

Statistical downscaling in the arid central Andes: uncertainty analysis of multi-model simulated temperature and precipitation

Maxime Souvignet · Jürgen Heinrich

Received: 15 June 2010 / Accepted: 3 March 2011 / Published online: 6 April 2011
© Springer-Verlag 2011

Abstract Statistical downscaling is a technique widely used to overcome the spatial resolution problem of General Circulation Models (GCMs). Nevertheless, the evaluation of uncertainties linked with downscaled temperature and precipitation variables is essential to climate impact studies. This paper shows the potential of a statistical downscaling technique (in this case SDSM) using predictors from three different GCMs (GCGM3, GFDL and MRI) over a highly heterogeneous area in the central Andes. Biases in median and variance are estimated for downscaled temperature and precipitation using robust statistical tests, respectively Mann–Whitney and Brown–Forsythe's tests. In addition, the ability of the downscaled variables to reproduce extreme events is tested using a frequency analysis. Results show that uncertainties in downscaled precipitations are high and that simulated precipitation variables failed to reproduce extreme events accurately. Nevertheless, a greater confidence remains in downscaled temperatures variables for the area. GCMs performed differently for temperature and precipitation as well as for the different test. In general, this study shows that statistical downscaling is able to simulate with accuracy temperature variables. More inhomogeneities are detected for precipitation variables. This first attempt to test uncertainties of statistical downscaling

techniques in the heterogeneous arid central Andes contributes therefore to an improvement of the quality of predictions of climate impact studies in this area.

1 Introduction

General Circulation Models (GCMs) are the principal instrument for making projections of future climatic conditions. They are designed to propose a large-scale vision of the evolution of climate in responses to natural and anthropogenic forcing. Because of the current limitations in computing resources, their spatial resolution is still limited. As a consequence, raw GCMs' outputs do not meet the needs of regional impact studies (Wilby et al. 2004). Another issue is that different GCMs return different climate projections for both the present day and future periods. As a consequence, using outputs from a range of GCMs is recommended for impacts studies (IPCC 2007). Multi-model approach has been therefore developed recently at the global scale using a Bayesian approach (Lopez et al. 2006), at the regional scale with a linear Bayesian linear model (Greene et al. 2006) or using more classic weighting methods (Giorgi and Mearns 2002).

Another issue is that projections in complex topographic areas, such as mountainous regions, cannot be rendered with enough accuracy (Hostetler 1994; IPCC 2007). Regional climate is also often affected by circulations that occur at smaller scales (e.g. Giorgi and Mearns 1991). Accordingly, recognising these needs for small scale projections, the scientific community developed during the last decade several so-called downscaling techniques. Among these, statistical downscaling is appreciated for its advantages of being simple and easy to implement when compared to other techniques (Wilby et al. 2004). Another

M. Souvignet · J. Heinrich
Institute of Geography, University of Leipzig,
Johannisallee 19a,
04103 Leipzig, Germany

M. Souvignet (✉)
Institute for Technology and Resources Management
in the Tropics and Subtropics (ITT),
Cologne University of Applied Sciences,
Betzdorfer Straße 2,
50679 Cologne, Germany
e-mail: maxime.souvignet@fh-koeln.de

advantage lies in the fact that spatial and temporal variability are conserved by involving historical records for the determination of the model's parameters. The Statistical Downscaling Model (SDSM) developed by Wilby et al. (2002) is a widely used linear regression model, especially in the field of the analysis of climate change impacts on water resources.

The interaction between the atmospheric circulation and the complex topography of the Andes mountain range has a strong influence on the regional climate. It makes also this region especially vulnerable to climate variability. Such mountainous arid regions, in their majority characterized by nival regimes, are highly dependent on precipitations but are as well strongly influenced by temperatures (Singh et al. 2006). If downscaling techniques, including statistical downscaling, have been mainly tested and applied in Northern America and Europe (e.g. Dibike et al. 2008; Diaz-Nieto and Wilby 2005; Boe et al. 2007), fewer contributions exist in Southern America. Solman and Nunez (1999) reported about one of the first attempts of statistical downscaling of temperature in central Argentina. In their conclusions, the authors estimated that if the interannual variability is not always well captured, the statistical downscaled data show a good agreement with observations. In a more recent work, Labraga (2010) proposed recently one of the few applications of a statistical downscaling model in western Argentina. The author concludes that significant useful relationships between precipitation and atmospheric patterns could be found. However, the scarcity of observational stations on the spatial and temporal levels adds up to the challenges of downscaling in mountainous areas (Labraga 2010). Recently, a preliminary investigation carried out in the arid Northern Chile by Souvignet et al. (2010) proposed a successful application of statistical downscaling for the projection of future precipitation and temperature scenarios. However, the authors concluded that uncertainties linked with the technique should be addressed.

In southern central Chile, Rojas (2006) first reported in a recent benchmark study about the application of downscaling. Even if dynamical and not statistical downscaling was involved, her conclusions indicate that biases were highly correlated with the station's elevation and the terrain representation in her model. More recently, Marengo et al. (2009) used the PRECIS regional climate modelling system to investigate the distribution of extreme temperature and precipitation in South America over the recent past as well as under future climate forcing. In their analysis, the authors note significant biases in the central Andes between observed and simulated values.

Uncertainties in downscaling techniques are manifold. A part of it could be associated with those intrinsic to the GCMs (Mearns et al. 2001). In addition, the downscaling

model structure contributes to the overall downscaling uncertainty. Nevertheless, several solutions exist in order to address the issue of uncertainty with statistical downscaling:

First, the use of multiple climate models is expected to reduce the uncertainty linked with GCMs (e.g. Doblaser-Reyes et al. 2003). It provides a measure of uncertainties as well as a better mean estimate, which is expected to mitigate possible biases of the different GCMs. Therefore, a multi-model approach should be used while implementing statistical downscaling techniques for impact studies. In the case of SDSM, climate model predictors are made available courtesy of the Canadian Institute for Climate Studies. However, predictors for South America and most of the world are available from only one GCM. As a consequence, most impact studies using SDSM rely only on the outcomes of one sole GCM (e.g. Chu et al. 2010).

Second, the quantification of the uncertainty linked with the downscaling model structure is made possible by several statistical methods. The most commonly used statistical techniques for assessing model uncertainty include the analysis of statistical properties of the model errors.

Hence, this paper aims at providing a comprehensive comparison of uncertainties linked with different GCMs using the SDSM in an arid mountainous watershed of the central Andes. The uncertainty assessment will include the deviations in terms of median and variance between downscaled variables and their corresponding observations using boundary data of three GCMs, which were never tested with SDSM in this region. In addition, and given the regional high climatic variability, the ability of SDSM to reproduce extreme events will be tested. Hence, with the quantification of biases linked with three GCMs, to the best of our knowledge never tested for SDSM in the central Andes, this work is expected to contribute to a reduction of uncertainties linked with impact studies in the region and in other similar areas in the world. In the case of the central Andes, it is expected that a better knowledge of uncertainties at stake will enhance the decision capacity of national and local water managers as well as provide decision makers with sound estimates of potential impacts of climate change in the region.

This paper is structured in four main sections. First, the study area and the data based used for this study will be presented. The next section will describe the methodology used in this work. After a description of the linear regression downscaling technique, the statistical tests and analysis used in the uncertainty analysis will be presented. Thereafter, results for the performance and associated uncertainties of downscaled variables will be explored. Eventually, general concluding remarks are presented in the last section.

2 Study area and database

2.1 Study area

This work focuses on the span between 29°20' and 32°17' S, the so-called Coquimbo Region (see Fig. 1). Located just south of the arid diagonal (Messerli et al. 1992), its climate ranges from arid to semi-arid from the north to the south (Miller 1976). The main reasons for the choice of the Coquimbo Region for an uncertainty analysis are described below.

A strong future climate change signal in the western side of the Andes is suggested by most climate models (Mata and Campos 2001). This makes this region a meaningful candidate for impact studies in arid mountainous regions (e.g. investigating the influence of climate variability on water resources).

As other arid and semi-arid areas, which cover more than 30% of the inland earth's land surface (Dregne 1985), the region has climatic, physiographic and ecological features of high vulnerability to climate variations (Downing et al. 1994; Holmgren et al. 2006). The annual precipitation shows a strong orographic dependence, going from ca. 80 to 300 mm a⁻¹ from the coastal area to the Andes (Favier et al. 2009). An increase in precipitation from north to south is also observed in the same amounts.

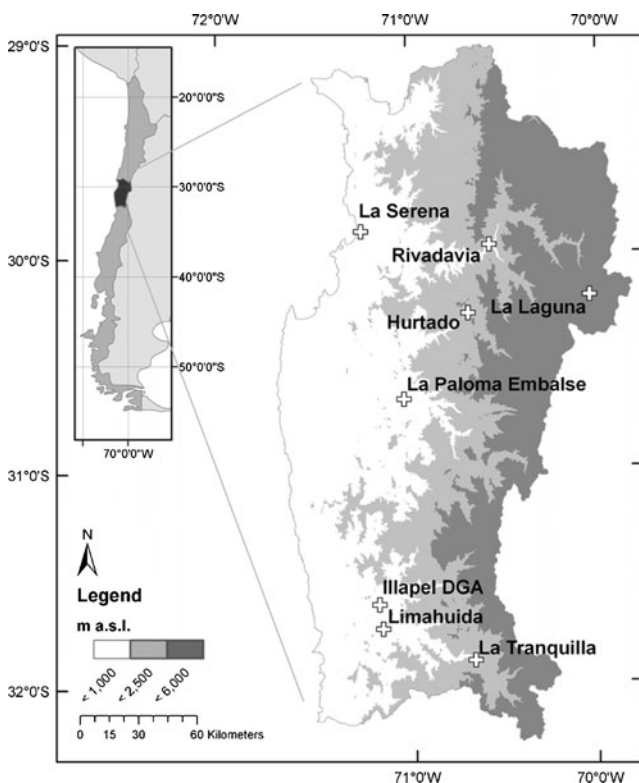


Fig. 1 Location of the meteorological stations within the study area

Moreover, precipitation is strongly influenced by several natural large-scale phenomena such as El Niño Southern Oscillation (ENSO) and the Pacific Decadal Oscillation (PDO) (Aceituno 1988; Garreaud et al. 2009; Verbist et al. 2010). Temperature, also influenced by these phenomena, reaches its minimum in winter (June–August), which coincides with precipitation maximum.

