



Multi-annual predictions of hot, dry and hot-dry compound extremes

Alvise Aranyosy^{1,2}, Paolo De Luca¹, Carlos Delgado-Torres¹, Balakrishnan Solaraju-Murali^{1,3}, Margarida Samso Cabre¹, and Markus G. Donat^{1,4}

¹Barcelona Supercomputing Center (BSC), Barcelona, Spain

²Universitat de Barcelona, Barcelona, Spain

³Zurich Insurance, Via Augusta 200, 08021, Barcelona

⁴Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Correspondence: Alvise Aranyosy (alvise.aranyosy@bsc.es)

Abstract. Hot-dry compound extremes have recently gained increasing attention due to their potential impacts on environments and societies. For these reasons, assessing climate predictions is essential to providing reliable information on such extremes. However, despite several studies focusing on compound extremes in the past and climate projections, little is known on a multi-annual timescale. At this regard, decadal climate predictions have been produced to provide useful information for this specific timescale. Thus, we evaluate the ability of the CMIP6 multi-model decadal climate hindcast in predicting hot-dry climate extremes, as well as their hot and dry univariate counterparts for the forecast years 2-5. The multi-model skillfully predicts hot-dry compound extremes and hot extremes over most land regions, while the skill is more limited for dry extremes. However, we find only small and spatially limited improvements from the initialisation of the hindcasts, especially for the hot-dry compound extremes, with most of the skill coming from external forcings, especially long-term trends. Finally, we find the decadal hindcast to be able to reproduce the connections between the compound extremes and their hot and dry univariate components. Evaluations such as this of decadal hindcast are an essential tool to establish the potential and the limits of these products, a necessary step to provide reliable and useful information regarding such impactful extremes.

1 Introduction

High-impact weather events, such as droughts, heatwaves, heavy precipitation and storms, often result from a combination of different physical processes that can interact on various spatial and temporal scales. The aggregation or sequence of these extreme events on the same or different temporal and spatial scales can be considered a compound extreme event (Bevacqua et al., 2021). The combination of these hazards can amplify the impacts on climate-vulnerable sectors and society, and recognising them only from a univariate perspective may lead to a severe underestimation of the risk (Leonard et al., 2014; Zscheischler and Fischer, 2020). In addition, the dependence among the univariate variables of a compound extreme can increase the likelihood and decrease the return period of such events, compared to situations where the univariate variables show no dependence (Zscheischler and Seneviratne, 2017).

Among these compound extremes, hot and dry compound extremes have gained increasing attention due to the potential of



elevated environmental, economic and societal losses, of which the compound extremes during 2003, 2010, 2015 and 2018 in Europe stand as an example (Grumm, 2011; Sutanto et al., 2020; Vogel et al., 2020; Zscheischler and Fischer, 2020). Several studies have focused on the variability of these extreme events in the past and how they are projected to change in the future (Zscheischler et al., 2018; Ridder et al., 2021; Bevacqua et al., 2022; De Luca and Donat, 2023). However, little is known about the predictability of the occurrence of these events on multi-annual time scales. Trustworthy predictions are essential for developing strategic planning to help mitigate any associated impacts. From this perspective, predictions that cover a multi-annual range could be more useful for potential users than projections for the long-term future.

Decadal predictions bridge the gap between short-term seasonal predictions and long-term climate projections. Unlike climate projections, in climate predictions, the initial state of the climate system is prescribed to closely represent the observed climate state at the time of starting the prediction, a procedure known as model initialisation (Meehl et al., 2021). To account for the observational uncertainty in the initial conditions, these climate predictions are initialised from a certain number of slightly perturbed states. Each of the outcoming simulations is called an ensemble member. Decadal prediction systems initially contributed to the Coupled Model Intercomparison Project Phase 5 (CMIP5; Taylor et al. (2012)) and Phase 6 (CMIP6; Eyring et al. (2016)) and have been shown to produce skilful predictions for several climate variables such as near-surface temperature and, to a lesser extent, precipitation (Smith et al., 2019; Delgado-Torres et al., 2022). Multi-annual predictions, which focus on the first period of the decadal forecasts, show the potential for providing useful information for a variety of socio-economic sectors. These forecasts show a significant prediction skill for the average of forecast years 2-5, which in some cases is comparable, especially in the prediction of temperatures, to the skill shown for seasonal predictions in the average of the corresponding forecast months (Kushnir et al., 2019; O’Kane et al., 2023). Further studies have also tested the predictability of extreme events on a multi-annual scale, providing comparable results: temperature-related extremes show stronger prediction skill compared to precipitation-related extremes (Solaraju-Murali et al., 2019; Delgado-Torres et al., 2023). To the best of our knowledge, decadal predictions have only been assessed for climate variables or univariate climate extremes. Here, we aim to assess the skill of hot-dry compound extreme multi-annual predictions in a multi-model setup. We also investigate the improvement of the skill provided due to model initialisation. Finally, given the importance of the interconnections between compound extremes and their univariate counterparts, we also analyse these correlations and compare the results between observations and models.

2 Data and Methods

We analyse the prediction skill of hot-dry compound extremes for a multi-model ensemble from the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring et al., 2016). The list of the models, the number of ensemble members and the corresponding horizontal resolutions are presented in Table A1. For each model, we use two different sets of experiments. The first set, hereafter referred to as DCPM MME, consists of 10-year initialised hindcasts from the Decadal Climate Prediction Project (DCPP) (Boer et al., 2016). This set of retrospective decadal predictions is produced every year from 1960 to 2009. For this study, we select and analyse from these decadal forecasts the forecast years 2-5. Thus, in this setup the initial forecast year



2 is 1962, while the last forecast year 5 is 2014. The second set of experiments, hereafter referred to as Hist MME, consists of CMIP6 historical forcing simulations from 1962 to 2014. To estimate the skill of the DCPM MME and Hist MME experiments, we use two observation-based reference datasets. The first reference dataset, which will be hereafter referred to as GPCC-BEST, combines the Global Precipitation Climatology Centre (GPCC, Schneider et al. (2016)) for monthly total precipitation and the Berkeley Earth Surface Temperatures (BEST, Rohde and Hausfather (2020)) for daily maximum and daily minimum temperature. Similarly, the second reference dataset combines daily total precipitation, daily maximum and daily minimum temperature and is obtained from the European Centre for Medium-Range Weather Forecasts (ECMWF) Reanalysis v5, and will be referred to as ERA5 (Hersbach et al., 2020).

2.1 Univariate and compound extreme indicators

To identify hot-dry compound extreme days, we first define univariate hot and dry extremes. For temperature extremes, named hereafter TX90p, we select the days above the 90th percentile of daily maximum temperature measuring the frequency of unusually warm days in each year, as per the definition of the Expert Team on Climate Change Detection and Indices (ETCCDI; Zhang et al. (2011)). The percentile is calculated for every day of the year, with a 5-day running window, and is based on the entire period of the study, which in the case of Hist MME, GPCC-BEST and ERA5 is the timeframe 1962-2014, while for DCPM MME depends on the forecast year, going from 1962-2011 for forecast year 2 to 1965-2014 for forecast year 5. In the case of the DCPM MME and Hist MME, the percentile is calculated using data from all the ensemble members for each model individually to account for the models' different variances. For dry extremes, we use two different indicators: the Standardized Precipitation Index (SPI, McKee et al. (1993)) and the Standardized Precipitation-Evapotranspiration Index (SPEI, Vicente-Serrano et al. (2010)). These two indicators are calculated through a two-step method, an accumulation and a standardization process. For the standardization process, we use a gamma distribution for the SPI, and a log-logistic distribution for the SPEI, the latter using unbiased probability-weighted moments (Beguería et al., 2014; Stagge et al., 2015). The result is a standardized index for the total precipitation (SPI) and one for the difference between precipitation and potential evapotranspiration (SPEI), each computed during selected accumulated periods, with the potential evapotranspiration (PET) computed with the Hargreaves method (Hargreaves, 1994). For computing these indices DCPM MME, we follow Solaraju-Murali et al. (2019). This method treats the accumulation and the standardization steps differently, due to the nature of decadal predictions. First, the accumulation is done from the first to the last forecast month of every initialisation. In the second step, the standardization is performed individually for each forecast month, using all initialisations together. From the resulting time series (one for each month) we form the distribution used to calculate SPI and SPEI. Consequently, the start and end of the reference period will depend on the forecast month analysed. This is done separately for each ensemble member. In this study, we analyse an accumulation period of 3 months. Thus, we analyse SPI3 and SPEI3 which are indicators of meteorological (SPI) and hydrological (SPEI) wet (positive values) or dry (negative values) conditions. Finally, at a grid-point level, we define months with drought conditions all months with SPI or SPEI ≤ -1 and name these dry indices as SPI3dry and SPEI3dry respectively, where 3 stands for the accumulation period in months (De Luca and Donat, 2023).

Hot-dry compound extremes are then identified following De Luca and Donat (2023), which retain all those TX90p days which



90 fall into SPI3dry or SPEI3dry months. With this method, we obtain two types of hot-dry compound extremes: one based on SPI (HDSPI3, Hot-Dry selected by SPI3dry) and one on SPEI (HDSPEI3, Hot-Dry selected by SPEI3dry).

