Vol. 54: 129–148, 2012
doi: 10.3354/cr01106

CLIMATE RESEARCH
Clim Res

Published online September 12

A multivariate multi-site statistical downscaling model for daily maximum and minimum temperatures

D. I. Jeong[1,*], A. St-Hilaire[2], T. B. M. J. Ouarda[2,3], P. Gachon[4]

[1]Centre ESCER (Étude et Simulation du Climat à l'Échelle Régionale), UQAM (Université du Québec à Montréal), 201 Ave. President-Kennedy, Montreal, Quebec H3A 2K6, Canada

[2]INRS-ETE, University of Quebec, 490 de la Couronne Street, Quebec G1K 9A9, Canada

[3]Water and Environmental Engineering, Masdar Institute of Science and Technology, PO Box 54224, Abu Dhabi, UAE

[4]Atmospheric Science and Technology Directorate, Canadian Centre for Climate Modeling and Analysis (CCCMA) section, Climate Research Division, Environment Canada, 800 de la Gauchetiere West, Office 7810, Montreal, Quebec H5A 1L9, Canada

ABSTRACT: A multivariate multi-site statistical downscaling model (MMSDM) was developed for simultaneous downscaling of climate variables including daily maximum and minimum temperatures (T_{max} and T_{min}) for multiple observation sites. The MMSDM employs multivariate multiple linear regression (MMLR) to simulate deterministic series from large-scale reanalysis data and adds spatially correlated random series to the deterministic series of the MMLR to complement the underestimated variance and to reproduce a spatial correlation of T_{max} and T_{min} from multiple sites and an at-site temporal correlation between T_{max} and T_{min}. The MMSDM model is called MMLRc. The downscaled results of the MMLRc were compared to those of MMLR without random noise (MMLRn) and MMLR with uncorrelated random noise (MMLRi) over the southern Quebec area of Canada. The MMLRc almost exactly reproduced the cross-site correlation of T_{max} and T_{min} among multiple observation sites, and it accurately reproduced the at-site temporal correlation between T_{max} and T_{min} at each observation site. The MMLRi and MMLRc reproduced monthly standard deviations of daily T_{max} and T_{min}, the 90th percentile of T_{max} ($T_{max}90$), the 10th percentile of T_{min} ($T_{min}10$), and the frost and thaw cycle (Fr-Th) more accurately than the MMLRn model. However, both MMLRc and MMLRi yielded a larger standard error for the monthly mean of daily T_{max} and daily T_{min}, frost season length (FSL), and growing season length (GSL). For the Fr-Th and diurnal temperature range, the MMLRc performed better than the MMLRn and MMLRi. We conclude that the MMLRn may serve as an alternative to downscaling deterministic signals of a predictand, consistent with global climate model predictors, and it may serve to project the averaged central tendency of a predictand. The MMLRc, however, is recommended for reproduction of variance, extreme events, and the inter-annual variability of the predictands.

KEY WORDS: Linear regression · Multi-site · Multivariate · Spatial and temporal correlations · Statistical downscaling · Mean and extreme temperatures

1. INTRODUCTION

Atmosphere-Ocean Global Climate Models (AOGCMs; Table 1 gives a list of abbreviations used in this article) provide simulated values for present and future climate variables such as wind, temperature, humidity, and air pressure, with forcing including scenarios of increases in greenhouse gas and aerosol concentrations in the atmosphere. AOGCM outputs are generated on a coarse scale grid, however, with horizontal resolutions generally coarser than 2° latitude by 2° longitude, which limits their use at the local scale. Despite the poor AOGCM performance at the local scale, they produce quite reliable atmo-

*Email: jeong@sca.uqam.ca

spheric circulation outputs for the large scale upper-air fields that are used as predictors or forcing variables (Huth 2002, Harding et al. 2011) in downscaling techniques applied to generate information and climate scenarios at a finer scale. Statistical downscaling models (SDMs) have frequently been adopted to fill this scaling gap in impact and adaptation studies (Huth 2002, Wilby et al. 2004).

SDMs are based on the fundamental assumption that empirical relationships between upper-air circulation variables (i.e. the predictors) and the local climate variable (i.e. the predictand) can be established (Wilby et al. 2002). Statistical downscaling approaches are often used due to their ease of implementation, low computation requirements, and their ability to provide climate information equivalent to point climate observations. Several studies have reviewed these methods (Xu 1999, Wilby et al. 2004), dividing them into 3 main groups: transfer function, stochastic weather generation, and weather typing approaches. Stochastic weather generator approaches — which reproduce the statistical characteristics of local climate variables based on random number generator, Markov chain model, and/or parametric (or nonparametric) distributions — are fairly simple but flexible and computationally economical methodologies for generating both single-site and multi-site daily weather data (Katz & Parlange 1995, Wilks 1998, 1999a, 1999b, Qian et al. 2002). Weather typing approaches generate local climate variables by resampling or by using a k-nearest neighbor approach from observed atmospheric data (Rajagopalan & Lall 1999, Zorita & von Storch 1999, Yates et al. 2003, Crimmins 2006). The common and relatively simple transfer function methods derive an empirical relationship between predictors and predictands using linear or non-linear transfer functions (von Storch et al. 2000). In this category, multiple linear regression (MLR) (Enke & Spekat 1997, Hellström et al. 2001, Huth et al. 2001, Palutikof et al. 2002, Wilby et al. 2002, Huth 2004, Hessami et al. 2008), canonical correlation regression (CCR) (Trigo & Palutikof 2001, Huth 2002, 2004, Xoplaki et al. 2004, Busuioc et al. 2006), principal component analysis regression (von Storch et al. 2000, Buishand & Brandsma 2001, Huth 2004), and artificial neural networks (Hewitson & Crane 1996, Trigo & Palutikof 1999, Mpelasoka et al. 2001, Stehlik & Bardossy 2002, Miksovsky & Raidl 2005, Tolika et al. 2007, Huth et al. 2008) have all been used as transfer functions. Most statistical downscaling approaches, including the transfer function approaches, are computationally low-cost (von Storch et al. 2000), adaptable to the local scale, and applicable to any AOGCM output (Hellström et al. 2001). The limited capability to reproduce the observed variability of the climate variables (Wilby et al. 2002) constitutes the main weakness of the transfer function approaches. By definition, single site downscaling usually does not account for the spatial coherence of the estimated variable over an area of interest. When dealing with multi-site downscaling, transfer function approaches seldom reproduce spatial dependence (Wilby et al. 2003, Harpham & Wilby 2005).

Table 1. List of abbreviations

Abbreviation	Expansion
AOGCM	Atmosphere-ocean global climate model
CCR	Canonical correlation regression
CGCM3	Canadian coupled global climate model, version 3
IQR	Interquartile range
MLR	Multiple linear regression
MMLR	Multivariate multiple linear regression
MMLRc	MMLR with spatially correlated random noise
MMLRi	MMLR with spatially independent random noise
MMLRn	MMLR without random noise
MMSDM	Multivariate multi-site statistical downscaling model
NCAR	National center for atmospheric research
NCEP	National centers for environmental prediction
MSE	Mean square error
RMSE	Root mean square error
SD	Standard deviation
SDM	Statistical downscaling models
SE	Standard error

Generally, AOGCM (and reanalysis) atmospheric predictors can explain only a part of the variance of the local predictands, i.e. the part associated with large-scale circulation variability. Studies must therefore implement a variance-increase methodology, and the use of either the inflation procedure or the randomization method addresses this issue (von Storch 1999, Huth 2002). The inflation procedure increases the downscaled temperature anomaly each day by adjusting a factor; however, the main drawback of this procedure is that all local variance originates from large-scale variability (Huth 2002). It is well known that the regional or local scale variability of low level air temperatures is also influenced by surface conditions, especially in large northern land masses, such as Canada, where snow cover and frost properties of the soil (among other factors) play a key role in modulating the surface radiative or low level diabatic fluxes over the year (see the study of the Hudson Bay area in winter in Gachon et al. 2003). With the randomization procedure, adding noise to

the downscaled predictand series supplements the underestimated variability (von Storch, 1999, Huth 2002, Hessami et al. 2008). The randomization procedure reproduces variability more accurately because the large-scale AOGCM predictors only partially control the spatial and temporal variability of downscaled variables (von Storch, 1999). An approach combining transfer function and stochastic randomization approaches can incorporate unexplained variability (i.e. from regional scale feedback) in addition to deterministic component (i.e. from large-scale feedback).

The most frequently used approaches for the simulation of climate variables at multiple sites are stochastic weather generation approaches based on autocorrelation (Wilks 1998, 1999a, 1999b, 2002, Qian et al. 2002, Apipattanavis et al. 2007, Mehrotra & Sharma 2007). More specifically, Wilks (1999b, 2002) developed a multivariate multi-site weather generation model to simultaneously generate daily T_{max} and T_{min}. He extended a single-site multivariate (i.e. T_{max} and T_{min}) weather generation model based on a first-order multivariate autoregressive approach to multi-site using the simultaneous cross-correlations and lagged autocorrelations between daily T_{max} and T_{min} at multiple sites. Drawing on the spatial autocorrelation defined as a correlation among the values of a single variable, Khalili et al. (2007 and 2009) employed multi-site daily precipitation and weather data (T_{max}, T_{min}, and solar radiation) generation models, that reproduced spatially correlated multi-site climate variables using the spatial moving average process based on a stochastic weather generation approach. Harpham & Wilby (2005) employed an artificial neural network technique as the transfer function to simulate precipitation at multiple sites using AOGCM predictors, but the transfer function tended to overestimate spatial dependency among the sites. In multi-site downscaling, the variance inflation procedure does not affect the spatial dependence of the series downscaled by transfer function approaches between the local observation sites (Cannon 2009). Wilby et al. (2003) developed a hybrid approach, combining transfer function and stochastic weather generator approaches for multi-site downscaling and this was successfully applied again by Harpham & Wilby (2005) for daily precipitation. Schoof et al. (2007) applied the hybrid approach to represent at-site correlation between daily T_{max} and T_{min}, but did not extend this approach to reproduce spatial dependency.

One relevant option for reproduction of the observed spatial coherence among multiple sites and temporal correlation between T_{max} and T_{min} at a site is to add the spatially correlated random series to the series downscaled by regression-based approaches. This study developed a multivariate multi-site SDM (MMSDM) based on an approach combining transfer function approach and stochastic randomization procedures for a limited regional area in which the spatial coherence of predictands is important for climate change impact analysis. This MMSDM is a multivariate and multi-site extension of previous univariate, single-site SDMs (e.g. Wilby et al. 2002, Hessami et al. 2008). The developed MMSDM was then applied to multiple climate variables, including T_{max} and T_{min} and their annual and seasonal extremes, at multiple observation sites in southern Quebec, Canada. As a first step, reanalysis products were used as predictor data series to test and validate this new model, before using AOGCM predictors in further applications. This study employed multivariate multiple linear regression (MMLR) for simultaneous downscaling of daily T_{max} and T_{min} for multiple observation sites. MMLR is a logical extension of univariate MLR, that enables different response variables from various atmospheric data series to be modeled simultaneously. In our study, the stochastic randomization procedure modeled variance that was not explained by MMLR. To reproduce spatial correlation of T_{max} and T_{min} and at-site temporal correlation between T_{max} and T_{min}, the correlated random noise was generated from a multivariate normal distribution and added to the downscaled deterministic series. The resultant MMSDM is called MMLRc in the present study. This paper compares the results of the MMLRc to those of MMLR without random noise (MMLRn) and MMLR with uncorrelated random noise (MMLRi). The only difference among the 3 models is the error part (i.e. what is not explained by the regression approach), which is (1) completely neglected by MMLRn, (2) assumed to be independent from site to site and from T_{max} to T_{min} by MMLRi and (3) correlated between sites and between T_{max} and T_{min} by MMLRc.

