Climatic Downscaling using R:

Application of the Delta Method to Minimum and Maximum Temperature from the HIRHAM RCM to Meteorological Stations in the Black Sea Catchment

Ian GUNDERSON
December 2011

Internship Report
Certificate of Specialization in Geomatics
Supervised by Anthony LEHMANN
Abstract

Climate models contain inherent biases due in part to their coarse resolution and to the uncertainty of future climate. Under the scope of the EnviroGRIDS project, part of the European Union's Seventh Research Framework Program (EU FP7), the Delta method is applied to minimum and maximum temperature datasets from the HIRHAM Regional Climate Model (RCM), run under the scope of the EU FP5 and the PRUDENCE project. These datasets correspond to the control period 1961-1990 and two runs for the future period 2071-2100, corresponding to the SRES A2 and B2 scenarios. The "deltas" were then applied onto observed time-series from meteorological station within the Black Sea Catchment. The outputs of the Delta method show that the Predicted minimum and maximum temperatures are more similar than the RCM's daily data to the observed time-series in terms of distribution and variability. The Delta method hence shows skill in generating realistic outputs, while maintaining the temperature shift due to the future scenarios presented by the RCM. Further research is suggested in order to determine the general applicability of the Delta method to different regions and different variables.

Key words
Climate, downscaling, Delta method, temperature, Black Sea, climate model, R

Résumé

Les modèles climatiques contiennent des biais systématiques dus en partie à leur résolution faible et à l'incertitude du climat futur. Dans le cadre du projet EnviroGRIDS, partie du Programme de Recherche de l'Union Européenne (UE FP7), la méthode Delta est appliquée sur des données de température minimum et maximum issues du modèle climatique régional (RCM) HIRHAM, tourné dans le cadre du projet PRUDENCE de l'UE FP5. Ces données correspondent à la période de contrôle 1961-1990 and deux versions pour la période future 2071-2000, correspondant aux scénarios A2 et B2 du SRES. Les "deltas" ont ensuite été appliqués sur des séries temporelles d'observations de stations météorologiques localisées dans le bassin versant de la Mer Noire. Les résultats de la méthode Delta montrent que les températures minimum et maximum sont plus similaires que les données journalières du RCM aux série-temps observées en terme de distribution et de variabilité. La méthode Delta présente une compétence pour générer des résultats réalisistes, tout en maintenant le décalage de température lié aux scénarios futures présents par le MCR. Plus de recherche est recommandée afin de déterminer l'applicabilité générale de cette méthode sur d'autres régions et d'autres variables.

Mots-clés
Climat, downscaling, méthode Delta, température, Mer Noire, modèle climatique, R
Acknowledgements

I would like to thank the entire UNEP/DEWA/GRID-Europe team for giving me a warm and pleasant welcome into the workplace.

I would also like to thank Dr. Stéphane Goyette, from the Climate Change and Climate Impacts (C³i) research group of the University of Geneva, for explaining his research and for providing the data to be manipulated during the internship, as well as for his availability.

Finally, I would particularly like to thank Dr. Anthony Lehmann, from UNEP/DEWA/GRID-Europe and the University of Geneva for the continuous, invaluable help and guidance he offered me throughout the duration of this study.
# TABLE OF CONTENTS

Abstract ........................................................................................................................................... ii  
Acknowledgements......................................................................................................................... iv  
Table of Contents............................................................................................................................ v  
List of Figures................................................................................................................................. vi  
List of Tables..................................................................................................................................... vii

1. Internship Presentation ............................................................................................................. 1 
   1.1 UNEP/DEWA/GRID-Europe ................................................................................................. 1  
   1.2 EnviroGRIDS ..................................................................................................................... 2  
2. Introduction ............................................................................................................................... 3  
   2.1 Climate Models ................................................................................................................... 3  
   2.2 Research Objectives .......................................................................................................... 4  
3. Data ........................................................................................................................................... 6  
   3.1 Regional Climate Model datasets ....................................................................................... 6  
   3.2 Meteorological station observation datasets ...................................................................... 6  
4. Methodology ............................................................................................................................. 8  
   4.1 Description of the Delta method ......................................................................................... 8  
   4.2 The Neighbor function ....................................................................................................... 9  
   4.3 Data extraction and R script run ......................................................................................... 10  
   4.4 The Delta function ............................................................................................................. 11  
   4.5 The Prediction function ..................................................................................................... 12  
   4.6 Output data correction and transformation ....................................................................... 15  
5. Results ....................................................................................................................................... 17  
6. Discussion .................................................................................................................................. 19  
7. Assessment of the internship ................................................................................................... 20  
   7.1 Problems and difficulties encountered .............................................................................. 20  
   7.2 Benefits of the internship ................................................................................................. 20  
8. Conclusion .................................................................................................................................. 21  
9. Bibliography .............................................................................................................................. 22  
   9.1 Manuals and articles .......................................................................................................... 22  
   9.2 Consulted Internet sites ...................................................................................................... 23
List of Figures

Figure 1 Locations of the 150 provided weather stations and the nearest grid point to each station. ................................................... ................................................... ......................................7

Figure 2 Zoom on the locations of the weather stations and associated grid points, with the position of the weather station ID 112700 and grid point ID 2654 used in the Results section. .................................................................................7

Figure 3 Delta Method organization chart. ......................................................................................................................8

Figure 4 Extract of the R script developed defining the Neighbor function .................................................................9

Figure 5 Extract from the R script detailing the extraction of T_{\text{max}} from the RCM Control period .................................10

Figure 6 Extract from the R script presenting the "main" loop that reads the variable data and applies the Delta and Prediction functions .................................................................................................11

Figure 7 Extract from the R script detailing the calculation of the delta for T_{\text{max}} in January. ........................................11

Figure 8 Extract from the R script presenting the output data frame of the Delta function ........................................12

Figure 9 Extract from the R script detailing how NAs are replaced in the Station observed T_{\text{max}} values. ............................................ ................................................... ...............................13

Figure 10 Extract from the R script detailing the calculation of decile categories for the altered Stations observation values, for T_{\text{max}} in January ........................................... ................................................... ............................13

Figure 11 Extract from the R script describing the pseudo-random scaling factor .................................................................13

Figure 12 Extract from the R script detailing the calculation of the "perturbed" T_{\text{max}} values for January ............................................. ................................................... ...............................14

Figure 13 Extract from the R script presenting the method to replace values of T_{\text{max}} where T_{\text{min}}> T_{\text{max}} ....................................................................................................................15

Figure 14 Extract from the R script detailing the date transformation of the "perturbed" T_{\text{min}} and T_{\text{max}} values. ........................................................................................................16

Figure 15 Averaged daily temperature distributions for weather station 112700 and associated RCM grid point 2654. ............................................................................................................17

