Predictor selection for downscaling GCM data with LASSO

Dorra Hammami,[1] Tae Sam Lee,[2] Taha B. M. J. Ouarda,[1,3] and Jonghyun Lee[4]

Received 29 March 2012; revised 1 August 2012; accepted 2 August 2012; published 14 September 2012.

[1] Over the last 10 years, downscaling techniques, including both dynamical (i.e., the regional climate model) and statistical methods, have been widely developed to provide climate change information at a finer resolution than that provided by global climate models (GCMs). Because one of the major aims of downscaling techniques is to provide the most accurate information possible, data analysts have tried a number of approaches to improve predictor selection, which is one of the most important steps in downscaling techniques. Classical methods such as regression techniques, particularly stepwise regression (SWR), have been employed for downscaling. However, SWR presents some limits, such as deficiencies in dealing with collinearity problems, while also providing overly complex models. Thus, the least absolute shrinkage and selection operator (LASSO) technique, which is a penalized regression method, is presented as another alternative for predictor selection in downscaling GCM data. It may allow for more accurate and clear models that can properly deal with collinearity problems. Therefore, the objective of the current study is to compare the performances of a classical regression method (SWR) and the LASSO technique for predictor selection. A data set from 9 stations located in the southern region of Québec that includes 25 predictors measured over 29 years (from 1961 to 1990) is employed. The results indicate that, due to its computational advantages and its ease of implementation, the LASSO technique performs better than SWR and gives better results according to the determination coefficient and the RMSE as parameters for comparison.

Citation: Hammami, D., T. S. Lee, T. B. M. J. Ouarda, and J. Lee (2012), Predictor selection for downscaling GCM data with LASSO, *J. Geophys. Res.*, *117*, D17116, doi:10.1029/2012JD017864.

1. Introduction

[2] Increasing attention is being devoted to the estimation of plausible scenarios of future climate evolution. The main source of information used for this purpose is derived from climate change scenarios developed using global climate models (GCMs). Because the resolution of GCMs is too coarse for regional and local climate studies, downscaling methods are one of the best alternatives for investigating GCM data in local impact studies. A number of approaches are used for downscaling. Regression models are regularly used due to their ease of implementation.

[3] Predictor selection is one of the most important steps in downscaling procedures. It can be considered as the basic step in realizing a successful climate scenario. Predictor selection involves an attempt to find the best model and to limit the number of independent variables when a number of potential independent variables exist. One downscaling technique is the stepwise regression (SWR) method. The first widely used algorithm summarizing the idea of SWR was proposed by *Efroymson* [1966] and developed by *Draper and Smith* [1966]. It is termed a variable selection method, which selects a particular set of independent variables.

[4] The first application of Efroymson's algorithm was reported by *Jennrich and Sampson* [1968] for non-linear estimation. *Lund* [1971] applied the SWR procedure to the problem of estimating precipitation in California. *Cohen and Cohen* [1975] investigated the two forms of the SWR method (forward and backward selection). *Hocking* [1976] described the stepwise method as one of the most important tools used for the analysis and selection of variables in linear regression. Despite the common use of this method in variable selection, *Flom and Cassell* [2007] expressed the limits of SWR and recommended that this method should not be used due to its weaknesses. In fact, the Fisher test and all other statistical tests are normally based on a single hypothesis under examination; however, with SWR, this assumption is violated in that it is intended for one to many tests.

[5] A possible alternative for overcoming the limits of SWR was suggested by *Tibshirani* [1996]. *Tibshirani* [1996] suggested a new method in variable selection and shrinkage

[1]INRS-ETE, University of Québec, Quebec, Quebec, Canada.
[2]Department of Civil Engineering, ERI, Gyeongsang National University, Jinju, South Korea.
[3]Water and Environmental Engineering, Masdar Institute of Science and Technology, Abu Dhabi, United Arab Emirates.
[4]Environmental Fluid Mechanics and Hydrology, Department of Civil and Environmental Engineering, Stanford University, Stanford, California, USA.

Corresponding author: T. Lee, Department of Civil Engineering, ERI, Gyeongsang National University, 501 Jinju-daero, Jinju, Gyeongnam 660-701, South Korea. (tae3lee@gnu.ac.kr)

0148-0227/12/2012JD017864

that retains the positive features of the most commonly used methods for improving ordinary least squares (OLS) estimates, i.e., subset selection and ridge regression. The method was named "Least Absolute Shrinkage and Selection Operator" (LASSO). A new algorithm for LASSO was proposed by *Fu* [1998] in a study on the structure of bridge estimators. Another algorithm for the LASSO method was suggested by *Grandvalet* [1998] and *Grandvalet and Canu* [1999] using the quadratic penalization, and they showed the outcomes of this equivalence. *Osborne et al.* [2000a] treated the LASSO method as a convex programming problem and derived its dual. In addition, *Osborne et al.* [2000b] proposed a new LASSO algorithm for solving constrained problems. Lasso overcomes some of the drawbacks of the most common methods currently used for the variable selection and shrinkage problem.

[6] Because the LASSO function has some non-differentiable points, *Schmidt* [2005] proposed assembling different optimization strategies to solve this problem. Different versions of the LASSO procedure have been developed based on researchers' different views of Tibshirani's theory. *Kyung et al.* [2010] proposed a LASSO method using the Bayesian formulation that encompasses most versions of LASSO.

[7] Despite its advantages, the LASSO approach remains unutilized in hydro-climatology. The objective of the current study is to present the suitability of the LASSO technique for predictor selection in downscaling compared to the traditional approach, stepwise regression. The maximum and minimum temperatures over 1961–1990 time window in the Québec region, Canada were used to show and compare the performances of the two models (LASSO and SWR).

[8] A mathematical description of the two methods is presented in the next section. In section 3, the methodology used for predictor selection in downscaling is explained. The data set used for this case study is described in section 4. In section 5, we present the results obtained for SWR and LASSO along with a comparison of their performances. In section 6, we discuss the results and provide an overview of the accomplished work and our main conclusions and recommendations concerning the two selection methods.

2. Theoretical Background

2.1. Stepwise Regression

[9] In statistical analyses, regression models are commonly used to find the combination of predictors x_i that best explains the dependent variable y. Regression models are often used for prediction [*Copas*, 1983]. The first model used is a simple linear model that allows for an estimation of the response variable y using a unique explanatory variable x, following a model of the form:

$$\hat{y} = ax + b \tag{1}$$

where $\hat{y}$ is the estimation of the dependent variable and a and b are the model parameters. However, this simple model is often inefficient in estimating the dependent variable, especially when more than one explanatory variable contributes to the dependent variable. In this case, multiple regression models must be applied. SWR is mainly used in selecting predictors from a large number of explicative variables. With SWR, the number of explicative variables is reduced by selecting the best performing variables. A comparison between different combinations of independent variables is generated step by step and validated by Fisher's test based on a comparison of the sum of the residual squares. However, other alternatives can also be used, such as the AIC and BIC criteria. SWR is considered to be a familiar, easily explained method and is widely used and implemented. Thus, it can easily be extended to other regression problems. It provides good results, especially for large data sets, and can be improved by complex stopping rules [*Weisberg*, 2010].

[10] There are three different SWR algorithms: (1) the addition of variables one by one according to a specific criterion (forward selection, FS), (2) the deletion of variables one by one according to another criterion (backward elimination, BE) and (3) the combined criteria of the two previous methods (SWR itself). See *Efroymson* [1966] or *Draper and Smith* [1966] for more details.

2.1.1. Forward Selection

[11] This method starts without the independent variable in the equation and adds the explanatory variables one by one until all explanatory variables are added or a stopping criterion is satisfied [*Hocking*, 1976]. The selection of the added variables is determined by a well-defined criterion because each predictor is evaluated according to its correlation with the dependent variable. In fact, FS first chooses the predictor that is most correlated to the dependent variable and then chooses among the remaining variables with the highest partial correlation, keeping the already selected variable constant. A succession of Fisher's tests is applied by adding variable i to the model if

$$F_i = \max \frac{\left(RSS_p - RSS_{p+i}\right)}{\sigma^2_{\mathrm{p+i}}} \geq F_{in} \tag{2}$$

where F_{in} denotes the Fisher value limit determined from Fisher tables, RSS_p denotes the residual sum of squares of the selected variables, RSS_{p+i} denotes the residual sum of squares obtained when variable i is added to the current p-term equation and $\boldsymbol{\sigma}^2_{p+i}$ denotes the variance of the model when variable i is added to the current p-term equation.

