Decadal Predictions of the Probability of Occurrence for Warm Summer Temperature Extremes

Leonard F. Borchert, Holger Pohlmann, Johanna Baehr, Nele-Charlotte Neddermann, Laura Suarez-Gutierrez, Wolfgang Müller

Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL085385

Decadal Predictions of the Probability of Occurrence for Warm Summer Temperature Extremes

Leonard F. Borchert[1,2], Holger Pohlmann[1,3], Johanna Baehr[4], Nele-Charlotte Neddermann[4,5], Laura Suarez-Gutierrez[1,5], and Wolfgang A. Müller[1]

[1]Max Planck Institute for Meteorology, Hamburg, Germany, [2]Sorbonne Universités (SU/CNRS/IRD/MNHN), LOCEAN Laboratory, Institut Pierre Simon Laplace, Paris, France, [3]Deutscher Wetterdienst, Hamburg, Germany, [4]Institute for Oceanography, CEN, Universität Hamburg, Hamburg, Germany, [5]International Max Planck Research School on Earth System Modelling, Hamburg, Germany

Abstract We use a decadal prediction system with the Coupled Model Intercomparison Project Phase 6 version of the coupled Max Planck Institute Earth System Model to predict the probability of occurrence for extremely warm summers in the Northern Hemisphere. An assimilation run with Max Planck Institute Earth System Model shows a robust response of summer temperature extremes in northern Europe and northeast Asia to North Atlantic sea surface temperature via a circumglobal Rossby wavetrain. When the North Atlantic is warm, warm summer temperature extremes occur with a probability of 20% and 24% in northern Europe and northeast Asia, respectively. In a cold North Atlantic phase, these probabilities are 0% and 8%. A similar difference in probability of occurrence is found in the initialized climate predictions. Consequently, the likelihood of a warm summer temperature extreme occurring in the examined regions in the next 10 years can be inferred from predictions of North Atlantic temperature.

Plain Language Summary Extremely warm summers can have a substantial impact on society. Trustworthy predictions of such events several years ahead could help in anticipating the extremes and managing their impacts. In this study, we show that the probability with which a warm summer temperature extreme occurs in northern Europe and northeast Asia can be linked to the surface temperature of the North Atlantic ocean. We further show that North Atlantic ocean surface temperature and the connection between ocean temperature and extreme summer temperature can be predicted. As a result, the probability for extremely warm summers to occur in northern Europe and northeast Asia in the next 10 years can be predicted.

Key Points:
- Extremely warm summers in northern Europe and northeast Asia occur more frequently when the North Atlantic is warm than when it is cold
- A set of initialized CMIP6 decadal hindcasts predicts this dependence of summer temperature extremes on North Atlantic temperature
- The likelihood of extremely warm summers to occur in the next 10 years can be inferred from predictions of North Atlantic ocean temperature

Correspondence to:
L. Borchert,
leonard.borchert@locean-ipsl.upmc.fr

Citation:
Borchert, L. F., Pohlmann, H., Baehr, J., Neddermann, N.-C., Suarez-Gutierrez, L., & Müller, W. A. (2019). Decadal predictions of the probability of occurrence for warm summer temperature extremes. *Geophysical Research Letters*, *46*. https://doi.org/10.1029/2019GL085385

Received 16 SEP 2019
Accepted 12 NOV 2019
Accepted article online 15 NOV 2019

1. Introduction

Prediction of extremely warm summers several years ahead could help in anticipating their occurrence and managing their impacts (e.g., Suarez-Gutierrez et al., 2018; Weisheimer et al., 2011). However, climate models struggle to skillfully predict warm summer temperature extremes on decadal time scales (e.g., Hanlon et al., 2013, 2014). Here, we examine the Coupled Model Intercomparison Project Phase 6 (CMIP6) version of the initialized Max Planck Institute Earth System Model (MPI-ESM-HR) and establish predictions of the probability with which warm summer temperature extremes in the Northern Hemisphere will occur.

Recently, studies have shown that several climatic parameters are predictable up to 10 years into the future (e.g., Boer et al., 2016; Marotzke et al., 2016; Yeager & Robson, 2017). High-impact events like tropical (Dunstone et al., 2011) and extratropical (Schuster et al., 2019) storms, Sahel summer rainfall (Sheen et al., 2017), or Eurasian precipitation and sea level pressure (SLP) (Smith et al., 2019) were shown to be predictable on this time scale. Seasonal predictions of temperature extremes were shown to be skillful (e.g., Eade et al., 2012), as well as decadal predictions of warm summer temperature extremes—skill in the latter, however, was attributed to the global warming trend in the analyzed data (Hanlon et al., 2013). Attempts to predict extremely warm summers on the decadal time scale beyond the externally forced trend were not successful so far.

A possible cause for the lack of skill in decadal predictions of temperature extremes beyond the global warming trend is that a robust link between the occurrence of such extremes and the ocean, particularly sea surface temperature (SST), was missing in the models: Often, high decadal climate prediction skill was shown to arise from the *memory* in the ocean via SST (e.g., Borchert et al., 2018; Matei et al., 2012; Robson et al., 2012; Yeager et al., 2012) and surface air temperature (SAT) (e.g., Müller et al., 2012; Wu et al., 2019). Studies showed that decadal predictability of northern European SAT (Årthun et al., 2017) and central European SAT variability (Duchez et al., 2016a; Wu et al., 2019) on the decadal time scale is connected to the North Atlantic ocean, indicating an important role of the ocean for decadal predictability of SAT variations. SAT variations in Europe were shown to be connected to North Atlantic SST (Gastineau & Frankignoul, 2015; Ghosh et al., 2016; Sutton & Hodson, 2005) through an atmospheric Rossby wave, the so-called *circumglobal wavetrain*. The wave originates over the North Atlantic and communicates surface temperature signals to Europe and further downstream, up to northeast Asia (NEA) (Monerie et al., 2017). The wavetrain is commonly identified in geopotential height anomalies at 500 hPa but can also be found in other parameters such as SLP. So far, no connection was found between SST and extreme summer temperature, and climate models struggle to predict the correct large-scale atmospheric conditions in the North Atlantic region (e.g., Borchert et al., 2018).