Figure 16 Frequency distributions of Minimum and Maximum temperature for weather station "112700" (altered time period 1976-2005). ........................................................................................................18

Figure 17 Frequency distributions of Minimum and Maximum temperature for RCM grid point "2654", the Control Period (HC1 – 1961-1990) and both SRES scenarios (HB1 (B2) and HS1 (A2) – 2071-2100). ........................................................................................................19

Figure 18 Frequency distributions of Predicted Minimum and Maximum temperature at weather station "112700", according to both SRES scenarios (HB1 (B2) and HS1 (A2) – 2071-2100) ........................................................................................................19
List of Tables

Table 1 Extract of the output of the Neighbor function, with information on the first 10 weather stations and associated grid point

Table 2 Extract of the output of the Delta function, representing the medians and standard deviation of each decile category of RCM $T_{\text{max}}$ and $T_{\text{min}}$ for January of the HB1 (SRES B2) scenario for grid point 2654

Table 3 Extract from the output of the Prediction function representing the Predicted $T_{\text{max}}$ and $T_{\text{min}}$ values for both scenarios (B = B2, S = A2), for the weather station 112700

Table 4 Central and dispersion values for the $R^2$ between all values of $T_{\text{min}}$ and $T_{\text{max}}$ of respectively the Observed time series, the B2 "Perturbed" time series and the A2 "Perturbed" time series

Table 5 Extract from the final output of the Delta method, containing the Corrected $T_{\text{min}}$ and $T_{\text{max}}$ values for both scenarios (B = B2, S = A2) for weather station 112700
1. Internship Presentation

This report presents the work that was accomplished under the scope of an internship with the UNEP/DEWA/GRID-Europe in their recent project EnviroGRIDS, part of the European Union's Seventh Framework Program (FP7). The mission was to downscale climatic data to available weather stations by generating long time series of temperature, which would be helpful in running climate impact and/or hydrological models. The compiled data resulting from this internship is intended to serve as input data for the Soil and Water Assessment Tool (SWAT).

The two-month internship working for the EnviroGRIDS project began on July 5th, 2010, and ended on September 10th, 2010. The internship was unconventional in the sense that time was split between the offices of the UNEP/DEWA/GRID-Europe at the Maison Internationale de l'Environnement/International Environment House (MIE/IEH) in Geneva, and the Institut des Sciences de l'Environnement (ISE) at the Battelle Research Center of the University of Geneva. In addition to the internship period, about two more months were necessary in order to complete the research objectives.

1.1 UNEP/DEWA/GRID-Europe

This acronym stands for United Nations Environment Program Division of Early Warning and Assessment Global Resource Information Database. DEWA/GRID-Europe forms a part of UNEP's global network of environmental information centers and is one of its major centers for data and information management. GRID-Europe, alongside Nairobi, was the first UNEP center to be launched in 1985, and has a partnership agreement with the Swiss Federal Office for the Environment (FOEN) and the University of Geneva. GRID-Europe focuses on providing and facilitating access to environmental data and information to improve the decision-making process and policy settings. By specializing in the manipulation and analysis of spatial and statistical data on environmental and natural resource issues, it is able to generate and disseminate information about the state of the environment.

GRID-Europe coordinates a number of activities and projects at global and regional levels in order to put into practice their objectives. At the regional level, its most notable activities are that it is related to the coordination of the European Program of the DEWA and as the main contact with diverse European agencies or institutions. It also carries out a certain number of tasks, including information networking between the different European GRID centers, supporting environmental assessment and reporting, such as the Global Environment Outlook (GEO) project, and facilitating access to environmental information by monitoring developments of new systems and tools.

At a global level, GRID-Europe participates actively in the UNEP publication GEO through the submission of reports and data and regular networking with international partner organizations. They also develop, collect and distribute global geo-referenced
environmental data and capacities, as well as provide an “early warning” and information service evaluating global environmental threats, and participate in many other such activities. One of these activities is a capacity building project centered on the Black Sea Catchment region named EnviroGRIDS, which is described below.

1.2 EnviroGRIDS

EnviroGRIDS is a project funded by the 7th Research Framework Program of the European Union and is jointly coordinated by the University of Geneva and UNEP/DEWA/GRID-Europe. It is a project that focuses its research on the Black Sea Catchment region, internationally known as having “ecologically unsustainable development and inadequate resource management, which has led to severe environmental, social and economic problems.”

The purpose of EnviroGRIDS is to address these issues by building the capacity of scientists to create an observation system of information technologies including data, models and scenarios to gather information about the past, present and future periods. EnviroGRIDS seeks in this way to also help build the capacity of decision-makers and of the general public to best use this system and understand the issues at hand.

One of the main objectives of the project is to bridge the gap between science and policy, making it possible to assess and predict the future sustainability and vulnerability of the region. It will also contain a system of shared information between partners in order to improve networking and to exchange data and methods more efficiently.

2. Introduction

2.1 Climate Models

Climate change issues are attracting more attention than ever before in many different contexts. There are "growing demands from decision-makers in the public and private sectors, from NGOs, from researchers and from the general public for detailed information on future climate" (Christensen et al., 2007a). In order to meet these demands, climatologists have developed models that simulate the climate at different scales. Climate models are useful tools that allow this foresight into what future atmospheric conditions might be. The main tool used for large-scale climate modeling is a General Circulation Model (GCM), whose spatial resolution of 200-300 km is insufficient to provide the required details for regional applications. Climates models often still contain inherent biases, notably due to their coarse spatial resolutions. Doubt has also been expressed concerning the reliability of some GCM output variables (Wilby and Wigley 2000). GCMs are therefore unsuitable for detailed regional analysis (Charles et al. 2007; Wilby et al. 1999). Therefore, the outputs generated by these models require further processing before being used.

Climate impact models and hydrological models require adequate input data allowing them to run conveniently and with sufficient accuracy so that their results can be used. These impact models are more often than not run at regional scales, for example at the scale of a catchment, as they are more realistic at this scale and must parameterize regional processes (Wilby et al. 1998). These require a certain degree of precision concerning the topography of the region or catchment and resulting regional inputs. An important issue when considering adaptation and mitigation responses to climate change is the uncertainty in the predictions of future climate (Christensen and Christensen, 2007b). One of the challenges is to incorporate climate change projection uncertainties in such a way that concerned groups may make informed decisions based on model projections (Blenkinsop and Fowler, 2007). Hence it is necessary to reduce as much as possible this uncertainty and the model biases (Murphy, 2000).

The method that provides the best current solution to this issue is downscaling GCM output data (Charles et al., 2007). Impact modelers have generally applied coarse resolution mean changes in climate, with some appropriate post-processing, to impact models calibrated by observed data (Christensen and Christensen, 2007b). There are two different methods of doing so: dynamical downscaling (DDS) and statistical downscaling (SDS).

