



## Empirical-statistical downscaling of reanalysis data to high-resolution air temperature and specific humidity above a glacier surface (Cordillera Blanca, Peru)

Marlis Hofer,<sup>1</sup> Thomas Mölg,<sup>1</sup> Ben Marzeion,<sup>1</sup> and Georg Kaser<sup>1</sup>

Received 29 May 2009; revised 8 January 2010; accepted 3 February 2010; published 25 June 2010.

[1] Recently initiated observation networks in the Cordillera Blanca (Peru) provide temporally high-resolution, yet short-term, atmospheric data. The aim of this study is to extend the existing time series into the past. We present an empirical-statistical downscaling (ESD) model that links 6-hourly National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) reanalysis data to air temperature and specific humidity, measured at the tropical glacier Artesonraju (northern Cordillera Blanca). The ESD modeling procedure includes combined empirical orthogonal function and multiple regression analyses and a double cross-validation scheme for model evaluation. Apart from the selection of predictor fields, the modeling procedure is automated and does not include subjective choices. We assess the ESD model sensitivity to the predictor choice using both single-field and mixed-field predictors. Statistical transfer functions are derived individually for different months and times of day. The forecast skill largely depends on month and time of day, ranging from 0 to 0.8. The mixed-field predictors perform better than the single-field predictors. The ESD model shows added value, at all time scales, against simpler reference models (e.g., the direct use of reanalysis grid point values). The ESD model forecast 1960–2008 clearly reflects interannual variability related to the El Niño/Southern Oscillation but is sensitive to the chosen predictor type.

**Citation:** Hofer, M., T. Mölg, B. Marzeion, and G. Kaser (2010), 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.

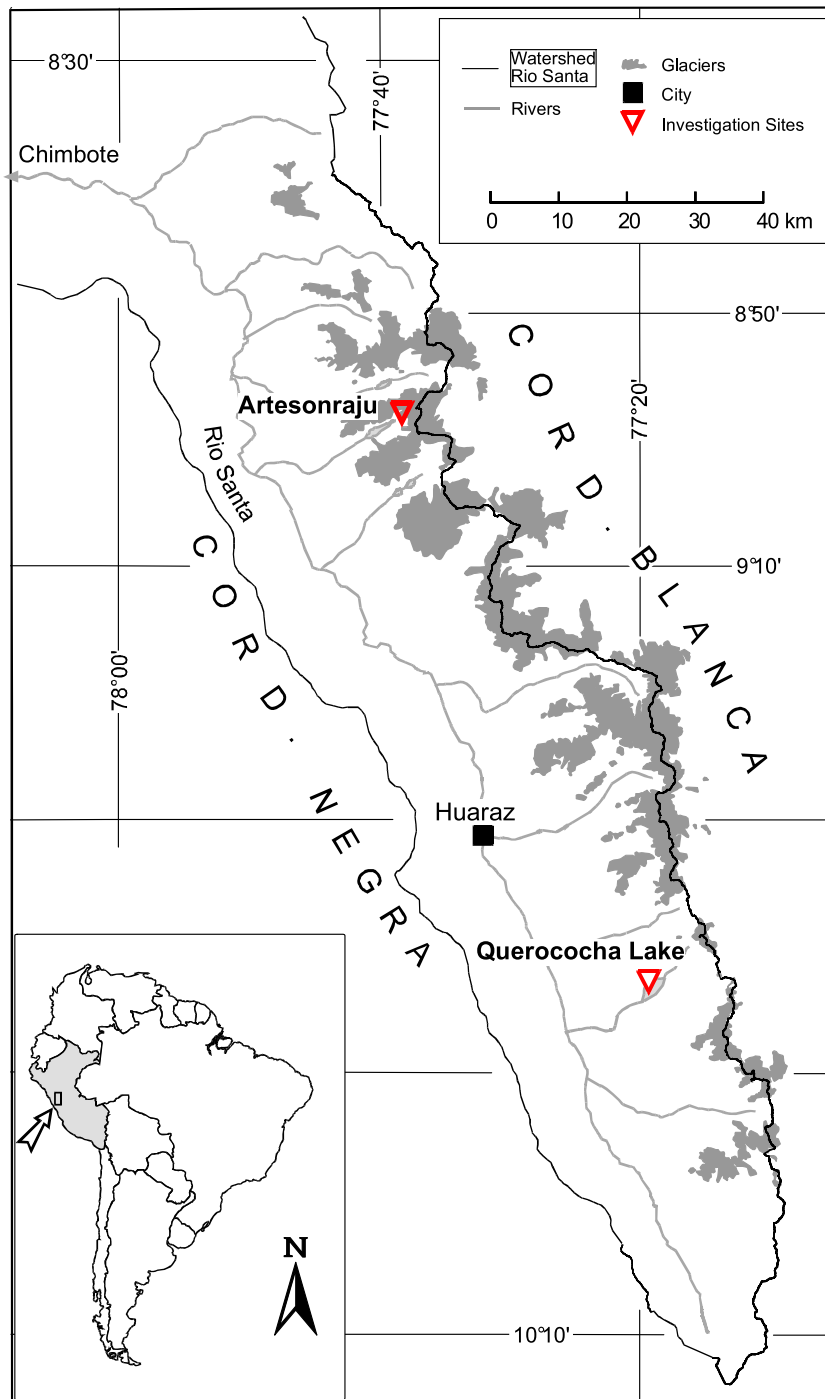
### 1. Introduction

[2] Mountain glaciers are widely recognized as sensitive indicators of climatic change and variability. Glaciers in the tropics are of particular interest, because they respond faster to climatic changes than glaciers in the midlatitudes or high latitudes [Kaser, 1995]. Located at high altitudes, tropical glaciers moreover provide important climate information about the tropical free troposphere. Knowledge about climate change in this zone is incomplete, because the observational data are sparse and controversial [Karl *et al.*, 2006; Kaser, 1995; Trenberth *et al.*, 2007]. More than 99% of all tropical glaciers (in terms of surface area) are located in the South American Andes and roughly 70% are in Peru [Kaser and Osmaston, 2002]. Peru's most extensively glacier-covered mountain range is the Cordillera Blanca that harbors 25% of the tropical glaciers. Figure 1 shows a map of the Cordillera Blanca with the main glaciers and the Rio Santa watershed [Kaser and Osmaston, 2002]. Similar to the

global trends, substantial glacier shrinkage in the Cordillera Blanca over the 20th century has been documented [e.g., Ames, 1998; Georges, 2004; Silverio and Jaquet, 2005]. Modeling studies predict that glaciers in the Cordillera Blanca will continue to shrink in the future, with drastic consequences for the runoff [Juen, 2006; Juen *et al.*, 2007]. The runoff from glaciers is of great socioeconomic concern in this region, because it provides an important service as seasonal water regulator [e.g., Kaser *et al.*, 2003]. During the dry season, almost all freshwater used by agriculture, industry, and households in the Rio Santa Valley originates from glaciers [e.g., Mark and Seltzer, 2003]. To estimate potential impacts of climate change on glacier systems in the Cordillera Blanca is thus of primary importance for water resource management in this extensively populated and developing region [e.g., Juen, 2006].

[3] The formulation of the energy balance at the glacier surface represents a deterministic approach to quantifying atmospheric controls of glacier mass loss [e.g., Kuhn, 1989]. Surface energy balance studies conducted on tropical glaciers [e.g., Favier *et al.*, 2004; Hastenrath, 1978; Mölg and Hardy, 2004; Wagnon *et al.*, 1999] reveal that, apart from air temperature, tropical glaciers are particularly sensitive to moisture [e.g., Kaser *et al.*, 2005]. Surface energy balance

<sup>1</sup>Centre for Climate and Cryosphere, Faculty of Geo- and Atmospheric Sciences, University of Innsbruck, Innsbruck, Austria.

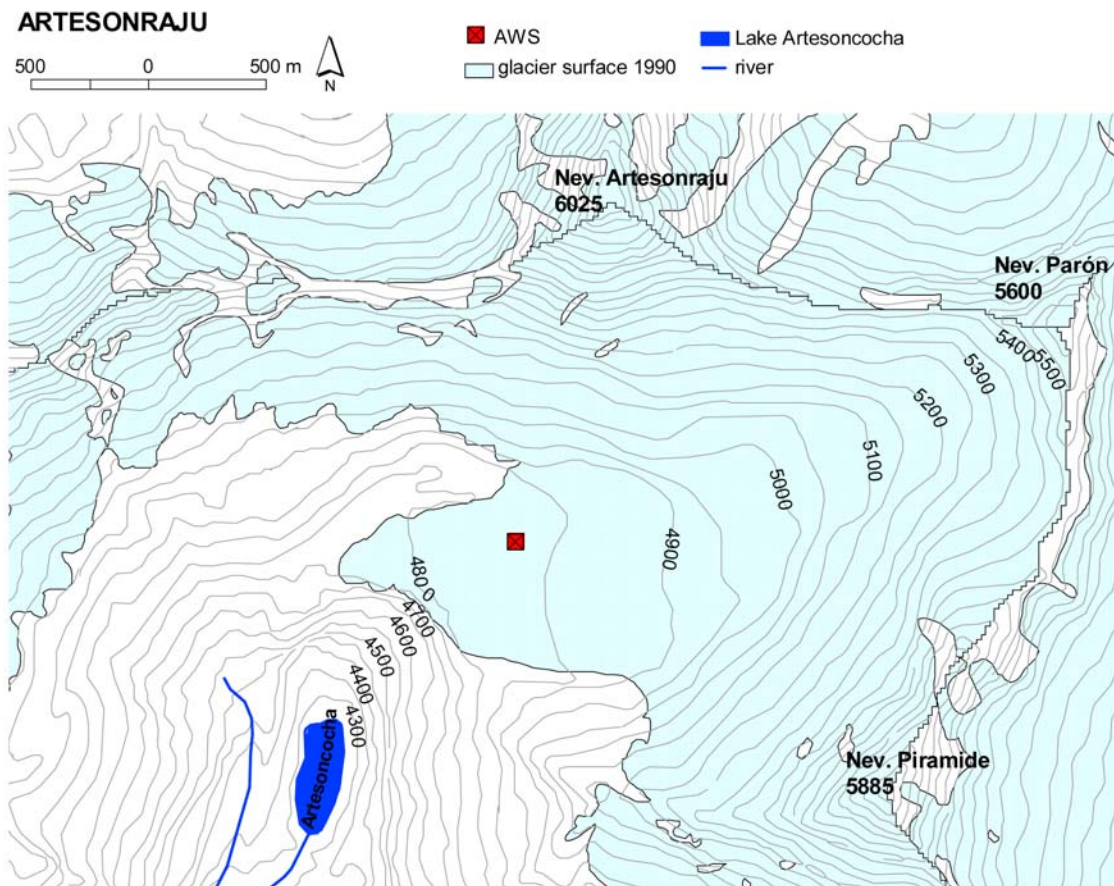


**Figure 1.** Map of the Cordillera Blanca with main glaciers and the Rio Santa watershed. The positions of the glacier Artesonraju and Lake Querococha are indicated.

investigations in the Cordillera Blanca are generally limited to short periods, because they require a variety of on-site observations at high temporal resolution for associated modeling. Measurements on the remote tropical high glaciers, however, are logistically difficult to carry out and have therefore remained rare. In 1999, the Innsbruck Tropical Glaciology Group (ITGG, Austria) and the Institut de Recherche pour le Développement (IRD, France) initiated an observation network on glaciers in the Cordillera Blanca

by installing automatic weather stations (AWSs) [see Juen, 2006].

