



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

A priori selection and data-based skill assessment of reanalysis data as predictors for daily air temperature on a glaciated, tropical mountain range

M. Hofer¹, B. Marzeion¹, and T. Mölg²

¹Institute of Meteorology and Geophysics, University of Innsbruck, Innsbruck, Austria

²Chair of Climatology, Technische Universität Berlin, Berlin, Germany

Received: 12 February 2012 – Accepted: 3 April 2013 – Published: 17 May 2013

Correspondence to: M. Hofer (marlis.hofer@uibk.ac.at)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

This study presents an empirical-statistical downscaling (ESD) method for high-altitude, glaciated mountain sites. In the ESD model validation emphasis is put on appropriately considering the pitfalls of small observational data records that are typical of high mountains. An application example is shown with daily mean air temperature time series on a glaciated mountain range (Cordillera Blanca, Peru) as target variables, and an ensemble of reanalyses air temperature time series as “a priori” predictor (i.e. a predictor selected without preceding data analysis). Results reveal strong seasonal variations of the predictor’s performance. With increasing data availability, the skill tends to increase. Similarly for lower temporal resolutions, the skill increases. Applied to a choice of different atmospheric reanalysis predictor variables, the ESD model identifies only air temperature and geopotential height as significant predictors with regard to local-scale air temperature variability. Accounting for natural periodicity in the data is vital in the ESD procedure to avoid spuriously high performances of certain predictors, which is demonstrated for 2 m air temperature versus air temperature in the pressure level close to the mountain station site.

1 Introduction

Ongoing developments in atmospheric modelling have made available choices of long-term, temporally high-resolution atmospheric data sets for the entire globe. These data, however, are still restricted in terms of spatial resolutions, such that their immediate application to study regional and local climate is not recommended. Especially over complex topography, such as glacier-covered mountains, atmospheric models often miss significant processes that characterize local weather and climate. So-called downscaling methods bridge this gap between the available data from global atmospheric models and the required local-scale information (for an overview see, e.g. Christensen et al., 2007). Generally two types of downscaling exist, namely dynamical downscaling

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(e.g. Hill, 1968; Giorgi and Bates, 1989; Mearns et al., 2003), and empirical-statistical downscaling (ESD; e.g. Klein et al., 1959; Wilby et al., 2004; Benestad et al., 2008). Since the early development of both downscaling classes a variety of different models and approaches has emerged.

5 An important step in general ESD procedures is the selection and assessment of global atmospheric model data for the input to the downscaling model (e.g. Von Storch, 1999; Wilby et al., 2004; Benestad et al., 2008). Given the increasing availability of atmospheric models and variables, the issue of predictor selection and assessment has become even more intricate over the last decades, concerning the choice of (i) the
10 physical variable type, (ii) the model grid points or spatial area (i.e. the downscaling domain), and (iii) of the data source (i.e. the type of global model). However, only few studies (e.g. Winkler et al., 1997; Cavazos and Hewitson, 2005) have systematically assessed the relevance of different predictors (in terms of variable types, spatial area, or predictor model). In fact, there is little consensus on the most appropriate choice (e.g.
15 Von Storch, 1999; Fowler et al., 2007), since it depends upon various factors (such as predictand variable, spatial and time scales, season, as well as geographical location). Wilby et al. (2002) propose a promising solution by providing regression-based, automated tools for predictor selection in ESD (see also Wilby and Dawson, 2007; Hesami et al., 2008). Yet these methods are suitable only if the observational data base
20 for model calibration is relatively large, i.e. daily time series for several decades. Beyond the field of atmospheric sciences the problem of predictor and model selection is also well known (e.g. Zucchini, 2000; Hastie et al., 2001); e.g. Bair et al. (2006) present an interesting avenue based on supervised principal components.

25 This study presents an ESD method for high-altitude, mountainous sites. The ESD method is designed (i) to be applicable when only short observational time series are available for model calibration (i.e. few years), (ii) to appropriately consider autocorrelation in the high-resolution time series, and (iii) to provide a solid tool for model (or predictor) assessment and selection by avoiding subjective choices. The ESD method, as presented here applicable to Gaussian variables only, is comprehensible and of

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

minimum complexity, in order to be easily transferred to different sites, predictors, or predictands. We show an application of the ESD method to quantify the skill of reanalysis data as predictors for local-scale, daily air temperature measured at high-altitude automatic weather stations (AWSs) in the tropical Cordillera Blanca, where only a few years of high-resolution air temperature measurements are available. Section 2 introduces the study site and observational data used in this study. Section 3 presents reanalysis data that are used as the predictors in this study. Section 4 gives a comprehensive description of the ESD model. The results of the ESD model application are discussed in Sect. 5 and summarized in Sect. 6.

2 Study site and observations: the predictands

The investigation site of the present study is the Cordillera Blanca, a glaciated mountain range located in the Northern Andes of Peru (Fig. 1) that harbors 25 % of all tropical glaciers (with respect to surface area; Kaser and Osmaston, 2002). Glaciers in the Cordillera Blanca have been shrinking since their last maximum extent in the late 19th century (e.g. Ames, 1998; Silverio and Jaquet, 2005; Georges, 2004) and have significantly shaped the socio-economic development in the region. During the 20th century, a series of the history's most catastrophic glacier disasters – i.e. outburst floods and avalanches – occurred (e.g. Carey, 2005, 2010). But Cordillera Blanca glaciers also have important positive impacts for water availability in industry, agriculture and households because they contribute to balancing the high runoff seasonality in the extensively populated Rio Santa valley (Juen, 2006; Juen et al., 2007; Mark and Seltzer, 2003; Kaser et al., 2003, 2010).

Located in the the outer tropical climate zone, atmospheric seasonality in the Cordillera Blanca is mainly characterized by precipitation variance, with the seasonal air temperature variance being small (Niedertscheider, 1990; Kaser and Osmaston, 2002; Georges, 2005; Juen, 2006). More than 50 % of the annual precipitation falls during the humid season (January–March), whereas during the dry season

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(June–August), less than 2% of the annual precipitation falls (annual precipitation amounts 770 mm in the Northern, and 470 mm in the Southern Cordillera Blanca; Niedertscheider, 1990). A detailed description about the underlying mechanisms is given in the work by Garreaud et al. (2003).

5 Since 1999, an observational network of several AWSs at and nearby glaciers in the Cordillera Blanca has been installed, primarily to provide high-resolution data for glacier mass balance and runoff modeling (Juen et al., 2007). Maintaining the AWSs to provide continuous and reliable atmospheric time series has represented a logistical and technical challenge. Field work has been costly in terms of time and materials, since the AWSs are located at very high altitudes (between 4700 and 5100 m a.s.l.) in remote areas. Further problems also include instrument theft and natural hazards (Juen, 2006). Thus ESD methods have been required that are able to provide reliable results also on the basis of limited measurement availability. Hofer et al. (2010) present a comprehensive ESD modeling procedure to investigate if the short-term AWS time series (air temperature and specific humidity) can be extended into the past, using reanalysis data as predictors. They find that the ESD model skill largely varies as function of season and daytime, and emphasize uncertainty in the exact choice of variables that constitute the mixed-field predictors, upon which the model results show large sensitivity. Hofer et al. (2012) use a simpler methodology, based on single linear regression, in order to determine the best reanalysis product for daily air temperature predictands measured in the Cordillera Blanca. In the present study, we present an ESD methodology which is similar in terms of complexity to the one used by Hofer et al. (2012), with more emphasis on important elements concerning the ESD model configuration and application (for example, of how many measured data – in terms of sample size – are needed exactly for the assessment to be significant, and how the targeted temporal resolution affects the model skill).

25 The target variables here are daily air temperature time series measured at two AWSs located in the Northern Cordillera Blanca (hereafter referred to as AWS1 and AWS2) and one AWS located in the Southern Cordillera Blanca (hereafter AWS3),

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



which include the longest time series available from all installed AWSs to date. Yet the measuring periods are still relatively short, ranging from July 2006 to July 2012 (AWS1), to August 2011 (AWS2), and to December 2009 (AWS3), with three months of missing data at AWS2. AWS1, AWS2, and AWS3 are situated at 5050, 4825, and 4950 m a.s.l., respectively, on rocky terrain (glacial polish and moraines) in the vicinity of retreating glacier tongues. Whereas AWS1 and AWS2 are located very nearby each other in the Paron valley (only about two km distance), AWS3 is located approximately 100 km more southwards in the Shallap valley. The sites are indicated in Fig. 1. In technical terms, the measurements are carried out with a HMP45 sensor by Väisälä and a ventilated radiation shield, described by Georges (2002).

Figure 2 shows statistics of AWS1, AWS2, and AWS3 daily mean air temperature for each month of the year (daily means are calculated from hourly samples measured at the AWSs), over the period for which data are available at all three AWSs (July 2006–December 2009, hereafter period 1). The air temperatures are approximately normally distributed (not shown). The seasonal cycles in the data are small ($< 2^{\circ}\text{C}$), showing multiple local minima and maxima throughout the year. The warmest months are November–January and April–May, and the coldest months March and July. Note, however, that these statistics (in particular the occurrence of multiple maxima and minima) should not be overvalued as a climatology, because they are based on only four years of measurements. The lowest-elevation AWS2 systematically shows overall slightly higher air temperatures, with the warmest months above 2°C , whereas at AWS1 and AWS2 monthly mean air temperatures are below 2°C throughout the year. At all AWSs, monthly mean air temperatures are above 0°C in all months, and daily means are above 0°C in 75 % of all cases. The interquartile ranges (blue bars in Fig. 2) show within-month daily mean air temperature variations of less than 2°C in 50 % of all cases at all AWSs. The highest within-month variabilities occur from December to January, which points to El Niño Southern Oscillation (ENSO) variability playing an important role in the region at this time of the year (e.g. Vuille et al., 2008b), whereas

variabilities are generally lower for the dry season months June–July. Also shown in Fig. 2 are statistics of the reanalysis data predictor that will be referred to later.

