Improved forecast skill in the tropics in the new MiKlip decadal climate predictions

H. Pohlmann,[1] W. A. Müller,[1] K. Kulkarni,[1] M. Kameswarrao,[1] D. Matei,[1] F. S. E. Vamborg,[1] C. Kadow,[2] S. Illing,[2] and J. Marotzke[1]

[1] We introduce an improved initialization to the decadal predictions performed for the Mittelfristige Klimaprognosen (MiKlip) project based on the Max-Planck-Institute Earth System Model and furthermore test the effect of increased ocean and atmosphere model resolutions. The new initialization includes both a more sophisticated oceanic initialization and additionally an atmospheric initialization. We compare the performance of retrospective decadal forecasts over the past 50 years with that of the previous system. The new oceanic initialization considerably improves the performance in terms of surface air temperature over the tropical oceans on the 2–5 years time scale, which also helps to improve the predictive skill of global mean surface air temperature on this time scale. The higher model resolution improves the predictive skill of surface air temperature over the tropical Pacific even further. Through the newly introduced atmospheric initialization, the quasi-biennial oscillation exhibits predictive skill of up to 4 years when a sufficiently high vertical atmospheric resolution is used. **Citation:** Pohlmann, H., W. A. Müller, K. Kulkarni, M. Kameswarrao, D. Matei, F. S. E. Vamborg, C. Kadow, S. Illing, and J. Marotzke (2013), Improved forecast skill in the tropics in the new MiKlip decadal climate predictions, *Geophys. Res. Lett.*, *40*, 5798–5802, doi:10.1002/2013GL058051.

1. Introduction

[2] Decadal climate prediction is a relatively new research field. After the first pioneering work [*Smith et al.*, 2007; *Keenlyside et al.*, 2008; *Pohlmann et al.*, 2009], only recently a comprehensive set of decadal climate predictions with different systems was performed as part of the Coupled Model Intercomparison Project Phase 5 (CMIP5) [*Taylor et al.*, 2009, 2012]. For CMIP5, retrospective forecasts, so-called hindcasts, were performed over the period between 1960 and 2010 to be assessed in the upcoming fifth assessment report of the Intergovernmental Panel on Climate Change [e.g., *Goddard et al.*, 2013; *Doblas-Reyes et al.*, 2013; *Meehl et al.*, 2013]. The Max Planck Institute for Meteorology contributes to CMIP5 with hindcasts from its decadal prediction system, which is based on the climate model Max-Planck-Institute Earth System Model (MPI-ESM)

[*Giorgetta et al.*, 2013; *Stevens et al.*, 2013; *Jungclaus et al.*, 2013]. This system (named here baseline-0) makes use of an oceanic initialization from a forced ocean simulation. Within the MiKlip project, we develop a coupled oceanic and atmospheric initialization and additionally test the effect of increased model resolution. In this paper, we show the improvements achieved with the new system (named here baseline-1) and the effect of model resolution on predictive skill.

[3] *Müller et al.* [2012] analyze the predictive skill of the baseline-0 (b0) system. They show that the initialization of MPI-ESM improves forecast skill with respect to the uninitialized experiment predominantly over the North Atlantic for all lead times and over parts of Europe for multiyear seasonal means. However, negative skill scores over the tropical Pacific reflect a systematic error in the initialization. As a consequence, the overall skill, for example, in terms of global mean temperature, is lower than in other systems [*Bellucci et al.*, 2012]. The reason for this problem is not fully understood. Flaws in the wind forcing [*Lee et al.*, 2013] of the ocean model may cause an overly strong ocean response that forces the coupled model to adjust by inducing unrealistic heat fluxes.

[4] Building on our experience from testing three different ocean initializations [*Kröger et al.*, 2012], we here initialize the ocean component with the newest oceanic reanalysis from the European Centre for Medium-Range Weather Forecasts (ECMWF) [*Balmaseda et al.*, 2012]. Additionally, the positive experience of other decadal prediction groups with the initialization of the atmosphere [e.g., *Smith et al.*, 2007] led us to also introduce an initialization of this component from ECMWF atmospheric reanalysis data [*Uppala et al.*, 2005; *Dee et al.*, 2011]. In the remainder of this paper, we give an overview of the two systems followed by an analysis of their differences in performance. Particular emphasis is placed on the tropical region, for which the skill in surface temperature prediction improves substantially. We particularly highlight predictive skill for the quasi-biennial oscillation (QBO) of equatorial stratospheric zonal winds [*Baldwin et al.*, 2001].