In this paper, we invoke the wavetrain mechanism to connect the probability of occurrence of warm summer temperature extremes in several regions in the Northern Hemisphere to North Atlantic SST variations. By linking the probability of occurrence for a warm summer extreme to North Atlantic SST, we aim to predict the probability of events instead of individual events. This approach eliminates the need to predict the right timing and phase of North Atlantic atmospheric conditions. Instead, the right tendency of atmospheric dynamics (Eade et al., 2012), based on ocean conditions, has to be predicted. We show in a global climate prediction system with CMIP6 forcing that North Atlantic SST is predictable on the decadal time scale. As a result, the probability for a warm summer temperature extreme to occur in the next 10 years can be predicted.

2. Data and Methods

2.1. Model

In this study we use the coupled model MPI-ESM1.2 (Mauritsen et al., 2019) to perform initialized decadal climate prediction experiments. The experiments are performed with a high-resolution version of MPI-ESM1.2 (MPI-ESM-HR; Müller et al., 2018). MPI-ESM-HR consists of the atmospheric component ECHAM6.3 in T127 (about 100 km) horizontal resolution and 95 vertical levels and the ocean component MPIOM in a nominal 0.4° horizontal resolution and 40 vertical levels. Because soil moisture is important for the prediction of extreme heat in Europe (e.g., MacLeod et al., 2016), the model used in this study features an improved five-layer soil moisture scheme (Bunzel et al., 2017).

A total of 10 ensemble members are run for the initialized experiments. The initialized runs (hereafter *hindcasts*) are started every year on 1 November for the period 1960–2018 and have a length of 10 years and 2 months each. The initial conditions are taken from an assimilation run with MPI-ESM-HR. The ensemble is generated using lagged initialization.

In the assimilation run the model state is nudged toward atmospheric (ERA-40/ERA-Interim; Dee et al., 2011) and oceanic reanalysis (ORAS4; Balmaseda et al., 2013), and National Snow and Ice Data Center (NSIDC) sea ice data (Fetterer et al., 2018). For the atmospheric component, the model state is nudged toward full values of divergence, vorticity, temperature, and SLP. For the ocean component, the model state is nudged toward anomalous three-dimensional temperature and salinity fields. This procedure is similar to a previous version of the decadal climate prediction system (Pohlmann et al., 2013). The anomalous sea ice concentration is nudged from 1979 onward (as in Bunzel et al., 2016). Prior to 1979 the climatology of sea ice concentration is used. CMIP6 forcing is used for all simulations (Eyring et al., 2016): historical forcing for the period 1960–2014 and the scenario SSP2-4.5 thereafter. A summary of the experiments is given in Pohlmann et al. (2019).

We examine climate variations in the assimilation run in this study, because internally consistent data are available for all atmospheric and oceanic variables that are needed for our analysis. That being said, the results of the analysis of SAT presented here do not substantially differ from the results of the same analysis using 10 randomly selected ensemble members of the HadCRUTv4 data set (Morice et al., 2012). We illustrate this in Figures 1a and 1b and use the assimilation run after that.

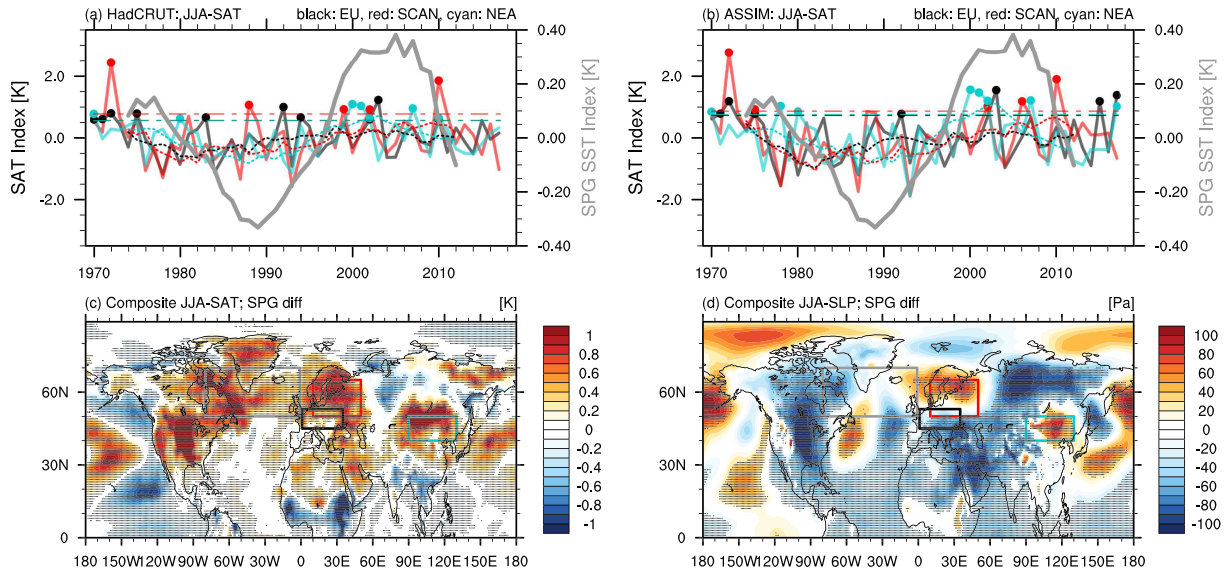

Figure 1. (a) Time series of summer (JJA) mean SAT in central Europe (EU, black), northern Europe and Scandinavia (SCAN, red), and northeast Asia (NEA, cyan) as well as the corresponding time series when a 10-year low-pass filter is applied (dotted lines) from HadCRUT4 observations, and summer mean subpolar gyre (SPG) SST (thick gray line, 10-year running mean) from the assimilation run. Dashed colored lines show the 2 standard deviation level above the 1980–2010 mean for the respective SAT time series. Whenever summer SAT exceeds this level, a dot shows the occurrence of a warm SAT extreme. (b) As in (a) but using SAT from the assimilation run. (c) Assimilation run-based SAT composite difference between warm and cold phases of SPG SST, defined as a half standard deviation below and above the 1980–2010 mean, respectively. Colored boxes show the area averages used to calculate the SPG index (gray), and the SAT indices for EU (black), SCAN (red), and NEA (cyan). Stippled areas show composite differences that are at the 1% level different from composite differences of randomly selected years, based on 5,000 bootstrap samples (Jolliffe & Stephenson, 2012). (d) As in (c), but for SLP instead of SAT.

