

Skilful sub-seasonal forecasts of aggregated temperature over Europe

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Baker, L. ORCID: <https://orcid.org/0000-0003-0738-9488>, Charlton-Perez, A. and Mattu, K. L. (2023) Skilful sub-seasonal forecasts of aggregated temperature over Europe. *Meteorological Applications*, 30 (6). e2169. ISSN 1469-8080 doi: <https://doi.org/10.1002/met.2169> Available at <https://centaur.reading.ac.uk/114443/>

It is advisable to refer to the publisher's version if you intend to cite from the work. See [Guidance on citing](#).

To link to this article DOI: <http://dx.doi.org/10.1002/met.2169>

Publisher: Royal Meteorological Society

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the [End User Agreement](#).

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online



Skilful sub-seasonal forecasts of aggregated temperature over Europe

Laura Baker¹  | Andrew Charlton-Perez²  | Kanzis L. Mattu³ 

¹National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Reading, UK

²Department of Meteorology, University of Reading, Reading, UK

³Department of Civil and Environmental Engineering, University of Strathclyde, Glasgow, UK

Correspondence

Laura Baker, National Centre for Atmospheric Science, Department of Meteorology, University of Reading, Earley Gate, Reading RG6 6UR, UK.
 Email: l.h.baker@reading.ac.uk

Funding information

Ceres-Agritech

Abstract

Subseasonal-to-seasonal (S2S) forecasts span the prediction range of weeks to 2–3 months ahead, bridging the gap between medium-range and seasonal weather forecasts. There has been growing interest in S2S forecasts in recent years, largely because of the many potential uses of forecasts spanning these timescales. However, the skill of S2S forecasts beyond the first 2 weeks or so is poor, potentially limiting the usability of these forecasts. We show in this study that when considering accumulated temperatures, there is in fact good forecasting skill over Europe for accumulation periods up to 30 days ahead. Using a set of S2S hindcasts, we show using both a deterministic and a probabilistic measure of skill that the accumulated 2-metre temperature forecasts out to 30 days are skilful over most of Europe. In summer, South West Europe has highest skill, while in winter North East Europe has highest skill. As an example application of such forecasts, we also evaluate the skill for summer cooling degree-days (CDD) and winter heating degree-days (HDD). For 30-day winter HDD, there is good skill in all four European regions; for 30-day summer CDD, the skill is limited in North West Europe, but still good in other regions.

KEYWORDS

ensemble forecasts, Europe, evaluation, forecasting, medium range, modelling, probabilistic forecasts, seasonal, sub-seasonal forecasting, temperature, verification

1 | INTRODUCTION

Sub-seasonal-to-seasonal (S2S) forecasts span the prediction range of weeks to 2–3 months ahead, bridging the gap between medium-range and seasonal weather forecasts. There has been growing interest in S2S forecasts in recent years, largely because of the many potential uses of forecasts spanning these timescales, for example, in the energy sector, insurance industry, public health, agriculture and water management (e.g., Büeler et al., 2021; Weigel et al., 2008; White et al., 2017, 2022). However,

forecasting on these timescales is challenging since it is a mixture between the more typical weather and seasonal forecasting problems. Sub-seasonal forecasts are influenced by both atmospheric initial conditions and more slowly varying boundary conditions. In the extratropics, the skill of S2S forecasts beyond the first 2 weeks or so is generally reported to be fairly limited (e.g., Büeler et al., 2021; Cortesi et al., 2021; Cui et al., 2021; Son et al., 2020; Vitart, 2014; Weigel et al., 2008). These studies often focus on forecasts of weekly means. Predictability on these timescales and at these lead-times is

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Meteorological Applications* published by John Wiley & Sons Ltd on behalf of Royal Meteorological Society.

inherently limited by the chaotic nature of the atmosphere, and therefore, the skill for forecasting weekly or sub-weekly timescales is very limited beyond the first 2 weeks. However, there is potentially more skill when considering longer averaging periods, such as means or accumulations over 1 month. Buizza and Leutbecher (2015) and Büeler et al. (2020) showed that good skill can be obtained over Europe for the monthly mean temperature (i.e., the mean temperature over the first month of the forecast). In the present study, we focus on the accumulated temperature (i.e., daily temperatures summed over a period of several days); it is worth noting that the skill obtained for these quantities is the same as that for the mean temperature averaged over the corresponding periods. The aim of reframing the results in terms of accumulated temperatures is to highlight potential applications of S2S forecasts that could utilise accumulated temperatures over lead-times of up to a month.

There are various applications that use aggregated temperature information over the sub-seasonal period, for example the accumulated temperature over 1 month. An example in the field of agriculture is forecasting crop growth rates and harvest times. Pearson et al. (1994) developed a model of curd growth of cauliflower crops based on the thermal time, or accumulated degree-days (as defined by the accumulated daily mean temperature), and the curd initiation date. Using this model, it was possible to produce forecasts of the dates that the crop would be ready to harvest. The ability to accurately forecast crop harvest dates and yields has the potential to optimise the usage of the crops and reduce wastage in the production and supply chain. A second example, in the field of energy production, is to estimate the heating and cooling degree-days (HDD and CDD, respectively) aggregated over the following weeks. HDD and CDD can have a strong impact on energy demand from buildings (Atalla et al., 2018). Accurate forecasts of heating and cooling degree-days could be used to make forecasts of energy demand, which would enable energy companies to preempt increased or decreased demand in advance (White et al., 2017). Another application in terms of energy production capability would be to forecast persistent heatwaves which will increase river temperatures and potentially impact availability of water for cooling in thermoelectric power generation (Chandramowli & Felder, 2014; van Vliet et al., 2012). HDD and CDD are also important considerations in the field of public health: being able to forecast periods of high HDD or CDD is beneficial in terms of planning for and mitigating the negative impacts that prolonged hot or cold periods can have on health (Charlton-Perez et al., 2019; Gasparrini et al., 2015).

While the focus of the present study is on accumulated temperature, it is worth noting that a similar methodology could be applied to other quantities such as precipitation. Forecasts of precipitation accumulated over a 1-month period would be of use for prediction of hydroelectric power generation, or for water management including drought planning or flood management and mitigation.

The aim of this paper is to demonstrate that useful skill can be obtained from S2S forecasts when aggregated, or accumulated, temperatures are considered. We focus on the 15- and 30-day accumulated daily-mean 2-metre temperatures over Europe. We also evaluate the skill of the forecasts of HDD and CDD over Europe, as an example of a more complex metric derived from 2-metre temperatures.

In Section 2 the S2S data, observation data and the skill metrics used are described, and the calculation of HDD and CDD are defined. In Section 3, the evaluation of skill for both the aggregated 2-metre temperature and the HDD and CDD are presented and discussed. Finally, conclusions are given in Section 4.

