Exploiting the signal to noise ratio in multi-system predictions of boreal summer~~time~~ precipitation and ~~maximum~~ temperature~~s in Europe~~

Juan C. Acosta Navarro[1] and Andrea Toreti[1]

[1]European Commission, Joint Research Centre, Ispra, Italy.

Correspondence to: Juan C Acosta Navarro (juan.acosta-navarro@ec.europa.eu)

Abstract

Droughts and heatwaves are among the most impactful climate extremes. Their co-occurrence can have adverse ~~devastating~~ consequences on natural and human systems. Early information on their possible occurrence on seasonal timescales is beneficial for many stakeholders. Seasonal climate forecasts have~~s~~ ~~gradually~~ become openly available to the community~~more widely used;~~ but a wider~~its~~ use is currently hindered by limited skill in certain regions and seasons. Here we show that a simple forecast metric from a multi-system ensemble, the signal to noise ratio, can help overcome some limitations~~ in the boreal summer~~. Forecasts of mean ~~maximum~~ daily near surface air temperature and precipitation in boreal summers with high signal to noise ratio tend to coincide with observed larger deviations from the mean than summers ~~years~~ with small signal to noise ratio. The signal to noise ratio of the ensemble predictions serves as a complementary measure of forecast reliability that could potentially benefit users of climate predictions. ~~The same metric also helps identify processes relevant to seasonal climate predictability. Here we show that a positive phase of boreal spring sea surface temperature dipole index in the North Atlantic may favor the occurrence of dry and hot summers in Europe.~~

1. Introduction

Droughts are typically slow onset climate extreme events (Mishra and Singh, 2010), yet they can be disruptive and affect millions of people every year (Below et al., 2007; Enekel et al., 2020). Heatwaves can intensify and trigger a faster drought evolution (Bevacqua et al., 2022). Compound drought and heatwaves can ~~have~~ strongly ~~devastating consequences on~~ impact socio-economic and ecological systems, and may even compromise our ability to reach the UN sustainable development goal on climate action while strongly reducing the Earth system's current natural capacity to absorb and store carbon (Yin et al., 2023). The use of seasonal climate forecasts can provide actionable information to reduce the risks and the impacts of these events on key sectors like agriculture, energy, transport, water supply (Buontempo et al 2018; Ceglar and Toreti 2021).

In the last couple of decades, climate predictions have shown important progress in anticipating the evolution of various components of the climate system across the subseasonal to decadal time range (Merryfield et al., 2020; Meehl et al., 2021). ~~In spite of this progress, climate predictions still have low to moderate skill in many regions and seasons (e.g. European summer; Mishra et al. 2019); this limits their use and represents a barrier for stakeholders.~~ A combination of multiple forecast systems has shown overall benefits as compared with single systems, and can improve forecast quality up to a certain extent (Hagedorn et al., 2005~~-~~; Mishra et al., 2019). In spite of the recent progress, climate predictions still exhibit low to moderate skill in many regions and seasons (e.g. European summer; Mishra et al. 2019), something that limits their use and represents a barrier for stakeholders. Furthermore, multiple studies have shown that large ensembles are required to achieve skillful predictions, something that seems to be related to the forecast systems being more skillful at predicting real climate than at predicting their own realizations (i.e. ensemble members). This odd phenomenon has been called the signal to noise paradox (Eade et al., 2014; Scaife and Smith, 2018; Smith et al., 2020). It is particularly evident in the Euro Atlantic region during winter both on seasonal and decadal timescales. However boreal summer predictions have been generally overlooked. A recent study based on a single forecasting system has shown that sampling years with high SNR results in more skillful predictions of monthly temperatures in Japan throughout the year (Doi et al., 2022).

In this study we exploit multi-system ensembles to test whether specific boreal summers ~~years~~ with higher than normal predictability can be detected through the local relation between skill and SNR. We explore this for near surface air temperature and precipitation predictions, both locally and on large aggregated mid-latitude regions of the Northern Hemisphere. ~~signal to noise ratio (SNR; section 3). We then use this proposed approach to explore sources of summer climate predictability in Europe (Section 4).~~

2. Methods

This~~e~~ analysis is based on seasonal re-forecasts (also known as hindcasts) of mean boreal summer precipitation and 2-meter mean daily ~~maximum~~ temperature (T2m~~Tmax~~) for the period 1993-2016 from ECMWF SEAS5 (S5, Johnson et al., 2019), UKMO GloSea6 (S600, MacLachlan et al., 2015), MeteoFrance (S8, Batté et al., 2017~~S8, Guérémy et al., 2021~~), CMCC (S35, Gualdi et al., 2020) and DWD (S21, Baehr et al., 2015), available from the Copernicus C3S Climate Data Store. The observationally based datasets to evaluate the re-forecasts are ERA5 (Hersbach al., 2020) for T2m~~Tmax~~ and GPCC (Schnider et al., 2011) for precipitation. The use of summer mean T2m~~Tmax~~ is not intended to characterize single heatwaves, but to estimate average ~~maximum~~ daily deviations from the mean on a seasonal scale. In a climatological sense, more intense, more frequent or longer heatwaves than usual generally define hot summers and hence average T2m~~Tmax~~ may be seen as a seasonal integrator of heatwave activity. Forecast skill is evaluated with the anomaly correlation coefficient (ACC) between the ensemble mean and the observational reference. To complement the skill estimates of ACC, two additional deterministic skill metrics are computed ~~in Figures 5 and 6~~: t~~T~~he mean squared skill score (MSSS, Murphy, 1988) and the Gilbert skill score (GSS, WMO, 2014). The mean squared skill score compares the mean square error of the forecasts with the mean square error of the climatological value. It ranges from minus infinity to 1 and values above~~over~~ 0 indicate skill in the predictions. The GSS measures the fraction of correctly predicted events over the total number of predicted events plus misses, and takes into consideration the randomly predicted events. The thresholds to define event/non event are the top and bottom 25% summers for T2m (hot) and precipitation (dry), respectively. ~~GSS values above 0 indicate skillful predictions.~~ Standardization of the anomalies of each ensemble member and the observational reference data is performed prior to the analysis. This step guarantees that each member from each system has a comparable year-to-year variability to the observed one. Additionally, the standardized T2m~~Tmax~~ anomalies are linearly detrended at the grid level and for each member of the re-forecasts and in ERA5 to isolate as much as possible the impact of the long term warming. ~~In Section 4, the Sea Surface Temperature (SST) and Geopotential Height (500 hPa, GPH500) fields are taken from ERA5 and ERSSTv5 (Huang et al., 2017), respectively.~~

Following Doi et al. (2022)~~In addition to the ACC~~, the SNR~~metric computed~~ is calculated as ~~the product of the average multi-system ensemble mean anomalies deviation from the long term mean and the intrinsic ensemble coherence (inverse of standard deviation) is calculated with the signal to noise ratio for both T2mTmax and precipitation as~~: $SNR = \frac{\mu_e}{\sigma_e}$, where μ_e is the multi-system ensemble mean and σ_e is the multi-system standard deviation after standardization, computed across ensemble members for every summer (June - August) and for each gridbox. 25 members per system are used to have an equal contribution from each system.