3 Reanalysis data: the predictors

In this study, we assess reanalysis data as the predictors for daily mean air temperature measured at AWS1, AWS2 and AWS3. Reanalysis data are a combination of general circulation model (GCM) “first guess” and quality-controlled observations, generated using a data assimilation system, similar to analysis data in numerical weather prediction (NWP). First proposed in the studies of Bengtsson and Shukla (1988), and Trenberth and Olson (1988) “re”-analyses have the advantage over NWP analyses that their production is based on a fixed modeling system for the entire assimilation period. Thus data discontinuities due to changes in atmospheric model and assimilation techniques are avoided. Today, global reanalysis data are available from four institutions worldwide (in cooperation with partner institutions not mentioned here for brevity): the NCEP (Kalnay et al., 1996), the European Centre for Medium-range Weather Forecasts (ECMWF; Uppala et al., 2005), the Japan Meteorological Agency (JMA; Kazutoshi et al., 2007), and the National Aeronautics and Space Administration (NASA; Bosilovich, 2008).

First-generation NCEP reanalyses have been a frequent choice in climate studies about the Cordillera Blanca and the South American Andes (e.g. Garreaud et al., 2003; Vuille et al., 2008a,b; Hofer et al., 2010). Hofer et al. (2012), however, show for the Cordillera Blanca that the interim reanalyses by the ECMWF, the MERRA (the Modern Era Retrospective-Analysis for Research and Applications from NASA), the NCEP Climate Forecast System Reanalysis, CFSR (the latest reanalysis product by the NCEP), as well as ensembles thereof, show considerably higher skill than the first generation NCEP reanalyses, or JMA reanalyses. In this study, we use ensembles constructed by the interim, the CFSR, and the MERRA, as described later in Sect. 4.2, as well as the interim reanalyses (who showed the overall highest performance in Hofer et al., 2012) as the predictors.

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



4 ESD model architecture

4.1 Accounting for seasonal periodicity

If atmospheric time series are considerably shorter than thirty years and the climatological seasonal cycle is not known, the problem arises how to strictly distinguish periodic, seasonal variations from aperiodic (or less periodic), day-to-day and inter-annual variability. Especially in statistical forecasting, periodicity must be accounted for to avoid that the periodic, seasonal variations dominate the model fit. When long enough data series are available, the problem is often avoided by subtracting the climatological seasonal cycle from the time series (e.g. Madden, 1976). This way seasonal periodicity is removed from the time series, but not necessarily from the model error.

In the present study we assume that seasonal atmospheric periodicity leads to changing relationships between large- and local-scale atmospheric variables throughout the year. Considering the atmospheric seasonal cycle in ESD models is important especially if the study site is located in the mountains. For example, local-scale atmospheric conditions can be affected by topographic shading that changes with the solar altitude throughout the year, but the topography is misrepresented and thus these effects can not be captured by the large-scale model. Due to the same effects related to the diurnal cycle (sub-daily data) different models for the different times of day are required (e.g. Hofer et al., 2010). By consequently using separate statistical predictor-predictand transfer functions for the different months of the year, seasonal periodicity is eliminated not only in the time series, but also in the model error. In practice in this study, each predictor-predictand pair is divided into twelve separate time series for each month, the number of observations in each time series consequently being approximately $n = N/12$, where N is the length of the complete data series. Then the modeling procedure is repeated identically for each calendar month's time series.

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



4.2 Predictor selection

In order to simplify the transference of our approach to different cases (in terms of target variables, locations, or model), we distinguish two ways of predictor selection, namely (1) a priori predictor selection, and (2) data-based predictor selection. More precisely, (1) means predictor selection based on knowledge outside the data – e.g. about physical mechanisms – that is available without (“prior” to) looking at the data or data analysis. Most downscaling studies more or less systematically use a combination of (1) and (2), by first pre-selecting a subset of potential predictors from an available pool based on process knowledge, and then choosing the definite, final predictors based on criteria derived from the data (i.e. data-based selection) (e.g. Klein and Glahn, 1974; Wilby et al., 2002). Yet, it is difficult to generalize data-based findings for different cases (in terms of variables, sites or models), and it is desirable to find objective, a priori criteria, that simplify predictor selection. Hofer et al. (2012), for example, use an a priori predictor for the assessment of different reanalysis models with regard to local, daily air temperature variations in the Cordillera Blanca.

What information of a large-scale atmospheric model would we use to represent local, daily air temperature, if no observations were available? The most intuitive choice is to relate the same physical predictor and target variables; thus here, to use large-scale air temperature as predictor for local-scale air temperature. Previous ESD studies focusing on air temperature have suggested air temperature predictors in combination with sea level pressure (e.g. Benestad et al., 2002), with geopotential height (e.g. Kidson and Thompson, 1998), or with zonal wind speed and specific humidity (e.g. Hofer et al., 2010). Similarly, Huth (2004) underline the necessity to use a combination of both circulation-based, and radiation-based predictors for air temperature. Von Storch (1999) recommends the use of air temperature predictors as surrogate indicator of atmospheric radiative properties that are not captured by circulation predictors alone. Beyond these recommendations, though, the definite choice of concrete variables (as well as horizontal extents, vertical levels or model) in every single case require data-based

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



assessments. Even if, by selecting air temperature as single a priori predictor here, we neglect any circulation-induced variations not included in the predictor air temperature, our choice is useful especially for inter-comparisons of different models, because it is reasonable to assume that the best model also shows the highest skill in representing the same variable.

As mentioned above, predictor selection includes not only the choice of a physical variable type, but also of geographical allocation in terms of model grid points. Since model topographies are smoothed representations of the real topography, the surface height at a particular location in the model does generally not correspond to the real surface height at that location. Therefore the question arises whether (1) surface or near surface predictors – to account for important surface processes, or (2) upper air predictors that are located at the same elevation as the predictands, are the more realistic choice for a predictand located at the surface. In the application example of this study, all three AWSs are located between 500 and 600 hPa, at 5050 (AWS1), 4825 (AWS2), and 4950 (AWS3) m a.s.l. For comparison, Table 1 shows coordinates, surface elevations (h), and geopotential heights (gph) for the 550 hPa level of the closest (relative to the study site) grid points in the interim, CFSR and MERRA models. The grid points are all located between 3000 and 3500 m a.s.l., thus about 1500 m lower than the AWS sites in reality. There is only a small difference between the geopotential heights of the 550 hPa levels of the different reanalyses, all located at about 5100 m a.s.l.

In terms of horizontal, or spatial predictor domain, ESD studies suggest that the optimum downscaling domain is generally not limited to the closest grid points around the study site, but includes important synoptic patterns around and upstream of the study area (e.g. Benestad et al., 2008). For studies that use grid point predictors, it is generally recommended to not using single, but rather ensembles of grid points as predictors. Grid point averaging of atmospheric models is necessary in order to minimize numerical model errors apparent in single grid point data (e.g. Grotch and MacCracken, 1991; Williamson and Laprise, 2000; Räisänen and Ylhäisi, 2011). Due to these numerical

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



errors, the minimum scale of a model (i.e. the distance of two neighbouring grid points) can not be regarded as the skillful scale of the same model (e.g. Von Storch et al., 1993; Zorita and Von Storch, 1997; Benestad et al., 2008; Hofer et al., 2012). For the Cordillera Blanca, Hofer et al. (2012) analyze the optimum spatial averaging scales of the interim, MERRA, CFSR, JCDAS, and the first NCEP/NCAR reanalyses. They find the optimum spatial averaging domains of the different reanalyses largely varying (from eight grid points for the NCEP/NCAR reanalyses, to 507 grid points of the modern CFSR reanalyses), with the minimum scales being only weak indicators of the optimum scales. Hofer et al. (2012) show further that the problem of determining the optimum scale of a model can be circumvented by using ensembles of grid points from different models, i.e. the mean of grid point data from different reanalyses. Hofer et al. (2012) find that there is no or only marginal difference whether the ensemble is constructed from the reanalyses at the closest grid points, or at their optimum scales. As a priori choice, it is thus most reasonable to use reanalysis ensembles as predictors, because this way no assessment is required in order to determine the optimum model for each case, or the optimum spatial domain of each model. Even if for our study site this choice is already evaluated (Hofer et al., 2012), the usefulness of ensemble predictors is justifiable also in cases with no data-based assessments. In our a priori selection, we do not consider remote grid point predictors, as recommended in several ESD studies for precipitation predictands (e.g. Wilby and Wigley, 2000; Brinkmann, 2002; Sauter and Venema, 2011).

Finally, synthesizing all our above considerations concerning a priori predictor selection, we define air temperature averaged over the three reanalyses that have been available most recently: interim, MERRA and CFSR, at the grid points located closest to the study site (the horizontally closest grid points in the reanalysis models, at the 550 hPa levels, shown in Table 1), hereafter abbreviated by rea-ens-air. Statistics of rea-ens-air are shown in Fig. 2 for each month of the year for the period, when data are available for all AWSs (i.e. the same period as shown for the AWSs in Fig. 2, period 1). rea-ens-air is proposed as a priori choice because it is based on simple assumptions

Skill assessment of reanalysis data for air temperatureM. Hofer et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

that apply equally for different sites, seasons and large-scale models, without preceding data analysis. Note however that we do not claim that this is necessarily the best choice in each individual case. More precisely, the coarser the large-scale model, the more complex a relationship between the simulated predictors and the local predictand might be, e.g. involving multiple large-scale variables, and other than linear relationships (e.g. Hofer et al., 2010). However, such selections include data-based inference. In practice, the relation between large- and local-scale variables can be investigated best with limited-area numerical atmospheric models (LAMs), as they certainly include the most complete framework of linkages between the different scales: i.e. expressed by the governing atmospheric equations (e.g. Mölg and Kaser, 2011). Though, LAMs are computationally expensive, by contrast to ESD.

