**Les Cahiers du GERAD** ISSN: 0711–2440

# Forecasting local warming: Missing data generation and future temperature prediction

A.M. Aboussalah
C. Neal

G–2016–76

October 2016
Revised: November 2016



**GERAD** HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

**Tél. : 514 340-6053**
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

# Forecasting local warming: Missing data generation and future temperature prediction

**Amine Mohamed Aboussalah** [a b c]
**Christopher Neal** [b c]

[a] *GERAD, Montréal (Québec) Canada, H3T 2A7*

[b] *Department of Mathematics and Industrial Engineering, Polytechnique Montréal (Québec) Canada, H3C 3A7*

[c] *Canada Excellence Research Chair in Data Science for Real-Time Decision-Making, Montréal (Québec) Canada, H3T 2A7*

amine-mohamed.aboussalah@polymtl.ca
christopher.neal@polymtl.ca

**October 2016**
Revised: November 2016

**Les Cahiers du GERAD**
**G–2016–76**

**Abstract:** Global warming is a much discussed topic as it sparks debate for shaping government policy and how humans should behave in reaction to climate change. Global warming can be considered at a local perspective when we look at temperature trends at an isolated region. In this work we aim to predict a local warming trend for Canada's capital city Ottawa, Ontario up to the year 2040 using optimization and machine learning techniques. We use data from the National Oceanic and Atmospheric Administration (NOAA)[1] which archives historical weather data from approximately 9000 weather stations from around the world [2]. Some of the datasets date back to 1955, however it is incomplete for a number of weather stations including Ottawa. In this work we first expand on a statistical based approach proposed by Robert J. Vanderbei to model local warming based on a previous day correlation. Then we present a forward algorithm which samples from the Laplace distribution to fill in the missing data. Lastly, we make predictions up to the year 2040 using the Neural Network toolbox within Statistica.

**Keywords:** Global warming, climate change, local warming, solar cycle, least absolute deviations, regression, least squares, median, linear programming, machine learning, neural networks

**Acknowledgments:** We would like to thank Professor Dominique Orban for advice with optimization techniques and Professor Doina Precup for assistance with machine learning concepts. Additionally we would like to express gratitude towards Professor Robert J. Vanderbei for some help during the process. This research was supported by the Canada Excellence Research Chair in Data Science for Real-Time Decision-Making at Polytechnique Montréal.

# 1 Introduction

## 1.1 Data

The data archived by the NOAA contains a number of different average daily measures from a weather station. This includes the average temperature, dew point, wind speed, etc. In this paper we consider only the daily average temperature for a weather station on a given date.


**Figure 1: Ottawa, Google Map 2016**

## 1.2 Local warming model formulation

We start by extending work done in [3] which proposes a statistical approach to model local warming temperature by considering three factors [4] : A linear trend, seasonal variations, and the 10.7 year solar cycle [5] given that the nature of data is a time series:

$$\begin{aligned} h_w(x_i) = w_0 + w_1 x_i + w_2 \cos(2\pi x_i/365.25) + w_3 \sin(2\pi x_i/365.25) \\ + w_4 \cos(2\pi x_i/10.7 \times 365.25) + w_5 \sin(2\pi x_i/10.7 \times 365.25), \end{aligned} \tag{1}$$

where:

- $x_i$ is the day,
- $h_w(x_i)$ is the temperature estimated for a given day $x_i$,
- $w_i$ are the weights associated with the model,
- $w_0 + w_1 x_i$ is the trend (linear long term direction),
- $w_2 \cos(2\pi x_i/365.25) + w_3 \sin(2\pi x_i/365.25)$ is the seasonal variation effect,
- $w_4 \cos(2\pi x_i/10.7 \times 365.25) + w_5 \sin(2\pi x_i/10.7 \times 365.25)$ is the cyclical effect (11-year solar cycle).

Note in [3] the author tried to add an additional weight coefficient in the original model for the solar cycle in order to constrain the solar cycle period. The solar cycle is known to be around 10.7 years [5] and models that produce an erroneous coefficient for the solar cycle period can be eliminated. In [3] it is concluded that using the $L_2$ norm (leads to a least-squares problem) provides unrealistic coefficient for the solar cycle period. From this we conclude that the $L_2$ norm is not a good metric for this formulation. This is because the $L_2$ norm takes the average as oppose to the median which results in a higher sensitivity to outliers. This motivates the use of the least-absolute deviations (LAD) norm as a metric for our formulation [7].

In [3], the daily average temperatures from the McGuire Air Force Base near Princeton, New Jersey were used to validate the effectiveness of the model by reformulating it into an optimization problem seeking the coefficients that minimize the difference between the observed temperatures and the temperature model equation in (1):

$$\min_{w} \sum_{i \in D} \mid h_w(x_i) - T_i \mid . \tag{2}$$

We extend the model to include correlation terms with previous days (a correlation length of 7 days was chosen, i.e. the temperature of a given day depends on the temperatures of the last seven days) as well as the evaluation of the nature of the noise present in the data. To consider this dependence, we chose to use the autoregressive model (AR model) to model the average temperature considered as a random process and output a set of correlated estimated temperatures. Thus we can represent our output variable (temperature) as a linear combination of temperatures from the last seven days along with a stochastic modeling term representing (white) noise:

$$\tilde{T_i} = \sum_{k=1}^{7} \alpha_k T_{i-k} + \epsilon. \tag{3}$$

Once this correlation is modeled, we solve an optimization problem to determine the optimal correlation weights for this problem (see **Appendix A.1** for AMPL implementation[6]).

In order to proceed with the LAD norm we used the Laplacian noise distribution for our model. Analysis of the noise present in the data will be done later in Section 3.2 and the variables $\alpha_k$ have been positively constrained:

$$\begin{aligned} &\min_{\alpha_1,\ldots,\alpha_7} \sum_{i \in D} \mid \sum_{k=1}^{7} \alpha_k T_{i-k} - T_i \mid, \\ &\text{s.t.} \qquad \alpha_k \geq 0, \ \forall k \in \{1, ..., 7\} \end{aligned} \tag{4}$$

where $T_i$ is the observed temperature for a given day. $D$ is the set of all the indexes of dates. Note that the indexing set D starts from the $8^{\text{th}}$ day to take into account the one week correlation length.

