Statistical downscaling of NCEP/NCAR reanalysis data to air temperature and specific humidity above an outer tropical glacier surface

Artesonraju (Peru)

A diploma thesis submitted to the
Department of Meteorology and Geophysics,
University of Innsbruck

for the degree of
Master of Natural Science

presented by
Marlis Hofer

November 2007
Für Karmen und Sarah
Acknowledgments

I would like to express my deep appreciation to my supervisor Dr. Georg Kaser, for the exceptional support and encouragement during the development of my master thesis.

I thank the members of the Innsbruck Tropical Glaciology Group Dr. Irmgard Juen, Dr. Javier Corripio and Dr. Thomas Mölg, who generously shared their insights, and for a wonderful and instructive experience in Peru.

Enormous gratitude to Dr. Michael Kuhn, Dr. Martin Ehrendorfer and all members of the Institute of Meteorology and Geophysics of Innsbruck for the excellent formation and the great support during my study.

I am grateful to Dr. Patrick Wagnon, Dr. Delphine Six and the members of the Laboratoire du Glaciologie et Géophysique de l’Environnement for their cordial guidance and support in the initial period of development of my master thesis in Grenoble.

I thank my colleagues Michael Winkler and Rainer Prinz, for numerous encouraging discussions.

I want to thank Roberta Pergher and Günther Burgstaller for proofreading my master thesis.

Thanks to all my study colleagues for the fun times at and around the university.

Thanks to my friends Mirjam Waldboth, Florian Robl, Annemarie Kompatscher and Stefan Vorhauser for encouraging me so often.

Thanks to my sisters Karmen and Sarah, who have always believed in me.

Thanks to Mattia for his patience and love.

Last but not least I want to thank my parents Notburga and Lorenz Hofer. Without their generous financial support, their trust and their openness this formation would never have been possible.
Abstract

Outer tropical climates feature a yearly cycle of air humidity and related meteorological variables. Glaciers reflect this seasonality in a yearly variability of their mass balances (GMB). In the outer tropical Cordillera Blanca (CB), however, reanalysis data (RA) often are the only available data input for GMB models. Yet, the horizontal resolution of the reanalysis model can not resolve the complex surface topography in the mountains, and put the applicability of RA to the CB into question. One way to extract local scale information from coarse scale general circulation models (GCM) is the statistical downscaling (SD) method. The objective of this master thesis is to find statistical relationships between the RA and meteorological variables on the Peruvian glacier Artesonraju. This SD is calibrated by two years of measurements at an automatic energy balance station (EBS) located on the ablation zone of Artesonraju glacier. The assessed predictand variables are air temperature and specific humidity at 1 m above the surface in hourly, daily, and monthly time scales. From the RA the same variables at three vertical levels and four grid points surrounding the study site are examined as potential predictors. The applied methods are mainly single and multiple linear regression analysis. The results show that RA from the SE grid point in the 500 hPa level correlate best to the observations. Yet, only monthly means exceed a correlation (r-square) of 0.5. However, standardizing the daily cycle of RA to the daily cycle measured at the EBS, and applying an optimal time shift between the data series in the hourly time scale improves the correlation for the predictand temperature. Specific humidity, though, is not a normally distributed variable in the hourly and daily time scale. Thus the applied methods are less successful for specific humidity than for temperature. Applying multiple linear regression to temperature and specific humidity as predictors increases the correlation for both predictands in all time scales. Yet, the residuals of all applied regression models feature a seasonal pattern. Hence it is necessary to standardize the RA to the seasonal means and variances of the observational data. Since the measurements have been initiated only in March 2004, the data series is not considered long enough for seasonal climatology.
Contents

Acknowledgments

Abstract

1. Background and objective

2. Structure of the master thesis

3. Abbreviations

I. General introduction to the topic

4. Outer tropical climate and glaciers
   4.1. General circulation at low latitudes
   4.2. Climate in the outer tropics
      4.2.1. Seasonal variations
      4.2.2. Inter-annual variability and El Niño
   4.3. The climate of the South American Altiplano
      4.3.1. Current rainfall regime
      4.3.2. Rainfall variability and the influence of atmospheric circulation patterns
      4.3.2.1. Inter-annual variability and El Niño
   4.4. Climatic delimitations of tropical glaciers
   4.5. Glacier-climate interaction in the outer tropics

5. The study site: Geographical and glaciological aspects

6. Methods of glacier mass balance modelling
6.1.1. The continuity of mass and energy on the glacier-atmosphere interface ........................................... 14
6.1.2. The specific mass balance ................................................................. 14
6.2. The surface energy budget ................................................................. 15
6.3. Melt modelling ..................................................................................... 16
6.3.1. Temperature-index melt models ...................................................... 16
6.3.1.1. The degree-day factor ................................................................. 17
6.3.2. Surface energy balance melt models ................................................ 17
6.4. Example for GMB modelling: ITGG-2.0 ............................................. 18
6.4.1. Model structure ................................................................................ 18
6.4.1.1. Calculation of specific mass balance $b$ ...................................... 18
6.4.2. Input variables of the ITGG-2.0 and the objective of this master thesis ......................................................................................... 19

7. The NCEP/NCAR reanalysis project ....................................................... 21
7.1. Introduction ......................................................................................... 21
7.2. Observational data input of the NCEP/NCAR reanalysis model ........ 21
7.3. NCEP/NCAR reanalysis model specifications and output variables .. 22

8. Statistical downscaling ......................................................................... 23
8.1. Introduction ......................................................................................... 23
8.2. Regression-based SD methods ............................................................ 24
8.3. General scheme of SD: a step-by-step process ..................................... 24
8.4. Advantages and limitations of SD ....................................................... 25

II. Downscaling ......................................................................................... 27

9. Data ........................................................................................................ 28
9.1. Observational data from automatic weather stations ....................... 28
9.1.1. Measuring period and data gaps ...................................................... 29
9.1.2. The predictands ............................................................................... 30
9.1.3. Calculations ..................................................................................... 30
9.1.3.1. Calculation of specific humidity ............................................... 30
9.1.3.2. Average building and time lag .................................................... 31
9.2. The NCEP/NCAR reanalysis data ..................................................... 31
9.2.1. The predictors ............................................................................... 31
9.2.2. Model topography and surface heights of the examined grid points

10. Statistical methods and applied software
10.1. Least squares regression
10.2. Model evaluation
   10.2.1. Graphical methods for model evaluation
      10.2.1.1. Correlation plot
      10.2.1.2. Residual analysis
   10.2.2. Numerical measures for model evaluation

11. Grid point selection
11.1. Discussion
   11.1.1. Time scales
   11.1.2. Vertical level
   11.1.3. Grid point
   11.1.4. Summary

12. Hourly values: 00, 06, 12 and 18 UTC
12.1. Are four values per day representative for the daily cycle?
12.2. Data statistics
   12.2.1. Discussion
   12.2.2. Standardization of X
      12.2.2.1. Discussion
   12.2.3. Normality check of hourly means
      12.2.3.1. Temperature
      12.2.3.2. Specific humidity
      12.2.3.3. Discussion
   12.3. The diurnal evolution of temperature and specific humidity of X and Y (case study)
      12.3.1. Introduction
      12.3.2. General evolution of temperature and specific humidity above the glacier surface (Y)
      12.3.2.1. Discussion (based on Figs. 12.2, 12.3, and 12.4)
      12.3.3. Comparison between RA and EBS
      12.3.3.1. Discussion (based on Figs. 12.2, 12.3, and 12.4)
   12.4. Mean daily cycle
12.4.1. Discussion of Fig. 12.5 ........................................... 49
  12.4.1.1. Comparison between temperature and specific hu-
     midity mean daily cycles at the EBS ......................... 49
  12.4.1.2. Comparison between X and Y daily cycles .......... 51
12.5. Time shift adjustment between X and Y: SD model c1 ........ 51
  12.5.1. Introduction ................................................. 51
  12.5.2. Results ................................................. 52
  12.5.3. Discussion ................................................. 52
  12.5.4. SD model c1 ........................................... 53
  12.5.5. Results and discussion (Fig. 12.5) ................. 53
12.6. Examination of r-square for single times of day ............. 53
  12.6.1. Introduction ................................................. 53
  12.6.2. Results ................................................. 54
  12.6.3. Discussion ................................................. 54
12.7. Standardization of the predictor’s daily cycle: SD model c2 .... 55
12.8. SD models c1 and c2: Summary and conclusions .............. 55
12.9. Evaluation of the results ....................................... 56
  12.9.1. Correlation plot .......................................... 56
    12.9.1.1. Temperature ........................................ 56
    12.9.1.2. Specific humidity ................................ 57
  12.9.2. Residual analysis ...................................... 58
    12.9.2.1. Temperature ........................................ 58
    12.9.2.2. Specific humidity ................................ 59
  12.9.3. Case study period ....................................... 60
12.10. Multiple linear and nonlinear regression ..................... 61
  12.10.1. Introduction .......................................... 61
  12.10.2. Specifications ........................................ 61
  12.10.3. Results ................................................. 61
  12.10.4. Discussion of Fig. 12.14 ............................... 62
    12.10.4.1. Temperature ........................................ 62
    12.10.4.2. Specific humidity ................................ 62

13. Daily Means .................................................. 65
  13.1. Data statistics .............................................. 65
    13.1.1. Discussion of Tab. 13.1 ............................. 65
      13.1.1.1. Comparison to the statistical properties of hourly
                  values (Tab. 12.1) ................................ 65
13.1.1.2. Comparison between Y and X ........................................ 66
13.1.2. Standardization of the predictor ........................................ 66
13.1.3. Normality check of daily means ........................................ 66
  13.1.3.1. Discussion of Fig. 13.1 ............................................. 68
13.2. Correlation analysis for daily means .................................... 68
  13.2.1. Correlation plot and linear fit ....................................... 68
  13.2.2. Residual analysis .................................................... 69
13.3. The course of temperature and specific humidity daily means throughout BP .................................................. 70
  13.3.1. Temperature ........................................................ 70
  13.3.2. Specific humidity .................................................. 70
  13.3.3. Discussion of Figs. 13.3.2 and 13.3.1 .......................... 72

14. Monthly Means .......................................................... 73
14.1. Data statistics ................................................................ 73
  14.1.1. Standardization of the predictor ................................... 74
  14.1.2. Normality check of monthly means .............................. 74
  14.1.3. Discussion of Fig. 14.1.2 .......................................... 75
14.2. Correlation analysis for monthly means .............................. 75
  14.2.1. Correlation plot ........................................................ 75
    14.2.1.1. Discussion of Fig. 14.2.1 ...................................... 75
  14.2.2. Residual analysis .................................................... 76
14.3. The course of temperature and specific humidity monthly means throughout BP .................................................. 77
  14.3.1. Description of the seasonal course of temperature and specific humidity over BP (Fig. 14.3) ...................... 77
  14.3.2. Discussion of Fig. 14.3 (left and right) ........................ 78
14.4. SD model c1 for individual seasons .................................... 78

15. Multiple regression analysis in the daily and monthly timescale 80
15.1. Introduction .................................................................. 80
15.2. Multiple linear regression: functional relationship ............... 80
15.3. Results ....................................................................... 81
  15.3.1. Discussion of the results ............................................. 81
    15.3.1.1. Multiple regression in the daily timescale ............... 81
    15.3.1.2. Multiple regression in the monthly timescale ........... 81
  15.3.2. Summary .............................................................. 83
15.4. Residual analysis of the multiple linear regression model for daily and monthly means ........................................... 83
   15.4.1. Temperature ..................................................... 83
   15.4.2. Specific humidity ............................................. 83

16. Summary, results and outlook 86
   16.1. Summary ............................................................ 86
   16.2. Results ............................................................... 87
   16.3. Outlook .............................................................. 88

Bibliography 93
1. Background and objective

Glacier mass balance (GMB) modelling has been investigated in the tropics on several sites. Since only few direct measurements of meteorological variables are available, it has been suggested to use reanalysis data as GMB model input. In fact, reanalysis data represent an opulent data pool for reconstructions of climate and meteorology of the last decades. Yet, the generally coarse horizontal grid resolution of the reanalysis model put their applicability to specific sites into question. One method to reconstruct (and predict) specific site weather and climates on the base of reanalysis data (or GCM output) is the statistical downscaling method.

The objective of this master thesis is to find and apply statistical regression models in order to reconstruct local meteorology on the ablation zone of the outer tropical glacier Artesonraju from the reanalysis data. Specifically, the diurnal, daily and monthly evolution of the variables air temperature and specific humidity is examined, because input data for GMB models are required at different temporal resolutions. The regression is calibrated by measurements at an automatic energy balance station (EBS) installed by the French IRD (Institut de Recherche pour le Développement) in 2004 on the ablation zone of Artesonraju glacier.
2. Structure of the master thesis

This master thesis is composed by two main parts:

**General introduction** Since one essential component of this master thesis has been literature review about the main topics that are linked together in this statistical downscaling approach, an overview thereof is given in the first part, containing

- the principals of climatic features in different time scales of the tropics in general and the outer tropics, and the response of outer tropical glaciers to this features, based on several studies,
- a general description of the study site (Cordillera Blanca and Artesonraju glacier),
- an introduction to mass balance modelling (with special attention to the ITGG-2.0, which has been used for mass balance and runoff reconstructions and projections in the Cordillera Blanca by Juen, 2006),
- an description of the NCEP/NCAR reanalysis data,
- a general overview about statistical downscaling, mainly based on Wilby et al. (2004).

**Downscaling** The downscaling approach in the second part of this master thesis is basically regression analysis of the observational data from Artesonraju glacier and the reanalysis data. The main parts are:

- the introduction describing the applied methods of regression analysis and the examined data from the energy balance station EBS and the reanalysis data RA,
- the main analysis part which is subdivided in three main sections, according to the three examined time scales: hourly (six-hourly) values, daily and monthly means,
- the final part with summary, results and outlook.
### 3. Abbreviations

In this section, abbreviations of terms (and *names*) applied in this master thesis are listed alphabetically. Further specifications follow in adjacent sections.

