

Spatial Bayesian Model for Statistical Downscaling of AOGCM to Minimum and Maximum Daily Temperatures

DOMINIQUE FASBENDER AND TAHA B. M. J. OUARDA

INRS-ETE, Quebec City, Quebec, Canada

(Manuscript received 8 September 2009, in final form 14 May 2010)

ABSTRACT

Atmosphere–ocean general circulation models (AOGCMs) are useful for assessing the state of the climate at large scales. Unfortunately, they are not tractable for the finer-scale applications (e.g., hydrometeorological variables). Downscaling methods allow the transfer of large-scale information to finer scales and they are thus relevant for the assessment of finer-scale variables. Among a wide range of downscaling methods, regression-based approaches are commonly used for downscaling AOGCM data because of their low computational requirements. However, downscaled variables are generally reproduced at gauged weather stations only. Results at the gauged stations can then be interpolated a posteriori at ungauged locations with kriging or other methods.

In this paper, a spatial Bayesian model is proposed for the downscaling of coarse-scale atmospheric data (i.e., either reanalysis or AOGCM) to minimum and maximum daily temperatures. This approach uses a Bayesian framework for mixing a prior distribution reflecting the monthly spatial dependence of the temperatures with the daily fluctuations induced by the atmospheric predictors. Local characteristics (i.e., altitude and latitude) are also taken into account in the mean of the prior distribution by using a geographical regression model. The posterior distribution thus reflects both monthly local patterns because of the prior and daily larger-scale fluctuations. Finally, the Bayesian approach also allows for the accounting of estimated parameter uncertainty, making it more stable to poor parameter fitting. The method is applied to the southern part of the province of Quebec, Canada. Results show that the downscaled distributions of the temperatures at gauged sites are in sufficient agreement with the validation dataset compared to a classical regression-based method. The proposed model has also the advantage of directly producing temperature maps.

1. Introduction and review

Over the last couple of decades, atmosphere–ocean general circulation models (AOGCMs) have been the main tools for assessing the evolution of the earth's climate. They allow the generation of scenarios for climate evolution at large scale for up to one or two centuries ahead. Unfortunately, the AOGCM data are generally produced on horizontal grids with a poor spatial resolution generally ranging from around $2.5^{\circ} \times 3.75^{\circ}$ [e.g., the third climate configuration of the Met Office Unified Model (HadCM3); see Gordon et al. 2000] to around $3.75^{\circ} \times 3.75^{\circ}$ [e.g., the Canadian Centre for Climate Modelling and Analysis (CCCma) Coupled General Circulation Model, version 3.1 (CGCM3.1); see Flato et al. 2000]. Although quite reasonable for large-scale

processes, this spatial resolution is not tractable for hydrological impacts and adaptation applications or modeling (i.e., skillful scale; see e.g., Grotch and MacCracken 1991; Huth and Kyselý 2000). There is thus a real need for methods and techniques that aim at enhancing the spatial resolution of such models to more relevant regional scales.

Precisely, statistical downscaling methods are dedicated to improve the spatial resolution. The main objective of these methods is to reproduce the local hydroclimatic variables (the predictands) at a finer scale from the temporal evolution of atmospheric variables (the predictors) available at a coarser scale mainly from AOGCM outputs. Among a wide range of techniques (see Benestad et al. 2008 for a detailed review), regression-based methods are commonly used for downscaling AOGCM data (see e.g., Wilby et al. 2002; Hessami et al. 2008; Jeong et al. 2010) because of both their ease of implementation and their low computational requirements (Hessami et al. 2008).

Corresponding author address: Dominique Fasbender, INRS-ETE, 490 de la Couronne, Quebec QC G1K 9A9, Canada.
E-mail: dominique.fasbender@ete.inrs.ca

Regression-based methods perform well for downscaling because the statistical relationships between the different predictors and the predictands are case specific. No matter the reasons for the observed differences between sites (e.g., different distances with respect to AOGCM grid points), the regression models will automatically adapt themselves. As it is the case for most empirical approaches, a major drawback of regression-based methods is that estimation is generally restricted to gauged sites, making it difficult to extend the results to ungauged locations (i.e., spatial estimation). The three main alternatives are (i) to interpolate the AOGCM predictors first on the gauged station locations before the regression (see Michelangeli et al. 2009), (ii) to interpolate the downscaled results (see e.g., Bárdossy and Van Mierlo 2000; Benestad 2002, 2007; Hundscha and Bárdossy 2005), or (iii) to interpolate the predictands at the gauged stations on the target regular grid (see Lim et al. 2007; Baigorria et al. 2008). When using the third alternative, the spatial and temporal patterns of both processes are captured first as empirical orthogonal functions (EOFs), which are then inserted in linear regression models in order to model the statistical relationships linking the predictors to the interpolated predictands. This third approach is quite powerful and reproduces adequately the observed climate anomalies. However, its main disadvantage is its large computational requirements that directly depend on the size of the target regular grid. Consequently, given the disadvantages of these three options, an alternative optimal solution would be to directly take into account the spatial dependence during the downscaling process.

Regression models generally reproduce the mean or the central part of the predictands. It is, however, more relevant to build the whole distribution in order to recapture the variability of the process as well. In this regard, probabilistic approaches have provided useful contributions in downscaling applications (see Hughes and Guttorp 1994; Hughes et al. 1999; Bates et al. 1998; Bellone et al. 2000; Vrac and Naveau 2007; Vrac et al. 2007). They enable us to model the whole distribution at once by modeling the influence of the AOGCM predictors on the parameters of the predictand distributions. Among probabilistic approaches, Bayesian formulations present a number of attractive characteristics (see e.g., Coelho et al. 2004; Tebaldi et al. 2004; Mendes et al. 2006). Bayesian approaches have a great potential in the context of downscaling applications as they enable us to integrate several sources of information and account for the uncertainty of the estimated parameters. However, these approaches have not gained wide acceptance, due mainly to their theoretical and practical difficulties. Moreover, spatial estimation at ungauged locations also remains an issue.

The main objectives of this paper are (i) to propose a spatial Bayesian model for the statistical downscaling of large-scale predictors to minimum and maximum daily temperatures, (ii) to apply the proposed model to a case study in order to produce maps of temperatures, and (iii) to make sure that (at least at the gauged stations) the proposed model provides similar results as other regression-based methods. The present work focused on using reanalysis products in the spatial downscaling model. Future efforts can focus on using AOGCM variables within the same context. The model relies on the Bayesian framework in order to combine a joint spatial model for the seasonal norms of both the minimum and maximum temperatures with the daily corrections induced by the atmospheric predictors. In this paper, the mean of the prior distribution is estimated with local characteristics (e.g., altitude and latitude) in a geographical regression model (GRM). The proposed model then mixes the finer scale monthly variations due to the GRM and the large-scale daily variations due to the atmospheric predictors. Consequently, the resulting posterior distribution reflects both the finer spatial scale and the daily evolutions. As a result of the Bayesian approach, prior distributions for the parameters of both the GRM and the large-scale regression are also accounted for in the model, thus making it more stable with respect to the uncertainty of the estimated parameters.

After a brief description of the notations, the proposed spatial Bayesian model is presented in detail with its properties and its possible future extensions. The proposed model is then applied to the case of minimum and maximum daily temperatures in the southern part of the province of Quebec, Canada. Reanalysis data are used here in order to assess the potential of the proposed method. The results are compared with those obtained with a multiple multivariate linear regression (MLLR) using the same set of predictors (see appendix A for the details of this model). We can thus assess the loss/gain of adequacy as a result of the spatial generalization of the Bayesian model. A couple of estimated maps directly provided by the Bayesian model are presented and the results are spatially validated on an independent dataset to assess the quality of the downscaled distributions.

2. Spatial Bayesian model

A spatial Bayesian model is presented in this section in order to circumvent the inability of classic regression methods to produce spatial estimations at ungauged sites. The general formulation as well as the modeling assumptions are described hereafter. A sampling procedure and potential generalizations to the method are also proposed in this section.

a. Notations

Let $\mathbf{x}_i \in \mathbb{R}^2$ be the coordinates of a specific location. At each location \mathbf{x}_i , we will consider m different predictands:

$$Z_{(j)}(\mathbf{x}_i), \quad j = 1, \dots, m$$

(e.g., $m = 2$ if only minimum and maximum temperatures are considered), so that we can define

$$\mathbf{Z}_i = [Z_{(1)}(\mathbf{x}_i) \dots Z_{(m)}(\mathbf{x}_i)]^T \quad (1)$$

as the corresponding vector of predictands of interest, where T denotes the transpose of the vector. Let the location \mathbf{x}_0 be the target location for the downscaling procedure. We denote

$$\mathbf{Z}_0 = [Z_{(1)}(\mathbf{x}_0) \dots Z_{(m)}(\mathbf{x}_0)]^T \quad (2)$$

as the vector of predictands at the target location (i.e., generally on a finer regular grid).

Finally, let $\mathcal{Y}_k(A_i)$ be the k th predictor at the i th AOGCM grid point. From $\mathcal{Y}_k(A_i)$, one can define

$$\mathcal{Y} = [\mathcal{Y}_1(A_1) \dots \mathcal{Y}_1(A_n) \dots \mathcal{Y}_p(A_1) \dots \mathcal{Y}_p(A_n)]^T \quad (3)$$

as the (np) vector of the p predictors at the n different AOGCM grid points and

$$\mathcal{Y}_k = [\mathcal{Y}_k(A_1) \dots \mathcal{Y}_k(A_n)]^T, \quad k = 1, \dots, p \quad (4)$$

as the p different subvectors reassembling the predictors of same nature at the n different AOGCM grid points. In a downscaling context, one typically tries to estimate the predictands \mathbf{Z}_0 with the predictors \mathcal{Y} (i.e., to reproduce the behavior of the finer-scale predictands with the large-scale predictors). In a probabilistic context, we are thus looking for the conditional distribution $f(\mathbf{Z}_0|\mathcal{Y})$.

b. General formulation

Let us first assume that each grid point A_i of the AOGCM model corresponds to a certain spatial area. Without loss of generality, we use the same notation for both the grid points and the corresponding areas. For each predictand j and each area A_i , let us define

$$Z_{(j)}(A_i) = \frac{1}{|A_i|} \int_{A_i} Z_{(j)}(\mathbf{x}) d\mathbf{x} \quad (5)$$

as an upscaled version of the predictand on A_i , where $|A_i|$ is the area of A_i . The predictand $Z_{(j)}(A_i)$ is thus an

averaged version of the predictand over the spatial area A_i . Finally, let us define the random vector $\mathcal{Z} \in \mathbb{R}^{mn \times 1}$ grouping each upscaled predictand at each area with

$$\mathcal{Z} = [Z_{(1)}(A_1) \dots Z_{(1)}(A_n) \dots Z_{(m)}(A_1) \dots Z_{(m)}(A_n)]^T. \quad (6)$$

The random elements of \mathcal{Z} are thus upscaled expressions of the different predictands $Z_{(j)}$ on the same scale and same locations as the predictors \mathcal{Y} . Such variables cannot be observed directly (i.e., hidden variables), but there exist accurate methods for upscaling stochastic variables over space (see e.g., Goovaerts 2008 and the appendix B for more details). Once the upscaled series of \mathcal{Z} have been estimated for a significant period of time (e.g., using 20–30 yr of historical data), one can model their relation to the predictors \mathcal{Y} with

$$\mathcal{Z} = \mathbf{g}(\mathcal{Y}; \boldsymbol{\beta}) + \boldsymbol{\epsilon}, \quad (7)$$

where $\mathbf{g}(\cdot)$ is some vector function, $\boldsymbol{\beta}$ are the parameters of the function $\mathbf{g}(\cdot)$, and $\boldsymbol{\epsilon}$ is a zero-mean vector of residuals with covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$. As the vectors \mathcal{Z} and \mathcal{Y} share the same coarser spatial resolution and the same locations, we expect that the information on the predictors \mathcal{Y} provided by the predictands \mathbf{Z}_0 is quite negligible compared to the information provided by the upscaled predictands \mathcal{Z} . In other words, it is reasonable to assume stochastic independence between predictors \mathcal{Y} and predictands \mathbf{Z}_0 conditionally to vectors \mathcal{Z} :

$$\mathbf{Z}_0 \perp \mathcal{Y} | \mathcal{Z}. \quad (8)$$

Thus, \mathcal{Z} can be seen as a vector of hidden variables allowing the transfer of information from the predictors \mathcal{Y} to the predictands \mathbf{Z}_0 , \mathbf{Z}_0 , and \mathcal{Z} being related by their spatial dependences [e.g., (cross) covariance functions or (cross) semivariograms].