2 | DATA AND METHODOLOGY

2.1 | Data

We use sub-seasonal hindcast (or re-forecast) data from the ECMWF forecasting system, obtained from the S2S database which is described in detail by Vitart et al. (2017). The version of the forecast system used is CY46R1, which was the operation system between 11 June 2019 and 30 June 2020. This system was chosen simply because it had a full year of forecast dates available, rather than due to any special features of the system or a suggestion that the sub-seasonal skill available was atypical. Forecasts were initialised twice a week, giving a total of 110 start-dates during this period. The hindcast data consists of 20 years of hindcasts initialised on the corresponding dates for each of the preceding 20 years; for example, corresponding to the 11 June 2019 forecast start-date, there are hindcasts initialised on 11 June in each year between 1999 and 2018. The hindcasts consist of 11 ensemble members and are run for 46 days from each initialisation date. We note that the operational forecasts have a much larger ensemble size of 51 ensemble members, and so it is expected that the results presented here, which are based on the hindcasts, represent a lower bound on the skill that would be obtained from the actual operational forecasts. The forecast model grid-point resolution is roughly 18 km for the first 15 days and then reduces to 36 km from day 15 onwards, with

91 vertical levels. The data in the S2S database are regridded to a $1.5^\circ \times 1.5^\circ$ grid, which is what is used in the present study. Full details of the forecast model are given by Vitart et al. (2017) and can also be found at <https://www.ecmwf.int/en/forecasts/documentation-and-support>.

The observation dataset used is E-Obs version 23.1e (Cornes et al., 2018). E-Obs is a gridded dataset obtained from interpolating daily station-based observations onto a regular grid. For each grid-point, an ensemble of estimates is produced using a stochastic technique. Full details are given in Cornes et al. (2018). In the present study, only the ensemble mean on the 0.25° grid is used. For the purposes of comparison with the hindcasts, the data were regridded to the 1.5° grid of the hindcasts.

For both the hindcasts and observations, the daily-mean 2-metre temperature field is used. For the calculation of HDD and CDD, the daily maximum and daily minimum 2-metre temperatures are also used.

2.2 | Skill evaluation metrics

In Section 3, an example deterministic and probabilistic skill metric are used, namely, the anomaly correlation coefficient (ACC) and the continuous ranked probability skill score (CRPSS), respectively. Different metrics are likely to give different results and highlight different aspects of the forecast skill. These metrics are chosen simply as broad illustrative examples.

For each hindcast start date, there are 20 years of hindcasts, and the ACC and CRPSS were calculated over these 20 years. Seasonal averages of these values were then computed as needed.

The ACC is a measure of ensemble mean correlation skill and is defined as

$$\text{ACC} = \frac{\sum_{i=1}^m (\bar{x}_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (\bar{x}_i - \bar{x})^2 \sum_{i=1}^m (y_i - \bar{y})^2}} \quad (1)$$

where y_i is the observed value in year i , \bar{x}_i is the ensemble mean hindcast in year i , \bar{x} is the mean over time of the hindcast ensemble mean value and m is the number of years. Since the ACC is based on anomalies of the observations and ensemble mean hindcast relative the time-means of the respective quantities, the data is effectively bias-corrected within this calculation. ACC can range from -1 to 1 , with 1 indicating a perfect skill score, and negative values indicating no skill. A threshold of 0.6 is often used to consider as skilful, although this depends somewhat on the application.

The ACC is useful for evaluating how well the hindcast ensemble mean represents variations in the observed quantity, but it is not a skill score in a classical sense because it does not compare the hindcasts against a reference hindcast; the CRPSS is a skill score, and in this context, we evaluate the hindcasts against the observed climatology.

Before calculating the CRPSS, the hindcast data is first bias-corrected. For each of the 110 hindcast start-dates, the ensemble climatology and observed climatology of the accumulated temperature were computed, using the leave-one-out method to exclude the year in question. The bias correction was done by subtracting the difference between the ensemble climatology and observed climatology for the corresponding date. The climatologies were calculated for the accumulated temperature values rather than individual daily temperatures.

The continuous ranked probability score (CRPS) is a measure of the difference between the hindcast and observed cumulative density functions (CDFs) (Hersbach, 2000). For a quantity x , we define $\rho(x)$ to be the probability density function (PDF) of forecasts of x , and x_{obs} to be the observed value of x . Then CRPS over m hindcast years is given by

$$\text{CRPS} = \frac{1}{m} \sum_{i=1}^m \int_{-\infty}^{\infty} (P(x) - P_{\text{obs}}(x))^2 dx \quad (2)$$

where P and P_{obs} are cumulative distributions given by

$$P(x) = \int_{-\infty}^x \rho(y) dy \quad (3)$$

and

$$P_{\text{obs}}(x) = H(x - x_{\text{obs}}) \quad (4)$$

where

$$H(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0. \end{cases} \quad (5)$$

The CRPSS is then calculated as

$$\text{CRPSS} = 1 - \frac{\text{CRPS}}{\text{CRPS}_{\text{clim}}} \quad (6)$$

where $\text{CRPS}_{\text{clim}}$ is the CRPS for the observed climatology. Values of CRPSS greater than 0 indicate that the forecast skill is an improvement over climatology, and a perfect forecast would have $\text{CRPSS} = 1$. As the CRPSS is sensitive

to ensemble size (Ferro et al., 2008), we use the fair CRPSS, which adjusts the CRPSS to correct for small ensemble sizes. The calculation of CRPSS was done using the Python package CRPS (<https://pypi.org/project/CRPS/>).

To provide an estimate of the uncertainty in the results, the 20 years of hindcasts were resampled 1000 times using bootstrap sampling with replacement, and the skill metrics were computed for each resampled set. The 5%–95% confidence interval over the 1000 samples was then computed.

2.3 | Heating and cooling degree-days

Heating and cooling degree-days (HDD and CDD, respectively) are useful metrics for estimating changes in weather-related energy demand and also for predicting temperature-related public health issues. HDD and CDD are essentially a measure of how much the temperature has deviated from a reference temperature (Spinoni

et al., 2015). HDD, which is typically calculated in the winter half-year or in cooler climates, gives an indication of the amount of energy required to heat the interior of a building to a specified base temperature, over a given period of one or more days; CDD, which is typically calculated in the summer half-year or in warmer climates, gives an indication of the amount of energy required to cool the interior of a building down to a specified base temperature, over a given period of one or more days. Here we use the definitions of HDD and CDD given in Spinoni et al. (2018), which are based on the UK Met Office equations (other studies may use different definitions). Following Spinoni et al. (2018), the base temperature T_{base} for HDD is 15.5 °C, and for CDD is 22 °C. The quantities HDD and CDD are given by

$$\text{HDD} = \begin{cases} T_{\text{base}} - T_{\text{mean}} & \text{if } T_{\text{max}} \leq T_{\text{base}} \\ (T_{\text{base}} - T_{\text{min}})/2 - (T_{\text{max}} - T_{\text{base}})/4 & \text{if } T_{\text{mean}} \leq T_{\text{base}} < T_{\text{max}} \\ (T_{\text{base}} - T_{\text{min}})/4 & \text{if } T_{\text{min}} < T_{\text{base}} < T_{\text{mean}} \\ 0 & \text{if } T_{\text{min}} \geq T_{\text{base}} \end{cases} \quad (7)$$

