Theor. Appl. Climatol. 91, 149–170 (2008)
DOI 10.1007/s00704-007-0299-z
Printed in The Netherlands

Theoretical and Applied Climatology

[1] OURANOS Consortium on Regional Climatology and Adaptation to Climate Change, Montreal (QC), Canada
[2] Adaptation and Impact Research Division (AIRD), Atmospheric Science and Technology Directorate, Environment Canada at Ouranos, Montreal (QC), Canada
[3] Institut National de la Recherche Scientifique Centre Eau, Terre & Environnement (INRS-ETE), University of Québec, Québec (QC), Canada
[4] Department of Civil Engineering and Applied Mechanics, McGill University, Montreal (QC), Canada

Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada

Y. B. Dibike[1], **P. Gachon**[1,2], **A. St-Hilaire**[3], **T. B. M. J. Ouarda**[3], and **Van T.-V. Nguyen**[4]

With 15 Figures

Received June 14, 2006; revised December 15, 2006; accepted December 30, 2006
Published online April 12, 2007

Summary

Uncertainty analysis is used to make a quantitative evaluation of the reliability of statistically downscaled climate data representing local climate conditions in the northern coastlines of Canada. In this region, most global climate models (GCMs) have inherent weaknesses to adequately simulate the climate regime due to difficulty in resolving strong land/sea discontinuities or heterogeneous land cover. The performance of the multiple regression-based statistical downscaling model in reproducing the observed daily minimum/maximum temperature, and precipitation for a reference period (1961–1990) is evaluated using climate predictors derived from NCEP reanalysis data and those simulated by two coupled GCMs (the Canadian CGCM2 and the British HadCM3). The Wilcoxon Signed Rank test and bootstrap confidence-interval estimation techniques are used to perform uncertainty analysis on the downscaled meteorological variables. The results show that the NCEP-driven downscaling results mostly reproduced the mean and variability of the observed climate very well. Temperatures are satisfactorily downscaled from HadCM3 predictors while some of the temperatures downscaled from CGCM2 predictors are statistically significantly different from the observed. The uncertainty in precipitation downscaled with CGCM2 predictors is comparable to the ones downscaled from HadCM3. In general, all downscaling results reveal that the regression-based statistical downscaling method driven by accurate GCM predictors is able to reproduce the climate regime over these highly heterogeneous coastline areas of northern Canada. The study also shows the applicability of uncertainty analysis techniques in evaluating the reliability of the downscaled data for climate scenarios development.

1. Introduction

The starting point for most climate scenarios development is the Global Climate Models (GCMs) which have horizontal resolutions of hundreds of kilometres (around 350 km in general). However, many impact assessments need point scale information or local climate variables which are highly sensitive to fine-scale climate variations and feedbacks that are parameterized in coarse-scale models (e.g. Wilby et al., 2004; Mearns et al., 2003). Hence, there is a clear need for more reliable high-resolution scenarios at a spatial scale much finer than that provided by global or even some regional climate models (see for example the recent European PRUDENCE project, http://www.dmi.dk/f+u/klima/prudence/, e.g. Christensen et al., 2002). Two major downscaling approaches, namely, *dynamical downscaling* and *statistical* *downscaling*, are commonly used for climate scenario development at higher resolution. Dynamical downscaling generates regional-scale information by developing and using Regional Climate Models (RCMs) with the coarse GCM data used as boundary conditions. Statistical downscaling (SD) methods, on the other hand, involve developing quantitative relationships between large-scale atmospheric variables, the predictors, and local surface variables, the predictands. Since they are derived from the historical observed data, they can provide site specific information as recommended in many climate change impacts studies.

Many types of statistical methods, such as regression techniques (e.g. Wilby et al., 2002), artificial neural networks (e.g. Coulibaly et al., 2005) or weather generator (e.g. Katz, 1996; Semenov and Barrow, 1997) have been developed and applied for downscaling climate variables. As downscaling predictability and skill varies seasonally, regionally and between different periods of record as well as according to the considered variable (e.g. Gachon et al., 2005), verification of the physical plausibility of the downscaling results is essential. It is also important to test the SD model with independent data and apply the model to a wide range of climate model outputs, to evaluate the uncertainties associated with the GCM structures and skills to reproduce the key atmospheric variables related to the local climate regime. Therefore, in order to have reasonable confidence on the reliability of the downscaled future climate scenarios, one necessary condition (not sufficient) is to make uncertainty analysis (with statistical objective criteria) on the downscaled data (with a comprehensive assessment of all SD methods, GCMs and RCMs) to evaluate its performance in reconstructing the observed climate regime for the baseline period (Khan et al., 2006a, b).

Whereas GCM simulated temperature changes in the north are higher than in southern areas of Canada (see Barrow et al., 2004), it is also evident that coarse-resolution GCMs have inherent difficulty to simulate a reliable climate regime in coastal or island regions, and/or of highly heterogeneous land/cover. Therefore, these anticipated changes are less reliable especially for Nordic areas of Canada where sea ice and snow cover are present over the major part of the year, and in which the temperature regime is highly sensitive to fine scale climate forcings from the underlying surface conditions (e.g. Gachon et al., 2003; Barrow et al., 2004). Gachon et al. (2005) have identified similar problem in the case of statistically downscaled information, especially in northern areas and around Hudson Bay in winter season. Accordingly, a careful evaluation of the reconstructed climate regimes (both from GCMs and downscaling outputs) is very important for northern applications, where reliable information on the magnitude and the timing of the thawing/freezing cycle is needed for the study of hydrologic, human health and environmental issues. Therefore, the main objective of this study is to evaluate the most widely used multiple regression-based statistical downscaling model (SDSM; Wilby et al., 2002) with respect to its potential to reproduce the mean, variability and the probabilities of extreme temperature and precipitation of the observed climate regime in northern Canada.

An uncertainty analysis is made on the downscaled data to evaluate the performance of the SDSM method in reproducing the observed climate variables for the baseline period (1961–1990) when provided with climate predictors simulated by two different GCMs, after calibrating and validating of the model with reanalysis predictors. The analysis focuses on monthly and seasonal mean values and variability of temperature and precipitation over different climate regimes in the northern Canada. Two complementary methods, namely the Wilcoxon Signed Rank test (Wilcoxon, 1945) and confidence intervals constructed using bootstrap resampling (Efron and Tibshirani, 1993) have been employed to perform the uncertainty analysis on the SD outputs. As suggested in Wilby et al. (2004), it is important to recognize that increased precision of downscaling results for the current period does not necessarily imply an increased confidence in regional or local climate scenario information compared to raw-GCM output. However, these methods will allow the use of objective statistical criteria to identify the most robust results, in term of the reliability and realism of GCMs and downscaled variables relative to observed climatology under present conditions, and help to evaluate the potential added values or new insight that have been gained through the use of downscaling methods.

This paper is structured in six sections. Section 2 describes the region of interest and the database used for the study, while Sect. 3 provides the methodology for the multiple linear regression-based SDSM downscaling model, and the uncertainty analysis techniques employed in this study. Section 4 presents the SDSM downscaling results along with the results of the uncertainty analysis on the downscaled precipitation and temperature data for the baseline period. Section 5 discusses the results and summarizes the findings of the study. General conclusions are provided in Sect. 6.

2. Region of interest and database

The region of interest in this study is the extreme northern part of Canada from Hudson and Baffin Bay areas in the east, to Beaufort and Canadian archipelago in the west (see Fig. 1). The main reasons for the choice of this region for uncertainty analysis of global climate models and statistical downscaling model outputs are as follows:

1. Stronger climate change signal are suggested in the north by most climate models (e.g. Houghton et al., 2001; Barrow et al., 2004);
2. Stronger biases in GCMs simulation of surface temperatures are identified in the northern areas of Canada and especially in winter as shown in Chapter 8 of IPCC (Intergovernmental Panel of Climate Change, e.g. Houghton et al., 2001; see also Barrow et al., 2004). Hence, there is relatively less confidence in the climate change signal in that region and season unless the uncertainties associated with the simulated data are explicitly assessed;
3. Stronger biases in SD results of temperature are reported for the Hudson Bay during winter months (e.g. Gachon et al., 2005). So it is important to analyze if this is systematic over all northern areas (i.e. not only limited to Hudson Bay region, but also over different areas near Hudson Bay, Baffin Bay, Canadian Archipelago and Beaufort Sea);
4. Climate change in the northern areas will potentially have strong impacts on ecosystem, permafrost, animals, biodiversity, human health, and transportation (e.g. the issues regarding the ice free conditions of the North West Passage). Hence, it is important to develop better knowledge and confidence on climate change regime in this particular region

Fig. 1. Geographic locations of the five climate stations in Northern Canada considered in this study. The stations are identified by red points and are all located over northern coastline areas

of Canada. Moreover, this is the first study that provides statistical downscaling results for northern Canada.

The downscaling models are developed (calibrated and validated) first using climate predictors derived from the National Centre for Environmental Prediction (NCEP) reanalysis data set (Kistler et al., 2001) for the recent past and current time period (1961–1990 and 1991–2000, respectively). Then, predictors simulated by two coupled GCMs, namely the Canadian global climate model version 2 (CGCM2; Flato et al., 2000; Flato and Boer, 2001) and the Hadley Center general circulation model (HadCM3; Gordon et al., 2000), are used to downscale the climate information for the baseline period (1961–1990). Besides the large scale climate variables (predictors), the viability of downscaling techniques also depends critically upon access to high-quality local predictands. Therefore, historical daily precipitation and temperature records at five climate stations located in various northern regions, from Hudson and Baffin Bay areas to the Beaufort Sea and the Canadian archipelago (Fig. 1, geographical coordinates are presented in Table 1) are used for the downscaling exercise. The Environment Canada homogenized temperature (i.e. daily minimum and maximum temperature, hereafter called T_{min} and T_{max}, respectively), and rehabilitated precipitation dataset (Vincent et al., 2002) are used for this study. This dataset minimizes the risk of introducing additional uncertainty due to the changes in climate monitoring practices or due to the non-homogeneity of individual site records. The preparation of the potential predictors from NCEP reanalyses and the GCMs' output involves data extraction, re-gridding and standardisation. Pre-processed data were obtained from CCSN (Canadian Climate Scenarios Network) project of Environment Canada (http://www.ccsn.ca). The NCEP and GCMs predictors available for the downscaling experiment are enumerated in Table 2.