Among the set of parameters, let us make the distinction between the vector $\boldsymbol{\beta}$ of the parameters of the regression and the vector $\boldsymbol{\alpha}$ of the parameters of the prior distribution of the predictands $(\mathbf{Z}_0, \mathcal{Z})$. We assume here that $\boldsymbol{\beta}$ only influences the link between \mathcal{Y} and \mathcal{Z} , while $\boldsymbol{\alpha}$ only influences the prior distributions. Additionally, we assume that $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are independent. These assumptions enable us to make several simplifications hereafter.

As stated at the end of section 2a, what is sought here is the conditional distribution $f(\mathbf{Z}_0|\mathcal{Y})$ of the local predictands \mathbf{Z}_0 given the coarse predictors \mathcal{Y} . Using the assumption of Eq. (8) and the Bayes theorem, we find that

$$f(\mathbf{Z}_0, \mathcal{Z}, \boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{Y}) = f(\mathbf{Z}_0, \mathcal{Z}|\mathcal{Y}; \boldsymbol{\alpha}, \boldsymbol{\beta})f(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{Y}), \quad (9)$$

$$= \frac{f(\mathcal{Y}|\mathbf{Z}_0, \mathcal{Z}; \boldsymbol{\alpha}, \boldsymbol{\beta})f(\mathbf{Z}_0, \mathcal{Z}|\boldsymbol{\alpha}, \boldsymbol{\beta})}{f(\mathcal{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta})}f(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{Y}), \quad (10)$$

$$= \frac{f(\mathcal{Y}|\mathcal{Z}; \boldsymbol{\alpha}, \boldsymbol{\beta})}{f(\mathcal{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta})}f(\mathbf{Z}_0, \mathcal{Z}|\boldsymbol{\alpha}, \boldsymbol{\beta})f(\boldsymbol{\alpha}, \boldsymbol{\beta}|\mathcal{Y}), \quad (11)$$

$$= \frac{f(\mathcal{Z}|\mathcal{Y}; \boldsymbol{\alpha}, \boldsymbol{\beta})}{f(\mathcal{Z}|\boldsymbol{\alpha}, \boldsymbol{\beta})}f(\mathbf{Z}_0, \mathcal{Z}|\boldsymbol{\alpha}, \boldsymbol{\beta})\frac{f(\boldsymbol{\alpha}, \boldsymbol{\beta})f(\mathcal{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta})}{f(\mathcal{Y})}, \quad (12)$$

where, for the sake of notation simplifications, we use the same $f(\cdot|\cdot)$ notation for the different probability density functions.

Moreover, because of the different hypotheses on both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, we have that $f(\mathcal{Y}|\boldsymbol{\alpha}, \boldsymbol{\beta})$ is in fact equal to $f(\mathcal{Y})$. After simplifications, we can conclude that $f(\mathbf{Z}_0|\mathcal{Y})$ is

$$f(\mathbf{Z}_0|\mathcal{Y}) \propto \iint \frac{f(\mathcal{Z}|\mathcal{Y}; \boldsymbol{\beta})}{f(\mathcal{Z}|\boldsymbol{\alpha})}f(\mathbf{Z}_0, \mathcal{Z}|\boldsymbol{\alpha})f(\boldsymbol{\alpha})f(\boldsymbol{\beta})d\mathcal{Z}d\boldsymbol{\alpha}d\boldsymbol{\beta}. \quad (13)$$

All the distributions in Eq. (13) are modeled monthly. Monthly modeled distributions enable us to account for the seasonal effects on both the prior distributions and the link between the predictors and the predictands. We consider here that these distributions are stationary. However, nonstationary fluctuations can still be induced by the nonstationary evolutions of the predictors.

Although Eq. (13) is general, it is important to make specific assumptions on the different distributions present

in this equation. Some particular choices are presented in the next subsections. However, these choices do not affect the general methodology proposed above.

c. Distributions of the predictands

Gaussian assumptions for $f(\mathbf{Z}_0, \mathcal{Z}|\boldsymbol{\alpha})$, $f(\mathcal{Z}|\mathcal{Y}; \boldsymbol{\beta})$ and $f(\mathcal{Z}|\boldsymbol{\alpha})$ are in good agreements with monthly temperature data. Moreover, these assumptions lead to significant computational simplifications and are thus particularly convenient. Indeed, using these assumptions, the function

$$L(\mathbf{Z}_0, \mathcal{Z}|\mathcal{Y}; \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{f(\mathcal{Z}|\mathcal{Y}; \boldsymbol{\beta})}{f(\mathcal{Z}|\boldsymbol{\alpha})}f(\mathbf{Z}_0, \mathcal{Z}|\boldsymbol{\alpha}) \quad (14)$$

is an exponential function of a quadratic form of $(\mathbf{Z}_0, \mathcal{Z})$. Consequently, $L(\mathbf{Z}_0, \mathcal{Z}|\mathcal{Y}; \boldsymbol{\alpha}, \boldsymbol{\beta})$ is proportional to a multivariate Gaussian distribution. Moreover, one can prove that the mean vector $\mathbf{M}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and the covariance matrix \mathbf{S} of this multivariate Gaussian distribution are given by

$$\begin{cases} \mathbf{S}^{-1} = \boldsymbol{\Sigma}^{-1} + \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\epsilon}^{-1} - \boldsymbol{\Sigma}_{\mathcal{Z}}^{-1} \end{pmatrix} \\ \mathbf{M}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathbf{S} \left\{ \boldsymbol{\Sigma}^{-1} \begin{bmatrix} \boldsymbol{\mu}_0(\boldsymbol{\alpha}) \\ \boldsymbol{\mu}_{\mathcal{Z}}(\boldsymbol{\alpha}) \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \boldsymbol{\Sigma}_{\epsilon}^{-1} \mathbf{g}(\mathcal{Y}; \boldsymbol{\beta}) - \boldsymbol{\Sigma}_{\mathcal{Z}}^{-1} \boldsymbol{\mu}_{\mathcal{Z}}(\boldsymbol{\alpha}) \end{bmatrix} \right\}, \end{cases} \quad (15)$$

where

- $\boldsymbol{\Sigma}$ is the covariance matrix of \mathbf{Z}_0 and \mathbf{Z} . It can be computed using, for example, the fitted semivariograms.
- $\boldsymbol{\Sigma}_{\mathcal{Z}}$ is the submatrix of $\boldsymbol{\Sigma}$ corresponding to \mathbf{Z} only.
- $\boldsymbol{\Sigma}_{\epsilon}$ is the covariance matrix of the residuals $\boldsymbol{\epsilon}$.
- $\boldsymbol{\mu}_0(\boldsymbol{\alpha})$ and $\boldsymbol{\mu}_{\mathcal{Z}}(\boldsymbol{\alpha})$ are the mean vectors of \mathbf{Z}_0 and \mathcal{Z} , respectively.

This result can be proven (details not shown here) by identification of the coefficients of the quadratic form with the corresponding coefficients of the multivariate Gaussian distribution.

A GRM using longitude, latitude, and altitude at the same location as \mathbf{Z}_0 is chosen for the estimation of $\boldsymbol{\mu}_0(\boldsymbol{\alpha})$ (see Benestad 2002; Benestad et al. 2008 for more details) with

$$\begin{aligned} \boldsymbol{\mu}_{0;d} &= \boldsymbol{\alpha}_{1;d} + \boldsymbol{\alpha}_{2;d}(\lambda - \mu_{\lambda}) + \boldsymbol{\alpha}_{3;d}(\phi - \mu_{\phi}) \\ &\quad + \boldsymbol{\alpha}_{4;d}h + \varepsilon_{\mu;d}, \quad d = 1, \dots, 12, \end{aligned} \quad (16)$$

where (i) λ , ϕ , and h are, respectively, the longitudes, latitudes, and altitudes of the location; (ii) μ_{ϕ} and μ_{λ} are the mean of the observed weather stations latitudes and

longitudes, respectively; and (iii) the vectors $\alpha_{1;d}, \dots, \alpha_{4;d}$ have size m and are subvectors of α . In accordance with the International Organization for Standardization (ISO 2533:1975; see the ISO Web site online at <http://www.iso.org/>), the parameters $\alpha_{4;d}$ are all set so that the lapse rate is $6.5^\circ\text{C km}^{-1}$, which also corresponds approximately to the mean rate of pseudoadiabatic near the ground at around a normal range of atmospheric pressures. The other parameters are fitted using ordinary least squares (OLS). Finally, as recommended by Benestad (2005) and Benestad et al. (2008), the monthly observed errors $\varepsilon_{\mu;d}$ are interpolated using inverse distance interpolation over space and are added in Eq. (16) in order to correct the observed offsets at the gauged stations. The parameters of the inverse distance interpolation are fitted using a leave-one-out procedure for each month.

As the GRM provides the monthly mean at any location, the elements of the prior mean $\mu_Z(\alpha)$ can be evaluated by averaging the GRM for each predictand j and on each area A_i using the same formula as in Eq. (5). This procedure ensures a more coherent estimation of $\mu_Z(\alpha)$.

In a Bayesian context, the function $L(\mathbf{Z}_0, \mathcal{Z}|\mathcal{Y}; \alpha, \beta)$ is known as the likelihood function and corresponds to the distribution of the predictands if the parameters were exactly known. Incidentally, sections 2d–f are dedicated to the parameters prior distributions and to the estimation of the posterior distribution of Eq. (13).

d. Distributions of the parameters α and β

There are two sources of uncertainty in the GRM: (i) the uncertainty of $\varepsilon_{\mu;d}$ (i.e., around the model itself) and (ii) the uncertainty due to the estimation of the parameters α . The fitted GRM is corrected by adding the interpolated errors $\varepsilon_{\mu;d}$ in order to respect the seasonal characteristics at gauged weather stations. A Gaussian semivariogram-like model using the distance to the closest station is chosen here in order to account for this source of uncertainty. The variance of this Gaussian semivariogram model is chosen as the observed variance of $\varepsilon_{\mu;d}$ and the range parameter is set to 100 km. This results in an increasing uncertainty with respect to the distance to the gauged stations, while preserving the observed seasonal characteristics. Let us define Σ_μ as the diagonal covariance matrix that is constructed using the Gaussian semivariogram model. As stated in section 2c, the parameters α are fitted monthly using an OLS procedure. The OLS estimators are generally considered to be asymptotically multivariate Gaussian, unbiased, and with a covariance matrix Σ_α that depends on the covariance matrix of the errors $\varepsilon_{\mu;d}$.

The distribution of β is linked to the choice of function $g(\cdot)$ in Eq. (7). Among other possibilities and similarly to

the model in appendix A, an MMLR model is assumed here

$$g(\mathcal{Y}; \beta) = \beta_{0,d} + \beta_{1,d}\mathcal{Y}, \quad d = 1, \dots, 12, \quad (17)$$

where $\beta_{0,d}$ is the subvector of β corresponding to the intercepts of the d th month and $\beta_{1,d}$ is a $m \times (np)$ matrix with the slopes corresponding to the different predictors \mathcal{Y} . First, as expressed in appendix A, this kind of relation is commonly used in downscaling methods because of its ease of implementation and its low computational rate. Second, similarly to the estimator of α and Γ (see appendix A), the OLS estimator for the parameters of this model is known to be unbiased and asymptotically multivariate Gaussian with a covariance matrix Σ_β that depends on the covariance matrix of the errors Σ_ε .

e. Estimation of the posterior distribution

By substituting all distributions in Eq. (12), it is possible to derive the expression of the posterior distribution $f(\mathbf{Z}_0|\mathcal{Y})$. It is not straightforward to find an analytical expression of this posterior. However, the structure of the model enables us to sample directly in the posterior with the following sampling algorithm

ALGORITHM

- 1) Draw a sample $\tilde{\beta}_i$ in $f(\beta)$ (i.e., in a Gaussian distribution with mean vector equal to the OLS estimator for the parameters and covariance matrix equal to Σ_β).
- 2) Similarly, draw a sample $\tilde{\alpha}_i$ in $f(\alpha)$ (i.e., in a Gaussian distribution with mean vector equal to the OLS estimator for the parameters and covariance matrix equal to Σ_α).
- 3) Draw a Gaussian perturbation $(\tilde{\varepsilon}_{\mu;d})_i$ with zero mean and covariance matrix Σ_μ .
- 4) Using the samples $\tilde{\alpha}_i$ and $(\tilde{\varepsilon}_{\mu;d})_i$, compute the perturbed values $(\mu_0(\tilde{\alpha}_i))_i$ for the GRM in Eq. (16).
- 5) Using the samples $\tilde{\beta}_i$ and $(\mu_0(\tilde{\alpha}_i))_i$, draw a sample $(\tilde{\mathbf{Z}}_0, \tilde{\mathcal{Z}})_i$ in the Gaussian distribution with the parameters given in Eq. (15).
- 6) Return to the first step until the number of simulations is reached.

