AGU PUBLICATIONS

Geophysical Research Letters

RESEARCH LETTER
10.1002/2015GL064833

Key Points:
- Modini initialization has decadal hindcast skill in the Pacific
- Modini initialization has skill at hindcasting the recent hiatus
- Forecast out to 2024 using Modini initialization suggests an end to the hiatus period

Supporting Information:
- Figures S1–S6

Correspondence to:
M. Thoma,
Malte.Thoma@awi.de

Citation:
Thoma, M., R. J. Greatbatch, C. Kadow, and R. Gerdes (2015), Decadal hindcasts initialized using observed surface wind stress: Evaluation and prediction out to 2024, *Geophys. Res. Lett.*, *42*, 6454–6461, doi:10.1002/2015GL064833.

Received 10 JUN 2015
Accepted 22 JUL 2015
Accepted article online 24 JUL 2015
Published online 6 AUG 2015

Decadal hindcasts initialized using observed surface wind stress: Evaluation and prediction out to 2024

Malte Thoma[1], Richard J. Greatbatch[2], Christopher Kadow[3], and Ruediger Gerdes[1]

[1]Alfred Wegener Institute Helmholtz Center for Polar and Marine Research, Bremerhaven, Germany, [2]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany, [3]Institute of Meteorology, Freie Universität Berlin, Berlin, Germany

Abstract We use surface air temperature to evaluate the decadal forecast skill of the fully coupled Max Planck Institut Earth System Model (MPI-ESM) initialized using only surface wind stress applied to the ocean component of the model (Modini: Model initialization by partially coupled spin-up). Our analysis shows that the greenhouse gas forcing alone results in a significant forecast skill on the 2–5 and 6–9 year range even for uninitialized hindcasts. For the first forecast year, the forecast skill of Modini is generally comparable with previous initialization procedures applied to MPI-ESM. But only Modini is able to generate a significant skill (correlation) in the tropical Pacific for a 2–5 year (and to a lesser extent for a 6–9 year) hindcast. Modini is also better able to capture the observed hiatus in global warming in hindcast mode than the other methods. Finally, we present forecasts for 2015 and the average of years 2016–2019 and 2020–2024, predicting an end to the hiatus.

1. Introduction

It is possible nowadays to forecast the weather a few days to a week ahead with growing confidence. This task is tackled by providing coherent initial conditions for numerical atmospheric models. In contrast, climate projections for a century ahead are based on coupled atmosphere-ocean models, which are forced with prescribed greenhouse gas (GHG) concentrations and aerosols based on future projections. These projections serve as the scientific basis for the International Panel on Climate Change (IPCC) reports and are getting more robust and authoritative with each update from the first report [*Houghton et al.*, 1990] to the most recent one [e.g., *Stocker et al.*, 2013]. However, providing reliable forecasts on an intermediate (seasonal to decadal) timescale is still challenging. Numerical simulations for this time frame require both initial conditions (like weather forecasts) and boundary conditions (like climate projections). This leads to so-called initialized climate predictions (or *forecasts*) [e.g., *Cox and Stephenson*, 2007; *Hawkins and Sutton*, 2009; *Kröger et al.*, 2012].

In general, model limitations (like resolution and the necessity to parameterize unresolved processes) fundamentally limit the ability of models to reproduce the observed climate leading to model biases [e.g., *Wang et al.*, 2014]. Thus, even if the model is initialized in the best possible way, the model forecast can be obscured by drift toward the model's own inherent climate. In spite of these inevitable limitations for decadal forecasts, it is possible to estimate the forecast skill of numerical models by applying statistical methods [e.g., *Goddard et al.*, 2013].

The ongoing German research project *Decadal Predictions* (MiKlip, Mittelfristige Klimaprognosen) aims to develop a system for climate forecasts for up to a decade ahead. This system is based on the *Max Planck Institute Earth System Model* (MPI-ESM). Three generations of decadal forecast systems have been developed and released during the last years, namely, *Baseline-0* (B0), *Baseline-1* (B1), and *Prototype* (PT) [e.g., *Pohlmann et al.*, 2013; *Kruschke et al.*, 2015]. Here we refer to B0, B1, and PT as MPI forecast systems. The only (but essential) difference between these systems is the initialization procedure prior to any forecast, which we shortly recall in section 2.

Here we present *model initialization by partially coupled spin-up* (Modini) as an alternative initialization procedure for MPI-ESM. *Thoma et al.* [2015] showed that Modini-MPI-ESM is able to reproduce the observed timing of climate events or shifts as for instance observed in the El Niño Southern Oscillation (ENSO), the Pacific Decadal Oscillation (PDO), or the Atlantic Meridional Overturning Circulation (AMOC) during the initialization phase. *Ding et al.* [2013] already demonstrated that Modini has potential as an initialization technique

Table 1. Compilation of Initialization Procedures Used by MPI-ESM Forecast Systems Described in the Text[a]