4.3 Downscaling process: linear model calibration and cross-validation

In this section the entire ESD modeling procedure, including data preprocessing, ESD model calibration, and skill estimation based on leave-one-out cross-validation is presented. Leave-one-out cross-validation is important especially in the case of short-term observational time series (as in the present study), because it allows each observation to be used in the model building process as well as in the model evaluation process (if time series are long, e.g. ten-fold cross-validation can be used instead of leave-one-out cross-validation, Hastie et al., 2001). The modification of leave-one-out cross-validation specifically presented here is appropriate for daily or sub-daily atmospheric time series, because it accounts for temporal autocorrelation (i.e. persistence, Madden, 1979).

First, the predictor and predictand time series, consisting of daily means, are separated into twelve different time series of daily means for the twelve months of the year. All steps described below are repeated separately and independently for each month's time series. The simplest way to relate the (a priori) predictor to a Gaussian target variable is a linear regression model. Note that the model is not appropriate for non

Gaussian target variables (e.g. precipitation). It applies

$$y_s(t) = \alpha \cdot x_s(t) + \epsilon(t). \quad (1)$$

ϵ is the model error, assumed to follow a Gaussian distribution with zero mean. α is the least-squares regression parameter. Note that least-squares regression does not account for the time ordering in data series, and is therefore not affected by the use of discontinuous (month-separated) time-series. Because $y_s(t)$ and $x_s(t)$ are standardized predictand and predictor time series (it applies $\bar{y}_s^t = \bar{x}_s^t := 0$ and $\sigma_t(y_s) = \sigma_t(x_s) := 1$, with the standardization parameters $\bar{\cdot}^t$ being the temporal mean, and $\sigma_t(\cdot)$ the temporal standard deviation of a variable), it can be shown that α is equal to the correlation coefficient (Von Storch and Zwiers, 2001).

With $\hat{y}(t) := y(t) - \epsilon(t)$, Eq. (1) can generally be rewritten including untransformed predictand $y(t)$, predictor $x(t)$, and $\bar{\cdot}^t$, and $\sigma_t(\cdot)$:

$$\hat{y}(t) = \bar{y}^t + \alpha \cdot \frac{\sigma_t(y)}{\sigma_t(x)} \left(x(t) - \bar{x}^t \right) \quad (2)$$

To estimate $\hat{y}(t)$ and $\epsilon(t)$ a modification of leave-one-out cross-validation (Michaelsen, 1987) is applied. Cross-validation is repeated n_{cv} times. Here we use $n_{cv} = n$ (n is the number of observations of each month-separated time series). Please note again (as defined above) that an observation in the month-separated time series is a daily mean value. Each cross-validation repetition, n_{lo} observations (thus, daily means) are excluded from the model fit (the “left-out” observations), with

$$n_{lo} = 2 \cdot \tau + n_{i0}. \quad (3)$$

τ is the temporal lag, for which the autocorrelation function of y is within the 95% confidence interval of Gaussian white noise (the 5% confidence interval is approximated with $2/n^{1/2}$). n_{i0} is the number of independent observations used to estimate the model error. In general n_{i0} can be chosen to be larger than one, then the leave-one-out

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



cross-validation becomes moving-block cross-validation (Kunsch, 1989). In this study we choose $n_{io} = 1$. In each cross-validation step (cv) $n_T = n - n_{io}$ (T for training) data pairs $\{y_T, x_T\} := \{y(t_T(cv)), x(t_T(cv))\}$ are used to estimate the parameters of the simple linear model. Thus in Eq. (2) it applies

$$5 \quad \alpha = \alpha(y_T, x_T), \quad \sigma_t(y)/\sigma_t(x) = \sigma_t(y_T)/\sigma_t(x_T), \quad (4)$$

and

$$\bar{y} = \bar{y}_T, \quad \bar{x} = \bar{x}_T.$$

$\{y_V, \hat{y}_V\} := \{y(t_V(cv)), \hat{y}(t_V(cv))\}$ (V for validation) are then used to estimate the model error ϵ (Eq. 1). y_V is the central of the withheld observations in each cross-validation step cv and can be considered as independent from the calibration process.

When the cross-validation process is completed, the skill score (SS) can be calculated (e.g. Wilks, 2006):

$$10 \quad SS = 1 - \frac{mse}{mse_r}, \quad (5)$$

and, with Eq. (1):

$$15 \quad mse = \frac{1}{n_{cv}} \cdot \sum \epsilon^2(cv), \quad (6)$$

where $\epsilon(cv) = y_V - \hat{y}_V$, and $cv = 1, \dots, n_{cv}$.

mse_r is the mean of squared errors of the reference model, \hat{y}_r , as follows:

$$mse_r = \frac{1}{n_{cv}} \cdot \sum (y_V - \hat{y}_r)^2 \quad y_r := \bar{y}_T. \quad (7)$$

20 SS, consisting of a contribution due to the correlation between the forecasts and observations, and two penalty terms relating to the reliability and bias of the forecast

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



(Murphy, 1988), is the more accurate skill measure than the correlation coefficient r^2 (Wilks, 2006). SS as specifically presented here can be considered as r^2 deflated by the conditional bias term (the unconditional bias is by construction constrained to be zero in least-squares regression; definitions of reliability and bias terms are given in the work of Murphy, 1988).

The ESD method and the skill assessment are appropriate also when only short measured time series are available for the model calibration and validation. Below, we show an application with about three to seven years' measurement series from the three AWSs introduced earlier. In order to give an exact quantification on whether the data are long enough for the ESD model and skill assessment to be reliable (or useful), the significance of SS is determined by performing a left-tailed T-test of the null hypothesis that the squared model error estimated by the cross-validation process, $\epsilon^2(\text{cv})$ in Eq. (5), is equal to the mean of squared errors of the reference model, mse_r , against the alternative that the mean of $\epsilon^2(\text{cv})$ is smaller than mse_r . SS is considered significant, when the null hypothesis is rejected at the 5 % significance level. In performing the T-test, we consider the autocorrelation of $\epsilon^2(\text{cv})$, by replacing the sample size $n_{\text{cv}} = n$ with the effective sample size n_{eff} (Wilks, 2006), approximated as

$$n_{\text{eff}} \approx n_{\text{cv}} \frac{1 - \tau_1}{1 + \tau_1}, \quad (8)$$

with τ_1 being the lag-1 autocorrelation of ϵ^2 .

Finally, let \bar{v}^{cv} be the mean, and $\sigma_{\text{cv}}(v)$ be the standard deviation of a variable (v) over all cv repetitions, then the final model \hat{y}_F , with model uncertainty estimated by cross-validation (not to be mistaken for the model error $\epsilon(t)$) is

$$\hat{y}_F(t) = \overline{\hat{y}(t)}^{\text{cv}} \pm \sigma_{\text{cv}}(\hat{y}(t)). \quad (9)$$

Equations (1)–(7) apply similarly in multiple regression, where x has multiple columns (i.e. $x \in R^{(n \times p)}$). Note that, even though not shown in this study, SS as defined above

is a powerful goodness-of-fit estimate especially in the case of multiple predictors, because it detects over-fitting (then SS is zero or not significant).

4.4 Example

Figure 3 provides a comprehensible example of the skill estimation procedure described above. The two plots show daily air temperature time series (y in Eq. 2) of the months March (top) and July (bottom) at AWS1 (blue line). The red line is the linear model \hat{y} based on rea-ens-air, as defined by Eq. (2). The example shows the model building and error estimation in an individual cross-validation repetition cv (Eqs. 6–7). The grey bar indicates the n_{lo} observations left out in the model calibration (note that each cross-validation round cv , the grey bar is shifted one observation to the right). The amount of observations left apart is determined by the cross-validation parameter τ (the temporal lag for which the autocorrelation of the time series can be assumed to be zero). The values of τ are 9 in March, and 2 in July. This means that, e.g. in the March time series an observation is considered independent from an other observation only if there is a shift of at least 9 time steps (in this case, days) between the two observations (thus, the grey bar includes $9 \cdot 2 + 1 = 19$ left-out observations). The error ϵ_{cv} for each repetition cv (Eq. 6) is estimated as the difference between the central (independent) observation (y_V , the blue star in the grey bar in Fig. 3) and the model value at this time step (\hat{y}_V , the red star in the grey bar in Fig. 3). y_r in Eq. (7) is the black star in Fig. 3, calculated as the mean of y_T , the observations used in the model training. Cross-validation is repeated until each observation is used once as y_V to determine the model error. This way independence between the observations used in the model training and the validation process is warranted, and at the same time all observations can be used to determine the final model (Eq. 9). Note that the results in Fig. 3 are discussed in Sect. 5.2.

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



5 ESD model application

5.1 Downscaling parameters

In this section the above presented ESD procedure is applied to daily mean air temperature time series of the AWSs in the Cordillera Blanca (introduced in Sect. 2). As mentioned in Sect. 4.3, the linear models are calibrated and validated independently for each calendar month and for each of the AWSs' time series. Thus the time series finally input to the ESD procedure (e.g. January at AWS1) consist of seven months (consecutive Januaries) of daily data from the seven years of available observations (July 2006–July 2012), consequently approximately 200 observations per month and AWS (data gaps included) to calibrate the ESD model (e.g. the AWS1 March time series consists of 186 values and the AWS1 July time series as well of 186 values, in Fig. 3).

