value of probability density or Mahalanobis distance) the total rainfall anomaly $g_1^T r$ for the first DCA pattern will be greater than or equal to that of the first PCA pattern. Once again it will only be equal in the degenerate case. One can say that this property is obvious, 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 4 and 5.

- If the first DCA pattern is scaled to have the same amount of total rainfall anomaly as the first PCA pattern, it will have a higher or equal log-likelihood (or Mahalanobis consistency), equal only in the degenerate case. Again one could say that this property is obvious from the definitions of PCA and DCA, but again it is proven in the supporting information, section 6.

- If the first DCA pattern is scaled to have the same log-likelihood, or Mahalanobis consistency, as the first PCA pattern (as in figure 1) it will have a greater or equal value for the total rainfall anomaly, equal only in the degenerate case. Again one could say this is obvious from the definitions, but is also proven carefully in the supporting information, section 7.

- In the non-degenerate case, the first DCA pattern can be scaled to the in-between case where it has *both* more rainfall *and* a higher log-likelihood (or Mahalanobis consistency) than the first PCA pattern. This is the most interesting property of DCA, and the property which justifies its use for identifying spatial extremes. It is proven in the supporting information, section 8.

We now illustrate these properties with an example.

4 Application of DCA to observed US rainfall

We now derive the first PCA and DCA patterns for an example dataset. We use the CRU data for US rainfall from 1901 to 2014 (Harris *et al.*, 2013) 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 overall pattern of highest rainfall anomalies irrespective of when they occur 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 patterns on the anomalies, and we assume the data is multivariate normal. The multivariate normal assumption is not essential, but allows us to relate the values of the Mahalanobis consistency to the log-likelihood / probability density, with which readers are likely to be more familiar.

4.1 Unit vector scaling

The first PCA and DCA 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. The explained variances are 43% and 35% respectively: as would be expected from the definition of PCA, the first PCA pattern has higher explained variance. The PCA pattern has a total rainfall anomaly of 16.6, and the DCA 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 pattern has more rain, but is less likely (since it has a higher Mahalanobis distance), which is not particularly remarkable: the interesting properties of the first DCA pattern only emerge after appropriate scaling.
The ratio of probability densities 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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

almost the same probability density. 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 origin. If the patterns were scaled to represent larger anomalies then the difference in probability densities would be larger: this is explored below in section 4.5.

4.2 Equal total rainfall anomaly scaling

The interesting properties of the first DCA pattern, as compared to the first PCA pattern, become apparent when we adjust the scaling of one or both of the patterns. One possibility is equal total rainfall anomaly scaling. Since the ratio of DCA rainfall to PCA rainfall is 2.3, if we multiply the DCA pattern by the inverse of 2.3, which is 0.43, then the two patterns will have the same rainfall. 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 pattern, and hence we have derived a pattern which has the same rainfall as the first PCA pattern, but is more likely.

4.3 Equal probability density scaling

Another possibility is equal probability density scaling, in which we scale the DCA pattern so that it has the same probability density as the PCA pattern. Since the ratio of the Mahalanobis distances for the unit vector patterns is 1.97, if we scale the DCA pattern by inverse of the square root of 1.97, which is 0.71, then the Mahalanobis distances of the PCA pattern and the scaled DCA pattern will be the same. The rainfall of the DCA 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 pattern. We have therefore derived a pattern which has the same probability density as the PCA pattern, but higher total rainfall anomaly.

4.4 Intermediate scaling

We can also scale the DCA pattern to an intermediate case in between equal total rainfall anomaly scaling and equal probability density scaling. For instance, if we apply a scaling of 0.57 to the DCA 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 pattern) and a Mahalanobis distance of 8.55×10^{-7} (0.65 times the Mahalanobis distance of the PCA pattern). We see that we have created a pattern which *both* contains more rainfall *and* is more likely (has a lower Mahalanobis distance) than the first PCA 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. The existence of such an intermediate scaling in all cases is proven in S8.

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 probability densities is very close to one, because the amplitudes of the patterns are very small. To create patterns with larger differences in probability density, 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 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 probability densities for the two patterns is then 1.42. In this case we have created a DCA pattern that has more rainfall than the scaled PCA pattern, and has a greater probability density by a clear margin.

