Calibrating Climate Change Time-Slice Projections with Estimates of Seasonal Forecast Reliability

M. Matsueda

Center for Computational Sciences, University of Tsukuba, Tsukuba, Japan and Department of Physics,
Atmospheric, Oceanic and Planetary Physics, University of Oxford, Oxford, United Kingdom

A. Weisheimer

National Centre for Atmospheric Science, Department of Physics, University of Oxford, Oxford, and European
Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

T. N. Palmer

National Centre for Atmospheric Science, Department of Physics, University of Oxford, Oxford,
United Kingdom

ABSTRACT

In earlier work, it was proposed that the reliability of climate change projections, particularly of regional rainfall, could be improved if such projections were calibrated using quantitative measures of reliability obtained by running the same model in seasonal forecast mode. This proposal is tested for fast atmospheric processes (such as clouds and convection) by considering output from versions of the same atmospheric general circulation model run at two different resolutions and forced with prescribed sea surface temperatures and sea ice. Here output from the high-resolution version of the model is treated as a proxy for truth. The reason for using this approach is simply that the twenty-first-century climate change signal is not yet known and, hence, no climate change projections can be verified using observations. Quantitative assessments of reliability of the low-resolution model, run in seasonal hindcast mode, are used to calibrate climate change time-slice projections made with the same low-resolution model. Results show that the calibrated climate change probabilities are closer to the proxy truth than the uncalibrated probabilities. Given that seasonal forecasts are performed operationally already at several centers around the world, in a seamless forecast system they provide a resource that can be used without cost to help calibrate climate change projections and make them more reliable for users.

1. Introduction

Providing society with reliable regional predictions of climate change is becoming more and more pressing, not least so that individuals, businesses, and national infrastructure can become well adapted to anticipated changes in climate. It is now widely recognized that such predictions must be framed in probabilistic language (e.g., Jenkins et al. 2010), reflecting inherent uncertainties arising from natural variability in climate, the numerical equations underlying climate models (including representations of physical processes), and future human-induced emissions of greenhouse gases.

Although making decisions under conditions of uncertainty can often be complex, there is little doubt that better decisions can be made with a knowledge of uncertainty. Indeed, by comparing the risk of some climatic event with the cost of taking precautionary action, one can demonstrate quantitatively the value of probabilistic forecasts for decision-making (Palmer 2002). However, realizing this value assumes that the forecast probabilities are reliable (Weisheimer and Palmer 2014).

Probability forecasts of precipitation are particularly problematic because precipitation is strongly linked with circulation, and climate models generically show

Corresponding author address: Mio Matsueda, Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, 305-8577, Japan.
E-mail: mio@ccs.tsukuba.ac.jp

substantial biases in circulation fields (IPCC 2013). This raises a difficult question: Given such biases and the fact that we cannot verify climate change predictions directly, how reliable are probabilistic predictions of precipitation climate change? In particular, are they sufficiently reliable to be used to inform decision makers about potential investments (e.g., in new adaptation infrastructure)?

In an attempt to address this problem, Palmer et al. (2008) suggested that the probabilistic reliability of seasonal forecasts, which can be tested or verified against observations, was a necessary (but not sufficient) condition for ensuring the reliability of longer-term climate change forecasts. More specifically, it was proposed that if seasonal forecasts using a particular model were not statistically reliable, then a quantitative measure of this unreliability should be used to calibrate climate change projections with the same model. Importantly, such seasonal forecasts would be available at no extra computational cost, in an operational center running seasonal and climate change forecasts (seamlessly).

This proposal is partially tested in the following sense. In making projections of climate change, there are many aspects of a climate model's representations of physical processes that are uncertain. Here we focus specifically on relatively fast time-scale processes in the atmosphere that are active on both seasonal and longer time scales. These would include, for example, convection and clouds. The representation of such moist processes is known to be crucial in a model's response to greenhouse gas forcing. To focus on such fast time-scale processes, all integrations in this paper are made using atmospheric climate models with prescribed sea surface temperatures (SST) and sea ice. First, climate simulations for the twentieth-century (20C) climate and climate change projections for the end of the twenty-first century (21C) were made using two versions of the same atmosphere-only model run at high and low resolution, both with the same specified SST and sea ice. Second, seasonal retrospective forecast (hindcast) integrations of the 20C climate were made using the low-resolution model with the same underlying SST and sea ice fields. Following Matsueda and Palmer (2011), this study treats output from the high-resolution model as "truth." Because we do not yet know the 21C climate, we cannot verify climate change projections for the 21C against observations. However, we can verify them against our hypothetical truth—a plausible estimate of reality. It is important to note that it is not necessary, for the validity of our hypothesis, that the bias with respect to real-world observations of the truth AMIP run be notably less than an equivalent AMIP run of the low-resolution model. The key point is rather that the truth simulation should be different from the "model" simulation. Of course, if the

bias of the truth runs was substantially worse than the bias of the model runs, then the relevance of our results to the real world could be called into question.

As such, the difference in the climate change signal between the low- and high-resolution model is a measure of the "error" of the low-resolution model in simulating the "true" climate change. To be consistent, we also verify the seasonal integrations of the low-resolution model against the high-resolution model (truth). This verification is used to estimate the reliability of seasonal hindcasts made with the low-resolution model, and these reliability estimates are further used to calibrate the low-resolution climate change projections.

The goal of this paper is, by comparison with the high-resolution simulations, to assess whether calibrated estimates of regional climate change from the low-resolution model have indeed smaller errors than the raw uncalibrated estimates if compared with our hypothetical truth.

In section 2 we discuss further the atmospheric models and the experimental design, and show the difference between the regional climate change signals at low and high resolution. In section 3 we describe the seasonal hindcast procedure and discuss the reliability of these seasonal forecasts. In section 4 we explain the calibration procedure and show results of a comparison of calibrated and uncalibrated climate change probabilities.