We define the standardization parameter R_σ as the ratio between the standard deviation of the predictand $\sigma_t(y)$ and the standard deviation of the predictor $\sigma_t(x)$

$$R_\sigma := \frac{\sigma_t(y)}{\sigma_t(x)}. \quad (10)$$

R_σ is a parameter in the ESD Eq. (2). Figure 4 shows the regression parameter α (equal to the correlation coefficient, Eq. 1) and the standardization parameter R_σ with uncertainties, estimated by cross-validation (Eqs. 4–9), for each calendar month, and data from AWS1. Most remarkably, both α and R_σ show a relatively high inter-monthly variability. Values of α are relatively high all year round, but highest for the months January–May (wet season in the Cordillera Blanca), with a second maximum in September, and slightly lower for the dry-season months June–August, and November–December. The largest coefficient uncertainties are evident for the months February–April. R_σ throughout the year varies from approximately 0.9–1.4. In December, the difference between the (day-to-day or year-to-year) variabilities of the predictand and rea-ens-air is largest, whereas in May–August the variability of the predictand is about the same as (or even

lower than) the variability of reanalysis-air. The rather high inter-monthly variations in the downscaling parameters clearly support the importance of using different models for the different calendar months.

Figure 5 (red circles) shows values of the cross-validation parameter τ of AWS1 daily air temperature means for each month of the year. As defined in Sect. 4.3, τ can be interpreted in terms of the temporal lag (in days), for which the time series values can be assumed independent. Values of τ for the different calendar months vary between 2 (small persistence) and 11 days (high persistence). Values of τ are 2 or 3 for all months despite the wet-season months February, March and the transitional-season month April, where values of τ are considerably higher (7, 9 and 11, respectively). The higher values of τ in these months are probably related to the prevailing intra-seasonal variability in the tropical Andes: i.e. rainy episodes in terms of sequences of wet days followed by sequences of dry days, with associated variances in air temperature (Garreaud et al., 2003). In the tropics, such synoptic episodes are known to typically range from 30 to 60 days (with the basic mechanism known as the Madden-Julian Oscillation, MJO; Madden and Julian, 1994), however for the Bolivian Altiplano (located nearby the Cordillera Blanca), Garreaud et al. (2003) report about shorter synoptic periods of approximately 15 days length. Please note that by examining τ of the month-separated time series, it is not possible to identify the full length of MJO cycles. Small values of τ especially for the austral winter months indicate not only small day-to-day (intra-seasonal), but also small inter-annual variability. Consequently there are no important differences amongst the different years of the respective austral winter months. In fact, ENSO, the most important source of inter-annual variability in the region, has its strongest and most widespread impacts during austral summer.

5.2 Skill assessment and significance analysis

Figure 6 shows values of SS of reanalysis-air for each month of the year, with AWS1 daily air temperature as the predictands. As mentioned earlier, data from AWS1 represent the longest high-resolution measurement series available at high altitude in the

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Cordillera Blanca (i.e. July 2006–July 2012, without data gaps, hereafter period 2; green circles in Fig. 6). Values of SS show a distinct seasonal pattern, with two maxima for April and for September ($SS \approx 0.6$), and two minima for June and for November ($SS \approx 0.3$). Overall, values of SS are lowest for the core dry season months June and July, and highest for the wet season months February–April. The high values of SS in the wet season imply that the largest portion of day-to-day variability of the AWS1 air temperature time series can be explained by rea-ens-air, indicating high spatial homogeneity of air temperature fluctuations for these months of the year. In fact, Garreaud et al. (2003) report spatially very coherent intra-seasonal weather patterns for the nearby Bolivian Altiplano, and also MJO is known to act on large spatial scales (local wavelengths of $1.2 - 2 \times 10^3$ km). By contrast in the core dry season, values of SS of rea-ens-air reach only half of the wet season values. We suggest that variability in these months must be governed by processes that act more locally (e.g. triggered by the strong radiation interacting with the complex topography), in a way that the generally weaker synoptic forcing in these months impacts the local-scale variability in a more complex way that is only partially captured by the single linear predictor rea-ens-air.

In Fig. 3 (lower panel), the model chosen by least-squares regression for the dry-season month July shows only minor variance, as the co-variability between rea-ens-air and the predictand data series is small. Year-to-year as well as day-to-day variability in the observational time series are evidently smaller than for the wet-season month March (upper panel in Fig. 3). However, the underestimation of observed variance is nevertheless smaller for the wet-season month March, indicating a higher co-variability between the predictor and predictand time series (note that, by construction of least-squares-regression, the difference in underestimation of observed variance is immediately related to the co-variability between predictand and predictor time series).

The grey bars in Fig. 6 show results of the skill assessment applied to the AWS1, AWS2, and AWS3 data over the period, where measurements are available for all three AWSs (i.e. from 2006/7 to 2009/12, period 1). Whereas for the same period, differences of SS for the different AWSs are rather small (all AWSs showing a similar

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



seasonal pattern, with similar minima and maxima like for AWS1 for the entire data period, and an additional minimum in the core wet season), differences in the values of SS between period 1 and period 2 are quite evident. Overall, but mostly for the core wet season, values of SS for AWS1 based on the shorter period 2 are considerably lower, than based on period 1. This is a somehow expected result, because the periods for the skill assessment are very limited. Thus, as the data base for the model training increases with an almost doubling from period 1 to period 2, the model parameters are estimated more accurately, and the cross-validation results are more positive. Values of SS are expected to be less dependent on the length of the measuring period with an increasing amount of available data.

The significance analysis as described in Sect. 4.3 reveals that the skill scores for all months and all AWSs shown in Fig. 6 are significant at the 5% significance level. By systematically assessing the significance of SS for decreasing time periods, a minimum time period can be identified for which the skill assessment is significant (and thus useful). This threshold time period can be quantified in terms of a minimum number of observations required in each calendar month for the skill assessment to be significant. An AWS time series can be considered long enough for the skill assessment, if for all months the number of available observations exceeds this minimum number of observations. Figure 5 shows the minimum number of observations required, n_{\min} , at the example of data from AWS1, for each month of the year. This minimum number of observations is a function of SS of each month, as well as of the autocorrelation of each month's time series (by considering the effective sample size of each time series n_{eff} , Eq. 8, also shown in Fig. 5). The minimum number of observations largely varies from about 40, for months with high values of SS (December–March, May, September), to 140, for months with low values of SS (July, November). Since the AWS1 time series includes more than 120–150 (period 2), and 210 (period 1), respectively, observations for each month (daily means), the minimum time length for the skill assessment to be significant is achieved for all months (see also Fig. 3); though only hardly for July and November for period 1.

Values of SS and r^2 averaged over all months for the different AWSs are shown in Table 2. The values are shown for the entire period of available data at each AWS, and additionally for AWS1 and AWS2 for period 1, the period of common available data at all AWSs. On average, SS is highest at AWS2 (mean SS = 0.48) and lowest at AWS3 (mean SS = 0.4). As discussed above for AWS1, values of SS are considerably lower for the shorter time period (period 1) also for AWS2. Table 2 further shows that values of r^2 overestimate the skill, compared to SS based on cross-validation, with up to 0.08 lower values (AWS3). The overestimation of skill by r^2 can be expected to increase strongly in the case of multiple predictors (multiple regression), due to overfitting.

5.3 Results for different time scales

In this study the short length of the data represents a lower limit of possible time resolutions for the skill assessment, because for lower temporal resolutions the number of observations n available for the model set-up decreases. More specifically, the seven-years observations at AWS1 include a time series of 84 monthly means, or 2555 daily means. As we use separate models for each month, the time series include only seven monthly means, but approximately 210 daily means. With this regard it makes sense to profit from the higher (daily) temporal resolution to have more observations for the model fit. Nevertheless it is worth to investigate if the reanalysis data show higher skill for temporal resolutions lower than the daily time resolution.

In this section we repeat the modeling procedure as introduced in Sect. 4.3, but for different temporal resolutions. The number of observations n in each time scale is set constant, in order to allow for a comparison of SS of the different time scales. The lowest temporal resolution for which the AWS1 time series (period 1) still include enough values for the skill assessment to be significant (as defined in Sect. 4.3), differ for the different calendar months. Figure 7 shows values of SS for decreasing time scales (from daily means to twelve-daily means), and the different calendar months of AWS1 data, for which SS is found to be significant at each respective time resolution.

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the CFSR variables spr and wwn between data prior and after December 2010 (the discrepancies are due to changes in the model configurations from CFSR available until December 2010, to the subsequently operationally available CFSV2).

The results of the skill assessment reveal that only three of the assessed variables show significant skill for the local measured daily air temperature predictand: air, t2m and gph. Values of SS of these variables for AWS1 data are shown in Fig. 8 for each month of the year. For all other variables listed in Table 3, values of SS are non-significantly different from zero. t2m clearly shows lower values of SS than air in all months. Thus the reanalysis variables that are less affected by the model surface actually emerge as the better predictors for surface predictand air temperature at AWS1. This is consistent with findings in earlier studies (e.g. Murphy, 1999; Rummukainen, 1997) and supports our a priori assumption that unrealistic surface information negatively affects the data. We conclude from Fig. 8 that (1) only air temperature predictors, or gph (which is physically closely linked with air temperature) show significant skill regarding day-to-day and year-to-year variability based on our linear model set up (presented in Sect. 4.3), (2) the a priori predictor choice is clearly supported by the data, and (3) even if values of SS show a distinct seasonal cycle, the same predictors show the highest skill throughout the year.