2.2 Forecast quality assessment

We use the Spearman rank correlation coefficient to assess the skill of the DCPM MME mean at forecasting the annual TX90p, HDSPI3 and HDSPEI3, and the SPI3dry and SPEI3dry for the average of the forecast years 2-5. This correlation coefficient
 95 measures the relationship between the forecasted and observed time series, respectively ranked from their lowest to their highest value (Wilks, 2011). We also use the residual correlation to quantify the skill improvement due to initialisation over the skill that comes from the model responses to external forcing (Smith et al., 2019). This method consists of linearly regressing out the ensemble mean of the Hist MME from the observations and the ensemble mean of the DCPM MME, obtaining the residual values. Finally, we calculate the correlation between the two residual time series, thus measuring how much of the observed
 100 variability is captured by the DCPM MME forecast that is not already captured by the Hist MME forecast. To try to keep the comparison between the DCPM MME and the Hist MME forecasts as consistent as possible, for the test between the residuals, we only use those models in the DCPM MME ensemble with at least one member also in the Hist MME ensemble (see Table A1). For all the correlations shown here, we consider only values statistically significant at the 95% confidence level (p-value < 0.05) using a one-tailed test for the prediction skill, and a two-tailed test for the residual skill deriving from initialisation.

105 3 Results

3.1 Multi-annual prediction skill of hot, dry and hot-dry compound extremes

In this section, we present the forecast quality assessment of multi-annual averages for the DCPM MME's forecast years 2-5. The correlation analysis is performed on the HDSPI3 and HDSPEI3 indicators and their univariate counterparts (TX90p, SPI3dry and SPEI3dry) between the DCPM MME mean and GPCC-BEST (Figure 1), as well as between DCPM MME mean
 110 and ERA5 as an alternative reference dataset (Appendix Figure S1).

For HDSPI3 (Figure 1a), we see significantly positive skill in the Mediterranean region, the Balkans, the Arabian peninsula and southern and eastern Asia, the western coast of North America, Central America, most South America, equatorial Africa and southern Australia. On the other hand, HDSPEI3 shows significantly positive skill over most of the globe, with only the Arctic and sub-Arctic regions, the western Sahara, parts of South America and portions of equatorial Africa showing non-significant
 115 correlations (Figure 1b). When we analyse the skill of the univariate extremes that contribute to these compound extremes, we find that TX90p (Figure 1c) shows statistically significant positive skill across the globe with only a few regions, such as the lower latitudes of South America, Western Sahara and some areas of the sub-Arctic regions, showing non-significant values. These results are in accordance with Delgado-Torres et al. (2023). On the other hand, SPI3dry and SPEI3dry show limited regions with significantly positive skill (Figure 1d,e). For SPI3dry, most of the positive skill is found in northern and
 120 equatorial Africa, Scandinavia, certain areas of eastern Europe, and some northern and central Asian regions. Positive skill is

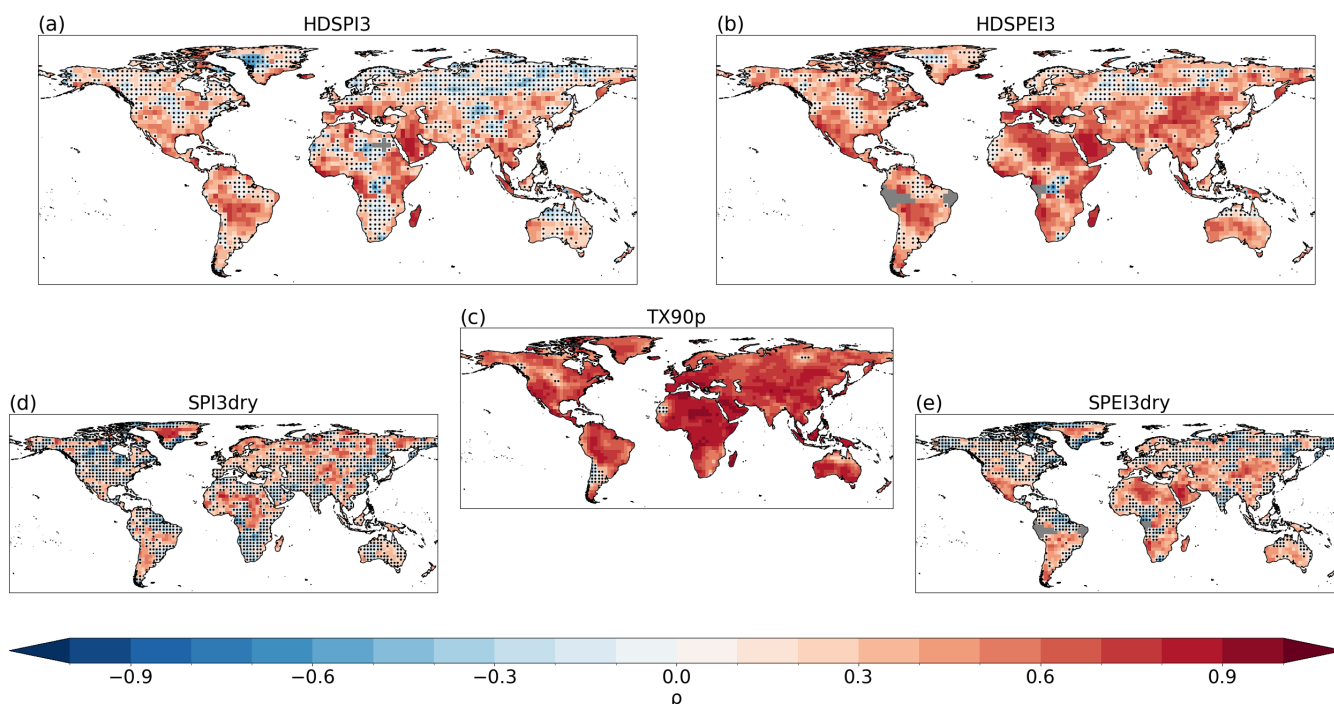


Figure 1. Spearman rank correlation for HDSPi3 (a), HDSPEI3 (b), TX90p (c), SPI3dry (d) and SPEI3dry (e) for the average forecast years 2-5 (1962-2014). Correlations are performed between the DCPM MME ensemble mean and GPCC-BEST. Dots indicate regions where the correlation is not statistically significant at the 95% confidence level (p-value < 0.05) using a one-tailed test. Grey areas indicate missing data.

also found in some parts of South America and Australia. Similarly, SPEI3dry shows positive and significant correlations over the Mediterranean region, large areas of Africa, the Arabian Peninsula, central and eastern Asia, central America and some parts of Southern America, as well as for the southern part of Australia. In general, it appears that SPEI3dry shows higher skill than SPI3dry, especially over the mid-latitudes. If we compare the skill of hot-dry compound extremes to their univariate counterparts, in general, we see an enhanced skill compared to the univariate dry extremes, but not as high as shown by the hot ones. In addition, dry and especially hot-dry compound extremes computed from SPEI show increased skill compared to the ones calculated with SPI. This characteristic could be partly explained by the contribution of Potential Evapotranspiration, which, relying on temperature, positively contributes to the general skill. This explanation appears to be supported by the fact that the more weight temperature has in the definition of the hot-dry compound extremes (for example HDSPEI3), the lower skill in dry condition (in this case SPEI3dry) seems to be a limiting factor in the overall skill.

By performing the analysis with ERA5 as the reference dataset (Appendix Figure A1), we obtain some differences among the analysed indices. In the compound extremes (HDSPi3 and HDSPEI3) we find similar areas that show significant skill compared to GPCC-BEST. In addition, as in the previous results for GPCC-BEST, HDSPEI3 shows an enhanced skill compared to HDSPi3. Regarding the univariate extremes, TX90p still shows the strongest skill, and SPEI3dry shows stronger skill than



135 SPI3dry, especially in the Northern Hemisphere. In general, compared to GPCC-BEST, the skill using ERA5 as a reference dataset shows a slight diminishment in the area where significant positive correlations are found. This is especially true for the dry extremes (SPI3dry and SPEI3dry), where large areas, specifically at the high latitudes and equatorial Africa, show strong negative correlations.

3.2 Residual correlations and trends

140 The skill of the decadal forecast shown above (Figure 1 and Appendix Figure A1) can be essentially divided into two components: the skill resulting from external forcings and the skill resulting from initialisation. In this section, we analyse to what extent the two elements contribute to the overall skill.

To quantify the added skill contributed by the models' initialisation in decadal predictions, we perform the correlation between the residuals of GPCC-BEST and the DCPM MME ensemble mean (Figure 2). HDSPI3 shows small areas of added skill around the globe. In addition, these are quite circumscribed and isolated, not showing a larger region where clear added skill can be found (Figure 2a). For HDSPEI3 on the other hand, despite showing a similar spatial pattern, areas of consistent added skill can be found in central Asia and southern Australia, and some regions of the high latitudes of North America (Figure 2b). Between the univariate extremes, TX90p (Figure 2d) appears to show the most benefit from the initialisation, with added skill located in central and eastern Asia, Greenland and southern Australia. In addition, we also find some added value over small parts of North and South America and Africa. For dry extremes, the areas of added skill are more limited. SPI3dry shows added skill in some small areas of South America, Asia and Scandinavia (Figure 2d). Moreover, northern Africa also shows a homogeneous area of added skill, which, at least in part, corresponds to the significant skill observed for SPI3dry (Figure 1d). On the other hand, SPEI3dry shows slightly fewer areas of added skill, limited to southern Australia and central Asia (Figure 2e), the latter being similar to the region where we find added skill for TX90p.

155

The residual correlation obtained with ERA5 as a reference dataset, leads to similar results, both in terms of general patterns and number of gridpoints showing significant residual correlations (Appendix Figure A2). However, for HDSPI3 and HDSPEI3 and SPI3dry and SPEI3dry, an increase in added skill can be noticed in the tropical region of South America, as well as in eastern Asia.

