AR4 Climate Model Performance in Simulating Snow Water Equivalent over Catskill Mountain Watersheds, New York, USA

AAVUDAI ANANDHI1 SONI M. PRADHANANG1 MARK S. ZION2 DONALD C. PIERSON2 ALLAN FREI1,3 ELIOT M. SCHNEIDERMAN2 RAJITH MUKUNDAN1 ADAO MATONSE1

ABSTRACT:

The ability of Global Climate Models (GCMs) to simulate observed meteorologic and hydrologic variables is an important indicator of the reliability of these models to project future climate conditions. In this study we evaluate the ability of GCMs participating in the Intergovernmental Panel for Climate Change’s (IPCC) Fourth Assessment Report (AR4) to simulate variability in the snow water equivalent (SWE) in New York City Water Supply watersheds located northwest of NYC in the Catskill Mountains. SWE is estimated using an empirical temperature-based degree day model. Inputs to this model are either measured historical meteorological (1961-2000) data or GCM model output for the same historical period. The evaluation of the GCMs is based on a skill score developed using probability distribution functions derived from the time series of simulated snowpack. From the skill scores (SS) calculated, the GCMs are ranked based on their ability to simulate the snowpack. These results have implications for selecting a subset of GCM simulations for climate change impact assessments in New York City’s water supply.

Results show that the GFDL 2.0 (gf001) model has the highest SS (0.93) and CCSM (ncc09) model has the lowest SS (0.26). Based on the SS, the GCM ensemble members are classified into three categories high, medium and low performance. The PDFs for the three performance classes show the largest between-model variability for models in low performance class. Differences between the means and medians of observation-based and GCM-based simulation were also greatest in the low performance class.

Keywords: Snow water equivalent (SWE), Evaluation AR4 models, Global climate models (GCMs), Probability based skill score, temperature based snowmelt algorithm, GWLF

INTRODUCTION

Snowmelt runoff is an important source of water in the watersheds of the New York City water supply that provide about ninety percent of the New York City’s (NYC) daily water demand. One hydrologic change that has been observed in this region during the period 1952-2005 is a shift in the timing of snowmelt runoff to earlier in the year (Burns et al., 2007) and more extreme winter/spring runoff events. If the climate continues to change in future, the contribution of spring

1 CUNY Institute for Sustainable cities, City University of New York, New York, NY, 10065.
2 Water Quality Modeling Group, New York City Department of Environmental Protection, 71 Smith Avenue, Kingston, NY 12401.
3 Hunter College, City University of New York, New York, NY, 10065.
snowmelt to streamflow may also change. Given the fact that changes in snowmelt runoff in Catskill Mountain (West of Hudson (WOH)) watersheds have potentially important implications for the water supply of New York City, there is a need to study the potential impacts of climate change on the quantity of snowmelt runoff in these watersheds. For this purpose, data derived from a suite of Global Climate Models (GCMs) are being used to drive watershed models to study snowmelt runoff in the absence of observed snow data.

Presently, the outputs from GCMs, related to snow and snowmelt are only available at monthly timescales. Snow cover fraction (SCF), an output from GCM is diagnostically derived from prognostic variables: snow water equivalent (SWE) or snow depth (SD). The details of the studies that have examined GCMs with respect to snow are given in Table 1. From the table it can be observed that snow simulations (SWE, surface albedo, SD, SCF, snow mass, snow cover area) from GCMs are evaluated at monthly, seasonal and annual timescales using methods such as annual cycle, frequency distribution, mean, median, decadal scale variability (Foster et al., 1996; Yang et al., 1999; Frei et al., 2003; Frei et al., 2005; Frei and Gong, 2005; Roesch, 2006; Roesch and Roeckner, 2006). No studies have focused on the evaluating the ability of GCM in simulating modeled snow accumulation and melt at daily timescales.