Table 1. Meteorological stations where observed precipitation and temperature data are used for the downscaling exercise

	Station location (in degrees)		
	Latitude (N)	Longitude (W)	Altitude (m)
Cape Dorset	64.2	76.5	51
Inuvik	68.3	133.5	68
Iqaluit	63.8	68.5	34
Resolute Bay	74.8	94.9	67
Shepherd Bay	68.8	93.4	51

Table 2. Predictor variables considered from NCEP, CGCM2 and HadM3 data sets

No.	Predictors	No.	Predictors
1	Mean sea level pressure	14	850 hPa divergence
2	Surface airflow strength	15	850 hPa airflow strength
3	Surface zonal velocity	16	850 hPa zonal velocity
4	Surface meridional velocity	17	850 hPa meridional velocity
5	Surface vorticity	18	850 hPa vorticity
6	Surface wind direction	19	850 hPa geopotential height
7	Surface divergence	20	850 hPa wind direction
8	500 hPa airflow strength	21	850 hPa divergence
9	500 hPa zonal velocity	22	Near surface relative humidity
10	500 hPa meridional velocity	23	Specific/relative humidity at 500 hPa
11	500 hPa vorticity	24	Specific/relative humidity at 850 hPa
12	500 hPa geopotential height	25	Near surface specific humidity
13	500 hPa wind direction	26	Mean temperature at 2 m

3. Methodology

3.1 Statistical down-scaling model (SDSM)

The multiple-regression statistical downscaling model, SDSM, implements a statistical representation of empirical relationships between local *predictands*, such as daily temperature and precipitation, and the large scale atmospheric variables or *predictors*. The downscaling procedure of the SDSM software (version 3.1 recently developed by Wilby and Dawson, 2004) is performed through a number of steps such as quality control and data transformation, screening of predictor variables, model calibration and weather generation (see more details in Wilby and Dawson, 2004). As suggested in Wilby et al. (2004), predictors have to be chosen based on both their relevance to the downscaled predictand, and their accurate representation by the climate models used for climate change simulation (e.g. Wilby and Wigley, 2000). Previous studies show that SDSM performs relatively well in simulating the main characteristics of temperature regime, but it explain only a fraction of the observed climate variability, especially in precipitation series (e.g. Wilby and Dawson, 2004; Dibike and Coulibaly, 2005; Gachon et al., 2005). Furthermore, downscaling climate variables using regression methods tends to be problematic since most extreme events lay towards the tails of the distribution or beyond the range of the calibration data set (e.g. Wilby et al., 2004). A recent study of Gachon et al. (2005) gives further details around the strengths and weaknesses of this tool in downscaling extreme indices of temperature and precipitation over eastern Canada. One of the main conclusions of that study is that SDSM has not performed well in reproducing the temperature regime in the north when GCMs (in that case, CGCM1 and HadCM3) are used as sources of predictors. In that region, SD models are strongly affected by biases in the underlying GCMs, and works are needed to systematically assess the accuracy of candidate predictors in that particular area (e.g. Gachon et al., 2005).

3.2 Selection of predictor variables

Once the SDSM downscaling model is calibrated and validated using NCEP predictors, then the corresponding GCM predictors are used to downscale the daily predictand variables. In the process, the downscaling model propagates the uncertainty from the driving GCM fields, and cannot correct biases or uncertainties existing in the GCM itself (e.g. Hewitson and Crane, in press). Validation of climate model outputs has shown that some potential CGCM2 and HadCM3 predictors, such as the near surface temperature and specific/relative humidity, have significant differences from the corresponding NCEP predictor variables (not shown here). This type of systematic biases and time shifts between some of NCEP and the corresponding GCMs predictors suggest that it is preferable not to use such variables as predictors so as to prevent the propagation of the discrepancies in these atmospheric fields into the statistical downscaling process from the host GCMs.

It is also necessary to specify the optimum location of the large scale predictor fields to achieve the best performance in downscaling local climate variables (e.g. Wilby et al., 2004). The optimal location of predictors may vary by region and the use of spatial correlation can be helpful in identifying the optimal geographic location for each potential predictor. For this study, such analysis was experimented for some of the potential predictor variables at one of the climate stations (not presented here). The main purpose was to analyze if it is necessary to use climate predictors from grid points other than the one which is the closest to the predictand. In most cases, the differences in the values of the correlation coefficients at the different grid point in the vicinity of the predictand are too small (in the order of 0.01 and 0.04 for temperature and precipitation, respectively) to make any significant difference in the downscaling performance. As a result, in all subsequent downscaling exercises in this study, NCEP and GCM predictors located only at the grid point closer to each particular predictand are used to calibrate and validate the SDSM model. The most commonly used predictors for the downscaling of temperature for the study area are the 500 and 850 hPa geopotential heights, relative and specific humidity at 850 hPa and vorticity at 500 hPa. This confirms the strong relationships between mid-tropospheric geopotential thickness and surface temperature regime, as this Canadian sector corresponds to an area where the presence of the upper level cold trough is very common, corresponding to the coldest air conditions at all tropospheric levels in Arctic, as suggested in Overland et al. (1997). On the other hand, the most commonly used predictors for the downscaling of precipitation are mean sea level pressure, 500 hPa geopotential height, 850 hPa zonal velocity, and specific humidity at 500 and 850 hPa.

As shown in Table 3, the amount of explained variance (R^2) obtained for the calibration period (using NCEP predictors) varies between 0.11 and 0.23 for precipitation, and between 0.52 and 0.59 for temperature. The relatively low explained variance for precipitation underlines the more stochastic nature of precipitation occurrence and magnitude, and the difficulty to capture the characteristics of the variability of the precipitation regime in the downscaling process (compared to temperature), as also suggested in other studies

Table 3. Relevant NCEP predictors selected for downscaling daily temperature and precipitation and the corresponding explained variances for the data at each of the five stations in Northern Canada. R^2 corresponds to explained variance. The corresponding number for each predictor is given in Table 2

Station	T_{max}		T_{min}		Precipitation	
	Predictors	R^2	Predictors	R^2	Predictors	R^2
Cape Dorset	11, 12, 19, 24	0.52	11, 12, 19, 24	0.53	1, 3, 5, 16, 23, 24	0.22
Resolute Bay	11, 12, 19, 24	0.54	11, 12, 19, 24	0.53	1, 9, 12, 16, 23, 24	0.17
Shepherd Bay	11, 12, 19, 24	0.59	11, 12, 19, 24	0.59	1, 5, 9, 12, 16, 23	0.11
Iqaluit	9, 12, 19, 24	0.53	11, 12, 19, 24	0.52	1, 12, 16, 21, 23, 24	0.23
Inuvik	11, 12, 19, 24	0.54	9, 11, 12, 24	0.56	1, 9, 12, 18, 23, 24	0.13

(e.g. Wilby and Wigley, 2000; Wilby et al., 2003; Gachon et al., 2005). However, the exclusion of the near surface temperature and specific/relative humidity from the list of potential predictors also contributed to the relatively low performance of the downscaling model (for example, the explained variance for temperature in this study for northern stations is around 0.55 compared to typical values of 0.75 in Gachon et al., 2005). Nevertheless, the significant difference between the NCEP and GCMs values of these predictors justifies their exclusion in order to avoid a considerable reduction in downscaling performance when SDSM is supplied with the corresponding GCMs predictors.

3.3 Uncertainty analysis

To have confidence on the climate scenarios downscaled from GCM outputs, one has to be at least convinced that the downscaled outputs can represent the current state of the temperature and precipitation regimes reasonably well. In other words, the downscaling outputs ability to represent the baseline climate is a necessary condition (not sufficient) to have reasonable confidence on the reliability of the climate change anomalies computed from the scenarios runs. The aim of the uncertainty analysis is, therefore, to evaluate the performance of the downscaling method in reproducing the mean value and variability of observed meteorological variables when provided with climate predictors for the baseline period.

Two complementary methods have been employed to analyze the uncertainty of the output of the statistical downscaling model, namely hypothesis testing and confidence intervals.

3.3.1 Hypothesis testing

The hypothesis testing method used in this study is the Wilcoxon Signed Rank test (Wilcoxon, 1945) which is a non-parametric method used to test the null hypothesis of no median difference in paired samples. It requires calculating the test statistic and p value for the null hypothesis, and either accept or reject the hypothesis at a given significant level α based on the p value. The p value is the probability of wrongly rejecting the null hypothesis if it is in fact true (type 1 error). An α value of 0.05 which corresponds to 5% significance level is used in this study. Small p-values suggest that the null hypothesis is unlikely to be true and the null hypothesis is rejected when $p < 0.05$. In this study, the analysis was performed by comparing observed climate variables with the corresponding downscaled variables. As three different set of SDSM simulations have been considered with predictor variables derived from NCEP, CGCM2 and HadCM3, and with 100 simulations for each case, the hypothesis tests are also repeated 100 times for each seasonal or monthly variable from which we calculate the rejection percentage of all simulated values.

3.3.2 Confidence intervals

A confidence interval is a measure of uncertainty regarding the true value of a statistics or estimate. Resampling is used to estimate confidence intervals for the statistics of a distribution. In the present analysis, the non-parametric technique known as Bootstrap simulation (Efron and Tibshirani, 1993) is used to estimate confidence intervals. Bootstrapping is used to generate a pseudo population of a test statistic by re-sampling from the original data set. In essence, the bootstrap method takes random samples, known as pseudo-samples, with replacement from the original one repeatedly. The statistic in question (mean values and standard deviations in our case) is then calculated for each pseudo-sample.

In this analysis, each time series of observed climate variable is paired with each of the corresponding downscaled variable taken from the one hundred SDSM simulations. One thousand boot strap samples are generated from each such pairs to calculate two sets of statistic for each of the bootstrap sample pairs, namely the differences between the *mean* and the *standard deviation* values of the observed and simulated variables. Then the bootstrap percentile method is used to calculate the confidence interval by ranking the one thousand statistics calculated from the bootstrap samples and selecting the appropriate percentile (5^{th} percentile for the lower confidence limit and 95^{th} percentile for the upper confidence limit) for the confidence interval required (90% in our case). This procedure is repeated for each of the one hundred SDSM simulation outputs. The overall upper bound and lower bound of the statistic's 90% confidence interval is then calculated by averaging the upper bound and lower bounds of the hundred simulations, respectively. All the above steps are repeated for the three sets of SDSM simulations driven from the three sets of predictors. The best simulation result is identified as the one with the smallest confidence interval and which includes zero between its upper and lower confidence limits.

4. Results

First, and before starting the uncertainty analysis, the $T_{\max}$, $T_{\min}$, and precipitation values downscaled from NCEP, CGCM2 and HadCM3 predictors are analyzed in terms of their basic distribution, mean and median values and of their variability in comparison with the observed time series. The downscaling results corresponding to the NCEP predictors are shown over the two separate periods of calibration and validation (i.e. 1961–1990 and 1991–2000, respectively). Box-plots of the mean values present graphically the simulated and observed distributions of the monthly climate statistics calculated for each year, both for the calibration and for the validation periods (Figs. 2 and 3). Such box-plots are used to assess the performance based on the monthly distribution of the one hundred SDSM simulations compared to the monthly distribution of the corresponding observed values. The size of the box indicates the spread around the median and the outliers outside the 1.5 IQR (inter quantile range) limits and gives a useful indication around the scattering of each simulated series compared to the observed ones (Figs. 2 and 3 for an example at one station). The downscaling results corresponding to the two series of GCMs predictors over the 1961–1990 baseline period are also presented. The probability density function plots at seasonal scale (Fig. 4) presents how good the downscaled data reproduced the statistical distribution of the observed data including the extremes in using both the NCEP and GCMs predictors. Histograms of monthly mean biases (MB) for $T_{\max}$ and $T_{\min}$ are also presented (Figs. 5 and 6) to show the models biases at one of the climate stations considered (e.g. Cape Dorset). Moreover, quantile–quantile plots and histograms of monthly mean errors of precipitation at the same station are presented in Figs. 7 and 8. The results of all the five stations are summarised by plotting the monthly biases at each of the five stations in Fig. 9. The mean and relative biases (MB and RB for temperature and precipitation, respectively) are computed from observed and simulated climate variables and averaged over the100 simulations, as follows:

$$MB = \frac{1}{n}\sum_{i=1}^{n}(X_{\mathrm{Est},i} - X_{\mathrm{Obs},i}) \tag{1}$$

or

$$RB = \frac{1}{n}\sum_{i=1}^{n}((X_{\mathrm{Est},i} - X_{\mathrm{Obs},i}) * 100/X_{\mathrm{Obs},i}) \tag{2}$$

where $X_{\mathrm{Obs},i}$ is the i^{th} observed value (monthly), $X_{\mathrm{Est},i}$ is the i^{th} estimated/simulated value (monthly), and n is the number of data.