2. Experimental design

In this study, 20C (1979–2003) simulations and 21C (2075–99) time-slice projections were performed with the Japanese Meteorological Research Institute's Atmospheric General Circulation Model (MRI-AGCM3.2; Mizuta et al. 2012) at two different resolutions—a resolution typical of contemporary climate models (TL95L64, 180 km, referred to as low resolution) and a resolution typical of contemporary numerical weather prediction models (TL959L64, 20 km, referred to as high-resolution). Here "TLx" refers to truncation at total wavenumber x using a triangular spectral truncation based on a Gaussian grid, and "Ly" refers to a vertical truncation with v vertical levels. Four-member initial-value ensemble simulations for each century were performed for the low-resolution model, whereas only one single simulation for each century was conducted for the high-resolution model due to limited computing resources.

In the 20C simulations, observed HadISST and sea ice concentrations (SICs) (Rayner et al. 2003) were used as lower boundary conditions. In the 21C simulations, the SST and SICs climate change signals were estimated from phase 3 of the Coupled Model Intercomparison Project (CMIP3) (Meehl et al. 2007), multimodel ensemble mean to which the detrended interannual

MRI—AGCM3.2 change in frequency of dry JJA (2075—2099)
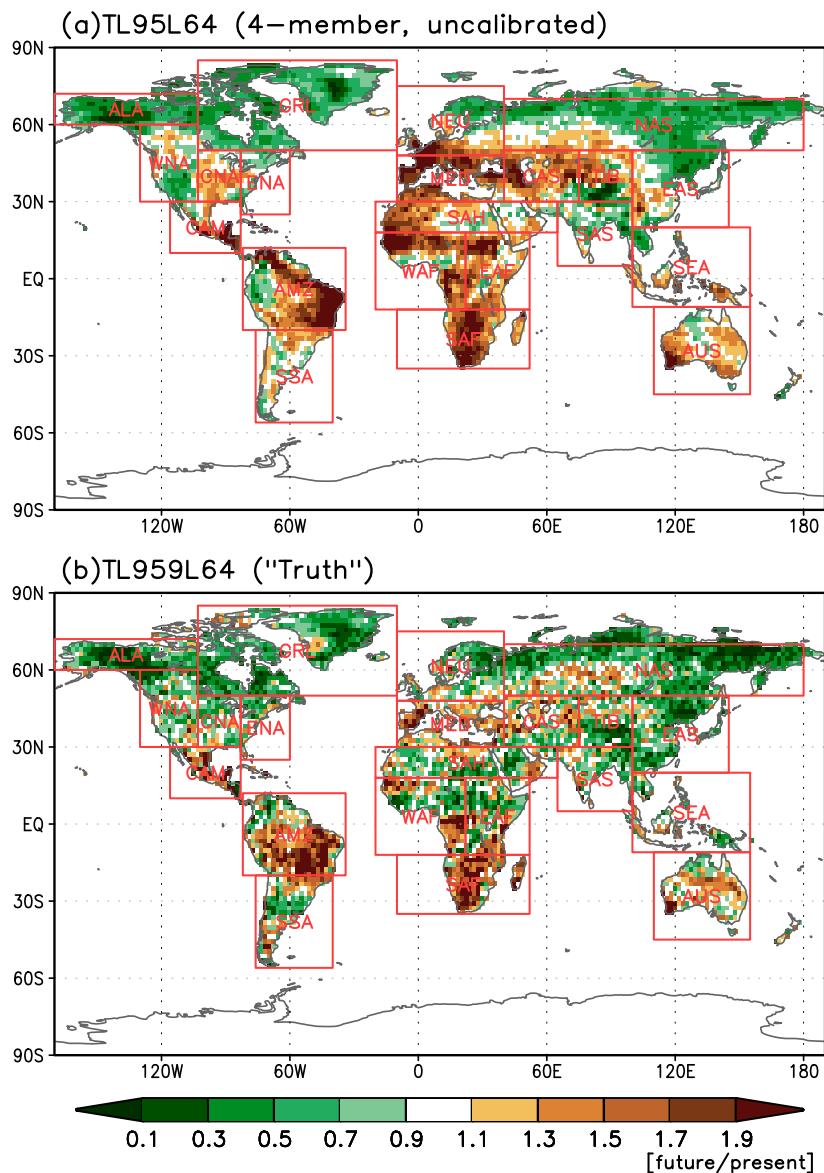
FIG. 1. Changes in the frequency of dry JJA for the period 2075–99, relative to the period 1979–2003, derived from MRI-AGCM3.2 with two different resolutions: (a) low-resolution TL95L64 and (b) high-resolution TL959L64 under the IPCC-A1B scenario.

variations in HadISST were added (Mizuta et al. 2008). The IPCC SRES A1B scenario was assumed for future emissions of greenhouse gasses.

Figure 1 illustrates the ratio of frequencies of dry June–August (JJA) at the end of the 21C relative to their reference frequencies in the 20C, for 21 standard land regions (the Giorgi regions; Giorgi and Francisco 2000). We refer to seasons in which precipitation falls below (above) the lower (upper) tercile of the corresponding 20C reference distribution as dry (wet). The tercile thresholds were estimated from the high- and

low-resolution model output for the 20C reference period and these were then used to calculate frequencies of exceeding the thresholds in 20C and 21C. By definition, the frequency of exceeding the threshold during the reference period is 1/3.

In some regions e.g., Alaska (ALA), Greenland (GRL), southern Africa (SAF), and South and North Asia (SAS and NAS), the climate change signals in the low-resolution model are similar to those in the high-resolution model. Other regions [e.g., the Mediterranean basin (MED), Sahara (SAH), Western and Eastern Africa (WAF and