2.2. Postprocessing

Summer means are in this study defined as June-July-August (JJA) means for SST, SAT, and SLP. We subtract the long-term linear trend from all time series analyzed. Extremely warm summer SAT is defined as two standard deviations above the mean SAT of the period 1980–2010. Because this reference includes a full cycle of multidecadal North Atlantic SST variations (Figures 1a and 1b), this method ensures that our results are not biased toward a higher or lower mean temperature.

For North Atlantic SST variations, we use average SST in the subpolar gyre (SPG), defined as a weighted area average in 80–0°W, 50–70°N (gray box in Figures 1c and 1d). We examine extremely warm summers in central Europe (EU: 0–35°E, 45–52°N; as in Duchez et al., 2016a), northern Europe and Scandinavia (SCAN: 10–50°E, 50–65°N), and NEA (90–130°E, 40–50°N, as in black, red, and cyan boxes in Figures 1c and 1d, respectively; Monerie et al., 2017). These regions were constructed to cover mostly land surface to minimize direct influence of the underlying ocean on the results. To illustrate the areas of highest influence of SPG temperature variations, we construct composite differences by subtracting SAT (Figure 1c) and SLP (Figure 1d) during cold SPG phases from warm SPG phases. Warm and cold SPG phases are defined as a half standard deviation below and above the 1980–2010 mean, respectively.

2.3. Pooling the Hindcast Ensemble

We use here an approach to hindcast ensemble analysis that pools all available realizations for any given year and lead time. This approach maximizes the sample size to 100 model realizations per year: 10 ensemble members times 10 lead years per year. This approach works for every year after 1970, because starting the first hindcasts in 1960 means that starting in 1970, 10 lead years are available. The pooling approach is illustrated in Figure 2 for the example Year 1980. Because a common number of realizations are available after Year 1970, we henceforth concentrate on the time period 1970–2018.

We define extremes for hindcasts relative to the base period 1980–2010 for the different lead year time series separately. This practice ensures that the definition of an extreme is consistent with the model development over time. We then pool the ensemble as described above and count the amount of extremes produced in the pooled ensemble for every given year. As a result, we can show the average probability with which a certain event, like a SAT extreme, was predicted up to 10 years in advance for any given year.

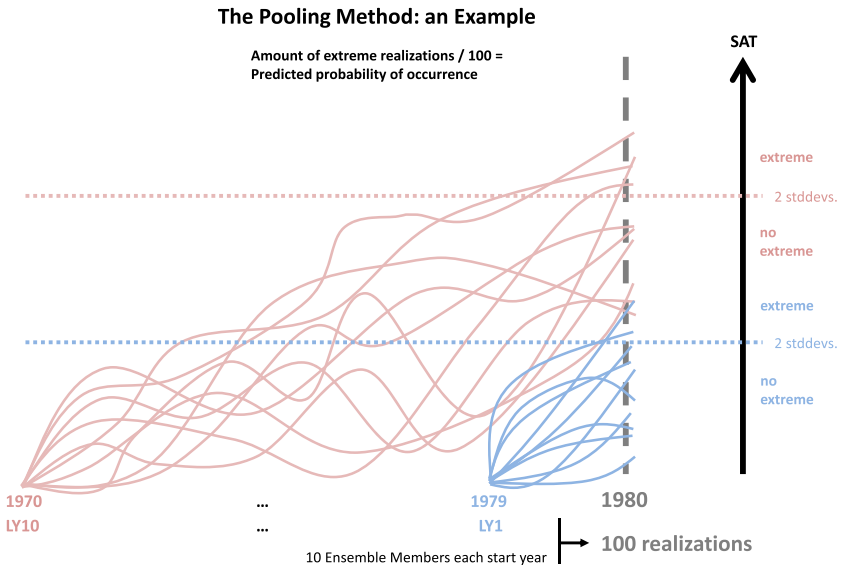

Figure 2. An illustration of how the pooling method used in this study works, exemplarily shown for the Year 1980. There are 100 realizations of the Year 1980, 10 ensemble members at 10 lead years. Whenever the predicted summer SAT in a realization lies above the lead time-dependent threshold of 2 standard deviations (*stddevs*), this realization counts toward the predicted likelihood of occurrence for a warm summer temperature extreme. All realizations are treated equally.

3. Results

3.1. North Atlantic Influence on Northern Hemisphere SAT

Both in HadCRUTv4 and the assimilation run, summer SAT in the three examined regions, EU, SCAN, and NEA, shows interannual variability alongside low-frequency fluctuations (Figures 1a and 1b). The low-frequency variations reach maxima in the 1970s and after 1995, while a minimum is found between 1980 and 1995. These maxima and minima match remarkably well with low-frequency variations in North Atlantic SPG temperature from the assimilation run: The time series of SPG temperature correlates with the 10-year low-pass filtered time series of observed SAT in the EU, SCAN, and NEA regions at values of 0.36, 0.71, and 0.88, respectively. For SAT in the assimilation run, these correlation values are 0.47, 0.7, and 0.92. All of these correlations are significant at the 5% level according to a Monte Carlo method of 1,000 bootstrap samples with replacement. In summary, SPG temperature shows maxima and minima that coincide with those in low-frequency variations of SAT in the three regions but with a stronger connection of SAT in SCAN and NEA to SPG temperature than in EU (Figures 1a and 1b). This finding indicates a potential link between SST variations in the North Atlantic SPG and SAT changes in the three regions, with warmer SAT in warm phases of the SPG and vice versa.