4.1 Downscaling results corresponding to NCEP predictors

The following results focus on comparing the monthly mean values between the observed and simulated results for the calibration (1961–1990) and validation (1991–2000) periods. In general, the SDSM downscaling results for daily $T_{\max}$ and $T_{\min}$ at each of the five stations reproduce the

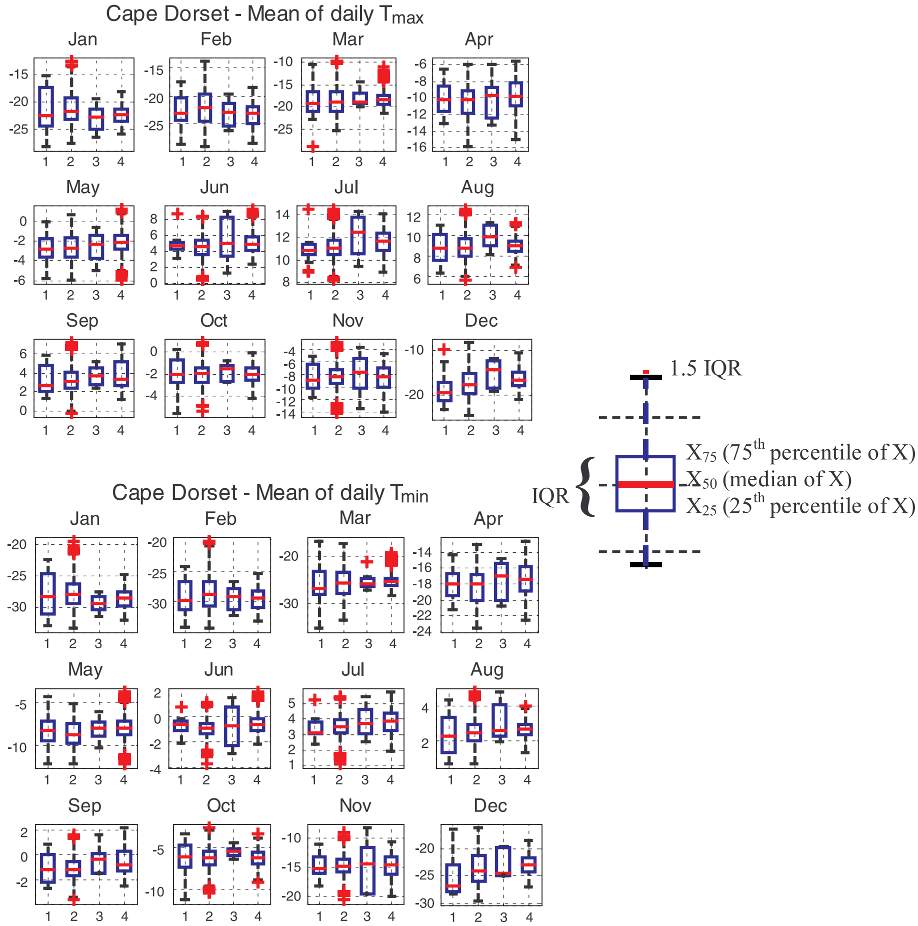

Fig. 2. Box-plots of SDSM downscaling results for daily T_{max} and T_{min} with NCEP predictors at Cape Dorset: *1* observed calibration (1961–1990), *2* downscaled calibration (1961–1990), *3* observed validation (1991–2000), and *4* downscaled validation (1991–2000)

observed values reasonably well. As representative example, box-plots of the monthly mean values of T_{max} and T_{min} at Cape Dorset are presented in Fig. 2, both for the calibration and validation periods. The plots clearly show that the performance of the SDSM model using NCEP predictors is very good and it is almost as good over the validation period as it is over the calibration one, except where huge changes in the observed climate regime appear between the two periods (as for June, July, August, and December for the IQR values) inducing some discrepancies in IQR for the validation period. But, the median values are well reproduced even for these particular months (i.e. SDSM is able to capture the change in median values of T_{max} and T_{min}), and the changes in all other months are well captured by the SD model. This suggests that the model is able to cope with such a change in atmospheric predictors, which results in a corresponding change in local predictand as well. Similar results are obtained for the downscaled data at the other four stations (Iqaluit, Inuvik, Resolute Bay and Shepherd Bay) which are not presented here.

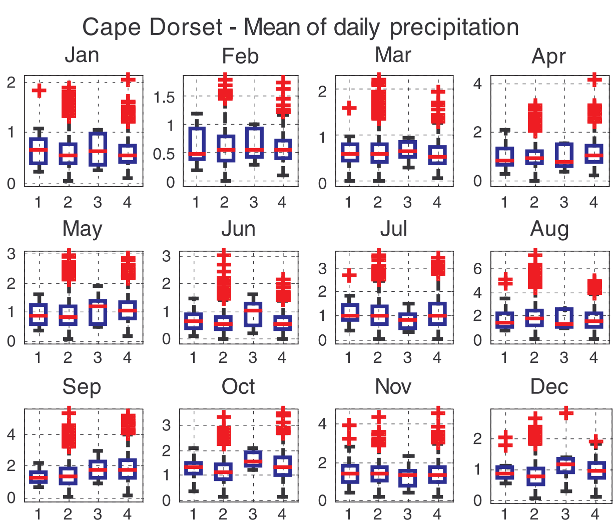

Fig. 3. Box-plots of SDSM downscaling results for daily precipitation with NCEP predictors at Cape Dorset: *1* observed calibration (1961–1990), *2* downscaled calibration (1961–1990), *3* observed validation (1991–2000), and *4* downscaled validation (1991–2000)

As shown in Fig. 3 for the Cape Dorset station, SDSM is also able to reproduce the median characteristics of monthly mean precipitation

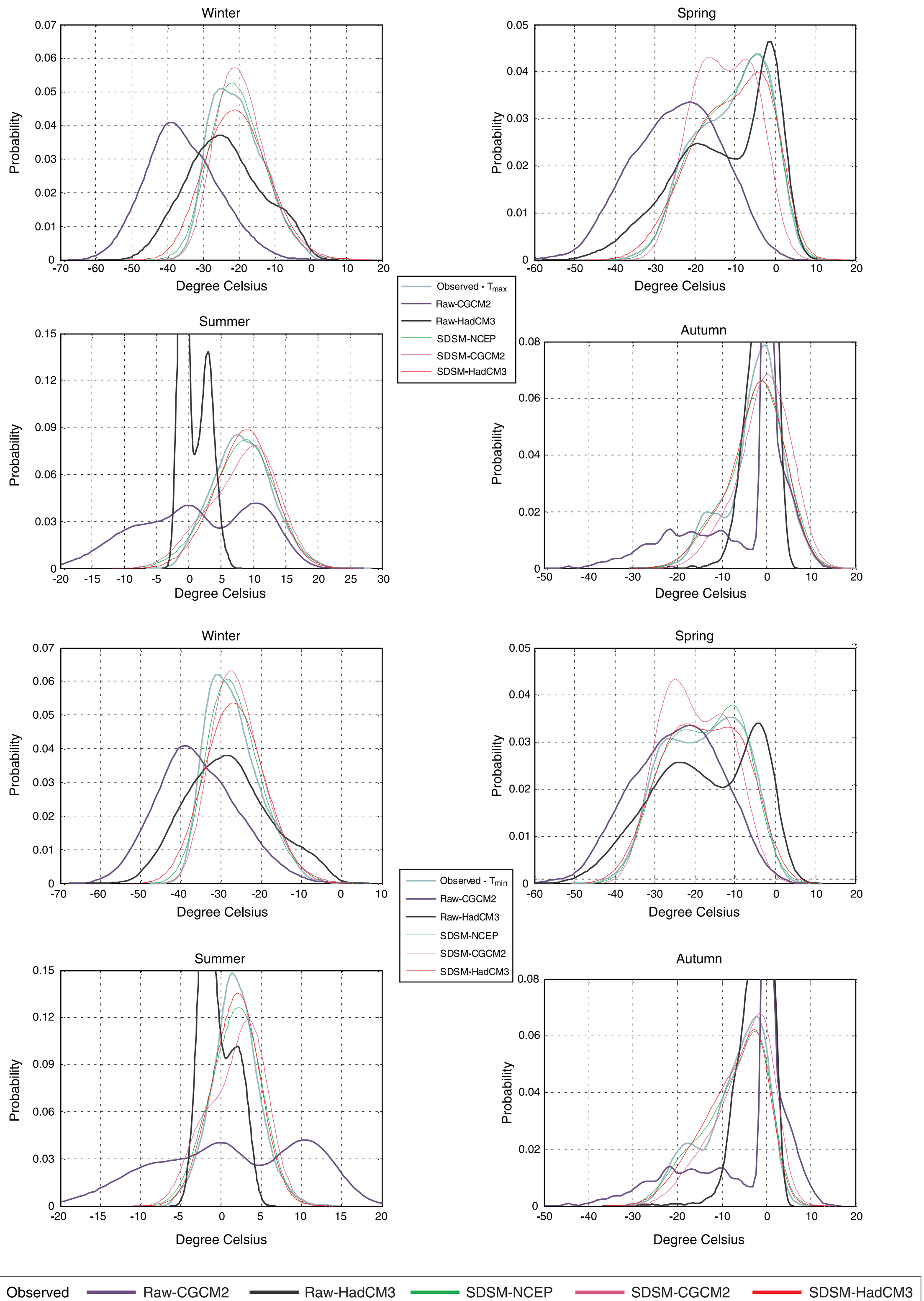

Fig. 4. Empirical seasonal probability density functions of (**a**) $T_{\max}$ and (**b**) $T_{\min}$ for the observed as well as the raw-GCMs, and SDSM downscaled data for the baseline period (1961–1990) at Cape Dorset

reasonably well. However, the SDSM model produces many more outliers than observed series, both in the calibration and in the validation periods. Even if the distribution of the observed values for almost all stations shows changes in the precipitation regime between the calibration

