Higher subseasonal predictability of extreme hot European summer temperatures as compared to average summers

C. Ole Wulff[1] **and Daniela I. V. Domeisen**[1]

[1]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland

Key Points:

- Summer warm extremes are more predictable at weekly lead times than average and cold events
- The prediction skill shows strong regional dependence within Europe
- The skill for hot extreme events is strongly increased by the most severe and persistent events

Corresponding author: C. Ole Wulff, ole.wulff@env.ethz.ch

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1029/2019GL084314

Abstract

Summer temperatures in the last decades were increasingly characterized by persistent extremes, and there is evidence that this trend will continue in a warming climate. The exact timing of these extremes is less well known and it is therefore crucial to consider their subseasonal predictability. We compare the prediction of summer 2m-temperature extremes in Europe with the prediction of average events for four subseasonal forecasting systems. We find higher prediction skill for warm extremes as compared to average events, with some regional dependence. The same is not true for cold extremes, indicating an asymmetry in the processes causing opposite summer temperature extremes. The forecast skill is strongly increased by the most severe and persistent events in the analyzed period. We hypothesize that the enhanced warm extreme skill is related to persistent flow patterns and land-atmosphere interaction. This could have implications for potentially enhanced predictability in a warming climate.

1 Introduction

Recent years have seen an increased number of extreme heat waves across the Northern Hemisphere, e.g. over central Europe in 2003 (Schär et al., 2004; Trigo et al., 2005; Fink et al., 2004; Fouillet et al., 2006), over Russia in 2010 (Dole et al., 2011; Barriopedro et al., 2011), and over Europe in 2018 (Schiermeier, 2018). It has been evident for decades (IPCC, 1990) that extreme events such as heat waves are exacerbated by climate change (Stott et al., 2004, 2015; Coumou & Rahmstorf, 2012; Sippel et al., 2016; Diffenbaugh et al., 2017; Mann et al., 2017; Imada et al., 2018). The time needed to prepare for an extreme event is often beyond the skillful prediction timescales of a few days that are currently available (White et al., 2017). The observed and projected increase in strength and frequency of heat waves therefore calls for reliable predictions on timescales of weeks to months. Currently available models cannot predict the onset, duration, or amplitude of a heat wave on subseasonal timescales, as e.g. for the 2010 Russian heat wave (Quandt et al., 2016).

On subseasonal timescales several potential predictors of summer near-surface temperatures have been identified. For instance, predictability could stem from persistent atmospheric flow patterns as a large fraction of the observed summer heat extremes are associated with atmospheric blocking (Pfahl & Wernli, 2012; Schaller et al., 2018; Brunner et al., 2018; Sousa et al., 2018). Further, warm and cold extremes can be related to the presence of upper-tropospheric Rossby wave packets (RWPs, Fragkoulidis, Wirth, Bossmann, and Fink (2018)) and a regional stalling of the jet stream (Röthlisberger et al., 2016). While RWPs are not generally more predictable, 500 hPa geopotential height forecasts initialized in the presence of specific types of RWPs display enhanced skill up to week 3 of the forecast (Grazzini & Vitart, 2015). Furthermore, regimes favouring local persistent temperature anomalies are also influenced by modes of low-frequency and possibly remotely forced variability like the summer North Atlantic Oscillation (Folland et al., 2009; Ossó et al., 2017) or the summer East Atlantic pattern (Wulff et al., 2017; Neddermann et al., 2019). This large-scale control on surface temperatures has the potential to enhance their predictability beyond the typical weather forecasting timescales.

Other studies highlight the potential of land-atmosphere interactions for seasonal (Weisheimer et al., 2011; Prodhomme et al., 2016; Ardilouze, Batté, Bunzel, et al., 2017; Bunzel et al., 2018) and subseasonal prediction (Koster et al., 2010; Ardilouze, Batté, & Déqué, 2017) by showing that temperature forecasts benefit from a realistic initialization of the land surface. Land-atmosphere interactions were especially important for the heat waves of 2003 in Europe (Ferranti & Viterbo, 2006; Fischer et al., 2007) and 2010 in Russia (Miralles et al., 2014; Hauser et al., 2015). During these events the successive drying of the soil under the persistent atmospheric forcing led soil moisture to drop below a critical value, triggering a positive feedback between soil dryness and near surface temperatures (Seneviratne et al., 2010).

Many studies consider predictors of warm summer temperatures and heat waves only (e.g. Cassou, Terray, and Phillips (2005); Ardilouze, Batté, and Déqué (2017)), but the aforementioned mechanisms do not work equally for hot and cold temperature extremes and thus their predictability could also differ (Quesada et al., 2012).

Motivated by the importance of predicting summer extreme events on subseasonal timescales, we test if near-surface extreme temperatures are more predictable than average temperatures. We evaluate the hindcast skill of different subseasonal forecasting systems for the near-surface temperature evolution on timescales of several weeks in Europe and specifically focus on the comparison between the skill of predicting warm and cold extremes in comparison to average temperatures. The definition of these events, the metrics applied to verify the hindcasts and the data used are described in Section 2. The model skill is described in Section 3, which also shows the comparison of the skill for the different event types. Sensitivity analyses are shown in the Supporting Information (SI). The results are summarized and discussed in Section 4.

2 Data and Methods

2.1 Hindcasts and Verification Data

We consider hindcasts from four subseasonal forecasting systems provided by the Australian Bureau of Meteorology (BoM), the Chinese Meteorological Agency (CMA), the European Centre for Medium-Range Weather Forecasts (ECMWF) and the National Centers for Environmental Prediction (NCEP). All data are made available through the subseasonal to seasonal (S2S) prediction project (Vitart et al., 2017). The systems differ in many aspects of their forecasting strategies, which are described in detail in the SI (Table S1). In this study, we focus mainly on the ECMWF system, which is initialized twice per week. The hindcasts were verified using the ECMWF's Interim reanalysis (ERA-Interim, Dee et al. (2011)).