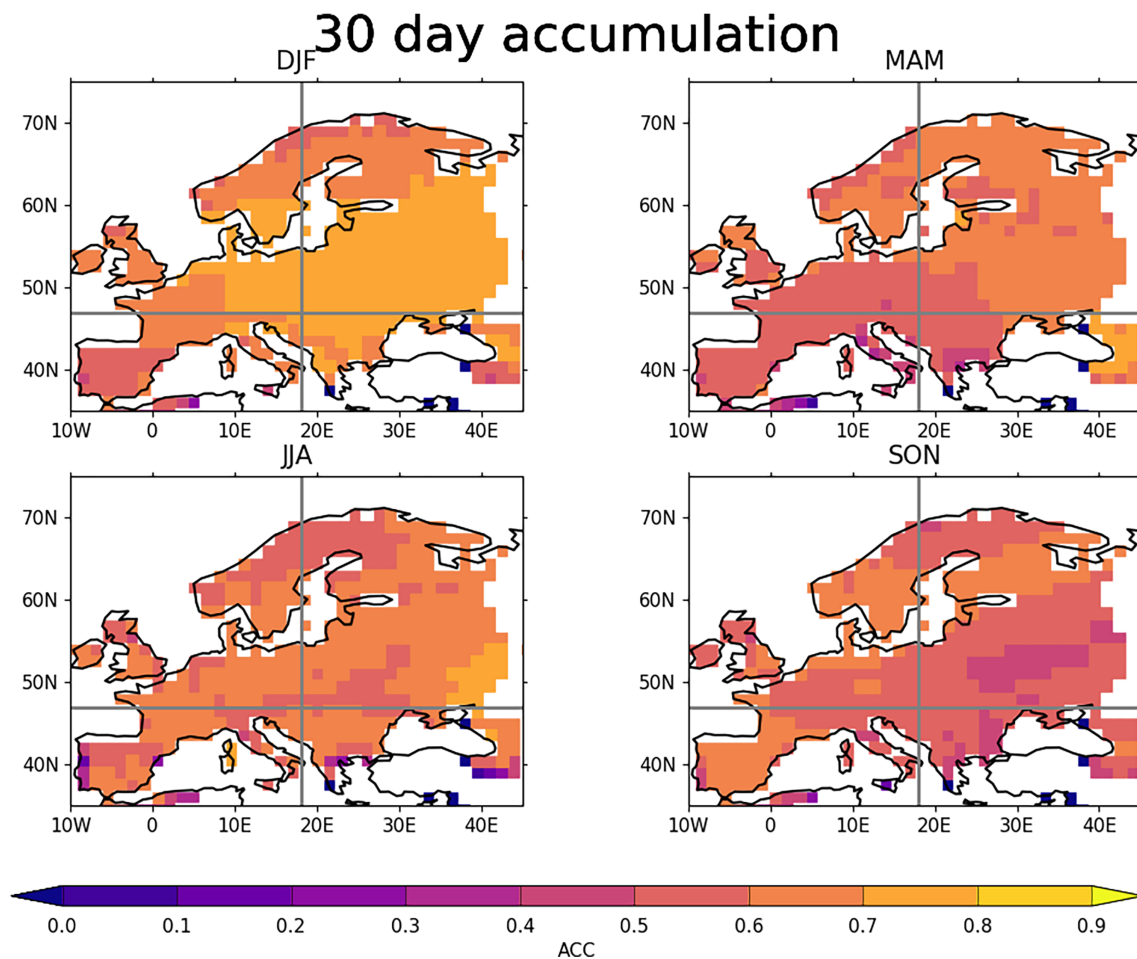


FIGURE 1 Anomaly correlation coefficient (ACC) for the 30-day accumulated 2-metre temperature in each season at each hindcast model land grid-point, for the hindcast ensemble mean evaluated against E-Obs observed temperatures. Grey lines show the division of the domain into four European sub-regions.

and

$$\text{CDD} = \begin{cases} 0 & \text{if } T_{\max} \leq T_{\text{base}} \\ (T_{\max} - T_{\text{base}})/4 & \text{if } T_{\text{mean}} \leq T_{\text{base}} < T_{\max} \\ (T_{\max} - T_{\text{base}})/2 - (T_{\text{base}} - T_{\min})/4 & \text{if } T_{\min} < T_{\text{base}} < T_{\text{mean}} \\ T_{\text{mean}} - T_{\text{base}} & \text{if } T_{\min} \geq T_{\text{base}}. \end{cases} \quad (8)$$

where T_{\min} , T_{\max} and T_{mean} are the daily minimum, maximum and mean temperatures, respectively. HDD and CDD for a period of, for example, 30 days, is computed by summing the daily HDD/CDD over the period.

3 | RESULTS

3.1 | Deterministic skill evaluation of the hindcast ensemble mean

The ACC for the 30-day accumulated 2-metre temperature for each season is shown in Figure 1. The 30-day accumulation is calculated by summing daily mean

temperatures over days 1 to 30, in both the hindcasts and observations. The seasons correspond to the start of each accumulation period; for example, the winter season DJF corresponds to dates with the accumulation period starting within the months December, January and February. In all seasons, there is significant skill over much of Europe. The skill in DJF is particularly high, with ACC exceeding 0.7 over much of Eastern Europe. SON is the season with the poorest skill overall, although for the Iberian Peninsula, this season has the highest ACC skill. Equivalent ACC for the 15-day accumulated 2-metre temperature is shown in Figure A.1. The skill is higher than for the 30-day accumulation (as expected), with ACC values exceeding 0.8 over most of Europe in all four seasons. Figure 2 summarises the ACC skill for each of the four sub-regions of Europe (as marked in Figure 1), for accumulation periods of 8, 15, 22 and 30 days (i.e., roughly week 1, weeks 1–2, weeks 1–3 and weeks 1–4). For the 8-day accumulation, the ACC is close to 1 for all regions and all seasons. As each additional week is added into the accumulation period, the skill decreases at a roughly constant rate for each region. In DJF and