Extremes of summer SAT occur in all three regions more often in phases of warm SPG than in phases of cold SPG (Figures 1a and 1b). In both HadCRUTv4 and the assimilation run this effect is stronger for the SCAN and NEA regions than for the EU region; qualitatively, however, the effect can be found in all regions and in both data sets. This can be attributed mainly to a warmer mean state in these regions when the SPG is warm than when the SPG is cold, because the amplitude of year-to-year SAT variability does not change much between warm and cold SPG phases (not shown). It is striking that warm SAT extremes in the SCAN and NEA regions occur almost exclusively when the SPG is warm, indicating a strong link between SPG temperature and the occurrence of extremes in these regions. In the EU region, this link is not as strong as in the other two, as some extremes do occur in cold phases of the SPG. However, all examined regions show a strong connection of both temperature fluctuations and the occurrence of warm summer temperature extremes to the phase of North Atlantic SST. It is interesting that HadCRUTv4 and the assimilation run agree well on these findings, because as the assimilation run is based on the ERA reanalyses, it can be regarded as largely independent of HadCRUTv4 (Simmons et al., 2016). Because the model simulation offers a more complete picture of the climate system, we use the assimilation run in all following analyses.

Spatially, SST changes in the SPG region are linked to fluctuations of summer air temperature over land in several regions in the Northern Hemisphere (Figure 1c). Regions of strong connection to SPG temperature include northern Europe and SCAN, NEA, the Arabian Peninsula, and the central United States. EU temperature shows a weak link to SPG SST, as do several regions in western and northern Russia and Alaska. SAT changes in the SCAN, NEA, and the central United States regions are particularly strongly connected to temperature fluctuations in the SPG (Figure 1c).

The circumglobal wavetrain connects temperature variations in the North Atlantic to the SAT anomalies in the SCAN and NEA regions, as revealed by SLP composite differences between warm and cold SPG phases (Figure 1d). An influence of the wavetrain can be assumed when a positive SAT anomaly coincides with a positive SLP anomaly and vice versa. This is not the case for the central United States. This is an indication that SST fluctuations in the SPG region play a role in SAT variations in SCAN and NEA via the circumglobal wavetrain, whereas the connection of central United States SAT to SPG temperature could be connected to other mechanisms like heat advection. Because the robust connection of SAT variations in SCAN and NEA to the North Atlantic (Årthun et al., 2017; Monerie et al., 2017) is generally in line with published literature, we will henceforth focus our analysis on these two regions alongside EU for comparison.

3.2. Predicting the Probability of Occurrence of Summer Temperature Extremes

Results presented above indicate that skillful predictions of the probability of occurrence for warm summer temperature extremes over land could be possible. In this section, we examine this hypothesis by studying the reproduction of the connection between SST in the SPG region and the occurrence of warm summer temperature extremes in initialized climate model simulations.

Accurate decadal predictions of SPG temperature are necessary for predictions of the probability of occurrence of warm summer temperature extremes in the examined regions. Hindcast skill for SPG SST, evaluated as anomaly correlation coefficients against the assimilation run, is significant at the 5% level throughout all lead years, according to a Monte Carlo test with 1,000 bootstraps (Figures 3a and 3b). The fact that the dynamical hindcasts outperform hindcasts based on persistence (Figure 3b) after lead Year 3 indicates that dynamical predictions are necessary for skillful predictions beyond this time horizon. Moreover, SPG states predicted to occur alongside SAT extremes in the EU, SCAN, and NEA regions correspond well to SPG temperature in the assimilation run (Figures 3c and 3d). As the predicted SPG states do not differ substantially between early and late lead times (Figure 3d), the pooling approach we use in this study is not strongly affected by a change in the model's capability of reproducing SPG SST variations over lead time.

The predicted probability of occurrence for warm summer temperature extremes in the different regions varies over time, as revealed by the pooled hindcasts (Figure 4a). Summer temperature extremes are for all regions predicted to occur with higher-than-average probability in the 1970s, lower-than-average probability in the 1980s and 1990s, and higher-than-average probability in the 2000s and early 2010s. This fits North Atlantic SST variations in the assimilation run (Figure 4a), with extremes being predicted to occur with a higher probability when the SPG is warm.

By assuming ergodicity we can compare the predicted probability to the observed probability of occurrence of warm summer temperature extremes. The probability of occurrence for observed summer temperature extremes can thus be estimated by dividing the amount of occurring extremes in either phase of the SPG by the total amount of years in the corresponding SPG phase. Warm summer temperature extremes occur with probabilities of 16%, 20%, and 24% during warm SPG phases in the EU, SCAN, and NEA regions, respectively. During cold SPG phases, these probabilities are 13%, 0%, and 9%. The model predicts extremes with probabilities of $21 \pm 1.5\%/13 \pm 1.5\%$, $21 \pm 1.4\%/14 \pm 1.45\%$, and $22 \pm 1.6\%/14 \pm 1.7\%$ during warm/cold SPG phases in the respective regions (Figure 4b).

The difference between the predicted probabilities of occurrence for warm and cold SPG states is significant in all three regions, predicting a higher probability for a heat extreme to occur in warm than in cold SPG phases. This relationship between SPG temperature and the probability of occurrence of warm summer temperature extremes is observed in SCAN and NEA as well, albeit much stronger than predicted (Figure 4b). While EU shows the same tendency as SCAN and NEA, the probability of an extremely warm summer to occur is less strongly connected to SPG temperature. This indicates that the model underestimates the connection of summer temperature extremes to SPG temperature in SCAN and NEA and overestimates this connection in EU. Because the observed difference in probability of occurrence between warm and cold

