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

Stephen Jewson

SUPPORTING INFORMATION

There are 8 sections of supporting information given below.

- Section 1 applies PCA and DCA to an example of a 2 × 2 covariance matrix
- Section 2 applies PCA and DCA to a general diagonal 2 × 2 matrix
- Section 3 contains a proof that the second DCA pattern is orthogonal to the first
- Section 4 contains a proof that the first DCA pattern contains more rain than the first PCA pattern, directly from the Lagrange cost functions
- Section 5 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
- Section 6 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.
- Section 7 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
- Section 8 contains a proof that there is always a scaling of the first DCA pattern that means it has both more rain and a greater Mahalanobis consistency than the first PCA pattern

1 Application of DCA to a 2x2 covariance matrix

We now give a worked example of PCA and DCA, applied to a 2x2 covariance matrix.
The direction vector we use is the uniform rainfall unit vector: $r = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$.
We will consider the covariance matrix: $C = \begin{pmatrix}8 & 6\\6 & 17\end{pmatrix}$ with inverse $C^{-1} = \frac{1}{100}\begin{pmatrix}17 & -6\\-6 & 8\end{pmatrix}$.
The first eigenvector of this matrix, which is the first PCA pattern, is $e_1 = \frac{1}{\sqrt{5}}\begin{pmatrix}1\\2\end{pmatrix}$.
This first PCA pattern has an angle to the x-axis of 63^o, and an angle to r of only 18^o. It is therefore already a reasonably rain-heavy pattern.
The second eigenvector, which is the second PCA pattern, is $e_1 = \frac{1}{\sqrt{5}}\begin{pmatrix}2\\-1\end{pmatrix}$
This second PCA pattern has an angle to the x-axis of -26^o, and an angle to the direction vector r of 72^o.
The eigenvalues (which are the explained variances) corresponding to these eigenvectors are 20 and 5.
The eigenvector matrix E is written as $E = \frac{1}{\sqrt{5}}\begin{pmatrix}1 & 2\\2 & -1\end{pmatrix}$ and the eigenvalue matrix is $\Lambda^2 = \begin{pmatrix}20 & 0\\0 & 5\end{pmatrix}$
Given the above it is easy to verify the relations $CE = E\Lambda^2$, $E^TE = I$ and $C = E\Lambda^2E^T$.
The first DCA vector g_1 is given by $g_1 = \frac{Cr}{|Cr|} = \frac{1}{\sqrt{725}}\begin{pmatrix}14\\23\end{pmatrix}$
This vector has an angle to the x-axis of 58^o, and an angle to r of 13^o.
We see that g_1 has a smaller angle to r than e_1.
g_1 has an explained variance of 19.9, slightly less than that of e_1, as would be expected since e_1 has the highest possible explained variance of any unit vector, by construction.
The rainfall amounts for the first PCA and DCA patterns e_1 and g_1 are given by $e_1^Tr = 0.9487$ and $g_1^Tr = 0.9717$, respectively. We see that g_1 has the greater rainfall amount, consistent with it having a smaller angle to r.
The Mahalanobis consistency values of e_1 and g_1 are -0.0500 and -0.0510 respectively. We see that e_1 has the higher likelihood, as would be expected since e_1 has the lowest possible Mahalanobis consistency value of any unit vector, again by construction.
The interesting properties of DCA versus PCA become apparent as we gradually scale the first DCA vector g_1 to make it shorter. As we reduce the length of g_1, the rainfall reduces and the Mahalanobis consistency, and hence likelihood, increases.
At around 0.99 of its original length the first DCA vector g_1 then becomes more likely than the first PCA vector e_1 since the Mahalanobis consistency of g_1 increases above that of e_1. The rainfall of g_1 at this point, is, however, still higher than that of e_1. We have therefore found a pattern that has both a higher likelihood and more rainfall than e_1 (in fact both $0.99g_1$ and $0.98g_1$ have this property). If we reduce the

length of g_1 further, then at around 0.97 of its original length the rainfall in the scaled g_1 drops below that of e_1, and the property is lost.

