Climate simulations of major estuarine watersheds in the Mid-Atlantic region of the United States

R. Najjar*, L. Patterson¹, and S. Graham²

¹Department of Meteorology, The Pennsylvania State University, University Park, PA
²Population Research Institute, The Pennsylvania State University, University Park, PA

*Corresponding author: Department of Meteorology, 503 Walker Building, University Park, PA, 16802-5013, 814-863-1586 (tele), 814-865-3663 (fax), najjar@meteo.psu.edu

Running head: Climate of Mid-Atlantic watersheds

Submitted to Climatic Change, August 15, 2007

Revised Version: June 23, 2008
Abstract

To better understand the implications of anthropogenic climate change for three major Mid-Atlantic estuaries (the Chesapeake Bay, the Delaware Bay, and the Hudson River Estuary), we analyzed the regional output of seven global climate models. The simulation given by the average of the models was generally superior to individual models, which differed dramatically in their ability to simulate 20th-century climate. The model average had little bias in its mean temperature and precipitation and, except in the Lower Chesapeake Watershed, was able to capture the 20th-century temperature trend. Weaknesses in the model average were too much seasonality in temperature and precipitation, a shift in precipitation’s summer maximum to spring and winter minimum to fall, interannual variability that was too high in temperature and too low in precipitation, and inability to capture the 20th-century precipitation increase. There is some evidence that model deficiencies are related to land surface parameterizations. All models warmed over the 21st century under the six greenhouse gas scenarios considered, with an increase of 4.7 ± 2.0° C (model mean ± 1 standard deviation) for the A2 scenario (a medium-high emission scenario) over the Chesapeake Bay Watershed by 2070-2099. Precipitation projections had much weaker consensus, with a corresponding increase of 3 ± 12% for the A2 scenario, but in winter there was a more consistent increase of 8 ± 7%. The projected climate averaged over the four best-performing models was significantly cooler and wetter than the projected seven-model-average climate. Precipitation projections were within the range of interannual variability but temperature projections were not. The implied research needs are for improvements in precipitation projections
and a better understanding of the impacts of warming on streamflow and estuarine ecology and biogeochemistry.
1. Introduction

Estuaries are heavily stressed due to a wide variety of anthropogenic activities, including nutrient over-enrichment, altered sediment delivery from land, wetland loss due to development, over-fishing, introduction of invasive species, altered streamflow, and increases in toxic substances (UNEP/GPA, 2006). These stresses have resulted in nuisance algal blooms, decreased oxygen concentrations, loss of benthic habitat, and dramatic shifts in species distributions, which are typically accompanied by a decrease in biodiversity. There are ongoing efforts to restore a number of estuaries, but only recently has it been recognized that climate change has the potential to dramatically alter the outcomes of such efforts. This awareness is the result of the growing consensus on the causes of long-term climate change (IPCC, 2007) combined with a body of research that has demonstrated the importance of climate-related forcing (e.g., freshwater flow, temperature and sea level) to estuaries. Freshwater flow is probably the largest driver of water quality and ecological variations in most estuaries (e.g., Cloern et al., 1983; Gillanders and Kingsford, 2002; Hagy et al., 2004; Howarth et al., 2000; Kimmel et al., 2006; Miller and Harding, 2007; Rabalais et al., 2001; Sharp et al., 1986). Temperature is also an important climate index, and there is substantial laboratory and field-based evidence of its influence on estuarine ecology and biogeochemistry (Kennedy and Mihursky, 1971; Lomas et al., 2002; Secor et al., 2000). Sea-level rise is yet another way in which climate affects estuaries; because bank erosion is an important source of sediment in estuaries (e.g., Cronin et al., 2003a) and because sea-level rise is likely to increase erosion, increased suspended sediment levels and decreased water clarity are likely consequences of sea-level rise accelerated by increased greenhouse gases. If
coastal managers want to account for climate change in their restoration efforts, they need a working knowledge of the main tools used to predict climate change: global climate models (GCMs). When using projections of these models, managers need to be aware of model limitations, which can be assessed through an evaluation of their ability to simulate the climate of the region of interest.