160

In general, the results in Figures 2 and A2 show that, for the univariate hot and dry extremes, we have certain limited areas where the initialisation significantly adds skill to the forecast (Figure 1, Figure A1). However, it appears that the added skill detected in certain areas for the univariate extremes is mostly lost when analysing the compound ones. Indeed, the results suggest that the skill relies strongly on the models' response to external forcings, such as the strong warming trend in temperatures, especially in the case of lower latitudes for HDSPEI3, where we find strong overall skill (in terms of correlation with observations) but little residual correlation. To better establish the role of external forcings and trends in the prediction skill, we thus explore the connection between significant correlations and the signs of the trends (Figure 3). From these results, it appears that positive significant skill in most regions is tied to an agreement in trend sign between DCPM MME and GPCC-

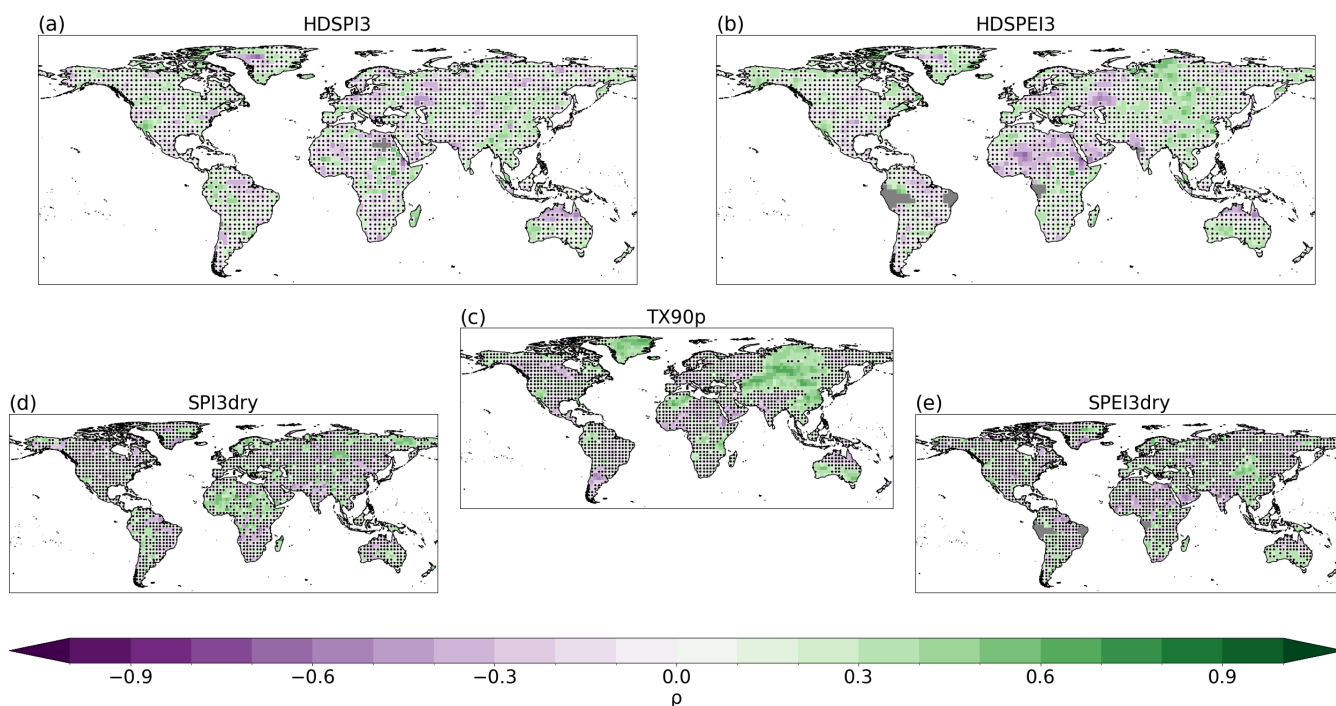


Figure 2. Spearman rank residual correlation for HDSPi3 (a), HDSPEi3 (b), TX90p (c), SPI3dry (d) and SPEI3dry (e) for the average forecast years 2-5. Correlations are performed between the residuals of the DCPM MME mean and GPCC-BEST. Dots indicate areas where the residual correlation is not statistically significant at the 95% confidence level (p -value < 0.05) using a two-tailed test. Grey areas indicate missing data.

BEST. In addition, in most areas where the correlation is significantly negative, we find that the trends between the two datasets are discordant. An exception can be seen for SPI3dry, where, especially in Eastern Europe, the Western Sahel and regions of mid-latitude South America the correlations are significantly positive, but the sign of the trend is discordant (Figure 3d). Two factors could help explain this. The first is the weak trend in precipitation in those areas for both datasets (Appendix Figure A4), which also explains the opposite scenario, a significantly negative correlation but coinciding trend signs, visible for example in South Africa and tropical South America. The second factor is the added skill from initialisation for those grid points, which despite the discordant trend, leads to a significant positive skill (Figure 2d). On the contrary, all the extreme indices that include temperature show a strong agreement between positive (negative) significant correlations and agreeing (discordant) trend signs, specifically evident in HDSPEi3 and SPEI3dry (Figure 3b and 3e), results which are in line with the findings of Donat et al. (2023). The same conclusions can be drawn concerning the results when using ERA5 as a reference dataset (Appendix Figure A3). Here the areas with opposite trends and significantly negative correlations are evident in tropical Africa for SPI3dry and SPEI3dry and in the high boreal latitudes for HDSPi3 and HDSPEi3. When looking at the trends, we see that these areas show, for the aforementioned extreme indices, opposite trends in DCPM MME and ERA5. However, in these regions, we can also see areas with significant negative correlations and agreeing trend signs (Appendix Figure A3d and A3e). These results could be

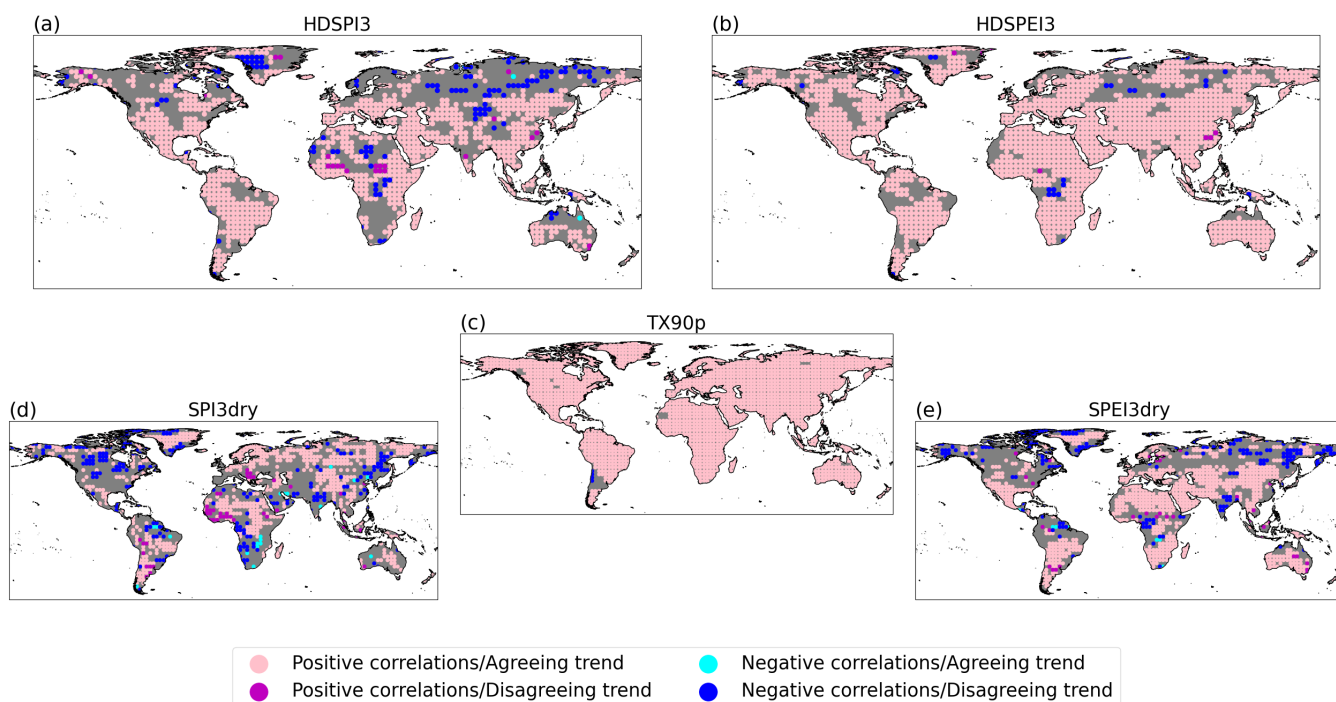


Figure 3. Statistically significant Spearman Rank correlations against the trend sign for HDSPI3 (a), HDSPEI3 (b), TX90p (c), SPI3dry (d) and SPEI3dry (e) for the average forecast years 2-5 (as in Figure 1). Correlations and trend's sign comparison are performed between the DCPM MME ensemble mean and GPCC-BEST. Pink (magenta) dots indicate areas where the correlation is positive and statistically significant at the 95% confidence level (p -value < 0.05) and where the trend of GPCC-BEST and DCPM MME show the same (opposite) sign. Light blue (Dark blue) dots indicate areas where the correlation is negative statistically significant at the 95% confidence level (p -value < 0.05) and where the trend of GPCC-BEST and DCPM MME show the same (opposite) sign. Grey areas indicate grid points where the correlation is non-significant.

explained by the weak trends in the region (Appendix Figure A5), making it a non-significant factor in determining significant skill. For SPI3dry, we can see the same situation observed for GPCC-BEST in tropical South America (Appendix Figure A3d). Here we have discordant trends between DCPM MME and ERA5 but significant positive correlations. We can also see, between DCPM MME and ERA5, that in that area, the precipitation trends are not strong (Appendix Figure A5) and that the forecast shows a large region of added skill for ERA5 (Appendix Figure A2d).