4.6 Discussion of pattern structure

The PCA pattern in figure 2 shows large regions of positive and negative anomalies. Some of the same regions appear in the DCA pattern in figure 3, but the large-scale eastern dipole has disappeared, and the DCA pattern shows more uniform rainfall over a larger area. We can measure this within-pattern variability using the spatial standard deviation of each pattern. The PCA pattern has a spatial standard deviation of 1.3×10^{-4} while the DCA pattern has a smaller spatial standard deviation of 7.0×10^{-5}.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

Combined with the fact that the DCA pattern for equal level of probability density has 1.64 times the total rainfall anomaly, one might imagine that the DCA pattern is the more relevant indicator of possible driving patterns for flood or drought on the spatial scale of this domain. The PCA pattern is, however, perhaps a better indication of the overall patterns of variability in rainfall and how they correlate in space. Taken together, the two patterns give more insight into the variability in the rainfall data than either pattern taken on its own.

5 Discussion

We have described a method for pattern identification in correlated multivariate space-time datasets, that we call DCA. For multivariate normal data the method finds patterns with the highest log-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 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 log-likelihood for any given level of total rainfall anomaly. Applying the method to US rainfall we were able to derive a pattern of rainfall that, within a certain range of scalings, is both more likely and contains more rainfall than the first PCA pattern. The PCA pattern is dominated by regions of positive and negative rainfall, while the DCA pattern is more uniform.

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 patterns seem like they may form part of a promising solution to this problem. The first DCA 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 200yr 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 pattern could be scaled to give exactly that rainfall amount. It would then represent the highest likelihood pattern for that rainfall total and as such would be an appropriate pattern to represent possible rainfall at that return period
- The DCA 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.

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 correlation 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 censured 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 (Wilks, 1995). 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 even capture the pattern that maximises the rainfall at a given probability density of occurrence. 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:

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

where the matrix G contains the DCA 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 total rain 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 rain. 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 variance subject to both a linear and a normalisation constraint, using the Lagrange function:

$$c = g^T Cg + \lambda_1 g^T r - \lambda_2 g^T g \tag{23}$$

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 $g^T Mg$, where M is a matrix, we have:

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

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

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

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

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 $f(g) \approx r^T g + g^T Mg$ we have

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

which has the solutions $g \propto (I - \lambda CM)^{-1}\lambda Cr$.

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. Finally, 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.

6 Supporting Information

There are 8 sections of supporting information.

- S1 applies PCA and DCA to an example of a 2x2 covariance matrix
- S2 applies PCA and DCA to a general diagonal 2x2 matrix
- S3 contains a proof that the second DCA pattern is orthogonal to the first
- S4 contains a proof that the first DCA pattern contains more rain than the first PCA pattern, directly from the Lagrange cost functions
- S5 contains an alternative proof of the result proven in section S4, but via the method of expanding the DCA pattern in terms of PCA patterns

 doi:10.20944/preprints202002.0073.v1

- S6 contains a proof that, if PCA and DCA patterns are scaled to a given total rainfall anomaly, then the DCA pattern has a higher Mahalanobis consistency.
- S7 contains a proof of the converse: that if the PCA and DCA patterns are scaled to a given value of the Mahalanobis consistency, then the DCA pattern contains more rain
- S8 contains a proof that there is always a scaling of the first DCA pattern than means it has both more rain and a greater Mahalanobis consistency than the first PCA pattern

7 Acknowledgements

Many thanks to Farid Ait-Chaalal who performed the calculations, created the figures and reviewed the text. Also thanks to Stephen Cusack who wrote the original computer code, and to the various other people I have discussed this work with, including Enrica Bellone, Arno Hilberts, Jo Kaczmarska and Christos Mitas.

References

Carter, L and Moss, S (2017). Drought Stress Testing: Making Financial Institutions More Resilient to Environmental Risks [Cited 2019 June 25]. Available from: https://naturalcapital.finance/wp-content/uploads/2018/11/Drought-Stress-Testing-Tool-FULL-REPORT.pdf.

Chen, H and Sun, J (2017). Characterizing present and future drought changes over eastern China. *International Journal of Climatology*, 37:138–156.

Fraedrich, K, McBride, J, Frank, W, and Wang, R (1997). Extended EOF Analysis of Tropical Disturbances: TOGA COARE. *Journal of Atmospheric Sciences*, 41:2363–2372.

Hannachi, A, Jolliffe, I, and Stephenson, D (2007). Empirical othogonal functions and related techniques in atmospheric science: A review. *International Journal of Climatology*, 27:1119–1152.

Harris, I, Jones, P, Osborn, T, and Lister, D (2013). Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology*, 34:623–642.

Jolliffe, I (2002). *Principal Component Analysis*. Springer.

Kim, J, Oh, H, Lim, Y, and Kang, H (2017). Seasonal precipitation prediction via data-adaptive principal component regression. *International Journal of Climatology*, 37:75–86.

Kurnik, B, Kajfez-Bogataj, L, and Horion, S (2015). An assessment of actual evapotranspiration and soil water deficit in agricultural regions in Europe. *International Journal of Climatology*, 35:2451–2471.

