Meteorologische Zeitschrift, PrePub DOI 10.1127/metz/2015/0601

MiKlip

UTLS temperature validation of MPI-ESM decadal hindcast experiments with GPS radio occultations

Torsten Schmidt[1*], Lena Schoon[1,2], Henryk Dobslaw[1], Katja Matthes[3], Maik Thomas[1] and Jens Wickert[1]

[1]GFZ German Research Centre for Geosciences, Potsdam, Germany
[2]now at Leibniz Institute of Atmospheric Physics, Kühlungsborn, Germany
[3]GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

(Manuscript received March 15, 2014; in revised form January 23, 2015; accepted June 12, 2015)

Abstract

Global Positioning System (GPS) radio occultation (RO) temperature data are used to validate MPI-ESM (Max Planck Institute – Earth System Model) decadal hindcast experiments in the upper troposphere and lower stratosphere (UTLS) region between 300 hPa and 10 hPa (8 km and 32 km) for the time period between 2002 and 2011. The GPSRO dataset is unique since it is very precise, calibration independent and covers the globe better than the usual radiosonde dataset. In addition it is vertically finer resolved than any of the existing satellite temperature measurements available for the UTLS and provides now a unique one decade long temperature validation dataset. The initialization of the MPI-ESM decadal hindcast runs mostly increases the skill of the atmospheric temperatures when compared to uninitialized climate projections with very high skill scores for lead-year one, and gradually decreases for the later lead-years. A comparison between two different initialization sets (b0, b1) of the low-resolution (LR) MPI-ESM shows increased skills in b1-LR in most parts of the UTLS in particular in the tropics. The medium resolution (MR) MPI-ESM initializations are characterized by reduced temperature biases in the uninitialized runs as compared to observations and a better capturing of the high latitude northern hemisphere interannual polar vortex variability as compared to the LR model version. Negative skills are found for the b1-MR hindcasts however in the regions around the mid-latitude tropospheric jets on both hemispheres and in the vicinity of the tropical tropopause in comparison to the b1-LR variant. It is interesting to highlight that none of the model experiments can reproduce the observed positive temperature trend in the tropical tropopause region since 2001 as seen by GPSRO data.

Keywords: decadal hindcast experiments, MPI-ESM, radio occultation, UTLS temperature

1 Introduction

The relatively new field of decadal climate predictions attempts to bridge the gap between seasonal predictions and climate projections. Statements about decadal changes are socio-economically highly relevant also with respect to adaptation and mitigation strategies for climate change. The scientific challenge for decadal climate prediction arises from the fact that decadal climate signals are difficult to distinguish from internally generated climate variability (Meehl et al., 2009). The low-frequency climate variability is mostly determined from natural variability in the climate system such as El Niño, fluctuations in the ocean thermohaline circulation, and anomalies in upper ocean heat content (Smith et al., 2007). Therefore, the knowledge of the current (initial) climate state and the annual variability of the ocean is essential.

The assessment of forecast skill of any decadal prediction system is typically based on validating extensive sets of hindcast experiments, where observations are already available (Smith et al., 2007). Beside surface air temperatures (Pohlmann et al., 2013) or precipitation (Goddard et al., 2012), the validation of upper air temperatures from decadal hindcast experiments in the upper troposphere and lower stratosphere (UTLS) region is important, because the UTLS is one of the key regions for troposphere-stratosphere interactions since it reacts very sensitive to climate change (Fueglistaler et al., 2009; Gettelman et al., 2011). Changes in the transport of trace gases entering the tropical stratosphere might directly affect the radiative balance and hence impact surface climate. The observation of long-term stratospheric temperature trends based on a wide range of different datasets has a long history (Randel et al., 2009) and even ambiguities as far as the interpretation of satellite instruments with coarse vertical resolution are concerned (Thompson et al., 2012).

The most important data source for atmospheric temperatures with high vertical resolution in the UTLS are radiosondes, where a global equally distributed coverage is generally impossible. Global Positioning System (GPS) radio occultation (RO) data enable atmospheric information with high vertical resolution (less than 0.5 km in the troposphere and about 1 km in the lower stratosphere). The GPSRO technique requires no

*Corresponding author: Torsten Schmidt, GFZ German Research Centre for Geosciences, Potsdam, Germany, e-mail: torsten.schmidt@gfz-potsdam.de

active calibration, is weather independent, and the occultations are almost uniformly distributed over the globe (Kursinski et al., 1997). The RO method exploits microwave GPS signals received onboard a Low Earth Orbiting (LEO) satellite for atmospheric limb sounding. The GPS signals are influenced by the atmospheric refractivity, resulting in a time delay and bending of the signal enabling calculation of refractivity and subsequently dry temperature as a function of height (Kursinski et al., 1997). The atmospheric excess phase is the basic observable that is measured with millimetric accuracy (Wickert et al., 2001) and is considered as a climate benchmark observable (Steiner et al., 2013).

The quality of the GPSRO temperature data were assessed by several comparison studies with radiosondes as well as cross-validations with other satellite sensors, [e.g., Hajj et al. (2004); Wang et al. (2004); Kuo et al. (2005); Steiner et al. (2007)] demonstrating the high precision of the GPSRO temperature measurements in the altitude range between 8 km and 30 km with variations less than 0.5 K. Recently Alexander et al. (2014) estimated an accuracy better than 0.1 K for GPSRO temperatures between 10 km and 30 km height for COSMIC (Constellation Observing System for Meteorology, Ionosphere, and Climate) temperature profiles (Anthes et al., 2008). Since 2006 leading weather service centers use RO data for the assimilation and forecast system (numerical weather prediction, NWP) and it is found that RO observations have a strong impact among all satellite observations (Cardinali and Healy, 2014).

The general potential of GPSRO data for climate monitoring has been shown first with simulation studies (Leroy et al., 2006). Recently, Steiner et al. (2013) compared CHAMP (CHAllenging Minisatellite Payload) bending angel, refractivity, and temperature data from different RO processing centers. The authors found that the structural uncertainty in temperature trends over 7 years (2001 to 2008) is less than 0.06 K in the tropics and mid-latitudes (50 ° S to 50 ° N) and the altitude range between 8 km and 25 km, low enough for detecting a climate change signal within about a decade (GCOS, 2006). Larger structural uncertainty (with a maximum of 0.2 K) above 25 km and at higher latitudes is attributable to differences in the processing schemes (Steiner et al., 2013). This results generally favor the RO temperature data for the validation of decadal hindcast experiments.

In this study which is part of the research project 'Mittelfristige Klimaprognosen' (MiKlip) we use the unique one decade long GPSRO dataset from CHAMP, GRACE (Gravity Recovery and Climate Experiment), and TerraSAR-X, processed at GFZ German Research Centre for Geosciences, Potsdam and use it to validate three different sets of decadal hindcast experiments performed with the coupled Earth System Model from Max Planck Institute for Meteorology (MPI-ESM) (Müller et al., 2012; Pohlmann et al., 2013). We follow the verification frame proposed by Goddard et al. (2012) to assess the prediction quality of the MPI-ESM decadal hindcasts against GPSRO temperature anomalies.

