



HAL
open science

Comparison of data-driven methods for downscaling ensemble weather forecasts

Xiaoli Liu, P. Coulibaly, N. Evora

► **To cite this version:**

Xiaoli Liu, P. Coulibaly, N. Evora. Comparison of data-driven methods for downscaling ensemble weather forecasts. *Hydrology and Earth System Sciences Discussions*, 2008, 12 (2), pp.615-624. hal-00305159

HAL Id: hal-00305159

<https://hal.science/hal-00305159>

Submitted on 18 Jun 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Comparison of data-driven methods for downscaling ensemble weather forecasts

Xiaoli Liu¹, P. Coulibaly¹, and N. Evora²

¹Department of Civil Engineering, School of Geography and Earth Sciences, McMaster University, 1280 Main Street West, Hamilton, Ontario, L8S 4L7, Canada

²Institut de Recherche d'Hydro-Quebec (IREQ), Varennes (Quebec) J3X 1S1, Canada

Received: 28 November 2006 – Published in Hydrol. Earth Syst. Sci. Discuss.: 1 February 2007

Revised: 7 December 2007 – Accepted: 27 February 2008 – Published: 20 March 2008

Abstract. This study investigates dynamically different data-driven methods, specifically a statistical downscaling model (SDSM), a time lagged feedforward neural network (TLFN), and an evolutionary polynomial regression (EPR) technique for downscaling numerical weather ensemble forecasts generated by a medium range forecast (MRF) model. Given the coarse resolution (about 200-km grid spacing) of the MRF model, an optimal use of the weather forecasts at the local or watershed scale, requires appropriate downscaling techniques. The selected methods are applied for downscaling ensemble daily precipitation and temperature series for the Chute-du-Diable basin located in northeastern Canada. The downscaling results show that the TLFN and EPR have similar performance in downscaling ensemble daily precipitation as well as daily maximum and minimum temperature series whatever the season. Both the TLFN and EPR are more efficient downscaling techniques than SDSM for both the ensemble daily precipitation and temperature.

1 Introduction

Downscaling methods were initially developed and used for global climate models (GCMs) outputs. In this study our goal is to develop data-driven methods for downscaling ensemble weather forecast data provided by the National Centers for Environmental Prediction (NCEP) medium range forecast (MRF) modeling system. It is well-known that these large scale numerical models are generally not accurate at modeling local climate, because they are unable to represent local sub-grid scale features and dynamics. However, in operational hydrology, hydrological models are usually used to simulate sub-grid scale phenomenon and therefore require input data (such as precipitation and temperature) at similar

sub-grid scale. For instance, precipitation scenarios at such finer temporal and spatial resolution are needed in order to improve the design and evaluate the future performance of urban drainage systems (Bronstert et al., 2002). In practical hydrologic applications, there is a need to convert the MRF forecasts into high resolution information useful at the local or watershed scale

There are various downscaling techniques available to convert coarse resolution climate model outputs into daily meteorological variables appropriate for hydrologic applications. The most widely used statistical downscaling models usually implement linear methods such as, multiple linear regression, Canonical correlation analysis or singular value decomposition (Conway et al., 1996). However, it is not yet clear which statistical downscaling method provides the most reliable estimates of daily rainfall and temperature series. Nevertheless, the interest in non-linear regression methods, namely, artificial neural networks (ANNs), is nowadays increasing because of their high potential for complex, non-linear and time-varying input–output mapping. Although the weights of ANN are similar to non-linear regression coefficients, the unique structure of the network and the non-linear transfer function associated with each hidden and output nodes allows ANNs to approximate highly non-linear relationships without a priori assumption. Moreover, while other regression techniques assume a functional form, ANNs allow the data to define the functional form. Therefore, ANNs are generally believed to be more powerful than the other regression-based downscaling techniques (von Storch et al., 2000). Genetic Programming (GP) is another well known data-driven technique that has shown promising potential for the downscaling of daily extreme temperatures (Coulibaly 2004). However, no study has fully investigated and compared the selected data-driven methods for downscaling ensemble weather forecasts.

Correspondence to: P. Coulibaly
(couliba@mcmaster.ca)