DDS implies that a Regional Climate Model (RCM) is nested within a GCM, providing additional skill due to a higher resolution and more detail (Murphy, 1999). As the boundary conditions are still given by the driving GCM, the GCM errors can be brought into the RCM (Hay and Clark; 2003). However, RCMs can provide substantial differences compared to their driving GCM (Jacob et al., 2007) as well as uncertainties of
their own (Christensen et al., 2007a), resulting notably from their formulation (Beniston et al., 2007). The advantage of DDS is, among others, its physical realism (Hay and Clark, 2003). SDS implies the use of statistical relationships between the GCM data and observed data and using a certain method of statistical analysis, and the resulting values help estimate future climate projections at a finer scale.

All studies advise using caution when using data produced by these methods, given their existing uncertainties (Hay et al., 2000; Haylock et al., 2006). It is also often suggested to use multiple sources and methods within the study, including multi-model ensemble scenarios (multiple RCMs driven by the same GCM, and RCMs driven by other GCMs) (Beniston et al., 2007) to minimize uncertainty, since even RCMs driven by the same GCM can produce very different results (Kjellstrom and Ruosteenoja, 2007; Jacob et al., 2007). Each method has its advantages, drawbacks and base assumptions determining its specific utility (Quilbe et al., 2008). It is therefore necessary to consider several approaches, tools and datasets to account for uncertainties when defining future meteorological series. Any one model simulation of future climate may represent only one of many possible future climate states (Blenkinsop and Fowler, 2007), and there are potential limitations to relying solely upon RCM-based information (Christensen et al., 2007a). Despite its limitations, downscaling remains one of the most appealing techniques for generating high-resolution climatic or hydrological datasets, at least for the foreseeable future (Wilby et al., 1998).

All studies encourage further studies and cooperation between modelers to compare and evaluate models in order to identify the causes of bias and remove them (Hay and Clark, 2003; Jacob et al., 2007; Christensen et al., 2007a). Hopefully these methods can be improved further, and could possibly lead to a "best" method (Murphy, 2000). However, the challenge in defining a “best” method is that there is no established way to tell which model represents the most probable version of the future (Deque et al., 2007), and that achieving skillful downscaling under present climate conditions does not guarantee that the methods will perform equally well when used to predict future conditions (Murphy, 1999).

### 2.2 Research Objectives

This study thus seeks to take this process a step further, even though only one method and data produced by only one model are used (notably due to time-constraint issues). The simple but effective “Delta” method (also termed Change Factor, Perturbation method) was used to relate the data from simulations to observations. The "Delta" method was applied onto available meteorological observations from weather stations within the Black Sea Catchment. Daily minimum (T_{min}) and maximum (T_{max}) temperature at two meters above the surface served as the meteorological variables to be downscaled. The assumption of the delta method is that GCMs, and hence RCMs, simulate relative changes more reliably than absolute values (Hay et al., 2000). In previous studies, GCM outputs were always used as the data to be downscaled (Hay et al., 2000; Diaz-Nieto and Wilby,
2005; Quilbe et al., 2008). In this study, climatic data was downscaled from an RCM and not from a GCM as is usually the case. The downscaled output was then applied to existing, spatially-explicit, meteorological stations. It is hoped that applying the delta method from RCM outputs to meteorological observations will provide more realistic and reliable results and serve to remove some of the bias present within the driving climate models.

The RCM outputs used in this study originate from the Danish Meteorological Institute’s HIRHAM model, which was run under the auspices of the EU Fifth Research Framework Program’s PRUDENCE project. The "Delta" method made it possible to determine the evolution of observed time-series beyond the observed time period and into the second half of the 21st century. The outputs of the "Delta" method would hopefully increase the accuracy of future scenarios in the Black Sea Catchment, and therefore improve other studies and tools that use the same type of data for their analyses and assessments such as climate impact and hydrological models. The results of this study will thus serve the Soil and Water Assessment Tool (SWAT), a hydrological model scaled to the size of a large watershed (Nietsch, 2011). The SWAT model requires reliable inputs in order to correctly assess the impacts of various climatic variations. However, outputs form the "Delta" method, as well as all downscaling techniques, are constrained by several factors: the GCM accuracy in simulating grid-box variable changes (Hay et al., 2000), the physical realism and internal consistency of the GCM forcing (Wilby et al., 1998) and the fact that the delta method does not take into account spatial variability (Quilbe et al., 2008).

A further improvement made was to not add the delta value directly to the observed time-series, but rather to apply certain statistics for representation issues and inducing scale factors in order to reproduce variability in the output data. This will ensure that the resulting distribution does not have the same range or variability at a different magnitude (Diaz-Nieto and Wilby, 2005). It is hoped that by representing a variation of a statistical-dynamical downscaling technique, it will be possible to eliminate a certain amount of bias inherent in the climatic model being used, reduce the uncertainty and therefore provide more accuracy and confidence in the output to be used in the impact or hydrological model. Through this, this study follows the recommendations of previous studies (Diaz-Nieto and Wilby, 2005; Murphy, 1999; Murphy, 2000) through the convergence of the two downscaling techniques by using the delta method alongside SDS techniques using their advantages in a complementary way. This will hopefully contribute to the acquisition of more reliable future scenarios useful for the SWAT model and therefore contribute to capacity building.

It is suggested that downscaling techniques should be globally applicable, and that further studies are needed concerning the generality of the findings, notably for other regions, variables and models (Wilby et al., 1998; Wilby and Wigley, 2000). Therefore, the final objective of this study was to create a methodology that could be applied for other studies, while requiring as few modifications as possible.
3. Data

3.1 Regional Climate Model datasets

The RCM used during this work was the HIRHAM model, driven by the United Kingdom's Hadley Center HadAM3H GCM. This model had been used under the scope of the Fifth Research Framework Program of the European Union and the PRUDENCE project (Christensen and Christensen, 2007b). The majority of the data produced by this project are now freely available.

The model was first run to simulate data for a control period (HC), 1961 to 1990 to evaluate the quality of the model. A number of potential future scenarios were also simulated for the period of 2071 to 2100 (HS and HB), representing the IPCC's SRES scenarios (Nakicenovic et al., 2000). The HS scenario corresponds to the IPCC's SRES A2 scenario, while the HB scenario represents the SRES B2 scenario. These two periods were run a number of times using slightly different parameters each time for PRUDENCE, but only the data from the first run was used, HC1, HS1 and HB1. Different time values were available, varying from seasonal, to monthly, to daily. Daily values were used in order to have a maximum amount of data possible. Eighteen different variables representing climatic data were also available. This study focuses on the minimum and maximum temperatures at 2 meters above the surface (respectively RCM $T_{\text{min}}$ and RCM $T_{\text{max}}$), and are in netCDF format. These variables were chosen to best represent the needs of the EnviroGRIDS project focusing on sustainable development and to be potentially useful for SWAT, developed for the project. In the case of the HIRHAM model, the netCDF for each variable and period contain 10,800 daily values, corresponding to 360 days for a 30 year period, for each of the 7560 grid points present in the RCM. The coordinates of the grid points were also included in the datasets and were used in order to match each weather station with its closest grid point, as presented by Figure 1.