Table 1 Literature review

<table>
<thead>
<tr>
<th>SN</th>
<th>Variable name</th>
<th>Region of study</th>
<th>Time scale</th>
<th>Evaluation metric</th>
<th>GCMs</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Snow cover, Snow mass</td>
<td>North America, Eurasia</td>
<td>Mean monthly</td>
<td>Climatology plots</td>
<td>Hadley centre, CGCM, GENESIS, ECHAM, GISS, GLA, ARIES</td>
<td>(Foster et al., 1996)</td>
</tr>
<tr>
<td>2</td>
<td>Snow mass, extent</td>
<td>Mid-latitude Grasslands in Russia, California</td>
<td>Month</td>
<td>Monthly time series plots</td>
<td>CCSM (BATS)</td>
<td>(Yang et al., 1999)</td>
</tr>
<tr>
<td>3</td>
<td>Snow cover area (SCA)</td>
<td>North America, Eurasia</td>
<td>Month</td>
<td>Interannual variability</td>
<td>18 GCMs participating in AMIP-1</td>
<td>(Frei et al., 2003)</td>
</tr>
<tr>
<td>4</td>
<td>SCA</td>
<td>North America</td>
<td>Annual, Decadal</td>
<td>mean, Decadal scale variability (DSV)</td>
<td>21 GCMs participating in IPCC-AR4</td>
<td>(Frei and Gong, 2005)</td>
</tr>
<tr>
<td>5</td>
<td>Snow water equivalent (SWE)</td>
<td>North America</td>
<td>Month, seasonal</td>
<td>Box and whisker plots, monthly mean and standard deviation, Pearson correlation coefficient</td>
<td>18 GCMs participating in AMIP-1</td>
<td>(Frei et al., 2005)</td>
</tr>
<tr>
<td>6</td>
<td>SWE, Snow cover fraction (SCF), Surface albedo</td>
<td>North America, Eurasia</td>
<td>Month, seasonal</td>
<td>Annual cycle, biases, frequency distribution</td>
<td>Most GCMs participating in IPCC-AR4</td>
<td>(Roesch, 2006)</td>
</tr>
<tr>
<td>7</td>
<td>SCF, snow depth (SD)</td>
<td>Eurasia</td>
<td>Month</td>
<td>Annual cycle, frequency distribution</td>
<td>ECHAM4, ECHAM5</td>
<td>(Roesch and Roeckner, 2006)</td>
</tr>
</tbody>
</table>

The watershed models used to study the hydrology of this region are run at daily timescales. Higher resolution snow data can however be obtained indirectly by modeling SWE at daily timescales using daily simulations derived from GCMs. Different approaches of varying complexity ranging from simple regression equations, blackbox approaches based only on temperature measurements to physics-based models containing equations for all the processes involved or complete multilayer models based on an energy balance (Stewart, 2009; Zeinivand...
and Smedt, 2009; Debele et al., 2010) have been used. In this study, daily snowmelt is estimated using the temperature based approach available in Generalized Watershed Loading Function (GWLF; Haith et al., 1992) watershed model.

Given that a relatively large number of GCMs that are presently available, using the results from all GCMs may result in an unreasonable number of watershed model simulations. To avoid this problem a subset of GCMs can be selected by GCM evaluation. Testing the GCM’s ability to simulate “present climate” (including variability and extremes) is an important part of model evaluation (Randall et al., 2007a). In this study GCMs are evaluated by examining the skill of models in simulating present-day climate (Raisanen, 2007; Johnson and Sharma, 2009). A number of studies have used methods such as skill scores, or other criteria statistics for evaluating the different meteorological variables available from GCMs simulations (Giorgi and Mearns, 2003; Tebaldi et al., 2004; Murphy et al., 2007; Perkins et al., 2007; Randall et al., 2007b; Maxino et al., 2008; Johnson and Sharma, 2009). A good review of these methods available to evaluate the performance of GCMs is found in (Johnson and Sharma, 2009). No studies have focused on the evaluating the ability of GCM in simulating modeled snow accumulation and melt using skill scores.