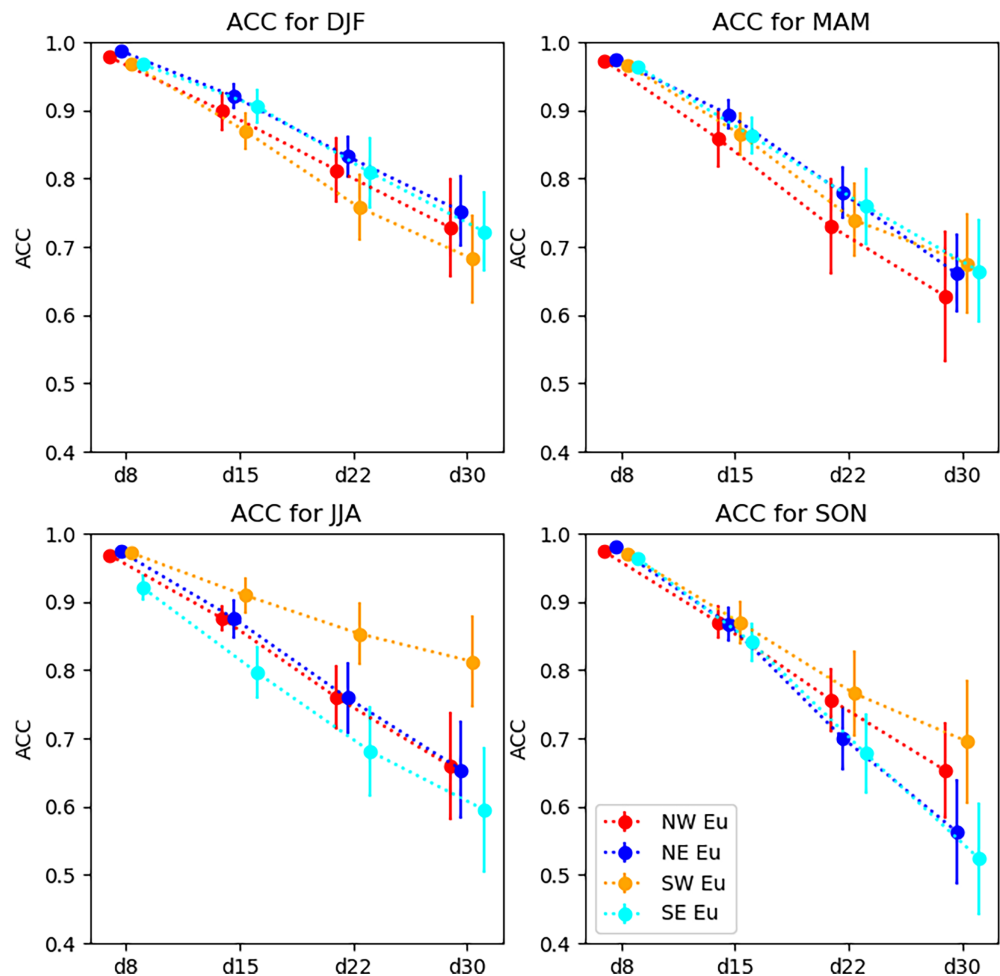


FIGURE 2 Anomaly correlation coefficient (ACC) for each European region, for each season: (a) DJF, (b) MAM, (c) JJA and (d) SON. The x-axis indicates the different accumulation periods (day 0–8, day 0–15, day 0–22 and day 0–30). The ACC values are for the hindcast ensemble mean evaluated against E-Obs observed temperatures. Colours indicate the four European sub-regions (see legend). Error bars show the 5%–95% confidence interval over 1000 bootstrap samples (see Section 2.2 for details).

MAM, all regions have fairly similar skill, and the mean skill remains above 0.6 even for accumulation periods of 30 days. In JJA and SON, SE Europe has the lowest skill. In these seasons, SW Europe has the highest skill, with notably high skill exceeding 0.8 in JJA even out to 30-day accumulations.

In order to confirm that the skill is not all simply coming from the first 2 weeks of the hindcasts, Figure A.3a shows the ACC for the day 1–30, day 1–15 and day 15–30 temperature accumulations in each region. The ACC for the day 15–30 accumulations is above 0 for all regions and all seasons, indicating that there is some skill for this 15-day accumulation period even 2 weeks after the hindcasts are initialised.

Overall, these results indicate that the ensemble mean hindcasts of 15- and 30-day accumulated 2-metre temperature perform well in all seasons, with significant skill at the regional scale in all four sub-regions of Europe, and at the grid-point scale almost everywhere.

3.2 | Probabilistic skill evaluation of the hindcast ensemble

To provide a probabilistic skill measure, the CRPSS was calculated for the hindcast ensemble evaluated against E-Obs. As a baseline, the observed climatology was used. The hindcasts were bias-corrected as described in Section 2.

Maps of CRPSS for the 30-day accumulated 2-metre temperature in each season are shown in Figure 3. CRPSS values above 0 indicate that there is more skill in the hindcast than simply using the observed climatology. CRPSS values exceed 0.3 almost everywhere in all seasons, with the exception of some points towards the south of the domain. The spatial and temporal patterns are similar to those seen for ACC (Figure 1), with the highest CRPSS in DJF, particularly in Eastern Europe (values exceeding 0.5). Equivalent maps of CRPSS for the 15-day temperature accumulation are shown in Figure A.2. These show a similar spatial and temporal

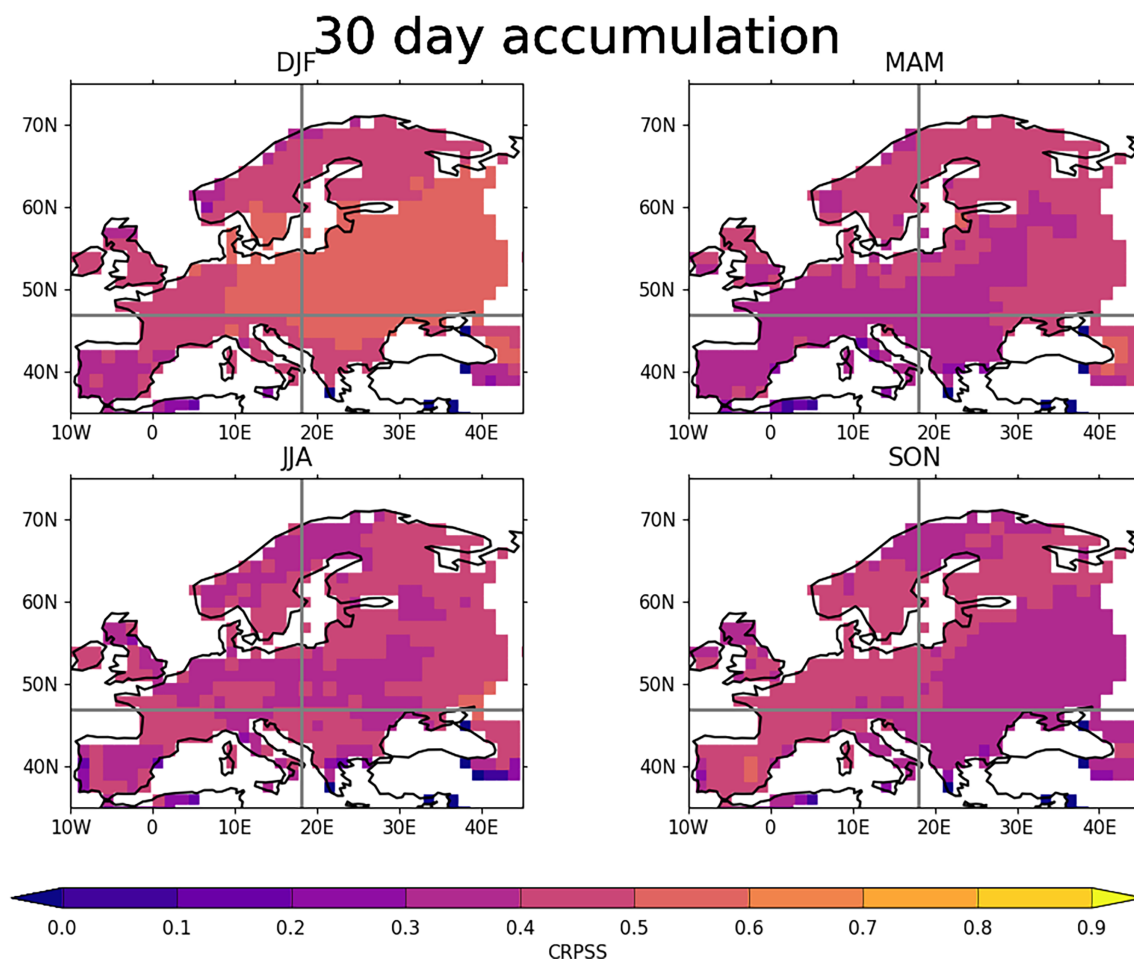


FIGURE 3 Continuous ranked probability skill score (CRPSS) for the 30-day accumulated 2-metre temperature in each season at each hindcast model land grid-point, for the hindcast ensemble evaluated against E-Obs observed temperatures. Grey lines show the division of the domain into four European sub-regions.

