Presenting Climate Projection Ensembles as Mean and Reasonable Worst Case, with Application to EURO-CORDEX Precipitation

Stephen Jewson (stephen.jewson@gmail.com)  
London, UK  https://orcid.org/0000-0002-6011-6262

Gabriele Messori  
Uppsala Universitet

Giuliana Barbato  
CMCC: Centro Euro-Mediterraneo sui Cambiamenti Climatici

Paola Mercogliano  
CMCC: Centro Euro-Mediterraneo sui Cambiamenti Climatici

Jaroslav Mysiak  
CMCC: Centro Euro-Mediterraneo sui Cambiamenti Climatici

Maximiliano Sassi  
RMS Ltd

Research Article

Keywords: climate projection, ensemble mean, worst-case, DCA, EURO-CORDEX, precipitation

DOI: https://doi.org/10.21203/rs.3.rs-494689/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Presenting Climate Projection Ensembles as Mean and Reasonable Worst Case, with Application to EURO-CORDEX Precipitation

Stephen Jewson¹, Gabriele Messori²,³, Giuliana Barbato⁴, Paola Mercogliano⁵, Jaroslav Mysiak⁴, Maximiliano Sassi⁵

¹London, UK
²Department of Earth Sciences and Centre of Natural Hazards and Disaster Science (CNDS), Uppsala University, Uppsala, Sweden
³Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, Stockholm, Sweden
⁴Euro-Mediterranean Center on Climate Change (CMCC) Foundation, Via Augusto Imperatore, 16, 73100, Lecce, Italy
⁵Risk Management Solutions Ltd, EC3R 7AG, London, UK

Corresponding author: Stephen Jewson (stephen.jewson@gmail.com)

Abstract

Users of ensemble climate projections have choices with respect to how they interpret and apply the ensemble. A simplistic approach is to consider just the ensemble mean and ignore the individual ensemble members. A more thorough approach is to consider every ensemble member, although for complex impact models this may be unfeasible. Building on previous work in ensemble weather forecasting we explore an approach in-between these two extremes, in which the ensemble is represented by the mean and a reasonable worst case. The reasonable worst case is calculated using Directional Component Analysis (DCA), which is a simple statistical method that gives a robust estimate of worst-case for a given linear metric of impact, and which has various advantages relative to alternative definitions of worst-case. We present new mathematical results that clarify the interpretation of DCA and we illustrate DCA with an extensive set of synthetic examples. We then apply the mean and worst-case method based on DCA to EURO-CORDEX projections of future precipitation in Europe, with two different impact metrics. We conclude that the mean and worst-case method based on DCA is suitable for climate projection users who wish to explore the implications of the uncertainty around the ensemble mean without having to calculate the impacts of every ensemble member.

Keywords: climate projection, ensemble mean, worst-case, DCA, EURO-CORDEX, precipitation

Declarations

Funding

G. Messori was partly supported by the Swedish Research Council Vetenskapsrådet (grant no. 2016-03724).

Conflicts of Interest

MS works for RMS Ltd, a private company involved in weather and climate risk modelling.

Data Availability Statement

The EURO-CORDEX data used in this study is freely available. Details are given at https://euro-cordex.net.

Code availability

Not applicable.

Author's Contributions

Stephen Jewson designed the study and the algorithms, wrote and ran the analysis code, produced the graphics and wrote the text. Gabriele Messori reviewed the study in detail, Giuliana Barbato, Paola Mercogliano and Jaroslav Mysiak extracted the EURO-CORDEX data and Maximiliano Sassi wrote the code to read the EURO-CORDEX data. All the authors contributed to proof-reading.

Acknowledgements
The authors would like to thank Francesco Repola from CMCC for assisting with data extraction, and Casper Christophersen, Marie Scholer and Luisa Mazzotta from EIOPA for arranging for the EURO-CORDEX data to be made available.

1. Introduction

There is an extensive literature on how users of climate information might process ensemble climate projections, including methods for bias correction (Marzouk, 2016; Knutti, et al., 2017; Chen, et al., 2019), methods for identifying the forced response (Barnes, et al., 2019; Sippel, et al., 2019; Wills, et al., 2020), analysis of the potential benefits of spatial smoothing (Raisanen & Ylhaisi, 2010; Masson & Knutti, 2011), methods for assigning weights to the different models in an ensemble (Knutti, et al., 2017; Sanderson, et al., 2017; Abramowitz, et al., 2019), methods for understanding uncertainty (Deser, et al., 2010; Hawkins & Sutton, 2009; Frankcombe, et al., 2015; Thompson, et al., 2015), methods for reducing uncertainty using emerging constraints (Hall, et al., 2019) and methods for creating smaller and hence more manageable ensembles (Evans, et al., 2013; Herger, et al., 2018).

In this article, we will revisit the question of how to create smaller and more manageable ensembles. We imagine a hypothetical user of climate projections who currently considers only the ensemble mean from a climate projection ensemble such as the CMIP5 project (Taylor, et al., 2012), the CMIP6 project (Eyring, et al., 2016) or the EURO-CORDEX project (Jacob, et al., 2014; Jacob, et al., 2020). They use an impact model which already incorporates simulations of weather variability, and the reason for using climate models is to understand possible changes in statistics of weather variability under climate change. They look at impacts of the ensemble mean for multiple cases based on the four main RCPs (Moss, et al., 2010) or the five SSPs (Riahi, et al., 2017), for different points in time and for different adaptation strategies. Their impact models are costly to run, and this is already a lot of calculations. They would like to consider the impact of uncertainty around the ensemble mean, but running every ensemble member for every case is not feasible. Is there, nevertheless, a way that they could take into account uncertainty within each ensemble?

As an example, this problem often arises when including the effects of climate change in insurance industry risk models for severe weather risks. These models, known as catastrophe models (Grossi & Kunreuther, 2005; Mitchell-Wallace, et al., 2017), simulate present-day weather risks using their own ensembles consisting of many tens of thousands of members, each of one year in length (a methodology developed in the 1960s by Don Friedman (Friedman, 1972)). The models can be adjusted to incorporate aspects of climate variability and change from climate models (Sassi, et al., 2019; Jewson, et al., 2019), but the catastrophe models are computationally intensive, because of the large size of their ensembles and their high spatial resolution, and running them many times to understand the impacts of every climate model ensemble member may not be possible for all users.

To bridge the gap between the two extremes of analyzing only the ensemble mean and analyzing every ensemble member, one might consider using only a subset of the members of the full ensemble, either chosen randomly, or according to some kind of algorithm (Evans, et al., 2013; Herger, et al., 2018). The smallest subset, apart from using just the ensemble mean, would be to use the ensemble mean and one additional pattern, so that the full impact model only has to be run twice. This approach has recently been studied by Scher, et al. (2021) (henceforth SJM) in the context of medium range ensemble forecasts of temperature. To determine which single additional pattern to use, SJM considered a linear function that gives a measure of which ensemble members are most important in terms of impact. They then compared four different methods that could be used to find a single reasonable worst-case deviation from the ensemble mean, given the linear impact function. ‘Reasonable worst’ in the phrase reasonable worst-case means that the probability of this pattern should not be too low, so that the pattern is realistic, and that the pattern should have a large positive value of the linear impact function relative to the members of the ensemble. Two of the four methods that SJM compared were based on Directional Component Analysis (DCA, Jewson (2020)), a third method consisted of taking the worst single member from the ensemble and a fourth method consisted of averaging the worst 10% of the members of the ensemble. They compared these four methods and found the two DCA methods to be the most robust.

We will expand on the ideas in Jewson (2020) and SJM and explore further the idea of defining a reasonable worst-case pattern from an ensemble using DCA. DCA is a method for deriving extreme patterns from space-time datasets such that the patterns have both a large linear impact and a high likelihood (where we use likelihood synonymously with probability density). We extend the previous work in four ways. First, we provide additional illustrative examples of how DCA identifies a worst-case pattern. For these examples we imagine applying DCA to climate projections of precipitation and include examples in which the definition of linear impact is based on increased precipitation in all regions, and other examples in which it is based on increased precipitation in some regions, and decreased precipitation in others. This latter case is motivated by the situation in Europe where parts
of northern Europe are projected to experience an increase in precipitation because of climate change, while parts of southern Europe are projected to experience a decrease in precipitation (European Environment Agency, 2017). In this situation, a Europe-wide worst-case might consist of greater than the ensemble mean precipitation in northern Europe and less than the ensemble mean precipitation in southern Europe. Second, we present various additional mathematical properties of DCA that were not described in Jewson (2020) or SJM, that help to clarify the interpretation of DCA. Third, we discuss in detail how the linear impact function used in DCA can be considered as a linearisation of a non-linear impact function, and how the method can be generalized to quadratic and cubic approximations. Fourth, we apply DCA to high-resolution multi-model climate projections from EURO-CORDEX for two different linear impact functions.