3. Signal to noise ratio and forecast~~s~~ skill

Figure 1 displays spatial maps of mean (boreal) summer T2m~~max~~ ACC, time averaged SNR, and a scatter plot which shows the local relation between ACC and SNR. On average, skill values over land increase with higher SNR values. Negative values of ACC are nearly non-existent when the threshold of SNR exceeds the value of about 0.5 in the same gridbox. Statistically significant skill in T2m~~Tmax~~ is mostly confined to the tropics and sub-tropics. However, significant skill is also found in western North America, the eastern Mediterranean, central Asia and southern South America. Notable exceptions in the tropics are ~~the~~ Congo and parts of the Amazon rainforests.

The patterns of SNR largely mirror those of ACC. Generally, there is a good agreement between areas of high skill (ACC) and areas with high SNR, something that is further confirmed by the local relation between ACC and SNR (Fig. 1c).

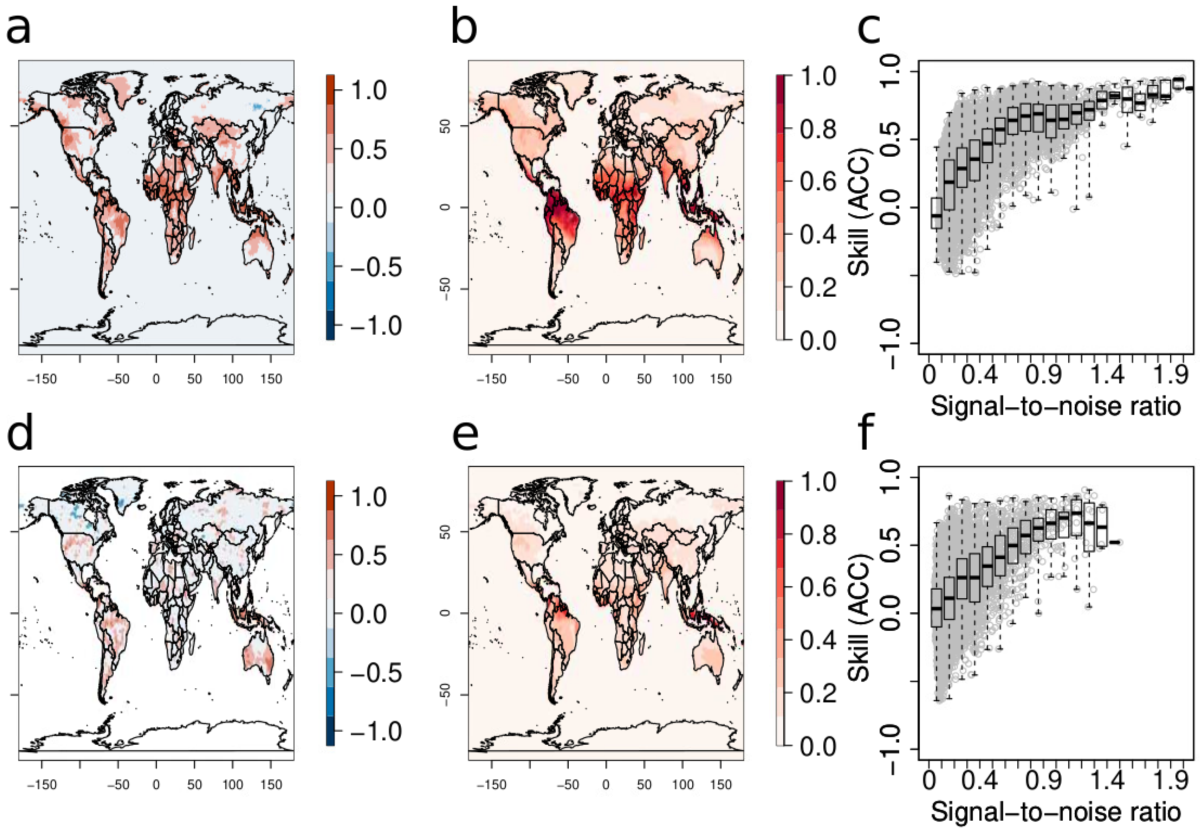


Figure 1: June-August Skill (ACC), time averaged SNR and scatterplots of local relation between ACC and SNR for T2m~~max~~ (a-c) and precipitation (d-f). Each gray dot in c,f represents the values of ACC and SNR at each gridbox. Only~~Gray dots in (a,d) indicate~~ statistically ~~non-~~significant values with a 90% confidence based on a t-test are displayed in (a,d). The re-forecasts are initialized every May.

Precipitation follows a similar behavior in terms of ACC and SNR, although statistically significant skill is less widespread (Fig. 1d-f). Areas under the influence of El Nino Southern Oscillation (ENSO; Lenssen et al., 2020) appear as regions with significant ACC and high SNR. Skillful values are mostly located in the Americas, the Maritime continent and Australia. Precipitation skill and SNR in Africa and Asia are much lower, making these the regions with the largest qualitative differences between the two variables.

Based on the observed link between skill and SNR, we use the latter one as the single criterion to exclude from the re-forecasts years with very low and very high values to understand their impact on skill. When 25% of the years (6 in total) with the highest SNR (Fig. 2a) are excluded, the results overall show much lower values of ACC than when only 25% of the years with the lowest SNR are excluded (Fig. 2b). Furthermore, differences between the latter and the former result (in many cases) in higher statistically significant values ~~and more statistical significance~~ than the ACC computed with~~hen~~ ~~only~~ selecting only years without the highest SNR (Fig. 2a,c). This result highlights the importance that these extreme SNR years can have on skill. In fact, only skill values computed by~~when~~ excluding the bottom 25% of SNR years (Fig. 2b) are comparable to the ones estimated when all years are used for the computation (Fig. 1a).