2.1.2. Backward Elimination

[12] BE is the reverse of FS. All of the explicative variables are included from the start in the model, and variables are eliminated one at a time. In each step, the variable with the smallest F-ratio is eliminated if the F-ratio is less than a specified threshold F_{out}. A succession of Fisher's tests is applied by eliminating variable i from the model if

$$F_i = \min \frac{\left(RSS_{p-i} - RSS_p\right)}{\sigma^2_p} \leq F_{out} \tag{3}$$

where $RSS_{p\text{-}i}$ denotes the residual sum of squares obtained when variable i is eliminated from the current equation and F_{out} denotes the limit value for determining whether a variable should be eliminated from the current equation.

2.1.3. Stepwise Regression

[13] This method combines the FS and BE algorithms. It consists of adding variables one at a time according to the criterion of partial correlation while checking whether the pre-selected variables are still significant in each step.

This variant uses two stopping criteria, which determine the introduction of a new variable and the elimination of an existing one.

2.2. LASSO

2.2.1. Introduction to LASSO

[14] The development of penalized regression has been the concern of many studies. In traditional regression, the ordinary least squares method (OLS) is used to minimize the residual squared errors, but these estimates have some critical drawbacks. If the number of independent variables is large or if the regressor variables are highly correlated, then the variance of the least squares coefficient estimates may be unacceptably high, which leads to a lack in interpretation and prediction accuracy; see *Tibshirani* [1996] for more details. Thus, for a large number of predictors, there could be problems with regard to multicollinearity and the selection of smaller subsets that include both the most important predictors that can fit and the whole set of variables [*Kyung et al.*, 2010]. For these reasons, OLS estimates may not always satisfy data analysts. Such concerns have led to the development of methods with different penalties to obtain more interpretable models and more accurate prediction methods. *Hoerl and Kennard* [1970] proposed ridge regression with an L2 quadratic penalty of the following form:

$$\sum_{i=1}^{p}\beta_i^2 \le t \tag{4}$$

where p denotes the number of independent variables, β_i denotes the regression coefficient of each variable and t denotes the tuning parameter, which is also called the shrinkage parameter. Ridge regression is a stable method that improves the prediction performance by overcoming the multicollinearity problem.

[15] *Frank and Friedman* [1993] introduced bridge regression, which minimizes the RSS subject to

$$\sum_{i=1}^{p}|\beta|_i^{\gamma} \le t \tag{5}$$

where γ denotes some number greater than or equal to 0. This constraint is called an L^{γ} norm. *Tibshirani* [1996] introduced the LASSO method which allows both continuous shrinkage and variable selection and minimizes the RSS subject to an L1 penalty corresponding to $\gamma = 1$.

[16] The most commonly used techniques for improving OLS estimates are subset selection and ridge regression. First, subset selection can give interpretable models, but because of its nature as a discrete process that can be influenced by the smallest change in the data set, it cannot give highly accurate prediction models. Second, ridge regression is considered to be a more stable technique but does not set any coefficients to 0; hence, it does not provide easily interpretable models. Thus, *Tibshirani* [1996] proposed a new technique (LASSO) that can retain the advantages of both subset selection and ridge regression.

2.2.2. Definition of LASSO

[17] Consider a data set($\mathbf{x}^i$, y_i), i = 1, 2, …, n, where y_i are the response variables, n is the sample size and $\mathbf{x}^i = (x_{i1}, \ldots, x_{ip})$ is the matrix of standardized regressors. Most penalized regression methods are used in the case where $n > p$.

[18] Considering $\hat{\beta} = \left(\hat{\beta}_1, \ldots, \hat{\beta}_p\right)^T$, the LASSO estimate $\left(\hat{\alpha}, \hat{\beta}\right)$ is defined by

$$\left(\hat{\alpha}, \hat{\beta}\right) = \arg\min\left\{\sum_{i=1}^{n}\left(y_i - \alpha - \sum_j \beta_j x_{ij}\right)^2\right\} \quad \text{subject to} \sum_j \left|\beta_j\right| \le t \tag{6}$$

where t is a positive constant called the tuning parameter, which basically determines the balance between model fitting and sparsity in the solution. To omit the parameter α, we can assume that $\hat{\alpha} = \bar{y}$ and $\bar{y} = 0$ without loss of generality. Using this formulation, the LASSO technique can provide values of exactly zero for some coefficients β_j, which results in interpretable models that are more stable than those obtained by subset selection using a low variance. The summation of squared difference between observations and its estimates (RSS) in equation (6) creates elliptical contours for the solution without constraints and the center of the contours is on the OLS estimates. The constraint in equation (6) illustrates the rhombus shape of the constraint region. The LASSO solution corresponds to the first point at which the contours intersect with the rhombus; this will sometimes occur at a corner, corresponding to a zero coefficient [*Tibshirani*, 1996].

[19] The parameter $t \ge 0$ is a shrinkage parameter used to control and limit the amount of shrinkage and elimination applied to the estimates. Let $\hat{\beta}_j^0$ be the full least squares estimates and $t_0 = \sum \left|\hat{\beta}_j^0\right|$. Note that all $t < t_0$ will cause the set of the coefficients to move towards 0, and some of these coefficients may be exactly equal to 0. For example, if $t = t_0/2$, the effect is roughly similar to finding the best subset of size $p/2$. Note also that the design matrix need not be of full rank [*Tibshirani*, 1996].

[20] The motivation for the LASSO technique came from Breiman's non-negative garotte method, as proposed by *Breiman* [1995]. It minimizes

$$\sum_{i=1}^{N}\left(y_i - \alpha - \sum_j c_j \hat{\beta}_j^0 x_{ij}\right)^2 \quad \textit{subject to } c_j \ge 0, \sum c_j \le t. \tag{7}$$

[21] The method begins with the OLS estimates and shrinks them by the non-negative factors c_j, whose sum is constrained. The work of *Breiman* [1995] showed that the garotte method behaves better than subset selection in terms of prediction error. It can be very competitive with regard to ridge regression in extensive simulation cases, except when the true model has many small non-zero coefficients. However, the garotte method presents some drawbacks. It depends closely on the sign and the magnitude of the OLS estimates [*Tibshirani*, 1996]. In fact, in some cases in which the settings are strongly correlated and the OLS estimates are inefficient, the garotte may be inaccurate as a result. In contrast, this is not the case with LASSO, which avoids the explicit use of OLS estimates. Briefly, the garotte function is very similar to the LASSO function but is significantly different when the design is not orthogonal [*Tibshirani*, 1996].

[22] Two kinds of LASSO formulations are possible: the constrained formulation shown above (equation (6)) and the unconstrained one. A matrix form of the unconstrained expression can be presented as follows:

$$\text{Min } \|\mathbf{x}\beta - \mathbf{y}\|_2^2 + \lambda\|\beta\|_1 \tag{8}$$

where **x** is the matrix of the standardized regressors, **y** represents the matrix of the response variables and λ represents a parameter equivalent to the tuning or shrinkage parameter *t*. Note that denotes the L^1 norm and that $\| \, . \, \|_1$ denotes the L^1 norm and that $\| \, . \, \|_2$ denotes the L^2 norm.

2.2.3. Computation of the LASSO

[23] To compute the LASSO, *Fan and Li* [2001] proposed an efficient alternative including the unconstrained LASSO formulation based on the following approximation:

$$|\beta| \approx \frac{\beta_i^2}{|\beta_i|} \tag{9}$$

[24] *Perkins et al.* [2003] proposed an optimization approach that computes the LASSO based on the unconstrained problem using the definition of the function gradient that gives coefficients exactly equal to 0 [*Perkins et al.*, 2003; *Schmidt*, 2005].

[25] *Efron et al.* [2002] proposed the least angle regression selection (LARS) method for a model selection algorithm. They showed that this method has a short computation time when implementing the LASSO. *Osborne et al.* [2000a, 2000b] proposed an active set method based on local linearization. A compact descent algorithm (as described in *Kyung et al.* [2010]) can solve the selection problem for a particular tuning parameter based on the constrained LASSO formulation [*Osborne et al.*, 2000a; *Schmidt*, 2005].