	Ocean (MPIOM)	A/F	Atmosphere (ECHAM6)	A/F	Ensemble Members
CMIP5	–	–	–	–	10
Baseline-0	NCEP	A	–	–	10(3)
Baseline-1	ORA-S4	A	3-D TVD and Surface Pressure (ERA40 and ERAI)	F	10
Prototype	ORA-S4	F	3-D TVD and Surface Pressure (ERA40 and ERAI)	F	15
Modini	Wind Stress (NCEPcfsr)	A	–	–	18

[a]A = *anomaly* initialization and F = *full-field* initialization; 3-D TVD stands for a three-dimensional initialization of temperature, vorticity, and divergence. CMIP5 refers to the MPI-ESM combination of the historical (until 2005) and the RCP4.5 scenarios (thereafter) without initialization. For Baseline-0 10(3) ensemble members exist for 1990, 1995, and 2000–2005 (1991–1994 and 1996–1999).

with the coarser resolved *Kiel Climate Model* (KCM) [*Park et al.*, 2009] showing some success at hindcasting the 1976/1977 and 1998/1999 climate shifts in the Pacific. In this study, we first describe the different initialization systems in section 2. In section 3, we then apply a standardized tool, developed within the MiKlip framework [*Illing et al.*, 2014; *Kadow et al.*, 2015], to evaluate and compare the Modini-MPI-ESM forecast skill with uninitialized forecasts and the initialized MPI forecast systems. Finally we provide a summary and discussion in section 4 and also present a prediction for the near-surface air temperature for the years 2015–2024.

2. Model and Experimental Setup

The common basis for all decadal forecast systems described in this article is the Max Planck Institute Earth System Model (MPI-ESM) in its LR (low resolution) configuration. This model has contributed successfully to the *Coupled Model Intercomparison Project Phase 5* (CMIP5), which has been the basis for the fifth IPCC report [*Stocker et al.*, 2013]. The ocean component (called Max Planck Institute Ocean Model (MPIOM)) [*Jungclaus et al.*, 2013] has 40 vertical levels and a horizontal resolution of about 12 to 150km on a curvilinear orthogonal grid with poles over Antarctica and Greenland. The atmospheric component ECHAM6 has a horizontal resolution of T63 (about 200km) with 47 vertical levels including the upper stratosphere up to 0.1hPa [*Stevens et al.*, 2013]. The MPI-ESM-LR is forced solely with changing radiative boundary conditions and known volcanic eruptions (identical to the CMIP5-*historical* and *RCP4.5* experiments) during hindcast/forecast periods. The essential differences between the evaluated decadal forecast systems is the initialization procedure prior to the forecast period.

In Baseline-0 (B0) the initial conditions are taken from an ocean-only model driven by the *National Centers for Atmospheric Prediction* (NCEP) reanalysis [*Kalnay et al.*, 1996; *Müller et al.*, 2012]. For Baseline-1 (B1) and Prototype (PT) thermodynamical self-consistent descriptions of the ocean state according to ORA-S4 from the European Centre for Medium-Range Weather Forecasts (ECMWF) [*Balmaseda et al.*, 2013] have been applied. Additionally, the atmospheric component is initialized with 3-D temperature, 3-D vorticity, 3-D divergence, and surface pressure fields from the ERA40 reanalysis [*Uppala et al.*, 2005]. The main difference is that B1 has been initialized with anomaly fields in the ocean, while PT uses a full-field initialization [*Kruschke et al.*, 2015] (compare Table 1). *Pohlmann et al.* [2013] demonstrated (with B1) that there is only little additional skill gain from using the higher resolution MPI-ESM-MR version of the MPI-ESM; therefore, we limit our analysis to results for the MPI-ESM-LR.

In contrast to the quite complex 3-D data assimilation procedures, necessary for the more recent MPI forecast systems, Modini-MPI-ESM uses a simpler approach: The ocean/sea ice component of the MPI-ESM is forced by the time series of observed wind stress anomalies added to the wind stress climatology from MPI-ESM. The atmosphere-ocean coupling of all other exchanged variables remains identical to that in the fully coupled MPI-ESM during the initialization [*Thoma et al.*, 2015]. This allows the atmosphere to respond to the ocean sea surface temperature and sea ice in a self-consistent way rather than being initialized by an external forcing field. This method is similar to that introduced by *Cane et al.* [1986] and, later, modified by *Chen et al.* [1997] but here applied to a fully coupled climate model. The numerical experiments have three periods: The preinitialization period before 1980 consists of the three original historical CMIP5 experiments performed with the fully coupled MPI-ESM-LR. During the initialization period the ocean and sea ice components of MPI-ESM-LR are forced with observed (reanalyzed) wind stress anomalies estimated from NCEPcfsr [*Saha et al.*, 2010] using

