ENVIRONMENTAL RESEARCH
LETTERS

ACCEPTED MANUSCRIPT • OPEN ACCESS

Constraining decadal variability regionally improves near-term projections of hot, cold and dry extremes

To cite this article before publication: Paolo De Luca *et al* 2023 *Environ. Res. Lett.* in press https://doi.org/10.1088/1748-9326/acf389

Manuscript version: Accepted Manuscript

Accepted Manuscript is “the version of the article accepted for publication including all changes made as a result of the peer review process, and which may also include the addition to the article by IOP Publishing of a header, an article ID, a cover sheet and/or an ‘Accepted Manuscript’ watermark, but excluding any other editing, typesetting or other changes made by IOP Publishing and/or its licensors”

This Accepted Manuscript is **© 2023 The Author(s). Published by IOP Publishing Ltd**.

As the Version of Record of this article is going to be / has been published on a gold open access basis under a CC BY 4.0 licence, this Accepted Manuscript is available for reuse under a CC BY 4.0 licence immediately.

Everyone is permitted to use all or part of the original content in this article, provided that they adhere to all the terms of the licence https://creativecommons.org/licences/by/4.0

Although reasonable endeavours have been taken to obtain all necessary permissions from third parties to include their copyrighted content within this article, their full citation and copyright line may not be present in this Accepted Manuscript version. Before using any content from this article, please refer to the Version of Record on IOPscience once published for full citation and copyright details, as permissions may be required. All third party content is fully copyright protected and is not published on a gold open access basis under a CC BY licence, unless that is specifically stated in the figure caption in the Version of Record.

View the article online for updates and enhancements.

This content was downloaded from IP address 84.88.53.170 on 30/08/2023 at 12:40

Constraining decadal variability regionally improves near-term projections of hot, cold and dry extremes

P. De Luca[1], C. Delgado-Torres[1], R. Mahmood[1], M. Samso-Cabre[1], and M.G. Donat[1,2]

1. Barcelona Supercomputing Center (BSC), Barcelona, Spain
2. Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Corresponding author: P. De Luca (paolo.deluca@bsc.es)

Abstract

Hot, cold and dry meteorological extremes are often linked with severe impacts on the public health, agricultural, energy and environmental sectors. Skillful predictions of such extremes could therefore enable stakeholders to better plan and adapt to future impacts of these events. The intensity, duration and frequency of such extremes are affected by anthropogenic climate change and modulated by different modes of climate variability. Here we use a large multi-model ensemble from the Coupled Model Intercomparison Project Phase 6 and constrain these simulations by sub-selecting those members whose global SST anomaly patterns are most similar to observations at a given point in time, thereby phasing in the decadal climate variability with observations. Hot and cold extremes are skillfully predicted over most of the globe, with also a widespread added value from using the constrained ensemble compared to the unconstrained full CMIP6 ensemble. On the other hand, dry extremes show skill only in some regions with results sensitive to the index used. Still, we find skillful predictions and added skill for dry extremes in some regions such as western north America, southern central and eastern Europe, southeastern Australia, southern Africa and the Arabian Peninsula. We also find that the added skill in the constrained ensemble is due to a combination of improved multi-decadal variations in phase with observed climate extremes and improved representation of long-term changes. Our results demonstrate that constraining decadal variability in climate projections can provide improved estimates of temperature extremes and drought in the next twenty years, which can inform targeted adaptation strategies to near-term climate change.

Keywords

Climate projections | Temperature and dry extremes | CMIP6 | Climate variability | Prediction skill | Constraint

Accepted Manuscript

1. Introduction

Hot and dry meteorological extremes are nowadays having significant impacts on societies, economies and ecosystems worldwide (Blauhut *et al* 2015, 2016, Brás *et al* 2021, Ebi *et al* 2021, Xu *et al* 2016, Wilhite *et al* 2007, García-León *et al* 2021). Such events are also projected to become stronger and more frequent in the future under anthropogenic climate change (Coumou and Robinson 2013, Cook *et al* 2018, Dai 2011, 2013, Fischer *et al* 2013, Fischer and Schär 2010, Sillmann *et al* 2013, De Luca and Donat 2023). In addition to hot and dry also winter cold extremes over the mid-latitudes can pose significant distress to infrastructures, emergency services, agricultural and energy sectors (Cheng *et al* 2019, Wang *et al* 2010, Guirguis *et al* 2011, Palmer 2014, Sillmann *et al* 2011). Given their potentially severe impacts, it is important to anticipate future changes of hot, cold and dry extremes with skillful climate predictions, so that their occurrence probabilities are correctly anticipated and suitable adaptation strategies can be implemented by governments and stakeholders. Information about near-term climate change (e.g. the next 10-30 years) is particularly important to inform strategic decisions to plan adaptation.

Anthropogenic climate change is expected to continue in the next decades as greenhouse gas concentrations are projected to rise due to continued emissions (Masson-Delmotte *et al* 2021) and is expected to drive further increases in hot and dry extremes (e.g. Sillmann *et al* 2013, De Luca and Donat 2023). On the other hand, the internal variability of the climate system plays a crucial role in shaping the climate on inter-annual to multi-decadal timescales (Dai *et al* 2015, Mann *et al* 2014, Meehl *et al* 2013). In fact, internal variability is the dominating source of uncertainty for projections of regional climate in the first few decades (Hawkins and Sutton 2009, Lehner *et al* 2020). For assessing future near-term climate change therefore both forced warming and climate variability need to be taken into account in order to provide the most accurate estimates of changes on these time scales.

Initialised climate predictions aim at reducing uncertainty from internal variability by synchronising the phasing of variability modes between the model simulations and the observations (Meehl *et al* 2021, Merryfield *et al* 2020, Meehl *et al* 2009). Initialised predictions show regionally improved skill when compared to uninitialised climate projections over some land regions for mean values of climatic variables (Smith *et al* 2019, Delgado-Torres *et al* 2022) and extreme indices in multi-annual predictions (Delgado-Torres *et al* 2023). However, because they are very computationally-expensive, especially if initialised every year, the time-span of these predictions is typically limited to ten years after initialization as in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Eyring *et al* 2016) Decadal Climate Prediction Project (DCPP) (Boer *et al* 2016). Moreover, decadal predictions are affected by initialization shocks and by their drift towards the model's preferred climate state which can negatively affect their skill (e.g. Bilbao *et al* 2021). Recently, several approaches have been developed that allow to obtain skillful climate prediction by constraining internal climate variability from large

Accepted Manuscript

ensembles of climate projections (Befort *et al* 2020, Mahmood *et al* 2021, 2022). Such methods select those ensemble members from large ensembles of transient climate simulations that are in closest agreement with for example a climate prediction (Befort *et al* 2020, Mahmood *et al* 2021) or observational (Mahmood *et al* 2022) reference dataset. This selection procedure is conceptually similar to initialisation of climate predictions (Meehl *et al* 2021) and has the main advantage to exploit initialisation information beyond the 10 years of decadal prediction, without much computational cost since it uses existing climate projections, and which in addition can provide seamless information until the end of the century. These constrained climate projections are consistent with the model-specific climate attractors and are therefore not affected by shock, drift and related artefacts (Hazeleger *et al* 2013, Bilbao *et al* 2021, Smith *et al* 2013).

Here we follow the approach of Mahmood *et al* (2022) for constraining decadal climate variability in a large multi-model ensemble, and we assess the prediction skill of hot, cold and dry extremes in these constrained projections over global land areas. With this method, we constrain climate variability based on the similarities, at a given point in time, between a large CMIP6 multi-model ensemble (MME) and multi-annual averages of observed sea surface temperature (SST) anomaly patterns. The method, for each year, sub-selects only those ensemble members which are most in agreement with the observed SST patterns. For the skill assessment we focus on the next 20-year period after applying the constraint, which is a time-scale where a previous study (Mahmood *et al* 2022) showed added value for some annual mean variables and where the role of internal variability is still large.