The data considered in this study are summer (June, July, August) 2m-temperature anomalies (T_{2m}) in the 12-year period 1999 – 2010, which is the longest common period that hindcasts from all four systems cover. Anomalies were computed with respect to each forecasting model's lead time dependent climatological seasonal cycle, which was computed from the first four harmonics of the daily 1999 – 2010 climatology. The lead time dependence of the climatology removes effects of drift in the models' climatologies with increasing lead time. Furthermore, as forecasts on subseasonal lead times are not expected to reproduce small-scale day-to-day variability, we applied a 5-day moving average to the daily temperature anomalies (T_{2m}^{5d}). Additionally, we averaged T_{2m}^{5d} over six regions in Europe (Figure 1b): Scandinavia (SC), Western and Eastern Europe (WEU and EEU, respectively), Russia and Ukraine (RUK), and the Western and Eastern Mediterranean (WMED and EMED, respectively). All analyses are based on hindcasts of the averages of T_{2m}^{5d} over one of these regions R ($\langle T_{2m}^{5d} \rangle_R$), where $\langle . \rangle$ indicates the spatial average using area weighting.

As a reference for the forecast skill, we further designed persistence hindcasts with a set-up mimicking the dynamical hindcasts. These were created by keeping $\langle T_{2m}^{5d} \rangle_R$ computed from ERA-Interim at each initialization date of the BoM system constant for 62 days (length of the BoM hindcasts). Note that both for the persistence forecast and for the verification we use anomalies with respect to the seasonal cycle.

2.2 Evaluation of Skill

The skill at reproducing the observed patterns of T_{2m}^{5d} is evaluated using the centered Anomaly Correlation Coefficient (ACC). The ACC for a given initialization time

i is defined following Wilks (2011, Chapter 8) as:

$$\mathrm{ACC}_i = \frac{\sum_{k=1}^{K}(y'_{ik} - \langle y' \rangle)(o'_{ik} - \langle o' \rangle)}{\sqrt{\sum_{k=1}^{K}(y'_{ik} - \langle y' \rangle)^2 \sum_{k=1}^{K}(o'_{ik} - \langle o' \rangle)^2}} \tag{1}$$

where k is an index for the grid point, y indicates the forecast and o the reanalysis. $\langle . \rangle$ indicates averaging over all K grid points within the chosen region, primes indicate anomalies with respect to the climatology.

When analyzing the skill of the hindcasts for a certain event type, we treat each individual ensemble member's hindcast as a deterministic binary forecast. We define an average event as a day on which $\langle T_{2m}^{5d} \rangle_R$ lies between the monthly 25th and 75th percentile of the distribution. An extreme warm (cold) event is detected when $\langle T_{2m}^{5d} \rangle_R$ exceeds (falls below) the 95th (5th) percentile. Thus, extreme (average) events in our analyses have a base rate of $p_x = 5\%$ ($p_a = 50\%$) Note that by using percentiles of the distribution of anomalies we eliminate contributions to the forecast skill resulting from a successful reproduction of the seasonal cycle. Furthermore, by defining an event based on the model's own climatological distribution, we eliminate the frequency bias in the hindcasts, i.e. each system's hindcast ensemble predicts an extreme (average) event on 5% (50%) of the considered days at any given lead time by design.

In order to be able to compare the skill for events with different base rates, we apply a skill measure that is base rate independent and does not degenerate when the base rate decreases (Jolliffe & Stephenson, 2012). The extremal dependence index (EDI, Ferro and Stephenson (2011)) fulfills these requirements and is defined as:

$$\mathrm{EDI} = \frac{\log F - \log H}{\log F + \log H} \tag{2}$$

where H is the hit rate, i.e. the number of hits divided by the number of observed events, and F the false alarm rate, i.e. the number of false alarms divided by the number of observed non-events. The EDI varies between -1 and 1 where 0 indicates no skill and 1 is the skill of a perfect forecast. The standard error of the EDI is given as (Ferro & Stephenson, 2011):

$$s_{\mathrm{EDI}} = \frac{2|\log F + \frac{H}{1-H}\log H|}{H(\log F + \log H)^2}\sqrt{\frac{H(1-H)}{np}} \tag{3}$$

with sample size n, i.e. the total number of days considered and base rate p. The EDI is always shown with an interval of 2 standard errors covering approximately the 95% uncertainty interval. The skill is here considered significant if this interval does not encompass 0.

We will analyze the skill of the forecasting systems in terms of their ACC and EDI. These scores measure different aspects of forecast skill. While the ACC quantifies how well the anomaly pattern of the verification field is reproduced, the EDI measures the deterministic skill at predicting a certain event type. We restrict ourselves to comparing the EDI between different forecasting systems, regions and event definitions. The EDI for extreme and average events will be referred to as xEDI and aEDI, respectively.

3 Results

3.1 Summer Temperature Skill on Subseasonal Timescales

First, we assess the prediction skill for summer surface temperatures over European land areas. As a reference for the general temperature skill in the six European regions, we compute the ACC of the ensemble mean hindcasts (Figure 1a) in the four analyzed systems as well as that of a persistence forecast as defined in Section 2.1.

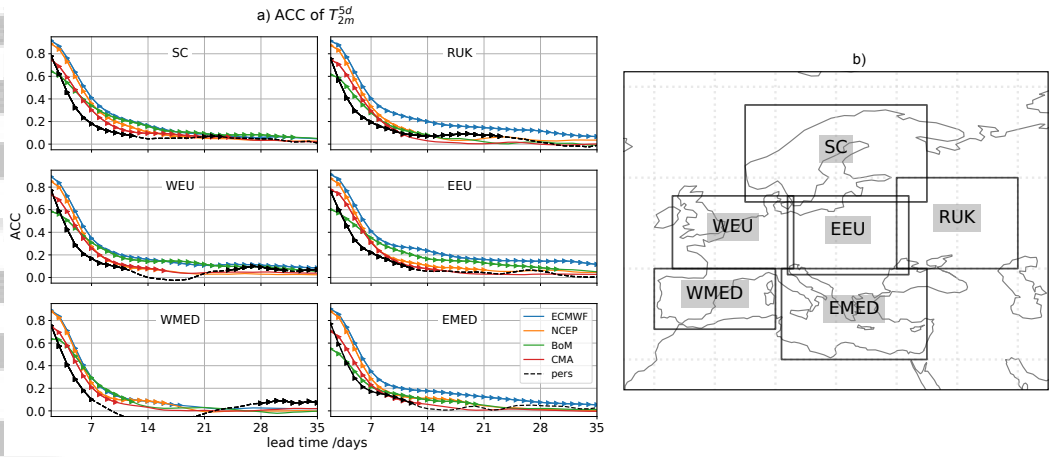

Figure 1. a) ACC of summer $\langle T_{2m}^{5d} \rangle_R$ as a function of lead time for the six European regions indicated in b). The ACC is shown for each of the four forecasting systems used in this study (see legend) and a persistence forecast based on ERA-Interim (black dashed line). Verification is with respect to ERA-Interim. Triangles indicate where the ACC is significantly different from zero. Note that the lead time corresponds to the central day of a 5 day running mean, thus the shortest lead time is two days.