<table>
<thead>
<tr>
<th>[abbreviation]</th>
<th>[term]</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS</td>
<td>automatic weather station</td>
</tr>
<tr>
<td>BP</td>
<td>baseline period</td>
</tr>
<tr>
<td>EBS</td>
<td>energy balance station</td>
</tr>
<tr>
<td>ENSO</td>
<td>El Niño Southern Oscillation</td>
</tr>
<tr>
<td>GCM</td>
<td>global circulation model</td>
</tr>
<tr>
<td>GMB</td>
<td>glacier mass balance</td>
</tr>
<tr>
<td>RA</td>
<td>NCEP/NCAR reanalysis data</td>
</tr>
<tr>
<td>RCM</td>
<td>regional climate model</td>
</tr>
<tr>
<td>r-square</td>
<td>coefficient of determination</td>
</tr>
<tr>
<td>SD</td>
<td>statistical downscaling</td>
</tr>
<tr>
<td>SEB</td>
<td>surface energy balance</td>
</tr>
<tr>
<td>shum</td>
<td>specific humidity</td>
</tr>
<tr>
<td>SST</td>
<td>sea surface temperature</td>
</tr>
<tr>
<td>temp</td>
<td>temperature</td>
</tr>
<tr>
<td>X</td>
<td>predictand</td>
</tr>
<tr>
<td>Y</td>
<td>predictor</td>
</tr>
<tr>
<td>ˆY</td>
<td>regression model</td>
</tr>
</tbody>
</table>

[**name/city**]

- **ITGG**  
  *Innsbruck Tropical Glaciology Group/Innsbruck*
- **LGGE**  
  *Laboratoire du Glaciologie et Geophysique de l’ Environnement/Grenoble*
- **NCAR**  
  *National Center of Atmospheric Research/Boulder*
- **NCEP**  
  *National Center for Environmental Prediction/Camp Springs*
Part I.

General introduction to the topic
4. Outer tropical climate and glaciers

4.1. General circulation at low latitudes

The two principal components in determining the global (or planetary) circulation pattern of the atmosphere are the meridional gradient of radiation (and energy) balance on the earth’s surface, caused by the meridional altering solar incident angle at different latitudes, and the Coriolis force, caused by the earth’s rotation. However, the global circulation pattern is modified throughout the seasonal cycle not only in terms of intensity (of highs and depressions), but also in a displacement of the whole pattern towards the pole in summer and the equator in winter. Moreover, earth surface conditions such as land-sea distribution and global snow (ice) cover further modify the general circulation pattern.

In low latitudes, three principal features may be mentioned to characterize the general circulation regime; namely, the Innertropical Convergence Zone (ITCZ) constituted by the ascending branches of the Hadley cells on their equatorward sides, secondly the subtropical highs approximately 30° from the equator (often used as natural boundary in the definition of the geographical extend of the tropics), and thirdly, the trade winds, which form the return flow to the equator at the earth’s surface.

[McGregor and Nieuwolt, 1998; Schönwiese, 2003]

4.2. Climate in the outer tropics

4.2.1. Seasonal variations

Since in the tropics, generally, temperature variations of daily means throughout the year are small, seasons can not be defined after thermal criteria. Still, the seasonality of the insolation induces a yearly oscillation of atmospheric circulation patterns, as, e.g., the propagation of the ITCZ throughout the year beneath the limits outlined in Fig. 4.1. This leads to a yearly oscillation of governing climate regimes in the outer part of the tropics, in contrast to the inner tropics where wet conditions dominate
4.2 Climate in the outer tropics

Figure 4.1.: The tropics and their delimitations from a glaciological point of view, and the distribution of the glacier areas according to countries’ ITCZ (after Kaser, 1996b)

all year round. Thus, the outer tropics can be defined as an an intermediate zone between the tropics and the subtropics where tropical conditions dominate during the humid season and subtropical conditions during the dry season (e.g., Kaser and Osmaston, 2002).

[McGregor and Nieuwolt, 1998; Kaser and Osmaston, 2002; Schönwiese, 2003]

4.2.2. Inter-annual variability and El Niño

In addition to the seasonal oscillation of climatic features in the tropics the Southern oscillation and El Niño (and la Niña, respectively), often referred to as ENSO event, have to be mentioned. These aperiodically occurring events are closely related to the Pacific sea surface temperature (SST) zonal gradients and anomalies and represent the two possible extreme modes of the Walker circulation. A further general description of the interaction between Walker circulation, Hadley circulation and El Niño is given, e.g., in Schönwiese (2003).

Since the 1970s ENSO events have been frequent, with an average time between events of around four years. Often referred to as warm events, ENSO events are characterized by a number of distinct changes in the atmospheric and oceanic circulation in the Pacific Basin. During ENSO (La Nina), the normally cool SST of the eastern Pacific are replaced by warm (cooler) ones as a result of a complex interplay between oceanic upwelling, the trade wind system, the oceanic equatorial
countercurrent and the pressure distribution across the Pacific Basin.

[McGregor and Nieuwolt, 1998; Schönwiese, 2003]

4.3. The climate of the South American Altiplano

This section summarizes the findings in Garreaud et al. (2003), where the large-scale controls on local climate in the South American Altiplano are examined using local observations, reanalysis data and model experiments. Garreaud et al. (2003) concentrate their studies on the climate of the 'Altiplano', which is the central and broadest portion of the South American Andes between 15° - 20°S. The Cordillera Blanca is located in the northern-most and narrower part of the Andes at 9°S. The study is considered particularly useful for the SD in this master thesis as it follows to assess the physical linkages between large scale climatic features and local meteorology on a site comparable to Artesonraju glacier.

4.3.1. Current rainfall regime

Rainfall over the South American Altiplano is largely produced by deep, moist convection. This is induced by three main factors:

- the heating of the surface which destabilizes the ABL;
- sufficiently high values of near surface humidity for the convection to become moist (typical threshold values are \( \sim 5 \) g/kg; Garreaud et al., 2003);
- and a dynamical lifting process.

On mountainous terrain at low latitudes such as the Andes both dynamical lifting processes and high values of midday solar radiation, two of the three ingredients for deep convection, are provided most of the time. Nevertheless, the Altiplano rainfall exhibits pronounced variations in all time scales (diurnal, seasonal and inter-annual) and is largely restricted to the austral summer season. Thus, rainfall fluctuations are thought to be triggered by varying moisture availability near the surface. In fact, near surface moisture in the Altiplano has shown to origin from the ABL of the eastern continental lowlands by several studies (e.g., Hardy et al., 1998).
4.3.2. Rainfall variability and the influence of atmospheric circulation patterns

The most obvious fluctuations in Altiplano rainfall are those associated with the diurnal and seasonal cycles. In a daily timescale, the maximum frequency and intensity occurs between afternoon and early night (after the diurnal maximum of insolation and the resulting surface heating). In the annual timescale rainfall is largely concentrated in the summer months.

However, precipitation anomalies in different timescales have been observed to coincide with anomalies of the upper-level zonal flow. Garreaud et al. (2003) explain the physical link between the dynamics and moisture variability on the Altiplano based on numerical model simulations as follows:

Moisture in the Altiplano is transported by regional circulation over the eastern slope of the Andes, drawing moist air from the ABL over the nearby continental lowlands. This confirms the continental origin of the near surface humidity over the Altiplano. So the moisture transport is primarily induced by surface heating and the resulting up-slope winds. This regional circulation can then be modulated in strength and duration by the large-scale flow aloft (by downward mixing of momentum at the top of the ABL). Thus, the link between the large-scale, upper-level flow anomalies over the central Andes and the local, near-surface moisture changes on the Altiplano occurs throughout the dynamical modulation of the regional slope circulation patterns (Fig. 4.2).

However, moisture variability in the continental lowlands, being always higher than about 12 g/kg, does not further effect the precipitation occurrence in the Andes.

4.3.2.1. Inter-annual variability and El Niño

Besides the diurnal and seasonal cycles of rainfall, the Altiplano experiences strong inter-annual fluctuations. Several studies agree that a significant fraction of this variability is related to the El Niño Southern Oscillation (ENSO). El Niño (La Niña) years are connected with negative (positive) precipitation anomalies.

4.4. Climatic delimitations of tropical glaciers

Kaser (1996b) suggested the following requirements to define an area as tropical (inner and outer) from a glaciological point of view:

**Radiative delimitation** The area must be within the astronomical tropics (the trop-
4.5 Glacier-climate interaction in the outer tropics

Figure 4.2: Schematic representation of the circulation patterns and different air masses over and adjacent to the central Andes, in a vertical-longitude section at the latitudes of the Altiplano, for (a) rainy episodes and (b) dry episodes. Large, open arrows indicate the direction of the upper-level, large-scale flow. Solid (dashed) curves represent the transport of moist (dry) air by the regional circulation over the Andean slopes. Thin vertical arrows represent the large-scale subsidence over the subtropical SE Pacific that maintains the strong trade inversion (solid line); figure by Garreaud et al. (2003).

Thermal delimitation On average the daily temperature variation, $\Delta T_d$, of the area exceeds the yearly temperature variation, $\Delta T_a$.

Hygric definition The area must be within the oscillation of the ITCZ.

The resulting delimitations for tropical glaciers and the distribution of glacier area per country are shown in Fig. 4.1.

4.5. Glacier-climate interaction in the outer tropics

Given a seasonality of humidity related variables but only small seasonal temperature variations (as discussed above), following considerations result for the yearly
course of outer tropical GMB (e.g., in Kaser, 2001): a notable accumulation occurs only during the wet season. During the dry season, there is only little accumulation, but ablation is also reduced. This is because of the dry air, which induces much of the available energy being consumed by sublimation. Sublimation, however, is an ineffective process of ablation compared to melting from the energetic point of view (2.835 MJ/kg is the energy needed for sublimation versus 0.334 MJ/kg for melting).

In the inter-annual timescale, GMB in the Cordillera Blanca is strongly influenced by the occurrence of ENSO events. In fact, during El Niño warm events, outer tropical glaciers usually experience a deficit of precipitation, an increase of air temperature, and thus a strongly negative mass balance (Wagnon et al., 2001; Francou et al., 2002; Francou et al., 2004). However, recent studies confirm distinctive breakdowns in this relationship between ENSO and GMB occurring from time to time, with an above average GMB during El Niño years or negative GMB during La Niña years (Vuille et al., 2007).
5. The study site: Geographical and glaciological aspects

5.1. The Cordillera Blanca

The Cordillera Blanca is located in the Peruvian tropical Andes (77°80′ – 78° 8′30″ – 78° S) and stretches about 180 km from NNW to SSE; extending only 30 km longitudinally. The high alpine mountain ridge includes 27 peaks reaching elevations above 6000 m a.s.l., the highest among them is Nevado Huascáran Sur (6,768 m a.s.l.). The highly glacierised Cordillera Blanca harbors more than 25% (with respect to surface area) of all tropical glaciers (Kaser, 2001).

Cordillera Blanca glaciers are, however, of major economic and environmental importance, releasing meltwater for the arid western part of the country during the dry season (May to September), when little to no rainfall occurs. In fact, much of the water resources consumed for agricultural, domestic and industrial purposes on the arid west coast of Peru originate from snow and ice in the high Andes.

Fig. 8.1 shows the location and glaciation (1970) of the Cordillera Blanca in Peru.

[Kaser, 2001; Georges, 2004; Juen, 2006; Vuille et al., 2007]

5.2. Artesonraju glacier

Artesonraju Glacier is situated in the northern Cordillera Blanca at 8°55′ S and 77°38′ W between 4750 and 5100 m a.s.l. and covers an area of around 5.7 m². It’s glacier basin gently slopes between 1 to 10°, ending up in heavily crevassed and steep flanks (up to 60°) of the surrounding mountain ridges. In fact, Artesonraju glacier is surrounded from the West, North and East by mountains reaching elevations between 5500 and 6025 m a.s.l.. At the end of the glacier tongue the valley plunges down steeply southwards to Lake Artesoncocha, where it becomes flat again, leading from south- to southwest toward the lake Laguna Parón. The catchment covers an
area of 7.7 km$^2$ with a glacierised area of 74%. [Juen et al. (2006)]

Figure 5.1.: The Cordillera Blanca in Northern Peru. Gray shadings refer to ice extension in 1970, with the names of the 12 mountain groups. The inlay shows the Cordillera’s position in South America (in Georges, 2004)
5.2 Artesonraju glacier

(a) Artesonraju glacier in humid season (photo taken by Marlis Hofer in February 2007)

(b) Artesonraju glacier in dry season (photo taken by Michael Winkler in June 2005)

Figure 5.2.
6. Methods of glacier mass balance modelling

6.1. Basic definitions and formulations

6.1.1. The continuity of mass and energy on the glacier-atmosphere interface

The principle of continuity requires a balance of incoming and outgoing vertical fluxes on the glacier surface. The conserved quantities are:

- Energy
- Mass
- Momentum

While the momentum balance is disregarded, the mass balance on the glacier-atmosphere interface consists in the water balance (Kuhn, 1981).

The balance of liquid water is formulated as follows (general water balance):

\[ P = R + E + S \] (6.1)

with \( P \) being precipitation, \( R \) runoff, \( E \) evaporation and \( S \) storage (Kuhn, 1981). For glaciers the storage term is the mass budget \( MB \):

\[ \text{storage} = MB = \text{accumulation} - \text{ablation} \] (6.2)

The mass and energy balances are explained in the adjacent sections.

6.1.2. The specific mass balance

The specific mass balance \( b \) (kg/m\(^2\), conventionally m w.e.) on a certain location \( \vec{x} \) of the glacier surface is defined as the difference between input \( c \) (accumulation) and outgoes \( a \) (ablation) over the period \( \Delta t \):
6.2 The surface energy budget

\[ b(\vec{x}, \Delta t) = c - a \] (6.3)

The accumulation \( c \) is the sum of different processes that contribute to the local ice or snow budget:

\[ c = p + t^+ \] (6.4)

where \( p \) is solid, or stored liquid precipitation and condensation, and \( t^+ \) positive transport of snow or ice by avalanches or wind (Kuhn, 1981). Another local source of mass can be refreezing melt water. Ablation \( a \), on the other hand, is the sum of processes that remove mass:

\[ a = m + t^- + \text{calving} \] (6.5)

where \( m \) is ice loss by melting or sublimation, \( t^- \) the negative transport by avalanches or wind, and calving the process of ice loss into the sea or a lake (Kuhn, 1981). Note that, in general, melting can occur on the surface of a glacier, but also on the base or internal (surface, basal or internal melting).