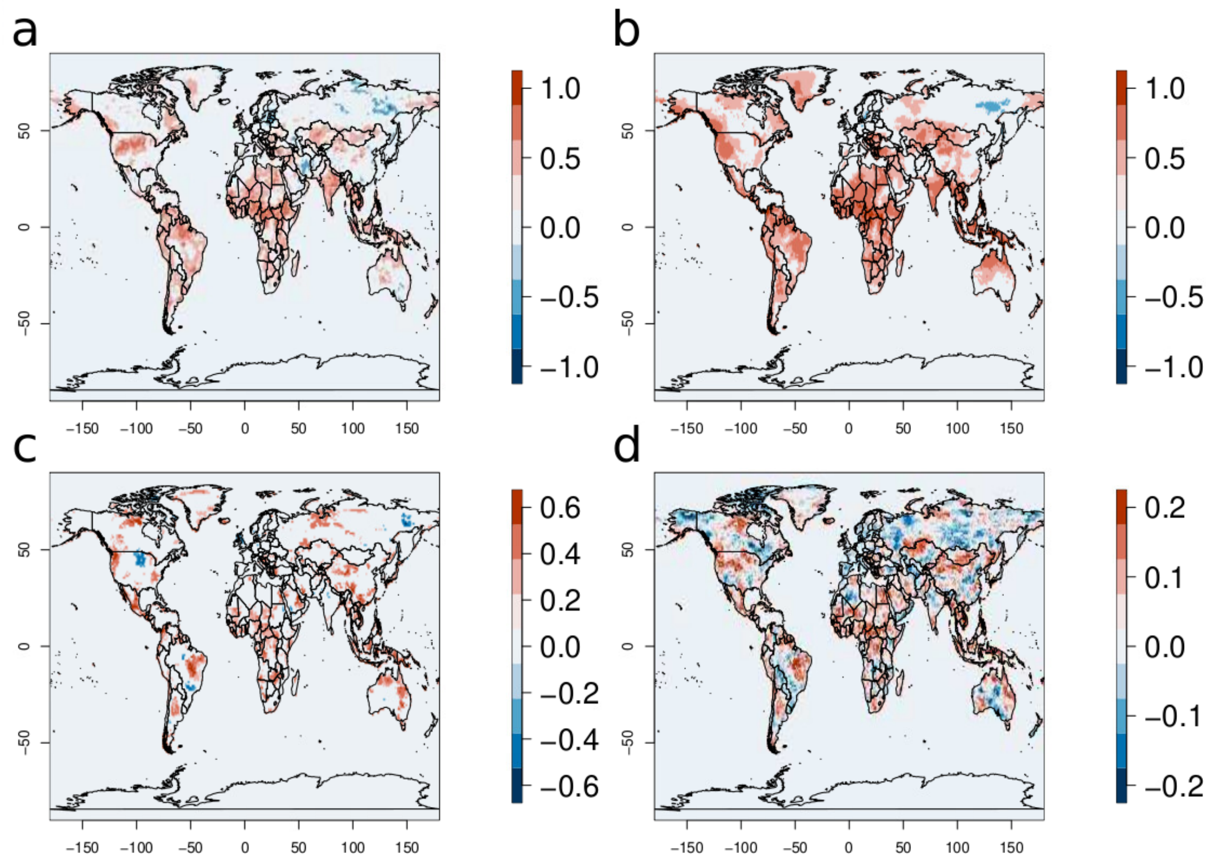


Figure 2: *Skill (ACC) of T2m~~Tmax~~ predictions excluding 25% of the years with highest (a) and lowest (b) local SNR. (c) Difference between (a) and (b). (d) Difference in the time-averaged absolute deviation from the mean in ERA5 T2m~~Tmax~~, excluding years having 25% of the lowest and highest local SNR, respectively. Only~~Gray dots in (a-c) indicate~~ statistically ~~non-~~significant values with a 90% confidence based on a t-test are displayed in (a-c). The re-forecasts are initialized every May.*

Interestingly, using the same criterion to select ERA5 T2m~~Tmax~~ values reveals that in general, excluding years with high ensemble SNR results in lower absolute deviations from the mean than when the low SNR years are excluded (Fig. 2d). Additionally, these differences overall coincide with regions with significant skill differences (Fig. 2c,d). This implies that years with more extreme deviations from the mean (in the observations/reanalysis) may be identified a priori by calculating the ensemble SNR of the forecast, and that forecast systems are in general more skillful when large deviations from the mean occur. ~~A notable exception is north-western Europe, where an opposite behavior is identified; however, it vanishes when a later initialization (June) is used.~~

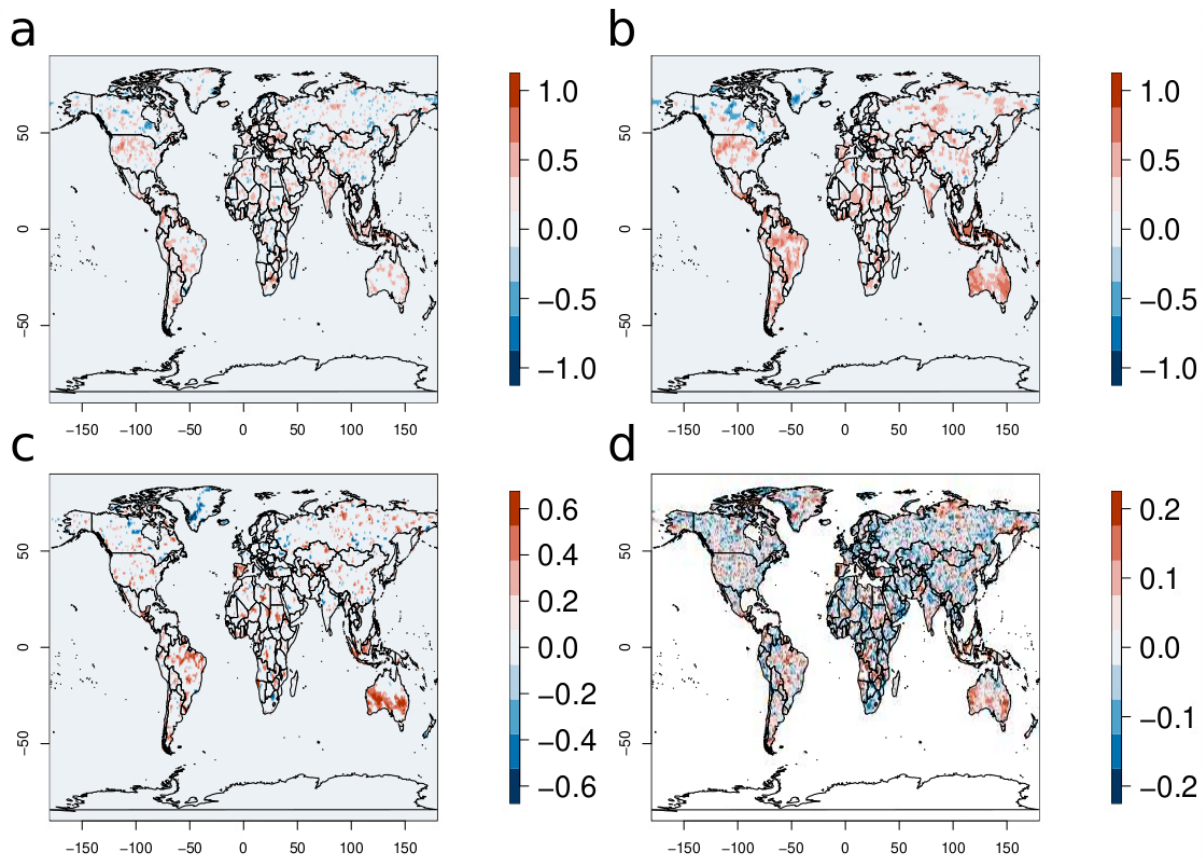


Figure 3: Same as Figure 2, but for precipitation.