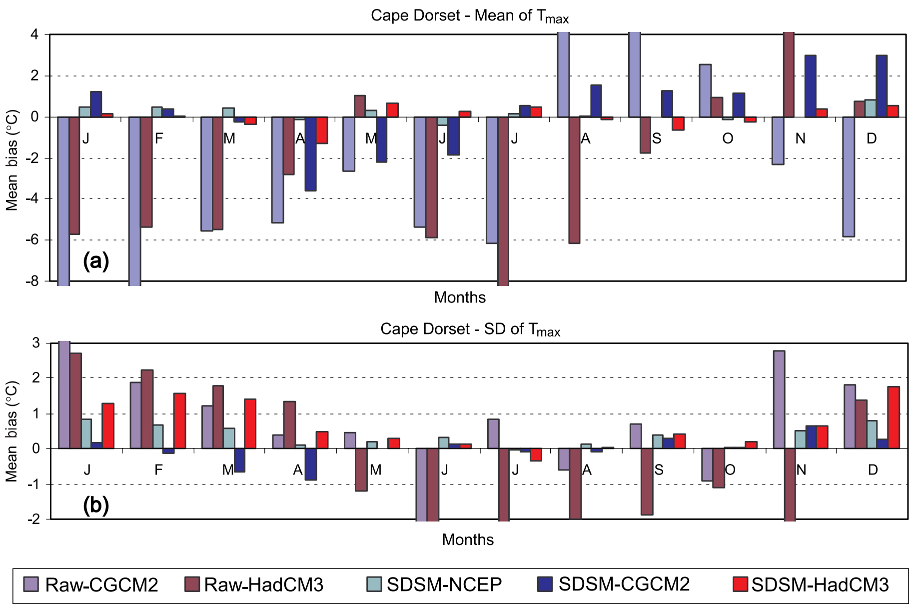

Fig. 5. Histograms of MB between (**a**) the monthly mean values and (**b**) standard deviation of observed data and the corresponding raw-GCMs and SDSM downscaled data of $T_{\max}$ (in °C) at Cape Dorset over the baseline period (1961–1990)

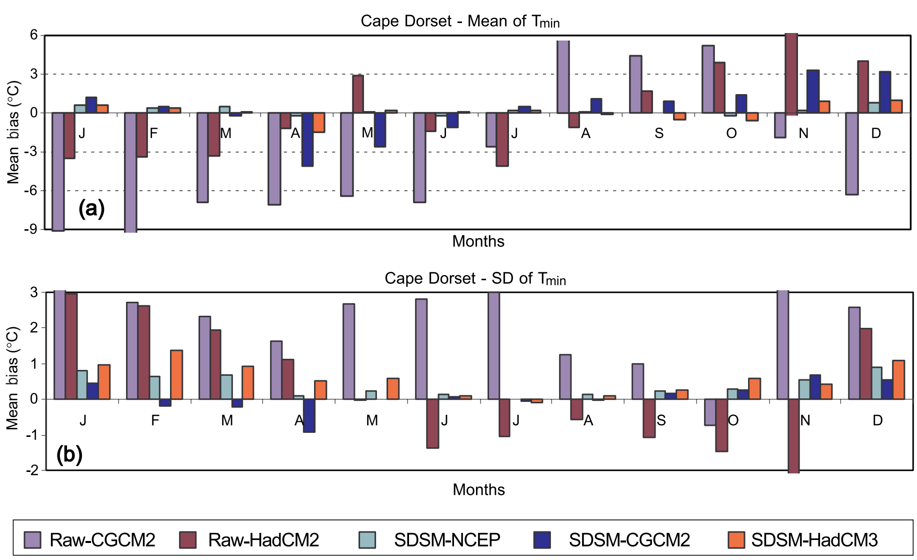

Fig. 6. Histograms of MB between (**a**) the monthly mean values and (**b**) standard deviation of observed data and the corresponding raw-GCMs and SDSM downscaled data of $T_{\min}$ (in °C) at Cape Dorset over the baseline period (1961–1990)

and validation periods, the main features of the monthly mean precipitation at each climate station are well reproduced by the NCEP predictors based SDSM downscaling results.

4.2 Downscaling result corresponding to GCMs predictors

Ensembles of synthetic daily weather series are generated based on the regression parameters calibrated with the predictor variables identified earlier (see Table 3), but in this case derived from the CGCM2 and HadCM3 simulations over the baseline 1961–1990 period. These downscaling results are compared with the raw-GCMs near surface temperature and precipitation outputs so that the value gained by downscaling is evaluated and compared to observed climatology. The downscaled results corresponding to the NCEP predictors are also included as a reference to evaluate the corresponding downscaled values based on GCMs ones.

4.2.1 Temperature downscaling

Empirical probability density functions (PDF) at seasonal scale are used to compare the raw-GCMs and SDSM downscaled $T_{\max}$ and $T_{\min}$ with the observed statistical distribution. These PDFs are shown in Fig. 4 for Cape Dorset station

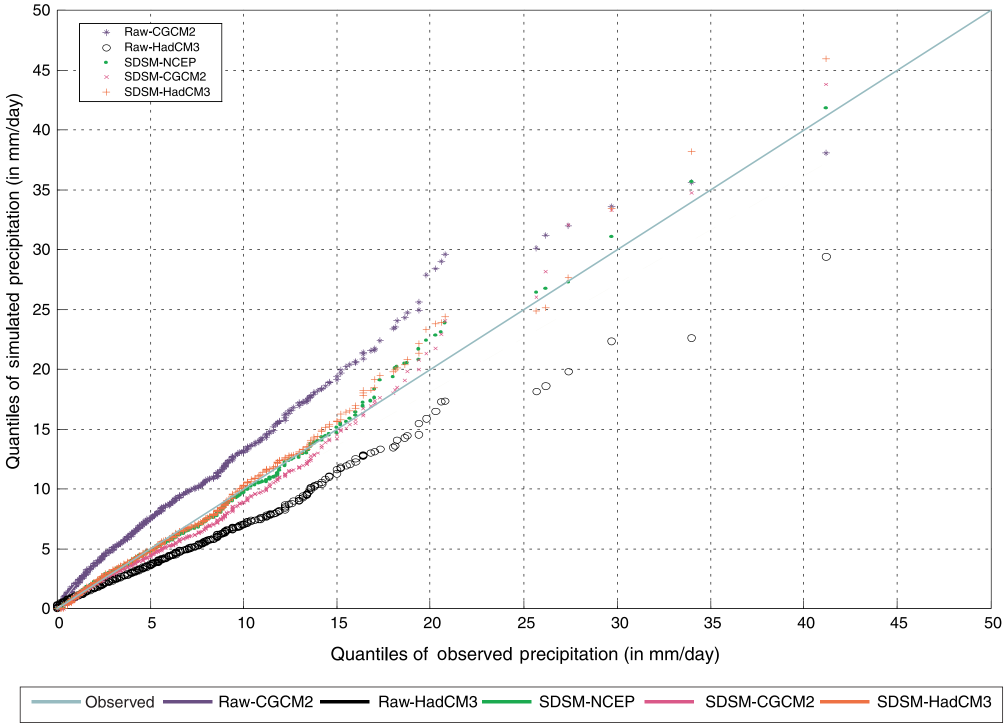

Fig. 7. Empirical quantile–quantile plots for the observed as well as the raw-GCM and downscaled precipitation data (in mm/day) for the baseline period (1961–1990) at Cape Dorset

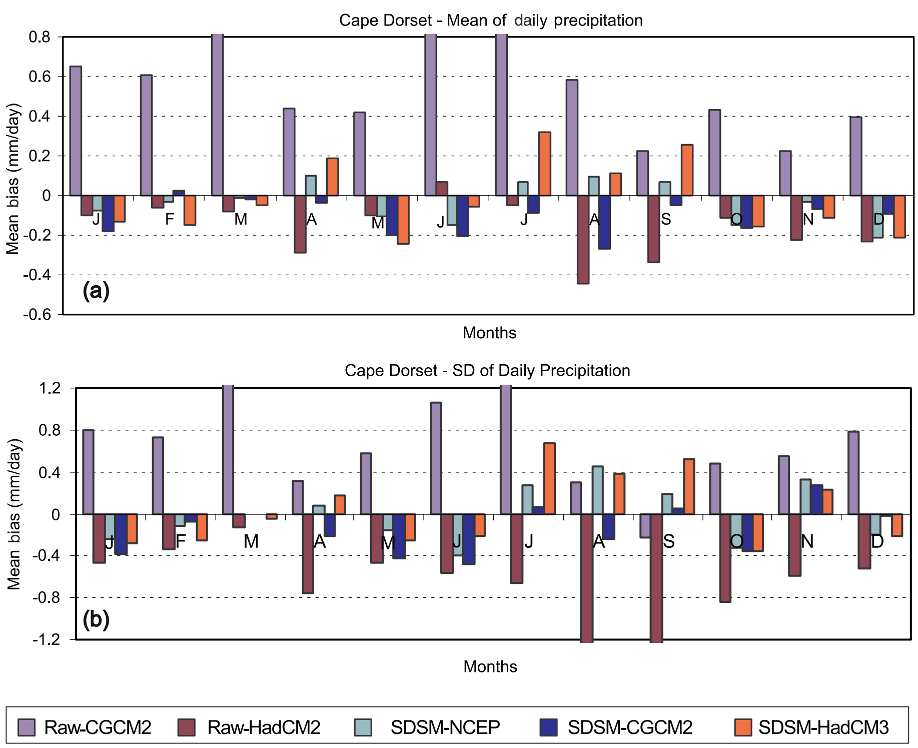

Fig. 8. Histograms of MB between (**a**) the monthly mean values and (**b**) standard deviation of observed data and the corresponding raw-GCMs and SDSM downscaled data of monthly mean precipitation (in mm/day) at Cape Dorset over the baseline period (1961–1990)

over the 1961–1990 period. Those clearly confirm the problems associated with the distribution of raw temperature data from GCMs, especially in simulating variability and extreme values with unrealistic high frequencies of temperature values near 0 °C over spring, summer and fall seasons. Also, the GCM raw data even exhibited a bimodal distribution both for minimum and maximum