2. Data

We use 149 ensemble members coming from a MME of 19 CMIP6 models (Table S1). From this MME we consider data of the historical simulations from 1960 to 2014 and concatenate them with the Shared Socioeconomic Pathway (SSP) 2-4.5 (O'Neill *et al* 2016) up to 2019 included. The data we analyse for calculating the extreme indices are monthly total precipitation (mm), daily and monthly minimum and maximum surface temperatures (°C). By the time of the analysis, these 149 members were all available members from the MME used in Mahmood *et al* (2022) that provided daily data required for computing the extremes indices. We evaluate these simulated extremes against observations-based datasets; to address sensitivity to the choice of reference dataset we use one observational and one reanalysis dataset for temperatures and two observational datasets for precipitation. The reference datasets we use are the gridded Berkeley Earth Surface Temperatures (BEST, https://climatedataguide.ucar.edu/climate-data/global-surface-temperatures-best-berkeley-earth-surface-temperatures) from which we obtain daily and monthly minimum and maximum surface temperatures, and the Global Precipitation Climatology Center (GPCC) (Becker *et al* 2013) from which we use monthly total precipitation. To test the robustness of the results related to the choice of the reference datasets we also replicate all the analyses using ERA5 reanalysis (Hersbach *et al* 2020) from where, similarly to BEST, we obtain minimum and maximum temperatures and Rainfall Estimates on a Gridded

Accepted Manuscript

Network (REGEN; Contractor *et al* (2020)) dataset from which, similarly as GPCC, we extract monthly total precipitation. We choose BEST and GPCC as the main reference for the retrospective predictions because they are both observational datasets and their combination allows us to end our hindcast evaluation in 2019, since this is the last year available for GPCC. On the other hand, ERA5 and REGEN represent our second reference because ERA5 is a reanalysis and REGEN ends in 2016. We combine the temperature and precipitation observational/reanalysis datasets to compute one of the two drought indices which is based on precipitation and potential evapotranspiration (mm) (see section 3.1). Since the dry extreme indices are computed from accumulated periods, we remove the year 1960 and base all our results within the 1961-2019 period.

To better understand some characteristics of how the variability constraint can improve near-term projections regionally, we also focus our analysis on four different regions located on four different continents. These regions are western north America (WN America, 25°N-45°N, 125°W-95°W), southern central and eastern Europe (SCE Europe, 35°N-55°N, 5°E-35°E), southeastern China (SE China, 20°N-40°N, 95°E-125°E) and southeastern Australia (SE Australia, 45°S-25°S, 135°E-155°E) (Figure S1). For all the regions we mask the oceans and consider only land grid-points. We choose these regions because they are the ones where the constraining method shows added skill over the raw CMIP6 ensemble, and to understand time-series characteristics that contribute to the skill in the constrained ensemble.

3. Methods

3.1 Extreme indices

We compute six extreme indices as measures for global hot, cold and dry extremes over land areas similarly to De Luca and Donat (2023). For hot extremes we calculate two ETCCDI indices (Zhang *et al* 2011), namely the percentage of days when daily maximum temperature exceeds the 90th percentile (*TX90p*) and the annual maximum value of daily maximum temperature (*TXx*). For cold extremes we also use two ETCCDI indices: the percentage of days when daily minimum temperature is below the 10th percentile (*TN10p*) and the annual minimum value of daily minimum temperature (*TNn*). These ETCCDI indices are computed using the R package *"climdex.pcic.ncdf"* (https://github.com/ARCCSS-extremes/climdex.pcic.ncdf). To quantify dry extremes we use the Standardized Precipitation Index (SPI, McKee *et al* (1993)) and the Standardized Precipitation Evapotranspiration Index (SPEI, Vicente-Serrano *et al* (2010)) with accumulation periods of 3-, 6- and 12-months.

The SPI is computed solely from monthly total precipitation and it is often used to measure meteorological drought, with lack of precipitation indicated by negative values. On the other hand, the SPEI is computed from monthly total precipitation and monthly mean of daily maximum and minimum temperatures , the last two used to compute potential evapotranspiration following the Hargreaves (1994) approximation; SPEI therefore represents drought in terms of

Accepted Manuscript

lack of water availability. We use the entire investigation period as baseline for the estimation of the distribution parameters (De Luca and Donat 2023, Vicente-Serrano *et al* 2020), i.e. 60 years (1960-2019) for the CMIP6 MME and BEST-GPCC datasets, and 57 years (1960-2016) for ERA5-REGEN. Since the SPI and SPEI indices do not directly indicate drought occurrences, we select from these indices only monthly values ≤ -1 which represent moderately dry conditions. We use -1 as threshold to make sure that a sufficient number of monthly values in the SPI and SPEI drought datasets are available. Our drought indices count the number of dry months per year and we named them *SPIn_dry* and *SPEIn_dry*, where *n* stands for the accumulation period of the index (i.e. 3, 6 and 12 months) (De Luca and Donat 2023). The SPI and SPEI indices are computed using the R package *"SPEI"* (Beguería *et al* 2014, Vicente-Serrano *et al* 2010).

We calculate all the indices on the native CMIP6 model, BEST, GPCC, ERA5 and REGEN grids and then re-grid them to a common latitude-longitude grid of 2.8°x2.8° (the resolution of the model with the coarsest resolution included in this study, CanESM5) to facilitate multi-model analysis. We then remove the ocean grid-points with a land-sea mask so that only land values are retained and exclude Antarctica.

3.2 Constraining internal climate variability

We follow the approach introduced by Mahmood *et al* (2022) to constrain the large MME of CMIP6 simulations. For this we used observational SST data from the Extended Reconstructed Sea Surface Temperature version 5 dataset (ERSSTv5; Huang *et al* (2017)) from the National Oceanic and Atmospheric Administration (NOAA). The monthly mean model and observed SST data were regridded to a common 3°x3° grid and the climatological mean (1981-2010) was removed to compute the anomalies.

Internal climate variability is constrained by comparing spatial distributions of global SST anomaly patterns between each of the 149 CMIP6 ensemble members and the observed anomaly averaged over a 9-year period preceding the start of the prediction. Such comparison is performed via area-weighted spatial pattern correlation. Similar to Mahmood *et al* (2022), we choose the top ranking 30 members (referred to as "Best30") for hindcasting up to 20 years after the initialization. The unconstrained ensemble consists of all 149 members (referred to as "All ensemble").

We use 9-year averages of SST anomalies since constraining based on this period showed high skill in constrained projections as shown by Mahmood *et al* (2022), who also tested sensitivity to using other averaging periods. To start a constrained prediction from January 1961, we use the 9-year mean SST anomalies from January 1952 to December 1960 to select the Best30 members. Such a procedure is repeated every year and the Best30 members selected based on SST anomaly comparison from 1953 to 1961 are used for predictions starting in 1962, 1954-1962 for predictions starting in 1963, etc. Here we focus on the hindcast period of 1 to 20 years after the

Accepted Manuscript

initialization. To evaluate the 20-year mean hindcasts against observational data sets, the final constraining period considered goes from January 1991 to December 1999 for predicting January 2000 to December 2019. Therefore, we use a total of 40 start dates for the retrospective predictions.

3.3 Evaluation metrics

We use a set of metrics that evaluate different aspects of the degree of agreement between the simulations and observations (e.g. Donat *et al* 2023, Mahmood *et al* 2021, 2022, Delgado-Torres *et al* 2022, 2023).

The Spearman Correlation Coefficient (Spearman 1904) estimates the linear relationship between the observational reference and the CMIP6 MME mean. It ranges between -1 (worst agreement) and 1 (best agreement). We use the Spearman rank correlation to avoid assumptions about distributional properties (e.g. normality). The Spearman correlation coefficient is defined as:

$$r = \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \qquad (eq.1)$$

where i corresponds to each time step (from 1 to n), and d_i is the difference between the ranks of x_i and o_i (simulated and observed value, respectively, for time step i).

In order to assess whether the Best30 ensemble captures more observed variability than the All ensemble, we use the residual correlation (Smith *et al* 2019, Mahmood *et al* 2022) using the Spearman's test (Corder and Foreman 2014). The residual correlation measures to what extent we can predict the variations around the forced signal and it therefore quantifies the added skill from aligning variability phases or "initialising" the predictions. We therefore remove the forced signal (using the All ensemble mean as best estimate of the forcing response) from the observed and Best30 mean time-series by subtracting their corresponding linear fits with the All ensemble mean (Smith *et al* 2019). This results in time-series of observed and Best30 residuals. The residual correlation is the correlation between the observed and Best30 residuals. Positive values of the residual correlation indicate that the Best30 ensemble captures some observed variability around the forced signal derived from the All ensemble mean, and negative values indicate the that the observed and predicted residuals are not in phase. The Spearman's correlations and residual correlations are computed from a total of forty 20-year averages, starting each year from 1961 to 2000.

