### Assessing Climate Impacts on Regional Water Resources in the Midwestern US

REU Site: Interdisciplinary Program in High Performance Computing

Joseph U. Emelike<sup>1</sup>, David Harper<sup>2</sup>, Charlotte Z. Mann<sup>3</sup>, Kwame Owusu-Boaitey<sup>4</sup>, Graduate assistant: Sai K. Popuri <sup>5</sup>, Faculty mentor: Nagaraj K. Neerchal<sup>5</sup>, Client: Amita Mehta<sup>6</sup>

<sup>1</sup>Department of Computer Science, Bowie State University,

<sup>2</sup>College of Engineering and Mathematical Sciences, The University of Vermont, <sup>3</sup>Department of Mathematics and Statistics, Carleton College,

<sup>4</sup>Department of Computer Science and Electrical Engineering, UMBC,

partment of Computer Science and Electrical Engineering, OMDC

<sup>5</sup>Department of Mathematics and Statistics, UMBC,

<sup>6</sup>Joint Center for Earth Systems Technology (JCET)

Technical Report HPCF-2015-22, hpcf.umbc.edu > Publications

#### Abstract

It is well documented that decadal climate variability (DCV) has a significant impact on water resources in the Missouri River Basin (MRB). This project aims to utilize multi-decadal simulations of Global Climate Models (GCM) from the Climate Model Inter-comparison Project (CMIP5) to assess the DCV impact on water yield and streamflow over the MRB using a widely utilized hydrology and crop model known as the Soil and Water Assessment Tool (SWAT). We use low-resolution (100km x 100km) data from MIROC5 and HadCM3 GCMs with 57 years of climate simulations at approximately 30,000 locations. The weather parameters included in the GCMs are monthly precipitation, maximum/minimum temperatures, sea-level pressure, relative humidity, and surface wind speed. We downscale all the parameters to match high resolution (12km x 12km) observed data using a two-step procedure. First, GCM-simulated weather parameters are spatially interpolated to the resolution of the observed data, and then multiple linear regression (MLR) is used to capture features of the observed data. The coefficients from regression are combined with hindcast data from the two GCMs to compute monthly predictions of maximum/minimum temperatures, and precipitation to input into SWAT. A Weather Generator tool in SWAT is used to generate the daily values necessary to input into SWAT using the monthly predictions and observed weather statistics. We modified a previously developed Graphical User Interface (GUI) in R to streamline the process and include more options for users. We explore if the use of different GCMs and the additional weather parameters in the regression models improve the accuracy of predicting the above-mentioned variables in the MRB. The procedures and GUI developed in this project will allow the client to conduct numerous studies with improved efficiency to assess sensitivity of water resources within the MRB resulting from climate variability and change scenarios.

Key words. MRB, GCM, GUI, SWAT, Regression

AMS subject classifications (2010). 62J05, 62J20

## 1 Introduction

The Missouri River Basin (MRB) is the largest river basin in the United States, covering more than 500,000 square miles and including parts of ten U.S. states, which amounts to about one-sixth of the United States. The MRB is a very important agricultural region, with 12% of all U.S. farms and 28% of all land used for farming in the basin [9]. Approximately 90% of the basin's cropland is not irrigated and is therefore entirely dependent on precipitation. As climate variability and change impact fresh water availability, and thus agricultural production, it is highly relevant to investigate expected climate conditions in the MRB, in order to develop strategies to manage production in this essential agricultural region.

The Missouri River Basin is an excellent region to understand the effect of climate variability, particularly temperature and precipitation, on water and crop yields long-term, as quality observations are available for validation. In order to assess crop and water yields, data from Global Climate Models (GCM), downscaled to 12km x 12km grid resolution from the original 100km x 100km, are used in conjunction with a hydrology and crop model, the Soil Water Assessment Tool (SWAT), to predict crop and water yields in the MRB. The primary objective of this project is to create an end-to-end tool that streamlines the computational procedures to downscale data, input into SWAT and output understandable results.