3.2 Meteorological station observation datasets

The meteorological data for maximum and minimum temperatures of 150 meteorological stations in the Black Sea Catchment were provided (respectively Station $T_{\text{min}}$ and Station $T_{\text{max}}$). These were in dBase format, and each station had its own file with values concerning maximum and minimum temperatures in two separate columns. The time frame for these values ranged from the 1st of January 1970 to a certain time in 2008, depending on the station. However, large sections of these files contained missing values. Therefore, a 30-year period, ranging from the 1st of January 1976 to the 30th of December 2005, was extracted and onto which we applied the Delta method. The coordinates of these 150 weather stations were provided in a separate file and are projected in the ED-1950:

\[ URL: \text{http://prudence.dmi.dk/} \]
Lambert Conformal Conic coordinate system. This data was provided by Dr. Karim Abbaspour, from the Swiss Federal Institute of Aquatic Science and Technology (EAWAG). These stations are all located in the western section of the Black Sea Catchment, as shown in Figure 1 and Figure 2. It must be mentioned that some of the nearest grid points to these weather stations are located outside of the Black Sea Catchment.

Figure 1 Locations of the 150 provided weather stations and the nearest grid point to each station.

Figure 2 Zoom on the locations of the weather stations and associated grid points, with the position of the weather station ID 112700 and grid point ID 2654 used in the Results section.
4. Methodology

4.1 Description of the Delta method

The methodology of running the "Delta" method on RCM $T_{\text{min}}$ and $T_{\text{max}}$ and applying them to meteorological observations from weather stations was developed in the R programming language.

It was thus chosen to perturb the observations to simulate a realistic scenario for future temperatures at the location of the weather station. As previously stated, this method would help to circumvent the problem generated by the use of model data, and remove the inherent bias of the driving climatic model. Although the “resolution”, or number of points represented by the stations, is not in excess of the modeled grid's resolution, it is still termed a downscaling method because we are scaling down the data to a different level.

The particular "Delta" method chosen to downscale the climatic data is based on the Deliverable 3.2 of WP3 "Scenarios of change", Task 3.2, "Climate" of the EnviroGRIDS project (Goyette, 2010, Perroud and Goyette, 2010). It consists of applying the difference of the desired climatic variable, originating from simulations from a climatic model, between the scenario period and the control period, and adding this value onto the observed data contemporary of the control period from weather stations. The result would be a "perturbed" value of the climatic variable in question for as long as desired, either corresponding to the scenario period or the evolution until the year 2100.

Figure 3 presents the organization of the Delta method and the different steps undertaken in order to obtain the final output result. These steps are described in more detail in the following sections.

![Figure 3: Delta Method organization chart.](image)
4.2 The Neighbor function

In order to achieve this result, the first step was to determine the nearest grid point to each weather station. The Neighbor function was developed that would be able to associate the geographic coordinates of existing meteorological stations in the Black Sea Region with those of the grid points of the RCM.

```r
# AUXILIARY Neighbor Function:
neighbor <- function(r_matrix, grid, dissolved_grid) {  # read weather station coordinates
  lat.w <- get.var.ncdf(r_matrix, "lat")
  lon.w <- get.var.ncdf(r_matrix, "lon")
  # read long and lat names from one of the netCDF files
  latu = c("LAT.ne", LAT.ne, pos(latu) = CORDON.lat(ne), LAT_ne, point)
  lonu = c("LONG.ne", LONG.ne, point)
  # create matrix for long and lat
  nameu = c("name", "name")
  # return column name
  closest.grid <- rep(-99, 100)  # to create empty vector
  distance.grid <- rep(99, 100)  # to distance closest Neighbor grid points
  RCM_grid[1:2] <- rownames(RCM_grid[2, 1])
  # create empty row to store alias and grid point name
  dist.station.grid <- rep(99, 100)
  # distance stations to all RCM grid points (5'x5')
  # loop to calculate closest grid neighbor to all 50 stations
  for (i in 1:nrow(grid)) {
    for (j in 1:nrow(grid)) {  
      dist.station.grid[i] <- sqrt((station coordenades[1, i] - RCM_grid[, i][1])^2 + (station coordenades[2, i] - RCM_grid[, i][2]))  
      # distance calculation
      distance.grid[i] <- min(dist.station.grid)
      # get closest distance
      position <- which(RCM_grid[, i] == closest.grid[i])
      # get the position for alias/grid from RCM
      RCM_grid[1:2][position] <- position
      # get position for alias/grid from RCM
      RCM_grid[1:2][position] <- position
      # get position for alias/grid from RCM
      # return to be returned by the Neighbor function
      return(neighbor)
    }
  }
}
```

Figure 4 Extract of the R script developed defining the Neighbor function

As presented by Figure 4, the station coordinates were provided in a table, while the grid point coordinates were extracted from the netCDF files containing the RCM data, using the "get.var.ncdf" function of the "ncdf" package. After creating a number of empty vectors to store various objects, a "for" loop is developed to run the function on all provided stations and grid points. This loop calculates the distance of each weather station to all grid points, according to the Pythagoras Theorem. This simple equation is justified by the fact that the stations and grid points are within a relatively short distance from each other. The grid point with the shortest distance to the weather station is then determined and stored in a data frame containing other relevant information (Table 1). This output then served for later manipulations.
Table 1 Extract of the output of the Neighbor function, with information on the first 10 weather stations and associated grid point.

