# The German Climate Forecast System: GCFS

**Kristina Fröhlich[1], Mikhail Dobrynin[2,1], Katharina Isensee[1], Claudia Gessner[4], Andreas Paxian[1], Holger Pohlmann[3,1], Helmuth Haak[3], Sebastian Brune[2], Barbara Früh[1], Johanna Baehr[2]**

[1]Deutscher Wetterdienst, Offenbach, Germany
[2]CEN, Universität Hamburg, Hamburg, Germany
[3]Max Planck Institute for Meteorology, Hamburg, Germany
[4]ETH, Zurich, Switzerland

**Key Points:**

- Seasonal forecasts
- Earth system model development

Corresponding author: Kristina Fröhlich, kristina.froehlich@dwd.de

## Abstract

Seasonal prediction is one important element in a seamless prediction chain between weather forecast and climate projections. After several years of common development in collaboration with Universität Hamburg and Max Planck Institute for Meteorology, the Deutscher Wetterdienst performs operational seasonal forecasts since 2016 with the German Climate Forecast System, now in Version 2 (GCFS2.0). Here, the configuration of previous system GCFS1.0 and the current GCFS2.0 are described and the performance of the two systems is compared over the common hindcast period of 1990-2014. In GCFS2.0, the forecast skill is improved compared to GCFS1.0 during boreal winter, especially for the Northern Hemisphere where the Pearson correlation has doubled for the North Atlantic Oscillation index. During boreal summer, overall a similar performance of GCFS2.0 in comparison to GCFS1.0 is assessed. Future developments for climate forecasts need a stronger focus on the performance of seasonal dependent processes in a model system.



## Plain Language Summary

An information about the expected departure from the climate of an upcoming season would be tremendously valuable for many sectors of the society. In Germany, three institutes join their expertise to build a forecast system for a climate prediction several months ahead. We use the Earth system model of the Max Planck Institute for Meteorology. It describes the atmosphere, land and rivers as well as the ocean and the sea ice in their interaction. This model is well designed for climate studies on a much longer timescale than a season. Universität Hamburg and the National Meteorological Service Deutscher Wetterdienst develop the methods that are necessary for such a forecast system and operationally perform the seasonal predictions. Here we compare two versions of our forecast system. We investigate especially the forecast quality during different seasons. The expectation, that the second model system is much better than the first system is not entirely fulfilled. We discuss possible reasons and suggest for future model development a stronger focus on the model quality for seasonal varying processes.


## 1 Introduction

Over the last two decades, seasonal climate prediction evolved from a scientific research topic into full-fledged operational systems. Today, seasonal prediction systems are mostly operationally run at large weather centres like ECMWF (European Centre for

Medium-Range Weather Forecasts) [*Johnson et al.*, 2019] or the British Met Office [*MacLachlan et al.*, 2015], issuing real-time seasonal forecasts. The number of global producing centres (GPC's) submitting long-range forecasts to the World Meteorological Organization WMO lead centre in South Korea currently counts 13 GPC's. Since 2011, the German National Meteorological Service Deutscher Wetterdienst (DWD), Universität Hamburg (UHH) and the Max Planck Institute for Meteorology (MPI-M) have been developing a system for seasonal climate prediction [*Baehr et al.*, 2015]. Since October 2016, the German Climate Forecast System (GCFS) operationally produces seasonal predictions, which are published every month at DWDs homepage. In 2017 DWD became the 13th GPC of the WMO's multi-model ensemble for long-range forecasts.

Predicting seasonal climate aims for a time horizon of more than one month up to one year, which is considerably longer than short-term weather prediction (1 to 10 days) and considerably shorter than climate projections (decades to centuries). As such, seasonal prediction systems lie between the initial value problem and the boundary value problem. State-of-the-art seasonal prediction systems are thus build on global coupled Earth system models (ESM) which are often also used for climate projections [*Ban et al.*, 2016]. In order to get the best estimate of the initial conditions for the seasonal forecasts, observations are assimilated into the ESM prior to the start of the predictions, a similar approach as for weather predictions.

Seasonal predictions are probabilistic climate forecasts. This means that the future guidance describes a probable deviation from a known climate [*White et al.*, 2017]. It implies model ensemble predictions for both, the chosen climate period (so called hindcasts) and the real-time forecasts. The hindcast data provide the basis for the assessment of the prediction skill in the past as well as the mean model climate. A reference climate is provided by a reanalysis data set.

The inherent uncertainty of seasonal predictions is adressed by the initialisation of an ensemble of predictions and the use of probabilistic evaluation methods. Furthermore, a multi-model ensemble, consisting of different prediction systems, may outperform any of the single ensemble models with respect to seasonal or longer term forecasts (e.g. *Stockdale* [2012], *Scaife et al.* [2014], *Athanisiadis and Bellucci* [2017]).

Any operational seasonal climate prediction system will try to skillfully predict variables, such as temperature, precipitation, wind and pressure, on a time scale of a few

months. These quantities may be related to large scale coupled physical processes such as El-Niño-Southern-Oscillation (ENSO) as well as the North Atlantic Oscillation (NAO), which impact the seasonal time scale.

Improving seasonal predictions bases on the understanding of predictable processes, the ability of the coupled model to represent them and a sufficient data base for statistical evaluation purposes. *Scaife et al.* [2014] showed an improved NAO prediction over its predecessor model for the boreal winter after elaborating on the ocean resolution, the representation of the stratosphere, and the sea ice physics in their model system. The current ECMWF system was able to improve the ENSO forecast quality considerably from their System4 to System5 [*Johnson et al.*, 2019]. Seasonal prediction skill can also be improved by a number of statistical and post processing methods, such as bias adjustment and recalibration [*Manzanas et al.*, 2019] or by using teleconnection patterns for ensemble subsampling [*Dobrynin et al.*, 2018].

Here, we describe the performance of first two German Climate Forecast Systems (GCFS) which are developed in cooperation between Max Planck Institute for Meteorology, Universität Hamburg and Deutscher Wetterdienst.

At the Max Planck Institute, the Earth system model (MPI-ESM, *Giorgetta et al.* [2013], *Mauritsen et al.* [2018]) is developed and tuned over many model-years under pre-industrial conditions (referring to climate conditions of 1850), until it is decided that a best match of the known state of the Earth's climate system is found while keeping the balance of the atmosphere's radiation [*Mauritsen et al.*, 2012]. This balanced model state serves as basis for any study on introduced anthropogenic changes within the following centuries. With increasing model resolution the goal of a well-adjusted model climate becomes a computational and time demanding challenge.

At Universität Hamburg, assimilation and ensemble generation methods are developed for the use in GCFS. Deutscher Wetterdienst adapts and maintains the whole system for operational performance and operationally issues the seasonal forecasts.

In the following, we will describe the configuration of the two versions GCFS1.0 and GCFS2.0, analyze both systems with respect to the representation and prediction skill of NAO, ENSO, surface temperature and geopotential height at 500 hPa and compare them with the focus on differences in the model physics and the respective hindcast skill.

We will discuss GCFS's strengths and weaknesses and possible approaches for the future development and conclude with our main findings.

## 2 Configuration of the operational systems

### 2.1 Model configuration

The first version of GCFS1.0 was based on MPI-ESM-LR [*Giorgetta et al.*, 2013], with an atmosphere resolution of T63 (corresponding to approximately 150 km at around 50° N) and 47 levels reaching up to 0.01 hPa and an ocean resolution of nominally 1.5° in the horizontal and 40 levels in the vertical down to approximately 5000 m [*Jungclaus et al.*, 2013]. The ocean model is connected to a dynamic and thermodynamic sea-ice model and the biogeochemical sub-model HAMOCC [*Ilyina et al.*, 2013] (the latter switched off in GCFS). Coupling between ocean and atmosphere was set to once a day for GCFS1.0. The land and vegetation model JSBACH also hosts a hydrological runoff model. In GCFS1.0, MPI-ESM uses the external forcing like greenhouse gases, ozone and aerosols based on Phase 5 of the Coupled Model Intercomparison Project CMIP5 [*Taylor et al.*, 2012] for historical data and future scenarios. The so-called historical period within CMIP5 ended in 2005, therefore all simulations of GCFS1.0 starting from 2006 onwards used the RCP4.5 scenario as the set up for the external data.

The version GCFS2.0 is based on the MPI-ESM-HR (*Müller et al.* [2018] and *Mauritsen et al.* [2018]) with a T127 spectral resolution in the atmosphere corresponding to approximately 70 km at around 50° N, with 95 levels covering the same vertical column up to 0.01 hPa and an ocean resolution of nominally 0.4° in the horizontal and, similar to GCFS1.0, 40 levels in the vertical. Coupling between ocean and atmosphere in GCFS2.0 takes places on hourly basis. The land and vegetation model JSBACH in MPI-ESM-HR includes vertical soil moisture transport [*Hagemann and Stacke*, 2015]. For the low-resolution version MPI-ESM-LR it has been shown, that this contributes to an improvement of European summer temperatures [*Bunzel et al.*, 2017]. A detailed description of all model developments of the new and higher resolved version is provided by *Mauritsen et al.* [2018]. For GCFS2.0 we use external forcing from Phase 6 of the Coupled Model Intercomparison Project CMIP6 [*Eyring et al.*, 2016], where the historical period has been extended until 2014. Scenario data were not available before mid 2018, there-

fore the external forcing has been set constant starting in 2015 up to present time. Table 1 gives an overview on both systems.