Similar to T2m~~Tmax~~, the exclusion of years with high SNR also results in lower overall precipitation skill values than the one obtained when excluding low SNR years (Fig. 3a,b). Important skill differences appear in the Iberian peninsula, Brazil, Australia and Indonesia (Fig. 3c), and in most cases imply a shift ~~n increase~~ from non-significant to significant skill (Fig. 3 a and b, respectively).

Contrasting with T2m~~Tmax~~, the relation between ACC and mean absolute deviation from the mean in the observations is not obvious for precipitation (Fig. 3c,d). To further investigate this behavior, we analyzed the relationship between skill differences and the differences in absolute deviation from the mean for T2m~~Tmax~~ and precipitation, as usual using the re-forecasts that exclude the 25% of the years with the lowest and the highest SNR, respectively. This analysis (not shown) confirms a statistically robust relationship between skill and large deviations from mean observed precipitation, but still weaker than for T2m~~Tmax~~.

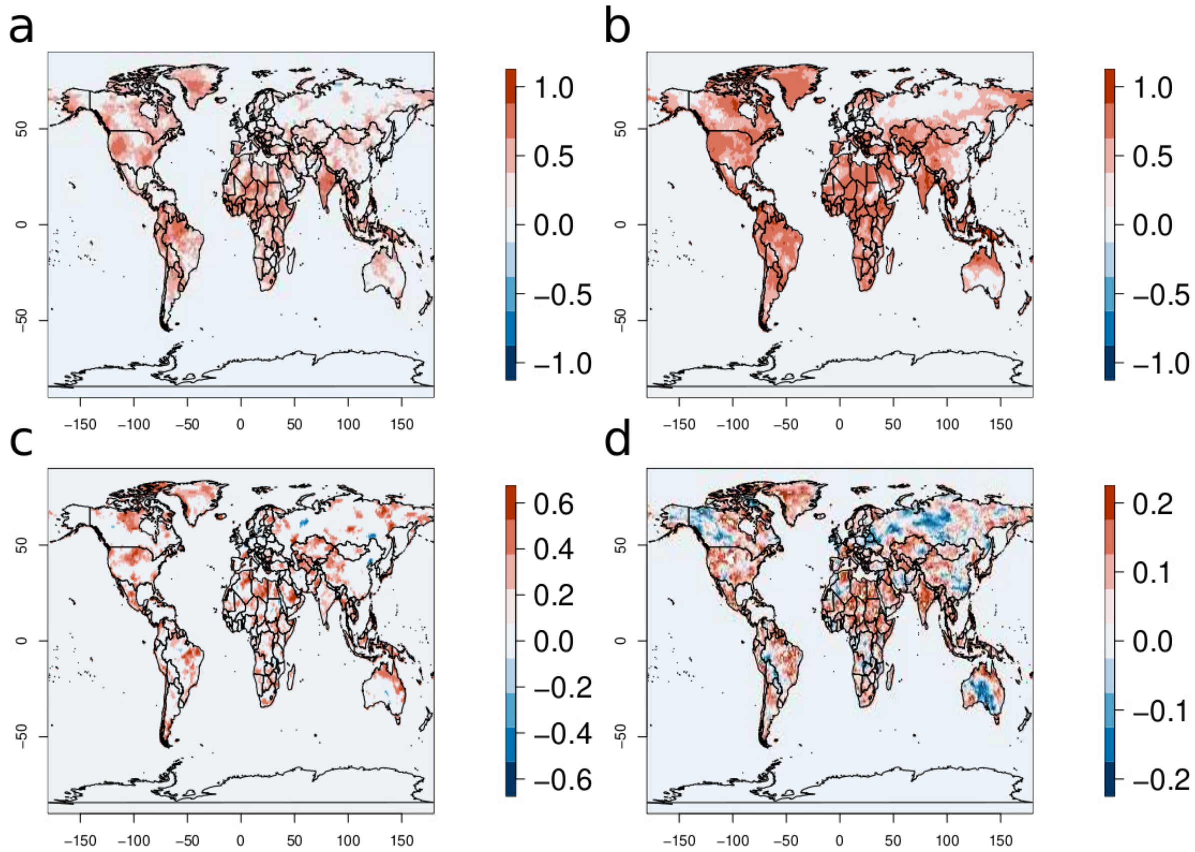


Figure 4: The same as Figure 2, but for re-forecasts initialized every June.

Figure 4 shows a clearer relation between the impact on skill of the most extreme years in terms of SNR and the absolute T2m~~Tmax~~ anomalies in ERA5, as compared with Figure 2. There is a good correspondence in all continents, including parts of Europe (Fig. 4 c,d).~~ as opposed to the results presented in Figure 2.~~ The only difference between the two F~~f~~igures is that they show the results from re-forecasts with different initialization dates. Both target the boreal summer months (June-August), but Figure 2 shows the results from the May initialization while ~~and~~ Figure 4 shows the results from the June initialization.~~ In addition,~~ S~~s~~imilar qualitative conclusions can be made for precipitation (not shown).

In Figure 5 we use the same methodology to sample years based on T2m~~-~~-SNR, but applied ~~in this case~~ to ~~the~~ specific northern hemisphere mid-latitude regions: the Mediterranean, North and Central Europe, north western Asia, east Asia, western North America and eastern North America. All the three skill metrics computed show that sampling the 18 years with highest SNR, generally results in more skillful T2m predictions than when sampling all 24 years or the 18 years with lowests SNR. The only exceptions are observed~~seen~~ in North and Central Europe where there is basically no skill or in eastern North America, where all the three selection methods show similar skill levels. Examples of successful prediction of extreme (high) T2m years and high SNR are 1999 and 2003 in the Mediterranean, 2002 in Northern/Central Europe, 1998 in northwestern Asia, 2006 and 1998 in western and eastern North America, respectively. There are also some examples of extreme (high) T2m and low SNR, such as 2012 in the Mediterranean, or 1994 and 2016 in East Asia. However, higher overall GSS for the top T2m positive anomalies indicates that on average, sampling years with high SNR results in better prediction of the extreme events.