In Sect. 2 we describe the data we will use and review the DCA methodology. In Sect. 3 we present new illustrations of DCA. In Sect. 4 we present new mathematical properties of DCA that help clarify the interpretation. We also discuss how to generalize DCA to quadratic and cubic impact functions. In Sect. 5 we illustrate the DCA methodology with results from the EURO-CORDEX climate model ensemble. In Sect. 6 we draw some conclusions. Appendices A to E include a number of related mathematical derivations.

2. Data and Methods

2.1 Synthetic Data

In section 3 we analyse synthetic ensemble data to illustrate the DCA method in various ways, and compare it with using just the worst member to define the worst case. The synthetic data was created by simulating from a bivariate normal distribution. The two dimensions are intended to represent changes in future annual mean precipitation values at two different but nearby locations and each ensemble member is intended to represent output from a different climate model. We simulate ensembles of both 10 and 1000 members, and we also simulate 500 repeats of the 10 member ensemble to understand the robustness of the results to the statistical sampling involved in creating the ensembles. 1000 is larger than the number of climate models in existence, but using 1000 member ensembles is helpful for illustrating how DCA works.

2.2 EURO-CORDEX data

The climate model data we use in Sect. 5 is annual mean precipitation data extracted from the EURO-CORDEX ensemble projections of future climate (Jacob, et al., 2014; Benestad, et al., 2017; Jacob, et al., 2020). We use data from 10 separate projections, simulated by 10 different combinations of global models and regional models: a list of the models used is given in Table 1. The model output is at 0.11-degree resolution (roughly 12km). Our analysis is based on the ensemble member by ensemble member difference between RCP4.5 annual mean precipitation in the period 2011-2040 and the baseline period 1981-2010, which we refer to as precipitation changes.

2.3 DCA Definition

We now give a definition and brief overview of DCA, based on Jewson (2020) and SJM. For the ensembles we are considering in this article, we define DCA as follows. The ensemble members, which are spatial patterns, are converted to anomalies from the ensemble mean and written as vectors, \( x \) with \( M \) components corresponding to the number of spatial points. Given an ensemble of \( N \) members we write the \( N \) vectors in a matrix \( X \). This matrix has dimensions \( M \times N \). The spatial covariance matrix of the vectors, which gives the covariance between every possible pair of spatial points, is given by \( C = XX^T/N \), which is an \( M \times M \) matrix. The total scalar impact \( a \) of a spatial pattern is defined as the sum of the impact of the ensemble mean \( a_0 \) and the deviation \( a' \) of the impact from \( a_0 \) due to the vector anomaly \( x \), giving \( a = a_0 + a'(x) \). We assume that the deviation \( a'(x) \) is a linear function of \( x \), and we will call it the linear impact. It can be written as a weighted sum of values in the anomaly pattern \( x \) as \( a' = x^T r \), where the vector \( r \) gives the weights at each location. The vector \( r \) can, equivalently, be interpreted as the direction of the gradient of the impact, pointing towards values of highest impact. A simple example of impact that can be expressed using \( a = a_0 + x^T r \) is the total precipitation change (change relative to the previous climate) across a precipitation field: if \( x \) is the spatial field of precipitation anomalies, \( a_0 \) is the total precipitation change of the ensemble mean and \( r \) is a vector of ones, then \( a \) is the total precipitation change due to the ensemble mean plus the anomaly pattern \( x \). A more complex example might involve using the components of the vector \( r \) to apply weighting as a function of spatial variation in population density.
Given the above definitions the DCA pattern \( g \) is then a vector with \( M \) components given by the proportional relationship:

\[
g \propto Cr
\]

(1)

The DCA vector \( g \) can be interpreted as a spatial pattern. Since \( C = XX^T/N \) this gives \( g \propto (XX^T/N)r = X(X^Tr)/N \), which can also be written as

\[
g \propto Xw
\]

(2)

where \( w = X^Tr \) is a vector that gives the linear impacts of the individual members of the ensemble. Based on these expressions we see that the DCA pattern \( g \) can be calculated either using the covariance matrix \( (g \propto Cr) \) or by calculating the linear-impact-weighted average of the ensemble members \( (g \propto Xw) \).

Equations (1) and (2) are proportional relationships and define the direction of the DCA vector \( g \) (i.e., the shape of the spatial pattern described by \( g \)), but not the magnitude of the vector (i.e., not the amplitude of the spatial pattern described by \( g \)). To make the definition of DCA unique, Jewson (2020) defines the first DCA pattern as a unit vector. The length of the unit vector DCA pattern can then be scaled to an appropriate value, depending on the application. Subsequent DCA patterns can also be derived, to create a set of orthogonal spatial patterns and a corresponding method for matrix factorisation, although in this study we will only consider the first pattern, which we will refer to simply as the DCA pattern or DCA vector.

### 2.3.1 Mathematical Properties

The mathematical properties of DCA depend on the statistical distribution of the ensemble. Jewson (2020) shows that if the ensemble is distributed as a MultiVariate Normal distribution (MVN), then DCA is the anomaly spatial pattern that maximises the probability density for a given value of the linear impact function (property 1). Part of the utility of DCA arises from the fact the direction of the DCA vector is then independent of the level of linear impact used in this definition, and depends only on the direction of the gradient of linear impact. Conversely, the DCA pattern is also the anomaly spatial pattern that maximises the linear impact for a given value of the probability density (property 2). These are the properties that justify the use of the DCA pattern as a pattern that is representative of extremes, where extreme is defined in terms of the linear impact function.

If the ensemble is not MVN, but is still elliptically distributed (for example, is distributed as a multivariate t distribution) then these properties may still hold: more precise details are given in Appendix A.

If the ensemble is not elliptically distributed it is useful to distinguish between what we will call *smoothable* and *non-smoothable* ensembles. Smoothable ensembles are those that have the property that weighted linear combinations of events in the ensemble are plausible alternative members of the ensemble, while non-smoothable ensembles are those which do not have this property. As examples of smoothable and unsmoothable ensembles: climate projections of temperature might be considered smoothable, while short-term weather forecasts of localised convective precipitation would probably not be, because averaging together convective events in different locations does not create a realistic field of convective events. For ensembles that are not elliptically distributed but are smoothable, DCA can be considered as a reasonable non-parametric method for creating extreme scenarios, based on Eq. (2) (which defines DCA as a weighted average of each ensemble member), but no longer possesses optimality properties related to impact and likelihood. For ensembles that are non-smoothable DCA could be calculated, but the resulting patterns are unlikely to be useful. The question of whether an ensemble is smoothable or not also affects the interpretation of the ensemble mean, and in general, if it is reasonable to consider using the ensemble mean as a plausible spatial pattern, then it is most likely reasonable to consider using DCA. This is because both DCA and calculating the ensemble mean involve weighted linear combinations of events in the ensemble (where for DCA the weights are given by Eq. (2) and for the ensemble mean the weights are equal).

### 2.3.2 Scaling

In several of the examples in this article we will set the length of the DCA vector to give a reasonable representation of the amplitude of the variability of the linear impact represented in the ensemble, as follows.

First, we determine the DCA direction \( g \) using Eq. (1) above and normalize \( g \) to be a unit vector. Second, we project all the members of the ensemble onto the unit vector \( g \) to create a series of projected values. Third, we
calculate the standard deviation of this series of projections and set the length of the DCA vector to a value of two standard deviations. This gives a pattern which is two standard deviations of ensemble variability away from the ensemble mean in the direction of the DCA pattern, and hence represents a reasonable level of extreme linear impact relative to the ensemble members. Within this method for scaling, the choice of two standard deviations is arbitrary, and could be replaced by any other number of standard deviations, or percentiles of the distribution of the projected values.

If the ensemble mean and the DCA pattern are used as inputs for a complex impact model, the output from the complex impact model when forced with the DCA pattern scaled in this way can be used as an approximation to two standard deviations of the distribution of impact as would be calculated using the brute-force and computationally intensive approach of running the complex impact model on each ensemble member. In this way DCA reduces the number of evaluations of the complex model from \( N \) to 2. Whether the approximation to the results of the complex model is a good one depends on whether the linear impact model is a good approximation to the complex impact model over the range of variability of the ensemble.