2. Simulations and Methods

[5] The initialization method is briefly summarized here: Estimates of the oceanic temperature and salinity fields for the period 1948–2012 are produced by forcing the Max Planck Institute ocean model [*Jungclaus et al.*, 2013] with daily fluxes of momentum, heat, and freshwater taken from the National Centers for Environmental Prediction/National Center for Atmospheric Research reanalysis [*Kalnay et al.*, 1996]. The anomaly technique [*Pierce et al.*, 2004; *Smith et al.*, 2013] is used to initialize the decadal hindcasts with these fields. An ensemble simulation of three decadal hindcasts is
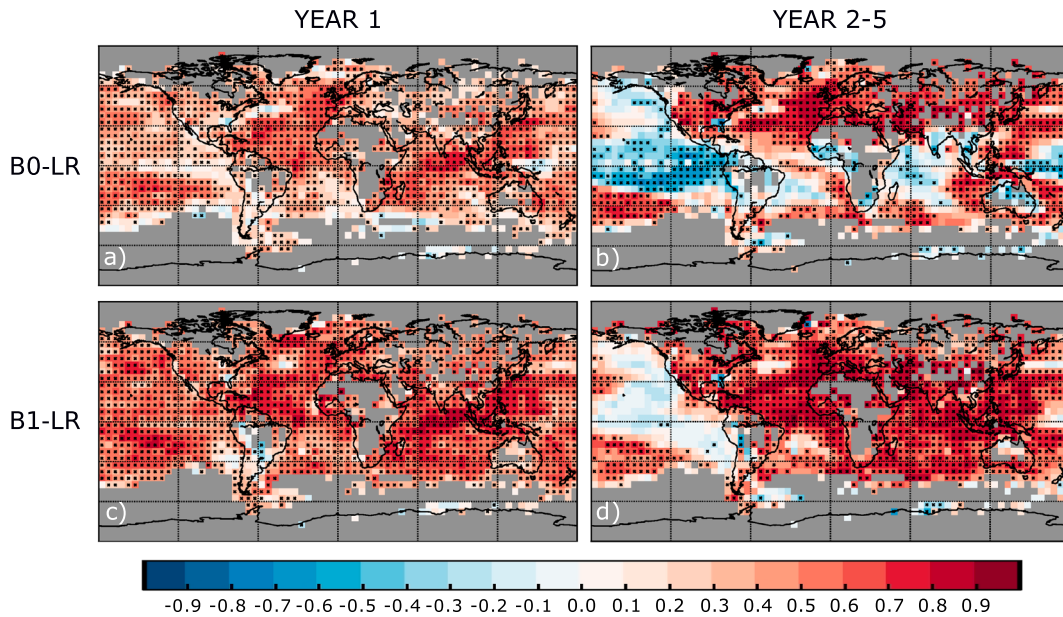

Figure 1. Maps of ensemble mean hindcast skill (anomaly correlation) of surface air temperature averaged over the (a and c) first prediction year and (b and d) years 2–5 for b0-LR in Figures 1a and 1b and b1-LR in Figures 1c and 1d against observation from HadCRUT3v over the period 1961–2012. Crosses denote skill exceeding the 5–95% confidence level.

started with MPI-ESM from consecutive days around 1 January each year from 1961 to 2012. The MPI-ESM in low resolution (LR, atmosphere: T63L47, ocean: 1.5°L40) is employed for this set of simulations.

[6] The baseline-1 (b1) system uses the same coupled model MPI-ESM as the baseline-0 (b0) system. However, the oceanic component is initialized with temperature and salinity anomalies from the ocean reanalysis system 4 (ORAS4) from ECMWF [*Balmaseda et al.*, 2012]. Additionally, the atmospheric component is initialized with full-field 3-D temperature, vorticity, divergence, and surface pressure fields with the data from ECMWF Re-Analysis (ERA)-40 [*Uppala et al.*, 2005] for the period 1960–1989 and ERA-Interim [*Dee et al.*, 2011] for the period 1990–2012, respectively. An ensemble of 10 decadal hindcasts is started around 1 January in the same way as in b0 (lagged initialization) over the period 1961–2012 with the LR system. However, a higher oceanic resolution would potentially improve the climate predictions [e.g., *Kirtman et al.*, 2012], and a higher vertical atmospheric resolution might resolve stratospheric processes more realistically [*Marshall and Scaife*, 2010; *Charlton-Perez et al.*, 2013]. Therefore, the method is repeated with the mixed resolution (MR, atmosphere: T63L95, ocean: 0.4°L40) version of MPI-ESM but with a smaller ensemble size of only five-ensemble members, around 1 January in the same way as in b0 (lagged initialization) over the period 1961–2012.