2. METHODOLOGY

2.1. Multivariate multiple linear regression

To analyze the relationship between different independent variables and multiple dependent variables, multivariate regression approaches have been used in many scientific areas. One of the simplest methods, MMLR is a set of univariate MLRs, which estimate the same coefficients and standard errors

obtained using the ordinary least squares method. In addition, MMLR estimates the between-equation covariance, enabling the verification of the adequacy of an estimated coefficient set for individual MLR equations.

Given a matrix, **X**, of multiple predictor variables of dimension $n \times k$ and a matrix, **Y**, of multivariate predictand variables of dimension $n \times m$—where the measurement record length n is larger than the dimension of the explanatory variables k—the parameter matrix **B** of dimension $k \times m$ that defines the linear relationship between the 2 matrices **X** and **Y** can be estimated. For the m dependent variables and k independent variables, the MMLR can be expressed as:

$$\mathbf{Y}_{[n \times m]} = \mathbf{X}_{[n \times k]} \cdot \mathbf{B}_{[k \times m]} + \mathbf{E}_{[n \times m]} \quad (1)$$

where **E** is the residual matrix of dimension $n \times m$.

The ordinary least squares estimates of parameter matrix $\hat{\mathbf{B}}$ in the $k \times m$ parameter space is given by:

$$\hat{\mathbf{B}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y} \quad (2)$$

where T denotes matrix transpose.

One of the problems with the parameter matrix $\hat{\mathbf{B}}$ is its multi-collinearity, which produces large standard errors of estimated parameters in the related independent variables. In order to limit the influence of multicollinearity, many methods such as latent variable regression methods (e.g. principal components regression, partial least squares regression, reduced rank regression and CCR), ridge regression, and stepwise regression have been employed. This study employed backward stepwise regression to select optimum predictors from a large number of potential predictors (126 in the present study), some of which are highly correlated. Backward regression was initiated with all predictors being part of the model, and redundant predictors were eliminated one after the other.

The F-test was used to confirm or void the null hypothesis that $(k - q)$ specified variables had regression coefficients equal to zero, where k and q are numbers of predictors in the regression models. The F-statistic for testing the above hypothesis at a given confidence level α in the MMLR approach is defined as:

$$F = \frac{(\bar{R}_k^{\,2} - \bar{R}_q^{\,2})/(k-q)}{(1-\bar{R}_k^{\,2})/(n-k-1)} \quad \text{d.f.} = k-q,\ n-k-1 \quad (3)$$

where $\bar{R}_k^{\,2}$ and $\bar{R}_q^{\,2}$ are averaged value of the R-squares of the downscaled daily T_{max} and T_{min} series from all sites for the MMLR equations containing all the k predictor variables and the q predictor variables, respectively. The regression abilities of MMLR model with the backward stepwise approach were compared to that of CCR, a popular latent variable multivariate regression (Burnham et al. 1996). CCR regressed the multivariate response matrix **Y** onto the first d-number of canonical latent variables in the **X** space. Orthogonal latent variables in the **X** and **Y** spaces are found by canonical correlation analysis (for more detail, see Burnham et al. 1996). If CCR employs all of the canonical variables ($d = m$ in this analysis) and MMLR employs all of the potential variables as predictors, then the CCR and MMLR will generate identical results for calibration data although both models may have an over-fitting problem and not produce robust outputs for validation data.

2.2. Transfer function methodology (MMLRn)

The daily maximum and minimum temperatures (T_{max}, T_{min}) at multiple observation sites in a region were simultaneously downscaled using the following MMLR equation:

$$\hat{\mathbf{T}} = \mathbf{X}\hat{\mathbf{B}} \quad (4)$$

where $\hat{\mathbf{T}}[n \times m]$ is the downscaled T_{max} and T_{min} series of MMLRn, $\mathbf{X}[n \times k]$ is the standardized (subtract the mean and divide by standard deviation) predictor variables matrix; and the parameter matrix $\hat{\mathbf{B}}[k \times m]$ has MMLR parameters estimated using the ordinary least squares estimation method. The total variances of the observed daily maximum and minimum temperatures **T** can be decomposed as follows:

$$\mathrm{Var}[\mathbf{T}] = \mathrm{E}(\mathrm{Var}(\mathbf{T}|\mathbf{X})) + \mathrm{Var}(\mathrm{E}(\mathbf{T}|\mathbf{X})) \quad (5)$$

where $\mathrm{E}(\mathbf{T}|\mathbf{X}) = \hat{\mathbf{T}}$ is the output of regression and $\mathrm{Var}(\mathbf{T}|\mathbf{X})$ is the error variance of the regression. As both terms on the right-hand side are positive, $\mathrm{Var}(\mathbf{T}) > \mathrm{Var}(\hat{\mathbf{T}})$. Note that the error variance $\mathrm{Var}(\mathbf{T}|\mathbf{X})$ is neglected by MMLRn, while it is complimented by independent random noise from site to site and from T_{max} to T_{min} by MMLRi, and by spatially correlated random noise between sites and between T_{max} and T_{min} by MMLRc.

2.3. Stochastic randomization procedure

The residual (or error) matrix $\mathbf{E}_{\mathrm{T}}$ $[n \times m]$ of the MMLRn model is described as:

$$\mathbf{E}_{\mathrm{T}} = (\mathbf{T} - \hat{\mathbf{T}}) \quad (6)$$

where **T** $[n \times m]$ is the observed matrix and $\hat{\mathbf{T}}[n \times m]$ is the matrix of daily T_{max} and T_{min} downscaled by the

MMLRn. As shown in Eq. (5), the variance of the MMLRn is smaller than that of observations, and variance of the MMLRn is increased by adding random noise.

The MMLRn series predict only deterministic components explainable by linear regression and the independent atmospheric variables **X** for the different multiple sites. Consequently, the MMLRn will probably overestimate the spatial correlation. Adding spatially uncorrelated random noises to the MMLRn of each site can help to better estimate the total variance. However, it is also equivalent to adding a nugget effect in the covariance functions (or in the semi-variograms). To reproduce the observed spatial coherence, an alternative option is to add correlated random noises to the MMLRn.

For the MMLRc, correlated random noise among the predictands (T_{max} and T_{min}) at multiple sites was generated from multivariate normal distribution and added to the series of the MMLRn. By adding the correlated random series, the MMLRc can reproduce variance of each predictand (T_{max} and T_{min}) at each site, the correlation between T_{max} and T_{min} at each site, and cross-correlation among T_{max} and among T_{min} at multiple sites. The cross-correlated error matrix $\mathbf{R_T}[n \times m]$ is generated from a multivariate normal distribution having zero error mean and an error covariance matrix [$\mathbf{\Sigma} = \mathbf{SCS}$] equal to that of the residual matrix $\mathbf{E_T}(\mathbf{R_T} \sim N_m(0,\mathbf{\Sigma}))$, where **S** is a diagonal matrix of standard deviations and **C** is a correlation matrix of the residual matrix $\mathbf{E_T}$. Generated residuals were then added to the downscaled temperature matrix $\hat{\mathbf{T}}$ as:

$$\tilde{\mathbf{T}} = \hat{\mathbf{T}} + \mathbf{R_T} \quad (7)$$

For the MMLRi, uncorrelated random noise was added to the deterministic series of the MMLRn. The uncorrelated random noise vector $\mathbf{r}_\mathbf{T}^j$ of the MMLRi was generated from a normal distribution having zero error mean and error standard deviation S^j equal to that of the residual vector $\mathbf{e}_\mathbf{T}^j(\mathbf{r}_\mathbf{T}^j \sim N(0,S^j))$. The MMLRi output vector $\tilde{\mathbf{T}}^j$ was produced by adding generated the residual vector $\mathbf{e}_\mathbf{T}^j$ to the downscaled deterministic temperature vector $\hat{\mathbf{t}}^j$. Note that the differences among the 3 models are error terms $\mathbf{E_T}$, which are neglected at MMLRn, assumed to be independent from site to site for the MMLRi, and correlated for site to site for the MMLRc. Fig. 1 schematizes the downscaling procedures of the MMLRn, MMLRc, and MMLRi.

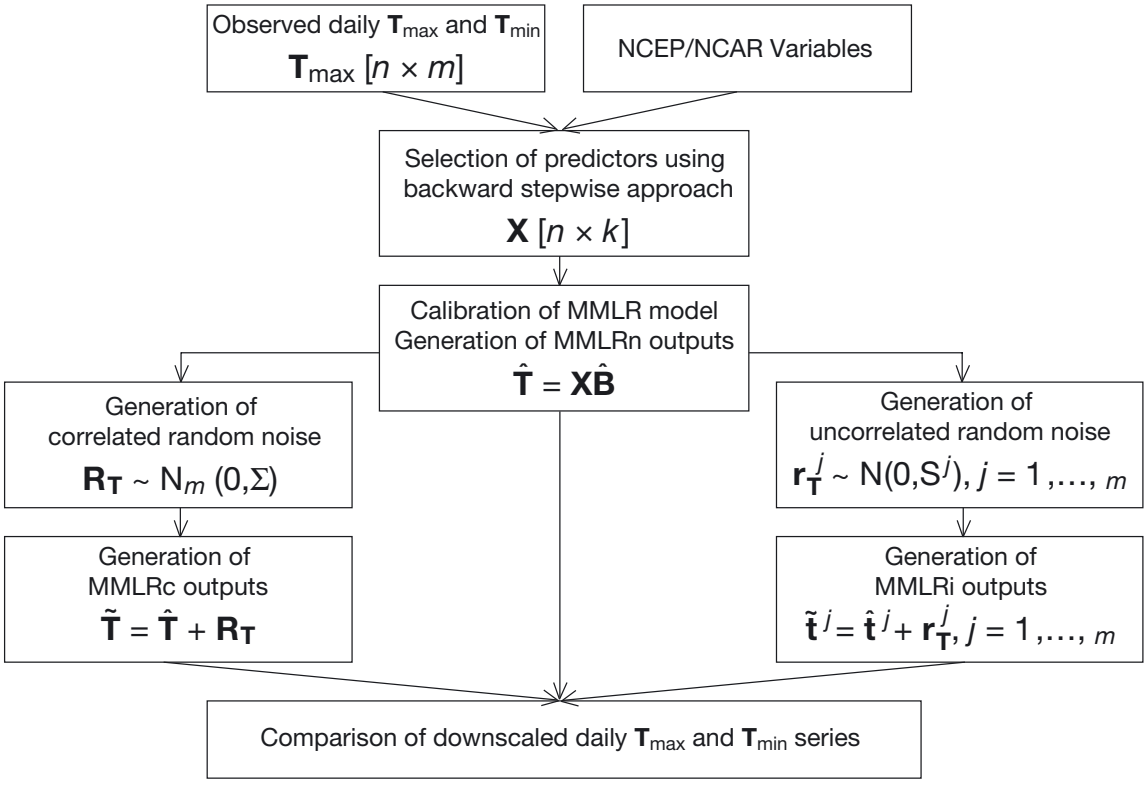

Fig. 1. Downscaling procedures of multivariate multiple linear regression (MMLR) models MMLRn, MMLRc, and MMLRi, which represent MMLR without random noise, with correlated random noise, and with uncorrelated random noise respectively (for details see Section 2)

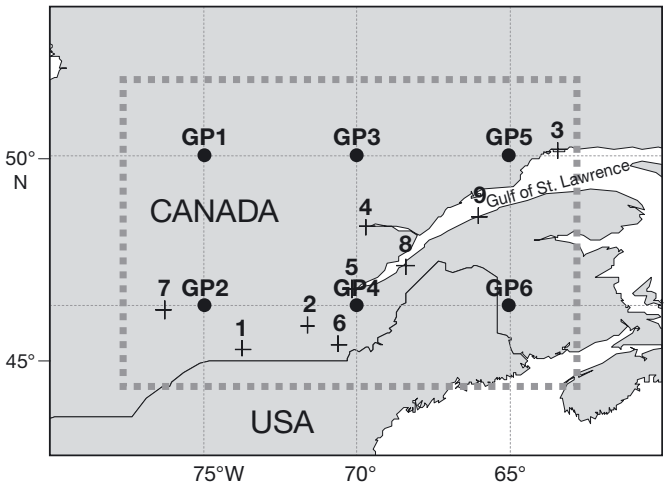

Fig. 2. Locations of 6 global scale predictor grid points and 9 observation stations of daily T_{max} and T_{min}. (+) Meteorological sites: (1) Cedars; (2) Drummondville; (3) Sept-Îles; (4) Bagotville A; (5) Quebec/Jean Lesage Intl.; (6) Sherbrooke A; (7) Maniwaki Airport; (8) La Pocatierre; (9) Mont-joli A. Dots GP1 to GP6 represent Canadian Coupled Global Climate Model, version 3 (CGCM3) grid points

3. MODEL APPLICATION

3.1. Study area and predictand data source

Fig. 2 shows the study area. We focused on southern Quebec, the main study region of a multi-partner project developing and comparing downscaling methods. The study area is relatively large (i.e. around 600 km from southwest to northeast stations and 350 km from south to north). It covers a broad range of climatic conditions from continental to maritime areas and from mild to colder climatic zones (mean annual temperature is 6.56°C at Cedars and 0.94°C at Sept-Îles). For predictands, this study used daily maximum and minimum temperature data from Environment Canada stations, with a homogenized time series developed by Vincent et al. (2002). 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. For the period 1961 to 2000, <4% of data are missing, and they were excluded from the analysis. Fig. 2 shows the names and locations of the 9 selected meteorological stations and the 6 grid points that provided global scale atmospheric predictor variables.