All models show a fast decrease in ACC in week 1 (up to day 7) and their skill tends strongly toward that of persistence by week 4 (lead time 21 – 28 days) at the latest. The ECMWF system's ACC curve lies above all other models for all considered regions and lead times with only few exceptions. For lead times up to 7 days only the NCEP model compares well but is consistently slightly below. For lead times longer than 7 days, the NCEP system's skill deteriorates more quickly and becomes indistinguishable from the ACC of the persistence forecast within week 2 (lead time 7 – 14 days) of the forecast. The ACC of the CMA system and the persistence forecast start at a similar value but the CMA outperforms the persistence forecast generally until the middle of week 2 and exhibits a comparable skill after. The BoM system shows a slightly different behavior from the other systems as its skill is lower than the persistence skill up to approximately 4 days lead time. However, the decrease in ACC of the BoM system is much more gradual than that of all other systems. In fact, despite having the lowest skill in week 1, the BoM outperforms all other systems except the ECMWF starting in week 2 in the regions SC, WEU and EEU, where it exhibits a skill comparable to the ECMWF system.

The ACC furthermore shows some dependence on the region under consideration. Focusing on the skill of the best performing model, the ACC in the SC, RUK and EEU regions drops below 0.4 approximately one day later than in the other regions. In RUK and EEU it remains above 0.2 until the end of week 2 whereas in SC it already falls below that threshold during week 2. Despite the slower decrease of the ACC in the first two weeks, in week 3 the skill is lowest in SC along with the WMED region and effectively drops to zero, while it remains above 0.1 for the other regions. For the RUK and EMED regions however, this is only the case for the ECMWF system. The ACC in the other systems is zero in these regions in week 3 as well. In WEU and EEU, both ECMWF and BoM keep an ACC different from zero until week 4. In the EEU region the ECMWF system remains an ACC above 0.1 until week 5.

Out of the four considered forecasting systems, the ECMWF system clearly performs best in terms of ensemble mean ACC at all lead times, but it is not able to outperform a persistence forecast beyond week 4 (except in the EEU region). Despite its lower skill in week 1, the BoM system performs better than NCEP and CMA after week

2 in two thirds of the regions. For subseasonal lead times the ACC of the ensemble mean forecasts to predict pentad means of 2m-temperature is generally very low with some dependence on the region under consideration.

3.2 Prediction Skill for Extreme versus Average Temperatures

Next, we consider how the EDI of predicting extreme near-surface temperature events (xEDI) compares to the EDI for average events (aEDI, for event definition see Section 2.2). Note that for simplicity we refer to events in the lower tail of the temperature distribution as "cold events". For warm events in the ECMWF system, there is a clear tendency of the extreme event skill to significantly exceed the average event skill in all regions in forecast week 1 as indicated by the grey areas in Figure 2a. At lead times up to 14 days, however, in regions EEU and EMED both the xEDI and the aEDI become effectively zero. In the remaining four regions, the average event skill drops much more quickly than the warm event skill such that in week 2 the warm event skill is significantly higher. In week 3, the EDI difference drops to zero in the WMED region but stays positive in the other three regions. In week 4 the warm event skill lies above the average event skill only in the WEU and RUK regions. These regions show the largest difference between warm and average event skill. Notably however, the absolute warm event EDI drops much slower in the RUK than in the WEU region.

In contrast to the warm events, the xEDI for cold events does not significantly exceed the aEDI at any lead time and in any region except WMED and EEU (Figure 2b). In these regions, we observe significantly higher skill for extreme than for average events in week 1. However, these values decrease to zero within week 2 in both regions showing that also for those regions — despite the slower skill decrease in week 1 — the EDI difference at lead times longer than one week is effectively zero. This implies that there is no extended skill for cold extremes over average temperatures at subseasonal timescales.

Generally, in the ECMWF system summer warm extremes in Europe are better predicted than climatological events, with a strong dependence on the region considered. Especially at subseasonal lead times, warm extremes in the RUK and WEU and to some degree in the SC regions are significantly better predicted than average events. This is in contrast to the prediction skill for cold extremes in the same regions which is of the same magnitude as that for average events throughout all forecast lead times. Even though the EDI for cold extremes is larger in the WMED and EEU regions, this difference vanishes at subseasonal lead times. A sensitivity test with respect to the percentile threshold for defining extreme events using base rates p_x of 10% and 2.5% yields that the skill difference between extreme and average events is relatively insensitive to the exact percentile threshold chosen for the definition of the events (Figure S1). Our main conclusions equally hold for an analysis of xEDI and aEDI for the ECMWF for an extended period of 20 years (1998 – 2017, see Figure S2).

In order to compare these results between different forecasting systems, we consider the EDI differences for the same regions but restrict the analysis to subseasonal lead times only, i.e. week 3 and 4 (Figure 3 a and b, respectively). The xEDI and aEDI are effectively equal in the WMED, EEU and EMED regions and across all models. The only exception is the warm event skill of the BoM system in the EEU region in week 3. In the SC region, Figure 3 confirms that the difference between xEDI and aEDI events is not significant for other models. The warm event xEDI in the WEU region exceeds the aEDI for all models even though the differences in the other systems are less pronounced than in the ECWMF system. Especially the BoM system shows a larger uncertainty in the skill estimates which is likely due to the higher climatological spread of its ensemble. Again, the skill for cold events is not significantly different from the average event skill. Across all models, the most pronounced differences in the skill arise for warm events in the RUK region, especially for ECMWF and NCEP.