<table>
<thead>
<tr>
<th>WS_IDpt</th>
<th>RCM_IDpt</th>
<th>WS_x</th>
<th>WS_y</th>
<th>rlong</th>
<th>rlat</th>
<th>RCM_x</th>
<th>RCM_y</th>
<th>Distance RCM to WS</th>
</tr>
</thead>
<tbody>
<tr>
<td>106870</td>
<td>3283</td>
<td>11.95</td>
<td>49.7</td>
<td>43</td>
<td>37</td>
<td>11.839</td>
<td>49.586</td>
<td>0.159</td>
</tr>
<tr>
<td>106880</td>
<td>3284</td>
<td>12.18</td>
<td>49.67</td>
<td>44</td>
<td>37</td>
<td>12.504</td>
<td>49.676</td>
<td>0.324</td>
</tr>
<tr>
<td>107710</td>
<td>3283</td>
<td>11.9</td>
<td>49.43</td>
<td>43</td>
<td>36</td>
<td>11.839</td>
<td>49.586</td>
<td>0.168</td>
</tr>
<tr>
<td>107715</td>
<td>3193</td>
<td>11.83</td>
<td>49.22</td>
<td>43</td>
<td>36</td>
<td>11.980</td>
<td>49.156</td>
<td>0.163</td>
</tr>
<tr>
<td>107760</td>
<td>3193</td>
<td>12.1</td>
<td>49.05</td>
<td>43</td>
<td>36</td>
<td>11.980</td>
<td>49.156</td>
<td>0.160</td>
</tr>
<tr>
<td>107880</td>
<td>3104</td>
<td>12.57</td>
<td>48.83</td>
<td>44</td>
<td>35</td>
<td>12.771</td>
<td>48.814</td>
<td>0.202</td>
</tr>
<tr>
<td>107910</td>
<td>3195</td>
<td>13.13</td>
<td>49.12</td>
<td>45</td>
<td>36</td>
<td>13.300</td>
<td>49.330</td>
<td>0.270</td>
</tr>
<tr>
<td>108270</td>
<td>3098</td>
<td>9</td>
<td>48.18</td>
<td>38</td>
<td>35</td>
<td>8.891</td>
<td>48.226</td>
<td>0.118</td>
</tr>
<tr>
<td>108370</td>
<td>3100</td>
<td>9.92</td>
<td>48.22</td>
<td>40</td>
<td>35</td>
<td>10.174</td>
<td>48.437</td>
<td>0.335</td>
</tr>
<tr>
<td>108380</td>
<td>3100</td>
<td>9.95</td>
<td>48.38</td>
<td>40</td>
<td>35</td>
<td>10.174</td>
<td>48.437</td>
<td>0.232</td>
</tr>
</tbody>
</table>

4.3 Data extraction and R script run

The following step is to extract the data representing temperature from the RCM netCDF files corresponding to the grid points associated to a weather station. The "open.ncl" and "get.var.ncl" functions from the "ncdf" package are used, and their results are stored in an object containing the values for all grid points associated to a weather station (Figure 5).

```r
# Load the required package
library(ncdf)

# Extract the data
Tmax <- ncdf["max"]

# Create a data frame for the extracted data
temp_data <- data.frame(
  WS_ID = 1,  # Example ID
  RCM_ID = 1, # Example ID
  WS_x = 1,   # Example value
  WS_y = 1,   # Example value
  rlong = 1,  # Example value
  rlat = 1,   # Example value
  RCM_x = 1,  # Example value
  RCM_y = 1,  # Example value
  Distance = 1 # Example value
)
```

Figure 5 Extract from the R script detailing the extraction of $T_{\text{max}}$ from the RCM Control period

A "for" loop is then created to run on all available weather stations to read the files containing the Station $T_{\text{min}}$ and Station $T_{\text{max}}$ data and store them using the "foreign" package. The data from the associated grid point is also stored. This data is then entered in the two main functions developed in this study, the Delta function and the Prediction function (Figure 6). This corresponds to the main loop that runs on each weather station.
The Delta function

The Delta function serves to calculate the daily average for each date of the complete 30-year period of the RCM data, for example calculating the average value of RCM $T_{\text{max}}$ for every 1st of January of the 30-year Control period. Figure 7 shows how a loop is developed over the value of each day of the year to calculate its mean value. For each month, the daily averages of each scenario period are then subtracted to the corresponding daily averages of the control period. The daily averages of the control period are then divided into deciles, and the value of each daily average is categorized in its corresponding monthly decile category. This calculation is repeated for each scenario of RCM $T_{\text{max}}$, each scenario of RCM $T_{\text{min}}$, and for each month.

### The Delta function

```r
# Define Delta function
delta <- function (scenario) {
  # Read data frame associated to selected Weather Station
  daily_temperatures <- daily_temperatures[scenario,]
  # Create empty vector to store daily monthly averages
  daily_average_RCM_Tmin <- NULL
  daily_average_RCM_Tmax <- NULL
  # Loops over each month of the 30-year period
  for (i in 1:12) {
    daily_average_RCM_Tmin[i] <- daily_average_RCM_Tmax[i] <- daily_average_RCM_Tmin / daily_average_RCM_Tmax
    # Calculate delta value between daily monthly average of Tmin and Tmax
    delta_RCM_Tmin[i] <- delta_RCM_Tmax[i] <- daily_average_RCM_Tmin[i] * daily_average_RCM_Tmax[i]
    # Create vector to store daily delta values
    delta_RCM_Tmin[i] <- delta_RCM_Tmax[i] <- daily_average_RCM_Tmin[i] * daily_average_RCM_Tmax[i]
    # Create vector to store daily delta values
    delta_RCM_Tmin[i] <- delta_RCM_Tmax[i] <- daily_average_RCM_Tmin[i] * daily_average_RCM_Tmax[i]
  }
}
```

Figure 6 Extract from the R script presenting the "main" loop that reads the variable data and applies the Delta and Prediction functions.

Figure 7 Extract from the R script detailing the calculation of the delta for $T_{\text{max}}$ in January.
Initially, an "Annual" version of the "Delta method was used where the daily averages were calculated over the complete year and not separated month by month. The "Annual" delta method consisted of organizing these daily averages into deciles calculated over the complete year, and not deciles calculated over each month. This resulted in 360 daily averages divided into ten decile categories, while the "Monthly" Delta method results in 30 daily averages divided into ten decile categories, each month stored separately. The latter method allows for an improved accuracy of the results and is therefore the chosen method described in this study.

The output of the Delta function is a data frame containing the median and standard deviation values for each decile category, for each scenario of each variable, month by month, as shown in Figure 8. These two basic statistical functions help determine the central and dispersion values of the distribution and are used to calculate the "Perturbed" \( T_{\text{min}} \) and \( T_{\text{max}} \) values in the Prediction function (Table 2).

### Table 2 Extract of the output of the Delta function, representing the medians and standard deviation of each decile category of RCM \( T_{\text{max}} \) and \( T_{\text{min}} \) for January of the HB1 (SRES B2) scenario for grid point 2654.