Lian, T and Chen, D (2012). An Evaluation of Rotated EOF Analysis and Its Application to Tropical Pacific SST Variability. *Journal of Climate*, pages 5361–5373.

Mestas-Nunez, A. M (2000). Orthogonality properties of rotated empirical modes. *International Journal of Climatology*, 20:1509–1516.

Smith, T and Reynolds, R (2002). Bias corrections for historical sea surface temperature based on marine air temperature. *Journal of Climate*, 15:73–87.

von Storch, H and Zwiers, F. W (1999). *Statistical Analysis in Climate Research*. CUP.

Wilks, D (1995). *Statistical Methods in the Atmospheric Sciences*. Academic Press.

 doi:10.20944/preprints202002.0073.v1

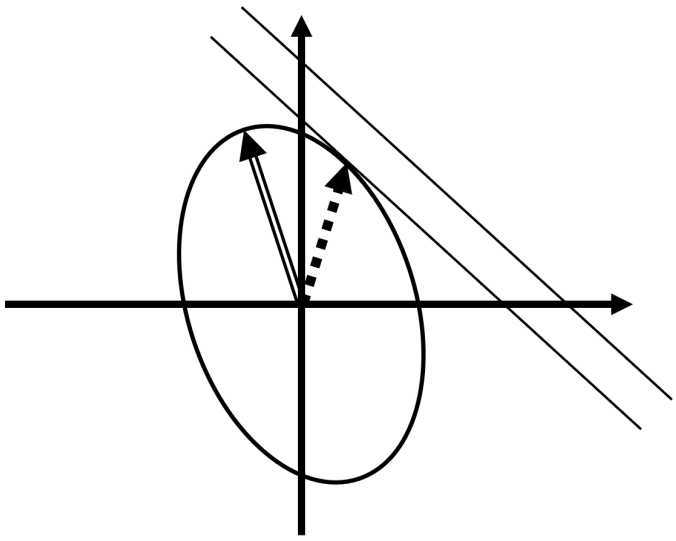

Figure 1: PCA and DCA patterns in a space with two dimensions. The axes are the two dimensions, which might be, for instance, rainfall at two locations. The diagonal lines then show lines of constant rainfall. Assuming that the two variables are multivariate normal distributed, the ellipse shows a contour of constant probability density or constant Mahalanobis distance. The double arrow shows the first PCA pattern, while the dotted arrow shows the first DCA pattern. In this case the two patterns are normalized to have the same likelihood. The PCA pattern is longer and captures more explained variance, while the DCA pattern is shorter but captures more total rainfall anomaly.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 February 2020 doi:10.20944/preprints202002.0073.v1

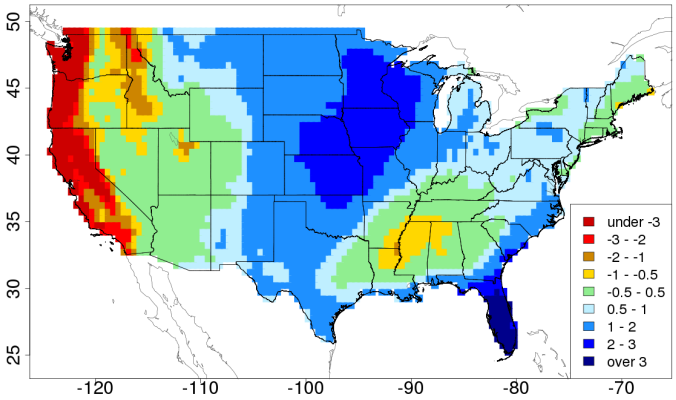

Figure 2: The first PCA mode for 6 month US rainfall anomalies.

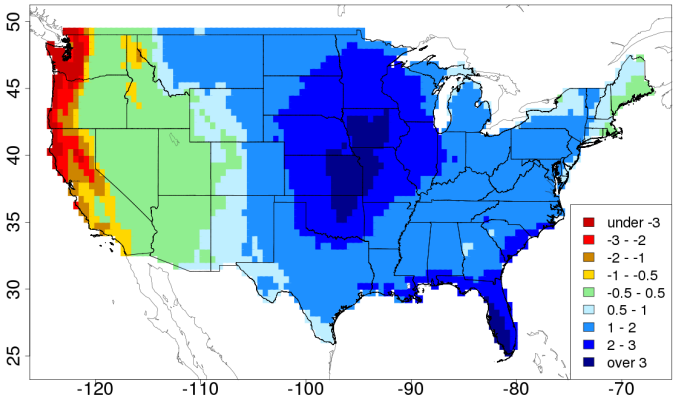

Figure 3: The first DCA mode for 6 month US rainfall anomalies.