Improving regional climate change projections of temperature for Halifax, Nova Scotia via statistical downscaling

LEE TITUS *

METEOROLOGICAL SERVICE OF CANADA/DALHOUSIE UNIVERSITY

RICHARD GREATBATCH

DALHOUSIE UNIVERSITY/ATMOSPHERIC SCIENCE, HALIFAX, NOVA SCOTIA, CANADA

IAN FOLKINS

DALHOUSIE UNIVERSITY/ATMOSPHERIC SCIENCE, HALIFAX, NOVA SCOTIA, CANADA

JINYU SHENG

DALHOUSIE UNIVERSITY/PHYSICAL OCEANOGRAPHY, HALIFAX, NOVA SCOTIA, CANADA

---

* *Corresponding author address:* Lee Titus, 16th Floor Queen Square, 45 Alderney Dr, Dartmouth NS, B2Y-2N6.

E-mail: Lee.Titus@ec.gc.ca

ABSTRACT

The output from the Canadian general circulation model version 3 (CGCM3) run on the A2 emission scenario was statistically downscaled to Halifax NS in this study. The maximum daily temperature (TMAX) was selected as the predictand in winter (DJF). The seasonal cycle of the predictand and predictors was removed from each. Data reduction was employed to remove predictors that are not useful or redundant. The principal components (PC's) of the remaining predictors were calculated from 1961-2000 daily National Center for Environmental Prediction (NCEP) values. A multiple linear regression was then developed using the predictand (TMAX) and the predictors (NCEP PC's). The regression was then used to hindcast 1961-2000 TMAX using PC's created from the 1961-2000 CGCM3 predictors by projecting them onto the NCEP derived eigenvectors. Finally the future CGCM3 predictors were transformed into their PC's via projection onto the NCEP eigenvectors to make projections of TMAX in future tri-decades (30 year periods).

The developed regression was found to have good predictive skill and overfitting was not an issue. The explained variance of the regression was 79 percent. The historical (1961-2000) prediction using both the PC's from NCEP and the CGCM3 was able to capture the mean and variance of observed TMAX. The PC that had the highest correlation with TMAX was examined for governing physics. During winter, temperature advection is the dominant forcing on TMAX and explains nearly forty percent of the variance in TMAX. Although the variance of TMAX remains near the historical value in the future predicted distribution, the

mean of the distribution increases by nearly 3.5 degrees in the 2080's (2071-2100) compared to the historical mean

1. Introduction

In order to best assess the expected climate change impacts on a species, ecosystem or natural resource in a region, climate variables and climate change scenarios must be developed on a regional or even site-specific scale (Wilby et al. 2002). To provide these values, projections of climate variables must be 'downscaled' from the general circulation model (GCM) results, utilizing either dynamical or statistical methods (Houghton et al. 2001). The typical grid spacing of a GCM is shown in figure 1. It is clear that with a typical resolution of 300 by 400 kilometers, it is not capable of resolving the small scale features that influence climate such as sea breezes and the like. Without downscaling, the climate is the same everywhere inside the grid box of the GCM. This is clearly not very realistic and does not give you confidence in the future climate projections at a specific location. The most important and fundamental part of statistical downscaling is to create a realistic statistical model from observations. This model should not only have predictive skill but also represent real physics governing the predictand. Once a realistic statistical model is obtained, predictors from a GCM can be used to hindcast the historical climate at the site. If the GCM predictors have a similar distribution to the observed predictors then the statistical model should produce a realistic predictand ditribution in the historical period. This certainly boosts confidence in the projections created using future GCM predictors.

2. DATA

In this study, predictors from the Canadian coupled general circulation model, version 3 (CGCM3) (Flato and Boer 2001) The A2 emission scenario were used as the forcing in the CGCM3. The historical predictor datasets, consist of 25 daily National center for environmental prediction (NCEP) and Canadian general circulation model version 3 (CGCM3) predictors from 1961-2000. The 25 CGCM3 daily predictors from 2001 to 2100 are the future predictors. Information on the creation of the predictors can be found at www.cccsn.ca. The predictand data is homogenized daily maximum temperature (TMAX) from 1961-2000 (Vincent et al. 2002) taken from Shearwater, Nova Scotia in Canada. Shearwater, depicted in figure 1 is about 4km east of the downtown core of Halifax was used as a proxy for Halifax.