3.2. Predictor data series

Potential predictors were obtained from the National Centers for Environmental Prediction and National Center for Atmospheric Research (NCEP/NCAR) reanalysis data (e.g. Kalnay et al. 1996, Kistler et al. 2001) for the period from 1961 to 2000. The NCEP/NCAR reanalysis uses a T62 (~209 km) global spectral model for consistent assimilation of observational data from a wide variety of observed sources. Kalnay et al. (1996) describes the reanalysis procedure in detail. The NCEP/NCAR reanalysis data, which originally have a horizontal resolution of 2.5° latitude × 2.5° longitude were linearly interpolated onto the grid of the Canadian Coupled Global Climate Model, version 3 (CGCM3) of the Canadian Centre for Climate Modeling and Analysis. This AOGCM and its gridded output are regularly used for downscaling and climate change impact research, both in Canada and abroad. The interpolation procedure from the NCEP/NCAR to the CGCM3 grids facilitates the use of AOGCM predictors for the climate change simulations. In subsequent work, we will use the CGCM3 predictors to generate daily climate change data for the same area, as is already being done by many other end users. It is therefore useful to apply the NCEP predictors redistributed on the CGCM3 grid. The CGCM3 has a horizontal resolution of 3.75° latitude × 3.75° longitude (about 400 km). Reanalysis variables are standardized with respect to their annual means and standard deviations for the period 1961 to 1990. All details on the preparation of the 25 atmospheric variables from NCEP/NCAR reanalysis and the pre-processed data developed and obtained from the Canadian Climate Change Scenarios Network project of Environment Canada (www.cccsn.ca) are available on the Data Access and Integration (DAI) web link (http://loki.qc.ec.gc.ca/DAI/), which is the main Canadian source for downscaling predictors.

Among the 25 atmospheric variables, wind directions at 500, 850, and 1000 hPa levels and temperature at 2 m are excluded from the set of potential predictors because wind direction is a circular variable and discontinuous at 360° (or 0°) and temperature at 2 m is known to be poorly represented in CGCM3 at the local scale—for example in the north (Gachon & Dibike 2007) and east (Jeong et al. 2011) of Canada. Therefore, the employed potential predictor set consists of atmospheric circulation variables (e.g. wind speed, U- and V-components of wind, divergence, and vorticity at 500, 850, and 1000 hPa levels), mean sea level pressure, geopotential heights (at 500 and 850 hPa levels), and specific humidities (at 500, 850, and 1000 hPa levels). Upper air-field temperatures and thicknesses have been shown to be effective predictors of the surface temperature (Huth 2002) but

they are not available as potential predictors from DAI. Temperature features and thickness are physically linked with pressure and humidity fields in the atmosphere, and these latter variables (i.e. geopotential height and specific humidity at various pressure levels) represent a good compromise substitute for the temperature fields and thickness at various pressure levels in the downscaling process. The thickness between pressure levels (e.g. difference between 500 and 850 hPa geopotential heights) is proportional to mean virtual temperature in that layer, so that the combination of pressure fields with the humidity content of the considered tropospheric level in the downscaling of temperature gives some quite accurate results (for further details in their use in the downscaling of temperatures in various areas of Canada, see Gachon et al. 2005, Gachon and Dibike 2007, Dibike et al. 2008, Hessami et al. 2008, Jeong et al. 2011).

3.3. Predictor selection

Following Hessami et al. (2008), a backward stepwise approach was used for the selection of the potential predictor group among the possible 126 predictors (6 grid points × 21 potential predictors). In the present study, the MMLR parameters for daily T_{max} and T_{min} time series were estimated separately for each month for the calibration period 1961 to 1990. The calibrated monthly MMLRn models were validated from 1991 to 2000. The MMLRc and MMLRi were also calibrated separately for each month and the same calibration and validation periods were used as for the MMLRn.

Table 2 shows the predictors selected for the MMLR equations each month (the choice of predictors varies during the year) from the initial 126 potential predictor variables on all 6 grid points. A backward stepwise approach was employed, using the *F*-test as criterion for predictor inclusion in the model, at a confidence level $\alpha = 0.01$. The same predictors were employed for both T_{max} and T_{min} for all sites to simultaneously predict all predictands under the assumption that the regional temperature at the observation sites is affected by the same large-scale atmospheric variables, although the strengths of influence of a large-scale variable on different stations might not be the same, and this could vary from month to month. Deterministic series downscaled from the same predictors for all predictands (i.e. MMLRn), and those from different predictor sets se-

Table 2. Selected numbers as predictors from 6 global scale grid points of each variable by backward stepwise regression for monthly multivariate multiple linear regression (MMLR) equations used in multivariate multi-site statistical downscaling modeling (MMSDM) of daily T_{min} and T_{max} at 9 obervation stations in Quebec, Canada (see Fig. 2). Note that same predictor set was used for both T_{max} and T_{min} at all stations

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Mean sea level pressure	2	3	2	3	3	2	1	2	3	3	3	2	29
1000 hPa													
Wind speed												3	3
U-component		1	2	2	1		1		1		1		9
V-component			1	1		1	1	1	1		1	1	8
Vorticity	1					1			1				3
Divergence		1					1			1		1	4
Specific humidity	2	2	1	1	1	2	2	2	2	1	3	1	20
850 hPa													
Wind speed	2	2				1							5
U-component				2	1		1				1		5
V-component				1		1	2				1	1	6
Vorticity							1		1				2
Divergence													
Specific humidity	1	1	1	1	2	1		1		1		1	10
Geopotential	2	2	4	3	3	1	2	2	2	2	1	4	28
500 hPa													
Wind speed				1									1
U-component				1						1	2		4
V-component						1	1			1			3
Vorticity				1	1						1		3
Divergence													
Specific humidity						1							1
Geopotential	1									2		1	4
Totals	11	12	11	17	12	12	13	8	11	12	14	15	

lected by the stepwise approach for each predictand (i.e. individual regression model) yielded statistically similar R^2 values at the 95% confidence level for all but 16 out of a total of 216 (i.e. cases for 12 months × 18 predictands). Different predictors were employed for each month, and approximately 8 to 17 atmospheric variables were selected as predictors of the MMLRn model for each month. Geopotential height at 850 hPa, specific humidity at 1000 hPa, and mean sea level pressure were the most frequently selected predictors among the 21 reanalysis variables. Variables near the surface and in the boundary layer (i.e. 1000 and 850 hPa) were selected as predictors more frequently than those at the upper level (500 hPa). The selected predictors presented in Table 2 are relevant for downscaling daily T_{max} and T_{min} variables because they include several atmospheric circulation predictors and thermodynamic fields (e.g. wind speeds, U- and V-components, specific humidity, and divergence at different levels) that are physically related to daily temperatures. Daily temperatures are known to be physically connected with warm or cold air advection and meridional circulation (Kozuchowski et al. 1992). Northerly winds generally determine low temperatures, while southerly winds determine high temperatures, especially in a continental area such as Canada where low-level cold or warm air advections, originating from the Arctic basin and the Gulf of Mexico/southern USA respectively, and strongly varying over the year, play a key role in the determining 2 m air temperature and its variability (see for example the effect of northerly and southerly winds on the low-level air temperature fluctuations during winter storm events in the Hudson Bay area in Gachon et al. 2003). The seasonally different effects of atmospheric variables on T_{max} and T_{min} in Canada were the main reason behind the use of different predictors and regression parameters for each month in this study. Wet or dry day conditions may have been indirectly linked or incorporated in the T_{max} and T_{min} downscaling model of this study by using specific humidities, geopotential heights, and atmosphere circulation variables at different pressure levels as predictors, since these affect both precipitation processes (Hessami et al. 2008) and regional or local diabatic fluxes, which in turn exert some influence on the temperature fields. Stepwise regression methodology occasionally automatically excludes predictor variables that may be crucial for climate change (IPCC 2001, 2007), but the selected predictor set included predictors sensitive to climate change such as 850 and 1000 hPa specific humidities and 500 and 850 hPa geopotential heights (e.g. Frías et al. 2006, Boé et al. 2007, Busuioc et al. 2007, Gachon & Dibike 2007, Dibike et al. 2008). One important assumption of statistical downscaling is that the selected predictors should be well simulated by the AOGCM. Jeong et al. (2011) reported, for example, that the CGCM3 AOGCM showed a fairly high level of skill with respect to reanalysis predictor variables in southern Quebec, especially for the mean sea level pressure, 850 hPa geopotential height, 850 and 1000 hPa specific humidities, which were most frequently selected as predictors in the present study.

3.4. Diagnostic and statistical criteria of analysis using temperatures indices

At-site relationships between T_{max} and T_{min} and cross-site correlations among T_{max} and among T_{min} at multiple sites were evaluated using linear correlation coefficients. The correlation coefficients of the MMLRn, MMLRc, and MMLRi models were compared to those of observations.

Monthly means and standard deviations (SDs) of downscaled daily T_{max} and T_{min} series were calculated for each month of every year and compared to those of observations. Although the 3 SDMs generate daily T_{max} and T_{min} series, the 2 basic statistics (i.e. mean and SD) of the 3 models was calculated for every month in every year and compared to the same statistics for the observed daily series. This procedure was adopted in order to reveal monthly differences in reproduction ability among the 3 SDMs. Additionally, 6 temperature indices were employed to evaluate the performance of the 3 models, using the methods shown in Table 3. For temperature indices, each daily downscaled series is used to compute monthly, seasonal, or yearly statistics. Warm and cold extremes are considered through the use of 90th percentile of T_{max} ($T_{max}90$), and 10th percentile of T_{min} (i.e. $T_{min}10$), respectively. The growing season length (GSL) index is calculated from daily mean temperatures (T_{mean}), which are defined as the average of T_{max} and T_{min}. The GSL starts when T_{mean} exceeds 5°C during 5 consecutive days and ends after 5 consecutive days of temperatures below 5°C. The number of days with freeze and thaw cycle (Fr-Th) and diurnal temperature range (DTR) can be inferred with reasonable accuracy when relationships between daily T_{max} and daily T_{min} series of a site are reproduced accurately. Because series downscaled by MMLRc and MMLRi produce different at-site correlations between T_{max} and T_{min}, they yield different values for Fr-Th, and DTR at each site. However, the other indices can only be estimated separately from daily T_{max} or daily T_{min}

Table 3. Diagnostic indices derived from daily T_{max} and T_{min} values used to evaluate the performance of statistical downscaling methods (for detailed definitions, see Gachon et al. 2005)

Statistics and indices	Definition	Unit	Time scale
Mean T_{max}	Mean of daily T_{max}	°C	Month
SD T_{max}	Standard deviation of daily T_{max}	°C	Month
Mean T_{min}	Mean of daily T_{min}	°C	Month
SD T_{min}	Standard deviation of daily T_{min}	°C	Month
$T_{max}90$	90th percentile of daily T_{max}	°C	Season
$T_{min}10$	10th percentile of daily T_{min}	°C	Season
FSL	Frost season length: T_{min} < 0 °C for >5 d	d	Year
GSL	Growing season length: Start: T_{mean} > 5 °C for >5 d End: T_{mean} < 5 °C for >5 d	d	Year
Fr-Th	Number of days with frost and thaw cycle (T_{max} > 0 °C, T_{min} < 0°C)	d	Month
DTR	Diurnal temperature range (T_{max} – T_{min})	°C	Season

series at a single site. Both MMLRi and MMLRc will therefore yield basically the same values for these indices, although values may vary slightly due to random sampling error. To assess the downscaling ability of the randomization models, all statistics and indices evaluated from the models were compared to those from the MMLR model without random noise (MMLRn). However, the results of the MMLRc and MMLRi are provided separately only when they yielded different results, i.e. for Fr-Th and DTR.