In summary: before scaling, the first DCA pattern has more rain, but a lower likelihood, than the first PCA pattern. If we scale it down by a large amount, it becomes less rainy, but has a higher likelihood. However, in between there is a region in which it is both has more rain and a higher likelihood.

For this particular example, the difference between the first PCA and DCA patterns is rather small, and the region in which the DCA pattern has both more rain and a higher likelihood is also small. However, in the example in the main text we show a case where the differences are large.

2 Application of DCA to a general diagonal 2x2 covariance matrix

As another simple example, both PCA and DCA can easily be applied to a general *diagonal* 2x2 covariance matrix.

Once again the direction we use for DCA is the uniform rainfall unit vector: $r = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}$.

We will consider the covariance matrix: $C = \begin{pmatrix}a^2 & 0\\0 & b^2\end{pmatrix}$ with inverse $C^{-1} = \begin{pmatrix}\frac{1}{a^2} & 0\\0 & \frac{1}{b^2}\end{pmatrix}$.

The eigenvectors of this matrix (the two PCA patterns) are $e_1 = \begin{pmatrix}1\\0\end{pmatrix}$ and $e_2 = \begin{pmatrix}0\\1\end{pmatrix}$.

Both eigenvectors have an angle to r of 45^o.

The eigenvalues (the explained variances) corresponding to these eigenvectors are a^2 and b^2.

The first DCA vector g_1 is given by $g_1 = \frac{Cr}{|Cr|} = \frac{1}{\sqrt{a^4+b^4}}\begin{pmatrix}a^2\\b^2\end{pmatrix}$.

g_1 has an explained variance of $\frac{a^6+b^6}{a^4+b^4}$, which is less than that of both eigenvectors unless $a = b$, as would be expected.

The rainfall amounts for the first PCA and DCA patterns e_1 and g_1 are given by $e_1^T r = \frac{1}{\sqrt{2}}$ and $g_1^T r = \frac{a^2+b^2}{\sqrt{2(a^4+b^4)}}$, respectively. We see that g_1 has the greater rainfall amount, unless $a = b$.

If we create scaled versions of e_1 and g_1, that have rainfall of 1, and call them f_1 and h_1, then:

$$\text{scaled first PCA pattern} = f_1 = \sqrt{2}e_1 \quad (1)$$

$$\text{scaled first DCA pattern} = h_1 = \frac{\sqrt{2(a^4+b^4)}}{a^2+b^2}g_1 \quad (2)$$

The Mahalanobis distance M^2 values for these scaled vectors are:

$$M^2(f_1) = \frac{2}{a^2} \quad (3)$$

$$M^2(h_1) = \frac{2}{a^2+b^2} \quad (4)$$

and we see that h_1 has the lower M^2 value (has a higher likelihood), unless $a = b$, as expected. In other words, if we scale the first PCA and DCA patterns to have the same rainfall, the DCA pattern is more likely.

If we create scaled versions of e_1 and g_1 that have $M^2 = 1$, and again call them f_1 and h_1, then:

$$\text{scaled first PCA pattern} = f_1 = ae_1 \quad (5)$$

$$\text{scaled first DCA pattern} = h_1 = \left(\frac{a^4+b^4}{a^2+b^2}\right)^{1/2} g_1 \quad (6)$$

The total rain values for these scaled vectors are then:

$$\text{rain}(f_1) = \frac{a}{\sqrt{2}} \quad (7)$$

$$\text{rain}(h_1) = \left(\frac{a^2+b^2}{2}\right)^{1/2} \quad (8)$$

and we see that h_1 contains more rain, unless $a = b$. In other words, if we scale the first PCA and DCA patterns so that they are equally likely, the DCA pattern has more rainfall.