[26] In the current study, the active set method is used because it does not require the number of variables in the problem to be doubled, does not require an exponential number of constraints, does not give degenerate constraints, has fast convergence properties and because the iteration cost can be kept relatively low through efficient implementation [*Schmidt*, 2005]. The basis of the implementation of Osborne's algorithm includes local linearization about β and the active set method operating only on non-zero variables and a single zero-valued variable. A new optimization problem is presented as follows, assuming that θ = *sign* (β) and that τ simply denotes members of the active set:

$$\begin{aligned} &\min_{h_\tau} f(\beta_\tau + h_\tau) \\ &s.t. \quad \theta_\tau^T(\beta_\tau + h_\tau) \le t \end{aligned} \tag{10}$$

where h_τ corresponds to the zero-valued elements outside the active set. The active set is initially empty, and the algorithm starts by assigning 0 to all of the elements. At the end of each iteration, one zero-valued element is added to the active set (one that is not already in τ) corresponding to the element with the largest violation. Using $\beta^+ = \beta + h$ and $r^+ = \mathbf{y} - \mathbf{x}\beta^+$, the violation function is defined as follows:

$$v^+ = \frac{\mathbf{x}^T r^+}{\|\mathbf{x}_\tau^T r^+\|_\infty} \tag{11}$$

where $\| \, . \, \|_\infty$ denotes the infinite norm. The solution to the KKT conditions of this problem is then

$$\begin{aligned} \mu &= \max\left(0, \frac{\theta_\tau^T\left(\mathbf{x}_\tau^T\mathbf{x}_\tau\right)^{-1}\mathbf{x}_\tau^T\mathbf{y} - t}{\theta_\tau^T\left(\mathbf{x}_\tau^T\mathbf{x}_\tau\right)^{-1}\theta_\tau}\right) \\ h_\tau &= \left(\mathbf{x}_\tau^T\mathbf{x}_\tau\right)^{-1}\left(\mathbf{x}_\tau^T(\mathbf{y} - \mathbf{x}_\tau\beta_\tau) - \mu\theta_\tau\right) \end{aligned} \tag{12}$$

[27] Because sign (0) is not well defined, the sign of the variable that will be introduced into the active set is the sign of its violation. Optimality is achieved when the magnitude of the violation for all elements outside the active set is less than 1.

[28] The cost in terms of iterations is small because the active set is small. However, the active set grows proportionally to the number of variables. The basis for maintaining efficient iterations with a large number of variables involves the maintenance and updating of a QR factorization of $\mathbf{x}_\tau^T\mathbf{x}_\tau$. For more details concerning the algorithm, the reader is referred to *Schmidt* [2005].

2.2.4. Standard Errors

[29] According to the original paper describing the LASSO technique, the manner for obtaining an accurate estimate of the standard errors of the LASSO is not straightforward due to its nature as a non-linear and non-differentiable function. An assessment of the standard errors can be performed using a bootstrap technique, by (1) fixing t, which requires one to select the best subset and then use the least squares error for that subset, or (2) proceeding by optimizing over t for each bootstrap sample.

2.2.5. Choice of the Tuning Parameter

[30] Typical approaches for estimating the tuning parameter include cross-validation; generalized cross-validation and analytical unbiased estimates of risk [*Tibshirani*, 1996]. Theoretically, cross-validation and generalized cross-validation methods are used for a random distribution of predictors. The analytical unbiased estimate of risk method is applicable when the distribution of the predictors is well known (the **X**-fixed case) [*Tibshirani*, 1996]. However, in real problems, there are often no clear differences between the results of the three methods; the most convenient method can be chosen. In the current study, we chose to work with the cross-validation method.

3. Predictor Selection

[31] The SWR and LASSO models for each station and month were tested with NCEP/NCAR predictors (see Table 3) over the 1961–1990 time window to select the best fitting predictors to apply in a downscaling context. The NCEP/NCAR reanalysis data constitute an updating gridded data set that represents the state of the Earth's atmosphere. The data set incorporates both observations and numerical weather prediction model outputs. These data were produced by the National Center for Environmental Prediction (NCEP) and the National Center for Atmospheric Research (NCAR).

[32] For each station, we investigated the predictors interpolated on the same grid-cell in which the station is located. The data sets for the predictands, the maximum and minimum temperature and predictors were divided by month to avoid the effect of seasonality, and missing values were then

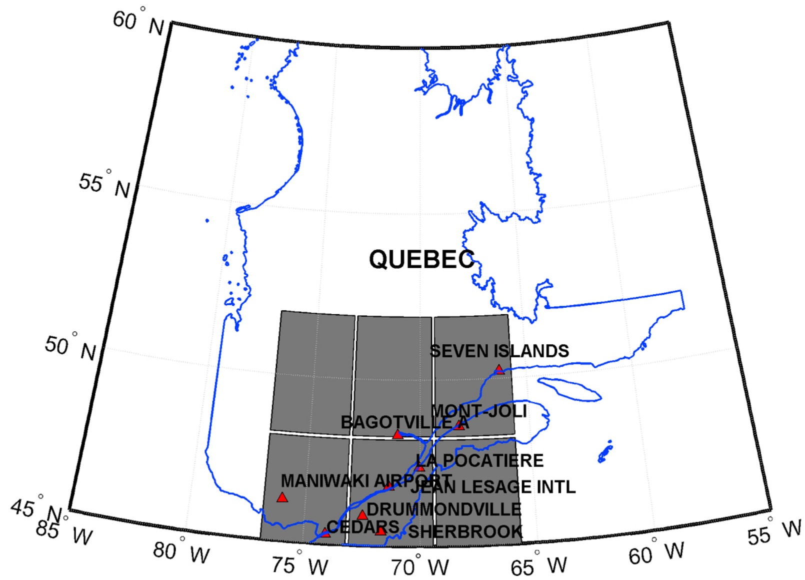

Figure 1. Meteorological stations located around the Gulf of St. Lawrence.

removed. When applying the LASSO technique, many steps are required for optimal results. For each month, a cross-validation technique is used to choose the best tuning parameter, i.e., the one corresponding to the minimum mean square error (MSE).

[33] To find the best tuning parameter for the active set method, we proceeded as in *Tibshirani* [1996]. We used 50 discrete t-values ($t_1 < t_2 < \ldots < t_{49} < t_{50}$) and 5-fold cross-validation, which, in essence, randomly breaks all of the data (**y**) up into 5 sets (y_1, y_2, y_3, y_4 *and* y_5), then applies the LASSO minimization using all of the data (**y**) except one set (for example, y_1, so y_2, y_3, y_4 and y_5 are used) with one t-value (say t_1) and obtains the regression coefficients (β_i). With the resulting coefficients, the values of the remaining set (say y_1) are estimates, and the MSE is computed separately for $y_{i,i=1,\cdots,5}$, defined by

$$MSE = mean\left((y_i - \hat{y}_i)^2\right) \tag{13}$$

where y_i is the dependent variable and $\hat{y}_i$ is the estimation of the dependent variable. Then, after calculating the MSE for each $y_{i,i=1,\cdots,5}$, the mean of the MSE for the 5 data sets ($y_1 \sim y_5$) is computed. For the 50 t-values, the same procedure was repeated, and graphs of the form $MSE = f(t)$ were plotted to obtain convex curves. The optimal penalty parameter chosen (t) corresponds to the MSE minimum. To avoid violating the choice of the penalty parameters obtained, 5-fold cross-validation was applied again with other randomly chosen sets, but the results led to similar t-values. In this case, the total number of optimization iterations is 50*5, i.e., 250 optimizations. After finding the best performing tuning parameter, LASSO minimization is applied, and the determination coefficient R^2 and the root mean square error (RMSE) are computed to compare the performances of LASSO and SWR. Thus, the best predictors were selected for further comparison with the results obtained for SWR.