The criteria for the comparison of results were the mean bias, standard error (SE), and root mean square error (RMSE) for estimated statistics and climate indices from statistical downscaling models. Note that the mean square error (MSE) is the sum of the variance of error and the square of bias ($MSE = SE^2 + bias^2$).

3.5. Comparison between MMLRn and CCR

Prior to implementing the MMLR approach for the SDM, the regression ability of the MMLR method (without random noise) was compared to that of the CCR, a popular latent variable multivariate regression (Burnham et al. 1996). Fig. 3 shows average R-squares for all T_{max} and T_{min} predictands of the MMLRn with selected predictors, and for CCR with a different number of canonical variables where the R-squares are calculated from the daily deterministic time series. The R-squares of CCR increased as the number of canonical variables increased during the calibration period, whereas in the validation period R-square reached a maximum with 8 canonical variables. Although MMLRn performed slightly better than CCR during validation period based on the study area and data, the difference in R-squares between the 2 approaches was modest. This study adopted the MMLR model because it is theoretically simpler than CCR. However, linear approaches based on canonical correlation analysis have frequently been employed to statistical downscaling problems (see Huth 2002, Busuioc et al. 2008). The multicollinearity among predictors is eliminated in the canonical variables of CCR. Compared to this, the MMLR could be unwieldy in predictor selection.

4. RESULTS

In this section, the results of MMLRc and MMLRi are provided separately for cross-site correlations of

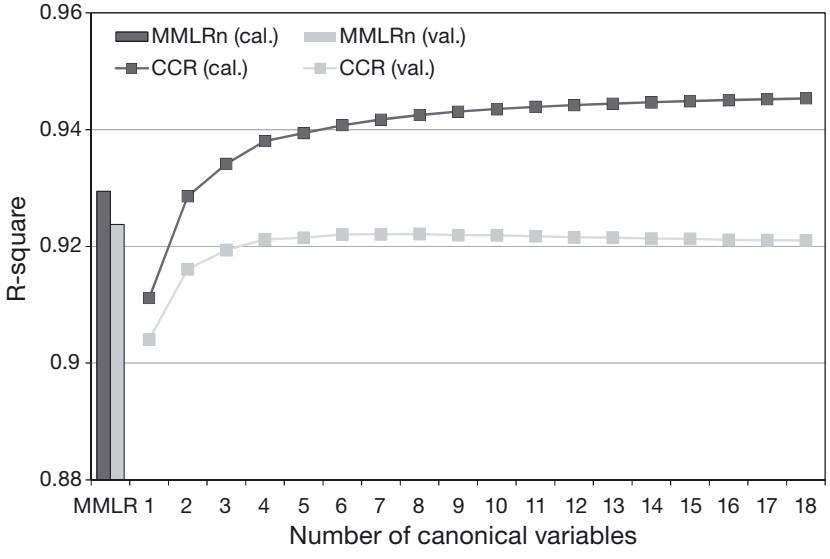

Fig. 3. Comparison of average R-squares over all predictands (i.e. T_{max} and T_{min} at 9 observation stations) between the MMLR without random noise (MMLRn) and canonical correlation regression (CCR) with different numbers of latent variables. R-squares were calculated from daily deterministic series for the entire calibration period 1961–1990 (cal.) and validation period 1991–2000 (val.)

T_{max} and T_{min} data, at-site correlation between T_{max} and T_{min}, Fr-Th, and DTR at each observation site. On the other hand, the results of the MMLRc and MMLRi are not provided separately for the monthly means and SDs of downscaled daily T_{max} and T_{min} series, $T_{max}90$, $T_{min}10$, frost season length (FSL), and GSL, because, as noted above, both models yielded nearly the same results in these cases.

Generally, when a model employs a randomization procedure, producing stable and robust results, this requires up to 100 or more repeated simulations. However, in our case the randomization part added a relatively small component to deterministic series of MMLRn. Therefore, this study provides the results of only one simulation for MMLRc and MMLRi, in order to compare the different model results more effectively and to avoid confusion. Before adding the random noise, a normality test was conducted for the residual series of the MMLRn model (data not shown): 71% of cases of residuals of the MMLR among a total of 216 cases (daily T_{max} and T_{min} × 12 months × 9 sites) fitted a normal distribution (from the Chi-square test, using the 90% confidence level).

4.1. Downscaled time series accuracy for each site

The downscaled time series of daily T_{max} and T_{min} from the MMLRi and the MMLRn were compared to the observed time series of the 2 variables for each site. Fig. 4 shows the R-squares of MMLRn and MMLRi for daily T_{max} and T_{min} over the 9 observation sites during the calibration and the validation periods. For MMLRn, R-square for T_{max} ranged from 0.928 to 0.946 during the calibration period and from 0.924 to 0.942 during the validation period. The R-squares of the corresponding T_{min} series were slightly smaller than those of the T_{max} series. The differences in R-squares between the calibration and the validation periods of the T_{max} and T_{min} series downscaled by the MMLRn and MMLRi were less than 2% (i.e. 0.02 of R-squares) over all sites. MMLRc basically yields the same results as those of MMLRi for this index (not shown) because both models employ same amount of random noise with respect to each predictand. Similar R-squares of T_{max} and T_{min} based on a linear transfer function and reanalysis predictors were previously reported by Hessami et al. (2008) over other sites in eastern Canada: R-squares ranged from 0.91 to 0.96 for T_{max} and from 0.91 to 0.95 for T_{min} in their applications. In our study, R-squares of the MMLRi for T_{max} and T_{min} were approximately 4.8 to 9.7% lower than for the MMLRn during both the calibration and the validation periods because added noises to the MMLRn T_{max} or T_{min} series decreased the temporal correlation between the MMLRi series and the observed series. R-squares for T_{max} were higher than average at Stns 4, 5, and 8, located in the middle of study area, with Stn 5 showing the highest values. R-squares for T_{max} were below average at Stns 6, 9, and 3, located in the south or northeast parts of the study area, with Stn 3 (Sept-Îles) showing the lowest values. These low values may be due to local influences being more

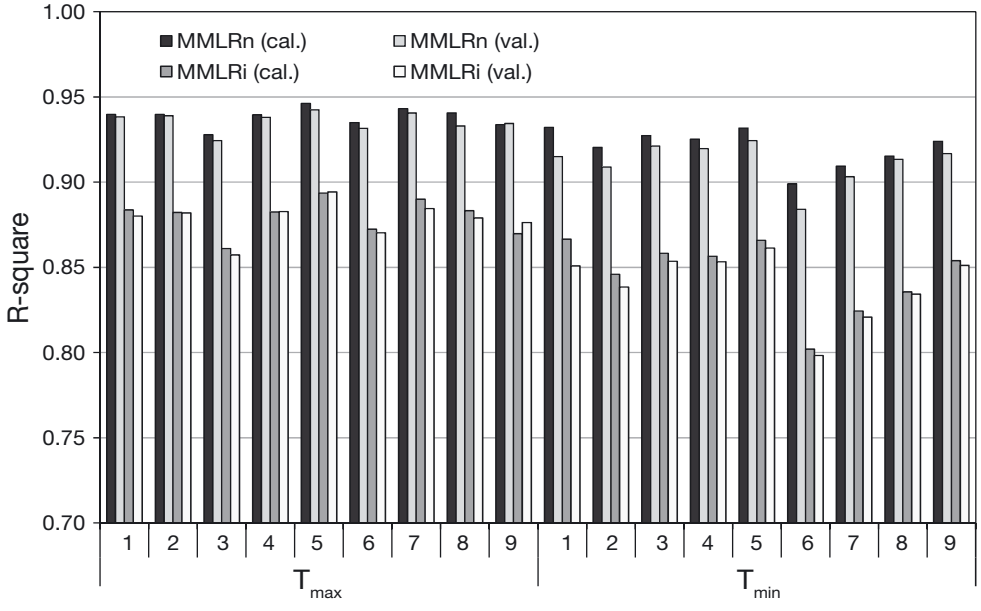

Fig. 4. The R-squares (annual scale) of the MMLRn and MMLRi for 9 observation stations for the entire daily series during the calibration period 1961–1990 (cal.) and validation period 1991–2000 (val.)

important at these 3 stations than the others, because of the topographical or land/sea variability or the presence of features such as the Appalachian mountains (Stn 6) and the Gulf of St. Lawrence (Stns 3 and 9). Stns 4 and 5, located in the middle of the study area, show higher than average R-squares for T_{min}. In contrast to their T_{max} results, Stns 3 and 9, located in the northeast portion of the study area, presented higher than average R-squares for T_{min}.

The variances of errors between observations and the series downscaled by the monthly based MMLRn or those by MMLRi (and MMLRc) varied more widely between months than between sites. One main cause for these monthly differences of the error variances was within-month natural variability for both daily T_{max} and T_{min} time series. To exemplify this, Fig. 5 shows box-whisker plots of observed T_{max} at one single station, Cedars (Stn 1), and the RMSE of MMLRn and MMLRi for each month during the calibration period. As shown from the RMSE values, the main source of error is within the variance because the MMLRn yields unbiased estimations from predictors and the MMLRi adds random noise with zero mean to the series of MMLRn. The differences in RMSEs between MMLRn and MMLRi were larger in winter than in summer, and the interquartile range (IQR) and extreme-value ranges were also larger in winter (December to February) than in summer (June to August). In fact, the monthly variability of RMSE originates from the monthly variability of T_{max} rather than from differences in the accuracy of the MMLR model (i.e. R-squares) for each month. Actually, R-squares of MMLRn in winter were 8% higher on average than those in summer (data not shown). The monthly variability of RMSE (or error variances) of MMLRn implies that the MMLRi and MMLRc require random series with different monthly variances.

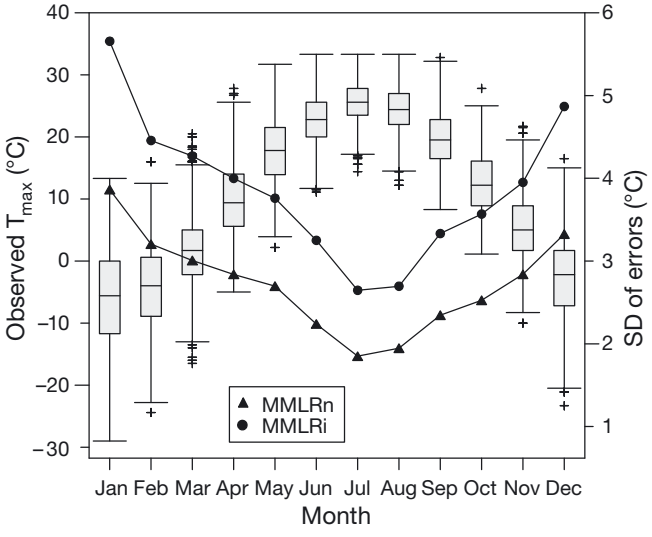

Fig. 5. Box-whisker plots of observed T_{max} at Cedars (Stn 1, see Fig. 2), and the root mean squared error (RMSE) of MMLRn and MMLRi, for each month during the calibration period. The box-whisker plot shows minimum, lower quartile (LQ), median, upper quartile (UQ), and maximum values. Outliers shown by crosses are identified when a value falls below LQ – 1.5 × IQR or above UQ + 1.5 × IQR, where IQR is the interquartile range

4.2. Spatial and temporal dependency of T_{max} and T_{min}

The relative performances of the 3 downscaling models in reproducing spatial coherence were as expected based on the theoretical structures of the models. MMLRn reproduces deterministic series using linear relationships between predictands and predictors for each individual site. Consequently, the MMLRn will most probably overestimate the spatial correlation between sites. For the MMLRi, total variance of predictand at each site is reproduced by adding uncorrelated random noise to the deterministic series of MMLRn. The uncorrelated random series reduce the original spatial coherence of deterministic series. For the MMLRc, the spatial correlation is properly reproduced by adding correlated random series to the deterministic series. For this reason, the MMLRc should appropriately reproduce the spatial correlation for daily T_{max} and T_{min} between sites.