The predictand was transformed into Z scores to be consistent with the predictors which were downloaded as Z scores. The Z scores were created using the following expression:

$$Z_i = \frac{X_i - \bar{X}}{\sigma} \tag{1}$$

where X_i is a particular observation, $\bar{X}$ is the mean from $1961-2000$ and σ is the standard deviation from $1961 - 2000$. The result is historical predictand data that has a mean of zero and a standard deviation of one. It should be noted that the predictor data, both historical and future Z scores were standardized based on the mean and standard deviation from $1961 - 1990$. The main purpose of this is to make the future trend $(2001 - 2100)$ match with the historical trend. This allows the trend in the predictors to be continuous. Z scores of the preditors are important in this analysis to remove any bias in the GCM predictors compared to NCEP.

3. METHODOLOGY

The basic methodology contains several steps; First, wind direction and speed and divergence at all three levels was removed. Next the seasonal cycle of the predictand and predictors was removed. The principal components (PC's) of the final NCEP predictor set were calculated. The highest correlated PC with the predictand were used to train and validate the regression. The first 30 winters of predictand and predictors were taken to aquire the regression coefficients. The regression coefficients were then used with the predictors from the final 9 winters to predict TMAX for the final 9 winters. This allows the regression to predict TMAX in the period independent of the training data. Once the best NCEP regression was made, the same predictors from the CGCM3 were transformed into their principal components via projection onto the NCEP derived eigenvectors. These were used in the regression to hindcast the CGCM3 TMAX from $1961-2000$. It was determined that the regression produced a much more realistic distribution of TMAX than the raw CGCM3 itself. Finally the CGCM3 predictors from the future were broken into three tri-decades (30 years) and turned into their principal components to predict TMAX distributions for the future.

4. RESULTS

The final predictors used in the regresssion and their selection process can be seen in Table 1. A correlation cutoff between the predictors and predictand of 0.05 was chosen which allowed 10 PC's in the regression. This gave an explained variance of near 79 percent.

The highest correlated principal component had a correlation of 0.62 and its weightings of the original predictors is shown in table 2. The regression accuracy was determined using γ^2 (Thompson and Sheng 1997) which is the variance in the prediction errors divided by the variance in the observations. γ^2 was determined to be 0.21. Validation of the regression showed a correlation of 0.89 between the observed TMAX and the predicted TMAX for the final 9 winters (independent of training data). Therefore the regression has skill and overfitting is not an issue. Regression information can be found in table 4. Figure 3 shows the observed TMAX distribution and the NCEP predicted distributions from 1961-2000. From a visual perspective the shape of the predicted distribution looks like observations. Further justification of the regression was done through an investigation of the physics. It makes sense that the physics should be dominated by the day to day synoptic forcing. A PC with a correlation of 0.62 explains nearly 40 percent of the variance in TMAX. From Table 2 it is clear that the high positive weightings of meridional windspeeds in a positively correlated PC are associated with warm advection. A positive meridional wind advects warmer air from the south and TMAX goes up. This gives confidence in our statistical model in that it is physically sensible.

In order to account for the unexplained variance of the regression an inflation factor (Huth 2002) of 10 percent was used. To quantify the accuracy of the predicted mean and variance, hypothesis testing was employed. A paired t test for the mean and an F test for the variance. The mean and the variance of the NCEP prediction turned out to be statistically the same as observations witin a 95 percent confidence interval.

Next the CGCM3 predictors were used to make a historical prediction as described in the methodology. Figure 4 shows the frequency distributions for 1961-2000 TMAX observed, CGCM3 regression predicted and the raw CGCM3 model output. It is clear visually that there has been a large improvement in the distribution from the raw CGCM3 output to the regression predicted distribution as compared the observed distribution. The same hypothesis test for the mean and variance was done to verify that the CGCM3 regression gave the same mean and variance as observed. It was again found that the mean and variance are statistically equivalent within the 95 percent confidence limits.

Finally, the developed regression was used on the predictors (PC's) in three future tri-decades. The tri-decades are 2011-2040, 2041-2070, 2071-2100. The numerical means and standard deviations of the future tri-decades can be found in table 3.