pattern of skill, but with higher values (exceeding 0.4 almost everywhere). The regional pattern of skill in DJF may be related to the North Atlantic Oscillation, which is the leading mode of variability in the North Atlantic region and impacts temperature in Northern and Eastern Europe. The higher skill found over Northern and Eastern Europe in DJF is in agreement with the results of Monhart et al. (2018) (their Figure 3, which shows skill for weekly-mean temperatures at 12–18 day lead-times). In other seasons, the spatial skill patterns are somewhat different however, and in particular Monhart et al. (2018) did not find the higher skill in South West Europe seen in Figure 3. These differences are likely due to the different forecast lead-times being evaluated. The CRPSS for the four European sub-regions in each season are summarised in Figure 4, for each weekly accumulation period. As seen for ACC (Figure 2), there is a steady decline in skill as each week is added to the accumulation period. SW Europe has the highest CRPSS in JJA, for all accumulation periods, while the hindcasts for SE Europe have the lowest skill. In DJF and MAM, all regions show similar values of CRPSS, with the best skill found in NE Europe. For the 30-day accumulation period, all regional-mean CRPSS values are above 0.3.

Figure A.3 shows that even for day 15–30 temperature accumulations, the CRPSS is positive, indicating that there is skill for 2-week accumulations even 15 days after the forecasts were initialised.

Overall, these results indicate that the ensemble hindcasts of 15-day and 30-day accumulated temperature are skilful at both the grid-point and regional scales, for all four seasons.

3.3 | Heating and cooling degree-days

The HDD and CDD were computed for the four sub-regions of Europe marked in Figure 1, for the winter half-year (October–March) and summer half-year (April–September), respectively. Figure 5 shows scatter plots of the hindcast against observed 30-day summer half-year CDD and winter half-year HDD. The hindcast data used to produce Figure 5 have not been bias-corrected. For the winter HDD (right column in Figure 5), there is very good correspondence between the observed and hindcast values in each of the four regions, with only a small positive bias seen in the hindcasts for SW Europe. These results suggest that even using the raw model output,

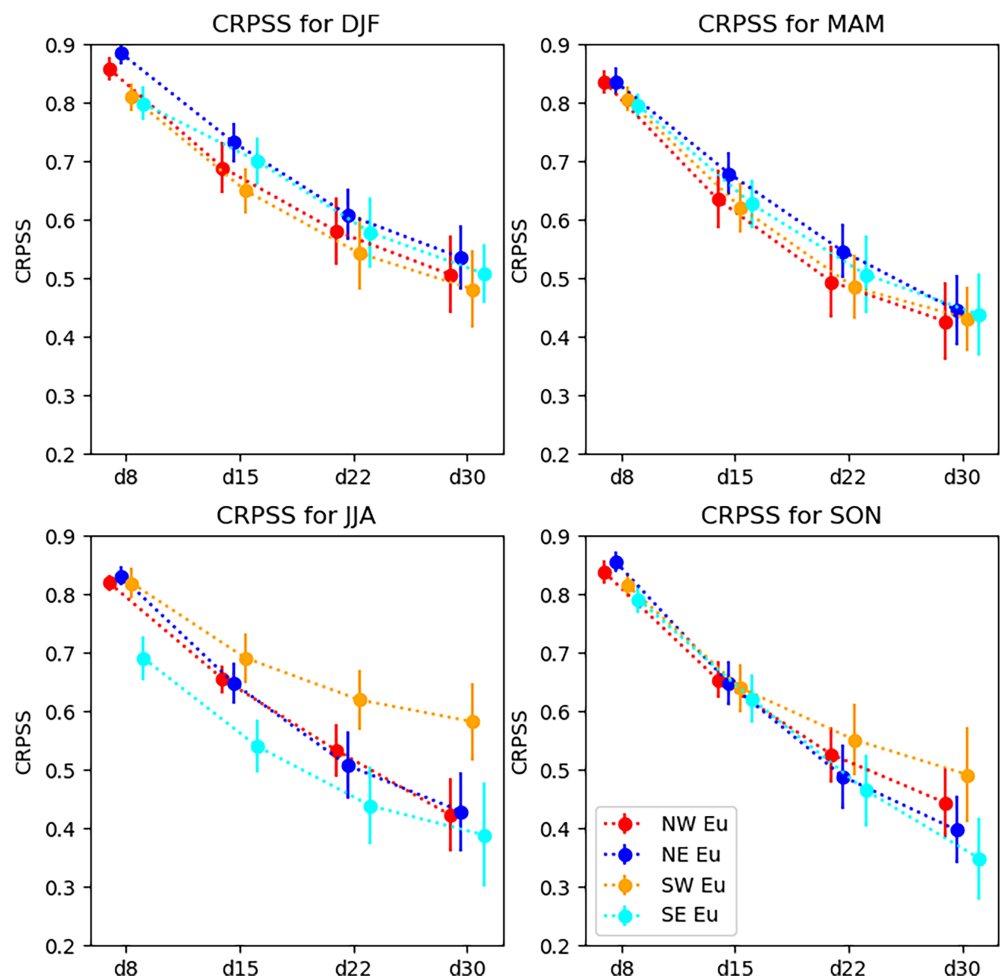


FIGURE 4 Continuous ranked probability skill score (CRPSS) for each European region, for each season: (a) DJF, (b) MAM, (c) JJA and (d) SON. The x-axis indicates the different accumulation periods (day 0–8, day 0–15, day 0–22 and day 0–30). The CRPSS values are for the hindcasts evaluated against E-Obs observed temperatures. Colours indicate the four European sub-regions (see legend). Error bars show the 5%–95% confidence interval over 1000 bootstrap samples (see Section 2.2 for details).