In this project we focus on: (1) generating high-resolution weather parameters (temperature and precipitation) by downscaling GCM data for monthly averages of these parameters and inputting these parameters into SWAT, (2) improving the efficiency and accuracy of model fitting and forecasting, and (3) improving the previously developed Graphical User Interface (GUI) tool to perform calculations on data from either the HadCM3 or MIROC5 GCM data sets, input the downscaled data to SWAT, and visualize SWAT outputs.

The paper is organized as follows: Section 2 gives a detailed explanation of the background of the project and the work done by the UMBC HPCREU 2014 team. Section 3 describes the statistical and computing methods used in the project and outlines the results obtained. Lastly, Section 4 summarizes the findings and gives suggestions for future work.

### 2 Background

Natural climate variability and climate change in regional temperature, precipitation, winds and humidity have significant impacts on fresh water availability and agricultural output. In order to better assess these challenges posed by the changing climate conditions, a multiinstitute (CRCES, Texas A&M, UMBC-JCET, NDMC) project supported by the US Department of Agriculture-National Institute for Food and Agriculture (USDA-NIFA) was created to assess the impacts of decadal climate variability on water availability and crop yields in the Missouri River Basin [3]. The client-team from UMBC-JCET has used both daily and monthly low resolution data for precipitation (pr), maximum, and minimum temperature provided by the two Global Climate Models HadCM3 [7] and MIROC5 [11] to generate high resolution data needed to run SWAT [4].

Downscaling the daily maximum/minimum temperatures, and precipitation in MRB using MIROC5 data was previously studied and implemented by the 2014 UMBC REU team [3]. They have used the daily data to model and their predictions were also at the daily level. We, on the other hand, have used the data at the monthly level (monthly aggregated daily data) and SWAT's weather generator to produce predictions at the daily level to be used for further processing by SWAT. There is interest in using monthly averaged data as the downscaling is computationally easier, and faster. Also, simpler linear regression models can be used, especially for precipitation.

The UMBC-JCET team previously focused on the MIROC5 data [8], but it came to their attention that while the downscaled minimum and maximum temperature forecasts are close to the observed temperature, this is not true for precipitation. This observation lead the 2014 UMBC REU team to investigate ways to improve the quality of downscaling, and ended up using the Tobit regression model [2] in their project as it is shown to be more appropriate for mixed data, which daily precipitation is as it contains many zero values.

The 2014 REU team improved the quality of downscaling using MIROC5 data, parallelized the downscaling code, and created a Graphical User Interface (GUI) in R that could call functions to downscale data. However, the GUI did not run the parallelized downscaling and could not directly input data into SWAT to get crop yield estimates.

We are interested in further improving the quality of downscaling by adding additional weather parameters such as wind, sea level pressure, and average temperature to the regression models. We investigate whether using monthly averages rather than daily data can improve forecasting as monthly averaged data is not entirely mixed, and simpler linear regression forms than Tobit regression could be used. We also investigate if using monthly data as an input for SWAT from R is feasible. Finally, we add the ability to choose different GCM models in the tool, namely MIROC5, and HadCM3 and explore the accuracy of predictions from each model.

### 3 Methodology and Results

#### 3.1 Data

For our study, we consider daily observed data for 57 years (1949-2005) provided by Maurer [5] for the MRB. The model data being used comes from two different GCMs, MIROC5 and HadCM3. We use the following variables from the HadCM3 model data: precipitation, near surface relative humidity, sea-level pressure, wind speed, maximum temperature, minimum temperature, and near surface air temperature. These variables are at every  $3.75^{\circ}$ longitude-2.5° latitude, making the resolution of HadCM3 data 417 km longitude, 277.8 km latitude. The precipitation is recorded in  $\frac{kg/m^2}{s}$ , which is converted to  $\frac{mm}{day}$  by multiplying by a factor of 86,400. The near surface temperature, maximum temperature, and minimum temperature are measured in K (Kelvin). The near surface wind is measured in  $\frac{m}{s}$  and the sea-level pressure is measured in pascals (Pa). Lastly, the relative humidity is recorded as %. HadCM3 simulated data is available covering the entire earth, spanning from 0° to 356.25° longitude and  $-90^{\circ}$  to 90° latitude. HadCM3 data follows a 360-day calendar whereas the observed data follows the regular calendar. Prior to analysis, HadCM3 data is aligned by repeating/discarding the last day(s) of each month depending on whether the month has 30 or 31 days, or for February, one or two values are discarded depending on whether the year