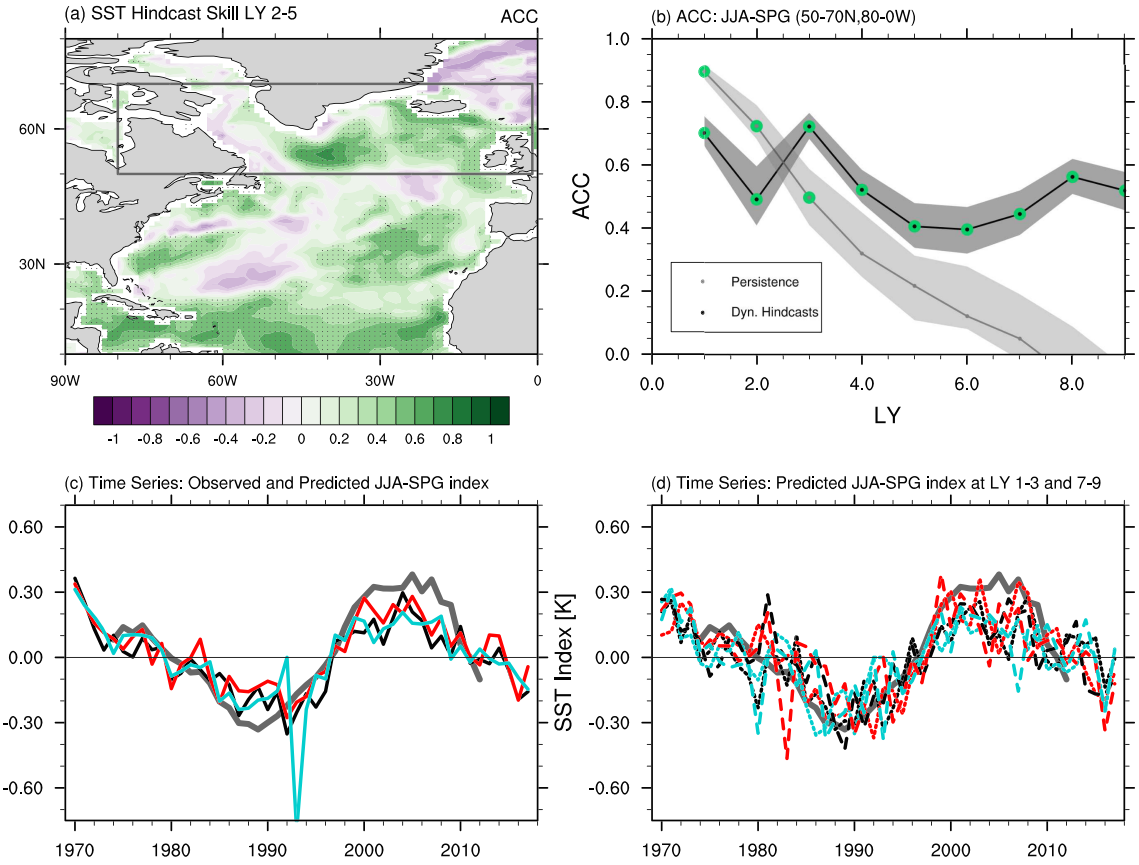

Figure 3. Summer (JJA) SST predictability and variability over time. (a) Ensemble mean hindcast skill in the North Atlantic evaluated as anomaly correlation coefficient (ACC) against HadISST at lead Years 2–5. Stippling indicates significant skill at the 5% level, based on 1,000 bootstrap samples with replacement. The gray box outlines the SPG region. (b) SPG hindcast skill for the ensemble mean hindcast model simulations (black) and persistence (gray) over lead time. Persistence skill is calculated from a lagged correlation of the observed time series with itself. Skill estimates marked in green are statistically significant at the 5% level, based on 1,000 bootstraps. Dark and light gray polygons mark skill between the 25th and 75th percentiles of 1,000 bootstraps of the dynamical and persistence hindcast time series around the corresponding mean, respectively. (c) The gray line shows the 10-year running mean SPG index from the assimilation run. Solid colored lines show for any given year the average SPG index in the pooled hindcast realizations that show a summer SAT extreme in the regions EU (black), SCAN (red), and NEA (cyan). (d) Colored lines show the same as solid ones in (c), but for lead Years 1–3 (dotted) and 7–9 (dashed).

phases in the North Atlantic is not significant in the EU region, this overestimation indicates that the model predicts a link that might not be there in reality. At the same time, decadal predictions of the probability of occurrence of warm summer temperature extremes in the other examined regions are linked to SPG temperature. This indicates a possibility for accurate predictions of the probability of occurrence of summer SAT extremes in hindcast simulations with the CMIP6 version of MPI-ESM-HR.

4. Discussion

We show in this study that the probability of occurrence for summer temperature extremes in several regions in the Northern Hemisphere is connected to SST in the North Atlantic SPG. This relation can also be found in initialized decadal hindcasts with the CMIP6 version of the MPI-ESM-HR. As a result, the likelihood of a warm summer temperature extreme occurring in one of the selected regions in the next 10 years can be inferred from predictions of North Atlantic SPG temperature.

However, the results presented here are somewhat limited by our definition of a summer extreme based on summer means. Extreme temperature often occurs on daily or even shorter time scales, which is why studies often use indices like daily maximum temperature to examine extreme temperature (e.g., Suarez-Gutierrez et al., 2018). However, ocean influence on the atmosphere is more long term than daily (e.g., Duchez et al., 2016b; Gulev et al., 2013), which justifies the use of summer means in this study. We do encourage future elaboration based on our findings with more detailed metrics like daily maximum temperature.

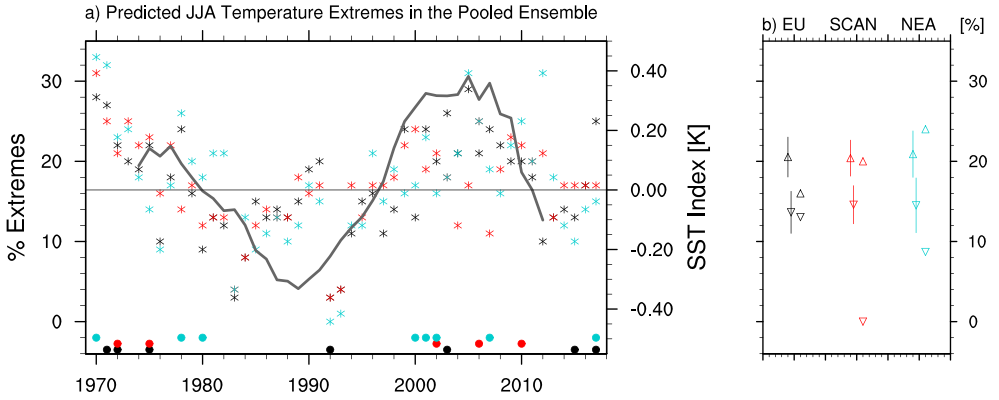

Figure 4. (a) Probability of occurrence for summer SAT extremes in the regions EU (black), SCAN (red), and NEA (cyan). Dots at the bottom show the occurrence of a summer SAT extreme in the respective region in the assimilation run (cf. Figure 1b). Asterisks indicate the percentage of pooled hindcast realizations that produce a warm summer SAT extreme in any given year in the respective region. The gray line shows the 10-year running mean SPG SST index from the assimilation run. (b) Triangles indicate average predicted (with uncertainty bar) and assimilation run-based (without uncertainty bar) probabilities of occurrence for summer SAT extremes in the respective regions. Probabilities of occurrence are shown for warm (upward pointing triangles) and cold (downward pointing triangles) SPG phases. Uncertainty bars show one standard deviation around the mean.