The Root Mean Squared Skill Score (RMSSS; Murphy (1988)) is also a deterministic skill measure computed from the MME mean and is used to assess whether the Best30 ensemble is more skillful than a reference hindcast. The RMSSS is based on the Root Mean Squared Error

Accepted Manuscript

(RMSE), which quantifies the agreement in terms of the error magnitude between the ensemble mean and the observational reference. For quantifying the RMSSS we compute the RMSE for the Best30 ensemble and the reference hindcast using 20-year averages with starting years ranging from 1961 to 2000. The reference hindcasts used to compute the RMSSS are the climatological hindcasts (i.e. no anomaly) for assessing the Best30 skill, and the All ensemble mean hindcasts for quantifying the added value in Best30 over the All ensemble. Positive RMSSS values indicate that the Best30 ensemble is more skillful than the reference hindcast and negative values indicate it is less skillful than the reference hindcast. The RMSSS is defined as:

$$RMSSS = 1 - \frac{RMS_{exp}}{RMS_{ref}} \qquad (eq.2)$$

where RMS_{exp} and RMS_{ref} correspond to the Root Mean Square (RMS) difference of the hindcasts and reference hindcast, respectively, from the observed value o_i, which is computed as:

$$RMS = \sqrt{\sum_{i=1}^{n} \frac{(x_i - o_i)^2}{n}} \qquad (eq.3)$$

The Ranked Probability Skill Score (RPSS; Wilks (2011)) is used to estimate the skill of probabilistic products from all members of the MME. The RPSS is based on the Ranked Probability Score (RPS) which evaluates the skill in terms of probabilities (computed as the percentage of members that fall into each equiprobable tercile category, with the three categories indicating below average, approximately average and above average conditions). For computing the RPSS, we first compute the RPS for each 20-year average with starting years from 1961 to 2000 and then quantify the temporal mean of these averages. As with the RMSSS, positive RPSS values indicate that the Best30 ensemble outperforms the reference hindcast, while negative RPSS values indicate that the reference hindcast is more skillful. The probabilistic climatological hindcast (defined as the same probability for all tercile categories, i.e., 33.3%) and the All ensemble are used as reference hindcasts. The RPSS is defined as:

$$RPSS = 1 - \frac{mean(RPS_{exp})}{mean(RPS_{ref})} \qquad (eq.4)$$

where RPS_{exp} and RPS_{ref} correspond to the RPS for each time step of the hindcasts and reference hindcast, respectively, which is computed as:

Accepted Manuscript

$$RPS = \sum_{m=1}^{J} \left[\left(\sum_{j=1}^{m} p_{xj} \right) - \left(\sum_{j=1}^{m} p_{oj} \right) \right]^2 \qquad (eq. 5)$$

where j corresponds to the probabilistic category (from 1 to J=3), and p_{xj} and p_{oj} are the hindcasted and observed probabilities, respectively, for the probabilistic category j.

We estimate the statistical significance of the correlation and residual correlation with a two-sided t-test (Wilks 2011) accounting for the time-series auto-correlation following (Zwiers and von Storch 1995) to assess whether the skill values are significantly different from zero. To assess the statistical significance of the RMSSS and RPSS, we apply a two-sided Random Walk test (DelSole and Tippett 2016) to the RMSE and RPSS time-series to assess whether the number of times that the Best30 ensemble is better or worse than the reference hindcast is statistically significant. To the p-values obtained with the two-sided t-test and Random Walk test we apply the False Detection Rate (FDR; Wilks (2016)) procedure using alpha_FDR = 0.1 to control the type I errors (or false positives).

4. Results

4.1 Hot and cold extremes

Best30 *TX90p* shows high skill in most global land regions, with correlations exceeding 0.9 in the majority of grid cells, and RMSSS > 0.8 and RPSS > 0.6 in large areas, respectively (Figure 1(a)-(c)). Improved skill from the constraint in Best30 in comparison to All ensemble as measured by positive residual correlations is found in the western USA, South and eastern North America, Africa, the Arabian Peninsula, Europe, most of Asia and northern Australia (Figure 1(d)), meaning that in these regions observed variability is captured better by Best30 than by All ensemble. Improved skill based on the RMSSS is found over central and northern South America, Greenland, most of the African continent, southeastern Europe, the Arabic Peninsula and most of central and southern Asia (Figure 1e), pointing out a good agreement between Best30 and the reference dataset. Improved skill measured by the RPSS is widespread and similar to the one of residual correlation (Figure 1(f)) and indicates that Best30 is more skillful than All ensemble when evaluating the skill in terms of probabilities.

Best30 *TXx* shows often weaker skill compared to *TX90p*, as also found for multi-annual predictions by Delgado-Torres *et al* (2023), but 20-year projections are still skillful over large areas of the globe for the three metrics. Lack of skill is found in some parts of North and South America, Scandinavia, western and southern Africa, central parts of Asia and northern Australia (Figure 1(g)-(i)). Improved skill as measured by residual correlation is found over Alaska, Canada, eastern North America, southwestern USA, Mexico, northern South America, eastern Europe, India, eastern Russia, southeastern Asia and western Australia (Figure 1(j)). RMSSS shows improved skill mainly over central Africa (Figure 1(k)), whereas the negative RMSSS

Accepted Manuscript

values in other regions (such as large parts of North and South America) are indicative of an increased mean bias in Best30 compared to All ensemble. RPSS improved skill is found in western North America, South America, central and eastern Europe, central Africa and in some localised parts of Asia (Figure 1(l)).

Similarly to hot extremes, we find high hindcast skill also for indices of cold extremes (Figure S2). *TN10p* and *TNn* show high Best30 skill over most of the globe, with the former having larger areas with significant skill than the latter (Figure S2(a)-(c), (g)-(i)). For both indices, we find added skill compared to All ensemble over southeastern North America, eastern Brazil, equatorial Africa, southeastern China and northern Australia (Figure S2(d)-(f), (j)-(l)). When using ERA5 as reference datasets we find similar spatial patterns for hot and cold extremes in both the Best30 skill and skill improvement (Figures S3-S4).