### 2.3.3 Generalisations to Non-linear Impact Functions

DCA uses a linear impact function. Real-world impact is likely, in most cases, to be non-linear, at least to some degree. The linear definition of impact used in DCA is the simplest possible formulation that allows us to explore the connections between the impact of a pattern and its likelihood, and to determine from that a straightforward methodology for defining reasonable worst-case patterns. Within the context of reducing an ensemble to just the mean and the reasonable worst case, using a linear impact function may be an appropriate level of complexity for many applications, given the other uncertainties inherent in the ensemble, the approach and the impact models, and given that the linear approximation only has to be a good approximation within the range of the ensemble. Alternatively, the impact function can be extended to include quadratic and cubic terms, as described in Sect. 4.2 below. If a cubic approximation is not sufficient then it may be necessary to run the full impact model with every member of the ensemble separately.

### 3 Illustrations and Examples

We now show a number of illustrations and examples of DCA applied to ensembles, extending on the illustrations and examples in Jewson (2020) and SJM. We will assume the ensemble is MVN.

#### 3.1 Illustrations

Figure 1 illustrates DCA in four different situations, in two dimensions. We will assume that the two dimensions represent changes in annual mean precipitation at two locations. In each case the ensemble is represented by a single elliptical contour of constant probability density, with higher densities inside the ellipse. The ensemble mean is shown by the red dot at the centre of the ellipse. Fig. 1a illustrates a situation in which the impact is simply the sum of the precipitation changes at the two locations, and so the impact vector \( r \) (illustrated by the black arrow) points towards the top right corner of the diagram where the total precipitation change is greatest. A line of constant impact, which is perpendicular to the impact vector \( r \), is shown by the straight diagonal line. The DCA pattern is shown by the blue dot at the point at which the straight-line is tangent to the ellipse. The DCA pattern is scaled to lie on the contour of probability density shown. The DCA pattern does not align with the impact vector because it also takes account of the covariance structure of the ensemble, i.e., the elliptical shape of the distribution. This diagram relates to the two properties of DCA given in section 2.3.1 above as follows. Considering property 1: we can see that of all points on the straight line, which have the same impact (in this case, the same total precipitation change), the point that touches the ellipse is the point with the highest probability density (since all the other points on the straight line lie outside the ellipse and are hence at lower probability densities). Considering property 2: we can see that of all the points on the ellipse, which are the same probability density, the point at which the straight line touches the ellipse is the point with highest impact (since it is the point which is furthest from the ensemble mean in the direction of the impact vector). Furthermore, we can see that if we move away from the DCA pattern in any direction then either the impact or the probability density will reduce: moving into the ellipse or along the ellipse reduces impact, while moving out of the ellipse reduces probability density.

Figure 1b illustrates the same ensemble (i.e., the same ellipse) as Fig. 1a but a situation in which the definition of impact puts twice as much weight on precipitation change at the second location (the vertical axis) as on the precipitation change at the first location (the horizontal axis). As a result, \( r \) now points more vertically than in...
Fig. 1a, and the line of constant impact is more horizontal. This changes the DCA vector, which now has more rain at location 2 than location 1. Fig. 1c illustrates a different ensemble, but for the same impact vector as in Fig. 1a. This also changes the DCA vector relative to Fig. 1a. Finally, Fig. 1d illustrates a more complex situation using a different ensemble in which the impact increases as precipitation changes increase at location 2 but impact decreases as precipitation changes increase at location 1. This would occur in the situation discussed in the introduction in which at location 2 the main concern with respect to climate change and future precipitation is related to increases in precipitation (e.g., such as concerns about increased flooding in the UK) while at location 1 the main concern is with respect to decreases in precipitation (e.g., such as concerns about increased drought in some parts of Southern Europe). In this case the vector $r$ points to the top left corner of the diagram, and the lines of constant impact are correspondingly different, and run from the lower left to the upper right. The DCA direction is also correspondingly very different, and selects a pattern that consists of increased precipitation at location 1 and decreased precipitation at location 2, relative to the ensemble mean, as the reasonable worst-case.

### 3.2 Examples

We now give some examples of DCA patterns calculated using synthetic ensemble data (Sect. 2.1). Our first set of examples are based on synthetic ensembles with 1000 members and are shown in Fig. 2. These examples correspond to the diagrams shown in Fig. 1. Figure 2a shows one ensemble with 1000 ensemble members (generated using a correlation of -0.4), the ensemble mean (red dot), the direction of the impact vector $r$ (black arrow), which puts equal weight on both precipitation variables (following Fig. 1a) and the DCA pattern scaled to two standard deviations of impact, as described in Sect. 2.3.2 (blue dot). The DCA pattern is calculated using Eq. (1). Figure 2b shows the same ensemble, with the same ensemble members, but now with a different impact vector $r$ which puts twice as much weight on the precipitation change on the vertical axis (following Fig. 1b). This shifts the DCA pattern. Figure 2c shows a different ensemble (now generated using a correlation of zero but unequal variances in the two directions), but with the same impact vector as Fig. 2a (following Fig. 1c). This leads to a different DCA pattern relative to Fig. 2a. Figure 2d shows a different ensemble again (generated using a correlation of -0.4), and an impact vector which puts equal weight on negative precipitation change on the horizontal axis and positive precipitation change on the vertical axis (following Fig. 1d). This leads to a DCA pattern which also has negative precipitation change on the horizontal axis and positive precipitation change on the vertical axis.

Overall, the illustrations in Sect. 3.1 above and these examples illustrate the way in which the DCA pattern is determined by both the impact function and the shape of the probability density function of the ensemble.

Our second set of examples, shown in Fig. 3, are similar, but are based on only 10 ensemble members, since our EURO-CORDEX ensemble only has 10 members. These ensembles were generated using a correlation of -0.4, and the elliptical structure of the ensemble is now much less clear from so few members.

Figure 4 explores the robustness of DCA for ensembles of size 10. It shows results from creating 500 ensembles, each of size 10, from the same underlying mean, variances and correlation as were used in Fig. 3. For each of the 500 ensembles, we identify the worst ensemble member (i.e., the one with the largest impact), and the DCA pattern, and these are illustrated in the figure. The mean of the 500 ensemble means is given by the red dots, and the individual ensemble members are not shown. The worst ensemble member shows large variability (Fig. 4a), while the DCA pattern shows much lower variability (Fig. 4b). This shows that the DCA pattern is a more robust estimate of a worst-case than the worst member of the ensemble, which is because the DCA pattern is based on more information: from Eq. (1) and Eq. (2) we see that it uses information from the entire ensemble, rather than from just one member. SJM studied the robustness of DCA in detail, comparing with the robustness of the worst ensemble member and the average of the worst 5 ensemble members, for a 50 member ensemble of ECMWF medium range temperature forecasts. They considered statistical robustness, i.e., robustness to the random sampling that occurs in the creation of the ensemble, robustness to small changes in the physical domain and robustness to the use of normal distributions. In all tests they found that DCA was more robust than the two alternative methods they considered.

### 4 Additional Mathematical Properties and Generalisations of DCA

In Sect. 4.1 below we present a number of additional mathematical properties of DCA for the MVN case, that we believe help with the interpretation. The idea that one pattern has multiple mathematical properties is a familiar one, since the ensemble mean itself has multiple mathematical properties, each of which helps with the
interpretation of the ensemble mean in different ways and in different situations (for example, the ensemble mean is the pattern that minimizes the in-sample mean squared error across the ensemble, and if the ensemble is symmetrically distributed then the ensemble mean is also the ensemble median). In Sect. 4.2. below we then discuss how DCA can be generalized to non-linear impact functions.

4.1 Additional Mathematical Properties of DCA

The optimality properties of DCA discussed in Sect. 2.3.1 above (property 1 and property 2) are of the form “the DCA vector maximises $X$ given $Y$”. In Sect. 4.1.1-4.1.2 below we will show that DCA also satisfies two statements of the simpler form “the DCA vector maximises $Z$”.

Equation (2) above shows that the DCA pattern can be considered as the linear-impact-weighted average of the members of the ensemble, which suggests that DCA may have properties related to the expectation over all possible patterns in the MVN. In Sect. 4.1.3-4.1.4 below we explore this idea and present two ways in which the DCA pattern can be considered as an expectation.

We will illustrate these additional properties in Fig. 5 with a simple two-dimensional example. Fig. 5a shows values of the probability density $p$ for a bivariate normal distribution with mean zero and a correlation of -0.4, representing the probability densities for an ensemble. Darker shading indicates higher probability density, and the black arrow (in this and subsequent panels in Fig. 5) shows the direction of the impact vector $r$, which is in the direction (1,1). The probability density has an elliptical shape, and the green arrow shows the principle axis of the ellipse, which is also the first eigenvector of the covariance matrix. The arrow is scaled to a length equal to the square root of the corresponding eigenvalue.