[7] To base our analysis on the same ensemble sizes, only results from the ensemble means of the first three-ensemble members (the maximum number with yearly initialization in b0-LR) are shown. The ensemble mean generally outperforms individual ensemble members [*Palmer et al.*, 2008]. We have convinced ourselves of the robustness of the results by comparison with other combinations wherever possible (b1-LR and b1-MR). The prediction skill is analyzed in the following section in terms of anomaly correlation [e.g., *Wilks*, 2011]. We also show root-mean-square error

(RMSE) skill scores [e.g., *Wilks*, 2011] using the uninitialized model simulations as a reference [*Goddard et al.*, 2013; *Matei et al.*, 2012] in the supporting information of this paper. Significance is estimated using a block bootstrap method [e.g. *Wilks*, 2011] considering for autocorrelation as in *Goddard et al.* [2013].

3. Results

[8] The variability of the ensemble mean 2 m air temperature from the b0-LR and b1-LR hindcasts is verified against observations from Hadley Centre and Climate Research Unit (HadCRUT3v) [*Brohan et al.*, 2006] for different prediction lead times in Figure 1. The anomaly correlation skill of the first prediction year (Figures 1a and 1c) is positive and significant almost everywhere, mainly reflecting that the observed warming trend over the period 1961–2012 is correctly represented in the first prediction year. This result is similar to findings in other studies [*Kim et al.*, 2012; *Hazeleger et al.*, 2013]. For hindcasts averaged over years 2–5, however, negative correlation skill appears in the b0-LR system in the tropics and eastern North Pacific with highest magnitudes in the tropical East Pacific (Figure 1b). A detailed analysis of the region with the negative predictive skill (not shown) reveals that the observed warming trend with relatively cool years in the 1960s and 1970s and relatively warm years in the 1990s and 2000s is reversed in the hindcasts over this prediction lead time. The problem is reduced in the b1-LR system: In the tropical Atlantic, Indian Ocean, and western Pacific, the correlation skill is positive almost everywhere (Figure 1d).

[9] For the first prediction year, b1-LR results in significant improvements over b0-LR in areas of the tropical and North Pacific, North Atlantic, and Southern Ocean (Figure 2a). For the hindcasts averaged over years 2–5, considerable improvement is achieved almost everywhere in the tropics (Figure 2b). A sensitivity study without assimilating the atmosphere
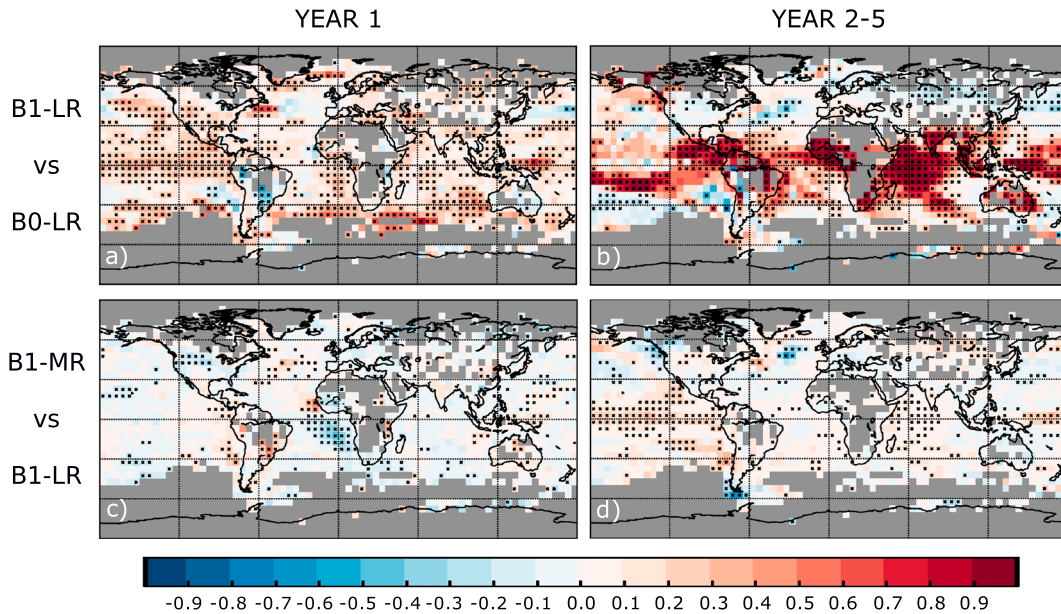
Figure 2. Differences of anomaly correlation skill. (a and b) b1-LR minus b0-LR. (c and d) b1-MR minus b1-LR. Year 1 in Figures 2a and 2c. Years 2–5 in Figures 2b and 2d. Crosses denote differences exceeding the 5–95% confidence level.