**Table 1.** Overview of GCFS1.0 and GCFS2.0 configuration

| Component | Subcomponent | GCFS1.0 | GCFS2.0 |
|---|---|---|---|
| **Model** | Atmosphere/ECHAM | T63L47 | T127L95 |
| | Land surface vegatation /JSBACH | bulk soil moisture | 5 layer soil moisture |
| | Ocean/MPIOM | GR15 | TP04 |
| | Coupler / OASIS3 MCT | once a day | hourly |
| **External data** | historical | CMIP5 1981-2005 | CMIP6 1981- 2014 |
| | scenarios | CMIP5 RCP 4.5 from 2006 onwards | CMIP6 constant 2014 values from 2015 onwards |
| **Ensemble generation** | Atmosphere | Perturbation of uppermost atmospheric layer | Perturbation of uppermost atmospheric layer |
| | Ocean | Bred vectors over 12 months | Bred vectors over 6 months |
| **Assimilation** | Atmosphere | nudging of ERA-Interim variable fields | nudging of ERA-Interim variable fields |
| | Ocean | nudging ORAS4 and NSIDC variable fields | nudging of ORAS5 variable fields |
| **Hindcast** | Period | 1981-2014 | 1990-2017 |
| | Forecast duration | 12 months | 6 months |
| | Ensemble member | 15 | 30 |
| **Forecast** | Forecast duration | 12 months | 6 months |
| | Ensemble member | 30 | 50 |

In order to give an idea of the different model behaviour purely due to the changed model configuration and physics, Figure 1 shows biases of 2m temperature of the two climate model versions CMIP5 MPI-ESM-LR and CMIP6 MPI-ESM-HR with respect to ERA-Interim for the time range 1990-2014. Displayed are ensemble means from the 10 member ensemble of the respective historical experiments (in total running from 1850-2014) and splitted up into the two seasons of December/January/February DJF in Fig. 1 (top) and June/July/August JJA Fig. 1 (bottom) months.

The striking feature of the CMIP6 version for both seasons are the much warmer tropical oceans compared to the reanalysis and the reduced cold tongue in the tropical Pacific in Fig. 1 (b,d). Although this error pattern has not disappeared completely, it is much smaller for the DJF months than for the JJA months. Further, the warm bias of the Amazon basin is considerably reduced in the CMIP6 model version. Seasonal differences also show up, as for instance the reduced cold boreal summer bias in northern Russia and northern Africa in the new and higher resolved model system. The bias pattern over northern America also changes between the model versions and the two different seasons, but no clear error reductions are visible there. Europe exhibits a cold bias in the CMIP6 version for DJF and JJA, while the CMIP5 version produced a warm DJF bias. The European cold JJA bias is slightly reduced in the CMIP6 version. Although some of these biases can be adressed during data assimilation, most will re-emerge in the seasonal hindcasts and forecasts.

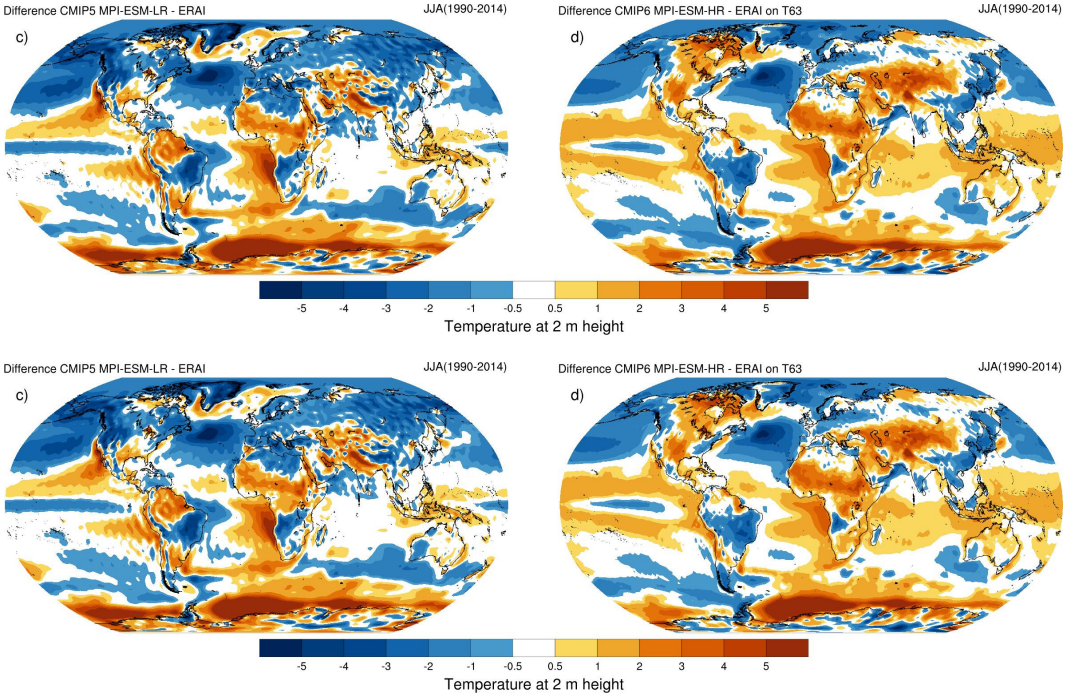



**Figure 1.** Difference of the 2 metre temperature between the historical experiment of CMIP5 MPI-ESM-LR (left) and CMIP6 MPI-ESM-HR (right) to ERA-Interim for DJF (top) and JJA (bottom) between 1990-2014.

### 2.2 Assimilation

In order to bring the model system close to the (re)analysed state of the climate system the assimilation method of a continuous nudging is used which produces the initial conditions for the retrospective and current forecasts. This method is the same as described in *Baehr et al.* [2015]. Assimilation for both versions starts in 1979.

The data requirements for initial conditions differ for hindcast and real-time forecast. Reanalysis (or other observational) data are used for the hindcasts while either near-real-time ocean analyses or real-time atmospheric analyses are taken for the forecast production. For the hindcast production both GCFS1.0 and GCFS2.0 use ERA-Interim [*Dee et al.*, 2011]. The model atmosphere is nudged towards vorticity, divergence, temperature and mean sea level pressure. For real-time forecasts initial conditions are taken from the analyses of the IFS weather-forecast model.

Ocean initial states for GCFS1.0 are provided by 3D ocean temperature and salinity of ORAS4 [*Mogensen et al.*, 2012] and sea ice concentration from the National Snow and Ice Data Center NSIDC [*Fetterer et al.*, 2002]. Also, the near-real-time (nrt) analysis system ORAS4 provided data for GCFS1.0 forecasts.

Pragmatical considerations for sea-ice data in terms of availability and consistency led to the choice of ORAS5 data [*Zou et al.*, 2017] for all oceanic variables in GCFS2.0. ORAS5 now also provides the near-real-time data for GCFS2.0 forecasts.

Additionally, GCFS2.0 makes use of a decadal assimilation with MPI-ESM-HR from the German decadal climate prediction project MiKlip [*Pohlmann et al.*, 2019]. Technical details are described in *Polkova et al.* [2019], where the authors used MPI-ESM-LR for their study. The decadal MPI-ESM-HR assimilation begins already in 1960. By using restart fields of this decadal assimilation GCFS2.0 starts in 1979 with an ocean and a land surface already adapted to permanently nudged atmospheric and oceanic conditions.

### 2.3 Ensemble generation

Ensemble members are generated by applying perturbations both to atmosphere and ocean. In the ocean the ensemble is initialised through bred vectors in all vertical levels [*Baehr and Piontek*, 2014]. The initial ensemble is generated by applying the lagged day initialisation. At the end of the first hindcast the ensemble members serve as perturba-

tions for the breeding cycle. After approximately two years the bred vectors have lost their memory of the initial perturbations [*Baehr et al.*, 2015].

For GCFS1.0 with the end of the forecast after one year, the generated ocean restart file of each of the 15 ensemble member provides the perturbation for the next start date. With the increase of GCFS2.0's ensemble size to 30 the duration of each single hindcast and forecast and therefore the breeding time is reduced to 6 months.

Within the atmosphere the perturbation of a physical parameter - the diffusion coefficient of the uppermost model layer - forces slightly different atmospheric conditions of the applied member. The perturbation value varies every fifth member in the hindcast ensemble and every tenth member in the forecast ensemble of GCFS2.0

### 2.4 Hindcasts

In GCFS1.0 the hindcast ensemble of 15 members started each month of the year beginning in 1981 up to 2015. The hindcast duration for each start date was 12 months.

The size of the GCFS2.0 ensemble has been increased from 15 to 30. The doubling of the ensemble size together with a higher resolution leads to a considerable increase in computational cost, therefore the period of hindcasts and the hindcast duration had to be shortened. The 30 members of GCFS2.0 hindcasts now start in 1990 and end in 2017 while performing 6 months of retrospective forecasts for each calendar month in all years. In this way more than 25 years of hindcasts are achieved.

The whole set of ensemble hindcast data for each calendar start month over many years (here 1990-2014) provides the basis for several purposes. First, for each of the 12 hindcast sets an ensemble and multi-year mean is derived and serves as model climate. Second, each of the 12 multi-year ensemble hindcast data sets can be assessed for its forecast quality in both, deterministic and probabilistic scores.