[34] The SWR and LASSO techniques were compared using the following criteria: (1) the RMSE, which incorporates the variance and the square of the bias of the estimates:

$$RMSE = \sqrt{mean\left((y_i - \hat{y}_i)^2\right)} \tag{14}$$

and (2) the model explained variance (R^2), defined as

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y}_i)^2} \tag{15}$$

where $\bar{y}_i$ denotes the mean of the dependent variables.

4. Data Set for Case Study

[35] Figure 1 shows the area over southern Quebec, Canada where the studied stations are located. We worked with data sets issued from the following 9 stations located near the Gulf of St. Lawrence: Cedars, Drummondville, Seven Islands, Bagotville, Jean Lesage Intl., Sherbrooke, Maniwaki Airport, La Pocatière and Mont-joli. For the predictor selection in statistical downscaling, the following data were employed in the current study: 9 series of minimum and maximum temperatures issued from the daily meteorological data from Environment Canada stations, which were homogenized and rehabilitated by *Vincent et al.* [2002] as predictands (see Table 1 for stations), and a series of daily normalized predictors from the NCEP-NCAR reanalysis spread over 6 grid-cells of longitude from −67.5°W to −75°W and latitude from 46.39°N to 50.10°N (see Table 2 for the CGCM3 grid-cells). The NCEP/NCAR reanalysis data have a grid spacing of 2.5° latitude by 2.5° longitude [*Gachon et al.*, 2008].

[36] The predictor data set for the period of 1961 to 1990 was employed. (The data set issued from the NCEP/NCAR reanalysis was already standardized for the period of 1961–1990, except for the wind direction.) A total of 25 normalized predictors issued from the NCEP/NCAR reanalysis were used in this study aiming to select the most important

Table 1. Geographical Information of Environment Canada Stations

	Station Name	Latitude	Longitude	Elevation
1	Cedars	45.30	−74.05	47.00
2	Drummondville	45.88	−72.48	82.00
3	Seven Islands	50.22	−66.27	55.00
4	Bagotville	48.33	−71.00	158.00
5	Jean Lesage Intl	46.79	−71.38	74.00
6	Sherbrooke	45.43	−71.68	240.00
7	Maniwaki Airport	46.27	−75.99	200.00
8	La Pocatière	47.36	−70.03	31.00
9	Mont-joli	48.60	−68.22	52.00

predictors that can fit as well as the whole data set using two methods, SWR and LASSO; these variables are presented in Table 3. We note that the relative humidity is absent from the predictor list. Due to the high correlation between relative humidity and specific humidity, the relative humidity could be eliminated from the predictor data. However, it is difficult to ensure that these two variables are interchangeable. The predictors are issued from data collected every 6 hours and then standardized on a daily basis using the average (μ) and the standard deviation (σ) for the reference period of 1961–1990 and at the end, the predictors are interpolated linearly on the CGCM3 grid-cells [*Gachon et al.*, 2008]. The CGCM3 data correspond to the third version of the Canadian Center for Climate Modeling and Analysis (CCCma)-coupled Canadian global climate model. The atmospheric component of the CGCM3 has 31 vertical levels and a horizontal resolution of approximately 3.75° latitude and longitude (approximately 400 km).

5. Results

[37] For all stations, Tables 4, 5, 6 and 7 show the results of the predictor selection obtained by SWR and LASSO for the minimum and maximum temperatures from 1961 to 1990. They present the most important predictors selected that can be used in further downscaling procedures. The predictor selection can be described by a subjective judgment depending on the analyzer. For all of the stations and for both the minimum and maximum temperature, the mean sea level pressure, the geopotential at 850 hPa, the geopotential at 500 hPa, the specific humidity at 850 hPa and the temperature at 2 m can be considered to be the most dominant variables. This seems plausible because these parameters are strongly associated with significant modifications to the temperature characteristics in the boundary layer [see *Hessami et al.*, 2008]. The selected predictors represent, at some level of confidence, almost all of the information provided by the whole data set.

Table 2. Longitude and Latitude of the CGCM3 Grid Cells

Box Number	Longitude	Latitude
77X, 11Y	−75.00	50.10
77X, 12Y	−75.00	46.39
78X, 11Y	−71.25	50.10
78X, 12Y	−71.25	46.39
79X, 11Y	−67.50	50.10
79X, 12Y	−67.50	46.39

Table 3. NCEP/NCAR Predictor Variables on CGCM3 Grid

Number	Predictor
1	Mean sea level pressure
2	Surface airflow strength
3	Surface zonal velocity
4	Surface meridional velocity
5	Surface vorticity
6	Surface wind direction
7	Surface divergence
8	500 hPa airflow strength
9	500 hPa zonal velocity
10	500 hPa meridional velocity
11	500 hPa vorticity
12	500 hPa geopotential
13	500 hPa wind direction
14	500 hPa divergence
15	850 hPa airflow strength
16	850 hPa zonal velocity
17	850 hPa meridional velocity
18	850 hPa vorticity
19	850 hPa geopotential
20	850 hPa wind direction
21	850 hPa divergence
22	500 hPa specific Humidity
23	850 hPa specific Humidity
24	Near surface specific Humidity
25	Temperature at 2 m

[38] Thus, depending on the location of a station relative to the Gulf of St. Lawrence, the selection of the most influential predictors can vary slightly. In fact, the selection of some meteorological variables at some stations depends on the location of the latter relative to the Gulf of St. Lawrence, such as the specific humidity, which represents the amount of water vapor in the air, defined as the ratio of water vapor to dry air at a particular mass. Based on the presence of water in a region, the specific humidity is highly related to temperature variations. In addition to the specific humidity, the temperature at 2 m may be affected by the presence of different air masses influencing temperature variations.

[39] For the 9 stations presented in this work, the mean sea level pressure appears as a common selected predictor for the maximum temperature for both methods. It can be considered to be the most effective predictor, which regroups almost all of the predictors' information needed for downscaling; this is somewhat expected because of its great influence on local climate.

[40] Furthermore, the selections by SWR and LASSO included the same predictors for the La Pocatière and

Table 4. Results of the Most Important Predictors Selected for the Maximum Temperature by SWR

Station	Predictors[a]
Cedars	1, 5, 19, 23, 24
Drummondville	1, 19, 23, 24, 25
Seven Islands	1, 19, 23, 24, 25
Bagotville	1, 19, 23, 24, 25
Jean Lesage Intl	1, 19, 23, 24, 25
Sherbrooke	1, 19, 23, 24, 25
Maniwaki Airport	1, 19, 23, 24, 25
La Pocatière	1, 12, 19, 23, 25
Mont-joli	1, 16, 19, 24, 25

[a]For each predictor, the number refers to the atmospheric variable defined in Table 3.

Table 5. Results of the Most Important Predictors Selected for the Maximum Temperature by LASSO[a]

Station	Predictors
Cedars	1, 19, 23, 24, 25
Drummondville	1, 12, 19, 23, 25
Seven Islands	1, 3, 19, 24, 25
Bagotville	1, 12, 19, 23, 25
Jean Lesage Intl	1, 12, 19, 23, 25
Sherbrooke	1, 12, 19, 24, 25
Maniwaki Airport	1, 19, 23, 24, 25
La Pocatière	1, 12, 19, 23, 25
Mont-joli	1, 19, 23, 24, 25

[a]For each predictor, the number refers to the atmospheric variable defined in Table 3.

Maniwaki Airport stations. Otherwise, there is a small difference in the predictor selection for the remaining stations between SWR (Table 4) and LASSO (Table 5). Despite this difference, both methods give similar combinations of selected variables, but LASSO has the advantage of its automatic aspect of selection. In addition, LASSO did not frequently select predictors 23 and 24 (see Table 3) at the same time as the most effective predictors. Indeed, the 850-hPa specific humidity and the near-surface specific humidity are highly correlated, and one of the specificities of LASSO is that it is not affected by correlations between predictors, which contributes to its robustness compared to SWR and may improve the selection quality.