Variables	MIROC5	HadCM3
Maximum Temperature	Κ	K
Minimum Temperature	Κ	Κ
Near Surface Air Temperature	Κ	Κ
Precipitation	$\frac{mm}{day}$	$\frac{mm}{day}$
Sea-level Pressure	Pa	Pa
Wind Speed	$\frac{m}{sec}$	$\frac{m}{sec}$

Table 3.1: Variable Units by Model

is a leap year.

MIROC5 also provides data for the same set of variables, recorded in the same units. MIROC5 data provides values every .125° latitude, .125° longitude. The range of spatial data is from  $34.9375^{\circ}$  to  $50.0625^{\circ}$  latitude and from  $-120.0625^{\circ}$  to  $-84.9375^{\circ}$  longitude. MIROC5 follows a 365-day calendar, which means that the MIROC5 data does not account for leap years and thus has missing dates. In order to match the dates of MIROC5 data to the observed data, data for the last day in February for each leap year is repeated.

Both of these GCMs contain ensembles which are essentially separate forecasts from different simulations of the model for each location and time. Each ensemble contains daily forecasts across a time range. These ensembles are averaged to find an average value from the model for each location and time at the daily level. Next, the daily data is averaged by month, resulting in monthly averaged data for 57 years, or 684 values for each location.

#### 3.2 Downscaling

As the data from MIROC5 and HadCM3 are at a low resolution, of about a 100 km x 100 km resolution and a 400 km x 200 km resolution respectively, while the observed data is at a 12 km x 12 km resolution, the downscaling procedure involves two steps: spatial interpolation and regression.

In order to estimate values for weather parameters at locations where values are missing, two different methods of interpolation are used, bilinear interpolation, and krigging. Bilinear interpolation calculates weighted averages for locations within squares of locations that have known values, using Equation 3.1 [10]. Given a square of certain latitudes and longitudes, the corners at which we have data (in Figure 3.1 the red points), values within the square can be estimated (in Figure 3.1 the green point). Then, the process is repeated to estimate values between the interpolated points in order to calculate values for each location that is needed to be at the same resolution as the observed data.

$$f(x,y) = \frac{1}{(x_2 - x_1)(y_2 - y_1)} (f(x_1, y_1)(x_2 - x)(y_2 - y) + f(x_2, y_2)(x - x_1)(y_2 - y) + f(x_1, y_2)(x_2 - x)(y - y_1) + f(x_2, y_2)(x - x_1)(y - y_1))$$
(3.1)

Kriging also estimates values for locations that do not have known values using a weighted average from values at locations with known values. The weights are determined by a

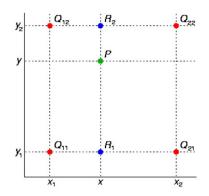


Figure 3.1: Schematic of bilinear interpolation [1]

semivariogram model, and thus spacial correlations between known values [6], rather than simply the distance between points, as in bilinear interpolation. Kriging is only used with the HadCM3 data as we did not have access to raw data for MIROC5, but only data that was already bilinearly interpolated, from the 2014 project.