MRI–AGCM3.2 precipitation biases against GPCC7.0
top: JJA(1979–2003), bottom: DJF(1979/80–2003/04)
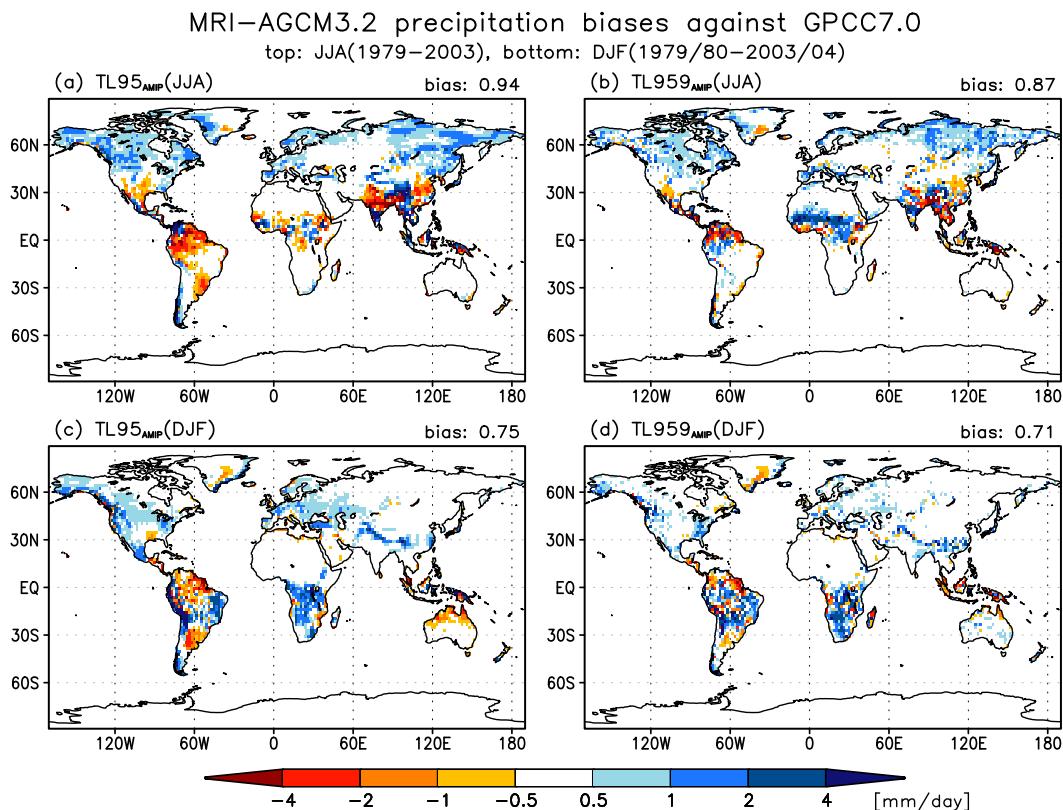
FIG. 2. Precipitation biases against Global Precipitation Climatology Centre (GPCC) precipitation version 7, for the 20C simulations by (a),(c) low-resolution TL95L64 and (b),(d) high-resolution TL959L64 for (a),(b) JJA in 1979–2003 and (c),(d) DJF in 1979/80–2003/04. Numbers at the top right of each panel indicate global mean biases.

EAF)] show large differences in climate change signals. In particular, over much of Europe, the low-resolution model shows a strong drying signal in summer, consistent with that found in CMIP3 and CMIP5 models (IPCC 2007, 2013). However, the high-resolution model (truth) only shows a weak dry or wet signal. Some higher-resolution climate models also tend to show a weaker drying signal (e.g., Delworth et al. 2012; Lau and Ploshay 2013; Demory et al. 2014). Rowell and Jones (2006) found that a larger land–sea contrast due to climatic warming, drying of soil moisture in spring, and large-scale atmospheric changes are all important drivers of projected summer drying over Europe, and that individual contributions to summer drying remain unclear, leading to a larger uncertainty in the magnitude of summer dry over Europe. One possible reason why the higher-resolution models show a weaker drying signal might be improved representations of circulation regimes [e.g., the North Atlantic Oscillation (NAO) and atmospheric blocking] due to increasing of horizontal resolution (Dawson and Palmer 2015; Dawson et al. 2012; Matsueda et al. 2010, 2009); this is currently under investigation.

3. Seasonal hindcasts: Measures of reliability

To investigate whether seasonal predictions can be used to calibrate probabilities of wet and dry seasons in the 21C climate change projections, seasonal retrospective forecasts were performed with the same low-resolution model as used in the 20C and 21C simulations, using the prescribed observed HadISST and SICs, which were also used in the 20C simulations. Note that the low-resolution seasonal forecasts were conducted without any changes in parameter settings.

The predictions were initialized with the Japanese 25-year Reanalysis Project (JRA-25; Onogi et al. 2007) around 1 May and 1 November of each year over the hindcast period 1979–2003. The forecast ensemble consists of 21 members and was run for 4 months to cover the JJA and December–February (DJF) seasons [e.g., the ensemble for JJA of 2003 was initialized at 1200 UTC 28 April to 1200 UTC 3 May (6 hourly)]. These predictions were then verified against the high-resolution model output for the 20C (truth). Figure 2 shows seasonal mean biases of precipitation against Global Precipitation Climatology Centre (GPCC)

Zonal mean T2m differences

(a)TL959$_{AMIP}$−JRA25 (JJA, 1979−2003)

(b)TL959$_{AMIP}$−JRA25 (DJF, 1979/80−2003/04)
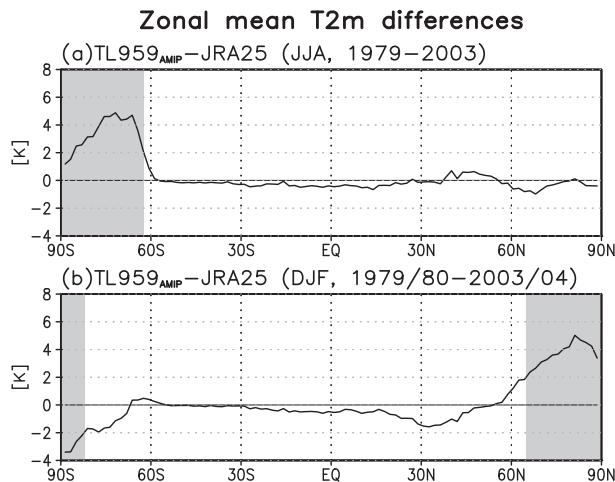
FIG. 3. Differences of zonal mean 2-m surface temperature between the TL959L64 model output and Japanese reanalysis (JRA-25) for (a) JJA in 1979–2003 and (b) DJF in 1979/80–2003/04. The gray shading indicates latitudes where the absolute value of differences in surface temperature is greater than 2 K.

precipitation, version 7 (Schneider et al. 2015), for the 20C simulations by the high- and low-resolution models. The high-resolution model has marginally lower bias for both JJA and DJF (especially in JJA). Following our comments in section 1, the high-resolution model output provides a reasonable estimate of truth, different from the low-resolution model, and somewhat closer to reality.