Fig. 6 shows the cross-site correlation between pairs of daily temperatures (T_{max} or T_{min}) versus the distance between stations, for all combinations of station pairs, for the calibration and the validation periods. As anticipated, MMLRn and the MMLRi did not reproduce the correlation accurately. The difference in the cross-site correlation coefficients between the MMLRn and observation increased as a function of distance between sites. In contrast, the difference in the cross-site correlation coefficients between the MMLRi and observations increased as the distance between the sites decreased. For daily T_{max}, differences in the averaged correlation coefficients between the model and observation during the validation period were 0.028 and 0.031 for MMLRn and MMLRi respectively. T_{min} showed similar differences. The overestimation of the cross-site correlations by MMLRn is not surprising because MMLR predicts only the deterministic components from the atmospheric predictor variables at multiple sites. Cannon (2009) and Harpham & Wilby (2005) reported the

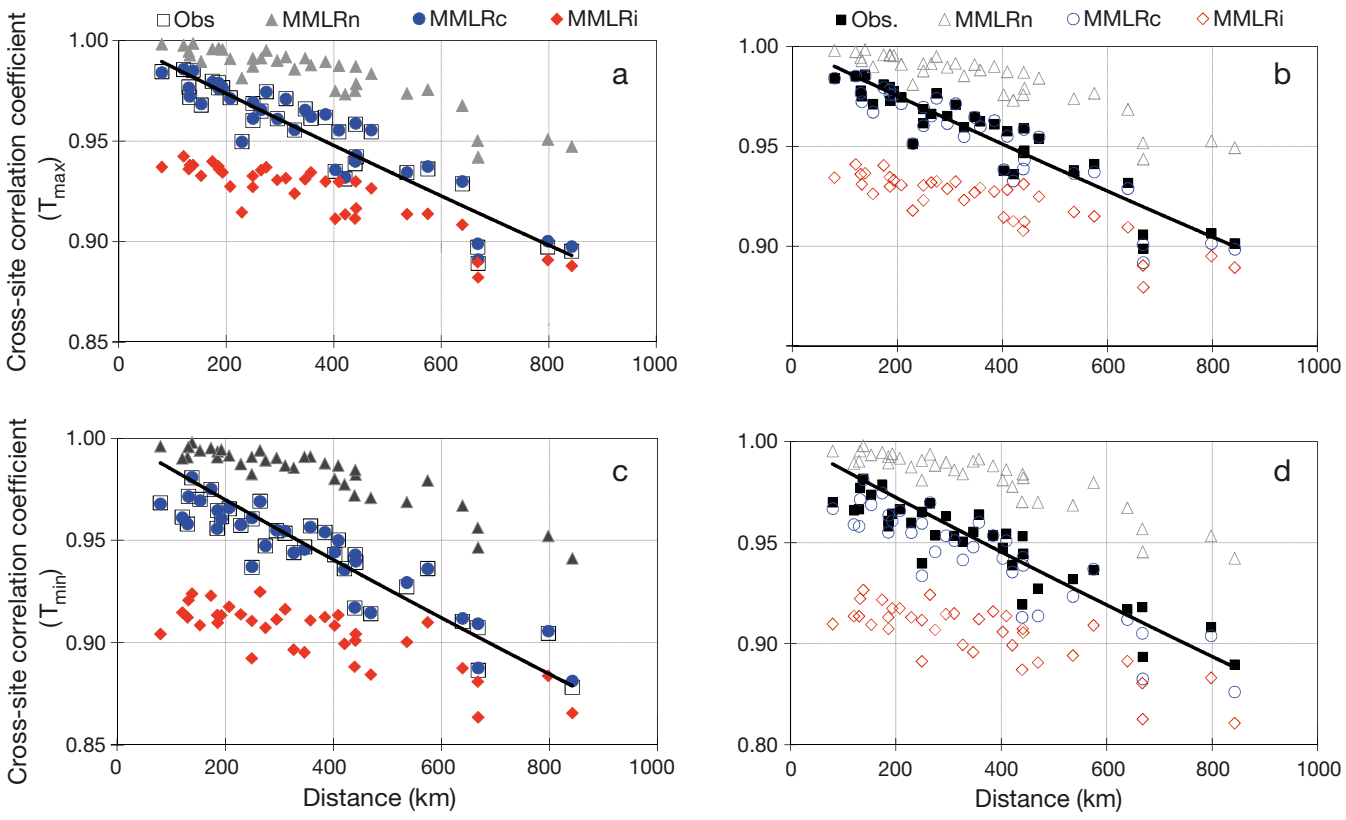

Fig. 6. Cross-site correlation coefficients of the MMLRn, MMLRc, and MMLRi between pairs of daily temperatures. The black solid line represents exponential decay functions fitted to the observation values. (a) T_{max} calibration period (1961–1990), (b) T_{max} validation period (1991–2000), (c) T_{min} calibration period, (d) T_{min} validation period

overestimation of cross-site correlations among multi-site temperature and/or precipitation series downscaled by transfer function techniques without randomization. By adding spatially uncorrelated random noises to the MMLRn of each site, MMLRi yields much smaller cross-site correlation coefficients than MMLRn. However, MMLRi cannot reproduce the observed cross-site correlations. Hence, the cross-site correlation coefficient values of the MMLRc showed better agreement with those of observations over the 2 periods. However, during the validation period, MMLRc slightly underestimated the cross-site correlation values of observations for both T_{max} and T_{min}, as these have changed over time with slightly larger values in the validation period than during the calibration period. MMLRc does not capture this difference and yields almost the same values for both periods. This is mainly due to this difference not being predicted by MMLRn from reanalysis predictors, and then this is propagated in the MMLRc model as the covariance structure of random noise in the validation period, inherited from the calibration period.

Fig. 7 shows at-site correlation between T_{max} and T_{min} over all observation sites for the MMLRn, MMLRc, and MMLRi. The MMLRn over-estimated the correlation values during the calibration and the validation periods. The MMLRi underestimated the observed correlations, but the MMLRc reproduced the at-site correlation between T_{max} and T_{min} quite well during both the calibration and the validation periods. The reasons for the over and under estimations by MMLRn and MMLRi of this statistic are the same as for the cross-site correlations. MMLRc, however, slightly underestimated the correlation at Stns 2, 4, 5, 6, 7 and 8 for the validation period. Observed correlations between the calibration and validation periods were similar except for Stns 6 and 7. However, MMLRn yielded smaller correlation values for the validation period than for the calibration period. Consequently, MMLRc showed a similar pattern of results between the two periods. This is mainly due to the deterministic component of the MMLR model as was the case for the spatial correlation. On average, the differences in averaged correlation coefficients between observation and the model generations were 0.042 for the MMLRn, 0.003 for the MMLRc, and 0.024 for the MMLRi during the validation period.

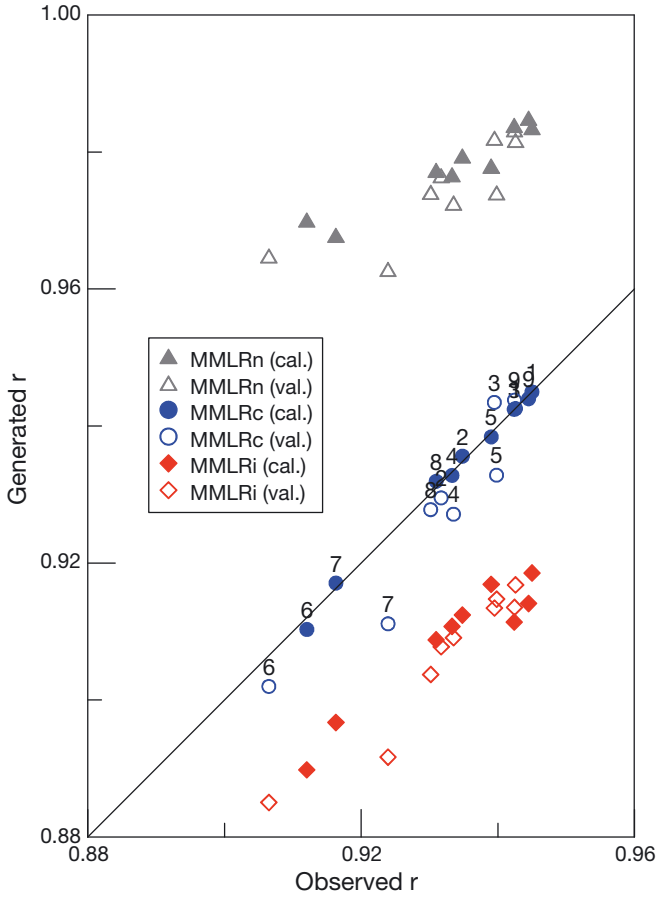

Fig. 7. Scatter plots of correlations between daily T_{max} and T_{min} observed at 9 stations versus correlations for the same stations modeled by MMLRn, MMLRc, and MMLRi. Solid symbols are the values of the calibration period and open symbols are the values of the validation period. Numbers represent the observation station numbers (see Fig. 2). Line: slope = 1

4.3. Statistics and climate indices

The monthly mean and SD of the daily T_{max} and T_{min} and climate indices shown in Table 3 were evaluated. As expected from theory, for monthly mean T_{max} and T_{min}, the MMLRn model performed clearly better than the randomization approaches (i.e. MMLRc and MMLRi), reflecting the better temporal agreement between MMLRn series and the observed series compared with the randomization approaches. On the other hand, the randomization approaches gave clearly better results for monthly SD T_{max} and T_{min}, $T_{max}90$, $T_{min}10$ than MMLRn, because they reproduce total variance of predictands better by adding random series to the deterministic series of MMLRn model.

Table 4 shows the bias, SE, and RMSE of the MMLRn and the MMLRc for each statistic and index for all observation sites during the calibration and the validation periods. The results for the MMLRi are not shown in Table 4, but the relevant results for Fr-Th and DTR are discussed in Figs. 10 & 11. The MMLRc showed lower RMSEs for SD T_{max}, SD T_{min}, $T_{max}90$, $T_{min}10$, and Fr-Th than the MMLRn for the validation period. For the same period, the MMLRn performed better in reproducing mean T_{max}, mean T_{min}, FSL, GSL, and DTR than the MMLRc. For the calibration period, the biases, SEs, and RMSEs of all statistics and climate indices showed almost identical patterns as those for the calibration period. This result indicates that the MMLRc can reproduce total variance of T_{max} and T_{min} series better than the MMLRn, although the latter is better at reproducing central tendencies and duration indices.