However by solving this linear programming problem to get weight coefficients ($\alpha_k$) associated with the previous 7 days for an observed temperature, we found that all of the weights positioned only around the first day's correlation:

$$\begin{aligned} \alpha_1 &= \quad 0.999999, \\ \alpha_2 &= \quad 7.12039 \times 10^{-10}, \\ \alpha_3 &= -2.64918 \times 10^{-10}, \\ \alpha_4 &= \quad 3.206 \times 10^{-10}, \\ \alpha_5 &= -2.96672 \times 10^{-10}, \\ \alpha_6 &= -2.94753 \times 10^{-10}, \\ \alpha_7 &= -3.95935 \times 10^{-10}. \end{aligned}$$

As a result we decided to consider only a one previous day correlation and our model becomes:

$$\begin{aligned} h_w(x_i) = w_0 + w_1 x_i + w_2 \cos(2\pi x_i/365.25) + w_3 \sin(2\pi x_i/365.25) \\ + w_4 \cos(2\pi x_i/10.7 \times 365.25) + w_5 \sin(2\pi x_i/10.7 \times 365.25) + w_6 y_{i-1}, \end{aligned} \tag{5}$$

where:

- $x_i$ is the day,
- $w_0 + w_1 x_i$ is the trend (linear long term direction),
- $w_2 \cos(2\pi x_i/365.25) + w_3 \sin(2\pi x_i/365.25)$ is the seasonal variation effect,

- $w_4 \cos(2\pi x_i/10.7 \times 365.25) + w_5 \sin(2\pi x_i/10.7 \times 365.25)$ is the cyclical effect (11-year solar cycle),
- $w_6 y_{i-1}$ is the one day correlation term.

The next step is to find the value of the unknown regression coefficients $w_0$, $w_1$, . . . , $w_5$, $w_6$ which minimize the sum of the absolute deviations.

$$\min_{w_0,w_1,\ldots,w_5,w_6} \sum_{i\in D} \mid w_0 + w_1 x_i + w_2 \cos(2\pi x_i/365.25) + w_3 \sin(2\pi x_i/365.25) + w_4 \cos(2\pi x_i/(10.7 \times 365.25)) + w_5 \sin(2\pi x_i/(10.7 \times 365.25)) + w_6 y_{i-1} - T_i \mid . \tag{6}$$

We model the mathematical programming problem, expressed in equation (5) using the AMPL modeling language, in **Appendix A.2**. AMPL models, with their associated user-supplied data sets, can be solved online using the Network Enabled Optimization Server (NEOS) at Argonne National Labs [8].

**Results:** The linear programming problem can be solved (using KNITRO solver) in only a few minutes on a modern laptop computer and the optimal values of the parameters together are:

$$\begin{aligned} w_0 &= 42.6673\,^\circ F, \\ w_1 &= 1.7008 \times 10^{-4}\,^\circ F/day, \\ w_2 &= 25.3573\,^\circ F, \\ w_3 &= 9.0692\,^\circ F, \\ w_4 &= 0.2359\,^\circ F, \\ w_5 &= 0.1310\,^\circ F, \\ w_6 &= 6.3811\ \times 10^{-8}. \end{aligned}$$

We notice that the seasonal variation is the most dominant effect.

Below is our prediction model using the hypothesis $h_w$ in which we add a Laplacian noise term $\epsilon_i$:

$$y_i = h_w(x_i) + \epsilon_i, \tag{7}$$

where: $\epsilon_i \sim \mathcal{L}(0, \sigma^2)$ i.e $\epsilon_i = \frac{1}{2\lambda} e^{\frac{|\epsilon_i|}{\lambda}}$ with $\sigma^2 = 2\lambda^2$ is the variance.

# 2 Construction of the missing data

## 2.1 Issue of missing data

The initial step is to generate the missing temperatures for our dataset. In Figure 2 we see the average daily recorded temperatures along with the dates that are missing temperatures for Ottawa, Ontario.

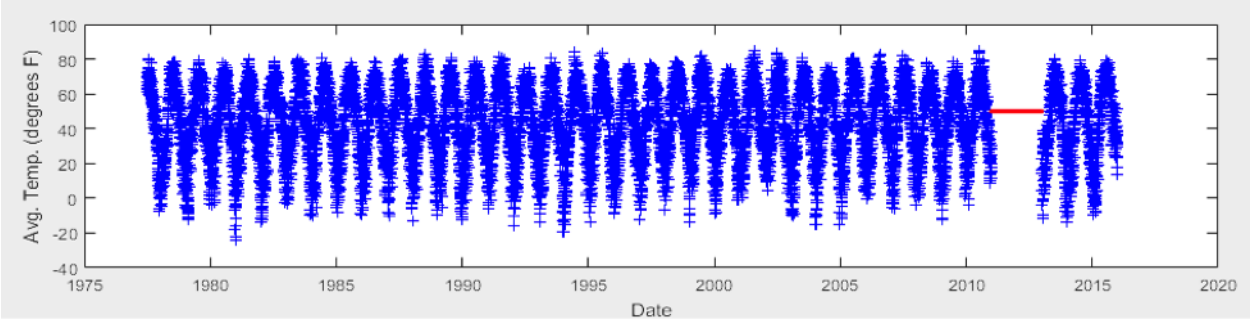



**Figure 2: Average daily temperatures (in °F) for Ottawa, Ontario with missing data**
Blue: Actual recorded temperatures
Red: Missing temperatures

As shown in equation (4) the correlation term $y_{i-1}$ is incorporated into the hypothesis $h_w(x_i)$. We can visualize this aspect as a 1st order Markov chain model (see Figure 3).

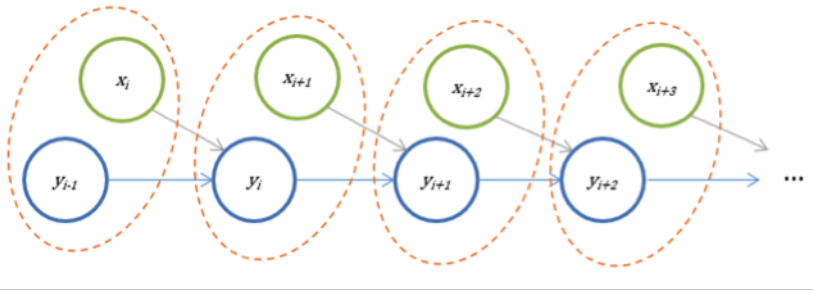



**Figure 3: Representation as 1st order Markov chain**

This representation allows us to assume that our data is independent and identically distributed ($i.i.d$).