[41] According to the results for the minimum temperature, there is a slight difference between the predictors selected by SWR and those selected by LASSO. The predictors selected by SWR and LASSO are the same for only the Bagotville station: the mean sea level pressure, the geopotential at 850 hPa, the specific humidity at 850 hPa, the near-surface specific humidity and the temperature at 2 m. Otherwise, the predictors selected by SWR randomly differ between stations due to the subjective aspect of the selection process, which depends on the analyzer. Meanwhile, the LASSO selection gives more accurate and interpretable models because the most important predictors chosen for all of the stations are nearly the same (see Tables 6 and 7). The most important predictors selected by LASSO for the minimum temperature are the mean sea level pressure, the geopotential at 850 hPa, the near-surface specific humidity and the temperature at 2 m, which is consistent with the predictor combinations found for the maximum temperature. The results found for the minimum temperature demonstrate the strength of the LASSO technique in dealing with correlations between predictors and in eliminating redundancy.

Table 6. Results of the Most Important Predictors Selected for the Minimum Temperature by SWR[a]

Station	Predictors
Cedars	1, 5, 23, 24, 25
Drummondville	5, 12, 23, 24, 25
Seven Islands	1, 12, 19, 24, 25
Bagotville	1, 19, 23, 24, 25
Jean Lesage Intl	4, 12, 23, 24, 25
Sherbrooke	1, 4, 19, 24, 25
Maniwaki Airport	1, 19, 23, 24, 25
La Pocatière	1, 7, 21, 24, 25
Mont-joli	1, 12, 19, 24, 25

[a]For each predictor, the number refers to the atmospheric variable defined in Table 3.

Table 7. Results of the Most Important Predictors Selected for the Minimum Temperature by LASSO[a]

Station	Predictors
Cedars	1, 4, 19, 24, 25
Drummondville	1, 7, 21, 24, 25
Seven Islands	1, 19, 21, 24, 25
Bagotville	1, 19, 23, 24, 25
Jean Lesage Intl	1, 19, 23, 24, 25
Sherbrooke	1, 19, 23, 24, 25
Maniwaki Airport	1, 12, 19, 24, 25
La Pocatière	1, 7, 19, 21, 24
Mont-joli	1, 5, 19, 24, 25

[a]For each predictor, the number refers to the atmospheric variable defined in Table 3.

[42] The differences between the LASSO and SWR results can be explained by the improved selection achieved by LASSO compared to other methods [*Grandvalet and Canu*, 1999]; this improved selection arises from our use of a large data set. Thus, LASSO is considered to be a method with enormous potential for extensions and modifications.

[43] To compare the performances of SWR and LASSO, a risk function was used (RMSE), incorporating the variance and the square of the estimate bias as well as the explained variance (R^2). Note that lower RMSE values and higher R^2 values imply better performance.

[44] For all stations over all months, LASSO has lower RMSE and higher R^2 values compared to SWR, as shown in Tables 8 and 9 for the maximum temperature and Tables 10 and 11 for the minimum temperature. Table 8 summarizes the RMSE results for LASSO and SWR for the maximum temperature at all stations. The RMSE of the maximum temperature varies from 2.01 (Cedars in July) to 4.39 (Bagotville in January) for SWR and from 1.95 (Cedars in July) to 4.25 (Bagotville in January) for LASSO. Table 8 indicates that LASSO performs better in terms of the RMSE for all stations and throughout all months. Figure 2 presents the RMSE for SWR and LASSO corresponding to the maximum temperature at the Bagotville and the La Pocatière stations, showing that the error found with SWR is always higher than the one corresponding to LASSO. The improved RMSE obtained by LASSO is quite clear for the La Pocatière station, showing that the error achieved by LASSO is consistently lower than that obtained with SWR.

[45] Table 9 presents the results of the explained variance (R^2) for LASSO and SWR for the maximum temperature at all stations. The R^2 value of the maximum temperature varies from 0.39 (Seven Islands in July) to 0.75 (Maniwaki Airport in March) for SWR and from 0.4 (Seven Islands in July) to 0.76 (Maniwaki Airport in March) for LASSO. For all stations and throughout all months, LASSO performs better than stepwise regression in terms of R^2. The improvement achieved by LASSO is clearly shown in Figure 3, which presents R^2 for both LASSO and SWR for the maximum temperature at the Cedars and Jean Lesage stations; the R^2 value for LASSO is always higher than the one corresponding to SWR. Figure 4 shows a comparison between LASSO and SWR in terms of R^2 for the maximum temperature at the Bagotville station, emphasizing the improvement obtained by LASSO.

Table 8. Results of the RMSE for LASSO and SWR for Maximum Temperature[a]

		RMSE											
	Method	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Cedars	LASSO	**4.01**	**3.51**	**3.21**	**3.02**	**2.89**	**2.34**	**1.96**	**2.10**	**2.53**	**2.71**	**3.00**	**3.64**
	Stepwise	4.14	3.58	3.27	3.06	2.95	2.37	2.01	2.12	2.56	2.77	3.07	3.69
Drummondville	LASSO	**4.18**	**3.52**	**3.36**	**3.19**	**3.12**	**2.70**	**2.33**	**2.31**	**2.81**	**2.84**	**2.99**	**3.63**
	Stepwise	4.22	3.58	3.40	3.25	3.16	2.74	2.38	2.34	2.81	2.88	3.02	3.67
Seven Islands	LASSO	**3.60**	**3.32**	**2.57**	**2.34**	**2.87**	**2.94**	**2.61**	**2.39**	**2.42**	**2.29**	**2.25**	**3.35**
	Stepwise	3.71	3.37	2.60	2.39	2.95	2.98	2.64	2.40	2.48	2.34	2.28	3.40
Bagotville	LASSO	**4.25**	**3.75**	**3.59**	**3.19**	**3.39**	**3.11**	**2.85**	**2.71**	**2.99**	**3.09**	**2.79**	**3.87**
	Stepwise	4.39	3.82	3.64	3.25	3.45	3.19	2.90	2.76	3.04	3.13	2.88	3.95
Jean Lesage Intl	LASSO	**3.86**	**3.19**	**2.78**	**3.06**	**3.26**	**2.93**	**2.43**	**2.34**	**2.56**	**2.66**	**2.62**	**3.29**
	Stepwise	3.74	3.24	2.85	3.12	3.35	3.00	2.48	2.40	2.61	2.69	2.66	3.36
Sherbrooke	LASSO	**3.96**	**3.59**	**3.17**	**3.18**	**3.13**	**2.56**	**2.08**	**2.26**	**2.68**	**2.99**	**3.17**	**3.67**
	Stepwise	4.03	3.64	3.20	3.22	3.19	2.59	2.10	2.28	2.70	3.06	3.26	3.72
Maniwaki Airport	LASSO	**3.78**	**3.27**	**2.98**	**3.15**	**3.03**	**2.58**	**2.16**	**2.38**	**2.53**	**2.74**	**2.79**	**3.45**
	Stepwise	3.82	3.34	3.01	3.24	3.13	2.62	2.21	2.31	2.64	2.78	2.84	3.53
La Pocatière	LASSO	**3.67**	**3.12**	**2.97**	**2.91**	**3.39**	**3.09**	**2.64**	**2.53**	**2.81**	**2.72**	**2.59**	**3.38**
	Stepwise	3.72	3.17	3.04	2.96	3.47	3.15	2.69	2.58	2.88	2.79	2.64	3.47
Mont-joli	LASSO	**3.73**	**3.47**	**2.99**	**2.72**	**3.34**	**3.06**	**2.58**	**2.42**	**2.78**	**2.87**	**2.62**	**3.30**
	Stepwise	3.79	3.51	3.03	2.76	3.43	3.11	2.61	2.46	2.83	2.90	2.69	3.37

[a]Bold means better result with LASSO.

[46] For the minimum temperature, the same trends were observed as for the maximum temperature. Table 10 presents the RMSE results for LASSO and SWR for the minimum temperature. The RMSE varies from 1.8 (Seven Islands in July) to 5.67 (Sherbrook in January) for SWR and from 1.79 (Seven Islands in July) to 5.58 (Sherbrook in January) for LASSO. Table 10 indicates that LASSO performs better in terms of the RMSE for all stations and throughout all months. In addition, Table 11 presents the R^2 results for the minimum temperature for LASSO and SWR. The R^2 values vary from 0.48 (Seven Islands in July) to 0.72 (Seven Islands in March) for SWR and from 0.49 (Seven Islands in July) to 0.73 (Seven Islands in March) for LASSO. The improvement achieved by LASSO is clearly shown in Figure 5, which presents R^2 for both LASSO and SWR for the minimum temperature at the Bagotville and the Maniwaki Airport stations; the R^2 values obtained by LASSO are higher than those found with SWR which emphasizes the improvement in the selection achieved by LASSO in terms of R2. Thus, LASSO performed well in all cases and for all stations.