Using the interpolated model data as the predictors and observed data as the response, linear regression is employed to capture the features of observed data in the forecasted values. Three separate models (Equations 3.2/3.3/3.4, 3.5, 3.6) are used with precipitation as the response to determine which model best predicts precipitation. Only Model 1 (Equations 3.2/3.3/3.4) is used to predict maximum temperature and minimum temperature. In Model 2 (Equation 3.5) average temperature is added to Model 1 as a covariate and in Model 3 (Equation 3.6) sea level pressure and surface wind are added as covariates.

$$y_{asi} = \beta_{0si} + \beta_{1si} x_{1asi} + \epsilon_{si} \tag{3.2}$$

$$y_{bsi} = \beta_{0si} + \beta_{1si} x_{1bsi} + \epsilon_{si} \tag{3.3}$$

$$y_{csi} = \beta_{0si} + \beta_{1si} x_{1csi} + \epsilon_{si} \tag{3.4}$$

$$y_{si} = \beta_{0si} + \beta_{1si}x_{1si} + \beta_{2si}x_{2si} + \epsilon_{si} \tag{3.5}$$

$$y_{si} = \beta_{0si} + \beta_{1si}x_{1si} + \beta_{2si}x_{2si} + \beta_{3si}x_{3si} + \beta_{4si}x_{4si} + \epsilon_{si}$$
(3.6)

Here,  $\{y_{asi}\}$  is the 57 × 1 vector of observed precipitation for the period 1949-2005,  $\{y_{bsi}\}$  is the 57 × 1 vector of observed maximum temperature for the period 1949-2005,  $\{y_{csi}\}$  is the 57 × 1 vector of observed minimum temperature for the period 1949-2005,  $\{x_{1asi}\}$  is the corresponding vector of model precipitation data,  $\{x_{1bsi}\}$  is the corresponding vector of model maximum temperature data,  $\{x_{1csi}\}$  is the corresponding vector of model minimum temperature data,  $\{x_{1csi}\}$  is the corresponding vector of model minimum temperature data,  $\{x_{2si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure data,  $\{x_{4si}\}$  is the corresponding vector of model sea-level pressure

Coefficients from the model for each latitude, longitude pair were combined with interpolated hindcast GCM data, for the years 1982-1990 to forecast precipitation, minimum temperature, and maximum temperature values to be inputted into SWAT for those years.

We used diagnostics such as plotting the residuals to ensure constant variance and that the residuals follow the expected N(0,1) distribution to initially check the fit of the models.

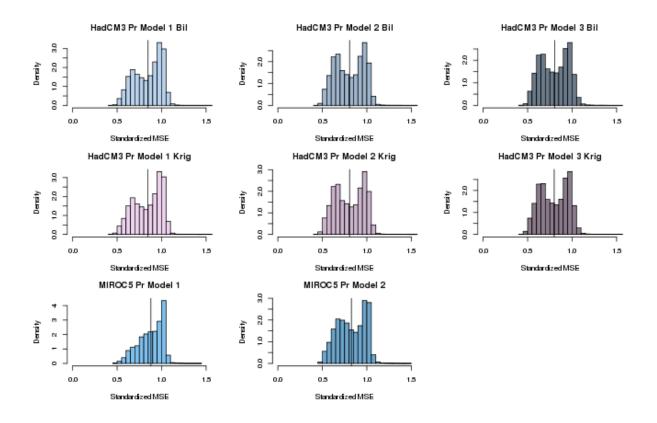


Figure 3.2: Distribution of standardized MSE for each precipitation model across all locations, with vertical line at the mean.

Next, in order to test the accuracy of the predictions and fit of the models, we calculated a standardized mean square error value defined in Equation 3.7. This is  $1 - R^2$ , thus, small values are desirable and values above one indicate that the model does a poor job of predicting precipitation or temperature.

$$\frac{SSE_s}{SS_{stot}} = \frac{\sum (y_{si} - \hat{y}_{si})^2}{\sum (y_{si} - \bar{y}_{si})^2}$$
(3.7)

Here,  $y_{si}$  is the observed value for month i,  $\hat{y}_{si}$  is the forecasted value for month i, and  $\bar{y}_{si}$  is the mean of all observed values at a given location s.