## 2.2 Parameter inference

In order to fill in the missing data we infer parameters (weights and variance) from the completed section of data (of size $m$) prior to the missing section (in Figure 2, $m$ is the number of temperature instances between 1977-07-01 and 2011-01-01).

We build the likelihood of the completed data section $D$ given the parameters:

$$L(w,\lambda) = P(D|w,\lambda) = \prod_{i=1}^{m} \frac{1}{2\lambda} e^{\frac{-|y_i - h_w(x_i)|}{\lambda}} \tag{8}$$

By taking the logarithm, we obtain:

$$\log(L(w,\lambda)) = \sum_{i=1}^{m} \log(\frac{1}{2\lambda}) - \frac{1}{\lambda} \sum_{i=1}^{m} \mid y_i - h_w(x_i) \mid \tag{9}$$

We remark that the expression of the log likelihood depends on the parameter $\lambda$ and the weights $w_i$. The idea is to find these parameters which maximize the log likelihood. We determine these parameters in two steps using block coordinate ascent as follows:

### Block coordinate ascent

- *Step 1* : we set $\lambda$ to a constant value

$$\max_w \log(L(w,\lambda)) \Longleftrightarrow \min_w \sum_{i=1}^{m} \mid y_i - h_w(x_i) \mid \tag{10}$$

- *Step 2* : we set all $w_i$ found in step 1 to a constant value

$$\frac{\partial \log(L(w,\lambda))}{\partial \lambda} = 0 \Longleftrightarrow \lambda^* = \frac{1}{m} \sum_{i=1}^{m} \mid y_i - h_w(x_i) \mid \tag{11}$$

Once we have determined a value for $\lambda$ which maximizes the log likelihood ($\lambda^*$), we can deduce the variance of the prior existing data:

$$\sigma^{*2} = 2\lambda^{*2}$$

Now we can suppose that the missing data follows the same Laplace distribution with the mean[1] calculated by our model $h_w$ and the same variance as the prior existing data:

$$\begin{aligned} y_{i+1} &\sim \mathcal{L}(h_w(x_{i+1}), \sigma^2) \\ &\vdots \\ y_{i+k} &\sim \mathcal{L}(h_w(x_{i+k}), \sigma^2) \end{aligned} \tag{12}$$

Since we have a 1st order Markov chain representation there is a one-step memory in the model. This allows us to use a forward algorithm to sample the missing data. This works up to the last point where we need to ensure the points $y_{i+k}$ and $y_{i+k+1}$ are relatively connected. This is done by adding a weight coefficient $\delta$ so that the last term becomes:

$$y_{i+k} \sim \delta * \mathcal{L}(h_w(x_{i+k}), \sigma^2) \tag{13}$$

Where:

$$\delta = P(y_{i+k+1}|x_{i+k+1}, y_{i+k})$$

This last missing point $y_{i+k}$ is determined using rejection sampling: we reject generated points until we find a $\delta$ near to 1. This implies a small mismatch.

## 2.3 Implementation and missing data generation results

We use the AMPL modeling language to model the problem and send a request to use the KNITRO solver [9]. This is done within the Matlab environment using AMPL's Matlab API [10]. With these returned set of weights, we make a set of temperatures based on our model $h_w(x_i)$ between 1977-07-01 and 2011-01-01.

This set of points are then used in equation (10) to determine the best parameter $\lambda^*$ which gives an estimation of the variance $\sigma^2$ within the data between 1977-07-01 and 2011-01-01. The final step is to use our proposed one day correlation model to sample temperatures from the Laplace distribution, along with a forward algorithm to fill in the missing data.

The final outcome of these computations is a completed set of temperatures for this missing values in the original dataset .This leads to a complete set of temperatures from 1977-07-01 to 2015-12-31 (see Figure 4). This fits reasonably well, however a visual inspection may indicate an overestimation of the values.

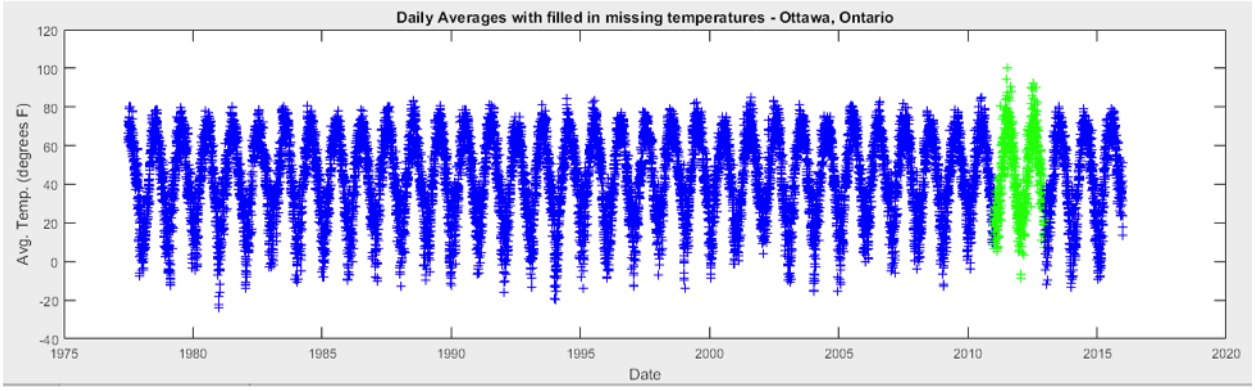



**Figure 4: Average daily temperatures (in ° F) for Ottawa, Ontario with completed data**
Blue: Actual recorded temperatures
Green: Generated Missing temperatures

[1]The Laplace location is the same as its mean and median.

# 3 Future temperature prediction

## 3.1 Building predictive models

To predict the evolution of the temperature in the Capital Region of Canada, Ottawa, we used artificial neural networks. The two types most commonly used in the family of "feedforward" algorithms are multilayer perceptrons (MLPs) and radial basis function (RBFs). The only fundamental difference between the two is the way which the units of the intermediate latent layers combine the values from the previous layers in the network. The MLPs use scalar products, while working RBFs Euclidean distance.