3 Proof that the second DCA pattern is orthogonal to the first

Following from the definitions, we have the relations:

$$
\begin{aligned}
C &= XX^T && (9)\\
g_1 &= \frac{Cr}{|Cr|} = \frac{Cr}{\sqrt{r^T C^2 r}} && (10)\\
X_2 &= X - g_1 g_1^T X && (11)\\
&= X - \frac{Crr^T CX}{r^T C^2 r} && (12)\\
C_2 &= X_2 X_2^T && (13)\\
&= \left(X - \frac{Crr^T CX}{r^T C^2 r}\right)\left(X^T - \frac{X^T Crr^T C}{r^T C^2 r}\right) && (14)\\
&= \left(XX^T - \frac{XX^T Crr^T C}{r^T C^2 r} - \frac{Crr^T CXX^T}{r^T C^2 r} + \frac{Crr^T CXX^T Crr^T C}{(r^T C^2 r)^2}\right) && (15)\\
&= \left(C^T - \frac{C^2 rr^T C}{r^T C^2 r} - \frac{Crr^T C^2}{r^T C^2 r} + \frac{Crr^T C^3 rr^T C}{(r^T C^2 r)^2}\right) && (16)\\
& && (17)
\end{aligned}
$$

The dot product of the first and second patterns is then given by:

$$
\begin{aligned}
g_1^T g_2 &\propto r^T C C_2 r && (18)\\
&\propto r^T C \left(C^T - \frac{C^2 rr^T C}{r^T C^2 r} - \frac{Crr^T C^2}{r^T C^2 r} + \frac{Crr^T C^3 rr^T C}{(r^T C^2 r)^2}\right) r && (19)\\
&\propto r^T C^2 r - \frac{r^T C^3 rr^T Cr}{r^T C^2 r} - \frac{r^T C^2 rr^T C^2 r}{r^T C^2 r} + \frac{r^T C^2 rr^T C^3 rr^T Cr}{(r^T C^2 r)^2} && (20)
\end{aligned}
$$

But the first and thirds terms are equal, as are the second and fourth, and so we find:

$$g_1^T g_2 = 0 \tag{21}$$

Similar derivations can be used to show the orthogonality of the entire set of patterns $g_1, ..., g_n$.

4 Proof that the first DCA pattern contains more rain than the first PCA pattern

Because the first PCA pattern maximises the cost function:

$$c = -g^T C^{-1} g - \lambda_1 g^T g \tag{22}$$

we know that the value of c for the first PCA pattern must be greater than the value of c for the first DCA pattern, and so:

$$-e_1^T C^{-1} e_1 - \lambda_1 e_1^T e_1 \geq -g_1^t C^{-1} g_1 - \lambda_1 g_1^T g_1 \tag{23}$$

but the patterns are normalized, so that $\lambda_1 e_1^T e_1 = \lambda_1 g_1^T g_1 = \lambda_1$ and so the expression above simplifies to:

$$-e_1^T C^{-1} e_1 \geq -g_1^T C^{-1} g_1 \tag{24}$$

or

$$g_1^T C^{-1} g_1 - e_1^T C^{-1} e_1 \geq 0 \tag{25}$$

Similarly, because the first DCA pattern maximises the cost function:

$$c = -g^T C^{-1} g + \lambda_2 g^T r \tag{26}$$

we know that the value of this new definition of c for the first DCA pattern must be greater than the value of c for the first PCA pattern, and so:

$$-g_1^T C^{-1} g_1 + \lambda_2 g_1^T r \geq -e_1^T C^{-1} e_1 + \lambda_2 e_1^T r \tag{27}$$

or

$$g_1^T C^{-1} g_1 - e_1^T C^{-1} e_1 \leq \lambda_2 g_1^T r - \lambda_2 e_1^T r \tag{28}$$

Combining these two inequalities for $g_1^T C^{-1} g_1 - e_1^T C^{-1} e_1$ gives:

$$\lambda_2 g_1^T r - \lambda_2 e_1^T r \geq 0 \tag{29}$$

or

$$g_1^T r \geq e_1^T r \tag{30}$$

which says that the first DCA pattern has more rainfall than the first PCA pattern.