Figure 5a can also be interpreted as showing a quantity known as the Mahalanobis distance, defined by $m = \sqrt{x^T C^{-1} x}$. Mahalanobis distance is a multivariate generalisation of the univariate concept of the number of standard deviations from the mean. For the MVN the Mahalanobis distance is proportional to minus the log of the probability density, and so (since numeric values are not specified for the colour scale) Fig. 5a can also be interpreted as showing the Mahalanobis distance, with lighter shading showing a greater distance. Fig. 5b shows values of the linear impact function, which is $a' = 30(x_1 + x_2)$, where $x_1$ and $x_2$ are the distance in the horizontal and vertical directions. The diagonal lines show constant values of this linear impact function. Fig. 5c, and all subsequent panels of Fig. 5, show the DCA pattern, given by a black cross, calculated by scaling $r$ to be a unit vector and then using Eq. (1). Figure 5d shows the DCA pattern, given by a red circle, calculated using Eq. (2), without further scaling. The red circle in Fig. 5d represents the same spatial pattern as shown by the black cross in Fig. 5c (i.e., has the same direction from the origin) but with a slightly different length scaling.

4.1.1 The DCA Pattern Maximises the Ratio of Linear Impact to Mahalanobis Distance

The first new property we describe is that the DCA vector maximises the ratio of the linear impact $a'$ to the Mahalanobis distance. We write this ratio as

$$c(x) = \frac{a'(x)}{m(x)} \quad (3)$$

This property of DCA emphasizes that in the MVN case DCA finds patterns that have both a high linear impact (a large value of $a'$) and at the same time have a high probability density, which corresponds to a low value of $m$.

The proof that maximising Eq. (3) leads to the same expression for the DCA pattern as Eq. (1) above is given in Appendix B.

The ratio can also be written using the fact that $m^2 = \ln(p_0) - \ln(p)$, where $p$ is the probability density of $x$ and $p_0$ is the probability density for the ensemble mean, giving $c = a' \sqrt{\ln(p_0) - \ln(p)}$. This expression also illustrates that DCA captures the idea of finding directions with large values of both $a'$ and $p$.

Figure 5e shows values of the ratio $c = a'/m$, with darker shading showing larger values. We can see that there is a direction along which this ratio has the largest values which aligns with the direction of the DCA pattern given by the black cross.

4.1.2 The DCA Pattern Maximises the Product of Linear Impact and Weighted Probability Density
The second new property we describe is that any DCA pattern, of any length scaling, can be described as the pattern that maximises the product of the linear impact $a'$ and the probability density $p$ to some positive power, i.e., the function

$$f(x) = p(x)^n a'(x)$$

(4)

In Appendix C we show that maximising Eq. (4) leads to a pattern which is a scaling of the DCA pattern given by Eq. (1), and that different values of $n$ give all possible scalings.

This property of DCA emphasizes more clearly than any of the other properties that for the MVN DCA finds patterns that have both a high linear impact, and a high probability density. One could also attempt to find patterns that maximise this function for probability distributions other than the MVN, although the solutions may not be DCA patterns, according to the definition of DCA given in Sect. 2.3.1 above.

Figure 5f shows values of $f(x)$ for $n = 1$, with darker shading showing higher values. We see that there is indeed a maximum in the direction of the DCA vector.

4.1.3 The DCA Vector is Parallel to the Tail Conditional Expectation

The third new property we describe is that, for the MVN, the direction of the DCA vector is the same as the direction of the vector defined as the expectation over all possible vectors, conditional on exceeding a certain value for the linear impact. We call this the tail conditional expectation. In Fig. 5b, this expectation is the expectation over the region above and to the right of one of the diagonal lines. In this sense DCA can be considered as a generalisation of the ensemble mean, since the tail conditional expectation gives the ensemble mean if we consider exceeding impacts of minus infinity.

This property is shown in Appendix D below.

Given this property one could estimate DCA as the ensemble mean, conditional on exceeding a certain threshold value for the linear impact, although this is unlikely to be a good estimator for large values of the threshold as it will only use a small amount of information from the ensemble. Fig. 5g shows patterns estimated in this way, for three different levels of linear impact, given by the three diagonal lines, which have impacts of -30, 0 and 30. We see that the patterns do indeed lie along the DCA direction, given by the black cross.

This property helps explain why SJM found that using the mean of the 5 worst members of the ensemble as a definition of reasonable worst case gave similar results to DCA: the mean of the 5 worst members is in fact an alternative (although less precise) estimator for DCA.

4.1.4 The DCA Vector is Parallel to the Weighted Tail Conditional Expectation

The fourth new property we describe is that, for MVN data, the direction of the DCA vector is the same as the direction of the vector defined as the expectation over all spatial patterns, weighted by the linear impact at each point, and conditional on exceeding a certain value for the linear impact. We call this the weighted tail conditional expectation. In Fig. 5b, the weighted tail conditional expectation is the expectation over the region above and to the right of one of the diagonal lines, weighted at each point by the linear impact.

This is shown in Appendix E below.

A special case of this property is that the expectation over all patterns, weighted by the linear impact, is parallel to DCA. This is the expectation version of Eq. (2).

Fig. 5h shows patterns estimated in this way, for three different levels of linear impact, given by the three diagonal lines, which have impacts of -30, 0 and 30. The patterns calculated as weighted tail conditional expectations do indeed lie on the DCA direction given by the black cross.

4.1.5 Summary of Properties

Given all the above, we can summarize the key properties of DCA, as applied to MVN distributed data, as follows:

1) When scaled appropriately, the DCA pattern gives the unique spatial pattern that maximises the probability density, for a given level of linear impact, for any level of linear impact.
2) When scaled appropriately, the DCA pattern gives the unique spatial pattern that maximises the linear impact, for a given level of probability density, for any level of probability density.

3) All scalings of the DCA pattern maximise the ratio of linear impact to Mahalanobis distance. All scalings give the same value for this ratio.

4) For any given scaling, the DCA pattern is the unique spatial pattern that maximises the linear impact multiplied by a positive power of the probability density. An implication of this is that for any given scaling of the DCA pattern, there is no other pattern which has both a higher probability density and a higher linear impact, and that any change to a DCA pattern would lead to either the probability density or the impact reducing.

5) When scaled appropriately, the DCA pattern is the unique spatial pattern that equals the expectation of all possible patterns, conditional on exceeding a certain linear impact.

6) When scaled appropriately, the DCA pattern is the unique spatial pattern that equals the weighted expectation of all possible patterns, conditional on exceeding a certain linear impact, and weighted by the linear impact.

7) When scaled appropriately, the DCA pattern equals the linear impact-weighted expectation of all possible patterns.

4.2 Linearisation

In general, the impact of a change in climate state is likely to be a complex non-linear function. Consider, for instance, the relationship between changes in precipitation and flood damage, or changes in wind-speed and wind damage. We will write the nonlinear impact of the climate state \( x \) as \( a = w(x) \). We make the simplifying assumption that \( a \) is a scalar, such as the total flood or wind damage. Considering \( x \) as an anomaly from the ensemble mean as before, we can then expand the function \( w \) around the ensemble mean climate state as:

\[
a = w(x) = w(0) + x^T \nabla w + x^T \nabla^2 w x + \cdots
\]

If we write \( r = \nabla w \) and \( M = \nabla^2 w \), and neglect higher terms, this equation becomes a quadratic equation for impact:

\[
a = w(0) + x^T r + x^T M x
\]

This quadratic approximation only has to be a good approximation across the range of variability in the single ensemble being considered, and not across the entire range of variability of \( x \).