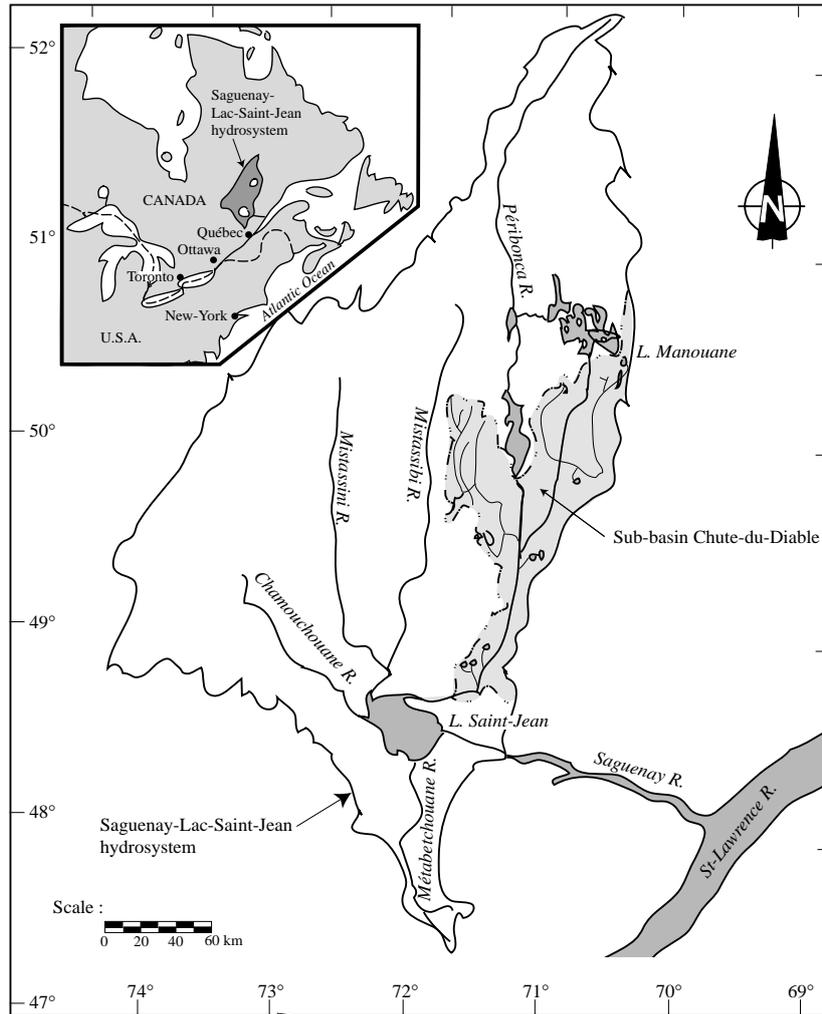


Fig. 1. Location of the study area in northern Quebec (Canada).

The purpose of this study is to identify optimal models that can capture the complex relationship between selected large-scale predictors and locally observed meteorological variables (or predictands) using three different methods, i.e. linear regression method, ANNs, and GP, so as to compare the performances of the three methods in downscaling daily precipitation and temperature. The paper specifically focuses on the time lagged feedforward neural networks (TLFN), which have temporal processing capability without resorting to complex and costly training methods, and on evolutionary polynomial regression (EPR) which is based on hybrid evolutionary paradigm. The results of these two models are compared with the well known multiple regression based downscaling tool namely statistical downscaling model (SDSM).

2 Study area and data

The study area selected in this research for the application and evaluation of downscaling methods is the Chute-du-Diable basin located in the Saguenay-Lac-Saint-Jean watershed (Fig. 1), which is a well-known flood prone region in Canada. There are a large number of reservoirs and dams in the Saguenay watershed and most of the large reservoirs are managed by the Aluminum Company of Canada (ALCAN) for hydroelectric power production. However, in this study, only the results from Chute-du-Diable are presented. Chute-du-Diable has an area of 9700 km² and is located in the eastern part of the Saguenay watershed (Fig. 1). Note that the study area is about one quarter of the grid size (40 000 km²) of the meteorological model.

Twenty-three years (1979–2001) of historical daily total precipitation (Prec.), daily maximum temperature (Tmax) and daily minimum temperature (Tmin) series are collected from the Chute-Du-Diable meteorological station located at

Table 1. NOAA reforecast ensemble variable fields.

| Variable field | Description | Surface level (mb) |
|----------------|---|--------------------|
| apcp | Accumulated precipitation (mm) | Surface |
| heating | Vertically integrated diabatic heating (K/s/mb) | Vertical average |
| pwat | Precipitable water | Surface |
| prmsl | Pressure reduced to mean sea-level (Pa) | Surface |
| t2 m | Temperature at 2meters (K) | Surface |
| rhum | Relative humidity (%) | 700 mb |
| u10 m | Zonal wind at 10 m (m/s) | Surface |
| v10 m | Meridional wind at 10 m (m/s) | Surface |

48.75° N and 71.7° W and used as predictands in this study. The NOAA-CIRES Climate Diagnostic Center has undertaken a reforecasting project providing retrospective numerical ensemble forecasts. An unchanged version of National Centers for Environmental Prediction's Global Forecast System (NCEP GFS, formerly known as MRF) at T62 resolution is used to generate 15-day real-time forecast scenarios (30 time steps of 12 h each). Forecasts are run every day from 00:00 UTC initial conditions from 1979 to present. There are 15-member ensemble forecasts that are generated from 15 initial conditions consisting of a reanalysis and seven pairs of bred modes (Hamill et al., 2004). The global latlon grid has a large-scale resolution of 2.5° both in longitude and latitude and contains 144×73-grid points. The global data were collected directly from the reforecast project ftp server. There are 12 files per day and the field variables are described in Table 1. These files are netCDF (network Common Data Format) files.

In order to get geographical subsets of grid points over the region of interest, we used an operator named “ncks” (netCDF Kitchen Sink) from NCO (netCDF Operators). This operator is executed through a Matlab Graphical User Interface we have developed. Geographical subsets are produced by “ncks” only from a global latlon grid. So, we have been able to process only the first eight variable fields shown on Table 1. The geographical subsets files are also netCDF files. So a second operation was necessary to transform the netCDF files for the geographical subsets into Matlab files using MexCDF conversion utilities. MexCDF is a mex-file interface between NetCDF and MATLAB. All the predictors are extracted from the grid point closest to the Chute-Du-Diable meteorological station located at 48.75° N and 71.7° W.