Ideally, the seasonal prediction should be initialized with the high-resolution model output for the 20C. However, the available high-resolution model output did not include enough variables to make initial conditions for the low-resolution seasonal prediction. Given that the boundary conditions provide greater contributions to seasonal simulations than initial conditions, the use of JRA-25 as initial condition for the low-resolution model seems a reasonable pragmatic solution.

Figure 3 shows differences in zonal mean 2-m surface temperature in both JJA and DJF between the hi-resolution model output (truth for 20C) and JRA-25. The differences are between −2 and 2 K (mostly −1 and 1 K) in the low- and midlatitudes, whereas the absolute values of differences in polar regions in winter are greater than 2 K. The large differences in polar regions are likely attributed to differences in surface conditions over land (i.e., snow) and can influence seasonal simulation, especially in the higher latitude. Therefore, land regions poleward of 60° were excluded in the following analyses for boreal winter.

As mentioned in the introduction, reliability is an essential characteristic for any climate forecasts to be useful in real-life decision-making (Weisheimer and Palmer 2014). A reliable forecasting system is one where
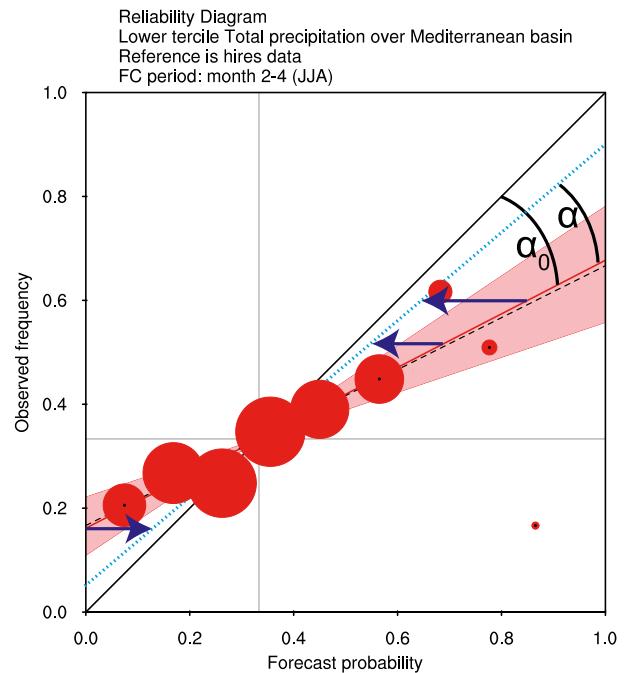
FIG. 4. Reliability diagram of dry JJA forecasts in 1979–2003 over the Mediterranean basin, simulated by the retrospective TL95L64 seasonal prediction. The TL95L64 seasonal prediction has been verified against the TL959L64 20C simulation (''truth'') for the same period.

the forecast probabilities for a certain event E match the corresponding observed frequency of occurrence of E, given the forecasts (Wilks 2011). Here, E can be any dichotomous meteorological event of interest. For this paper we follow common practice in seasonal forecasting and use precipitation events E based on terciles of their climatological distribution of the seasonal-mean precipitation. By definition, these events have a climatological probability of 1/3 in both the forecasts and observations (indicated by the gray lines in Fig. 4).

The reliability of a forecasting system can be graphically displayed in a reliability diagram where the observed frequencies of E are plotted as a function of the binned forecast probabilities (e.g., see Fig. 4). Here, the size of the red data points is proportional to the number of forecasts falling into that probability bin. To get a best estimate of the linear relationship between forecast probabilities and observed frequencies, a weighted linear regression was applied to these data (red line). The red shaded area around the best-estimate regression line is an estimate of the inherent sampling uncertainty (here the 75% confidence limit) derived from a bootstrapping resampling procedure.

We use the slope of the regression line and its uncertainty range to classify the reliability into five simple categories, following Weisheimer and Palmer (2014).
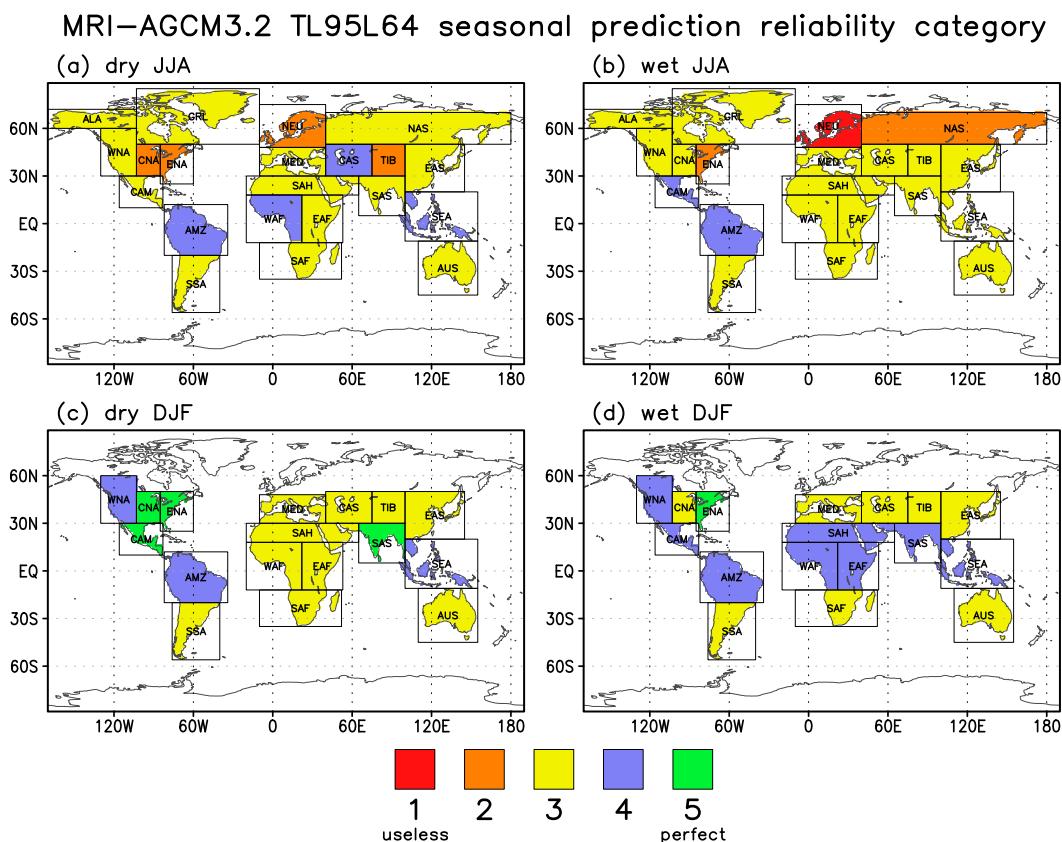
FIG. 5. Reliability categories for (a) dry JJA, (b) wet JJA, (c) dry DJF, and (d) wet DJF, simulated by the retrospective TL95L64 seasonal predictions. Category 5 indicates perfect reliability, whereas category 1 indicates dangerously useless reliability.