5.5 Accounting for the effects of diurnal periodicity

Finally we show a simple analysis based on AWS1 air temperature data to demonstrate the effects of periodicity (in this case diurnal) for regression analysis based solely on r^2 (an often applied criterium for predictor selection). Figure 9 shows values of r^2 between six-hourly (all-month) time series of AWS1 air temperature, and the predictors assessed in Sect. 9: shm, gph, air, uwn, vwn, spr, vor, t2m, and wwn. The same analysis is shown at a daily time scale (thus r^2 between the same time series, but the six-hourly data averaged to daily means). Results show for the six-hourly data (left hand side in Fig. 9) that t2m emerges as the best predictor, showing a relatively high covariance ($r^2 > 0.7$), whereas all other predictors show only minor covariance ($r^2 < 0.1$)

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



to the predictand (including air, which emerged as the best predictor in Sect. 9). This pattern significantly changes in the daily time scale (right hand side in Fig. 9): now the predictor air shows the highest covariance (r^2 is almost 0.5), but t2m shows only very small covariance ($r^2 = 0.2$ – yet still higher than all other assessed predictors). In the diurnal analysis t2m appears as important predictor only because it has the most pronounced diurnal variations which explain the largest portion of variability also in the predictand data (i.e. high values during day-time and low values in the night). However this is achieved easily with a constant diurnal cycle. For example, the covariance between the hourly air temperature series and a time series composed by consecutive constant diurnal cycles is $r^2 = 0.7$. This shows why predictor selection that does not account for diurnal (or other) periodicity is not meaningful. This analysis is by no means innovative in statistics, but given here as an illustration, since the issue of periodicity is not accounted for appropriately in numerous studies.

6 Summary and conclusions

We have presented an ESD technique that links large-scale atmospheric model predictors to Gaussian, local-scale target variables measured over complex terrain. The method is appropriate for temporally high-resolution, sub-daily to daily time series, and is designed to give a significant estimate of large-scale model skill on the basis of small observation availability (i.e. few-years) as is often the case at mountainous sites. We highlight the importance of systematically eliminating seasonal (or diurnal) periodicity in the meteorological time series, and at the same time in the ESD model error, by using separate models for the different times of year (and times of day, respectively). The presented skill estimation is based on leave-one-out cross-validation that accounts for autocorrelation in the observational data.

We have shown an application of the ESD model with reanalysis data as the predictors, and daily mean air temperature measured at three high-altitude sites in the glaciated Cordillera Blanca (Peru) as the target variables. First, the skill of the a priori

**Skill assessment of
reanalysis data for air
temperature**

M. Hofer et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

5 predictor reanalysis-air is assessed for the longest AWS air temperature series. High seasonality of statistical data properties (e.g. persistence) and ESD model parameters emphasizes the importance of using different models for different times of year. The ESD model skill shows high seasonality as well, with generally higher skill in the wet
10 season, and lower skill in the dry season. Whereas differences in skill for the different AWSs are rather small, the seasonal pattern of skill shows an increasing tendency for increasing numbers of available observations. The skill assessment is shown to be significant at a 5 % test level for a minimum of 40–140 daily observations available for each calendar month (depending on calendar month). For the same number of observations
15 at different temporal resolutions (i.e. from one-day to twelve-day averages), values of skill averaged over all months increase with increasing averaging time windows. Consequently we suggest switching to lower temporal resolutions when the ESD model skill is low, given that long enough data series are available.

The predictor air clearly shows higher skill than other potential predictors, such as
20 t2m or gph. We suggest that large-scale surface variables are weak predictors if the model surface is not representative for the real surface, because unrealistic boundary layer variability masks the relevant synoptic forcing and variability. By skill assessments that do not account for periodicity, the remarkably lower performance of t2m, compared to air, can not be identified. For six further assessed reanalyses predictors the skill assessment reveals no significant skill. The presented ESD model can be generalized to
25 non-linear, multiple regression problems (not shown here). The validation process is especially useful in multiple predictor fitting because it detects over-fitting. The method is not restricted to reanalysis data and can be applied to any atmospheric model predictors.

30 *Acknowledgements.* This study is funded by the Austrian Science Foundation (P22106-N21). Thomas Mölg is supported by the Alexander von Humboldt Foundation. ERA-interim data are provided by the ECMWF. MERRA are disseminated by the GMAO and the Goddard Earth Sciences Data and Information Services Center (GES DISC). CFSR are obtained by the Research Data Archive (RDA) which is maintained by the Computational and Information Systems Laboratory (CISL) at the National Center for Atmospheric Research (NCAR).

References

- Ames, A.: A documentation of glacier tongue variations and lake development in the Cordillera Blanca, Peru, *Zeitung für Gletscherkunde und Glazialgeologie*, 34, 1–36, 1998. 2886
- Bair, E., Hastie, T., and Tibshirani, R.: Prediction by supervised principal components, *J. Am. Stat. Assoc.*, 101, 119–137, doi:10.1198/016214505000000628, 2006. 2885
- Benestad, R. E., Førland, E. J., and Hanssen-Bauer, I.: Empirically downscaling temperature scenarios for Svalbard, *Atmos. Sci. Lett.*, 3, 71–93, doi:10.1006/asle.2002.0050, 2002. 2891
- Benestad, R. E., Hanssen-Bauer, I., and Deliang, C.: Empirical-statistical downscaling, *World Scientific*, Singapore, 2008. 2885, 2892, 2893
- Bengtsson, L. and Shukla, J.: Integration of space and in situ observations to study global climate change, *B. Am. Meteorol. Soc.*, 69, 1130–1143, 1988. 2889
- Bosilovich, M. G.: NASA's Modern Era Retrospective-analysis for Research and Applications: Integrating Earth Observations, available at: <http://www.earthzine.org/2008/09/26/nasas-modern-era-retrospective-analysis/> (last access: 5 April 2013), 2013. 2889
- Brinkmann, W. A. R.: Local versus remote grid points in climate downscaling, *Clim. Res.*, 21, 27–42, doi:10.3354/cr021027, 2002. 2893
- Carey, M.: Living and dying with glaciers: people's historical vulnerability to avalanches and outburst floods in Peru, *Global Planet. Change*, 47, 122–134, doi:10.1016/j.gloplacha.2004.10.007, 2005. 2886
- Carey, M.: *In the Shadow of Melting Glaciers*, *Climate Change and Andean Society*, Oxford University Press, New York, USA, 2010. 2886
- Cavazos, T. and Hewitson, B. C.: Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation, *Clim. Res.*, 28, 95–107, 2005. 2885
- Christensen, J., Hewitson, B., Busuioc, A., Chen, A., Gao, X., Held, I., Jones, R., Kolli, R., Kwon, W.-T., Laprise, R., Rueda, V. M., Mearns, L., Menendez, C., Räisänen, J., Rinke, A., Sarr, A., and Whetton, P.: *Regional Climate Projections*, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2007. 2884
- Fowler, H., Blenkinsop, S., and Tebaldi, C.: Review: linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *Int. J. Climatol.*, 27, 1547–1578, 2007. 2885

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Garreaud, R., Vuille, M., and Clement, A.: The climate of the Altiplano: observed current conditions and mechanisms of past changes, *Palaeogeogr. Palaeoclimatol.*, 194, 5–22, doi:10.1016/S0031-0182(03)00269-4, 2003. 2887, 2889, 2900, 2901

Georges, C.: Ventilated and unventilated air temperature measurements for glacier-climate studies on a tropical high mountain site, *J. Geophys. Res.*, 107, 4775, doi:10.1029/2002JD002503, 2002. 2888

Georges, C.: 20th-century glacier fluctuations in the tropical Cordillera Blanca, Peru, *Arct. Antarct. Alp. Res.*, 36, 100–107, 2004. 2886, 2917

Georges, C.: Recent Glacier Fluctuations in the Tropical Cordillera Blanca and Aspects of the Climate Forcing, Ph. D. thesis, Leopold-Franzens University Innsbruck, Innsbruck, Austria, 2005. 2886

Giorgi, F. and Bates, G. T.: The climatological skill of a regional model over complex terrain, *Mon. Weather Rev.*, 117, 2325–2347, doi:10.1175/1520-0493(1989)117<2325:TCSOAR>2.0.CO;2, 1989. 2885

Grotch, S. L. and MacCracken, M. C.: The use of global climate models to predict regional climatic change, *J. Climate*, 4, 286–303, 1991. 2892

Hastie, T., Tibshirani, R., and Friedman, J.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer series in statistics, 1 Edn., Springer, New York, 2001. 2885, 2894

Hessami, M. P. G., Ouarda, T. B. M. J., and St-Hilaire, A.: Automated regression-based statistical downscaling tool, *Environ. Modell. Softw.*, 23, 813–834, doi:10.1016/j.envsoft.2007.10.004, 2008. 2885

Hill, G. E.: Grid telescoping in numerical weather prediction, *J. Appl. Meteorol.*, 7, 29–38, doi:10.1175/1520-0450(1968)007<0029:GTINWP>2.0.CO;2, 1968. 2885

Hofer, M., Mölg, T., Marzeion, B., and Kaser, G.: Empirical-statistical downscaling of reanalysis data to high-resolution air temperature and specific humidity above a glacier surface (Cordillera Blanca, Peru), *J. Geophys. Res.*, 115, D12120, doi:10.1029/2009JD012556, 2010. 2887, 2889, 2890, 2891, 2894

Hofer, M., Marzeion, B., and Mölg, T.: Comparing the skill of different reanalyses and their ensembles as predictors for daily air temperature on a glaciated mountain (Peru), *Clim. Dynam.*, 39, 1969–1980, doi:10.1007/s00382-012-1501-2, 2012. 2887, 2889, 2891, 2893, 2904, 2916