3 Time lagged feedforward neural network (TLFN)

A neural network is characterized by its architecture, which is represented by the network topology and pattern of connections between the nodes, its method of determining the connection weights, and the activation functions that it em-

ploy. Multi-layer perceptrons (MLPs), which constitute probably the most widely used network architecture, are composed of a hierarchy of processing units organized in a series of two or more mutually exclusive sets of neurons or layers. The information flow in the network is restricted to a flow, layer by layer, from the input to the output, hence also called feedforward network.

TLFN is a neural network that can be formulated by replacing the neurons in the input layer of an MLP with a memory structure, which is sometimes called a tap delay-line. The size of the memory layer (the tap delay) depends on the number of past samples that are needed to describe the input characteristics in time and it has to be determined on a case-by-case basis. TLFN uses delay-line processing elements, which implement memory by simply holding past samples of the input signal. The output (y) of such a network with one hidden layer is given by:

$$\begin{aligned}
 y(n) &= \varphi_1 \left(\sum_{j=1}^m w_j y_j(n) + b_0 \right) \\
 &= \varphi_1 \left(\sum_{j=1}^m w_j \varphi_2 \left(\sum_{j=0}^k w_{ji} x(n-1) + b_j \right) + b_0 \right) \quad (1)
 \end{aligned}$$

where m is the size n is the weight vector for the connection between the hidden and output layers, k is the memory depth, w_{ji} is the weight matrix for the connection between the input and hidden layers, φ_1 and φ_2 are transfer functions at the output and hidden layers respectively; b_j and b_0 are additional network paed biases) to be determined during training of the networks with observed input/output case of multiple inputs (of size p), the delay-line with a memory depth k can be represented by

$$\chi(n) = [X(n), X(n-1), \dots, X(n-k+1)] \quad (2)$$

where $X(n) = (x_1(n), x_2(n), \dots, x_p(n))$ and represents the input pattern at time step n , $x_j(n)$ is an individual input at the n th time step and is the combined input matrix to the processing elements at time step n . Such delay-line only “remembers” k samples in the past. The advantage of TLFNs is that they share some of the nice properties of feedforward

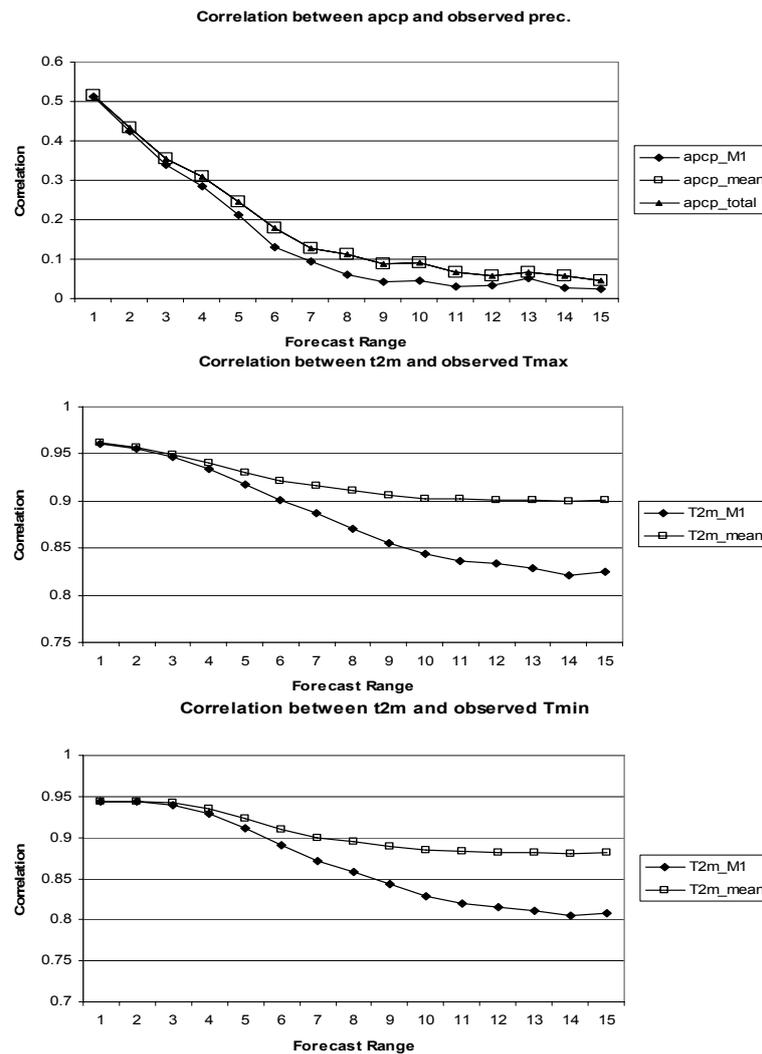


Fig. 2. Correlation plots between predictors and observed Precipitation, Tmax and Tmin.

neural networks, but they can capture the information present in the input time signals. An interesting feature of the TLFN is that the tap delay-line at the input does not have any free parameters; therefore the network can still be trained with the classical backpropagation algorithm. The TLFN topology has been successfully used in non-linear system identification, time series prediction, and temporal pattern recognition (Principe et al., 2000). A major advantage of the TLFN is that it is less complex than the conventional time delay and recurrent networks and has the similar temporal patterns processing capability (Coulibaly et al, 2001).