Our analysis of the assimilation run is also limited by the relatively small sample size, as we base our analysis of extreme events on 57 summers. The small sample is one of the reasons why we chose a comparatively low threshold to define a SAT extreme. Since reliable observations of the atmosphere only go back to 1960, this is the time horizon that is available at the moment. This study should, however, be repeated when longer observational data sets become available to assess the robustness of our findings.

The linear detrending we apply to account for the global warming trend might introduce some error, because the global warming trend is likely not linear (e.g., Maher et al., 2019). For example, some of the SAT extremes identified in the early 2000s in the assimilation run might actually originate from the global warming trend and be artifacts of the linear detrending. While there are more elegant approaches to subtract the warming signal from a climate data set, like using the ensemble mean of a very large ensemble (e.g., Maher et al., 2019), these techniques cannot be applied in this case because only one realization of the assimilation run is available. If the post-2010 extremes are in fact artifacts of the linear detrending, a cleaner subtraction of the warming trend might lead to a stronger connection of the probability of occurrence of summer temperature extremes to SPG temperature. The phenomenon we describe here might therefore be more robust than what we show in this paper.

On the other hand, a stronger separation of the probability of occurrence of extremely warm summers between warm and cold North Atlantic phases in the assimilation run than in HadCRUTv4 in Figures 1a and 1b suggests that the model overestimates the impact of SPG temperature on the probability of occurrence of a warm summer temperature extreme in the examined regions. The same is suggested by Figure 4b. This is a further limitation that needs to be considered when interpreting the results presented here. As the different probabilities of occurrence are qualitatively similar in HadCRUTv4 and the assimilation run, our results are robust in observations as well as the model.

We base the selection of studied regions on a physical mechanism, the circumglobal wavetrain, and only examine regions in which surface temperature shows a positive anomaly coinciding with a high pressure anomaly with respect to SPG temperature (cf. Figures 1c and 1d). An exception to this is the EU region, that only partly shows this physical connection. As a result, EU temperature extremes show a weaker dependence on SPG temperature than SCAN and NEA (cf. Figure 4). Here we examine EU as well because it is a high-impact region. Remarkably, predictions of the probability of occurrence of warm summer temperature extremes in the EU region show a similarly large dependence on North Atlantic temperature as in the other regions. This might indicate a limited capability of the examined model to predict the probability of occurrence of extremely warm summers in the EU region, which would agree with the lower correlation of EU SAT to North Atlantic SST.

Most other regions do not show the observed connection of probability of occurrence for warm summer temperature extremes to SPG temperature. This indicates that the circumglobal wavetrain actually links SPG temperature to warm summer temperature extremes. As a consequence, predicting the wavetrain is key to predicting the actual occurrence of a warm summer temperature extreme beyond its probability of occurrence. Representation of the wavetrain in climate models should thus be subject to future research to improve decadal predictions of extreme summer SAT on the Northern Hemisphere. It is noteworthy that the phenomenon we describe in this paper works for cold summer temperature extremes as well: Extremely cold summers occur and are predicted to occur more often when the SPG is cold than when SPG is warm (not shown).

Our pooling approach combines all lead times into one set of simulations and treats all realizations equally. This can only work if the connection of SPG temperature to the probability of occurrence for summer temperature extremes does not change with lead time and if predictions of SPG SST show skill for all lead times. In other words, the mechanisms examined in this paper need to be relatively insensitive to hindcast initialization so that the pooling approach can be used. In the atmosphere, it is reasonable to assume that the initialization method does not sufficiently disrupt model variability to leave a long-lasting imprint because of the quick turnover time of the atmosphere (e.g., Boer et al., 2016). On the other hand, it was demonstrated that systematic biases in the ocean, resulting from improper hindcast initialization, can lead to long-standing impact of the initialization on North Atlantic SST predictability (Kröger et al., 2017). Predictions of SPG variability for lead Years 1–3 and 7–9 shown in Figure 3, however, indicate that the capability of the model to reproduce SPG variability on the decadal time scale does not strongly depend on lead time. Moreover, predictions of SPG temperature show significant skill between lead Years 1 and 10 (cf. Figures 3a and 3b), which is in line with published literature (e.g., Borchert et al., 2019), and outperform persistence hindcast after lead Year 3 (cf. Figure 3b). This suggests that there is no fundamental disturbance of the ocean by hindcast initialization and that dynamical predictions are needed to make this prediction at lead times beyond 3 years. Therefore, the pooling approach we introduce in this study can be used to study not only the subject matter of this paper but potentially also other processes linked to North Atlantic SST.

5. Conclusion

We present in this study a new approach to predict warm summer temperature extremes on the decadal time scale. By not focusing on individual extremes but predicting a statistical value—the probability of occurrence of a summer temperature extreme—we harvest the full current potential of a state-of-the-art initialized global climate model. We identify a link of the probability with which a warm summer temperature extreme occurs in different regions in the Northern Hemisphere to SST variations in the North Atlantic in the HadCRUTv4 data set: Extremes occur more often when the North Atlantic SPG is warm than when it is cold. We derive a physical mechanism that explains this link in a MPI-ESM-HR-based assimilation model run. Initialized decadal climate predictions with the CMIP6 version of MPI-ESM-HR produce an extremely warm summer in the examined regions in the Northern Hemisphere with a higher probability when the SPG is warm than when it is cold and skillfully predict SPG temperature. As a result, there is potential for predicting the probability of warm summer temperature extremes to occur in the next 10 years in the examined prediction system.

This study is the first to demonstrate skillfull predictions of the probability of occurrence of warm summer temperature extremes. Our study indicates potential for predicting socially relevant parameters in the upcoming CMIP6 decadal climate prediction models and encourages further investigation into the development of decadal climate prediction models.

