NCEP-NCAR Reanalyses Hydroclimatic Data for Rainfall-Runoff Modeling on a Watershed Scale

By Tarana A. Solaiman¹ and Slobodan P. Simonovic²

¹Ph.D Candidate, Department of Civil and Environmental Engineering, The University of Western Ontario, London, Ontario, Canada, N6A5B9, (Email: tsolaima@uwo.ca)
²Professor, Department of Civil and Environmental Engineering; Director, Institute for Catastrophic Loss Reduction, The University of Western Ontario, London, Ontario, Canada, N6A5B9, (Email: simonovic@uwo.ca)

Abstract
The main goal of this paper is to assess the impact of climate change on a watershed scale. The most common approaches of assessing the climate change impact are limited by the uncertainties related to the availability of proper spatial and temporal hydroclimatic data. Global circulation models provide long records but on a coarse spatial scale. The historical observed information, although has better spatial resolution, is limited temporally; hence, cannot be used as an indicator of the future climate change. One way to address these drawbacks is to use reanalysis data and scale it down to local scales. In the presented research, an analysis has been made of the correspondences and/or discrepancies between observed precipitation and temperature data, and the data from the National Centers for Environmental Prediction-National Center for Atmospheric Research (NCEP-NCAR) (a) global (NNGR) and (b) regional (NARR) reanalysis project. The following data between 1980 and 2005 has been extracted for the analysis: daily precipitation, maximum, mean and minimum temperature. The extracted data at several grid points in and around the Upper Thames River watershed in Southwestern Ontario, Canada have been used with a continuous hydrologic model to generate low flows. Both NARR and NNGR temperature (Tmax, Tmin and Tmean) data show a good synopsis of the climate conditions within the study area. The precipitation data from NNGR is less reliable than the NARR. The stream flows generated from the NARR dataset show encouraging result; however, some overestimations are also seen. The uncertainty estimations of the outputs are plotted using variation in mean and variance. The results indicate that the NARR dataset can be used as a good source for interpreting climate variation; nonetheless, a complete knowledge and careful investigations of the differences are necessary for application with the hydrologic models.

1. Introduction
The presence of uncertainty often complicates understanding of the local impacts of climate change. One of the most significant potential impacts of climate change is the intensification of hydrological cycle, i.e the change of river flows, available regional water resources and increase in the magnitude of global and potential runoff (Labat et al. 2004; Zhang et al. 2008). For the long time, climatologists and hydrologists have faced challenges in understanding the link between climate variability and stream flow. A specific synoptic-scale circulation regime creates spatial pattern of the precipitation and stream discharge whose frequency can be either enhanced or suppressed which in turn may generate irregular hydrologic conditions (Keables 1991). These changes in the precipitation patterns combined with natural and anthropogenically-induced climate variations have enormous ecological, societal and economic impacts. An increased near
surface temperature, on the other hand, increases the evaporation rates and accelerates the transport of water vapour in the atmosphere and thus causes change in the hydrologic cycle. Analysis of hydroclimatic variability therefore can provide an insight into the current climatic system.

Unfortunately, weather stations are often sparsely distributed over areas of interest, especially in the mountainous and high latitude regions. Data obtained from these weather stations have limited value for the efficient analysis of the entire climate system in a region. In such instances, alternate data sources, such as, remotely sensed data, data generated by global and regional climate models (e.g. GCMs and RCMs), and reanalysis data such as the National Center for Environmental Prediction – National Center for Atmospheric Research (NCEP-NCAR) Global Reanalysis – NNGR (Kalnay et al. 1996) and North American Regional Reanalysis – NARR (Mesinger et al. 2006) can be viable alternatives to bridge the gap (Choi et al. 2007). The two latter databases are relatively new with finer spatial and temporal resolution than most GCMs.

Several studies have compared the reanalysis precipitation and temperature data with other available databases. Neito et al. (2004) compared the NNGR data with ECHAM4/OPYC3 and HadCAM3 models to analyze the correspondences and/or the discrepancies between the observed winter precipitation data during 1949-2000 for the Iberian Peninsula. NNGR precipitation data captured well the spatial and temporal variability and showed a good agreement with the observed precipitation. However, Tolika et al. (2006) found an inferior agreement but closer interannual variability when NNGR was compared with the GCM-HadAM3P data to examine the suitability of the averaged distributions and the spatial and temporal variability of the winter precipitation in Greece. In many applications the NNGR resolution appeared to be coarse. The North American Regional Reanalysis (NARR) developed by Mesinger et al. (2006) provides climate data in a spatial resolution lower than the NNGR. Castro et al. (2007) applied 57 years of NNGR data with dynamic downscaling using the Regional Atmospheric Modeling System (RAMS) to generate regional climate model (RCM) climatology of the contiguous US and Mexico. They compared the RAMS simulated data with that of the NARR, the observed precipitation and temperature data, and found a good agreement of the NARR data in some parts of the Great Plains.