Table 4. Biases, standard deviations of errors (SEs), and root mean square errors (RMSEs) of the estimated statistics and climatic indices shown in Table 3, produced by multivariate multiple linear regression (MMLR) models MMLRn and MMLRc (for details see Section 2) during the calibration period (1961–1990) and validation period (1991–2000). The statistics and index values were calculated every year from 9 observation stations in Quebec, Canada (see Fig. 2) at the time scales shown in Table 3. **Bold:** smaller RMSEs between the 2 models. FSL: frost season length; GSL: growing season length; Fr-Th: frost and thaw cycle; DTR: diurnal temperature range

	Calibration period						Validation period					
	Bias		SE		RMSE		Bias		SE		RMSE	
	MMLRn	MMLRc	MMLRn	MMLRc	MMLRn	MMLRc	MMLRn	MMLRc	MMLRn	MMLRc	MMLRn	MMLRc
Mean T_{max} (°C)	−0.001	0.015	0.655	0.857	**0.655**	0.857	−0.212	−0.205	0.759	0.967	**0.788**	0.988
SD T_{max} (°C)	−0.946	0.073	0.542	0.779	1.091	**0.783**	−0.896	0.118	0.562	0.742	1.058	**0.751**
Mean T_{min} (°C)	−0.020	0.002	0.807	1.012	**0.807**	1.012	−0.322	−0.348	0.978	1.127	**1.030**	1.179
SD T_{min} (°C)	−1.012	0.106	0.693	0.857	1.227	**0.863**	−0.946	0.203	0.758	0.934	1.213	**0.956**
$T_{max}90$ (°C)	−1.013	0.043	0.946	1.206	1.386	**1.207**	−1.282	−0.273	1.029	1.323	1.644	**1.350**
$T_{min}10$ (°C)	1.081	0.042	1.445	1.598	1.805	**1.598**	0.410	−0.543	1.836	1.585	1.881	**1.675**
FSL (d)	−1.396	−0.764	15.424	16.634	**15.487**	16.651	5.407	4.148	18.673	21.268	**19.440**	21.669
GSL ()	1.205	0.184	14.251	14.883	**14.301**	14.884	1.956	2.000	16.794	17.748	**16.908**	17.861
Fr-Th (d)	−0.186	−0.153	2.568	2.533	2.575	**2.494**	0.069	−0.183	2.847	2.829	2.848	**2.783**
DTR (°C)	0.004	0.157	0.557	0.710	**0.557**	0.727	0.101	0.315	0.628	0.714	**0.636**	0.781

Box-whisker plots were used to compare observed monthly SD of T_{max} and T_{min} and of percentiles ($T_{max}90$ and $T_{min}10$) to those of downscaling models. The box-whisker plots show results for representative Stn 4 (Bagotville A, see Fig. 2), which is located in the middle of the study area. Fig. 8, plots the observed values of the monthly SD of T_{max} and downscaled values produced by the MMLRn and the MMLRc in the validation period 1991 to 2000, for 6 months: February, April, June, August, October, and December. The MMLRn underestimated this statistic for these months. Harpham & Wilby (2005) reported the underestimation of SD of daily precipitation series downscaled by artificial neural network techniques without randomization and the underestimation was up to 35% in their study area. The MMLRc showed better reproduction ability for the lower and upper quartiles and median of observed values than MMLRn for all months. Note that the main source of the RMSE of the MMLRn model is bias whereas that of the MMLRc is SE (see Table 4). The box-whisker plots of monthly SD of T_{min} for observations and for both models (not shown) were similar to those shown in Fig. 8.

Fig. 9 shows the box-whisker plots of seasonal $T_{max}90$ and $T_{min}10$ for the observed values and downscaled values produced by the MMLRn and the MMLRc on Stn 4 during the validation period. Fig. 9a–d shows that MMLRn underestimated the observed $T_{max}90$. However, the MMLRc displayed a much smaller bias and reproduced this extreme index quite well. As shown in Table 4, the RMSEs of the MMLRc were smaller than those of the MMLRn for all sites for both the calibration and the validation periods, even though it had larger SEs than the MMLRn. Results for $T_{min}10$ (Fig. 9e–h) were basically similar to those for $T_{max}90$. The MMLRn tended to overestimate $T_{min}10$, compared to the observations. Note that, for this index, the RMSEs of the MMLRc were smaller than those of the MMLRn for all sites for both the calibration and the validation periods (see Table 4).

As shown in Table 4, the MMLRn model slightly outperformed the MMLRc for FSL and GSL. These 2 duration indices might be mainly affected by the mean tendency of temperature rather than daily temperature variability. Thus, better performance can be expected from the MMLRn than the MMLRc for these 2 indices. The MMLRn predicts deterministic components of T_{max} and T_{min} using linear regression and the independent atmospheric predictors, and the

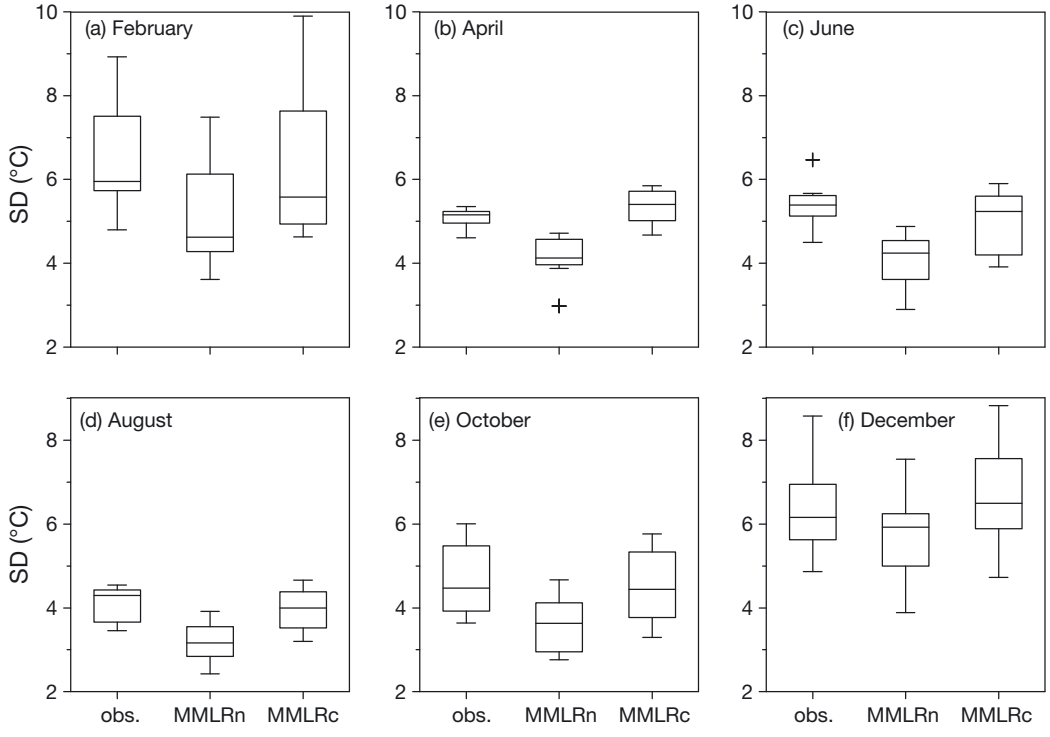

Fig. 8. Box-whisker plots of the monthly standard deviation of T_{max} for the observed values and downscaled values produced by the MMLRn and the MMLRc at Stn 4 (Bagotville A, see Fig. 2) during the validation period. For brevity, only results for (a) February, (b) April, (c) June, (d) August, (e) October and (f) December are shown. The box-whisker plot shows minimum, lower quartile (LQ), median, upper quartile (UQ), and maximum values. Outliers shown by crosses are identified when a value falls below LQ – 1.5 IQR or above UQ + 1.5 IQR, where IQR is the interquartile range

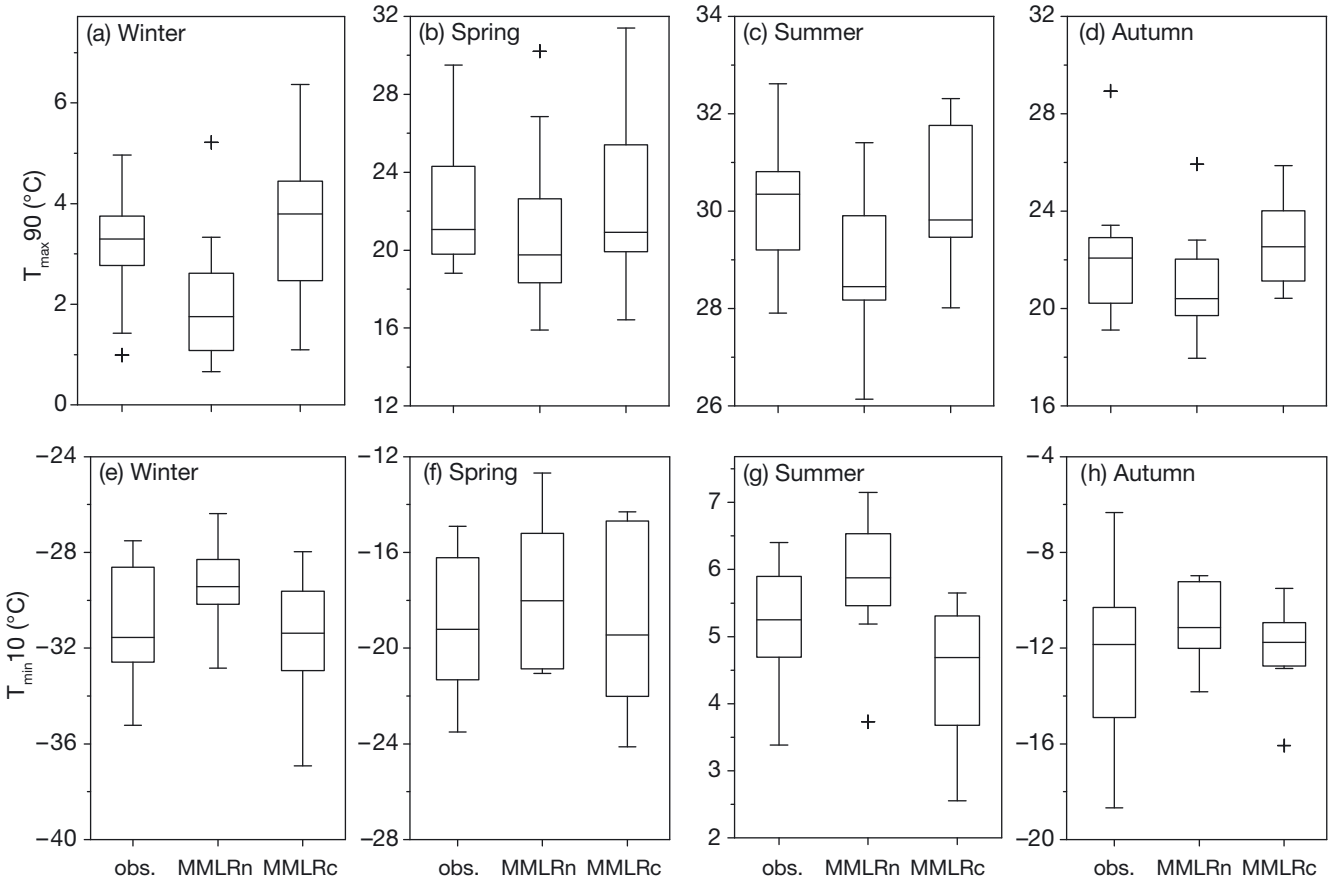

Fig. 9. Box-whisker plots of the seasonal (a–d) $T_{max}90$ and (e–h) $T_{min}10$, for observed values and downscaled values produced by MMLRn and MMLRc at Stn 4 (Bagotville A, see Fig. 2) during the validation period. For box-whisker plot definitions, see Fig. 8

temporal correlations between observations and downscaled series were larger for MMLRn than for MMLRi (see Fig. 4). The RMSEs of FSL were 19.4 d for the MMLRn and 21.7 d for the MMLRc during the validation period. However, the difference in RMSEs of this index between the 2 models is 2.3 d, which accounts for only 1.3% of the averaged observed FSL. The results for GSL were similar to those for FSL. The RMSEs of both models were modest at 16.9 d for the MMLRn and 17.9 d for the MMLRc during the validation period (data not shown).