The database is described in Section 2, whereas Section 2.1 gives an overview of satellite data processing and Section 2.2 outlines a brief summary on the available hindcast experiments. Section 3 sketches the validation approach applied (Section 3.1) and the results for deterministic skill scores (Section 3.2), followed by conclusions in the final section of this paper.

2 Database

Within Miklip three different ensemble sets of decadal hindcast experiments are available from two different versions of the coupled MPI-ESM climate model. The low resolution (LR) model includes the ocean-sea-ice component MPIOM (Max Planck Institute ocean model) (Jungclaus et al., 2013) of the MPI-ESM on a 1.5 ° grid with 40 vertical layers and the atmospheric general circulation model ECHAM6 with T63 horizontal resolution, i.e. 1.875 ° grid, and 47 vertical levels up to 0.1 hPa. A second medium resolution (MR) variant of MPI-ESM has a finer horizontal resolution in the ocean (0.4 °) and a finer vertical resolution in the atmosphere (95 instead of 47 vertical levels, so that the quasi-biennial oscillation (QBO) develops self-consistently in the latter version).

By using temperature observations from GPSRO an attempt is made to identify how skillful different sets of hindcasts from the MPI-ESM are in terms of predicting temperatures in the UTLS. The comparison between the hindcast experiments and GPSRO observations is finally performed on the basis of annual mean temperatures (2002 to 2011) on 10 standard pressure levels between 300 hPa and 10 hPa and 18 latitude bins with a meridional resolution of 10 ° centered at 85 ° N, 75 ° N, ..., 85 ° S, 75 ° S. The annual means are based on monthly means for the according pressure level and latitude bin.

Because GPSRO data are available since May 2001 we chose a comparable decadal time period from 2002 to 2011 for the observations and the model experiments.

2.1 Observations: GPS radio occultation temperature data

The proof-of-concept GPSRO experiment GPS/MET (GPS/METeorology) performed between 1995 and 1997 has demonstrated for the first time the potential of GPS based limb sounding from LEO satellites for deriving atmospheric temperature profiles (Kursinski et al., 1997). The CHAMP mission has generated the first long-term GPSRO dataset (2001 to 2008). Beside one complete month of missing data (July 2006) CHAMP delivered continuously between 150 and 200 temperature profiles daily (Schmidt et al., 2005). In addition to the CHAMP dataset ending in September 2008 RO data from GRACE-A and TerraSAR-X are available since 2006 and 2009, respectively, delivering atmospheric profiles with the same daily data rate and comparable error characteristics as from CHAMP (Wickert et al., 2009).

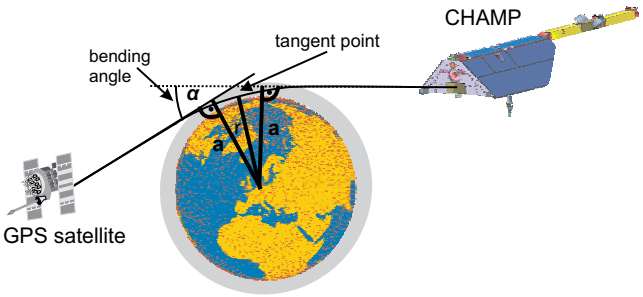

Figure 1: Radio occultation method. a denotes the impact parameter and r the tangent radius (Eq. (2.1)).

Worldwide, there are only few GPSRO data processing centers providing RO products. GFZ Potsdam has delivered RO data for CHAMP from 2001 to 2008, and is still in charge of the RO processing for GRACE (since 2006) and TerraSAR-X (since 2009), and more recently for the Tandem-X (since 2014) mission. UCAR (University Corporation for Atmospheric Research) and EUMETSAT (European Organisation for the Exploitation of Meteorological Satellites) provide GPSRO data from the COSMIC and Metop missions since 2006.

GPSRO data analysis

A detailed general description of the derivation of vertical atmospheric profiles from RO measurements is presented by Kursinski et al. (1997). Here, a brief summary is given: The GPS receiver onboard the LEO records phase and amplitude variations with high temporal resolution (e.g., 50 Hz) during an occultation event. By using high precision orbit information from the LEO and the occulting GPS satellite (König et al., 2002) the atmospheric excess phase can be extracted which is related to a bending angle profile (α) as a function of the impact parameter a (Figure 1):

$$a = nr = n(z + R_c) \tag{2.1}$$

(n: refractive index, r: tangent radius, z: geometric height, and R_c: local radius of curvature). Assuming spherical symmetry the bending angles can be related to the refractive index n using an Abel transform. Finally, the atmospheric refractivity N_{ref} in the neutral atmosphere is given by (Smith and Weintraub, 1953):

$$N_{\text{ref}} = (n-1)\cdot 10^6 = 77.6\frac{p}{T} + 3.73\cdot 10^5 \frac{e_w}{T^2} \tag{2.2}$$

(p: total air pressure, T: air temperature, and e_w: water vapor pressure). To convert the refractivity profiles into pressure and temperature profiles the assumption of dry air has to be made because of the ambiguity between the dry and wet part in the resulting refractivity (Eq. (2.2)). Further on, by applying the hydrostatic equation pressure and temperature profiles can be calculated.

The assumption of dry air in the RO retrieval (Eq. (2.2)) leads to lower (dry) RO temperatures in relation to physical temperatures if atmospheric water vapor

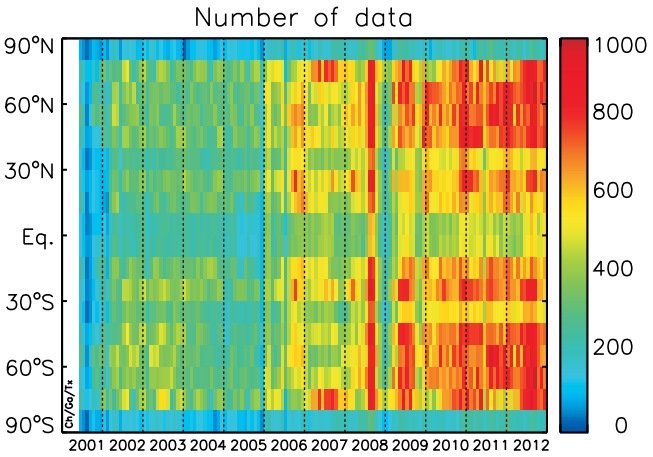

Figure 2: Number of monthly zonal CHAMP, GRACE, and TerraSAR-X temperature data for different latitude bands processed at GFZ.

is dominant and not negligible. After Kursinski et al. (1997), the assumption of dry air is justified if atmospheric temperatures are below 250 K which is a suitable approach in the UTLS above 300 hPa. For this reason we restrict the validation of the decadal hindcast experiments with RO temperatures to the altitude range between 300 hPa and 10 hPa.