Here we present an evaluation of climate models and an analysis of their future projections over the watersheds of three major Mid-Atlantic estuaries: the Chesapeake Bay, the Delaware Bay and the Hudson River Estuary (Figure 1). The ecological, economic and cultural importance of these estuaries stems from their large size, the extensive fisheries they are (or have been) home to, their location along major bird migratory routes, and the large human population living inside or in close proximity to their watersheds, including many major metropolitan areas (New York City, Philadelphia, Baltimore and Washington D.C.). The estuaries can be briefly characterized by a few superlatives: the Chesapeake Bay is the largest estuary in the United States (Schubel and Pritchard, 1986) and supports the largest military port in the world; the Delaware Bay is home to the world’s largest freshwater port and largest population of horseshoe crabs; and the Hudson River Estuary borders the largest U.S. city (New York) and has the highest nutrient loading per estuarine volume or area among large estuaries in the United States (NRC, 1993).

These estuaries are at various stages of degradation resulting from anthropogenic activity. For much of the 20th century, the Hudson River Estuary and the Delaware Bay suffered severe water quality problems, exemplified by low dissolved oxygen levels resulting from substantial organic matter (sewage) inputs from New York City and
Philadelphia, respectively. As a result of improvements in wastewater treatment, oxygen concentrations have rebounded toward much healthier levels in both estuaries (Albert, 1988; O'Shea and Brosnan, 2000). Other indicators of estuarine health generally show improvements in these two estuaries, though there are still significant concerns, such as high sediment toxicity (Steinberg et al., 2004; USEPA, 2000). The Chesapeake Bay has also suffered from low dissolved oxygen levels, but has shown no sign of improvement, presumably because excess inorganic nutrient loading has been the main cause, and this has continued to increase (Hagy et al., 2004). Other pollution-related problems in the Chesapeake Bay include excess chlorophyll concentrations, chemical contaminants, poor water clarity, degraded benthic habitats, and low fish and shellfish abundance (Chesapeake Bay Program, 2007).

Our focus is on watersheds because their climates (temperature and precipitation, in particular) determine freshwater inflow to estuaries, which is closely linked to the loading of nutrients and sediment. Furthermore, estuarine temperature at interannual timescales often tracks regional air temperature (Preston, 2004; Secor and Wingate, 2008), and so projections of the air temperature change of moderate sized watersheds (such as those considered here) may be applicable to their corresponding estuaries. Sea-level rise, a third way in which climate affects estuaries, is considered in a companion study (Wu et al., 2008).

The bulk of the results are presented for the Chesapeake Bay Watershed (CBW), which makes up 73% of the study region. Results for other watersheds are shown when they differ substantially from the CBW. Because of the importance of the Susquehanna River, which provides about half of the freshwater to the Chesapeake Bay and is the only
river to empty directly into the Bay’s mainstem, we further subdivide the CBW into the Susquehanna River Basin and the Lower Chesapeake Watershed (Figure 1) in some presentations. This separation also allows some of the spatial variability in this large watershed to be captured.

This study complements and extends previous work examining climate simulations of the Mid-Atlantic region. Several studies utilized projections of GCMs forced by a doubling of atmospheric CO$_2$ over different domains: the Delaware River Basin (McCabe and Ayers, 1989), the New England/Mid-Atlantic region (Moore et al., 1997), and the Susquehanna River Basin (Najjar, 1999). Two transient studies have been conducted: two GCMs forced by CO$_2$ increases of 1% yr$^{-1}$ were analyzed in the Mid-Atlantic (Polsky et al., 2000); and nine GCMs forced by up to three CO$_2$ emissions scenarios were analyzed for the U. S. Northeast (Hayhoe et al., 2007). In only two of these studies were the GCMs evaluated with regional observations: McCabe and Ayers (1989) evaluated the models’ mean annual cycle in temperature and precipitation, and Hayhoe et al. (2007) compared multi-model average trends in climate and hydrology with observations. Here, we evaluate the ability of seven models to simulate the mean annual cycle, interannual variability, and long-term trends in watershed-averaged temperature and precipitation. A simple scheme is developed for ranking the models, which is used as a basis for selecting, in some presentations, the models used for climate projection. Projections for two emissions scenarios are presented and extrapolation to four other scenarios is made using globally averaged surface air temperature projections.