Interestingly, most of the previous studies were focused on the spatial distributions of the seasonal and/or interannual variability of hydro-meteorological variables only. There have been only a few studies relevant to hydrologic modeling. Woo and Thorne (2006) used temperature and precipitation data from the ERA 40, NNGR and NARR as input to a macroscale hydrologic model to estimate the snowmelt contribution to discharge in the Subarctic Canada and found (i) a cold bias resulting in a later snowmelt peaks and (ii) NARR providing a better representation of the relative flow contribution from different sections of the watershed. Recently Choi et al. (2007) evaluated the monthly and daily NNGR and NARR temperature and precipitation datasets to examine their potential as alternative data source for hydrological modeling in Manitoba. Their study revealed satisfactory performance of the temperature data but a weaker performance of the precipitation data and better performance of NARR than it’s NNGR counterpart. Zhang et al. (2008) applied NNGR data to investigate spatial and temporal patterns of the trends of precipitation maxima in the Yangtze River watershed for 1960-2005 and found a significant increase in the summer precipitation intensity and changing rainfall frequency in the middle and lower reaches of Yangtze River with higher probability of flood and drought hazards in the region. In this study an attempt has been made to evaluate applicability of the global and regional reanalysis temperature and precipitation data for hydrologic modeling in the Upper Thames.
River (UTR) watershed in Southwestern Ontario, Canada. This is, however, an important step towards examining impact of climate change on water resources. A reliable alternative database can reduce uncertainties involved in hydrologic modeling.

2. Study Area and Data
The Upper Thames River watershed (42°35’24”N, 81°8’24”W, 3500 km² area) in the Southwestern Ontario, Canada is the study area for this investigation. Several weather stations around the watershed provide point measurements of weather variables including daily temperature and precipitation. Unfortunately only a few of them have operated over the years to make a reliable database and provide an adequate spatial coverage of variable climatic conditions within the watershed. The spatial distribution of the weather stations is also sparse and does not cover the entire watershed.

For the purpose of comparison, three different datasets have been used. The observed precipitation and temperature data covering the UTR watershed for the period of 1980 – 2005 has been collected from the Meteorological Services of Canada. The observed daily total precipitation, mean, maximum and minimum temperature have been obtained from Environment Canada weather stations in and around the Upper Thames River watershed. The stations are located in Blyth, Dorchester, Exeter, Foldens, Glen Allan, London Airport, St. Thomas, Stratford, Waterloo-Wellington, Woodstock and Wroxeter. The west-central part of the watershed is less well represented due to lack of the data. During the 26 year study period there have been several days with missing records. Those days were eliminated from both station and reanalysis datasets in order to maintain consistency. The Physical Sciences Division of the Earth System Research Laboratory of National Oceanic and Atmospheric Administration (NOAA) provided the NNGR daily precipitation rate (kg m⁻¹ sec⁻¹), mean, maximum and minimum temperature (°K) data on a 2.5° X 2.5° grid. The daily average of three hour mean, maximum and minimum temperature and daily totals of three hour precipitation rate (mm day⁻¹) NARR data at 0.3° X 0.3° grid have been made available through the Data Access Integration of the Canadian Climate Change Scenarios Network of Environment Canada.

3. Hydrologic Modelling
A continuous hydrologic model accounts for long-term water balance in the watershed considering the detailed movement of moisture between conceptualized canopy, surface, soil and groundwater storage reservoirs (Cunderlik and Simonovic, 2007). The model used in this study is a semi-distributed rainfall-runoff model developed based on the computational engine of HEC-HMS (USACE, 2006). In order to simulate stream flow, the observed weather station data and regularly spaced reanalyses hydroclimatic data have been interpolated by inverse distance method to the irregularly spaced sub-catchments within the watershed which take precipitation and temperature as input. The model is divided into seven components/modules as shown in Figure 1a. The continuous based hydrologic model applied to the Upper Thames River watershed is based on the work of Cunderlik and Simonovic (2004, 2005, and 2007). The model consists of thirty two special units, twenty one river reaches and three reservoirs (Figure 1b). The parameter sets for summer and winter seasons are presented in Cunderlik and Simonovic (2004) and Prodanovic and Simonovic (2007). The model has been properly calibrated and verified with extensive sensitivity analyses. The continuous hydrologic model used is bi-seasonal in nature; a
different parameter set is used for summer and winter seasons. Consideration of seasonality provided superior calibration results (Cunderlik and Simonovic, 2004, 2005).