5 Alternative proof that the first DCA pattern contains more rain than the first PCA pattern

First we discuss how to expand the first DCA pattern using PCA patterns, and then we prove the main result.

5.1 Expanding the First DCA pattern using PCA patterns

Any pattern can be written as a weighted sum of PCA patterns, and any pattern can be written as a weighted sum of DCA patterns. For instance, we can expand the first DCA pattern in terms of PCA patterns as follows.

First, we write the direction vector r in terms of the n PCA patterns as:

$$r = \sum_{i=1}^{n} \alpha_i e_i \tag{31}$$

then multiplying by C gives:

$$Cr = \sum_{i=1}^{n} \alpha_i C e_i = \sum_{i=1}^{n} \alpha_i \mu_i e_i \tag{32}$$

and

$$|Cr|^2 = \left(\sum_{i=1}^{n} \alpha_i \mu_i e_i \right)^2 = \sum_{i=1}^{n} \alpha_i^2 \mu_i^2 \tag{33}$$

giving the first DCA pattern as:

$$g_1 = \frac{Cr}{|Cr|} = \frac{\sum_{i=1}^{n} \alpha_i \mu_i e_i}{\sqrt{\sum_{i=1}^{n} \alpha_i^2 \mu_i^2}} \tag{34}$$

We see that g_1 combines information about the direction vector (from the α_i) with information about the covariance matrix (from the μ_i).

5.2 Main result

The derivations of the first PCA and DCA patterns guarantee that the first DCA pattern has more, or the same, total rain as the first PCA pattern. However, it is of interest to prove this result bottom-up.

We first expand r in terms of eigenvectors of C, which we write as $e_1, ..., e_n$, with eigenvalues $\mu_1, ..., \mu_n$, giving $r = \sum_{i=1}^{n} \alpha_i e_i$.

Then the rainfall in the first PCA pattern e_1 is given by:

$$e_1^T r = e_1^T \sum_{i=1}^{n} \alpha_i e_i = \alpha_1 \tag{35}$$

We also have the following relations:

$$\begin{aligned}
Cr &= C\sum_{i=1}^{n}\alpha_i e_i = \sum_{i=1}^{n}\alpha_i\mu_i e_i && (36)\\
r^T Cr &= \left(\sum_{j=1}^{n}\alpha_j e_j\right)\left(\sum_{i=1}^{n}\alpha_i\mu_i e_i\right) = \sum_{i=1}^{N}\alpha_i^2\mu_i && (37)\\
|Cr| &= \sqrt{\left(\sum_{i=1}^{n}\alpha_i\mu_i e_i\right)\left(\sum_{i=1}^{n}\alpha_i\mu_i e_i\right)} = \left(\sum_{i=1}^{N}\alpha_i^2\mu_i^2\right)^{1/2} && (38)
\end{aligned}$$

and so the rainfall in the first DCA pattern g_1 is given by:

$$\begin{aligned}
g_1^T r &= \frac{r^T Cr}{|Cr|} && (39)\\
&= \frac{\sum_{i=1}^{N}\alpha_i^2\mu_i}{\left(\sum_{i=1}^{N}\alpha_i^2\mu_i^2\right)^{1/2}} && (40)
\end{aligned}$$

The ratio of these rainfall amounts (first DCA pattern to first PCA pattern) is:

$$\begin{aligned}
\frac{g_1^T r}{e_1^T r} &= \frac{\sum_{i=1}^{N}\alpha_i^2\mu_i}{\left(\sum_{i=1}^{N}\alpha_i^2\mu_i^2\right)^{1/2}}\frac{1}{\alpha_1} && (41)\\
&= \frac{\sum_{i=1}^{N}\frac{\alpha_i^2}{\alpha_1^2}\frac{\mu_i}{\mu_1}}{\left(\sum_{i=1}^{N}\frac{\alpha_i^2}{\alpha_1^2}\frac{\mu_i^2}{\mu_1^2}\right)^{1/2}} && (42)\\
&= \frac{1+\sum_{i=2}^{N}\frac{\alpha_i^2}{\alpha_1^2}\frac{\mu_i}{\mu_1}}{\left(1+\sum_{i=2}^{N}\frac{\alpha_i^2}{\alpha_1^2}\frac{\mu_i^2}{\mu_1^2}\right)^{1/2}} && (43)\\
&\geq \frac{1+\sum_{i=2}^{N}\frac{\alpha_i^2}{\alpha_1^2}\frac{\mu_i}{\mu_1}}{\left(1+\sum_{i=2}^{N}\frac{\alpha_i^2}{\alpha_1^2}\frac{\mu_i}{\mu_1}\right)^{1/2}} && (44)\\
&\geq 1 && (45)
\end{aligned}$$

and so we see that the first DCA pattern has more rainfall than the first PCA pattern, except in the case $r = e_1$, when the patterns and rainfall are equal.

6 Proof that the first DCA pattern has a higher log-likelihood than the first PCA pattern, when both are scaled to a given total rainfall

The derivations of the first PCA and DCA patterns guarantee that the first DCA pattern has a higher, or the same, log-likelihood, when both are scaled to have the same total rainfall. Again, it is of interest to prove this result bottom-up.

Consider a given total rain amount c. We can scale both the first PCA pattern and the first DCA pattern to give exactly that rain amount using:

$$\begin{aligned}
\text{scaled first PCA pattern} &= f_1 = \frac{ce_1}{e_1^T r} && (46)\\
\text{scaled first DCA pattern} &= h_1 = \frac{cg_1}{g_1^T r} && (47)
\end{aligned}$$

It can easily be verified that $f_1^T r = c$ and $h_1^T r = c$ (i.e. that both patterns contain a total rainfall of c).

The M^2 values of these two scaled patterns are then given by:

$$M^2(f_1) = f_1^T C^{-1} f_1 = \frac{c^2}{(e_1^T r)^2} e_1^T C^{-1} e = \frac{c^2}{\mu_1 (e_1^T r)^2} \tag{48}$$

$$M^2(h_1) = h_1^T C^{-1} h_1 = \frac{c^2}{(g_1^T r)^2} g_1^T C^{-1} g_1 = \frac{c^2}{\lambda g_1^T r} \tag{49}$$

The ratio of the M^2 values of these two scaled patterns (PCA to DCA) is then given by:

$$\frac{M^2(f_1)}{M^2(h_1)} = \frac{f_1^T C^{-1} f_1}{g_1^T C^{-1} g_1} \tag{50}$$

$$= \frac{g_1^T r}{\mu_1 \lambda (e_1^T r)^2} \tag{51}$$

If we now expand r using the eigenvectors of C, as in the previous section, and use the relations:

$$r^T e_i = \alpha_i \tag{52}$$

$$g_1^T r = \lambda \sum_{i=1}^{n} \alpha_i^2 \mu_i \tag{53}$$

then:

$$\frac{M^2(f_1)}{M^2(h_1)} = \frac{\lambda \sum_{i=1}^{n} \alpha_i^2 \mu_i}{\lambda \mu_1 (e_1^T r)^2} \tag{54}$$

$$= \frac{\sum_{i=1}^{n} \alpha_i^2 \mu_i}{\mu_1 \alpha_1^2} \tag{55}$$

$$= \sum_{i=1}^{n} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \frac{\mu_i}{\mu_1} \tag{56}$$

$$= 1 + \sum_{i=2}^{n} \left(\frac{\alpha_i}{\alpha_1}\right)^2 \frac{\mu_i}{\mu_1} \tag{57}$$

and so h_1 is has the lower M^2 value, and a higher log-likelihood, except in the case where $r = e_1$, when they are equally likely.

7 Proof that the first DCA pattern contains more rain than the first PCA pattern, when both are scaled to a given likelihood

The derivations of the first PCA and DCA patterns guarantee that the first DCA pattern has a higher, or the same, rain when both are scaled to have the same log-likelihood. Once more, it is of interest to prove this result bottom-up.