In summary, from the results discussed above we see that decadal predictions show some added skill compared to Hist MME. However, most of the skill seen in Section 3.1 is still tied to the signal associated with external forcings. In particular, we see that the presence of the skill, in the absence of added skill from the initialisation, is tied to the trends of the models and the reference datasets, and whether these two trends are agreeing in their signs or not, especially if the selected variable is affected by strong temperature trends. In this regard, the question now lies to which extent the different univariate extremes influence the detection and the variability of the compound extremes in the model and the reference datasets.



3.3 Correlations between univariate and compound extremes

195 With the method used in this study, relying on the co-occurrence of hot and dry conditions at the same location, one of the two univariate extremes may play a more important role in the overall detection of the events. In addition, this co-occurrence may vary between different datasets, for example, between the DCPM MME and the observation-based datasets, leading to a different count of compound extremes and, consequently, different interannual variations. Thus, we perform a correlation analysis between the annual sum of univariate and compound extremes to assess the representation of the underlying intercon-

200 nections. Since the HDSPI3 and HDSPEI3 were estimated for every ensemble member, we calculated the correlations between univariate and compound for every ensemble member and picked the median correlation across DCPM MME. Finally, we consider a significant correlation in those areas where at least 50% of the ensemble members show significant positive or negative correlations. The results for the correlations between HDSPI3 and HDSPEI3 and the univariate extremes are shown in Figures 4 and 5, respectively. If we look at the correlations between HDSPI3 and TX90p in GPCC-BEST and the DCPM MME (Figure

205 4a,b), we can see that they both show significantly positive correlations across most of the globe. In addition, the DCPM MME can reproduce the large areas where the connections between HDSPI3 and TX90p are weaker, such as the high latitudes of the Northern Hemisphere. However, it is less effective in areas such as Africa or central Asia, where correlations in DCPM MME are still significant, unlike GPCC-BEST. On the other hand, when looking at the correlations between HDSPI3, SPI3dry and SPEI3dry we see similar patterns (Figures 4c,4d and 4e,4f respectively). Both GPCC-BEST and DCPM MME show strong

210 positive correlations between drought occurrence and hot-dry compound extremes in most regions globally, except mainly in northern Africa. However, as seen in the correlations with TX90p (Figures 4a, 4b) DCPM MME partly fails to reproduce the high spatial differences of the connections seen in GPCC-BEST, with only larger areas, such Northern Africa, showing differences in the correlation values (Figures 4d, 4e).

215 For the correlations between HDSPEI3 and the univariate extremes (Figure 5), we see patterns very similar to Figure 4. In this case, the correlations between HDSPEI3 and TX90p appear stronger and more widely significant in GPCC-BEST and the DCPM MME, with the high latitudes showing the weakest correlations (Figures 5a, 5b). The correlations between HDSPEI3 and SPEI3dry in GPCC-BEST have strong positive correlations globally, with DCPM MME being able to reproduce them significantly well (Figures 5c, 5d). On the other hand, the correlations between HDSPEI3 and SPI3dry in GPCC-BEST show

220 high spatial differences, with large areas showing non-significant correlations. However, DCPM MME shows non-significant correlations uniquely across northern Africa, the Arabian Peninsula and central Asia (Figures 5e, 5f).

Analysing the correlations between compound and univariate extremes in ERA5 we can observe a very similar pattern as in the results seen above for GPCC-BEST (Appendix Figure A6). In general, both reference datasets (GPCC-BEST and ERA5) show stronger correlations for TX90p with HDSPEI3, especially in the equatorial and mid-latitudes. Still, correlations are

225 found for the correlations between hot-dry compounds and dry extremes, which appear to be an important driving factor in the variability of both HDSPI3 and HDSPEI3, and more specifically for the dry extreme with which the compound extreme is being calculated (i.e. SPI3dry for HDSPI3, SPEI3dry for HDSPEI3). The DCPM MME is able to reproduce these general

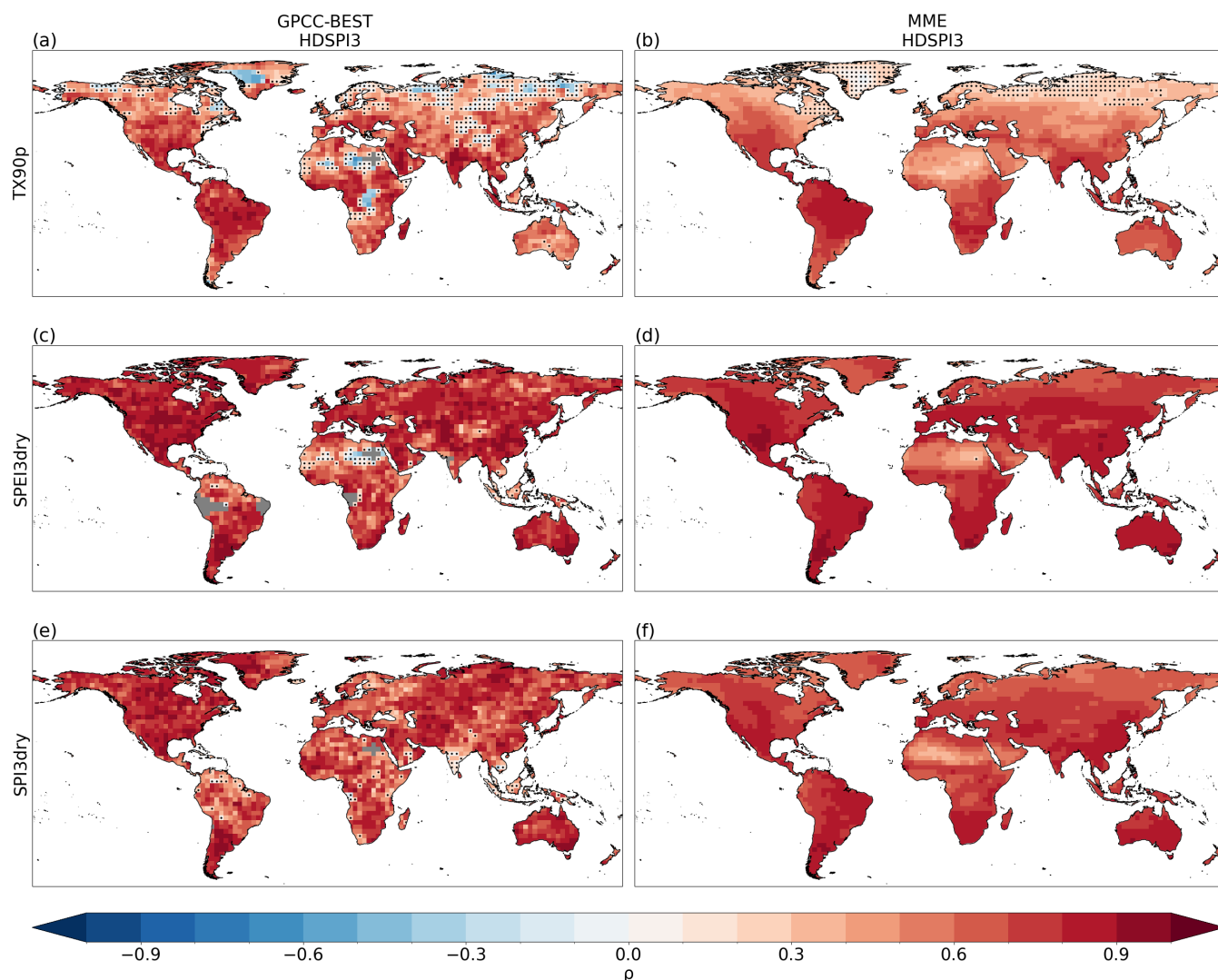


Figure 4. Spearman rank correlation between the annual sum of HD SPI3 and TX90p (a,b), SPEI3dry (c,d) and SPI3dry (e,f), respectively. For a,c, and e: correlations are calculated with GPCC-BEST, and dots indicate correlations not statistically significant at the 95% confidence level (p -value < 0.05) using a two-tailed test. For b, d and f: the correlations represent the median (ensemble-wise) correlation for the DCPM MME ensemble, and dots indicate grid points where more than 50% of the ensemble members provide correlations not statistically significant at the 95% confidence level (p -value < 0.05) using a two-tailed test. Grey areas indicate missing data.

patterns, particularly the weaker correlations in the high North Hemisphere latitudes for TX90p and in Northern Africa for the dry extremes. However, it fails to represent the high spatial differences in the correlation values, which instead can be seen in the reference datasets. This can be attributed to the smoothing that inevitably happens while calculating the ensemble mean, causing the variability of the single ensemble members to be lost.