In general, the MLP is the most common structure. It requires iterations in the training phase, its structure is more compact, it runs quickly once trained, and in most problems performs better compared to other types of neural networks. The RBF structure tends to be larger compared to that of MLP and often does not perform as well. RBF can be trained faster than MLP in the case where there is a linear output of activation functions and contains large amounts of data.

We decided to train and test four types of MLP and 4 types of RBF networks by only modifying the number of neurons in the intermediate layer. Other parameter settings were not changed such as: the optimization algorithm used (sum of squares), hiddent and output activation functions. For the 4 MLPs, we decided on the Broyden-Fletcher-Goldfarb-Shanno (BFGS) as an optimization algorithm, the hyperbolic tangent function as the hidden activation function and the activation identity as the output function. For 4 RBFs we used the radial basis function training algorithm (RBFT), the Gaussian function for the hidden activation function, and the identity function for the output activation. The goal is to choose the structure having the best performance. In Table 1 we present the composition of the neural networks and in Table 2 we present their performance.

**Table 1: Neural Net model composition**

| Number of NN | Net. name | Training algorithm | Error function | Hidden activation | Output activation |
|---|---|---|---|---|---|
| 1 | MLP 4-2-1 | BFGS 70 | SOS[2] | Tanh | Identity |
| 2 | MLP 4-3-1 | BFGS 33 | SOS | Tanh | Identity |
| 3 | MLP 4-4-1 | BFGS 73 | SOS | Tanh | Identity |
| 4 | MLP 4-5-1 | BFGS 50 | SOS | Tanh | Identity |
| 5 | RBF 4-9-1 | RBFT | SOS | Gaussian | Identity |
| 6 | RBF 4-10-1 | RBFT | SOS | Gaussian | Identity |
| 7 | RBF 4-11-1 | RBFT | SOS | Gaussian | Identity |
| 8 | RBF 4-11-1 | RBFT | SOS | Gaussian | Identity |

**Table 2: Neural Net model performance**

| Number of NN | Net. name | Training perf. | Test perf. | Validation perf. | Training error | Test error | Validation error |
|---|---|---|---|---|---|---|---|
| 1 | MLP 4-2-1 | 0.985948 | 0.992269 | 0.983482 | 3.997897 | 2.217925 | 5.660753 |
| 2 | MLP 4-3-1 | 0.985601 | 0.990045 | 0.984933 | 4.099442 | 2.761998 | 5.101899 |
| 3 | MLP 4-4-1 | 0.986886 | 0.991260 | 0.982900 | 3.740690 | 2.464385 | 5.784824 |
| 4 | MLP 4-5-1 | 0.987486 | 0.990058 | 0.985032 | 3.564053 | 2.645464 | 5.056469 |
| 5 | RBF 4-9-1 | 0.926400 | 0.921397 | 0.936095 | 20.34604 | 20.53535 | 29.37388 |
| 6 | RBF 4-10-1 | 0.969780 | 0.953815 | 0.978785 | 8.534303 | 12.35420 | 7.300723 |
| 7 | RBF 4-11-1 | 0.987006 | 0.987334 | 0.985995 | 3.698330 | 3.249343 | 4.928045 |
| 8 | RBF 4-11-1 | 0.959391 | 0.927318 | 0.942155 | 11.40842 | 18.75901 | 19.53311 |

Figure 5 and Figure 6 show the performance of the different networks (in logarithmic scale) showing the change in the training and test errors at each given iteration:

[2]Sum-Of-Squares

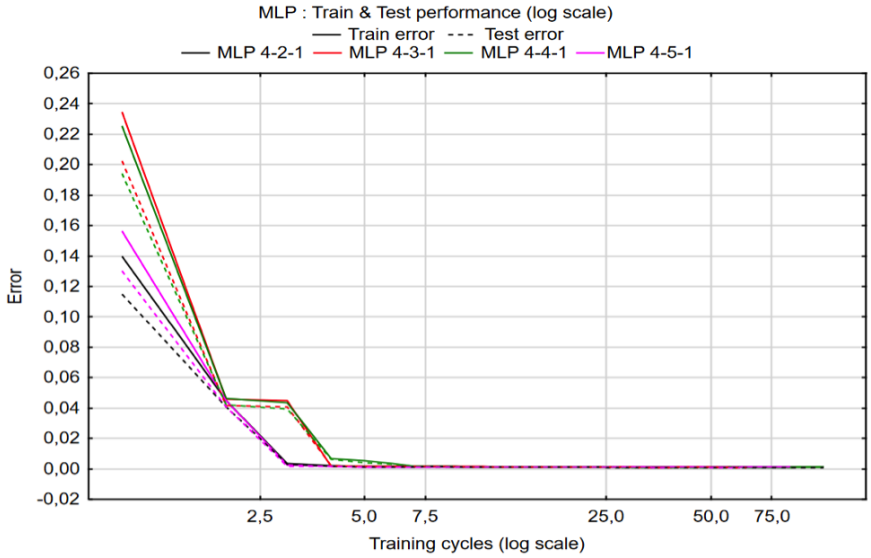



**Figure 5: MLP models performance**

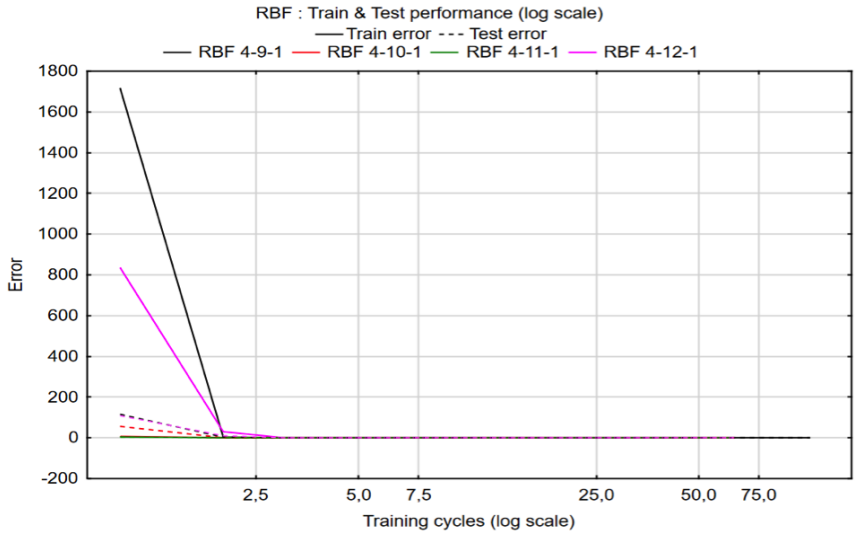



**Figure 6: RBF models performance**

The first point to mention by looking at the scale of the y-axis, is that the error of the RBF models is higher than that of the MLP models. This confirms the remark mentioned above. The best performance is given by the 4-2-1 MLP regarding the MLPs models, and RBF 4-11-1 of RBFs models. With a completed set of data we can predict future temperatures. An initial pre-processing step is to smooth-out the data into 101 day moving averages (like in [3]) in order to reduce the extreme fluctuations which will allow us to see the linear warming trend quite more easily.

Now that we have selected the best model for both types of neural networks, we will deploy and use them as predictive models to predict the evolution of the temperature in the Ottawa area, shown in Figure 7 and Figure 8.

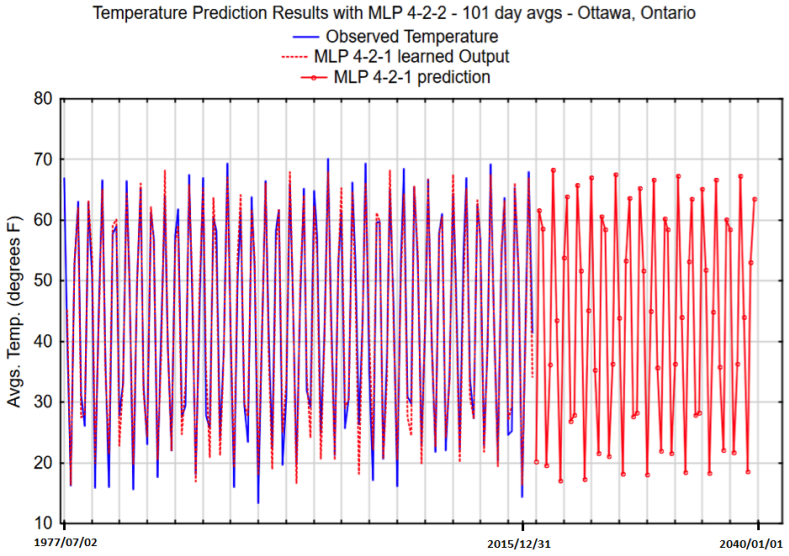



**Figure 7: Temperature prediction with MLP 4-2-2 until 2040**

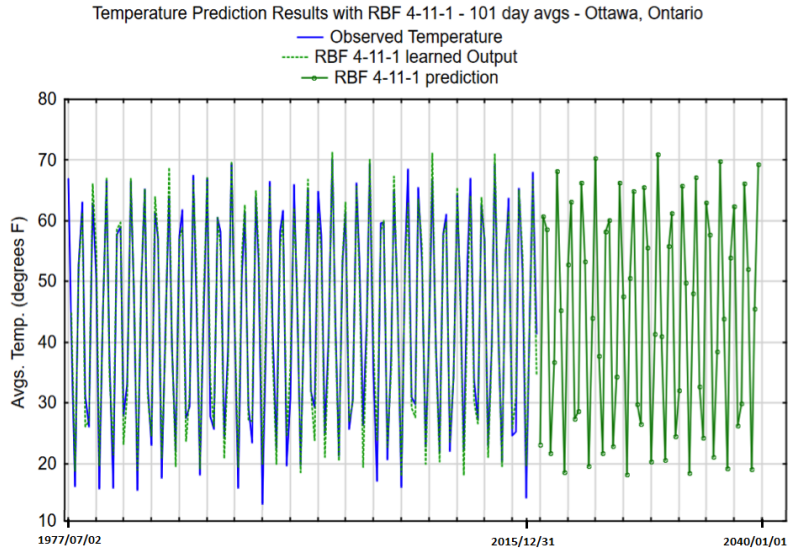


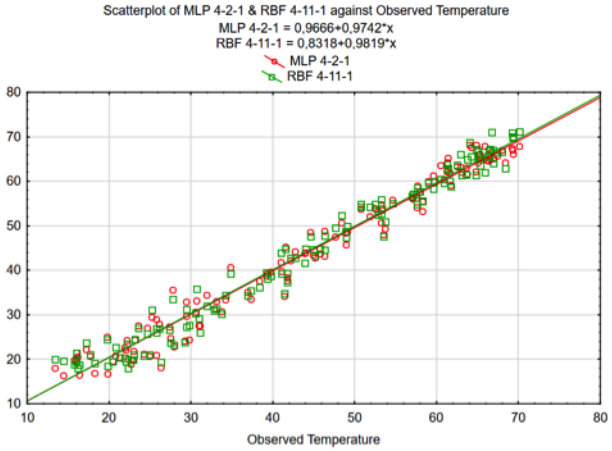

**Figure 8: Temperature prediction with RBF 4-11-1 until 2040**

## 3.2 Noise analysis and reliability of the models

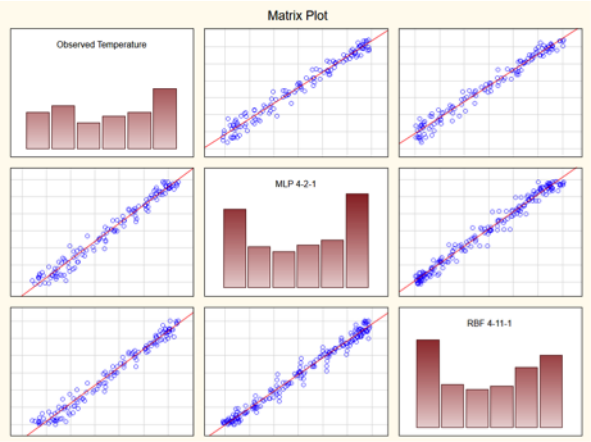
Predictions are prone to quickly accumulate errors, thus multi-step projections are reliable only if the predictive capacity of the network can be shown to be accurate. Even in these cases, trying to predict a large number of forward projections is risky as the overall accuracy of a network vastly declines as the length of projection increases.