In this expression the impact for the climate state \( x \) is made up of three parts: the impact due to the ensemble mean \( w(0) \), the linear impact \( x^T r \), and the quadratic impact \( x^T M x \). The quadratic impact can then itself be split into two parts: the local impact relating to the diagonal elements of \( M \) and the non-local impact relating to the off-diagonal elements of \( M \), which contains quadratic cross-terms such as \( x_1 M_{12} x \) which model the impact as depending on the product of climate variables at different locations. Writing \( M = M^{diag} + M^{offdiag} \), Eq. (6) becomes

\[
a = w(0) + x^T r + x^T M^{diag} x + x^T M^{offdiag} x
\]

The derivation for DCA in Jewson (2020) is based on maximising likelihood for a given impact, and solves a Lagrangian maximisation problem for the linear impact \( x^T r \) with the Lagrangian given by:

\[
c = -x^T C^{-1} x + \lambda (w(0) + x^T r - a)
\]

Which has the solution \( x \propto Cr \). The direction of the solution does not depend on the value of the Lagrange multiplier \( \lambda \). For quadratic nonlinear impact we can extend the Lagrangian to

\[
c = -x^T C^{-1} x + \lambda (w(0) + x^T r + x^T M^T x - a)
\]

Which has the solution

\[
x \propto (I - \lambda CM)^{-1} Cr = A Cr
\]

Where we have written \( A = (I - \lambda CM)^{-1} \). This is still a simple expression that could be readily evaluated given an ensemble, the linear impact vector \( r \) and the quadratic impact matrix \( M \). The pattern \( x \) is no longer a weighted sum of the ensemble members, but is the weighted sum \( S = Xw/N \) transformed by the matrix \( A \):
weight on positive precipitation at all locations. We see positive precipitation anomalies distributed throughout two standard deviations everywhere, this is also not a realistic pattern of climate uncertainty and is also not an ensemble mean is negative there may be concern about decreasing precipitation. However, as with simply adding standard deviations using the method described in Sect. 2.3.2, based on a linear impact function that puts equal that where the ensemble mean is positive there may be concern about increasing precipitation, and where the standard deviations in the regions where the ensemble mean is positive, and the ensemble mean minus two standard deviations pattern shown here is only suitable for understanding possible precipitation, in which the differences between ensemble members are likely to exhibit complex spatial variability, it would likely not be reasonable. This is because it is highly unlikely that any future realisation of climate would consist of two standard deviations from the mean at every location at the same time. As a result, the ensemble mean plus two standard deviations pattern shown here is only suitable for understanding possible precipitation risks at each location separately, but not for understanding precipitation risk over the whole domain, and the pattern does not form a useful candidate for reasonable worst-case for the domain as a whole. Figure 6d shows an alternative way to combine the ensemble mean with the standard deviation. It shows the ensemble mean plus two standard deviations in the regions where the ensemble mean is positive, and the ensemble mean minus two standard deviations in the regions where the ensemble mean is negative. This is perhaps more relevant than adding two standard deviations everywhere, and reflects the idea (as discussed in the introduction and in Sect. 3.1 above) that where the ensemble mean is positive there may be concern about increasing precipitation, and where the ensemble mean is negative there may be concern about decreasing precipitation. However, as with simply adding two standard deviations everywhere, this is also not a realistic pattern of climate uncertainty and is also not a candidate for worst-case over the whole domain.

Figure 7 shows the results of the DCA analysis. Figure 7a shows the DCA pattern for this ensemble, scaled to two standard deviations using the method described in Sect. 2.3.2, based on a linear impact function that puts equal weight on positive precipitation at all locations. We see positive precipitation anomalies distributed throughout two standard deviations everywhere, and reflects the idea (as discussed in the introduction and in Sect. 3.1 above) that where the ensemble mean is positive there may be concern about increasing precipitation, and where the ensemble mean is negative there may be concern about decreasing precipitation. However, as with simply adding two standard deviations everywhere, this is also not a realistic pattern of climate uncertainty and is also not a candidate for worst-case over the whole domain.

\[ x \propto Acr = AXX^T r/N = AX(X^T r)/N = AXw/N = AS \]  

The pattern \( x \) is also no longer constant in direction as a function of the Lagrange multiplier. In Eq. (10), different values of the Lagrange multiplier \( \lambda \) lead to solutions with different values for the probability density, lengths of the vector and directions.

We can simplify Eq. (10) in two stages. First, we might imagine a situation in which the quadratic impact is just the weighted sum of the impacts at each location, with no cross-terms, i.e., \( M_{\text{offdiag}} = 0 \). This could be appropriate for wind damage as a function of wind speed, since wind damage is generally local. It would be less appropriate for flood damage as a function of precipitation, since hydrological processes may make the relationship between precipitation and flood damage non-local. Under this first simplification the matrix \( M \) is then diagonal, \( M = M_{\text{diag}} \), and represents the quadratic curvature of the impact at each location.

Second, we might imagine a situation in which there is no quadratic impact at all, in which case the matrix \( M \) is zero. In this case the solution reverts to the linear DCA solution.

A similar, although slightly more complicated, analytical solution also exists for a cubic approximation to the impact, based on the analytical method for solving vector quadratic equations.

4.2.1 Quadratic Impact Example

Fig. 5i shows patterns generated from Eq. (10), using a quadratic impact function. These are not DCA patterns, according to our definition, but are generalisations of DCA. These patterns correspond to the same linear impact vector as used in the previous panels, along with a quadratic impact matrix \( M \) given by the identity matrix, with values of \( \lambda \) of -2, 0 and 2. We see that the patterns for different values of \( \lambda \) are no longer parallel, as expected.
the domain, with larger values over the Alps and parts of South Eastern Europe. There are also some negative precipitation anomalies: this indicates that there are spatial anti-correlations of precipitation within the ensemble, such that the occurrence of a pattern with only positive anomalies is low probability. Unlike the local two standard deviation pattern shown in Fig. 6c, this pattern, if the statistical assumptions are correct, is a realistic pattern of variability that has a reasonable likelihood of occurrence given the ensemble. It has all the properties of DCA discussed in Sect. 4: for this level of total precipitation anomaly, there is no other pattern with higher likelihood; also, there is no pattern with higher total precipitation anomaly for the same likelihood; this pattern maximises both total precipitation anomaly divided by Mahalanobis distance and total precipitation anomaly multiplied by probability density to some positive power; the pattern is proportional to (i.e., the vector is parallel to) the expectation over all possible patterns weighted by their linear impact, and finally there is no possible adjustment to this pattern which could increase both the total precipitation anomaly and the likelihood. Figure 7c shows the sum of the ensemble mean with this DCA pattern. We see that precipitation amounts have increased over much of Northern Europe, relative to the ensemble mean, but to a much lesser extent than was seen in the ensemble mean plus two standard deviations pattern in Fig. 6c, which emphasizes further that Fig. 6c is not a realistic pattern. The ensemble mean plus the DCA pattern would be an appropriate forcing for complex flood risk models, to estimate a reasonable worst-case for flood risk in Europe.

Figure 7b shows the DCA pattern for the EURO-CORDEX ensemble based on a linear impact function that puts equal weight on positive precipitation change at locations where the ensemble mean is positive, and negative precipitation change at locations where the ensemble mean is negative. This relates to the cases illustrated in Fig. 1d and Fig. 2d, and to the mean plus/minus two standard deviations example shown in Fig. 6c. In this case the DCA pattern shows positive precipitation anomalies in much of Northern Europe, but negative anomalies in Spain and Northern Italy. Figure 7d shows the sum of the ensemble mean and this DCA pattern. In combination they again show increasing precipitation in Northern Europe and decreasing precipitation in Southern Europe, but to a lesser extent than the pattern given by adding and subtracting two standard deviations, as shown in Fig. 6d. Based on the impact function, and the variability and correlations in the climate model ensemble, the ensemble plus this DCA pattern is an indication of what the reasonable worst-case for changes in European precipitation might look like. This pattern would be an appropriate forcing for complex models for combined flood and drought risk, to estimate a reasonable worst-case outcome in Europe.

6 Conclusion

Ensemble climate projections present a large amount of information and ideally, given infinite time and resources, users of climate projections would consider the possible impact of every ensemble member in their analysis. However, this may not always be feasible since some impact models are simply too computationally intensive to run for every ensemble member, every RCP or SSP, every socio-economic scenario (such as migration scenarios), every adaptation scenario (such as flood defence building scenarios), and every time point required. In these situations, an alternative to considering the whole ensemble is to use a smaller ensemble, and an extreme case of using a smaller ensemble is to use the ensemble mean and just one other pattern, carefully selected to be representative of possible extremes of impact. We refer to this as the mean and reasonable worst-case approach (Scher et al. (2021)). One way to create a single worst-case pattern is by adding (or subtracting) two standard deviations locally. However, the spatial pattern that is created in this way is not, in most situations, a reasonable pattern of uncertainty since it assumes unrealistically high correlations of the uncertainty between locations. Another alternative would be to consider the worst member of the ensemble. However, the worst member is highly affected by the randomness of the ensemble generation process, and hence is not robust. In a multi-model ensemble, in which some models may be less realistic than others, there is also a risk that the model that determines the worst member may be one of the less realistic models. Following Scher et al., (2021), we have investigated a third option, which is to use the statistical method of DCA (Jewson (2020)).

We have extended the work of Jewson (2020) and Scher et al., (2021) in a number of ways. We have given additional illustrations and simulated examples of DCA, including a case where the highest linear impacts are defined by increasing precipitation in some locations and decreasing precipitation in others. We have presented several new properties of DCA for multivariate normally distributed data, including (a) optimality properties that avoid the use of Lagrangian calculus and show that DCA maximises certain simple functions related to probability and linear impact and (b) properties that show that DCA can be interpreted as an expectation in several different ways, including as a generalisation of the ensemble mean from the expectation over all patterns to the expectation over all patterns with linear impact over a threshold. These new properties give additional insight and help with
interpretation. We have also discussed how the linear impact function used in DCA can be derived as an
approximation to a more general non-linear impact function, and shown that a higher order quadratic
approximation is possible, and is scarcely more complex to apply than DCA. A cubic approximation is also
possible. Finally, we have applied DCA to a high-resolution EURO-CORDEX climate projection ensemble, and
shown that it gives a reasonable worst-case that is materially less severe than adding or subtracting two standard
deviations from the mean locally.