In addition, hydrological processes are expected to be strongly affected. Snowfall covers large areas, up to 50% of total surface in certain watersheds at high altitude (Favier et al. 2009). With a relatively low glacier coverage (ca. 7 km² between 29–32° S; see Garin 1987), river discharge is then mainly driven by the melting process of winter-accumulated snowpack. Therefore, a better understanding of uncertainties linked with downscaled temperature and precipitation for future scenarios will give more confidence in the outcomes of future impacts studies.

Last but not the least, this region, as other areas in developing countries, has been given very little attention with regard to climate change. Most downscaling technique comparative studies and impact assessment investigations concentrate on areas within Western Europe and Northern America (Kundzewicz et al. 2008).

2.2 Predictands

The database used in this study is threefold. It includes (1) a set of observed daily data records, (2) daily variables based on the National Center for Environmental Prediction (NCEP) re-analysis data set and (3) daily variables based on three GCMs data set.

The predictands, i.e. observed data, were provided by the Dirección General de Agua (DGA) for height stations and are presented in Table 1. Daily maximum temperatures (referred to as Tmax), daily minimum temperature (referred to as Tmin), and daily precipitation for a 22-year period (1979–2000) were considered. Because of data relative scarcity and quality constraints, a longer time period was not available.

Data quality assessment is an important preliminary step in uncertainty analysis. Because the failure of the input data to meet certain quality standards enhances the biases introduced into the simulation, a special emphasis was given to data quality control. The homogenous data series used in this study were provided by Souvignet et al. (2011) for a trend analysis of temperature and precipitation records in central–northern Chile.

2.3 Predictors

In statistical downscaling, the relevance of the relationship between large-scale predictors (i.e. variables from NCEP re-analysis and GCMs datasets) and the small

scale predictands (i.e. precipitations, Tmax, Tmin at the stations' scale) will determine the model ability to produce good climate projections for the research area. This is based on the assumption that the predictor–predictand relationships under the current conditions remain valid under future climate conditions. This assumption allows the statistical downscaling of global future climate projections. Therefore, the choice of predictors is of particular importance to linear regression downscaling techniques.

The NCEP re-analysis dataset (Kalnay et al. 1996; Kistler et al. 2001) was re-gridded to the same coordinate system as the selected GCMs (listed in Table 2) and normalized with respect to their respective 1961–1990 mean. This modified dataset will be referred to here as NCEP.

When the best fit between predictands and predictors is validated for observed climate, a simulated control signal is introduced by substituting observed large-scale predictors from the NCEP dataset with the corresponding simulated GCMs predictors.

GCMs predictors were obtained from the World Climate Research Programme Coupled Model Inter-comparison Project Phase 3 (CMIP3) dataset for the climate of the twentieth century experiment (20C3M) scenario. This scenario corresponds to a change in greenhouse gases as observed until 2000. Thereafter, variables were standardized according to their 1961–1990 variance and mean values. The modified datasets for GCMs predictor variables will be referred to as to using GCMs short names described in Table 2.

Among all predictors available from the CMIP3 database, only 16 were available at the daily resolution for NCEP and the three GCMs considered in this study. Subsequently, re-gridding and standardization of data of NCEP and the GCMs were processed using the PostScript-

based language of the International Research Institute for Climate and Society (IRI) Data Library (<http://iridl.ldeo.columbia.edu/SOURCES/.WCRP/.CMIP3/>). The 16 predictors considered for the calibration exercise are displayed in Table 3.

3 Methodology

3.1 Statistical downscaling technique

The Statistical Downscaling Model was developed by Wilby et al. (2002). As Wilby et al. (2002) already propose a complete and detailed description of SDSM, the model will be depicted hereafter only briefly. Additional details could be found in the literature cited in this section. SDSM is designed to produce high-resolution daily climate information from coarse-resolution GCM simulations. It is best described as a hybrid of the stochastic weather generator and multiple linear regression methods. The underlying idea of the model is that large-scale circulation patterns and atmospheric moisture variables (e.g. observed measures of vorticity or relative humidity) are used to condition local-scale weather generator parameters (e.g. precipitation occurrence and intensity; Penlap et al. 2004). It is then assumed that this relationship remains constant under climatic change. This allows the generation of future local climate scenarios.

Among statistical downscaling techniques, SDSM has become increasingly used, mostly because it proposes a set of pre-processed predictors (NCEP and GCCMs) available for most regions of the world. Although this model has been used in several countries, uncertainties linked with its downscaled variables appear to be highly dependent on the local climate regimes.

Lately, Chu et al. (2010) tested SDSM in semi-arid China. Their results suggest that temperature data are well simulated, whereas downscaled precipitation introduced systematic errors in extreme events.

In a recent study, Dibike et al. (2008) investigated the uncertainties linked with SDSM downscaled precipitation and temperature regimes in Northern Canada. Their approach is based on the systematic analysis of uncertainties related to hypothesis test of median and variance along with their corresponding confidence intervals. The authors concluded that biases introduced by the sole SDSM technique were acceptable for impact studies in the region. In central Sweden, Wetterhall et al. (2007) investigated four statistical downscaling methods in terms of their ability to capture statistical properties of daily precipitation in different seasons. Their conclusions showed that although SDSM outperformed the other analogue models, it was unable to capture well differences between wet and dry

Table 1 Location of meteorological stations used in the uncertainty analysis

	Altitude (m a.s.l.)	Lat (S)	Lon (W)	Gauge
La Serena	15	29°54	71°15	P
Illapel DGA	290	31°38	71°11	T
Limahuida	295	31°45	71°10	P
La Paloma Embalse	320	30°41	71°02	P, T
Rivadavia	820	29°58	70°34	P, T
La Tranquilla	975	31°54	70°40	P, T
Hurtado	1,100	30°17	70°41	P, T
La Laguna	3,160	30°12	70°02	P, T

P precipitation, *T* temperature

Table 2 GCMs used for uncertainty analysis

Official name	Short name	Institution	Country	Resolution (Lon×Lat)	Grid box centroid		Reference
					Lon (W)	Lat (S)	
CGCM3_1	CGCM3	Canadian Centre for Climate Modelling and Analysis	Canada	3.75°×3.75°	71.25°	31.54°	(Flato and Boer 2001)
GFDL_CM2_1	GFDL	NOAA/Geophysical Fluid Dynamics Laboratory	USA	2.5 °×2°	71.25°	31.35°	(Delworth et al. 2006)
MRI_CGCM2	MRI	Meteorological Research Institute	Japan	2.81°×2.81°	70.31° 70.31°	29.30° 32.09°	(Yukimoto et al. 2001)

Coordinates selected for the study area are displayed along with institution, country and reference. In this study, GCMs will be referred to as using their short names

summers. Khan et al. (2006) compared three different downscaling techniques in northern Canada using a similar approach later used by Dibike et al. (2008). They concluded that SDSM was one of the best performers in simulating temperatures and precipitation in this region. In addition, Wetterhall et al. (2006) investigated the ability of four statistical downscaling methods to simulated rainfall in southern eastern and central China. The authors conclude that the annual cycle was well captured with SDSM.

3.2 Screening predictor variables

In order to define the statistical downscaling model parameters, a screening of available predictors is necessary. To ensure a proper calibration, NCEP predictor variable (re-gridded for the different GCMs) were selected according to their robust statistical significance, i.e. significant partial correlation (*r*) at the 95% confidence level with the local predictand (i.e. precipitation and temperature). This statistical measure helps in identifying the amount of explanatory power of each predictor once the influence of other predictors have been removed.

In order to allow a comparison among the three GCMs, the whole set of predictors listed in Table 3 is tested for each GCMs against observed values. Predictors with non-significant correlations are discarded until a set of predictors with the highest partial correlation is chosen (from

two up to six, depending on the predictand). This procedure, although compromising between the three grid box outputs, allows a sound comparison between the different GCMs, attributing them the same set of statistically significant predictand–predictors relationships. In addition, based on the data statistical characteristics, a lag 1-day auto-regression was introduced in the model for temperature simulation. The outcome of the predictor screening for precipitation, maximum and minimum temperature are displayed in Table 4.

Predictor variables such as sea level pressure (mslp) as well as specific humidity at different level pressure (sh85 and sh50) return indeed high partial correlation values for precipitation. Nevertheless, at high altitude (e.g. La Laguna), mslp does not play a prevailing role anymore, ceding to a greater influence of winds (p5_u). As shown by Kalthoff et al. (2002), westerly winds prevail at high altitude (4,000 m.a.s.l.) while below this height, winds flow southward along the Andes. These winds appears to have an influence on both precipitation and temperature with meridional and zonal velocity at appropriate pressure levels (surface, 850 hPa, 500 hPa) returning relative high partial correlations. However, few predictor variables, with the above-mentioned regional physical characteristics, returned satisfactory partial correlation values for precipitation. In contrast, a larger set of physically based predictors were found correlated with Tmin and Tmax.