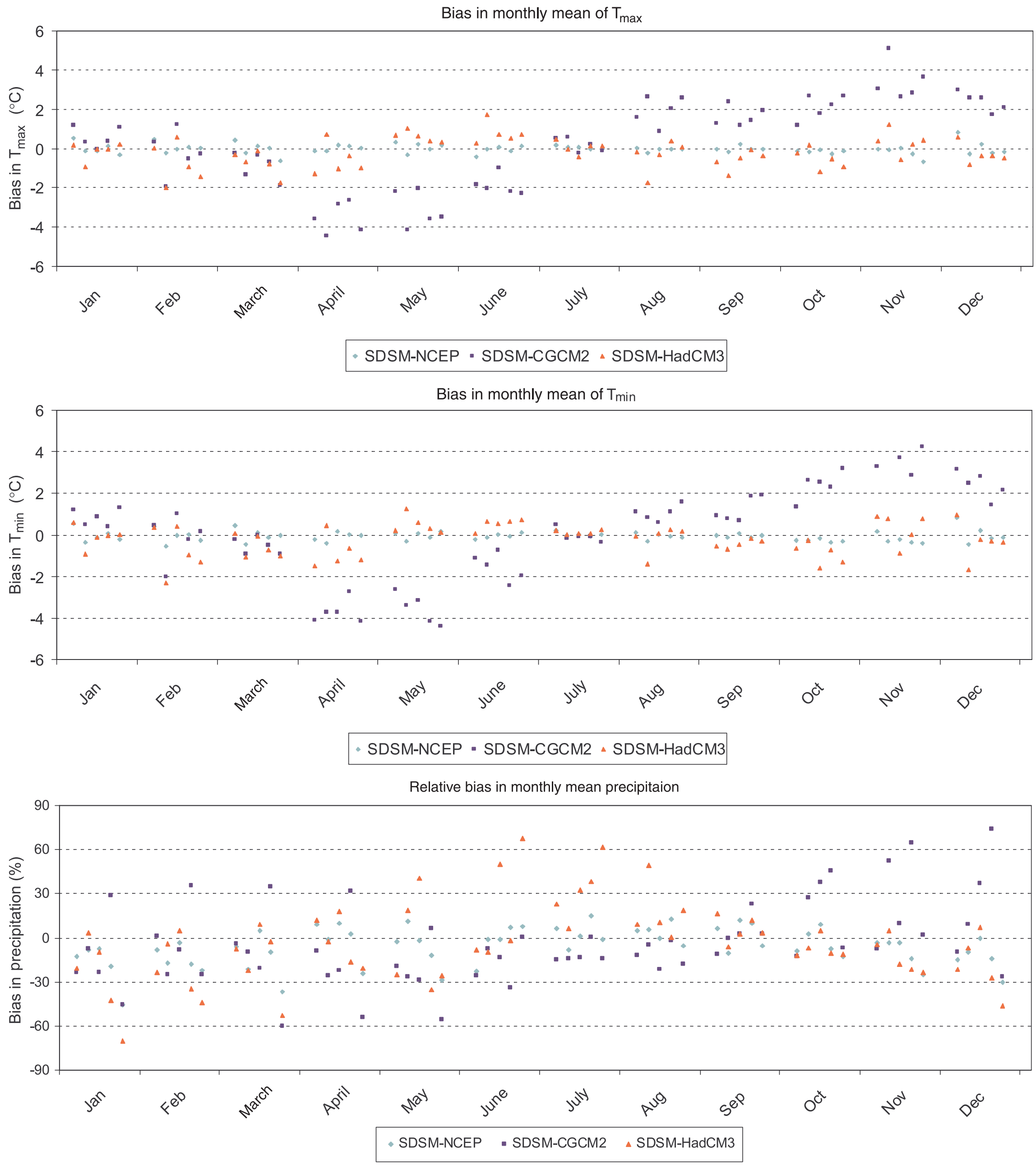

Fig. 9. Monthly absolute biases of T_{max}, T_{min}, and monthly relative biases of precipitation at each of the five climate stations for the baseline period (1961–1990). Cape Dorset, Inuvik, Iqaluit, Resolute Bay and Shepherd Bay correspond to each symbol from left to right for each month, respectively

temperature which is not observed in the study area (i.e. systematic for all the five stations, not shown here). These biases do not only result from coarse scale resolution of the two GCMs and/or from the land/sea surface conditions of the GCM grid point(s) closer to the predictands, but also are issued from the surface physical processes parameterized in the GCMs. In that case, the oceanic and/or land surface conditions are not well reproduced by the two GCMs during the frost-thaw period inducing a strong shift and delay in the overlying simulated air temperatures from the surface diabatic fluxes (i.e. radiative, latent and sensible heat) as suggested for example over ice free conditions in Hudson Bay in Gachon et al. (2003). The results also show that the downscaling has definitely improved the simulated distribution of T_{max} and T_{min} which are (after downscaling) more close to the observed data than the raw-GCMs outputs. Similar results to the one shown in Fig. 4 are also obtained for the remaining four climate stations and are generally more accurate with HadCM3 predictors than those obtained with CGCM2 ones.

Histograms of mean biases (MB) are also plotted for the monthly mean and standard deviation values of the climate variables to show the monthly biases associated with both the raw-GCMs and the SDSM downscaled data at each of the five climate stations. Because of the close similarity between the results corresponding to the different climate stations, the histograms of only Cape Dorset are presented hereafter. Figures 5 and 6 reveal that T_{max} and T_{min} data simulated by the two GCMs have strong biases (in terms of monthly mean and standard deviation) for most of the months. The raw-GCM data shows a warm bias for the autumn months while the rest of the seasons have cold bias, as also suggested by the seasonal PDFs. In general, the monthly biases in the raw CGCM2 temperature are higher than those corresponding to HadCM3. Over all the five climate stations, the CGCM2 monthly temperature biases range between 10 and 30 °C (with the highest biases being in summer and winter seasons at Inuvik, Iqaluit and Shepherd Bay) while that of HadCM3 are in the order of 5–10 °C, suggesting a more systematic problem with the surface processes representation in the CGCM2 model compared to HadCM3 (as suggested also in the CGCM1 model studied over the Hudson Bay area in Gachon et al., 2005). Moreover, the monthly errors in standard deviation between the observed and simulated values shown in Figs. 5 and 6 reveals that the raw CGCM2 overestimates the temperature variability for almost all seasons, while HadCM3 underestimates the variability for summer and autumn seasons, and overestimates for winter and spring seasons. As noted in the seasonal PDF curves, Figs. 5 and 6 also shows that the downscaling has definitely improved the GCM outputs by strongly reducing the temperature biases compared to the raw-GCM, in terms of both monthly mean and standard deviation values. In general, while the downscaling done with the NCEP predictors give the best agreement with the observed data in terms of mean monthly temperature and their standard deviations, the downscaling from the HadCM3 also gave very good results. However, the downscaled data from CGCM2 still contain some negative biases in spring season and positive biases in the autumn months comparatively larger that the one downscaled from HadCM3. This is illustrated in more detail with the uncertainty analysis results presented in Sect. 4.3.

4.2.2 Precipitation downscaling

Figure 7 shows the empirical quantile–quantile plot for daily precipitation at Cape Dorset with observed as well as the raw-GCMs and SDSM downscaled data. The GCM-driven downscaling results have definitely improved the frequency distribution in reducing the over and underestimation of medium range and higher quantiles of daily precipitation, compared to the raw CGCM2 and HadCM3, respectively. Similar results are also obtained for the remaining four climate stations (not shown). The histograms of MB for the monthly mean and standard deviation values of precipitation at Cape Dorset are shown in Fig. 8. Those reveal that the raw precipitation data simulated by CGCM2 has shown strong bias for most months, as this GCM overestimates the precipitation for most of the year. On the contrary, the raw HadCM3 precipitation data is more close to the observed with smaller negative biases for few months. While the raw-CGCM2 overestimates the precipitation variability, the raw-HadCM3 underestimate this values for almost all seasons. The results in Fig. 8 also show that the statistical downscaling has definitely improved the GCMs outputs by reducing the biases observed in the monthly mean and standard deviations of the raw-CGCM2 precipitation data. It has also improved the HadCM3 data especially for those months and climate stations where the raw HadCM3 has stronger biases. Surprisingly, there is no significant difference between the three sets of downscaled precipitation series which uses the NCEP or the two GCMs predictors, as also noted in Gachon et al. (2005).

Figure 9 summarizes the statistical downscaling performances for T_{max}, T_{min} and precipitation over all the five climate stations considered in this study. The figure presents absolute biases for T_{max} and T_{min} and relative biases for precipitation, on monthly basis and for each station. The

Table 4. Rejected percentage of the null hypothesis (in Wilcoxon Signed Rank test at 5% significance level by season) that median difference in observed and SDSM downscaled data of T_{max} or T_{min} is zero. NCEP, CGCM2 and HadCM3 correspond to downscaled values with NCEP, CGCM2, and HadCM3 predictors, respectively

	Winter			Spring			Summer			Autumn		
	NCEP	CGCM2	HadCM3	NCEP	CGCM2	HadCM3	NCEP	CGCM2	HadCM3	NCEP	CGCM2	HadCM3
T_{max}												
Cape Dorset	0	33	0	0	67	33	0	69	4	0	100	2
Inuvik	0	63	26	0	71	29	0	71	67	0	100	45
Iqaluit	0	32	0	0	67	17	0	66	34	0	100	39
Resolute Bay	0	33	0	0	68	0	0	67	4	1	100	0
Shepherd Bay	10	47	0	1	92	35	1	67	7	0	100	2
T_{min}												
Cape Dorset	0	39	0	1	67	33	1	75	0	0	98	5
Inuvik	4	62	50	1	67	33	1	64	46	3	77	11
Iqaluit	0	33	0	0	67	9	0	67	26	0	100	57
Resolute Bay	0	33	0	1	67	6	0	67	20	4	100	2
Shepherd Bay	0	67	5	0	91	48	0	71	17	0	100	14

Table 5. Rejected percentage of the null hypothesis by seasons (in Wilcoxon Signed Rank test at 5% significance level) that median difference in observed and SDSM downscaled precipitation data is zero. NCEP, CGCM2 and HadCM3 correspond to downscaled values with NCEP, CGCM2, and HadCM3 predictors, respectively

	Winter			Spring			Summer			Autumn		
	NCEP	CGCM2	HadCM3	NCEP	CGCM2	HadCM3	NCEP	CGCM2	HadCM3	NCEP	CGCM2	HadCM3
Cape Dorset	9	8	23	2	7	29	14	10	8	9	14	9
Inuvik	15	23	0	12	12	10	3	3	34	4	63	3
Iqaluit	6	30	0	5	31	34	1	30	62	10	34	9
Resolute Bay	30	86	62	15	38	22	27	24	28	34	72	10
Shepherd Bay	20	13	65	12	91	19	2	8	63	4	0	3

results for T_{max} and T_{min} show that the ones downscaled from NCEP give the best performance. The ones downscaled from HadCM3 also have smaller biases in most of the cases, except some overestimation of T_{max} and T_{min} in May and June and underestimation in March, April, September and October. However, the temperature data downscaled from CGCM2 show consis-