4 Evolutionary Polynomial Regression (EPR)

EPR is a hybrid evolutionary regression technique based on genetic programming (GP) introduced by Koza (1992). GP is

a method for constructing populations of mathematical models using stochastic search methods namely evolutionary algorithms. For multivariate time series modeling using the GP approach, the ultimate objective of the evolutionary process is to discover an optimal equation (or model) for relating dependent variable (or predictand) and independent variables (or predictors). However, as the search space of all possible equations is extremely large particularly for multivariate time series, the heuristic search needs to be optimized in term of computational efficiency and parsimonious solution (i.e. model structure). The evolutionary polynomial regression (EPR) technique recently proposed by Giustolisi and Savic (2005) aims to provide optimal solution by exploiting both the numerical and symbolic regression. Essentially, EPR uses a GA to find the form of the polynomial expressions and least squares optimization to determine the values of the parameters in the expressions. The description of the EPR

Table 2. Selected predictors for all the three downscaling methods.

| Predictand | Downscaling method | Selected lags of predictors | | | | | | | |
|------------|--------------------|-----------------------------|---------------------------------------|---|------------|---|------------|------------|------------|
| | | apcp | heating | pwat | prmsl | t2m | rhum | u10m | v10m |
| PREC | SDSM | lag0,1,2,3 | lag1,2,3 | lag0,1,2 | | | lag0,1 | | |
| | TLFN | lag0,1,2,3 | lag1,2,3, 4,5,7,8,9, 10,11,12 | lag0,1,2, 3,5,6,7,9, 10,11,12 | lag0,1,2,3 | lag0,1,2, 3,4,5,6,8, 9,10,11,12 | lag0,1,2 | lag0,1,2,3 | lag0,1,3 |
| Tmax | EPR | lag0,1,2 | lag0,1,2 | lag0,1,2 | lag0,1,2 | lag0,1,2 | lag0,1,2 | lag0,1,2 | lag0,1,2 |
| | SDSM | | | | | lag0,1,2, 3,4,5,6, 7,8,9,10,12 | | | |
| Tmin | TLFN | lag0,1, 2,3 | lag0,1,2, 3,4,6,7, 8,9,10,11,12 | lag0,2,3, 5,6,7,9, 10,11,12 | lag0,1,2,3 | lag0,1,2, 3,4,5,6, 7,8,9,10,11,12 | lag0,1,2 | lag0,1,2,3 | lag0,1,2,3 |
| | EPR | lag0,1 | lag0,1 | lag0,1 | lag0,1 | lag0,1 | lag0,1 | lag0,1 | lag0,1 |
| Tmin | SDSM | | | | | lag0,1,2, 3,4,5,6, 7,8,9,10,11 | | | |
| | TLFN | lag0,1,2,3 | lag1,2,3, 4,5,6,7, 8,9,10,11,12 | lag0,1,2, 3,4,5,6, 7,8,9,10,11,12 | lag0,1,2,3 | lag0,1,2, 3,4,5,6, 7,8,9,10,11,12 | lag0,1,2 | lag0,1,2,3 | lag0,1,2,3 |
| | EPR | lag0,1,2,3 | lag0,1,2,3 | lag0,1,2,3 | lag0,1,2,3 | lag0,1,2,3 | lag0,1,2,3 | lag0,1,2,3 | lag0,1,2,3 |

Lag: indicates the forecast range of the predictor variable. Lag 0 indicate the MRF ensemble forecasts (15 members) for day 1; Lag 1 stands for ensemble forecasts for day 2 and so one.

Table 3. Optimal TLFN and EPR model structure and parameters.

| TLFN | | EPR | |
|--------------------------------|------------------|-----------------|---|
| parameter | value | parameter | value |
| Processing Element | 20 | EPR type | $Y = \sum(a_i \hat{X}_1 \hat{X}_2 \dots f(X_1) \dots f(X_2)) + a_0$ |
| Epoches | 2000 | Regression type | Dynamical |
| Memory | GammaAxon | Generation | 20 |
| Number of hidden layer | 1 | Function | Secant Hyperbolic |
| Hidden layer transfer function | SoftMaxAxon | Terms | [1:5] |
| Output layer transfer function | LinearAxon | Exponent | [-1,0.5,1,2,3] |
| Stopping criteria | Cross validation | Solution | Linear programming |
| Learning rule | DeltaBarDelta | Strategy | Multi-objective genetic algorithm |

method is limited herein to the needs of the present study. For more detailed description of the EPR method, the readers are referred to other sources, such as Giustolisi and Savic (2003, 2004, 2005).

Although the EPR technique is similar to the rule-based symbolic regression (Davidson et al., 2000), there is a key difference in the search for model structure. While the latter uses rules to simplify symbolic expressions, the former employs a simple GA to search in the model structure space.

In general, the rule-based symbolic regression limits the range of operators normally used in symbolic regression to addition, multiplication and non-negative integer powers. The expressions that result from applying the limited set of

operators are usually in the form of polynomials such as

$$y = \sum_{j=1}^m a_j z_j + a_0 \tag{3}$$

where y is the least squares estimate of the target value, a_j is an adjustable parameter for the j th term, a_0 is an optional bias, m is the number of terms/parameters of the expression, and z_j is a transformed variable. In EPR method, it is useful to transform Eq. (4) into the following vector form (Giustolisi and Savic, 2005)

$$Y_{N \times 1}(\theta, Z) = [I_{N \times 1} Z_{N \times m}^j] [a_0 a_1 \dots a_m]^T = Z_{N \times d} \theta_{d \times 1} \tag{4}$$

where $Y_{N \times 1}(\theta, Z)$ is the least squares estimate vector of the N target values; $\theta_{1 \times d}$ is the vector of $d=m+1$ parameters a_j