Table 3 Available predictor variables considered for NCEP, CGCM3, GFDL and MRI models

Predictor	Code	Predictor	Code
Mean sea level pressure	mslp	500 hPa meridional velocity	p5_v
Latent heat flux	lhflx	850 hPa zonal velocity	p8_u
Sensible heat flux	shflx	850 hPa meridional velocity	p8_v
Near surface specific/relative humidity	shum	1,000 hPa zonal velocity	10_u
Mean temperature at 2 m	temp	1,000 hPa meridional velocity	10_v
Surface zonal velocity	p_u	500 hPa specific/relative humidity	sh50
Surface meridional velocity	p_v	850 hPa specific/relative humidity	sh85
500 hPa zonal velocity	p5_u	1,000 hPa specific/relative humidity	sh10

Table 4 Selected NCEP predictor variables and their respective partial correlation for precipitation, Tmax and Tmin corresponding to the investigated GCMs

Precipitation	Maximum temperature				Minimum temperature							
	Re-gridded NCEP predictors				Re-gridded NCEP predictors							
	Partial r^*		Partial r^*		Partial r^*		Partial r^*					
	CGCM3	GFDL	MRI	CGCM3	GFDL	MRI	CGCM3	GFDL	MRI			
La Serena	mslp	-0.47	-0.316	-0.181	p_u	-0.154	-0.148	-0.227	p_u	0.17	0.171	0.162
	temp	-0.316	-0.227	-0.466	sh10	0.052	0.066	0.051	mslp	-0.095	-0.099	-0.255
Limahuida	p8_u	-0.237	-0.475	-0.306	shflx	0.282	0.268	0.465	sh50	0.083	0.083	0.188
	mslp	-0.477	-0.288	-0.52	temp	0.128	0.133	0.108	sh85	0.192	0.193	0.242
	sh85	-0.208	-0.176	-0.185	p5_v	0.225	0.22	0.158	temp	0.347	0.345	0.236
La Paloma Emb.	p8_u	-0.185	-0.5	-0.218	p8_v	0.06	0.073	0.305	mslp	-0.14	-0.182	-0.187
	mslp	-0.504	-0.174	-0.34	p_v	0.143	0.113	0.101	sh85	0.158	0.148	0.12
Rivadavia	sh85	-0.186	-0.527	-0.254	shflx	0.137	0.277	0.103	temp	0.183	0.206	0.258
	mslp	-0.54	-0.187	-0.533	temp	0.308	0.165	0.267	p8_u	-0.075	-0.08	-0.098
	sh85	0.226	0.202	0.21	p8_v	0.16	0.162	0.153	p_v	0.088	0.087	-0.054
La Tranquilla	shflx	0.202	0.139	0.367	p5_v	0.288	0.265	0.256	sh50	0.168	0.164	-0.048
	mslp	-0.176	-0.232	-0.203	p_v	0.086	0.095	0.102	temp	0.345	0.347	0.362
	p5_u	0.236	0.557	0.182	shflx	0.07	0.055	0.048	p5_v	0.179	0.18	0.227
Hurtado	mslp	-0.477	-0.48	-0.381	temp	0.184	0.188	0.169	p_v	-0.073	-0.075	-0.213
	sh50	0.24	0.245	0.236	p8_u	-0.049	-0.043	-0.104	sh50	0.192	0.185	0.197
	sh85	-0.277	-0.263	-0.242	p5_v	0.348	0.345	0.329	sh85	-0.117	-0.115	0.211
La Laguna	lhflx	-0.277	-0.45	-0.292	p_v	0.164	0.066	0.062	temp	0.393	0.39	0.139
	sh50	0.215	0.184	0.237	lhflx	0.046	0.062	0.075	p5_u	-0.127	-0.137	-0.117
	p5_v	0.353	0.282	0.319	sh50	-0.198	-0.035	-0.105	p8_v	0.076	0.075	0.128
				temp	0.379	0.271	0.256	p_v	0.072	0.071	0.061	
				p5_v	0.353	0.319	0.329	mslp	-0.061	-0.065	-0.045	
				p5_v	0.261	0.199	0.08	sh50	0.147	0.139	0.143	
				sh85	0.195	0.167	0.162	sh85	-0.123	-0.123	-0.11	
				shflx	0.108	0.23	0.051	temp	0.363	0.361	0.307	
				temp	0.285	0.319	0.274	mslp	-0.116	-0.149	-0.046	
				p8_v	0.348	0.46	0.316	sh50	-0.176	-0.202	-0.161	
				mslp	-0.043	-0.119	-0.155	temp	0.462	0.542	0.499	
				sh50	-0.232	-0.209	-0.13	p5_u	0.033	0.039	0.034	
				shflx	0.054	0.108	0.109					
				temp	0.385	0.442	0.414					
				p5_u	-0.089	-0.092	-0.082					

* $p < 0.05$

In conclusion, among the predictors considered for SDSM calibration in Table 3, sea level pressure plays a major role in the prediction of precipitation with high partial correlations. If this predictor is also significant for temperature (especially for Tmin), its influence is lower in the region. Meridional and zonal velocity appears to influence both precipitation and temperature in the region, with graduated pressure levels correlated with the station's altitude. Equally, relative humidity predictors at higher pressure levels return significant partial correlations for both precipitation and temperature. On the other hand, surface relative humidity and high pressure winds (10_u and 10_v) do not prove to be predictors appropriate for the region.

A 16-year calibration period (1979–1994) was chosen for temperature. The statistical model's parameters were then validated during a 6-year period (1995–2000). Because rainy days are very scarce in the region, a longer validation period was necessary for precipitation. Therefore, a 10-year (1979–1988) period for calibration was chosen, along with a 12-year (1989–2000) validation period for precipitation. As a result of these data availability constraints, the calibration period is relatively short for both precipitation and temperature variables. However, as the region's climate is strongly affected by phenomena such as ENSO and the PDO, it should be verified that the calibration phase is representative of the regional climate variability. Figure 2 compares the calibration and validation period for precipitation and temperature against the Oceanic Nino Index (ONI) from 1950 to 2000. The PDO phenomenon is not represented on this graph because, as shown by Biondi et al. (2001), after 1977, the PDO entered a so-called warm phase which is homogenous over the calibration and validation period presented here. Nevertheless, it should be noted that the “cold phase” of the PDO is consequently not taken into consideration. Concerning the

ENSO phenomena, the ONI is represented by a dark grey shaded area from 1950 to 2000. Standardized regional averaged annual precipitations (for all stations considered in this study) are represented by a black line. The dashed line shows standardized regional average mean temperature for the region of interest. For both temperature and precipitation, medium and extreme warm/cold phases of the ENSO phenomenon are included within the calibration period. Hence, although the calibration period is relatively short, the regional climate variability is well captured by the calibration period.

3.3 Uncertainty analysis

The conservation of the mean value and variability of observed events, for a baseline period, into the simulation of future events is necessary in order to enable a reasonable confidence into the statistical model outputs. Therefore, the uncertainty analysis aims at quantifying the model's ability to reproduce the current state of precipitation and temperature. This is a prerequisite for the simulation of future climate based on the outputs of GCMs.

The methodology to identify uncertainties of downscaled variables was first proposed by Khan et al. (2006). Given the regional climate characteristics of the central Andes, a modified approach is used for the region and an analysis of biases linked with the reproduction of extreme events was introduced.

Hence, the uncertainty linked with the statistical downscaling of daily temperatures (Tmax and Tmin) and precipitation is determined in terms of (1) model's biases in the estimates of median, (2) model's biases in the estimates of variance and (3) model's ability in the simulation of extreme events.

The uncertainty analysis is carried out in three steps. The first step describes the analysis of the data basic statistical

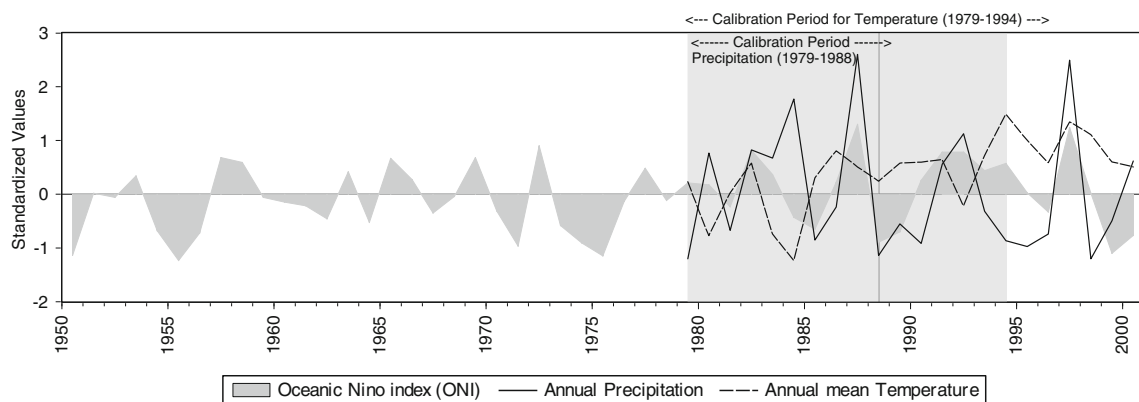


Fig. 2 Long-term regional climate variability against the calibration and validation periods. The grey shaded area represents the ONI, black line displays standardized average precipitation and the dashed

line represents the standardized mean temperature. The calibration phases for temperature are indicated by a light grey area

characteristics. This step, commonly called the exploratory data analysis (EDA) is recommended while handling statistical tests in natural sciences (Helsel and Hirsch 2002). A hypothesis test requires the population's distribution to be characterized by certain parameters. For example, many tests rely on the assumption that the population follows a normal distribution, is outlier free and is not auto-correlated when values were collected at regular time intervals. Non-parametric tests, for instance, do not make these assumptions, so they are useful for data not meeting the above-mentioned characteristics. The EDA helps in determining the statistical characteristics of the data. Therefore, it is useful in order to identify what tests are best suited. The outcomes of the EDA, not detailed here, showed that temperature data (Tmax and Tmin) return a normal distribution, with few outliers, whereas precipitation data is heavily right skewed (i.e. returns a strong asymmetrical probability distribution, with the mass of the distribution concentrated on the left tail of the figure), with many outliers (as this is an arid area, outliers are in fact records of precipitation, most records being zeros). In addition, the EDA shows the presence of autocorrelation for temperature, which disappears after a few lags. However, no evidence of autocorrelation was found for precipitation records. Hence, the outcomes of the EDA suggest that specific statistical tests should be used.