In summary, the ECMWF system performs significantly better at forecasting extreme warm near-surface temperature events than average events on subseasonal timescales

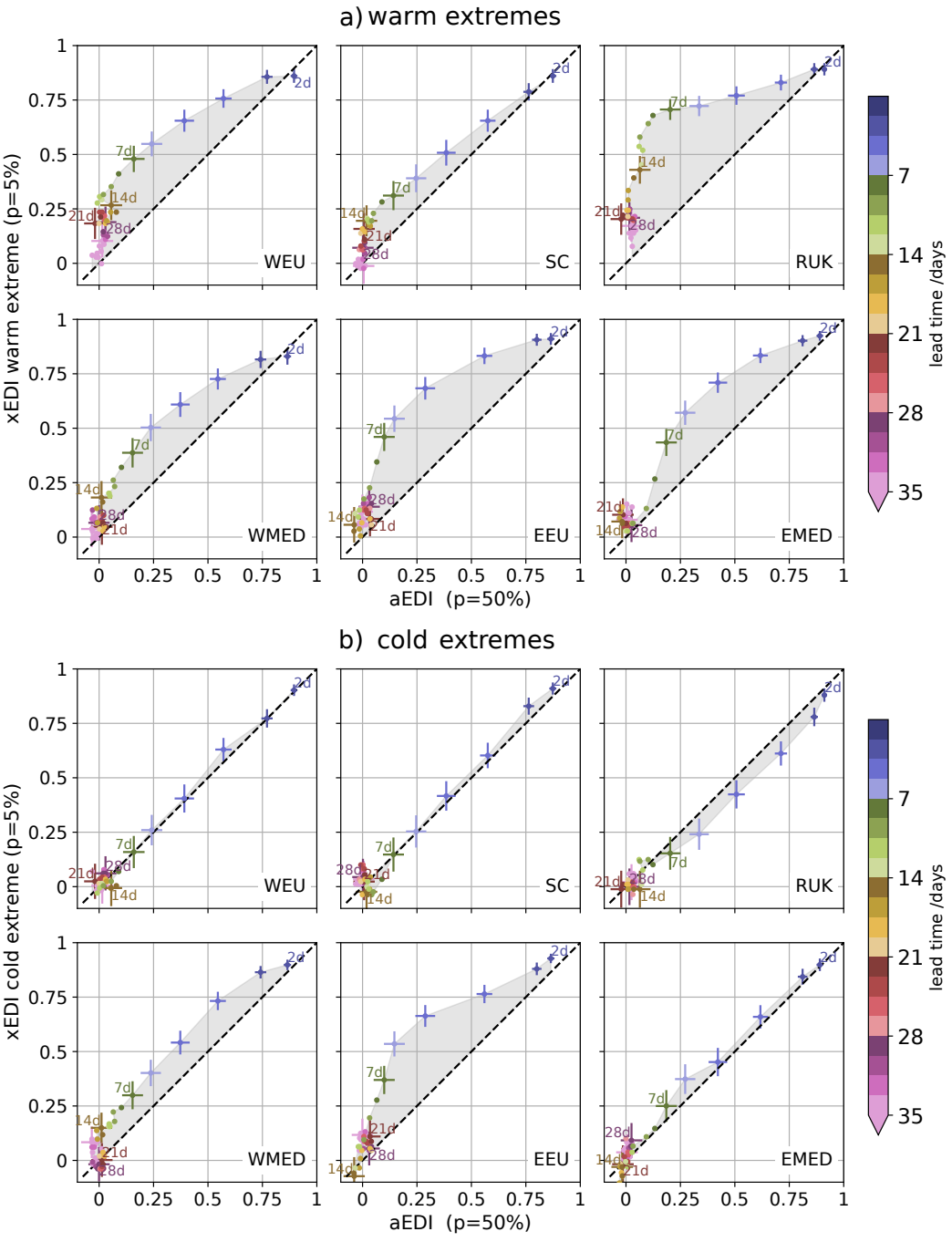

Figure 2. aEDI against xEDI for (a) extreme hot and (b) extreme cold events in the six regions for the ECMWF system. Bars indicate two standard errors around the aEDI and xEDI and are shown for lead times of 2-7, 14, 21, 28, and 35 days. Values are colored according to lead time. The black dashed diagonal indicates where aEDI = xEDI.