Perfect reliability would be achieved if there was, within the uncertainty range, a one-to-one correspondence between the forecast probability and the observed frequency (black diagonal line). We classify such forecasts as the highest category 5. Forecast in category 4 can still be very useful, while forecast in category 3 are considered marginally useful. If the regression line is flat, the forecasting system shows no correspondence between the forecast probabilities of E and the observed frequencies of occurrence of E. We classify such unreliable and thus useless forecasts as category 2. Category 1 is reserved for those few cases where the slope of the regression line is negative implying dangerously useless forecasts as decision-makers can be seriously misled by such forecast probabilities.

In Fig. 5, results are shown in terms of the reliability categories of the low-resolution seasonal forecasts using the high-resolution model as truth for the two events $E = precipitation\ above\ (below)\ the\ upper\ (lower)\ tercile$ in the JJA and DJF seasons for the same global land Giorgi regions of Fig. 1. Forecast reliability varies between regions, seasons, and the events considered. While

some events can be classified as perfectly reliable (parts of North and Central America in DJF), most events fall in the marginally reliable category 3. Wet conditions over northern Europe (NEU) in JJA classify as the least reliably predicted event.

4. Calibration of climate change predictions

We now use the reliability information to calibrate the climate change projections. To describe the procedure, we return to Fig. 4. Calibration to improved reliability is achieved by projecting the data points, and thus the best-guess regression line, toward the perfect reliability diagonal as indicated by the blue arrows. The calibration leads to calibrated forecast probabilities (while leaving the observed frequencies unchanged) and a steeper slope of the regression line. In this study we consider partial calibration: that is, calibrations that not necessarily project onto the perfect reliability diagonal (black line) but can take any value of the slope between the raw uncalibrated and the full perfectly calibrated slope. We introduce the calibration factor $\varepsilon = \alpha/\alpha_0$ which describes

Optimal ε minimising RMSD between "truth" and calibrated TL95
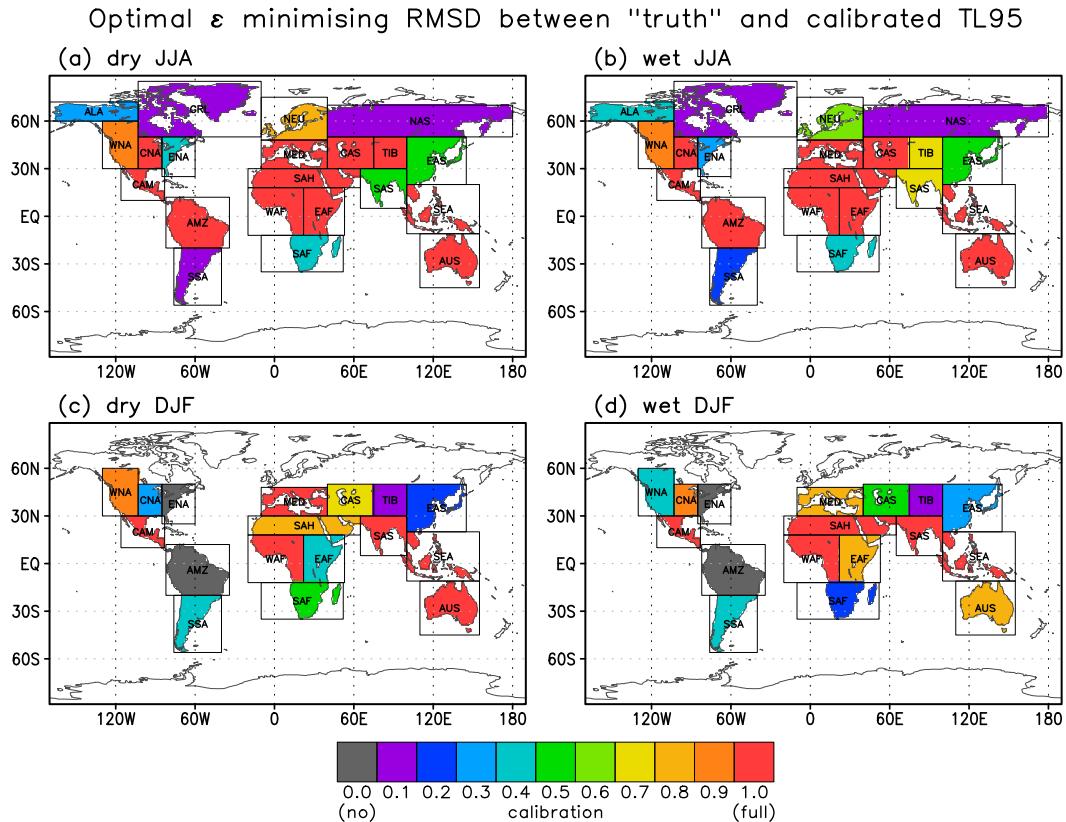
FIG. 6. Optimal ε that minimizes RMSD between the TL959L64 (truth) and the calibrated TL95L64 probabilities of (a) dry JJA, (b) wet JJA, (c) dry DJF, and (d) wet DJF for each Giorgi region.

the fraction of the full calibration that is used. Here α_0 corresponds to the calibration angle in case of full calibration ($\varepsilon = 1$) and α describes the partial calibration angle. Hence, if $\alpha = 0$, then no calibration is applied ($\varepsilon = 0$). The light blue dotted line in Fig. 4 indicates such a partially calibrated reliability curve.