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Huth, R.: Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors, *J. Climate*, 17, 640–652, doi:10.1175/1520-0442(2004)017<0640:SOLDTC>2.0.CO;2, 2004. 2891
- Juen, I.: Glacier mass balance and runoff in the Cordillera Blanca, Peru, Ph. D. thesis, Leopold Franzens University Innsbruck, Innsbruck, Austria, 2006. 2886, 2887
- Juen, I., Georges, C., and Kaser, G.: Modelling observed and future runoff from a glacierized tropical catchment (Cordillera Blanca, Peru), *Global Planet. Change*, 59, 37–48, 2007. 2886, 2887
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *B. Am. Meteorol. Soc.*, 77, 437–471, 1996. 2889
- Kaser, G. and Osmaston, H.: *Tropical Glaciers*, International Hydrology Series, Cambridge University Press, Cambridge, UK, 2002. 2886
- Kaser, G., Juen, I., Georges, C., Gomez, J., and Tamayo, W.: The impact of glaciers on the runoff and the reconstruction of mass balance history from hydrological data in the tropical Cordillera Blanca, Peru, *J. Hydrol.*, 282, 130–144, 2003. 2886
- Kaser, G., Grosshauser, M., and Marzeion, B.: The contribution potential of glaciers to water availability in different climate regimes, *P. Natl. Acad. Sci. USA*, 107, 20223–20227, 2010. 2886
- Kazutoshi, O., Tsutsui, J., Koide, H., Sakamoto, M., Kobayashi, S., Hatsushika, H., Matsumoto, T., Yamazaki, N., Kamahori, H., Takahashi, K., Kadokura, S., Wada, K., Kato, K., Oyama, R., Ose, T., Mannoji, N., and Taira, R.: The JRA-25 reanalysis, *J. Meteorol. Soc. Jpn.*, 85, 369–432, 2007. 2889
- Kidson, J. W. and Thompson, C. S.: A comparison of statistical and model-based downscaling techniques for estimating local climate variations, *J. Climate*, 11, 735–753, doi:10.1175/1520-0442(1998)011<0735:ACOSAM>2.0.CO;2, 1998. 2891
- Klein, W. H. and Glahn, H. R.: Forecasting local weather by means of model output statistics, *B. Am. Meteorol. Soc.*, 55, 1217–1227, 1974. 2891
- Klein, W. H., Lewis, B. M., and Enger, I.: Objective prediction of five-day mean temperature during winter, *J. Meteorol.*, 16, 672–682, 1959. 2885
- Kunsch, H. R.: The jackknife and the bootstrap for general stationary observations, *Ann. Stat.*, 17, 1217–1241, 1989. 2896

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Madden, R. A.: Estimates of the natural variability of time-averaged sea-level pressure, *Mon. Weather Rev.*, 104, 942–952, doi:10.1175/1520-0493(1976)104<0942:EOTNVO>2.0.CO;2, 1976. 2890
- Madden, R. A.: A simple approximation for the variance of meteorological time averages, *J. Appl. Meteorol.*, 18, 703–706, doi:10.1175/1520-0450(1979)018<0703:ASAFTV>2.0.CO;2, 1979. 2894
- Madden, R. A. and Julian, P. R.: Observations of the 40 to 50 day tropical oscillation: a review, *Mon. Weather Rev.*, 122, 814–837, doi:10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2, 1994. 2900
- Mark, B. G. and Seltzer, G. O.: Tropical glacier meltwater contribution to stream discharge: a case study in the Cordillera Blanca, Peru, *J. Glaciol.*, 49, 271–281, 2003. 2886
- Mearns, L. O., Giorgi, F., Whetton, P., Pabon, D., Hulme, M., and Lal, M.: Guidelines for Use of Climate Scenarios Developed from Regional Climate Model Experiments, IPCC Supporting Material, available at: http://ipcc-data.org/guidelines/dgm_no1_v1_10-2003.pdf, last access: 7 May 2013, 2003. 2885
- Michaelsen, J.: Cross-validation in statistical climate forecast models, *J. Clim. Appl. Meteorol.*, 26, 1589–1600, 1987. 2895
- Mölg, T. and Kaser, G.: A new approach to resolving climate-cryosphere relations: downscaling climate dynamics to glacier-scale mass and energy balance without statistical scale linking, *J. Geophys. Res.*, 116, D16101, doi:10.1029/2011JD015669, 2011. 2894
- Murphy, A. H.: Skill scores based on the mean square error and their relationships to the correlation coefficient, *Mon. Weather Rev.*, 116, 2417–2424, doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, 1988. 2897
- Murphy, J.: An evaluation of statistical and dynamical techniques for downscaling local climate, *J. Climate*, 12, 2256–2284, doi:10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2, 1999. 2905
- Niedertscheider, J.: Untersuchungen zur hydrographie der Cordillera Blanca (Peru), M.S. thesis, Leopold Franzens University, Innsbruck, Austria, 1990. 2886, 2887
- Räisänen, J. and Ylhäisi, J. S.: How much should climate model output be smoothed in space?, *J. Climate*, 24, 867–880, doi:10.1175/2010JCLI3872.1, 2011. 2892
- Rummukainen, M.: Methods for statistical downscaling of GCM simulations, no. 80 in: SMHI Reports Meteorology and Climatology, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden, 1997. 2905

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Sauter, T. and Venema, V.: Natural three-dimensional predictor domains for statistical precipitation downscaling, *J. Climate*, 24, 6132–6145, doi:10.1175/2011JCLI4155.1, 2011. 2893
- Silverio, W. and Jaquet, J.-M.: Glacial cover mapping (1987–1996) of the Cordillera Blanca (Peru) using satellite imagery, *Remote Sens. Environ.*, 95, 342–350, 2005. 2886
- 5 Trenberth, K. E. and Olson, J. G.: An evaluation and intercomparison of global analyses from the National Meteorological Center and the European Centre for Medium Range Weather Forecasts, *B. Am. Meteorol. Soc.*, 69, 1047–1057, doi:10.1175/1520-0477(1988)069<1047:AEAIOG>2.0.CO;2, 1988. 2889
- Uppala, S., Källberg, P., Simmons, A., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J., 10 Haseler, J., Hernandez, A., Kelly, G., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Hólm, E., Hoskins, B., Isaksen, L., Janssen, P., Jenne, R. M. A., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N., Saunders, R., Simon, P., Sterl, A., Trenberth, K., Untch, A., 15 Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, *Quarterly Journal of the Royal Meteorological Society*, 131, 2961–3012, doi:10.1256/qj.04.176, 2005. 2889
- Von Storch, H.: On the use of inflation in statistical downscaling, *J. Climate*, 12, 3505–3506, 1999. 2885, 2891
- Von Storch, H. and Zwiers, F.: *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, UK, 2001. 2895
- 20 Von Storch, H., Zorita, E., and Cubasch, U.: Downscaling of global climate change estimates to regional scales: an application to Iberian rainfall in wintertime, *J. Climate*, 6, 1161–1171, doi:10.1175/1520-0442(1993)006<1161:DOGCCCE>2.0.CO;2, 1993. 2893
- Vuille, M., Francou, B., Wagnon, P., Juen, I., Kaser, G., Mark, B. G., and Bradley, R. S.: Climate change and tropical Andean glaciers: past, present and future, *Earth-Sci. Rev.*, 89, 79–96, 25 doi:10.1016/j.earscirev.2008.04.002, 2008a. 2889
- Vuille, M., Kaser, G., and Juen, I.: Glacier mass balance variability in the Cordillera Blanca, Peru and its relationship to climate and large scale circulation, *Global Planet. Change*, 62, 14–28, 2008b. 2888, 2889
- 30 Wilby, R. and Dawson, C.: *SDSM 4.2 A decision support tool for the assessment of regional climate change impacts*, Loughborough University, Leicestershire, UK, 2007. 2885
- Wilby, R. and Wigley, T.: Precipitation predictors for downscaling: observed and general circulation model relationships, *Int. J. Climatol.*, 20, 641–661, 2000. 2893

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Wilby, R., Dawson, C., and Barrow, E.: SDSM – a decision support tool for the assessment of regional climate change impacts, *Environ Modell. Softw.*, 17, 145–157, 2002. 2885, 2891
- Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P., and Mearns, L.: Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling Methods, Tech. rep., IPCC Supporting Material, available at: <http://www.narccap.ucar.edu/doc/tgica-guidance-2004.pdf>, last access: 7 May 2013, 2004. 2885
- Wilks, D. S.: Statistical methods in the atmospheric sciences, vol. 91 of International Geophysics Series, 2 Edn., Academic Press, Oxford, UK, 2006. 2896, 2897
- Williamson, D. L. and Laprise, R.: Numerical approximations for global atmospheric GCMs, in: Numerical Modeling of the Global Atmosphere in the Climate System, Kluwer Academic, Castelvechio Pascoli, Italy, 147–219, 2000. 2892
- Winkler, J. A., Palutikof, J. P., Andresen, J. A., and Goodess, C. M.: The simulation of daily temperature time series from GCM output, Part II: Sensitivity analysis of an empirical transfer function methodology, *J. Climate*, 10, 2514–2532, doi:10.1175/1520-0442(1997)010<2514:TSODTT>2.0.CO;2, 1997. 2885
- Zorita, E. and Von Storch, H.: A survey of statistical downscaling techniques, Tech. rep., GKKS, Geesthacht, Germany, 1997. 2893
- Zucchini, W.: An Introduction to model selection, *J. Math. Psychol.*, 44, 41–61, doi:10.1006/jmps.1999.1276, 2000. 2885

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 1. Specifications of the reanalyses' data grid points applied as predictors: coordinates, surface heights (h), and mean geopotential heights (gph), with standard deviations in brackets, during the investigation period (all values are in units meters above sea level).

	h [m a.s.l.]	gph 550 hPa [m a.s.l.]
interim (77.5° W 9° S)	3324	5113 (± 11.4)
CFSR (77.5° W 8.5° S)	3409	5106 (± 11.6)
MERRA (77.3° W 8.5° S)	3120	5103 (± 11.4)

GMDD

6, 2883–2925, 2013

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Table 2. Values of SS and r^2 for the three assessed AWSs, averaged over all months of the year, estimated using the entire measuring periods from each AWS, and in brackets the respective values based on period 1 (the shortest, AWS3 measuring period) for AWS1 and AWS2.