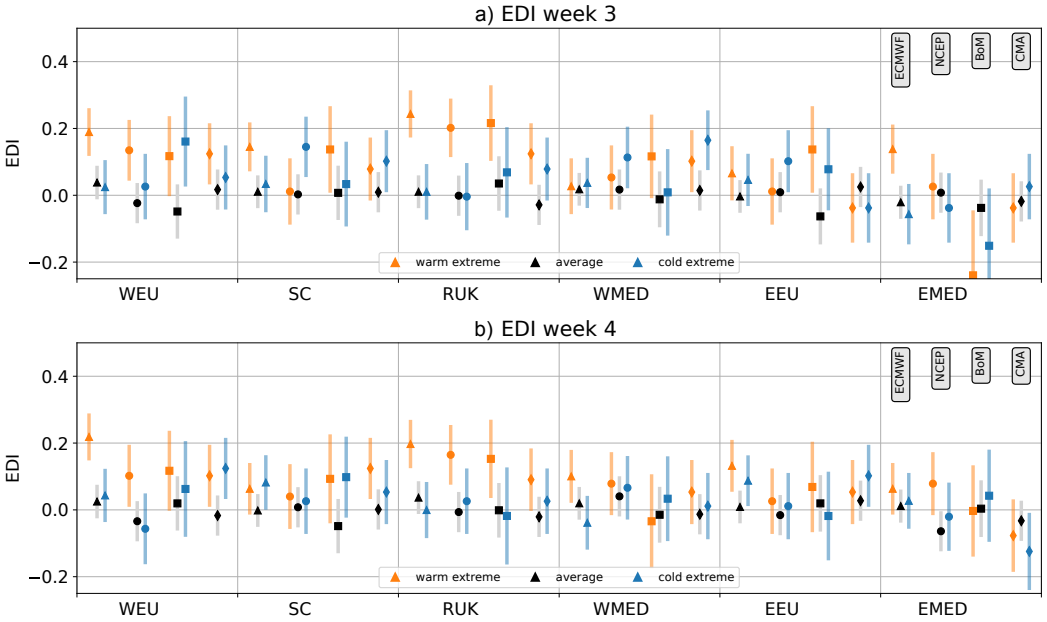

Figure 3. EDI by region and model on subseasonal lead times for (a) 3 and (b) 4 weeks. Orange, black, and blue dots and intervals indicate the EDI and two standard errors around it for extreme hot, average and extreme cold 2m-temperature events, respectively. For each region, the EDI for each of the four considered forecasting systems is shown, where triangles, circles, squares and diamonds indicate the ECMWF, NCEP, BoM and CMA system respectively.

in three (WEU, SC and RUK) out of six European regions. In all those regions the xEDI for warm events is significantly different from zero for all models indicating some prediction skill of extreme hot events on subseasonal timescales. The EDI difference is only robust in two (WEU and RUK) out of those regions when taking the other forecasting systems into consideration. For cold events, there is no enhanced skill over average temperature events at subseasonal lead times. This finding is consistent for all considered forecasting systems. To further test for the robustness of these results, we consider another base-rate independent skill measure. The odds ratio skill score (ORSS, Text S1 & Figure S3) confirms the overall findings of Figures 2 and 3 but at subseasonal lead times of 3 and 4 weeks the ORSS is only significantly different between extreme warm and average events for the ECMWF system. It should additionally be noted that the ORSS — despite being base-rate independent — increases the rarer the event ismaking its interpretation more difficult (Ferro & Stephenson, 2011; Jolliffe & Stephenson, 2012).

3.3 Sensitivity to Most Severe Heat Waves

Due to the limitations of the hindcast period for the S2S model systems, we are only able to sample 12 summers from 1999 to 2010. Since these years contain the two most severe heat waves in two of the study regions (European heat wave 2003 dominant for WEU region, Russian heat wave 2010 dominant for RUK region) we check the sensitivity of our skill estimates to removing these events from the sample for those two regions. Removing 2003 (2010) from the xEDI computation in the WEU (RUK) region reduces the sample to three quarters (one third) of its original size (orange percentages in Figure 4 b and d). Indeed, there is a dependence in both regions to removing the most severe events (Figure 4) such that the xEDI at subseasonal lead times is strongly reduced. While in the WEU region the xEDI of the full sample remains significantly different from

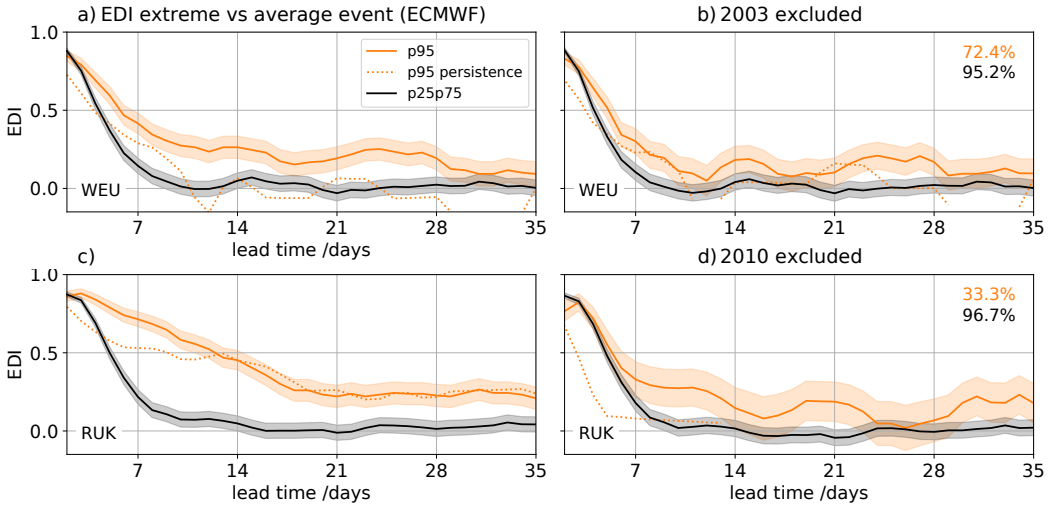

Figure 4. EDI for average (black) and extreme hot (orange) temperatures as a function of lead time taking into account ECMWF hindcasts in all available years (a,c) and excluding the most extreme years (b,d) in the respective region (2003 for WEU, 2010 for RUK). The shading around the curves indicates two standard errors around the EDI. The orange dotted line shows the xEDI of a persistence forecast. Percentages in b and d indicate the fraction of extreme (orange) and average (black) events that are retained in the estimation of the curves in the respective panel.

the aEDI out to week 5, the errorbars for aEDI and xEDI start overlapping already after 10 days when removing the summer of 2003. In general, the xEDI at lead times up to week 3 is most strongly affected by the removal of 2003 from the sample. In the RUK region the effect on the xEDI of removing the strongest event is even more pronounced but also note that the sample is more strongly reduced in the case of 2010. When 2010 is removed for the estimation in the RUK region the xEDI drops as quickly as the aEDI during the first week. It then levels off more gradually but due to the larger uncertainty it is no longer significantly different from the aEDI by the end of week 2.

These differences in the extreme event skill imply that there was larger skill at predicting the two events that were removed in Figures 4b and d. To answer the question why these events show extended prediction skill we also considered the xEDI of a persistence forecast (orange dotted line in Figure 4). The persistence xEDI is lower than that of the dynamical forecasts and rather insensitive to removing the 2003 event from the sample in the WEU region. In contrast, the persistence xEDI in the RUK region is comparable to the xEDI of the dynamical hindcasts when all years are taken into account, especially at subseasonal lead times. When removing 2010 from the sample, the xEDI of the persistence forecast drops to effectively zero already in week 1. This highlights the importance of the temperature persistence during the Russian heat wave of 2010 and the ability of the hindcast models to capture this aspect of the event.