Figure 2 shows the monthly number of RO profiles per 10 degree zonal latitude band from the CHAMP (2001 to 2008), GRACE (since 2006), and TerraSAR-X (since 2009) missions.

Data binning and sampling error estimation

Zonal monthly mean temperatures from RO data ($\overline{T_{\text{RO}}}$) at standard pressure levels between 300 hPa and 10 hPa were calculated for the 18 latitude bins (see above) by using cosine weighted averages for the time period 2002 to 2011. In a first step uncorrected zonal means were calculated according to the availability of the monthly number of RO temperature profiles per latitude bin:

$$\overline{T_{\text{ROuncorrected}}(\varphi_i, p_j)} = \frac{\sum_{i=1}^{N} T(\varphi, p)\cdot\cos(\varphi)}{\sum_{i=1}^{N}\cos(\varphi)} \tag{2.3}$$

with $i = 1\ldots18$ (number of latitude bins) and $j = 1\ldots10$ (number of pressure levels). In Eq. (2.3) φ and p denote the latitude and pressure level of the observation whereas N represents the number of observations at the according latitude range.

Due to the orbit constellation between the LEO (CHAMP, GRACE, TerraSAR-X) and GPS satellites the global distribution of RO temperature profiles is non-uniform in time and space meaning that the zonal monthly mean temperature climatologies are based on a different number of spatio-temporal observations in each month. This causes a sampling error which has to be corrected. The sampling error can be estimated if the real state of the atmosphere is known continuously at any time, which might be approximated by output

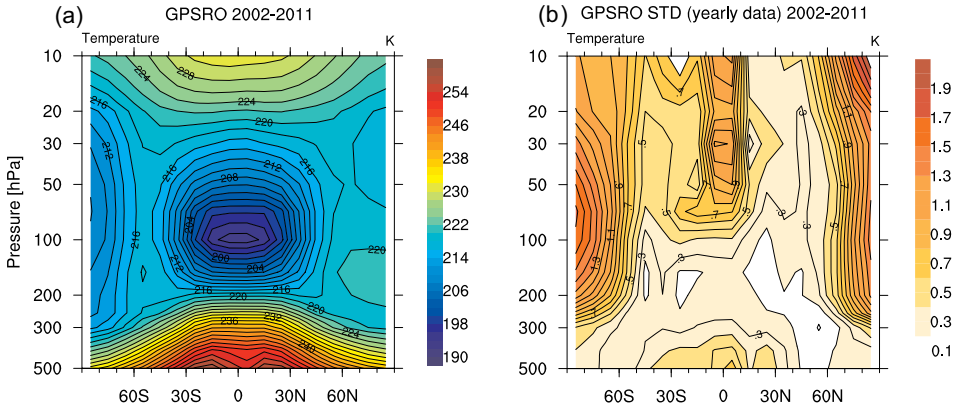

Figure 3: (a) Zonal mean temperature climatology from GPSRO data based on the time interval 2002 to 2011; contour interval: 2 K. (b) Same as (a) but for the standard deviation; contour interval: 0.2 K. The meridional resolution is 10 °.

from NWP models. Following Fölsche et al. (2007) a (simulated-true) temperature climatology $\overline{T_{\mathrm{SimTrue}}(\varphi, p)}$ is calculated by using 6-hourly European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-Interim reanalysis data (Dee et al., 2011) with a horizontal resolution of 1 ° by 1 ° similar to Eq. (2.3). An additional (simulated-observed) climatology $\overline{T_{\mathrm{SimObs}}(\varphi, p)}$ is calculated from ERA-Interim by using only data points interpolated to the time and location of the radio occultations. If it is considered that the ERA-Interim data have the same statistical properties as the RO data, the $\overline{T_{\mathrm{SimObs}}(\varphi, p)}$ climate data will be affected by the same sampling errors as the real observations. The true data are unknown, but with the knowledge of the $\overline{T_{\mathrm{SimTrue}}(\varphi, p)}$ data, errors introduced by an incomplete sampling can be estimated (Fölsche et al., 2007). The difference:

$$\overline{T_{\mathrm{SamplingError}}(\varphi, p)} = \overline{T_{\mathrm{SimTrue}}(\varphi, p)} - \overline{T_{\mathrm{SimObs}}(\varphi, p)} \tag{2.4}$$

is attributed to the zonal monthly mean sampling error. Finally the zonal monthly mean sampling error corrected RO temperature climatologies are represented by the difference:

$$\overline{T_{\mathrm{RO}}(\varphi, p)} = \overline{T_{\mathrm{ROuncorrected}}(\varphi, p)} - \overline{T_{\mathrm{SamplingError}}(\varphi, p)} \tag{2.5}$$

Figure 3a shows the sampling error corrected GPSRO temperature climatology for the UTLS region between 2002 and 2011. Well-known features are the temperature minimum at the tropical tropopause with temperatures less than 194 K, the cold polar regions with lower temperatures on the Southern Hemisphere (SH) within the Antarctic polar vortex. The standard deviation (Figure 3b) exhibits the highest values in the polar vortex region on both the SH and the Northern Hemisphere (NH). Enhanced temperature variability is also observed above the tropical tropopause (< 100 hPa) due to the QBO.

It is also evident from Figure 3a that the mean GPSRO temperature during the considered time period is less than 250 K between 300 hPa and 10 hPa. As many previous studies have shown (we refer to Kursinski et al. (1997) and selected paper in the introduction) it is save to interpret the (dry) RO temperatures as physical temperatures in this altitude region.

2.2 Model: Decadal hindcasts from MPI-ESM

For both, the LR and MR types of the coupled MPI-ESM climate models (see above), historical runs in line with the CMIP5 (Coupled Model Intercomparison Project Phase 5) protocol (Taylor et al., 2012) are available until 2006. For the subsequent years, the projections follow the Representative Concentration Pathway RCP4.5 emission scenario. In the following those runs are denoted as 'projections' or 'uninitialized model runs'. In total, three ensemble members of projections are available from both versions of the MPI-ESM for the years 2002 through 2011.

Both, the MPI-ESM projections and the mean GPSRO temperature climatology serve as a reference forecast for the calculation of the mean-squared skill score and since we use anomaly timeseries (Section 3), the mean GPSRO temperature climatology converts into the zero anomaly forecast (ZAF).

Figure 4 exhibits the mean temperature differences between the ensemble mean of the projections (uninitialized model runs) and GPSRO data for the time period 2002 to 2011 for low (Figure 4a) and medium (Figure 4b) resolution of the MPI-ESM projections. The LR model version has a cold bias with maximum differences up to 6 K in the polar upper troposphere (300 hPa to 150 hPa). A cold bias of more than 2 K occurs also in the tropical and subtropical lower stratosphere, the region where the QBO dominates the variability. At the edge of the polar vortex (˜60 °) on both hemispheres above 30 hPa the temperature bias for this version is positive instead.