The models presented here were used by the Consortium for Atlantic Regional Assessment (CARA), a project which began in 2000 and was funded by the U.S.
Environmental Protection Agency. The main goal of CARA was to assist decision
makers in the middle and upper Atlantic region of the U. S. by providing them with
information about changes in climate and land use (see www.cara.psu.edu). We used the
most up-to-date climate information available to us at that time, including model output
published in support of the Third Assessment Report (TAR) of the Intergovernmental
Panel on Climate Change (IPCC). The recently published Fourth Assessment Report
(AR4) uses models that are slightly more updated (Randall et al., 2007). Hayhoe et al.
(2007) used nine of those models in their U.S. Northeast climate assessment. The models
presented in this study are older by roughly four years and of coarser atmospheric
resolution by roughly 30% in each horizontal direction and 10% in the vertical direction
than those used by Hayhoe et al. (2007). Despite these differences in age and resolution,
many of the results are similar, as we will show.

2. Methods

2.1. Models

Climate model output was obtained from the online archive in support of the TAR
(Cubasch et al., 2001). We used monthly 2-m temperature and precipitation output from
seven GCMs (Table 1) run under the A2 and B2 greenhouse gas emission scenarios
(Nakićenović and Swart, 2000) to the year 2100. Historical model output was also
analyzed beginning in 1911 for five of the models and 1971 for the remaining two, based
on availability. The seven models differ in many ways, some of which are noted in Table
1: resolution, land surface hydrological parameterization, and surface flux adjustment.
Among the six scenario groups considered to be “equally sound” (Nakićenović and
Swart, 2000), A2 and B2 were chosen because they were available for the greatest number of models and because they bracket the middle range of the scenarios used for the TAR (Figure 2a). A2 and B2 correspond to atmospheric CO\textsubscript{2} levels of about 850 and 620 ppm, respectively, in 2100, which can be compared to the pre-industrial (prior to 1800) level of 280 ppm and present-day (2000) level of 370 ppm (Figure 2b) (Prentice et al., 2001). Global mean temperature increases by 2100 for A2 and B2 with respect to 1990 are 3.8 and 2.7° C, respectively, which are averages over all of the TAR models (Figure 2c) (Cubasch et al., 2001). Following Hewitson (2003), we smoothed the model output to obtain a coarse resolution data set that more closely matches the skill resolution of the models. The data were then fit with a splined surface at 1/8° resolution for mapping and to obtain model output at United States Historical Climate Network (USHCN) stations in the region. Model output and observations were processed identically to compute temporal and spatial averages, described below.

2.2. Observations

The models were evaluated with observations of temperature and precipitation from the most recent version of the USHCN database. We started with the version of the USHCN station data set (a) that contains monthly means of temperature and precipitation, (b) that is derived from a data set in which missing daily values were filled in using surrounding stations, and (c) in which no urban heat island corrections were made (Williams et al., 2005). Figure 1 shows the station distribution in the study region.

The monthly USHCN data were processed to compute seasonal, annual and 30-year averages over climate divisions and watersheds. Climate divisions, which are also shown in Figure 1, were developed to describe regions within a given state that have a
relatively homogeneous climate (Guttman and Quayle, 1996). For a given month and year, climate division averages were computed as the mean of all stations within a climate division. Seasonal (December-February, March-May, June-August and September-November) and annual averages for a climate division were computed from the monthly means for each climate division only if all monthly values were defined. Watershed averages for each time interval (monthly, seasonal and annual) were computed by weighting each climate division average by the area of the climate division within the watershed. In the rare cases when a climate division did not have a defined value, that climate division was not included in the watershed average. Thirty-year (1971-2000) means and standard deviations of watershed averages were computed on seasonal and annual bases.