(not shown) reveals that the skill improvement is mainly due to the different oceanic initializations. The step from b1-LR to b1-MR has only a small effect for the first prediction year (Figure 2c). However, for hindcasts averaged over years 2–5, an additional improvement is achieved in the tropical Pacific with the higher model resolution (Figure 2d). Very similar results are obtained for the RMSE skill scores (Figures S1 and S2 in the supporting information).

[10] The negative correlation skill over the tropical Pacific in b0-LR also affects the correlation skill of the global mean surface air temperature. The analysis of annual averages shows that the hindcasts with lead times of 2 and 3 years are problematic in b0-LR (Figure 3a). Even the 1–4 and 2–5 years lead time global mean temperature averages are clearly affected by this problem (Figure 3b). As a consequence, the skill of the b0-LR system is lower than in other decadal prediction systems [*Bellucci et al.*, 2012]. Apart from this problem, for the global mean surface air temperature, the hindcasts of the different baseline systems are relatively similar to the uninitialized simulations for all lead times.

[11] In the following, we analyze the effect of the initialization of the atmosphere in the baseline-1 system as compared to the baseline-0 hindcasts with an uninitialized atmosphere. A candidate for multiannual predictive skill of the atmosphere is the quasi-biennial oscillation in the stratosphere. The winds in the equatorial stratosphere change direction with a period of roughly 28 months [*Baldwin et al.*, 2001]. We here define a QBO index as the time series of monthly and zonal mean zonal wind anomalies at 20 hPa averaged between 10°S and 10°N. The relatively low vertical atmospheric resolution in MPI-ESM-LR does not allow for spontaneous QBO variability; however, the MPI-ESM-MR produces a spontaneous QBO due to its higher vertical atmospheric resolution [*Schmidt et al.*, 2013; *Krismer et al.*, 2013]. Given this situation, it is clear that the LR runs without the atmospheric initialization do not reveal any QBO variability (Figure 4a). With atmospheric initialization in the b1-LR model, the initial phase of the QBO is in alignment with observations, but thereafter the QBO amplitude decays on time scales of the order of months (Figures 4a and 4c), as the model at low vertical resolution cannot simulate wave
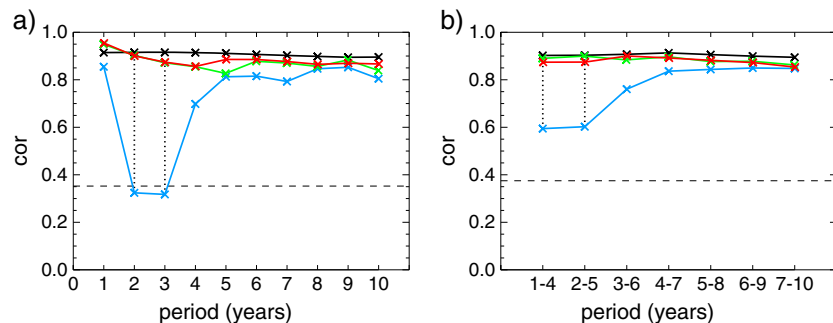