6.2. The surface energy budget

The three types of heat transfer to or from any location are:

- radiance (not bound to matter)
- heat conduction (energy transfer between particles)
- or convection (energy transport by particles).

The thermal exchange at the glacier surface, the interface between ground and atmosphere, is determined by vertical energy fluxes (\( W/m^2 \)) on a horizontal plane. Conventionally, flows towards the surface are defined positive. According to the three types of energy transfer listed above, the surface energy balance, \( SEB \), can be formulated as follows:

\[ R + H + LE + Q_G + Q_P = Q_M \] (6.6a)

\[ R = SW + LW \] (6.6b)
with

\[ SW = SW_{in} - SW_{out} \]
\[ LW = LW_{in} - LW_{out} \]

\( R \) is the sum of shortwave (SW) and longwave (LW) radiation balances; \( H \) and \( LE \) are the sensible and latent heat fluxes (convective); \( Q_g \) is the conductive heat flux, \( Q_P \) the heat supplied by precipitation, and \( Q_M \) the resulting imbalance between the left hand side energy fluxes in Eq. 6.6a.

### 6.3. Melt modelling

Apart from calving or snow transport by avalanches or wind, ablation \( a \) in Eq. 6.5 occurs through sublimation (an ineffective ablation process as it consumes about 8 times as much energy as melting per kg mass), and melting (based on the assumption that melt water immediately runs off without refreezing on- or sub-surface). Based on their way in assessing ablation by melting, glacier melt models generally are classified into two categories (Hock, 2003):

- the *temperature-index* models (which assume an empirical relationship between air temperatures and melting rates),
- and the more sophisticated *energy balance* models, which examine the SEB in order to quantify the energy available for melting.

#### 6.3.1. Temperature-index melt models

The temperature-index or degree-day method has firstly been used by Finsterwalder (1887) and then been widely applied and further refined. Due to following reasons, temperature-index models have still remained a common approach for melt modelling (Hock, 2003):

1. Temperature is the only variable required for model input, and temperature data are available for many sites in long time series.
2. Temperature-index models are of low computational demand.
3. They have proven a good model performance, despite simplicity.
6.3 Melt modelling

6.3.1.1. The degree-day factor

The most basic formulation of temperature index models relates the amount of ice or snow melt, \( m \), in mm w.e., during a period of \( n \) time intervals, \( \Delta t \), to the sum of positive air temperatures of each time interval, \( T^+ \), in °C, during the same period; the factor of proportionality being the degree-day factor, \( DDF \) (Hock, 2003):

\[
\sum_{i=1}^{n} M = DDF \sum_{i=1}^{n} T^+ \Delta t
\]  

(6.7)

Commonly, daily or monthly time intervals \( \Delta t \) are used for temperature integration. Though, the use of daily temperature means can be misleading: mean temperature may be negative indicating no melt, whereas melt conditions may have prevailed during part of the day.

In practice, the degree-day approach often assumes the form

\[
M = f_m(T_d - T_0) \iff T_d > T_0 \\
M = 0 \iff T_d \leq T_0
\]

where \( M \) is the daily melt, \( T_d \) is daily mean temperature, \( T_0 \) a threshold temperature beyond which melt is assumed to occur, and \( f_m \) the melt factor (Hock, 2003). Note that, if \( T_0 \neq 0^\circ C \), \( f_m \) is not the same as \( DDF \) (the inclusion of a threshold temperature accounts for the fact that melt does not necessarily occur at air temperatures > 0°C; Kuhn, 1986).

6.3.2. Surface energy balance melt models

A physically based approach to compute melt is to assess the energy fluxes to and from the glacier surface (Eq. 6.6a). At surface temperatures exceeding 0°C, the SEB melt models assume any surplus of energy at the surface (the imbalance term \( Q_m \) in Eq. 6.6a) being consumed for mass loss by melting.

Energy balance models fall into two categories: point studies and distributed models. The former assess the SEB at one location, usually the site of a climate station. The latter involve estimating the budget over an area, usually on a square grid. However, promoted by increasing availability of digital terrain models and computational power, recently increasing efforts have been devoted to areal melt modelling and increasing temporal resolution modelling; e.g. hourly time steps (Hock, 2003).
6.4. Example for GMB modelling: ITGG-2.0

The GMB model presented here has been developed by the Innsbruck Tropical Glaciology Group (ITGG) and referred as ITGG model, version 2.0. It is an extended version of the vertical balance profile (VBP) model presented earlier (Kaser, 2001), which in turn is based on the principles of a model published by Kuhn (1981, 1989). It has been applied in Juen (2006) for mass balance reconstructions at several sites in the Cordillera Blanca.

6.4.1. Model structure

The calculation of absolute GMB (defined as the GMB of a glacier in kg, in contrast to the specific balance $b$ in kg/m$^2$ at one location of the glacier) in the ITGG-2.0 can be summarized basically in four steps:

1. the calculation of the specific net mass balance $b_r$ [mm w.e.] at one reference level $z_r$ over a defined period (the time step of the model),

2. the calculation of the vertical balance profile $VBP$ and the absolute vertical balance profile $VBP_{abs}$ [mm w.e.] as a function of $b_r$ and $\Delta b$ (discrete altitude steps $\Delta z$):

\[
VBP \equiv \frac{db(z)}{dz}
\]

and

\[
VBP_{abs}(z) \equiv b_r + \Delta b
\]

\[
\Delta b = \frac{db(z)}{dz} \cdot \Delta z
\]

\[
z = z_r + \Delta z
\]

3. the calculation of the absolute mass balance $MB$ [kg] of the glacier by multiplying the $VBP_{abs}(z)$ with vertical distribution of the glacier surface area $A(z)$ [m$^2$] and integration over the altitude range of $z$:

\[
MB = \sum_z VBP_{abs}(z) \cdot A(z)
\]

6.4.1.1. Calculation of specific mass balance $b$

Analogous to the $VBP$ model presented in Kaser (2001), the specific net mass balance at the reference level in the ITGG-2.0 is calculated as the difference between $c_r$ and $a_r$, accumulation and ablation at one reference level:
\[ b_r = c_r - a_r \]  \hspace{1cm} (6.11a)

and further:

\[ a_r = \tau F(f) \cdot [R|_{z_r} + H|_{z_r}] \]  \hspace{1cm} (6.11b)

with \( \tau \) being the duration of the ablation period, \( R \) the radiation balance and \( H \) the sensible heat flux, and

\[ F = \frac{1-f}{L_m} + \frac{1}{L_s} \]  \hspace{1cm} (6.11c)

\[ f = \frac{L_s S}{L_m M + L_s S} \]  \hspace{1cm} (6.11d)

\( F(f) \) converts the available energy, resulting from the imbalance between \( R \) and \( H \), into ablation rates due to melting and sublimation (\( f \) is the ratio of energy that is consumed for sublimation to the total amount of available energy). Compared to the full SEB in Eq. 6.6a, two fluxes are neglected in the ITGG-2.0: the conductive heat flux and the heat supplied by precipitation.

See Kaser (2001) and Juen (2006) for more details about the ITGG-2.0, e.g. the theoretical basis and applications of the \( VBP \) model.

### 6.4.2. Input variables of the ITGG-2.0 and the objective of this master thesis

This master thesis is a try to fit reanalysis data to the surface conditions of Artesonraju glacier, with the objective to provide input data for mass balance models. To solve the ITGG-2.0, many input variables are necessary. In practice, only few data sets are available in the Cordillera Blanca for a limited number of sites. Thus, the ITGG-2.0 is constructed to run also with air temperature and one moisture related variable as the only input data. From one moisture related variable (e.g. precipitation) all other variables have to be derived. More precisely, the range of required input variables is predefined for one very 'wet' and one very 'dry' scenario. Intermediate values are then interpolated linearly according to the measured precipitation amounts in a monthly time scale (see Juen (2006) for a detailed description of how the input variables of the ITGG-2.0 are derived).
The objective of this master thesis is to derive air temperature and specific humidity above the glacier surface of Artesonraju in different temporal resolutions. If this approach is successful, it is possible to run the ITGG-2.0 in the examined time scales over the whole period when RA are available.
7. The NCEP/NCAR reanalysis project

7.1. Introduction

Similar to the ECMWF (European Center for Medium-Range Weather Forecasts) 40 Year Re-analysis, ERA-40, the NCEP (National Centers for Environmental Prediction) and NCAR (National Center for Atmospheric Research) have cooperated in a project to reanalyze observations for the entire globe from 1948 up to the present. By using a fixed model, artificial "climate jumps" in the data series, related to changes in the operational data assimilation system, could be avoided. Nevertheless, changes in the observing system (such as the implementation of satellite data) still affect the reliability of the data series. Referring to Kistler et al., 2001, the first part of this section gives an overview of the observational input collected by NCAR; whereas the second part discusses the reanalysis model and output provided by NCEP.

7.2. Observational data input of the NCEP/NCAR reanalysis model

The observations assimilated within the NCEP/NCAR reanalysis project include

- upper air radiosonde observations of temperature, horizontal wind and specific humidity;
- operational Television Infrared Observation Satellite (TIROS) Operational Vertical Sounder (TOVS) vertical temperature sounder from NOAA polar orbits over ocean excluded between 20° N and S;
- TOVS temperature soundings over land above 100 hPa;
- cloud-tracked winds from geostationary satellites;
• aircraft observations of wind and temperature;

• land surface reports of surface pressure;

• oceanic reports of surface pressure, temperature, horizontal wind and specific humidity.

These observations have been collected by the NCAR. However, throughout the reanalysis period starting from 1948, the global observing system has improved substantially. Generally, three major phases can be distinguished: an "early" period up to 1957 when the first upper-air observations were established; the "modern rawinsonde network" from 1958 to 1978; and the "modern satellite" era from 1979 to the present. The impact of the major changes in the observing system is complex and has been assessed in several studies.

7.3. NCEP/NCAR reanalysis model specifications and output variables

The reanalysis data assimilation system includes the NCEP global spectral model operational in 1995. To be specific, the analysis scheme is a three-dimensional variational (3DVAR) scheme cast in spectral space denoted spectral statistical interpolation. The model runs 28 vertical levels and has a triangular truncation of 62 waves T62 (about 210 km horizontal resolution).

The model output are gridded variables, including six-hourly instantaneous values, daily and monthly means of meteorological variables from the surface to the top of the atmosphere. Depending on the extend to which the reanalyzed variables include observational data, they have been classified into three types: A, B and C; where type A (such as upper-air temperature and geopotential height) is generally strongly influenced by observations and the most reliable product; type B (e.g. moisture variables, surface parameters) is attained both by observations and the model and is therefore less reliable; and type C, such as precipitation, is completely determined by the model (subject to the constraint of the assimilation of other observations) and should be used with caution.

All products of the reanalysis project can be obtained from NCAR, NCEP, and from the National Oceanic and Atmospheric Administration/Climate diagnostics Center (NOAA/CDC). (Their Web page addresses can be accessed from the Web page of the NCEP-NCAR reanalysis at http://wesley.wwb.noaa.gov/Reanalysis.html)
8. Statistical downscaling

8.1. Introduction

Within the development of coarse-scale general circulation models GCM (spatial resolutions typically in the order of $\sim 100$ km), the requirement arose to use GCM output for analyzing and forecasting meteorological variables in regional scales (Leung et al., 2003) and at individual sites. Though the inability of GCMs to resolve important sub-grid scale features, such as soil properties and topography, caused the direct use of GCM variables be inadequate for regional or local climate research. As a consequence, two groups of techniques have emerged in order to bridge the gap between coarse-scale GCM information and regional climate or local-scale surface variables. On the one hand, the Statistical Downscaling (SD) technique (Wilby et al., 2001; Wilby et al., 2004) is based on empirical relationships between GCM output and observed weather and climate on one or more specific sites. Alternatively so-called Dynamical Downscaling techniques have developed consisting principally in nesting regional climate models RCM (spatial resolutions typically in the order of $\sim 10$ km) into the GCM (Hay and Clark, 2003; Mearns et al., 2003).

SD involves developing quantitative relationships between large-scale atmospheric variables from GCMs and measured surface variables. In current SD literature the large-scale variable is generally referred to as predictor and the local-scale variable as predictand. Predictor sets are typically derived from sea level pressure, geopotential height, wind fields, humidity variables, and temperature variables (Wilby et al., 2004). Concerning the choice of predictand variable, however, current SD studies have concentrated mainly on mean, minimum or maximum temperature and precipitation (e.g., Katz and Parlange, 1996; Hay and Clark, 2003; Schmidli, 2006) in daily or monthly timescales, at mid-latitudinal sites of the northern hemisphere (Wilby et al., 2004).
8.2. Regression-based SD methods

There are several studies who examine the relative skill of different SD techniques, or SD against Dynamical Downscaling approaches. The SD techniques are classified among others by Wilby et al. (2004) into three main types; they are *weather classification*, *regression models*, and *weather generators*. In this master thesis the regression-based technique is applied.

Principally regression-based SD relies on empirical relationships between predictor and predictand. However, individual approaches differ generally by the mathematical transfer function, the choice of predictor variables, or the statistical fitting method. To date, linear or non-linear single or multiple regression (e.g., Kidson and Thompson, 1998; Schoof and Pryor, 2001; Khan et al., 2006), artificial neural networks (e.g., Schoof and Pryor, 2001), canonical correlation and principal components analysis (e.g., Kim et al., 1984; Benestad et al., 2002) have been applied to find functional relationships between predictor and predictand.

In this master thesis, generally single and multiple regression analysis is applied examining linear and non-linear functional relationships between the predictor variable(s) and the predictand. A further description of applied method and mathematical tools is given in Chap. 10.

8.3. General scheme of SD: a step-by-step process

Referring to the ”Guidelines for Use of Climate Scenarios Developed from Statistical Downscaling (IPCC supporting material)” by Wilby et al. (2004), which mainly correspond to the proceeding of other authors cited here, this section summarizes the general steps of a SD approach, are:

**Choice of the adequate SD method** The choice of the statistical method to find a predictor-predictand relationship is to a great extend dependent on statistical properties of the predictand variables. For normally distributed variables, such as temperature (the first predictand examined in this master thesis), linear regression might be a satisfying tool (because, e.g., model evaluators such as the coefficient of determination assume normality of the input data). For non-gaussian variables like precipitation, more complicated models need to be applied. Specific humidity, the second predictand in this master thesis,
8.4 Advantages and limitations of SD

is often considered difficult to reanalyze and forecast for specific sites because of its heterogeneity in space and time.