Table 4. Seasonal model performance for daily Prec., Tmax, and Tmin for the validation period using SDSM, TLFN and EPR downscaling models.

| Predictand | Model | Season | | | | | | | | | | | |
|------------|-------|--------|------|----------|--------|------|----------|--------|------|----------|--------|------|----------|
| | | Winter | | | Spring | | | Summer | | | Autumn | | |
| | | MSE | NMSE | <i>r</i> |
| PREC | SDSM | 32.89 | 1.86 | 0.03 | 27.09 | 1.33 | 0.09 | 78.77 | 1.66 | 0.04 | 56.48 | 1.47 | 0.04 |
| | TLFN | 9.51 | 0.54 | 0.71 | 12.17 | 0.60 | 0.64 | 34.17 | 0.72 | 0.56 | 25.13 | 0.65 | 0.61 |
| | EPR | 9.53 | 0.54 | 0.71 | 11.66 | 0.57 | 0.66 | 34.53 | 0.73 | 0.55 | 24.37 | 0.63 | 0.64 |
| Tmax | SDSM | 84.94 | 1.82 | 0.12 | 89.40 | 1.02 | 0.62 | 34.41 | 1.67 | 0.15 | 51.54 | 0.78 | 0.70 |
| | TLFN | 8.27 | 0.18 | 0.91 | 13.55 | 0.15 | 0.94 | 8.55 | 0.42 | 0.81 | 7.54 | 0.11 | 0.94 |
| | EPR | 9.17 | 0.20 | 0.91 | 15.60 | 0.18 | 0.94 | 8.60 | 0.42 | 0.81 | 7.64 | 0.12 | 0.94 |
| Tmin | SDSM | 43.08 | 0.53 | 0.75 | 25.26 | 0.31 | 0.86 | 8.49 | 0.55 | 0.71 | 11.36 | 0.24 | 0.88 |
| | TLFN | 22.68 | 0.28 | 0.85 | 11.64 | 0.14 | 0.93 | 5.42 | 0.35 | 0.81 | 8.61 | 0.18 | 0.91 |
| | EPR | 26.56 | 0.33 | 0.82 | 12.80 | 0.16 | 0.93 | 5.42 | 0.35 | 0.82 | 8.49 | 0.18 | 0.93 |

and a_0 (θ^T is the transposed vector); and $Z_{N \times d}$ is a matrix formed by I , unitary vector for bias a_0 , and m vectors of variables Z_j that for fixed j are a product of the independent predictor vectors of inputs, $X=(X_1 X_2 \dots X_k)$. The key idea behind the EPR is to use evolutionary search for exponents of polynomial expressions by means of a GA engine (Giustolisi and Savic, 2004, 2005). This allows: (a) easy computational implementation of the algorithm; (b) efficient search for an expression (formula); (c) improved control of the complexity of the expression generated; and (d) a small number of search parameters to be pre-specified (Giustolisi and Savic, 2005).

5 Model design

When applying the three downscaling methods in this study, data from 1979 to 1996 are used to construct the models, and data from 1997 to 2001 are used for validation to test the model performance. There are two major steps in designing the models, the first step is to select the input predictors, and the second step is to determine the model parameters.

5.1 Selection of predictors

As described before, the predictor variables are derived from the 3-D ensemble forecasts. This means each variable has 15 time delays (forecast range), and in each delay, there are 15 members. First, the correlations between the predictands and the members of certain predictors are calculated to decide which members should be selected as inputs in the downscaling models. Among all the possible predictor variables for precipitation, the predicted accumulated precipitation (apcp) appears the most correlated to observed daily precipitation (Prec.). Similarly, predicted temperature at 2 m (t2m) is most correlated to observed Tmax and Tmin. To further investigate the time delay effect, the correlation between the observed precipitation and the first member (M1) of the 15

members of apcp, the mean of the 15 members, and the total of the 15 members of the variable apcp are compared (see Fig. 2). Similar analysis is done for Tmax and Tmin, correlations between the observed temperature (Tmax, Tmin) and M1 (of the 15 members of t2m) are compared with the correlations between the observed temperature and the mean of the 15 members of t2m (Fig. 2). It appears from Fig. 2 that the mean of the 15 members is most correlated with the observed values for apcp and t2m respectively. Therefore, the mean of each predictor variable is preferred rather than using any single member from its ensemble.

Then partial auto-correlation analysis (PACF) is performed for each predictor variable to find significant time lags. Then all the 8 predictors (Table 1) with their significant lags from PACF analysis are used as input variables in TLFN models to perform sensitivity analysis. Sensitivity analysis provides a measure of the relative importance among the predictors by calculating how the model output varies in response to variation of an input. The relative sensitivity of the model to each input is calculated by dividing the standard deviation of the output by the standard deviation of the input, which is varied to create the output. The results provide a measure of the relative importance of each input (predictor) in the particular input-output transformation. Based on sensitivity analysis results, the most relevant input variables are then selected. The final selected predictor variables used for the three models are presented in Table 2. Note that the downscaling is performed for all the forecast range (i.e. 15 days ahead) by relating the 15 days ahead weather forecasts with the 15 days ahead observations (precipitation or temperature).

5.2 Model parameters determination

Once the input variables are selected, they are used to construct downscaling models with the three different methods.