Figure 6 illustrates the optimal ε that minimizes root-mean-square distances (RMSD) between the high-resolution and the ε-calibrated low-resolution probabilities of dry and wet JJA and DJF in the 21C for each Giorgi region. In JJA (Figs. 6a and 6b), the optimal value of ε is greater than 0.5 for most of the Giorgi regions, especially for lower latitudes. The regions at higher latitudes, especially GRL, NAS, and southern South America (SSA), tend to show a smaller value of ε. In DJF (Figs. 6c and 6d), the optimal ε for each region tends to be smaller than that in JJA. The optimal ε tends to be relatively large at low latitudes. The optimal ε for eastern North America (ENA) and the Amazon basin (AMZ) in DJF is 0.0, indicating that calibrations cannot reduce the RMSD of probabilities. However, note that ENA and AMZ in winter already have good reliability in categories 5 and 4, respectively, where any calibration will have a relatively small impact on probability.

Figure 7 shows the change in RMSD between the high-resolution (truth) and low-resolution probabilities of dry and wet JJA and DJF in the 21C by the optimal ε calibrations.

As discussed above, regions at higher latitudes in the boreal winter have been excluded. The key result is that the calibrations reduce the RMSD for all the 21 Giorgi regions, in JJA more than in DJF. The largest reduction in RMSD is seen for dry JJA over MED. The calibration reduced the RMSD over MED by 34%. The reductions over MED are also dominant for the other events (i.e., wet JJA, dry DJF, and wet DJF). The optimal values of ε for MED are 0.9 and 1.0 for wet DJF and the other events, respectively. A larger ε (i.e., the calibrated reliability category is expected to be 5) tends to lead to a larger reduction in RMSD for both JJA and DJF. The reliability category before the calibrations above each bar (also shown in Fig. 5) does not seem to be connected with the amount of the reduction in RMSD.

It is important to understand why calibration has a bigger impact in boreal summer than in boreal winter. We believe that an important reason may be that there are strong indications that atmospheric initial conditions are more important for an accurate prediction of the

Change in RMSD of probability
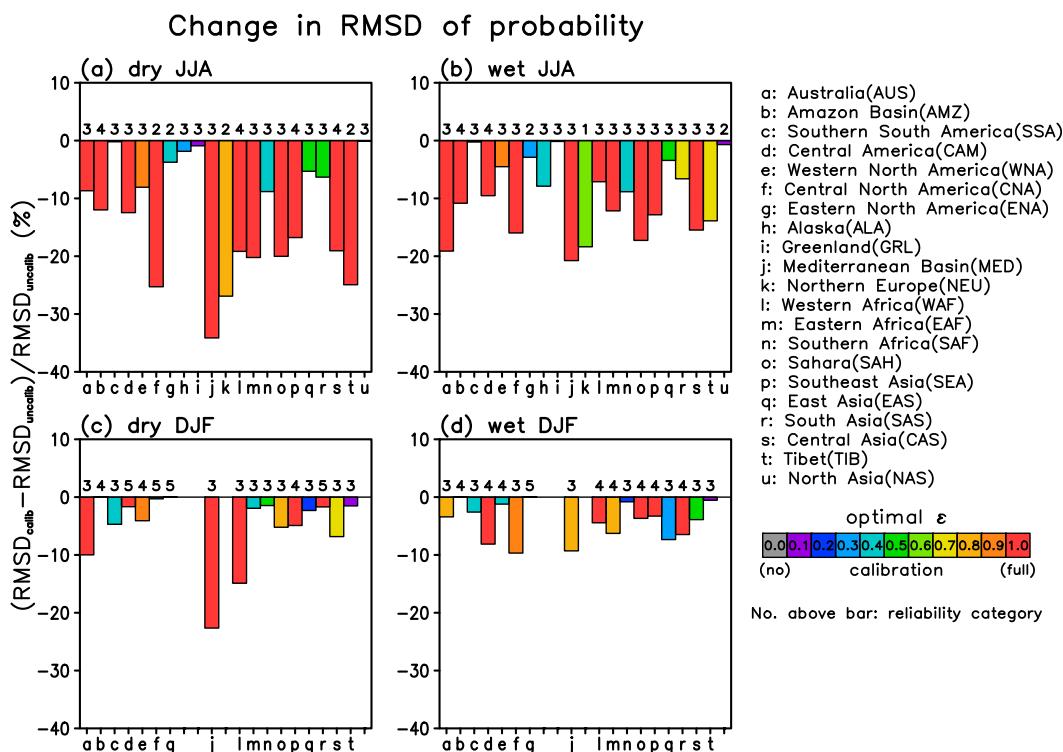
FIG. 7. Change in root-mean-square distances (RMSD) between the TL959L64 (truth) and TL95L64 probabilities of (a) dry JJA, (b) wet JJA, (c) dry DJF, and (d) wet DJF by the optimal ε calibrations. The color changes reflect the optimal ε. The number above each bar indicates the reliability category before the calibrations.

NAO/Arctic Oscillation, which can influence surface temperature and precipitation on continental scales, in boreal winter than in boreal summer (e.g., Stockdale et al. 2015; Scaife et al. 2014; Ineson and Scaife 2009). As noted above, in our study it was not possible for technical reasons to initialize our seasonal integrations with high-resolution output (truth). This might suggest that initializing the low-resolution model with the high-resolution output in the seasonal predictions would potentially lead to further reductions in RMSD, especially in boreal winter.

Figure 8 shows differences of changes in the frequency of dry JJA in 21C between the uncalibrated and calibrated low-resolution model simulations. As show in Fig. 7, the calibrated low-resolution frequencies with the optimal ε are closer to truth, especially at low latitudes, than the uncalibrated values.