<table>
<thead>
<tr>
<th>Decile category</th>
<th>Med.CB_MAX_JAN</th>
<th>SD.CB_MAX_JAN</th>
<th>Med.CB_MIN_JAN</th>
<th>SD.CB_MIN_JAN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.291</td>
<td>0.079</td>
<td>2.338</td>
<td>0.077</td>
</tr>
<tr>
<td>2</td>
<td>2.232</td>
<td>0.169</td>
<td>2.558</td>
<td>0.528</td>
</tr>
<tr>
<td>3</td>
<td>1.872</td>
<td>0.244</td>
<td>2.097</td>
<td>0.119</td>
</tr>
<tr>
<td>4</td>
<td>2.080</td>
<td>0.493</td>
<td>2.114</td>
<td>0.243</td>
</tr>
<tr>
<td>5</td>
<td>1.956</td>
<td>0.238</td>
<td>1.810</td>
<td>0.103</td>
</tr>
<tr>
<td>6</td>
<td>1.734</td>
<td>0.145</td>
<td>1.391</td>
<td>0.225</td>
</tr>
<tr>
<td>7</td>
<td>1.960</td>
<td>0.429</td>
<td>2.152</td>
<td>0.324</td>
</tr>
<tr>
<td>8</td>
<td>1.672</td>
<td>0.552</td>
<td>1.743</td>
<td>0.352</td>
</tr>
<tr>
<td>9</td>
<td>1.780</td>
<td>0.159</td>
<td>1.314</td>
<td>0.425</td>
</tr>
<tr>
<td>10</td>
<td>1.276</td>
<td>0.163</td>
<td>0.955</td>
<td>0.403</td>
</tr>
</tbody>
</table>

Figure 8 Extract from the R script presenting the output data frame of the Delta function.

### 4.5 The Prediction function

The Prediction function is the next step of the Delta method. This function first renames the column headers of the Station's observations in order to ease further processing, while all -99 values, representing missing values, are replaced by NAs. As the RCM data contains "regular" dates (30 days for each month), an object is created in order to transform the Stations' observations, which are dated according to the Gregorian...
calendar, into regular dates corresponding to 1\textsuperscript{st} of January 1976 until the 30\textsuperscript{th} of December 2005, as previously explained.

The remaining missing values are then replaced by using a sample corresponding to the same date. In other words, if the value for the 1\textsuperscript{st} of January 1976 is missing, then it is replaced by the existing value of the 1\textsuperscript{st} of January from another year. The missing values for the 29\textsuperscript{th} and 30\textsuperscript{th} February are replaced respectively by the value of the 28\textsuperscript{th} of February or the 1\textsuperscript{st} of March of the same year, as shown in Figure 9. This is repeated for both $T_{\text{min}}$ and $T_{\text{max}}$.

```r
# Loop to replace NAs by Sampled value from the same date
for (i in 1:360) {
  if (i == 58) j <- 38 : if (i == 60) j <- 61
  x[i] <- x[j]
}
```

*Figure 9 Extract from the R script detailing how NAs are replaced in the Station observed $T_{\text{max}}$ values.*

The next step in the Prediction function is to calculate the decile categories for each month. Unlike in the Delta function, the decile categories are calculated over the complete monthly distributions, and not on the daily averages (Figure 10). This step is repeated for each variable and for each month, and the results are stored in a data frame.

```r
Station.X1:JAN <- Station.X1:JAN[Station.X1:JANMonth==1]  # select Jan values and desired month from Station.X1 data frame
q_Station.X1:JAN <- quantile(Station.X1:JAN, [0:10]/10)  # calculate decile values for complete monthly distribution
q_Station.X1:JAN <- NULL
for (i in 1:31) {q_Station.X1:JAN[Station.X1:JAN[i]] <- q_Station.X1:JAN[1]} \# Station.X1:JAN[i] <- q_Station.X1:JAN[i]
```

*Figure 10 Extract from the R script detailing the calculation of decile categories for the altered Stations observation values, for $T_{\text{max}}$ in January.*

In order to reproduce a notion of variability, a pseudo-random factor between 1 and -1 is introduced. This value is modified in order to follow a normal distribution, with most values close to 0, and fewer near the extremes of 1 and -1 (Figure 11). The pseudo-random factor is later used in the calculation of the "perturbed" temperature values.

```r
# Set scale factors and seed
Scale1 <- 5
Scale2 <- 1
y1 <- cumsum(cumsum(0:1,1) * 1)
Alcm <- alcm * Scale2 * cry || (Scale1 * cumsum(0:1,1) * 1/2)
```

*Figure 11 Extract from the R script describing the pseudo-random scaling factor.*
The final step of the Prediction function calculates the "Perturbed" \( T_{\text{min}} \) and \( T_{\text{max}} \). The number of years that the method is intended to run is entered, in this case 110. Empty vectors are identified in order to store the results. An incremental value is introduced, whose objective is to increase the variability year by year. This incremental value begins at 16/110 (or 0.145), as while the altered Station observations being in 1976, the modeled results begin in 1991 and end in 2100. This increase factor therefore ends at 126/110 (or 1.145).