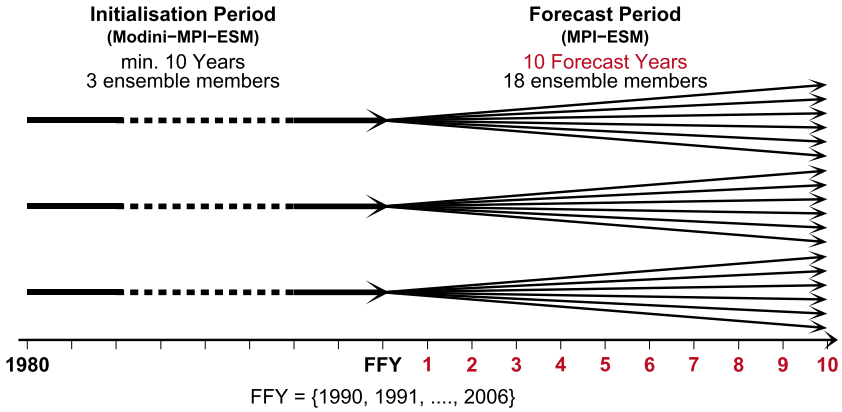

Figure 1. Sketch illustrating the Modini forecast system: During the initialization period three ensemble members are forced. These serve as seeds for the 18 ensemble members for each forecast period, starting at the individual First Forecast Years (FFYs).

bulk formulas according to *Large and Yeager* [2009] to convert wind velocity into wind stress. After a minimum of 10years (from 1990 onward) of wind stress anomaly forcing the Modini mode is switched off on six consecutive days starting with 1 January of the *First Forecast Year* (FFY). These different switch-off dates generate six ensemble members per original CMIP5 experiment, resulting in 18 ensemble members (Figure 1). With 2006 being the last FFY, we generate 17 historical forecasts, which are evaluated in the following section. It is known that a larger ensemble size increases the significant prediction skill of a forecast system. We present results for the highest number of ensemble members available in this study (section 3). Random sampling of the ensemble members to generate ensembles of different size has shown that the main differences in predictive skill between the different forecast systems can be attributed to the initialization method rather than the ensemble member size (not shown).

3. Results

The skill of the different initialization methods is assessed by correlating the *surface air temperature* (SAT) against observations from HadCRUT4 median [*Morice et al.*, 2012]. We estimate the correlation and the corresponding significance using the *Murphy-Epstein decomposition and Continuous Ranked Probability Skill Score* (MurCSS) tool, which analyzes decadal hindcast experiments in a deterministic and probabilistic way following and extending the framework suggested by *Goddard et al.* [2013]. This tool has been developed as part of the MiKlip evaluation system to improve the comparability within the project [*Illing et al.*, 2014; *Kadow et al.*, 2015] and has already been applied in *Pohlmann et al.* [2013]. The significance is tested using a 500-fold nonparametric bootstrap approach, taking account of autocorrelation. The skill assessment implies spatial averaging on a $5 \times 5^{\circ}$ grid, temporal aggregation and lead time-dependent bias adjustment including an implicit drift reduction [*International CLIVAR Project Office*, 2011]. In particular, we focus on the accuracy of the hindcasts using the anomaly correlation and the *Mean Squared Error Skill Score* (MSESS) between the initialized forecasts and the observations as skill for the average over the 2–5 and 6–9 year periods for FFYs from 1990 to 2006 (Figure 2). The MSESS represents the improvement in the accuracy of a hindcast over the climatology with respect to the observations. It takes the *anomaly correlation* and the *conditional bias* into account and ranges from $-\infty$ (incredibly bad) to 1 (perfect hindcast). The anomaly correlation represents the potential skill of a prediction system and measures the (linear) dependency between the modeled and observed values. It ranges from −1 (reversed skill) over 0 (no correlation) to 1 (great potential skill).

Note that already the uninitialized CMIP5 experiment (Figure 2, first row) has a decent skill, in particular, for the forecast years 2–5. This indicates how important the climate background GHG forcing is for multiyear forecasts. The B0 and, in particular, the subsequent B1 and PT initializations show some improvements compared to uninitialized CMIP5 forecasts for the first forecast year [*Pohlmann et al.*, 2013], but there is no significant skill gain for the years 2–5 or for the second pentade (years 6–9). We also do not see any obvious skill gain of PT compared to B1 for any forecast period (2–5 or 6–9) in the analyzed time frame. (In addition to the full-field maps shown in Figure 2, we present difference maps between individual prediction systems in Figures S1 and S2 in the supporting information).