The properties of DCA can be summarized as follows. If we assume that the ensemble is multivariate normally
distributed then DCA factors in both the probability density of different patterns, and their linear impacts, to derive
a pattern that has both high probability density and high linear impact, relative to other patterns. As such it is a
good candidate for a pattern that could be used as the definition of reasonable worst-case. The DCA pattern is
realistic, robust and possesses a number of optimal mathematical properties. Some of the mathematical properties
also hold if the ensemble is not multivariate normal distributed, but has a distribution which is one of the
commonly used members of the wider class of distributions known as elliptical distributions, which includes, for
example, the multivariate t distribution. If the ensemble is not elliptically distributed but the variability in the
ensemble is such that averaging together patterns gives realistic alternative patterns, which we call a smoothable
ensemble, then DCA is realistic and robust but loses the optimal mathematical properties. If the variability in the
ensemble is such that averaging together patterns does not give realistic alternative patterns, i.e., the ensemble is
not smoothable, then neither DCA, nor the ensemble mean itself, are likely to be very useful, since they are both
created by averaging patterns. Unsmoothable ensembles may arise, for instance, when considering an ensemble
of daily precipitation values, but are unusual when considering climate time averages.

Using the ensemble mean and one additional pattern does not give a full picture of the uncertainty around the
ensemble mean: that can only be achieved by a deeper analysis, involving all the ensemble members. However,
if there are practical reasons due to resource limitations for why it is only possible to consider the ensemble mean
and one additional pattern, DCA is a strong candidate for the most useful pattern to look at.

Appendix A: Mathematical Properties of DCA for Elliptical Distributions

Elliptical distributions are those multivariate distributions for which the probability density for anomalies $x$ from
the mean can be written as

$$ p(x) = k f(x^T C^{-1} x) $$  \hspace{1cm} (A1)

where $k$ is a normalising constant. Examples are the multivariate normal distribution and the multivariate t
distribution. If the function $f$ is strictly decreasing (as it is for the multivariate normal and multivariate t) then
maximising $p(x)$ is equivalent to minimising $x^T C^{-1} x$. The derivations of property 1 and property 2 (given in
Jewson (2020)) are therefore still valid. Alternatively, we can rederive DCA from Eq. (A1) as follows.

In Jewson (2020) DCA is derived using the Lagrange function

$$ c = -x^T C^{-1} x + 2\lambda (x^T r - a') $$  \hspace{1cm} (A2)

Using Eq. (A1) the Lagrange function becomes:

$$ c = -k f(x^T C^{-1} x) + 2\lambda (x^T r - a') $$  \hspace{1cm} (A3)

which has the maximum if

$$ k f'(x^T C^{-1} x) C^{-1} x = \lambda r $$  \hspace{1cm} (A4)

If $f$ is strictly decreasing then $f'$ is never zero, and this equation gives the same solution as Eq. (1) ($x \propto Cr$).

To cover all elliptical distributions, including those that are not strictly decreasing, property 1 has to be amended
to become “DCA is the spatial pattern that minimises the Mahalanobis distance for a given value of the linear
impact function”.

Appendix B: The DCA Pattern Maximises the Ratio of Linear Impact to Mahalanobis Ratio

We will now show that the DCA pattern provides a solution to the question: given an ensemble with covariance
matrix $C$, what is the spatial pattern $x$ that maximises the ratio of the linear impact $a' = x^T r$ to the Mahalanobis
distance from the ensemble mean \( m(x) = \sqrt{x^TC^{-1}x} \). The Mahalanobis distance represents the distance between the pattern \( x \) and the origin, measured using a metric that takes into account the covariance structure: it can be loosely described as being proportional to the number of equally-spaced contour lines of probability density between the origin and the pattern \( x \). In one dimension it is the number of standard deviations from the mean. This derivation applies to any probability distribution, although it is most meaningful for elliptically distributed ensembles.

We will consider the function

\[
c(x) = \frac{a'(x)}{m(x)} = \frac{x^Tr}{\sqrt{x^TC^{-1}x}} \quad (B1)
\]

The length of the vector \( x \) does not affect the ratio \( c \) because it cancels between the numerator and the denominator.

Differentiating Eq. (B1) by the vector \( x \) gives

\[
\frac{dc}{dx} = \left( m \frac{da'}{dx} - a' \frac{dm}{dx} \right) / m^2 = \left( mr - a'C^{-1}x \right) / m^2 \quad (B2)
\]

Setting this equation to zero gives

\[ m^2r = a'C^{-1}x \quad (B3) \]

And hence

\[ m^2Cr = a'x \quad (B4) \]

This is a vector equality, and the two sides of this equation must be parallel, from which we see that \( x \) must be parallel to \( Cr \), which means it is parallel to the DCA pattern given by Eq. (1). Since the length of the vectors on both sides of Eq. (B4) are proportional to the length of \( x \) squared, this equation says nothing about the length of \( x \): any length of \( x \) is a solution. This is confirmed by the example shown in Fig. 5e.

Appendix C: The DCA Pattern Maximises the Product of Linear Impact and Weighted Probability Density

We now show that, for MVN ensembles, DCA provides a solution to the question: what pattern maximises the product of the linear impact and the weighted probability density? This property emphasizes directly that DCA finds patterns that both have a high linear impact and have a high probability density. We consider the function:

\[
f(x) = p^n(x)a'(x) \quad (C1)
\]

where \( n \) is positive. Changing the value of \( n \) varies the weight on the probability density, relative to the impact.

Differentiating with respect to \( x \) gives

\[
\frac{df}{dx} = \frac{d(p^n a')}{dx} = p^n \frac{da'}{dx} + na' p^{n-1} \frac{dp}{dx} = p^n r - nx^Trp^nC^{-1}x = p^n(r - nx^TrC^{-1}x) \quad (C2)
\]

Setting this equation equal to zero gives

\[ nx^Tr = Cr \quad (C3) \]

This is a vector equality, and the two sides of this equation must be parallel, from which we see that \( x \) must be parallel to \( Cr \), which means it is parallel to the DCA pattern given by Eq. (1). The lengths of the vectors on both sides must also be equal which gives

\[ x = \frac{cr}{\sqrt{nr^Tcr}} \quad (C4) \]

This is the unique solution to Eq. (C1) and is a scaled version of DCA. Large values of \( n \) make the solution shorter.

Every scaling of DCA is a solution of Eq. (C1) for a different value of \( n \).

Appendix D: The DCA Vector is Parallel to the Tail Conditional Expectation
To derive this property, first we need to show that the DCA pattern is parallel to the impact conditional expectation, i.e., the vector derived as the expectation over all possible vectors, conditional on a certain value for the linear impact. In Fig. 5b, this expectation is the expectation along one of the diagonal lines. In greater than two dimensions, lines of constant linear impact become planes and hyperplanes, and this expectation is then the expectation over one of these planes or hyperplanes.

This initial property can be shown as follows. Property 1 from Sect. 2.3.1 above states that DCA maximises probability density for a given level of linear impact. For an MVN in two or more dimensions, the probability density for a given level of linear impact is also an MVN, in a space with one fewer dimension. For instance, in two dimensions the subspace defined by a given level of linear impact is a line (see the lines in Fig. 5b) and the probability density of points on this line is a univariate normal. In three dimensions the subspace defined by a given level of linear impact is a plane, and the probability density of patterns on this plane is a bivariate normal.

To demonstrate this, consider a two-dimensional case, with dimensions \( x \) and \( y \), and consider a line of constant impact \( x + y = \text{const} \). We can parameterize distance along this line using the value of \( x \) and so the probability density, conditional on being on this line, can be written as \( p(x|y) \). Using Bayes' theorem we can factorize this into

\[
p(x|y) = \frac{p(x+y|y)p(x)}{p(x+y)}  \tag{D1}
\]

The terms on the RHS are all normal distributions and combine to form a normal distribution. As a result, the LHS must also be a normal distribution.

This demonstration can be extended to other linear impact functions, and to multivariate distributions. It does not work analogously for the multivariate t distribution, since the distributions on the RHS of Eq. (D1) would not combine to be a multivariate t distribution.