[4] The goal of this study is to extend high-resolution but short-term meteorological time series on tropical glaciers to the past. Therefore, we present an empirical-statistical downscaling (ESD) scheme for translating coarse-scale reanalysis data to the local meteorological conditions. Our approach is particular in the context of ESD for mainly two reasons. First, the data series for calibration cover only a very short period (approximately 2 years, see further



**Figure 2.** Glacier Artesonraju and surroundings topography, glacier surface of 1990, and position of AWS.

section 2). Second, unlike most ESD studies [e.g., *Benestad et al.*, 2008], we aim to forecast variables at a very high temporal resolution (i.e., 6-hourly values). Our target variables are two of the key drivers in the surface energy balance of tropical glaciers: air temperature and specific humidity, here above a tropical glacier surface in a very complex orographic setting. These two variables govern turbulent heat exchange to a large extent [e.g., *Mölg and Hardy*, 2004] and also influence radiative energy [e.g., *Mölg et al.*, 2009].

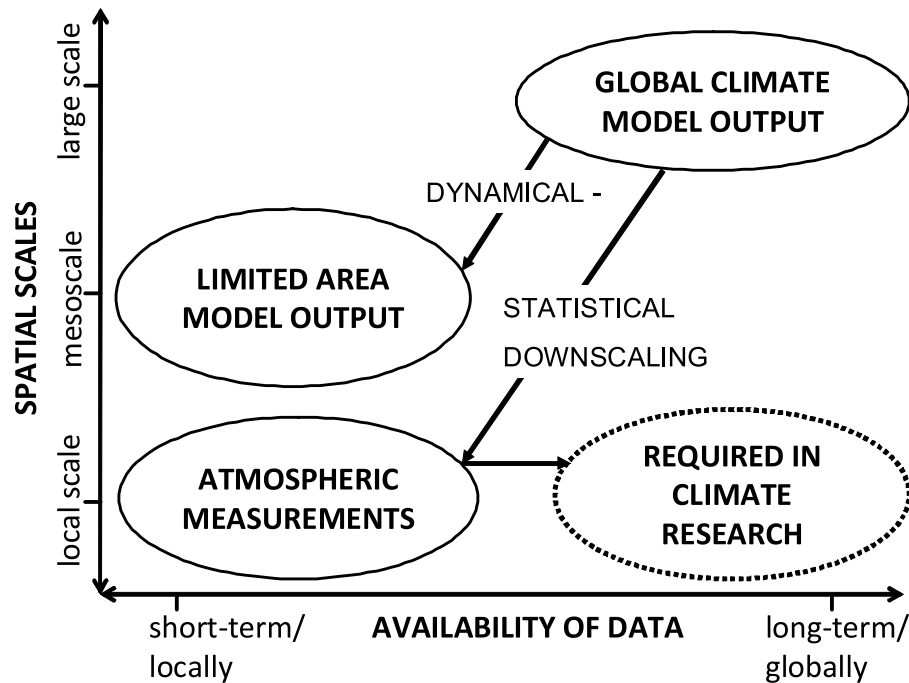
[5] We want to explore the possibilities and limitations of ESD with regard to reaching the aims stated above. Study site and observational data are introduced in section 2. We describe the employed ESD modeling procedure (including the predictor selection) in section 3. The results of the ESD model are presented and evaluated in section 4. Finally, we summarize our conclusions in section 5.

## 2. Study Site and Observations

[6] The investigation area in this study is the glacier Artesonraju, located at 8.9°S and 77.4°W in the Paron Valley (northern Cordillera Blanca, cf. Figure 1). The glacier covers an area of about 5.7 km<sup>2</sup> and reaches from 6025 m above sea level (asl) down to approximately 4750 m asl [see *Juen*, 2006]. In 2004, the ITGG and the IRD installed four

AWSs at and around the glacier. The maintenance of these stations has been difficult due to logistical problems. To date, approximately 2 year records are available from each station, and it is not decided yet if they can be maintained in the future. In this study, we mainly focus on data from one AWS at approximately 4850 m asl operated by the IRD, because it is the only one of the four AWSs that is located on the glacier and that is equipped with air temperature and humidity sensors (additionally to wind speed/direction, incoming/outgoing shortwave and longwave radiation). Figure 2 shows the glacier Artesonraju and the position of the AWS on the glacier surface (hereafter simply referred to as AWS). The measurement period at AWS extends from March 2004 to May 2006 (hereafter referred to as calibration period). Because of several instrument failures, the data series is not continuous in time. Within the calibration period, data are missing from November 2004 to February 2005 (approximately 15% of the data; periods of missing data are marked in Figures 9 and 10 presented later). We refer to the work of *Juen* [2006] for a detailed description of the study site and measurement setup.

[7] In the entire Cordillera Blanca, only one long-term air temperature record exists from a station operated by Electroperu at Lake Querococha (3980m asl, 77.3°W 9.72°S, about 100 km distant from the glacier Artesonraju, see Figure 1). The observations at Querococha are available



**Figure 3.** The role of downscaling techniques to transfer atmospheric data with high availability in time and space (abscissa) but low spatial resolution (ordinate) to the higher spatial resolutions required for atmospheric research. Here, we use the more general term “global climate model output,” but it is true also for reanalysis data.

from 1965 to 1994 (only monthly means) and are presented in section 4.5, where annual means of the ESD model forecast 1960–2008 are discussed.

### 3. The Empirical-Statistical Downscaling (ESD) Model

#### 3.1. What Is ESD and Why Do We Need It?

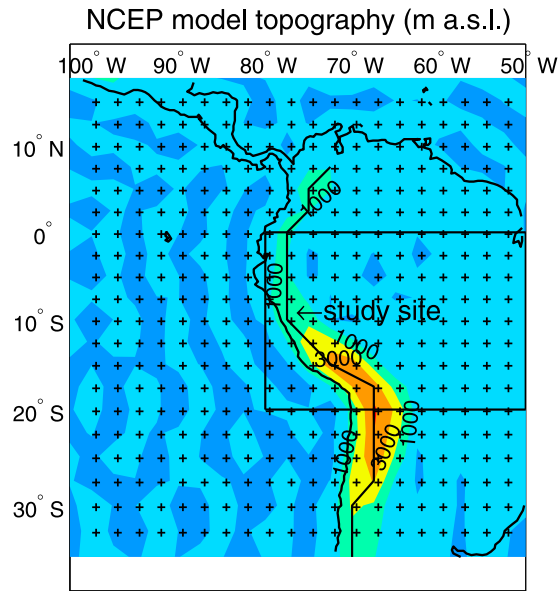
[8] General circulation models (GCMs) are powerful tools to simulate past, present, and future climate and weather, but their spatially coarse resolution does not realistically resolve local and regional atmospheric states [Grotch and MacCracken, 1991]. This is true also for reanalysis data (introduced later), which are closely related to GCM output. Hence, techniques have emerged to downscale GCM output to higher spatial resolutions as required for climate change impact studies. Figure 3 depicts the role of downscaling to transfer GCM output (and reanalysis data) to higher spatial scales. The development of downscaling strategies is an important focus in regional climate research [Christensen *et al.*, 2007]. Two different approaches exist: empirical-statistical downscaling (ESD) and dynamical downscaling. Dynamical downscaling is comprehensive limited area modeling with initial and lateral boundary conditions provided by the GCM. Yet the limited area models are computationally expensive: for long-term simulations, their spatial resolutions are restricted to  $\sim 10$  km. The added value of regional climate models over GCMs is still under debate [e.g., Prömmel *et al.*, 2009], and usually the regional climate model output needs further downscaling steps before it can be applied in impact studies [e.g., Früh *et al.*, 2006; Salzmann, 2006].

[9] ESD, the technique applied in this study, is a statistical model that transfers large-scale model output to the local scale. One advantage of ESD against dynamical downscaling is its much lower computational demand. However, it can be applied only for sites where observations are available for model calibration. A number of ESD techniques exist for widely differing applications, and some of them focus on mountain or glacier environments [e.g., Matulla, 2005; Radic and Hock, 2006; Reichert and Bengtsson, 2002; Salzmann, 2006]. However, most ESD studies concentrate on variables in the monthly time resolution [Benestad *et al.*, 2008]. We aim for high temporal resolutions, because a further scope of this study is to provide input data for process-based glacier mass balance models required at subdaily time steps [e.g., Mölg *et al.*, 2009]. Thus, in terms of temporal resolution, our ESD approach is novel in climate research. In the context of numerical weather prediction, similar techniques have been explored for decades (i.e., the perfect prog technique) [see Klein and Glahn, 1974]. With regard to downscaling at a subdaily time scale in complex and glacier-covered mountain environments, however, this study represents a completely new approach.

[10] In its simplest form, ESD can be expressed as the random or deterministic function  $f$ , such that

$$Y = f(X) + \varepsilon(t), \quad (1)$$

where  $Y$  is the predictand (the local scale target variable),  $X$  is the large-scale predictor, and  $\varepsilon$  is a random error (or noise) [e.g., von Storch *et al.*, 2000]. In this study, the predictands  $Y$  are 6-hourly variations of air temperature and specific humidity above the surface of the glacier Arte-



**Figure 4.** NCEP model topography over South America. The crosses are grid points. The black line connects the highest locations in the NCEP model topography. The black rectangle shows the horizontal area finally included in the ESD model.

sonraju (measured at AWS). As predictors ( $X$ ), we use reanalysis data.