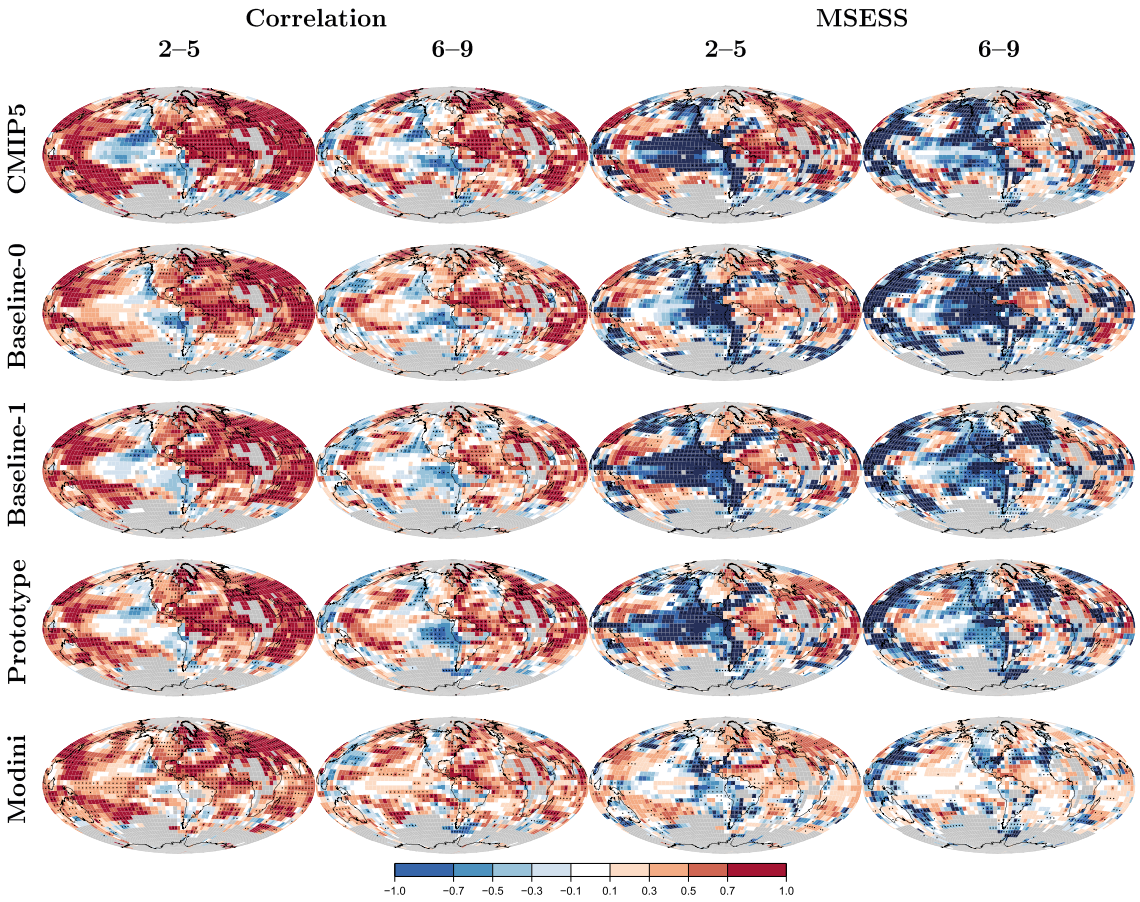

Figure 2. Maps of (first and second columns) anomaly correlation and (third and fourth columns) Mean Squared Error Skill Score (MSESS) of ensemble mean historical forecast skills of SAT calculated against HadCRUT4 median, averaged over the 2–5 and 6–9 year periods, respectively, for FFYs from 1990 to 2006. Black regions in MSESS maps indicate areas with a value below −1. Crosses denote values significantly different from zero exceeding at a 5% level applying 500 bootstraps. Gray shaded areas mark missing values with less than 90% data consistency in the observations.

The Modini forecast system gives a similar picture for the first forecast year (1–1, Figures 2, S3, and S4): The skill gain is, in general, somewhat between B0 and B1 (or the nearly identical PT); only over the Atlantic Ocean do the MPI forecast systems perform better during this first year. However, on the 2–5 year range, only Modini is able to generate (correlation) skill in the tropical Pacific, the key region for ENSO, and many teleconnected climate signals. Despite the simpler initialization method, the skill in the Atlantic Ocean is comparable to the MPI forecast systems. Only in the Indian Ocean do the MPI forecast systems have a higher correlation. Considering the MSESS, which also takes the conditional bias into account, Modini has less areas with nonexisting (negative) skill, compared to the MPI forecast system (right part of Figures 2, S1, and S2). However, compared to climatology, there is still room for improvement in all cases, including Modini. The Modini advantages persist (although to a lesser degree) even until forecast years 6–9.

Since about 1998 the globally averaged surface temperature has increased relatively little compared to the long-term trend. This period is often referred to as *global warming hiatus* [*Kosaka and Xie*, 2013]. Compared to the MPI forecast systems, Modini is better able to capture this hiatus in hindcast mode. Figure 3 shows that the temperature anomalies estimated with Modini are much closer to observations than those of B0 or PT, which feature a strong warming during the 2–5 and 6–9 year hindcast periods.

4. Discussion and Conclusion

We have compared the hindcast performance of the MPI-ESM (LR version) using different initialization schemes; that is, the B0, B1, and PT schemes of the German MiKlip project [*Pohlmann et al.*, 2013; *Kruschke et al.*, 2015] and what we have called Modini-MPI-ESM. For the latter, initial states are created by running the MPI-ESM using a time series of observed wind stress anomalies applied only to the ocean/sea ice component