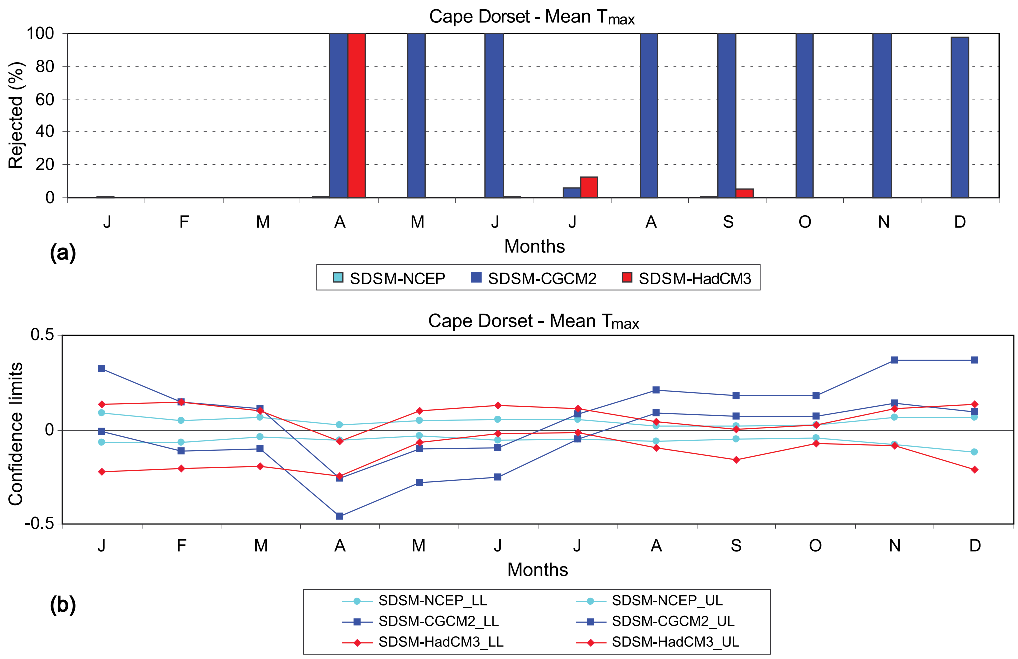

Fig. 10. Histograms of rejected percentage in Wilcoxon Signed Rank hypothesis test that the difference between observed and downscaled monthly mean values of T_{max} (in °C) is zero (top) and the corresponding 90% confidence intervals (bottom, UL means upper limit and LL means lower limit) over the baseline period at Cape Dorset

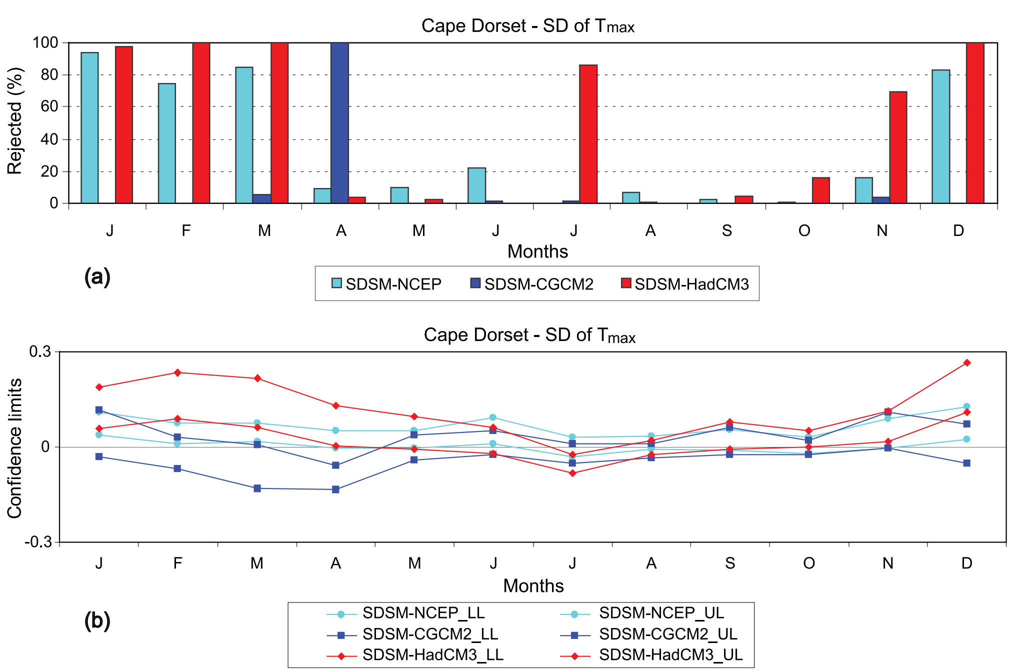

Fig. 11. Histograms of rejected percentage in Wilcoxon Signed Rank hypothesis test that the difference between observed and downscaled monthly standard deviation values of T_{max} (in °C) is zero (top) and the corresponding 90% confidence intervals (bottom, UL means upper limit and LL means lower limit) over the baseline period at Cape Dorset

tent biases for most of the stations and for most months except January, February and March. On the other hand, the results of the precipitation downscaling show that while the biases associated with that downscaled from NCEP are relatively smaller, the ones downscaled from CGCM2 and HadCM3 show more or less similar results. There is also no major difference in the precipitation downscaling performances from season to season, or from station to station.

4.3 Results of the uncertainty analysis

The results presented below mainly focus on the uncertainty analysis of the SDSM downscaled temperature and precipitation data, with a special emphasis on the GCM-driven results. The rejected percentage in Wilcoxon Signed Rank hypothesis test is presented on a seasonal basis in Tables 4 and 5 for all the stations. The monthly values of rejected percentages and the corresponding confidence intervals are also presented on Figs. 10–15, but only for the Cape Dorset station.

4.3.1 Uncertainty in downscaling $T_{\max}$

Figure 10a shows histograms of rejected percentage in Wilcoxon Signed Rank test that median difference in observed and SDSM downscaled data of monthly mean values of $T_{\max}$ at Cape Dorset is zero, while Fig. 10b shows the corresponding 90% confidence interval. For $T_{\max}$ downscaled from NCEP and HadCM3 predictors, the rejected percentage (at 5% significance level) in each month is either zero or very closes to zero except in April. However, for $T_{\max}$ downscaled from CGCM2 predictors, the rejected percentage is over 95% for most of the months except January, February, March and July (see also in Table 4 for all the stations at seasonal scale). The 90% confidence interval in Fig. 10b also shows that while most of the confidence intervals calculated from the downscaled data based on NCEP and HadCM3 predictors contains zero for almost all months, the downscaled results from CGCM2 contain zero for only few months (January, February, March and July). For the rest of the year, the confidence intervals are on either side of zero. This confirms the problem that exists in the data downscaled from CGCM2 predictors in capturing the monthly mean values in most of the stations, as also suggested by the Wilcoxon Signed Rank test. The figures also show that data downscaled from NCEP predictors resulted in smaller confidence interval (mostly between -0.1 and $+0.1\,^{\circ}\mathrm{C}$) than the one downscaled from HadCM3 (mostly between -0.2 and $+0.2\,^{\circ}\mathrm{C}$). This, of course, is not a surprise as the downscaling model is calibrated with NCEP predictors which are reanalysis products. One can also see from the figure that the confidence intervals are smaller during summer months.

Similarly, Fig. 11a shows the histograms for the rejected percentage in Wilcoxon Signed Rank test that median difference in observed and SDSM downscaled data of monthly values of the standard deviation of $T_{\max}$ is zero. Here the histograms show mixed pictures where the hypothesis is rejected at different level for the different cases, and the standard deviation of large proportion of data downscaled from NCEP and HadCM3 were rejected for the winter and spring months. Figure 11b shows the corresponding 90% confidence interval for the difference between observed and downscaled monthly standard deviations of $T_{\max}$. The figures reveal that while some of the confidence intervals calculated from the data downscaled from NCEP predictors have the smallest confidence interval, it has not contain zero for most of the winter and spring months showing overestimation of the variability during these seasons. In general, the data downscaled from CGCM2 seems to better represent the variability of the $T_{\max}$, than the one downscaled from HadCM3.

4.3.2 Uncertainty in downscaling $T_{\min}$

Figure 12a shows histograms of rejected percentage in Wilcoxon Signed Rank test that median difference in observed and SDSM downscaled data of monthly values of $T_{\min}$ at Cape Dorset is zero, while Fig. 12b shows the corresponding 90% confidence interval. The results for $T_{\min}$ are quite similar to the one for $T_{\max}$ in that, the rejected percentage for each month (at 5% significance level) of $T_{\min}$ downscaled from NCEP and HadCM3 predictors is either zero or very closes to zero except in April. However, for $T_{\min}$ downscaled from CGCM2 predictors, the rejected percentage is over 95% for most of the

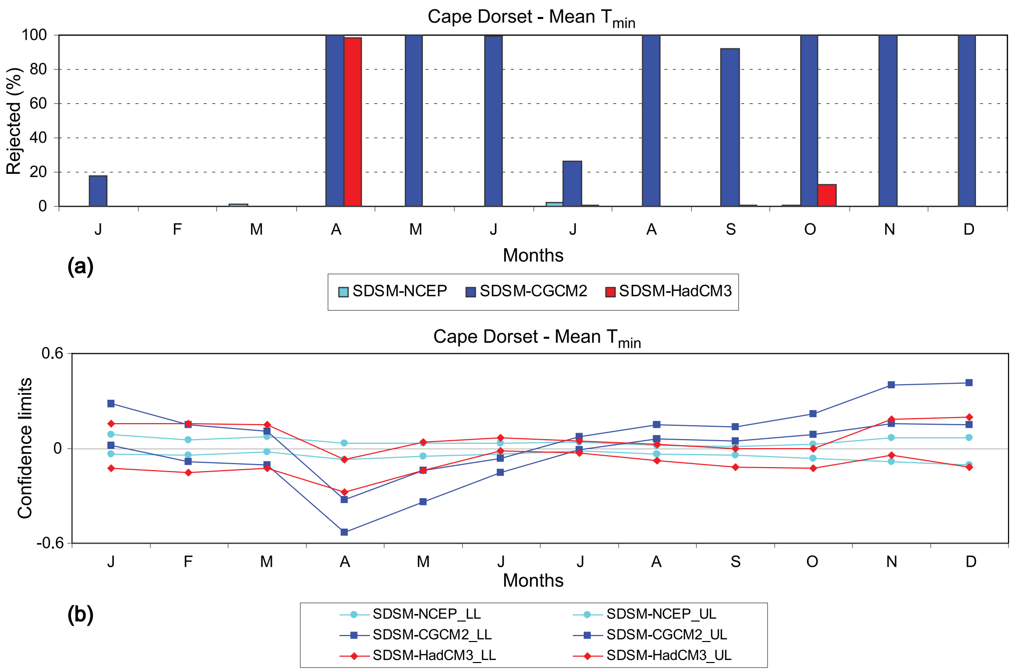

Fig. 12. Histograms of rejected percentage in Wilcoxon Signed Rank hypothesis test that the difference between observed and downscaled monthly mean values of T_{min} (in °C) is zero (top) and the corresponding 90% confidence intervals (bottom, UL means upper limit and LL means lower limit) over the baseline period at Cape Dorset

months, except January, February, March and July (see also in Table 4 for all the stations at seasonal scale). The corresponding 90% confidence interval in Fig. 12b for the difference between the observed and downscaled monthly mean values of T_{min} shows that while most of the confidence intervals calculated from the downscaled data based on NCEP and HadCM3 predictors contains