### 2.5 Forecasts

GCFS1.0 forecasts ran with 30 ensemble members started at the first of each month, with a forecast duration of one year.

With GCFS2.0 50 ensemble members are integrated over half a year, again all members starting on the first day of the month.

Figure 2 sketches the workflow of the seasonal forecast system GCFS2.0 from the generation of the initial conditions to the hindcast and forecast ensembles. This figure and a simple description of the system can also be found under `https://www.dwd.de/EN/ourservices/seasonals_forecasts/project_description.html`.

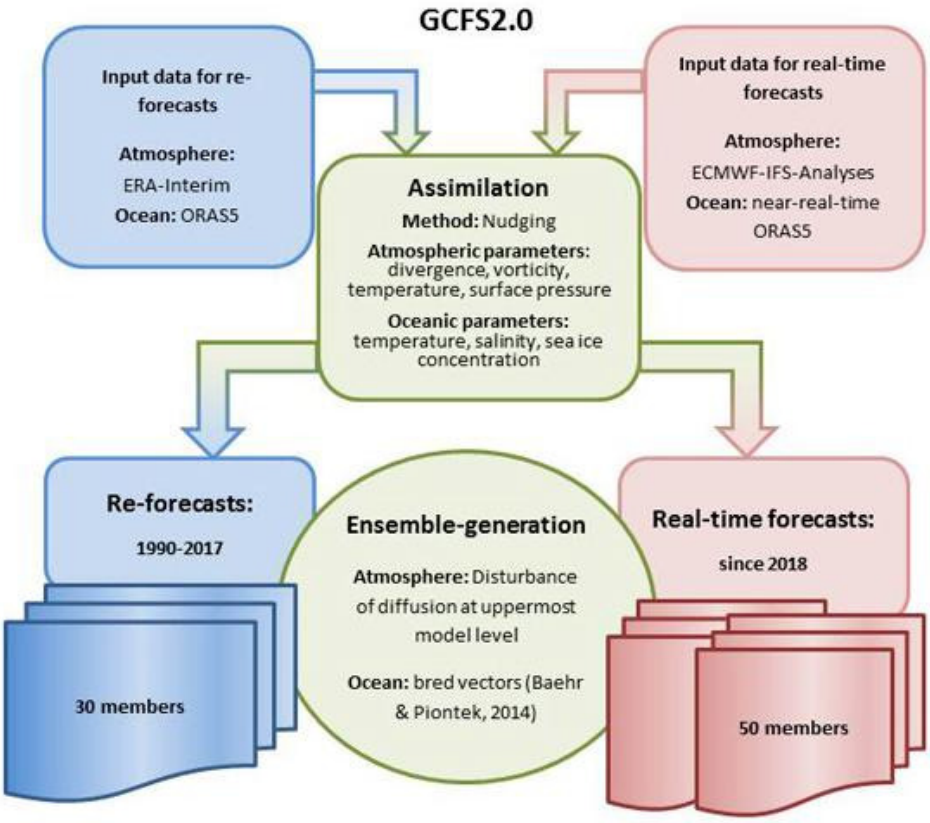



**Figure 2.** Workflow of GCFS2.0 seasonal forecasts.

### 2.6 Communication of seasonal predictions

Seasonal predictions are provided as anomaly forecasts, with respect to a defined model climatology. Ensemble mean anomalies, for instance, are created by subtracting the lead-time dependent ensemble-mean model-climate from the chosen reference hindcast period. A bias corrected forecast is obtained by adding the observational based climatology of the same reference period to each ensemble member-anomaly.

A probabilistic outlook checks, how many members cluster in a defined event category. For seasonal forecasts, commonly terciles out of the climatological reference period define the cold/normal/warm or dry/normal/wet events.

Seasonal forecasts are usually averaged over 3 months. The first month is discarded in most cases to account for an initialisation shock so that the first forecast period starts with the second to the fourth forecast month, and so on. Therefore, a DJF forecast is issued in November predicting anomalies for the upcoming December, January and February. Likewise, the May forecasts give the outlook for the months June, July and August.

To account for the low skill and huge uncertainty in seasonal predictions, information about the forecast quality is best provided along with the forecast itself. Uncertainty estimates are provided via probabilistic representations.

### 2.7 Evaluation metrics of seasonal hindcasts

A simple but instructive first measure is the difference between the reanalysis and each hindcast data set for a given variable. The reduction of differences between the simulations and a reference data set is usually considered as a success in a better description of climate processes which is expected to result in improved prediction skill.

To further evaluate the predictive skill of each forecast system, deterministic as well as probabilistic measures are used.

As a conventional deterministic score we apply the Pearson correlation or so-called anomaly correlation coefficient (ACC) to the ensemble mean hindcasts averaged over 3 months. This measure of linear correlation between the ensemble mean and the reanalysis varies between $-1$ (anti-correlation) and 1 (high correlation), where 0 means no correlation. Here, bias corrected hindcast data in cross-validation are used for the ACC. As such, the anomaly for each member is created by subtracting the model climatology where the corresponding member and year is excluded.

As probabilistic measures for this study we chose the fair ranked probability skill score RPSS [*Ferro*, 2014]. The RPSS is calculated over all defined events or categories (usually 3) of the ensemble hindcasts. While the ranked probability score RPS is the mean square error of probabilistic multi-category forecasts, the RPSS shows the improvement of using probabilistic forecasts versus using a climatological value. The score is called *fair* as

it is adapted to the finite ensemble size. The perfect score is 1, values below 0 denote that the climatology performes better than the respective hindcast ensemble. Further metrics to estimate the reliability, resolution and sharpness of the ensemble system are reliability diagrams, Brier scores or ROC curves or maps. These are as well calculated for GCFS but not shown here for lack of space. For more details on skill measures and scores we refer to *Wilks* [1995].

## 3 Comparison of GCFS1.0 and GCSF2.0 hindcasts

### 3.1 Comparison setup

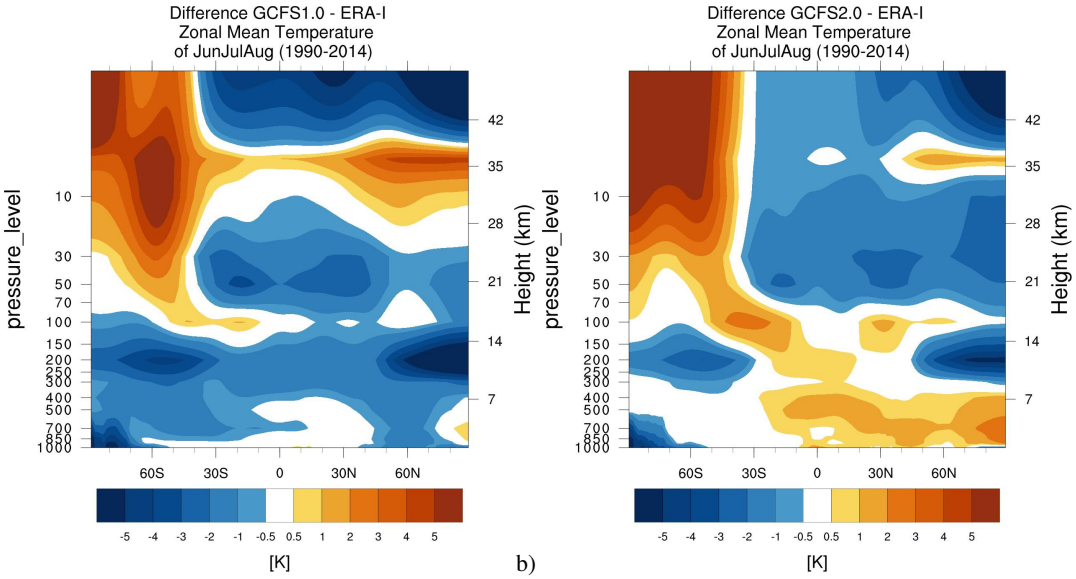
In order to have a common hindcast period of GCFS1.0 and GCFS2.0 the reference time of 1990 – 2014 is chosen. We focus our assessment on temperature and geopotential fields. Mean sea level pressure is used for the evaluation of the NAO prediction. The reference data set is the ERA-Interim reanalysis except for the evaluation of the Niño3.4 regions, where the NOAA Optimum Interpolation Sea Surface Temperature Version 2 NCEP OIv2 is used.

### 3.2 JJA mean state and hindcast skill

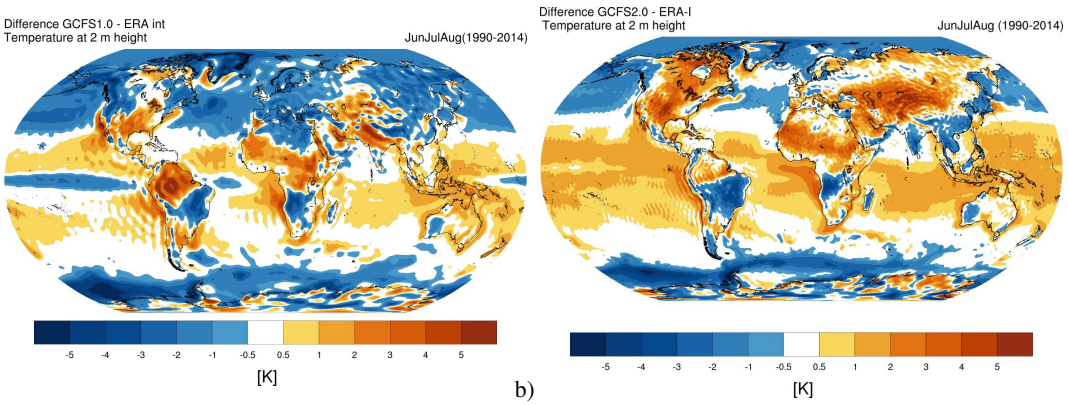
We present the mean state of the models by looking at the vertical structure of the atmosphere as well as at the surface and the level of 500 hPa. We consider the ensemble mean and time mean of the respective 3 months hindcasts over the time period of 1990-2014.