In a second step, hypothesis tests of median and variance explored the model's ability to reproduce daily observed data (Tmax, Tmin and precipitation) for three different GCMs during the 1979–2000 baseline period. SDSM is able to produce for a single run several simulations (in this case, 20) with the same probability of appearance. One set of 20 simulations is called an ensemble. The hypothesis tests were based on the ensemble means. Analyses were performed on a seasonal and an annual basis. Seasons are defined as follows: summer (December–February), fall (March–May), winter (June–August) and spring (September–November). A complementary technique, namely frequency analysis, was used in a third step to analyse the models ability to accurately simulate extreme events. The next sections will shortly address the theory related to the steps mentioned above.

3.3.1 Test of equality of median

The Mann–Whitney test is one of the most powerful non-parametric hypothesis tests of the equality of medians of two populations (Mann and Whitney 1947). This test is also known as the two-sample rank test or the two-sample Wilcoxon rank sum test. The Mann–Whitney test is based on the idea that the sum of the ranks for the samples above and below the median should be similar. In this study, a

significance level of 5% ($p < 0.05$) is used. As other non-parametric statistics, the Mann–Whitney test uses the ranks of the sample data, instead of their specific values, to detect the p value. This allows this test to be particularly robust against outliers, non-normal distributed, and auto-correlated data.

3.3.2 Test of equality of variance

The Levene's test is usually used to test the equality of two populations' variance. This test is based on an analysis of variance of the absolute difference from the mean and is most appropriate in cases where data are normally distributed (Levene 1960). However, in this study, as precipitation series displayed a strong skewness in the preliminary EDA, the more robust Brown–Forsythe's test is used (Brown and Forsythe 1974). This is a modification of the Levene's test in which the absolute mean difference is replaced with the absolute median difference. This test appears to be more robust and powerful for skewed data. Using the sample median rather than the sample mean makes also the test more robust for smaller samples (Conover et al. 1981).

In this study, the equality of variances between observed and downscaled Tmin will be tested for four different GCMs and a significance level of 5% ($p < 0.05$) is used.

3.3.3 Frequency analysis

Frequency analysis allows fitting a statistical distribution to observed and downscaled data. It can be used in order to interpret extreme events. Different statistical distributions can be fitted to the selected data. In this case, as the distribution of extreme events (Tmax, Tmin and precipitation) is of interest, a generalised extreme value (GEV) distribution was used. The data are fitted a GEV distribution of the form:

$$F(x) = \exp \left\{ -\exp \left[\frac{-(x - \xi)}{\beta} \right]^k \right\},$$

Where x is the random variable, ξ , β and k are respectively, location, scale and shape parameters estimated from the sample.

This allows interpreting the ability of the statistical downscaling model to simulate the extreme events with regard to observed data sets at various return periods. When plotting extreme value distributions for modelled data with a number of ensembles, it is possible to plot confidence limits (lower and upper percentiles) around the fitted line. In this study, 20 ensembles were simulated for the 1979–2000 period and the 95% confidence interval boundary is chosen.

4 Calibration and validation

Before beginning with the uncertainty analysis per se, the performance of the model shall be explored through classical statistics indicators for the calibration and validation phases.

The analysis of the model's performance is twofold: first, its ability to reproduce observed data by simulations based on the NCEP predictors is investigated. Second, its ability to reproduce observed data by simulation based on the three GCMs predictor variables is analysed. The overall performance of SDSM at investigated stations is summarized in Table 5.

It is clear that Tmax and Tmin are simulated more accurately than precipitation for both NCEP and GCMs variables. Nevertheless, the calibration process returns similar statistics for both temperature and precipitation, with only slightly better values for Tmax and Tmin. However, when the coefficients of determination (R^2) of

precipitation and temperature are compared for the validation period (NCEP and all GCMs), the superior accuracy of simulation for temperature is obvious. In contrast, R^2 values for simulation of precipitation with NCEP and GCMs variables are rather low. Likewise, a decrease (increase) of the standard error values for temperature (precipitation) is observed at most stations.

These general observations fit the conclusions of most works about SDSM and statistical downscaling in a broader sense: the model tends to underestimate precipitation simulations, while reproducing accurately temperatures.

Using yet another perspective of analysis, the performance of the model could be assessed based on which predictor dataset is used. This is possible for the validation period only, the model being calibrated with the NCEP dataset. First, concerning R^2 values for temperature, no significant difference is observed between NCEP-based and GCMs-based simulations. The parameters, derived from the NCEP predictor variables, seem to be well reproduced by

Table 5 Performance indicators for calibration and validation of the model for precipitation, Tmax and Tmin

Predictands	Altitude m a.s.l.	Calibration		Validation							
		R^2	SE	NCEP		CGCM3		GFDL		MRI	
				R^2	SE	R^2	SE	R^2	SE	R^2	SE
Precipitation											
La Serena	15	0.44	0.43	0.33	0.63	0.34	0.30	0.28	0.23	0.15	2.60
Limahuida	295	0.45	0.38	0.49	1.08	0.31	1.02	0.33	1.22	0.42	1.51
Paloma Embalse	320	0.56	0.44	0.52	0.90	0.40	1.13	0.29	1.02	0.41	1.23
Rivadavia	820	0.47	0.42	0.48	0.75	0.42	0.93	0.26	0.82	0.29	1.21
La Tranquilla	975	0.44	0.48	0.51	1.76	0.36	0.51	0.25	0.54	0.44	0.64
Hurtado	1,100	0.48	0.40	0.43	0.91	0.35	1.12	0.30	0.37	0.30	0.47
La Laguna	3,160	0.19	0.35	0.15	1.93	0.24	2.03	0.27	0.55	0.28	0.73
Tmax											
Illapel DGA	290	0.33	2.51	0.85	1.48	0.87	1.32	0.77	2.02	0.88	1.34
Paloma Embalse	320	0.28	2.64	0.85	1.47	0.89	1.24	0.82	1.88	0.89	1.30
Rivadavia	820	0.41	2.62	0.81	1.28	0.72	1.50	0.70	1.30	0.59	1.42
La Tranquilla	975	0.53	2.74	0.66	2.34	0.82	1.65	0.82	1.92	0.77	1.53
Hurtado	1,100	0.46	2.51	0.64	2.00	0.77	1.64	0.79	1.63	0.62	1.64
La Laguna	3,160	0.65	2.29	0.58	3.20	0.84	1.98	0.85	2.47	0.85	2.42
Tmin											
Illapel DGA	290	0.46	1.67	0.80	1.22	0.83	1.06	0.86	0.96	0.81	1.03
Paloma Embalse	320	0.49	1.38	0.90	0.81	0.81	1.05	0.87	0.94	0.80	1.17
Rivadavia	820	0.39	1.82	0.76	1.08	0.71	1.10	0.73	0.98	0.71	1.25
La Tranquilla	975	0.52	1.73	0.62	1.61	0.71	1.25	0.76	1.40	0.52	1.26
Hurtado	1,100	0.46	1.87	0.68	1.50	0.72	1.35	0.72	1.77	0.74	1.48
La Laguna	3,160	0.66	1.64	0.59	2.50	0.86	1.34	0.87	1.79	0.88	1.86

Calibration, 16 years (1979–1994) for temperature and 10-year (1979–1988) period for calibration; validation, year period (1995–2000) for temperatures and 12-year (1989–2000) validation period for precipitation

R^2 coefficient of determination, SE standard error

the GCMs predictors. In the same way, no major difference could be determined between the different GCMs. In contrast, precipitation simulations based on the NCEP dataset return greater R^2 and lesser SE values than those based on the GCMs predictor variables. This indicates that parameters estimated during the calibration process do not fit entirely for the different GCMs predictor variables. Now, it seems important to scrutinize whether the model reports different accuracy with regards to which station was used for the simulation. Comparison studies tend to rely on one or two stations for their analysis.

Table 5 shows that certain stations return a R^2 significantly different from one another. In the case of La Laguna (3,160 ma.s.l.), the R^2 value during the calibration process is distinctly smaller when compared to other stations, especially those located at lower elevations. This variability among stations is also observed for temperature, with high elevation stations such as La Laguna performing better (in term of R^2) than the lower elevation stations Illapel DGA (290 ma.s.l.), La Paloma Embalse (320 ma.s.l.) and Rivadavia (820 ma.s.l.). This confirms a behaviour observed for simulations of temperature and precipitation with SDSM in the region (Rojas 2006; Souvignet et al. 2010). The important eastward slope in the region causing important gradient in precipitation and temperature seems to affect the calibration. Nevertheless, during the validation period, the variability of performance between stations disappears.