The objective of this study is to evaluate the ability of daily GCM-derived SWE to simulate daily observation-based SWE using a probability based skill score. Daily snow accumulation or snowmelt is estimated using the temperature based snow algorithm in GWLF watershed model.

STUDY REGION AND DATA USED

The study area encompasses a watershed area of about 4100 km². It consists of six reservoir watersheds namely Cannonsville, Ashokan, Nerversink, Schoharie, Rondout and Pepacton (Figure 1). These watersheds are part of the Eastern Plateau Climate Region of New York. The regional climate is characterized as humid continental with cool summers, cold winters, abundant precipitation and snowfall. It experiences a uniform distribution of precipitation throughout the year. Typically, total precipitation in the region is about 1000-1200 mm per year, with snowfall accounting for approximately 20 percent of total precipitation. In addition, orography influences the spatial distributions of precipitation and temperature (Frei et al., 2002; Burns et al., 2007).

Figure 1. Study region. The six reservoir watersheds provide approximately 90% of NYC drinking water needs.

For each of the six WOH watersheds, daily observed data for precipitation from National Climate Data Center cooperator stations was obtained from the Northeast Regional Climate Center (NRCC). Each watershed was broken into Theissen polygons based on the location of the nearest precipitation stations. The proportion of the area of each polygon representing a precipitation
station to the total watershed area is the weight given for averaging the value from that station. Using this method the watershed average precipitation is calculated. Average air temperatures are derived from four stations measuring this variable, Cooperstown, Liberty, Slide Mountain, and Walton. Each of these stations has been active since 1965 or earlier. The averaging method includes the application of an environmental lapse rate to correct for elevation differences between the station and the mean watershed elevation and use of inverse distance squared weighting averaging of the four stations (NYCDEP, 2004). After processing the observed daily precipitation and average temperatures, a single time series for a variable and watershed is obtained and used in this study. The period of observed data used in this study is 1960-2000.

GCM simulations are obtained from the World Climate Research Programme's (WCRP's) Coupled Model Intercomparison Project phase 3 (CMIP3) multi-model dataset. The daily baseline scenario (20C3M) GCM simulations are from 20 GCMs (Table 2), and for two meteorological variables (precipitation and average temperatures at the surface). A list of the GCM simulations (name and realization number), used in the study are provided in Table 2. The data from all the GCMs for the region surrounding the study region are extracted and interpolated to a common 2.5º grid using bilinear interpolation (this is implemented with NCAR Command Language www.ncl.ucar.edu).

### Table 2 Names of the Climate models, their versions, realization numbers, acronym used in the study.

The GCMs are classified based on their performance in simulating the snow water equivalent and are shown in different colors in column realization number. All snow simulations were made using baseline runs associated with these models