A "for" loop is then developed which selects a random year from the altered Station observations, and calculates the "perturbed" values according to the equation presented in Figure 12. The equation first selects the altered Station observations from its first decile category, and adds a factor combining the incremental value, the median of the corresponding decile category from the output of the Delta function, the pseudo-random factor and standard deviation of the corresponding decile category. This step is repeated within the loop for each decile category, each variable, each scenario, and each month. The output of the Prediction function is then a data frame combining the "Perturbed" \( T_{\text{min}} \) and \( T_{\text{max}} \) values for both scenarios. The end of the Prediction function marks the end of the "main" loop described in Figure 6, and results in 150 tables, each corresponding to a "Perturbed" weather station (Table 3).

```r
# Predict the enjoiced values

Ndays <- 110
Result_pred <- rep(NA, Ndays)

# number of days corresponding to the number of years to be predicted

RESULT_B_MIN <- data.frame("ResultB"="rep([1,11], each=0.1, Ndays), "Pred"="rep([0, Ndays])")
RESULT_B_MAX <- data.frame("ResultB"="rep([1,11], each=0.1, Ndays), "Pred"="rep([0, Ndays])"

# create empty data frame to store final predictions

# repeat for each variable and each scenario

for (j in 1:10) {
  # select monthly values for each year
  Month <- seq(1,12, by=1)
  # select month's median values for each year
  Median <- median(Month)
  # select month's median values for each year
  StdDev <- sd(Month)
  # calculate predicted monthly values for each scenario
  Perturbed <- median(Month)*perturbation + Median*Delta_perturbation + StdDev*RandomFactor + (1 + median(Month)*perturbation + Median*Delta_perturbation + StdDev*RandomFactor))
  # store predicted monthly values in corresponding month
  RESULT_B_MIN[Month, RESULT_B_MAX[Month, [perturbation + median(Month)*perturbation + Median*Delta_perturbation + StdDev*RandomFactor]]]
  # store corresponding month's prediction in

Figure 12 Extract from the R script detailing the calculation of the "perturbed" \( T_{\text{max}} \) values for January

Ian Gunderson
Table 3 Extract from the output of the Prediction function representing the Predicted $T_{\text{max}}$ and $T_{\text{min}}$ values for both scenarios ($B = B_2$, $S = A_2$), for the weather station 112700.

<table>
<thead>
<tr>
<th>Day</th>
<th>$B_{\text{MAX}}$</th>
<th>$B_{\text{MIN}}$</th>
<th>$S_{\text{MAX}}$</th>
<th>$S_{\text{MIN}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.059</td>
<td>0.146</td>
<td>2.365</td>
<td>0.354</td>
</tr>
<tr>
<td>2</td>
<td>1.770</td>
<td>-0.191</td>
<td>1.901</td>
<td>-0.019</td>
</tr>
<tr>
<td>3</td>
<td>-0.277</td>
<td>-2.828</td>
<td>-0.206</td>
<td>-2.543</td>
</tr>
<tr>
<td>4</td>
<td>-0.415</td>
<td>-4.787</td>
<td>-0.228</td>
<td>-4.635</td>
</tr>
<tr>
<td>5</td>
<td>0.451</td>
<td>-1.911</td>
<td>0.570</td>
<td>-1.663</td>
</tr>
<tr>
<td>6</td>
<td>3.886</td>
<td>1.639</td>
<td>4.046</td>
<td>1.842</td>
</tr>
<tr>
<td>7</td>
<td>3.474</td>
<td>0.518</td>
<td>3.633</td>
<td>0.700</td>
</tr>
<tr>
<td>8</td>
<td>2.394</td>
<td>-1.141</td>
<td>2.664</td>
<td>-0.790</td>
</tr>
<tr>
<td>9</td>
<td>1.185</td>
<td>-1.809</td>
<td>1.409</td>
<td>-1.557</td>
</tr>
<tr>
<td>10</td>
<td>0.151</td>
<td>-2.649</td>
<td>0.271</td>
<td>-2.400</td>
</tr>
</tbody>
</table>

4.6 Output data correction and transformation

Running the model on the "Annual" Delta method resulted in a large number of occasions where the "Perturbed" $T_{\text{min}}$ on a given day was higher than the "Perturbed" $T_{\text{max}}$. The "Monthly" Delta method model significantly decreased the number of times this result occurred (down to < 10% of values). In order to correct these errors, the coefficient of determination for all predicted results was calculated between the "Perturbed" $T_{\text{min}}$ and $T_{\text{max}}$ values. As these values were significantly important (Table 4), all values of "Perturbed" $T_{\text{max}}$ that were lower than the "Perturbed" $T_{\text{min}}$ were replaced by the theoretical value given by the distributions' linear equation, as in Figure 13. This step is repeated for each "Perturbed" weather station, and results in "Corrected" weather station values.

Figure 13 Extract from the R script presenting the method to replace values of $T_{\text{max}}$ where $T_{\text{min}} > T_{\text{max}}$.

Table 4 Central and dispersion values for the $R^2$ between all values of $T_{\text{max}}$ and $T_{\text{min}}$ of respectively the Observed time series, the B2 "Perturbed" time series and the A2 "Perturbed" time series.

<table>
<thead>
<tr>
<th></th>
<th>Obs</th>
<th>B</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min</td>
<td>0.524</td>
<td>0.593</td>
<td>0.593</td>
</tr>
<tr>
<td>Max</td>
<td>0.933</td>
<td>0.938</td>
<td>0.936</td>
</tr>
<tr>
<td>Mean</td>
<td>0.702</td>
<td>0.764</td>
<td>0.766</td>
</tr>
<tr>
<td>Med</td>
<td>0.696</td>
<td>0.762</td>
<td>0.764</td>
</tr>
<tr>
<td>St. dev.</td>
<td>0.086</td>
<td>0.059</td>
<td>0.059</td>
</tr>
</tbody>
</table>
Input files in SWAT require a certain format in order to be accessed correctly. The data generated by the Delta method described above contains date-specific values. However, these dates correspond to the format present in the RCM data, i.e. 30 days/month, 360 days/year. SWAT only functions with dates corresponding to the Gregorian calendar. In order to resolve this issue, and as SWAT is able to interpolate missing values, the "Corrected" weather station data were merged with a file containing dates following the Gregorian calendar until the end of the model run period, the 31st of December 2100, as shown in Figure 14. Therefore, missing values are created where the dates do not correspond (for example on the last day of each month with 31 days). Instead of having results containing only 39 600 values (corresponding to 110 years of "regular" dates), each result now contains 40178 days (actual number of days between the 1st of January 1991 and the 31st of December 2100), with 578 missing values.

Figure 14 Extract from the R script detailing the date transformation of the "perturbed" T\textsubscript{min} and T\textsubscript{max} values.

The result of this transformation corresponds to the final output of the developed Delta method, with tables containing the "Predicted" T\textsubscript{min} and T\textsubscript{max} values for the scenarios A2 and B2 for all 150 meteorological stations that were available (Table 5).

Table 5 Extract from the final output of the Delta method, containing the Corrected T\textsubscript{min} and T\textsubscript{max} values for both scenarios (B = B2, S = A2) for weather station 112700.

<table>
<thead>
<tr>
<th>Date</th>
<th>B MAX</th>
<th>B MIN</th>
<th>S_MAX</th>
<th>S_MIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>19910101</td>
<td>2.059</td>
<td>0.146</td>
<td>2.365</td>
<td>0.354</td>
</tr>
<tr>
<td>19910102</td>
<td>1.770</td>
<td>0.191</td>
<td>1.901</td>
<td>0.019</td>
</tr>
<tr>
<td>19910103</td>
<td>-0.277</td>
<td>-2.828</td>
<td>-0.206</td>
<td>-2.543</td>
</tr>
<tr>
<td>19910104</td>
<td>-0.415</td>
<td>-4.787</td>
<td>-0.228</td>
<td>-4.635</td>
</tr>
<tr>
<td>19910105</td>
<td>0.451</td>
<td>-1.911</td>
<td>0.570</td>
<td>-1.663</td>
</tr>
<tr>
<td>19910106</td>
<td>3.886</td>
<td>1.639</td>
<td>4.046</td>
<td>1.842</td>
</tr>
<tr>
<td>19910107</td>
<td>3.474</td>
<td>0.518</td>
<td>3.633</td>
<td>0.700</td>
</tr>
<tr>
<td>19910108</td>
<td>2.394</td>
<td>-1.141</td>
<td>2.664</td>
<td>-0.790</td>
</tr>
<tr>
<td>19910109</td>
<td>1.185</td>
<td>-1.809</td>
<td>1.409</td>
<td>-1.557</td>
</tr>
<tr>
<td>19910110</td>
<td>0.151</td>
<td>-2.649</td>
<td>0.271</td>
<td>-2.400</td>
</tr>
</tbody>
</table>
5. Results

The result of this script is a set of 150 datasets, one for each meteorological station, including the $T_{\text{min}}$ and $T_{\text{max}}$ values for the A2 and B2 SRES scenarios as well as the corresponding date for each day. These datasets will then serve as input files for the SWAT program.

The following Figures present various characteristics of the input data used during this study as well as for the predicted temperatures generated by the Monthly Delta Method. All Figures present the data corresponding to the randomly-picked Weather Station 112700 and associated RCM grid point 2654.