To quantify the climate change over the 20th century, we computed changes in temperature and precipitation averaged over each watershed between two thirty-year periods: 1911-1940 and 1971-2000. To prevent spurious trends associated with the changing station distribution (e.g., Allard and Keim, 2007), we used only those stations that were present for both time periods. Specifically, we computed a seasonal average for each station for a given year only if all three months of the season had defined values. A thirty-year average was computed for each station and season only if at least 25 of the years had defined seasonal averages. The difference between the two thirty-year periods was then computed. Watershed averages of the differences were computed using the same climate-division binning and weighting described above. Annual averages for each watershed were then computed from the seasonal averages.

2.3. Statistics
Confidence bounds on Pearson correlation coefficients computed for some model analyses (e.g., annual precipitation range as a function of model resolution) were determined using bootstrap resampling (Efron and Tibshirani, 1993, pp. 49-50) with a sample size of 1000. Confidence estimates of model results as a function of model parameterization were determined using a permutation approach. For example, in comparing the absolute error in interannual variability of temperature between the two models with bucket hydrology to the five with complex hydrology, we determined the error difference of all possible bucket-complex pairs \((2 \times 5 = 10)\), and computed the fraction of those pairs whose error difference was positive. Given the small number of models considered in this study, we used relatively generous confidence intervals of 80% to determine significance of the correlation coefficient and the impact of model parameterization.

3. Results

3.1. Climate Model Evaluation

3.1.1. Mean State

Figure 3 shows the seasonal and annual temperature and precipitation for each model, the seven-model average, and the observations averaged over the CBW and the 1971-2000 period. For the annual average, model errors range from \(-3.0^\circ\) C to \(+2.4^\circ\) C, with a Root-mean-square error (RMSE) of 2.0° C (Table 2). The overall bias, as given by the seven-model average error, is \(-0.8^\circ\) C. On a seasonal basis the RMSEs are somewhat larger, particularly in winter \((3.8^\circ\) C), which is mainly due to the models being too cool by 3.2° C on average. The bias is smaller in spring and summer and is
particularly small in fall. Some models are clear outliers, such as the GFDL model in summer (8.9° C too warm) and CSIR in winter (6.7° C too cold). The amplitude of the annual cycle, as given by the summer-winter temperature difference, is 4.3° C larger (on average) than the observed value of 21.5° C, an overestimate of 19%.

Annual precipitation in the CBW is simulated well by the models, with an RMSE of 10% and an overall bias of 0%. However, RMSEs are larger for individual seasons, between 14 and 36%. There is a weak annual cycle in observed precipitation over the watershed with a summer maximum, a winter minimum, and an annual range (summer minus winter) that is 27% of the annual mean. The models, on the other hand, typically have strong annual cycles that are advanced with respect to the observations, with a spring maximum, a fall minimum, and an annual range (spring minus fall) for the model average that is 54% of the annual mean.

Model errors in simulating the mean state of other watersheds are very similar to those of the CBW. As an example of the differences among the watersheds, the mean observed and seven-model average climate of each watershed is presented in Figure 4. The observed-modeled differences are very similar (small) among the watersheds. The figure also shows that the seven-model average captures the spatial variation in temperature and the lack thereof in precipitation among the four watersheds.

3.1.2. Interannual variability

To evaluate model skill at simulating interannual variability, standard deviations of seasonal and annual temperature and precipitation for the CBW are presented in Figure 5. RMSEs at the annual time scale are 65% for temperature and 21% for precipitation (Table 3). Most of this error is due to the models being systematically too variable in
temperature and not variable enough in precipitation, with respective biases of +49% and -17%. The results are generally similar on the seasonal time scale, but RMSEs are particularly large for temperature in the summer (95%) and precipitation in the fall (43%), with the main outliers being GFDL and CCSR in both cases. The models are, in general, able to capture the larger interannual variability in winter temperature compared to summer temperature. Results for the Delaware and Hudson River Basins are similar (Table 3).

3.1.3. Long-term trend

Figure 6 shows seasonal and annual temperature and precipitation change from the 1911-1940 period to the 1971-2000 period for the CBW. All models exhibit warming, but the five-model-average temperature increase of 0.39° C is much greater than the observed increase of 0.12° C. There is better agreement in winter and spring but little agreement in the summer and fall. The models generally do not capture the observed increase in precipitation of 8%; the five-model average is 2%, with only one model coming close (HADC) and the other four changing little or getting drier. The discrepancies are particularly dramatic in spring and fall.