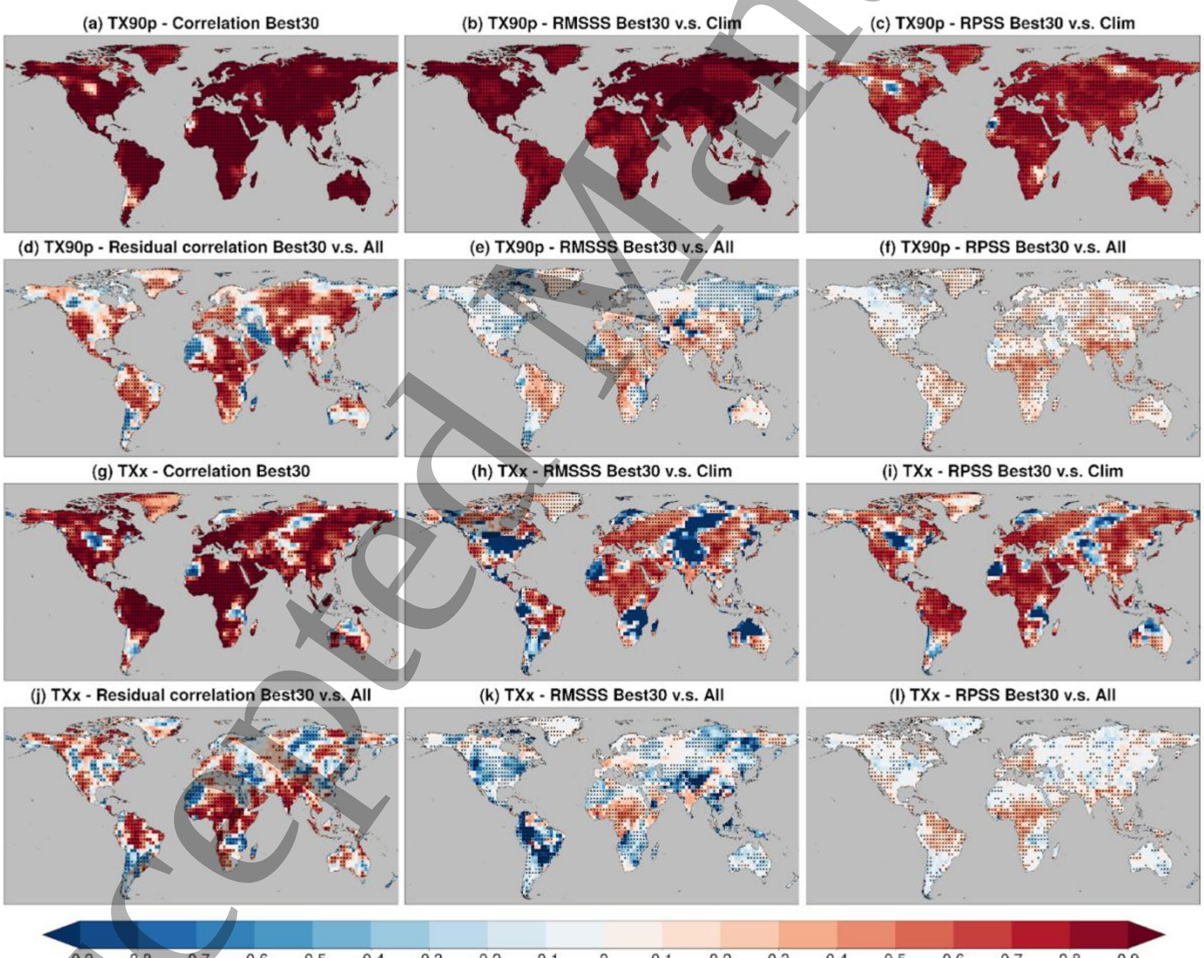


Figure 1 Skill measures obtained with the Best30 ensemble for *TX90p* (first row) and *TXx* (third row), and added skill of the Best30 ensemble in comparison to the All ensemble for *TX90p* (second row) and *TXx* (fourth row). The first column shows the correlation between the Best30 ensemble mean and observations (a, g) and the correlation between the residuals of the Best30 ensemble mean and

Accepted Manuscript

observations calculated by linearly regressing out the All ensemble mean (d, j). The second column shows the RMSSS of Best30 using the climatological hindcast (b, h) and the All ensemble (e, k) as the reference hindcast. The third column is similar to the second column, but for the RPSS. Stippling indicates grid points where the skill measures are statistically significant controlling the FDR with alpha_FDR = 0.1. The observational reference dataset is BEST.

We next inspect the regional average time series for three regions in which the Best30 ensemble shows improved skill over the All ensemble, namely western North America, southern central and eastern Europe, and southeastern China (Figure 2). We use these time series plots to illustrate some of the characteristics that help explain the improved skill in Best30 compared to the unconstrained ensemble. While all time series indicate a long-term warming over the analysis period for both *TX90p* and *TXx* in all three regions, there are also some noteworthy differences.

In SCE Europe and SE China (Figure S1) the Best30 ensemble mean has lower values than the All ensemble mean for both *TX90p* and *TXx* in the first two decades of the investigation period. These values are closer to the observed temperature values, contributing to the improved skill. Overall this leads to a stronger long-term warming of hot extremes in these regions in Best30 compared to All, and more similar to observations. In addition, The Best30 ensemble also captures some of the observed decadal-scale variations with accelerated warming in the 1980s and early 1990s and reduced warming rates from the mid 1990s, whereas the All ensemble mean features temporally more homogeneous increases. In WN America (Figure S1) the Best30 ensemble mean also has lower values than the All ensemble mean during the first two decades of the investigation period. In this case this makes it more different to the observed time series, as also reflected by the negative RMSSS values when using the All ensemble as reference. However, the positive Residual Correlation (and positive RPSS for the *TXx* index) indicate some added skill in Best30 over the All ensemble, and this is indicative of correctly predicting some aspects of the decadal-scale variations in the warming rates (such as the reduced warming rates in the 1990s). Similar time-series are also found when using the ERA5 reference datasets as shown in Figure S5.

Accepted Manuscript

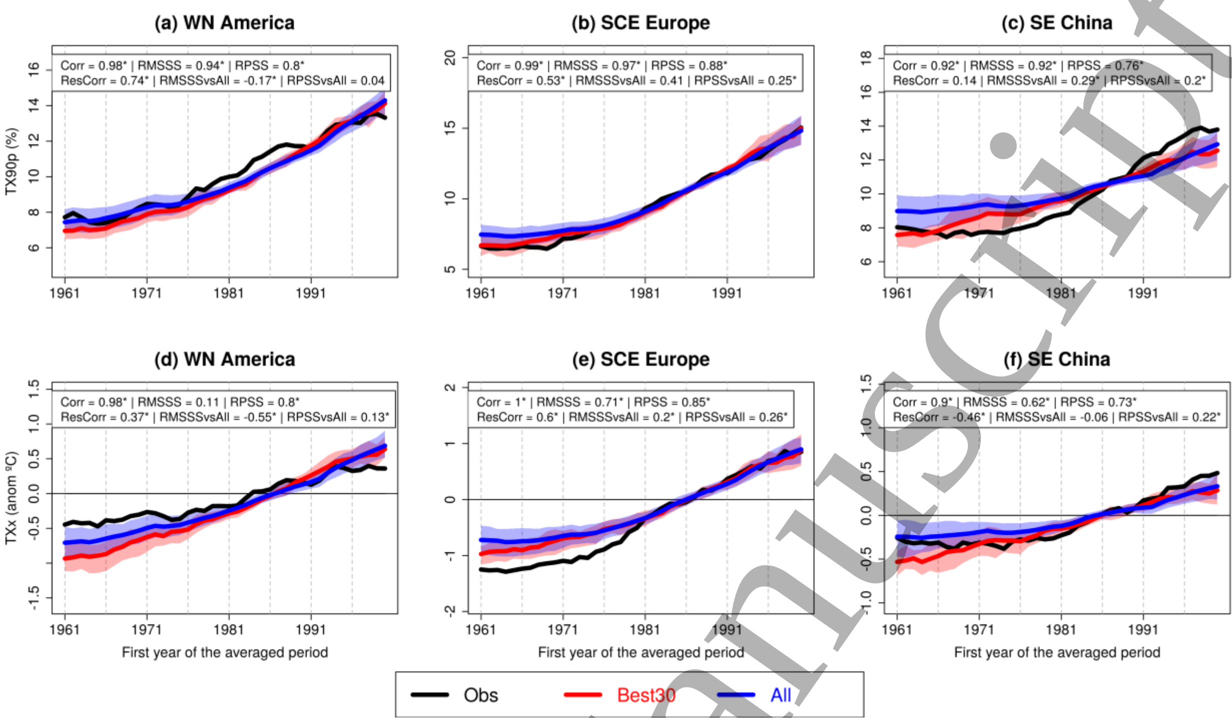


Figure 2 Regional 20-year average time-series of hot extreme indices from 1961 to 2000 initialised years for Best30 (red), All ensemble (blue) and observational (black) datasets. Regions are western North America (WN America), southern central and eastern Europe (SCE Europe) and southeastern China (SE China). Shaded coloured bands represent the interquartile range (25th and 75th percentiles) of the Best30 and All ensemble. We also show the evaluation metrics averaged over each single region, in the same order as Figure 1(a)-(f). Asterisks indicate metrics statistically significant. The observational reference dataset used is BEST.

4.2 Dry extremes

Skill for dry extremes is overall spatially more limited when compared to the skill for hot extremes. However, there are some areas where near-term projections of dry extremes are skillful, and where our constraint adds skill.

Best30 *SPI3_dry* correlations are locally significant over the southwestern USA, central and southern South America, Greenland, northern Europe, central Africa, parts of Asia and southeastern Australia (Figure 3(a)). Whereas RMSSS and RPSS show similar patterns of positive skill over central South America, northern Europe, central Africa and central and northern Asia (Figure 3(b),(c)). Residual correlations indicate skill improvements from the constraint for *SPI3_dry* in a few regions, e.g. over the southern USA, central Africa and in other localised areas of the globe (Figure 3(d)). Also RMSSS and RPSS indicate some added value for the constrained ensemble in similar regions, e.g. the southern USA, some scattered areas in South America, the Arabian Peninsula and southern Australia (Figure 3(e)-(f)).