The point with maximum probability density of an MVN is also the mean of that MVN (just as the mode and mean coincide for a univariate normal distribution) and so the DCA pattern is the mean of the reduced dimension MVN.

We write this property for a pattern \( x \), for any value of \( a' \), as the proportional relationship:

\[
E(x|x^T r = a') \propto Cr  \tag{D2}
\]

We will integrate this expression over a range of levels of linear impact \( a' \), and since the integral of DCA patterns is a DCA pattern, we find that DCA is also the expectation conditional on a range of linear impacts. Integrating Eq. (D2) over a range of linear impacts gives

\[
\int_u^v E(x|x^T r = a') \, da' \propto \int_u^v Cr \, da'  \tag{D3}
\]

The LHS is equal to \( E(x|x^T r \in (u, v)) \). Since \( Cr \) does not depend on \( a' \) the RHS simplifies to \( Cr \), giving

\[
E(x|x^T r \in (u, v)) \propto Cr  \tag{D4}
\]

For instance, we could integrate over the region between two of the diagonal lines in Fig. 5b.

Setting \( v = \infty \) gives the special case

\[
E(x|x^T r \in (u, \infty)) \propto Cr  \tag{D5}
\]

in which the range extends to infinity on one side, which we refer to as 'tail conditioning', showing that DCA is the tail conditional expectation. If the range extends to infinity on both sides, then the expectation becomes the ensemble mean, which corresponds to the DCA pattern with a length of zero. In this sense the DCA pattern is a generalisation of the ensemble mean.

**Appendix E: The DCA Vector is Parallel to the Weighted Tail Conditional Expectation**

This property also follows from Eq. (D2). Since DCA is the expectation of the spatial pattern, conditional on a certain value for the linear impact, it is also the expectation of the spatial pattern weighted by the linear impact, conditional on a certain value for the linear impact. Integrating over a range of impacts then shows that DCA is
the expectation of the spatial pattern, weighted by the linear impact at each point, conditional on exceeding a
certain value for the linear impact.

Starting with Eq. (D2), multiplying both sides by any function of \(a'\), \(h(a')\) and integrating over a range of values
of \(a'\) gives:

\[
\int_u h(a') E(x|x^T r = a') \, da' \propto \int_u h(a') Cr \, da' \tag{E1}
\]

Considering the LHS first: writing the expectation as an integral:

\[
LHS = \int_{x \in V_1} h(a') xp(x) \, dV_1 \, da' = \int_{x \in V_1} h(a') xp(x) \, dV_1 \, da' \tag{E2}
\]

where \(V_1\) is the volume defined by the values of \(x\) that satisfy \(x^T r = a'\), \(dV_1\) is a volume element in this volume,
and \(h(a')\) can be taken inside the \(dV_1\) integral because it is a constant within this volume.

The two integrals can then be combined into one, giving:

\[
LHS = \int_{x \in V_2} h(a')xp(x) \, dV_2 = E(h(a')x|x \in V_2) \tag{E3}
\]

where \(V_2\) is the volume defined by the values of \(x\) that satisfy \(x^T r \in (u, v)\), \(dV_2\) is a volume element in this
volume, and \(dV_2 = dV_1 \, da'\). Considering the RHS of Eq. (E1), and integrating in the same way

\[
RHS = \int_u h(a')Cr \, da' = Cr \int_u h(a') \, da' \propto Cr \tag{E4}
\]

Combining Eq. (E3) and Eq. (E4) gives:

\[
E(h(a')x|x \in V_2) \propto Cr \tag{E5}
\]

Setting \(v = \infty\) gives the special case

\[
E(h(a')x|x \in V_2) \propto Cr \tag{E6}
\]

where \(V_2\) is the volume defined by the values of \(x\) that satisfy \(x^T r \in (u, \infty)\). This is what we set out to prove: the
weighted tail conditional expectation of over all possible patterns is parallel to DCA.

Setting both \(u = -\infty\) and \(v = \infty\) gives the special case

\[
E(h(a')x) \propto Cr \tag{E7}
\]

which says that the expectation over all patterns, weighted by the linear impact, is parallel to DCA. This is the
expectation version of Eq. (2).

**Bibliography**

Abramowitz, G. et al., 2019. Model dependence in multi-model climate. *Earth Syst. Dynam.*, Volume 10, pp.
91-105.

Barnes, E. et al., 2019. Viewing Forced Climate Patterns Through an AI Lens. *GRL*, 46(22), pp. 13389-13398.

Benestad, R. et al., 2017. *Guidance for EURO-CORDEX*. [Online]
Available at: https://www.euro-cordex.net/imperia/md/content/csc/cordex/euro-cordex-guidelines-version1.0-
2017.08.pdf [Accessed 9 1 2021].

Chen, J. et al., 2019. Bias correcting climate model multi-member ensembles to assess climate change impacts
on hydrology. *Climatic Change*, 153(3), pp. 361-377.

Deser, C., Phillips, A., Bourdette, V. & Teng, H., 2010. Uncertainty in climate change projections: the role of
internal variability. *Climate Dynamics*, Volume 38, p. 527–546.

European Environment Agency, 2017. *Indicator Assessment: Mean Precipitation*. [Online]
Available at: https://www.eea.europa.eu/data-and-maps/indicators/european-precipitation-2/assessment
[Accessed 15 4 2020].
Evans, J., Ji, F., Abramowitz, G. & Ekstrom, M., 2013. Optimally choosing small ensemble members to produce robust climate simulations. *Environ. Res. Lett.*, Volume 8.

Eyring, V. et al., 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.*, Volume 9, pp. 1937-1958.

Frankcombe, L., England, M., Mann, M. & Steinman, B., 2015. Separating Internal Variability from the Externally Forced Climate Response. *J. Clim.*, 28(20), p. 8184–8202.

Friedman, D., 1972. Insurance and the Natural Hazards. *ASTIN*, Volume 7, pp. 4-58.

Grossi, P. & Kunreuther, H., 2005. *Catastrophe Modelling: A new approach to managing risk*. NY: Springer.

Hall, A., Cox, P., Huntingford, C. & Klein, S., 2019. Progressing emergent constraints on future climate change. *Nature climate change*, Volume 9, pp. 269-278.

Hawkins, E. & Sutton, R., 2009. The Potential to Narrow Uncertainty in Regional Climate Predictions. *BAMS*, 90(8), p. 1095–1108.

Hawkins, E. & Sutton, R., 2012. Time of emergence of climate signals. *GRL*, 39(1), pp. 1-6.

Herger, N. et al., 2018. Selecting a climate model subset to optimise key ensemble properties. *Earth Syst. Dynam.*, Volume 9, pp. 135-151.

Jacob, D., Petersen, J. & authors, 2014. EURO-CORDEX: new high-resolution climate change projections for European impact research. *Reg Environ Change*, Volume 14, pp. 563-578.

Jacob, D., Teichmann, C. & authors, 2020. Regional climate downscaling over Europe: perspectives from the EURO-CORDEX community. *Regional Environmental Change*, Volume 20.

Jewson, S., 2020. An Alternative to PCA for Estimating Dominant Patterns of Climate Variability and Extremes, with Application to U.S. and China Seasonal Rainfall. *Atmosphere*, 11(4), p. 354.

Jewson, S., Barnes, C., Cusack, S. & Bellone, E., 2019. Adjusting catastrophe model ensembles using importance sampling, with application to damage estimation for varying levels of hurricane activity. *Met Apps*, Volume 27, pp. 1-14.

Knutti, R. et al., 2017. A climate model projection weighting scheme accounting for performance and interdependence. *GRL*, Volume 44, pp. 1909-1918.

Maraun, D., 2016. Bias Correcting Climate Change Simulations. *Curr. Clim. Change Rep.*, Volume 2, p. 211.

Masson, D. & Knutti, R., 2011. Spatial-Scale Dependence of Climate Model Performance in the CMIP3 Ensemble. *J. Clim*, Volume 24, pp. 2680-2692.

Mitchell-Wallace, K., Jones, M., Hillier, J. & Foote, M., 2017. *Natural Catastrophe Risk Management and Modelling*. s.l.:Wiley Blackwell.

Moss, R. et al., 2010. The next generation of scenarios for climate change research and assessment. *Nature*, Volume 463, pp. 747-756.

Raisanen, J. & Ylhaisi, J., 2010. How Much Should Climate Model Output Be Smoothed in Space?. *J. Clim*, Volume 24, pp. 867-880.

Riahi, K. et al., 2017. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, Volume 42, pp. 153-168.

Sanderson, B., Wehner, M. & Knutti, R., 2017. Skill and independence weighting for multi-model assessments.. *Geosci. Model Dev.*, Volume 10, pp. 2379-2395.