Figure 3.2 shows the distribution of standardized MSE for each of the models for precipitation for all locations, and Figure 3.3 shows the distribution of standardized MSE for each of the minimum and maximum temperature. Table 3.2 summarizes this information, displaying the mean standardized MSE for each model. For the SLR model, Model 1, predictions using HadCM3 are slightly more accurate than MIROC5, with a .03 lower standardized MSE on average. This is the same for Model 2. The standardized MSE value is improved with the predictions using Model 2 for both GCMs. Predicted values from Model 3 are also slightly more accurate than Model 2. The fit of the models using kriged HadCM3 data are essentially identical to those of the models using bilinearly interpolated data. The SLR model is able to predict minimum and maximum temperature with very high accuracy, as the standardized

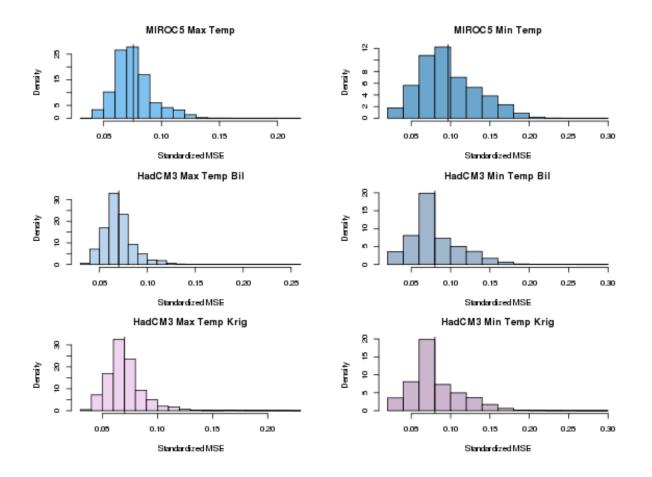


Figure 3.3: Distribution of standardized MSE for each temperature model across all locations, with vertical line at the mean.

MSE values are all below 0.1. As these results indicate, the models seem to perform better at predicting maximum, and minimum temperatures than for precipitation.

Model:	MIROC5	HadCM3 Bil	HadCM3 Krig
Pr Model 1	0.8793	0.8477	0.8452
Pr Model 2	0.8262	0.8063	0.8065
Pr Model 3	NA	0.8033	0.7987
Min Temp	0.0967	0.0800	0.0800
Max Temp	0.0757	0.0701	0.0702

We additionally explored how the fit of the models differed for the approximately 30,000 locations for which we forecasted. Figure 3.4 shows maps of the standardized MSE by location for predicted precipitation with Model 2 from MIROC5 data, predicted precipitation with Model 2 from HadCM3 data, predicted maximum temperature from MIROC5 data and predicted maximum temperature from HadCM3 data. Darker values depict lower, and thus better standardized MSE and better fitted models, while lighter values depict higher standardized MSE, meaning the model does not fit well at that location. The states outlined

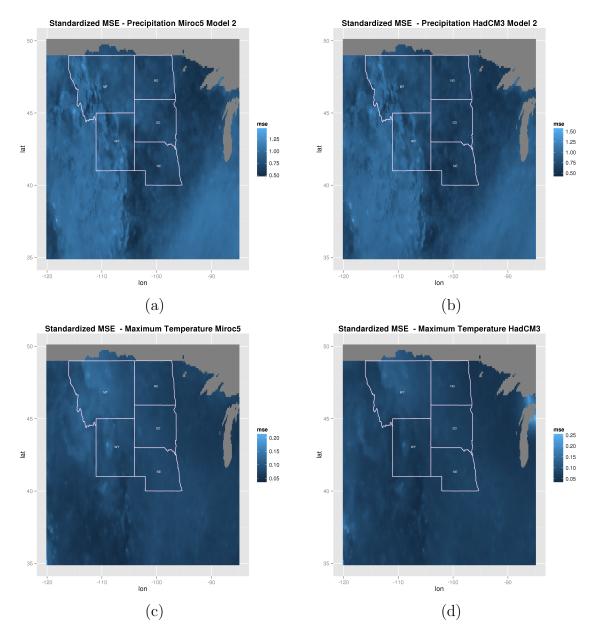
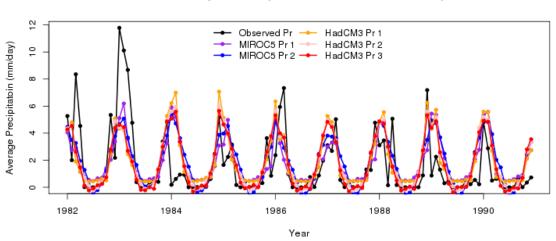