Accepted Manuscript

Best30 *SPEI3_dry* shows significant skill based on the three metrics over southwestern USA, central and northern Mexico, central and southern South America, northern and southern Africa, the Iberian peninsula, southeastern Europe, the Middle East, western and central Asia and southeastern Australia (Figure 3(g)-(i)). Residual correlations indicate improved skill over the western USA, northern South America, the Balkans, parts of central and southern Africa, the Arabian Peninsula and southeastern Australia (Figure 3(j)). Also here RMSSS and RPSS indicate some skill improvements for *SPEI3_dry* in the southwestern USA and northern Mexico, parts of South America and Africa, the Arabian Peninsula and in a few areas of central Asia (Figure 3(k)-(l)).

Overall similar results are obtained when considering the drought indices with longer accumulation periods, such as *SPI6_dry*, *SPEI6_dry* (Figure S6), *SPI12_dry* and *SPEI12_dry* (Figure S7) and different reference datasets (i.e. ERA5-REGEN, Figures S8-S10).

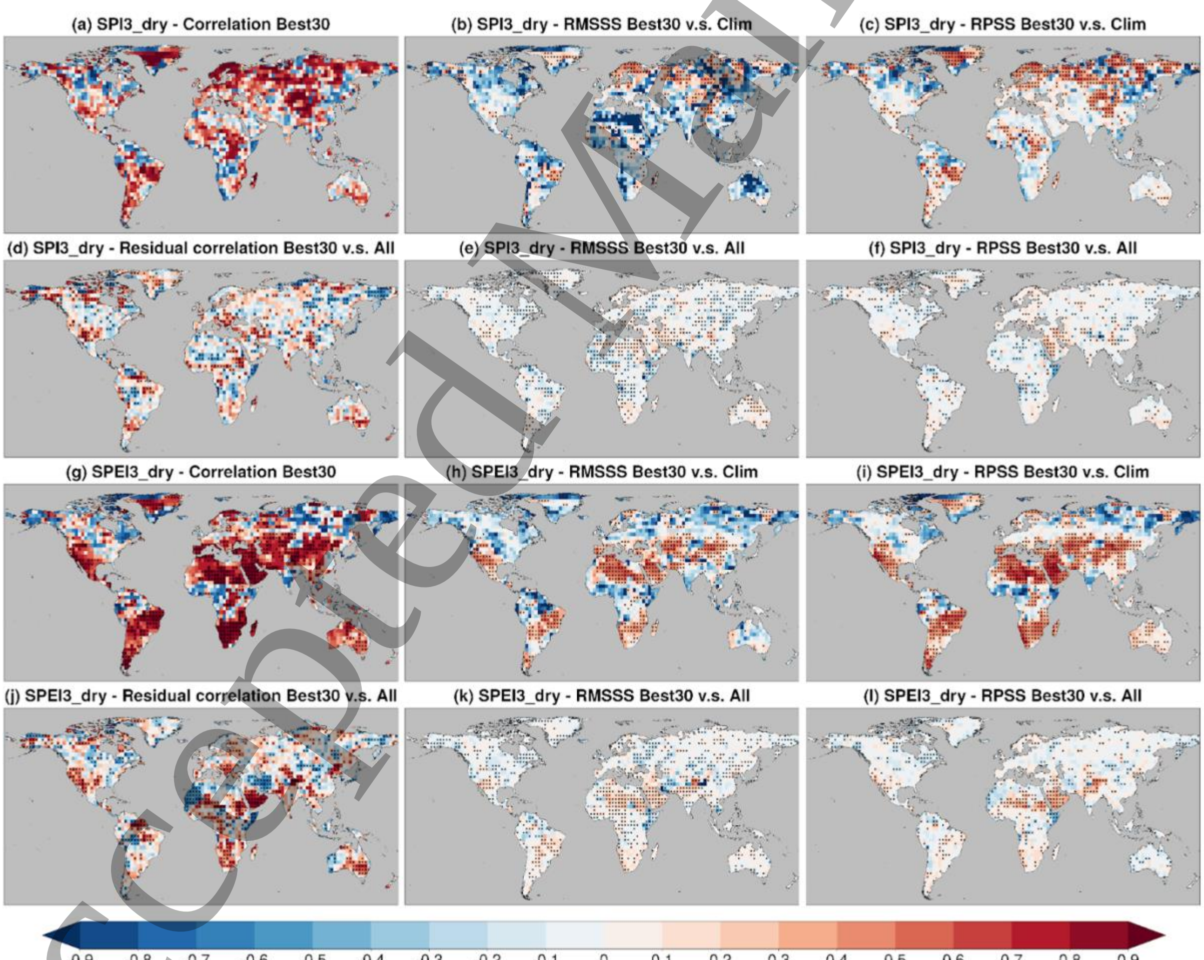
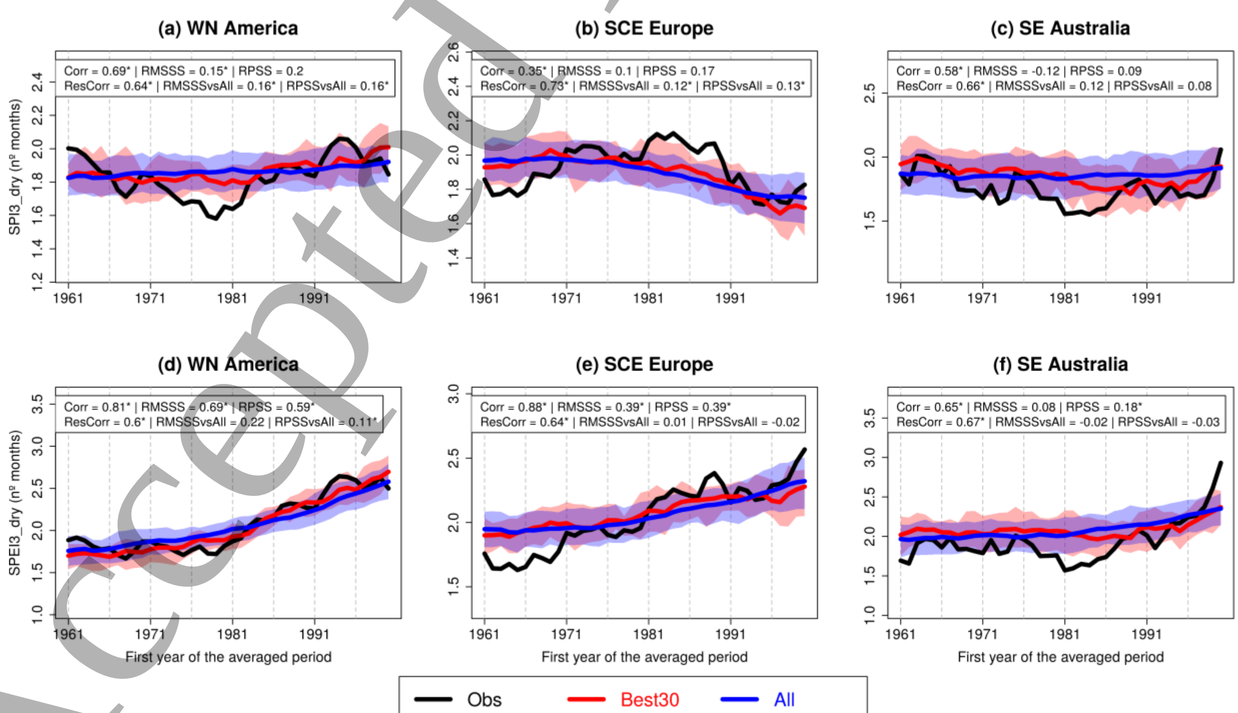


Figure 3 Same as Figure 1 but for *SPI3_dry* and *SPEI3_dry*. The observational reference datasets used are GPCC (precipitation) and BEST (to compute potential evapotranspiration).