### 3.2. Choice of Predictor Fields

[11] Reanalysis data are the state-of-the-art archive of past meteorological variables spanning the entire atmosphere. They are produced by data assimilation, with a modeling system kept frozen over the simulation period and using quality-controlled observations. Reanalysis data are available, e.g., from the National Centers for Environmental Prediction (NCEP)/National Center for Atmospheric Research (NCAR) [Kalnay *et al.*, 1996], the European Center for Medium-Range Weather Forecasts (ECMWF) [Uppala *et al.*, 2005], or the Japan Meteorological Agency (JMA) [Kazutoshi *et al.*, 2007]. In this study, we use the NCEP/NCAR reanalysis data, because they have shown skill in several studies focusing on the South American Andes [e.g., Garreaud *et al.*, 2003]. NCEP/NCAR reanalysis data are available in subdaily time resolution (6-hourly) for the period from 1948 to present, with a spatial resolution of approximately 210 km (T62) and 28 “sigma hour” levels in the vertical.

[12] Only a few studies have systematically assessed the skill of different predictors in terms of variable types [e.g., Cavazos and Hewitson, 2005] or downscaling domain [e.g., Brinkmann, 2002], and there is no consensus on the most appropriate choice [Fowler *et al.*, 2007]. ESD is based on the following three assumptions for suitable predictors [e.g., Benestad *et al.*, 2008]. The predictors must (1) have a physical relationship to the predictand, (2) be reliably represented by the reanalysis data or GCM, and (3) reflect climate change. In this study, criterion 1 is specified further as a linear physical relationship, because we apply linear regression techniques. Criterion 2 refers to the reliability of variables in the NCEP/NCAR reanalysis data. According to

the availability of observations in the data assimilation, the NCEP/NCAR reanalysis variables are classified as type A, B, and C [Kalnay *et al.*, 1996]. Type A variables (such as upper air temperature, wind, and geopotential height) are most reliable, because they include measurements. Type C variables (e.g., surface fluxes, heating rates, or precipitation) are determined completely by the model and therefore associated with larger uncertainty. Moisture and surface variables are classified as type B; they partially include direct measurements. If the ESD model is based on the assumption that the predictors represent the true large-scale atmospheric state (analogous to the Perfect Prog approach [Klein and Glahn, 1974]), type C variables should not be included. However, MOS approaches (model output statistics [Klein and Glahn, 1974]) also apply type C variables as the predictors [Widmann *et al.*, 2003]. In this study, we use type A and type B variables as predictors (presented later in this section). Condition 3 implies that the predictor includes climatic trends or variability changes that affect the predictand; otherwise, these changes cannot be captured by the ESD model. This postulation is linked to the stationarity assumption, which is a major uncertainty in ESD [e.g., Benestad *et al.*, 2008]. If the local predictand is subject to temporal changes not present in the predictors, the ESD transfer function becomes less appropriate with time. Several studies suggest to use multiple variable types as predictors to account for each changing variable that affects the local climate [e.g., Fowler *et al.*, 2007; Hewitson and Crane, 2006]. In this study, both single-field and mixed-field predictors are applied (see section 4.1).

[13] Apart from the choice of variable types as predictors, ESD results also depend on the definition of the downscaling domain (i.e., the grid point area of the variable fields included in the ESD model) [Fowler *et al.*, 2007]. The optimum choice of the predictor domain is generally not limited to the closest grid points around the study site but includes the most important synoptic patterns around and upstream of the study area. Unrelated atmospheric variability though has shown to negatively impact the results in ESD [see Benestad *et al.*, 2008]. In terms of the climatic situation, the Cordillera Blanca is located in the outer tropics [Kaser *et al.*, 1996; Kaser and Osmaston, 2002] and the large-scale flow is dominated by easterly wind directions all year round [e.g., Georges, 2005]. Anomalies of the large-scale flow, related to the strength and location of the Bolivian High, are associated with precipitation variability in the Andes [Vuille *et al.*, 2008]. Near-surface moisture in the tropical lowlands also plays an important role for precipitation mechanisms [Garreaud *et al.*, 2003; Vuille and Keimig, 2004].

[14] Since the Cordillera Blanca is a very complex mountain range, it is important to consider the NCEP model topography. Figure 4 shows the location of the glacier Artesonraju within the NCEP model topography: it is located east of the mountain crest in the model, whereas in reality, the glacier faces mainly toward west and ends up in high peaks in the north and east. The four grid points closest to the study site are located between 1000 and 2000 m asl, whereas the measurement site (AWS) is located at 4850 m asl, with surrounding peaks above 6000 m asl. Thus, the NCEP topography differs from the real orographic situation of the study site and therefore, at subdaily time scales, the

direct use of grid point variables from NCEP/NCAR re-analyses for impact studies is not deemed adequate.

[15] For all variable fields, we define first a preliminary horizontal grid point domain that extends from 17.5°N to 35°S and 102.5°W to 50°W (entire area in Figure 4). Due to the coarse NCEP model topography, the grid points at the model surface are not located in the pressure level of the study site. Consequently, the vertical allocation of the study site in the NCEP model is undetermined. In this study, we therefore consider both variable fields in the lower troposphere (1000 and 800 hPa) and the mid troposphere (600 and 400 hPa). Midtropospheric variables have the advantage to be more reliable in the NCEP/NCAR reanalysis data (type A), whereas surface variables are influenced both by the model and by measurements (type B) [Kalnay *et al.*, 1996]. We define the definite downscaling domain in this study by means of correlation maps between the preliminary predictor fields and the predictands to identify regions with high correlation. From a theoretical perspective, this is suggested by the similarity of principal component–multiple regression analysis (the technique applied here, introduced later) with one-dimensional maximum covariance analysis. The latter uses the regression map (which has a spatial structure very similar to the correlation map) to define the weights of the local predictors [e.g., Widmann, 2005].

[16] In this study, we select a preliminary predictor set from the NCEP/NCAR reanalysis data, including the variables air temperature  $a$ , specific humidity  $s$ , the horizontal wind components  $u$  and  $v$ , and the geopotential height  $h$ . The variables  $a$  and  $s$  are the first candidates to be included in the multiple predictor set, because the physical linkage to the local predictands is most obvious for  $a$  and  $s$ , and they are important variables with regard to climate change [e.g., Trenberth *et al.*, 2007]. We also include a set of circulation-related predictors ( $u$ ,  $v$ , and  $h$ ), which are more frequently applied variable types in ESD [Cavazos and Hewitson, 2005; Fowler *et al.*, 2007] not only because they explain a large proportion of variance in the predictands but also because of the long temporal records and the reliability of GCMs in simulating these fields. Starting from this preliminary predictor set, we try to find out the most appropriate fields to be included in the ESD model based on multiple regression analyses. The results of the downscaling domain and predictor variable selection are presented in section 4.1.

### 3.3. Transfer Functions for Individual Months and Times of Day

[17] In the outer tropical Cordillera Blanca, seasonality is characterized by variations in atmospheric moisture rather than by temperature variations. The core wet season is defined from January to March, the core dry season is from June to August, and the two transitional seasons span April to May and September to December (definitions are based on precipitation climatology [Juen, 2006; Niedertscheider, 1990]). It is important to note that the seasons are not of equal length over the year: this is a typical behavior of outer tropical climates and is governed by oscillation of the Intertropical Convergence Zone [e.g., Kaser and Osmaston, 2002].

[18] To account for seasonal (and diurnal) changes, ESD variables are generally standardized individually for each

month (and time of day, in the case of ESD in subdaily time scales) [e.g., Kim *et al.*, 1984; Schoof and Pryor, 2001; Xoplaki *et al.*, 2003]. In this study, statistical transfer functions are derived separately for distinct months and times of day, because we assume that physical processes determining micrometeorological conditions change for different months and times of day. In particular, diurnal air temperature maxima above temperate ice surfaces generally show small variations due to the stabilizing effect of the ice underneath, which cannot exceed 0°C. By contrast, nighttime air temperatures exhibit high variability due to the high thermal emissivity of ice [e.g., Paterson, 1994]. By applying separate transfer functions for different months and times of day, we include changes of the relationships between local and large-scale variables in addition to changes of means and variances of the variables.

[19] Linear transfer functions are derived for each month of the year and the 4 times of the day when reanalysis data are available: 6, 12, 18, and 0 UTC (corresponding to 1, 7, 19, and 13 h Peruvian local time, LT), hereafter referred to as “month/hour-models” (here, we use the term “hour” instead of “time of day” for brevity; please note that “hour” does not refer to “hourly mean,” but to instantaneous time of day, as defined above). This way, the calibration data set from AWS (March 2004 to May 2006) is split up into 48 subsamples (corresponding to the 48 month/hour-models) of sample sizes  $n$ , each of them including approximately 30–70 observations (data gaps excluded).

### 3.4. Empirical Orthogonal Function (EOF) Analysis and Predictor Preprocessing

[20] Empirical orthogonal function (EOF) analysis is a convenient multivariate technique used in earth sciences to identify dominant variation patterns of random variables [e.g., Hannachi *et al.*, 2008; von Storch and Zwiers, 2001]. Various forms of EOFs are commonly applied in ESD for dimension reduction and to eliminate colinearity in the predictors [e.g., Benestad *et al.*, 2008]. The suitability of ESD techniques based on EOF analysis and related techniques (such as canonical correlation analysis for multivariate predictands [see von Storch and Zwiers, 2001]) has been reported in several studies [e.g., Benestad, 2001; Hertzog and Jacobeit, 2008]. In this study, we use the common terminology EOFs for the spatial and principal components (PCs) for the temporal patterns, e.g., as proposed by Hannachi *et al.* [2008].

[21] We employ EOF analysis for dimensionality reduction in the predictors (the definite predictor fields used are presented later, section 4.1). The predictors’ time series over the entire forecasting period (here, 1960–2008) are divided into 48 subsamples corresponding to the 48 month/times of day (as defined before). The 48 subsamples are transformed individually into single-field and mixed-field EOFs [Kutzbach, 1967]. By applying EOF analysis separately for different months and times of day, the corresponding PCs do not contain annual or diurnal cycles any more but anomalies thereof. Mixed-field EOFs are applied for the predictors containing multiple variable fields (because colinearity appears not only spatially but also amongst the different variable types). Before the calculation of EOFs, the variables are converted into dimensionless  $z$  scores. The  $z$  transformation is a step of data preprocessing especially

important in the case of mixed-field EOFs in order to remove the different physical units of the variables. In this study, the variables at each grid point are normalized individually (there is also the option of removing the field variance for each variable type). Geographical weighting is applied to the horizontal fields to account for increasing grid point density with latitude.