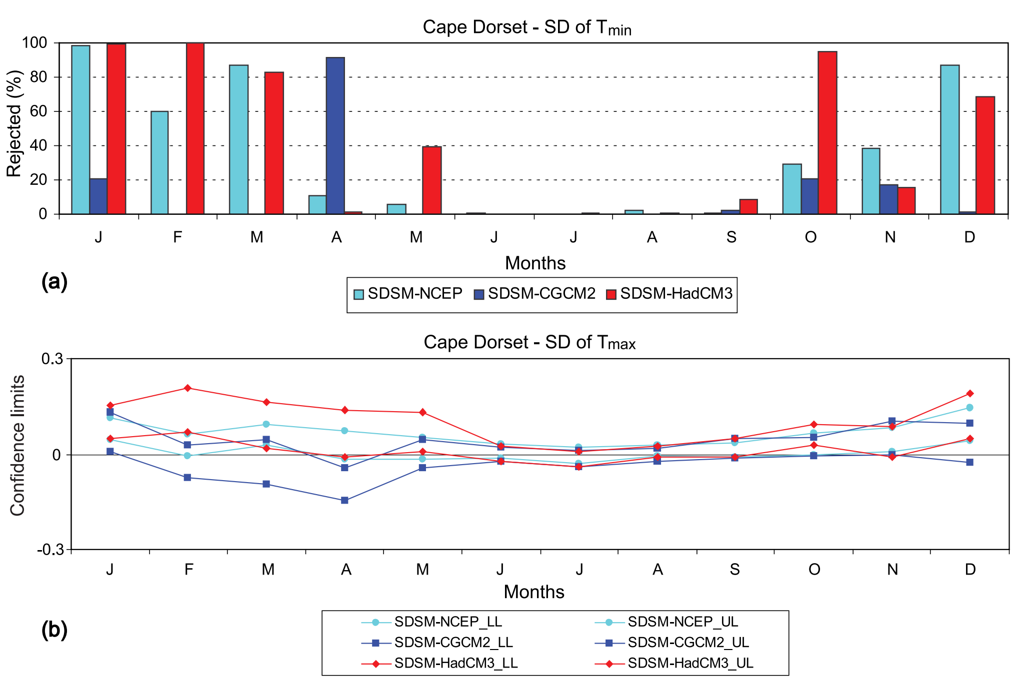

Fig. 13. Histograms of rejected percentage in Wilcoxon Signed Rank hypothesis test that the difference between observed and downscaled monthly standard deviation values of T_{min} (in °C) is zero (top) and the corresponding 90% confidence intervals (bottom, UL means upper limit and LL means lower limit) over the baseline period at Cape Dorset

zero for almost all months, the downscaled results from CGCM2 contain zero for only few months (January, February, March and July). For the rest of the year, the confidence intervals are on either side of zero. Figure 13a, on the other hand, shows the histograms for the rejected

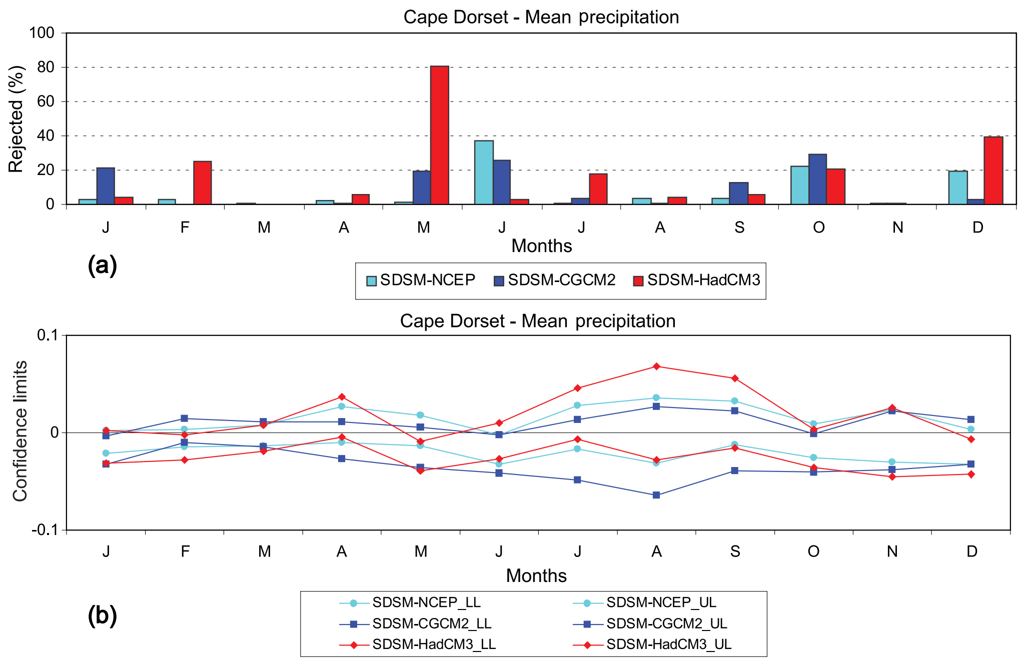

Fig. 14. Histograms of rejected percentage in Wilcoxon Signed Rank hypothesis test that the difference between observed and downscaled monthly mean precipitation (in mm/day) is zero (top) and the corresponding 90% confidence intervals (bottom, UL means upper limit and LL means lower limit) over the baseline period at Cape Dorset

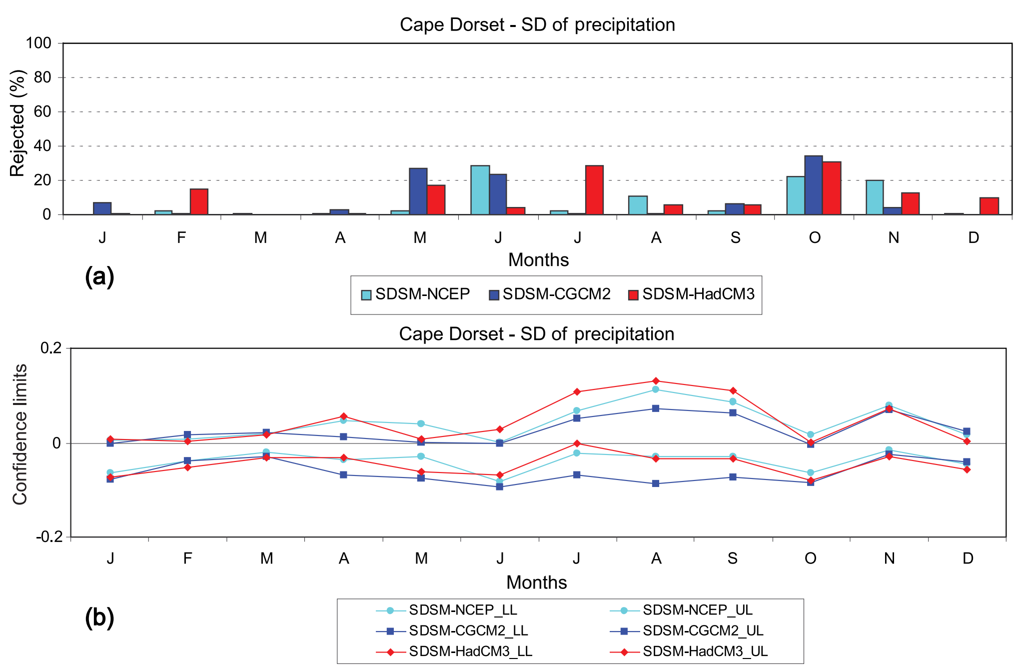

Fig. 15. Histograms of rejected percentage in Wilcoxon Signed Rank hypothesis test that the difference between observed and downscaled monthly standard deviation of precipitation (in mm/day) is zero (top) and the corresponding 90% confidence intervals (bottom, UL means upper limit and LL means lower limit) over the baseline period at Cape Dorset

percentage in Wilcoxon Signed Rank test that median difference in observed and SDSM downscaled data of monthly values of the standard deviation of $T_{\min}$ is zero. Here the histograms show that the hypothesis for standard deviation is rejected in the winter and spring months for the data downscaled from NCEP and HadCM3. Figure 13b shows the corresponding 90% confidence interval for the difference between observed and downscaled monthly standard deviations of $T_{\min}$. As for $T_{\max}$, the data downscaled from CGCM2 seems to better represent the variability of the $T_{\min}$ than the one downscaled from HadCM3.

4.3.3 Uncertainty in downscaling precipitation

Figure 14a shows histograms of rejected percentage for monthly values of precipitation for Cape Dorset, while Table 5 presents the rejected percentage by season for all stations. In general, the rejected percentages are zero or close to zero for most of the cases. Precipitation downscaled with both CGCM2 and HadCM3 predictors have rejection of 20% or more for about four months, while those downscaled with NCEP has three months with similar rejection percentage. The rejection percentages seem to be higher during spring and autumn seasons for CGCM2 and during winter and summer seasons for HadCM3. Figure 14b shows the corresponding 90% confidence interval for the difference between the observed and downscaled monthly values of precipitation. Most of the confidence intervals calculated from the downscaled data contains zero. Similar to the case of temperature, the precipitation downscaled from NCEP predictors result in the smallest confidence interval (mostly between -0.03 and $+0.03$ mm/day), than the one downscaled from CGCM2 and HadCM3. The confidence intervals for the precipitation downscaled from CGCM2 are comparable to the one downscaled from HadCM3.

Figure 15a and b show similar test results for the monthly standard deviation of precipitation. In addition to showing similar level of uncertainties in terms of rejected percentage in the hypothesis testing, there seems to be no big difference in the magnitude of the confidence interval of the data downscaled from NCEP and the two GCMs. Most of the confidence intervals also include the zero value indicating that the downscaled precipitation has captured the variability in the observed precipitation.

5. Discussion

First of all, the study confirms that temperature data for the extreme northern part of Canada simulated by the two GCMs (CGCM2 and HadCM3) generally exhibit strong biases in terms of monthly mean and standard deviation. These discrepancies are not only due to incorrect reconstruction of the local climate regime in GCMs due to their coarse resolution, with inherent difficulties to reproduce the land/sea contrasts as for all areas studied here, but they are also due to the major problem in physical parameterization in most climate models. This problem is related to surface processes over heterogeneous conditions such as snow cover, sea ice and more generally, frost/thaw length characteristics of the soil, including water content and its phase distribution during the year (i.e. solid versus liquid phase). These processes have strong influence on surface energetic budget and the overlying air temperature and in particular over arctic and sub-arctic regions (e.g. Covey et al., 2000; Walsh et al., 2002; Barrow et al., 2004; Gachon et al., 2005). In general, the study shows that the regression-based statistical downscaling (with SDSM) is able to capture most of the precipitation and thermal regimes in northern Canada provided that the large scale climate predictors used in the process are well simulated by the GCMs considered. The downscaling exercise has definitely improved the GCMs outputs by reducing the biases found in the raw CGCM2 and HadCM3 values, and improving the variability. However, the data downscaled from CGCM2 still contain negative temperature biases during the spring and positive biases in the autumn, relatively larger than the one downscaled from HadCM3. The statistical downscaling is also able to reproduce the precipitation regime in general, and especially in terms of median characteristics of monthly mean precipitation and its standard deviation. However, inter-quantile ranges and higher percentiles are sometimes overestimated, and the downscaling produces many more outliers than observed series. Moreover, no significant difference is observed in the performance of SDSM in downscaling precipitation, either with the NCEP predictors or with the predictors derived from the two GCMs.