Figure 3. Hindcast skill (anomaly correlation) of the global and ensemble mean surface air temperature as a function of lead time verified against observations from HadCRUT3v for b0-LR (blue), b1-LR (green), b1-MR (red), and uninitialized (black) for (a) annual averages and (b) 4-year averages. Dashed black and dotted black lines denote skill and difference of skill between b1-MR and b0-LR exceeding the 95% confidence level.
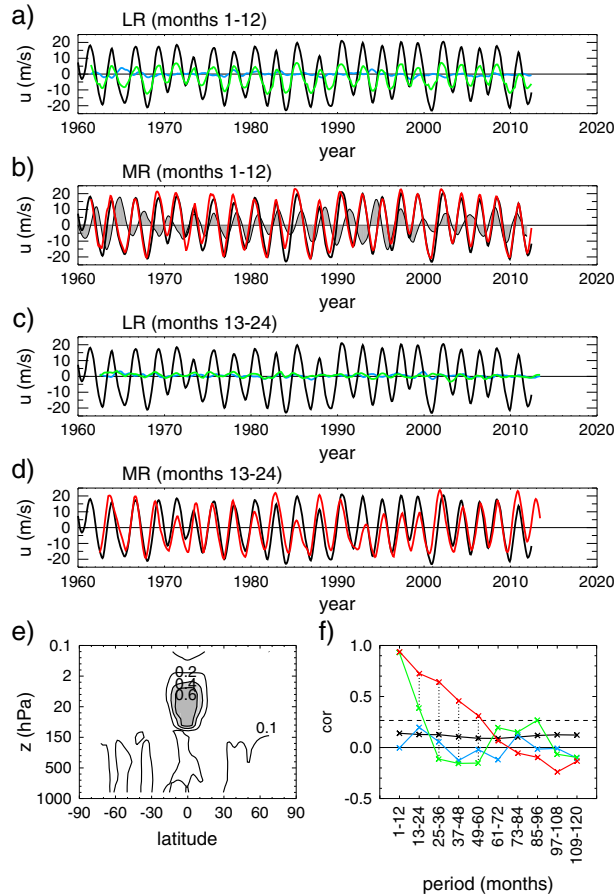
Figure 4. (a and b) Time series of monthly and zonal mean zonal wind anomalies at 20 hPa averaged between 10°S and 10°N (QBO) from ERA-40 and ERA-Interim reanalyses (black) together with ensemble means of the first 12 months of the hindcasts (combined into one time series) for b0-LR (blue) and b1-LR (green) in Figure 4a and b1-MR (red) together with the uninitialized (MR) simulation (grey shaded) in Figure 4b. (c and d) As in Figures 4a and 4b but for the hindcast months 13–24. (e) Hindcasts skill (anomaly correlation) of ensemble mean QBO of b1-MR for hindcast months 13–24 as a function of latitude and height. The shaded area exceeds the 95% confidence level. (f) Hindcast skill (anomaly correlation) of ensemble mean QBO at 20 hPa as a function of lead time for b0-LR (blue), b1-LR (green), b1-MR (red), and uninitialized (black). Dashed black and dotted black lines denote skill (being different to zero) and difference of skill between b1-MR and b1-LR exceeding the 95% confidence level. All data are smoothed with a 1-year running mean.

mean flow interaction that is essential to simulate the QBO. This is similar to the time scale found in earlier studies [e.g., *Hamilton and Yuan*, 1992]. With the higher vertical atmospheric resolution in the b1-MR system, however, the QBO is simulated and due to the initialization remains in alignment with observations well beyond the first 12 months (Figure 4d). The region with significant predictive skill extends from about 15°S to 15°N and 10 to 70 hPa for the prediction lead time of 13–24 months (Figure 4e). The analysis of all lead times shows that only in the b1-MR system the

atmospheric initialization does lead to predictive skill that remains significant for up to 4 years (Figure 4f). All other systems: b0-LR, b0-MR (not shown), and b1-LR lack either the ability to simulate the QBO or the initialization of the atmosphere and hence the QBO.

4. Discussion and Conclusions

[12] Although progress in terms of prediction quality is shown for the new MiKlip decadal prediction system baseline-1 compared to baseline-0, some smaller issues like the relatively low predictive skill in the tropical East Pacific are still present. Additionally, the absence of the pause in global warming (so-called "hiatus" period) after the year 2000 in historical runs in many climate models (including ours) [*Meehl et al.*, 2011; *Guemas et al.*, 2013; *Kosaka and Xie*, 2013] may also affect decadal forecasts. In both of our systems (b0 and b1), the hindcasts that are started during the hiatus period initially stay close to observations, but they drift after only a few years toward a too warm state. Caution is therefore required when actual forecasts are issued, as this drift may limit skill to much less than a decade. Since decadal climate prediction builds a bridge between climate projections and observations through initialization, studying their drift may help to understand and solve this problem.