Using this algorithm, the resulting sequence $(\tilde{\mathbf{Z}}_0)_i$ is a random sample of the distribution $f(\mathbf{Z}_0|\mathcal{Y})$ so that any properties of this distribution can be directly estimated from the random sample $(\tilde{\mathbf{Z}}_0)_i$ (e.g., mean, standard deviation, mode, and median).

f. Properties and potential extensions of the approach

The use of upscaled predictands \mathcal{Z} can be questioned. There are, however, several justifications for this particular choice. Given the change of spatial scale between predictors and predictands, one can only expect the predictors

to provide information on the smoother part of the physical process. Indeed, the small structures at finer scale (i.e., the observed predictands) are more likely to be induced by the prior mean (i.e., the GRM regression). On the other hand, upscaled predictands \mathcal{Z} share the same scale and the same location with the predictors \mathcal{Y} . One can thus expect to observe the same spatiotemporal patterns in the upscaled predictands as in the predictors. The statistical link between the predictors and the upscaled predictands should thus be more direct than when using the original predictands. Thus, as the predictors and the original predictands do not share the same location (nor the same scale), upscaling the predictands is preferred to interpolating in order to account explicitly for the change of scale. Moreover, thanks to the spatial averaging, the upscaled variables are more stable than the interpolated ones. As a direct consequence of this stability, the estimation of the parameters β in the function $\mathbf{g}(\cdot)$ gains also in both precision and relevance.

It is important to underline that the conditional independence assumption in Eq. (8) is rather weak (i.e., it is not too restrictive). It is at least weaker than imposing a direct relation between \mathbf{Z}_0 and \mathcal{Y} as the location of \mathbf{Z}_0 changes along the estimation map. Moreover, in the absence of clear information about this conditional relation, the maximum entropy solution (i.e., the solution that minimizes the effects of a particular choice of link) is to assume the conditional independence (see e.g., Bogaert and Fasbender 2007 for a discussion of this property). Consequently, the applicability of this Bayesian spatial model is potentially wide. Even if linear relations are assumed here, the use of this conditional independence assumption enables us to consider nonlinear relations, which is generally not possible when using classical spatial methods, such as cokriging approaches. It also enables us to separate the spatial component [i.e., $f(\mathbf{Z}_0, \mathcal{Z}|\alpha)$] from the information concerning the injection of the predictors [i.e., $f(\mathcal{Z}|\mathcal{Y}; \alpha)$] and the estimation of the uncertainty of the parameters [i.e., $f(\alpha)$ and $f(\beta)$]. Consequently, each of these components is focused separately, which is much easier than to model the whole distribution at once.

As a direct consequence of the underlying spatial model, it is straightforward to include observed series at other weather stations. Setting \mathbf{Z}_S as the observed values corresponding to these new weather stations, the main difference would be to replace the prior distribution $f(\mathbf{Z}_0, \mathcal{Z}|\alpha)$ with the conditional prior $f(\mathbf{Z}_0, \mathcal{Z}|\mathbf{Z}_S; \alpha)$ in Eqs. (11) and (12). To make this substitution, one must assume

$$f(\mathcal{Y}|\mathbf{Z}_0, \mathbf{Z}_S, \mathcal{Z}; \alpha, \beta) = f(\mathcal{Y}|\mathcal{Z}; \alpha, \beta) \quad (18)$$

instead of the assumption in Eq. (8). Thanks to this property, it is possible to account for information with

different spatial resolutions, namely at-site observations \mathbf{Z}_S and AOGCM predictors \mathcal{Y} , within a unified theoretical spatial framework. Of course, in a forecasting context, this property is not of great support as at-site observations will no longer be available. The assessment of this ability of the proposed approach will be addressed in future research.

A second direct consequence of the spatial model is that one could also consider the variable \mathbf{Z}_0 to be multisite as well. Again, the main change is the conditional independence assumption of Eq. (8). If one considers \mathbf{Z}_0 as a vector of multisite predictands, then the assumption is stronger, since it implies more variables. However, the expressions in Eqs. (9)–(12) and the simulation procedure remain the same. Similarly to multiple multivariate linear regression, the main advantage of this extension is that jointly considering multisite predictands will enable us to account for their spatial dependence. Although this approach is not tractable in practice when estimating a whole map for a large area, this could be particularly interesting when downscaling AOGCM predictors at only a small number of sites (e.g., weather stations).

By construction, the proposed method can provide spatial estimates for the predictands. This is due to the spatial covariance model that is more general than site specific models (e.g., the MMLR models in appendix A). However, at gauged locations, site specific models will generally perform more adequately since they are incidentally specific to these weather stations. As an example, the spatial model ensures that the covariance between two predictands is always the same for a given spatial lag. This is not necessarily respected in real-case applications. This leads to a drop of precision that is expected to grow as the actual covariances between series diverge from the spatial covariance model. We empirically evaluate this loss of precision at gauged locations in the results section.

3. Case study

To illustrate the potential of the proposed Bayesian spatial model of section 2, a case study is proposed in this paper. In subsections 3a–c, details are provided concerning the study area, the variables at hand and the data processing. Section 3c is also dedicated to the definition of the different quality criteria that are used in section 4.

a. Data and study area

The study area is located in the southern part of the province of Quebec, Canada. Nine series of observed daily minimum and maximum temperatures are selected as predictands: Cedars, Drummondville, Seven Islands, Bagotville, Quebec, Sherbrooke, Maniwaki, La Pocatiere, and Mont-Joli, Canada (Data Access Integration 2009).

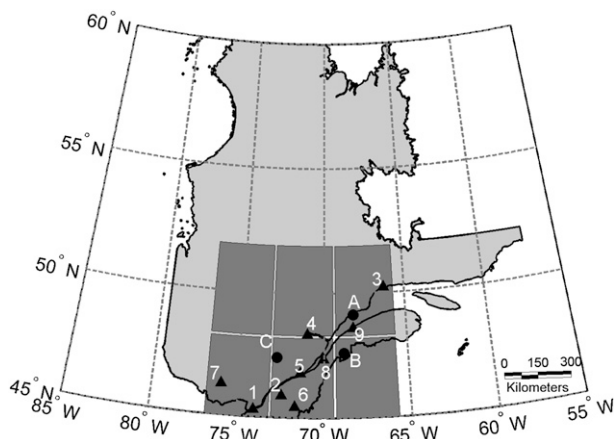


FIG. 1. Weather stations and AOGCM grid points located around the Gulf of St. Lawrence. Calibration stations: 1) Cedars, 2) Drummondville, 3) Seven Islands, 4) Bagotville A, 5) Quebec, 6) Sherbrooke A, 7) Maniwaki Airport, 8) La Pocatière, and 9) Mont-Joli A. Validation stations: (a) Baie-Comeau A, (b) Notre Dame du Lac, and (c) La Tuque.

Three additional stations are also selected in the same region for the spatial validation of the proposed model: Baie-Comeau A, Notre Dame du Lac, and La Tuque, Canada. Figure 1 shows the location of these stations around the St. Lawrence River and the Gulf of St. Lawrence. All

homogenized temperature data series (from Environment Canada weather stations) are provided by Vincent et al. (2002), except Baie Comeau and Notre Dame du Lac, which come from the National Climate Data and Information Archive of Environment Canada (see online at http://climate.weatheroffice.gc.ca/climateData/canada_e.html). All series cover the period between 1 January 1961 and 31 December 2000 (i.e., 14 610 days). The number of unobserved daily values in the original time series ranges between 0 and 478 values depending on the location. Days with at least one missing value are removed in the analysis. For the validation of the method, the total dataset is divided into two independent subsets: a calibration period between 1961 and 1990 (i.e., 10 957 days) and a validation period between 1991 and 2000 (i.e., 3653 days).

The region covered by the 9 weather stations is rather homogeneous in latitudes (between 45° and 50°N), in longitudes (between 65° and 77.5°W), and in altitudes (between 0 and 1025 m above sea level; see Fig. 2; more information is available online at <http://www.geogratis.ca/geogratis/fr/index.html>). The region's climate is continental, distributed from moderate in the south to sub-polar in northern areas, with cold winters and hot-mild summers modulated by the presence of water masses as the Gulf of St. Lawrence (see further details in Gerardin

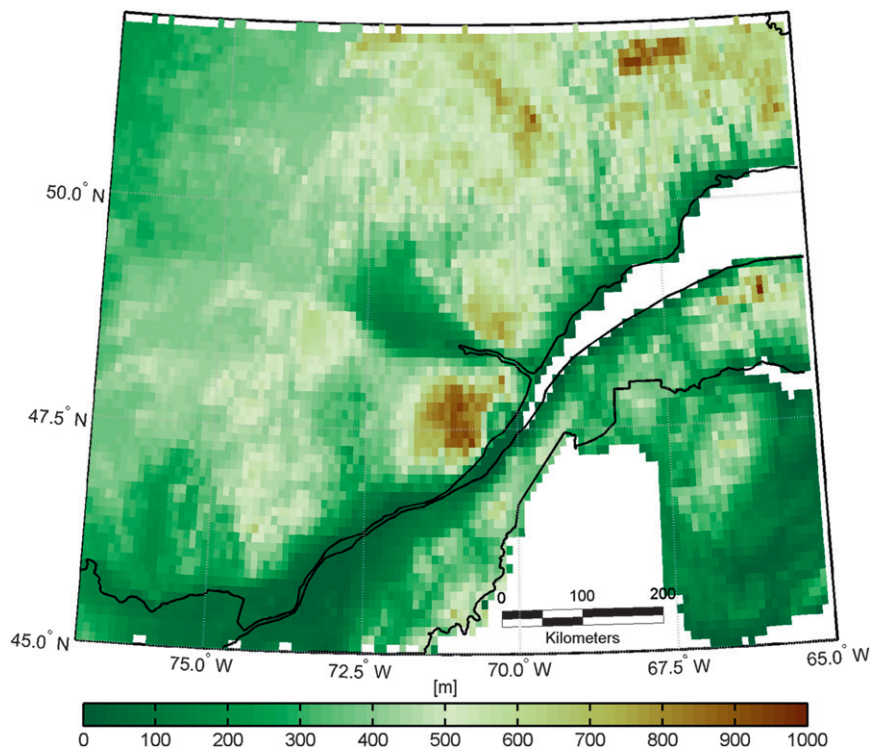


FIG. 2. Digital Elevation Model for the study area. The spatial resolution is 0.1° in both latitude and longitude (approximately 6–10 km).

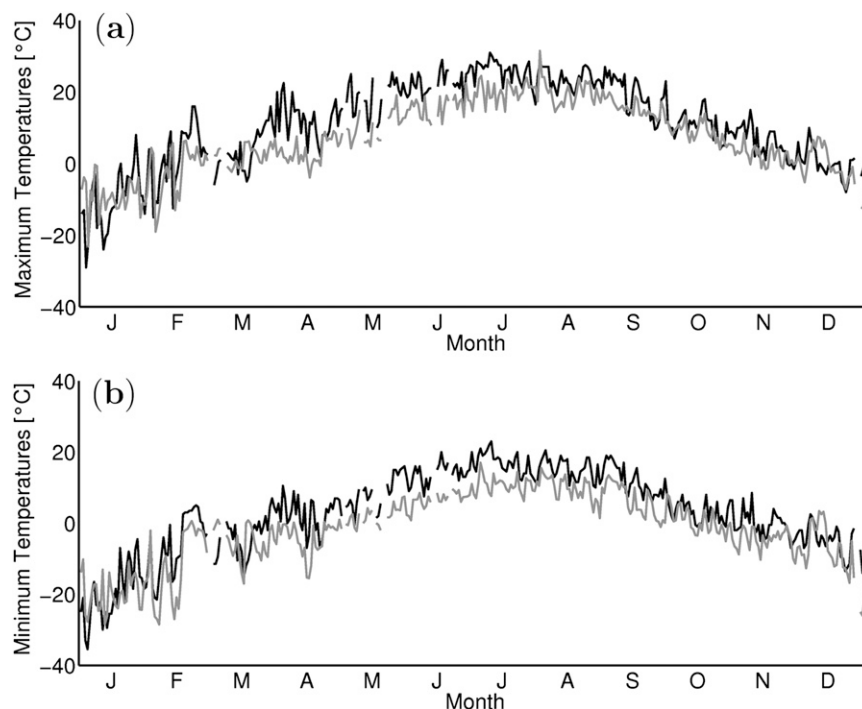


FIG. 3. Evolution of (a) maximum and (b) minimum daily temperatures for Cedars (south; in black) and Seven Islands (north; in gray) for the year 1981.

and McKenney 2001). For a given day, we observe that the nine temperature series are rather homogeneous with only few differences along the location, except when prominent cyclonic systems are moving from southern areas or in development over the Atlantic coastlines especially in the winter season. Figure 3 illustrates two examples of maximum and minimum daily temperatures evolutions in the case of Cedars and Seven Islands stations for the year 1981. One can check that both series share in most cases similar daily variations, while the annual mean values of the temperatures are lower at Seven Islands (i.e., the most Nordic station of the study area). Table 1 confirms this general observation by comparing the annual means and annual standard deviations in

decreasing latitudes. The presence of the water masses (i.e., estuary and Gulf of St. Lawrence) tends to reduce the continental conditions of the climate and the diurnal and annual temperature range (see Gerardin and McKenney 2001).