Finally, we focus on MED where the biggest differences of changes in the frequency of dry JJA between high- and low-resolution models are seen and are reduced by the calibrations the most. Figure 9 shows the changes in the frequency of dry JJA events over MED, for 21C relative to 20C, derived from uncalibrated low resolution (Fig. 9a), with ε = 1.0, which is the optimal value for dry JJA events in that region, (Fig. 9b), and the

truth value from high-resolution (Fig. 9c). It can be seen that the calibration has substantially reduced the overestimation in the probability of dry events by 34% (i.e., the ε-calibrated low-resolution frequency with the optimal ε has become closer to truth).

5. Conclusions

Through the use of high- and low-resolution seasonal and climate integrations with prescribed sea surface temperature and sea ice at two different resolutions, this paper provides support for developing seamless weather and climate prediction models (i.e., where the climate models use essentially the same computer code for representing processes on interannual and shorter time scales as the weather and seasonal forecast models). However, it is important to note that we are not advocating that climate change scientists should be performing separate atmosphere-only or coupled seasonal integrations. Rather, the practical importance of the results in this paper derives from the fact that in an operational center running seasonal forecasts and climate change projections seamlessly, important information that can be used to improve the reliability of the climate change projections is essentially available at no extra computational cost.
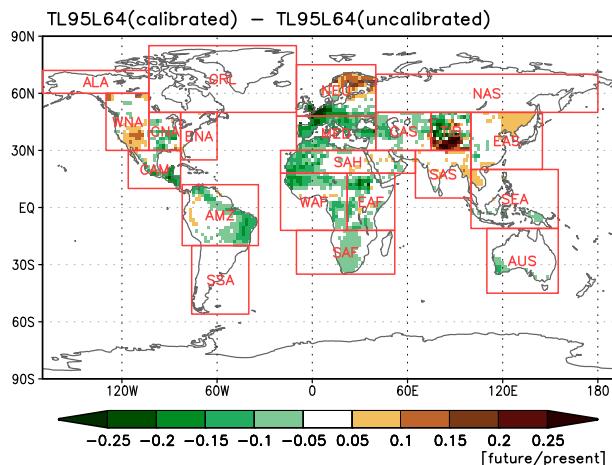
FIG. 8. Differences of changes in the frequency of dry JJA for the period 2075–99, relative to the period 1979–2003, between uncalibrated and calibrated TL95L64.

FIG. 9. Changes in the frequency of dry JJA over the Mediterranean basin, for the period 2075–99, relative to the period 1979–2003, derived from (a) uncalibrated TL95L64, (b) calibrated TL95L64 with $\varepsilon = 1.0$ (the optimal value for dry JJA events in that region), and (c) TL959L64 (truth). The hatching in (b) indicates significant changes in frequency by the calibrations.

In particular, it has been shown that information about the reliability of seasonal forecast ensembles can help improve the reliability of regional climate change projections of precipitation. By using a T959 high-resolution model as a surrogate for "truth." it has been demonstrated that future projections of precipitation from a T95 low-resolution model improve if they are calibrated for regions where seasonal retrospective forecasts with the low-resolution model are unreliable. In particular, it has been shown that the root mean squared distance of probabilities for wet and dry winter and summer seasons at the end of the twenty-first century is reduced if such a calibration was applied. The largest reduction in RMSD was found for dry JJA events over the Mediterranean basin. Interestingly, the drying Mediterranean area is exactly the event and region where climate projections of precipitation using uncalibrated low-resolution and high-resolution atmospheric models differ the most in terms of the strength of the signal indicating large uncertainties in the projections.

The methodology proposed here does not guarantee that the climate change predictions will be reliable—the proposed calibration scheme should be considered necessary but not sufficient for ensuring reliable climate change projections (Palmer et al. 2008, 2009). In particular, in this paper we have only tested our hypothesis in terms of the fast processes occurring in the atmosphere. For example, the application of this methodology to include uncertainties in slower oceanic processes (e.g., Andrejczuk et al. 2016) would require a study with coupled ocean–atmosphere models—this is work for the future. Also, it is important to note that we are not proposing to calibrate climate change projections based on the skill of the seasonal forecast results: a seasonal forecast system can be perfectly reliable and yet show little or no skill.
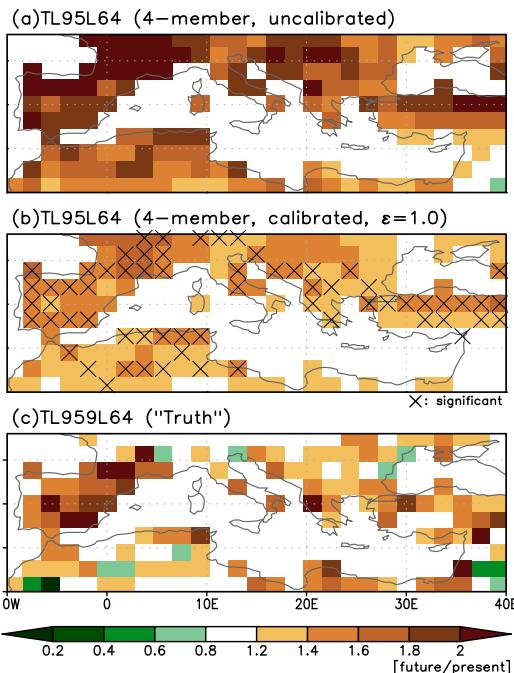
The results here show modest but consistent improvements in climate change skill with calibration. We expect to obtain more substantial increases in climate change skill if the seasonal forecast set was initialized with data from the high-resolution model. We aim to test this speculation in the future. In this paper, we have tested the seamless prediction idea by using an atmosphere-only model with prescribed SST and sea ice. Studies with a wider range of models, including coupled models, would be desirable to test the robustness of the results.

(University of Oxford) for his efforts to do verifications of the seasonal hindcasts and Mr. Eiki Shindo (MRI) for his efforts to provide codes to make the initial conditions for MRI-AGCM3.2.

REFERENCES

Andrejczuk, M., F. C. Cooper, S. Juricke, T. N. Palmer, A. Weisheimer, and L. Zanna, 2016: Oceanic stochastic parameterizations in a seasonal forecast system. *Mon. Wea. Rev.*, **144**, 1867–1875, doi:10.1175/MWR-D-15-0245.1.