Another problematic issue for SD is the handling of extreme values who are often misrepresented.

Selecting appropriate predictor variables In order to systematically choose powerful predictor variables, a profound knowledge of the GCM model and the driving forces of local- and regional scale meteorology is fundamental. However, some of the earliest downscaling studies used GCM grid box values of the predictand to derive station-scale values of the same variable (e.g., local surface temperature). The availability of reanalysis data sets has significantly increased the number and variety of candidate predictors. Unfortunately, there have been relatively few systematic assessments of different predictors (e.g., Huth, 2005).

Standardization of data Standardization is widely used prior to SD to reduce systematic biases in the mean and variance of GCM predictors relative to observations. The procedure typically involves subtraction of the mean and division by the standard deviation of the predictor for a predefined baseline period. The main issues here relate to the choice of the baseline and averaging window.

Model evaluation using independent data A standard approach to model validation is to use one portion of the available records for model calibration and one portion for the evaluation. Further evaluation involves an inter-comparison of different SD techniques (e.g., B., 2001; Khan et al., 2006) or the assessment of SD skill relative to RCMs (e.g., Kidson and Thompson, 1998; Murphy, 1998; Schmidli, 2006).

8.4. Advantages and limitations of SD

Probably the most important advantage of SD techniques versus, e.g., the use of RCM, is their low computational demand. This represents, above all in developing nations with limited computational resources, an important pragmatic argument for SD. Moreover, SD is considered particularly useful in heterogeneous environments with complex physiography or steep environmental gradients (as islands, mountainous or land/sea contexts). On the other hand, one major limitation of SD is that,
8.4 Advantages and limitations of SD

Figure 8.1.: General statistical downscaling scheme, Figure by Wilby et al. (2001)

Unlike RCMs, most SD methods cannot be applied unless station data are available for model calibration. Moreover, the majority of SD schemes are unable to incorporate land-surface feedbacks to climate change; this is because when using SD, the output is entirely driven by the free atmosphere predictor supplied by the GCM. Thus it is doubtful whether the results obtained by SD are valid under different climatic conditions. This is one major uncertainty of SD when applied in climate change impact assessments.
Part II.

Downscaling
9. Data

9.1. Observational data from automatic weather stations

In 2004, two automatic weather stations (AWS) were installed by the ITGG on moraines near the glacier. One of them is located on the boarder moraine South to the glacier tongue and the other on a moraine at the opposite site of the valley. In addition, there are one automatic energy balance station (EBS) by the IRD and one automatic radiation balance station (RBS) by the ITGG on the ablation zone of Artesonraju glacier. See Fig. 9.1 for the exact positions of the stations within the area at and around Artesonraju glacier.

Figure 9.1.: Map of Artesonraju Glacier with indicated stations and stake positions (Juen, 2006)
9.1 Observational data from automatic weather stations

In this master thesis, data from the EBS at 4880 m a.s.l on the ablation zone of Artesonraju glacier are examined. Specifically, temperature and specific humidity, at the measuring height of one meter above the surface, are the variables to be ‘predicted’ by the RA. Tab. 9.1 shows instrument specifications of the EBS. At the RBS no humidity measurements are taken and the accuracy of temperature measurements is limited (Judd Com./Ultrasonic depth sensor, accuracy $\pm 1^\circ C$ according to manufacturers).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Sensor</th>
<th>Sensor h.</th>
<th>Sampling i.</th>
<th>Storage i.</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>temp</td>
<td>Vaisala</td>
<td>1 m</td>
<td>10 s</td>
<td>30 min</td>
<td>$\pm 0.2^\circ C$</td>
</tr>
<tr>
<td>rhum</td>
<td>HMP45C</td>
<td></td>
<td></td>
<td></td>
<td>$\pm 2%$</td>
</tr>
</tbody>
</table>

Table 9.1.: Specifications of temperature and humidity measurements at EBS (abbreviations: temp temperature; rhum relative humidity; h. height; i. interval; and acc accuracy according to manufacturers).

Figure 9.2.: EBS on the ablation zone of Artesonraju glacier; February 2007

9.1.1. Measuring period and data gaps

At the beginning of this study in September 2006, EBS temperature and humidity records were available from 24 March 2004 (when the stations were installed) to 5 April 2006 (last field trip). This period is defined as calibration period for this SD approach, in the following referred to as BP (baseline period). In February
2007, another half year’s records were taken during an ITGG field trip, and are now available for the validation of the downscaling results. Because of several logger breakdowns and instrument failures throughout the measuring period, however, none of the stations exhibits a complete data series. At the EBS, for instance, out of the 795 day calibration period only 631 days remain after data gaps are retaken. The main periods during which temperature and humidity measurements at the EBS failed, are the following:

- 5 to 23 November 2004
- 15 December 2004 to 10 February 2005
- 11 to 20 March 2006

### 9.1.2. The predictands

The predictand variables of the SD approach are temperature and specific humidity, measured at one point (location of EBS) one meter above the surface 4880 m a.s.l., in the ablation zone of the outer tropical glacier Artsonraju. Given the existing GMB models input data in different temporal resolutions, all possible time steps (according to the availability of RA data) are examined:

- six-hourly instantaneous values at the main synoptic times 00, 06, 12 and 18 UTC (according to the availability of the reanalysis data),
- monthly means,
- and daily means.

### 9.1.3. Calculations

#### 9.1.3.1. Calculation of specific humidity

Specific humidity is not directly measured at the EBS. Hence it has been calculated as a function of relative humidity, temperature and atmospheric pressure as follows:

\[ sh = \frac{6.22}{p} \cdot e \]

\[ e = 6.11213 \cdot rh \cdot \exp \left( \frac{17.08085 \cdot T}{234.175 + T} \right) \]
with \( sh \) being the specific humidity in g/kg; \( p \cong 560\text{hPa} \) approximated for atmospheric pressure at the EBS (not measured); \( e \) the water vapor content in hPa; \( rH \) the relative humidity measured at the EBS in \( \% \); and \( T \) the temperature measured at the EBS in \( ^\circ\text{C} \).

### 9.1.3.2. Average building and time lag

The data logger at the EBS calculates half-hourly means from values (temperature, or relative humidity, respectively) measured every 10 seconds and stores them at the point in time after the measuring period. For instance, the 00h30’00” UTC value is the average of 180 values from 00h00’10” UTC to 00h30’00” UTC. The predictand values at 00, 06, 12 and 18 UTC are hourly averaged values of the preceding hour.

Daily means are not calculated unless hourly means for the whole day are available. Monthly means are calculated when at least 20 daily means are available.

Data are stored at the data logger in Peruvian local time:

\[
UTC(h) = LT + 5
\]  

### 9.2. The NCEP/NCAR reanalysis data

#### 9.2.1. The predictors

Following variables are examined as potential predictors for 1m - temperature and specific humidity at the EBS:

<table>
<thead>
<tr>
<th>Variables</th>
<th>Time step</th>
<th>Vertical levels</th>
<th>Grid points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air temperature, specific humidity</td>
<td>6-hourly values</td>
<td>500, 600 and 700hPa</td>
<td>77.5(^\circ\text{W}) and 75(^\circ\text{W}), 7.5 and 10(^\circ\text{S})</td>
</tr>
<tr>
<td></td>
<td>daily and monthly means</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 9.2.** Reanalysis data from the NCEP/NCAR examined for SD in this master thesis

According to the classification of NCEP/NCAR reanalysis, the variable air temperature belongs to type A (which is strongly influenced by measurements; the ”most reliable”). There again humidity variables are classified into type B, being influenced both by the model and measurements (Kistler et al., 2001).
9.2.2. Model topography and surface heights of the examined grid points

Unrealistic GCM topography due to a coarse grid resolution is, particularly in mountainous regions, one major constraint for predicting local weather and climate (Leung et al., 2003). The reanalysis model topography (see Fig. 9.3), in fact, only poorly represents the narrow but high Cordillera Blanca mountain range. Consequently, even if screen variables (2 m above the surface) are available from the reanalysis data, they are not taken into consideration here. Tab. 9.2.2 lists the exact surface heights at the four grid points examined in this study in the model topography.

<table>
<thead>
<tr>
<th>Grid point</th>
<th>Surface height (m a.s.l.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>75°W 10°S (SE)</td>
<td>1719</td>
</tr>
<tr>
<td>75°W 7.5°S (NE)</td>
<td>712</td>
</tr>
<tr>
<td>77.5°W 7.5°S (NW)</td>
<td>1910</td>
</tr>
<tr>
<td>77.5°W 10°S (SW)</td>
<td>1993</td>
</tr>
</tbody>
</table>

Table 9.3.: Surface height of the reanalysis model topography at the four examined grid points

Figure 9.3.: NCEP model topography in the South American Andes (the crosses indicate the four grid points closest to the study site)
10. Statistical methods and applied software

In this diploma thesis, the Curve Fitting Toolbox of MATLAB R2006a is applied for fitting the predictor variables to the predictands; the methods are basically linear and non-linear single and multiple regression analysis. Moreover, statistical properties of the variables are examined and results of the SD approach evaluated. This section gives an overview of the basic methods and definitions used hereafter. In this section, the term "model" is used generally for a regression model. In the case of SD in this master thesis, the model is the SD model (or transfer function) with the predictors being the model input and the predictands the model output.

10.1. Least squares regression

MATLAB Curve Fitting Toolbox uses the method of least squares when fitting data. To obtain coefficient estimates of the model, the least squares method minimizes the summed square of residuals. The residual for the i-th data point, $R_i$, is identified as the difference between model and observational data.

$$R_i = \hat{Y}_i - Y_i$$  
$$\hat{Y}_i \equiv Y(X_i)$$  
$$i = 1 : n$$

In Eq. 10.1b, $Y_i$ is the i-th data point of the observations and $\hat{Y}_i$ the model output for the i-th data point; $n$ is the size of $X$, $Y$, and $R$. More precisely, the model $\hat{Y}$ is a functional relationship with $X_i$ being the independent variable: the predictor variable. The functional relationship between the predictor variable $X_i$ and the observational data $Y_i$ (the predictands) can thus be determined in the following form:
In Eq. 10.2 $\beta$ is the vector of regression coefficients. These are, in the case of single linear regression, slope $a$ and intercept $b$ in the linear equation:

$$Y = \hat{Y}(\beta, X) + R$$

\[
\hat{Y}_{\text{lin}} = a \cdot X + b \quad (10.3a)
\]

$$\beta = \{a, b\} \quad (10.3b)$$

In the case of multiple regression, $X$ is a matrix ($n \times m$) with $m$ predictors and $\beta$ has the size ($1 \times m + 1$).

10.2. Model evaluation

Several definitions are common in statistical literature for goodness of fit of a regression model. A model can be evaluated based on the uncertainty of the model coefficients estimate (e.g. $a$ and $b$) and the proportion of the variability in the data explained by the model ($r$-square).

If independent data are available, it can be validated whether the model predicts new observations with high certainty.

Generally, methods to check the goodness of fit are both graphical and numerical.

10.2.1. Graphical methods for model evaluation

10.2.1.1. Correlation plot

The correlation plot displays the values predicted by the model versus the observations. It is a useful method to analyze the goodness of fit "quantitatively": by visualizing the form of statistical relationship, and it is also useful for the detection of outliers.

10.2.1.2. Residual analysis

Another graphical method for model evaluation is the residual analysis. The residuals are defined as model errors (Eq. 10.1b) and thus underlie the basic assumptions about an error:
10.2 Model evaluation

- to be normally distributed,
- and to have constant variance.

Furthermore the residuals are assumed to be independently randomly distributed over time, and over the range of $Y$.

In this master thesis, for the purpose of model evaluation, the following two methods of residual analysis are applied:

- to check the normality of the residuals via normal probability plots,
- to check the independence of residuals over time by displaying them as a function of time (run-sequence plots, Chambers et al., 1983; if the residuals display a systematic pattern over time, the model fits the data poorly).

10.2.2. Numerical measures for model evaluation

Numerical measures are a "quantitative" way of model evaluation, and useful to compare one model against others, or to compare the different stages of development of a model. The quantities used in this master thesis are:

- $r$-square,
- and the root mean squared error (RMSE).

In the case of one single regressor, fitted by least squares, $R$-square is the square of the Pearson product-moment correlation coefficient relating the regressor and the response variable. More generally, $R$-square is defined as the ratio of the sum of squares of the regression (SSR) and the total sum of squares (SST).

\[
SSR = \sum w_i (\hat{Y}_i - \bar{Y})^2 \quad (10.4)
\]
\[
SST = \sum w_i (X_i - \bar{X})^2 \quad (10.5)
\]
\[
r - square = \frac{SSR}{SST} \quad (10.6)
\]

In this diploma thesis $r$-square is applied for several steps of the downscaling approach:

- The choice of predictor (vertical level and grid point from the reanalysis data) (Chap. 11),
• evaluating the SD steps and finding the coefficients in the SD models (Chap. 12.5),

• and evaluation of the SD results.

Applying r-square for model evaluation assumes the predictor and predictand variables to be normally distributed. How far this assumption holds for the temperature and specific humidity in this SD approach is validated via Normal probability plots for hourly values, and daily and monthly means.

[Schlittgen, 1996; Ross, 2006; MATLAB R2006a Help]
11. Grid point selection

The examined predictor variables from the reanalysis data already represent a selection from numerous possibilities. In this chapter the examined predictor variables are further selected by choosing the grid point which corresponds best to the predictands. The method here is examining the percentage of explained variance by linear least-square regression between the predictor data at 12 grid points (4 horizontal grid points in 3 vertical levels) and the observational data. Furthermore, r-square of multiple linear regression with the 4 grid points as predictors is examined for every vertical level as well as a spatially interpolated grid point to the coordinates of the study site. Fig. 11.1 displays r-square of the predictor-predictand relationship for the examined cases.