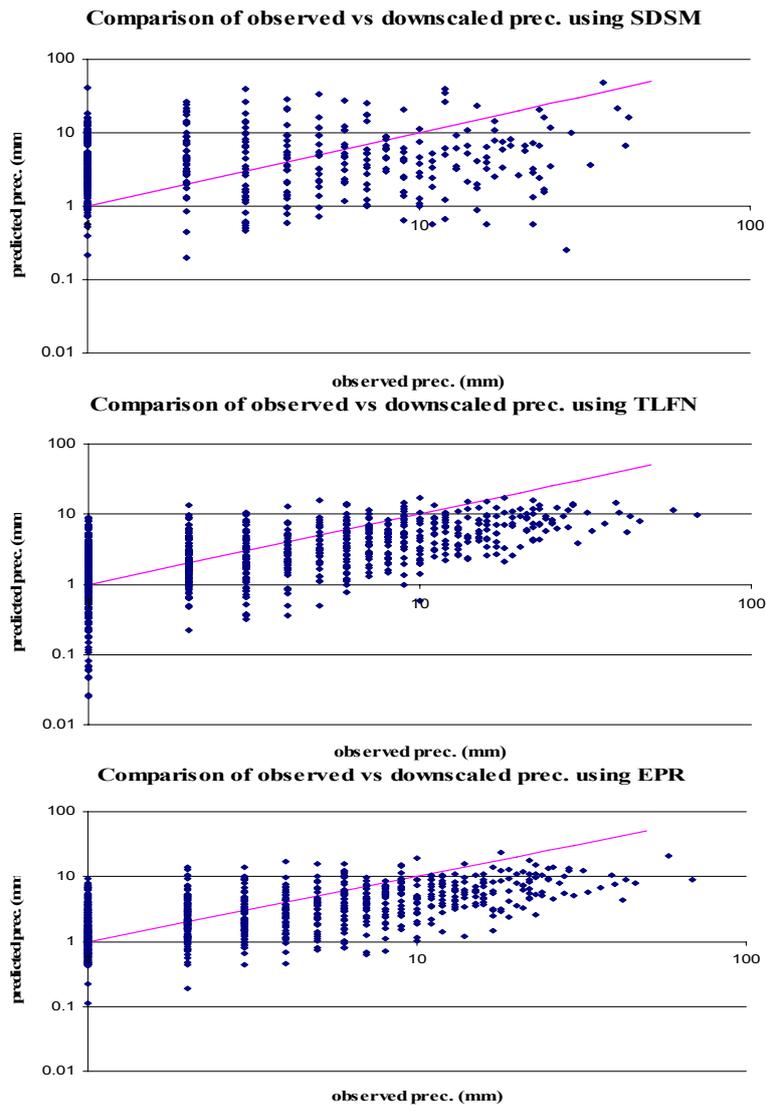


Fig. 3. Scatter plots of observed versus downscaled Prec. using SDSM, TLFN, EPR.

Because SDSM is a well-known multiple regression technique, here only structure of TLFN and EPR are described. All potential optimal combinations of the parameters of TLFN and EPR are trained to get the best models. The performance of the models is evaluated by three statistics: mean squared error (MSE), normalized mean squared error (NMSE, $NMSE = MSE / \text{variance of desired output}$), and correlation (r) between model output and desired output for validation period. The optimal model parameters identified for TLFN and EPR are shown in Table 3.

6 Downscaling results

Validation statistics in terms of seasonal model MSE, NMSE, and r are used (Table 4). Surprisingly, the comparative re-

sults indicate that TLFN and EPR have very similar performance in downscaling precipitation, Tmax and Tmin, and they perform much better than SDSM, especially in downscaling precipitation. Precipitation is always more difficult to downscale than temperature, but the TLFN and EPR perform well in winter, and then in spring and autumn. While the results are not so good in summer, this may be caused by heavier precipitation in form of convective storms and thunderstorms that are difficult to model with large scale weather models. All the three methods have better model results for temperature in spring and autumn, this is because the temperature usually has much higher variance in winter and summer, which makes it more difficult to model.

Moreover, scattered plots of downscaled versus observed Prec., Tmax and Tmin data during the whole validation