Figure 3 visualises in a latitude-height-plot the bias of temperature of GCFS1.0 (left) and GCFS2.0 (right) with respect to ERA-Interim. The vertical and the latitudinal temperature structure in the two panels is quite different. In stratospheric levels around/above 35 km a dipole feature of the boreal summer stratospheric temperature bias of GCFS1.0 (Fig. 3a) points to a wrong position of the ozone layer. This feature is weakened in GCFS2.0 (Fig. 3b). However, the middle atmosphere is characterized by a cold bias in the summer Northern Hemisphere and a strong warm bias in the winter Southern Hemisphere, pointing to difficulties in the description of the ozone layer and the non-orographic gravity wave parameterization.

In lower layers, the comparison shows, that the tropospheric cold bias in GCFS1.0 during the JJA months is reduced in GCFS2.0, revealing almost a bias-free Southern Hemi-




277 **Figure 3.** Temperature differences of *a*) GCFS1.0 ensemble mean and *b*) GCFS2.0 ensemble mean to
278 ERA-Interim with respect to JJA hindcasts during 1990-2014.




283 **Figure 4.** Difference of the 2m temperature forecasts a) GCFS1.0 and b) GCFS2.0 to ERA-Interim for JJA
284 hindcasts in 1990-2014.

296 sphere. In the Northern Hemisphere however, a considerable warm bias in the middle tro-
297 posphere extends from the northern polar latitudes into the tropics. The maximum of the
298 positive bias is placed above the North Pole while in the lowermost layers the cold bias
299 of the Arctic remains. The warm-bias Northern Hemisphere region is accompanied by a
300 positive humidity bias (not shown here).

301 JJA hindcast temperature biases at 2m height are displayed in Figure 4. The cold
302 bias of Northern Hemisphere land masses in GCSF1.0, which can also be seen in the 10
303 member ensemble of [*Baehr et al.*, 2015], is greatly reduced, but is in some regions re-
304 placed by a warm bias in GCFS2.0, as e.g. over North-America. The error in the Amazon

basin is strongly reduced in GCFS2.0. The cold tongue in the tropical Pacific, present in GCFS1.0, completely vanished thanks to the nudging but is replaced by a strong warm bias in the upwelling region west of South American and South African coasts.

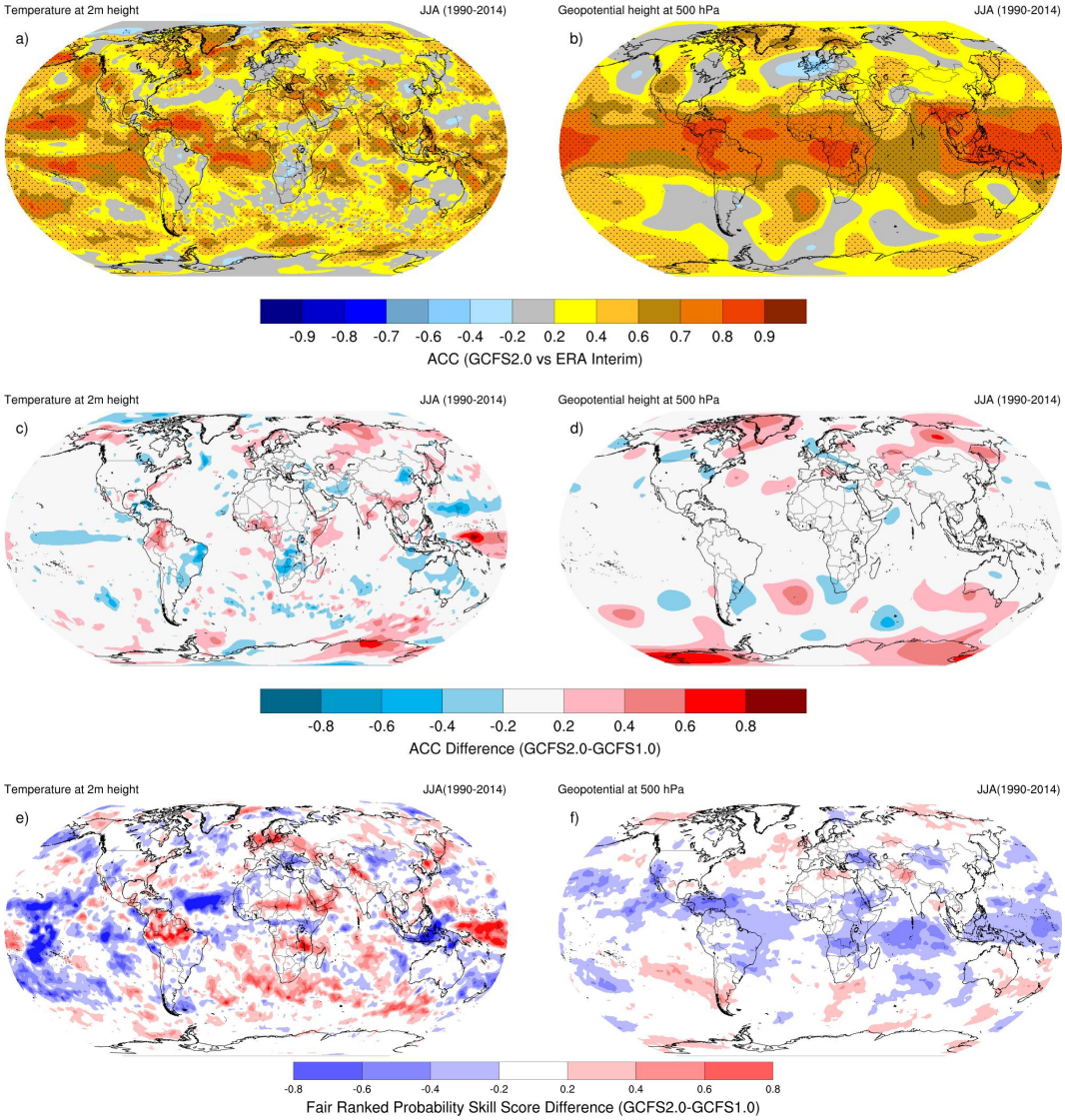


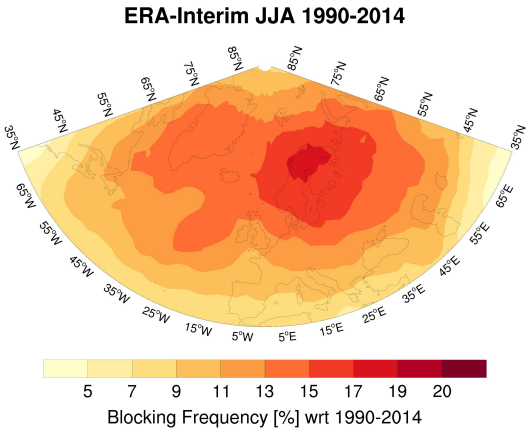

**Figure 5.** Top panel: Anomaly correlation coefficient (ACC) of hindcasts for June, July, August with respect to ERA-Interim during 1990 – 2014: a) 2m temperature and b) geopotential height at 500 hPa pressure level. Dots represent significant values at the 95% confidence level. Middle panels show the difference of ACC skill between GCFS1.0 and GCFS2.0 c) 2m temperature and d) geopotential height at 500 hPa. Bottom panels show the difference of GCFS1.0 and GCFS2.0 for the Fair Ranked Probability Skill Score RPSS for e) 2m temperature and f) geopotential height at 500 hPa.

The change of the bias patterns seems to lead to a neutral behaviour of the GCFS2.0 hindcast skill, the ACC. Figure 5a) shows, that for central Europe no hindcast skill in temperature can be expected, while the situation is better in North-America, the Mediterranean and some Asian regions, where significant correlation values appear, represented by dots if at the 95% confidence level. In the middle troposphere the ACC for the geopotential height (Figure 5b) shows negative values over Europe. Again, western US, Greenland and Central Asia exhibit significant positive skill outside the tropical regions.