4 Summary and Discussion

We investigated the prediction skill of near-surface summer temperatures in Europe from ensemble hindcasts of four subseasonal forecasting systems for the years 1999 – 2010. The skill of the ensemble mean hindcasts was evaluated using the ACC, which measures the pattern correlation between the hindcast and reanalysis fields. The ensemble mean ACC for 5-day means of 2m-temperature lies below 0.4 for lead times of more

than one week. The ECMWF performs best out of the four considered forecasting systems. While the BoM system performs poorest in the first week of the forecast it outperforms the CMA and NCEP systems in multiple regions on subseasonal lead times. This could be due to the generally larger climatological spread of the BoM system, which could be an advantage in capturing a weak predictable signal on these time scales, especially for cases where the NCEP and CMA systems are overconfident. The model skill also exhibits regional dependence. The regions with largest skill beyond week 1 are Russia/Ukraine (RUK), Eastern Europe (EEU) and Western Europe (WEU). The Extremal Dependence Index (EDI) for different event types (see Section 2.2) was applied to evaluate how the prediction skill for extremes compares to that for average events. Our analysis shows that in most regions of Europe, the prediction of warm near-surface temperature extremes is significantly better than for events close to the mean of the temperature distribution. This holds for week 1 in all regions, but the difference between the EDI for average events (aEDI) and for extreme events (xEDI) is strongest in the RUK, EEU and WEU regions. At subseasonal lead times the xEDI is significantly higher than the aEDI only in the RUK and WEU regions. In the RUK region, this result is robust across all considered forecasting systems. Excluding the European heat wave 2003 and the Russian heat wave 2010 from the sample for the computation of the EDI yields that these events contributed most strongly to the extended skill in the WEU and RUK regions, respectively. While warm extremes clearly show a higher predictability than average events, this does not hold for cold extremes with the same base rates. The cold event xEDI is only significantly different from the aEDI in the EEU and the Western Mediterranean (WMED) regions up to approximately 10 days lead time and not distinguishable from zero in the remaining regions at all lead times.

Although the differences between systems are small we find the ECMWF to have the largest skill at subseasonal lead times. Note however that we use ERA-Interim (i.e. a product of the ECMWF's Integrated Forecasting System used to generate the hindcasts) for the verification of the hindcasts, which potentially favors the ECMWF system in the skill comparison.

The asymmetry in the prediction skill between warm and cold extremes points to different processes for these opposing events. The large-scale control on near-surface temperatures could offer one explanation for the higher skill in forecasting extremes. Although both the onset and the duration of atmospheric blocking are rather poorly predicted in forecast models (Tibaldi & Molteni, 1990; Matsueda & Palmer, 2018), Grazzini and Vitart (2015) show that subseasonal forecast skill of the synoptic-scale circulation is enhanced when long, coherent RWPs from the Pacific to the Atlantic are present in the initial conditions. Although we did not explicitly test the occurrence of such wave packets in the initial conditions of our extreme event forecasts, some of these situations are captured in the analyzed period (Fragkoulidis et al., 2018). This could explain part of the observed extended skill for extreme temperatures. Land-atmosphere interactions have also been suggested to extend the predictability of near-surface temperatures to subseasonal timescales (Koster et al., 2010). In particular, the soil moisture-temperature feedback has been shown to be crucial for temperature extremes (Seneviratne et al., 2010) and could force temperatures to rise. To the extent that land-atmosphere fluxes are correctly represented in the models, they may yield extended prediction skill for the most long lasting warm extremes. Note that the regions considered here mostly have soil moisture above the critical value (Teuling et al., 2009). Thus the process described above only applies for extreme warm temperatures and could account for part of the asymmetry in the prediction skill between warm and cold extremes. Furthermore, the feedback would likely only act under strong and persistent atmospheric forcing as was the case in 2003 (Fischer et al., 2007) and 2010 (Miralles et al., 2014).

The results obtained here are specifically relevant in the context of a warming climate with potentially different predictability characteristics (Scher & Messori, 2019). Under global warming the temperature distribution exhibits both a shift and a broadening towards warmer values (Schär et al., 2004) and temperature extremes become more

frequent (Coumou & Rahmstorf, 2012). While an increased number of heat events can have harmful impacts on ecosystems and society, the potential for a better prediction of the most severe heat events as presented here is promising.

Acknowledgments

ERA-Interim and S2S data were obtained from the ECMWF. The S2S data is provided through the S2S Prediction Project. Funding by the Swiss National Science Foundation through project PP00P2_170523 is gratefully acknowledged. We thank Sonia Seneviratne and Mathias Hauser for helpful discussions on soil moisture-atmosphere interactions. All analyses and visualizations were conducted in python.

References

Ardilouze, C., Batté, L., Bunzel, F., Decremer, D., Déqué, M., Doblas-Reyes, F. J., ... Prodhomme, C. (2017). Multi-model assessment of the impact of soil moisture initialization on mid-latitude summer predictability. *Clim. Dyn.*, *49*(11-12), 3959–3974. doi: 10.1007/s00382-017-3555-7

Ardilouze, C., Batté, L., & Déqué, M. (2017). Subseasonal-to-seasonal (S2S) forecasts with CNRM-CM: a case study on the July 2015 West-European heat wave. *Adv. Sci. Res.*, *14*, 115–121. doi: 10.5194/asr-14-115-2017

Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., & Garcia-Herrera, R. (2011). The Hot Summer of 2010: Redrawing the Temperature Map of Europe. *Science (80-.).*, *332*, 220–224. doi: 10.1080/10255842.2015.1069566

Brunner, L., Schaller, N., Anstey, J., Sillmann, J., & Steiner, A. K. (2018). Dependence of Present and Future European Temperature Extremes on the Location of Atmospheric Blocking. *Geophys. Res. Lett.*, *45*. doi: 10.1029/2018GL077837