Hence, despite a careful screening of predictors variables, the overall accuracy of downscaled precipitations remains poor, whereas it returns good results for downscaled temperature. The use of several stations at different elevations allows drawing conclusion about the influence of slope gradient on the accuracy of downscaled variable. Indeed, results show that downscaled precipitations are influenced by the slope gradient.

5 Uncertainty analysis

To have confidence in simulations of future temperature and precipitation, one should be convinced of the ability of the downscaling model to simulate accurately observed data based on GCMs derived parameters. This section will explore how the downscaling results corresponding to the GCMs predictors reproduce mean, median and extreme events of daily observed records for the 1979–2000 period. The selection of the hypothesis tests of equality was based on the statistical characteristics of observed temperature and precipitation. These characteristics were determined according to the EDA and the conclusions suggested that the uneven distribution of precipitation and possible outliers due to extreme events, robust statistical tests should be used. The next section presents the outcomes of the hypothesis tests of equality of

median and variance. Subsequently, a frequency analysis will determine whether extreme events are accurately simulated for downscaling results based on the GCMs predictors.

5.1 Equality of median

The uncertainty linked with the ability of downscaled data, based on three different GCMs predictors, to reproduce observed median values has been quantified by applying the non-parametric Mann–Whitney test. This test was applied to all stations on a seasonal basis at the 95% confidence level for temperature and precipitation. The results of the test of equality of median are presented for respectively daily precipitation, Tmax and Tmin in Fig. 3a–c. In this case, stations satisfying the hypothesis of equality of median between observed and simulated variables are represented by a black dot, which magnitudes correspond to their respective p values. Stations which did not satisfy the hypothesis of equality are not represented. In addition, the respective percentages of rejected tests of equality of median for precipitation, Tmax and Tmin, summarized in all stations, are displayed in Fig. 4a–c. The results are discussed below.

In the case of precipitation, Fig. 3a shows that the statistical downscaling model reproduces daily observed data median values accurately for the investigated stations at the annual level with the exception of the station La Laguna. Nevertheless, the comparison of median values in spring indicates that daily observed precipitations are not well reproduced by the model in almost all stations (with the exception of La Tranquilla and La Laguna). In addition, lower p values in the northern part of the study area tend to indicate less confidence in the simulation of median values by downscaled precipitation. The graphical comparison in Fig. 4a shows that no significant differences on how median is reproduced by daily downscaled precipitations exist between the three different GCMs. GCGM3, GFDL and MRI return a 14% rejection rate at the annual level and in fall, whereas all three models display a 71% rejection rate in spring. Hence, the comparison of daily median value indicates homogeneity among stations and GCMs with respect to their ability to reproduce daily observed data. This homogeneity exists also at the seasonal level, with the exception of spring. In addition, with the exception of the spring season, precipitation might not be reproduced adequately at high elevations as indicated by the lower performance of the La Laguna station.

In the case of daily Tmax, Fig. 3b shows that observed daily temperature median values are accurately reproduced in all stations by downscaled precipitations at the annual and seasonal level. Comparatively to downscaled precipitation, downscaled Tmax returns a better ability to reproduce median values. Nevertheless, lesser confidence

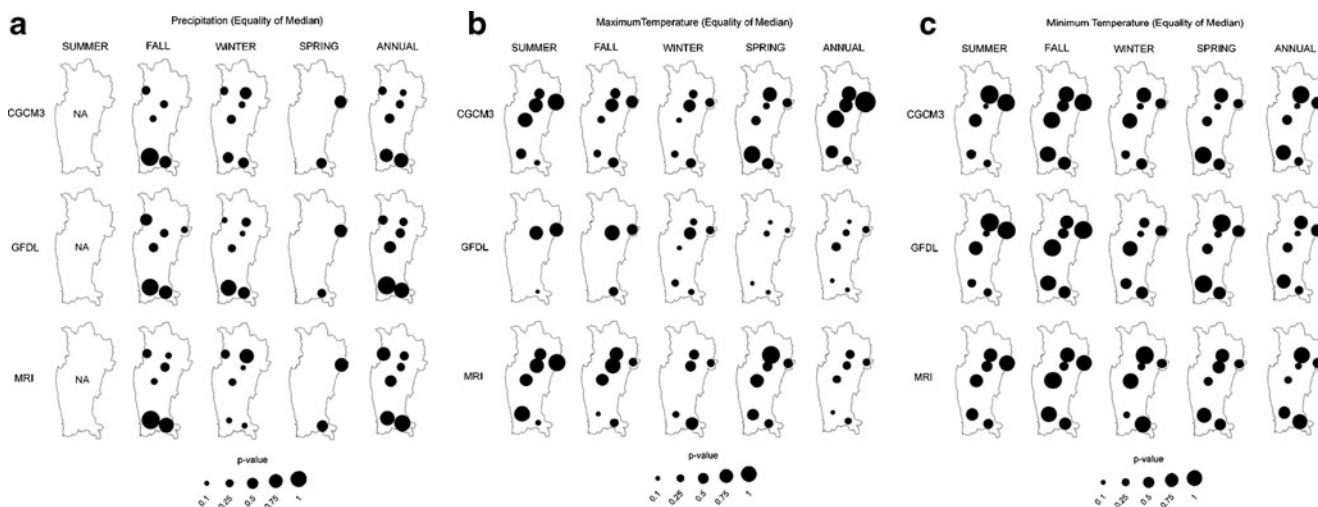


Fig. 3 a–c Results of the Mann–Whitney test for the difference of median between observed and downscaled daily precipitation, Tmax and Tmin. Black dots represent no significant difference at the 95%

confidence level. The magnitude of the dots is proportional to the test p values for the different stations

should be given to certain stations (La Paloma Embalse, Rivadavia and Illapel DGA) which display non-significant p values indicating that the hypothesis of equality of variance should be rejected, especially in summer and fall. The graphical comparison in Fig. 4b indicates that disparities exist in how the different GCMs-based downscaled Tmax reproduce observed median. Where CGCM3 and MRI-based simulations return low rejection rates, GFDL shows a rejection percentage of the hypothesis of equality of median of 67% (50%) for summer (fall). This is also confirmed by the lower p values displayed in Fig. 3b for GFDL. Thus, the comparison of median values between

downscaled and observed Tmax shows that uncertainty with SDSM downscaling exists both for stations, as well as for the different GCM. Consequently, the choice of the GCM and the station selected for analysis is expected to influence the uncertainty linked with the downscaled values.

In the case of Tmin, Fig. 3c shows that all p values are significant, so that all downscaled Tmin median are considered equal to their respective observed values. In the same vein, Fig. 4c shows a rejection rate of zero for all season and all models. In addition, p values are larger relatively to those of Tmax and precipitation, which points

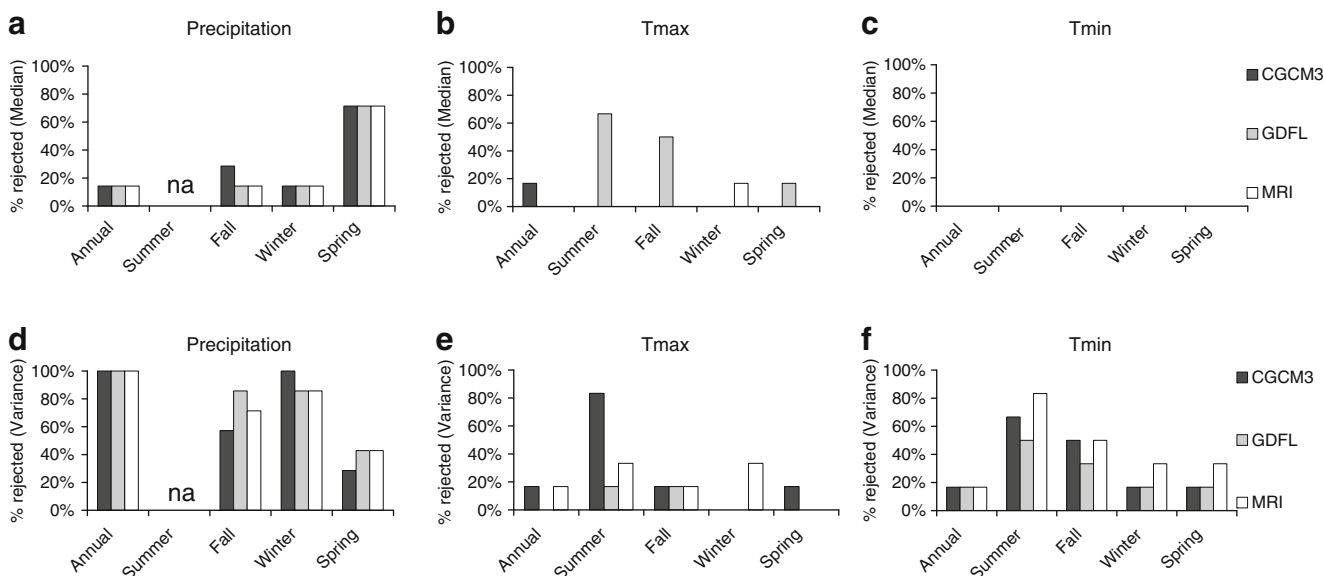


Fig. 4 Percentage of rejected tests for a–c the Mann–Whitney test (equality of median) and d–f the Brown–Forsythe test (equality of variance) for daily precipitation, Tmax and Tmin. Note: summer is not displayed for the tests with precipitation

out an excellent agreement between observed and downscaled values. This indicates that a high confidence level exists in the ability of SDSM to simulate T_{min} in the region, and this independently of the season, the model or the station selected.