As can be seen by Figure 15 below, the RCM averaged daily values for $T_{\text{min}}$ and $T_{\text{max}}$ create a relatively smooth curve with a seasonal variation of a maximum of 20°C for $T_{\text{max}}$ and 15°C for $T_{\text{min}}$. It is to be noticed that both scenarios provide higher average temperatures than the control period. The temperatures predicted using the Delta Method present a distribution with much more variation. There is more fluctuation over the averaged daily values, and a larger seasonal variation. The predicted temperatures follow the weather station distribution for observed temperatures. The temperature gap between the different periods and scenarios are similar to those in the RCM data.

![Figure 15 Averaged daily temperature distributions for weather station 112700 and associated RCM grid point 2654.](image-url)
The frequency distributions of the various data that was provided for both $T_{\text{min}}$ and $T_{\text{max}}$ (observed temperatures at a Weather Station, RCM temperatures) and the predicted $T_{\text{min}}$ and $T_{\text{max}}$ temperatures, shown in Figures 15 to 17, all present a bimodal distribution. The RCM frequency distributions show a higher frequency peak at the lower temperature ranges for both minimum and maximum temperatures, while the Weather Station frequency distributions provide a slightly higher frequency peak at the higher temperature ranges. The Predicted frequency distributions however show that the minimum temperatures reach their highest frequency peak at the lower first mode, while the maximum temperatures reach theirs at the higher second mode.

It is apparent that values originating from the RCM data have a much smaller range than the observed time-series. This suggests that the RCM data does not reproduce accurately temperature at the weather station to which the grid point is closest.

*Figure 16 Frequency distributions of Minimum and Maximum temperature for weather station "112700" (altered time period 1976-2005).*
Figure 17 Frequency distributions of Minimum and Maximum temperature for RCM grid point "2654", the Control Period (HC1 – 1961-1990) and both SRES scenarios (HB1 (B2) and HS1 (A2) – 2071-2100).

Figure 18 Frequency distributions of Predicted Minimum and Maximum temperature at weather station "112700", according to both SRES scenarios (HB1 (B2) and HS1 (A2) – 2071-2100).

6. Discussion

As shown by the Figures above, the Predicted Minimum and Maximum temperatures seem to observe similar distributions as those of the observed Weather Station values, while taking into account the temperature shift corresponding to the RCM scenario runs. In addition, our hypothesis that RCMs simulate relative values better than absolute values seems to be confirmed due to the shorter range of the RCM distributions compared to the observed temperature distributions from weather stations.

The Delta Method developed during this internship seems coherent and thus appears to be a valid method to downscale RCM data and provide relevant results for SWAT.
7. Assessment of the internship

7.1 Problems and difficulties encountered

During the internship, there were a number of occurrences where I was unable to continue working as either the R scripts were incorrect or there were other underlying problems.

- Code-writing and programming: a long adaptation period was necessary in order to become familiar with the R programming environment.
- Computer capacity: computers under Windows 32-bit were unable to load the complete RCM netCDF files used for this study in R. However, by extracting only the data relative to the grid points associated to the closest weather stations, this issue was resolved.
- Data compatibility: it was necessary to alter the input data in order for them to be compatible on a number of levels, most notably to correspond the column names from data using different formats, as well as finding a method to account for the different time periods between the observed temperatures (ranging from 1970 to 2008 with a large amount of missing values) and the RCM data (ranging from 1961 to 1990 with only 360 days a year).
- Original "Annual" Delta Method: the Annual Delta Method proved to be inefficient and was discarded as the results were not accurate enough to be useful. On a large number of occasions, the derived Minimum temperature was higher than the Maximum temperature for the same day. The "Monthly" Delta Method was then created, and considerably reduced the occurrence of this error.
- Output compatibility with SWAT: the "date" format of the input files for SWAT need to be in "real" dates, and not a regular, 30-day a month, 360-day a year file. Therefore it was necessary to add a certain number of gaps in order to transform the predicted temperature datasets into the required format for the complete time run of the method.

7.2 Benefits of the internship

The internship was overall an enriching experience, during which I learned new tools and methods. For obvious reasons, the internship was an invaluable experience on a personal level. I learned a great deal about the organization and the structure of the Geneva offices of UNEP, and was allowed to attend staff meetings and presentations which offered insight into the work environment. The actual tasks of my internship were also very informative, I was able to learn a new software, R, and refine my skills working with ArcGIS, which I had previously worked with. Working with climatic data of the type encountered during this internship will most definitely be useful as well for my future
studies. On a professional level, the internship will undoubtedly enrich my CV, and will hopefully open new doors for a future career that I could very well see go in this direction.

I hope that my internship with the team of the UNEP/DEWA/GRID-Europe, and especially with those involved in the EnviroGRIDS project, was useful for their work and research, and that I was able to produce meaningful and useful results.

8. Conclusion

As mentioned previously, the Delta Method provides relevant results that may be applicable to SWAT and contain more realistic information that corresponds more accurately to weather stations in the area of study in the Black Sea Catchment than using raw RCM data. As the results were calculated using certain statistical characteristics of the RCM data, and applied onto observed time-series, the Delta method is able to render data that correspond more closely to reality. This is evidenced by the fact that the control period data from the RCM is not all that similar to the observed time series. However, it is necessary to be cautious with these results, as a certain amount of uncertainty stills remains. Notably, the daily archives of only one RCM were used, while the validation of the Delta method developed in this study would require further study.

One of the objectives that was not satisfactorily resolved was the general applicability of the method to different studies. Undoubtedly the general “delta” method can be and has been used for other studies, but the lack of uniformity when it came to the data formats and organization made it very difficult to get a notion of the potential of the version of the method developed for this study, in particular concerning the universality of the delivered code and its applicability to different datasets.

Further research is recommended, notably to apply this method using different climatic variables as well as data generated by different climatic models. A study on precipitation using the same Delta method is currently being developed at the University of Geneva. The Delta method is also expected to be developed further to be applicable as a tool for web processing services.
9. Bibliography

9.1 Manuals and articles


GOYETTE, Stéphane, (2010) "Existing data access ad compilation on regional climate, historical records and prospects model", University of Geneva, Geneva


KJELLSTROM, Erik, RUOSTEENOJA, Kimmo, (2007), “Present-day and future precipitation in the Baltic Sea region as simulated in a suite of regional climate models” in Climatic Change,


9.2 Consulted Internet sites

http://www.envirogrids.net/
http://www.grid.unep.ch/
http://prudence.dmi.dk/