Accepted Manuscript

In the following, we focus on regional timeseries of the drought measures in three regions where the constraint adds skill (i.e. WN America, SCE Europe and SE Australia), to better understand the characteristics of the improved hindcasts (Figure 4). Here, Best30 and All ensemble correctly capture both the stationarity (Figure 4(a),(c)) and long-term changes in the observations (Figure 4(b),(d)-(f)) for both indices. There is also added value in Best30 compared to All, especially for WN America where Best30 captures some of the observed decadal-scale variations around the CMIP6 (All) mean, although with smaller magnitude. We obtain similar results with *SPI6_dry*, *SPEI6_dry*, *SPI12_dry* and *SPEI12_dry* (Figure S11), or when using the ERA5-REGEN reference datasets (Figures S12-S13). In summary, these results illustrate how the constraint can improve near-term projections of drought, by enhancing the representation of both decadal-scale variations and long-term changes in WN America, SCE Europe and SE Australia.

When comparing the skill for these drought indices (i.e. *SPIn_dry* or *SPEIn_dry*) against the skill in predicting the entire distributions of *SPI* or *SPEI* (i.e. including dry and wet conditions), we note some interesting differences (Figure S14). While for most regions the patterns of skill are reasonably similar between predicting *SPI*/*SPEI* and the corresponding drought indices, in particular in SE Australia (where both *SPI3_dry* and *SPEI3_dry* had skill and added skill), there is no skill (nor added skill) for *SPI3* or *SPEI3*. This indicates some asymmetry in the predictability of accumulated precipitation (*SPI*) or water availability (*SPEI*), with dry conditions being more predictable than wet conditions.



Accepted Manuscript

Figure 4 As Figure 2 but for *SPI3_dry* and *SPEI3_dry* and southeastern Australia (SE Australia) instead of southeastern China (SE China).

5. Discussion and Conclusions

In this work we presented the first evaluation of multi-decadal prediction skill of CMIP6 projections of hot, cold and dry extremes in global land regions with decadal variability constrained based on observations. We performed our analysis within the 1961-2019 period with 20-year predictions started each year from 1961 to 2000 (i.e. generating retrospective predictions of 20-year windows ranging from 1961-1980 to 2000-2019). We showed that the constrained ensemble (Best30) has high skill for hot and cold extremes over large parts of the globe, with also added value compared to the unconstrained ensemble (All) in several regions. Dry extremes, on the other hand, showed lower skill compared to temperature extremes but drought predictions are skillful in some regions. These regions include e.g. western North America, Southeastern Europe and Southeastern Australia, which were affected by prominent dry and hot extremes in recent decades.

This work builds on recent studies, which investigated the predictability of extremes in multi-annual predictions from the DCPP MME against observations for temperature extremes (Delgado-Torres *et al* 2023) and on recent work developing the methods to constrain variability in large projection ensembles with the goal to provide multi-decadal climate predictions (Mahmood *et al* 2022). The former study showed high skill in predicting average and extreme temperatures with DCPP when compared to observations, and added value when compared to an historical CMIP6 MME. The latter investigation, on the other hand, showed high skill in average temperature for Best30 compared to observations and added value when this is compared to the All ensemble. These studies reflect our findings, but their geographical patterns of DCPP and Best30 added value against the Historical and unconstrained ensemble respectively do not necessarily reflect our maps, since for example we obtained more positive and significant skill in *TX90p* and *TXx* than Delgado-Torres *et al* (2023), especially in western north America, central south America, central Africa, southern central and eastern Europe, India and southeastern China. Similarly to our results, Delgado-Torres *et al* (2023) found higher skill for *TX90p* than *TXx*. This is because the former is a more moderate extreme occurring several days in a year and for which modulation related to climate variability modes more detectable, whereas the latter represents only the one most extreme day per year whose is intensity is affected by different processes (e.g. specific atmospheric circulation patterns on that day), which may not be predictable with our method.

We envisage future work on assessing the multi-decadal prediction skill of other impact-relevant climate phenomena, such as compound hot-dry (e.g. De Luca and Donat 2023, Bevacqua *et al* 2022) and wet-windy (e.g. De Luca *et al* 2020, Martius *et al* 2016) extremes derived from a large MME of CMIP6 models. In addition, identifying the sources of predictability driving good skill

Accepted Manuscript

in selected regions of the globe, as done for example by Patricola *et al* (2020) and Imada and Kawase (2021), can further extend the understanding of the physical processes at play and improve the prediction. In particular, applying the constraint only to specific ocean regions can help to attribute the predictability to specific modes of variability or climate system components (e.g. Mahmood *et al* 2022).

Our work demonstrates that constraining internal climate variability with observations, leads to more trustworthy predictions of hot and dry extremes on multi-decadal time-scales, and we believe that such predictions can be useful for stakeholders to develop targeted adaptation strategies to climate change over the next 20 years.

Acknowledgments

This research has been partly supported by the Horizon2020 LANDMARC project (grant agreement No. 869367) and the Horizon Europe ASPECT project (grant number 101081460). PDL has received funding from the European Union’s Horizon Europe Research and Innovation Programme under grant agreement No 101059659. CDT acknowledges financial support from the Spanish Ministry for Science and Innovation (FPI PRE2019-509 08864 financed by MCIN/AEI/10.13039/501100011033 and by FSE invierte en tu futuro). MGD is grateful for support by the AXA Research Fund. The authors are further grateful for the support by the Department of Research and Universities of the Government of Catalonia to the Climate Variability and Change Research Group (Code: 2021 SGR 00786).

References

Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Schamm K, Schneider U and Ziese M 2013 A description of the global land-surface precipitation data products of the Global Precipitation Climatology Centre with sample applications including centennial (trend) analysis from 1901–present *Earth Syst. Sci. Data* **5** 71–99 Online: https://essd.copernicus.org/articles/5/71/2013/

Befort D J, O’Reilly C H and Weisheimer A 2020 Constraining Projections Using Decadal Predictions *Geophys. Res. Lett.* **47** e2020GL087900 Online: https://doi.org/10.1029/2020GL087900

Beguería S, Vicente-Serrano S M, Reig F and Latorre B 2014 Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring *Int. J. Climatol.* **34** 3001–23 Online: https://doi.org/10.1002/joc.3887

Bevacqua E, Zappa G, Lehner F and Zscheischler J 2022 Precipitation trends determine future occurrences of compound hot–dry events *Nat. Clim. Chang.* **12** 350–5 Online: https://doi.org/10.1038/s41558-022-01309-5

Bilbao R, Wild S, Ortega P, Acosta-Navarro J, Arsouze T, Bretonnière P-A, Caron L-P, Castrillo M, Cruz-Garc\'\ia R, Cvijanovic I, Doblas-Reyes F J, Donat M, Dutra E, Echevarr\'\ia P, Ho A-C, Loosveldt-Tomas S, Moreno-Chamarro E, Pérez-Zanon N, Ramos A, Ruprich-Robert Y, Sicardi V, Tourigny E and Vegas-Regidor J 2021 Assessment of a full-field

Accepted Manuscript

initialized decadal climate prediction system with the CMIP6 version of EC-Earth *Earth Syst. Dyn.* **12** 173–96 Online: https://esd.copernicus.org/articles/12/173/2021/

Blauhut V, Gudmundsson L and Stahl K 2015 Towards pan-European drought risk maps: quantifying the link between drought indices and reported drought impacts *Environ. Res. Lett.* **10** 14008 Online: https://dx.doi.org/10.1088/1748-9326/10/1/014008

Blauhut V, Stahl K, Stagge J H, Tallaksen L M, De Stefano L and Vogt J 2016 Estimating drought risk across Europe from reported drought impacts, drought indices, and vulnerability factors *Hydrol. Earth Syst. Sci.* **20** 2779–800 Online: https://hess.copernicus.org/articles/20/2779/2016/

Boer G J, Smith D M, Cassou C, Doblas-Reyes F, Danabasoglu G, Kirtman B, Kushnir Y, Kimoto M, Meehl G A, Msadek R, Mueller W A, Taylor K E, Zwiers F, Rixen M, Ruprich-Robert Y and Eade R 2016 The Decadal Climate Prediction Project (DCPP) contribution to CMIP6 *Geosci. Model Dev.* **9** 3751–77 Online: https://gmd.copernicus.org/articles/9/3751/2016/

Brás T A, Seixas J, Carvalhais N and Jägermeyr J 2021 Severity of drought and heatwave crop losses tripled over the last five decades in Europe *Environ. Res. Lett.* **16** 65012 Online: https://dx.doi.org/10.1088/1748-9326/abf004