[13] The improvements achieved from the transition from baseline-0 toward baseline-1 include the global mean surface air temperature. The problem of negative predictive skill over the tropical oceans, which is present for the 2–5 years prediction lead time in the baseline-0 system, is largely reduced in the baseline-1 system. This is mainly achieved by an improved oceanic initialization with data from the ORAS4 reanalysis. Additionally, the new atmospheric initialization in baseline-1 allows a skillful prediction of the QBO up to 48 months when the MR model is used. The predictability of the QBO bears a large potential impact on the variability of the circulation in the tropics but possibly also in the (stratospheric) extratropics [e.g., *Baldwin et al.*, 2001].

References

Baldwin, M. P., et al. (2001), The quasi-biennial oscillation, *Rev. Geophys.*, *39*, 179–229, doi:10.1029/1999RG000073.

Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2012), Evaluation of the ECMWF ocean reanalysis system ORAS4, *Q. J. R. Meteorol. Soc.*, *139*, 1132–1161, doi:10.1002/qj.2063.

Bellucci, A., et al. (2012), An assessment of a multi-model ensemble of decadal climate predictions performed within the framework of the COMBINE project. COMBINE Technical Report No. 2, available from http://www.combine-project.eu/fileadmin/user_upload/combine/tech_report/COMBINE_TECH_REP_n02.pdf.

Brohan, P., J. J. Kennedy, I. Harris, S. F. B. Tett, and P. D. Jones (2006), Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850, *J. Geophys. Res.*, *111*, D12106, doi:10.1029/2005JD006548.

Charlton-Perez, A. J., et al. (2013), On the lack of stratospheric dynamical variability in low-top versions of the CMIP5 models, *J. Geophys. Res. Atmos.*, *118*, 2494–2505, doi:10.1002/jgrd.50125.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.

Doblas-Reyes, F. J., I. Andreu-Burillo, Y. Chikamoto, J. García-Serrano, V. Guemas, M. Kimoto, T. Mochizuki, L. R. L. Rodrigues, and G. J. van Oldenborgh (2013), Initialized near-term regional climate change prediction, *Nat. Commun.*, *4*, 1715, doi:10.1038/ncomms2704.

Giorgetta, M. A., et al. (2013), Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5, *J. Adv. Model. Earth Syst.*, *5*, 572–597, doi:10.1002/jame.20038.

Goddard, L., et al. (2013), A verification framework for interannual-to-decadal predictions experiments, *Clim. Dyn.*, *40*(1–2), 245–272, doi:10.1007/s00382-012-1481-2.

Guemas, V., F. J. Doblas-Reyes, I. Andreu-Burillo, and M. Asif (2013), Retrospective prediction of the global warming slowdown in the past decade, *Nat. Clim. Change*, *3*, 649–653, doi:10.1038/NCLIMATE1863.

Hamilton, K., and L. Yuan (1992), Experiments on tropical stratospheric mean-wind variations in a spectral general circulation model, *J. Atmos. Sci.*, *49*, 2464–2483, doi:10.1175/1520-0469(1992)049<2464:EOTSMW>2.0.CO;2.

Hazeleger, W., B. Wouters, G. J. van Oldenborgh, S. Corti, T. Palmer, D. Smith, N. Dunstone, J. Kröger, H. Pohlmann, and J.-S. von Storch (2013), Predicting multi-year North Atlantic Ocean variability, *J. Geophys. Res. Oceans*, *118*, 1087–1098, doi:10.1002/jgrc.20117.

Jungclaus, J. H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, and J. S. von Storch (2013), Characteristics of the ocean simulations in MPIOM, the ocean component of the MPI-Earth system model, *J. Adv. Model. Earth Syst.*, *5*, 422–446, doi:10.1002/jame.20023.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Am. Meteorol. Soc.*, *77*, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Keenlyside, N. S., M. Latif, J. Jungclaus, L. Kornblueh, and E. Roeckner (2008), Advancing decadal-scale climate prediction in the North Atlantic sector, *Nature*, *453*, 84–88, doi:10.1038/nature06921.

Kim, H.-M., P. J. Webster, and J. A. Curry (2012), Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts, *Geophys. Res. Lett.*, *39*, L10701, doi:10.1029/2012GL051644.

Kirtman, B. P., et al. (2012), Impact of ocean model resolution on CCSM climate simulations, *Clim. Dyn.*, *39*, 1303–1328, doi:10.1007/s00382-012-1500-3.

Kosaka, Y., and S.-P. Xie (2013), Recent global-warming hiatus tied to equatorial Pacific surface cooling, *Nature*, *501*, 403–407, doi:10.1038/nature12534.