11.1. Discussion

11.1.1. Time scales

For monthly means, r-square between $X$ and $Y$ exceeds 0.5, for humidity even 0.7. This shows that monthly means of predictor data represent climatic features even for the local scale on a Cordillera Blanca glacier surface.

Yet, in a daily timescale r-square is lower than 0.5; in a hourly time scale around 0.2. The decrease of r-square from the monthly to the daily timescale is larger than it is from the daily to the hourly time scale.

The low correlation between $X$ and $Y$ in the daily and hourly time resolution indicates that the use of only linearly transformed reanalysis data for GMB models does not introduce local scale meteorologic information for the site of Artesonraju glacier and can therefore not be further recommended.

11.1.2. Vertical level

For both temperature and specific humidity in all time scales, the 500 hPa level model data correlates best to the station data, despite monthly humidity means,
Figure 11.1.: Correlation coefficient r-square between EBS and reanalysis temperature temp (left) and specific humidity shum (right) data in different time scales (6-hourly instantaneous values top, daily second row, and monthly means bottom), grid points (the four grid points surrounding the study site NW NE SW SE, as described in Chap. 9.2, and INT; the four grid points interpolated to the study site, weighted by distance; MR is a linear multiple regression fit with the four grid points as predictors, only for the 500 hPa level) and vertical levels 500, 600 and 700 hPa for 6-hourly values and 500 and 600 hPa for daily and monthly means.
where the 600 hPa level shows out slightly higher values of r-square. The 700 hPa level, however, has shown a weak correlation to the locally measured values in an hourly time scale. A possible explanation is that the 700 hPa level is already in the lower ABL of the model atmosphere, close to the surface. Yet, the model topography, being completely unrepresentative for the examined study site, can not provide information about local meteorological features on Artesonraju glacier. Pressure at Artesonraju glacier has been measured during several field trips and amounts approximately 560 hPa.

### 11.1.3. Grid point

In terms of temperature, the SE grid point shows the highest correlation to the observational data. However, the NW grid point correlates systematically worse of all, even if it is closest to the study site. Generally, data from the two southern grid points fit better to the EBS data better than the northern. Yet, if the grid points are geometrically interpolated to the study site, the correlation does not increase with respect to the SE grid point. Also the multiple regression model, with the four grid points as model predictors, does not increase the correlation more then 1% of explained variance.

In terms of specific humidity, the differences between the grid points are smaller, even though the SE grid point outmatches the other in terms of r-square.

### 11.1.4. Summary

To sum up, the 500 hPa data is (according to the coefficient of determination) the best choice from the examined levels for further downscaling. Concerning the grid points, in sum SE seems to represent best the local conditions on the glacier surface. This is in good agreement with the outcomes of Georges (2004), who examines seasonal flow patterns over the Cordillera Blanca in relation to locally measured surface precipitation series (monthly means). He finds a consistently easterly flow direction over the tropical Andes at the 500 hPa level. On the other hand, he finds no proof for a distinct southern flow direction, but notes that the (small) meridional flow component varies between N and S throughout the seasons.

In terms of multiple regression at the 500 hPa level, with the four grid points as predictors, the resulting coefficients of determination do not show a distinct improvement with respect to the SE grid point (only few percent increase of explained variance). Moreover, the regression coefficients vary highly between the different
timescales, with physically unviable values (such as negative slopes and intercepts much higher than the variables’ dimensions). Therefore it is not applied any further here.

Hereafter the abbreviation X for predictor refers to the RA 500 hPa temperature, and specific humidity, respectively, of the SE grid point.
12. Hourly values: 00, 06, 12 and 18 UTC

12.1. Are four values per day representative for the daily cycle?

Reanalysis data in the hourly time scale are available only for four times per day, namely at the main synoptic times at 00, 06, 12 and 18 UTC. In this master thesis, I try to fit data measured at the EBS for these synoptic times, but no approach is done to extract a daily cycle out of the four values per day. However, in order to extract a daily mean, I investigated a comparison between different means calculated from the four values per day (linear and weighted), to the daily mean from the measurements taken every 10 s for 24 hours, on the example of EBS temperature over BP, as follows:

\[
T_d = \frac{aT_{d,01} + bT_{d,07} + cT_{d,13} + dT_{d,19}}{4} + R_d
\]

(12.1)

\[
a + b + c + d = 4
\]

(12.2)

\[
a, b, c, d \in [0, 4]
\]

(12.3)

where \(T_d\) is the daily mean temperature of the EBS at day \(d\); \(T_n\) the measured temperatures at the respective times of day; \(a, b, c, d\) the weighting coefficients constraint by the requirements \(a + b + c + d = 4\) and \(a, b, c, d \in [0, 4]\); and \(R_d\) the resulting residual. This has been examined for various coefficients \(a, b, c, d\) (steps of 0.01 through all possible values). The results give a correlation coefficient r-square for \(a = b = c = d = 1\) of 0.91. By weighting the four values with different combinations of coefficients, however, r-square could be improved only up to 0.93.
12.2. Data statistics

Tab. 12.1 shows statistical properties of temperature and specific humidity values in the hourly (Y) and six-hourly (X) timescale throughout BP.

<table>
<thead>
<tr>
<th></th>
<th>temp/[°C]: Y</th>
<th>temp/[°C]: X</th>
<th>shum/[g/kg]: Y</th>
<th>shum/[g/kg]: X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.4</td>
<td>-4.9</td>
<td>5.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Std</td>
<td>2.1</td>
<td>1.1</td>
<td>1.2</td>
<td>0.9</td>
</tr>
<tr>
<td>Min</td>
<td>-7.7</td>
<td>-8.9</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Max</td>
<td>6.0</td>
<td>-1.3</td>
<td>7.9</td>
<td>5.2</td>
</tr>
<tr>
<td>95p</td>
<td>3.5</td>
<td>-3.1</td>
<td>6.6</td>
<td>3.7</td>
</tr>
<tr>
<td>75p</td>
<td>2.0</td>
<td>-4.1</td>
<td>6.1</td>
<td>2.9</td>
</tr>
<tr>
<td>Median</td>
<td>0.5</td>
<td>-4.9</td>
<td>5.3</td>
<td>2.3</td>
</tr>
<tr>
<td>25p</td>
<td>-0.8</td>
<td>-5.7</td>
<td>4.4</td>
<td>1.5</td>
</tr>
<tr>
<td>5p</td>
<td>-3.6</td>
<td>-6.8</td>
<td>2.7</td>
<td>0.7</td>
</tr>
<tr>
<td>n</td>
<td>15207</td>
<td>3180</td>
<td>15207</td>
<td>3180</td>
</tr>
</tbody>
</table>

Table 12.1.: Statistics of the predictand’s (X) and predictor’s (Y) temperature (temp) and specific humidity (shum) data for the baseline period 24 March 2004 to 5 April 2006: mean; standard deviation (std); absolute minimum and maximum values (min and max); 95th, 75th, 50th (median), 25th and 5th percentiles; and n, number of not NaN elements of the sample.

12.2.1. Discussion

On average, the temperature from the RA, $X_{temp}$, is 5.3 °C lower than the temperature at the EBS, $Y_{temp}$. The specific humidity is on average 5.1 g/kg at the EBS, and 2.2 g/kg in the RA. Whereas the standard deviation of $Y_{temp}$ is twice the standard deviation of $X_{temp}$ (indicating the higher diurnal variability of temperature at the EBS than in the RA), the standard deviation of specific humidity near the glacier surface, $Y_{shum}$, is only 0.3 g/kg higher than in the RA.

The range of values of $Y_{temp}$ is 13.7 °C, with an absolute maximum of 6 °C (a rather high value one meter above the glacier surface), but only 5% of the values exceed 3.5 °C. However, 90% of the Y temperature values are between ±3.5 °C, but the 25th and 75th percentiles are not symmetrical. This indicates that the 1m - temperature above the glacier surface is not normally distributed at least for the outer 10% of the values. A possible explanation is the upper boundary condition that the ice (or snow) surface can not exceed 0°C.
12.2 Data statistics

12.2.2. Standardization of \( X \)

Given the 500 hPa level not corresponding to the EBS height of 4880 m a.s.l, the climatological means and variances of \( X \) and \( Y \) differ by elevation bias (Tab. 12.1). This bias is removed by standardizing the distributions of \( X \) to the mean \( \overline{Y}_{BP} \), and variance \( \sigma(Y) \) of \( Y \):

\[
X_s^h = \frac{\sigma(Y_{h,BP})}{\sigma(X_{h,BP})} \cdot \left[ X^h - \overline{X}_{BP} \right] + \overline{Y}_{BP}
\]  

(12.4)

This conversion is linearly applied to all values and does not change the correlation between \( X \) and \( Y \). Hereafter, the lower index \( s \) in \( X_s \) refers to this conversion; the upper index indicates the time scale of the values, \( h \) for hourly.

12.2.2.1. Discussion

Figs. 12.1 (b) and (d) are the equivalent to Figs. 12.1 (a) and (b) with standardized predictor \( X_s \), \( Y \) remaining the same. Whereas for temperature, deviations between the two curves are visible above the 90th and below the 10th percentile (20%), for specific humidity, only about 50% of the values (in the range between 4 and 6 g/kg) of \( X \) and \( Y \) show the same distribution.

12.2.3. Normality check of hourly means

12.2.3.1. Temperature

Fig. 12.1 is a normal probability plot to see in greater detail whether \( X \) and \( Y \) are Gaussian variables. If the values fit the linear function (straight line in the plot), they can be considered normally distributed (the probability scale on the y-axe in Fig. 12.1 is logarithmical). While in (Fig. 12.1 a) the \( X_{temp} \) values fit the normal distribution well, the \( Y_{temp} \) values differ from the straight line for at least 20% of the values, namely the values above 3 °C, and and below -2 °C.

12.2.3.2. Specific humidity

\( Y_{shum} \) values deviate strongly from the normal distribution (Fig. 12.1 c: around 40% corresponding to values of specific humidity above 6.2, and and below 4, respectively, g/kg). Of \( X_{shum} \), the extreme values above the 90th and below the 10th percentiles do not correspond to the normal distribution, neither.
Figure 12.1.: Probability plot of temperature (a and b) and specific humidity (c and d) hourly X and Y values compared to normal distribution (straight lines) before (a and c) and after (b and d) standardization of mean and standard deviation of X to Y (c3).
12.3 The diurnal evolution of temperature and specific humidity of X and Y (case study)

12.2.3.3. Discussion

The normal probability plots show that the reanalysis data are to a large percentage normally distributed, while deviations of the observational data from a normal distribution are visible. These deviations are very clearly for specific humidity. Thus linear regression analysis validated by r-square might be an appropriate tool for temperature hourly values, but not for specific humidity.

The probabilities of the EBS data can be explained by considering the boundaries of possible temperature values on the glacier surface: strong cooling of the ice surface during the nights due to the high emissivity of snow and ice, but a limited heating because the (pure) mix of ice and water can not exceed the melting point 0 °C.

The upper limit of specific humidity values of both the EBS and the RA is the saturation humidity. The lower limit of the RA values is 0 g/kg (in Fig. 12.1 c).

12.3. The diurnal evolution of temperature and specific humidity of X and Y (case study)

12.3.1. Introduction

Figs. 12.2 and 12.3 display diurnal variations (in local time) of temperature and specific humidity from the RA and the EBS. Nine days with complete data in humid (01.-10.02.2006) and nine days in dry season (29.06.-07.07.2005) are chosen in order to examine

- the general evolution of temperature and specific humidity of X and Y, in order to discuss diurnal meteorology on the glacier surface and in the model atmosphere;

- differences between the diurnal cycles in wet and dry season; to see whether it is necessary to define seasons for the SD (different statistical relationships for defined seasons);

- deviations between model and observations: to examine whether the modeled data predict the locally observed daily cycle well, and if and where systematic deviations occur.

The difference between specific humidity in dry and humid season (Fig. 12.3) indicates that the chosen periods are good representatives for the respective season.
12.3 The diurnal evolution of temperature and specific humidity of X and Y (case study)

12.3.2. General evolution of temperature and specific humidity above the glacier surface (Y)

Fig. 12.4 displays the EBS temperature and specific humidity in the same plots for humid season (top) and dry season (bottom). This allows a better comparability of the temporal evolution of both variables relative to each other.

12.3.2.1. Discussion (based on Figs. 12.2, 12.3, and 12.4)

Humid season The daily amplitudes of the EBS temperature in the humid season (red line in Fig. 12.2 top) vary between 3 and 6 °C. Over the whole period, the temperature decreases in the afternoon, accompanied by an increase of the specific humidity. This indicates humidity advection via anabatic winds from the valley, but also clouds (formation or advection) and probably precipitation.
12.3 The diurnal evolution of temperature and specific humidity of X and Y (case study)

Figure 12.3.: Hourly variations of X (black points) and Y (red pointed line) specific humidity for a case study period in the humid season (top), and the dry season (bottom).

In fact, the stepwise and sharp decrease in temperature (compared to the dry season) after reaching the maximum at around 14 LT, indicates precipitation. In fact, precipitation events are often accompanied by an immediate strong cooling on the surface (even without frontal activities): this, firstly, because cloud and rain droplets absorb or reflect the incoming shortwave radiation; secondly, due to energy loss via melting, sublimation or evaporation, when rain or snow falls into a layer of dryer or warmer air (Mair, 2005). Moreover, if the precipitation is convective, it is often accompanied by down drafts of cold air (Mair, 2005). During the nights, the cooling is limited due to high values of (absolute) humidity in the atmosphere. If there is no humidity advection (e.g., in the mornings), the specific humidity pattern in the humid season follows the temperature evolution with a little time delay. The daily amplitude of specific humidity is about 2 g/kg, values vary between 5 and 7 g/kg. However,
12.3 The diurnal evolution of temperature and specific humidity of X and Y (case study)

Garreau et al. (2003) find that for the Altiplano (see Chap. 4.3), values of specific humidity at the surface exceeding the threshold value of 5 g/kg are always combined with deep moist convection.