5. CONCLUSIONS

The methodology described in this study is a method that works well in this case. It cannot be stated with any confidence that the method will work equally well for another location. The strength of this method is that it maintains the assumptions on which multiple linear regression is based. Removing the seasonal cycle certainly reduces serial correlation and allows for independent regression errors. Also actually comparing the NCEP and CGCM3 predictor distributions is essential. Even though you may end up throwing out highly correlated predictor for TMAX, if the GCM does not reproduce the NCEP distribution for that predictor it is useless in this process.

Statistical downscaling allows you to get a local climate projection using a large scale GCM. The method requires minimal computational power which makes it very attractive to scientists needing site specific climate information in their research. The major downfall is that you must assume the regression holds in the future. This assumption seems plausible since predictors like 500hpa geopotential height are used. 500hpa height is directly related to mean layer temperature (if you assume surface pressure is constant) which in turn can be forced by anthropogenic warming. The question becomes, do the dynamic predictors capture all of the warming associated with this anthropogenic forcing? The regression mean shift has the local climate forcing in it and hence predicts a smaller mean shift compared to what the GCM gridbox would suggest. However it is unlikely that the predictors capture all the physics influencing the warming. The actual mean shift for this model and emission scenario probably lies in between the method described in the paper and the trend from the raw gridbox. Future work will hopefully sort out exactly where. Another problem with the linear regression is it underestimates the variance. The probability in the tails of the projected distribution has a large influence on the type of world we will live in. A more advanced non-linear regression is probably needed to improve the extremes prediction.

In closing it is important to point out that this work is still in it's infancy. This method is the best method found for this location with the best method likely being different in another location. Also this technique has only been done for one model with a specific emission scenario, but different GCM's run of different emission scenarios produce slightly different but equally palausible results. This work is important because as GCM's progress and start to narrow their solution range, we need to have the best methods possible to get high quality and trustworthy projections for decisions on adaptation to climate change.

REFERENCES

Flato, G. and G. Boer, 2001: Warming asymmetry in climate change simulations. *Geophysical research letters*, **28 (1)**, 195–198.

Houghton, J., Y. Ding, D. Griggs, M. Noguer, P. van der Linden, X. Dai, and K. Maskell, 2001: *Climate Change 2001: The Scientific Basis*. elsevier.

Huth, R., 2002: Statistical Downscaling of Daily Temperature in Central Europe. *Journal of Climate*, **15 (13)**, 1731–1742.

Thompson, K. and J. Sheng, 1997: Subtidal circulation on the Scotian Shelf: Assessing the hindcast skill of a linear, barotropic model. *Journal of Geophysical Research*, **102 (C11)**, 24 987–25 004.

Vincent, L., X. Zhang, B. Bonsal, and W. Hogg, 2002: Homogenization of Daily Temperatures over Canada. *Journal of Climate*, **15 (11)**, 1322–1334.

Wilby, R., C. Dawson, and E. Barrow, 2002: sdsma decision support tool for the assessment of regional climate change impacts. *Environmental Modelling and Software*, **17 (2)**, 145–157.

List of Figures

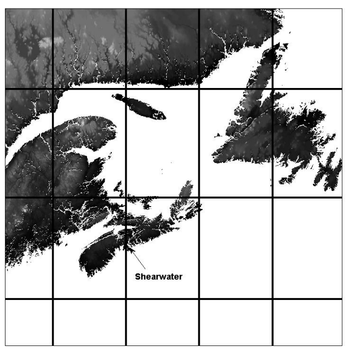

FIG. 1. Typical Grid boxes from a general circulation model are about 300km by 400km plotted over Atlantic Canada. This study takes it's observations from Shearwater NS which is used as a proxy for Halifax. Shearwater is about 4KM East of Halifax's downtown core.

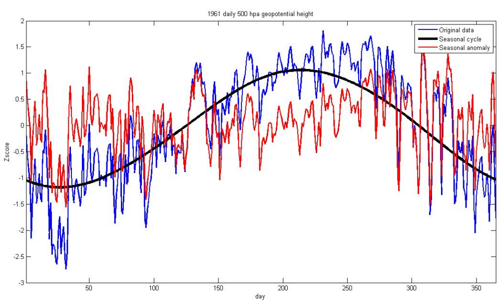

Fig. 2. One year (1961) of NCEP geopotential height are plotted here (blue). The 1961-2000 fitted seasonal cycle is plotted on top in black. The actual predictor used in the study are the seasonal anomalies which is the actual data minus the seasonal cycle plotted in red.