Bunzel, F., Müller, W. A., Dobrynin, M., Fröhlich, K., Hagemann, S., Pohlmann, H., ... Baehr, J. (2018). Improved Seasonal Prediction of European Summer Temperatures With New Five-Layer Soil-Hydrology Scheme. *Geophys. Res. Lett.*, *45*, 346–353. doi: 10.1002/2017GL076204

Cassou, C., Terray, L., & Phillips, A. S. (2005). Tropical Atlantic influence on European heat waves. *J. Clim.*, *18*(15), 2805–2811. doi: 10.1175/JCLI3506.1

Coumou, D., & Rahmstorf, S. (2012). A decade of weather extremes. *Nat. Clim. Chang.*, *2*(7), 491–496. doi: 10.1038/nclimate1452

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, *137*, 553–597. doi: 10.1002/qj.828

Diffenbaugh, N. S., Singh, D., Mankin, J. S., Horton, D. E., Swain, D. L., Touma, D., ... Rajaratnam, B. (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci.*, *114*(19), 4881–4886. doi: 10.1073/pnas.1618082114

Dole, R., Hoerling, M., Perlwitz, J., Eischeid, J., Pegion, P., Zhang, T., ... Murray, D. (2011). Was there a basis for anticipating the 2010 Russian heat wave? *Geophys. Res. Lett.*, *38*(6), L06702. doi: 10.1029/2010GL046582

Ferranti, L., & Viterbo, P. (2006). The European summer of 2003: Sensitivity to soil water initial conditions. *J. Clim.*, *19*(15), 3659–3680. doi: 10.1175/JCLI3810.1

Ferro, C. A. T., & Stephenson, D. B. (2011). Extremal Dependence Indices: Improved Verification Measures for Deterministic Forecasts of Rare Binary Events. *Weather Forecast.*, *26*(5), 699–713. doi: 10.1175/waf-d-10-05030.1

Fink, A. H., Bruecher, T., Krueger, A., Leckebusch, G. C., Pinto, J. G., & Ulbrich, U. (2004). The 2003 European summer heatwaves and drought-synoptic diagnosis and impacts. *R. Meteorol. Soc.*, *59*(8), 209–216. doi: 10.1256/wea.73.04

Fischer, E. M., Seneviratne, S. I., Lüthi, D., & Schär, C. (2007). Contribution of land-atmosphere coupling to recent European summer heat waves. *Geophys. Res. Lett.*, *34*, L06707. doi: 10.1029/2006GL029068

Folland, C. K., Knight, J., Linderholm, H. W., Fereday, D., Ineson, S., & Hurrel, J. W. (2009). The summer North Atlantic oscillation: Past, present, and future. *J. Clim.*, *22*(5), 1082–1103. doi: 10.1175/2008JCLI2459.1

Fouillet, A., Rey, G., Laurent, F., Pavillon, G., Bellec, S., Guihenneuc-Jouyaux, C., ... Hémon, D. (2006). Excess mortality related to the August 2003 heat wave in France. *Int. Arch. Occup. Environ. Health*, *80*, 16–24. doi: 10.1007/s00420-006-0089-4

Fragkoulidis, G., Wirth, V., Bossmann, P., & Fink, A. H. (2018). Linking Northern Hemisphere temperature extremes to Rossby wave packets. *Q. J. R. Meteorol. Soc.*, *144*, 553–566. doi: 10.1002/qj.3228

Grazzini, F., & Vitart, F. (2015). Atmospheric predictability and Rossby wave packets. *Q. J. R. Meteorol. Soc.*, *141*, 2793–2802. doi: 10.1002/qj.2564

Hauser, M., Orth, R., & Seneviratne, S. I. (2015). Role of soil moisture vs . recent climate change for heat waves in western Russia. *Geophys. Res. Lett.*, *43*, 2819–2826. doi: 10.1002/2016GL068036

Imada, Y., Shiogama, H., Takahashi, C., Watanabe, M., Mori, M., Kamae, Y., & Maeda, S. (2018). Climate Change Increased the Likelihood of the 2016 Heat Extremes in Asia, In: Explaining Extremes of 2016 from a Climate Perspective. *Bull. Am. Meteorol. Soc.*, *99*, S97–S101. doi: 10.1175/bams-d-17-0109.1

IPCC. (1990). *The IPCC Scientific Assessment* (J. Houghton, G. Jenkins, & J. Ephraums, Eds.). Cambridge University Press. Retrieved from `https://www.ipcc.ch/report/ar1/wg1/`

Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast Verification: A Practitioner's Guide in Atmospheric Science* (2nd ed.; I. T. Jolliffe & D. B. Stephenson, Eds.). University of Exeter, UK: John Wiley and Sons, Ltd.

Koster, R. D., Mahanama, S. P., Yamada, T. J., Balsamo, G., Berg, A. A., Boisserie, M., ... Wood, E. F. (2010). Contribution of land surface initialization to subseasonal forecast skill: First results from a multi-model experiment. *Geophys. Res. Lett.*, *37*, L02402. doi: 10.1029/2009GL041677

Mann, M. E., Rahmstorf, S., Kornhuber, K., Steinman, B. A., Miller, S. K., & Coumou, D. (2017). Influence of Anthropogenic Climate Change on Planetary Wave Resonance and Extreme Weather Events. *Sci. Rep.*, *7*, 45242. doi: 10.1038/srep46822

Matsueda, M., & Palmer, T. (2018). Estimates of flow-dependent predictability of wintertime Euro-Atlantic weather regimes in medium-range forecasts. *Q. J. R. Meteorol. Soc.*(July 2017), 1012–1027. doi: 10.1002/qj.3265

Miralles, D. G., Teuling, A. J., Van Heerwaarden, C. C., & De Arellano, J. V. G. (2014). Mega-heatwave temperatures due to combined soil desiccation and atmospheric heat accumulation. *Nat. Geosci.*, *7*(5), 345–349. doi: 10.1038/ngeo2141

Neddermann, N. C., Müller, W. A., Dobrynin, M., Düsterhus, A., & Baehr, J. (2019). Seasonal predictability of European summer climate re-assessed. *Clim. Dyn.*. doi: 10.1007/s00382-019-04678-4