4. Analysis of Results

The analysis of the results contains an evaluation and comparison of daily discharge generated by the HEC-HMS model. First performances of the NNGR and NARR inputs into the hydrologic model have been compared with the statistical goodness-of-fit measures: the root mean square error (RMSE), correlation coefficient (r), normalized mean squared error (NMSE), mean absolute error (MAE) and relative bias (RB). The outputs (daily discharge) have been assessed by flow comparison graphs, scatter plots and confidence interval plots. Because of the existing model and data errors it’s necessary to use appropriate criteria for estimating the relevant uncertainties (Sorooshian, 1993). In this study, only data uncertainty arising from the (i) inconsistency and non-homogeneity and (ii) inadequate representation of the reanalysis data due to space and time limitations has been assessed. The errors arising from the data sources are evaluated by estimating the means and variances.

4.1 Changes in precipitation and temperature

Figure 2a presents the cumulative daily precipitation graphs of NARR and NNGR at Foldens station for year 2000. When all stations are taken into consideration it can be observed that at Foldens and Stratford (not in Figure), NARR is fairly close to the observed precipitation. While at Woodstock (Not in Figure), NNGR data performs well. Although none of the methods has been able to get estimates in Wroxeter (Not in Figure), still NARR values are closer than the NNGR. Figure 2b presents Tmean values interpolated at the Foldens weather station in the watershed. Both reanalysis datasets have shown a tendency to over-estimate the observed values.
in all three cases, especially in summer. However, the NARR data has also overestimated observed values in spring. Except these over-estimations, both datasets are in close agreement with observed temperatures.

Figure 2: Comparison of (a) Cumulative Daily Precipitation and (b) mean monthly temperature in 2000

4.2 Hydrologic model results

Table 1 compares the statistical performance measures of the daily discharge obtained for evaluating the performances of the reanalyses data. The root mean square error for both NNGR and NARR varies considerably, from 4 cumec (NNGR) and 3.44 cumec (NARR) at Ingersoll to 28.097 cumec (NNGR) and 24.37 cumec (NARR) at Byron. The correlation coefficients produced by NARR (0.59-0.65) are significantly higher than those from NNGR (0.41-0.44). The normalized mean square error is also slightly higher in case of NNGR. The absolute mean error is differentiable both in terms of the data types and locations. NNGR produces higher errors. MAE seems to decrease with the increase of subwatersheds that are contributing to the total runoff. The values of the relative bias differ greatly at the selected locations with the NARR producing negative bias unlike it’s counterpart. The bias produced by the NNGR data is much higher, ranging from 26% to 45% to that of -12% to -7% from NARR.

<table>
<thead>
<tr>
<th>Location</th>
<th>RMSE (cumec)</th>
<th>r</th>
<th>NMSE (cumec)</th>
<th>MAE (cumec)</th>
<th>RBias (%)</th>
<th>RMSE (cumec)</th>
<th>r</th>
<th>NMSE (cumec)</th>
<th>MAE (cumec)</th>
<th>RBias (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Byron</td>
<td>28.097</td>
<td>0.44</td>
<td>1.03</td>
<td>15.73</td>
<td>31</td>
<td>24.37</td>
<td>0.65</td>
<td>0.77</td>
<td>9.95</td>
<td>-12</td>
</tr>
<tr>
<td>Ingersoll</td>
<td>4.2875</td>
<td>0.41</td>
<td>1.25</td>
<td>2.62</td>
<td>45</td>
<td>3.44</td>
<td>0.63</td>
<td>0.80</td>
<td>1.57</td>
<td>-7</td>
</tr>
<tr>
<td>St. Marys</td>
<td>10.08</td>
<td>0.44</td>
<td>0.97</td>
<td>5.23</td>
<td>26</td>
<td>10.04</td>
<td>0.59</td>
<td>0.97</td>
<td>3.72</td>
<td>-9</td>
</tr>
</tbody>
</table>

The poor model performance at Byron can be attributed to the fact that this part of the watershed suffers from inadequate meteorological data which may have restricted adequate representation of the situation. Apparently, Byron is the outlet of the watershed with a contributing area of 3,110 km² (Cunderlik and Simonovic, 2004) with 32 subwatersheds. Although there were over
estimations and underestimations, it appears that the NARR data successfully captures the spatial representation of the model.