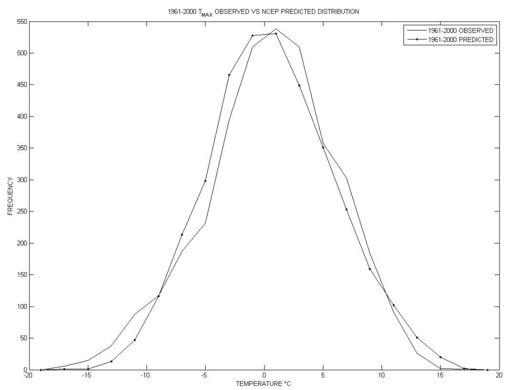

Fig. 3. Comparison of the NCEP predicted distribution of TMAX for 1961-2000 winters compared to the observed distribution. Regression information can be found in Table 4 and the highest correlated PC can be viewed in Table 2

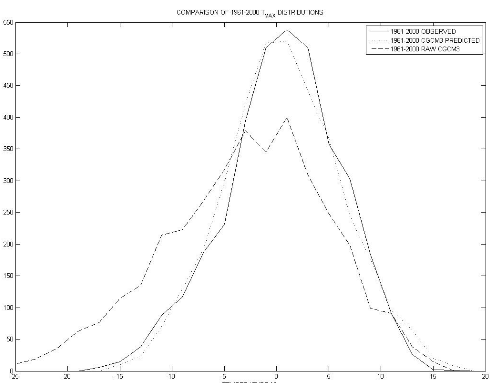

Fig. 4. Comparison of the 1961-2000 observed distribution for TMAX in winter VS the CGCM3 regression prediction VS the distribution taken directly from the CGCM3 gridbox over Halifax.

List of Tables

TABLE 1. The names of the original predictors and the names of the predictors not including wind direction, total windspeed or divergence are shown. Predictors the CGCM3 handles well in winter has a Y. The predictors not handled well by the CGCM3 have an N.

PREDICTORS	*WINTER*
mean sea level Pressure	Y
500hpa geopotential height	Y
500hpa zonal windspeed	Y
500hpa meridional windspeed	Y
500hpa vorticity	Y
850hpa geopotential height	Y
850hpa zonal windspeed	Y
850hpa meridional windspeed	Y
850hpa vorticity	Y
surface zonal windspeed	Y
surface meridional windspeed	Y
surface vorticity	Y
500hpa specific humidity	N
850hpa specific humidity	N
surface specific humidity	N
surface mean temperature	N

TABLE 2. A list of the final predictors used to create the PC's as determined in table 1 and their associated weights in the highest correlated principal component which has a correlation of 0.62 with TMAX.

PREDICTORS	*PC*1
mean sea level pressure	-0.17
500hpa geopotential height	0.14
500hpa zonal windspeed	-0.05
500hpa meridional windspeed	0.48
500hpa vorticity	-0.14
850hpa geopotential height	-0.01
850hpa zonal windspeed	-0.08
850hpa meridional windspeed	0.58
850hpa vorticity	0.10
surface zonal windspeed	-0.13
surface meridional windspeed	0.52
surface vorticity	0.24

Table 3. The mean and standard deviation of the distribution from the observed 1961-2000 period and all three projection periods.

$PERIOD$	$MEAN(celcius)$	$STD(celcius)$
OBSERVED (1961-2000)	0.64	5.48
Tri-decade1 (2011-2040)	1.42	5.48
Tri-decade2 (2041-2070)	2.64	5.52
Tri-decade3 (2071-2100)	4.08	5.18

Table 4. Information on the regression between the TMAX seasonal anomaly and the PC's of the NCEP predictor seasonal anomalies

Parameter	*Value*
Explained variance (percent)	79
Regression error variance	0.07
Number of predictors (PC's)	10
Gamma squared	0.21
Inflation factor (percent)	10