In contrast, the major differences between model and observations are smaller in the MR version of the uninitialized ensemble mean in comparison with the LR version and range between ±2 K with slightly positive values poleward of about 40 ° on both hemispheres above

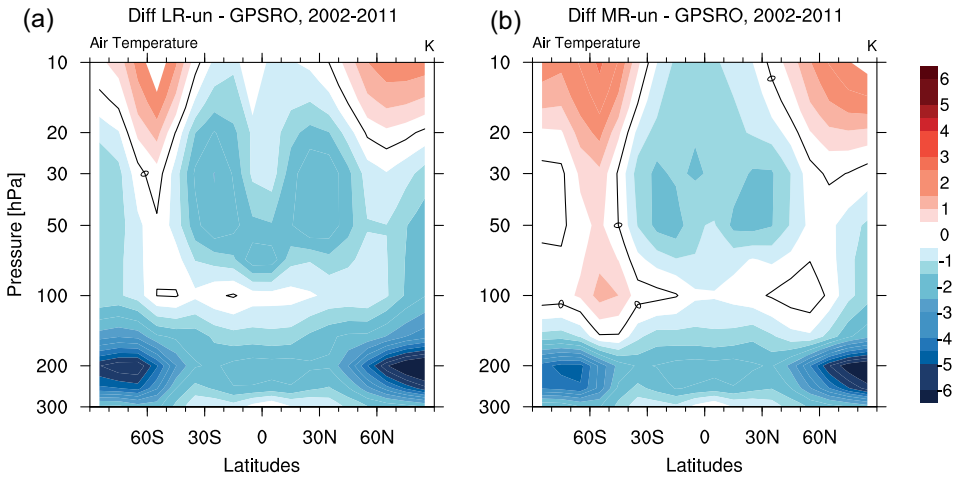

Figure 4: Annual zonal mean temperature differences between projections (uninitialized model runs) and GPSRO data for the time period 2002 to 2011 for low (a) and medium (b) resolution, contour interval: 1 K.

150 hPa. Also in the tropical lower stratosphere the temperature differences are smaller in the MR compared to the LR version.

For the initialization of the so-called baseline 0 (b0) hindcasts, anomalies are used (Smith et al., 2013) with sea surface temperatures (SSTs) and salinity anomaly fields taken from a stand-alone simulation with MPIOM that is forced with NCEP/NCAR atmospheric reanalysis data (Kalnay et al., 1996). Baseline 1 (b1) instead uses atmospheric fields from ECMWF ERA-Interim reanalysis and the oceanic anomaly fields from ECMWF ORAS4 (ocean reanalysis and derived ocean heat content) data (Balmaseda et al., 2013).

The following sets of decadal hindcasts are considered in this study for same time period (2002 to 2011) as the RO observations:

- b0-LR: baseline 0 at low resolution (3 ensemble members).
- b1-LR: baseline 1 at low resolution (10 ensemble members).
- b1-MR: baseline 1 at medium resolution (5 ensemble members).

We are primarily interested in assessing the hindcast quality in relation to lead-time which can be understood as the time-lag to the initialized year. Thus we resort data and build lead-year time series for the ensemble mean of each set of hindcasts. Resorting the decadal hindcasts implies the consideration of data starting with the initialized year 1993 and ending up with 2010 taking into account that b0-LR consists of 3 ensemble members until the initialized year 1999 and since 2000 of 10 ensemble members. b1-LR is composed of 10 ensemble members and b1-MR is averaged over 5 members.

3 Validation

3.1 Approach

As a deterministic metric for the assessment of the different hindcast experiments the mean-squared skill score (MSSS) according to Murphy (1988) is applied:

$$MSSS(P,R,O) = 1 - \frac{MSE(P,O)}{MSE(R,O)}, \tag{3.1}$$

that compares the mean-squared errors (MSE)

$$MSE(P,O) = \frac{1}{n}\sum_{j=1}^{n}(P_j - O_j)^2, \tag{3.2}$$

of two sets of predictions P and R with respect to the observations O. In Eq. (3.1) and Eq. (3.2) P_j is the prediction for a certain time step j, O_j the corresponding observation, and R the reference forecast represented by the uninitialized model runs (projections) or the ZAF based on GPSRO data. Both the P_j and O_j are anomalies relative to their respective climatologies.

As a deterministic score, the MSE is typically calculated from ensemble mean averages obtained from the resorted lead-year dependent time series from the different sets of ensemble hindcasts as described in the previous section. Besides showing MSSS for the single lead-year one, also averaged MSSS over lead-years two to five, and six to nine are given as suggested by Goddard et al. (2012).

3.2 Results

3.2.1 MSSS for lead-year one

By using the GPSRO dry temperatures as the observational basis, MSSS maps are calculated for the three different MPI-ESM decadal prediction ensemble hindcast sets on an annual zonal mean basis (Figure 5). By utilizing the ZAF as the reference forecast, skill scores for the uninitialized runs are generally negative, indicating that projection experiments typically used for the assessment of long-term climate changes do not provide forecast information about the year-to-year variability (Figure 5a,b). Notable exceptions are positive scores in