4.3 Error evaluation

The absolute model errors (absolute difference between the observed and simulated data) in the estimates of mean daily discharge for the dry months (May-October) are illustrated in Figure 3a for Ingersoll station. The graphical analyses of the mean for all stations indicate that the error caused by NNGR data is significant especially during the summer with highest error during July; whereas NARR has produced an impressive performance during the summer. The error eventually increases at the beginning and end of summer period with the increase of discharge from snowmelt during the spring and winter seasons. The results indicate how the choice of these data sources affects the generation of future climate scenarios. Overall, for current period (2001-2005), NNGR precipitation is likely to produce discharge with significant prediction error compared to the observed discharge in calculating the monthly means. Next, comparison analyses of the variance of the observed and predicted discharge for the dry months are plotted (Figure 3b) for Ingersoll. The graphical representation at Ingersoll indicates that NNGR does not capture the variability that is closer to the observed discharge. Higher variability is observed during the late summer. Significantly lower variability has been detected at St. Marys station (Not in Figure). In case of NARR data, results are better; lower variability is observed during the summer months with the start and end of summer providing slightly worse result.

Figure 3: (a) Absolute error and (b) variance in daily discharge from reanalysis data at Ingersoll (2001-2005)

5. Conclusion

In this paper, an attempt has been made to verify the performance of the NCEP global and regional reanalysis data under present climate conditions for precipitation, maximum, minimum and mean temperatures in the Upper Thames River watershed in the South-western Ontario, Canada. The results from this study are in general agreement with the literature. Overall, the results for the reanalysis precipitation data considered are as reasonable as previous work performed in different regions of Canada and around the globe (Choi et al, 2007;) which indicates that precipitation data from the reanalyses projects are fairly consistent. The performances of NARR and NNGR data are less satisfactory than expected. In some occasions the difference between observed and reanalyses data is -8 to 7 °C for NNGR and -8 to 12 °C for
NARR for Tmean. Such discrepancies may lead to substantial errors in simulating snowmelt runoff during winter and spring time. This is however more prominent in NARR than NNGR. Radiational cooling is important for the minimum temperature that results in an inversion into the lowest layers. In areas such as UTR watershed with significant presence of wind, it is speculated that the reanalysis might not allow for enough radiational cooling (Rusticucci and Kousky, 2002) and may result in such overestimation. The hydrological model results indicate that NARR and NNGR produced similar RMSE and NMSE in simulating daily discharges. The correlation coefficients of NARR are higher than that of NNGR. However, the differences between the two datasets are more prominent in comparisons of their relative bias; NNGR is associated with significant bias than it’s NARR counterpart. The NARR, whereas, produced insignificant negative bias at all locations which may be due to insufficient meteorological inputs that has restricted the representation of the real watershed conditions. However, the variance comparison does not show the same pattern for NNGR data set; the peak error in estimating mean occurs during the month of July while the higher variability is shifted forward towards August. In the case of NARR data set, results are better; lower variability is observed during the summer months with slightly deteriorating results at the start and at the end of summer. Based on the following observations, it can be added that the main difference in simulating discharge using NARR and NNGR data sets lies in their different process of generating precipitation. NARR data has been produced by assimilating high quality and detailed precipitation observations into the atmospheric analysis, thus making the forcing into the land surface model component of the system more accurate by enabling the interaction of the land hydrology and land-atmosphere. The NNGR precipitation is far from the observed values quite often during the year. The issue of the frequency of precipitation events also needs to be taken care of. It is important to note here that the continuous model is developed to generate low flows, taking the soil moisture balance into account. So the higher flows generated due to a single large event precipitation will show poor agreement, which may lead to misinterpretation of the results. This issue is more important in the regions like Upper Thames River watershed which are prone to heavier precipitation in the form of convective storms from nearby Great Lakes in winter and severe thunderstorms in summer and fall, that makes the whole prediction process with large weather models difficult.

Based on the above analysis and discussion, the following conclusions can be made:

1) The performance of NARR data set in simulating low flows in the Upper Thames River watershed is satisfactory;
2) NNGR data set with its coarse resolution is not appropriate for small watersheds such as the study area considered;
3) The reanalysis data needs careful investigation before their application with hydrologic models and can be used only with a complete explanation of the possible sources of difference between them and the observed data.

Acknowledgements
Financial assistance of the Canadian Foundation for Climate and Atmospheric Sciences is thankfully acknowledged. The authors also wish to gratefully thank Mr. Patrice Constanza and Mr. Louis Lafaivre of DAI/CCCSN/EC for providing NARR data for the work. Thanks are also due to Mr. Don Hooper of PSD/NOAA for his valuable suggestions.
References