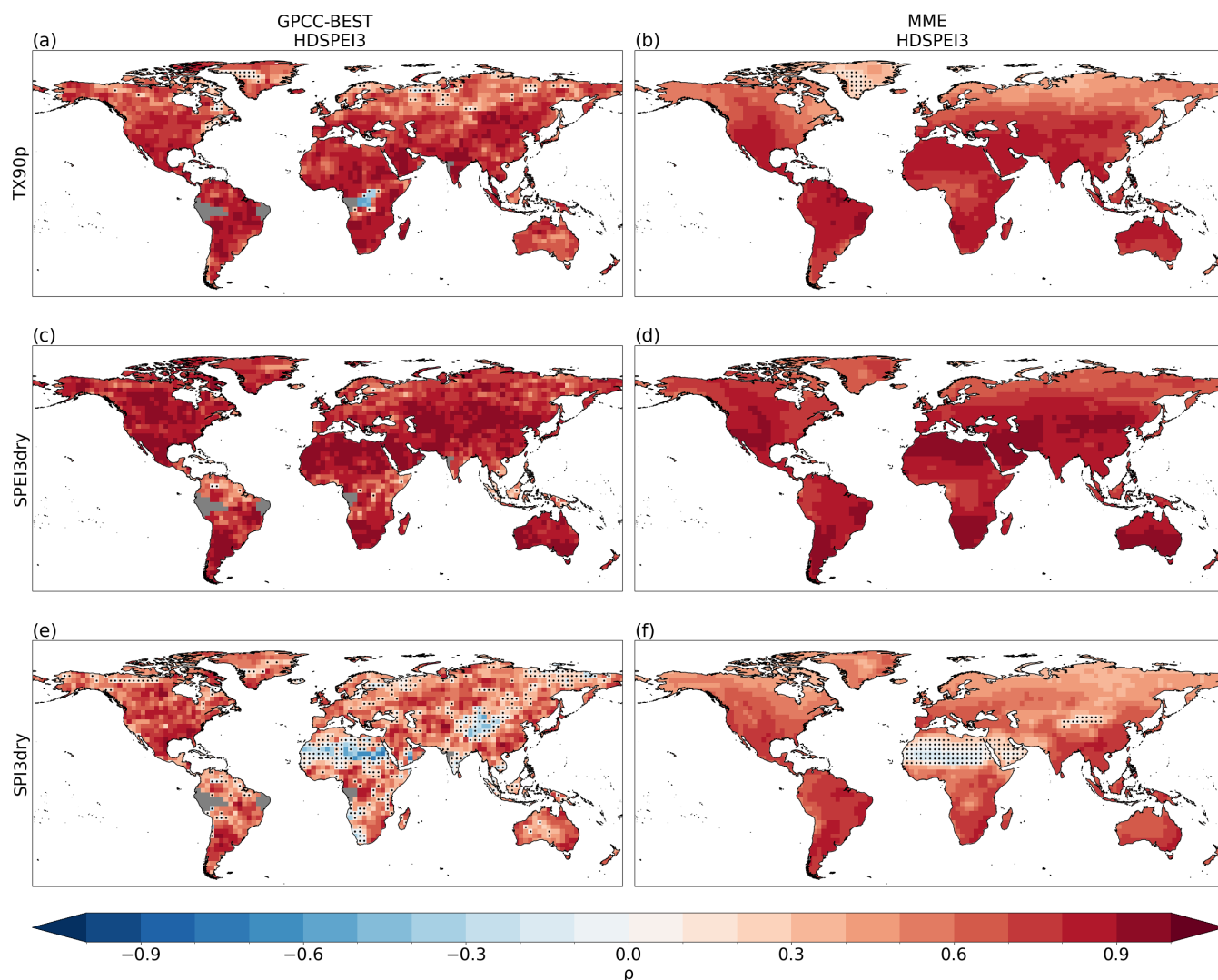


Figure 5. Same as Figure 4, but for HDSPEI3.

4 Summary and Conclusions

This study provides an overall assessment of hot, dry and hot-dry compound extremes in multi-annual predictions. The results show that the predictive capabilities of the decadal forecast systems are not only limited to the univariate extremes but also show value for the compound extreme occurrences. Compared to the skill shown in the univariate extremes, compound extremes show higher or lower predictability: global compound skill is smaller compared to hot extremes, especially at high latitudes; on the other hand, they show higher skill compared to their dry univariate counterparts. This is likely associated with the fact that temperature-related extremes have been seen shown to be more predictable than their precipitation-related coun-



terparts, with the latter's predictability still representing a bottleneck that could damp the signal (Solaraju-Murali et al., 2019; Bellucci et al., 2015; Delgado-Torres et al., 2023; Smith et al., 2019; De Luca and Donat, 2023). From this perspective, the compound extremes partly benefit from the higher predictability of hot extremes. An analysis of different metrics of the skill also shows that some areas, especially for the univariate extremes, benefit from added skill from the initialisation of the decadal forecast systems. However, for compound extremes, most areas of the globe show that the skill of Hist MME is similar to the one of DCPM MME. These results show how most of the skill of the compound extremes comes from external forcings, which cause long-term trends in some extreme indices. This is less true for dry extremes, where we can find specific regions with a weak trend, and where the initialisation actively plays a role in providing significant skill. In fact, we see that in the variables where temperatures play a role in the definition of the extreme indices (i.e. TX90p, SPEI3dry, HDSPI3 and HDSPEI3), the tie with the trend is nearly ubiquitous (Donat et al., 2023). In addition to these results, the analysis of the interconnections between compound and univariate extremes shows that the decadal forecast systems are generally able to reproduce the connections between the compound extremes and their univariate counterparts based on observations, attributing a leading role to the dry conditions, and specifically to the one on which the compound extreme is calculated (e.g. drought defined based on SPI or SPEI). Thus, if we compare these results with the prediction skill, we can identify where the main limiting factor in the prediction of hot-dry compound extremes may lie: dry extremes show to be the leading factor determining the variability of hot-dry extremes, but on the other hand it is the dry extremes that show the least prediction skill on a multi-annual time range. Lastly, we also ascertain that the use of two different reference datasets (GPCC-BEST and ERA5) leads to results which differ slightly, but not significantly, and can easily be attributed to observational uncertainties.

We conclude that CMIP6 decadal forecast systems can skillfully predict the compound hot and dry extremes over large parts of the world at multi-annual scale. While the added skill from initialisation compared to the skill in response to global forcings is limited, they can still provide useful information for these high-impact extremes. Finally, the decadal prediction systems can detect the underlying connections between compound and univariate extremes, even if at a large-scale level, showing the leading role of dry conditions. Similar forecast quality assessments should be conducted on a regional and seasonal scale to provide even more insightful and specific information on such extreme phenomena. These further steps could provide the information necessary to enable targeted climate information to support mitigation and adaptation policies against potential damages caused by these events.

Data availability. Gridded ERA5 data ranging from 1961 to 2014 was downloaded from <https://cds.climate.copernicus.eu/>. Gridded GPCC data ranging from 1961 to 2014 were downloaded from https://opendata.dwd.de/climate_environment/GPCC/html/fulldata-monthly_v2020_doi_download.html. Gridded BEST data ranging from 1961 to 2014 were downloaded from <http://berkeleyearth.lbl.gov/>. CMIP6 dcpa hindcasts and historical simulations were downloaded from <https://esg-dn1.nsc.liu.se/projects/esgf-liu>.



270 *Author contributions.* AA, MD and PDL conceptualized and designed the study. MSC provided the data for the simulations, observations and reanalysis. CDT and BSM contributed in the analyses. AA wrote the first draft. PDL, CDT and MD contributed to interpreting the results, discussing the findings and improving the final version of the paper.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. This research was funded by the Horizon Europe project ASPECT (Grant 101081460). We are also grateful for support
275 by the Departament de Recerca i Universitats de la Generalitat de Catalunya for the Climate Variability and Change (CVC) Research Group (Reference: 2021 SGR 00786), and support by the AXA Research Fund. A.A. was funded by MCIN/AEI/10.13039/501100011033 and ESF Investing in Your Future (Grant PRE2022-104391). P.D.L. was funded by Horizon Europe Marie Skłodowska-Curie Actions (Grant 101059659).



References

- 280 Beguería, S., Vicente-Serrano, S. M., Reig, F., and Latorre, B.: Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring, *International journal of climatology*, 34, 3001–3023, 2014.
- Bellucci, A., Haarsma, R., Gualdi, S., Athanasiadis, P., Caian, M., Cassou, C., Fernandez, E., Germe, A., Jungclaus, J., Kröger, J., et al.: An assessment of a multi-model ensemble of decadal climate predictions, *Climate Dynamics*, 44, 2787–2806, 2015.
- 285 Bethke, I., Wang, Y., Counillon, F., Keenlyside, N., Kimmritz, M., Fransner, F., Samuelsen, A., Langehaug, H., Svendsen, L., Chiu, P.-G., et al.: NorCPM1 and its contribution to CMIP6 DCP, *Geoscientific Model Development*, 14, 7073–7116, 2021.
- Bevacqua, E., De Michele, C., Manning, C., Couasnon, A., Ribeiro, A. F., Ramos, A. M., Vignotto, E., Bastos, A., Blesić, S., Durante, F., et al.: Guidelines for studying diverse types of compound weather and climate events, *Earth's Future*, 9, e2021EF002 340, 2021.
- Bevacqua, E., Zappa, G., Lehner, F., and Zscheischler, J.: Precipitation trends determine future occurrences of compound hot–dry events, *Nature Climate Change*, 12, 350–355, 2022.
- 290 Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., Kushnir, Y., Kimoto, M., Meehl, G. A., Msadek, R., et al.: The decadal climate prediction project (DCPP) contribution to CMIP6, *Geoscientific Model Development*, 9, 3751–3777, 2016.
- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., et al.: Presentation and evaluation of the IPSL-CM6A-LR climate model, *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002 010, 2020.
- 295 De Luca, P. and Donat, M. G.: Projected changes in hot, dry, and compound hot-dry extremes over global land regions, *Geophysical Research Letters*, 50, e2022GL102 493, 2023.
- Delgado-Torres, C., Donat, M. G., Gonzalez-Reviriego, N., Caron, L.-P., Athanasiadis, P. J., Bretonnière, P.-A., Dunstone, N. J., Ho, A.-C., Nicoli, D., Pankatz, K., et al.: Multi-model forecast quality assessment of CMIP6 decadal predictions, *Journal of Climate*, 35, 4363–4382, 2022.
- 300 Delgado-Torres, C., Donat, M. G., Soret, A., González-Reviriego, N., Bretonnière, P.-A., Ho, A.-C., Pérez-Zanón, N., Cabré, M. S., and Doblas-Reyes, F. J.: Multi-annual predictions of the frequency and intensity of daily temperature and precipitation extremes, *Environmental Research Letters*, 18, 034 031, 2023.
- Donat, M. G., Delgado-Torres, C., De Luca, P., Mahmood, R., Ortega, P., and Doblas-Reyes, F. J.: How credibly do CMIP6 simulations capture historical mean and extreme precipitation changes?, *Geophysical Research Letters*, 50, e2022GL102 466, 2023.
- 305 Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arneth, A., Arsouze, T., Bergmann, T., Bernadello, R., Bousetta, S., Caron, L.-P., et al.: The EC-earth3 Earth system model for the climate model intercomparison project 6, *Geoscientific Model Development Discussions*, 2021, 1–90, 2021.
- Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., and Taylor, K. E.: Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization, *Geoscientific Model Development*, 9, 1937–1958, 2016.
- 310 Grumm, R. H.: The central European and Russian heat event of July–August 2010, *Bulletin of the American Meteorological Society*, 92, 1285–1296, 2011.
- Hargreaves, G. H.: Defining and using reference evapotranspiration, *Journal of irrigation and drainage engineering*, 120, 1132–1139, 1994.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al.: The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, 146, 1999–2049, 2020.
- 315