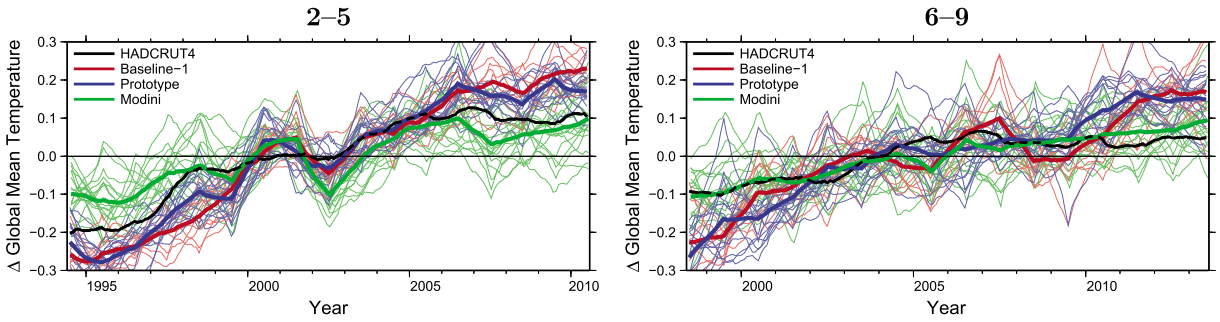

Figure 3. Time series (4 year running means) of global mean temperature anomalies for the hindcast years (left) 2–5 and (right) 6–9 for FFYs ranging from 1990 to 2006. Anomalies are estimated with respect to the respective mean value of each individual graph. Green (red, blue) thick line indicates ensemble mean result for *Modini* (Baseline-1, Prototype). Thin lines show individual ensemble members. Black line indicates 4 year running mean observations as reference according to HadCRUT4 median. Note that the value assigned to the abscissa corresponds to the center of the hindcast period.

of MPI-ESM. The atmospheric component of the MPI-ESM responds, in turn, to the sea surface temperature (SST) and sea ice given to it by the ocean component of the MPI-ESM using the thermodynamic coupling of the fully coupled model. Despite its simplicity, initialization using Modini-MPI-ESM leads to a better hindcast performance for both SAT and SST (not shown) in the Pacific sector than any of the other schemes, especially after the first year. We attribute this improvement to the fact that in Modini-MPI-ESM, the fully coupled model is initialized, rather than the oceanic and atmospheric components being initialized separately, as in the B0, B1, and PT initializations. Furthermore, using a long time series of observed wind stress for the initialization ensures that the ocean component of the MPI-ESM is dynamically balanced with the wind stress used for the initialization, an important consideration in the equatorial oceans [*Chen et al.*, 1997; *Bell et al.*, 2004; *Luo et al.*, 2005]. Here we used a spin-up time of at least 10years. This enables the tropical Pacific Ocean to adjust to wind forcing before commencing the hindcasts. It should be noted that the performance of the B0 and B1 is very similar to that shown in *Pohlmann et al.* [2013], even though the hindcasts were carried out over a different time period (1961–2012 compared to 1990–2006 here).

A weak point of the hindcasts initialized using Modini-MPI-ESM is the generally weaker performance in the Atlantic sector compared to the other schemes in the first forecast year (1–1, Figure 2), although after year 1, the performance of Modini-MPI-ESM is no worse in this sector than the other cases. We attribute this to the fact that wind stress alone is used for the initialization in Modini-MPI-ESM. It is well known that variability in the Atlantic on time scales of decadal and longer (for example, associated with the Atlantic Multidecadal Variability) is strongly influenced by surface heat flux forcing [e.g., *Eden and Jung*, 2001], which is missing in Modini-MPI-ESM. On the other hand, the ability of Modini-MPI-ESM to capture decadal variability in the Pacific sector has been shown by *Thoma et al.* [2015] in initialization mode and by *Ding et al.* [2013, 2014] using the Modini approach applied to the Kiel Climate Model. Indeed, that initialization using the Modini approach can lead to hindcast skill out to decadal time scales in the Pacific sector has already been anticipated by *Ding et al.* [2013]. A related feature of Modini is its much better performance at capturing, in hindcast mode, the hiatus

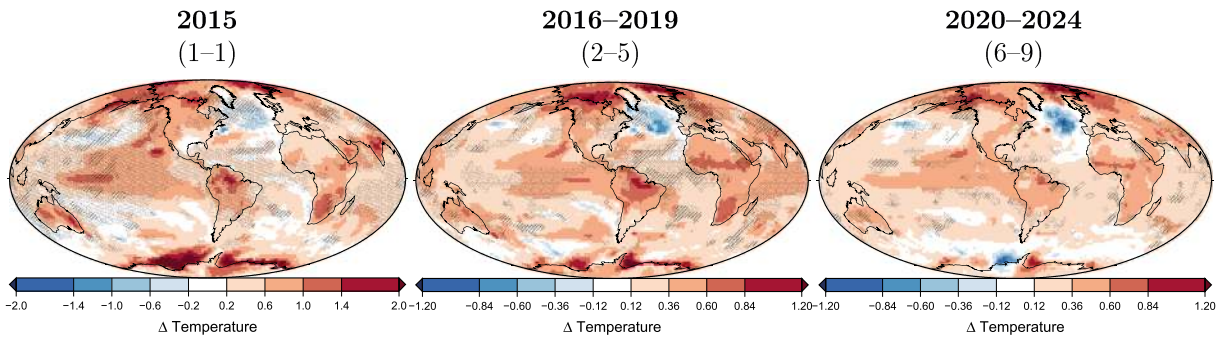

Figure 4. Predicted average SAT change for (left) year 2015, (middle) years 2016–2019, and (right) years 2016–2019 with respect to the Modini-MPI-ESM climatology from 1990 to 2006, including an inherent model drift correction. This forecast has been initialized with Modini-MPI-ESM and started with FFY 2015. Dotted (striped) areas indicate regions of significance in positive correlation (MSESS).

in global warming than the MPI forecast systems. We attribute this success to the improved performance of Modini in the Pacific sector compared to the other systems, while noting the importance of the Pacific sector in the dynamics of the hiatus as argued by *Kosaka and Xie* [2013] and *England et al.* [2014].