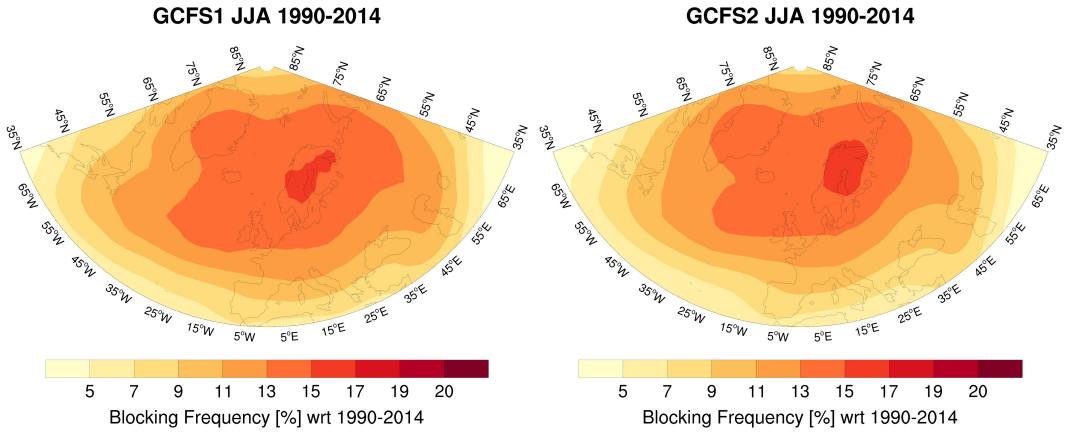
The middle panels, Figures 5c,d), show the change in skill from GCFS1.0 to GCFS2.0. For the inter-model comparison GCFS2.0 is mapped onto the coarse grid of GCFS1.0. If regions appear in reddish colours GCFS2.0 is superior over its predecessor GCFS1.0. At the surface relevant improvements for the temperature correlation can be seen over Alaska, the west Sibirien Plain and the Amazon region. A slight improvement is gained in northern Europe. At 500 hPa the correlation for the geopotential height is now stronger over Greenland and parts of Siberia as well as over parts of Antarctica. The negative skill over Europe is a pattern which has not changed much in comparison to GCFS1.0.

The difference in the probabilistic measure RPSS between the two model systems is shown in the bottom panels of Figure 5, again for temperature on the left in Figure 5e) and geopotential on the right in Figure 5f). As before, the red-coloured regions highlight the domains where the probabilistic hindcasts of GCFS2.0 for all event categories are better than the GCFS1.0 ensemble. The differences between the systems are stronger at the surface than in 500 hPa. GCFS2.0 probabilistic hindcasts have improved over the North and Baltic sea and its surroundings, in the Sahel zone and again over the Amazon region. Degradation is seen for the central tropical Pacific, tropical Atlantic and the Indonesian Archipelago. For geopotential at 500 hPa probabilistic hindcasts are now slightly worse around the tropics, while the North Atlantic shows neutral to slightly improved behaviour.

The structure of one of the prominent European summer features, the summer blocking events, remains also similar between the two systems as seen in Figure 6. Blocking is diagnosed from daily values of geopotential at 500 hPa of the hindcasts started in May by using the a combination of two methods based on *Tibaldi and Molteni* [1990] and *Barriopedro et al.* [2010]. The ERA-Interim reanalysis shows (Fig. 6a) that the European blocking has its maximum over northern Scandinavia. The region where events are present




a)




b) c)

**Figure 6.** Analysis of blocking frequency in JJA during 1990-2014 for a) ERA-Interim reanalysis, b) GCFS1.0 ensemble mean and c) GCFS2.0 ensemble mean.

for more than 15% of the summer days during the considered time range extends from Greenland to the Ural mountains and from Spitsbergen down south to Poland and Ukraine. While the general shape and the location of the central core of summer blockings match quite well between reanalysis and forecast systems, the extension and especially the amplitude of the GCFS1.0 and GCFS2.0 blockings are smaller. GCFS1.0 blockings reach farther south than GCFS2.0, while the newer system exhibits more Greenland events. This result is consistent with Fig. 5d), where the ACC of geopotential at 500 hPa for GCFS2.0 over Europe shows a neutral behaviour or a slight degradation but a considerable improvement over Greenland. The underestimation of this phenomenon is however not unexpected

as blocking processes are known to evolve properly only at horizontal grid resolutions of about 40 km as shown in, e.g. *Jung et al.* [2012].

### 3.3 DJF mean state and hindcast skill

As before, the ensemble and time means of GCFS1.0 and GCFS2.0 are compared by using the corresponding ERA-Interim period of December, January and February 1990-2014, computed from the November start date. The year is related to the start of the season, namely December.

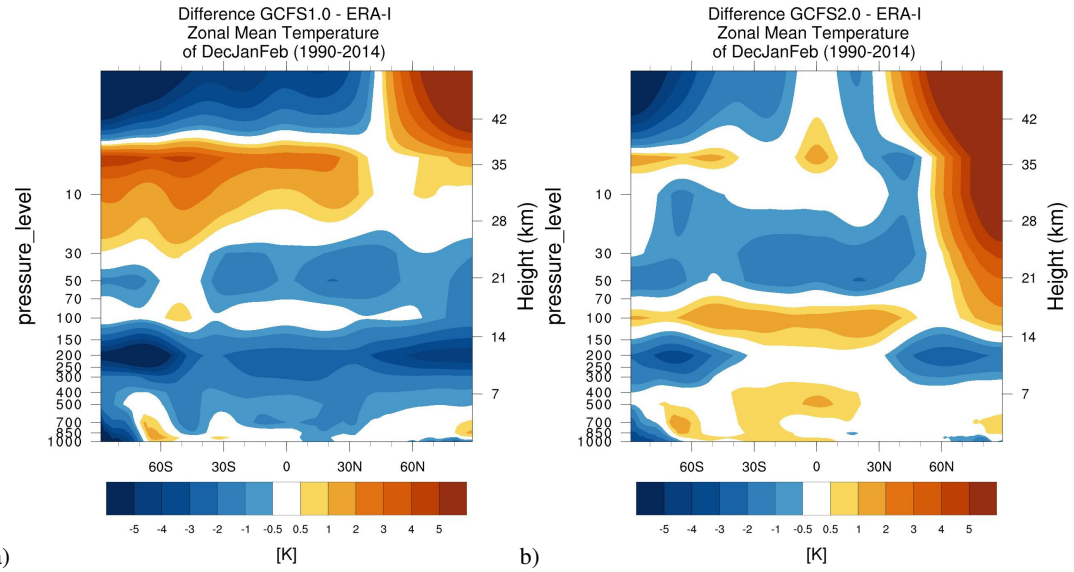



**Figure 7.** Temperature differences of a) GCFS1.0 ensemble mean and b) GCFS2.0 ensemble mean to ERA-Interim with respect to DJF hindcasts during 1990-2014.

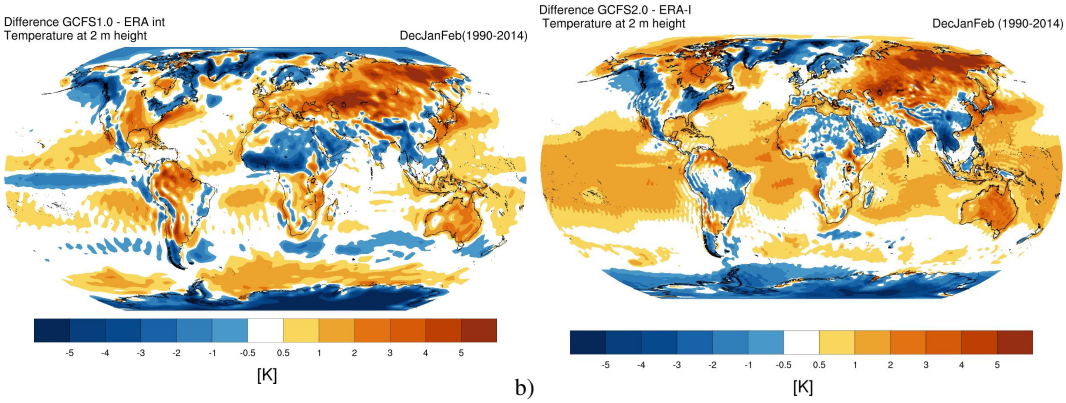



**Figure 8.** Difference of the T2m hindcasts a) GCFS1.0 and b) GCFS2.0 to ERA-Interim for DJF hindcasts between in 1990-2014.

Again, we show first the vertical structure of the temperature differences in Figure 7 for the DJF climate. For GCFS1.0 (Figure 7a) a strong cold bias is seen in the upper troposphere extending with weaker values up into the stratosphere. In the GCFS2.0 troposphere (Figure 7b), the cold bias is considerably reduced. In tropical and subtropical latitudes a minor warm bias is now present with a maximum at the tropopause.

The middle atmosphere of the GCFS1.0 Southern Hemisphere is too warm up until 35 km, while above a strong cold bias is visible suggesting again a misplacement of the ozone layer. In contrast, the winter stratosphere in the Northern Hemisphere in the same height appears to be too warm. This dipole structure is strongly reduced in GCFS2.0, due to the new ozone climatology and the higher resolution. Unfortunately, the warm bias in the winter polar stratosphere extends now from about 100 hPa up to 1 hPa. This seems to be related to a too strong (gravity) wave activity decelerating the polar vortex by wave breaking (personal communication with M. Giorgetta).

At the surface the bias distribution has also changed as is displayed in Figure 8. The tropical oceans now show a general warm bias for GCFS2.0 (8b), as seen before in the historical experiment (Fig. 1), but in the Pacific the cold tongue again vanishes as simulations start from an analysed climate state. Other strong biases from the GCFS1.0 are now also considerably reduced, like the warm bias over Europe, the Amazon region and southern Africa or the cold bias over northern Africa.