6. Discussion and Conclusions

[47] Despite the positive features of the SWR method, the LASSO technique performed better with the data set employed herein for selecting predictors for downscaling of the maximum and minimum temperature data issued from 9 stations located in eastern Canada near the Gulf of St. Lawrence. Some limitations of SWR are overcome by LASSO. SWR is based only on correlations and uses only one model throughout treatment of the whole data set. Hence, if the model does not perform well, the selection may not be optimal. Furthermore, with SWR, if a variable has already been eliminated, it cannot be reintroduced to the model, even if it becomes significant. SWR is considered as a highly

Table 9. Results of R^2 for LASSO and SWR for Maximum Temperature[a]

		R2											
	Method	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Cedars	LASSO	**0.69**	**0.71**	**0.7**	**0.74**	**0.7**	**0.67**	**0.63**	**0.65**	**0.67**	**0.72**	**0.67**	**0.68**
	Stepwise	0.67	0.7	0.69	0.73	0.69	0.67	0.61	0.65	0.67	0.71	0.66	0.68
Drummondville	LASSO	**0.68**	**0.72**	**0.7**	**0.73**	**0.71**	**0.65**	**0.6**	**0.66**	**0.66**	**0.72**	**0.72**	**0.7**
	Stepwise	0.68	0.72	0.7	0.72	0.7	0.64	0.59	0.65	0.66	0.71	0.71	0.7
Seven Islands	LASSO	**0.68**	**0.69**	**0.74**	**0.6**	**0.53**	**0.51**	**0.4**	**0.47**	**0.52**	**0.59**	**0.71**	**0.72**
	Stepwise	0.67	0.68	0.73	0.59	0.5	0.5	0.39	0.46	0.5	0.57	0.7	0.72
Bagotville	LASSO	**0.66**	**0.7**	**0.7**	**0.7**	**0.7**	**0.63**	**0.58**	**0.63**	**0.65**	**0.67**	**0.71**	**0.68**
	Stepwise	0.64	0.69	0.69	0.69	0.69	0.61	0.57	0.62	0.64	0.66	0.7	0.67
Jean Lesage Intl	LASSO	**0.68**	**0.71**	**0.71**	**0.67**	**0.68**	**0.6**	**0.57**	**0.62**	**0.66**	**0.7**	**0.7**	**0.69**
	Stepwise	0.68	0.71	0.7	0.66	0.66	0.58	0.56	0.6	0.65	0.69	0.69	0.68
Sherbrooke	LASSO	**0.69**	**0.72**	**0.75**	**0.74**	**0.71**	**0.68**	**0.67**	**0.67**	**0.68**	**0.72**	**0.71**	**0.69**
	Stepwise	0.69	0.72	0.75	0.74	0.7	0.67	0.66	0.66	0.68	0.71	0.7	0.69
Maniwaki Airport	LASSO	**0.7**	**0.73**	**0.76**	**0.76**	**0.75**	**0.68**	**0.67**	**0.64**	**0.73**	**0.76**	**0.73**	**0.72**
	Stepwise	0.7	0.72	0.75	0.75	0.73	0.67	0.65	0.66	0.7	0.75	0.73	0.71
La Pocatière	LASSO	**0.69**	**0.72**	**0.68**	**0.66**	**0.63**	**0.59**	**0.56**	**0.59**	**0.6**	**0.68**	**0.71**	**0.68**
	Stepwise	0.68	0.72	0.67	0.65	0.61	0.58	0.55	0.58	0.59	0.67	0.7	0.67
Mont-joli	LASSO	**0.66**	**0.68**	**0.69**	**0.67**	**0.63**	**0.62**	**0.61**	**0.65**	**0.63**	**0.66**	**0.73**	**0.69**
	Stepwise	0.65	0.68	0.69	0.66	0.61	0.61	0.6	0.64	0.62	0.66	0.71	0.68

[a]Bold means better result with LASSO.

Table 10. Results of RMSE for LASSO and SWR for Minimum Temperature[a]

		RMSE											
	Method	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Cedars	LASSO	**4.50**	**4.48**	**3.80**	**2.40**	**2.18**	**2.15**	**1.92**	**1.93**	**2.45**	**2.55**	**2.75**	**3.98**
	Stepwise	4.56	4.54	3.99	2.49	2.21	2.20	1.96	1.98	2.49	2.65	2.83	4.12
Drummondville	LASSO	**5.10**	**4.67**	**4.09**	**3.29**	**2.71**	**2.41**	**2.26**	**2.33**	**2.56**	**2.80**	**3.05**	**4.45**
	Stepwise	5.20	4.77	4.25	2.75	2.72	2.43	2.30	2.36	2.61	2.84	3.07	4.56
Seven Islands	LASSO	**4.10**	**4.05**	**3.74**	**2.43**	**1.85**	**2.00**	**1.79**	**1.94**	**2.25**	**2.55**	**3.10**	**4.00**
	Stepwise	4.15	4.13	3.82	2.48	1.88	2.03	1.80	1.96	2.27	2.57	3.15	4.12
Bagotville	LASSO	**4.89**	**4.78**	**4.40**	**2.81**	**2.45**	**2.53**	**2.21**	**2.32**	**2.58**	**2.60**	**3.50**	**4.72**
	Stepwise	4.99	4.94	4.48	2.86	2.50	2.59	2.26	2.36	2.68	2.65	3.59	4.80
Jean Lesage Intl	LASSO	**4.24**	**4.03**	**3.65**	**2.28**	**2.13**	**2.19**	**2.00**	**1.99**	**2.29**	**2.30**	**3.00**	**4.07**
	Stepwise	4.36	4.12	3.68	2.29	2.19	2.20	2.02	2.01	2.37	2.34	3.03	4.22
Sherbrooke	LASSO	**5.58**	**5.21**	**4.49**	**2.80**	**2.80**	**2.67**	**2.52**	**2.42**	**2.95**	**2.85**	**3.47**	**4.81**
	Stepwise	5.67	5.37	4.54	2.82	2.88	2.67	2.56	2.48	2.97	2.87	3.51	4.90
Maniwaki Airport	LASSO	**5.05**	**5.39**	**4.81**	**3.09**	**2.58**	**2.54**	**2.34**	**2.25**	**2.58**	**2.84**	**3.56**	**4.83**
	Stepwise	5.17	5.54	4.91	3.14	2.64	2.61	2.35	2.29	2.64	2.90	3.59	4.90
La Pocatière	LASSO	**4.20**	**4.14**	**3.98**	**2.58**	**2.75**	**3.07**	**2.61**	**2.61**	**2.90**	**2.72**	**2.89**	**4.10**
	Stepwise	4.31	4.21	4.07	2.66	2.79	3.11	2.65	2.68	2.95	2.77	2.96	4.18
Mont-joli	LASSO	**3.98**	**3.76**	**3.31**	**2.32**	**2.30**	**2.48**	**2.14**	**2.21**	**2.36**	**2.35**	**2.59**	**3.55**
	Stepwise	4.06	3.92	3.35	2.37	2.37	2.53	2.18	2.21	2.37	2.36	2.61	3.62

[a]Bold means better result with LASSO.

instable method because the selection can vary strongly if the data are changed even slightly.

[48] The LASSO technique combines the positive features of subset selection and ridge regression using stable algorithms as ridge regression and by shrinking some coefficients and setting others to zero as the subset selection. Lasso provides easily interpretable models and improves the prediction accuracy.

[49] Moreover, this technique works well with large data sets, mainly when p ≫ n (the number of predictors is much higher than the predict and number). There are numerous reports of extensions and modifications of this method, which explains the existence of more than 8 formulations for the LASSO technique [see *Schmidt*, 2005]. The usefulness of the LASSO method depends on the choice of the tuning parameter as the appropriate choice of t will allow to avoid "over fitting" or "under fitting" of LASSO and successful development of the statistical theory.