### 3.5. PC Screening Based on Cross Validation

[22] Once the selected predictors are prepared and transformed into EOFs and PCs, respectively, the number of PCs to be finally included as predictors for the month/hour models need to be determined (i.e., PC truncation and screening [e.g., *Wilks*, 2006]). In many studies a priori screening (or prescreening) is applied, according to the percentage of explained variance of PCs in the predictor data (i.e., an a priori determined number of leading PCs are selected). However, such choices are subjective. In this study, the PC selection is based on cross-validated estimates of model error variances.

[23] The procedure applied in this study is moving blocks cross validation [e.g., *Wilks*, 1997; 2006]. The subsamples for each month/hour model are further divided into  $m = n - L + 1$  subsamples of consecutive observations of length  $L$  ( $n$  is the number of observations of each month/hour model). Ordinary least squares regression is then repeated  $m$  times to derive the predictor-predictand relationships by omitting each subsample once. The number of omitted observations  $L$  is determined for each month/hour model based on the autocorrelation functions of the predictand time series, such that  $L = 2(\text{dct} - 1) + 1$ , where  $\text{dct}$  (decorrelation time) is the number of lags for which the serial correlation in the data is close to zero. Note that  $L$  has to be determined individually for each month/hour model, because  $\text{dct}$  varies for the different subsamples.

[24] We define the mean square of the  $m$  differences between the  $m$  developed models and corresponding central withheld data points as forecast mean square error,  $\text{mse}_f$ . By repeating the cross-validation procedure for different numbers of included PCs ( $k$ ) and computing  $\text{mse}_f$  as a function of  $k$ , the optimum  $k$  can be determined. As model selection criterion, we use the Akaike Information Criterion (AIC) [*Akaike*, 1973], given by:

$$\text{AIC}(k) = \ln(\text{mse}_f) + 2\frac{n}{k}, \quad (2)$$

where  $n$  is the number of observations. The optimum number  $k$  of PCs to be included ( $k_{\text{opt}}$ ) is found at the minimum of  $\text{AIC}(k)$ .  $k_{\text{opt}}$  is determined individually for each month/hour model. The use of  $\text{AIC}(k)$  rather than  $\text{mse}_f(k)$  for model selection has the advantage that, for a given value of  $\text{mse}_f$ , the more parsimonious model is selected (i.e., the model with less predictors).

[25] The ESD model (cf. equation (1)) can be expressed now in terms of the month/hour (md) transfer functions, such that

$$y_{\text{md}}(t) = \alpha_{\text{md}} + \sum_{\text{npc}} [\beta_{\text{md}}^{\text{npc}} \text{PC}_{\text{md}}^{\text{npc}}(t)] + \varepsilon_{\text{md}}(t), \quad (3)$$

where  $y$  is the predictand (air temperature or specific humidity) as a function of time  $t$ ,  $\alpha$  and  $\beta$  are intercept and

coefficients of the linear multiple fit for each PC of the respective predictors (single field or mixed-field), and  $\text{npc}$  is the order of included PCs, reaching from 1 to  $k_{\text{opt}}$ .

### 3.6. Model Calibration and Double Cross Validation

[26] The transfer functions (equation (3)) are derived by least squares regression. We use cross validation to estimate the forecast skill of the regression equations. The data are resampled as described in section 3.5. The model fitting procedure, including the determination of  $k_{\text{opt}}$  based on cross validation, is then repeated for each sample, i.e., by double cross validation [e.g., *Michaelsen*, 1987]. The regression coefficients (and uncertainties) are determined as the means (and variances) of coefficients from each repetition of the model fitting procedure in the double cross-validation experiment.

[27] Again, as independent goodness-of-fit measure, for each month/hour model,  $\text{mse}_f$  is calculated based on data withheld from the entire modeling procedure (including the cross-validated PC screening). We define the skill score (SS) as

$$\text{SS} = 1 - \frac{\text{mse}_f}{\text{mse}_{\text{rf}}}, \quad (4)$$

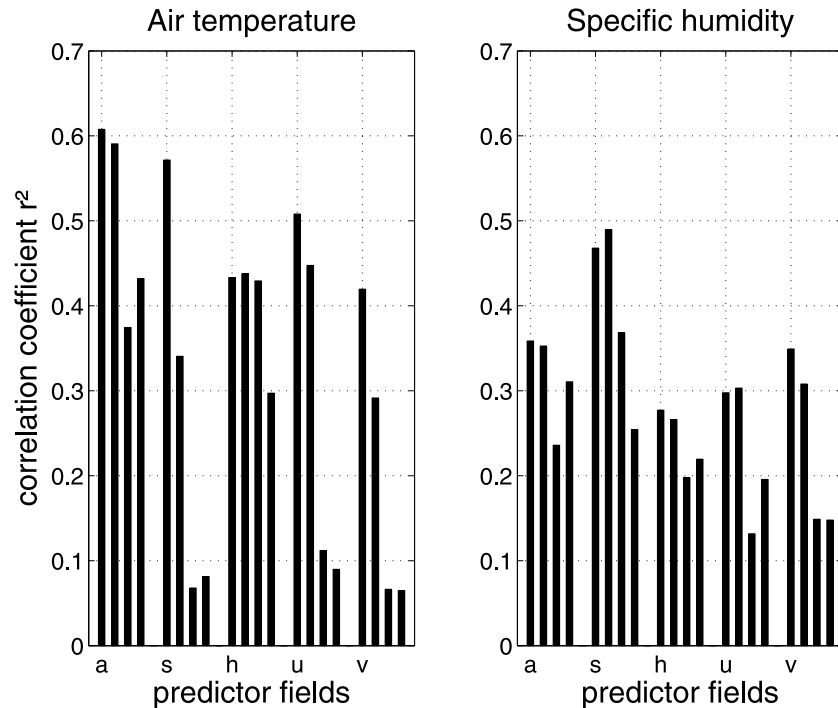
where  $\text{mse}_{\text{rf}}$  is the mse obtained by fitting the mean of each month/time of day to the values. The subscript f (forecast) in  $\text{mse}_{\text{rf}}$  refers to the estimation of mean values using the same cross-validation procedure as for  $\text{mse}_f$ . SS can be interpreted as the added value over fitting the values to a reference model and used to evaluate the month/hour-models in this study, with the reference model being the mean value of the respective month/time of day estimated by cross validation. SS is also known as reduction of error [e.g., *Wilks*, 2006].

## 4. Results and Discussion

### 4.1. Predictor Analysis

[28] We use correlation maps to identify regions with high correlation between predictors and predictands (introduced in section 3.2). The results reveal maximum values above the Bolivian Altiplano of the reanalysis model, the southernmost parts of the selected domain, and Brazil (results not shown). The Bolivian Altiplano, because of its larger horizontal extent, is represented in the NCEP topography more realistically than the Cordillera Blanca (in terms of surface elevation). Consequently, the model climate above the Bolivian Altiplano also fits better to the conditions in the Cordillera Blanca, even if located about  $10^\circ$  farther south. We reduce the preliminary downscaling domain (as defined in section 3.2) to cover an area from  $80^\circ\text{W}$  to  $50^\circ\text{W}$  and  $0$  to  $20^\circ\text{S}$  (black rectangle in Figure 4). The rectangular domain is not symmetric around the study area but shifted upstream, toward the east. This way, the area covers the significant synoptical patterns around and upstream of the study site (section 3.2): the Bolivian High and the Intertropical Convergence Zone [e.g., *Garreaud et al.*, 2003; *Kaser and Osmaston*, 2002].

[29] Figure 5 shows the correlation of the examined predictor fields  $a$ ,  $s$ ,  $h$ ,  $u$ , and  $v$  in the vertical levels 1000, 800, 600, and 400hPa, with the target variables air temperature and specific humidity.  $r^2$  is the adjusted coefficient of



**Figure 5.** Adjusted correlation of determination ( $r^2$ ) between the leading 10 PCs of the fields  $a$ ,  $s$ ,  $h$ ,  $u$ , and  $v$  for the predictands air temperature and specific humidity. The four bars of each variable correspond to the vertical locations of the predictor fields: 1000, 800, 600 and 400 hPa levels from left to right.

determination obtained by multiple linear regression between the predictands and the first 10 PCs of each predictor field ( $\bar{X}$ ). On the basis of Figure 5, the relative performances of the predictors can be evaluated. In the following discussion,  $a_{1000}$  denotes the variable  $a$  (air temperature) in the 1000 hPa level, and so on. The predictand air temperature generally shows higher correlations to the large-scale predictors ( $>0.5$ ) than specific humidity ( $<0.5$ ). The best five predictors in terms of  $r^2$  for air temperature are  $a_{1000}$ ,  $a_{800}$ ,  $s_{1000}$ ,  $u_{1000}$ , and  $u_{800}$  (in that order). The differences in  $r^2$  of the predictor  $a_{1000}$  and other best-performing predictors (e.g.,  $s_{1000}$ ) are small, which indicates considerable covariability amongst these predictor fields. The variable fields in the lower troposphere generally show higher correlations than the predictors from the mid troposphere (possibly because of the larger diurnal variations in the lower troposphere that represent a major part in the predictand variances as well). For specific humidity, the predictor fields  $s_{1000}$  and  $s_{800}$  are the best predictors (out of the examined) with values of  $r^2$  exceeding 0.4. Regression analyses performed to test all possible combinations of two and three predictors out of the examined fields (the 10 leading PCs of the mixed-field EOFs) reveal several variable combinations with comparable high correlations  $r^2$  with the predictands.