Dry season In the dry season, the specific humidity is clearly reduced, maxima reaching only up to 5.5 g/kg. One very dry event occurs on July, 06th, with a minimum of 1.5 g/kg (in the other days the minimum is around 2.5 g/kg). Again an increase in specific humidity is visible in the afternoon, but absolute values are lower, even though the daily amplitude is higher. Temperature values in the (austral) winter period vary between -5°C to +4°C. The daily cycles of temperature are larger than in the humid season, due to the reduced values of (absolute) humidity (explained above).
12.3.3. Comparison between RA and EBS

12.3.3.1. Discussion (based on Figs. 12.2, 12.3, and 12.4)

**Temperature** In the humid season, the predictor (X, black points in Figs. 12.2 and 12.3) shows a small daily cycle (2 to 4°C). The largest diurnal temperature increase of both predictor X and predictand Y occurs between 7 and 13 LT, but it is smaller for X. However, the largest diurnal temperature decrease of Y occurs between 13 and 19 LT, while X decreases only after reaching the highest value at 19 LT. The small daily temperature cycles of the X indicates that the selected grid point from the RA is located in the upper ABL.

In the dry season, daily amplitudes of X remain of the same magnitude from 2 to 4°C, only in the last day it is 6°C. However, the maximum temperature of the predictor is reached earlier than in the humid season: between 13 and 19 LT (rather than at the 19 LT point in time).

**Specific humidity** The course of specific humidity of X throughout the case study period shows no regular diurnal variations, but variations occur rather from day to day. The largest deviations between X and Y are 2 g/kg, but no regular pattern of deviations is indicated. Specific humidity of X ranges from 4 to 8 g/kg in the period in the humid season and from 2.5 to 6 g/kg in the dry season.

12.4. Mean daily cycle

Fig. 12.5 displays the diurnal course of temperature and specific humidity, for the predictor and the predictand, averaged over BP.

12.4.1. Discussion of Fig. 12.5

12.4.1.1. Comparison between temperature and specific humidity mean daily cycles at the EBS

On average over the BP the temperature minimum at the EBS is reached at 6 LT, and the specific humidity minimum at 7 LT. Thereafter temperature and specific humidity both rise. The temperature maximum is reached at 14 LT, contemporaneously with the secondary maximum of specific humidity. However, specific humidity reaches the absolute diurnal maximum only 4 hours later, at 18 UTC. Thus specific
Figure 12.5.: Mean daily cycles of temperature (above) and specific humidity (below) of X and Y over BP: (a) X and Y in local time, (b) X temperature with a time shift of +4 hours (in respect to LT), specific humidity -10 hours according to conversion c1.
humidity follows the diurnal temperature cycle, except for the afternoon when temperature decreases contemporarily to the humidity maximum. This is in agreement to the diurnal evolution for single days in the case study period, Chap. 12.3.

12.4.1.2. Comparison between X and Y daily cycles

**Temperature** The standardization of the predictor X, (Eq. 12.4), adapted the variance of X to the variance of the predictand Y. This resulted in an increase of the daily temperature cycle of X, so that there is no big difference in the daily amplitudes between X and Y. Though, the two daily cycles throughout the day indicate a time delay: the predictand reaches the maximum value on average at 19 LT: this is 5 hours later than the predictor.

**Specific humidity** The daily amplitude of the predictand’s specific humidity averaged over BP is more than twice the daily amplitude of the predictor. The missing daily cycle of the predictor’s specific humidity has already been discussed in Chap. 12.3. In contrast to the daily temperature cycle, the standardization in Eq. 12.4 could not increase the daily specific humidity cycle of X. This indicates that the variable specific humidity is not affected by the surface in the reanalysis model. Thus it is questionable whether variances in the daily cycle at the EBS can be determined by the specific humidity of the RA.

12.5. Time shift adjustment between X and Y: SD model c1

12.5.1. Introduction

The time shift between the diurnal temperature cycles of X and Y has been indicated both in the average daily cycle and the diurnal evolutions in the case study period. This is assessed in greater detail, examining r-square between X and Y as a function of time shift $dt$ of X as follows:

$$r^2(dt) = r^2(X^{h+dt}, Y^h)$$  

(12.5)

$dt$ is the time shift in hours up to 24 h; $Y^h$ and $X^h$ are the time series of X and Y throughout BP, $h$ for 6-hourly values.
12.5.2. Results

The resulting r-square as a function of time shift $dt$ is displayed in Fig. 12.6.

![Rsquare between Y and time shifted X](image)

**Figure 12.6.** Variation of the coefficient of determination between model and station hourly values as a function of time shift (in hours) applied to the hourly values of the predictor

12.5.3. Discussion

Significant variations of r-square with $dt$ are visible for temperature. r-square at the starting point ($dt = 0h$) is less than half of the maximum r-square, which is reached at $dt = +4h$. Interestingly, r-square decreases to values $< 0.01$ for a time shift of around half a day, indicating that there is a significant daily cycle represented in the model data as well. Yet, for specific humidity, as expected, such a clear signal of daily evolution is not visible. Still, the initial value of r-squared is increased by around 15% of explained variance for $dt = 14h$. 
12.5.4. SD model c1

According to the results, conversion c1 is applied to the hourly values of X.

\[ X_s(h) = X_s(h + dt) \]  \hspace{1cm} (12.6)

with \( dt \) being the time shift in hours: +4 for temperature and \( a = -10 \) for specific humidity. Hereafter, index c1 refers to the normalized (after Eq. 12.4) and time shifted (after Eq. 12.6) predictor X.

12.5.5. Results and discussion (Fig. 12.5)

SD model c1 has been applied to the 6-hourly values of X with the time shift that induces the highest r-square between the single values of X and Y throughout BP. One way of evaluating the results is to picture again the mean daily cycles, this time for \( X_{c1} \) and Y, in order to examine if they fit better. This is done in Fig. 12.5 (b) for temperature and specific humidity. The results are as follows:

- For temperature the average daily cycles of \( X_{c1} \) and \( X \) are almost identical, deviations are smaller than 0.5°C.

- For specific humidity though, as expected, the mean diurnal cycles do not conform better after the time shift adjustment.

12.6. Examination of r-square for single times of day

12.6.1. Introduction

In order to examine whether for certain times of day the correlation between X and Y is much lower than for the whole day, the coefficient of determination is examined for single times of day when X is available, in comparison to the coefficient of determination of the whole day. These times differ between temperature and specific humidity, because \( X_{c1} \) is examined. So for temperature, hourly data are available at 3, 9, 15, and 21 LT; and for specific humidity, at 5, 11, 17 and 23 LT.
12.6 Examination of r-square for single times of day

12.6.2. Results

12.6.3. Discussion

For temperature values, the r-square between X and Y varies more depending on the time of day than for specific humidity values. Moreover, for temperature values a clear improvement is visible, when the whole data sample is examined (DAY), while for specific humidity values the 11 LT point in time shows the strongest correlation, even stronger than for the whole data sample (DAY). This indicates again that specific humidity from the RA does not provide information about the daily cycle at the EBS, even if temperature from the RA does.
12.7. Standardization of the predictor’s daily cycle: SD model c2

Since, particularly with respect to specific humidity, daily cycles of the predictand and the predictor do not conform, a standardization equivalent to Eq. 12.4 is applied to X, but now with an additional correction of the bias between the two daily cycles. This is a standardization of the daily cycles by applying 12.4 for the different times of day, as outlined in Eq. 12.7.

\[
X_{c2}^h(td) = \frac{\sigma[Y_{h,BP}^h(td)]}{\sigma[X_{h,BP}^h(td)\cdot] \cdot [X^h(td) - X_{BP}^h(td)] + Y_{BP}^h(td)}
\]  

(12.7)

Here \(X_{c2}^h(td)\) is the predictor X at the time of day \(td\) after \(c2\), normalized to the means and variances of Y for every examined time of day. \(c2\) has been applied to the predictor X both before and after the time shift correction \(c1\).

12.8. SD models c1 and c2: Summary and conclusions

Fig. 12.8 displays the coefficients of determination between X and Y for temperature and specific humidity after the different conversions of X (s, c1, c2, c1 and c2).

SD models c1 and c2 both try to add information about the diurnal variability of temperature and specific humidity at the EBS to X. In fact, the raw (or generally standardized) predictors determine the daily cycle at the EBS poorly. However, at the beginning, r-square between the 6-hourly values of X and Y amounts 0.19 for temperature and 0.3 for specific humidity (Fig. 12.8). SD model c1 increases the correlation between X and Y for temperature significantly to 0.39. However, for specific humidity, no such improvement is visible: after c2, the correlation increases only by about 3% of explained variance. SD model c2 improves the correlation between X and Y for temperature further: r-square amounts 0.5 after c1 and c2. For specific humidity, however, a time shift seems not to be effective, since there is no daily specific humidity cycle of X. Still, by increasing the variance of specific humidity as a function of time of day (SD model c2), r-square could be increased by more than 10% to values of about 0.42.
Figure 12.8.: r-square between X and Y specific humidity and temperature after different downscaling steps applied to the predictor X, that are: s (standardization, this conversion of X is linearly applied to all values and does not influence r-square, thus conforms to r-square of the raw data), c1 (time shift), c2 (standardization in respect to the daily cycle), and c1 and c2 (both c1 and c2 are applied to X, now specific humidity and temperature not available at the same times of day any more).

12.9. Evaluation of the results

12.9.1. Correlation plot

12.9.1.1. Temperature

Correlation plots are useful tools for the graphical, qualitative evaluation of a regression model. Fig. 12.11 shows the correlation plot of for the temperature, before and after the downscaling conversions c1 and c2 applied to X. Moreover, the numerical values of r-square, the regression coefficients in the linear equation $a$ and $b$, and the root mean square error $rmse$ are displayed in the graph for a quantitative evaluation of the fit. A comparison between the two plots shows a clear improvement. $a$ is
12.9 Evaluation of the results

more close to 1 which means less systematic under- or overestimation of the station values by $X_c$, and $rmse$ is smaller than before.

Low night temperatures at the EBS, however, can not be captured by the predictors even after the SD corrections.

12.9.1.2. Specific humidity

For specific humidity, the correlation plot shows only small indications of a linear functional relationship between predictor and predictand. Even if r-square increases after c1 and c2, it is difficult to see this improvement in the correlation plots. However, an upper boundary of 7 g/kg is visible for the EBS values: the RA here cover a range from almost 2 to 9 g/kg; thus no dependency of Y to X or $X_c$.

![Figure 12.9.: Temperature: Correlation plot [X, Y] before (left) and after (right) the SD transformations c1 and c2](image-url)
12.9 Evaluation of the results

12.9.2. Residual analysis

12.9.2.1. Temperature

Residual analysis is a method to examine the weaknesses of a regression model (Chap. 10). The residuals are plotted in the course of BP to examine whether they show a systematic pattern (in this plot visible only in the seasonal timescale). The time series covers BP (almost two years), but it is not continuous because data gaps are not displayed.

In fact, the residuals of temperature show a seasonal pattern. Thus the model is poor in determining seasonal fluctuations of the observed data.

The normal probability plot indicate deviations from a normal distribution for negative residuals (lower than -2 °C). As it has been visible in the correlation plot before also, the model does not catch the negative values of Y (lower than -2 °C). The normal probability plot for temperature, in Fig. 12.1, show deviations from normality in same range, too. This influences the effectiveness of c2 for these tem-

Figure 12.10.: Specific humidity: Correlation plot [X, Y] before (left) and after (right) SD model c2
12.9 Evaluation of the results

temperatures, as the correction does not take into consideration the asymmetry of the distribution by weighting the values according to their probability.

12.9.2.2. Specific humidity

The residuals of specific humidity show a distinctive seasonal cycle. This is even more obvious than in the temperature residual series. Thus, for specific humidity, a normalization of seasonal variability is necessary, too. However, the probability plots show out that the residuals are normally distributed. The probability plot in Fig. 12.1 for the specific humidity of x, where steeper slopes are visible in the outer parts of the curve than those between 2 and 6 g/kg, indicates that data points are accumulating in the range between ~ 1 and 1.5 g/kg, and 6.2 to 8 g/kg. The same limit for the predictor is indicated at 9 g/kg, but there are only few data points and thus not do influence the normality of the variable. However, a lower boundary being reached at 2.3 g/kg is visible for the reanalysis specific humidity as well.
12.9 Evaluation of the results

12.9.3. Case study period

In order to study the effectiveness of SD models c1 and c2 in more detail, X and Y for same period as in Chap. 12.3 are examined again. Fig. 12.8 is equivalent to Figs. 12.2 and 12.3 with the additionally displayed predictor $X_c$ (X after c1 and c2 for temperature, and after c2 for specific humidity). For temperature, a clear improvement is visible. There is no time delay visible any more, and even if deviations of 1°C can occur in the dry season (above all in the night), the model seems to capture daily fluctuations on the glacier surface. However, deviations between model and station temperature are still larger in the dry season (neither daily maximum nor minimum are ever captured, and deviations go up to 2°C).

For humidity, however, the improvement seems to be larger in the dry season, where the adjusted model values catch the daily cycle on the station. Nevertheless in humid season, e.g. on the 4th of February, model values show no visible relation to the station measurements and deviations up to 2 g/kg occur.

**Figure 12.12.:** Specific humidity: Time series and normal probability plot of the residuals before and after SD model c2
12.10. Multiple linear and nonlinear regression

12.10.1. Introduction

In addition to conversions c1 and c2 to the predictor X, the aptitude of multiple regression is examined, applying not only linear and constant but squared and cross product terms of the predictors temperature and specific humidity. This extends the downscaling approach to the assumption that the relationship between RA and glacier 1m-temperature is dependent on specific humidity, also. In fact, it is not clear that temperature of the 500 hPa level at one single grid point is the best predictor for screen temperature on a glacier surface. On the other hand, taking specific humidity into additional consideration is only one possibility.

12.10.2. Specifications

Multiple regression is applied to the predictors $X_s$ and $X_{c2}$ without $c1$. This is because after the time shift conversion temperature and specific humidity are no longer available at the same points in time and they can not be applied for multiple regression.