Finally, we present a prediction for 2015 (corresponding to the first forecast year) and averaged over the years 2016–2019 (corresponding to forecast years 2–5) and 2020–2024 (years 6–9) based on a Modini-MPI-ESM initialization until January 2015. The colors in Figure 4 indicate temperature anomalies with respect to the forecast year-dependent climatology from the Modini initialized hindcasts with FFY 1990–2006, i.e., warmer or colder than the period 1991–2006 (1992–2010 and 1996–2014) for the forecast of 2015 (2016–2019 and 2020–2024). This method implicitly corrects for an inherent forecast year-dependent model drift, although the 16 years are shorter than the 30 years usually used to compute such a climatology.

The shaded areas in Figure 4 indicate areas where significance is estimated in correlation and/or MSESS (and the corresponding values are positive), from the previous hindcasts according to the results described in section 3. In general, we predict the global surface air temperature in 2015 to be 0.53K warmer than the 1990–2006 mean which would make 2015 the warmest year ever in the instrumental record, exceeding the previous record in 2014. However, because of the short Modini climatology from 1990 to 2006, the model drift might not be completely compensated in this forecast, and the real (to be measured) warming might be slightly different in 2015. The warming is predicted to be most pronounced in high latitudes, but also over most parts of the continents. Note that the warm tropical Pacific prediction for 2015 is consistent with statements from the NOAA Climate Prediction Center who published the following synopsis on 5 March 2015: There is an approximately 50–60% chance that El Niño conditions will continue through Northern Hemisphere summer 2015 (http://www.cpc.noaa.gov/products/analysis_monitoring/enso_advisory/ensodisc.html), and also with the ENSO Tracker of the Australian Bureau of Meteorology, who raised the El Niño status on 12 May 2015 (product code IDCKGEWW00, http://www.bom.gov.au/climate/enso/tracker/\#tabs=History). As is usually the case when the tropical Pacific is warm, colder than normal SATs are predicted in the North Pacific, associated with a deepened Aleutian low (Figure 4). These colder than normal SSTs in the North Pacific persist into years 2–5 and 6–9 of the forecast, with a pattern resembling the Pacific Decadal Oscillation (PDO). Given the importance of the PDO for influencing global mean surface air temperature [*Kosaka and Xie*, 2013; *England et al.*, 2014], it is no surprise that our forecast suggests that the current hiatus in the rise of global mean surface air temperature will come to an end, leading to accelerated warming compared to the past decade. Indeed, according to our prediction, the global mean temperature from 2016 to 2019 (2020–2024) will be 0.35K (0.28K) warmer than the 2–5 (6–9) year forecast mean for the 1990–2006 FFY hindcasts (cf. Figure 3).

The intensified cooling in the North Atlantic to the west of Europe from 2015 to years 2016–2019 is a consequence of a weakening of the Atlantic Meridional Overturning Circulation (AMOC, not shown). Although we have noted the weakness of Modini-MPI-ESM for capturing decadal and longer time scale variability in the Atlantic sector, Modini-MPI-ESM does have skill at capturing variability associated with wind forcing of the AMOC [*Thoma et al.*, 2015]. Furthermore, we have seen that greenhouse gas forcing is important for changes in SAT in years 2–5 and 6–9 of the hindcast experiments, suggesting that anthropogenic forcing may also be a factor in the predicted weakening of the AMOC and the associated North Atlantic cooling [*Rahmstorf et al.*, 2015].

Baseline-1 participates in a multimodel assessment for decadal climate predictions [*Smith et al.*, 2013]. Within this context forecasts for 2015 and the 2015–2019 average are available online (http://www.metoffice.gov.uk/research/climate/seasonal-to-decadal/long-range/decadal-multimodel). We have added the Baseline-1 forecast and the multimodel average (MMA) forecast for temperature (five models) to the supporting information (Figure S6) for comparison with Figure 4. Although these forecasts are based on a 1971–2000 climatology, a qualitative assessment is possible: A common feature of the Modini, B1, and the MMA forecast is a warmer than average Northern Hemisphere for 2015, a trend that continues until 2019. The results for the Southern Hemisphere are not as consistent. Here only Modini shows a spatially consistent warmer-than-average surface air temperature for the Southern Ocean and Antarctica, while there are contradicting results between B1 and MMA for 2015. However, on pentadal timescales (until 2019) all models agree on a general warming of the Southern Hemisphere as well, although qualitative differences in the models for the Southern Ocean remain. On a regional scale the features of all models differ quite a lot, which indicates that further studies are needed until reliable decadal predictions can be used as basis for decision making on regional scales.