A similar analysis on precipitation is shown in Figure 6. The results of precipitation qualitatively agree with those of T2m. Precipitation skill is highest for years with highest SNR and lowest for years with lowest SNR, the only exception being North and Central Europe, again a region with no skill in either precipitation or T2m predictions. Years of successful predictions of low precipitation and high SNR are 1994 and 2000 in the Mediterranean, 2015 in Northern/Central Europe, 1997 and 2001 in East Asia, 2003 in western North America, and 2011 in eastern

North America. Similar to T2m, GSS for low precipitation summers is generally higher for the top 18 years (in terms of SNR) than for the bottom 18 years or for all 24 years. Overall precipitation predictability is lower than T2m predictability in the regions analyzed, since skill scores for precipitation are generally lower than those of T2m. Note also that the same conclusions are obtained for both T2m and precipitation when separately sampling only the half of years with highest and lowest SNRs and/or when varying the threshold to define the most extreme years used in the GSS calculations (not shown).

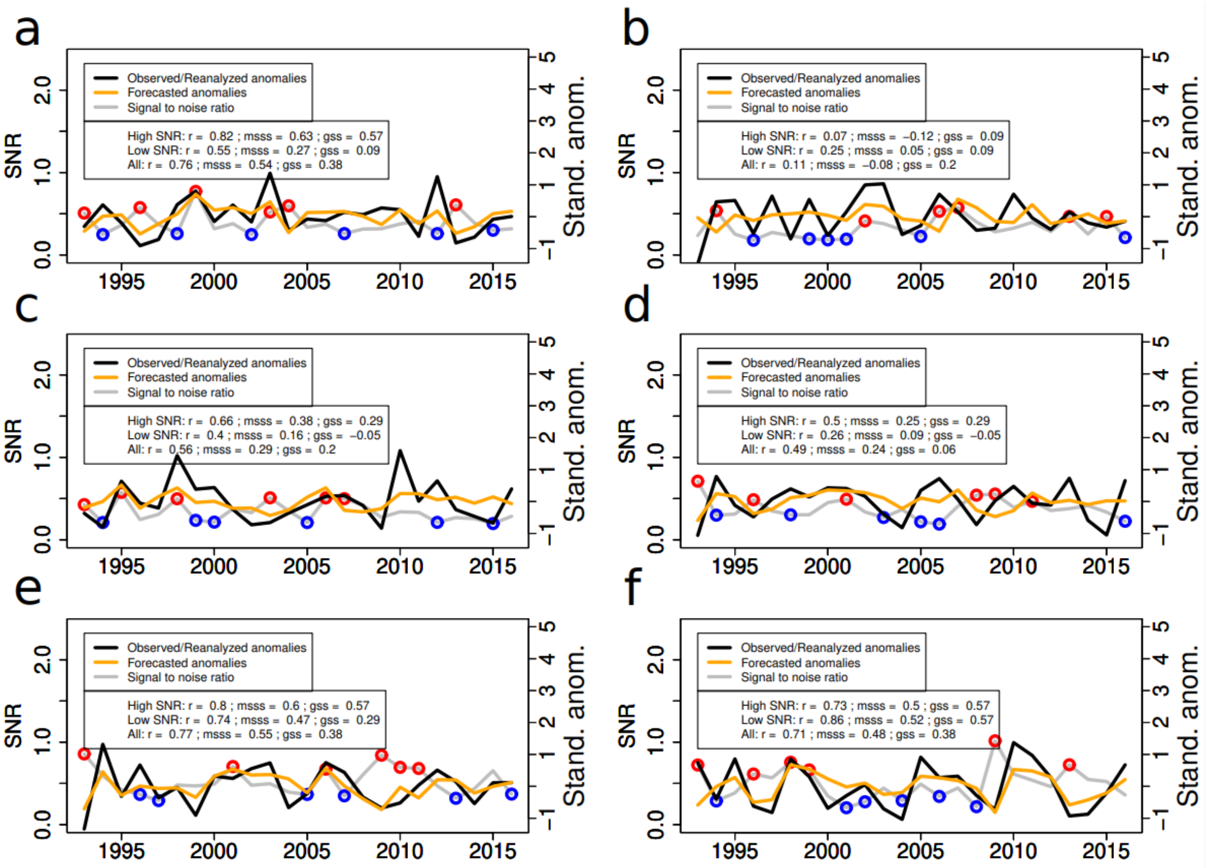


Figure 5: Area-averaged time series of observed and predicted, detrended and standardized mean summer T2m (right axis) and SNR (left axis) in (a) the Mediterranean (10W-35E, 30-45N), (b) North and Central Europe (10W-35E, 45-65N), (c) northwestern Asia (35-70E, 40-65N), (d) East Asia (90-130E, 25-45N), (e) western North America (123-100W, 30-50N) and (f) eastern North America (90-70W, 30-55N). Skill metrics provided separately for the 18 years with highest SNR (excluding blue circles), the 18 years with the lowest SNR (excluding red circles) and for all 24 years. The skill metrics are linear correlation, mean square skill score and Gilbert skill score (See methods). The values are taken from the re-forecasts initialized in June.