With these improvements comes a better representation of the jet in the storm track region (compare with Figure 10 of *Müller et al.* [2018]). This leads to a good skill pattern in ACC of 2m temperature as seen in Figure 9a and also for the geopotential height at 500 hPa pressure level (Figure 9b). While ACC values over Europe in DJF are not apparent, the skill in the Arctic region, over Greenland and northern America shows significant values up to 0.8 (Figure 9a). However, directly south of Greenland appears an oceanic region with negative skill. This is a new pattern in GCFS2.0 and is most probable not a model feature but stems from assimilating the ocean reanalysis data ORAS5 from ECMWF. The same negative Pearson correlation appears in the DJF hindcast skill of ECMWF's forecast model System 5 [*Johnson et al.*, 2019], see their Figure 19a. The authors also discuss ORAS5 as the reason of skill degradation. At the level of 500 hPa again the strongest positive values are found in the tropics. However, regions with considerable positive and significant skill are found over the eastern North-Pacific extending into North-America,

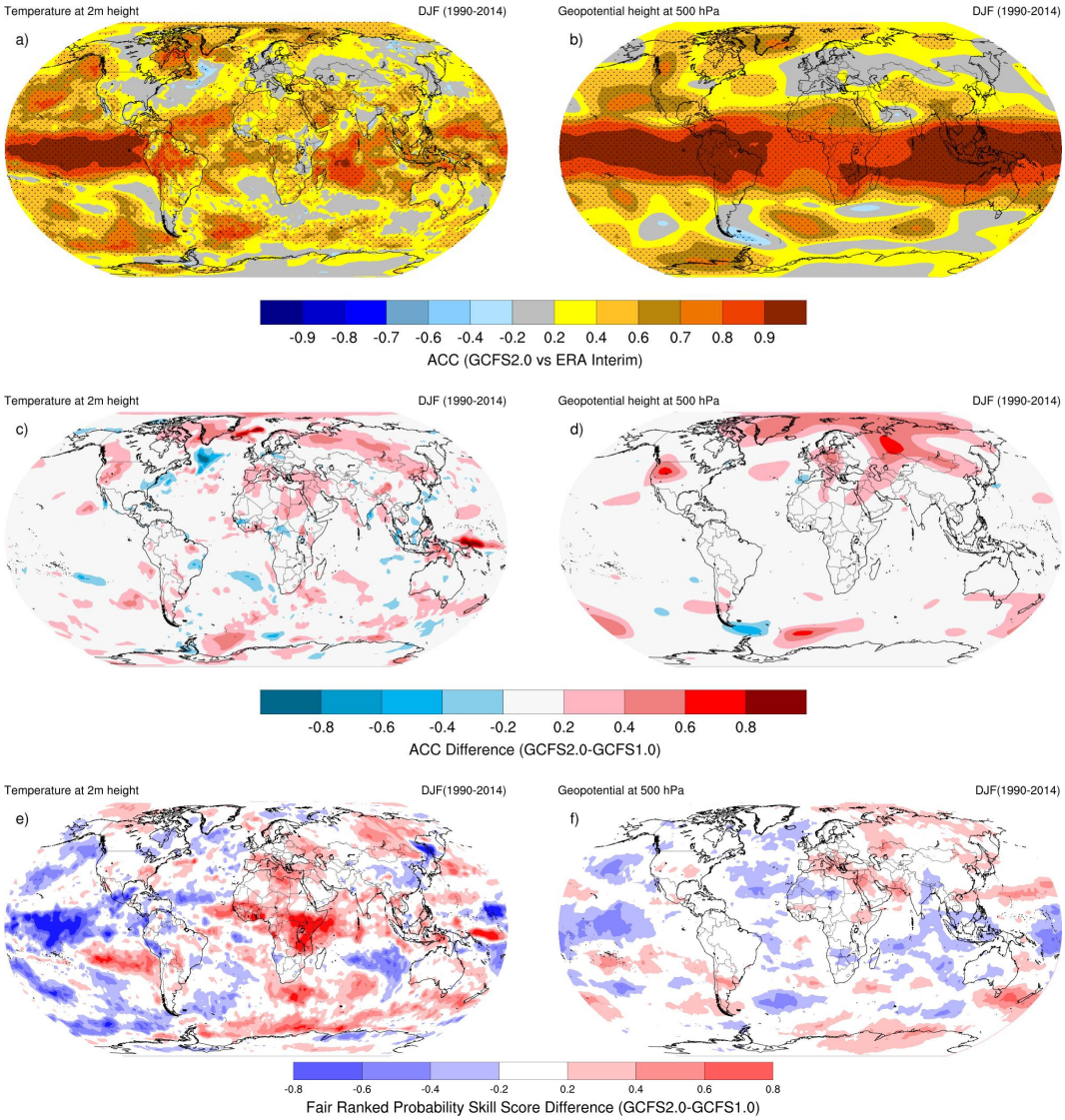



**Figure 9.** Top panel: Anomaly correlation coefficient (ACC) of hindcasts for December, January, February with respect to ERA-Interim: a) 2m temperature and b) geopotential height at 500 hPa pressure level. Dots represent significant values at the 95% confidence level. Middle panels show the difference of skill between GCFS1.0 and GCFS2.0 for the same variables c) 2m temperature and d) geopotentialxs height at 500 hPa. Bottom panels show the difference of GCFS1.0 and GCFS2.0 for the Fair Ranked Probability Skill Score for e) 2m temperature and f) geopotential height at 500 hPa.

as well as over Greenland and Arctic regions. Also the North-Atlantic shows a patch of significant positive skill.

The middle panels of Figure 9 highlight as before the change in skill in the ACC between the two versions GCFCS1.0 and GCFS2.0. Considerable skill has been gained

at the surface, represented by the $2m$ temperature (Figure 9c). Greenland and large parts of Eurasia benefit from the new version. Europe partly gains skill, especially for the very north and the southern regions and partly loses skill, especially over central Europe. A substantial gain in skill is further evident for the geopotential in the storm track level in the North Atlantic, Arctic regions (Figure 9d), as well as over Eurasia. The improvement over central Europe is gained by replacing negative ACC values with weak positive ones.

The bottom panels in Figure 9 present the difference of the two systems in terms of the probabilistic hindcast skill score RPSS as previously shown in Figure 5. Skill differences are stronger at the surface for temperature than for the middle troposphere in 500 hPa. During DJF GCFS2.0 probabilistic hindcasts are more skillful for eastern and southern Europe as well as eastern Russia. The skill amendment over Europe is seen as well in 500 hPa. A region, where GCFS2.0 forecasts are worse than its predecessor is the central tropical Pacific. The strongest gain in skill is seen over central Africa extending into the Indian Ocean. Improvements over the ocean are also visible for the subtropical southeastern Pacific and the Agulhas basin.

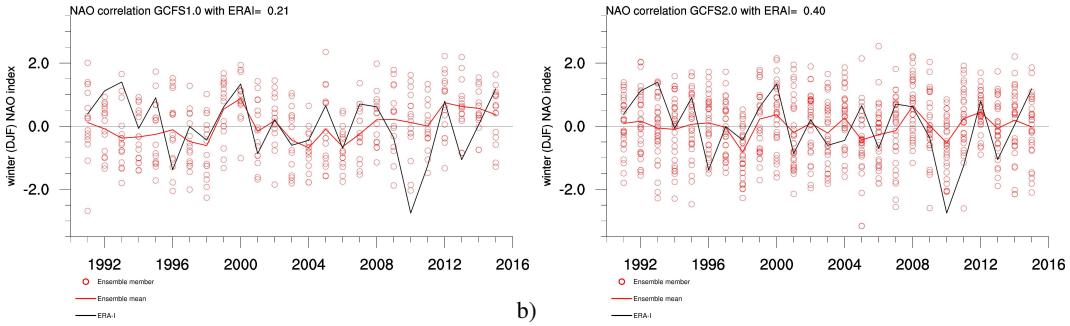



**Figure 10.** NAO time series of ERA-Interim, Model ensemble mean and the single ensemble members averaged over December, January, February during 1990 – 2014, a) GCFS1.0 and b) GCFS2.0. The labeled years refer to January of the considered DJF.

The improved conditions in the Northern Hemisphere show up as well in a doubling of skill in GCFS2.0 when predicting the North Atlantic Oscillation (NAO) for the upcoming DJF or boreal winter season, as shown in Figure 10. The NAO index has been calculated using an empirical orthogonal function (EOF) like in *Dobrynin et al.* [2018]. A considerably higher skill is also present for a reduced ensemble size down to the same number as in GCFS1.0 (not shown here). This highlights again, that beside enhancing the

ensemble size the improvements of the model dynamics and physics have the potential for long range forecasts in the mid-latitudes as shown by *Scaife et al.* [2014].

### 3.4 ENSO hindcasts

The assessment of ENSO hindcasts in Figure 11 for the Niño3.4 region shows a strong seasonal dependence of skill in GCFS2.0. However, considering all 12 start months predictive skill has shrunk in GCFS2.0. The general structure of difficult hindcast start months March, April, May and June can be seen as well in GCFS1.0 and is also known for other models as "spring predictability barrier" (see e.g. *Wang-Chun Lai et al.* [2017]), making ENSO predictions difficult for forecast systems in general. This feature has not been improved in GCFS2.0. However, in all other start months from July to December GCFS2.0 performs with a comparable forecast quality to GCFS1.0.