[50] Two different types of methods were presented in the current study for selecting predictors to compare their performances. Both methods appear to be appropriate to select the smallest number of predictors that can fit the data as well as if we had used the whole data set. Due to its sparseness and its computational advantages, LASSO presented a better alternative. It is an automated method that is unaffected by collinearity problems, such as correlations between predictors in the regression model, unlike SWR, for which collinearity problems are exacerbated. Thus, LASSO achieved lower errors and higher R^2 values. SWR behaved well in this case, but its results are strongly dependent on the data set used, while LASSO can be considered as a more stable

Table 11. Results of R^2 for LASSO and SWR for Minimum Temperature[a]

		RMSE											
	Method	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Cedars	LASSO	**0.62**	**0.62**	**0.64**	**0.68**	**0.71**	**0.64**	**0.65**	**0.70**	**0.67**	**0.66**	**0.66**	**0.67**
	Stepwise	0.61	0.62	0.61	0.66	0.70	0.62	0.64	0.68	0.66	0.64	0.64	0.65
Drummondville	LASSO	**0.60**	**0.64**	**0.64**	**0.48**	**0.66**	**0.65**	**0.60**	**0.67**	**0.68**	**0.64**	**0.66**	**0.65**
	Stepwise	0.59	0.63	0.62	0.63	0.66	0.65	0.59	0.66	0.67	0.63	0.66	0.63
Seven Islands	LASSO	**0.71**	**0.71**	**0.73**	**0.69**	**0.56**	**0.56**	**0.49**	**0.58**	**0.61**	**0.57**	**0.69**	**0.73**
	Stepwise	0.71	0.71	0.73	0.68	0.55	0.56	0.48	0.58	0.61	0.57	0.69	0.72
Bagotville	LASSO	**0.62**	**0.64**	**0.67**	**0.64**	**0.65**	**0.58**	**0.57**	**0.64**	**0.64**	**0.59**	**0.64**	**0.68**
	Stepwise	0.61	0.62	0.66	0.64	0.64	0.56	0.55	0.63	0.61	0.58	0.62	0.67
Jean Lesage Intl	LASSO	**0.66**	**0.69**	**0.69**	**0.68**	**0.68**	**0.61**	**0.62**	**0.69**	**0.70**	**0.67**	**0.67**	**0.69**
	Stepwise	0.64	0.68	0.69	0.68	0.67	0.61	0.62	0.69	0.68	0.66	0.66	0.67
Sherbrooke	LASSO	**0.61**	**0.63**	**0.64**	**0.63**	**0.66**	**0.62**	**0.59**	**0.67**	**0.65**	**0.64**	**0.62**	**0.65**
	Stepwise	0.60	0.60	0.64	0.63	0.65	0.62	0.58	0.66	0.64	0.64	0.61	0.64
Maniwaki Airport	LASSO	**0.69**	**0.65**	**0.66**	**0.63**	**0.70**	**0.65**	**0.63**	**0.70**	**0.70**	**0.63**	**0.62**	**0.70**
	Stepwise	0.67	0.63	0.65	0.62	0.69	0.63	0.63	0.69	0.69	0.61	0.62	0.69
La Pocatière	LASSO	**0.60**	**0.62**	**0.61**	**0.56**	**0.53**	**0.50**	**0.54**	**0.58**	**0.55**	**0.57**	**0.63**	**0.64**
	Stepwise	0.58	0.61	0.60	0.53	0.52	0.49	0.53	0.56	0.53	0.56	0.62	0.63
Mont-joli	LASSO	**0.60**	**0.65**	**0.69**	**0.61**	**0.54**	**0.55**	**0.51**	**0.57**	**0.56**	**0.60**	**0.67**	**0.68**
	Stepwise	0.59	0.62	0.69	0.60	0.51	0.54	0.50	0.57	0.56	0.60	0.67	0.67

[a]Bold means better result with LASSO.

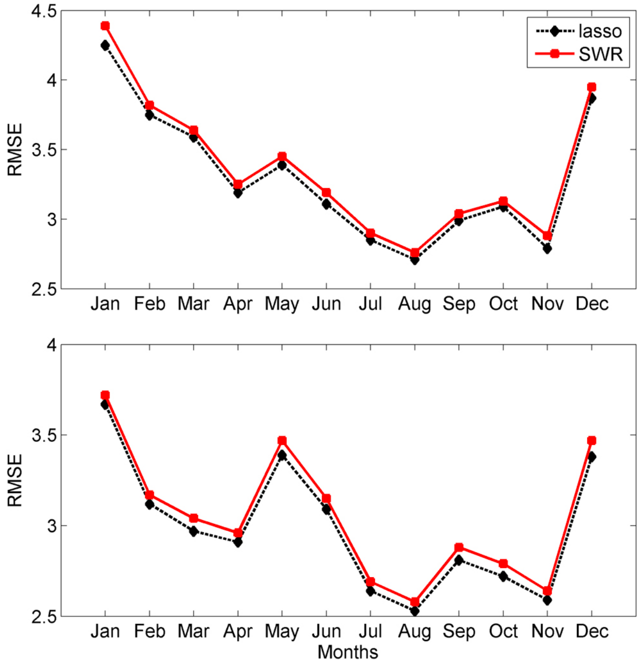

Figure 2. RMSE for LASSO and SWR represented for (top) Bagotville and (bottom) La Pocatière stations for maximum temperature.

method that provides accurate predictions and easily interpretable models.

[51] Both methods obtained good results, but more confidence can be placed in LASSO. Many researchers have described the drawbacks of stepwise regression; for example, the R^2 values have a high bias, the Fisher and χ^2 test statistics do not have the claimed distribution, and the standard errors of the parameter estimates are low, causing the confidence intervals around the parameter estimates to be too narrow. Hence, LASSO presented a better alternative for predictor selection: it can be implemented in statistical downscaling models (SDSMs) that use SWR for predictor selection, so LASSO may improve the accuracy of model outputs. The limitations of LASSO include the difficulty encountered in choosing the regularization parameter, which defines the shrinkage rate as well as the set of some coefficients to zero. This all depends on the development of the statistical theory.

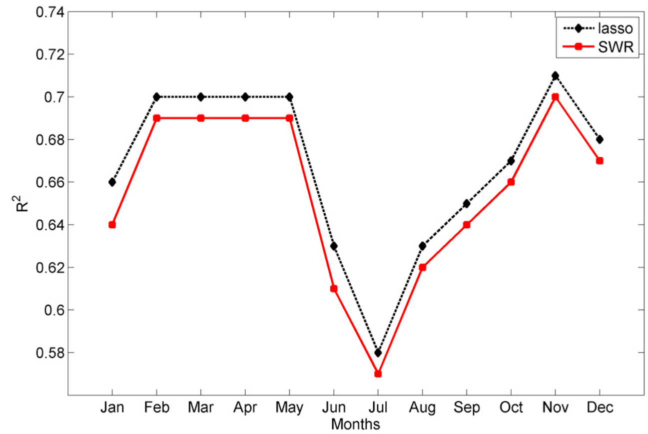

Figure 4. R^2 for LASSO and SWR represented for Bagotville station for maximum temperature.

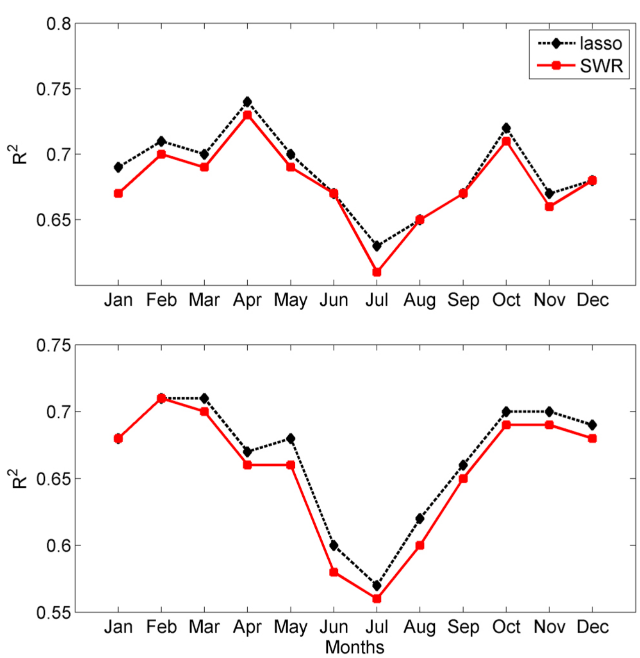

Figure 3. R^2 for LASSO and SWR represented for (top) Cedars and (bottom) Jean Lesage stations for maximum temperature.