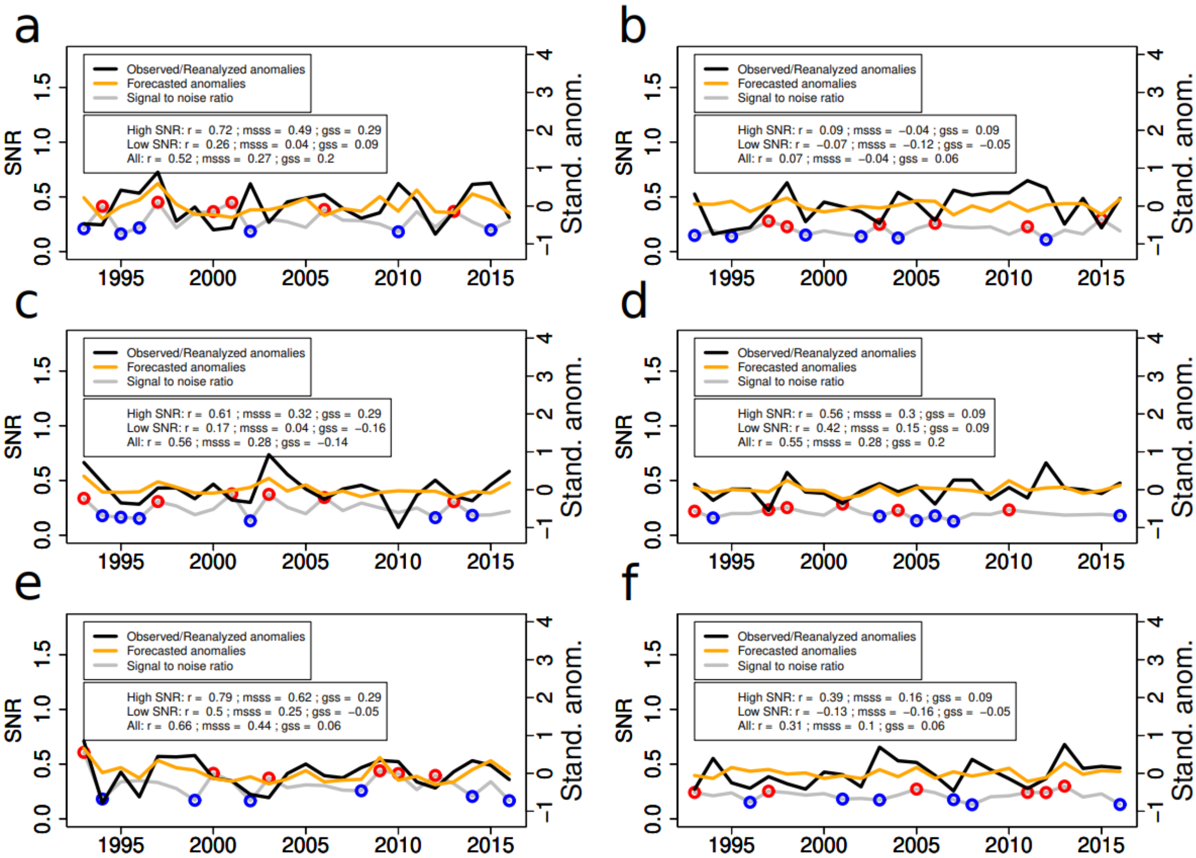
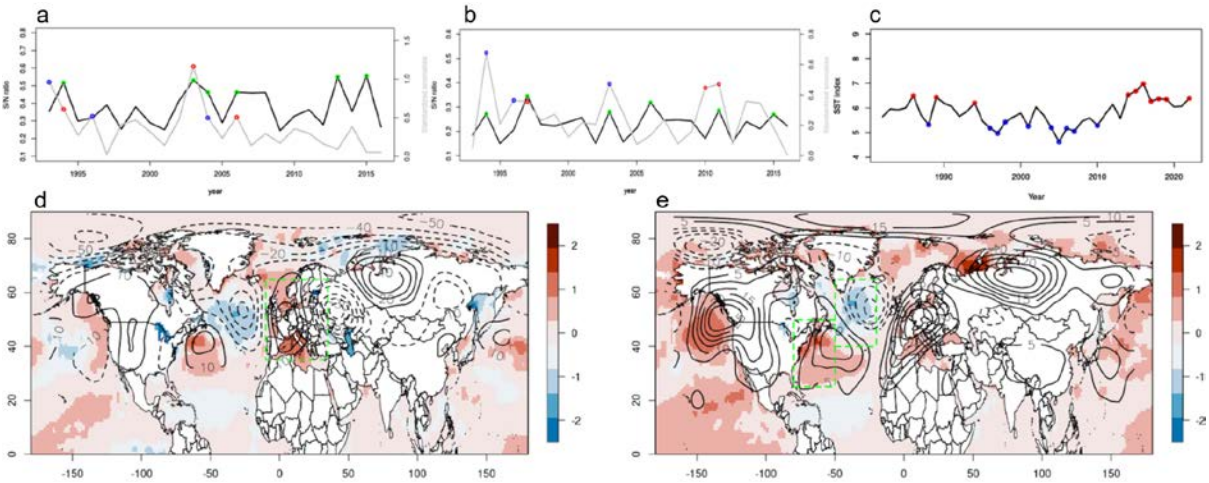


Figure 6: The same as Figure 5, but for precipitation.

~~3.2 **Sources of climate predictability in Europe**~~

~~Figure 5 shows how the ensemble SNR can also be applied to explore and understand sources of predictability and related climate processes. Figure 5a displays the time series of the SNR ratio (black) of the June initialized re-forecasts and the absolute value of the standardized ERA5 T2mTmax anomalies (gray) over Europe (defined in the area within 35-65N - 10W-35E, green box in Fig. 5d). The six years with the highest T2mTmax SNR in Europe are 1994, 2003, 2004, 2006, 2013 and 2015 (green dots in Fig. 5a), while the years with the lowest and the highest T2mTmax anomalies in Europe (after detrending) are 1993, 1996 and 2004, and 1994, 2003, and 2006, respectively (blue and red dots in Fig. 5a, respectively).~~

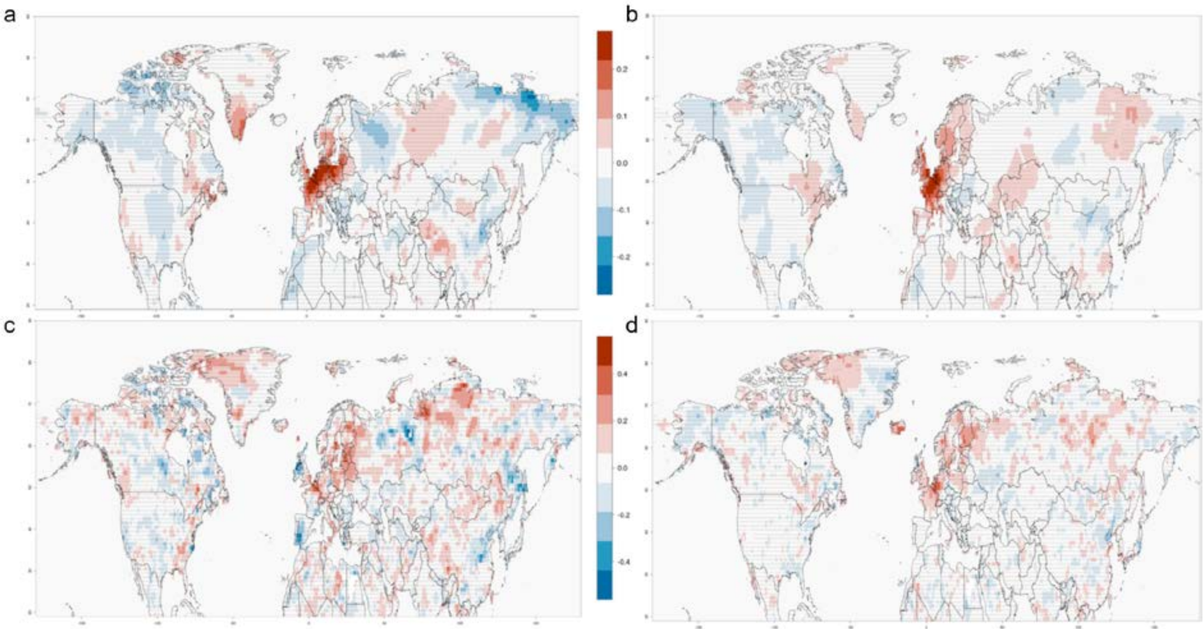
~~In terms of precipitation the largest SNR values are reached in 1994, 1997, 2003, 2006, 2011 and 2015, while the highest and lowest observed precipitation anomalies occur in 1997, 2010 and 2011 and 1994, 1996 and 2003, respectively. Common years with high absolute anomalies and high ensemble SNR are 1994, 2003, 2004, and 2006 for T2mTmax and 1994, 1997, 2003 and 2011 for precipitation. The summers of 1994 and 2003 have been documented as both dry and hot in Europe (e.g. Toreti et al., 2019) and also show high ensemble SNR for both T2mTmax and precipitation. This makes these years good candidates to explore possible sources of predictability. Anomalies of 1994 and 2003 of observed summer SST and GPH500 reveal a dipole of positive SST anomalies in the western North Atlantic and negative SST anomalies in the central/eastern North Atlantic (Fig. 5d), and a stationary Rossby wave pattern in the summer with anticyclonic anomalies in the western North Atlantic, western/central Europe and central Russia, and cyclonic anomalies in the central/eastern North Atlantic, eastern Europe/western Asia and northeastern Asia.~~