Figure 3.4: Plots of standardized MSE by location from (a) predicting precipitation by Model 2 with MIROC5 data, (b) predicting precipitation by Model 2 with HadCM3 data, (c) predicting maximum temperature with MIROC5 data, (d) predicting maximum temperature with HadCM3 data.

represent the main area of the Missouri River Basin (although parts of Colorado, Iowa, and other neighboring states are included in the basin's area).

We can see that Model 2 has more accurate predictions in the eastern half of this region, and does a poorer job in the western end of the region. This area is where the Rocky Mountains begin, so there is an opportunity for future research to explore how the geography of an area effects the ability of both Global Climate Models and the regression models used in this study to predict precipitation. The prediction of maximum temperature are more



Observed Monthy Ave Precipitation v. Forecasted Ave Precipitation

Figure 3.5: Plot of observed precipitation values and forecasted precipitation values for Model 1, 2, and 3 for MIROC5 and HadCM3

homogeneous across all locations, although there is a similar trend of the predictions being slightly less accurate in the western region of the MRB.

Figure 3.5 compares the forecasted values of precipitation for each month from each regression model for MIROC5 and HadCM3 (excluding Kriging), to the observed values of precipitation for each month for the years 1982-1990, at one location. The location used for Figure 3.5 is  $40.1875^{\circ}$  latitude,  $-102.6875^{\circ}$  longitude, which is in the upper-eastern corner of Colorado. We can see that the models are unable to predict sharp spikes in temperature, but rather follow a relatively regular oscillation through the months. Note, the forecasts generate some negative values, which does not actually make sense, but we decided to keep these values, rather then set them to zero. The five models produce similar forecasts, although they are not completely identical.

Finally, we were able to compare the accuracy of predicting precipitation using monthly averaged data with the predictions using daily data by the 2014 REU team. The standardized MSE calculated from the 2014 prediction, which used the SLR Model 1, is 0.77999999. Referring back to Table 3.2 the standardized MSE values for Model 1 generated with monthly data are 0.8793, 0.8477, and 0.8452. Thus, the predictions using SLR are more accurate when using daily data, but we are able to predict at almost the same accuracy by adding additional covariates (as in Model 2 and Model 3).

#### 3.3 SWAT Linkage

SWAT needs input data at the daily level. Since we predict precipitation, maximum/minimum temperatures at the monthly level, we need to generate predictions for these variables at the daily level to input into SWAT. In other words, we need to disaggregate monthly predictions

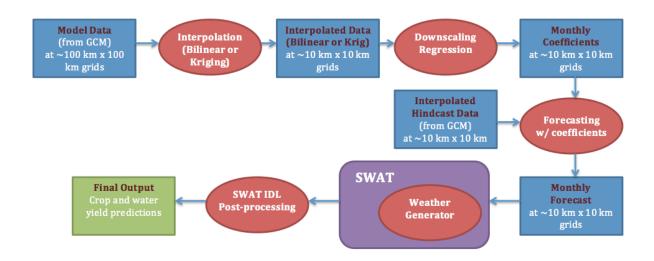


Figure 3.6: Methodology used to process data from raw model data to crop and water yields

to daily. We use SWATs Weather Generator to generate daily values using the monthly predictions produced by our method. SWAT can be used with multiple different weather generators as long as the necessary daily data is generated. SWATs default weather generator is the WXGEN weather generator model. The WXGEN weather generator model generates daily data from monthly data for precipitation, maximum half hour rainfall, solar radiation, maximum and minimum temperature, relative humidity, and wind speed. SWAT needs daily data for six variables: precipitation, maximum and minimum temperature, solar radiation, relative humidity, and wind speed. By using the weather generator, we capture the daily climate variability from the observed data while monthly variability is captured from our predictions. The daily data produced for precipitation, maximum and minimum temperature are combined with the observed daily data for the other three variables using a R procedures to generate .wgn files, which are then sent to SWAT as input.