Ossó, A., Sutton, R., Shaffrey, L., & Dong, B. (2017). Observational evidence of European summer weather patterns predictable from spring. *Proc. Natl. Acad. Sci.*, *115*, 59–63. doi: 10.1073/pnas.1713146114

Pfahl, S., & Wernli, H. (2012). Quantifying the relevance of atmospheric blocking for co-located temperature extremes in the Northern Hemisphere on (sub-)daily time scales. *Geophys. Res. Lett.*, *39*, L12807. doi: 10.1029/2012GL052261

Prodhomme, C., Doblas-Reyes, F., Bellprat, O., & Dutra, E. (2016). Impact of land-surface initialization on sub-seasonal to seasonal forecasts over Europe. *Clim. Dyn.*, *47*(3), 919–935. doi: 10.1007/s00382-015-2879-4

Quandt, L.-A., Keller, J. H., Martius, O., & Jones, S. C. (2016). Forecast Variability of the Blocking System over Russia in Summer 2010 and Its Impact on Surface Conditions. *Weather Forecast.*, *32*, 61–82. doi: 10.1175/waf-d-16-0065.1

Quesada, B., Vautard, R., Yiou, P., Hirschi, M., & Seneviratne, S. I. (2012). Asymmetric European summer heat predictability from wet and dry southern winters and springs. *Nat. Clim. Chang.*, *2*(10), 736–741. doi: 10.1038/nclimate1536

Röthlisberger, M., Pfahl, S., & Martius, O. (2016). Regional-scale jet waviness modulates the occurrence of midlatitude weather extremes. *Geophys. Res. Lett.*, *43*, 10,989–10,997. doi: 10.1002/2016GL070944

Schaller, N., Sillmann, J., Anstey, J., Fischer, E. M., Grams, C. M., & Russo, S. (2018). Influence of blocking on Northern European and Western Russian heatwaves in large climate model ensembles. *Environ. Res. Lett.*, *13*, 054015. doi: 10.1088/1748-9326/aaba55

Schär, C., Vidale, P. L., Lüthi, D., Frei, C., Häberli, C., Liniger, M. A., & Appenzeller, C. (2004). The role of increasing temperature variability in European summer heatwaves. *Nature*, *427*, 332–336. doi: 10.1038/nature02300

Scher, S., & Messori, G. (2019). How Global Warming Changes the Difficulty of Synoptic Weather Forecasting. *Geophys. Res. Lett.*, *46*. doi: 10.1029/2018GL081856

Schiermeier, Q. (2018). Climate as culprit. *Nature*, *560*, 20–22. Retrieved from `https://www.nature.com/magazine-assets/d41586-018-05849-9/d41586-018-05849-9.pdf` doi: 10.1038/d41586-018-05849-9

Seneviratne, S. I., Corti, T., Davin, E. L., Hirschi, M., Jaeger, E. B., Lehner, I., ... Teuling, A. J. (2010). Investigating soil moisture-climate interactions in a changing climate: A review. *Earth-Science Rev.*, *99*, 125–161. doi: 10.1016/j.earscirev.2010.02.004

Sippel, S., Otto, F. E. L., Flach, M., & Van Oldenborgh, G. J. (2016). The Role of Anthropogenic Warming in 2015 Central Euopean Heat Waves, In: Explaining Extremes of 2015 from a Climate Perspective. *Bull. Am. Meteorol. Soc.*, *97*, S51–S55. doi: 10.1175/BAMS-D-16-0149

Sousa, P. M., Trigo, R. M., Barriopedro, D., Soares, P. M., & Santos, J. A. (2018). European temperature responses to blocking and ridge regional patterns. *Clim. Dyn.*, *50*(1), 457–477. doi: 10.1007/s00382-017-3620-2

Stott, P. A., Christidis, N., Otto, F. E., Sun, Y., Vanderlinden, J. P., van Oldenborgh, G. J., ... Zwiers, F. W. (2015). Attribution of extreme weather and climate-related events. *Wiley Interdiscip. Rev. Clim. Chang.*, *7*, 23–41. doi: 10.1002/wcc.380

Stott, P. A., Stone, D. A., & Allen, M. R. (2004). Human Contribution to the European Heatwave of 2003. *Lett. to Nat.*, *432*, 610–614. doi: 10.1029/2001JB001029

Teuling, A. J., Hirschi, M., Ohmura, A., Wild, M., Reichstein, M., Ciais, P., ... Seneviratne, S. I. (2009). A regional perspective on trends in continental evaporation. *Geophys. Res. Lett.*, *36*, L02404. doi: 10.1029/2008GL036584

Tibaldi, S., & Molteni, F. (1990). On the operational predictability of blocking. *Tellus*, *42A*, 343–365. doi: 10.1034/j.1600-0870.1990.t01-2-00003.x

Trigo, R. M., García-Herrera, R., Díaz, J., Trigo, I. F., & Valente, M. A. (2005). How exceptional was the early August 2003 heatwave in France? *Geophys. Res. Lett.*, *32*, L10701. doi: 10.1029/2005GL022410

Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C., ... Zhang, L. (2017). The subseasonal to seasonal (S2S) prediction project database. *Bull. Am. Meteorol. Soc.*, *98*, 163–173. doi: 10.1175/BAMS-D-16-0017.1

Weisheimer, A., Doblas-Reyes, F. J., Jung, T., & Palmer, T. N. (2011). On the predictability of the extreme summer 2003 over Europe. *Geophys. Res. Lett.*, *38*,

L05704. doi: 10.1029/2010GL046455

White, C. J., Carlsen, H., Robertson, A. W., Klein, R. J., Lazo, J. K., Kumar, A., ... Zebiak, S. E. (2017). Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorol. Appl.*, *24*(3), 315–325. doi: 10.1002/met.1654

Wilks, D. S. (2011). *Chapter 8: Forecast Verification* (3rd ed.). Academic Press. doi: 10.1016/B978-0-12-385022-5.00008-7

Wulff, C. O., Greatbatch, R. J., Domeisen, D. I. V., Gollan, G., & Hansen, F. (2017). Tropical Forcing of the Summer East Atlantic Pattern. *Geophys. Res. Lett.*, *44*, 11,166–11,173. doi: 10.1002/2017GL075493