~~*Figure 5: (a) Time series of mean spatial SNR (black line) and absolute deviation from mean (gray) for T2mTmax over Europe in ERA5 (green box in panel d). Blue and red dots in (a) show the top three coldest and hottest summers in Europe (after detrending), while green dots indicate the top six years in terms of T2mTmax SNR. (b) The same as (a) but for precipitation. Blue and red dots in (b) show the three driest and wettest summers in Europe, while green dots indicate the top six years in terms of precipitation SNR. The re-forecasts used in (a-b) are from the June initialization. (c) Time series of the index estimated as the difference between the western and central/eastern North Atlantic SST in spring (March-May). Blue and red dots indicate the 25% lowest and highest values, respectively. (d) Mean summer anomalies of SST and GPH500 for the years 1994 and 2003. (e) Composites of summer SSTs and GPH500 for years with high minus low March-May SST index in the period 1982-2022.*~~

~~We hypothesize that years with a strong dipole in North Atlantic SST anomalies could precondition atmospheric flow, affecting hydroclimatic summer conditions in Europe. To test this hypothesis, we created an observed spring SST index (Fig. 5c) measuring the dipole strength defined as the difference in mean SST in the western and central/eastern centers of action (green boxes in Fig. 5e). Between 1982 and 2022, the years with the strongest dipole are identified before 1994 and after 2014, while years with the weakest dipole are almost exclusively found in the period 1995-2010, pointing to decadal/multi-decadal variability. A composite of summer SSTs and GPH500 (Fig. 5e), defined as the respective difference between the top 25% and the bottom 25% years based on the spring index, reveals very similar patterns than those observed in 1994 and 2003 (Fig. 5d). The SST index estimated in spring is associated with persistent SST anomalies well into the summer. These long lasting SST anomalies appear to force (or reinforce) a stationary Rossby wave train that induces both dry and hot summer conditions over most of Europe.~~

~~To further demonstrate the importance of this North Atlantic dipole for European summer climate, Figure 6 displays the added value of selecting each year the 60% of ensemble members that better reproduce the North Atlantic dipole index in the summer. The ranking is based on the values of the squared error of the index from each member with respect to ERA5. The reduced ensemble shows a clear, consistent and statistically significant improvement of skill of summer T2mTmax (Fig. 6a,b) and precipitation (Fig. 6c,d) in central and northwestern Europe for re-forecasts initialized in May (Fig 5a,c) and June (Fig 5b,d) as compared to the full ensemble. These improvements are only achieved by subsampling the members based on the summer dipole index for re-forecasts initialized in May and June. When the subsampling of members is based on the May index of the May initialized re-forests, there are no improvements of summer T2mTmax or precipitation skill in Europe, most likely because there is neither an improvement in the representation of the dipole in the summer (not shown).~~



~~*Figure 6: Skill difference (ACC) between a selection of 60% of the members with the best JJA SST index score (lowest RMSE) and the full ensemble for summer a) T2mTmax in forecasts initialized in May, b) T2mTmax in forecasts initialized in June, c) precipitation in forecasts initialized in May and d) precipitation in forecasts initialized in June. Gray dots indicate statistically non-significant values with a 90% confidence based on a t-test.*~~

4. Discussion

The SNR measures the relative weight of the ensemble mean anomalies with respect to the ensemble coherence. Its close resemblance in terms of spatial patterns with a skill metric like ACC, indicates that it can provide complementary information related to seasonal climate predictability. We have shown that in regions where the forecasts are skilful, years with high SNR exhibit on average larger observed deviations from the mean than years with low SNR, both for T2m and precipitation. This means that forecast systems are on average more reliable at predicting extremes when excluding years with low SNR. This has been further demonstrated for several Northern Hemisphere mid-latitude regions during boreal summer.

Despite the well known limitations of climate forecast systems (e.g. the signal to noise paradox), w~~W~~e have shown that in a multi-system ensemble, the SNR may provide valuable~~contains valuable~~ information as it represents an intrinsic measure of reliability for T2m and precipitation forecast.~~ which can be used to inform in advance on possible exceptional years with large temperature and precipitation anomalies.~~ The short span of 24 years defining the common hindcast period is a limitation of this study. Hence, longer hindcasts would be necessary to obtain more robust results, but are currently unavailable for most of the multiple systems analyzed.

~~The SNR also provides valuable information to detect potential sources of predictability. We have shown that, despite overall low skill, impactful events (i.e. anomalously dry and hot European summers) seem to be favored by a preceding dipole of high and low surface temperature anomalies in the western and central/eastern North Atlantic. These anomalies are identified in spring, persist through the summer and are associated with an anomalous stationary wave pattern showing anticyclonic conditions over most of Europe, a prime driver of hot/dry summer conditions. Dunstone et al. (20189) associate precipitation anomalies in central/northern Europe with a tripole pattern of North Atlantic SSTs in spring which has the two northernmost centers of action partially collocated with the two centers of action here identified, hence qualitatively agreeing with our findings. Nedderman et al. (2019) also show that ensemble subsampling selecting members that better reproduce a process involving North Atlantic Sea surface temperatures in spring followed by a Rossby wave train in late summer largely improves temperature forecasts in central/south-western Europe. Finally, the findings presented here also~~

~~agree with the ones reported by Acosta Navarro et al. (2022), which show that improved forecasts of central North Atlantic Sea surface temperatures in late spring/early summer increase skill in Europe during late summer thanks to a better simulated atmospheric circulation.~~

~~Significant skill improvements of~~ T2m~~Tmax and precipitation can be achieved in central and north-western Europe by subsampling ensemble members that better follow the evolution of the observed North Atlantic dipole temperature index during summer. Selecting members of the re-forecasts initialized in May that better agree with the observed dipole index in May, results in no clear improvement in the summertime dipole index or in the European climate. This points to the need for further efforts and analyses to understand this unexpected behavior. The proposed detection method based on ensemble SNR and North Atlantic SST pattern found here is nonetheless useful as a means to explore sources of atmospheric predictability for summer forecasts in Europe and could likely be applied to other regions and seasons.~~