Acronyms

AMOC	Atlantic Meridional Overturning Circulation
B0	Baseline-0
B1	Baseline-1
PT	Prototype
CMIP5	Coupled Model Intercomparison Project Phase 5
CRUTEM4	Climatic Research Unit Temperature data set, version 4
ECHAM	Acronym from ECMWF and Hamburg
ECMWF	European Centre for Medium-Range Weather Forecasts
ENSO	El Niño Southern Oscillation
EOF	empirical orthogonal function
ERA40	ECMWF-40 Year Re-analysis
ERAI	ERA-Interim reanalysis
FFY	First Forecast Year
GHG	greenhouse gas
GPHA	Geopotential Height Anomaly
HadSST	Hadley Centre SST data set
HadISST	Met Office Hadley Centre's sea ice and sea surface temperature data set
HadCRUT4	Blended data set from the CRUTEM4 surface air temperature data set and the HadSST3 sea surface temperature data set
KCM	Kiel Climate Model
IPCC	Intergovernmental Panel on Climate Change
LR	Low resolution: Atmospheric resolution: T63L47, default; Ocean-Sea-Ice resolution GR15L40, default ≈1.5°
MiKlip	Decadal Climate Predictions (Mittelfristige Klimaprognosen)
Modini	Model initialization by partially coupled spin-up
MPI-ESM	Max Planck Institute Earth System Model
MPIOM	Max Planck Institute Ocean Model
MurCSS	Murphy-Epstein decomposition and Continuous Ranked Probability Skill Score
MR	Mixed resolution: Atmospheric resolution: T63L95, highly resolved middle atmosphere; Ocean-Sea-Ice resolution GR04L40, eddy permitting ≈0.4°
MSESS	Mean Squared Error Skill Score
NCEP	National Centers for Atmospheric Prediction
NCEPcfsr	National Center for Environmental Prediction, Climate Forecast System Reanalysis
ORA-S4	Ocean Reanalysis System 4
PDO	Pacific Decadal Oscillation
RCP	Representative Concentration Pathway
SAT	surface air temperature
SST	sea surface temperature

Acknowledgments

This work was funded by the BMBF MiKlip (Modini and Integration) projects under grants 01LP1134A&B and 01LP1160A, which is gratefully acknowledged. We thank all the people of the Max Planck Institute for Meteorology (MPI) and the German Climate Computing Center (DKRZ) who worked on the development of MPI-ESM used in the historical CMIP5 and decadal prediction experiments. The simulations have been performed at the German High Performance Computing Centre for Climate and Earth System Research (DKRZ). All data are stored at the DKRZ on long-term archives and can be made accessible on request. We thank the Editor, two anonymous reviewers, and Holger Pohlmann for suggestions which improved the manuscript. M.T. thanks Klaus Grosfeld and all members of the sea ice group at AWI for their support and fruitful discussions. R.J.G. is grateful for continuing support from GEOMAR. C.K. likes to thank Sebastian Illing.

The Editor thanks two anonymous reviewers for their assistance in evaluating this paper.

References

Balmaseda, M. A., K. Mogensen, and A. T. Weaver (2013), Evaluation of the ECMWF ocean reanalysis system ORAS4, *Q. J. R. Meteorol. Soc.*, *139*(674), 1132–1161, doi:10.1002/qj.2063.

Bell, M. J., M. J. Martin, and N. K. Nichols (2004), Assimilation of data into an ocean model with systematic errors near the equator, *Q. J. R. Meteorol. Soc.*, *130*(598), 873–893, doi:10.1256/qj.02.109.

Cane, M. A., S. E. Zebiak, and S. C. Dolan (1986), Experimental forecasts of El Niño, *Nature*, *321*(6073), 827–832, doi:10.1038/321827a0.

Chen, D., S. E. Zebiak, M. A. Cane, and A. J. Busalacchi (1997), Initialization and predictability of a coupled ENSO forecast model, *Mon. Weather Rev.*, *125*(5), 773–788, doi:10.1175/1520-0493(1997)125<0773:IAPOAC>2.0.CO;2.

Cox, P., and D. Stephenson (2007), A changing climate for prediction, *Science*, *317*(5835), 207–208, doi:10.1126/science.1145956.

Ding, H., R. J. Greatbatch, M. Latif, W. Park, and R. Gerdes (2013), Hindcast of the 1976/77 and 1998/99 climate shifts in the Pacific, *J. Clim.*, *26*(19), 7650–7661, doi:10.1175/JCLI-D-12-00626.1.

Ding, H., R. J. Greatbatch, W. Park, M. Latif, V. A. Semenov, and X. Sun (2014), The variability of the East Asian summer monsoon and its relationship to ENSO in a partially coupled climate model, *Clim. Dynam.*, *42*(1–2), 367–379, doi:10.1007/s00382-012-1642-3.