The correlation coefficient between the target data and the predictions (outputs) produced by the network can be used as a performance measure of the network. A correlation coefficient can have any value between -1 and 1, where 1 represents a perfect fit. Having a correlation coefficient too close to 1 on the training samples isn't necessarily preferred since the target data not only holds actual measurements, but also an amount

of additive noise. This correlation value of 1 would imply that overfitting occurred. Overfitting is when a network finds values to perform very well on training data creating a maximum correlation coefficient. This will generally cause the network to achieve poor results on testing and validation samples. As a result, when selecting a network, the test and validation correlation coefficients should be used. Having a low correlation coefficient from the training phase does not implicitly mean that the network was poorly trained. It can be an indication of a network being implemented conservatively, so that it takes steps to avoid too closely fitting the noise. The below scatterplot (Figure 9a) shows the observed and predicted values for the target variable:




(a) Scatterplot of MLP 4-2-2 and RBF 4-11-1




(b) Matrix plot of MLP 4-2-2 and RBF 4-11-1

**Figure 9**

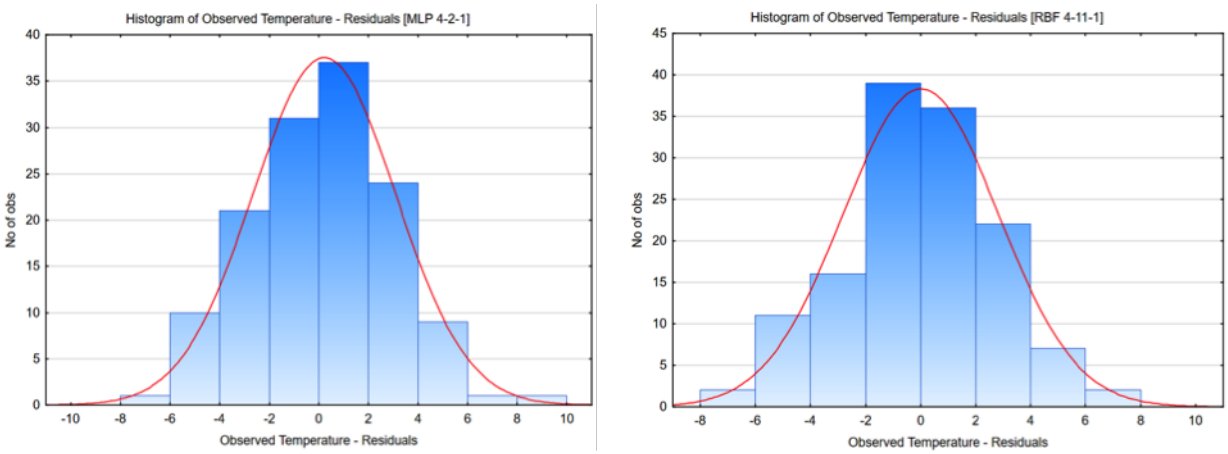
These graphs can be used to visually validate the effectiveness of a target in comparison to the output of the network by looking at how near is the graph relative to the line at 45-degrees also in the scatterplot. This graph serves as a visualization of how well the correlation coefficient performs. Many of the points on the scatterplot for the given networks do not lie precisely along the 45-degree line because noise on the objective values have not been modeled as actual signals, as is desired. In the matrix plot (Figure 9b) we only see positive correlation values.

### 3.2.1 Residual diagnosis

#### 3.2.1.1 Histogram residuals

It is important to analyze the dispersal of the network residuals of the observed temperature (i.e. the difference between predicted and target values).

Analyzing the histograms indicate the residuals are near the normal distribution and a mean of zero. This is a good sign showing that the network has taken into consideration the noise model. Larger variance in the histogram indicates more noise. A histogram with a width thinner than the actual variance is a sign of overfitting. Note that the amount of noise (i.e. the variance) cannot be known ahead of time.



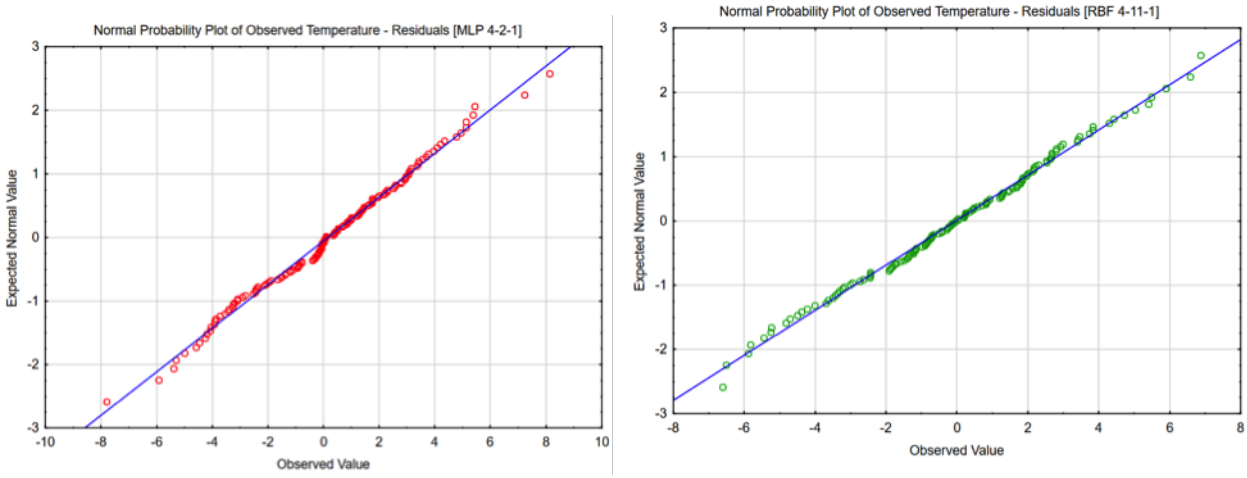

(a) Histogram of MLP 4-2-2

(b) Histogram of RBF 4-11-1

**Figure 10**

#### 3.2.1.2 Normal plot of residuals

An assumption for multiple regressions is that the observed values minus the predicted values (residual values) follow the normal distribution. Further, the regression function is linear. Should any of these assumptions be violated enough, the regression coefficients (B coefficients) will likely be altered (deflated or inflated) causing the statistical significance tests to be skewed. A normal distribution of the residual values is a good sign. A quick way to visually check the degree to which the residual patterns follow a normal distribution is through normal probability plots. Residuals will deviate from the line if the residuals do not follow a normal distribution. This plot can also help identify outliers. When there is not a good fit and if the data shows a clear pattern (e.g. S shape) in relation to the line, there may have been a transformation on the dependent variable.