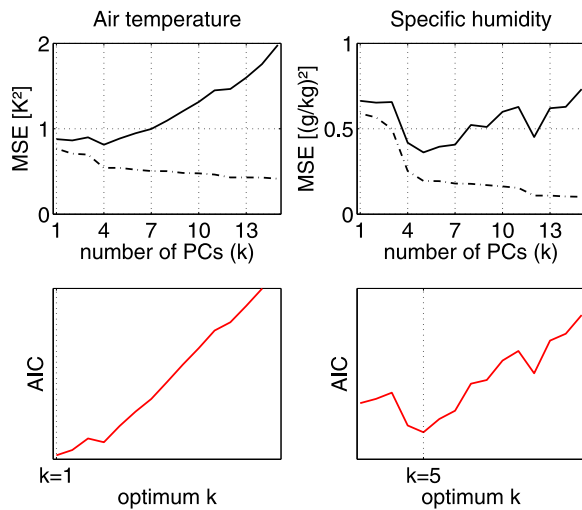
[30] Synthesizing the exploratory analyses above and the climatological criteria discussed in section 3.2, the pool of most appropriate predictors is more constrained but the definitive choice remains ambiguous. In this study, we repeat the modeling procedure (sections 3.3–3.6) with two predictors for each predictand:  $a_{1000}$  for the predictand air temperature,  $s_{1000}$  for the predictand specific humidity, and the combined field of  $a_{1000}$ ,  $s_{1000}$  and  $u_{400}$  for both pre-

dictands. This way, we want to examine the relative merits, as far as possible, of using multiple predictors against single-field predictors in ESD. Even though it does not show to be an important single predictor in Figure 5,  $u_{400}$  shows the best performance as third predictor. The influence of mid and upper tropospheric zonal winds to the local scale variability in the Altiplano is described, e.g., by *Garreaud et al.* [2003], who state that the winds trigger the lower tropospheric moisture transport by downward mixing of zonal momentum. Note that we do not use different predictor variable types or domains for the different month/hour models.

#### 4.2. Discussion of the Month/Hour-Models

[31] In this study, autocorrelation is taken into account by defining the block length of withheld samples in the cross-validation procedure,  $L$ , as described in section 3.5. Most of the month/hour subsamples (more than 90% of all cases) are serially uncorrelated for time lags (dct) of 1–4. The lag-1 autocorrelation is smaller than 0.5 (0.6) for air temperature (specific humidity) for more than 75% of the 48 month/hour subsamples.  $L$  is chosen to range from 1 to 7 in more than 90% of all cases. For  $L = 1$ , the time series is not serially correlated and the moving blocks cross validation reduces to leave-one-out cross validation. The resulting  $m$ , number of independent residuals in the double cross validation (for the  $mse_f$  and  $mse_{rf}$  statistics), ranges from 20 to 60 and is larger than 40 in more than 75% of all cases (data gaps considered).

[32] Figure 6 is an illustration of the PC screening procedure applied in this study.  $mse_f$  and  $mse_h$  (i.e., the hind-cast mse, defined as mse derived from the developmental sample) are plotted against the number of included PCs,  $k$ .



**Figure 6.** Illustration of the PC screening procedure applied in this study using the example of the monthly/hourly case December/13LT for the predictand (left) air temperature and (right) specific humidity and the single-field predictor PCs of  $a_{1000}$  and  $s_{1000}$ , respectively. (top)  $mse_f$  (solid lines) and  $mse_h$  (dashed lines) as a function of  $k$ . (bottom)  $AIC(k)$  (arbitrary units).

In the experimental set up of this study, the number of predictors ( $k$ ) means that PCs of the orders 1 to  $k$  are included as predictors in the multiple regression. As an example, the case of December/13LT is shown, for both predictands air temperature and specific humidity and corresponding single-field predictors ( $a_{1000}$  and  $s_{1000}$ , respectively). While  $mse_h$  generally decreases as  $k$  increases, minimum  $mse_f$  is found at  $k = 4$  (air temperature) and at  $k = 5$  (specific humidity). In the example for air temperature, the minimum AIC is found at  $k = 1$ . Thus, in this case, AIC is the more parsimonious selection criterion than  $mse_f$  and the PCs 1 (and 1–5) are included as predictors in the December/13LT transfer functions for air temperature (and specific humidity).

[33] The entire double cross-validation procedure, including the above-illustrated PC screening, is repeated 192 times for the determination of transfer functions for each of the 48 months/times of day of all four predictand-predictor combinations (single-field and mixed-field predictors for the two predictands, cf. equation (2)). Here, we want to give an overview about the resulting functional relationships. About 50% of all relationships include four or less PCs ( $k_{opt}$  smaller than 4), and about 25% include seven PCs or more. In particular, the more parsimonious models are found for predictand air temperature and for the single-field predictors than for specific humidity or mixed-field EOFs. We refer to the absolute values of regression coefficients for the standardized predictors as weights. The weights of selected PCs are smaller than 0.2 K in about 60% (single-field) and 90% (mixed-field predictors) of all month/hour models of the predictand air temperature and smaller than 0.2 g/kg in about 60% (single-field) and 97% (mixed-field predictor) of all month/hour-models of specific humidity. Relative uncertainties of the regression coefficients are defined here as standard deviations divided by

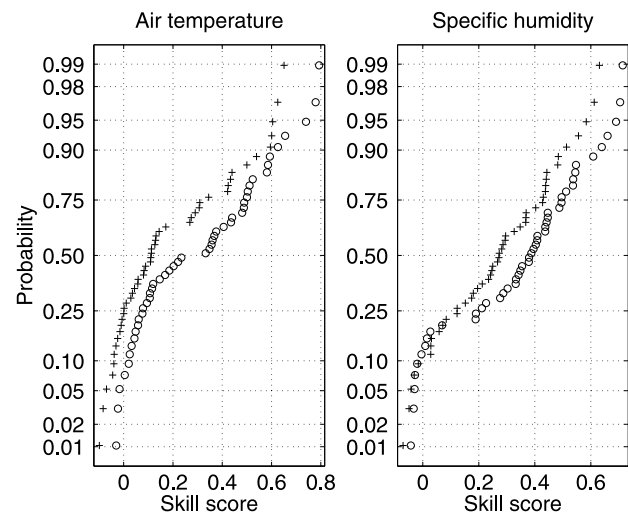
mean values of the coefficients estimated in the double cross-validation procedure (section 3.6). The resulting uncertainties of about 90% of all estimated coefficients are smaller than 0.1%. In more than 70% of all cases, the respective uncertainties of the estimated intercepts are smaller than 0.1% (for both predictands air temperature and specific humidity), and the uncertainties are slightly smaller in the case of mixed-field predictors than single-field predictors.

### 4.3. Cross Validation of the Month/Hour-Models

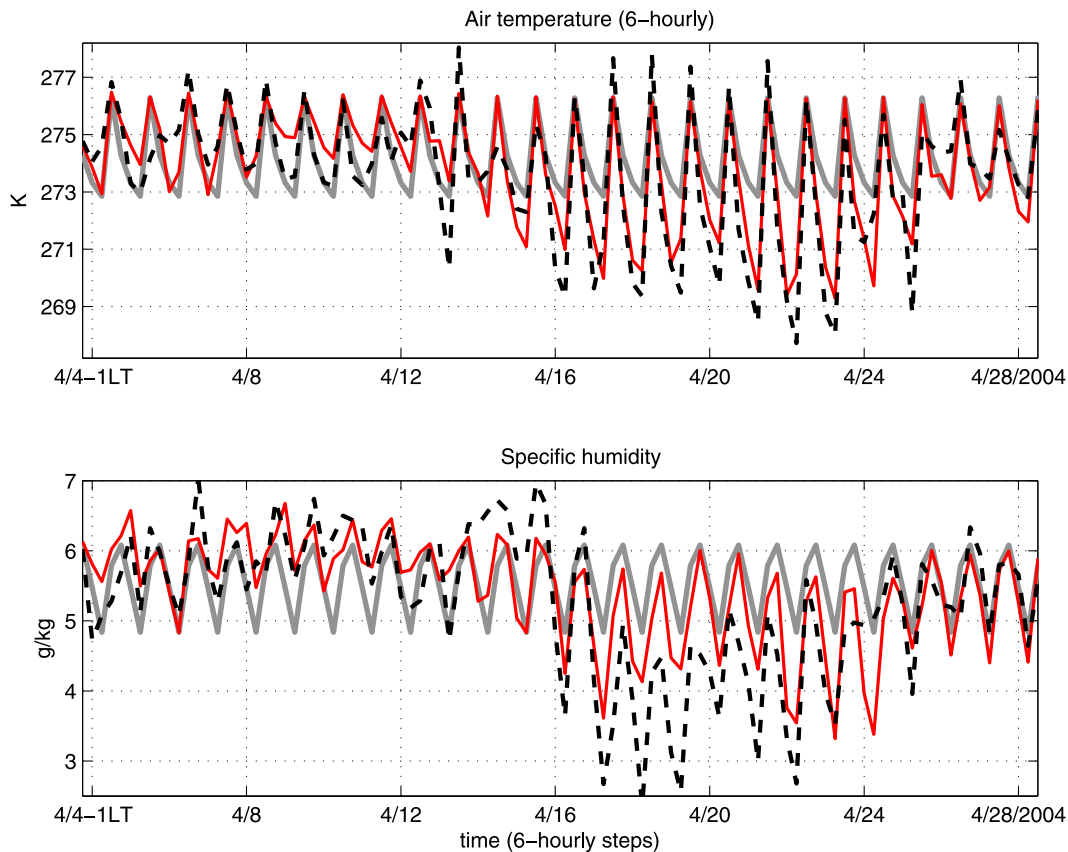
[34] The month/hour-models are tested by moving blocks cross validation to evaluate the forecast skill (i.e., based on data not used in the calibration process). Skill scores (SS) are calculated based on  $mse_f$  as defined in section 3.6. As mentioned earlier, the SS can be interpreted as reduction of error against (in this experiment) the mean of the respective month/hour, estimated by cross validation. The resulting values are displayed in a normal probability plot (Figure 7) to facilitate the discussion of all 48 month/hour-models for each of the four predictor/predictand combinations.

[35] For the combination air temperature and single-field predictor ( $a_{1000}$ ), more than 60% of the month/hour-models SS is smaller than 0.2 (with even slightly negative values), indicating no or only few added value of the respective transfer function over a constant value. Some regression functions also show high values of SS (larger than 0.6). For the mixed-field predictors, however, the respective values are clearly more positive (cf. Figure 7): 50% higher than 0.3, 25% higher than 0.5, and maximum values reaching up to 0.8. The combination specific humidity with mixed-field predictor also shows higher SSs than with the single-field predictor (Figure 7); e.g., about 65% are higher than 0.2 for the single-field predictors and 75% for the mixed-field predictors. Generally, there are less SSs smaller than 0.2 for the predictand specific humidity than for air temperature.

[36] To summarize, the forecast skill of the developed transfer functions quantified by SS varies from 0 to 0.8 but



**Figure 7.** Skill scores of the month/hour transfer functions estimated by double cross validation for the predictand (left) air temperature and (right) specific humidity single-field (crosses) and mixed-field (circles) predictors.