To assess the potential of the proposed method, we use in this study the National Centers for Environmental Prediction–National Center for Atmospheric Research (NCEP–NCAR) reanalysis products (e.g., Kalnay et al. 1996; Kistler et al. 2001). The $2.5^{\circ} \times 2.5^{\circ}$ original NCEP data are interpolated on the $3.75^{\circ} \times 3.75^{\circ}$ CGCM3 grid (Gaussian), corresponding to the third generation of the coupled Canadian Global Climate Model (see Scinocca et al. 2008). In addition to the interpolation procedure,

TABLE 1. Annual statistics for the nine weather stations in decreasing latitudes.

Station	Lat ($^{\circ}$)	Max temperature ($^{\circ}$ C)		Min temperature ($^{\circ}$ C)	
		Annual mean	Annual SD	Annual mean	Annual SD
Seven Islands	50.22	5.13	10.90	−4.03	11.81
Mont-Joli A	48.60	7.12	11.80	−1.21	10.49
Bagotville A	48.33	7.42	13.45	−2.93	12.74
La Pocatière	47.36	8.38	12.20	−0.68	11.29
Quebec	46.79	8.71	12.62	−0.56	11.49
Maniwaki Airport	46.27	9.49	12.72	−2.54	12.69
Drummondville	45.88	10.31	12.74	0.64	12.12
Sherbrooke A	45.43	9.88	12.23	−1.57	11.92
Cedars	45.30	10.21	12.32	1.89	11.80

TABLE 2. NCEP predictor variables on the CGCM3 grid.

No.	Predictors	No.	Predictors
1	Mean sea level pressure	14	500-hPa divergence
2	1000-hPa wind speed	15	850-hPa wind speed
3	1000-hPa <i>U</i> component	16	850-hPa <i>U</i> component
4	1000-hPa <i>V</i> component	17	850-hPa <i>V</i> component
5	1000-hPa vorticity	18	850-hPa wind vorticity
6	1000-hPa wind direction	19	850-hPa geopotential
7	1000-hPa divergence	20	850-hPa wind direction
8	500-hPa wind speed	21	850-hPa divergence
9	500-hPa <i>U</i> component	22	500-hPa specific humidity
10	500-hPa <i>V</i> component	23	850-hPa specific humidity
11	500-hPa vorticity	24	1000-hPa specific humidity
12	500-hPa geopotential	25	Mean temperature at 2 m
13	500-hPa wind direction		

all NCEP–NCAR data has been averaged on a daily basis from 6 hourly data, before being linearly interpolated to match the CGCM3 data (see all details about the preparation of NCEP–NCAR and CGCM3 predictors in DAI Team 2008), as the final objective is to use AOGCM predictors. However, the calibration procedure of all statistical downscaling models requires first reanalysis datasets before to use properly the AOGCM predictor variables. At each gridpoint location, 25 different NCEP predictors are provided (see Table 2). The six closest grid points in the southern part of the province of Quebec are selected (see Fig. 1). For each date, 150 covariates are thus available for the downscaling process. A principal component analysis (PCA) is first performed in order to reduce this number to 40. Indeed, using the first 40 principal components (PCs) enables us (i) to remove correlations between the predictors, (ii) to obtain uncorrelated estimates of parameters β for a given month, and (iii) to prevent overfitted regression while preserving more than 97% of predictor variance. It is important to note that for areas located in high absolute latitudes, the corresponding grid cells cannot be assumed to be rectangular. Indeed, in the Northern Hemisphere, the top edge of the grid pixels is smaller than the bottom edge (and reversely in the Southern Hemisphere) and these edges are not linear. Figure 1 illustrates this phenomenon on the six selected grid pixels. The direct consequence for the present study is that higher latitude grid points correspond to smaller areas. This has an effect on the values of the regularized covariance functions.

b. Application of the model to the case study

The prior means are modeled using the GRM proposed in section 2c. However, the linear spatial distribution of the stations does not permit to distinguish between

TABLE 3. Monthly estimation for the inverse distance power parameters for both minimum and maximum temperatures.

	J	F	M	A	M	J	J	A	S	O	N	D
Tmax	1.18	8.24	6.44	9.00	9.84	1.62	7.96	3.07	2.09	5.52	1.89	8.38
Tmin	4.61	4.39	8.95	8.49	8.72	3.71	4.66	3.49	5.95	5.73	9.50	2.83

the effects of the latitude and the longitude (i.e., colinearity). In this study, GRM are thus assumed to vary only with the latitude and the altitude [i.e., all the parameters $\alpha_{2,d}$ in Eq. (16) are set to 0]. Residuals of this regression model are then interpolated over space and added using an inverse distance interpolation. The power parameters for the inverse distance are fitted monthly using a leave-one-out procedure (see Table 3) while the variance of this interpolation is modeled as a Gaussian-like semivariogram function with a variance equal to the observed variance of the GRM residuals and a range equal to 100 km. This approach enables us to consider the uncertainty as proportional to the distance to the closest observed weather station. The final GRM is then spatially validated using a leave-one-out cross-validation procedure. Figure 4 shows the comparison between the monthly modeled standard deviations of the GRM and the observed monthly RMSE of the cross validation (averaged over space). One can see that the GRM standard deviations are in satisfactory agreements with the cross-validation RMSE, the differences between the lines being rather negligible.

The joint spatial model for both minimum and maximum daily temperatures is fitted monthly using the linear model of coregionalization (LMC; see e.g., Chilès and Delfiner 1999) on the observed covariances between series and the corresponding spatial lags (see Fig. 5). To prevent the spatial deformations due to high latitudes, Universal Transverse Mercator (UTM) coordinates are used for the computation of spatial lags instead of the original (λ, ϕ) coordinates of the Gaussian grid. As the temperatures present a smooth spatial behavior, a combination of a Gaussian model with a range equal to 1500 km and of a nugget model is selected. The 12 different models are validated using 2000 Gaussian random simulations at the same locations. As the experimental semivariograms belong to the 95% confidence interval (see Fig. 5 for January), this validation procedure confirms that the chosen spatial model is in acceptable agreement with the variability of the empirical semivariograms observed in equivalent conditions (i.e., same 9 locations, same theoretical spatial model, etc.). Figure 6 shows the monthly evolution of the different parameters. Generally speaking, there is a stronger spatial dependence during the summer than during the winter since the proportion of the nugget effect is smaller during the summer. This

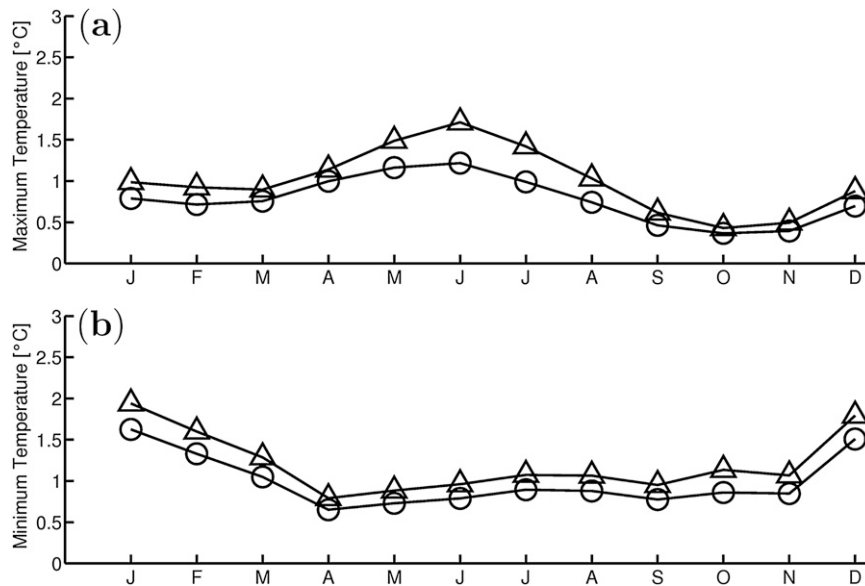


FIG. 4. Visual comparison between the estimated standard deviation provided by the GRM (circles) and the observed RMSE computed using a leave-one-out cross validation (triangles) for (a) the maximum temperatures and (b) the minimum temperatures.

is compatible with higher variability in winter than in summer for both the minimum and maximum daily temperatures across all regions. This seasonal difference in the variability is even more pronounced in northern areas (see Gerardin and McKenney 2001; Gachon et al. 2005). For the maximum temperatures, the variance is higher between April and June. This is due to the fact that the maximum daily temperatures in the southern part of the study area increase earlier than in the northern part, where prevalent winter conditions can persist over spring months. This results in a larger spatial variance in the LMC model. Similarly, the spatial variances of minimum temperatures are larger during winter because the northern minimum temperatures can drop significantly lower than in the southern part, with more persistent or longer duration of colder conditions in northern areas (see further details in Gachon et al. 2005).

Using the fitted spatial model, the upscaled variables \mathcal{Z} are estimated using the theory of regularized covariance functions (Goovaerts 2008; see appendix B for more details). The means of the upscaled predictands are computed by averaging the GRM on the corresponding area so that Eq. (B5) can be used instead of Eq. (B6). As expected, these upscaled variables reflect smoother time series than the original series of observations since the upscaled variables result from a spatial average (not shown here).

The estimated upscaled predictands are then used to fit the MMLR model of Eq. (7) with the decorrelated predictors \mathcal{Y} . Monthly covariance matrices Σ_{ϵ} for the

regression are then estimated based on the observed residuals. The covariance matrices Σ_{β} for the estimated parameters are computed using the classic properties of MMLR models. This completes the expressions of both the conditional pdf $f(\mathcal{Z}|\mathcal{Y}; \beta)$ and the prior $f(\beta)$. Similarly, the MMLR model presented in appendix A is applied to the dataset using the same decorrelated predictors \mathcal{Y} . One can thus compare the results of the proposed spatial model with those of this classical regression-based approach.

Finally, the sampling algorithm proposed in section 2e is applied using the modeled conditional and prior distributions. In this paper, the number of simulations is set to 1000. The samples are then used for the estimation of the posterior mean vector and the posterior covariance matrix of the predictands.

c. Quality assessment

A first quality assessment is carried out through a direct comparison of the downscaled distribution $f(\mathbf{Z}_0|\mathcal{Y})$ and the true observations. The mean error (ME) and the root-mean-square error (RMSE) are computed using the difference between the observation and posterior mean. The ME measures the accuracy of the downscaled distribution while the RMSE measures the precision. One can then compare the RMSE values with the posterior standard deviations, as they depict the same property of the distribution. Indeed, close RMSE and posterior standard deviation values indicate that the proposed spatial model correctly depicts the variability. Additionally,

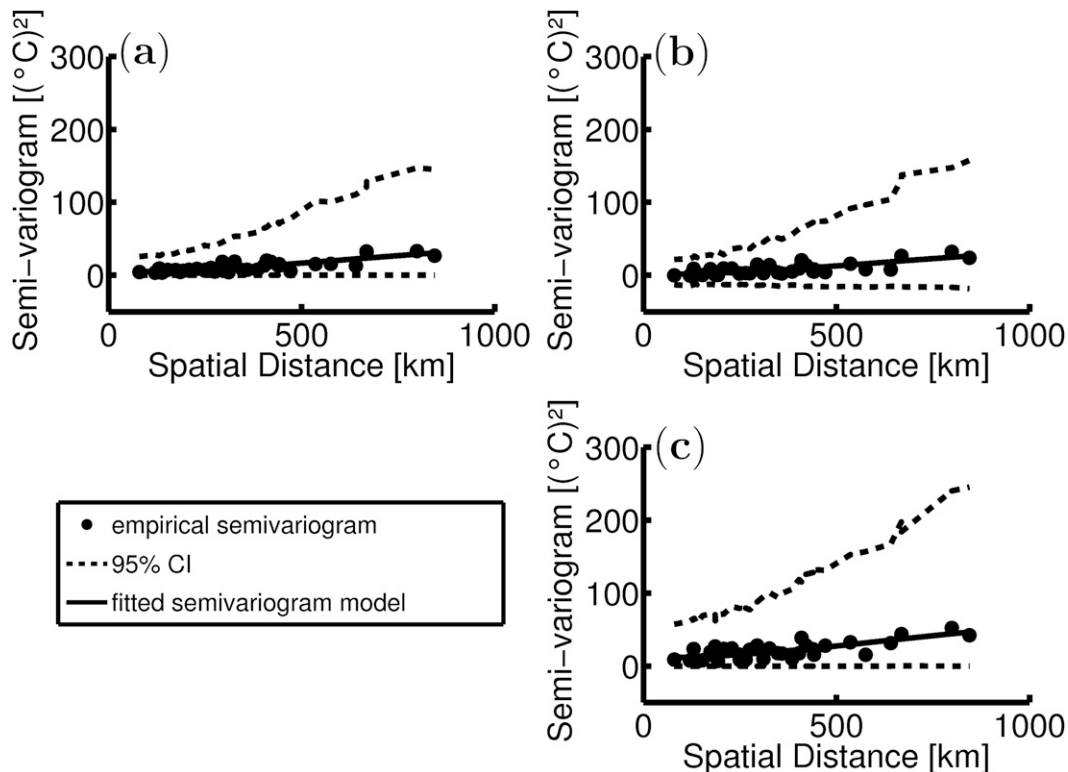


FIG. 5. Visual validation of the fitted multivariate semivariogram model (bold lines) for January. The 95% confidence intervals (dashed lines) are computed using the 2000 simulations. The fitted model is in agreement the observations as the empirical semivariograms (stars) belong to the 95% confidence intervals for (a) the maximum temperatures, (b) the couple maximum–minimum temperatures, and (c) the minimum temperatures.

building the 95% credibility intervals and checking that approximately 95% of the observations fall within the interval also somewhat validate the modeled distribution of the downscaled predictands.