Cheng J, Xu Z, Bambrick H, Su H, Tong S and Hu W 2019 Impacts of heat, cold, and temperature variability on mortality in Australia, 2000–2009 *Sci. Total Environ.* **651** 2558–65 Online: https://www.sciencedirect.com/science/article/pii/S0048969718340774

Contractor S, Donat M G, Alexander L V, Ziese M, Meyer-Christoffer A, Schneider U, Rustemeier E, Becker A, Durre I and Vose R S 2020 Rainfall Estimates on a Gridded Network (REGEN) – a global land-based gridded dataset of daily precipitation from 1950 to 2016 *Hydrol. Earth Syst. Sci.* **24** 919–43 Online: https://hess.copernicus.org/articles/24/919/2020/

Cook B I, Mankin J S and Anchukaitis K J 2018 Climate Change and Drought: From Past to Future *Curr. Clim. Chang. Reports* **4** 164–79 Online: https://doi.org/10.1007/s40641-018-0093-2

Corder G W and Foreman D I 2014 *Nonparametric Statistics: A Step-by-Step Approach* (Wiley)

Coumou D and Robinson A 2013 Historic and future increase in the global land area affected by monthly heat extremes *Environ. Res. Lett.* **8** 0–6

Dai A 2011 Drought under global warming: A review *Wiley Interdiscip. Rev. Clim. Chang.* **2** 45–65

Dai A 2013 Increasing drought under global warming in observations and models *Nat. Clim. Chang.* **3** 52–8 Online: https://doi.org/10.1038/nclimate1633

Dai A, Fyfe J C, Xie S-P and Dai X 2015 Decadal modulation of global surface temperature by internal climate variability *Nat. Clim. Chang.* **5** 555–9 Online: https://doi.org/10.1038/nclimate2605

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, Paxian A, Pérez-Zanón N, Cabré M S, Solaraju-Murali B, Soret A and Doblas-Reyes F J 2022 Multi-Model Forecast Quality Assessment of CMIP6 Decadal Predictions *J. Clim.* **35** 4363–82 Online: https://journals.ametsoc.org/view/journals/clim/35/13/JCLI-D-21-0811.1.xml

Delgado-Torres C, Donat M G, Soret A, González-Reviriego N, Bretonnière P-A, Ho A-C, Pérez-Zanón N, Samsó Cabré M and Doblas-Reyes F J 2023 Multi-annual predictions of the frequency and intensity of daily temperature and precipitation extremes *Environ. Res. Lett.*

Accepted Manuscript

18 34031 Online: https://dx.doi.org/10.1088/1748-9326/acbbe1

DelSole T and Tippett M K 2016 Forecast Comparison Based on Random Walks *Mon. Weather Rev.* **144** 615–26 Online: https://journals.ametsoc.org/view/journals/mwre/144/2/mwr-d-15-0218.1.xml

Donat M G, Delgado-Torres C, De Luca P, Mahmood R, Ortega P and Doblas-Reyes F J 2023 How Credibly Do CMIP6 Simulations Capture Historical Mean and Extreme Precipitation Changes? *Geophys. Res. Lett.* **50** e2022GL102466 Online: https://doi.org/10.1029/2022GL102466

Ebi K L, Capon A, Berry P, Broderick C, de Dear R, Havenith G, Honda Y, Kovats R S, Ma W, Malik A, Morris N B, Nybo L, Seneviratne S I, Vanos J and Jay O 2021 Hot weather and heat extremes: health risks *Lancet* **398** 698–708 Online: https://www.sciencedirect.com/science/article/pii/S0140673621012083

Eyring V, Bony S, Meehl G A, Senior C A, Stevens B, Stouffer R J and Taylor K E 2016 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization *Geosci. Model Dev.* **9** 1937–58

Fischer E M, Beyerle U and Knutti R 2013 Robust spatially aggregated projections of climate extremes *Nat. Clim. Chang.* **3** 1033

Fischer E M and Schär C 2010 Consistent geographical patterns of changes in high-impact European heatwaves *Nat. Geosci.* **3** 398

García-León D, Casanueva A, Standardi G, Burgstall A, Flouris A D and Nybo L 2021 Current and projected regional economic impacts of heatwaves in Europe *Nat. Commun.* **12** 5807 Online: https://doi.org/10.1038/s41467-021-26050-z

Guirguis K, Gershunov A, Schwartz R and Bennett S 2011 Recent warm and cold daily winter temperature extremes in the Northern Hemisphere *Geophys. Res. Lett.* **38** Online: https://doi.org/10.1029/2011GL048762

Hargreaves G H 1994 Defining and Using Reference Evapotranspiration *J. Irrig. Drain. Eng.* **120** 1132–9 Online: https://doi.org/10.1061/(ASCE)0733-9437(1994)120:6(1132)

Hawkins E and Sutton R 2009 The Potential to Narrow Uncertainty in Regional Climate Predictions *Bull. Am. Meteorol. Soc.* **90** 1095–108 Online: https://journals.ametsoc.org/view/journals/bams/90/8/2009bams2607_1.xml

Hazeleger W, Wouters B, van Oldenborgh G J, Corti S, Palmer T, Smith D, Dunstone N, Kröger J, Pohlmann H and von Storch J-S 2013 Predicting multiyear North Atlantic Ocean variability *J. Geophys. Res. Ocean.* **118** 1087–98 Online: https://doi.org/10.1002/jgrc.20117

Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, Simmons A, Soci C, Abdalla S, Abellan X, Balsamo G, Bechtold P, Biavati G, Bidlot J, Bonavita M, De Chiara G, Dahlgren P, Dee D, Diamantakis M, Dragani R, Flemming J, Forbes R, Fuentes M, Geer A, Haimberger L, Healy S, Hogan R J, Hólm E, Janisková M, Keeley S, Laloyaux P, Lopez P, Lupu C, Radnoti G, de Rosnay P, Rozum I, Vamborg F, Villaume S and Thépaut J-N 2020 The ERA5 global reanalysis *Q. J. R. Meteorol. Soc.* **146** 1999–2049 Online: https://doi.org/10.1002/qj.3803

Huang B, Thorne P W, Banzon V F, Boyer T, Chepurin G, Lawrimore J H, Menne M J, Smith T M, Vose R S and Zhang H-M 2017 Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons *J. Clim.* **30** 8179–205 Online: https://journals.ametsoc.org/view/journals/clim/30/20/jcli-d-16-0836.1.xml

Imada Y and Kawase H 2021 Potential Seasonal Predictability of the Risk of Local Rainfall Extremes Estimated Using High-Resolution Large Ensemble Simulations *Geophys. Res.*

Accepted Manuscript

Lett. **48** e2021GL096236 Online: https://doi.org/10.1029/2021GL096236

Lehner F, Deser C, Maher N, Marotzke J, Fischer E M, Brunner L, Knutti R and Hawkins E 2020 Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6 *Earth Syst. Dyn.* **11** 491–508 Online: https://esd.copernicus.org/articles/11/491/2020/

De Luca P and Donat M 2023 Projected changes in hot, dry and compound hot-dry extremes over global land regions *Geophys. Res. Lett.* **50**

De Luca P, Messori G, Pons F M E and Faranda D 2020 Dynamical Systems Theory Sheds New Light on Compound Climate Extremes in Europe and Eastern North America *Q. J. R. Meteorol. Soc.* 1636–50 Online: https://doi.org/10.1002/qj.3757

Mahmood R, Donat M G, Ortega P, Doblas-Reyes F J, Delgado-Torres C, Samsó M and Bretonnière P-A 2022 Constraining low-frequency variability in climate projections to predict climate on decadal to multi-decadal timescales – a poor man's initialized prediction system *Earth Syst. Dynam.* **13** 1437–50 Online: https://esd.copernicus.org/articles/13/1437/2022/

Mahmood R, Donat M G, Ortega P, Doblas-Reyes F J and Ruprich-Robert Y 2021 Constraining decadal variability yields skillful projections of near-term climate change *Geophys. Res. Lett.* **48** e2021GL094915

Mann M E, Steinman B A and Miller S K 2014 On forced temperature changes, internal variability, and the AMO *Geophys. Res. Lett.* **41** 3211–9 Online: https://doi.org/10.1002/2014GL059233

Martius O, Pfahl S and Chevalier C 2016 A global quantification of compound precipitation and wind extremes *Geophys. Res. Lett.* **43** 7709–17