Acknowledgments
This research was supported by the German Ministry of Education and Research under MiKlip FlexForDec (Grant 01LP1519A; L. F. B., H. P., and W. A. M.) and the Second Regional Atlantic Circulation and Global Change Project (RACE II; N. C. N. and J. B.); by the IAFE project DAKLIM by the Deutscher Wetterdienst (H. P.); by the International Max Planck Research School on Earth System Modelling, Hamburg (L. S. G.); by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy—EXC 2037 "Climate, Climatic Change, and Society"—Project 390683824, contribution to the Center for Earth System Research and Sustainability (CEN) of Universität Hamburg (J. B.); and by the European Union under Horizon 2020 Project EUCP (Grant Agreement 776613; L. F. B.). The model simulations were performed at the German Climate Computing Centre. Data and scripts are available from the Climate and Environmental Retrieval and Archive (CERA) of the World Data Center for Climate as part of the ESGF project (https://esgf-data.dkrz.de/projects/esgf-dkrz/).

The authors thank two anonymous reviewers as well as Sebastian Milinski, Tim Kruschke, Elina Plesca, Sebastian Brune, and André Düsterhus for helpful discussions and comments on this manuscript, and Kameswarrao Modali and Klaus Pankatz for performing the model simulations.

References

Årthun, M., Eldevik, T., Viste, E., Drange, H., Furevik, T., Johnson, H. L., & Keenlyside, N. S. (2017). Skillful prediction of northern climate provided by the ocean. *Nature Communications*, *8*, 15875. https://doi.org/10.1038/ncomms15875

Balmaseda, M. A., Mogensen, K., & Weaver, A. T. (2013). Evaluation of the ECMWF ocean reanalysis system ORAS4. *Quarterly Journal of the Royal Meteorological Society*, *139*, 1132–1161. https://doi.org/10.1002/qj.2063

Boer, G. J., Smith, D. M., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, *9*, 3751–3777. https://doi.org/10.5194/gmd-9-3751-2016

Borchert, L. F., Düsterhus, A., Brune, S., Müller, W. A., & Baehr, J. (2019). Forecast-Oriented Assessment of Decadal Hindcast Skill for North Atlantic SST. *Geophysical Research Letters*, *46*, 11,444–11,454. https://doi.org/10.1029/2019GL084758

Borchert, L. F., Müller, W. A., & Baehr, J. (2018). Atlantic Ocean heat transport influences interannual-to-decadal surface temperature predictability in the North Atlantic region. *Journal of Climate*, *31*, 6763–6782. https://doi.org/10.1175/JCLI-D-17-0734.1

Bunzel, F., Müller, W. A., Stacke, T., Hagemann, S., Dobrynin, M., Baehr, J., et al. (2017). Improved seasonal prediction of European summer temperatures with new 5-layer soil-hydrology scheme. *Geophysical Research Letters*, *45*, 346–353. https://doi.org/10.1002/2017GL076204

Bunzel, F., Notz, D., Baehr, J., Müller, W. A., & Fröhlich, K. (2016). Seasonal climate forecasts significantly affected by observational uncertainty of Arctic sea ice concentration. *Geophysical Research Letters*, *43*, 852–859. https://doi.org/10.1002/2015GL066928

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA–Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597. https://doi.org/10.1002/qj.828

Duchez, A., Courtois, P., Harris, E., Josey, S. A., Kanzow, T., Marsh, R., et al. (2016b). Potential for seasonal prediction of Atlantic sea surface temperatures using the RAPID array at 26°N. *Climate Dynamics*, *46*, 3351–3370. https://doi.org/10.1007/s00382-015-2918-1

Duchez, A., Frajka-Williams, E., Josey, S. A., Evans, D. G., Grist, J. P., Marsh, R., et al. (2016a). Drivers of exceptionally cold North Atlantic Ocean temperatures and their link to the 2015 European heat wave. *Environmental Research Letters*, *11*, 074004. https://doi.org/10.1088/1748-9326/11/7/074004

Dunstone, N. J., Smith, D. M., & Eade, R. (2011). Multi-year predictability of the tropical Atlantic atmosphere driven by the high latitude North Atlantic Ocean. *Geophysical Research Letters*, *38*, L14701. https://doi.org/10.1029/2011GL047949

Eade, R., Hamilton, E., Smith, D. M., Graham, R. J., & Scaife, A. A. (2012). Forecasting the number of extreme daily events out to a decade ahead. *Journal of Geophysical Research*, *117*, D21110. https://doi.org/10.1029/2012JD018015

Eyring, V., Bony, S, Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*, 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fetterer, F., Knowles, K., Meier, W., Savoie, M., & Windnagel, A. K. (2018). Sea ice index version 3. NSIDC, accessed 2018.

Gastineau, G., & Frankignoul, C. (2015). Influence of the North Atlantic SST variability on the atmospheric circulation during the twentieth century. *Journal of Climate*, *28*, 1396–1416. https://doi.org/10.1175/JCLI-D-14-00424.1

Ghosh, R., Müller, W. A., Bader, J., & Baehr, J. (2016). Impact of observed North Atlantic multidecadal variations to European summer climate: A linear baroclinic response to surface heating. *Climate Dynamics*, *48*, 3547–3563. https://doi.org/10.1007/s00382-016-3283-4

Gulev, S. K., Latif, M., Keenlyside, N., Park, W., & Koltermann, K. P. (2013). North Atlantic Ocean control on surface heat flux on multidecadal timescales. *Nature*, *499*, 464–467. https://doi.org/10.1038/nature12268

Hanlon, H. M., Hegerl, G. C., Tett, S. F. B., & Smith, D. (2013). Can a decadal forecasting system predict temperature extreme indices? *Journal of Climate*, *26*, 3728–3744. https://doi.org/10.1175/JCLI-D-12-00512.1

Hanlon, H. M., Hegerl, G. C., Tett, S. F. B., & Smith, D. (2014). Near-term prediction of impact-relevant extreme temperature indices. *Climate Change*, *132*(1), 61–76. https://doi.org/10.1007/s10584-014-1191-3

Jolliffe, I. T., & Stephenson, D. B. (2012). *Forecast verification: A practitioner's guide in atmospheric science* (2nd ed.). Oxford: Wiley and Sons, Ltd.