**Figure 8.** Diurnal cycles of (top) air temperature and (bottom) specific humidity at AWS: observations (black dashed), ESD model hindcast (red), and reference model (gray solid). A case study in April 2004 (1LT in the first x-tick means 1 Peruvian local time).

is higher than 0.3 in more than 50% of the cases, and the mixed-field predictors show generally larger skill than the single-field predictors for both predictands.  $SS = 0$  (as is the case for some month/hour-models) means that the deterministic part in equation (1) is 0. In this case, the resulting model reduces to the mean of the respective case. In section 4.4, we want to give a more comprehensive discussion on how the varying skills of the different month/hour models affect the ESD model as complete time series.

#### 4.4. Performance of the ESD Model Against Simpler Reference Models

[37] The modeled time series of the individual month/hour models are put together to form a complete time series over the entire forecasting period 1960–2008 (hereafter referred to as ESD model). Note that in this section, unlike the cross-validation experiment in section 3.6, the performance of the ESD model is assessed based on data used in the calibration process (i.e., hindcast performance). We define reference model 1 (RM1) as the mean diurnal cycles of air temperature and specific humidity for each month of the year, calculated from the AWS observations. Since the predictor PCs in the month/hour-models are standardized, RM1 is equal to the respective intercepts of the month/hour transfer functions (equation (2)). Note that RM1 is not a climatology, given the short averaging period (2004–2006).

[38] Figure 8 shows a case study of 6-hourly observations, RM1, and the ESD model in April 2004. The example

shows a period where melting at the glacier surface occurred [Mölg *et al.*, 2009]. Within the first few days, nighttime air temperatures above  $0^{\circ}\text{C}$  and high values of specific humidity (up to 6 g/kg) indicate that increased longwave incoming radiation constrained the cooling of the ice surface and thus air temperature variability. In the second half of the case study, melt intervals were shorter and air temperature varied more strongly due to lower minima. The linkage between melt occurrence and atmospheric moisture in the tropics is well known: if humidity is low, sublimation consumes much energy and reduces or prevents melting [e.g., Kaser and Osmaston, 2002]. In fact, specific humidity is clearly lower in the second half of the case study period. The ESD model is able to simulate the small diurnal cycles of air temperature associated with high values of specific humidity in the first few days and the drop in humidity accompanied by larger diurnal cycles of air temperature in the second half of the case study. This case study also exemplifies how high-resolution air temperature and specific humidity are linked to glacier melt, which can be simulated in energy balance-based mass balance models [Mölg *et al.*, 2009]. RM1, defined as constant diurnal cycle for each month, cannot show the drop in specific humidity on 16 April. In this context, it is important to see that the ESD model (which per definition reduces to RM1 when the SSs of all the month/hour-models are zero) captures the above-described day-to-day variability that significantly affects glacier mass balance.

**Table 1.** Hindcast Skill Scores (Equation (5))<sup>a</sup>

Predictand Time Scale	Air Temperature			Specific Humidity		
	Subdaily	Daily	Monthly	Subdaily	Daily	Monthly
RM1	0.37	0.47	0.63	0.43	0.56	0.57
RM2	0.75	0.63	0.93	0.67	0.71	0.79

<sup>a</sup>Of the ESD model (mixed-field predictors) against reference models RM1 and RM2 for the two predictands air temperature and specific humidity and different temporal scales.

[39] In order to quantify the added value of the ESD model versus the reference model over the entire calibration period, the hindcast skill score (SSh) is defined as:

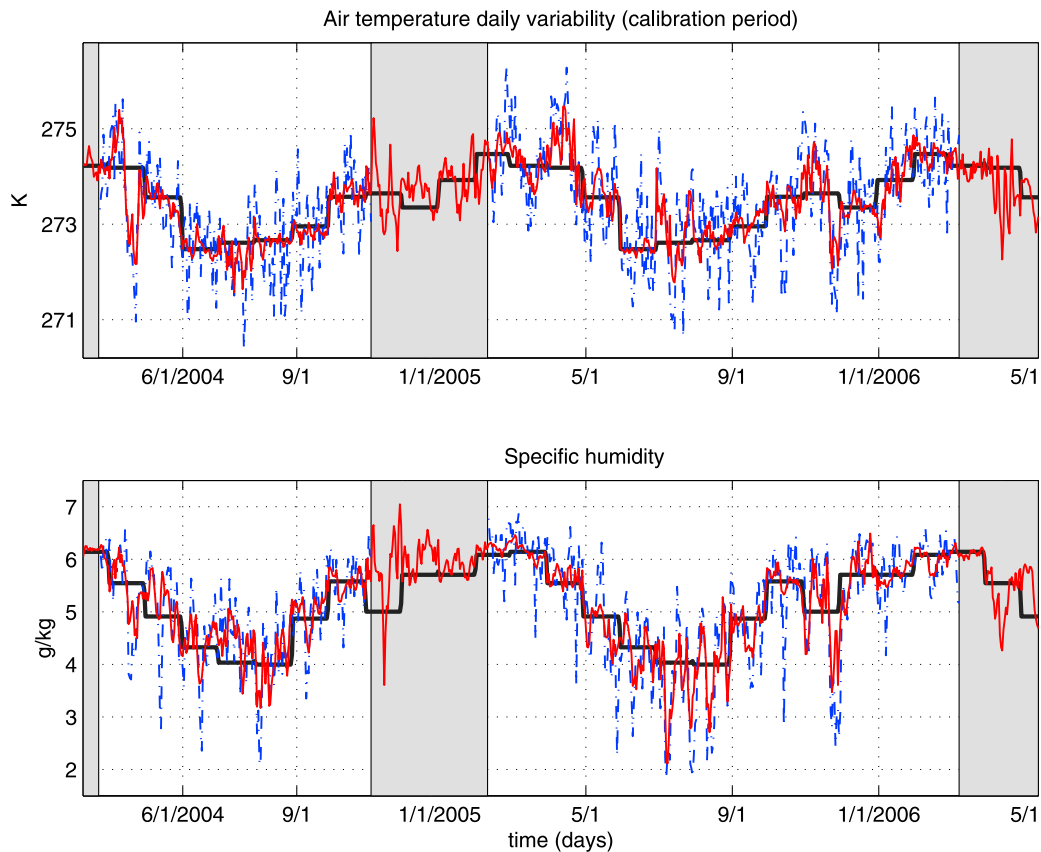
$$\text{SSh} = 1 - \frac{\text{mse}_{h,\text{ESD}}}{\text{mse}_{h,\text{ref}}}, \quad (5)$$

where  $\text{mse}_{h,\text{ESD}}$  and  $\text{mse}_{h,\text{ref}}$  are the hindcast mean square errors of the ESD model and the reference model, respectively (i.e., the mean square differences to the observations over the entire calibration period). The values of SSh are shown in Table 1. In the hourly time scale, values of SSh range from 0.2 to 0.4. SSh is higher for specific humidity than for air temperature and higher in the case of mixed-field predictors than for single-field predictors.

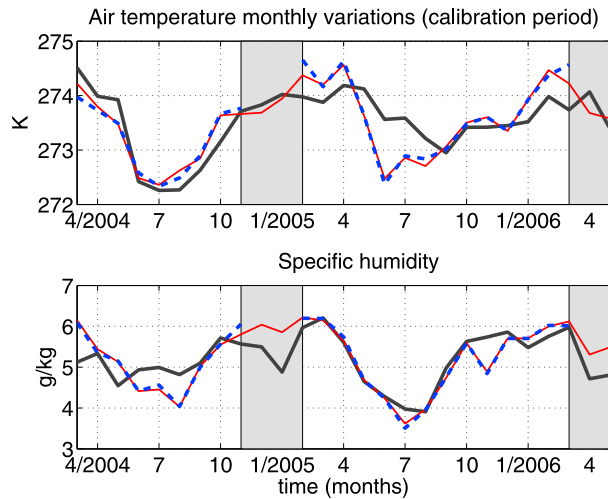
[40] To evaluate the hindcast performance of the ESD model for different temporal scales, daily and monthly

means are built from the 6-hourly observations and ESD model in the calibration period. The respective reference models in the daily time scale is reduced to seasonal cycles represented by monthly means of the observations. Figure 9 shows observations, reference model, and ESD model daily means as a function of time over the entire calibration period. The added value of the ESD model over the reference model is most evident in the daily time scale, since, per definition, in the reference model, no day-to-day variations occur within 1 month. Values of SSh are higher than in the hourly time scale, ranging from about 0.45–0.55 (see Table 1).

[41] In the monthly time scale, the ESD model fits the observations almost perfectly (Figure 10). The resulting values of SSh show high variations for the four predictand-predictor combinations, ranging from 0.32 (for air temperature/single-field predictor) up to 0.93 (for specific humidity/single-field predictor; please note that the values SSh for the single-field predictors are not shown in Table 1). SSh is a relative measure of skill and must be considered with care. In the monthly time scale in particular, both the ESD model and the reference model fit the observations very closely, because the calibration period is so short (12 transfer functions and monthly means based on about 24 months of data). Consequently, the values of mse are very low (<0.1) and minor changes in mse result in high variations of SSh.



**Figure 9.** (top) Daily mean air temperature and (bottom) specific humidity at AWS: observations (blue dashed), ESD model hindcast (red), and reference model (black solid) over the calibration period (March 2004 to May 2006). Periods with missing data are gray shaded.



**Figure 10.** (top) Monthly means of air temperature and (bottom) specific humidity: observations (blue dashed), ESD model hindcast (red), and reanalysis grid point (gray solid) in the calibration period (March 2004 to May 2006). Periods with missing data are gray shaded.

[42] Figure 10 also shows monthly means of NCEP/NCAR reanalysis data from the nearby grid point at 500 hPa, 77.5°W and 10°S (note that this grid point is not the closest of the eight grid points surrounding the study site in three dimensions but the one that shows the highest correlation to the observations at AWS [Hofer, 2007]). The curve representing the NCEP grid point values is shifted by its mean bias to the observations (given the 500 hPa level is located far from the NCEP model surface, temperature, and humidity values are systematically too low). NCEP grid point values are not shown in Figures 8 and 9 (6-hourly and daily values) in order to keep Figures 8 and 9 simple; however, there is much less agreement to the observations than in the monthly time scale. To evaluate the added value of the ESD model hindcast (using mixed-field predictors) against the simple use of grid point values, we define  $SSh$  similarly as in equation (5), but the reference model is represented by the NCEP grid point values (hereafter referred to as RM2). Hence,  $mse_{h,ref}$  is the mse between RM2 and the observations. Since RM2 is defined with zero bias,  $mse_{h,ref}$  is an estimate for the error variance.  $SSh$  amounts to 0.75 (6-hourly), 0.63 (daily), and 0.93 (monthly) for the predictand air temperature and to 0.67 (6-hourly), 0.71 (daily), and 0.79 (monthly) for specific humidity. Even if the NCEP grid point values fit the seasonal cycle already closely (cf. Figure 10), values of  $SSh$  in the monthly time scale are still very high because  $mse_{h,ESD}$  is close to zero. Overall,  $SSh$  is higher than 0.6 for both predictands at all time scales.