<table>
<thead>
<tr>
<th>S.N</th>
<th>GCM I.D *</th>
<th>Acronym</th>
<th>Country</th>
<th>GCM name</th>
<th>Realization number**</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BCCR-BCM2.0</td>
<td>bcc</td>
<td>Norway</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>2</td>
<td>CCSM3</td>
<td>nc</td>
<td>USA</td>
<td></td>
<td>01,03,05,06,07,08,09</td>
</tr>
<tr>
<td>3</td>
<td>CGCM3.1(T47)</td>
<td>cc4</td>
<td>Canada</td>
<td></td>
<td>01,02,03,04,05</td>
</tr>
<tr>
<td>4</td>
<td>CGCM3.1(T63)</td>
<td>cc6</td>
<td>Canada</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>5</td>
<td>CNRM-CM3</td>
<td>cnr</td>
<td>France</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>6</td>
<td>CSIRO-Mk3.0</td>
<td>cs0</td>
<td>Australia</td>
<td></td>
<td>01,02,03</td>
</tr>
<tr>
<td>7</td>
<td>CSIRO-Mk3.5</td>
<td>cs5</td>
<td>Australia</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>8</td>
<td>ECHAM5/MPI-OM</td>
<td>mpi</td>
<td>Germany</td>
<td></td>
<td>01,04</td>
</tr>
<tr>
<td>9</td>
<td>ECHO-G</td>
<td>miu</td>
<td>Germany=Korea</td>
<td></td>
<td>01,02,03</td>
</tr>
<tr>
<td>10</td>
<td>FGOALS-g1.0</td>
<td>iap</td>
<td>China</td>
<td></td>
<td>01,03</td>
</tr>
<tr>
<td>11</td>
<td>GFDL-CM2.0</td>
<td>gfl</td>
<td>USA</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>12</td>
<td>GFDL-CM2.1</td>
<td>gfl</td>
<td>USA</td>
<td></td>
<td>02</td>
</tr>
<tr>
<td>13</td>
<td>GISS-AOM</td>
<td>gao</td>
<td>USA</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>14</td>
<td>GISS-ER</td>
<td>gir</td>
<td>USA</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>15</td>
<td>INGV-SXG</td>
<td>ing</td>
<td>Italt</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>16</td>
<td>IPSL-CM4</td>
<td>ips</td>
<td>France</td>
<td></td>
<td>01,02</td>
</tr>
<tr>
<td>17</td>
<td>MIROC3.2(hires)</td>
<td>mih</td>
<td>Japan</td>
<td></td>
<td>01</td>
</tr>
<tr>
<td>18</td>
<td>MIROC3.2(medres)</td>
<td>mim</td>
<td>Japan</td>
<td></td>
<td>01,02</td>
</tr>
<tr>
<td>19</td>
<td>MRI-CGCM2.3.2</td>
<td>mri</td>
<td>Japan</td>
<td></td>
<td>01,02,03,04,05</td>
</tr>
</tbody>
</table>


** Realization number highlighted in yellow are classified as models having high skill scores, the red numbers represent the models classified as medium skill score and the numbers in black represent models having low skill scores. This classification is subjective

### METHODOLOGY

The methodology followed in this study is shown in Figure 2 and explained below.
SWE estimation using Generalized Watershed Loading Functions (GWLF) model

Due to scarcity of measured snow data in the study area and comparable snow parameters in the GCM simulations, comparisons presented in this paper are based on simulated snow parameters. These simulations are driven using either observed daily mean measurement of air temperature and precipitation or daily GCM data for these variables from the GCM grid cell nearest to the study area and are referred as “observation-based” SWE and “GCM-based simulated” SWE respectively.

GWLF is a lumped parameter hydrologic model coupled to simple water quality model and details of the model may be found in Haith et al (1992) and Schneiderman et al (2002). Runoff is further distributed by topographic index. GWLF is driven by daily precipitation and temperature data. For six reservoir watersheds, six separate GWLF model applications are driven using watershed averaged precipitation and air temperature, and are calibrated and validated by comparing simulated and measured streamflow at the watershed outlet. In the absence of snow measurements, these simulations are used to provide a surrogate for observed SWE and snowmelt corresponding to present day conditions. They are referred as “observation-based” SWE and snowmelt.

In GWLF, snow water equivalent (SWE) at a given time \( t \), is a function of SWE at a previous time \( t-1 \), snowfall \([P(t)]\) and snowmelt \([P_s(t)]\) at time \( t \) (cm). If the mean daily temperature \( T(t) \) is less than or equal to \( 0^\circ C \) precipitation is assumed to be snowfall. If \( T(t) > 0^\circ C \), snowmelt \( P_s(t) \) is calculated based on equation 1.

\[
P_s(t) = M \cdot T(t)
\] (1)
Snowmelt is a function of mean daily temperature and a snowmelt parameter or degree day factor \( (M) \) given in equation (1). \( M \) depends on basin geographical location, time of year, vegetation and topography (Maidment, 1993). The snowmelt degree day factor for the six validated models one for each six WOH watersheds are given in Table 3. The calibrated parameter varies between 0.29 and 0.48 based on the comparison of simulated and measured streamflow. They are within the range of values typical of this region (Maidment, 1993).

\[
\text{Table 3 Melt coefficients calibrated for the six WOH watersheds}
\]