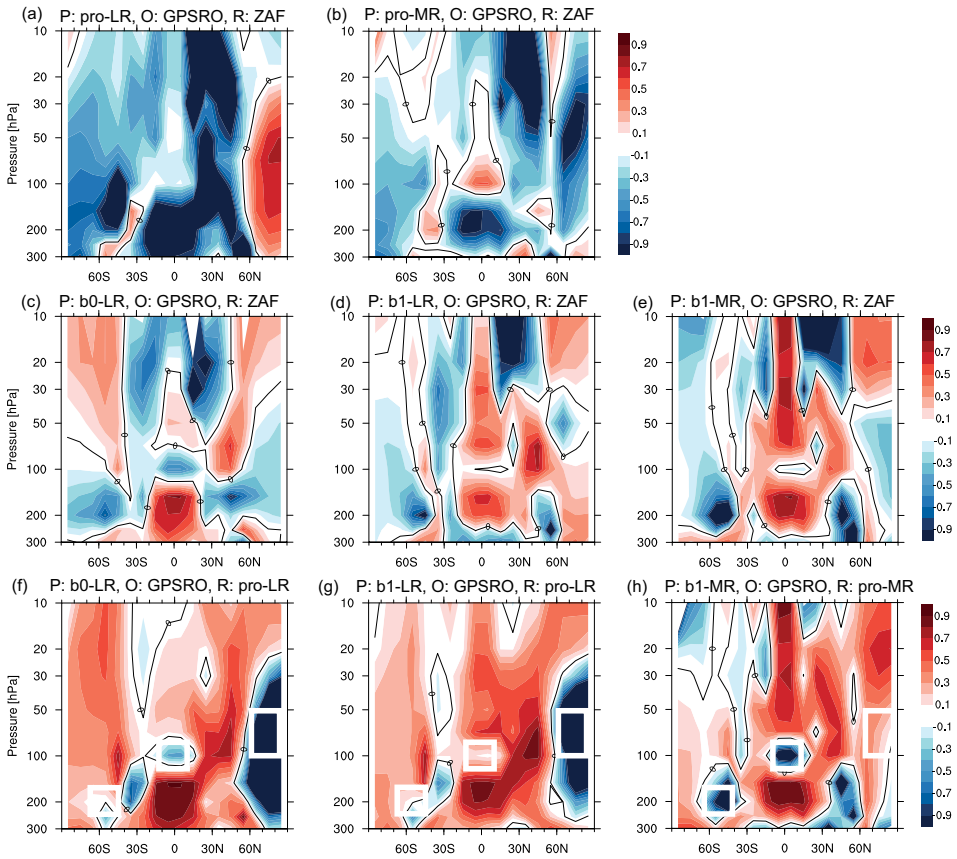

Figure 5: Mean squared skill score (MSSS) for three different hindcast experiments (Section 2.2) for **lead-year 1** based on the period 2002 to 2011. P, O, and R are according Eq. (3.1) with P: b0-LR (c, f), b1-LR (d, g), b1-MR (e, h) as the different hindcast experiments, O: GPSRO as the according observations, and R: pro-LR, pro-MR, and ZAF as the projections and zero anomaly forecast (2002 to 2011) used as reference forecasts. (a, b) MSSS of the projections (LR and MR) vs. ZAF; contour interval: 0.2 K. For the marked areas in f–h annual time series are discussed in Figure 8.

the projection of the LR model version (pro-LR) in the northern polar vortex in the lower stratosphere, as well as slightly positive scores in the upper subtropical troposphere and tropical tropopause region in the MR model version of MPI-ESM. We will return to those features later in this section.

Scores comparing initialized climate predictions of lead-year one with GPSRO observations suggest a positive impact of the initialization in several regions for all three hindcast experiments available (Figure 5c–e). Positive scores are consistently found in the upper tropical troposphere (100 hPa to 300 hPa) below the tropopause region, in the middle troposphere (200 hPa to 300 hPa) around mid-latitudes and in the equatorial lower stratosphere above 70 hPa. When comparing skill between b0-LR and b1-LR hindcasts, we note an increase in skill in the region surrounding the tropical tropopause all the way up into the lower and middle stratosphere, as well as in the subtropical lower stratosphere and the polar vortex at the NH. Decreasing skills are obtained in the stratospheric polar vortex of the SH and in the tropical upper troposphere below 250 hPa.

When comparing scores for b1-LR with b1-MR, we note substantially increased scores in the tropical and subtropical stratosphere that is probably related to the ability of the MR model to generate a spontaneous QBO signal. When initialized at the right phase (Scaife et al., 2014), this ability of the model contributes to the substantially higher scores. We further find increased scores in the surrounding of the tropical and subtropical tropopause on both hemispheres that benefit from smaller temperature biases of the MR version (see also Figure 4).

Similar conclusions are obtained from analyzing the skill scores for initialized predictions over the uninitialized projections for lead-year one (Figure 5f–h). Apart from the NH polar vortex in the LR model runs, the tropical and SH subtropical tropopause region, and upper tropical troposphere, skills are largely positive at all latitudes and all height levels. For b1-MR (Figure 5h), however, we note prominent negative scores in the areas above the tropospheric mid-latitude jets (~ 200 hPa) in both hemispheres and at the tropical tropopause (~ 100 hPa). At those positions, the ensemble mean of the MR uninitialized runs shows already slightly positive skill (Figure 5b) not reproduced after the initialization. The negative skill scores for the b1-MR variant above the mid-latitude jets on both hemispheres and in the vicinity of the tropical tropopause may demonstrate the difficulty to capture and model all relevant

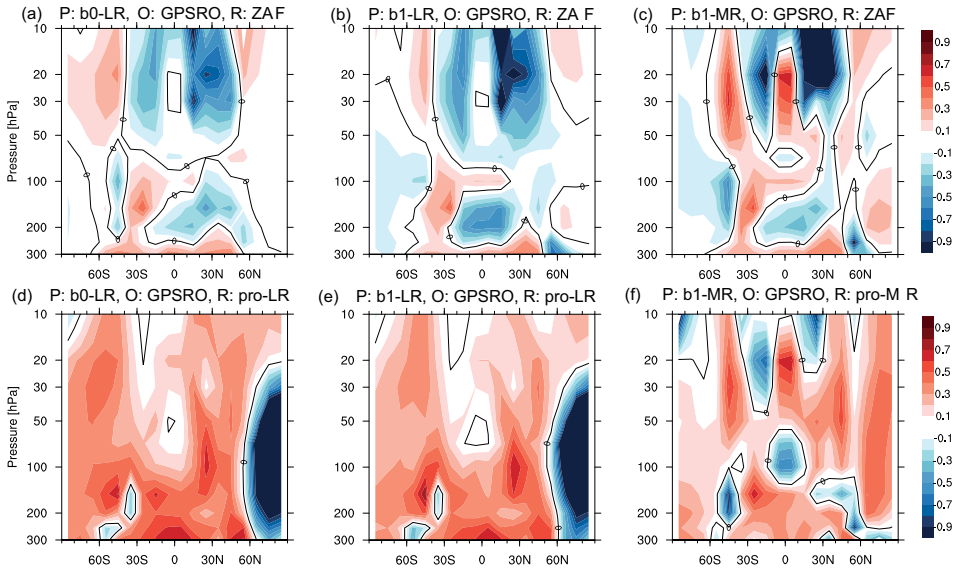

Figure 6: same as Figure 5c–h, but for **lead-years 2 to 5**.

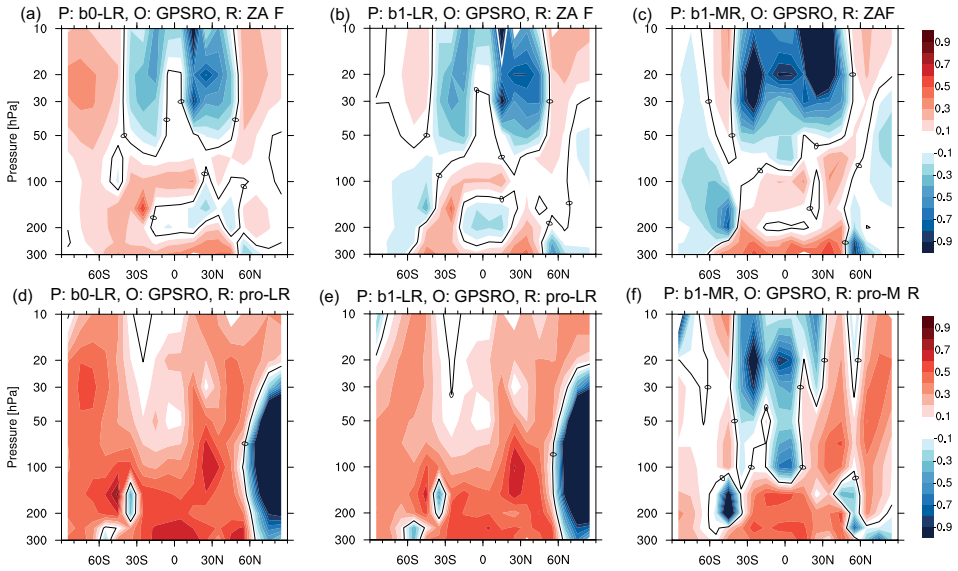

Figure 7: same as Figure 5c–h, but for **lead-years 6 to 9**.

processes responsible for the variability in that regions. Special attention is given to the tropical tropopause region (100 hPa to 70 hPa) where GPSRO data show a positive temperature trend since 2001 (Schmidt et al., 2010; Wang et al., 2013) that is not reproduced in the hindcast experiments (see also Figure 8a and the discussion below).