The multiple regression models in Fig. 12.14 are defined as follows:

$$\hat{Y}^k_l = \sum_i (\alpha X^i + b)$$ (12.8)

$$\hat{Y}^k_i = \sum_{i \neq j} (\alpha X^i \cdot X^j + \beta X^i + c)$$ (12.9)

$$\hat{Y}^k_p = \sum_i (\alpha (X^i)^2 + \beta X^i + c)$$ (12.10)

$$\hat{Y}^k_q = \sum_{i,j} (\alpha X^i X^j + \beta X^i + c)$$ (12.11)

Here the indices $l$, $i$, $p$, and $q$ are linear, interactive, pure quadratic and quadratic model; $\alpha$ and $\beta$ are vector, $b$ and $c$ scalar regression coefficients. $X^i$ is the $i$-th predictor and $Y^k$ the $k$-th predictand. As predictors temperature and specific humidity (RA 500 hPa level) are applied.

12.10.3. Results

Fig. 12.14 show the results.
12.10.4. Discussion of Fig. 12.14

12.10.4.1. Temperature

Taking specific humidity as additional predictor for EBS temperature increases the correlation by 7% of explained variance. Yet, other than linear multiple regression models do not show further effects on r-square.

However, applying multiple regression models without c2 shows out only a very small improvement (< 2%). r-square obtained by multiple linear regression with predictors standardized by c2 amounts to 0.53.

12.10.4.2. Specific humidity

For specific humidity, multiple regression with standardized predictors shows out only minor improvements compared to single regression. Again, there no significant difference between the multiple regression models.

Yet, the improvement of r-square from single to multiple regression amounts to almost 10% when the unstandardized (no c2) predictor is applied! Nevertheless, the effects of c2 and single linear regression outmatches the multiple regression without c2.
Figure 12.13.: Case study period for temperature and specific humidity with predictor $X$ before and after $c_1$ and $c_2$, $X_c$ and $X_s$, and predictand $Y$. 
Figure 12.14.: r-square statistics for single linear (slinear) and multiple regression with several models (linear, interactive, pure quadratic and quadratic) between X and Y temperature *temp* and specific humidity *shum* before and after c3
13. Daily Means

13.1. Data statistics

Tab.13.1 shows statistical properties of the daily means of temperature and specific humidity, of the predictor X and the predictand Y. The data of the EBS are averaged to daily means only if 24 hourly means per day are available, otherwise set to NaN.

<table>
<thead>
<tr>
<th></th>
<th>temp/[°C]: Y</th>
<th>temp/[°C]: X</th>
<th>shum/[g/kg]: Y</th>
<th>shum/[g/kg]: X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.4</td>
<td>-4.9</td>
<td>5.1</td>
<td>2.2</td>
</tr>
<tr>
<td>Std</td>
<td>1.0</td>
<td>0.8</td>
<td>1.0</td>
<td>0.9</td>
</tr>
<tr>
<td>Min</td>
<td>-2.2</td>
<td>-7.5</td>
<td>1.9</td>
<td>0.1</td>
</tr>
<tr>
<td>Max</td>
<td>3.2</td>
<td>-2.8</td>
<td>6.9</td>
<td>4.9</td>
</tr>
<tr>
<td>95p</td>
<td>2.0</td>
<td>-3.6</td>
<td>6.4</td>
<td>3.6</td>
</tr>
<tr>
<td>75p</td>
<td>1.1</td>
<td>-4.3</td>
<td>5.9</td>
<td>2.8</td>
</tr>
<tr>
<td>Median</td>
<td>0.5</td>
<td>-4.9</td>
<td>5.3</td>
<td>2.3</td>
</tr>
<tr>
<td>25p</td>
<td>-0.2</td>
<td>-5.4</td>
<td>4.5</td>
<td>1.6</td>
</tr>
<tr>
<td>5p</td>
<td>-1.2</td>
<td>-6.4</td>
<td>3.1</td>
<td>0.8</td>
</tr>
<tr>
<td>n</td>
<td>619</td>
<td>795</td>
<td>619</td>
<td>795</td>
</tr>
</tbody>
</table>

Table 13.1.: Statistics of the predictand’s (X) and predictor’s (Y) temperature (temp) and specific humidity (shum) daily means for the baseline period 24 March 2004 to 5 April 2006: mean; standard deviation (std); absolute minimum and maximum values (min and max); 95th, 75th, 50th (median), 25th and 5th percentiles; and n, number of not NaN elements of the sample.

13.1.1. Discussion of Tab. 13.1

13.1.1.1. Comparison to the statistical properties of hourly values (Tab. 12.1)

The standard deviation of hourly values of a two years data sample reflects the diurnal, but also the seasonal variance of a meteorological variable. In the standard deviation of daily means the diurnal variance is filtered out.
Y temperature standard deviation amounts to less than half of standard deviation in the hourly time scale. The difference between the absolute extreme daily means is 5.4 °C, less than half of the range in the hourly time scale (13.7 °C). This is typical for tropical sites, where the diurnal variability of temperature exceeds the seasonal variability (Chap. 4.2).

Specific humidity at the EBS has the same standard deviation in the hourly and monthly time scale; thus the seasonal variance is larger than the diurnal, or has the same size.

13.1.1.2. Comparison between Y and X

The mean daily means of X and Y temperature differ by 5.3 °C, but the standard deviation only by 0.2 °C. This is in contrast to the hourly time scale, where the standard deviation of Y is twice the standard deviation of X. Thus X exhibits variations in the same order of magnitude like Y from day to day, but smaller diurnal variations. This indicates that the predictor X represents meteorological conditions from the upper ABL rather than the surface.

The mean temperature at the EBS over BP is 0.4 °C.

Due to the lower temperatures, also the specific humidity is (in average) 2.9 g/kg lower in the RA (X) than at the EBS (Y), but the variances of X and Y are similar (standard deviations of 1.0 g/kg at the EBS versus 0.9 g/kg of X).

13.1.2. Standardization of the predictor

The standardization of the predictor (equivalent to Eq. 12.4 in the hourly timescale) is done by Eq. 13.1 for daily means:

\[ X_d^s = \frac{\sigma(Y_d^{BP})}{\sigma(X_d^{BP})} \cdot [X^d - X^{BP}] + Y^{BP} \]  

with upper index \( d \) referring to daily mean values of X and Y.

13.1.3. Normality check of daily means

Similar to Fig. 12.1 in Chap. 12, Fig. 13.1 is to show whether daily temperature and specific humidity means of the predictor X and the predictand Y are normally distributed variables.
Figure 13.1.: Probability plot of temperature (a and b) and specific humidity (c and d) X and Y daily means compared to normal distribution (straight lines) before (a and c) and after (b and d) standardization of mean and standard deviation of X to Y (c3).
13.1.3.1. Discussion of Fig. 13.1

Temperature values of Y exhibit only small deviations from a normal distribution. The deviations appear only for the extreme values of Y, that are smaller than -1 °C or larger than 2 °C. This concerns only about 1% of the data. This is less than in the case of hourly instantaneous values in Fig. 12.1.

About 35% of the specific humidity values of Y, though, deviate from the normal distribution. The deviations appear for values of specific humidity smaller then about 3.5 g/kg or larger than 6 g/kg. This is similar to the results in the hourly time scale.

13.2. Correlation analysis for daily means

13.2.1. Correlation plot and linear fit

![Correlation plots](image)

Figure 13.2.: Correlation plot [X Y] with standardized predictors X and numerical values of $r^2$, $a$, $b$, and rmse of the linear fit

In Fig. 13.2.1 large deviations from the linear fit are visible for temperature and...
specific humidity, and low values of r-square confirm the poor correlation between RA and EBS data. Even if r-square is higher (10%) for specific humidity than for temperature daily means, the correlation plot shows many outliers, concerning principally the low values of specific humidity at the EBS (Y). The \textit{rmse} is lower than compared to the hourly timescale, by 1.12 °C for daily temperature means and by 0.22 g/kg for specific humidity means.

\subsection*{13.2.2. Residual analysis}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{residuals.png}
\caption{Residuals course through BP and normal probability plot of residuals}
\end{figure}

Fig. 13.2.2 shows, analogous to Fig. 15.4.2 in the hourly timescale, residuals plotted as a function of time over BP (top; Temp left, shum right), and the normal probability plot of the residuals (bottom; Temp left, shum right). Even more so than in the hourly time scale, a clear seasonal pattern of residuals is visible, for both temperature and specific humidity. However, the normal probability plot shows a good agreement with the Gaussian distribution for both variables. Slight deviations from normality are visible for the highest negative values of residuals.
13.3 The course of temperature and specific humidity daily means throughout BP

13.3.1. Temperature

Fig. 13.3.1 plots the course of daily temperature means of $X_s$ and $Y$ over BP (top), and the difference between $X_s$ and $Y$ (bottom; not equivalent to the residuals time series in absolute values and amplitude).

![Daily temperature means of $X_s$ and $Y$ throughout BP](image1)

![Temperature difference $Y - X_s$](image2)

**Figure 13.4.** Top: daily temperature means throughout BP of $Y$ and $X_s$; bottom: difference between the two curves.

13.3.2. Specific humidity

Fig. 13.3.2 displays the course of specific humidity daily means of $X_s$ and $Y$ throughout BP.
13.3 The course of temperature and specific humidity daily means throughout BP

Figure 13.5.: Top: daily means of specific humidity throughout BP of Y and $X_s$; bottom: difference between the two curves.
13.3.3. Discussion of Figs. 13.3.2 and 13.3.1

On a first view on Figs. 13.3.2 and 13.3.1 (top) indicates that the RA represent the seasonal variations of temperature and specific humidity at the EBS well. In the two years of BP, the dry season with low values of specific humidity can be delimited roughly between July and September. In 2004, however, the dry season is less pronounced (daily means of specific humidity generally above 4 g/kg). High daily means of specific humidity throughout BP occur between November and May. The humidity threshold suggested by Garreaud et al. (2003) for convection activity over the Altiplano, 5 g/kg, corresponds to the initial presumption to distinguish between dry and humid season according to humidity daily means above the glacier surface. This confirms that (above) surface humidity in the Andes is influenced by large-scale dynamics rather than by local meteorology, despite the highly alternating terrain. Thus the findings in Garreaud et al. (2003) and their applicability to the Cordillera Blanca are confirmed here. A comparison between Fig. 13.3.2 and Fig. 13.3.1 shows that low values of specific humidity coincide with low temperatures during the austral summer season (Chap. 4.2).

The differences between predictor and predictand, Figs. 13.3.2 and 13.3.1 (bottom), show a distinct seasonal cycle in the case of temperature. For specific humidity, this is not visible.
14. Monthly Means

14.1. Data statistics

<table>
<thead>
<tr>
<th></th>
<th>temp/[^{C}: Y</th>
<th>temp/[^{C}: X</th>
<th>shum/[^{g/kg}: Y</th>
<th>shum/[^{g/kg}: X</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.4</td>
<td>-4.9</td>
<td>5.0</td>
<td>2.2</td>
</tr>
<tr>
<td>Std</td>
<td>0.6</td>
<td>0.6</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>Min</td>
<td>-0.6</td>
<td>-6.2</td>
<td>3.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Max</td>
<td>1.5</td>
<td>-4.0</td>
<td>6.2</td>
<td>3.4</td>
</tr>
<tr>
<td>n</td>
<td>19</td>
<td>27</td>
<td>19</td>
<td>27</td>
</tr>
</tbody>
</table>

Table 14.1.: Statistics of the predictand’s (X) and predictor’s (Y) temperature (temp) and specific humidity (shum) monthly means for the baseline period 24 March 2004 to 5 April 2006: mean; standard deviation (std); absolute minimum and maximum values (min and max); and n, number of not NaN elements of the sample.

Tab. 14.1 shows statistical properties of monthly means of X and Y, temperature and specific humidity. Compared to hourly and daily means (Tabs. 12.1 and 13.1), the standard deviation of monthly temperature means is small: 0.6 °C versus 1.0 °C (daily means) and 2.1 °C (hourly values). The variability of specific humidity also decreases with the increasing averaging window, as expected. This is because higher frequency fluctuations are filtered out. Yet, the decrease is smaller than for temperature: from the hourly to the monthly time scale the standard deviation of specific humidity decreases by one third (X) and less (Y). The mean temperature over BP at the EBS is 0.4 °C, and the average specific humidity amounts to exactly 5 g/kg. This is slightly different from the mean daily mean values, because monthly means are taken only when at least 20 daily means of the month are available.
14.1 Data statistics

14.1.1. Standardization of the predictor

The standardization of the predictor’s mean and standard deviation (equivalent to Eq. 12.4 in the hourly time scale and Eq. 13.1 in the daily time scale) is calculated for monthly means as follows:

\[
X^m_s = \frac{\sigma(Y^{m,BP})}{\sigma(X^{m,BP})} \left[ X^m - \overline{X^{m,BP}} \right] + \overline{Y^{m,BP}}
\]

(14.1)

with upper index \( m \) referring to monthly means of \( X \) and \( Y \).

14.1.2. Normality check of monthly means

\[\text{Figure 14.1.:} \quad \text{Probability plot of temperature (a and b) and specific humidity (c and d) monthly X and Y means compared to normal distribution (straight lines) before (a and c) and after (b and d) standardization of mean and standard deviation of X to Y (c3).}\]
14.1.3. Discussion of Fig. 14.1.2

In the hourly and daily time scale, temperature can be considered a Gaussian variable, but specific humidity not (Chap. 12). However, in the monthly time scale 30% of $X_s$ temperature means deviate from the normal distribution, whereas specific humidity fits well with normality, as well as temperature.

14.2. Correlation analysis for monthly means

14.2.1. Correlation plot

![Correlation plot](image)

**Figure 14.2.**: Correlation plot $[X \ Y]$ with standardized predictors $X$ and numerical values of $r - square$, $a$, $b$, and $rmse$ of the linear fit

14.2.1.1. Discussion of Fig. 14.2.1

Fig. 14.2.1 shows a clear accordance of the scatter plot with the linear fit. One outlier is visible for temperature, but in general, the deviations are small ($rmse =$
0.38 °C for temperature, and 0.34 g/kg for specific humidity) and equally distributed over the whole ranges of values.

The numerical measures of correlation underline the quantitative results based on the correlation plot: \( r - square \) is clearly higher than in the hourly and daily time scale (this we have already seen in Chap. 11). \( a \) is more close to 1 and \( b \) is more close to 0 than in the hourly and daily time scale.