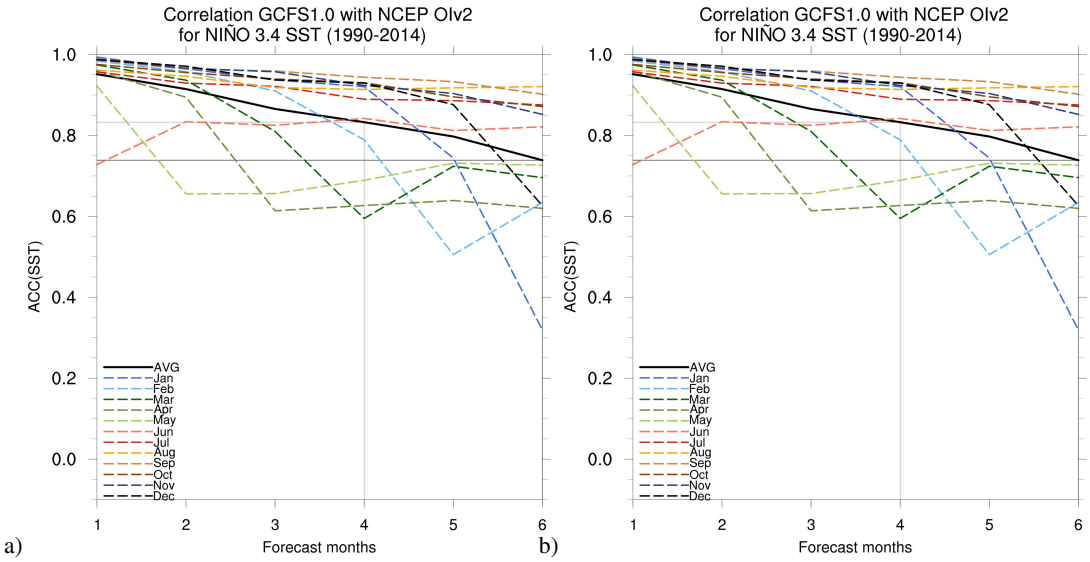



**Figure 11.** Anomaly correlation of the SST forecasts for the ENSO 3.4 region for a) GCFS1.0 and b) GCFS2.0, both wrt NCEP reanalysis for all 12 start months and all 6 lead months in 1990-2014.

The decreased ENSO skill in GCFS2.0 together with the strong tropical warm bias brings up the question, whether a code correction of the ocean mixing mentioned in *Mauritsen et al.* [2018] is able to alleviate the warm bias and also to improve the ENSO predictive skill. This 'bugfix' was implemented and tuned for the low resolution version CMIP6 MPI-ESM-LR but for the high resolution version MPI-ESM-HR it was decided to leave out this further tuning connected with the ocean correction.

*Johnson et al.* [2019] also report a warm bias of the ECMWF's seasonal forecast model System 5 during JJA hindcasts in the eastern Pacific basin (compare with their Figure 1d). The warm structure of System 5 is stronger pronounced north of the equator and much more confined to the South American coast. Their ENSO predictive quality improved with the higher resolution of the model.

To investigate the impact of the ocean code correction onto the CMIP6 MPI-ESM-HR climate an experiment is set up. Here, MPI-ESM-HR is run again for approximately 100 years under pre-industrial control conditions with correct ocean mixing in order to allow a spin-up for the ocean. Afterwards, a historical experiment simulates the time range from 1850 up to 2014. However, the comparison between the original historical run and the bugfixed version did not show the desired result of a cooler tropical belt. The revised ocean mixing in CMIP6 MPI-ESM-HR revealed a small improvement but no change of error pattern in the tropical Pacific similar to the change seen in Figure 1. Except for Arctic regions temperature differences between these two CMIP6 versions remain mostly below $1K$ and are hard to recognise. The reason for the tropical warm bias is thus suspected to derive from the revised atmospheric parameterizations. Required diagnostics and tuning experiments are currently beyond available resources.

## 4 Conclusions

To summarize, we have shown that the second version of the German Climate Forecast System brings some improvements over its predecessor. However, a version change does not necessarily lead to improvement everywhere and for every variable, which is certainly true here. During JJA, overall a similar behaviour of GCFS2.0 in comparison to GCFS1.0 is assessed. Skill degradation in JJA forecasts is prominent in the tropical Pacific and in the skill for ENSO. During DJF, hindcast quality is improved, especially for Northern Hemisphere where e.g. NAO skill for the winter months has doubled.

We have learned again with this version, increasing the model resolution is not per se a solution to many forecast problems as it requires an intense work and evaluation of the model physics on the new grid. Similar challenges are observed for the new system 5 at ECMWF [*Johnson et al.*, 2019], where many issues come up with the new resolution. *Scaife et al.* [2019] even recommend to invest more into the ensemble size, vertical resolution or ocean resolution than in increasing the atmospheric horizontal resolution. As can

be seen from our results, any changes on the model grid need carefull adaptation of the model physics.

For a future system, besides taking into account a well tuned climate and climate sensitivity of the Earth-System-Model, which is the key for climate experiments and climate projections, a comprehensive performance testing is needed depending on different time-scale applications of this model [*Schmidt et al.*, 2017]. In that way, processes active in different seasons or timescales can be accounted for during the model tuning.

**Acknowledgments**

Large parts of this work have been performed under the Copernicus Service Contracts C3S-433 and C3S-330. The version of ECHAM6 in MPI-ESM is `https://svn.zmaw.de/svn/echam6/tags/echam-6.3.04p1`, the version of MPIOM in MPI-ESM is `https://svn.zmaw.de/svn/mpiom/tags/mpiom-1.6.3`. Hindcast data used for this study and scripts to generate the figures are available by request to the first author under `https://doi.org/10.5281/zenodo.3697080`. GCFS2.0 data used for this study are available on the C3S climate data store `https://climate.copernicus.eu/seasonal-forecasts`.

## References

Athanisiadis, P. J., and A. Bellucci (2017), A multisystem view of wintertime nao seasonal predictions, *Journal Of Climate*, *30*, 1461–1475, doi:10.1175/JCLI-D-16-0153.1.

Baehr, J., and R. Piontek (2014), Ensemble initialization of the oceanic component of a coupled model through bred vectors at seasonal-to-interannual timescales, *Geoscientific Model Development*, *7*(1), 453–461, doi:10.5194/gmd-7-453-2014.

Baehr, J., K. Fröhlich, M. Botzet, D. I. V. Domeisen, L. Kornblueh, D. Notz, R. Piontek, H. Pohlmann, S. Tietsche, and W. A. Müller (2015), The prediction of surface temperature in the new seasonal prediction system based on the mpi-esm coupled climate model, *CLIMATE DYNAMICS*, *44*(9-10), 2723–2735, doi:10.1007/s00382-014-2399-7.

Ban, R. J., E. Dunlea, A. Ravishankara, D. Halpern, J. Francis, and A. Staudt (2016), *Next Generation Eart System Prediction: Strategies for Subseasonal to Seasonal Forecasts*, National Academies of Science, Engineering and Medicine, Washington, DC: The National Academic Press, doi:10.17226/21873.

Barriopedro, D., R. Garcia-Herrera, and R. Trigo (2010), Application of blocking diagnosis methods to general circulation models. part i: a novel detection scheme, *Climate*

*Dynamics*, *35*(7-8), 1373–1391, doi:10.1007/s00382-010-0767-5.

Bunzel, F., W. A. Müller, M. Dobrynin, K. Fröhlich, S. Hagemann, H. Pohlmann, T. Stacke, and J. Baehr (2017), Improved seasonal prediction of european summer temperatures with new five-layer soil-hydrology scheme, *Geophysical Research Letters*, *45*(1), 346–353, doi:10.1002/2017GL076204.

Dee, D. P., S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hólm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, and F. Vitart (2011), The era-interim reanalysis: configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, *137*(656), 553–597, doi:10.1002/qj.828.

Dobrynin, M., D. I. V. Domeisen, W. A. Müller, L. Bell, S. Brune, F. Bunzel, A. Düsterhus, K. Fröhlich, H. Pohlmann, and J. Baehr (2018), Improved teleconnection-based dynamical seasonal predictions of boreal winter., *Geophys. Res. Lett.*, *45*, 3605–3614, doi:doi.org/10.1002/2018GL077209.

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor (2016), Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization, *Geoscientific Model Development*, *9*(5), 1937–1958, doi:10.5194/gmd-9-1937-2016.

Ferro, C. (2014), Fair scores for ensemble forecasts. , *Q.J.R. Meteorol. Soc.*, *140*, 1917–1923.

Fetterer, F., K. Knowles, W. Meier, and M. Savoie (2002), Sea ice index, doi:10.7265/N5QJ7F7W.

Giorgetta, M. A., J. Jungclaus, C. H. Reick, S. Legutke, J. Bader, M. Böttinger, V. Brovkin, T. Crueger, M. Esch, K. Fieg, K. Glushak, V. Gayler, H. Haak, H.-D. Hollweg, T. Ilyina, S. Kinne, L. Kornblueh, D. Matei, T. Mauritsen, U. Mikolajewicz, W. Mueller, D. Notz, F. Pithan, T. Raddatz, S. Rast, R. Redler, E. Roeckner, H. Schmidt, R. Schnur, J. Segschneider, K. D. Six, M. Stockhause, C. Timmreck, J. Wegner, H. Widmann, K.-H. Wieners, M. Claussen, J. Marotzke, and B. Stevens (2013), Climate and carbon cycle changes from 1850 to 2100 in mpi-esm simulations

for the coupled model intercomparison project phase 5, *Journal of Advances in Modeling Earth Systems*, *5*(3), 572–597, doi:10.1002/jame.20038.