For the DTR and Fr-Th, the results of MMLRc and MMLRi are provided separately because the correlated daily T_{max} and T_{min} series on each site are required to calculate the 2 indices. The scatter plots of observed versus downscaled DTR for MMLRn, MMLRc, and MMLRi are shown in Fig. 10a,b & c, respectively, for all sites during the validation period, and for the 4 seasons. Among the 3 models, the MMLRn performed best for this index. The biases were 0.101, 0.315, and 0.396°C, and the RMSEs were 0.639, 0.781, and 0.952°C for the MMLRn, MMLRc, and MMLRi, respectively, during the validation period. Without randomization, the DTR was reproduced most accurately by the MMLRn. MMLRn can be expected to perform best for this index because DTR is calculated for each season and the MMRLn series show stronger temporal correlations with observed daily T_{max} and T_{min} series than those of MMLRc at each site. Since MMLRn reproduced autocorrelation in each season more adequately, DTR results were better than for MMLRi and MMLRc. When randomized errors were added to the downscaled series by the MMLRn, the MMLRc reproduced this index better than the MMLRi, as the MMLRc accurately reproduced the at-site temporal correlation between T_{max} and T_{min}. The difference in RMSE between the 2 models was 0.140°C during the validation period. Observed and downscaled values of this index showed larger variances in winter and spring than those in summer and autumn. MMLRc could perform as well as MMLRn if it were able to adequately reproduce the correlation between daily T_{max} and T_{min} time series at each site.

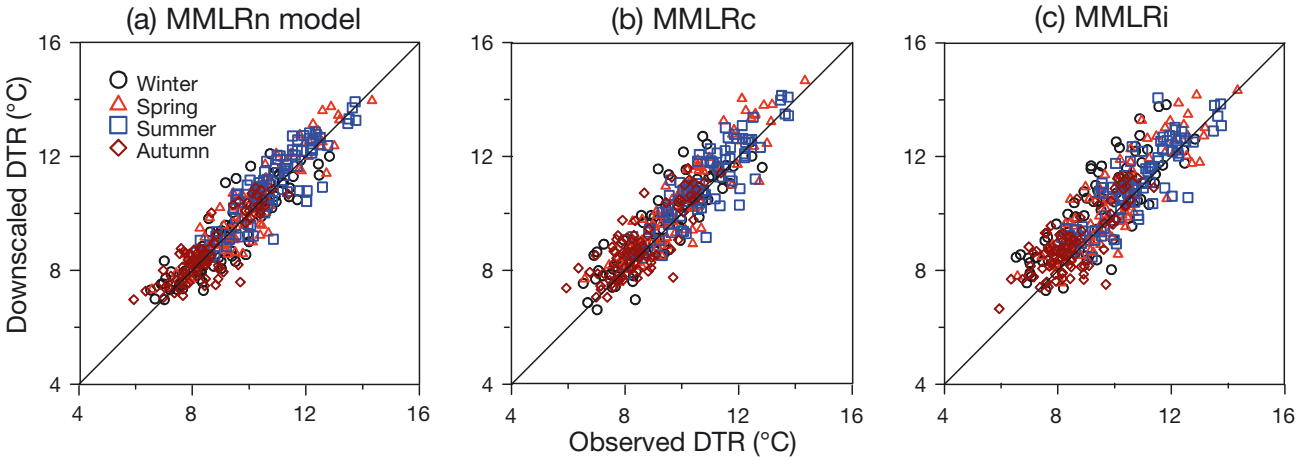

Fig. 10. Scatter plots of the observed diurnal temperature range (DTR) versus DTR values produced by (a) MMLRn, (b) MMLRc, and (c) MMLRi, for all months (symbols distinguish between seasons) and sites during the validation period (1991–2000)

Fig. 11a,b & c shows the bubble plots of the Fr-Th for the MMLRn, MMLRc , and MMLRi, respectively, for all sites during the validation period. The *x*-axis represents observed Fr-Th. RMSEs of the MMLRn, MMLRc, and MMLRi for this index were similar, with values of 2.848, 2.783, and 2.821 d, respectively, during the validation period. However, the MMLRn tended to underestimate this index when the observed Fr-Th was <10 d, and it tended to overestimate the Fr-Th when the observed value was within the range of 15 to 20 d. In the study area, the observed Fr-Th is relatively small during winter (e.g. December, January, and February) and from May to October. In contrast, the Fr-Th has a large value when most daily T_{max} and T_{min} values are near zero (e.g. March, April, and November). MMLRn produces temporal series that vary less than observed daily T_{max} and T_{min} series. Therefore, MMLRn can be expected to yield underestimated values during months with small Fr-Th and overestimated values when Fr-Th monthly values are large. Furthermore, the Fr-Th is more affected by the daily temperature variability than FSL and GSL around the threshold of 0°C (i.e. Fr-Th is strongly sensitive to the concomitant daily T_{max} and T_{min} variability), as these 2 latter indices are only dependent on a fixed threshold of T_{min} or T_{mean}, respectively.

5. CONCLUSIONS

This study describes a MMSDM procedure based on a hybrid structure combining a transfer function downscaling approach and a stochastic randomization procedure. The resulting MMSDM can downscale multiple predictands at multiple observation sites simultaneously, and can therefore be a multisite and multivariate extension of previous single site and single variable statistical downscaling models (e.g. Wilby et al. 2002, Hessami et al. 2008). In the MMSDM, the MMLR was employed for simultaneously downscaling daily T_{max} and T_{min} at 9 observation sites in southern Quebec, Canada. The unexplained total variance, spatial (i.e. inter-site) dependency of a single predictand and the temporal dependency of multiple predictands at a given site were reproduced by adding correlated random noise generated from multivariate normal distribution to the deterministic series downscaled using the MMLR approach. The downscaled daily T_{max} and T_{min} series generated by the MMSDM, called MMLRc in this study, were compared to the observed series for the following 3 aspects: (1) spatial dependency of T_{max} and T_{min}, (2) at-site correlation between T_{max} and T_{min}, and (3) monthly statistics and climate extreme indices. The results of the MMLRc were also compared to those of MMLR without random noise (MMLRn) and MMLR with uncorrelated random noise (MMLRi) to determine the reliability of the stochastic randomization procedure in the MMSDM.

The MMLRn showed better agreement than MMLRc and MMLRi with observed predictands during the calibration and the validation periods, although it underestimated the total variance of predictands. The R-squares between observations and series downscaled by the randomization approach models (i.e. MMLRc and MMLRi) were 4.8 to 9.7% lower than those downscaled by the MMLRn because of the random noise added by the statistical randomization procedure. This result simply implies

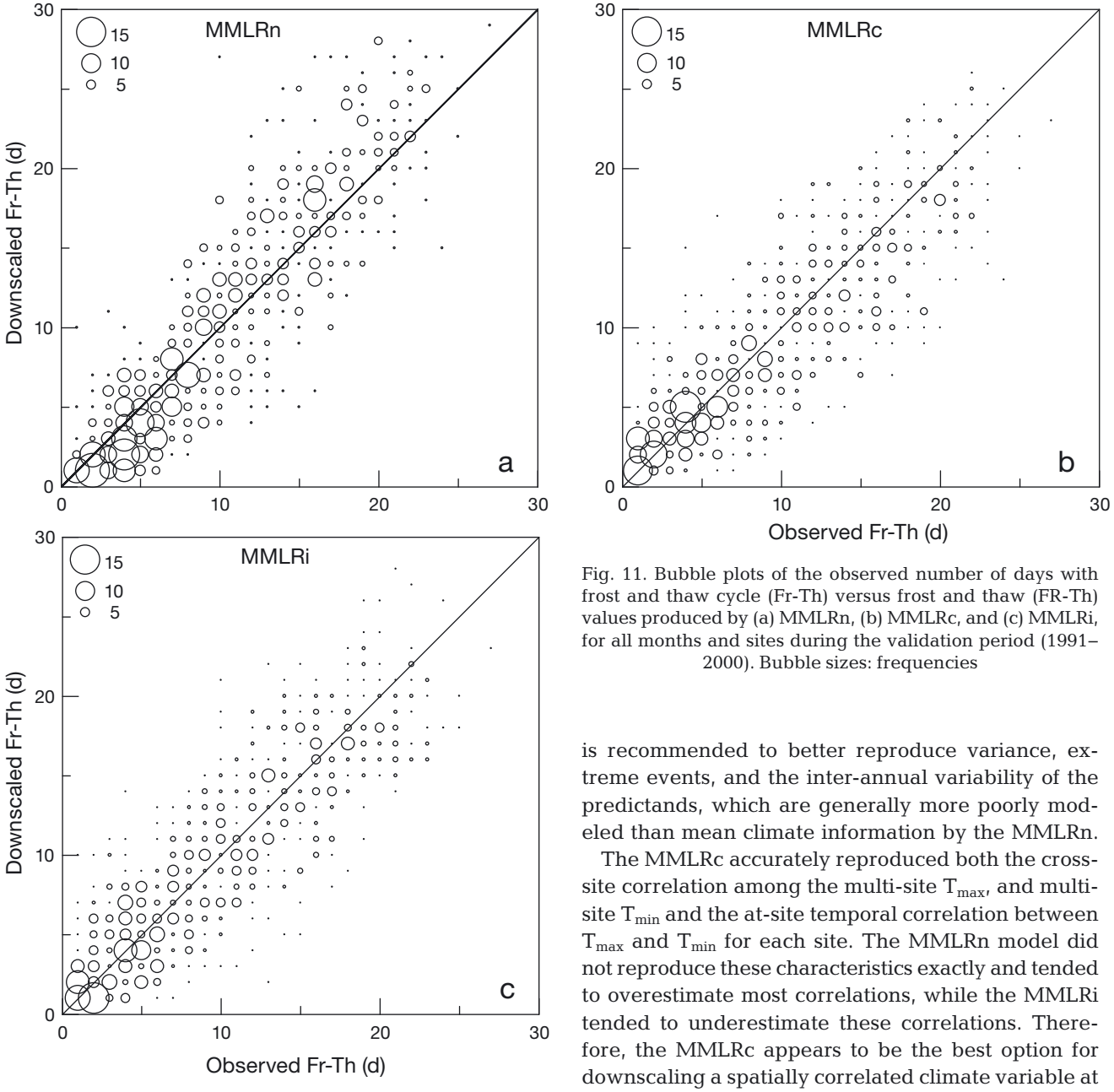

Fig. 11. Bubble plots of the observed number of days with frost and thaw cycle (Fr-Th) versus frost and thaw (FR-Th) values produced by (a) MMLRn, (b) MMLRc, and (c) MMLRi, for all months and sites during the validation period (1991–2000). Bubble sizes: frequencies

that downscaled predictand series generated by the randomization approach are less consistent with the observed series for a single site; i.e. they may be less consistent with the NCEP/NCAR reanalysis predictors and with future AOGCM atmospheric predictors using methods other than the MMLRn. Hence, the linear-regression downscaling approach without a randomization procedure may serve as an alternative to downscaling a predictand, a method more consistent with AOGCM predictors, and it could be used to project the averaged central tendency of a predictand. However, the use of a randomization procedure is recommended to better reproduce variance, extreme events, and the inter-annual variability of the predictands, which are generally more poorly modeled than mean climate information by the MMLRn.

The MMLRc accurately reproduced both the cross-site correlation among the multi-site T_{max}, and multi-site T_{min} and the at-site temporal correlation between T_{max} and T_{min} for each site. The MMLRn model did not reproduce these characteristics exactly and tended to overestimate most correlations, while the MMLRi tended to underestimate these correlations. Therefore, the MMLRc appears to be the best option for downscaling a spatially correlated climate variable at multiple observation sites, and/or temporally correlated multiple climate variables at a single observation site.