3.2.2 MSSS for longer lead-times

For longer lead-times as depicted for the average of lead-years two to five (Figure 6) and lead-years six to nine (Figure 7), scores become gradually smaller when compared to the corresponding scores for lead-year one, which is in particular apparent in the upper tropical troposphere. Moreover, we note declining differences between b0 and b1 hindcast experiments with the LR model version for the longer lead-times, which has been expected since the impact of initial conditions generally fades out with growing forecast time. For the MR hindcast experiment, we note decreased skill in the QBO regions for lead-years two to five, which even turns into no skill at all for the later years. More work is required to better understand forcing mechanisms and dynamics of the QBO and their representation in numerical coupled climate models.

3.2.3 Time series for selected regions

To discuss in more detail the variability in both, observations and model experiments, at certain locations mentioned above, we exemplarily present time-series for three selected regions (Figure 8). For convenience, results from the most recent ECMWF re-analysis ERA-Interim have been included into these figures as well.

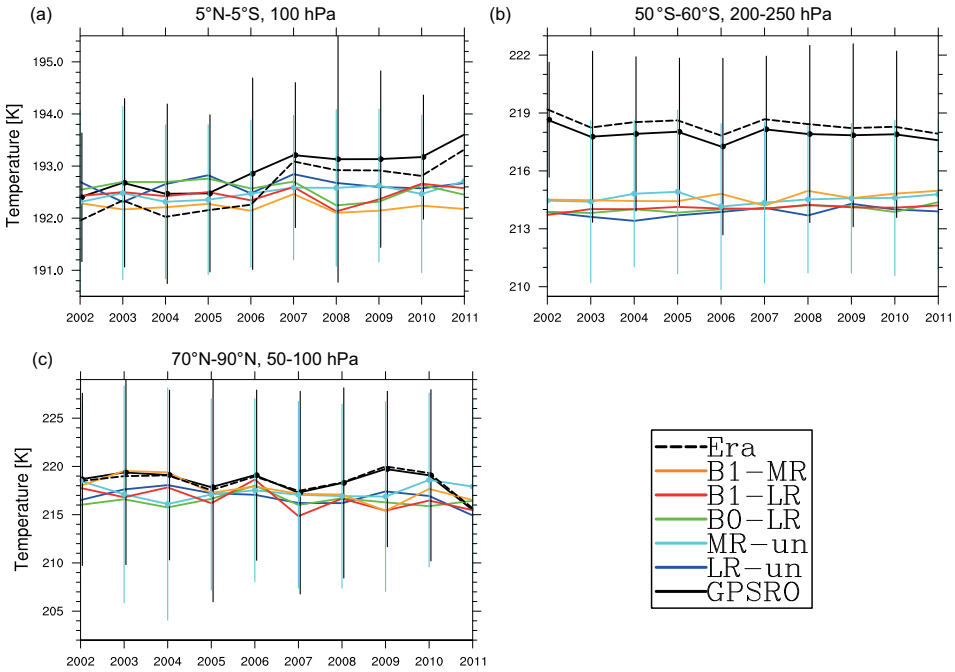

Figure 8: Annual mean temperature timeseries of the different hindcast experiments for **lead-year 1** for different latitude bands and the UTLS region in comparison to GPSRO observations and ERA-Interim data. For the GPSRO and the uninitialized MR model data the standard deviation is shown.

First, we concentrate on the tropical tropopause region (5 ° N to 5 ° S, 100 hPa; Figure 8a). For this region, both the GPSRO as well as ERA-Interim data show a slight temperature increase of about 1 K per decade that is neither captured by the initialized nor the uninitialized model runs. Since temperature variability apart from this trend is generally small in this region, skill scores are negative for all model experiments considered. For the upper tropospheric region 50 ° S to 60 ° S and 200 hPa to 250 hPa (Figure 8b), we note a large bias of 4 K between MPI-ESM model results and observations. This bias is slightly smaller in the MR version due to the substantially higher vertical resolution, but is nevertheless still large enough to make skill scores negative in this area in all experiments considered.

Finally, we concentrate on the NH stratospheric polar vortex region (90 ° N to 70 ° N, 50 hPa to 100 hPa; Figure 8c) which is characterized by a large interannual variability (Figure 3b). Large differences are apparent here between both the uninitialized runs of LR and MR, and also the different initialized hindcasts. A cold bias of the model is still apparent, but since internal variability is much larger in this region, it does not dominate the skill scores. Instead, positive temperature anomalies related to sudden stratospheric warmings in the years 2002, 2003, 2005, and 2006, which are well predicted by the b1-MR version contribute to the high skill scores obtained here. The b1-MR model version fails to predict the stratospheric warmings in 2008, 2009 and 2010. In particular the warming in 2009 was unexpected with respect to the prevailing stratospheric conditions (Labitzke and Kunze, 2012). b1-LR also predicts peak temperatures in 2006 but the level of correspondence is much lower than for the MR version. Since variability in the later years are not captured properly by the ensemble mean of the hindcasts from the LR version, average skill scores remain small. It is interesting to note that the uninitialized ensemble mean of the LR model version reproduces the dominating low-frequency variability generally well, which leads to the exceptionally high skill scores of the LR projection in that region as mentioned earlier.

4 Conclusions and outlook

The GPSRO technique is a well-established method to derive atmospheric temperatures in the UTLS region. The RO data of several missions are assimilated at leading weather service centers since 2006, thereby contributing to the increased forecast accuracy on a daily basis. Since GPSRO data from different receiving satellites do not show systematic differences, and since future missions carrying GPSRO receivers as primary or secondary payloads are currently prepared for launch (GRACE-follow on in 2017; COSMIC-2 in 2016 and 2018), we expect that GPSRO will continue to provide long-term stable atmospheric observations, which have the potential to serve as benchmark data for the detection of climate change signals (Steiner et al., 2013).

Therefore it is straightforward to use this dataset also for the quality assessment of decadal hindcast experiments. In this contribution, we make use of one decade of UTLS temperatures from CHAMP, GRACE,

and TerraSAR-X processed at GFZ Potsdam that are applied for the validation of three different decadal hindcast experiments with the MPI-ESM coupled climate model following the validation framework of Goddard et al. (2012).