<table>
<thead>
<tr>
<th>S.N</th>
<th>Name of reservoir watershed</th>
<th>Melt coefficient cm/day</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ashokan</td>
<td>0.29</td>
</tr>
<tr>
<td>2</td>
<td>Cannonsville</td>
<td>0.41</td>
</tr>
<tr>
<td>3</td>
<td>Neversink</td>
<td>0.48</td>
</tr>
<tr>
<td>4</td>
<td>Pepacton</td>
<td>0.39</td>
</tr>
<tr>
<td>5</td>
<td>Rondout</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>Schoharie</td>
<td>0.38</td>
</tr>
</tbody>
</table>

The precipitation and average temperature obtained from GCM simulations were also input into each of the six calibrated GWLF models to obtain snowmelt and SWE for the various combinations of GCM/realizations. These simulations are referred to as GCM-based SWE.

**Estimation of probability density functions (PDFs) of SWE:**

The PDFs of SWE are estimated using “observation-based” and “GCM-based simulated” SWE for the months December to March (winter to early spring) using MatLab (http://www.mathworks.com). Six simulated time series of SWE were developed based on observed data (daily basin-wide averaged precipitation and air temperature) for each of the six watersheds. All the values in the six time series are used to construct the representative distribution in the observation-based PDF. The PDFs were also calculated for each reservoir watershed using daily grid cell air temperature data for each GCM / realization.

To estimate the PDFs used in this study, we require bin sizes \( (S_b) \) and number of bins \( (N_b) \). For each variable, a common value of \( S_b \) is used for all analyses; \( N_b \) is then determined based on the range of values (equation 2). \( S_b \) selected for this study are 0.5 mm/day for snow water equivalent, 0.5 mm/day for snowmelt and 0.5°C for average temperature.

\[
N_b = \frac{(V_{\text{max}} - V_{\text{min}})}{S_b}
\]

where \( V_{\text{max}} \) and \( V_{\text{min}} \) is the maximum and minimum value of the variable and vary for the different combinations. The frequencies of values within each bin (\( n \)) is then calculated for GCM-based (\( F_{g,n} \)) and observation-based (\( F_{o,n} \)) data.

**Skill score (SS)**

The ability of the GCMs to estimate snowmelt and SWE was judged using the skill score (SS) developed by Perkins et al, (2007) which computes the empirical PDFs derived from observation-based and GCM-based simulation. The advantage of this skill score is its simplicity and applicability across variables, spatial scales and seasons. For each bin \( n \), in the SWE frequency distribution the minimum frequency associated with either the GCM-based (\( F_{g,n} \)) or observation-based (\( F_{o,n} \)) data is recorded. SS is the summation of these minimum frequency values over all bins (equation 3).

\[
SS = \sum_{n=1}^{N_b} \min(F_{g,n}, F_{o,n})
\]
across variables and is easily interpreted. For SWE, each GCM/realization gets one SS. The skill scores are then ranked.

RESULTS AND DISCUSSION

The skill scores for snow water equivalent (SWE) are estimated for 41 GCM/realizations used in the study. The ranks of the GCM are provided in Figure 3 where, the x-axis denotes the rank and y-axis the SS. The GFDL 2.0 (gf001) has the highest SS (0.93) and CCSM (ncc09) has the lowest SS (0.26). Based on the SS, the GCMs are classified into three categories high, medium and low performance. The classification is based on the changes in the SS and is subjective. In Table 2, the GCMs and realization number which are classified as high, medium and low performance are differentiated by color. It can be observed that the SSs are generally consistent between ensemble members of each GCM, however in some cases they fall between adjacent classes and no GCM has one ensemble member in the highest performance group and one in the lowest performance group. The range of the SS in the three categories are high skill score: 0.87-0.93, medium skill score: 0.72 - 0.83 and low skill score: < 0.72 (0.26 - 0.72).