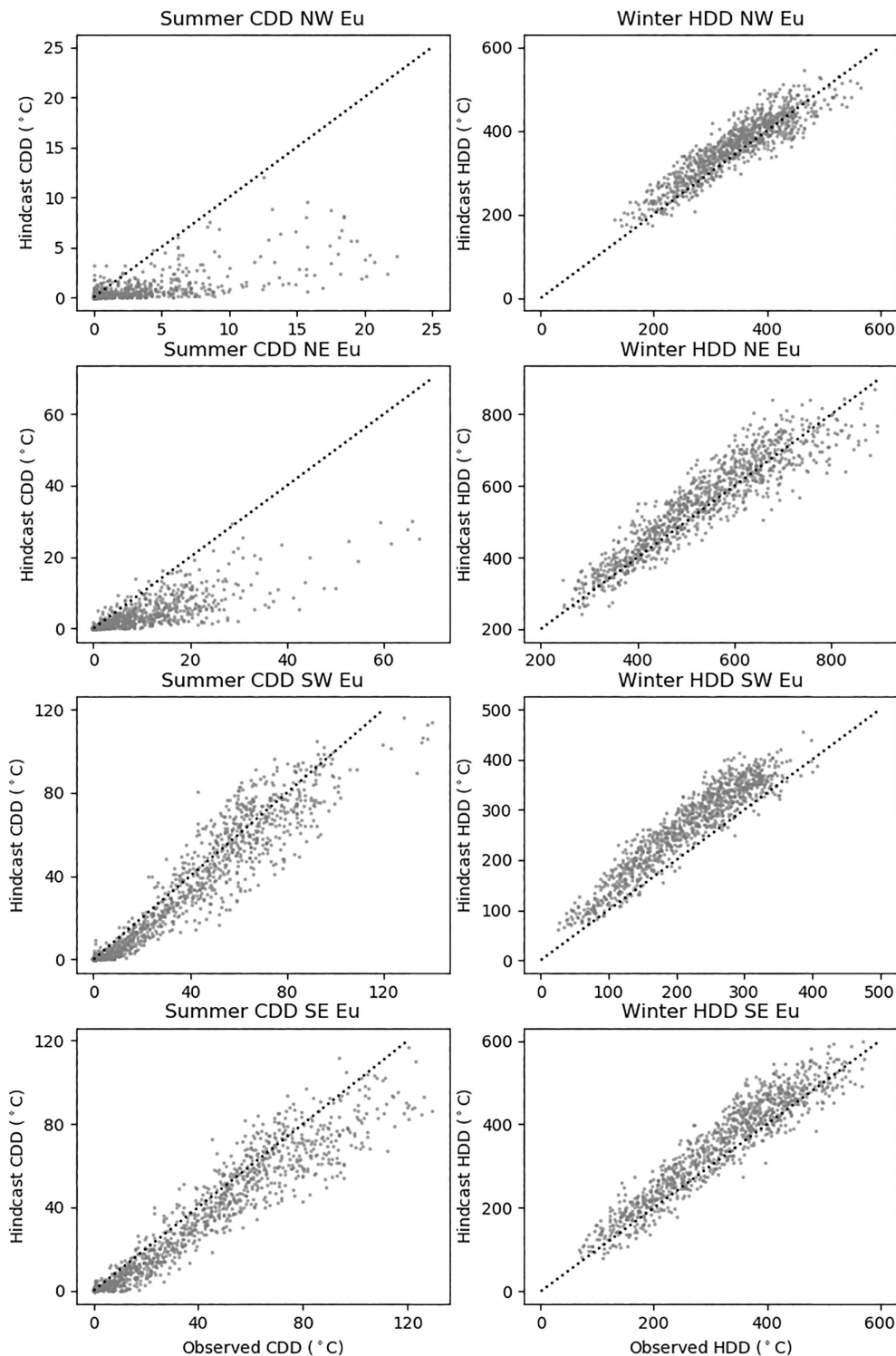
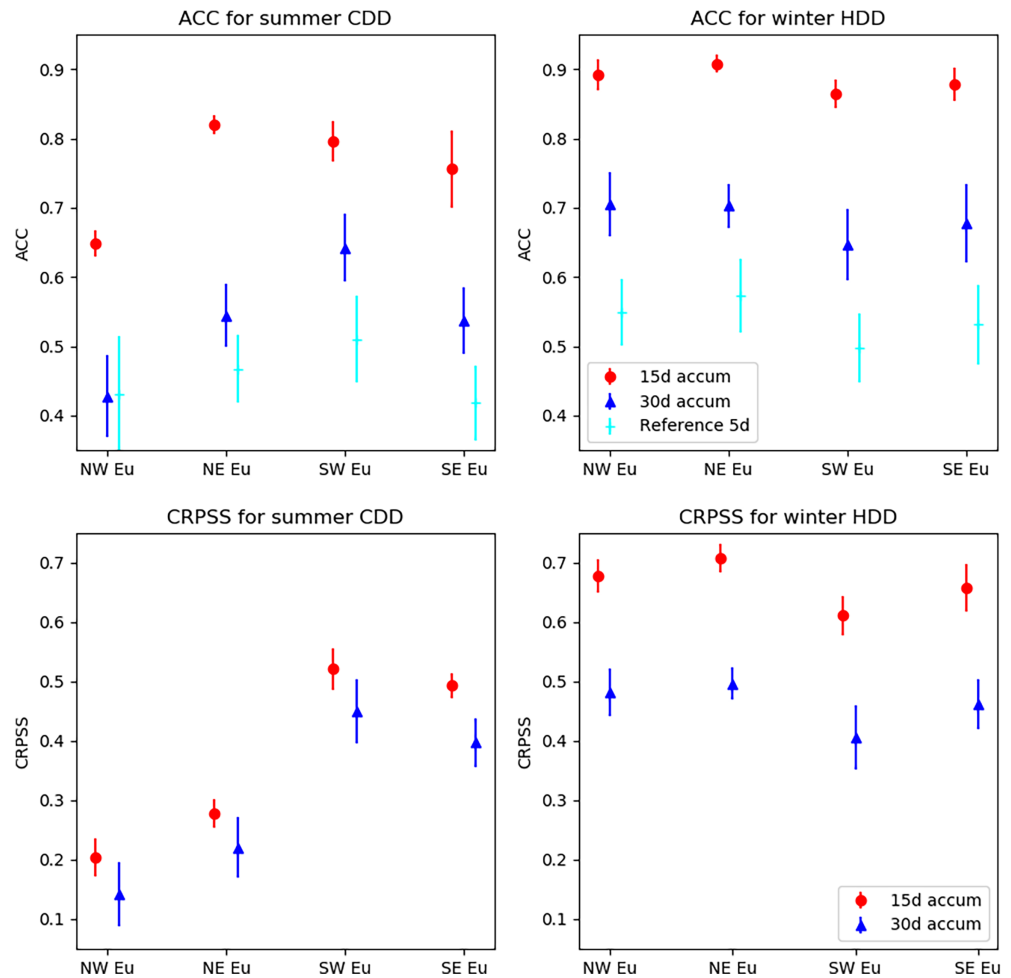


FIGURE 5 Scatter plots of hindcast ensemble mean versus observed summer half-year 30-day CDD (left column) and winter half-year 30-day HDD (right column), for four European regions.