Hagemann, S., and T. Stacke (2015), Impact of the soil hydrology scheme on simulated soil moisture memory, *Clim. Dyn.*, *44*, 1731–1750, doi:10.1007/s00382-014-2221-6.

Ilyina, T., K. D. Six, J. Segschneider, E. MaierReimer, H. Li, and I. NunezRiboni (2013), Global ocean biogeochemistry model hamocc: Model architecture and performance as component of the mpi earth system model in different cmip5 experimental realizations, *J. Adv. Modelling Earth Syst. (JAMES)*, *5*, 287–315, doi:10.1029/2012MS000178.

Johnson, S. J., T. N. Stockdale, L. Ferranti, M. A. Balmaseda, F. Molteni, L. Magnusson, S. Tietsche, D. Decremer, A. Weisheimer, G. Balsamo, and S. P. E. Keeley (2019), Seas5: the new ecmwf seasonal forecast system, *Geosci. Model Dev*, *12*, 1087–1117, doi.org/10.5194/gmd-12-1087-2019.

Jung, T., M. Miller, T. Palmer, P. Towers, N. Wedi, D. Achuthavarier, J. Adams, E. Altshuler, B. Cash, J. Kinter, L. Marx, C. Stan, , and K. Hodges (2012), High-resolution global climate simulations with the ecmwf model in project athena: Experimental design, model climate, and seasonal forecast skill., *J. Climate*, *25*, 3155–3177, doi:10.1175/JCLI-D-11-00265.1.

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, *Journal of Advances in Modeling Earth Systems*, *5*(2), 422–446, doi:10.1002/jame.20023.

MacLachlan, C., A. Arribas, K. Peterson, A. Maidens, D. Fereday, A. Scaife, M. Gordon, M. Vellinga, A. Williams, R. E. Comer, J. Camp, and P. Xavier (2015), Global Seasonal forecast system version 5 (GloSea5): A high resolution seasonal forecast system , *Quarterly Journal of the Royal Meteorological Society*, *141*, 1072–1084, doi:10.1002/qj.2396.

Manzanas, R., J. Gutierrez, J. Bhend, S. Hemri, F. Doblas-Reyes, V. Torralba, E. Penabad Ramos, and A. Brookshaw (2019), Bias adjustment and ensemble recalibration methods for seasonal forecasting: A comprehensive intercomparison using the c3s dataset., *Climate Dynamics*, doi:10.1007/s00382-019-04640-4.

Mauritsen, T., B. Stevens, E. Roeckner, T. Crueger, M. Esch, M. A. Giorgetta, H. Haak, J. Jungclaus, D. Klocke, D. Matei, U. Mikolajewicz, D. Notz, R. Pincus, H. Schmidt, and L. Tomassini (2012), Tuning the climate of a global model, *Journal of Advances in*

*Modeling Earth Systems*, *4*, 1–18.

Mauritsen, T., J. Bader, T. Becker, J. Behrens, M. Bittner, R. Brokopf, V. Brovkin, M. Claussen, T. Crueger, M. Esch, I. Fast, S. Fiedler, D. Fläschner, V. Gayler, M. Giorgetta, D. S. Goll, H. Haak, S. Hagemann, C. Hedemann, C. Hohenegger, T. Ilyina, T.Jahns, D. J. de la Cuesta Otero, J. Jungclaus, T. Kleinen, S. Kloster, D. Kracher, S. Kinne, D. Kleberg, G. Lasslop, L. Kornblueh, J. Marotzke, D. Matei, K. Meraner, U. Mikolajewicz, K. Modali, B. Möbis, W. A. Müller, J. E. M. S. Nabel, C. C. W. Nam, D. Notz, S.-S. Nyawira, H. Paulsen, K. Peters, R. Pincus, H. Pohlmann, J. Pongratz, M. Popp, T. Raddatz, S. Rast, R. Redler, C. H. Reick, T. Rohrschneider, V. Schemann, H. Schmidt, R. Schnur, U. Schulzweida, K. D. Six, L. Stein, I. Stemmler, B. Stevens, J.-S. von Storch, F. Tian, A. Voigt, P. de Vrese, K.-H. Wieners, S. Wilkenskjeld, A. Winkler, and E. Roeckner (2018), Developments in the mpi-m earth system model version 1.2 (mpi-esm1.2) and its response to increasing $co_2$, *J. Adv. Modelling Earth Syst. (JAMES)*, *11*, 998–1038, doi:10.1029/2018MS001400.

Mogensen, K., M. A. Balmaseda, and A. Weaver (2012), The nemovar ocean data assimilation system as implemented in the ecmwf ocean analysis for system 4, Shinfield Park, Reading, 59 pp.

Müller, W. A., J. H. Jungclaus, T. Mauritsen, J. Baehr, M. Bittner, R. Budich, F. Bunzel, M. Esch, R. Ghosh, H. Haak, T. Ilyina, T. Kleine, L. Kornblueh, H. Li, K. Modali, D. Notz, H. Pohlmann, E. Roeckner, I. Stemmler, F. Tian, and J. Marotzke (2018), A higher-resolution version of the max planck institute earth system model (mpi-esm1.2-hr), *Journal of Advances in Modeling Earth Systems*, *10*(7), 1383–1413, doi:10.1029/2017MS001217.

Pohlmann, H., W. A. Müller, M. Bittner, S. Hettrich, K. Modali, K. Pankatz, and J. Marotzke (2019), Realistic quasibiennial oscillation variability in historical and decadal hindcast simulations using cmip6 forcing., *Geophys. Res. Lett.*, *46*, 14,118–14,125, doi:10.1029/2019GL084878.

Polkova, I., S. Brune, C. Kadow, V. Romanova, G. Gollan, J. Baehr, R. Glowienka Hense, R. J. Greatbatch, A. Hense, S. Illing, A. Köhl, J. Kröger, W. A. Müller, K. Pankatz, and D. Stammer (2019), Initialization and ensemble generation for decadal climate predictions: A comparison of different methods, *Journal of Advances in Modeling Earth Systems*, *11*, 149–172, doi:10.1029/2018MS001439.

Scaife, A., J. Camp, R. Comer, P. Davis, N. Dunstone, M. Gordon, C. MacLchlan, N. Martin, Y. Nie, H.-L. Ren, M. Roberts, W. Robinson, D. Smith, and P. Vidale (2019), Does increased atmospheric resolution improve seasonal climate predictions?, *Atmospheric Science Letters*, *20*, 1–10, doi:10.1002/asl.922.

Scaife, A. A., A. Arribas, E. Blockley, A. Brookshaw, R. T. Clark, N. Dunstone, R. Eade, D. Fereday, C. K. Folland, M. Gordon, L. Hermanson, J. R. Knight, D. J. Lea, C. MacLachlan, A. Maidens, M. Martin, A. K. Peterson, D. Smith, M. Vellinga, E. Wallace, J. Waters, and A. Williams (2014), Skillful long-range prediction of european and north american winters, *Geophysical Research Letters*, *41*(7), 2514–2519, doi:10.1002/2014GL059637.

Schmidt, G. A., D. Bader, L. J. Donner, G. S. Elsaesser, J.-C. Golaz, C. Hannay, A. Molod, R. B. Neale, and S. Saha (2017), Practice and philosophy of climate model tuning across six us modeling centers, *Geoscientific Model Development*, *10*(9), 3207–3223, doi:10.5194/gmd-10-3207-2017.

Stockdale, T. N. (2012), The eurosip system - multi-model approach, *conference paper*.

Taylor, K. E., R. J. Stouffer, and G. A. Meehl (2012), An overview of CMIP5 and the experiment design, *Bull. Amer. Meteor. Soc.*, *93*(4), 485–498, doi:10.1175/bams-d-11-00094.1.

Tibaldi, S., and F. Molteni (1990), On the operational predictability of blocking, *Tellus A*, *42*(3), 343–365, doi:10.1034/j.1600-0870.1990.t01-2-00003.x.

Wang-Chun Lai, A., M. Herzog, and H. Graf (2017), Enso forecasts near the spring predictability barrier and possible reasons for the recently reduced predictability., *J. Climate*, *31*, 815–838, doi:10.1175/JCLI-D-17-0180.1.

White, C., H. Carlsen, A. Robertson, R. Klein, J. Lazo, A. Kumar, F. Vitart, E. Coughlan de Perez, A. Ray, V. Murray, S. Bharwani, D. MacLeod, R. James, L. Fleming, A. Morse, B. Eggen, R. Graham, E. Kjellström, E. Becker, K. Pegion, N. Holbrook, D. McEvoy, M. Depledge, S. Perkins-Kirkpatrick, T. Brown, R. Street, L. Jones, T. Remenyi, I. Hodgson-Johnston, C. Buontempo, R. Lamb, H. Meinke, B. Arheimer, and S. Zebiak (2017), Potential applications of subseasonal-to-seasonal (s2s) predictions, *Met. Apps.*, *24*, 315–325, doi:10.1002/met.1654.

Wilks, D. (1995), *Statistical Methods in the Atmospheric Sciences*, 467 pp., Academic Press and San Diego.

645 Zou, H., M. Balmaseda, and C. Mogensen (2017), The new eddy-permitting orap5 ocean
646 reanalysis: description, evaluation and uncertainties in climate signals, *Climate Dynam-*
647 *ics*, *49*, 791–811.