Sassi, M. et al., 2019. Impact of climate change on European winter and summer flood losses. *Advances in Water Resources*, Volume 129, pp. 165-177.
Scher, S., Jewson, S. & Messori, G., 2021. Robust worst-case scenarios from ensemble forecasts. *Wea. Forecasting*, Volume (in press).

Sippel, S. et al., 2019. Uncovering the Forced Climate Response from a Single Ensemble Member Using Statistical Learning. *J. Clim.*, 32(17), p. 5677–5699.

Stainforth, D. et al., 2005. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. *Nature*, 433(7024), pp. 403-406.

Taylor, K., Stouffer, R. & Meehl, G., 2012. An Overview of CMIP5 and the Experiment Design. *BAMS*, 93(4), pp. 485-498.

Thompson, D. et al., 2015. Quantifying the Role of Internal Climate Variability in Future Climate Trends. *J. Clim.*, 28(16), p. 6443–6456.

Wills, R. et al., 2020. Pattern Recognition Methods to Separate Forced Responses from Internal Variability in Climate Model Ensembles and Observations. *J. Clim.*, 33(20), p. 8693–8719.

Yip, S., Ferro, C., Stephenson, D. & Hawkins, E., 2011. A Simple, Coherent Framework for Partitioning Uncertainty in Climate Predictions. *J. Clim.*, 24(17), p. 4634–4643.

### Tables

| Model | Driving GCM | GCM Member | RCM |
|-------|-------------|------------|-----|
| M1    | CNRM-CM5    | r1i1p1     | ALADIN53 |
| M2    | IPSL-CM5A-MR| r1i1p1     | RCA4  |
| M3    | CNRM-CM5    | r1i1p1     | RCA4  |
| M4    | CNRM-CM5    | r1i1p1     | CCLM4-8-17 |
| M5    | EC-EARTH    | r1i1p1     | CCLM4-8-17 |
| M6    | EC-EARTH    | r12i1p1    | RACMO22E |
| M7    | EC-EARTH    | r12i1p1    | RCA4  |
| M8    | EC-EARTH    | r1i1p1     | RACMO22E |
| M9    | EC-EARTH    | r3i1p1     | HIRHAM5 |
| M10   | IPSL-CM5A-MR| r1i1p1     | WRF331F |

Table 1: The climate models used in this study.
Figure 1 Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2 Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3 As Fig. 2a, but now for two randomly generated 10 member ensembles

Figure 4 Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble
Figure 5 Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities $p$ for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance $m$ from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function $\alpha$ increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a black cross, calculated using Eq. (1) with $r$ normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio $a'/m$, with darker shading indicating higher values. Panel (f) shows values of the product $pa'$, with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.
Figure 6 Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations.
Figure 7 Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern.
Figures

Figure 1

Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2

Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3

As Fig. 2a, but now for two randomly generated 10 member ensembles.

Figure 4

Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Figure 5

Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with $\mathbf{x}$ normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio $\mathbf{y}/\mathbf{x}$, with darker shading indicating higher values. Panel (f) shows values of the product $\mathbf{y}'$, with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figures

Figure 1

Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2

Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3
As Fig. 2a, but now for two randomly generated 10 member ensembles.

Figure 4
Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with  normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio \( \frac{\mathbb{E}^r}{\mathbb{E}^t} \), with darker shading indicating higher values. Panel (f) shows values of the product \( \mathbb{E}^r \mathbb{E}^t \), with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations.

Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figures

Figure 1

Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2

Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3
As Fig. 2a, but now for two randomly generated 10 member ensembles

Figure 4
Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with \( \mathcal{I} \) normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio \( \mathcal{I} / \mathcal{I} \), with darker shading indicating higher values. Panel (f) shows values of the product \( \mathcal{I} \mathcal{I} \), with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figures

Figure 1

Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2

Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1
Figure 3

As Fig. 2a, but now for two randomly generated 10 member ensembles

Figure 4

Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Figure 5

Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with $\overline{D}$ normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio $\overline{D}/\overline{A}$, with darker shading indicating higher values. Panel (f) shows values of the product $\overline{D}\overline{A}$, with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.

Figure 6
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figures

Figure 1

Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2

Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3

As Fig. 2a, but now for two randomly generated 10 member ensembles.

Figure 4

Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with $\bar{z}$ normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio $\bar{z}/\bar{z}$, with darker shading indicating higher values. Panel (f) shows values of the product $\bar{z}^2$, with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.

Figure 6
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Figure 2

Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3

As Fig. 2a, but now for two randomly generated 10 member ensembles

Figure 4

Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Figure 5

Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with $\mathcal{B}$ normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio $\mathcal{B}'/\mathcal{B}$, with darker shading indicating higher values. Panel (f) shows values of the product $\mathcal{B}'\mathcal{B}$, with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Four illustrations of Directional Component Analysis (DCA), applied to climate variables at locations 1 and locations 2. The ellipse in each case illustrates a contour of probability density for an ensemble prediction for these two locations. The ensemble is assumed to be bivariate normally distributed. The red dot in each case illustrates the ensemble mean, the arrow indicates the gradient of increasing impact, the straight line, which is perpendicular to the arrow, indicates a line of constant impact and the blue dot indicates the resulting DCA pattern. Examples (a) and (b) show the same ensemble, but different directions of impact, and hence different DCA patterns. Examples (a) and (c) show different ensembles, but the same direction of impact, and hence different DCA patterns. Example (d) shows a different ensemble and a new direction of impact.
Four randomly generated 1000 member examples and related quantities. In each case the ensemble members are shown by black dots, the ensemble mean is shown by a red dot, the gradient of increasing impact is shown by the arrow and the ensemble mean plus the DCA pattern, scaled to two standard deviations of impact, is shown by a blue dot. Examples (a) and (b) use the same ensemble, while examples (a) and (c) use the same impact vector. These examples correspond to the illustrations in Fig. 1.
Figure 3

As Fig. 2a, but now for two randomly generated 10 member ensembles.

Figure 4

Worst members and DCA patterns for 500 randomly generated ensembles, for the same parameters as the ensembles shown in Fig. 3. The red dots show the mean of the ensemble means. Panel (a) shows 500 worst members, one for each ensemble. Panel (b) shows 500 DCA patterns, one for each ensemble.
Figure 5

Illustrations of mathematical properties of DCA. In each panel, the gradient of increasing impact is shown by the black arrow that points towards the top right hand corner. Panel (a) shows probability densities for a bivariate normal, with darker shading indicating higher densities, but lower values of the Mahalanobis distance from the origin. The density is elliptical, and the principal axis of the ellipse is given by the green arrow. Panel (b) shows values of a linear impact function increasing towards the top right-hand corner, with darker shading indicating higher impact. Panel (c) shows the DCA pattern as a
black cross, calculated using Eq. (1) with $\mathbf{v}$ normalized to unit length. This black cross is repeated in all subsequent panels. Panel (d) shows the DCA pattern as a red circle, calculated using Eq. (2). Panel (e) shows values of the ratio $\mathbf{u} / \mathbf{v}$, with darker shading indicating higher values. Panel (f) shows values of the product $\mathbf{u} \cdot \mathbf{v}$, with darker shading showing higher values. Panel (g) shows three examples of the conditional mean pattern, conditional on exceeding a certain level of impact, as red circles. The impact thresholds are given by the three diagonal lines. Panel (h) shows three examples of the impact-weighted conditional mean pattern as red circles, conditional on exceeding a certain level of impact. The impact thresholds are given by the diagonal lines. Panel (f) shows three examples of quadratic DCA patterns, as red circles.

Figure 6
Annual mean precipitation change from the EURO-CORDEX ensemble, based on the models listed in Table 1, for the difference between the periods 2011-2040 and 1981-2010, in mm/year. EM is ensemble mean and ESD is ensemble standard deviation. Panel (a) shows the ensemble mean change, panel (b) shows twice the standard deviation of change. Panel (c) shows the ensemble mean change plus twice the standard deviation of change and panel (d) shows the ensemble mean change plus twice the standard deviation of change in regions where the ensemble mean change is positive, and the ensemble mean change minus twice the standard deviation of change in regions where the ensemble mean change is negative. Non-linear colour scales are used to highlight the spatial variations. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.
Figure 7

Reasonable worst-case precipitation scenarios calculated using DCA, based on the EURO-CORDEX data used to create Fig. 6. Panel (a) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points, normalized to two standard deviations. Panel (c) shows the ensemble mean change plus this pattern. Panel (b) shows a DCA pattern derived from an impact function defined as the sum of precipitation at all points where the ensemble mean change shows positive precipitation, minus the sum of precipitation at all points where the ensemble mean change shows negative precipitation. Panel (d) shows the ensemble mean change plus this pattern. Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.