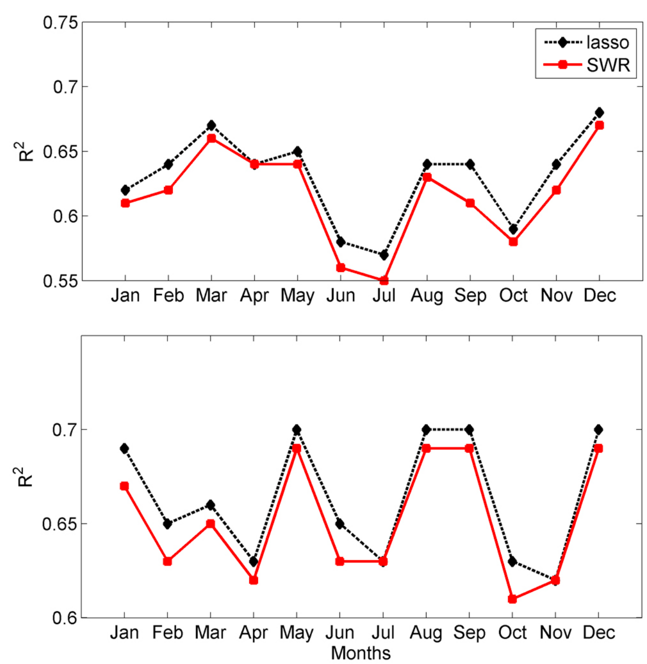

Figure 5. R^2 for (top) LASSO and (bottom) SWR represented for Bagotville and Maniwaki Airport station for minimum temperature.

Further work may be directed towards implementing this technique in statistical downscaling models and applying LASSO to other hydrological variables, such as precipitation.

Notation

OLS	Ordinary Least Squares method.
AIC	Akaike Information Criterion.
BIC	The Bayesian Information Criterion.
GCM	Global Climate Model
SWR	Stepwise Regression
LASSO	Least Absolute Shrinkage and Selection Operator
FS	Forward Selection
BE	Backward Elimination
R^2	Determination coefficient
LARS	Least Angle Regression Selection
NCEP	National Center for Environmental Prediction
NCAR	National Center for Atmospheric Research
MSE	Mean Square Errors
RMSE	Root Mean Square Errors
SDSM	Statistical Downscaling Model

[52] **Acknowledgments.** The authors acknowledge the funding from the National Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chair Program, and the Ministère du Développement Économique Innovation et Exportation du Québec. The authors acknowledge also the help provided by Mark Schmidt from l' École Normale Supérieure of Paris and Environment Canada. The second author also acknowledges that he was supported for this work by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MEST) (2012-0434).

References

Breiman, L. (1995), Better subset selection using the non-negative garotte, *Technometrics*, *37*(4), 738–754.

Cohen, J., and P. Cohen (1975), Analytic strategies: Simultaneous, hierarchical, and stepwise regression, report, 18pp., Dep. of Sociol., Univ. of Notre Dame, Notre Dame, Ind. [Available at http://www.nd.edu/~rwilliam/stats1/x95.pdf.]

Copas, J. B. (1983), Regression, prediction and shrinkage, *J. R. Stat. Soc., Ser. A*, *45*(3), 311–354.

Draper, N. R., and H. Smith (1966), *Applied Regression Analysis*, John Wiley, Hoboken, N. J.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani (2002), Least angle regression, technical report, Stanford Univ., Stanford, Calif.

Efroymson, M. A. (1966), Stepwise regression—A backward and forward look, paper presented at Eastern Regional Meeting, Inst. of Math. Stat., Florham Park, N. J.

Fan, J., and R. Li (2001), Variable selection via nonconcave penalized likelihood and its oracle properties, *J. Am. Stat. Assoc.*, *96*(456), 1348–1360, doi:10.1198/016214501753382273.

Flom, P. L., and D. L. Cassell (2007), Stopping stepwise: Why stepwise and similar selection methods are bad, and what you should use, paper presented at NESUG 2007, NorthEast SAS User Group, Baltimore, Md.

Frank, I. E., and J. H. Friedman (1993), A statistical view of some chemometrics regression tools, *Technometrics*, *35*, 109–135, doi:10.1080/00401706.1993.10485033.

Fu, W. J. (1998), The Bridge versus the LASSO, *J. Comput. Graph. Stat.*, *7*(3), 397–416.

Gachon, P., M. Radojevic, and A. Harding (2008), DAI MCCG3 Prédicteurs: Ensembles de Données de Prédicteurs issus de la Réanalyse du NCEP/NCAR et du MCCG3.1 T47, version 1.0, report, 17 pp., Environ. Canada, Montréal, Quebec, Canada.

Grandvalet, I. (1998), Least absolute shrinkage is equivalent to quadratic penalization, in *Perspectives in Neural Computing*, pp. 201–206, Springer, Berlin.

Grandvalet, I., and S. Canu (1999), Outcomes of the equivalence of adaptive ridge with least absolute shrinkage, in *Proceedings of the 1998 Conference on Advances in Neural Information Processing Systems II*, pp. 445–451, MIT Press, Cambridge, Mass.

Hessami, M., P. Gachon, T. B. M. J. Ouarda, and A. St-Hilaire (2008), Automated regression-based statistical downscaling tool, *Environ. Modell. Software*, *23*, 813–834, doi:10.1016/j.envsoft.2007.10.004.

Hocking, R. R. (1976), A Biometrics invited paper: The analysis and selection of variables in linear regression, *Biometrics*, *32*(1), 1–49, doi:10.2307/2529336.

Hoerl, E. A., and W. R. Kennard (1970), Ridge regression: Applications to nonorthogonal problems, *Technometrics*, *12*(1), 69–82.

Jennrich, R. I., and P. F. Sampson (1968), Application of stepwise regression to non-linear estimation, *Technometrics*, *10*(1), 63–72, doi:10.1080/00401706.1968.10490535.

Kyung, M., J. Gill, M. Ghosh, and G. Casella (2010), Penalized regression, standard errors, and Bayesian LASSO, *Bayesian Anal.*, *5*(2), 369–412.

Lund, I. A. (1971), An application of stagewise and stepwise regression procedures to a problem of estimating precipitation in California, *J. Appl. Meteorol.*, *10*, 892–902, doi:10.1175/1520-0450(1971)010<0892:AAOSAS>2.0.CO;2.

Osborne, M. R., B. Presnell, and B. A. Turlach (2000a), A new approach to variable selection in least squares problems, *IMA J. Numer. Anal.*, *20*, 389–403, doi:10.1093/imanum/20.3.389.

Osborne, M. R., B. Presnell, and B. A. Turlach (2000b), On the LASSO and its dual, *J. Comput. Graph. Stat.*, *9*(2), 319–337.

Perkins, S., K. Lacker, and J. Theiler (2003), Grafting: Fast, incremental feature selection by gradient descent in function space, *J. Mach. Learn. Res.*, *3*, 1333–1356.

Schmidt, M. (2005), Least squares optimization with L1-norm regularization, *CS542B Project Rep.*, Dép. d'Informatique École Normale Supérieure, Paris. [Available at http://www.di.ens.fr/~mschmidt/Software/lasso.pdf.]

Tibshirani, R. (1996), Regression shrinkage and selection via the LASSO, *J. R. Stat. Soc., Ser. B*, *58*(1), 267–288.

Vincent, L., X. Zhang, B. Bonsal, and W. Hogg (2002), Homogenization of daily temperatures over Canada, *J. Clim.*, *15*, 1322–1334, doi:10.1175/1520-0442(2002)015<1322:HODTOC>2.0.CO;2.

Weisberg, S. (2011), Variable selection and regularization (revised), report, U. of Minn.-Twin Cities, Minneapolis. [Available at http://users.stat.umn.edu/~sandy/courses/8053/handouts/lasso.pdf.]