Hence, results from the Mann–Whitney test show that very high confidence could be given to the ability of downscaled T_{min} to reproduce the median of observed data. This remains valid at the seasonal and annual levels independently of the GCM used or to which stations it applies. However, more uncertainties arise in the case of downscaled T_{max} and precipitation. If downscaled precipitation returns a relative homogeneity among stations and GCMs with regard to their ability to reproduce variance, significant biases are observed in spring and possibly for high elevations. In the case of T_{max} , the choice of the GCMs from which the predictor variables are used for downscaling appears to have an influence on the accuracy of the median simulation. In this case CGCM3 and MRI return the best estimates.

5.2 Equality of variance

The uncertainty linked with the ability of downscaled data to reproduce the variance of observed values has been quantified by applying the Brown–Forsythe test. The test was applied to all stations on a seasonal basis at the 95% confidence level for temperature and precipitation. The results of the test of equality of variance are presented for respectively daily precipitation, T_{max} and T_{min} in Fig. 5a–c. In this case, stations satisfying the hypothesis of equality of variance between observed and simulated variables are represented by a black dot, which magnitudes correspond to their respective p values. Stations which did

not satisfy the hypothesis of equality are not represented. The respective percentages of rejected tests of equality variance for precipitation, T_{max} and T_{min} , summarized in all stations, are displayed in Fig. 4d–f. The results are discussed below.

In the case of daily downscaled precipitation, Fig. 5a shows that SDSM does not reproduce observed variance values accurately for all investigated stations at the annual. However, the variance is well reproduced by downscaled precipitations in spring for the majority of stations. Nevertheless, in some stations located in the southern part of the region of interest (La Tranquilla and Limahuida), there is no evidence that downscaled precipitations variance is reproduced accurately. The graphical comparison in Fig. 4d shows that differences between the three different GCMs exist on how variance is reproduced by daily downscaled precipitations. For instance, CGCM3 returns the lowest rejection rate in fall (57%) and spring (29%). On the other hand, with a 100% rejection rate, downscaled precipitations from the same GCM seem unable to reflect the variance of observed data at the annual level. In the same way, GFDL has the highest rejection rate (86%) in fall. Consequently, the histogram indicates that levels of rejection for the test of equality of variance are high for all GCMs and for all seasons. In addition, there are some differences between the simulations of variance by the GCMs. Hence, the variance of daily observed precipitation is not accurately reproduced, independently of the GCMs chosen. Slightly better figures exist for spring. However, this season is of lesser importance for arid mountainous region where precipitations occur mainly in winter.

In the case of daily T_{max} , Fig. 5b shows that observed variance is well reproduced in all stations. Nevertheless, in Hurtado, p values are inferior to 0.05 for all GCMs in summer and fall. In addition, comparatively to downscaled

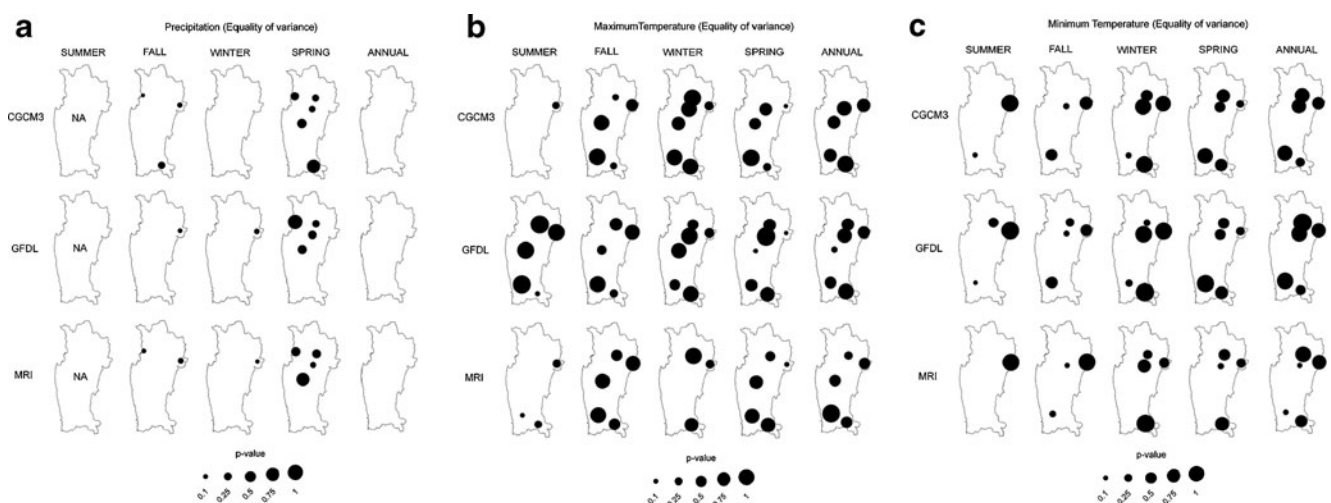


Fig. 5 a–c Results of the Brown–Forsythe test (modified Levene's test) for the difference of variance between observed and downscaled daily precipitation, T_{max} and T_{min} . *Black dots* represent no

significant difference at the 95% confidence level. The magnitude of the dots is proportional to the test p values for the different stations

precipitation, there is a higher confidence in the ability of downscaled Tmax to reproduce variance of observed daily data. The graphical comparison in Fig. 4e shows that the variance is well reproduced by downscaled Tmax for all seasons. Nevertheless, the simulations based on CGCM3 return a high rejection rate (83%) of the test of equality of variance in summer. Consequently, uncertainty linked with the variance of downscaled Tmax is influenced by the GCM choice. Overall, GFDL is the best suited model to reproduce variance of observed data for Tmax in the region.

In the case of daily downscaled Tmin, Fig. 5c shows that the test of hypothesis was less rejected in only three stations (Hurtado, La Laguna and Illapel DGA). It indicates that GCMs-based predictors permitted to reproduce well the variance of observed data. Nevertheless, the lesser performance with regard to the statistical test in other stations suggests that downscaled Tmin are subject to an important variability among stations. Fig. 4f shows that the higher rejection rate is for summer with values ranging from 50% to 83% respectively for GFDL and MRI. MRI-based predictors appear to introduce the most biases in downscaled Tmin variance in all seasons. GFDL is the best performer in terms of non-rejection of the hypothesis of the equality of variance. Hence, these results indicate that a moderate confidence could be placed in the ability of downscaled daily Tmin to reproduce variance of observed data in the region. In addition, the GCMs from which predictors are retrieved have an influence on the uncertainty linked with variance simulation. Moreover, disparities between seasons exist, with higher biases in summer for all GCMs.

Hence, considering the outcomes of the Brown–Forsythe test, downscaled precipitation failed to reproduce observed data variance, independently on which GCM is chosen. On the other hand, the choice of the GCM appears to introduce biases for downscaled Tmax. In this case, GFDL was found to return the best estimates for variables compared with observed data. With regards to Tmin, a moderate confidence could be placed in the downscaled variances. A high uncertainty exists in summer and the choice of the GCM still has an influence, GFDL appearing again to be the best choice for the region.

5.3 Simulation of extreme events

The following section focuses on the model's ability to simulate extreme events. The results of the frequency analysis, performed for the different GCMs for the 1979–2000 time period, are displayed in Fig. 6a–f for precipitation, in Fig. 6g–l for Tmax and in Fig. 3m–r for Tmin. Because not all results for the 19 stations at three GCMs could be intelligibly displayed, a selection of investigated stations was made for the display of the frequency analysis.

As we have seen previously that the performance indicator (R^2) of simulations for stations located at different elevations vary significantly, two stations were selected according to their respective elevation: at high altitude (La Laguna, 3,160 m.a.s.l.) and low altitude (La Paloma Embalse, 320 m.a.s.l.) are investigated. Because the authors understand that results of the frequency analysis for other stations in the region could be of interest for comparison purposes, outcomes for other stations could be delivered upon request.

First, simulations of extreme events will be analysed with regards to precipitation. Subsequently, an analysis of extreme Tmax and Tmin will follow.

The extreme events prediction ability of SDSM for precipitation is displayed for La Laguna in Fig. 6a–c and in La Paloma Embalse in Fig. 6d–f. At both stations and for all three GCMs, downscaled precipitations clearly underestimate historical records. Nevertheless, extreme events with a return period inferior to 20 years are accurately simulated in La Laguna. In contrast, downscaled precipitations in La Paloma return underestimation for shorter return periods (5 years). These results seem to indicate that extreme events with large return period (>20 years) could not be simulated with confidence in the upper part of the Andes, where the largest rainfall amount occurs. Also, the systematic underestimation of rainfall amounts and the inability to simulate short term extreme events (events with a 5-year return period) at lower elevation is problematic with regard to the ENSO phenomena. In addition, no evident difference is observed between precipitations downscaled with the three different models. This indicates that a multi-model approach to the simulation of future precipitations in the region might not enhance the model response. Thus, lower reliance levels should be set for extreme precipitation values. This behaviour is explained by the inherent characteristics of statistical downscaling procedures. Indeed, as mentioned by Kim et al. (2000), this type of method performs better “if two conditions are met: (1) there are sufficient historical data for generating probability distribution functions and (2) variables of interest have well defined statistical patterns”. Thus, in the case of our study, in an arid regime with rather few precipitation events per year, it seems that the first condition is hardly met, i.e. there are not enough values to allow for an accurate regression with the NCEP and the CGM data. Second, the occurrence within the validation data set of a particularly strong El Niño event (1997) and a severe, La Niña-related, drought period (1994–1996) is likely an additional source of variability. This is supported by the studies of Barnston et al. (1999), Kim et al. (2000) and Wood et al. (2002) that show the effect of the 1997 El Niño event, one of the strongest in the last century, on the performance of several climate and downscaling models.