A set of several monthly and seasonal climatic indices are also chosen for assessing the performance of the downscaling models. The definitions of the climatic indices are presented in Table 4. They were proposed in the context of Nordic climates. They reflect the frequency, intensity, and duration temperature extremes. They, thus, ensure that the downscaled series depict correctly the observed climate (see Hessami et al. 2008; Gachon et al. 2005; Wilby et al. 1998, 2002). The DTR index measures the mean of the diurnal temperature range (i.e., the difference between maximum and minimum temperatures) and is computed on a seasonal basis. The frost season length (FSL) index measures the number of days per year separating 5 consecutive days of freezing conditions (i.e., $T_{\text{mean}} < 0^{\circ}\text{C}$) to 5 consecutive days of thaw conditions (i.e., $T_{\text{mean}} > 0^{\circ}\text{C}$), where $T_{\text{mean}} = (T_{\text{min}} + T_{\text{max}})/2$. Similarly, the growing season length (GSL) index measures the number of days per year separating 5 consecutive days for which $T_{\text{mean}} > 5^{\circ}\text{C}$ to 5 consecutive days

for which $T_{\text{mean}} < 5^{\circ}\text{C}$ and represents the growing season. The FR-Th index measures the number of days per month of freeze and thaw cycle (i.e., $T_{\text{max}} > 0^{\circ}\text{C}$ and $T_{\text{min}} < 0^{\circ}\text{C}$). Finally, the Tmax90 and Tmin90 indices are the 90th seasonal percentiles of T_{max} and T_{min} , respectively.

4. Results

The proposed spatial model was applied on the study area. Figure 7 illustrates the maximum (left panels) and minimum (right panels) predictions of temperatures for a couple of specific dates corresponding to some large observed anomalies in both the calibration and the validation periods. Figures 7a,b show the estimated maps for 3 January 1981. Even if this day was characterized by very low temperatures, the influence of the predictors enabled us to reproduce the very low temperatures in the estimated maps. Similarly, Figs. 7c,d show that the spatial Bayesian approach was able to correctly estimate the temperatures for 28 March 1998, even if the actual observed temperatures were quite large for this period of the year (approximately 10°C above the seasonal averages). The last example corresponds to 31 August 1994

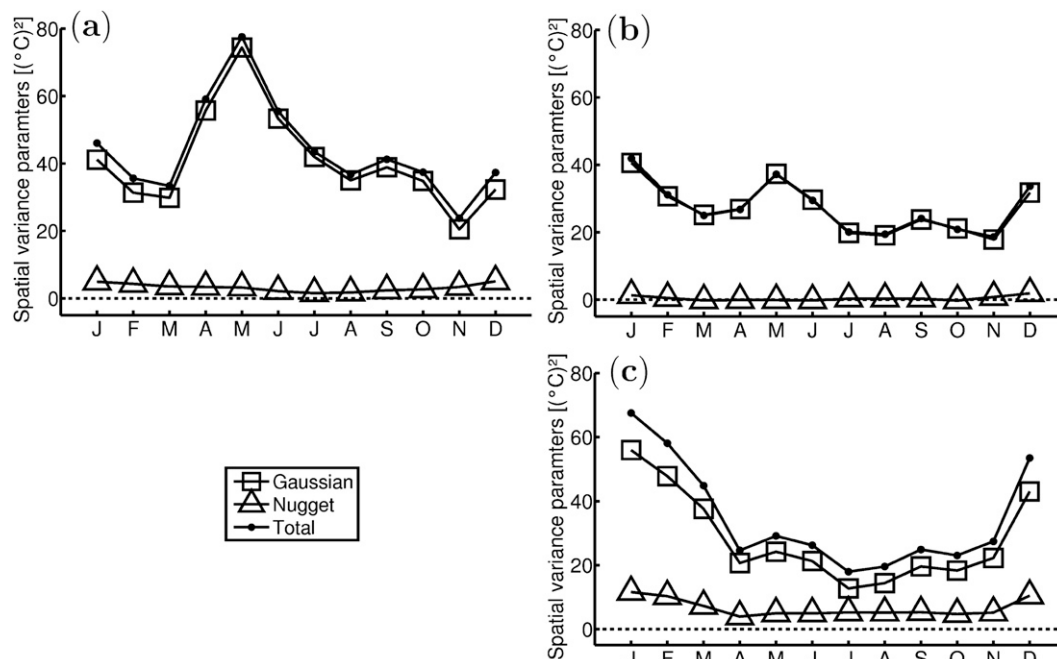


FIG. 6. Evolution of the fitted parameters for the different components of the spatial model and for each month: (a) monthly variances of maximum temperature, (b) monthly covariances between maximum and minimum temperatures, and (c) monthly variances of minimum temperature.

for which the proposed approach correctly estimated very low temperatures ($5^{\circ}\text{--}10^{\circ}\text{C}$ below the seasonal averages) with freezing conditions in higher altitudes and latitudes (Figs. 7e,f).

The proposed approach allowed to model not only the mean of the conditional pdf $f(\mathbf{Z}_0|\mathcal{Y})$ (see Fig. 7) but also the whole distribution. It was thus interesting to focus on the conditional standard deviation. Figure 8 illustrates the map of the estimated standard deviation for 31 August 1994. It was observed that the standard deviation of the model is smaller for locations close to the gauged weather stations. This was mainly due to the modeled uncertainty of the GRM. On the other hand, the large-scale differences were mainly due to the influence of the NCEP predictors over both gauged and ungauged sites, which cannot capture all atmospheric features related to local-scale temperature characteristics. Indeed, the covariance matrix of the residuals in Eq. (7) contained different variances depending on the upscaled predictand location, so that the precision of the regression was not homogeneous over the study area. Finally, it was computed that 10%–30% of the total variance for the maximum temperatures and 5 to 10% of the total variance for the minimum temperatures was due to parameter uncertainties. However, the difference for the 95% credibility intervals with and without consideration of the parameter uncertainty ranged between 0.5° and 2°C in both

cases. Therefore, parameter uncertainty was almost non-influent in this case.

Table 5 shows the mean RMSE values between the observed and the 1000 daily simulated series based on the proposed Bayesian method for each of the climatic indices proposed in Table 4. Table 6 presents the same indices but for the MMLR model of appendix A. One can see that the results of the MMLR were in general slightly

TABLE 4. Definition of the climatic indices used for the performance assessment of downscaled temperatures. Here T_{mean} is defined here as $(T_{\text{min}} + T_{\text{max}})/2$.

Indices	Definition	Unit	Time scale
DTR	Mean of diurnal temperature range	$^{\circ}\text{C}$	Season
FSL	Frost season length Days between 5 consecutive $T_{\text{mean}} < 0^{\circ}\text{C}$ and 5 consecutive $T_{\text{mean}} > 0^{\circ}\text{C}$	Days	Year
GSL	Growing season length Days between 5 consecutive $T_{\text{mean}} > 5^{\circ}\text{C}$ and 5 consecutive $T_{\text{mean}} < 5^{\circ}\text{C}$	Days	Year
FR-Th	Days with freeze and thaw cycle ($T_{\text{max}} > 0^{\circ}\text{C}$, $T_{\text{min}} < 0^{\circ}\text{C}$)	Days	Month
Tmax90	90th percentile of daily T_{max}	$^{\circ}\text{C}$	Season
Tmin90	90th percentile of daily T_{min}	$^{\circ}\text{C}$	Season

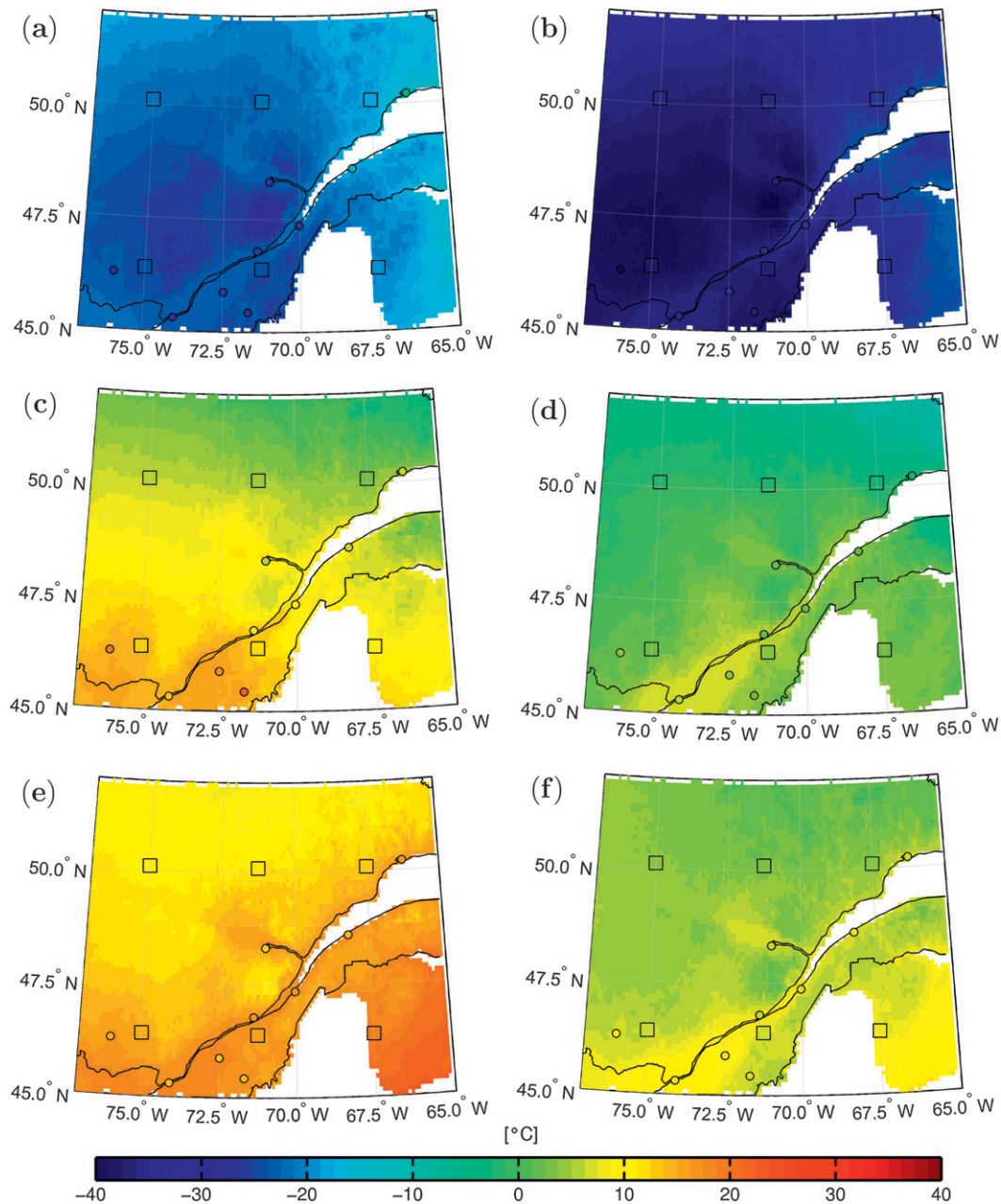


FIG. 7. Examples of (left) maximum and (right) minimum temperatures estimation maps for (a),(b) 3 Jan 1981; (c),(d) 28 Mar 1998; (e),(f) 31 Aug 1994. Circles represent the true observations and squares are for the upscaled temperatures on the AOGCM grid points.