Masson-Delmotte V, Zhai P, Pirani A, Connors S L, Péan C, Berger S, Caud N, Chen Y, Goldfarb L, Gomis M I, Huang M, Leitzell K, Lonnoy E, Matthews J B R, Maycock T K, Waterfield T, Yelekçi O, Yu R and Zhou B 2021 *IPCC, 2021: Summary for Policymakers. In: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA)

McKee T, Doesken N and Kleist J 1993 The relationship of drought frequency and duration to time scales *AMS 8th Conference on Applied Climatology* (Anaheim: AMS 8th Conference on Applied Climatology) pp 179–84

Meehl G A, Goddard L, Murphy J, Stouffer R J, Boer G, Danabasoglu G, Dixon K, Giorgetta M A, Greene A M, Hawkins E, Hegerl G, Karoly D, Keenlyside N, Kimoto M, Kirtman B, Navarra A, Pulwarty R, Smith D, Stammer D and Stockdale T 2009 Decadal Prediction: Can It Be Skillful? *Bull. Am. Meteorol. Soc.* **90** 1467–86 Online: https://journals.ametsoc.org/view/journals/bams/90/10/2009bams2778_1.xml

Meehl G A, Hu A, Arblaster J M, Fasullo J and Trenberth K E 2013 Externally Forced and Internally Generated Decadal Climate Variability Associated with the Interdecadal Pacific Oscillation *J. Clim.* **26** 7298–310 Online: https://journals.ametsoc.org/view/journals/clim/26/18/jcli-d-12-00548.1.xml

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, Kim H, Kirtman B P, Kushnir Y, Lovenduski N S, Mann M E, Merryfield W J, Nieves V, Pegion K, Rosenbloom N, Sanchez S C, Scaife A A, Smith D, Subramanian A C, Sun L, Thompson D, Ummenhofer C C and Xie S-P 2021 Initialized Earth System prediction from subseasonal to decadal timescales *Nat. Rev. Earth Environ.* **2**

Accepted Manuscript

340–57 Online: https://doi.org/10.1038/s43017-021-00155-x

Merryfield W J, Baehr J, Batté L, Becker E J, Butler A H, Coelho C A S, Danabasoglu G, Dirmeyer P A, Doblas-Reyes F J, Domeisen D I V, Ferranti L, Ilynia T, Kumar A, Müller W A, Rixen M, Robertson A W, Smith D M, Takaya Y, Tuma M, Vitart F, White C J, Alvarez M S, Ardilouze C, Attard H, Baggett C, Balmaseda M A, Beraki A F, Bhattacharjee P S, Bilbao R, de Andrade F M, DeFlorio M J, Díaz L B, Ehsan M A, Fragkoulidis G, Grainger S, Green B W, Hell M C, Infanti J M, Isensee K, Kataoka T, Kirtman B P, Klingaman N P, Lee J-Y, Mayer K, McKay R, Mecking J V, Miller D E, Neddermann N, Justin Ng C H, Ossó A, Pankatz K, Peatman S, Pegion K, Perlwitz J, Recalde-Coronel G C, Reintges A, Renkl C, Solaraju-Murali B, Spring A, Stan C, Sun Y Q, Tozer C R, Vigaud N, Woolnough S and Yeager S 2020 Current and Emerging Developments in Subseasonal to Decadal Prediction *Bull. Am. Meteorol. Soc.* **101** E869–96 Online: https://journals.ametsoc.org/view/journals/bams/101/6/bamsD190037.xml

Murphy A H 1988 Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient *Mon. Weather Rev.* **116** 2417–24 Online: https://journals.ametsoc.org/view/journals/mwre/116/12/1520-0493_1988_116_2417_ssbotm_2_0_co_2.xml

O'Neill B C, Tebaldi C, van Vuuren D P, Eyring V, Friedlingstein P, Hurtt G, Knutti R, Kriegler E, Lamarque J-F, Lowe J, Meehl G A, Moss R, Riahi K and Sanderson B M 2016 The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6 *Geosci. Model Dev.* **9** 3461–82 Online: https://www.geosci-model-dev.net/9/3461/2016/

Palmer T 2014 Record-breaking winters and global climate change *Science (80-.).* **344** 803–4 Online: https://doi.org/10.1126/science.1255147

Patricola C M, O'Brien J P, Risser M D, Rhoades A M, O'Brien T A, Ullrich P A, Stone D A and Collins W D 2020 Maximizing ENSO as a source of western US hydroclimate predictability *Clim. Dyn.* **54** 351–72 Online: https://doi.org/10.1007/s00382-019-05004-8

Sillmann J, Croci-Maspoli M, Kallache M and Katz R W 2011 Extreme Cold Winter Temperatures in Europe under the Influence of North Atlantic Atmospheric Blocking *J. Clim.* **24** 5899–913 Online: https://journals.ametsoc.org/view/journals/clim/24/22/2011jcli4075.1.xml

Sillmann J, Kharin V V., Zwiers F W, Zhang X and Bronaugh D 2013 Climate extremes indices in the CMIP5 multimodel ensemble: Part 2. Future climate projections *J. Geophys. Res. Atmos.* **118** 2473–93

Smith D M, Eade R and Pohlmann H 2013 A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction *Clim. Dyn.* **41** 3325–38 Online: https://doi.org/10.1007/s00382-013-1683-2

Smith D M, Eade R, Scaife A A, Caron L-P, Danabasoglu G, DelSole T M, Delworth T, Doblas-Reyes F J, Dunstone N J, Hermanson L, Kharin V, Kimoto M, Merryfield W J, Mochizuki T, Müller W A, Pohlmann H, Yeager S and Yang X 2019 Robust skill of decadal climate predictions *npj Clim. Atmos. Sci.* **2** 13 Online: https://doi.org/10.1038/s41612-019-0071-y

Spearman C 1904 "General intelligence," objectively determined and measured. *Am. J. Psychol.* **15** 201–93

Vicente-Serrano S, Beguería S and López-Moreno J 2010 A Multiscalar Drought Index Sensitive to Global Warming: The Standardized Precipitation Evapotranspiration Index *J. Clim.* **23** 1696–718

Vicente-Serrano S M, Domínguez-Castro F, McVicar T R, Tomas-Burguera M, Peña-Gallardo

Accepted Manuscript

M, Noguera I, López-Moreno J I, Peña D and El Kenawy A 2020 Global characterization of hydrological and meteorological droughts under future climate change: The importance of timescales, vegetation-CO2 feedbacks and changes to distribution functions *Int. J. Climatol.* **40** 2557–67 Online: https://doi.org/10.1002/joc.6350

Wang C, Liu H and Lee S-K 2010 The record-breaking cold temperatures during the winter of 2009/2010 in the Northern Hemisphere *Atmos. Sci. Lett.* **11** 161–8 Online: https://doi.org/10.1002/asl.278

Wilhite D A, Svoboda M D and Hayes M J 2007 Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness *Water Resour. Manag.* **21** 763–74 Online: https://doi.org/10.1007/s11269-006-9076-5

Wilks D S 2011 *Statistical methods in the atmospheric sciences* ed Elsevier (Amsterdam, the Netherlands, Boston: Elsevier)

Wilks D S 2016 “The Stippling Shows Statistically Significant Grid Points”: How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It *Bull. Am. Meteorol. Soc.* **97** 2263–73 Online: https://journals.ametsoc.org/view/journals/bams/97/12/bams-d-15-00267.1.xml

Xu Z, FitzGerald G, Guo Y, Jalaludin B and Tong S 2016 Impact of heatwave on mortality under different heatwave definitions: A systematic review and meta-analysis *Environ. Int.* **89–90** 193–203 Online: https://www.sciencedirect.com/science/article/pii/S0160412016300411

Zhang X, Alexander L, Hegerl G C, Jones P, Tank A K, Peterson T C, Trewin B and Zwiers F W 2011 Indices for monitoring changes in extremes based on daily temperature and precipitation data *WIREs Clim. Chang.* **2** 851–70 Online: https://doi.org/10.1002/wcc.147

Zwiers F W and von Storch H 1995 Taking Serial Correlation into Account in Tests of the Mean *J. Clim.* **8** 336–51 Online: https://journals.ametsoc.org/view/journals/clim/8/2/1520-0442_1995_008_0336_tsciai_2_0_co_2.xml

Accepted Manuscript