GCM ensemble members with high skill score are gfdl2.0(gf001), ipsl(ips:01,02), csiro(cs0:01,02), cgem3(T47)(cc4:02,04), gissam(gao01), mrieam(mri02, mri03), cnrm-cm3 (cnr01).

GCM ensemble members with medium skill score cgem3(T63)(cc601), ECHAM (mpi:01,04), echo (miu: 01,02,03), ingv(ing01), mri-cgem(mri:01,04,05), ccs3 (ncc:01,03,05,06,07,08), fgoals-g1.0(iap01), bccrm2(bcc01).

GCM ensemble members with low skill score miroc hires (mih01), miroc medres(mim:01,02), csiro3.5(cs501), giss-er(gir01), fgoals-g1.0(iap03), ccs3(ncc09).

In Figure 4, the PDFs of SWE, snowmelt and air temperature for the three categories defined in Figure 3 are plotted. In Figure 4, a separate row is plotted for each variable and columns represent different performance class. In each panel, the range of the GCM-based PDFs (shaded region) is shown along with the observation-based PDF (bold line). For snowmelt and SWE (row 2 and 3) the x-axis is transformed using a natural log. These figures suggest that:

The largest between-model variability, are found for models in low performance class and lowest between-model variability, are found for models in high performance class.

For the three variables, the median and mean lines obtained from all GCM-based simulations are more representative to observation-based snowmelt, snow water equivalent and mean temperature for high performance class followed by medium and then low performance class.

The mean and median lines in each class are closer to each other in high performance class when compared to low performance class.
The differences between the high and medium performance class is less when compared to the differences between medium and low performance class.

From the basic statistics (such as mean, median, standard deviation, interquartile range) estimated for all the GCM-based SWE and observation-based SWE, it can be inferred that the mean statistics in the GCM-based SWE in high performance class were more representative of observation-based SWE when compared to the other two performance classes.

Figure 4. The shaded region represents the variation in Probability Density Functions (PDFs) for average temperature, snowmelt and snow water equivalent (SWE) for the various AR4 climate models in the three performance classes considered in the study. The PDFs are estimated for the period 1962-1999 for the DJFM months. In each of the plots, the black bold line represents the PDF obtained using daily observation-based simulation for the study region. The red dashed line represents the median PDF and the red line shows the mean PDFs for the GCM-based simulation.
CONCLUSIONS

Snowmelt runoff is an important source of water for New York City’s (NYC) water supply. The GCMs participating in the IPCC’s AR4 report are evaluated for their performance in simulating snow water equivalent in the water supply watersheds using probability based skill scores. In the absence of observed daily snow water equivalent (SWE) or comparable GCM simulated daily SWE, SWE is estimated using a simple watershed model, which includes a degree-day snow melt parameterization.

Results show that SSs are generally consistent between ensemble members of each GCM. The GFDL 2.0 (gf001) has the highest SS (0.93) and CCSM (ncc09) has the least SS (0.26). Based on the SS, the GCM ensemble members are classified into three categories high, medium and low performance. It is observed that the SSs are generally consistent between ensemble members of each GCM. The range of the SS in the three categories are 0.87- 0.93 for high skill score, 0.72-0.83 for medium skill score and <0.72 (0.26 - 0.72) for low skill score.

The PDFs of snowmelt, SWE and mean temperature for the three performance classes show the largest between-model variability for models in low performance class. Differences between the mean and median from GCM-based PDFs and observation-based PDFs were also greatest in the low performance class.

The statistics (such as mean, median, standard deviation, interquartile range) from the GCM-based SWE simulation were more representative of observation-based SWE in high performance class when compared to the other two performance classes.

ACKNOWLEDGEMENT

We acknowledge the modeling groups, the Program for Climate Model Diagnosis and Intercomparison (PCMDI) and the WCRP's Working Group on Coupled Modelling (WGCM) for their roles in making available the WCRP CMIP3 multi-model dataset. Support of this dataset is provided by the Office of Science, U.S. Department of Energy. The GIS support from David G. Lounsbury and data support from Don Kent is also acknowledged.

REFERENCES