The MMLRc and MMLRi performed better than the MMLRn model in reproducing monthly SDs of daily T_{max} and T_{min}, as well as $T_{max}90$ and $T_{min}10$. These results are reasonable because the downscaled daily T_{max} and T_{min} series generated by the 2 randomization approach models represent the variability of the observed daily T_{max} and T_{min} series better than those by the MMLRn model. However, the 2 randomization approach models yielded larger standard errors than the MMLRn for the monthly mean of

daily T_{max} and daily T_{min}. This suggests that the randomization approach might be more appropriate for the reproduction of extreme events (e.g. annual maximum and minimum temperatures, $T_{max}90$ and $T_{min}10$) of a predictand than the simplified regression model. For the FSL, GSL, and DTR, the MMLRn model performed better than the other 2 models. For the 2 duration indices (i.e. FSL and GSL), a regression approach may be preferable to the randomization models, because the former reproduces temporal variations, such as seasonality, in the observed daily T_{max} and T_{min} time series from series that are more closely correlated with observations than those produced by randomization models. However, the differences in the indices between randomization and regression approaches were modest: differences between the RMSEs of MMLRn and MMLRc for these 2 indices were <1.3% of the averaged values of the observed indices for the validation period. Although MMLRn performed best for mean DTR, the MMLRc also performed fairly well for this index, and for Fr-Th, because the MMLRc accurately reproduced at-site temporal correlation between daily T_{min} and T_{max} series. The MMLRc also performed better for these indices than the MMLRi. The MMLRc is therefore relevant when the objective is to downscale temporally correlated multiple climate variables at an observation site from AOGCM predictors.

The randomization procedure was adopted for the MMSDM because it can easily reproduce the spatial coherence among multiple observation sites and correlation between multiple variables. Cannon (2009) and Bürger & Chen (2005) reported that the variance inflation approaches misrepresent spatial correlation for multi-site downscaling of daily temperature and/or precipitation because the variance inflation factors in the transfer function do not account for the spatial dependence between sites. In this respect, variance inflation is not within the scope of this study, which focuses on multi-site downscaling. However, variance inflation approaches can be expected to perform better than MMLRn, and similarly to MMLRc, in reproducing the variance and extreme percentiles of daily temperatures for single-site downscaling.

In future work, the performance of this randomization approach will be evaluated by comparing it to the variance inflation approach in terms of at-site downscaling ability for temperatures. The proposed MMLRc model will be tested for other predictands such as daily precipitation series from multiple observation sites. It is a well known problem that transfer function models with atmospheric predictors can explain only 20 to 40% of observed variance of daily precipitation (see Wilby et al. 2002, Hessami et al. 2008 for reviews). Therefore, larger variance of random noise should be required for precipitation than for the temperature fields. Additional weather generator techniques such as Markov chain model and probability distribution functions might be employed to properly downscale daily precipitation series using the MMLRc.

Acknowledgements. We acknowledge the financial support provided of the National Science and Engineering Research Council (NSERC) of Canada, Special Research Opportunity (SRO) program. We are also grateful to L. Vincent and E. Mekis from Environment Canada for providing observed data sets of homogenized temperature. The authors also acknowledge the Data Access Integration (DAI, see http://loki.qc.ec.gc.ca/DAI/) Team for providing the predictors data and technical support. The DAI data download gateway is made possible through collaboration among the Global Environmental and Climate Change Centre (GEC3), the Adaptation and Impacts Research Division (AIRD) of Environment Canada, and the Drought Research Initiative (DRI).

LITERATURE CITED

Apipattanavis S, Podestá G, Rajagopalan B, Katz RW (2007) A semiparametric multivariate and multisite weather generator. Water Resour Res 43:W11401, doi:10.1029/2006WR005714

Boé J, Terray L, Habets F, Martin E (2007) Statistical and dynamical downscaling of the Seine basin climate for hydrometeorological studies. Int J Climatol 27:1643–1655

Buishand TA, Brandsma T (2001) Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling. Water Resour Res 37: 2761–2776

Bürger G, Chen Y (2005) Regression-based downscaling of spatial variability for hydrologic applications. J Hydrol (Amst) 311:299–317

Burnham AJ, Viveros R, Macgregor JF (1996) Frameworks for latent variable multivariate regression. J Chemometr 10:31–45

Busuioc A, Giorgi F, Bi X, Ionita M (2006) Comparison of regional climate model and statistical downscaling simulations of different winter precipitation change scenarios over Romania. Theor Appl Climatol 86:101–123

Busuioc A, Tomozeiu R, Cacciamani C (2007) Statistical downscaling model for winter extreme precipitation events in Emilia-Romagna region. Int J Climatol 28: 449–464

Busuioc A, Tomozeiu R, Cacciamani C (2008) Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region. Int J Climatol 28:449–464

Cannon AJ (2009) Negative ridge regression parameters for improving the covariance structure of multivariate linear downscaling models. Int J Climatol 29:761–769

Crimmins MA (2006) Synoptic climatology of extreme fire-weather conditions across the southwest United States. Int J Climatol 26:1001–1016

Dibike YB, Gachon P, St-Hilaire A, Ouarda TBMJ, Nguyen

VTV (2008) Uncertainty analysis of statistically downscaled temperature and precipitation regimes in Northern Canada. Theor Appl Climatol 91:149–170

Enke W, Spekat A (1997) Downscaling climate model outputs into local and regional weather elements by classification and regression. Clim Res 8:195–207

Frías MD, Zorita E, Fernández J, Rodríguez-Puebla C (2006) Testing statistical downscaling methods in simulated climates. Geophys Res Lett 33:L19807. doi:10.1029/2006GL027453

Gachon P, Dibike Y (2007) Temperature change signals in northern Canada: convergence of statistical downscaling results using two driving GCMs. Int J Climatol 27: 1623–1641

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

Gachon P, St-Hilaire A, Ouarda TBMJ, Nguyen VTV and others (2005) A first evaluation of the strength and weaknesses of statistical downscaling methods for simulating extremes over various regions of eastern Canada. Subcomponent, Climate Change Action Fund (CCAF), Environment Canada, Montreal

Harding AE, Gachon P, Nguyen VTV (2011) Replication of atmospheric oscillations, and their patterns, in predictors derived from Atmosphere-Ocean Global Climate Model output. Int J Climatol 31:1841–1847

Harpham C, Wilby RL (2005) Multi-site downscaling of heavy daily precipitation occurrence and amounts. J Hydrol (Amst) 312:235–255

Hellström C, Chen D, Achberger C, Raisanen J (2001) Comparison of climate change scenarios for Sweden based on statistical and dynamical downscaling of monthly precipitation. Clim Res 19:45–55

Hessami M, Gachon P, Ouarda TBMJ, St-Hilaire A (2008) Automated regression-based statistical downscaling tool. Environ Model Softw 23:813–834

Hewitson BC, Crane RG (1996) Climate downscaling: techniques and applications. Clim Res 7:85–95

Huth R (2002) Statistical downscaling of daily temperature in central Europe. J Clim 15:1731–1742

Huth R (2004) Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors. J Clim 17:640–652

Huth R, Kliegrová S, Metelka L (2008) Nonlinearity in statistical downscaling: Does it bring an improvement for daily temperature in Europe? Int J Climatol 28:465–477

IPCC (2001) Climate change 2001: the scientific basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge

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

Jeong DI, St-Hilaire A, Ouarda TBMJ, Gachon P (2012) CGCM3 predictors used for daily temperature and precipitation downscaling in southern Québec, Canada. Theor Appl Climatol 107:389–406

Kalnay E, Kanamitsu M, Kistler R, Collins W and others (1996) The NCEP/NCAR 40-year reanalysis project. Bull Am Meteorol Soc 77:437–471

Katz RW, Parlange MB (1995) Generalizations of chain-dependent processes: application to hourly precipitation. Water Resour Res 31:1331–1341

Khalili M, Leconte R, Brissette F (2007) Stochastic multisite generation of daily precipitation using spatial autocorrelation. J Hydrometeorol 8:396–412

Khalili M, Brissette F, Leconte R (2009) Stochastic multi-site generation of daily weather data. Stochastic Environ Res Risk Assess 23:837–849

Kistler R, Kalnay E, Collins W, Saha S and others (2001) The NCEP/NCAR 50-year reanalysis. Bull Am Meteorol Soc 82:247–267

Kozuchowski K, Wibig J, Maheras P (1992) Connections between air temperature and precipitation and the geopotential height of the 500 hPa level in a meridional cross-section in Europe. Int J Climatol 12:343–352

Mehrotra R, Sharma A (2007) Preserving low-frequency variability in generated daily rainfall sequences. J Hydrol (Amst) 345:102–120

Miksovsky J, Raidl A (2005) Testing the performance of three nonlinear methods of time series analysis for prediction and downscaling of European daily temperatures. Nonlinear Process Geophys 12:979–991

Mpelasoka FS, Mullan AB, Heerdegen RG (2001) New Zealand climate change information derived by multivariate statistical and artificial neural network approaches. Int J Climatol 21:1415–1433

Palutikof JP, Goodess CM, Wathkins SJ, Holt J (2002) Generating rainfall and temperature scenarios at multiple sites: example from the Mediterranean. J Clim 15: 3529–3548

Qian B, Corte-Real J, Xu H (2002) Multisite stochastic weather models for impact studies. Int J Climatol 22: 1377–1397

Rajagopalan B, Lall U (1999) A *k*-nearest-neighbor simulator for daily precipitation and other weather variables. Water Resour Res 35:3089–3101

Schoof JT, Pryor SC, Robeson SM (2007) Downscaling daily maximum and minimum temperatures in the midwestern USA: a hybrid empirical approach. Int J Climatol 27: 439–454

Stehlik J, Bardossy A (2002) Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. J Hydrol (Amst) 256: 120–141

Tolika K, Maheras P, Vafiadis M, Flocas HA, Arseni-Papadimitriou A (2007) Simulation of seasonal precipitation and raindays over Greece: a statistical downscaling technique based on artificial neural networks (ANNs). Int J Climatol 27:861–881

Trigo RM, Palutikof JP (1999) Simulation of daily temperature for climate change scenarios over Portugal: a neural network model approach. Clim Res 13:45–59

Trigo RM, Palutikof JP (2001) Precipitation scenarios over Iberia: a comparison between direct GCM output and different downscaling techniques. J Clim 14:4422–4446

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

von Storch H (1999) On the use of 'inflation' in statistical downscaling. J Clim 12:3505–3506

von Storch H, Hewitson B, Mearns L (2000) Review of empirical downscaling techniques. In: Iversen T, Høiskar BAK (eds) Regional climate development under global warming. General Tech Rep 4, Norwegian Meteorological Institute, Oslo, p 29–46

Wilby RL, Dawson CW, Barrow EM (2002) SDSM—a deci-

sion support tool for the assessment of regional climate change impacts. Environ Model Softw 17:147–159

Wilby RL, Tomlinson OJ, Dawson CW (2003) Multisite simulation of precipitation by conditional resampling. Clim 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, Intergovernmental Panel on Climate Change. www.ipcc-data.org/guidelines/index.html

Wilks DS (1998) Multisite generation of a daily stochastic precipitation generation model. J Hydrol (Amst) 210: 178–191

Wilks DS (1999a) Multisite downscaling of daily precipitation with a stochastic weather generator. Clim Res 11: 125–136

Wilks DS (1999b) Simultaneous stochastic simulation of daily precipitation, temperature, and solar radiation at multiple sites in complex terrain. Agric For Meteorol 96: 85–101

Wilks DS (2002) Realizations of daily weather in forecast seasonal climate. J Hydrometeorol 3:195–207

Xoplaki E, González-Rouco JF, Luterbacher J, Wanner H (2004) Wet season Mediterranean precipitation variability: influence of large-scale dynamics and trends. Clim Dyn 23:63–78

Xu C (1999) From GCMs to rivers flow: a review of downscaling methods and hydrologic modeling approaches. Prog Phys Geogr 23:229–249

Yates D, Gangopadhyay S, Rajagopalan B, Strzepek K (2003) A technique for generating regional climate scenarios using a nearest-neighbor algorithm. Water Resour Res 39:1199. doi:10.1029/2002WR001769

Zorita E, von Storch H (1999) The analog method as a simple statistical downscaling technique: comparison with more complicated methods. J Clim 12:2474–2489

Editorial responsibility: Balaji Rajagopalan, Boulder, Colorado, USA

Submitted: April 15, 2011; Accepted: April 10, 2012
Proofs received from author(s): August 14, 2012