Based on the deterministic scores presented above we conclude that the initialization of decadal hindcast runs largely increases the skill of atmospheric temperatures when compared to uninitialized climate projections. Those skills are largest and close to one for lead-year one, and gradually decrease for the later lead-years, as expected. While focussing on lead-year one, we note increased skills in particular for the upper tropical troposphere when moving from b0-LR to the b1-LR experiment. We therefore conclude that - from the GPSRO point of view - the changes applied to the initialization strategy after b0 lead to better predictions of the upper air temperature field in several regions of the world (lower stratosphere in the tropics and subtropics and tropical tropopause region; Figure 5f,g).

None of the MPI-ESM experiments is able to reproduce the observed positive trend in tropical tropopause temperatures since 2001.

For the b1-MR hindcasts, we identify increased skills in the tropical stratosphere, which is related to the ability of the model to generate a spontaneous QBO. The b1-MR version of the model is the only one reproducing the high-latitude NH interannual variability at least before 2006 and is therefore a good choice forecasting stratospheric dynamics. For other regions, mainly in the vicinity of the tropospheric mid-latitude jets on both hemispheres and the tropical tropopause improvements for the b1-MR variant are necessary.

The use of (dry) RO temperatures restricts the atmospheric temperature validation to altitudes where water vapor is insignificant, i.e. to the upper troposphere and lower stratosphere. To overcome this restriction, validation and assessment of model parameters should focus on refractivity and bending angles (Ringer and Healy, 2008), which are the most fundamental observable of the RO technique. To facilitate routine validation based on bending angles the MiKlip central evaluation system should be extended to derive atmospheric refractivity and bending angles out of simulated temperature and humidity fields. By means of refractivity and bending angles, the assessment of the decadal hindcast experiments with GPSRO data will be then also realistically possible in the lower troposphere.

Acknowledgments

The GEOCLIM study has been supported by the German Federal Ministry of Education and Research (BMBF) within the FONA research program under grants 03F0654A and 01LP1151A.

References

Alexander, P., A. de la Torre, P. Llamedo, R. Hierro, 2014: Precision estimation in temperature and refractivity profiles retrieved by GPS radio occultations. – J. Geophys. Res., published online, DOI: 10.1002/2013JD021016.

Anthes, R.A., D. Ector, D.C. Hunt, Y-H. Kuo, C. Rocken, W.S. Schreiner, S.V. Sokolovskiy, S. Syndergaard, T-K. Wee, Z. Zeng, P.A. Bernhardt, K.F. Dymond, Y. Chen, H. Liu, K. Manning, W.J. Randel, K.E. Trenberth, L. Cucurull, S.B. Healy, S-P. Ho, C. McCormick, T.K. Meehan, D.C. Thompson, N.L. Yen, 2008: The COSMIC/FORMOSAT-3 mission: Early results. – Bull. Amer. Meteor. Soc. **89**, 1–21, DOI: 10.1175/BAMS-89-3-313.

Balmaseda, M.A., K. Mogensen, A.T. Weaver, 2013: Evaluation of the ECMWF ocean reanalysis system ORAS4. – Quart. J. Roy. Meteor. Soc. **139**, 1132–1161, DOI: 10.1002/qj.2063.

Cardinali, C., S.B. Healy, 2014: Impact of GPS radio occultation measurements in the ECMWF system using adjoint-based diagnostics. – Quart. J. Roy. Meteor. Soc. **140**, 2315–2320, DOI: 10.1002/qj.2300.

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. Holm, L. Isaksen, P. Kållberg, M. Köhler, M. Matricardi, A.P. McNally, B.M. Monge–Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thépaut, F. Vitart, 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. – Quart. J. Roy. Meteor. Soc. **137**, 553–597, DOI: 10.1002/qj.828.

Foelsche, U., M. Borsche, A.K. Steiner, A. Gobiet, B. Pirscher, G. Kirchengast, J. Wickert, T. Schmidt, 2007: Observing upper troposphere-lower stratosphere climate with radio occultation data from the CHAMP satellite. – Climate Dyn. **31**, 49–64, DOI: 10.1007/s00382-007-0337-7.

Fueglistaler, S., A.E. Dessler, T.J. Dunkerton, I. Folkins, Q. Fu, P.W. Mote,2009: Tropical tropopause layer. – Rev. Geophys., published online, DOI: 10.1029/2008RG000267.

GCOS, 2006: Systematic observation requirements for satellite-based products for climate. – GCOS-107, (WMO/TD No. 1338).

Gettelman, A., P. Hoor, L.L. Pan, W.J. Randel, M.I. Hegglin, T. Birner, 2011: The extratropical upper troposphere and lower stratosphere. – Rev. Geophys., published online, DOI: 10.1029/2011RG000355.

Goddard, L., A. Kumar, A. Solomon, D. Smith, G. Boer, P. Gonzalez, V. Kharin, W. Merryfield, C. Deser, S.J. Mason, B.P. Kirtman, R. Msadek, R. Sutton, E. Hawkins, T. Fricker, G. Hegerl, C.A.T. Ferro, D.B. Stephenson, G.A. Meehl, T. Stockdale, R. Burgman, A.M. Greene, Y. Kushnir, M. Newman, J. Carton, I. Fukumori, T. Delworth, 2012: A verification framework for interannual-to-decadal predictions experiments. – Climate Dyn. **40**, 245–272, DOI: 10.1007/s00382-012-1481-2.

Hajj, G.A., C.O.Ao, B.A. Iijima, D. Kuang, E.R. Kursinski, A.J. Mannucci, T.K. Meehan, L.J. Romans, M. de la Torre Juarez, T.P. Yunck, 2004: CHAMP and SAC-C atmospheric occultation results and intercomparisons. – J. Geophys. Res. **109**, published online, DOI: 10.1029/2003JD003909.

Jungclaus, J.H., N. Fischer, H. Haak, K. Lohmann, J. Marotzke, D. Matei, U. Mikolajewicz, D. Notz, 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., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, R. Reynolds, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K.C. Mo, C. Ropelewski, J. Wang, R. Jenne, D. Joseph, 1996: The NCEP/NCAR 40-year reanalysis project. – Bull. Amer. Meteor. Soc. **77**, 437–471, DOI: 10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

König, R., S. Zhu, C. Reigber, K.-H. Neumayer, H. Meixner, R. Galas, G. Baustert, P. Schwintzer, 2002: CHAMP Rapid Orbit Determination for GPS Atmospheric Limb sounding. – Adv. Space Res. **30**, 289–293, DOI: 10.1016/S0273-1177(02)00297-1.

Kuo, Y.-H., W.S. Schreiner, J. Wang, D.L. Rossiter, Y. Zhang, 2005: Comparison of GPS radio occultation soundings with radiosondes. – Geophys. Res. Lett. **32**, published online, DOI: 10.1029/2004GL021443.

Kursinski, E.R., G.A. Hajj, J.T. Schofield, R.P. Linfield, K.R. Hardy, 1997: Observing Earth's atmosphere with radio occultation measurements using the Global Positioning System. – J. Geophys. Res. **102**, 23429–23465, DOI: 10.1029/97JD01569.