References

- Acosta Navarro, J. C., García-Serrano, J., Lapin, V., & Ortega, P. (2022). Added value of assimilating springtime Arctic sea ice concentration in summer-fall climate predictions. Environmental Research Letters, 17(6), 064008.
- Baehr, J., Fröhlich, K., Botzet, M., Domeisen, D. I., Kornblueh, L., Notz, D., ... & Müller, W. A. (2015). The prediction of surface temperature in the new seasonal prediction system based on the MPI-ESM coupled climate model. Climate Dynamics, 44(9), 2723-2735.
- Below, R., Grover-Kopec, E., & Dilley, M. (2007). Documenting drought-related disasters: A global reassessment. The Journal of Environment & Development, 16(3), 328-344.
- Batté L., L. Dorel, C. Ardilouze, and J.-F. Guérémy, 2017: Documentation of the METEO- FRANCE seasonal forecasting system 8. Météo-France, 36 pp., https://www.umr-cnrm.fr/IMG/pdf/system8-technical.pdf.
- Bevacqua, E., Zappa, G., Lehner, F., & Zscheischler, J. (2022). Precipitation trends determine future occurrences of compound hot–dry events. Nature Climate Change, 12(4), 350-355.
- Buontempo, C., Hanlon, H. M., Soares, M. B., Christel, I., Soubeyroux, J. M., Viel, C., ... & Liggins, F. (2018). What have we learnt from EUPORIAS climate service prototypes?. Climate Services, 9, 21-32.
- Doi T, Nonaka M and Behera S (2022). Can signal-to-noise ratio indicate prediction skill? Based on skill assessment of 1-month lead prediction of monthly temperature anomaly over Japan. Front. Clim. 4:887782. doi: 10.3389/fclim.2022.887782

- ~~Dunstone, N., Smith, D., Scaife, A., Hermanson, L., Fereday, D., O'Reilly, C., ... & Belcher, S. (2018). Skilful seasonal predictions of summer European rainfall. Geophysical Research Letters, 45(7), 3246-3254.~~
- Ceglar, A., & Toreti, A. (2021). Seasonal climate forecast can inform the European agricultural sector well in advance of harvesting. npj Climate and Atmospheric Science, 4(1), 1-8.
- Eade, R. et al. Do seasonal to decadal climate predictions underestimate the predictability of the real world? Geophys. Res. Lett. 41, 5620–5628 (2014).
- Enenkel, M., Brown, M. E., Vogt, J. V., McCarty, J. L., Reid Bell, A., Guha-Sapir, D., ... & Vinck, P. (2020). Why predict climate hazards if we need to understand impacts? Putting humans back into the drought equation. Climatic Change, 162(3), 1161-1176.
- Gualdi, S., A. Sanna, A. Borrelli, A. Cantelli, M. del Mar Chaves Montero, S. Tibaldi, 2020: The new CMCC Operational Seasonal Prediction System SPS3.5. Centro Euro-Mediterraneo sui Cambiamenti Climatici. CMCC Tech. Note RP0288, 26pp.
- ~~Guérémy, J. F., Dubois, C., Viel, C., Dorel, L., Ardilouze, C., Batte, L., ... & Le Breton, M. (2021, March). Assessment of Météo-France current seasonal forecasting system S7 and outlook on the upcoming S8. In EGU General Assembly Conference Abstracts (pp. EGU21-10185).~~
- Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. Tellus A: Dynamic Meteorology and Oceanography, 57(3), 219-233.
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society, 146(730), 1999-2049.
- ~~Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., ... & Zhang, H. M. (2017). NOAA extended reconstructed sea surface temperature (ERSST), version 5. NOAA National Centers for Environmental Information, 30, 8179-8205.~~
- Johnson, S. J., Stockdale, T. N., Ferranti, L., Balmaseda, M. A., Molteni, F., Magnusson, L., ... & Monge-Sanz, B. M. (2019). SEAS5: the new ECMWF seasonal forecast system. Geoscientific Model Development, 12(3), 1087-1117.
- Lenssen, N. J., Goddard, L., & Mason, S. (2020). Seasonal forecast skill of ENSO teleconnection maps. Weather and Forecasting, 35(6), 2387-2406.
- MacLachlan, C., Arribas, A., Peterson, K. A., Maidens, A., Fereday, D., Scaife, A. A., ... & Madec, G. (2015). Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. Quarterly Journal of the Royal Meteorological Society, 141(689), 1072-1084.
- Meehl, G. A., Richter, J. H., Teng, H., Capotondi, A., Cobb, K., Doblas-Reyes, F., ... & Xie, S. P. (2021). Initialized Earth System prediction from subseasonal to decadal timescales. Nature Reviews Earth & Environment, 2(5), 340-357.
- Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., Butler, A. H., Coelho, C. A., ... & Yeager, S. (2020). Current and emerging developments in subseasonal to decadal prediction. Bulletin of the American Meteorological Society, 101(6), E869-E896.
- Mishra, A. K., & Singh, V. P. (2010). A review of drought concepts. Journal of hydrology, 391(1-2), 202-216.
- Mishra, N., Prodhomme, C., & Guemas, V. (2019). Multi-model skill assessment of seasonal temperature and precipitation forecasts over Europe. Climate Dynamics, 52(7), 4207-4225.
- Murphy, A. H. (1988). Skill scores based on the mean square error and their relationships to the correlation coefficient. Monthly weather review, 116(12), 2417-2424.
- ~~Neddermann, N. C., Müller, W. A., Dobrynin, M., Düsterhus, A., & Baehr, J. (2019). Seasonal predictability of European summer climate re-assessed. Climate Dynamics, 53(5), 3039-3056.~~
- Scaife, A.A., Smith, D. A signal-to-noise paradox in climate science. npj Clim Atmos Sci 1, 28 (2018). https://doi.org/10.1038/s41612-018-0038-4.
- Smith, D.M., Scaife, A.A., Eade, R. et al. North Atlantic climate far more predictable than models imply. Nature 583, 796–800 (2020). https://doi.org/10.1038/s41586-020-2525-0.

- Schnider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., & Ziese, M. (2011). GPCC Full Data Reanalysis Version 6.0 at 1.0◦: Monthly Land-Surface Precipitation from Rain-Gauges built on GTS-based and Historic Data.
- Toreti, A., Belward, A., Perez-Dominguez, I., Naumann, G., Luterbacher, J., Cronie, O., ... & Zampieri, M. (2019). The exceptional 2018 European water seesaw calls for action on adaptation. Earth's Future, 7(6), 652-663.
- Yin, J., Gentine, P., Slater, L., Gu, L., Pokhrel, Y., Hanasaki, N., ... & Schlenker, W. (2023). Future socio-ecosystem productivity threatened by compound drought–heatwave events. Nature Sustainability, 1-14.
- World Meteorological Organization (WMO, 2014). Forecast verification for the African severe weather forecasting demonstration projects; No. 1132. Geneva, Switzerland: World Meteorological Organization.