Consider a value for the Mahalanobis consistency of c. We can scale both the first PCA pattern and the first DCA pattern to give exactly that Mahalanobis consistency using:

$$\text{scaled first PCA pattern} = f_1 = \left(\frac{c}{\mu_1}\right)^{1/2} e \tag{58}$$

$$\text{scaled first DCA pattern} = h_1 = \left(\frac{c}{\lambda g_1^T r}\right)^{1/2} g_1 \tag{59}$$

The rainfall amounts of the scaled patterns are then given by:

$$f_1^T r = \left(\frac{c}{\mu_1}\right)^{1/2} e^T r \tag{60}$$

$$h_1^T r = \left(\frac{c}{\lambda g_1^T r}\right)^{1/2} g_1^T r = \left(\frac{c g_1^T r}{\lambda}\right)^{1/2} \tag{61}$$

The ratio of the rainfall amounts (DCA to PCA) is then:

$$
\begin{aligned}
\frac{h_1^T r}{f_1^T r} &= \left(\frac{g_1^T r}{\lambda \mu_1 e^T r^2}\right) && (62)\\
&= \left(\frac{\sum_{i=1}^n \alpha_i^2 \mu_i}{\mu_1 \alpha_1}\right) && (63)\\
&= \left(1 + \sum_{i=2}^n \left(\frac{\alpha_i}{\alpha_1}\right)^2 \frac{\mu_i}{\mu_1}\right) && (64)
\end{aligned}
$$

and so h_1 contains more rainfall, expect in the case where $r = e_1$, when they contain the same amount.

8 Proof that there's a scaling of the first DCA pattern that always is higher likelihood, and has more rain, than any scaling of the first PCA pattern

Consider a scaling of the first PCA pattern:

$$f = ce_1 \tag{65}$$

with rain R_f and M^2 values of M_f^2.
We know that the DCA1 pattern can be scaled so that it has more rain, and the same likelihood, compared to f. We write this pattern as:

$$h_a = ag_1 \tag{66}$$

with rain=$h_a^T r = a g_1^T r = R_a \geq R_f$, and $M^2 = h_a^T C^{-1} h_a = a^2 g_1 C^{-1} g_1 = M_f^2$.
We also know that the DCA1 pattern can be scaled so that it has the same rain, but a higher likelihood, compared to f. We write this pattern as:

$$h_b = bg_1 \tag{67}$$

with rain=$h_b^T r = b g_1^T r = R_b = R_f$, $M^2 = h_b^T C^{-1} h_b = b^2 g_1 C^{-1} g_1 \leq M_f^2$, and $a \geq b$.
We can then consider any pattern with a scaling in between these two cases, such as:

$$h_m = \frac{1}{2}(a+b)g_1 \tag{68}$$

The rain in h_m is:

$$h_m^T r = \frac{1}{2}(a+b)g_1^T r = \frac{1}{2}a g_1^T r + \frac{1}{2} b g_1^T r = \frac{1}{2}R_a + \frac{1}{2}R_b \geq R_f \tag{69}$$

and so we see that the rain in h_m is greater or equal to that in f.
The M^2 value of h_m is:

$$
\begin{aligned}
M_m^2 &= h_m^T C^{-1} h_m && (70)\\
&= \frac{1}{2}(a+b) g_1^T C^{-1} \frac{1}{2}(a+b) g_1 && (71)\\
&= \frac{1}{4}(a^2 + 2ab + b^2) g_1^T C^{-1} g_1 && (72)\\
&= \frac{1}{4}(M_a^2 + M_b^2 + 2M_a M_b) && (73)\\
&= \frac{1}{4}(M_a + M_b)^2 && (74)\\
&\leq \frac{1}{4}(2M_f)^2 = M_f^2 && (75)
\end{aligned}
$$

and so we see that h_m has a greater or equal likelihood to that in f.

9 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.