- Kataoka, T., Tatebe, H., Koyama, H., Mochizuki, T., Ogochi, K., Naoe, H., Imada, Y., Shiogama, H., Kimoto, M., and Watanabe, M.: Seasonal to decadal predictions with MIROC6: Description and basic evaluation, *Journal of Advances in Modeling Earth Systems*, 12, e2019MS002 035, 2020.
- Kushnir, Y., Scaife, A. A., Arritt, R., Balsamo, G., Boer, G., Doblas-Reyes, F., Hawkins, E., Kimoto, M., Kolli, R. K., Kumar, A., et al.: Towards operational predictions of the near-term climate, *Nature Climate Change*, 9, 94–101, 2019.
- Leonard, M., Westra, S., Phatak, A., Lambert, M., van den Hurk, B., McInnes, K., Risbey, J., Schuster, S., Jakob, D., and Stafford-Smith, M.: A compound event framework for understanding extreme impacts, *Wiley Interdisciplinary Reviews: Climate Change*, 5, 113–128, 2014.
- McKee, T. B., Doesken, N. J., Kleist, J., et al.: The relationship of drought frequency and duration to time scales, in: *Proceedings of the 8th Conference on Applied Climatology*, vol. 17, pp. 179–183, Boston, 1993.
- Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., Donat, M. G., England, M. H., Fyfe, J. C., Han, W., et al.: Initialized Earth System prediction from subseasonal to decadal timescales, *Nature Reviews Earth & Environment*, 2, 340–357, 2021.
- Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., Bunzel, F., Esch, M., Ghosh, R., Haak, H., et al.: A higher-resolution version of the max planck institute earth system model (MPI-ESM1. 2-HR), *Journal of Advances in Modeling Earth Systems*, 10, 1383–1413, 2018.
- Nicoli, D., Bellucci, A., Ruggieri, P., Athanasiadis, P. J., Materia, S., Peano, D., Fedele, G., Hénin, R., and Gualdi, S.: The Euro-Mediterranean Center on Climate Change (CMCC) decadal prediction system, *Geoscientific Model Development*, 16, 179–197, 2023.
- O’Kane, T. J., Scaife, A. A., Kushnir, Y., Brookshaw, A., Buontempo, C., Carlin, D., Connell, R. K., Doblas-Reyes, F., Dunstone, N., Förster, K., et al.: Recent applications and potential of near-term (interannual to decadal) climate predictions, *Frontiers in Climate*, 5, 1121 626, 2023.
- Ridder, N. N., Pitman, A. J., and Ukkola, A. M.: Do CMIP6 climate models simulate global or regional compound events skillfully?, *Geophysical Research Letters*, 48, e2020GL091 152, 2021.
- Rohde, R. A. and Hausfather, Z.: The Berkeley Earth land/ocean temperature record, *Earth System Science Data*, 12, 3469–3479, 2020.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., and Ziese, M.: GPCC full data reanalysis version 7.0: Monthly land-surface precipitation from rain gauges built on GTS based and historic data, 2016.
- Sellar, A. A., Walton, J., Jones, C. G., Wood, R., Abraham, N. L., Andrejczuk, M., Andrews, M. B., Andrews, T., Archibald, A. T., de Mora, L., et al.: Implementation of UK Earth system models for CMIP6, *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001 946, 2020.
- Smith, D., Eade, R., Scaife, A., Caron, L.-P., Danabasoglu, G., DelSole, T., Delworth, T., Doblas-Reyes, F., Dunstone, N., Hermanson, L., et al.: Robust skill of decadal climate predictions, *Npj Climate and Atmospheric Science*, 2, 13, 2019.
- Solaraju-Murali, B., Caron, L.-P., Gonzalez-Reviriego, N., and Doblas-Reyes, F. J.: Multi-year prediction of European summer drought conditions for the agricultural sector, *Environmental Research Letters*, 14, 124 014, 2019.
- Stagge, J. H., Tallaksen, L. M., Gudmundsson, L., Van Loon, A. F., and Stahl, K.: Candidate distributions for climatological drought indices (SPI and SPEI), *International Journal of Climatology*, 35, 4027–4040, 2015.
- Sutanto, S. J., Vitolo, C., Di Napoli, C., D’Andrea, M., and Van Lanen, H. A.: Heatwaves, droughts, and fires: Exploring compound and cascading dry hazards at the pan-European scale, *Environment international*, 134, 105 276, 2020.
- Swart, N. C., Cole, J. N., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., Anstey, J., Arora, V., Christian, J. R., Hanna, S., et al.: The Canadian earth system model version 5 (CanESM5. 0.3), *Geoscientific Model Development*, 12, 4823–4873, 2019.



- Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An overview of CMIP5 and the experiment design, *Bulletin of the American meteorological Society*, 93, 485–498, 2012.
- 355 Vicente-Serrano, S. M., Beguería, S., and López-Moreno, J. I.: A multiscalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index, *Journal of climate*, 23, 1696–1718, 2010.
- Vogel, M. M., Hauser, M., and Seneviratne, S. I.: Projected changes in hot, dry and wet extreme events' clusters in CMIP6 multi-model ensemble, *Environmental Research Letters*, 15, 094 021, 2020.
- Wilks, D. S.: *Statistical methods in the atmospheric sciences*, Academic press, 2011.
- 360 Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., Tsujino, H., Deushi, M., Tanaka, T., Hosaka, M., et al.: The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2. 0: Description and basic evaluation of the physical component, *Journal of the Meteorological Society of Japan. Ser. II*, 97, 931–965, 2019.
- Zhang, X., Alexander, L., Hegerl, G. C., Jones, P., Tank, A. K., Peterson, T. C., Trewin, B., and Zwiers, F. W.: Indices for monitoring changes in extremes based on daily temperature and precipitation data, *Wiley Interdisciplinary Reviews: Climate Change*, 2, 851–870, 2011.
- 365 Zscheischler, J. and Fischer, E. M.: The record-breaking compound hot and dry 2018 growing season in Germany, *Weather and Climate Extremes*, 29, 100 270, 2020.
- Zscheischler, J. and Seneviratne, S. I.: Dependence of drivers affects risks associated with compound events, *Science advances*, 3, e1700 263, 2017.
- Zscheischler, J., Westra, S., Van Den Hurk, B. J., Seneviratne, S. I., Ward, P. J., Pitman, A., AghaKouchak, A., Bresch, D. N., Leonard, M.,
- 370 Wahl, T., et al.: Future climate risk from compound events, *Nature climate change*, 8, 469–477, 2018.



Appendix

Table A1. List of models and number of ensemble members used for the DCPM MME and the Hist MME.

Forecast systems	DCPM members	Historical members	Spatial resolution	Month initialisation	Reference
CMCC-CM2-SR5	10	1	0.9°x1.25°	November	Nicoli et al. (2023)
CanESM5	20	20	2.8°x2.8°	January	Swart et al. (2019)
EC-Earth3-i1	10	-	0.7°x0.7°	November	Döscher et al. (2021)
EC-Earth3-i2	5	-	0.7°x0.7°	November	Döscher et al. (2021)
EC-Earth3-i4	10	9	0.7°x0.7°	November	Döscher et al. (2021)
HadGEM3-GC31-MM	10	-	0.55°x0.83°	November	Sellar et al. (2020)
IPSL-CM6A-LR	10	10	1.25°x2.5°	January	Boucher et al. (2020)
MIROC6	10	10	1.4°x1.4°	November	Kataoka et al. (2020)
MPI-ESM1-2-HR	10	1	0.9°x0.9°	November	Müller et al. (2018)
MRI-ESM2-0	10	10	1.125°x1.125°	November	Yukimoto et al. (2019)
NorCPM1-i1	10	-	1.9°x2.5°	October	Bethke et al. (2021)
NorCPM1-i2	10	10	1.9°x2.5°	October	Bethke et al. (2021)
MME	125	80	2.8°x2.8°		