Figure 3.6 illustrates the entire process of preparing the data to input into SWAT and get crop yield outputs from weather parameter data. The blue boxes represent the stages of the data, while the red ovals represent the methodology used to process each step of the data.

### **3.4** Graphical User Interface (GUI)

The Graphical User Interface (GUI), Figure 3.7, is a medium in which a user can communicate with and utilize a device. For our research, it was developed to output various statistical calculations including downscaling and regression on the given GCM data set. This interactive tool allows the user to streamline the process of downscaling GCM data, performing regression, forecasting, inputting the data into SWAT and visualizing the output. The different models can more quickly be compared to one another. Also, the modeling process was generalized so that it can be implemented for any GCM, whether it is from MIROC5 or HadCM3.

First, the GUI allows the user to choose which GCM data to use. Next, the user has the

Model: MIROC5		X		
Select variable type ) pr () tasmax () tasmin				
Downscale				
Select Regression method				
SLR				
○ MLR*	Start Downscale	Start Forecast		
*MLR is exclusive to pr				
MIROC5 data is interpolated exclusively with bilinear.				
Run SWAT from month  year  to month  year    Plot	Run SWAT	odel Make map		
Time series from				
year				
to				
year				
Plot Time Series				

Figure 3.7: GUI Screenshot

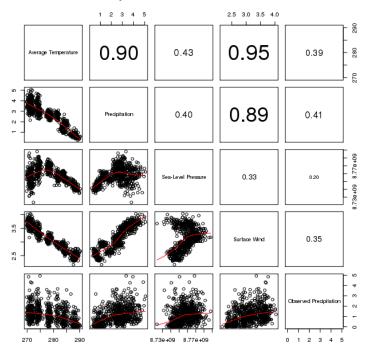
ability to pick the locations for which they wish to make calculations for. This option allows the user to cut down the run time if they are only interested in predictions for a certain area. The GUI loads the appropriate data which can then be downscaled from the chosen GCM. Before running any regressions, the user has the option to plot a correlation matrix of the variables that are used in the modeling for each GCM, as in Figure 3.8. This allows the user to better understand which model they may want to use, based on the correlations between the variables.

The GUI provides more than one option for interpolation as well as more than one option for regression modeling. For each month and each variable type, the user can use bilinear interpolation to generate high-resolution data, or Kriging to do the same (only for HadCM3). Furthermore, the user has the option to predict precipitation, minimum temperature, or maximum temperature, as well as which model they wish to use to predict precipitation.

Next, the user can use the downscaled coefficients created in the previous step to forecast their chosen weather parameter. Once the predictions are made, the user can visualize the predictions in relation to the observed data in the form of a time series plot.

# 4 Concluding Remarks

Our results show that predictions based on monthly data are similar to those based on daily data. Using monthly averaged data makes the computations simpler and much faster. It takes much less time to run the GUI using monthly data and as a result, there is not a need for a parallelization option in the GUI. The main reason is that using monthly data only has 684 data points for the 57 year time span compared to roughly 20,520 data points for the daily data. Using monthly data additionally allowed us to make predictions for the entirety of the 57 years, without having to divide the process by month. Our predictions for



HadCM3 Scatterplot Matrix for Variables and Correlations

Figure 3.8: GUI Screenshot

minimum temperature and maximum temperature were strong, yielding standardized MSE values less than 0.1 at many locations. However, the predictions were weak for precipitation, yielding high standardized MSE values of 0.8-0.9 at many locations. However, we did see an increase in prediction strength when adding covariates to the regression models, which is an improvement from previous regression models. Also, both GCM's offered similar overall predictions, showing that the user is not only limited to using one specific GCM. One last thing to note is that the prediction accuracy appears dependent on geographic features of the various locations. For example, the predictions were strongest at locations that are geographically flat, while the locations with the weakest predictions occur densely in the Rocky Mountains. Overall, it may prove beneficial to investigate further research in using different GCMs as well as exploring the impact of geographic characteristics and possibly incorporating this into a future regression model.