The uncertainty analysis shows that the percentage by which the null hypothesis (that there is no median difference between the observed and downscaled data) is rejected by Wilcoxon Signed Rank test (at 5% significance level) for almost all temperature downscaling with NCEP and HadCM3 predictors is relatively small. However, for temperatures downscaled from CGCM2 predictors, the percentage rejected is very high for most of the months except in January, February, March and July indicating a higher level of uncertainty associated with temperature downscaled with CGCM2 predictors. The inter-annual variability of the mean monthly values of temperature is slightly underestimated in the downscaled data irrespective of the source of predictors. In general, the confidence intervals are smaller during summer months, indicating a better predictability during this season. The uncertainty analysis for precipitation suggests that, in general, the rejected percentage is small for most of the cases, and there is no specific pattern in terms of locations or seasons. The confidence intervals for precipitation downscaled with CGCM2 predictors are better (smaller) than the ones downscaled from HadCM3, while those downscaled from NCEP predictors resulted in the smallest confidence interval. All the results confirm that the downscaling performance and predictability of the climate variables strongly vary with the source of predictors, with seasons and the location of the climate station considered for the analysis.

6. Conclusions

Confidence on future climate scenarios at a regional or local scale depends to a large extent on the ability of the downscaled data to reconstruct the observed climate regimes. Hence, this study has shown, in general, the potential of the regression-based statistical-downscaling technique in order to develop reliable climate information using GCM predictors, over highly heterogeneous surface conditions in northern coastlines of Canada. However, uncertainty analysis has reveals that, even after careful screening of the most relevant predictors, some of the downscaled data is still significantly different from the observed values. This has been specially the case for CGCM2-driven downscaling of temperature data over spring and fall seasons. With such uncertainties in the downscaled data for the baseline period, it will be difficult to have great confidence on climate variables downscaled for future climate scenarios in order to use them for a meaningful climate change impact studies. Nevertheless, the added values from downscaled results confirm that some form of downscaling (i.e. SDSM in our case) is still preferred to using directly raw-GCM outputs at the local scale, for impacts studies in northern Canada.

In general, the study presented in this paper shows one way of evaluating the reliability of the downscaled data, and the added values of downscaling using objective hypothesis testing criteria and evaluating confidence intervals to quantify the uncertainty associated with climate scenarios development, as recommended in various studies or guidelines (e.g. Goodess et al., 2003; Wilby et al., 2004). This approach may help to better identify GCM predictors and downscaling techniques which are more appropriate for generating reliable climate scenario information. However, further work may be needed to identify the more reliable GCMs and develop more relevant predictors which can be used effectively to better reconstruct the temperature and precipitation regime. For example, Gachon et al. (2005) have suggested that test with direct thermal advection terms responsible for the main part of the temperature regime in the Nordic regions, especially in northern Canada in winter, must be done in order not only to improve the variability but also to better reconstruct the tails of the temperature distribution. Moreover, the use of the Canadian Regional Climate Model outputs as candidate predictors for statistical downscaling should be explored, not only for the downscaling of temperature but also to improve the downscaling of precipitation as this variable needs mesoscale forcing and feedback better resolved in regional than in global climate models. Other statistical downscaling techniques should also be explored in conjunction with the use of RCM simulations to evaluate the uncertainties associated with the choice of downscaling techniques.

Acknowledgements

This study was completed with the financial support of Ouranos. The authors would like to thank Dimitri Parishkura, Nathalie Gauthier, Diane Chaumont and Marie Claude Simard for their valuable contribution at various stages of this research. We would like to acknowledge the Hadley Centre in UK and the Canadian Centre for Climate Modelling and Analysis (CCCma) for giving us access to daily values of HadCM3 and CGCM2 outputs, respectively. We are also grateful to Lucie Vincent and Eva Mekis from Environment Canada for providing observed data sets of homogenized temperature and rehabilitated precipitation.

References

Barrow EM, Maxwell B, Gachon P (eds) (2004) Climate variability and change in Canada: past, present and future. ACSD Science Assessment Series No. 2, Meteorological Service of Canada, Environment Canada, Toronto, Ontario, 114 pp

Christensen JH, Carter TR, Giorgi F (2002) PRUDENCE employs new methods to assess European climate change. EOS Transactions 83: 147

Coulibaly P, Dibike YB (2004) Downscaling of global climate model outputs for flood frequency analysis in the Saguenay river system. Sub-component, Climate Change Action Fund (CCAF), Environment Canada, 84 pp

Coulibaly P, Dibike YB, Anctil F (2005) Downscaling precipitation and temperature with temporal neural networks. J Hydrometeorol 6(4): 483–496

Covey C, Abe-Ouchi A, Boer GJ, Boville BA, Cubasch U, Fairhead L, Flato GM, Gordon H, Guilyardi E, Jiang X, Johns TC, Le Treut H, Madec G, Meehl GA, Miller R, Noda A, Power SB, Roeckner E, Russell G, Schneider EK, Stouffer RJ, Terray L, von Storch J-S (2000) The seasonal cycle in coupled ocean atmosphere general circulation models. Clim Dynam 16: 775–787

Dibike YB, Coulibaly P (2005) Hydrologic impact of climate change in the Saguenay Watershed: comparison of downscaling methods and hydrologic models. J Hydrol 307(1–4): 145–163

Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. New York: Chapman & Hall/CRC, 436 pp

Flato GM, Boer GJ (2001) Warming asymmetry in climate change simulations. Geophys Res Lett 28: 195–198

Flato GM, Boer GJ, Lee W, McFarlane N, Ramsden D, Weaver A (2000) The CCCma global coupled model and its climate. Clim Dynam 16: 451–467

Gachon P, Laprise R, Zwack P, Saucier JF (2003) The effects of interactions between surface forcings in the development of a model-simulated polar low in Hudson Bay. Tellus 55A: 61–87

Gachon P, St-Hilaire A, Ourda T, Nguyen VTV, Lin C, Milton J, Chaumont D, Goldstein J, Hessami M, Nguyen TD, Selva F, Nadeau M, Roy P, Parishkura D, Major N, Choux M, Bourque A (2005) A first evaluation of the strength and weaknesses of statistical downscaling methods for simulating extremes over various regions of eastern Canada. Sub-component, Climate Change Action Fund (CCAF), Environment Canada, Montreal, Quebec, Canada, 209 pp

Goodess CM, Osborn TJ, Hulme M (2003) The identification and evaluation of suitable scenario development methods for the estimation of future probabilities of extreme weather events. Tyndall Centre Technical Report 4: Norwich, 69 pp

Gordon CC, Cooper CA, Senior H, Banks JM, Gregory TC, Johns JF, Mitchell B, Wood RA (2000) The simulation of SST, sea ice extents and ocean heat transport in a version of the Hadley Centre coupled model without flux adjustments. Clim Dynam 16: 147–168

Hewitson BC, Crane RG (2004) Global climate change: Questions of downscaled climate scenarios for impact assessment. Bull Amer Meteor Soc (in press)

Houghton JT, Ding Y, Griggs DJ, Noguer M, van der Linden PJ, Dai X, Maskell K, Johnson CA (eds) (2001) Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC): Cambridge University Press, Cambridge, UK and New York, NY, USA, 84 pp

Katz RW (1996) Use of conditional stochastic models to generate climate change scenarios. Climatic Change 32(3): 237–255

Khan MS, Coulibaly P, Dibike Y (2006a) Uncertainty analysis of statistical downscaling methods. J Hydrol 319(1–4): 357–382

Khan MS, Coulibaly P, Dibike Y (2006b) Uncertainty analysis of statistical downscaling methods using Canadian Global Climate Model predictors. Hydrol Process 20(14): 3085–3104

Kistler R, Kalnay E, Collins W, Saha S, White G, Woollen J, Chelliah M, Ebisuzaki W, Kanamitsu M, Kousky V, Dool HVD, Jenne R, Fiorino M (2001) The NCEP/NCAR 50-year reanalysis. Bull Amer Meteor Soc 82(2): 247–267

Mearns LO, Giorgi F, Whetton P, Pabon D, Hulme M, Lal M (2003) Guidelines for use of climate scenarios developed from regional climate model experiments. Data Distribution Centre of the International Panel on Climate Change (IPCC), 38 pp

Overland JE, Adams JM, Bond NA (1997) Regional variation of winter temperatures in the Arctic. J Climate 10(5): 821–837

Semenov MA, Barrow EM (1997) Use of stochastic weather generator in the development of climate change scenarios. Climatic Change 35: 397–414

Vincent LA, Zhang X, Bonsal BR, Hogg WD (2002) Homogenization of daily temperatures over Canada. J Climate 15: 1322–1334

Walsh JE, Kattsov VM, Chapman WL, Govorkova V, Pavlova T (2002) Comparison of Arctic Climate Simulations by uncoupled and coupled global models. J Climate 15: 1429–1446

Wilby RL, Dawson CW (2004) Using SDSM Version 3.1 – a decision support tool for the assessment of regional climate change impacts. King's College London: UK, 67 pp

Wilby RL, Wigley TML (2000) Precipitation predictors for downscaling: observed and general circulation model relationships. Int J Climatol 20: 641–661

Wilby RL, Wigley TML, Conway D, Jones PD, Hewitson BC, Main J, Wilks DS (1998) Statistical downscaling of general circulation model output: a comparison of methods. Water Resour Res 34: 2995–3008

Wilby RL, Dawson CW, Barrow EM (2002) SDSM – a decision support tool for the assessment of regional climate change impacts. Environ Modell Softw 17: 145–157

Wilby RL, Tomlinson OJ, Dawson CW (2003) Multi-site simulation of precipitation by conditional resampling. Climate Res 23: 183–194

Wilby RL, Charles SP, Zorita E, Timbal B, Whetton P, Mearns LO (2004) Guidelines for use of climate scenarios developed from statistical downscaling methods. Data Distribution Centre of the International Panel on Climate Change (IPCC) TGCIA, 27 pp

Wilcoxon F (1945) Individual comparisons by ranking methods. Biometrics 1: 80–83

Authors' addresses: Dr. Yonas B. Dibike (e-mail: Dibike.Yonas@ouranos.ca), NSERC Research Fellow, OURANOS Consortium, 550 Sherbrooke Street West, 19th Floor, Montreal (QC) H3A 1B9, Canada; Philippe Gachon, Adaptation and Impact Research Division (AIRD), Atmospheric Science and Technology Directorate, Environment Canada at Ouranos, Montreal (QC), Canada; André St-Hilaire and Taha B. M. J. Ouarda, Institut National de la Recherche Scientifique Centre Eau, Terre & Environnement (INRS-ETE), University of Québec, 490 Rue de La Couronne, Québec (QC) G1K 9A9, Canada; Van T.-V. Nguyen, Department of Civil Engineering and Applied Mechanics, McGill University, 817 Sherbrooke Street West, Montreal (QC) H3A 2K6, Canada.