Dawson, A., and T. N. Palmer, 2015: Simulating weather regimes: Impact of model resolution and stochastic parameterization. *Climate Dyn.*, **44**, 2177–2193, doi:10.1007/s00382-014-2238-x.

——, ——, and S. Corti, 2012: Simulating regime structures in weather and climate prediction models. *Geophys. Res. Lett.*, **39**, L21805, doi:10.1029/2012GL053284.

Delworth, T. L., and Coauthors, 2012: Simulated climate and climate change in the GFDL CM2.5 high-resolution coupled climate model. *J. Climate*, **25**, 2755–2781, doi:10.1175/JCLI-D-11-00316.1.

Demory, M.-E., P. L. Vidale, M. J. Roberts, P. Berrisford, J. Strachan, R. Schiemann, and M. S. Mizielinski, 2014: The role of horizontal resolution in simulating drivers of the global hydrological cycle. *Climate Dyn.*, **42**, 2201–2225, doi:10.1007/s00382-013-1924-4.

Giorgi, F., and R. Francisco, 2000: Uncertainties in regional climate change prediction: A regional analysis of ensemble simulations with the HADCM2 coupled AOGCM. *Climate Dyn.*, **16**, 169–182, doi:10.1007/PL00013733.

Ineson, S., and A. A. Scaife, 2009: The role of the stratosphere in the European climate response to El Niño. *Nat. Geosci.*, **2**, 32–36, doi:10.1038/ngeo381.

IPCC, 2007: *Climate Change 2007: The Physical Science Basis*. Cambridge University Press, 996 pp.

——, 2013: *Climate Change 2013: The Physical Science Basis*. Cambridge University Press, 1535 pp., doi:10.1017/CBO9781107415324.

Jenkins, G. J., J. M. Murphy, D. M. H. Sexton, J. A. Lowe, P. Jones, and C. G. Kilsby, 2010: UK Climate Projections: Briefing Report. Met Office Hadley Centre, Exeter, United Kingdom, 10 pp.

Lau, N.-C., and J. J. Ploshay, 2013: Model projections of the changes in atmospheric circulation and surface climate over North America, the North Atlantic, and Europe in the twenty-first century. *J. Climate*, **26**, 9603–9620, doi:10.1175/JCLI-D-13-00151.1.

Matsueda, M., and T. N. Palmer, 2011: Accuracy of climate change predictions using high resolution simulations as surrogates of truth. *Geophys. Res. Lett.*, **38**, L05803, doi:10.1029/2010GL046618.

——, R. Mizuta, and S. Kusunoki, 2009: Future change in wintertime atmospheric blocking simulated using a 20-km-mesh atmospheric global circulation model. *J. Geophys. Res.*, **114**, D12114, doi:10.1029/2009JD011919.

——, H. Endo, and R. Mizuta, 2010: Future change in Southern Hemisphere summertime and wintertime atmospheric blockings simulated using a 20-km-mesh AGCM. *Geophys. Res. Lett.*, **37**, L02803, doi:10.1029/2009GL041758.

Meehl, G. A., C. Covey, K. E. Taylor, T. Delworth, R. J. Stouffer, M. Latif, B. McAvaney, and J. F. B. Mitchell, 2007: The WCRP CMIP3 multimodel dataset: A new era in climate change research. *Bull. Amer. Meteor. Soc.*, **88**, 1383–1394, doi:10.1175/BAMS-88-9-1383.

Mizuta, R., Y. Adachi, S. Yukimoto, and S. Kusunoki, 2008: Estimation of the future distribution of sea surface temperature and sea ice using CMIP3 multi-model ensemble mean. Tech. Rep. 56, Meteorological Research Institute, Tsukuba, Japan, 28 pp., doi:10.11483/mritechrepo.56.

——, and Coauthors, 2012: Climate simulations using MRI-AGCM3.2 with 20-km grid. *J. Meteor. Soc. Japan*, **90A**, 233–258, doi:10.2151/jmsj.2012-A12.

Onogi, K., and Coauthors, 2007: The JRA-25 Reanalysis. *J. Meteor. Soc. Japan*, **85**, 369–432, doi:10.2151/jmsj.85.369.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quart. J. Roy. Meteor. Soc.*, **128**, 747–774, doi:10.1256/0035900021643593.

——, F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, 2008: Toward seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bull. Amer. Meteor. Soc.*, **89**, 459–470, doi:10.1175/BAMS-89-4-459.

——, ——, ——, and ——, 2009: Reply. *Bull. Amer. Meteor. Soc.*, **90**, 1551–1554, doi:10.1175/2009BAMS2916.1.

Rayner, N. A., D. E. Parker, E. B. Horton, C. K. Folland, L. V. Alexander, D. P. Rowell, E. C. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, doi:10.1029/2002JD002670.

Rowell, D. P., and R. G. Jones, 2006: Causes and uncertainty of future summer drying over Europe. *Climate Dyn.*, **27**, 281–299, doi:10.1007/s00382-006-0125-9.

Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519, doi:10.1002/2014GL059637.

Schneider, U., A. Becker, P. Finger, A. Meyer-Christoffer, B. Rudolf, and M. Ziese, 2015: GPCC full data reanalysis version 7.0 at 0.5°: Monthly land-surface precipitation from rain-gauges built on GTS-based and historic data. Deutscher Wetterdienst, accessed 16 April 2016, doi:10.5676/DWD_GPCC/FD_M_V7_050.

Stockdale, T. N., F. Molteni, and L. Ferranti, 2015: Atmospheric initial conditions and the predictability of the Arctic Oscillation. *Geophys. Res. Lett.*, **42**, 1173–1179, doi:10.1002/2014GL062681.

Weisheimer, A., and T. N. Palmer, 2014: On the reliability of seasonal climate forecasts. *J. Roy. Soc. Interface*, **11**, 9620131162, doi:10.1098/rsif.2013.1162.

Wilks, D., 2011: *Statistical Methods in the Atmospheric Science.* 3rd ed. International Geophysics Series, Vol. 100, Academic Press, 704 pp.