The GUI allows a user to make various statistical calculations such as downscaling and regression, and perform operations such as forecasting, inputting data into SWAT, and visualizing the data, with a few button clicks. It should offer the ability to perform many different tests, with different GCM models, interpolation methods, and regression models, quickly without having to write code. Thus, allowing for efficient testing in the future.

# Acknowledgments

These results were obtained as part of the REU Site: Interdisciplinary Program in High Performance Computing (hpcreu.umbc.edu) in the Department of Mathematics and Statistics at the University of Maryland, Baltimore County (UMBC) in Summer 2015. This program is funded by the National Science Foundation (NSF), the National Security Agency (NSA), and the Department of Defense (DOD), with additional support from UMBC, the Department of Mathematics and Statistics, the Center for Interdisciplinary Research and Consulting (CIRC), and the UMBC High Performance Computing Facility (HPCF). HPCF is supported by the U.S. National Science Foundation through the MRI program (grant nos. CNS-0821258 and CNS-1228778) and the SCREMS program (grant no. DMS-0821311), with additional substantial support from UMBC. Co-author Kwame Owusu-Boaitey was supported, in part, by the UMBC National Security Agency (NSA) Scholars Program through a contract with the NSA. Graduate assistant Sai K. Popuri was supported during Summer 2015 by UMBC.

# References

- [1] Bilinear interpolation, 2015. https://en.wikipedia.org/wiki/Bilinear\_ interpolation.
- [2] T. Amemiya. Advanced Econometrics. Harvard University Press, Cambridge, Massachusetts, 1985.
- [3] Christopher Evans, Abigail Gartrell, Lauren Gomez, Moise Mouyebe, Darius Oxley, Sai Kumar Popuri, Nagaraj K. Neerchal, and Amita Mehta. Improving the computational efficiency of downscaling GCM data for use in SWAT. Technical Report HPCF– 2014–12, UMBC High Performance Computing Facility, University of Maryland, Baltimore County, 2014. (HPCF machines used: maya.).
- [4] P.W. Gassman, M.R. Reyes, C.H. Green, and J.G. Arnold. The Soil and Water Assessment Tool: Historical development, applications, and future research directions. http://www.card.iastate.edu/environment/items/asabe\_swat.pdf.
- [5] E.P. Maurer, A.W. Wood, J.C. Adam, and D.P. Lettenmaier. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states. *Journal of Climate*, 15(22):3237–3251, 2002.
- [6] Steven P. Millard and Nagaraj K. Neerchal. Environmental Statistics with S-Plus. CRC Press, 2001.
- [7] V.D. Pope, M.L. Gallani, P.R. Rowntree, and R.A. Stratton. The impact of new physical parametrizations in the Hadley centre climate model — HadCM3. *Climate Dynamics*, 2000.
- [8] S.K. Popuri, N.K. Neerchal, and A. Mehta. Comparison of Linear and Tobit modeling of downscaled daily precipitation over the Missouri River Basin using MIROC5. In

Machine Learning and Data Mining Approaches to Climate Science: Proceedings of the 4th International Workshop on Climate Informatics. Springer, 2015.

- [9] United States Department of Agriculture Natural Resources Conservation Service. Assessments of the Effects of Conservation Practices on Cultivation Cropland in the Missouri River Basin, 2012. http://www.nrcs.usda.gov/Internet/FSE\_DOCUMENTS/ stelprdb1048712.pdf.
- [10] R. Wagner. Multi-Linear Interpolation. http://bmia.bmt.tue.nl/people/BRomeny/ Courses/8C080/Interpolation.pdf, 2008.
- [11] M. Watanabe, T. Suzuki, et al. Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity. *Journal of Climate*, 2010.