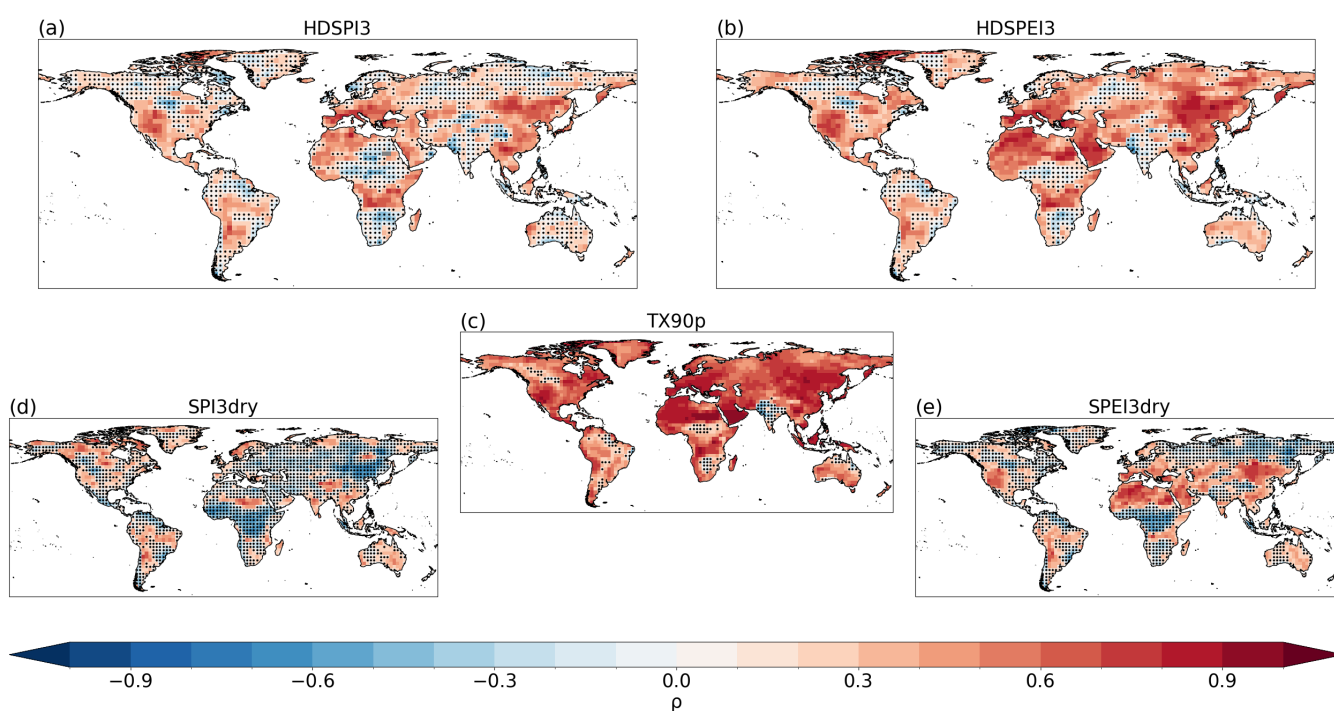


Figure A1. Spearman rank correlation for TX90p(a), SPI3dry and SPEI3dry (b,c) and HD SPI3 and HDSPEI3 (d,e) for the average forecast years 2-5. Correlations are performed between the DCP MME ensemble mean and the ERA5 dataset. Hatchings indicate correlations not statistically significant at the 95% confidence level (p-value < 0.05) using a one-tailed test.

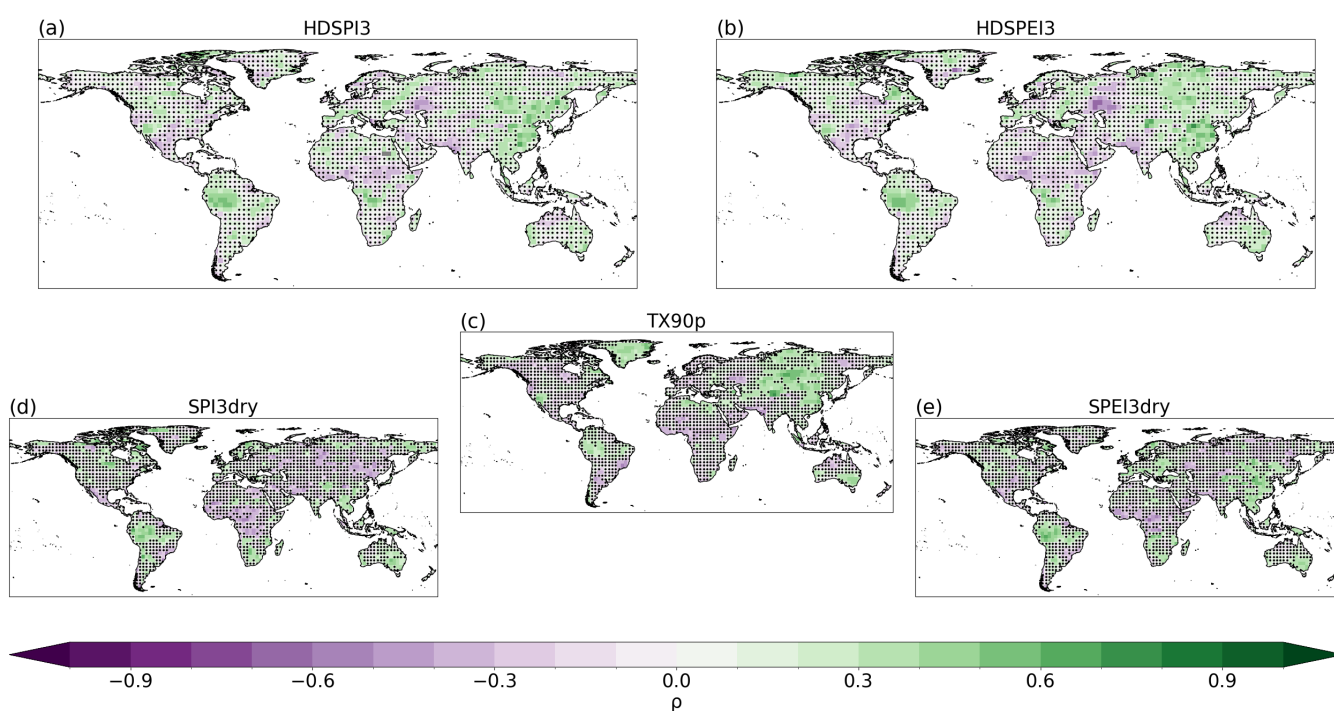


Figure A2. Spearman rank residual correlation for TX90p(a), SPI3dry and SPEI3dry (b,c) and HD SPI3 and HDSPEI3 (d,e) for the average forecast years 2-5. Correlations are performed between the DCPM MME ensemble mean and the ERA5 dataset. Hatchings indicate correlations not statistically significant at the 95% confidence level (p -value < 0.05) using a two-tailed test.

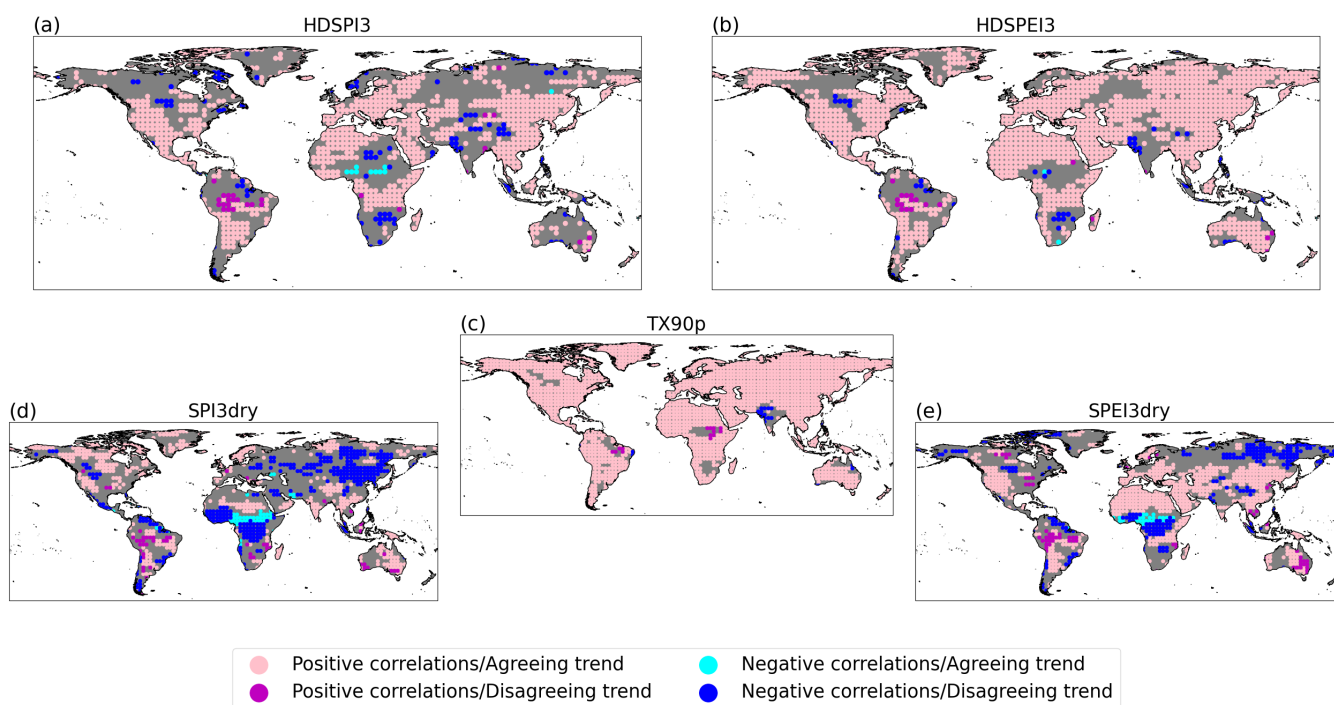


Figure A3. Statistically significant Spearman Rank correlations against the trend sign for HDSPi3 (a), HDSPEi3 (b), TX90p (d), SPI3dry (c) and SPEI3dry (e) for the average forecast years 2-5 (as in Figure 1). Correlations and trend sign comparison are performed between the DCPM MME ensemble mean and the ERA5 dataset. Pink (magenta) dots indicate areas where the correlation is positive and statistically significant at the 95% confidence level (p -value < 0.05) and where the trend of GPCC-BEST and DCPM MME show the same (opposite) sign. Light blue (Dark blue) dots indicate areas where the correlation is negative statistically significant at the 95% confidence level (p -value < 0.05) and where the trend of GPCC-BEST and DCPM MME show the same (opposite) sign. Grey areas indicate grid points where the correlation is non-significant.