### 14.2.2. Residual analysis

![Residual Analysis](image)

**Figure 14.3.** Residuals time series (over BP) and normal probability plot of residuals

Despite the high coefficient of determination for both temperature and specific humidity monthly means, Fig. 14.2.2 illustrates the weaknesses of the model: the normality assumption is not true for both temperature and specific humidity residuals; and they feature a systematic pattern with the time (above all the temperature residuals).
14.3. The course of temperature and specific humidity monthly means throughout BP

Fig. 14.3 displays the course of temperature, and specific humidity, respectively, monthly means of $X_s$ and $Y$ throughout BP, as well as differences $Y - X_s$ of both variables.

14.3.1. Description of the seasonal course of temperature and specific humidity over BP (Fig. 14.3)

BP started in March 2004 (the end of the humid season), and covers two dry and two humid seasons until April 2006.

In 2004, the humid season is characterized by lower (monthly averaged) temperatures and higher specific humidities than in the following year. The temperature difference between the two dry seasons is even more pronounced in the RA. A comparison between the temperature and specific humidity course at the EBS ($Y$) shows a time delay between the coldest month (June in both years) and the ‘wettest’ month.
(July in both years). This is not the case for X with the lowest monthly mean temperature in June (2004), but only in September in 2005. The most humid monthly mean in the RA occurs in July (2004) and in August (2005). To sum up, it can be stated that 2004 experiences (in average) a colder, but more humid dry season, both at the EBS and the RA.

14.3.2. Discussion of Fig. 14.3 (left and right)

Fig. 14.3 left (top and bottom) shows the course of differences of temperature $dT$ and specific humidity $dH$ between the two curves (again the differences are not equivalent to the residuals in absolute values and amplitudes). This is to examine whether systematic deviations between X and Y occur, compared to the absolute values of X and Y. In fact, the temperature deviations $dT$ show a pattern similar to the course of specific humidity over BP. Roughly summarized, large negative deviations occur in the dry season and large positive deviations in the humid season. However, the lowest negative deviations between Y and X temperature (May 2004 and June 2005) do not occur contemporarily to the lowest values of specific humidity (July 2004 and 2005, Y; August 2005, Y); rather they are delayed by one to two months. The coefficient of determination between X specific humidity and the temperature difference $Y - X_s$ is 0.33. This value is rather low, possibly because of the discussed time delay between the extremes of differences and specific humidity. Unfortunately, the full period (BP) can not be examined because of gaps in the data series.

14.4. SD model c1 for individual seasons

Seen the seasonality of residuals in all time scales (hourly values, daily means and monthly means; Chap. 12.9.2, 13.2 and 14.2), the following conversion is suggested in order to correct for the 'seasonal' bias between X and $\hat{Y}$:

$$X_s(m, s) = \frac{\sigma(Y(m, s))}{\sigma(X(m, s))} \cdot \left[ X(m, s) - X^{BP}(m, s) \right] + Y^{BP}(m, s) \quad (14.2)$$

This conversion is similar to SD model c1 in the hourly time scale, where the daily cycle of X is standardized to Y in order to assimilate mean and variance of every hour of day. In the hourly time scale, this conversion has shown a significant improvement in terms of correlation between X and Y. However, c1 can not applied
here for monthly means, because BP is too short for representative seasonal statistics (only up to two monthly means per month).
15. Multiple regression analysis in the daily and monthly timescale

15.1. Introduction

The seasonal bias of the SD model, determined by residual analysis at all time scales (Chap. 12.9.2, 13.2 and 14.2), and the similarity between specific humidity monthly means and monthly mean differences between X and Y (Chap. 14.3.2), are related by the fact that the yearly cycle in the outer tropics is given by the seasonality of humidity related variables (Chap. 4.2).

Consequently, I try to fill the gap of missing data for representative seasonal statistics (Chap. 14.4) by relating seasonality to the variable specific humidity. This is achieved by determining specific humidity systematically as the second predictor variable for multiple regression analysis in the monthly time scale.

In the daily time scale, up to now, the same multiple regression analysis is examined only with daily means of specific humidity as additional predictor (which can not be considered representative for determining seasons).

15.2. Multiple linear regression: functional relationship

Eq. 15.2 shows the functional relationship in the case of multiple linear regression:

\[ \hat{Y}_k = \sum_{i=1}^{n} \{ a \cdot X(i, 1) + b \cdot X(i, 2) \} + c \]

\[ \iff \sum_{i=1}^{n} \sum_{j=1}^{2} \{ a + \beta(j) \cdot X(i, j) \} \]

with \( \hat{Y}_k \) being the SD model output for the \( k \)-th predictand \( Y \) (temperature or specific humidity); \( X(i, 1) \) the first predictand (temperature), and \( X(i, 2) \) the
second predictand (specific humidity); both from the RA SE grid point in the 500 hPa level.

15.3. Results

Tab. 15.1 shows the coefficients of multiple linear regression $a$, $b$, and $c$, described in Eq. 15.2; as well as coefficients of determination (r-square) and mean squared error (mse) for temperature and specific humidity daily and monthly means.

<table>
<thead>
<tr>
<th></th>
<th>temp d</th>
<th>shum d</th>
<th>temp m</th>
<th>shum m</th>
</tr>
</thead>
<tbody>
<tr>
<td>rsquare</td>
<td>0.47</td>
<td>0.48</td>
<td>0.89</td>
<td>0.83</td>
</tr>
<tr>
<td>$a$</td>
<td>-1.41</td>
<td>1.90</td>
<td>-1.67</td>
<td>1.17</td>
</tr>
<tr>
<td>$b$</td>
<td>0.55</td>
<td>0.24</td>
<td>0.62</td>
<td>0.20</td>
</tr>
<tr>
<td>$c$</td>
<td>0.34</td>
<td>0.62</td>
<td>0.39</td>
<td>0.78</td>
</tr>
<tr>
<td>mse</td>
<td>0.49</td>
<td>0.55</td>
<td>0.05</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 15.1.: Coefficient of determination (r-square), regression coefficients (intercept $a$, and $b$ and $c$, the slopes related to temperature and specific humidity), and mean root square error for daily (d) and monthly (m) means of temperature (temp) and specific humidity (shum).

Fig. 15.3 shows the results of multiple linear and non linear regression (the same models like in the hourly time scale, Chap. 12.10, are applied) in the daily and monthly time scale.

15.3.1. Discussion of the results

15.3.1.1. Multiple regression in the daily timescale

Applying multiple regression instead of simple linear regression increases r-square by more than 10% of explained variance for the temperature predictor, but the improvement for specific humidity is minor (+ 4% of explained variance). Moreover, in contrast to the daily and monthly timescale, there is a small but visible difference for specific humidity, when quadratic regression models are applied instead of linear or interactive models.

15.3.1.2. Multiple regression in the monthly timescale

Applying multiple regression instead of simple linear regression increases r-square by 10% of explained variance for the temperature prediction (as expected after the
15.3 Results

Figure 15.1.: Coefficient of determination (r-square statistics) between temperature and specific humidity daily (top) and monthly (bottom) means applying different models of single and multiple regression.
15.4 Residual analysis of the multiple linear regression model for daily and monthly means

discussion in Chap. 14.3.2). However, the improvement for the prediction of specific humidity is minor (+ 4% variance of explained variance). Moreover, similar to the outcomes of multiple regression analysis in the hourly time scale (Chap. 12.10), no further improvement is visible when other than linear models are taken into consideration. In fact, for different models the coefficient of determination varies only in the order of 1%.

15.3.2. Summary

To summarize, in the daily time scale, regression coefficients slightly below 0.5 result from multiple linear regression, for both temperature and specific humidity predictands. This is slightly below the value which could be achieved for temperature hourly predictors via multiple linear regression and the SD conversions c1 and c2 (0.53). In the monthly time scale, r-square is 0.89 for temperature and 0.83 for specific humidity. The residual analysis in the next section examines the aptitude of the multiple regression model in more detail. In the case of for monthly means, the residual analysis is to examine if the seasonality of residuals can be corrected, when specific humidity from the RA is applied as additional predictor.

15.4. Residual analysis of the multiple linear regression model for daily and monthly means

15.4.1. Temperature

The normal probability plots show that temperature can be approximated as normal variable for both daily and monthly means. However, the time series of residuals still displays a systematic pattern throughout BP. Thus, even if the multiple regression model increases the explained variance of predictands by the model, it does not correct for the seasonal bias.

15.4.2. Specific humidity

More than 10% of specific humidity residuals in the daily time scale deviate from normality. In the monthly time scale, almost 50% of residuals can not be considered Gaussian. A seasonal oscillation of residuals is visible for daily means, but also for monthly means.
15.4 Residual analysis of the multiple linear regression model for daily and monthly means

**Figure 15.2.** Multiple regression goodness of fit evaluation: residuals throughout BP (left) for daily (top) and monthly (bottom) means of temperature and probability plot of residuals compared to normal variables (right).
15.4 Residual analysis of the multiple linear regression model for daily and monthly means

![Graphs showing residual analysis for daily and monthly means of specific humidity](image)

**Figure 15.3.:** Multiple regression goodness of fit evaluation: residuals throughout BP (left) for daily (top) and monthly (bottom) means of specific humidity and probability plot of residuals compared to normal variables (right).
16. Summary, results and outlook

16.1. Summary

The main steps of the SD approach in this master thesis have been:

- the selection of predictor variables (grid points) from the RA by single linear regression analysis in the hourly, daily, and monthly time scale;

- statistical analysis of the predictand and the selected predictor for hourly (six-hourly) values, daily, and monthly means;

- the standardization of the predictor to the predictand in each time scale;

- a comparison between the diurnal cycles of X and Y, based on the example of two case studies from the dry and humid season;

- analysis of the mean daily temperature and specific humidity cycles;

- analysis of daily and monthly means of X and Y throughout BP;

- definition and application of SD model c1: time shift adjustment applied to temperature data (hourly values);

- definition and application of SD model c2: standardization of the predictor’s daily cycle to the predictand’s;

- the application of non linear and multiple regression models with temperature and specific humidity from the selected grid point for both predictands in the hourly time scale, for daily and monthly means;

- and the validation of the SD model in all time scales by correlation and residual analysis.
16.2 Results

The outcomes of the SD approach in this master thesis are summarized below:

- Linear regression analysis of untransformed data shows that correlations (values of r-square) are lower than 0.5 for hourly values and daily means, but higher than 0.5 for monthly means. In all examined cases the correlation for specific humidity (X and Y) is higher than for temperature.

- Reanalysis data from SE grid point (75°W 10°S) of the 500 hPa level shows the highest correlation to the EBS data, for both temperature and specific humidity in all time scales. This is in good agreement with previous studies (Georges, 2005).

- Whereas temperature from the RA and at the EBS can be considered normally distributed, for specific humidity, deviations from normality can be neglected only in the monthly time scale.

- By standardizing the predictor to the predictand individually for every hour of the day (SD model c1), the correlation between X and Y can be significantly increased for both variables temperature and specific humidity.

- The analysis of the diurnal temperature cycle indicates a time delay between X and Y. By time shifting the X sample by +4 hours (SD model c2), the correlation improves from 19% to 38% of explained variance. After this conversion no time delay is visible any longer in the mean daily cycles, as well as in the case study period. A time shift of 11 or 18 hours decreases the correlation to almost 0.

- Correlation and residual analysis shows that the SD models c1 and c2 can be successfully applied to the variable temperature, but not to specific humidity. Even if specific humidity has shown higher correlations in the beginning, c1 and c2 can not increase the correlation significantly, and the correlation plots indicate poor evidence of a linear relationship between X and Y in the hourly and daily time scale.

- Applying multiple regression to temperature and specific humidity from the selected grid point as predictors can increase the correlations, in all time scales, and for both predictands temperature and specific humidity.
• For both predictands temperature and specific humidity, in all time scales, residuals feature a seasonal pattern. This indicates that the SD models do not cover seasonal fluctuations. Thus it is recommended to standardize the seasonal cycle of the predictor to the predictand, equivalent to SD model c1 for the daily cycle.

• By applying other than linear models for SD, the results can not further be improved.

16.3. Outlook

In order to stay within the time schedule of this master thesis and due to the restricted amount of available data (observations), the following important steps have not been included here:

• finding and applying an appropriate SD technique for the non Gaussian variable specific humidity;

• the standardization of the RA to the yearly cycle of temperature and specific humidity at the EBS;

• and the validation of the results using independent data.

The next step after the SD approach in order to reanalyze local meteorology on the glacier surface of Artesonraju is to take the same approach using the dynamical downscaling technique.
Curriculum Vitae

Marlis Hofer
St. Stefan 98
I - 39040 Villanders

Born on 17 March 1982 in Brixen, Italy

Education:

2006-2007 Diploma thesis under the guidance of Dr. Kaser, Department of Geography, Innsbruck, and Dr. Wagnon, Laboratoire de Glaciologie et Géophysique de l’Environnement, Grenoble: "Statistical downscaling of NCEP/NCAR reanalysis data to air temperature and specific humidity on an outer tropical glacier surface". Erasmus, UJF Grenoble.

2004 - 2007 Diploma study at the University of Innsbruck. Master of Natural Science (Magister rerum naturalium) in Meteorology.

2001-2004 Diploma study of Physics, University of Innsbruck.


Participation in field experiments: Field work in the Cordillera Blanca, Peru, February 2007.


J.-W. Kim et al. The statistical problem of climate inversion: Determination of
the relationship between local and large-scale climate. *Monthly Weather Review*,
1984.

R. Kistler et al. The NCEP-NCAR 50-year reanalysis: Monthly means cd-rom and

M. Kuhn. Climate and glaciers. In *Sea level, ice and climatic change*. International
association of hydrological sciences, 1981.

M. Kuhn. Micro - meteorological conditions for snow melt. *Journal of glaciology*,
1986.

M. Kuhn. The response of the equilibrium line altitude to climatic fluctuations:
theory and observations. In *Oerlemans, Glacier fluctuations and climatic change*.

L. Leung, O. Mearns, F. Giorgi, and R. Wilby. Regional climate research: needs


G. R. McGregor and S. Nieuwolt. *Tropical Climatology*. John Wiley and Sons Ltd.,
1998.


J. Murphy. An evaluation of statistical and dynamical techniques for downscaling


Oldenbourg Verlag München Wien, 1996.