Krismer, T. R., M. A. Giorgetta, and M. Esch (2013), Seasonal aspects of the quasi-biennial oscillation in the Max Planck Institute Earth System Model and ERA-40, *J. Adv. Model. Earth Syst.*, *5*, 406–421, doi:10.1002/jame.20024.

Kröger, J., W. A. Müller, and J. S. von Storch (2012), Impact of different ocean reanalyses on decadal climate prediction, *Clim. Dyn.*, *39*, 795–810, doi:10.1007/s00382-012-1310-7.

Lee, T., D. E. Waliser, J.-L. F. Li, F. W. Landerer, and M. M. Gierach (2013), Evaluation of CMIP3 and CMIP5 wind stress climatology using satellite measurements and atmospheric reanalysis products, *J. Clim.*, *26*, 5810–5826, doi:10.1175/JCLI-D-12-00591.1.

Marshall, A. G., and A. A. Scaife (2010), Improved predictability of stratospheric sudden warming events in an atmospheric general circulation model with enhanced stratospheric resolution, *J. Geophys. Res.*, *115*, D16114, doi:10.1029/2009JD012643.

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

Meehl, G. A., J. M. Arblaster, J. T. Fasullo, A. Hu, and K. E. Trenberth (2011), Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods, *Nat. Clim. Change*, *1*, 360–364, doi:10.1038/nclimate1229.

Meehl, G. A., et al. (2013), Decadal climate prediction: An update from the trenches, *Bull. Am. Meteorol. Soc.*, doi:10.1175/BAMS-D-12-00241.1.

Müller, W. A., J. Baehr, H. Haak, J. H. Jungclaus, J. Kröger, D. Matei, D. Notz, H. Pohlmann, J. S. von Storch, and J. Marotzke (2012), Forecast skill of multi-year seasonal means in the decadal prediction system of the Max Planck Institute for Meteorology, *Geophys. Res. Lett.*, *39*, L22707, doi:10.1029/2012GL053326.

Palmer, T. N., F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell (2008), Toward seamless prediction calibration of climate change projections using seasonal forecasts, *Bull. Am. Meteorol. Soc.*, *89*, 459–470, doi:10.1175/BAMS-89-4-459.

Pierce, D. W., T. P. Barnett, R. Tokmakian, A. Semtner, M. Maltrud, J. Lysne, and A. Craig (2004), The ACPI project, element 1: Initializing a coupled climate model from observed conditions, *Clim. Change*, *62*, 13–28, doi:10.1023/B:CLIM.0000013676.42672.23.

Pohlmann, H., J. H. Jungclaus, A. Köhl, D. Stammer, and J. Marotzke (2009), Initializing decadal climate predictions with the GECCO oceanic synthesis: Effects on the North Atlantic, *J. Clim.*, *22*, 3926–3938, doi:10.1175/2009JCLI2535.1.

Schmidt, H., et al. (2013), Response of the middle atmosphere to anthropogenic and natural forcings in the CMIP5 simulations with the Max Planck Institute Earth system model, *J. Adv. Model. Earth Syst.*, *5*, 98–116, doi:10.1002/jame.20014.

Smith, D. M., S. Cusack, A. W. Colman, C. K. Folland, G. R. Harris, and J. M. Murphy (2007), Improved surface temperature prediction for the coming decade from a global climate model, *Science*, *317*, 796–799, doi:10.1126/science.1139540.

Smith, D. M., R. Eade, and H. Pohlmann (2013), A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction, *Clim. Dyn.*, doi:10.1007/s00382-013-1683-2.

Stevens, B., et al. (2013), Atmospheric component of the MPI-M Earth System Model: ECHAM6, *J. Adv. Model. Earth Syst.*, *5*, 146–172, doi:10.1002/jame.20015.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2009), A summary of the CMIP5 experiment design. PCDMI Rep., 33 pp, available from http://cmip-pcmdi.llnl.gov/cmip5/docs/Taylor_CMIP5_design.pdf.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An oveview of CMIP5 and the experiment design, *Bull. Am. Meteorol. Soc.*, *93*, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, *131*, 2961–3012, doi:10.1256/qj.04.176.

Wilks, D. S. (2011), *Statistical Methods in the Atmospheric Sciences*, International geophysics series, 3rd ed., vol. 100, Academic Press, Elsevier, Amsterdam, The Netherlands.