Kröger, J., Pohlmann, H., Sienz, F., Marotzke, J., Baehr, J., Köhl, A., et al. (2017). Full-field initialized decadal predictions with the MPI Earth System Model: An initial shock in the North Atlantic. *Climate Dynamics*, *51*, 2593–2608. https://doi.org/10.1007/s00382-017-4030-1

MacLeod, D. A., Cloke, H. L., Pappenberger, F., & Weisheimer, A. (2016). Improved seasonal prediction of the hot summer of 2003 over Europe through better representation of uncertainty in the land surface. *Quarterly Journal of the Royal Meteorological Society*, *142*(694), 79–90. https://doi.org/10.1002/qj.2631

Maher, N., Milinski, S., Suarez-Gutierrez, L., Botzet, M., Dobrynin, M., Kornblueh, L., et al. (2019). The Max Planck Institute Grand Ensemble: Enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems*, *11*, 2050–2069. https://doi.org/10.1029/2019MS001639

Marotzke, J., Müller, W. A., Vamborg, F. S., Becker, P., Cubasch, U., Feldmann, H., et al. (2016). MiKlip: A national research project on decadal climate prediction. *Bachelor of Ayurveda, Medicine, and Surgery*, *97*, 2379–2394. https://doi.org/10.1175/BAMS-D-15-00184.1

Matei, D., Pohlmann, H., Jungclaus, J., Müller, W. A., Haak, H., & Marotzke, J. (2012). Two tales of initializing decadal climate prediction experiments with the ECHAM5/MPI-OM model. *Journal of Climate*, *25*, 8502–8523. https://doi.org/10.1175/JCLI-D-11-00633.1

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*, 998–1038. https://doi.org/10.1029/2018MS001400

Monerie, P.-A., Robson, J., Dong, B., & Dunstone, N. (2017). A role of the Atlantic Ocean in predicting summer surface air temperature over north East Asia? *Climate Dynamics*, *51*, 1–19. https://doi.org/10.1007/s00382-017-3935-z

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 dataset. *Journal of Geophysical Research*, *117*, D08101. https://doi.org/10.1029/2011JD017187

Müller, W. A., Baehr, J., Haak, H., Jungclaus, J. H., Kröger, J., Matei, D., et al. (2012). Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology. *Geophysical Research Letters*, *39*, L22707. https://doi.org/10.1029/2012GL053326

Müller, W. A., Jungclaus, J. H., Mauritsen, T., Baehr, J., Bittner, M., Budich, R., et al. (2018). A high-resolution version of the Max Planck Institute Earth System Model (MPI-ESM1.2-HR). *Journal of Advances in Modeling Earth Systems*, *10*, 1383–1413. https://doi.org/10.1029/2017MS001217

Pohlmann, H., Müller, W. A., Bittner, M., Hettrich, S., Modali, K., Pankatz, K., & Marotzke, J. (2019). Realistic quasi-biennial oscillation variability in historical and decadal hindcast simulations using CMIP6 forcing. *Geophysical Research Letters*, *46*. https://doi.org/10.1029/2019GL084878

Pohlmann, H., Müller, W. A., Kulkarni, K., Modali, K., Matei, D., Vamborg, F. S. E., et al. (2013). Improved forecast skill in the tropics in the new MiKlip decadal climate predictions. *Geophysical Research Letters*, *40*, 5798–5802. https://doi.org/10.1002/2013GL058051

Robson, J. I., Sutton, R. T., & Smith, D. M. (2012). Initialized decadal predictions of the rapid warming of the North Atlantic Ocean in the mid-1990s. *Geophysical Research Letters*, *39*, L19713. https://doi.org/10.1029/2012GL053370

Schuster, M., Grieger, J., Richling, A., Schartner, T., Illing, S., Kadow, C., et al. (2019). Improvement in the decadal prediction skill of the Northern Hemisphere extra-tropical winter circulation through increased model resolution. *Earth System Dynamics Discuss*, *10*, 171–187. https://doi.org/10.5194/esd-2019-18

Sheen, K. L., Smith, D. M., Dunstone, N. J., Eade, R., Rowell, D. P., & Vellinga, M. (2017). Skilful prediction of Sahel summer rainfall on inter-annual and multi-year timescale. *Nature Communications*, *8*, 14966. https://doi.org/10.1038/ncomms14966

Simmons, A. J., Berrisford, P., Dee, D. P., Hersbach, H., Hirahara, S., & Thépaut, J.-N. (2016). A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *Quarterly Journal of the Royal Meteorological Society*, *143*, 101–119. https://doi.org/10.1002/qj.2949

Smith, D. M., Eade, R., Scaife, A. A., Caron, L.-P., Danabasoglu, G., DelSole, T. M., et al. (2019). Robust skill of decadal climate predictions. *Npj Climate and Atmospheric Science*, *1*, 13. https://doi.org/10.1038/s41612-019-0071-y

Suarez-Gutierrez, L., Li, C., Müller, W. A., & Marotzke, J. (2018). Internal variability in European summer temperatures at 1.5°C and 2°C of global warming. *Environmental Research Letters*, *13*(064026). https://doi.org/10.1088/1748-9326/aaba58

Sutton, R. T., & Hodson, D. L. (2005). Atlantic Ocean forcing of North American and European summer climate. *Science*, *309*, 115–118. https://doi.org/10.1126/science.1109496

Weisheimer, A., Doblas-Reyes, F. J., Jung, T., & Palmer, T. N. (2011). On the predictability of the extreme summer 2003 over Europe. *Geophysical Research Letters*, *38*, L05704. https://doi.org/10.1029/2010GL046455

Wu, B., Zhou, T., Li, C., Müller, W. A., & Lin, J. (2019). Improved decadal prediction of Northern-Hemisphere summer land temperature. *Climate Dynamics*, *53*, 1–13. https://doi.org/10.1007/s00382-019-04658-8

Yeager, S., Karspeck, A., Danabasoglu, G., Tribbia, J., & Teng, H. (2012). A decadal predictions case study: Late twentieth–century North Atlantic Ocean heat content. *Journal of Climate*, *25*, 5173–5189. https://doi.org/10.1175/JCLI-D-11-00595.1

Yeager, S. G., & Robson, J. I. (2017). Recent progress in understanding and predicting Atlantic decadal climate variability. *Current Climate Change Reports*, *3*, 112–127. https://doi.org/10.1007/s40641-017-0064-z