#### 4.5. ESD Model Forecasts 1960–2008

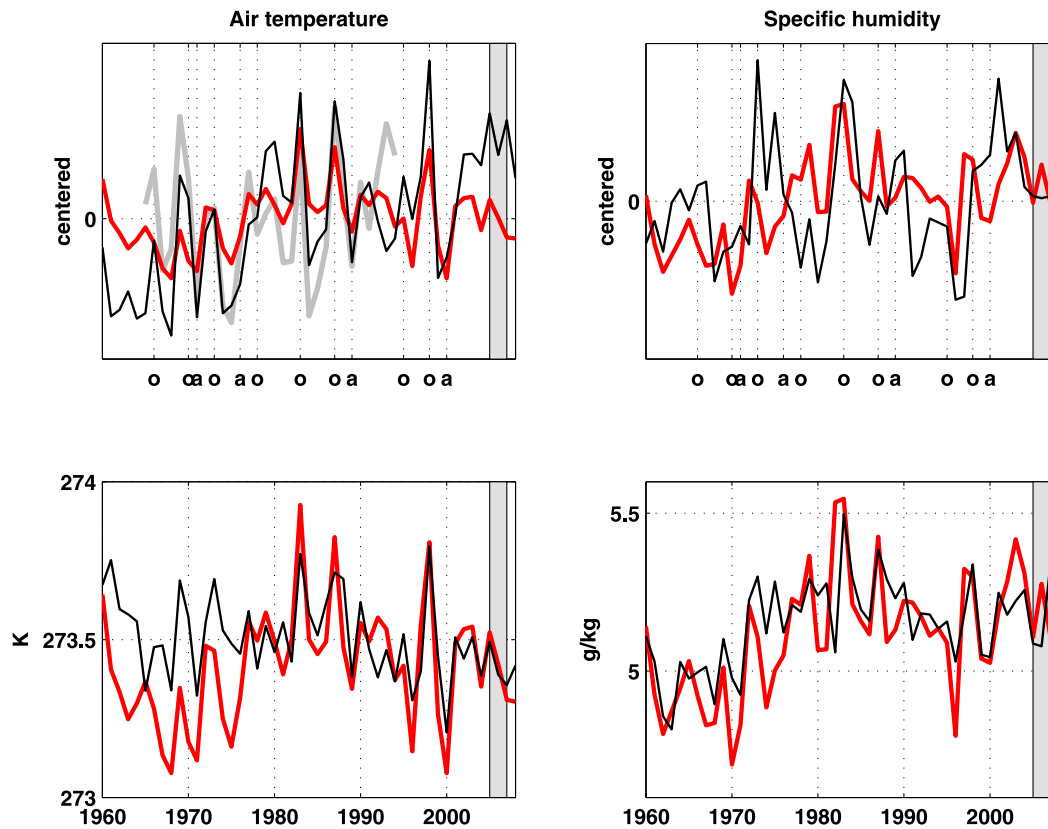
[43] The training period of the developed ESD model (calibration and validation) represents less than 3 years, about 5% of the reanalysis period. The question arises whether the ESD model is valid outside the training period. In particular, is the model able to reproduce variations that it has not experienced within the training period? If the

deterministic part in equation (1) is zero, the ESD model is equal to the RM1 and consists of constant values for each month and time of day. In this case, the interannual variations are zero. If the random part in equation (1) is zero, the ESD model fits the observations perfectly. The double cross-validation procedure yields skill scores  $SS$  ranging from 0 to 0.8 (Figure 7), indicating that the random part in equation (1) for some months/times of day is large and the resulting transfer functions are constant (or almost constant). Consequently, we can expect the ESD model to underestimate the variability at interannual time scales (similarly as it does in the time scale of model fit [see von Storch, 1999]).

[44] Interannual atmospheric variability in the Cordillera Blanca is mainly governed by the El Niño Southern Oscillation (ENSO); positive temperature anomalies are connected with El Niño and negative ones are connected with La Niña events. Concerning atmospheric moisture, El Niño years often tend to be dry and La Niña years tend to be wet, but the reverse is not uncommon as well [e.g., Garreaud et al., 2003; Vuille et al., 2008]. Figure 11 (top) shows annual means of the ESD model forecast (single-field predictors) 1960–2008 for both predictands (air temperature and specific humidity). As a reference, annual means of the NCEP grid point values (77.5°W/10°S in 500h Pa of both air temperature and specific humidity) and the observational time series from Querococha (only air temperature from 1965 to 1994) are displayed (centered values). Note that these time series, even if expected to be similar, must not be identical (i.e., differences do not necessarily come from errors in the ESD model).

[45] For air temperature (Figure 11, top left) ESD model, NCEP grid point, and observational time series generally show similar patterns that clearly reflect ENSO. While the NCEP grid point time series shows a rising trend over the period, no such trend is evident in the Querococha time series and the ESD model. A comparison of ESD model forecast to the annual means in the calibration period (gray-shaded area in Figure 11) shows larger variability in the forecast period than in the calibration period, supporting the stationarity of the ESD model. However, interannual variability is smaller in the ESD model than in the NCEP grid point and Querococha time series. As discussed above, the ESD model underestimates the real variability by construction. However, the permanent exchange of latent heat due to sublimation, melting, and refreezing smoothes air temperature variations above a glacier surface at all time scales and could add to the smaller variance in the ESD model. Specific humidity variances (Figure 11, right top), by contrast, show similar magnitudes in both the ESD model and the NCEP grid point time series, but the temporal patterns over the years differ. Also, the relation of specific humidity to El Niño/La Niña events is less evident in both time series, and sometimes reverses.

[46] Figure 11 (bottom) again shows the ESD model forecast 1960–2008 for air temperature and specific humidity, but this time with two different predictors (single and mixed-field). We want to assess if the chosen model is sensitive to the variable fields used as predictors. In fact, the two time series show different patterns. For air temperature, the models are very similar in recent years, but prior to the late 1970s the two solutions diverge. Still, they show more



**Figure 11.** (top) Annual means of ESD model forecast (solid red), NCEP grid point values (black), and observations at Querococha (solid gray, only air temperature) of (left) air temperature and (right) specific humidity for the period 1960–2008. El Niño (o) and La Niña (a) events are indicated in the abscissa. (bottom) ESD model forecast of (left) annual mean air temperature and (right) specific humidity for the period 1960–2008, with two different predictor fields applied for each variable: the single-field predictors  $a_{1000}$  for air temperature and  $s_{1000}$  for specific humidity (solid red) and the mixed-field predictors including  $a_{1000}$ ,  $s_{1000}$ , and  $u_{400}$  (black). The model training period (2004–2006) is gray shaded.

or less the same pattern of maxima and minima. For specific humidity, the ESD model based on mixed-field predictors shows less interannual variability than the ESD model based on single-field predictors, but the general pattern over time is similar (e.g., both time series show an increasing trend until the early 1980s and a slight decrease thereafter).

## 5. Conclusions

[47] We have presented a computationally cheap method to transfer coarse scale NCEP/NCAR reanalysis data to local air temperature and specific humidity in a glacier environment at high temporal resolution. The developed nonparametric technique is appropriate for short time series and applicable wherever high-resolution observations are available. The modeling procedure includes an automated cross-validation scheme, but appropriate choices need to be made concerning the exact definition of predictor domain and variable fields. So far, we have not assessed the performance of NCEP/NCAR reanalysis data against other reanalysis products [e.g., Kazutoshi *et al.*, 2007; Uppala *et al.*, 2005]. Also, EOF analysis is used in this study only for dimensionality reduction and not for pattern interpretation.

[48] The developed high-resolution ESD model shows added value over simpler techniques, such as the use of reanalysis grid point values or constant annual/diurnal cycles calculated from the calibration data set. We are therefore confident that this study represents an important progress for physically based mass balance studies on data-sparse tropical glaciers. As a next step, we schedule case study simulations using a regional atmospheric model to increase the knowledge about significant dynamic processes above a tropical glacier surface. By combining statistical and dynamical techniques, we target for the complete set of atmospheric key variables (including spatial gradients) required for long-term, process-based glacier mass balance modeling in the Cordillera Blanca.

[49] **Acknowledgments.** We thank P. Wagnon (IRD) and I. Juen (ITGG) for kindly providing the data from the glacier Artesonraju (AWS). NCEP Reanalysis data were made available by the NOAA/OAR/ESRL PSD, Boulder, CO, USA, from their Web site at <http://www.cdc.noaa.gov/>. This work is funded by the “Verein zur Förderung der wissenschaftlichen Tätigkeit von Südtirolern an der Landesuniversität Innsbruck” and the Austrian Science Foundation FWF (P20089-N10, P22106-N21). We are especially grateful to the three reviewers. Their comments helped to improve our work substantially.