better than those of the proposed spatial model. This was not surprising since the MMLR model accounted for more specific relationships between the NCEP predictors and the predictands at the nine weather stations while the proposed spatial model took into account the NCEP predictors through the upscaled predictands. Moreover, even the links between the predictands themselves were more specific when using MMLR. Indeed, in the case of

the MMLR, two couples of stations sharing the same spatial lag did not especially share the same covariances, while covariances were uniquely specified by the spatial lag within the proposed approach (see Fig. 5). A second explanation could be that the independence hypothesis between \mathbf{Z}_0 and \mathcal{Y} given \mathcal{Z} was not fulfilled. However, as the true conditional dependence between these variables could not be directly observed, the conditional

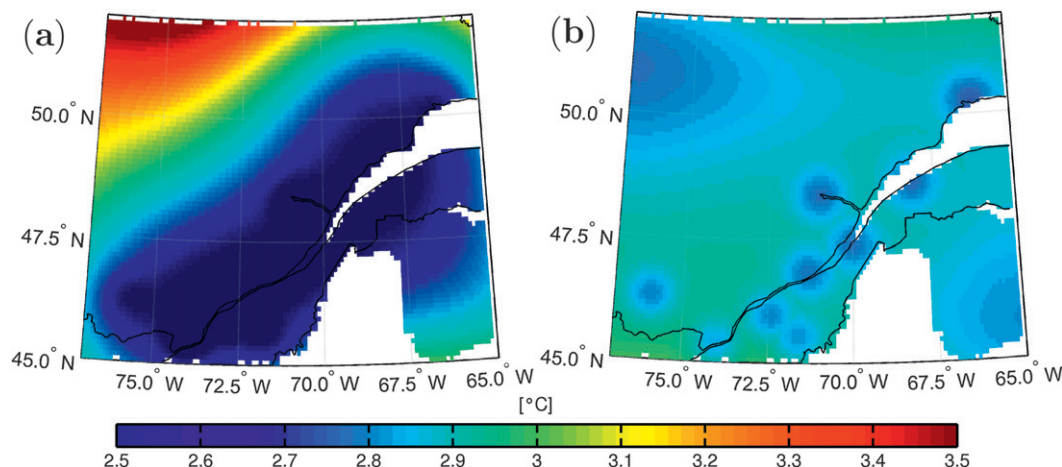


FIG. 8. Maps of the estimated standard deviation for (a) maximum and (b) minimum temperatures for 31 Aug 1994.

independence hypothesis was the maximum entropy solution (see section 2f). On the other hand, the main advantage of the proposed approach was that it directly provided spatial estimation at ungauged locations, which was not possible with the MMLR method precisely because it was too specific to the observed predictands. Consequently, the loss of precision when using the proposed spatial model could be considered negligible compared to the advantage of providing estimation maps.

The results were then spatially validated by comparison of the observed and estimated temperatures at the three validation stations (see Fig. 1 for the locations of these stations). Figures 9, 10, and 11 show, for instance, this comparison for the year 1995. Results indicated that the estimated series were close to the true observed ones. Moreover, the majority of the observations belonged to the 95% credibility intervals based on the simulations, which ensured that the proposed model adequately depicted the natural process and its fluctuations. Table 7 confirms these visual quality assessments by means of

the ME, the RMSE, and the percentage of observations within the 95% credibility intervals calculated at the 3 validation stations during the validation period (1991–2000). It indicates that there was a small estimation bias (between 0.5° and 1.95°C in absolute value) while the RMSE ranged between 3° and 5°C, which was less than half the original standard deviation based on the fitted spatial model. The model thus permitted to significantly reduce the variability. Moreover, the estimated distributions were in satisfactory agreements with the observed values as approximately 95% of the observations belonged to the 95% credibility intervals except for the minimum temperatures at La Tuque, which seemed to be overestimated. Aside from this exception, the validation using CI ensured that the uncertainty was accurately depicted by the model.

5. Discussion and conclusions

A spatial Bayesian model is proposed in this paper for the downscaling of AOGCM predictors. This model

TABLE 5. RMSE between observed and simulated series for the spatial Bayesian model on the basis of climatic indices for each of the nine gauge weather stations during the validation period (1991–2000).

Climatic Indices	RMSE								
	1	2	3	4	5	6	7	8	9
Monthly mean Tmax (°C)	1.15	0.99	1.15	1.15	1.09	1.05	1.15	1.20	1.04
Monthly SD Tmax (°C)	0.88	0.77	0.76	0.83	0.84	0.87	0.77	0.82	0.79
Monthly mean Tmin (°C)	1.22	1.24	1.22	1.30	1.27	1.26	1.58	1.13	1.18
Monthly SD Tmin (°C)	1.19	0.97	0.94	0.86	1.02	1.01	0.92	1.00	1.18
DTR (°C)	0.92	0.84	0.70	0.75	0.71	0.84	0.94	1.01	0.64
FSL (day)	2.77	3.21	1.78	0.65	2.23	4.93	2.53	2.09	1.79
GSL (day)	21.84	14.82	13.31	19.33	13.62	15.25	17.92	14.28	12.57
FR-Th (day)	2.70	2.89	2.67	2.48	2.69	2.61	3.11	2.69	2.88
Tmax90 (°C)	1.28	1.15	1.17	1.23	1.18	1.30	1.30	1.27	1.39
Tmin90 (°C)	1.41	1.23	1.58	1.35	1.18	1.47	1.60	1.30	1.21

TABLE 6. RMSE between observed and simulated series for the MMLR model on the basis of climatic indices for each of the nine gauge weather stations during the validation period (1991–2000).

Climatic Indices	RMSE								
	1	2	3	4	5	6	7	8	9
Monthly mean Tmax (°C)	1.12	0.98	1.07	1.12	1.07	1.05	1.13	1.23	1.07
Monthly SD Tmax (°C)	0.77	0.79	0.72	0.77	0.69	0.83	0.75	0.75	0.75
Monthly mean Tmin (°C)	1.20	1.27	1.01	1.22	1.20	1.21	1.50	1.14	1.17
Monthly SD Tmin (°C)	0.82	0.90	0.80	0.88	0.74	0.95	1.10	0.79	0.77
DTR (°C)	0.90	0.92	0.64	0.74	0.66	0.87	0.95	1.08	0.63
FSL (day)	2.86	3.11	1.77	1.01	2.13	3.50	2.30	1.98	1.76
GSL (day)	21.61	14.67	12.09	19.68	12.67	14.23	19.00	13.86	12.73
FR-Th (day)	2.69	3.02	2.52	2.37	2.56	2.52	3.16	2.64	2.86
Tmax90 (°C)	1.26	1.18	1.18	1.15	1.00	1.26	1.27	1.27	1.29
Tmin90 (°C)	1.12	1.21	1.34	1.26	1.18	1.14	1.27	1.26	1.20

relies on the Bayesian framework in order to link the predictors to the predictands through unknown upscaled predictands (i.e., hidden variables) at the same horizontal resolution as the AOGCM predictors. The links between the predictors and the upscaled predictands are assumed to be constant for each month. On the other hand, the links between the upscaled and the original predictands change with their relative position by using the geostatistics theory of regularization of covariance and semi-variogram functions (see e.g., Goovaerts 2008). This approach enables the proposed model to provide estimates of the predictands at any given spatial location. The proposed model was applied to the estimation of minimum and maximum temperatures in the southern part of the province of Quebec, Canada. NCEP reanalysis data were used as predictors in order to assess the potential of the method, although the final objective is to use AOGCM predictors. The obtained results were found satisfactory compared to a regression-based downscaling approach (e.g., MMLR) using the same predictors (see appendix B for more details about the comparison method).

The proposed model provides interesting results for the downscaling of coarse-scale atmospheric predictors, specifically for the estimation of predictands such as minimum and maximum temperatures at ungauged locations. It uses Bayesian updating of seasonal norms at a finer spatial resolution (the prior distribution; computed using local information sources such as the altitude and the latitude) with temporal fluctuations induced by the coarse-resolution predictors (NCEP in our case). The method thus enables us to take advantage of both information sources in order to produce the final estimation maps.

The Bayesian approach also enables the proposed model to account for the uncertainty of both the estimated prior mean and the estimated parameters linking the NCEP (or AOGCM) predictors to the upscaled predictands. An algorithm is also proposed in order to jointly generate simulated predictands and parameters for the

evaluation of the integrals over unknown parameters and unknown upscaled predictands. The resulting estimations are thus based on a complete probabilistic formulation, which enables us to directly compute the probability of events on the predictands (e.g., intervals of temperatures). It is important to mention that accounting for such uncertainties mainly influences the variability (i.e., the variance) of the estimated conditional distribution, while the mean of this distribution remains almost unchanged. This is not surprising since (i) the distributions of the estimated parameters are assumed to be Gaussian, (ii) the mean of each Gaussian distribution is the corresponding observed parameter, and (iii) the variances of the estimated parameters are small thanks to the long period of observations.

Another important fact is that any model is a simplified representation of a complex reality. There is thus no true model. Either because of its adjusted parameters and/or to its specific hypotheses, any model carries a certain amount of uncertainty (e.g., variables or effects that are not accounted for in the model). Consequently, as the AOGCM values are themselves outputs of mathematical models, they are also a source of uncertainty for downscaling applications. In this study, this source of uncertainty was not accounted for even if it is expected to induce more variability in the mean of the predictands distribution. Again, working within a Bayesian framework is an advantage when trying to account for uncertainty. Because of space limitations, this element was left out in the present work. However, it is clear that such uncertainty should be taken into account by the downscaling techniques.

Another advantage of the proposed approach is that it sets the problem in a complete probabilistic framework. The different assumptions are stated and discussed in detail so that the limits of the approach are clearly defined. Moreover, by construction, it is the whole conditional distribution that is sought, thus modeling the mean of the downscaled predictands as well as their variability.

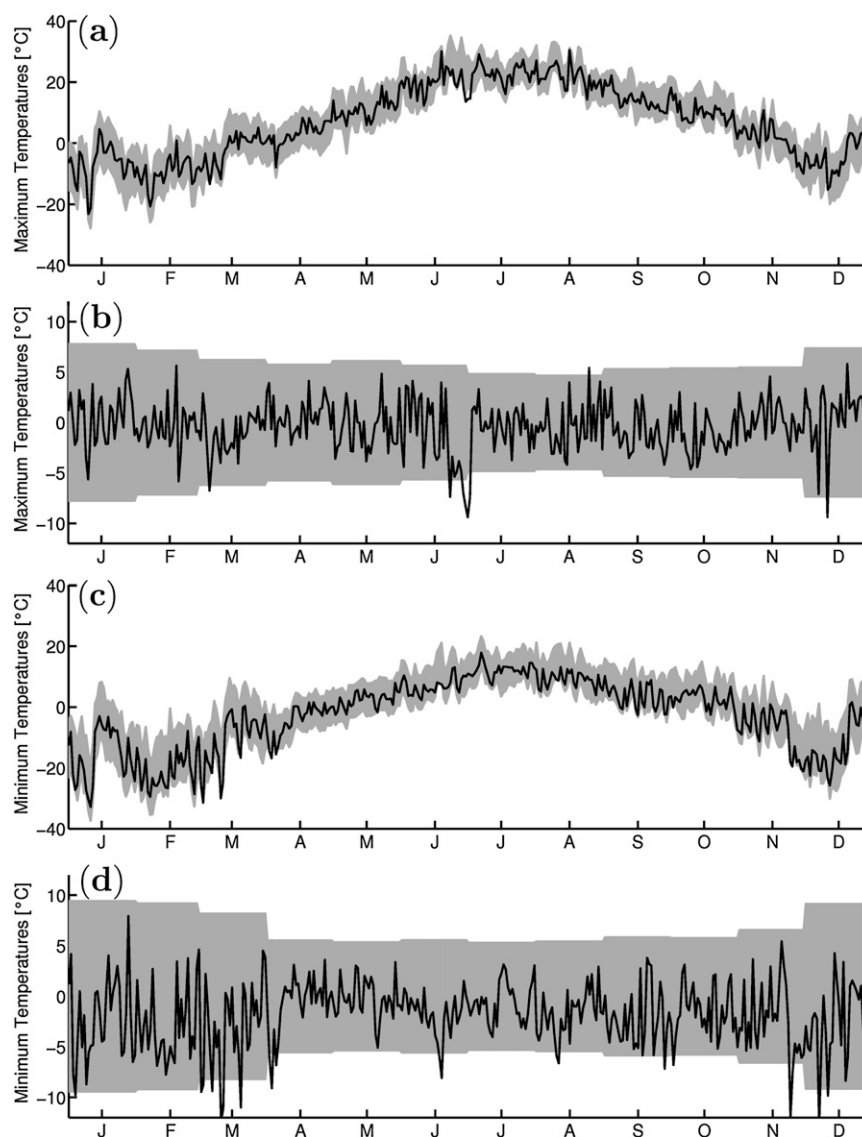


FIG. 9. Visual validation of the Bayesian model for the year 1995 at the weather station of Baie-Comeau. (a),(c) The evolutions of both the observed temperatures (bold lines) and the 95% credibility intervals (gray zones). (b),(d) The evolutions of both the estimation errors and the 95% credibility intervals centered on 0°C.