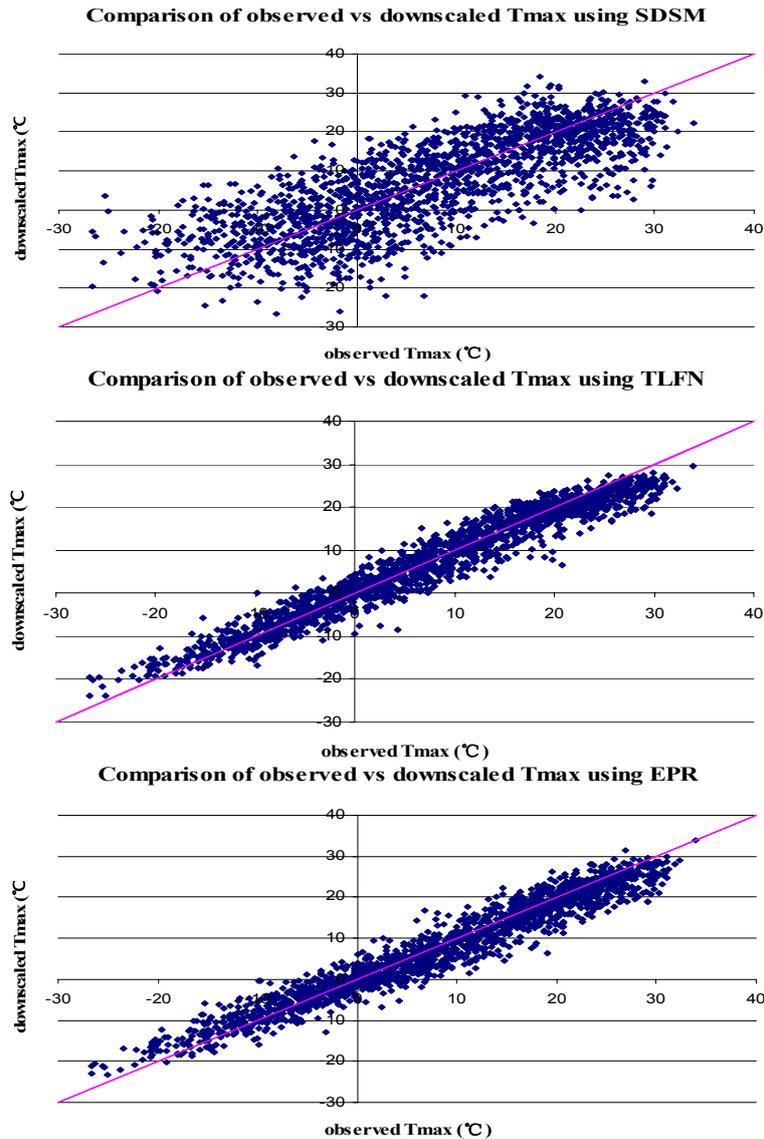


Fig. 4. Scatter plots of observed versus downscaled Tmax using SDSM, TLFN, EPR.

period are used to show the model results. All the output data from validation period are plotted and a comparison line which represent the perfect model is also shown on the plot. Figure 3 shows the logarithmic scale plots of downscaled Prec. versus observed Prec. using SDSM, TLFN, and EPR respectively. It can be seen that all the spots from TLFN and EPR are distributed more closely around the perfect model line than those of SDSM. Similar results for Tmax and Tmin can be seen in Fig. 4 and Fig. 5. All the three methods demonstrated good performance in downscaling Tmax and Tmin, but TLFN and EPR are still better than SDSM. Although SDSM appears to perform poorly as compared to the other models, it consistently uses a much smaller number of parameters whatever the predictand of concern. The EPR

also appears more parsimonious than the TLFN in term of number of input variables. In general, the comparative results suggest TLFN and EPR have a good potential for downscaling ensemble weather forecasts. However, further improvement is needed for the downscaling of precipitation series.

7 Conclusions

This study investigates and compares three data-driven methods, SDSM, TLFN, and EPR, for downscaling ensemble daily precipitation and temperature series at the Chute-du-Diable station located in the Saguenay watershed in north-eastern Canada. The comparative results show that TLFN and EPR have quite similar performance in downscaling

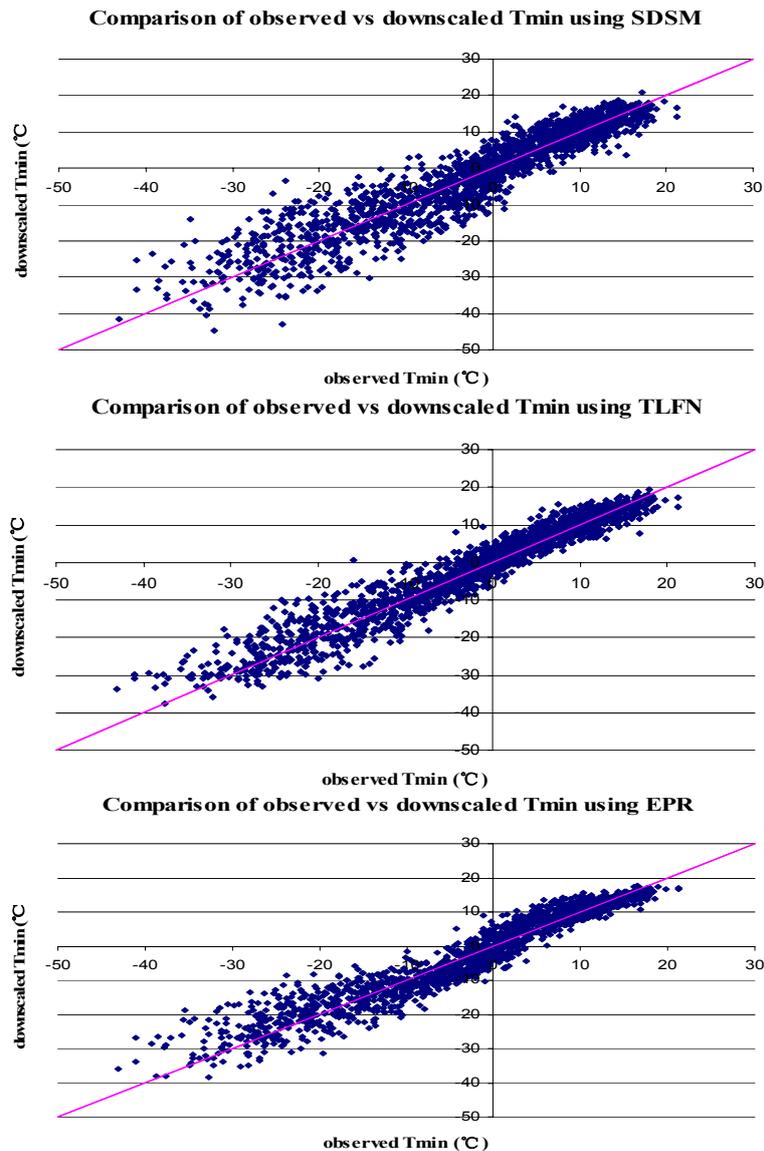


Fig. 5. Scatter plots of observed versus downscaled Tmin using SDSM, TLFN, EPR.