Labitzke, K., M. Kunze, 2012: In: Climate And Weather of the Sun-Earth System (CAWSES): Highlights from a priority program, F.-J. Lübken (Ed.). – Springer, Dordrecht, The Netherlands, ISBN:978-94-007-4348-9.

Leroy, S.S., J.G. Anderson, J.A. Dykema, 2006: Testing climate models using GPS radio occultation: A sensitivity analysis. – J. Geophys. Res. **111**, published online, DOI: 10.1029/2005JD006145.

Meehl, G.A., L. Goddard, J. Murphy, R.J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M.A. Giorgetta, A.M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer, T. Stockdale, 2009: Decadal Prediction. – Bull. Amer. Meteor. Soc. **90**, 1467–1485, DOI: 10.1175/2009BAMS2778.1.

Müller, W.A., J. Baehr, H. Haak, J.H. Jungclaus, J. Kröger, D. Matei, D. Notz, H. Pohlmann, J.S. von Storch, 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**, published online, DOI: 10.1029/2012GL053326.

Murphy, A.H., 1988: Skill scores based on mean square error and their relationship to the correlation coefficent. – Mon. Weath. Rev. **116**, 2417–2424, DOI: 10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2.

Pohlmann, H., W.A. Müller, K. Kulkarni, M. Kameswarrao, D. Matei, F.S.E. Vamborg, C. Kadow, S. Illing, 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.

Randel, W.J., K.P. Shine, J. Austin, J. Barnett, C. Claud, N.P. Gillett, P. Keckhut, U. Langematz, R. Lin, C. Long, C. Mears, A. Miller, J. Nash, D.J. Seidel, D.W.J. Thompson, F. Wu, S. Yoden, 2009: An update of observed stratospheric temperature trends. – J. Geophys. Res. **112**, published online, DOI: 10.1029/2008JD010421.

Ringer, M.A., S.B. Healy, 2008: Monitoring twenty-first century climate using GPS radio occultation bending angles. – Geophys. Res. Lett. **35**, published online, DOI: 10.1029/2007GL032462.

Scaife, A.A., M. Athanassiadou, M. Andrews, A. Arribas, M. Baldwin, N. Dunstone, J. Knight, C. MacLachlan, E. Manzini, W.A. Müller, H. Pohlmann, D. Smith, T. Stockdale, A. Williams, 2014: Predictability of the quasi-biennial oscillation and its northern winter teleconnection on seasonal to decadal timescales. – Geophys. Res. Lett. **41**, 1752–1758, DOI: 10.1002/2013GL059160.

Schmidt, T., J. Wickert, G. Beyerle, R. König, R. Galas, C. Reigber, 2005: The CHAMP atmospheric processing system for radio occultation measurements. – In: Reigber C., H. Lühr H., P. Schwintzer, J. Wickert (Eds): Earth Observation with CHAMP, Springer, Berlin, ISBN 3-540-22804-7, 597–602.

Schmidt, T., J. Wickert, A. Haser, 2010: Variability of the upper troposphere and lower stratosphere observed with GPS radio occultation bending angles and temperatures. – Adv. Space Res. **46**, 150–161, DOI: 10.1016/j.asr.2010.01.021.

Smith, E., S. Weintraub, 1953: The constants in the equation for atmospheric refractive index at radio frequencies. – Proceedings of the I.R.E. **41**, 1035–1037, DOI: 10.1109/JRPROC.1953.274297.

Smith, D.M., S. Cusack, A.W. Colman, C.K. Folland, G.R. Harris, 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, H. Pohlmann, 2013: A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. – Climate Dyn. **41**, 3325–3338, DOI: 10.1007/s00382-013-1683-2.

Steiner, A.K., G. Kirchengast, M. Borsche, U. Foelsche, T. Schoengassner, 2007: A multi-year comparison of lower stratospheric temperatures from CHAMP radio occultation data with MSU/AMSU records. – J. Geophys. Res. **112**, published online, DOI: 10.1029/2006JD008283.

Steiner, A.K., D. Hunt, S.-P. Ho, G. Kirchengast, A.J. Mannucci, B. Scherllin-Pirscher, H. Gleisner, A. von Engeln, T. Schmidt, C. Ao, S.S. Leroy, E.R. Kursinski, U. Foelsche, M. Gorbunov, S. Heise, Y.-H. Kuo, K.B. Lauritsen, C. Marquardt, C. Rocken, W. Schreiner, S. Sokolovskiy, S. Syndergaard, J. Wickert, 2013: Quantification of structural uncertainty in climate data records from GPS radio occultation. – Atmos. Chem. Phys. **13**, 1469–1484, DOI: 10.5194/acp-13-1469-2013.

Taylor, K.E., R.J. Stouffer, G.A. Meehl, 2012: An Overview of CMIP5 and the Experiment Design. – Bull. Amer. Meteor. Soc. **93**, 485–498, DOI: 10.1175/BAMS-D-11-00094.1.

Thompson, D.W.J., D.J. Seidel, W.J. Randel, C.-Z. Zou, A.H. Butler, C. Mears, A. Osso, C. Long, R. Lin, 2012: The mystery of recent stratospheric temperature trends. – Nature, **491**, 692–697, DOI: 10.1038/nature11579.

Wang, D.Y., G.P. Stiller, T. von Clarmann, H. Fischer, M. López-Puertas, B. Funke, N. Glatthor, U. Grabowski, M. Höpfner, S. Kellmann, M. Kiefer, A. Linden, G.M. Tsidu, M. Milz, T. Steck, J.H. Jiang, C.O. Ao, G. Manney, K. Hocke, D.L. Wu, L.J. Romans, J. Wickert, T. Schmidt, 2004: Cross-validation of MIPAS/ENVISAT and GPS-RO/CHAMP temperature profiles. – J. Geophys. Res. **109**, published online, DOI: 10.1029/2004JD004963.

Wang, W., K. Matthes, T. Schmidt, L. Neef, 2013: Recent variability of the tropical tropopause inversion layer. – Geophys. Res. Lett. **40**, published online, DOI: 10.1002/2013GL058350.

Wickert, J., C. Reigber, G. Beyerle, R. König, C. Marquardt, T. Schmidt, L. Grunwaldt, R. Galas, T.K. Meehan, W.G. Melbourne, K. Hocke, 2001: Atmosphere sounding by GPS radio occultation: First results from CHAMP. – Geophys. Res. Lett. **28**, 3263–3266, DOI: 10.1029/2001GL013117.

Wickert, J., G. Michalak, T. Schmidt, G. Beyerle, C.-Z. Cheng, S.B. Healy, S. Heise, C.-Y. Huang, N. Jakowski, W. Köhler, C. Mayer, D. Offiler, E. Ozawa, A.G. Pavelyev, M. Rothacher, B. Tapley, C. Arras, 2009: GPS radio occultation: Results from CHAMP, GRACE and FORMOSAT-3/COSMIC. – Terr. Atmos. Ocean. Sci. **20**, 35–50, DOI: 10.3319/TAO.2007.12.26.01(F3C).