without any post-processing, would give a reasonable estimate of 30-day accumulated HDD in the winter half-year. The summer CDD (left column in Figure 5) shows good correspondence between observed and hindcast values in SW and SE Europe, with a slight negative bias in the hindcasts for SE Europe. However, for NE Europe and NW Europe, there is a clear negative bias in the hindcasts. Similar results are also seen for the 15-day

accumulated CDD and HDD (not shown), and in particular, the negative bias in NE and NW Europe summer CDD is still present. We attribute this bias in summer CDD for the Northern European regions partly to the low proportion of days in which the CDD is non-zero: in NW Europe, only 15% of days, and in NE Europe, only 35% of days, have an observed CDD >0, compared with 77% in each of the Southern Europe regions. This means that

FIGURE 6 Skill scores for summer half-year cooling degree-days (left column) and winter half-year heating degree-days (right column) for four European regions. Top row: ACC; bottom row: CRPSS. Red circles show 15-day accumulations; blue triangles show 30-day accumulations. Cyan plus symbols in the top panels show ACC for 30-day accumulation reference forecasts, using the first 5 days of the hindcasts and the remaining 25 days from model climatology. Error bars show the 5%–95% confidence interval based on 1000 bootstrap samples (see Section 2.2).



the hot days included are extreme events, which are generally less predictable and in particular are less likely to be captured by the ensemble mean. The small number of days contributing to the CDD in the Northern European regions means that any small biases in temperature will lead to relatively large biases in the CDD. Reducing the CDD base threshold for these Northern Europe regions may reduce these biases.

The results from the ACC and CRPSS skill metrics are summarised in Figure 6, for the four European sub-regions and for both the 15- and 30-day accumulations. Here the CRPSS is calculated using bias-corrected hindcast values. As expected from the scatter plots discussed above, the skill for the winter HDD hindcasts is good, with ACC values exceeding 0.85 for the 15-day accumulations, and exceeding 0.65 for the 30-day accumulations. The CRPSS shows that the hindcasts also have good probabilistic skill, with values exceeding 0.55 for the 15-day accumulations and exceeding 0.35 for the 30-day accumulations. For the summer CDD, the ACC is lower for NW Europe than for other regions, but is still around 0.65 for the 15-day accumulations, and still reasonable at above 0.4 for the 30-day accumulations. The CRPSS for

summer CDD is lower in NW and NE Europe than in SW and SE Europe, but still shows considerable skill above climatology in these regions, even for the 30-day accumulations.

It is interesting to consider these results from an energy forecasting perspective. For this application, in the absence of S2S forecast data, typically the short-range (day 1–5) forecasts would be used, followed by climatology for the remainder of the required forecast period. In order to assess how much added value is obtained by using the sub-seasonal forecasts out to 30 days, instead of just the short-range forecasts, the top panels in Figure 6 include the ACC for a reference forecast of 30-day accumulated summer CDD and winter HDD, which uses the hindcasts for day 1–5, and the hindcast climatology for the remaining 25 days. For winter HDD, using the full 30-days of hindcasts clearly adds skill compared with reference forecast, in all regions. For summer CDD, this is true in both Southern Europe regions and NE Eu. However, in NW Eu the full 30-day hindcast gives very similar results to the reference forecast. This is likely due to the low bias in the CDD in this region seen in Figure 5. A bias correction before the calculation of CDD

may correct this and improve the skill obtained by the hindcasts, but this is not tested in the present study. Figure A.4 shows the equivalent results for additional reference forecasts, with hindcasts for the first 10 days and the first 15 days, followed by climatology for the remainder of the forecasts. These show that, for the winter HDD and summer CDD, there is only a small amount of skill gained by using the hindcast data instead of climatology after the first 10 days, and no skill gained by using the forecast data after 15 days instead of climatology; and in the case of NW Eu, the reference forecasts actually do better than the full 30-day hindcasts.

The results presented here are for HDD and CDD with specific values of the base temperatures for each quantity. Improved results might be obtained by optimising these base temperature thresholds for different regions. In addition, a similar method could be used for other applications, such as forecasting the cumulative number of days the temperature is above or below zero, which would be useful for forecasting crop growth or optimising crop planting times.

4 | CONCLUSIONS

In this paper, we have shown that for a typical sub-seasonal forecasting system (ECMWF), there is significant skill for forecasting 15- and 30-day accumulated 2-metre temperature over Europe. Both a deterministic skill measure (ACC) of the ensemble mean forecasts, and a probabilistic skill measure (CRPSS) were evaluated. For both measures, good skill was found over most of Europe even out to 30-day accumulations. As an example of a specific application of these results, the skill for forecasting winter HDD and summer CDD over the same accumulation periods was also assessed. High levels of skill were found for these quantities in each sub-region of Europe, especially for winter HDD. In summer, the hindcasts showed a low bias in the CDD in Northern Europe, due to the low frequency of events with temperatures exceeding the CDD base temperature in these regions, meaning that these events were more extreme and less likely to be captured by the ensemble mean hindcasts. Skill for summer CDD in SW Europe was, however, much higher.

While other publications showing sub-seasonal forecast skill generally focus on weekly means and show a decrease in skill over time, in particular generally finding very little skill beyond about the first 2 weeks, an important message from the present study is that skill for forecasting the accumulated quantities is still relatively high even out to accumulation periods of 30 days. Even at lead-times of 15 days, there was still reasonable skill found for

15-day temperature accumulations (i.e., accumulation over days 15–30).

It is worth considering the reason for the relatively higher skill for 30-day accumulated temperatures, compared with the weekly or sub-weekly forecast skill for these longer lead-times. A sub-seasonal forecast might correctly predict, for example, that a series of cyclones or cold air outbreaks will pass over Europe over the forecast period, but not when exactly the individual systems will cross Europe. Thus, the accuracy on shorter periods of the forecasts would be low, but the overall temperature impact of these events over the whole period would be more accurate.

Given the potential applications of these accumulated temperature forecasts detailed in the Introduction, including crop growth forecasts and energy demand forecasts, it is hoped that these results will encourage confidence in the skill of such sub-seasonal forecasts, and prompt greater use of these forecasts.

AUTHOR CONTRIBUTIONS

Laura Baker: Conceptualization (equal); formal analysis (lead); funding acquisition (supporting); investigation (lead); methodology (lead); software (lead); visualization (lead); writing – original draft (lead); writing – review and editing (lead). **Andrew Charlton-Perez:** Conceptualization (lead); project administration (equal); supervision (lead); writing – review and editing (equal). **Kanzis L. Mattu:** Formal analysis (supporting); methodology (supporting); writing – review and editing (supporting).

ACKNOWLEDGMENTS

The authors would like to thank two anonymous reviewers for their helpful comments which greatly improved the paper. We also thank David Brayshaw for his input and comments on an earlier draft of the paper.

FUNDING INFORMATION

LB and KLM were funded by Ceres Agri-tech, under the VegCast project.

DATA AVAILABILITY STATEMENT

The S2S data can be obtained from the S2S data portal on the ECMWF website: <https://apps.ecmwf.int/datasets/>. E-Obs data can be obtained from <https://www.ecad.eu/download/ensembles/download.php>.