## References

- Akaike, H. (1973), Information theory and an extension of the maximum likelihood principle, in *Proceedings of the 2nd International Symposium on Information Theory*, edited by B. N. Petrov and F. Csaki, pp. 267–281, Akademiai Kiado, Budapest.
- Ames, A. (1998), A documentation of glacier tongue variations and lake development in the Cordillera Blanca, Peru, *Zeitung für Gletscherkunde und Glazialgeologie*, 34(1), 1–36.
- Benestad, R. E. (2001), A comparison between two empirical downscaling strategies, *Int. J. Climatol.*, 21(13), 1645–1668.
- Benestad, R. E., I. Hanssen-Bauer, and C. Deliang (2008), *Empirical-statistical downscaling*, World Sci., Singapore.
- Brinkmann, W. A. R. (2002), Local versus remote grid points in climate downscaling, *Clim. Res.*, 21(1), 27–42.
- Cavazos, T., and B. C. Hewitson (2005), Performance of NCEP-NCAR reanalysis variables in statistical downscaling of daily precipitation, *Clim. Res.*, 28, 95–107.
- Christensen, J. H., et al. (2007), Regional Climate Projections, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate*, edited by S. Solomon et al., Cambridge Univ. Press, Cambridge, UK.
- Favier, V., P. Wagon, J.-P. Chazarin, L. Maisincho, and A. Coudriaun (2004), One-year measurements of surface heat budget on the ablation zone of Antizana Glacier 15, Equadorian Andes, *J. Geophys. Res.*, 109 (D18), D18105, 1–15, doi:10.1029/2003JD004359.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi (2007), Review Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling, *Int. J. Climatol.*, 27, 1547–1578.
- Früh, B., J. W. Schipper, A. Pfeiffer, and V. Wirth (2006), A pragmatic approach for downscaling precipitation in alpine-scale complex terrain, *Meteorol. Z.*, 15(16), 631–646.
- Garreaud, R., M. Vuille, and A. C. Clement (2003), The climate of the Altiplano: Observed current conditions and mechanisms of past changes, *Palaeogeogr., Palaeoclimatol., Palaeoecol.*, 194, 5–22.
- Georges, C. (2004), 20th Century Glacier fluctuations in the Tropical Cordillera Blanca, Peru, *Arct. Antarct. Alp. Res.*, 36(1), 100–107.
- Georges, C. (2005), Recent Glacier Fluctuations in the Tropical Cordillera Blanca and Aspects of the Climate Forcing, Ph.D. Thesis, Leopold Franzens Univ., Innsbruck, Austria.
- Grotch, S. L., and M. C. MacCracken (1991), The use of global climate models to predict regional climatic change, *J. Clim.*, 4, 286–303.
- Hannachi, A., I. T. Jolliffe, and D. B. Stephenson (2008), Empirical orthogonal functions and related techniques in atmospheric science: A review, *Int. J. Climatol.*, 27, 1119–1152.
- Hastenrath, S. (1978), Heat budget measurements on Quelccaya Ice Cap, Peruvian Andes, *J. Glaciol.*, 20(82), 85–97.
- Hertig, E., and J. Jacobeit (2008), Assessments of Mediterranean precipitation changes for the 21st century using statistical downscaling techniques, *Int. J. Climatol.*, 28, 1025–1045.
- Hewitson, B. C., and R. G. Crane (2006), Consensus between GCM climate change projections with empirical downscaling: precipitation downscaling over South Africa, *Int. J. Climatol.*, 26, 1315–1337.
- Hofer, M. (2007), Statistical downscaling of NCEP/NCAR reanalysis data to air temperature and specific humidity above an outer tropical glacier surface; Artesonraju (Peru), MS thesis, 93 pp, Leopold Franzens Univ., Innsbruck, Austria.
- Juen, I. (2006), Glacier mass balance and runoff in the Cordillera Blanca, Peru, Ph.D. thesis, Leopold Franzens Univ., Innsbruck, Austria.
- Juen, I., C. Georges, and G. Kaser (2007), Modelling observed and future runoff from a glacierized tropical catchment (Cordillera Blanca, Peru), *Global Planet. Change*, 59, 37–48.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-Year Reanalysis Project, *Bull. Am. Meteorol. Soc.*, 77(3), 437–471.
- Karl, T. R., S. J. Hassol, S. J. Miller, and W. L. Murray (2006), Temperature Trends in the Lower Atmosphere: Steps for Understanding and Reconciling Differences, *Rep.*, 180 pp, Washington, D.C.
- Kaser, G. (1995), Some notes on the behaviour of tropical glaciers., *Bull. Inst. Fr. Etud. Andines*, 24(3), 671–681.
- Kaser, G., and H. Osmaston (2002), *Tropical Glaciers*, Cambridge Univ. Press, Cambridge, UK.
- Kaser, G., S. Hastenrath, and A. Ames (1996), Mass Balance Profiles on Tropical Glaciers, *Zeitung für Gletscherkunde und Glazialgeologie*, 32, 75–81.
- Kaser, G., I. Juen, C. Georges, J. Gómez, and W. Tamayo (2003), 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(1–4), 130–144.
- Kaser, G., C. Georges, I. Juen, and T. Mölg (2005), Low-latitude glaciers: Unique global climate indicators and essential contributors to regional fresh water supply. A conceptual approach., in *Global Change and Mountain Regions: A State of Knowledge Overview*, edited by U. Huber, H. K. M. Bugmann and M. A. Reasoner, pp. 185–196, Kluwer, New York.
- Kazutoshi, O., et al. (2007), The JRA-25 reanalysis, *J. Meteorol. Soc. Jpn.*, 85(3), 369–432.
- Kim, J.-W., J.-T. Chang, N. L. Baker, D. S. Wilks, and W. L. Gates (1984), The statistical problem of climate inversion: Determination of the relationship between local and large-scale climate, *Mon. Weather Rev.*, 112(10), 2069–2077.
- Klein, W. H., and H. R. Glahn (1974), Forecasting local weather by means of model output statistics, *Bull. Am. Meteorol. Soc.*, 55(10), 1217–1227.
- Kuhn, M. (1989), The response of the equilibrium line altitude to climate, in *Glacier Fluctuations and Climatic Change*, edited by J. Oerlemans, pp. 407–417, Kluwer Acad., Dordrecht, Netherlands.
- Kutzbach, J. E. (1967), Empirical of sea-level pressure, surface temperature and precipitation complexes over North America, *J. Appl. Meteorol.*, 6, 791–802.
- Mark, B. G., and G. O. Seltzer (2003), Tropical glacier meltwater contribution to stream discharge: a case study in the Cordillera Blanca, Peru, *J. Glaciol.*, 49(165), 271–281.
- Matulla, C. (2005), Regional, seasonal and predictor-optimized downscaling to provide groups of local scale scenarios in the complex structured terrain of Austria, *Meteorol. Z.*, 14(1), 31–45(15).
- Michaelsen, J. (1987), Cross-validation in statistical climate forecast models, *J. Clim. Appl. Meteorol.*, 26, 1589–1600.
- Mölg, T., and D. R. Hardy (2004), Ablation and associated energy balance of a horizontal glacier surface on Kilimanjaro, *J. Geophys. Res.*, 109 (D16), D16104, 1–13, doi:10.1029/2003JD004338.
- Mölg, T., N. J. Cullen, and G. Kaser (2009), Solar radiation, cloudiness and longwave radiation over low-latitude glaciers: Implications for mass balance modeling, *J. Glaciol.*, 55, 292–302.
- Niedertscheider, J. (1990), Untersuchungen zur Hydrographie der Cordillera Blanca (Peru), M.S. thesis, Dept. of Geographie, Univ. of Innsbruck.
- Paterson, W. S. P. (1994), *The Physics of Glaciers*, 3rd ed., Butterworth-Heinemann, Oxford.
- Prömmel, K., B. Geyer, J. M. Jones, and M. Widmann (2009), Evaluation of the skill and added value of a reanalysis-driven regional simulation for Alpine temperature, *Int. J. Climatol.*, 30(5), 760–773.
- Radic, V., and R. Hock (2006), Modeling future glacier mass balance and volume changes using ERA-40 reanalysis and climate models: A sensitivity study at Storglaciären, Sweden, *J. Geophys. Res.*, 111(F3), F03003, 1–12, doi:10.1029/2005JF000440.
- Reichert, B. K., and L. Bengtsson (2002), Recent Glacier Retreat Exceeds Internal Variability, *J. Clim.*, 15(23), 3069–3081.
- Salzmann, N. (2006), The use of results from regional climate models for local-scale permafrost modelling in complex high-mountain topography – possibilities, limitations and challenges for the future. Ph.D. thesis, Geographisches Institut der Universität Zürich, Zürich, CH.
- Schoof, J. T., and S. C. Pryor (2001), Downscaling temperature and precipitation: A comparison of regression-based methods and artificial neural networks, *Int. J. Climatol.*, 21, 773–790.
- Silverio, W., and J.-M. Jaquet (2005), Glacial cover mapping (1987–1996) of the Cordillera Blanca (Peru) using satellite imagery, *Remote Sens. Environ.*, 95, 342–350.
- Trenberth, K. E., et al. (2007), Observations: Surface and Atmospheric Climate Change, *Rep.*, Cambridge, UK.
- Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, 131, 2961–3012.
- von Storch, H. (1999), On the use of “inflation” in statistical downscaling, *J. Clim.*, 12, 3505–3506.
- von Storch, H., and F. Zwiers (2001), *Statistical Analysis in Climate Research*, 484 pp., Cambridge Univ. Press, Cambridge, UK.
- von Storch, H., B. Hewitson, and L. Mearns (2000), Review of empirical downscaling techniques, in *Regional Climate Development Under Global Warming*, edited by T. Iversen and B. A. K. Hoiskar.
- Vuille, M., and F. Keimig (2004), Interannual variability of summertime convective cloudiness and precipitation in the Central Andes derived from ISCCP-B3 data, *J. Clim.*, 17, 3334–3348.
- Vuille, M., G. Kaser, and I. Juen (2008), Glacier mass balance variability in the Cordillera Blanca, Peru and its relationship to climate and large scale circulation, *Global Planet. Change*, 62, 14–28.
- Wagon, P., P. Ribstein, B. Francou, and B. Pouyaud (1999), Annual cycle of energy balance of Zongo Glacier, Cordillera Real, Bolivia, *J. Geophys. Res.*, 104(D4), 3907–3923.
- Widmann, M. (2005), One-dimensional CCA and SVD, and their relationship to regression maps, *J. Clim.*, 18(14), 2785–2792.

- Widmann, M., C. S. Bretherton, and E.P. Salathé (2003), Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor, *J. Clim.*, *16*(5), 799–816.
- Wilks, D. S. (1997), Resampling hypothesis tests for autocorrelated fields, *J. Clim.*, *10*(1), 65–82.
- Wilks, D. S. (2006), *Statistical Methods in the Atmospheric Sciences*, 2 ed., Academic.
- Xoplaki, E., J. F. Gonzalez-Rouco, J. Luterbacher, and H. Wanner (2003), Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs, *Clim. Dyn.*, *20*(7–8), 723–739.
- 
- M. Hofer, G. Kaser, B. Marzeion, and T. Mölg, Center for Climate and Cryosphere, Geo- and Atmospheric Sciences, Geography, Tropical Glaciology Group, University of Innsbruck, Innrain 52, 6020 Innsbruck, Austria 0512 50700. (Marlis.Hofer@uibk.ac.at)