	AWS1	AWS2	AWS3
r^2	0.47 (0.4)	0.5 (0.38)	0.48
SS	0.45 (0.37)	0.48 (0.32)	0.4

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Table 3. List of predictors (and their abbreviations) assessed in Sect. 9. All predictors are from the interim reanalyses at their optimum spatial domain (as determined in the work of Hofer et al., 2012).

shm	specific humidity at 550 hPa
gph	geopotential height at 550 hPa
air	air temperature at 550 hPa
uwn	zonal wind speed at 550 hPa
vwn	meridional wind speed at 550 hPa
spr	surface pressure
vor	potential vorticity at 550 hPa
t2m	air temperature at 2 m a.s.l.
wwn	vertical pressure velocity at 550 hPa

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



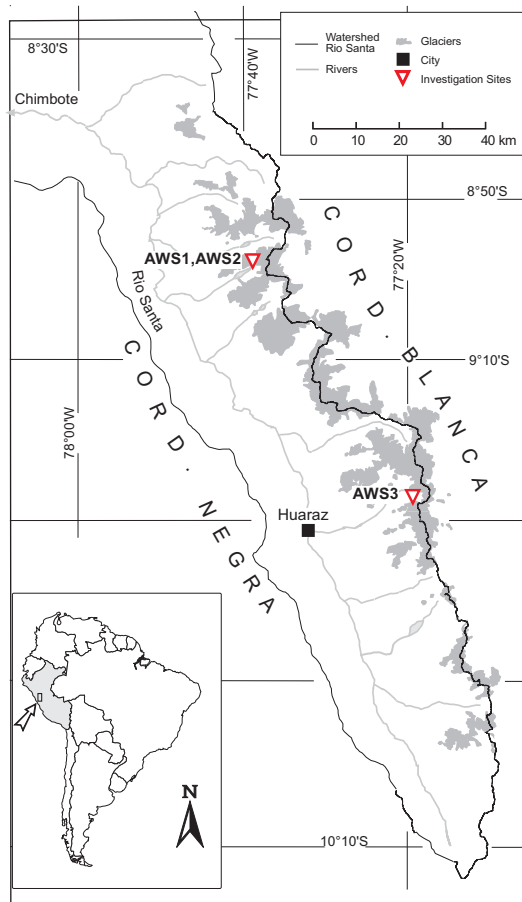


Fig. 1. The map shows the Rio Santa watershed with the Cordillera Blanca mountain range with the positions of AWS1, AWS2 and AWS3 (mentioned in the text). Also indicated is the 1990 glacier extent (grey shaded area, Georges, 2004).

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



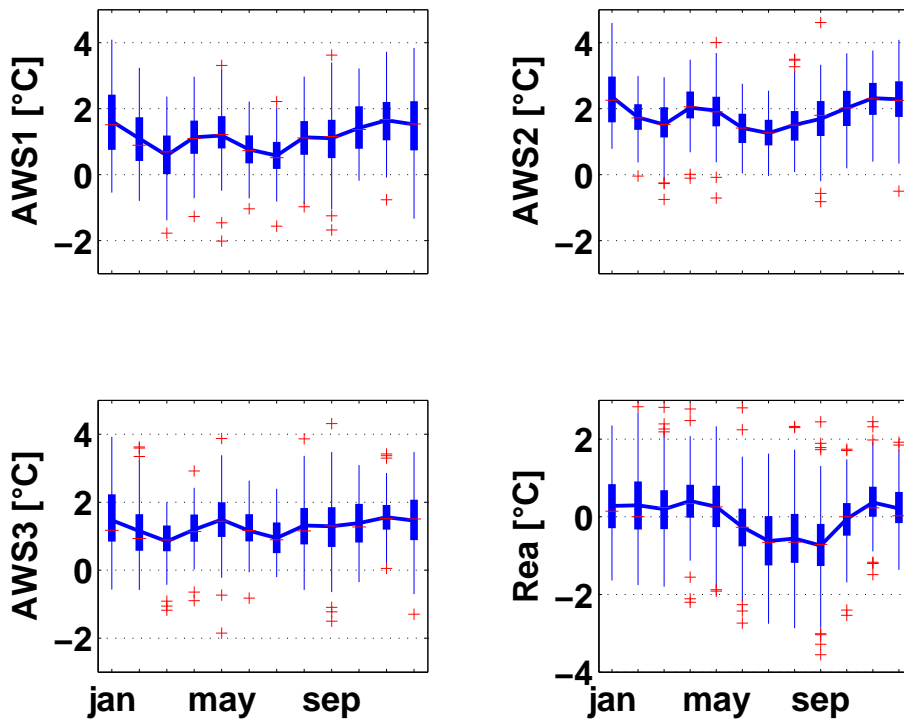


Fig. 2. Statistics of daily air temperature time series at AWS1 (5050 m a.s.l.), AWS2 (4820 m a.s.l.), and AWS3 (4950 m a.s.l.), and of the a priori selected predictor rea-ens-air (as defined in the text), for each month of the year (abscissa: January–December). Shown are the means (blue solid line) and the medians (red dashes). The edges of the thick blue bars are the 25th and the 75th percentiles. The thin blue bars extend to the most extreme data not considered as outliers, and the red crosses are the outliers. All statistics are computed for period 1 (July 2006–December 2009).

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



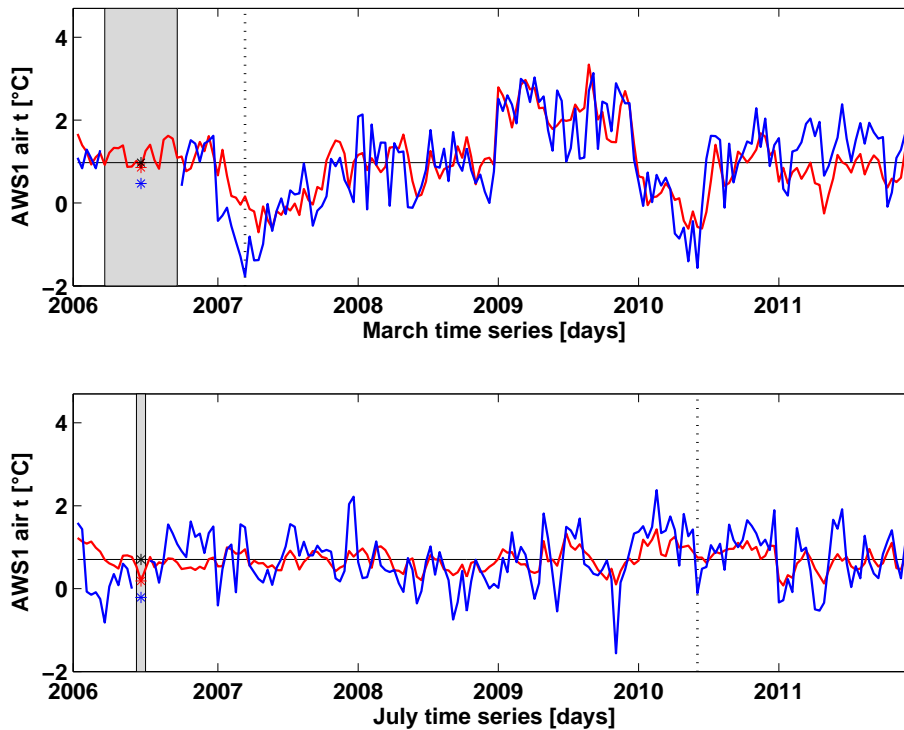


Fig. 3. Example of the skill estimation procedure using cross-validation as described in Sects. 4.3 and 4.4, for AWS1 predictands and the rea-ens-air predictor in March (top) and July (bottom): $y(t)$ (blue line), $\hat{y}(t)$ (red line), y_r (black line), $y_V(cv)$ (blue star), $\hat{y}_V(cv)$ (red star), and $y_r(cv)$ (black star). The grey shaded area highlights the observations left out in the model calibration process. The vertical dashed line indicates the minimum number of observations necessary for obtaining statistical significant skill for the respective calendar month (the same numbers as given in Fig. 5).