(a) Normal Probability Plot of MLP 4-2-2

(b) Normal Probability Plot of RBF 4-11-1

**Figure 11**

It is clear that our two neural network models were able to detect noise in the data. This demonstrates the robustness of the two neural network models when faced with the noise present in the observed data to ensure good predictive performance of the temperature.

## 3.3 Longterm forecasted trends

A final result is to look at a longterm trend in the forecasted temperatures. The results from Part 1 show that the seasonal variation effect is the most dominant one. In order to see linear trends, we have to ignore in equation (5) the seasonal variation effect and the 11-year solar cycle as well. Figure 12 shows the linear trends for our two predictive models. In both cases we see an upward trend suggesting the occurrence of local warming in Canada's capital city Ottawa. The RBF linear trend produces a more modest increase in comparison with the MLP linear trend:

**Table 3: Starting and ending temperatures for longterm trends**

| | Initial Temperature (1977) | Final Temperature (2040) |
|---|---|---|
| **MLP 4-2-1** | 43.1074 °F | 44.0228 °F |
| **RBF 4-11-1** | 43.1141 °F | 44.5571 °F |

It is interesting to note the occurrence of local warming since in all cases the slope shows a positive trend. The linear trend coming from the observed temperature in the data set indicates a coefficient $w_1$ of 0.0117 °F/101day, which scales to 4.23 °F per century. This is similar to other global climate change models which estimate a per century climate rate of warming between 2.0 and 2.4 °C (see [11],[12]). However, the linear trend results coming from the two neural network models MLP 4-2-1 and RBF 4-11-1 generates a more modest upward local warming of 1.68 °F per century and 2.42 °F per century respectively. This difference can be explained by the propagation of errors over the years. The overall accuracy of the models vastly declines as the length of projection increases in time and therefore the predictive models fail to properly capture the hidden trends. Additionally, this data set starts at 1977, by running on a larger dataset we may achieve better performance learning and therefore better predicted values.

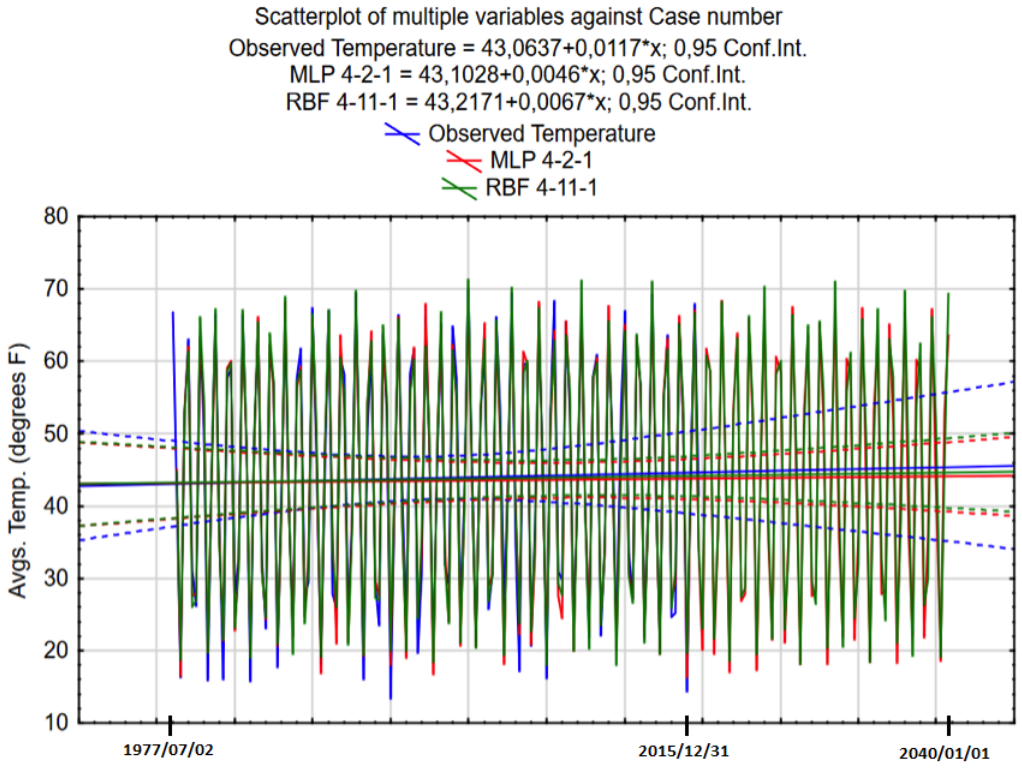



**Figure 12: Linear trend temperature predictions**

# 4 Conclusion

The results obtained demonstrate that our proposed method does in forecast a local warming trend for Ottawa, Ontario up to the year 2040. This is done by using existing recorded daily average temperatures from 1977-07-01 to 2010-12-31 to generate missing data from 2011-01-01 to 2012-12-31. Then the completed set of data from 1977-07-01 to 2015-12-31 is used to perform the learning phase to generate a longterm forecast up to the year 2040. The work presented here did not evaluate the overall performance of the generated missing data. The missing temperature generation could be quantitatively validated by producing a subset of temperatures on a dataset which has complete data and comparing the difference. This approach could be evaluated against another method which generates this type of data. For example, one could try the Expectation-Maximization (EM) algorithm which performs a forward- backward approach instead of only a forward algorithm (as done in this work).

In [3] the author suggests a few of ways to improve the model which we did not build upon here. Firstly, that the sinusoidal seasonal variation is an approximation that does not work for data in tropical locales. Secondly, that the linear trend could also be modelled via a function of of population density. Lastly, that the longterm trend is likely close to being linear, but it is definitely not exactly linear. This work could also be extended by adding a correlation with other locations which are geographically nearby. This could be used to get a local warming trend over a larger area and perhaps be further generalized to infer a global warming trend. The results obtained here could eventually be compared with other models of global warming to get a better understanding of the phenomena.