There is thus no need to rely on “inflation” techniques (e.g., von Storch 1999) in order to reproduce the total variability of the process. Indeed, the regression models generally reproduce the mean of the process conditionally to the selected independent variables. As a consequence, one can prove that the variability of the regression is always smaller than the initial variability (not shown here). In this regard, it is more relevant to rely on simulation techniques in order to compare the observed and estimated series of predictands since this is equivalent to considering that the observed series are random realizations of a theoretical probabilistic model.

This study was conducted for a relatively small area (cf. the scale of the AOGCM predictors). Consequently, the estimated parameters of the model for this study cannot be used for larger areas. Moreover, it is likely that parameters for two different regions will be different, as the climate conditions are sufficiently different and the combination of predictors and its links with the local or regional predictand should vary. It is thus recommended to rely on locally estimated parameters in order to avoid such risk. On the other hand, the number of AOGCM grid points in this study could easily be increased, which would improve the precision of both downscaling methods: the

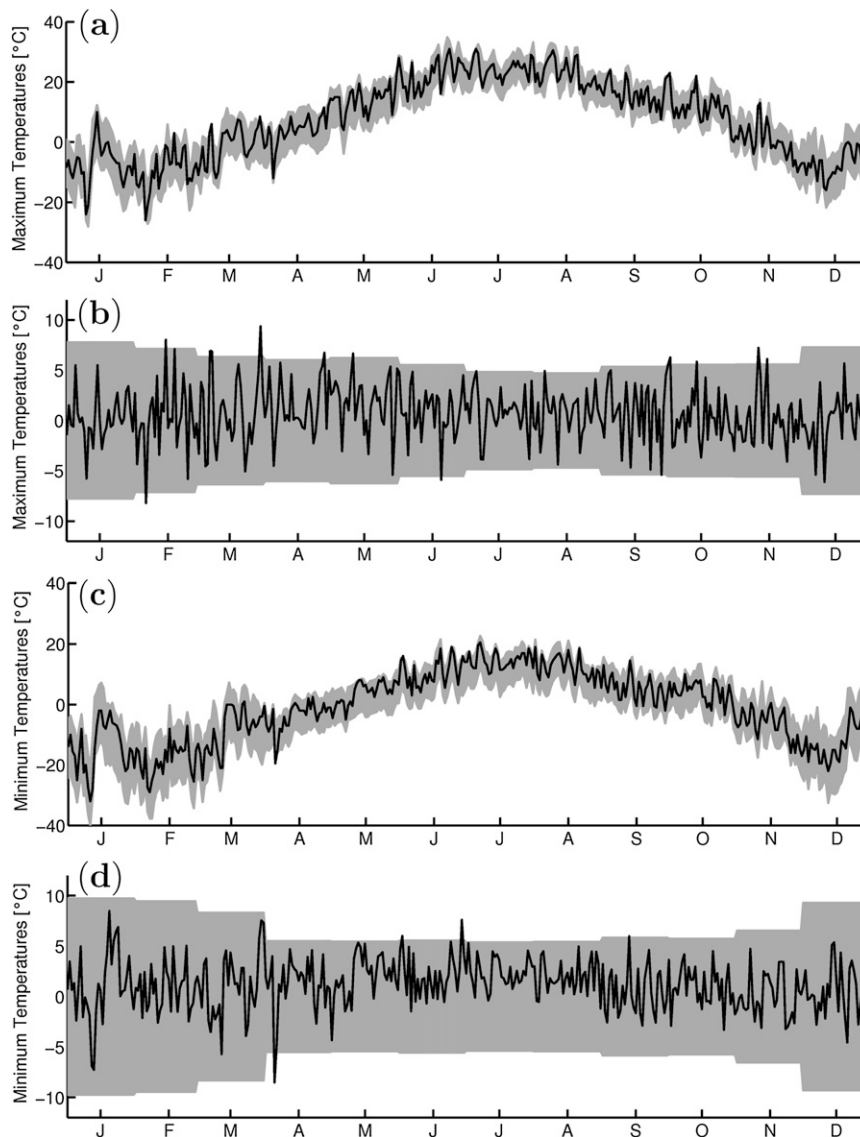


FIG. 10. Visual validation of the Bayesian model for the year 1995 at the weather station of Notre Dame du Lac. (a),(c) The evolutions of both the observed temperatures (bold lines) and the 95% credibility intervals (gray zones). (b),(d) The evolutions of both the estimation errors and the 95% credibility intervals centered on 0°C.

MMLR and the spatial Bayesian model. In that context, the use of regional-scale predictors from regional climate model (RCM) instead of coarse-scale AOGCM will be strongly beneficial for the downscaling process, and the precision or the spatial representatives of the upscaled predictands. Ongoing works will explore these issues.

The proposed GRM could certainly be improved. Benestad (2005) and Benestad et al. (2008) also proposed to account for the distance from the coast and the local slopes as independent variables. However, in our study, the small number of weather stations did not enable us

to account for numerous independent variables. We deliberately chose to limit the model to the latitudinal position ϕ and the altitude h . They were indeed the more significant independent variables explaining the local seasonal characteristics of temperatures in this region. However, the influence on regional and local features of temperatures from the distance of water masses (i.e., the estuary and the Gulf of St. Lawrence) as well as the orientation of the local slope and/or valley (ex. main rivers as Saguenay, St. Lawrence, and others) will be ones of the factors to include in future improved GRM model, in the context of the southern Quebec climate. A locally

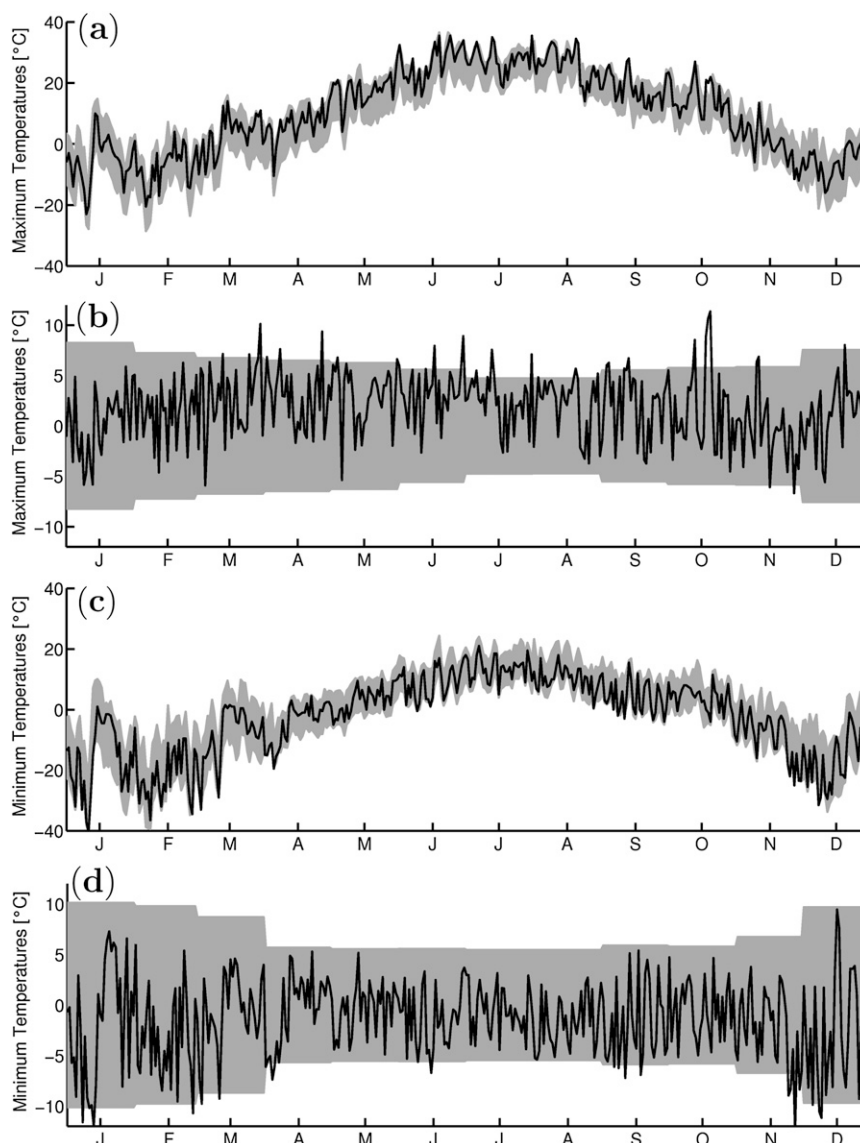


FIG. 11. Visual validation of the Bayesian model for the year 1995 at the weather station of La Tuque. (a),(c) The evolutions of both the observed temperatures (bold lines) and the 95% credibility intervals (gray zones). (b),(d) The evolutions of both the estimation errors and the 95% credibility intervals centered on 0°C.

estimated GRM is not necessarily a limitation of the proposed approach as downscaling methods are generally produced in relatively restricted areas (the size of, e.g., a hydrological basin).

Finally, the method could be adapted and extended in order to account for other sources of information. Basically, the cost for this would be to assume additional conditional hypotheses. The use of the spatial model and

TABLE 7. Quality assessment of the estimated series for the validation period (1991–2000) on the three validation stations. Criteria are ME, RMSE, and percentage of observations in the 95% credibility interval (% in CI).

Station Ref	Max temperature			Min temperature		
	ME (°C)	RMSE (°C)	% in CI	ME (°C)	RMSE (°C)	% in CI
A	−0.42	2.99	0.95	−1.95	4.24	0.93
B	0.67	3.42	0.94	1.05	3.67	0.95
C	1.24	3.81	0.91	−1.74	5.18	0.88

the regularization of covariance functions, also allows us to take into account other predictors at a third scale (intermediate or larger than the first AOGCM spatial resolution). This possibility might be of great interest for downscaling applications since the method would not only account for the change of spatial resolution but also for multiple scales or multiple AOGCM or RCM models at once. Future research efforts can focus on these factors and key components.

Acknowledgments. The financial support provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Ministère du Développement Economique, de l'Innovation et de l'Exportation du Québec (MDEIE), and the Canada Research Chair (CRC) Program is gratefully acknowledged. The authors would like to acknowledge the Data Access Integration (DAI, see online at <http://loki.qc.ec.gc.ca/DAI/DAI-e.html>) Team for providing the data and technical support. The Digital Elevation Model was provided by the Department of Natural Resources Canada (all rights reserved; more information is available online at <http://www.geogratis.ca/geogratis/fr/index.html>). The valuable comments provided by Dr. Patrick Bogaert (Université catholique de Louvain, Belgium), Dr. Dae Il Jeong (INRS-ETE), Dr. Philippe Gachon (Environment Canada), and the three anonymous reviewers noticeably contributed to the improvement of this paper.

APPENDIX A

Regression-Based Statistical Downscaling

Using the proposed notations of section 2a, one can summarize the regression-based downscaling approaches with the following equation:

$$\mathbf{Z}_0 = \mathbf{h}(\mathcal{Y}; \mathbf{\Gamma}) + \boldsymbol{\varepsilon}, \quad (\text{A1})$$

where the function $\mathbf{h}(\cdot)$ can be calibrated with historical data, $\mathbf{\Gamma}$ is the vector of parameters of the regression, and $\boldsymbol{\varepsilon}$ is a vector of residuals depicting the unexplained part of the regression.

The choice for the function $\mathbf{h}(\cdot)$ is not straightforward. It should be motivated by the different predictors \mathcal{Y} at hand and by their correspondences with \mathbf{Z}_0 . Among other possibilities, monthly MMLR models have been commonly implemented and have proven to provide accurate results in practice (Wilby et al. 2002; Benestad et al. 2008; Hessami et al. 2008; Jeong et al. 2010). Equation (19) then becomes

$$\mathbf{Z}_0 = \mathbf{\Gamma}_{d,0} + \mathbf{\Gamma}_{d,1}\mathcal{Y} + \boldsymbol{\varepsilon}, \quad (\text{A2})$$

where (i) $d = 1, \dots, 12$ corresponding to the actual month; (ii) $\mathbf{\Gamma}_{d,0}$ is the vector of dimension m corresponding to the intercepts for each predictand; and (iii) $\mathbf{\Gamma}_{d,1}$ is a $m \times (np)$ matrix of parameters. Such parameters have first to be estimated from historical data (generally from 20 to 30 yr of calibration data). It is a classic result of MMLR that using OLS for the estimation of such parameters leads to unbiased estimators of $\mathbf{\Gamma}$. Moreover, if the errors $\boldsymbol{\varepsilon}$ are multivariate Gaussian then the estimated vector $\mathbf{\Gamma}$ is also Gaussian with a covariance matrix that depends on the predictors \mathcal{Y} and on the covariance matrix of the errors $\boldsymbol{\varepsilon}$.