ORCID

Laura Baker  <https://orcid.org/0000-0003-0738-9488>

Andrew Charlton-Perez  <https://orcid.org/0000-0001-8179-6220>

Kanzis L. Mattu  <https://orcid.org/0000-0001-8109-9484>

REFERENCES

- Atalla, T., Gualdi, S. & Lanza, A. (2018) A global degree days database for energy-related applications. *Energy*, 143, 1048–1055. URL. Available from: <https://www.sciencedirect.com/science/article/pii/S0360544217318388>
- Büeler, D., Beerli, R., Wernli, H. & Grams, C.M. (2020) Stratospheric influence on ECMWF sub-seasonal forecast skill for energy-industry-relevant surface weather in European countries. *Quarterly Journal of the Royal Meteorological Society*, 146, 3675–3694. Available from: <https://doi.org/10.1002/qj.3866>
- Büeler, D., Ferranti, L., Magnusson, L., Quinting, J.F. & Grams, C.M. (2021) Year-round sub-seasonal forecast skill for Atlantic-European weather regimes. *Quarterly Journal of the Royal Meteorological Society*, n/a. URL, 147, 4283–4309. Available from: <https://doi.org/10.1002/qj.4178>
- Buizza, R. & Leutbecher, M. (2015) The forecast skill horizon. *Quarterly Journal of the Royal Meteorological Society*, 141, 3366–3382. Available from: <https://doi.org/10.1002/qj.2619>
- Chandramowli, S.N. & Felder, F.A. (2014) Impact of climate change on electricity systems and markets—a review of models and forecasts. *Sustainable Energy Technologies and Assessments*, 5, 62–74.
- Charlton-Perez, A.J., Aldridge, R.W., Grams, C.M. & Lee, R. (2019) Winter pressures on the UK health system dominated by the Greenland blocking weather regime. *Weather and Climate Extremes*, 25, 100218. URL. Available from: <https://www.sciencedirect.com/science/article/pii/S2212094718302147>
- Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M. & Jones, P.D. (2018) An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research-Atmospheres*, 123, 9391–9409. URL. Available from: <https://doi.org/10.1029/2017JD028200>
- Cortesi, N., Torralba, V., Lledó, L., Manrique-Suñén, A., Gonzalez-Reviriego, N., Soret, A. et al. (2021) Yearly evolution of euro-Atlantic weather regimes and of their sub-seasonal predictability. *Climate Dynamics*, 56, 3933–3964.
- Cui, J., Yang, S. & Li, T. (2021) How well do the S2S models predict intraseasonal wintertime surface air temperature over mid-high-latitude Eurasia? *Climate Dynamics*, 57, 503–521.
- Ferro, C.A.T., Richardson, D.S. & Weigel, A.P. (2008), On the effect of ensemble size on the discrete and continuous ranked probability scores. *Meteorological Applications*, 15, 19–24. <https://doi.org/10.1002/met.45>
- Gasparrini, A., Guo, Y., Hashizume, M., Lavigne, E., Zanobetti, A., Schwartz, J. et al. (2015) Mortality risk attributable to high and low ambient temperature: a multicountry observational study. *The Lancet*, 386, 369–375. URL. Available from: <https://www.sciencedirect.com/science/article/pii/S0140673614621140>
- Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15, 559–570. URL. Available from: https://journals.ametsoc.org/view/journals/wefo/15/5/1520-0434_2000_015_0559_dotcrp_2_0_co_2.xml
- Monhart, S., Spirig, C., Bhend, J., Bogner, K., Schär, C. & Liniger, M.A. (2018) Skill of subseasonal forecasts in Europe: effect of bias correction and downscaling using surface observations. *Journal of Geophysical Research-Atmospheres*, 123, 7999–8016. Available from: <https://doi.org/10.1029/2017JD027923>
- Pearson, S., Hadley, P. & Wheldon, A. (1994) A model of the effects of temperature on the growth and development of cauliflower (Brassica oleracea L. botrytis). *Scientia Horticulturae*, 59, 91–106. URL. Available from: <https://www.sciencedirect.com/science/article/pii/0304423894900760>
- Son, S.-W., Kim, H., Song, K., Kim, S.-W., Martineau, P., Hyun, Y.-K. et al. (2020) Extratropical prediction skill of the subseasonal-to-seasonal (S2S) prediction models. *Journal of Geophysical Research-Atmospheres*, 125, e2019JD031273. URL. Available from: <https://doi.org/10.1029/2019JD031273>
- Spinoni, J., Vogt, J. & Barbosa, P. (2015) European degree-day climatologies and trends for the period 1951–2011. *International Journal of Climatology*, 35, 25–36. Available from: <https://doi.org/10.1002/joc.3959>
- Spinoni, J., Vogt, J.V., Barbosa, P., Dosio, A., McCormick, N., Bigano, A. et al. (2018) Changes of heating and cooling degree-days in Europe from 1981 to 2100. *International Journal of Climatology*, 38, e191–e208. URL. Available from: <https://doi.org/10.1002/joc.5362>
- van Vliet, M.T., Yearsley, J.R., Ludwig, F., Vögele, S., Lettenmaier, D.P. & Kabat, P. (2012) Vulnerability of us and european electricity supply to climate change. *Nature Climate Change*, 2, 676–681.
- Vitart, F. (2014) Evolution of ECMWF sub-seasonal forecast skill scores. *Quarterly Journal of the Royal Meteorological Society*, 140, 1889–1899. Available from: <https://doi.org/10.1002/qj.2256>
- Vitart, F., Ardilouze, C., Bonet, A., Brookshaw, A., Chen, M., Codorean, C. et al. (2017) The subseasonal to seasonal (S2S) prediction project database. *Bulletin of the American Meteorological Society*, 98, 163–173. URL. Available from: <https://journals.ametsoc.org/view/journals/bams/98/1/bams-d-16-0017.1.xml>
- Weigel, A.P., Baggenstos, D., Liniger, M.A., Vitart, F. & Appenzeller, C. (2008) Probabilistic verification of monthly temperature forecasts. *Monthly Weather Review*, 136, 5162–5182. Available from: <https://doi.org/10.1175/2008MWR2551.1>
- White, C.J., Carlsen, H., Robertson, A.W., Klein, R.J., Lazo, J.K., Kumar, A. et al. (2017) Potential applications of subseasonal-to-seasonal (S2S) predictions. *Meteorological Applications*, 24, 315–325. Available from: <https://doi.org/10.1002/met.1654>
- White, C.J., Domeisen, D.I.V., Acharya, N., Adefisan, E.A., Anderson, M.L., Aura, S. et al. (2022) Advances in the application and utility of subseasonal-to-seasonal predictions. *Bulletin of the American Meteorological Society*, 103, E1448–E1472. URL. Available from: <https://journals.ametsoc.org/view/journals/bams/103/6/BAMS-D-20-0224.1.xml>

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Baker, L., Charlton-Perez, A., & Mattu, K. L. (2023). Skilful sub-seasonal forecasts of aggregated temperature over Europe. *Meteorological Applications*, 30(6), e2169. <https://doi.org/10.1002/met.2169>