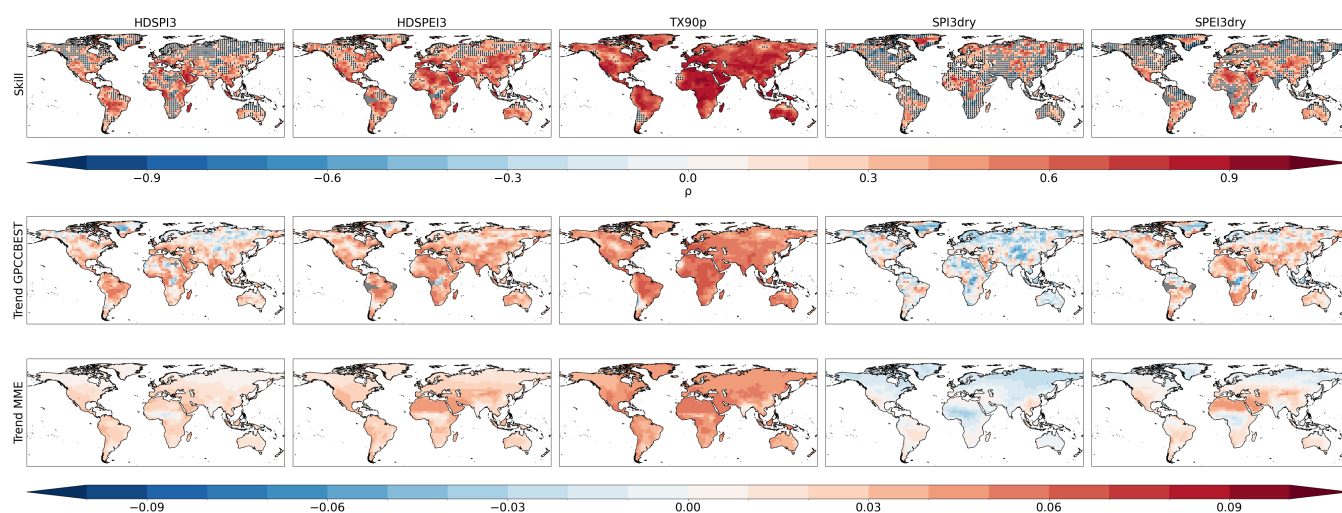


Figure A4. First row: Spearman rank correlation between the GPCC-BEST dataset and the DCPP MME ensemble mean for HDSP13, HDSPEI3, TX90p, SPI3dry and SPEI3dry respectively. Hatchings indicate correlations not statistically significant at the 95% confidence level (p -value < 0.05) using a one-tailed test. Second row: Normalized trends in the GPCC-BEST dataset for HDSP13, HDSPEI3, TX90p, SPI3dry and SPEI3dry respectively. Third row: Median normalized trends, calculated ensemble-wise, in the DCPP MME ensemble for HDSP13, HDSPEI3, TX90p, SPI3dry and SPEI3dry respectively. Grey areas indicate missing data.

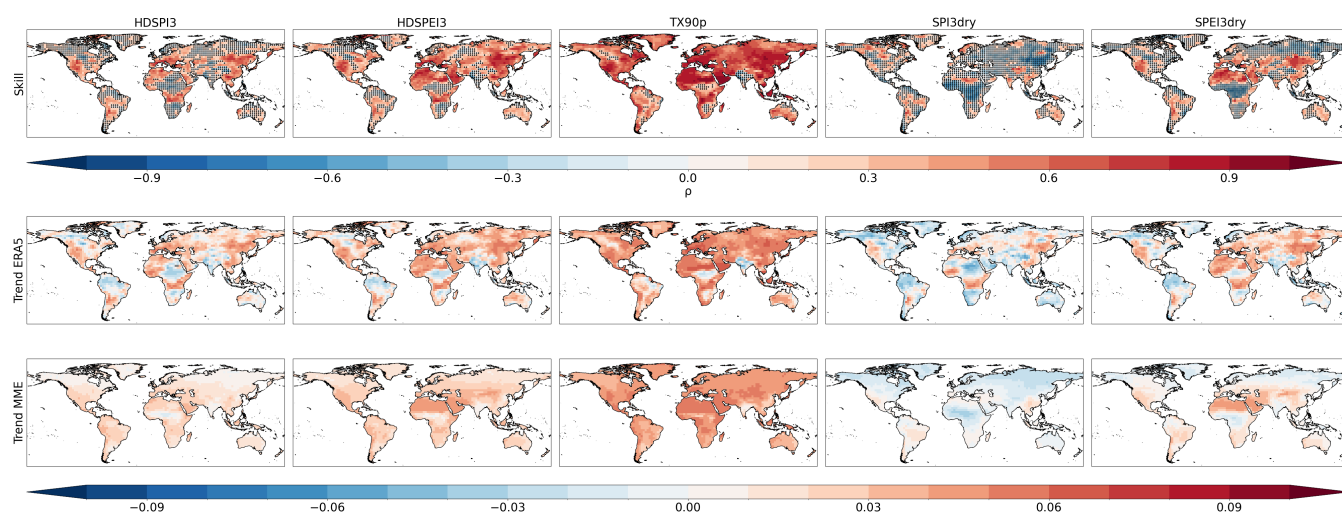


Figure A5. First row: Spearman rank correlation between the ERA5 dataset and the DCP MME ensemble mean for HD SPI3, HDSPEI3, TX90p, SPI3dry and SPEI3dry respectively. Hatchings indicate correlations not statistically significant at the 95% confidence level (p -value < 0.05) using a one-tailed test. Second row: Normalized trends in the ERA5 dataset for HD SPI3, HDSPEI3, TX90p, SPI3dry and SPEI3dry respectively. Third row: Median normalized trends, calculated ensemble-wise, in the DCP MME ensemble for HD SPI3, HDSPEI3, TX90p, SPI3dry and SPEI3dry respectively.

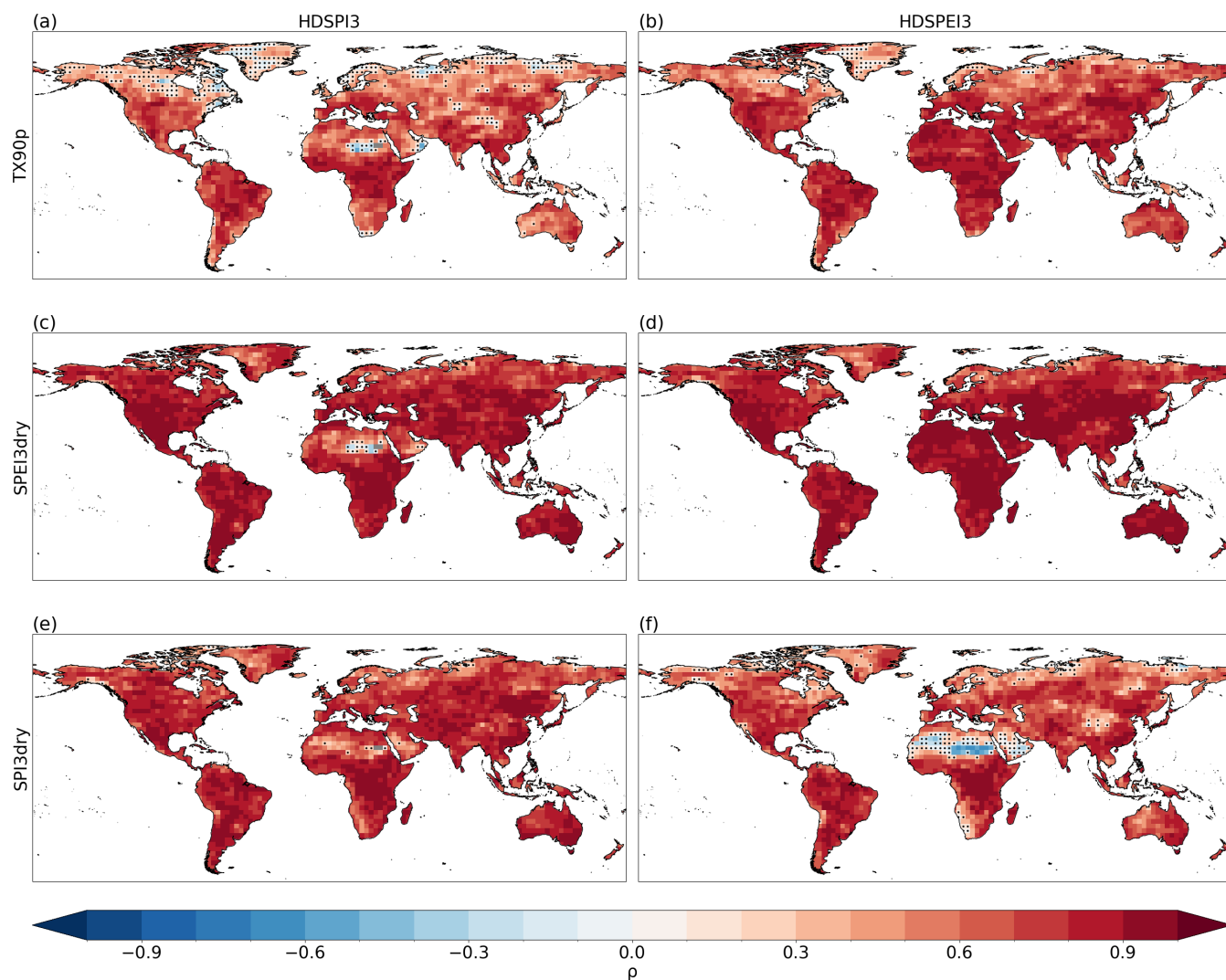


Figure A6. Spearman rank correlation between the annual sum of HD SPI3 (a,c,e) and HD SPEI3 (b,d,f) and TX90p, SPEI3dry and SPI3dry respectively, in the ERA5 dataset. Hatchings indicate correlations not statistically significant at the 95% confidence level (p -value < 0.05) using a two-tailed test.