APPENDIX B

Regularization of Covariance Functions and Statistical Upscaling

Let $c(\cdot)$ be the covariance function of a given second order stationary random field on \mathbb{R}^d , that is, for each \mathbf{x}_i and $\mathbf{x}_j \in \mathbb{R}^d$:

$$c(h) = \text{Cov}[Z(\mathbf{x}_i), Z(\mathbf{x}_j)], \quad (\text{B1})$$

where $h = \|\mathbf{x}_i - \mathbf{x}_j\|$ is the distance between the two locations (e.g., Euclidean distance) and where $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_j)$ are the random variables corresponding to these locations.

Now defining $Z(A_k)$ as the upscaled variable on the spatial area $A_k \subseteq \mathbb{R}^d$ with

$$Z(A_k) = \frac{1}{|A_k|} \int_{A_k} Z(\mathbf{x}) d\mathbf{x}, \quad (\text{B2})$$

where $|A_k|$ is the area of A_k , it is straightforward to prove that, within the same random field, the block-to-point covariance between any random variable $Z(\mathbf{x}_i)$ and the upscaled variable $Z(A_k)$ is

$$\text{Cov}[Z(\mathbf{x}_i), Z(A_k)] = \frac{1}{|A_k|} \int_{A_k} c(\|\mathbf{x} - \mathbf{x}_i\|) d\mathbf{x} \quad (\text{B3})$$

because of the linearity of the integrals. Similarly, the block-to-block covariance between two upscaled variables $Z(A_k)$ and $Z(A_j)$ is defined by

$$\begin{aligned} \text{Cov}[Z(A_k), Z(A_j)] &= \frac{1}{|A_k||A_j|} \\ &\times \int_{A_j} \int_{A_k} c(\|\mathbf{x}_k - \mathbf{x}_j\|) d\mathbf{x}_k d\mathbf{x}_j. \end{aligned} \quad (\text{B4})$$

Regularized covariance functions are thus averaged covariance functions on a corresponding spatial area. Their behavior is generally smoother and their variances (i.e., values at lag $h = 0$) are smaller than the underlying covariance function at finer scale. For a complete description and for more details concerning their properties, the reader is referred to Goovaerts (2008).

Using the theory of regularized covariance function, one can build the linear relations between different scales of the process at different locations. As direct by-product of this property, one can also build estimates of an unobserved upper-scale process $Z(A_k)$ from observed finer-scale processes $\mathbf{Z} = (Z(\mathbf{x}_1) \dots Z(\mathbf{x}_n))^T$. If $\boldsymbol{\mu}_Z$ and μ_{A_k} , respectively the mean vector of \mathbf{Z} and of $Z(A_k)$, are both known, then one can use the aggregation simple kriging (ASK) formula:

$$\widehat{Z(A_k)} = \mu_{A_k} + \boldsymbol{\sigma}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}_Z), \quad (\text{B5})$$

where $\boldsymbol{\Sigma}$ is the covariance matrix of \mathbf{Z} and $\boldsymbol{\sigma}$ is the vector of covariances between $Z(A_k)$ and different elements of \mathbf{Z} . In this case (i.e., estimation of upper scale process), the vector $\boldsymbol{\sigma}$ only has to be computed using regularized covariance functions, since $\boldsymbol{\Sigma}$ is directly computed from the covariance function $c(\cdot)$. If μ_{A_k} is unknown, then aggregation ordinary kriging (AOK) can be used instead with

$$\widehat{Z(A_k)} = \boldsymbol{\lambda}^T \mathbf{Z}, \quad (\text{B6})$$

where $\boldsymbol{\lambda}$ is the solution of the kriging system:

$$\begin{pmatrix} \boldsymbol{\Sigma} & \mathbf{1} \\ \mathbf{1}^T & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{\lambda} \\ \nu \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma} \\ 1 \end{pmatrix}, \quad (\text{B7})$$

where $\mathbf{1}$ is a column of 1 and where ν is the Lagrangian multiplier that ensures that

$$\sum_{i=1}^n \lambda_i = 1.$$

The two main advantages of using regularized covariance functions for the estimation of the upper-scale process are the following:

- 1) As already mentioned, this is a spatial estimation. One can thus estimate the process on the spatial areas that do not contain any finer scale observation. This is not possible when using (e.g., simple averaging on the spatial area).
- 2) The second interesting property is that this estimation accounts for the spatial dispersion of the observed data. The issue with spatial data is that two

closely located observations are likely to have similar values. The partial information of the second observation is thus small. Knowing the value of the second observation does not improve significantly our total knowledge. This clustering issue is accounted for when using regularized covariance functions in opposition to using simple averaging.

On the other hand, the main disadvantage of the regularized covariance functions is that in practice, one has to discretize the spatial area for the evaluation of the integrals. The resolution of the discretized area should be (i) finer enough in order to accurately estimate the integrals and (ii) not too fine in order to limit the computational requirements. This choice of discretization resolution can be motivated by the shape of the finer-scale covariance function. For instance, smoother covariance functions (e.g., with higher range or lower nugget effect) can accept coarser discretization. This is because the variations of a smooth covariance function on a spatial area is small, the averaged covariance is thus even more stable.

It is important to note that the same methodology can be used for the estimation of finer-scale processes from upper ones. Again, the reader is referred to Goovaerts (2008) for more details about “area-to-point” kriging. However, it is not likely that the upper-scale observations will correctly reproduce the finer-scale behavior of the process. The main advantage of using this change of scale method for the estimation of the finer-scale process is that it ensures the conservation of the information at the coarser scale. In other words, averaging the estimation on the same scale will provide the original block value.

REFERENCES

- Baigorria, G. A., J. W. Hanse, N. Ward, J. W. Jones, and J. J. O'Brien, 2008: Assessing predictability of cotton yields in the Southeastern United States based on regional atmospheric circulation and surface temperatures. *J. Appl. Meteor. Climatol.*, **47**, 76–91.
- Bárdossy, A., and J. Van Mierlo, 2000: Regional precipitation and temperature scenarios for climate change. *Hydrol. Sci. J.*, **45**, 559–575.
- Bates, B., S. Charles, and J. Hughes, 1998: Stochastic downscaling of numerical climate model simulations. *Environ. Model. Software*, **13** (3–4), 325–331.
- Bellone, E., J. Hughes, and P. Guttorp, 2000: A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. *Climate Res.*, **15**, 1–12.
- Benestad, R. E., 2002: Empirically downscaled temperature scenarios for Northern Europe based on a multi-model ensemble. *Climate Res.*, **21**, 105–125.
- , 2005: Climate change scenarios for northern Europe from multi-model IPCC AR4 climate simulations. *Geophys. Res. Lett.*, **32**, L17704, doi:10.1029/2005GL023401.
- , 2007: Novel methods for inferring future changes in extreme rainfall over Northern Europe. *Climate Res.*, **34**, 195–210.

- , D. Chen, and I. Hanssen-Bauer, 2008: *Empirical-Statistical Downscaling*. World Scientific Publishing Company, 228 pp.
- Bogaert, P., and D. Fasbender, 2007: Bayesian data fusion in a spatial prediction context: A general formulation. *Stochastic Environ. Res. Risk A*, **21**, 695–709.
- Data Access Integration, 2009: DAI data catalogue version 1.0. April 2009, Montreal, QC, Canada, 24 pp. [Available online at http://loki.qc.ec.gc.ca/DAI/doc/catalogue/DAI_Catalogue_January2010_En.pdf.]
- Chilès, J.-P., and P. Delfiner, 1999: *Geostatistics: Modeling Spatial Uncertainty*. John Wiley and Sons, Inc., 720 pp.
- Coelho, C. A. S., S. Pezzulli, M. Balmaseda, F. J. Doblas-Reyes, and D. B. Stephenson, 2004: Forecast calibration and combination: A simple Bayesian approach for ENSO. *J. Climate*, **17**, 1504–1516.
- DAI Team, 2008: Sets of predictor variables derived from CGCM3 T47 and NCEP/NCAR reanalysis. Tech. Rep. 1.1, Montreal, QC, Canada, 15 pp.
- Flato, G., G. Boer, W. Lee, N. McFarlane, D. Ramsden, M. Reader, and A. Weaver, 2000: The Canadian center for climate modeling analysis global coupled model and its climate. *Climate Dyn.*, **16**, 451–467.
- Gachon, P., and Coauthors, 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), Environnement Canada, Montreal, QC, Canada, 209 pp.
- Gerardin, V., and D. McKenney, 2001: Une classification climatique du Québec à partir de modèles de distribution spatiale de données climatiques mensuelles: Vers une définition des bioclimats du Québec. Contribution du Service de la cartographie écologique 60, Ministère de l'Environnement, 48 pp.
- Goovaerts, P., 2008: Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Math. Geol.*, **40**, 101–128.
- Gordon, C., C. Cooper, C. A. Senior, H. Banks, J. M. Gregory, T. C. Johns, J. F. B. Mitchell, and R. A. Wood, 2000: The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Climate Dyn.*, **16** (2–3), 147–168.
- Grotch, S. L., and M. C. MacCracken, 1991: The use of general circulation models to predict regional climatic change. *J. Climate*, **4**, 286–303.
- Hessami, M., P. Gachon, T. B. M. J. Ouarda, and A. St.-Hilaire, 2008: Automated regression-based statistical downscaling tool. *Environ. Model. Software*, **23**, 813–834.
- Hughes, J., and P. Guttorp, 1994: Incorporating spatial dependence and atmospheric data in a model of precipitation. *J. Appl. Meteor.*, **33**, 1503–1515.
- , —, and S. Charles, 1999: A non-homogeneous hidden Markov model for precipitation occurrence. *J. Roy. Stat. Soc. C: Appl. Stat.*, **48**, 15–30.
- Hundechea, Y., and A. Bárdossy, 2005: Trends in daily precipitation and temperature extremes across western Germany in the second half of the 20th century. *Int. J. Climatol.*, **25**, 1189–1202.
- Huth, R., and J. Kysely, 2000: Constructing site-specific climate change scenarios on a monthly scale. *Theor. Appl. Climatol.*, **66** (1–2), 13–27.
- Jeong, D., A. St-Hilaire, T. Ouarda, and P. Gachon, 2010: Development of multisite statistical downscaling model and projection of future temperature extremes over Montréal, Canada. *Proc. 44th Annual Canadian Meteorological and Oceanographic Society (CMOS) Congress*, Ottawa, ON, Canada, CMOS, 3C14.1.
- Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–267.
- Lim, Y.-K., D. W. Shin, S. Cocke, T. E. LaRow, J. T. Schoof, J. J. O'Brien, and E. P. Chassignet, 2007: Dynamically and statistically downscaled seasonal simulations of maximum surface air temperature over the southeastern United States. *J. Geophys. Res.*, **112**, D24102, doi:10.1029/2007JD008764.
- Mendes, J. M., K. F. Turkman, and J. Corte-Real, 2006: A Bayesian hierarchical model for local precipitation by downscaling large-scale atmospheric circulation patterns. *Environmetrics*, **17**, 721–738.
- Michelangeli, P.-A., M. Vrac, and H. Loukos, 2009: Probabilistic downscaling approaches: Application to wind cumulative distribution functions. *Geophys. Res. Lett.*, **36**, L11708, doi:10.1029/2009GL038401.
- Scinocca, J. F., N. A. McFarlane, M. Lazare, J. Li, and D. Plummer, 2008: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, **8** (23), 7055–7074.
- Tebaldi, C., L. Mearns, D. Nychka, and R. Smith, 2004: Regional probabilities of precipitation change: A Bayesian analysis of multimodel simulations. *Geophys. Res. Lett.*, **31**, L24213, doi:10.1029/2004GL021276.
- Vincent, L., X. Zhang, B. Bonsal, and W. Hogg, 2002: Homogenization of daily temperatures over Canada. *J. Climate*, **15**, 1322–1334.
- von Storch, H., 1999: On the use of inflation in statistical downscaling. *J. Climate*, **12**, 3505–3506.
- Vrac, M., and P. Naveau, 2007: Stochastic downscaling of precipitations: From dry events to heavy rainfalls. *Water Resour. Res.*, **43**, W07402, doi:10.1029/2006WR005308.
- , M. Stein, and K. Hayhoe, 2007: Statistical downscaling of precipitations through nonhomogeneous stochastic weather typing. *Climate Res.*, **34**, 169–184.
- Wilby, R., H. Hassan, and K. Hanaki, 1998: Statistical downscaling of hydrometeorological variables using general circulation model output. *J. Hydrol.*, **205**, 1–19.
- , C. Dawson, and E. Barrow, 2002: SDSM—A decision support tool for the assessment of regional climate change impacts. *Environ. Model. Software*, **17**, 145–157.