Prec., Tmax and Tmin, and they perform much better than SDSM in all the downscaling experiments. For precipitation, TLFN and EPR have relatively good results in winter than in other 3 seasons, while SDSM performs poorly in all the four seasons. TLFN and EPR have very good results in mapping Tmax and Tmin, especially in spring and autumn. SDSM also performs well in downscaling temperature, but not as good as TLFN and EPR. The comparative study indicates that EPR and TLFN exhibit good potential for downscaling ensemble weather forecasts. Further study will include the downscaling of all the 15 members to better assess the variance of downscaled data as compared to the observed data. This will lead to ensemble hydrologic modeling using the downscaled ensemble variables.

Acknowledgements. This work was made possible through a joint grant (CRD) from NSERC and Hydro-Quebec. The authors wish to thank O. Giustolisi (Technical University of Bari) and D. Savic (University of Exeter) for providing the EPR software used in this study. The authors are grateful to A. Doglioni for his precious suggestions and help in applying the EPR software. The authors thank the Aluminum Company of Canada (Alcan) for providing the experiment data. The ensemble reforecast data is kindly made available by NOAA at: <http://www.cdc.noaa.gov/reforcast/>.

Edited by: D. Solomatine

References

- Bronstert, A., Niehoff, D., and Burger, G.: Effects of climate and land-use change on storm runoff generation: Present knowledge and modeling capabilities, *Hydrol. Processes*, 16(2), 509–529, 2002.
- Burger, G. and Chen, Y.: Regression-based downscaling of spatial variability for hydrologic applications, *J. Hydrol.*, 311(1–4), 299–317, 2005.
- Conway, D., Wilby, R. L., and Jones, P. D.: Precipitation and air flow indices over the British Isles, *Climate Res.*, 7, 169–183, 1996.
- Coulibaly, P., Anctil, F., and Bobee, B.: Multivariate reservoir inflow forecasting using temporal neural networks, *J. Hydrol. Eng.*, ASCE, 6(5), 367–376, 2001.
- Coulibaly, P., Dibike, Y. B., and Anctil, F.: Downscaling Precipitation and Temperature with Temporal Neural Networks, *J. Hydrometeorol.*, 6(4), 483–496, 2005.
- Coulibaly, P.: Downscaling daily extreme temperatures with genetic programming, *Geophys. Res. Lett.*, 31, L16203, doi:10.1029/2004GL020075, 2004.
- Davidson, J. W., Savic, D. A., and Walters G. A.: Approximators for the Colebrook-White Formula Obtained through a Hybrid Regression Method, *Computational Methods in Water Resources*, Bentley, 2000.
- Dibike, Y. B. and Coulibaly, P.: Temporal neural networks for downscaling climate variability and extremes, *Neural Networks*, 19, 135–144, 2006.
- Giustolisi, O. and Savic, D. A.: Evolutionary Polynomial Regression (EPR): Development and Applications, Report 2003/01, School of Engineering, Computer Science and Mathematics, Centre for Water Systems, University of Exeter, 2003.
- Giustolisi, O. and Savic, D. A.: A novel strategy to perform genetic programming: Evolutionary Polynomial Regression, 6th Int. Conf. on Hydroinformatics, Singapore, edited by: Liong, S. Y., Phoon, X., and Babovic, V., World Scientific Publishing Company, 1, 787–794, 2004.
- Giustolisi, O. and Savic, D. A.: A Symbolic Data-driven Technique Based on Evolutionary Polynomial Regression, *J. Hydroinform.*, 8(3), 207–222, 2005.
- Hamill, T. M., Whitaker, J. S., and Wei, X.: Ensemble Reforecasting: Improving Medium-Range Forecast Skill Using Retrospective Forecasts, *Mon. Weather Rev.*, 132(6), 1434–1447, 2004.
- Koza, J. R.: Genetic Programming: On the Programming of Computers by Natural Selection, MIT Press, Cambridge, MA, 1992.
- Principe, J. C., Euliano, N. R., and Lefebvre, W. C.: Neural and adaptive systems: Fundamentals through simulations, New York, 2000.
- Salathe, E. P.: Comparison of various precipitation downscaling methods for the simulation of streamflow in a rainshadow river basin, *Int. J. Climatol.*, 23, 887–901, 2003.
- Von Storch, H., Hewitson, B., and Mearns, L.: Review of empirical downscaling techniques, in: Regional climate development under global warming, edited by: Iversen, T. and Høiskar, B. A. K., General Technical Report No. 4, Conference Proceedings Regclim Spring Meeting Jevnaker, Torbjørnrud, Norway, 29–46, 2000.
- Wilby, R. L., Dawson, C. W., and Barrow, E. M.: SDSM – a decision support tool for the assessment of regional climate change impacts, *Environ. Modell. Softw.*, 17, 147–159, 2002.
- Wilby, R. L. and Wigley, T. M. L.: Downscaling general circulation model output: a review of methods and limitation, *Prog. Phys. Geog.*, 21, 530–548, 1997.