Eden, C., and T. Jung (2001), North Atlantic interdecadal variability: Oceanic response to the North Atlantic Oscillation (1865–1997), *J. Clim.*, *14*(5), 676–691, doi:10.1175/1520-0442(2001)014<0676:NAIVOR>2.0.CO;2.

England, M. H., S. McGregor, P. Spence, G. A. Meehl, A. Timmermann, W. Cai, A. S. Gupta, M. J. McPhaden, A. Purich, and A. Santoso (2014), Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus, *Nat. Clim. Change*, *4*(3), 222–227.

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.

Hawkins, E., and R. Sutton (2009), The potential to narrow uncertainty in regional climate predictions, *Bull. Am. Meteorol. Soc.*, *90*(8), 1095–1107, doi:10.1175/2009BAMS2607.1.

Houghton, J. T., G. J. Jenkins, and J. J. Ephraums (Eds.) (1990), *Climate Change: The IPCC Scientific Assessment*, 1–410 pp., Cambridge Univ. Press, New York.

International CLIVAR Project Office (2011), Decadal and bias correction for decadal climate predictions, *CLIVAR Publ. Ser., 150*, 6 pp.

Illing, S., C. Kadow, K. Oliver, and U. Cubasch (2014), MurCSS: A tool for standardized evaluation of decadal hindcast systems, *J. Open Res. Software*, *2*(1), e24, doi:10.5334/jors.bf.

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 the Max Planck Institute Ocean Model (MPIOM) the ocean component of the MPI-Earth system model, *J. Adv. Model. Earth Syst.*, *5*(2), 422–446, doi:10.1002/jame.20023.

Kadow, C., S. Illing, O. Kunst, H. Rust, H. Pohlmann, W. A. Müller, and U. Cubasch (2015), Evaluation of forecasts by accuracy and spread in the MiKlip decadal climate prediction system, *Meteorol. Z.*, 1–13, doi:10.1127/metz/2015/0639.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kosaka, Y., and S. -P. Xie (2013), Recent global-warming hiatus tied to equatorial Pacific surface cooling, *Nature*, *501*(7467), 403–407.

Kruschke, T., H. W. Rust, C. Kadow, W. A. Müller, H. Pohlmann, G. C. Leckebusch, and U. Ulbrich (2015), Probabilistic evaluation of decadal prediction skill regarding Northern Hemisphere winter storms, *Meteorol. Z.*, 1–18, doi:10.1127/metz/2015/0641.

Kröger, J., W. A. Müller, and J. -S. von Storch (2012), Impact of different ocean reanalyses on decadal climate prediction, *Clim. Dynam.*, *39*(3–4), 795–810, doi:10.1007/s00382-012-1310-7.

Large, W. G., and S. G. Yeager (2009), The global climatology of an interannually varying air-sea flux data set, *Clim. Dynam.*, *33*, 341–364, doi:10.1007/s00382-008-0441-3.

Luo, J. -J., S. Masson, S. Behera, S. Shingu, and T. Yamagata (2005), Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecasts, *J. Clim.*, *18*(21), 4474–4497, doi:10.1175/JCLI3526.1.

Morice, C. P., J. J. Kennedy, N. A. Rayner, and P. D. Jones (2012), Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set, *J. Geophys. Res.*, *117*, D08101, doi:10.1029/2011JD017187.

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*(22), L22707, doi:10.1029/2012GL053326.

Park, W., N. Keenlyside, M. Latif, A. Ströh, R. Redler, E. Roeckner, and G. Madec (2009), Tropical Pacific climate and its response to global warming in the Kiel Climate Model, *J. Clim.*, *22*(1), 71–92, doi:10.1175/2008JCLI2261.1.

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.

Rahmstorf, S., J. E. Box, G. Feulner, M. E. Mann, A. Robinson, S. Rutherford, and E. J. Schaffernicht (2015), Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation, *Nat. Clim. Change*, *5*(5), 475–480, doi:10.1038/nclimate2554.

Saha, S., et al. (2010), The NCEP climate forecast system reanalysis, *Bull. Am. Meteorol. Soc.*, *91*(8), 1015–1057, doi:10.1175/2010BAMS3001.1.

Smith, D. M., et al. (2013), Real-time multi-model decadal climate predictions, *Clim. Dyn.*, *41*(11–12), 2875–2888, doi:10.1007/s00382-012-1600-0.

Stevens, B., et al. (2013), Atmospheric component of the MPI-M Earth System Model: ECHAM6, *J. Adv. Model. Earth Syst.*, *5*(2), 146–172, doi:10.1002/jame.20015.

Stocker, T. F., D. Qin, G. -K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P. Midgley (Eds.) (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, pp. 1–1535, Cambridge Univ. Press, Cambridge, U. K.

Thoma, M., R. Gerdes, R. J. Greatbatch, and H. Ding (2015), Partially coupled spin-up of the MPI-ESM: Implementation and first results, *Geosci. Model Dev.*, *8*(1), 51–68, doi:10.5194/gmd-8-51-2015.

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.

Wang, C., L. Zhang, S. -K. Lee, L. Wu, and C. R. Mechoso (2014), A global perspective on CMIP5 climate model biases, *Nat. Clim. Change*, *4*(3), 201–205.