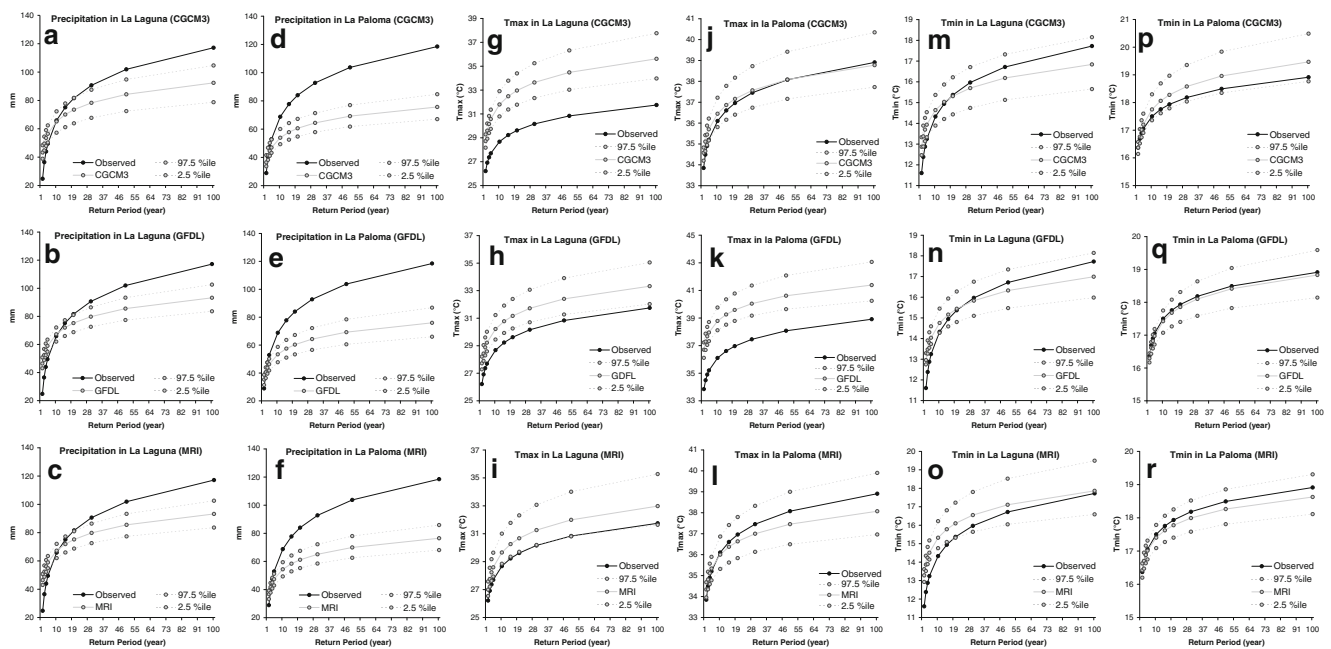


Fig. 6 Frequency analysis for precipitation (a–f), maximum temperature (g–l) and minimum temperature (m–r) in La Laguna (3,160 m.a.s.l.) and La Paloma Embalse (320 m.a.s.l.) for CGCM3, GFDL and

MRI models (1979–2000). Note: all statistical ensembles (20) are included for the GCMs simulations

The ability to simulate extreme downscaled maximum temperature is displayed for La Laguna in Fig. 6g, i and La Paloma Embalse in Fig. 6j, l. It appears that the extreme events prediction ability, based on all ensembles, is not equally well reproduced by the GCM models at the two stations. In addition, it does not appear that the three GCMs have the same ability to simulate downscaled values for Tmax. Indeed, downscaled Tmax is overestimated in La Laguna for the CGCM3 model and for the GFDL model in La Paloma Embalse. In contrast, downscaled Tmax based on MRI predictors performs well in both stations. This indicates that the choice of the model might be of some importance with regard to the ability of the model to reproduce downscaled extreme Tmax.

Downscaled extreme minimum temperatures are plotted against observed records in Fig. 6m–o for La Laguna and in Fig. 6p–r for La Paloma Embalse. Results suggest that extreme minimum temperatures are simulated accurately by all models and in both stations. However, for different return periods (<10 years for GCM3 and GFDL; <20 years for MRI), downscaled Tmin are overestimated in La Laguna. This does not appear in La Paloma Embalse, where extreme events are simulated accurately for all return periods. Hence, in the case of Tmin, the elevation of the station seems to influence the model ability to simulated extreme events.

Hence, outcomes from the frequency analysis show that the ability of SDSM to simulated extreme events is not always secured. In the case of precipitation, a systematic

underestimation of extreme events is observed. Therefore, limited confidence should be placed on the ability of the model to simulate future ENSO phenomena, and this is independent of the model chosen. With regard to Tmax and Tmin, higher confidence exists in the model ability to simulate extreme events. Nevertheless, biases persist, with an overestimation of Tmax, a dependency upon what GCM predictors are used and biases introduced by altitude's gradients. This has consequently important implications for arid mountainous regions, where snow and glacier melting are heavily influence by temperature changes (Pouyaud et al. 2005; Singh et al. 2006).

6 Conclusions

The evaluation of uncertainties linked with downscaled temperature and precipitation variables is crucial for regional and local impact studies. This paper showed the potential of statistical downscaling technique using predictors from three different CGMs over the highly heterogeneous area of the central Andes in Chile. Biases in median and variance were estimated for downscaled temperature and precipitation. In addition, the ability of the downscaled variables to reproduce extreme events was tested using a frequency analysis.

Despite a careful screening of predictor variables, the overall accuracy of downscaled precipitations remains poor compared to downscaled temperatures, which return more

accurate simulations of observed records. Extreme events, of particular relevance in the central Andes, where climate is heavily influenced by the ENSO phenomena, were systematically underestimated for precipitation. Whereas a higher confidence lies in the simulation of extreme temperature conditions, the existence of biases due to altitude gradients questions the accuracy of such predictions in regions where snow and ice processes are strongly influenced by temperature gradients. Besides, no evidence was found that a particular GCM was best adapted to the region. This underlines therefore the importance of a multi-model approach for impact studies in the region and generally in areas with heterogeneous surface conditions. Such uncertainties in the downscaled results diminish the confidence that one should have in simulation of future climate scenario, and raise the question of whether local climate impact studies based on the outcomes of SDSM in the region (e.g. on water resources) are meaningful.

Nevertheless, in general, the simulations produced with the statistical downscaling approach (viz. SDSM) still outperformed raw GCMs outputs, unable to reproduce the complexity of the central Andes and similar regions.

Testing the outcomes of downscaling is important and as stated by Blöschl and Montanari (2010), climate impact studies tend to be over-optimistic about their own reliability and over-pessimistic concerning the potential impacts on society. Therefore, evaluating the reliability and quantifying uncertainties linked with statistical downscaling allows a more serene approach to local climate impact studies. In addition this study contributed to evaluate which predictors are the most appropriate to the region. However, in this case, further work is needed to develop predictors better adapted to precipitation and temperature. Eventually, other downscaling techniques should be explored. For instance, a multi-site artificial neural network approach using a non-linear transfer function might be introduced to map the predictor–predictand relationships. Moreover, the multi-objective fuzzy logic-based classification approach offers the possibility to identify large-scale atmospheric or oceanic patterns, which are responsible for wet and dry phenomena in the research area. This method has already been successfully used to identify droughty and wet weather patterns in West Africa (Laux et al. 2009)

Acknowledgments The authors express also their gratitude to the DGA for providing the meteorological variables used in this study. We acknowledge the modelling groups for making their model output available for analysis, and the Program for Climate Model Diagnosis and Inter-comparison for collecting and archiving this data. In addition, the authors would like to thank the IRI for providing access to its data library. SDSM was supplied by Robert Wilby and Christian Dawson on behalf of the Environment Agency of Wales. Finally, we would like to acknowledge one anonymous reviewer for his constructive comments, which greatly helped in improving our original manuscript.