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

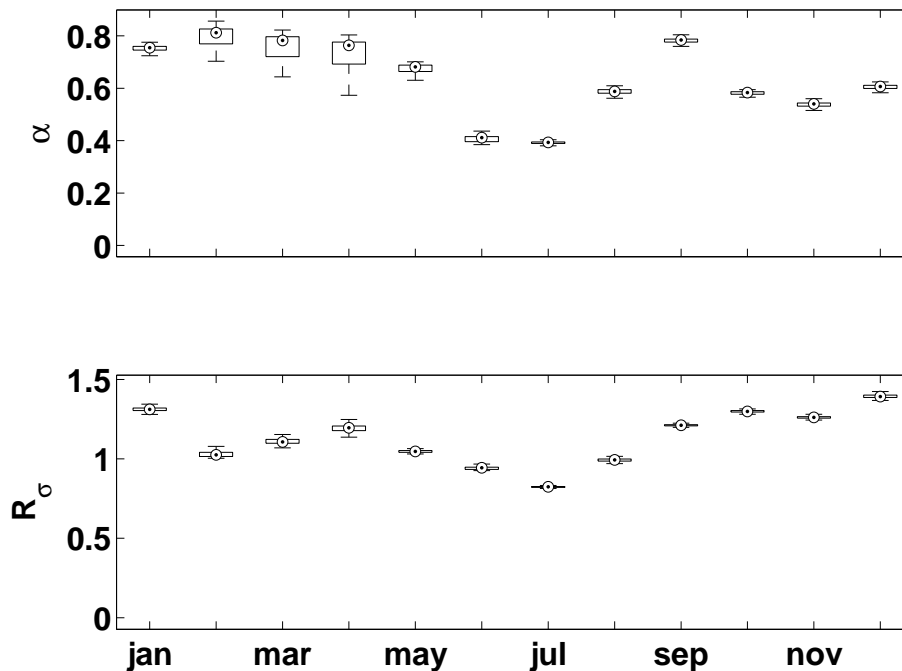


Fig. 4. Medians (dots) of the downscaling model parameters α (top) and R_σ (bottom) estimated by cross-validation for the AWS1 predictands and the rea-ens-air predictor, for all calendar months. The edges of the boxes are the 25th and the 75th percentiles. The dashes extend to the most extreme data not considered as outliers.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

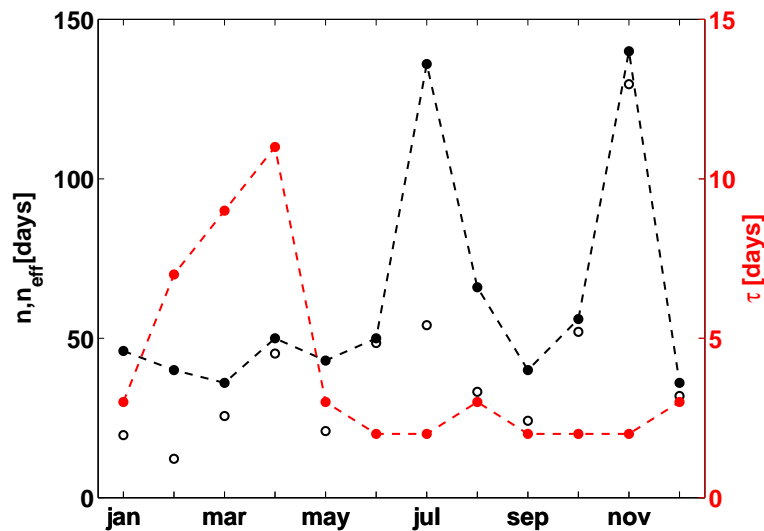


Fig. 5. The black filled circles show the minimum number of AWS1 air temperature observations (n) required for each calendar month's model (based on rea-ens-air) to obtain statistically significant skill (as described in the text). The black empty circles show the effective sample size, n_{eff} , associated with n . The red circles (with axes to the right) show the decorrelation time (τ), for each month's daily air temperature time series measured at AWS1.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

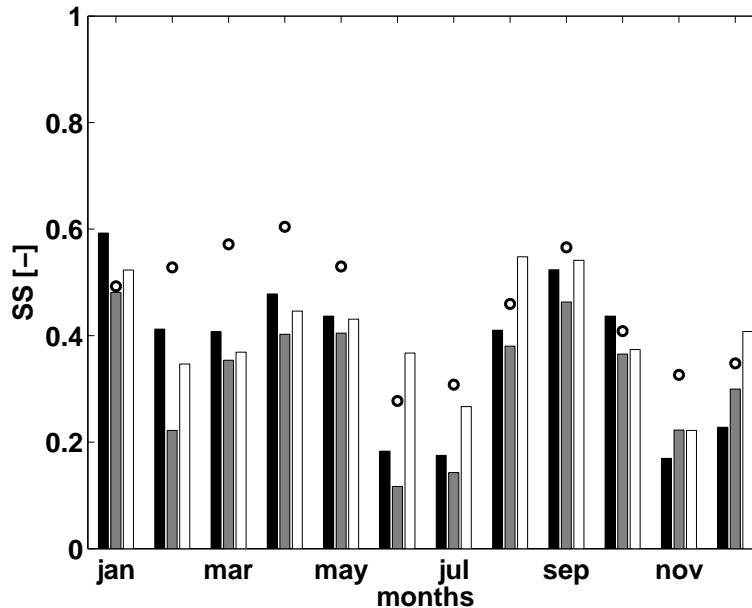


Fig. 6. Values of SS for AWS1 (black), AWS2 (gray), and AWS3 (white) for the predictor re-ens-air for period 1. The circles show the respective values of SS at AWS1 for its entire data period (period 2).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



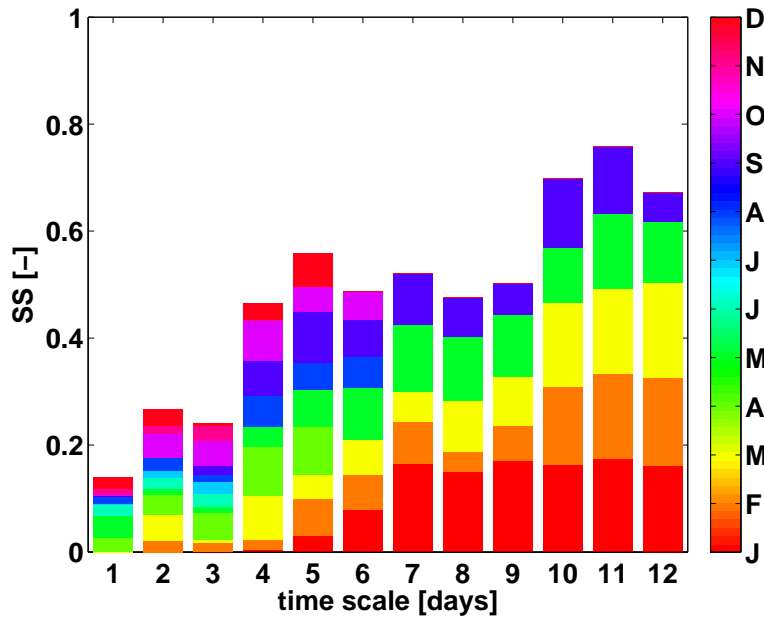


Fig. 7. SS for the different time scales (from one-days averages to twelve-days averages) and the different months (January–December), for which SS is found to be significant at each time scale, for the AWS1 predictand and the predictor rea-ens-air. The values are divided by the number of months, for which SS is found to be significant at each time scale, such that the sum of the bars is equivalent to the average SS at each time scale.

Skill assessment of reanalysis data for air temperature

M. Hofer et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

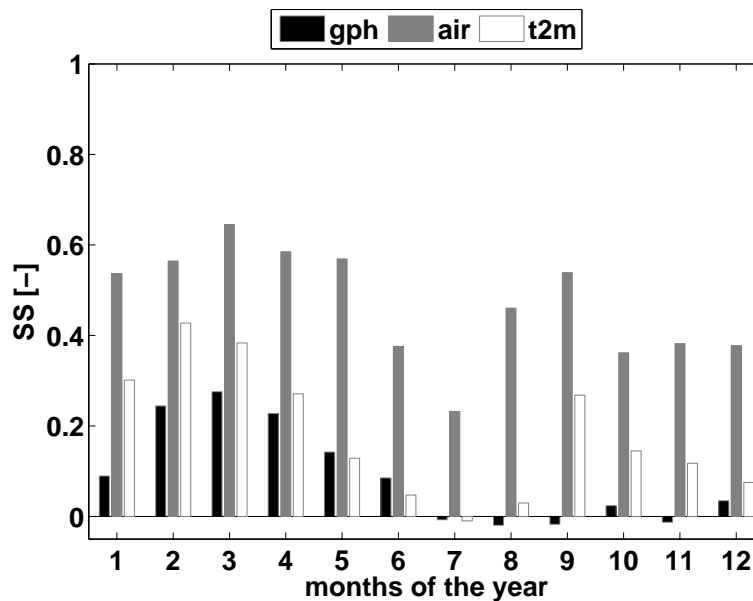


Fig. 8. Values of SS for each month (here for AWS1), and the three predictors assessed in Sect. with significant results.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Skill assessment of reanalysis data for air temperature

M. Hofer et al.

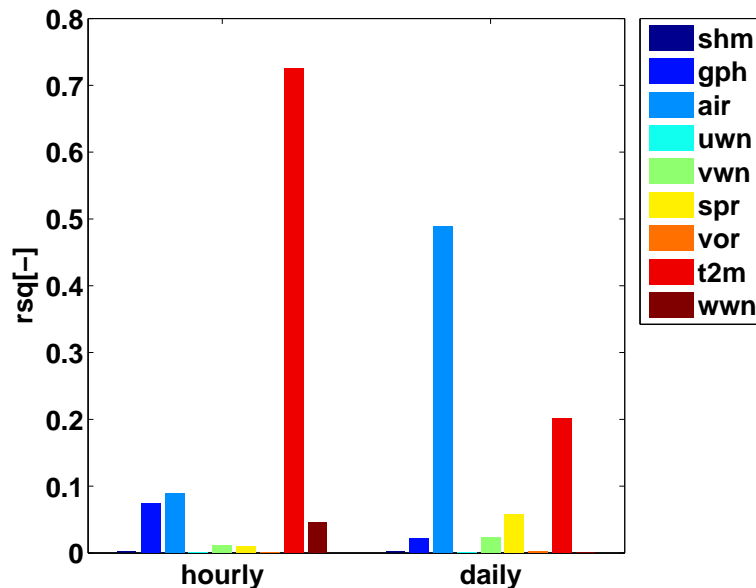


Fig. 9. r^2 for six-hourly and daily (all-month) time series of AWS1 air temperature and all predictors assessed in Sect. .

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