# Appendices

## Appendix A AMPL modeling language

### A.1 Optimal correlation weights

```
set DATES ordered;
param avg {DATES};
param day {DATES};
param p1 {DATES};
param p2 {DATES};
param p3 {DATES};
param p4 {DATES};
param p5 {DATES};
param p6 {DATES};
param p7 {DATES};

var y {j in 1..7};
var dev2 {DATES} >= 0, := 1;

minimize sumdev: sum {d in DATES} dev2[d];
subject to def_pos_dev {d in DATES}:
c[1]*p1[d] +
c[2]*p2[d] +
c[3]*p3[d] +
c[4]*p4[d] +
c[5]*p5[d] +
c[6]*p6[d] +
c[7]*p7[d] - avg[d]
<= dev2[d];
subject to def_neg_dev {d in DATES}:
-dev2[d] <=
c[1]*p1[d] +
c[2]*p2[d] +
c[3]*p3[d] +
c[4]*p4[d] +
c[5]*p5[d] +
c[6]*p6[d] +
c[7]*p7[d] - avg[d];

subject to cons1: c[1] >= 0;
subject to cons2: c[2] >= 0;
subject to cons3: c[3] >= 0;
subject to cons4: c[4] >= 0;
subject to cons5: c[5] >= 0;
subject to cons6: c[6] >= 0;
subject to cons7: c[7] >= 0;

data;

set DATES := include "data\McGuireAFB1955_2015_dates2.dat";
```

```
param: avg := include "data\Ottawa_1977_2015.dat";
param: p1 := include "data\predictedQuadLAD\predict1.dat";
param: p2 := include "data\predictedQuadLAD\predict2.dat";
param: p3 := include "data\predictedQuadLAD\predict3.dat";
param: p4 := include "data\predictedQuadLAD\predict4.dat";
param: p5 := include "data\predictedQuadLAD\predict5.dat";
param: p6 := include "data\predictedQuadLAD\predict6.dat";
param: p7 := include "data\predictedQuadLAD\predict7.dat";

#param: predicted := include "data\data\Ottawa_1977_2015_predicted_withDates.dat";
#param: predicted  := include "data\data\Ottawa_1977_2015_2.dat";

let {d in DATES} day[d] := ord(d,DATES);

option solver "data\ampl\loqo\loqo\loqo";
solve;

display c[1], c[2], c[3], c[4], c[5], c[6], c[7]
```

## A.2 One day correlation model

```
set DATES ordered;
param avg {DATES};
param day {DATES};
param prev {DATES};
param genT {DATES};
param pi := 4*atan(1);

var x {j in 0..6};
var dev {DATES} >= 0, := 1;

minimize sumdev: sum {d in DATES} dev[d];
subject to def_pos_dev {d in DATES}:
w[0] + w[1]*day[d]
+ w[2]*cos( 2*pi*day[d]/365.25)
+ w[3]*sin( 2*pi*day[d]/365.25)
+ w[4]*cos( 2*pi*day[d]/(10.7*365.25))
+ w[5]*sin( 2*pi*day[d]/(10.7*365.25))
+ w[6]*prev[d]
- avg[d]
<= dev[d];

subject to def_neg_dev {d in DATES}:
-dev[d] <=
w[0] + w[1]*day[d] + w[2]*cos( 2*pi*day[d]/365.25)
+ w[3]*sin( 2*pi*day[d]/365.25)
+ w[4]*cos( 2*pi*day[d]/(10.7*365.25))
+ w[5]*sin( 2*pi*day[d]/(10.7*365.25))
+ w[6]*prev[d]
- avg[d];

subject to cons1: x[6] >= 0;
subject to cons2: x[6] <= 1;
```

```
data;

set DATES := include "data\data\Ottawa_1977_2010_step2.dat";
param: avg := include "data\data\Ottawa_1977_2010_step2.dat";
param: genT := include "data\data\Ottawa_1977_2010_generated.dat";
param: prev := include "data\data\Ottawa_1977_2010_previous.dat";

let {d in DATES} day[d] := ord(d,DATES);
#Initialization
let w[0] := 60;
let w[1] := 0;
let w[2] := 20;
let w[3] := 20;
let w[4] := 0.01;
let w[5] := 0.01;

option solver "data\ampl\loqo\loqo\loqo";
solve;

display w[0], w[1], w[2], w[3], w[4], w[5], w[6]
```

## References


[1] NOAA. Climate data format and download instruction, 2011. ftp://ftp.ncdc.noaa.gov/pub/data/gsod/readme.txt

[2] NOAA. List of weather stations, 2011. ftp://ftp.ncdc.noaa.gov/pub/data/gsod/ish-history.txt

[3] R. J. Vanderbei. Local warming. SIAM Review, 2012.

[4] Australian Bureau of Statistics. Time series analyis: The basics. http://www.abs.gov.au/websitedbs/D3310114.nsf/home/TimeSeriesAnalysis:TheBasics

[5] R.C. Willson and H.S. Hudson. The sun's luminosity over a complete solar cycle. Nature, 1991.

[6] D. Gay R. Fourer and B. Kernighan. Ampl: A modeling language for mathematical programming. Duxbury Press, Belmont, CA, 1993.

[7] Y. Li and G. R. Arce. A maximum likelihood approach to least absolute deviation regression. EURASIP Journal on Applied Signal Processing, 2004.

[8] M. Mesnier J. Czyzyk and J.J. Moré. The neos server. IEEE Computational Science & Engineering, 1998.

[9] Knitro for AMPL. http://ampl.com/products/solvers/solvers-we-sell/knitro/

[10] AMPL API. http://ampl.com/products/api

[11] D. Griggs J. Houghton, L. M. Filho and K. Maskell. An introduction to simple climate models used in the IPCC second assessment report, IPCC technical paper ii. Intergovernmental Panel on Climate Change, 1997.

[12] R. Knight T. Karl and B. Baker. The record breaking global temperatures of 1997 and 1998: Evidence for an increase in the rate of global warming? Geophysical Research Letters, 2000.