References

- Aceituno P (1988) On the functioning of the southern oscillation in the south american sector. Part I: surface climate. *Mon Weather Rev* 116(3):505–524
- Barnston AG, He Y, Glantz MH (1999) Predictive skill of statistical and dynamical climate models in sst forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bull Am Meteorol Soc* 80(2):217–243
- Biondi F, Gershunov A, Cayan DR (2001) North Pacific decadal climate variability since 1661. *J Climate* 14:5–10
- Blöschl G, Montanari A (2010) Climate change impacts—throwing the dice? *Hydrolog Process* 24(3):374–381
- Boe J, Terray L, Habets F, Martin E (2007) Statistical and dynamical downscaling of the seine basin climate for hydro-meteorological studies. *Int J Climatol* 27(12):1643–1655
- Brown MB, Forsythe AB (1974) Robust tests for the equality of variances. *J Am Stat Assoc* 69:364–367
- Chu J, Xia J, Xu CY, Singh V (2010) Statistical downscaling of daily mean temperature, pan evaporation and precipitation for climate change scenarios in Haihe River, China. *Theor Appl Climatol* 99(1):149–161
- Conover WJ, Johnson ME, Johnson MM (1981) A comparative study of tests for homogeneity of variance with applications to the outer continental shelf bidding data. *Technometrics* 23:351–361
- Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Cooke WF, Dixon KW, Dunne J, Dunne KA, Durachta JW, Findell KL, Ginoux P, Gnanadesikan A, Gordon CT, Griffies SM, Gudgel R, Harrison MJ, Held IM, Hemler RS, Horowitz LW, Klein SA, Knutson TR, Kushner PJ, Langenhorst AR, Lee H-C, Lin S-J, Lu J, Malyshev SL, Milly PCD, Ramanaswamy V, Russell J, Schwarzkopf MD, Shevliakova E, Sirutis JJ, Spelman MJ, Stern WF, Winton M, Wittenberg AT, Wyman B, Zeng F, Zhang R (2006) GFDL's CM2 global coupled climate models. Part I: formulation and simulation characteristics. *J Climate* 19(5):643–674
- Diaz-Nieto J, Wilby RL (2005) A comparison of statistical downscaling and climate change factor methods: impacts on low flows in the River Thames, United Kingdom. *Climatic Change* 69(2):245–268
- Dibike YB, Gachon P, St-Hilaire A, Ouarda T, Nguyen VTV (2008) Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada. *Theor Appl Climatol* 91(1–4):149–170
- Doblas-Reyes FJ, Pavan V, Stephenson DB (2003) The skill of multi-model seasonal forecasts of the wintertime North Atlantic Oscillation. *Clim Dynam* 21(5):501–514
- Downing ET, Santibáñez F, Romero H, Pena HT, Gwynne RN, Ihl M, Rivera A (1994) Climate change and sustainable development in the Norte Chico, Chile: climate, water resources and agriculture. Environmental Change Unit Research Report No. 6. University of Oxford, Oxford
- Dregne HE (1985) Aridity and land degradation. *Environment* 27(8):16–20
- Favier V, Falvey M, Rabatel A, Praderio E, Lopez D (2009) Interpreting discrepancies between discharge and precipitation in high-altitude area of Chile's Norte Chico region (26–32°s). *Water Resour Res* 45 (W02424). doi:10.1029/2008WR006802
- Flato GM, Boer GJ (2001) Warming asymmetry in climate change simulations. *Geophys Res Lett* 28(1):195–198
- Garreaud RD, Vuille M, Compagnucci R, Marengo J (2009) Present-day South American climate. *Palaeogeogr Palaeoclimatol Palaeoecol* 281(3–4):180–195
- Giorgi F, Mearns LO (1991) Approaches to the simulation of regional climate change: a review. *Rev Geophys* 29:191–216
- Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from aogcm simulations via the

- reliability ensemble averaging (REA) method. *J Clim* 15(10):1141–1158. doi:10.1175/1520-0442(2002)015<1141:COAURA>2.0.CO;2
- Greene AM, Goddard L, Lall U (2006) Probabilistic multimodel regional temperature change projections. *J Clim* 19(17):4326–4343
- Helsel DR, Hirsch RM (2002) *Statistical methods in water resources*. Elsevier, Amsterdam
- Holmgren M, Stapp P, Dickman CR, Gracia C, Graham S, Gutierrez JR, Hice C, Jaksic F, Kelt DA, Letnic M, Lima M, Lopez BC, Meserve PL, Milstead WB, Polis GA, Previtali MA, Richter M, Sabate S, Squeo FA (2006) Extreme climatic events shape arid and semiarid ecosystems. *Front Ecol Environ* 4(2):87–95. doi:10.1890/1540-9295(2006)004[0087:ECESAA]2.0.CO;2
- Hostetler SW (1994) Hydrologic and atmospheric models: the (continuing) problem of discordant scales. *Climatic Change* 27(4):345–350
- IPCC (2007) *Climate change 2007: the physical science basis*. Contribution of Working Group I to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Jenne R, Joseph D (1996) The NCEP-NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77(3):437–471
- Kalthoff N, Bischoff G, Inge Fiebig-Wittmaack M, Fiedler F, Thrauf J, Novoa E, Pizarro C, Castillo R, Gallardo L, Rondanelli R, Kohler M (2002) Mesoscale wind regimes in Chile at 30°s. *J Appl Meteorol* 41(9):953–970
- Khan MS, Coulibaly P, Dibike Y (2006) Uncertainty analysis of statistical downscaling methods using Canadian global climate model predictors. *Hydrolog Process* 20(14):3085–3104
- Kim J, Miller NL, Farrara JD, Hong S-Y (2000) A seasonal precipitation and stream flow hindcast and prediction study in the western United States during the 1997/98 winter season using a dynamic downscaling system. *J Hydrometeorol* 1(4):311–329
- Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, van den Dool H, Jenne R, Fiorino M (2001) The NCEP-NCAR 50-year reanalysis: monthly means CD-ROM and documentation. *Bull Am Meteorol Soc* 82(2):247–267
- Kundzewicz ZW, Mata LJ, Arnell NW, Döll P, Jimenez B, Miller K, Oki T, Şen Z, Shiklomanov I (2008) The implications of projected climate change for freshwater resources and their management. *Hydrol Sci J* 53(1):3–10
- Labraga J (2010) Statistical downscaling estimation of recent rainfall trends in the eastern slope of the Andes mountain range in Argentina. *Theor Appl Climatol* 99(3):287–302
- Laux P, Wagner S, Wagner A, Jacobeit J, Bárdossy A, Kunstmann H (2009) Modelling daily precipitation features in the Volta Basin of West Africa. *Int J Climatol* 29(7):937–954
- Levene H (1960) Robust tests for the equality of variances. In: Olkin I, Ghurye S, Hoëfding W, Madow WG, Mann HB (eds) *Contribution to probability and statistics*. Stanford University Press, Palo Alto
- Lopez A, Tebaldi C, New M, Stainforth D, Allen M, Kettleborough J (2006) Two approaches to quantifying uncertainty in global temperature changes. *J Climate* 19(19):4785–4796. doi:10.1175/JCLI3895.1
- Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. *Ann Math Stat* 18:50–60
- Marengo JA, Jones R, Alves LM, Valverde MC (2009) Future change of temperature and precipitation extremes in South America as derived from the PRECIS regional climate modeling system. *Int J Climatol* 29:2241–2255
- Mata LJ, Campos M (2001) Latin America. In: McCarthy JJ, Canziani OF, Leary NA, Dokken DJ, White KS (eds) *Climate change 2001: impacts, adaptation, and vulnerability*. Contribution of Working Group II to the third assessment report of the intergovernmental panel on climate change. University Press, Cambridge, pp 693–734
- Mearns LO, Easterling W, Hays C, Marx D (2001) Comparison of agricultural impacts of climate change calculated from high and low resolution climate change scenarios: part I. The uncertainty due to spatial scale. *Climatic Change* 51(2):131–172
- Messerli B, Grosjean M, Graf K, Schotterer U, Schreier H, Vuille M (1992) Die veränderungen von klima und umwelt in der region acatama (chile) seit der letzten kaltzeit. *Erdkunde* 46(3/4):257–272
- Miller A (1976) The climate of Chile. In: Schwerdfeger W (ed) *World survey of climatology: climates of Central and South America*. Elsevier, Amsterdam, pp 113–145
- Penlap EK, Matulla C, von Storch H, Kamga FM (2004) Downscaling of GCM scenarios to assess precipitation changes in the little rainy season (March–June) in Cameroon. *Clim Res* 26(2):85–96
- Pouyaud B, Zapata M, Yerren J, Gomez J, Rosas G, Suarez W, Ribstein P (2005) Avenir des ressources en eau glaciaire de la cordillère blanche. *Hydrol Sci J* 50(6):999–1022
- Rojas M (2006) Multiply nested regional climate simulation for southern South America: sensitivity to model resolution. *Mon Weather Rev* 134(8):2208–2223
- Singh P, Arora M, Goel NK (2006) Effect of climate change on runoff of a glacierized Himalayan Basin. *Hydrolog Process* 20:1979–1992
- Solman SA, Nunez MN (1999) Local estimates of global climate change: a statistical downscaling approach. *Int J Climatol* 19:835–861
- Souvignet M, Gaese H, Ribbe L, Kretschmer N, Oyarzún R (2010) Statistical downscaling of precipitation and temperature in north-central Chile: an assessment of possible climate change impacts in an arid Andean watershed. *Hydrol Sci J* 56(1):41–57
- Souvignet M, Oyarzún R, Verbist K, Gaese H, Heinrich J (2011) Trends in hydro-meteorological extremes in semi-arid North-Central Chile (20–32°s): water resources implications for a fragile Andean region. *Hydrolog Sci J* (in press)
- Verbist K, Robertson AW, Cornelis WM, Gabriels D (2010) Seasonal predictability of daily rainfall characteristics in central northern Chile for dry-land management. *J Appl Meteorol Clim* 49(9):1938–1955. doi:10.1175/2010JAMC2372.1
- Wetterhall F, Bárdossy A, Chen D, Halldin S, Xu C-Y (2006) Daily precipitation-downscaling techniques in three Chinese regions. *Water Resour Res* 42(11):13. doi:10.1029/2005WR004573
- Wetterhall F, Halldin S, Xu CY (2007) Seasonality properties of four statistical-downscaling methods in central Sweden. *Theor Appl Climatol* 87(1–4):123–137
- Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Supporting material of the Intergovernmental Panel on Climate Change. Available from the DDC of IPCC TGCI, 27pp.
- Wilby RL, Dawson CW, Barrow EM (2002) SDSM—a decision support tool for the assessment of regional climate change impacts. *Environ Model Software* 17(2):145–157
- Wood AW, Maurer EP, Kumar A, Lettenmaier DP (2002) Long-range experimental hydrologic forecasting for the eastern United States. *J Geophys Res* 107(D20):4429. doi:10.1029/2001JD000659
- Yukimoto S, Noda A, Kitoh A, Sugi M, Kitamura Y, Hosaka M, Shibata K, Maeda S, Uchiyama T (2001) The new Meteorological Research Institute coupled GCM (MRI-CGCM2). Model climate and variability. *Pap Meteorol Geophys* 51:47–88