For all of the watersheds, the models show the same bias in precipitation trend (lack of increase) but not in temperature trend (Figure 7). Whereas the Northeast U.S. has been warming on average over the 20th century, the southern part of the study region has more closely followed the Southeast U.S., which has cooled (Allard and Keim, 2007; Hansen et al., 2001). The result is that the Lower Chesapeake Watershed shows no long-term temperature change. The models, however, do not capture this spatial variation in temperature trends; instead they warm uniformly over the study region. Thus there is
reasonably good agreement between the five-model average and the observed temperature trend in the Delaware, Susquehanna and Hudson River Basins, but poor agreement in the Lower Chesapeake Watershed (Figure 7).

3.1.4. Model Ranking

As a model simulation of the past climate improves, its future projections gain credibility. Because the models presented here differ dramatically in their ability to simulate the past, we developed a simple ranking system to quantify their relative skill. Model ranking is based on skill at simulating the temperature and precipitation averaged over each watershed for the three metrics already discussed: the 1971-2000 mean, the 1971-2000 standard deviation, and the change in the mean from 1911-1940 to 1971-2000. Models are ranked only at the annual time scale, amounting to a total of six categories (three each for temperature and precipitation). The ranking is based on a normalized error index, $e_{ij}$, for the $i^{th}$ model in the $j^{th}$ watershed:

$$e_{ij} = \frac{1}{N} \sum_{k=1}^{N} \frac{E_{ijk}}{\bar{E}_{jk}},$$

where $E_{ijk}$ is the absolute value of the error (the “absolute error,” for short) for category $k$ and $\bar{E}_{jk}$ is the model-mean absolute error. A perfect model has $e = 0$, an above-average model has $e < 1$, and a below-average model has $e > 1$. Because trend output was not available for two models (GFDL and NCAR), we computed two scores ($N = 4$ and $N = 6$) for each model and watershed (Table 4). Additionally, we computed the normalized error index for the five- and seven-model averages (AVG in Table 4).
If only the mean and standard deviation are considered, then the seven-model average is superior to the individual models in all of the watersheds. The individual models that are above average for at least two watersheds are CCCM, ECHM, HADC and NCAR. When the trend is included, the model average (now of five models), is no longer superior to all of the individual models, but is still above average and in the top two for all watersheds. The individual models that are above average in at least two of three watersheds are CCCM, ECHM and HADC. The weakest models, overall, are GFDL, CCSR and CSIR.

To evaluate model performance over an area representative of the Mid-Atlantic region of the U.S., we also present in Table 4 the normalized error index for the states considered by CARA: Virginia, Maryland, Delaware, Pennsylvania, New Jersey, New York, Connecticut, Rhode Island, and Massachusetts (Figure 1). We find that mid-Atlantic model skills generally agree well with the model skills in the individual basins. In particular, the multi-model average continues to be most skillful, the best individual models are ECHM and CCCM, and the worst individual models are GFDL and CSIR.

3.2. Model projections

We now present future changes of watershed-averaged temperature and precipitation with respect to 1971-2000, starting with the CBW under the A2 scenario (Figure 8). All models predict warming throughout the 21st century with little consensus on its seasonality, which is also true for B2 and the other watersheds (not shown). Differences among models greatly exceed the difference between the two scenarios, particularly for the first two thirty-year periods, where the model means differ by only
0.2° C between the two scenarios. Among both scenarios and all models, the range in the annual mean temperature increase by 2070-2099 is 2.0 to 8.7° C, with the mean ± 1σ equal to 4.7 ± 2.0° C for A2 and 3.5 ± 1.4° C for B2 (Table 5).

Precipitation projections vary dramatically among the models. For the CBW, changes in annual precipitation by 2070-2099 range from -17% to +19% under the A2 scenario, with mean ± 1σ changes of 3 ± 12% for A2 and 3 ± 7% for B2 (Table 5). The range and lack of agreement among models increase for the summer and fall but decrease considerably for the winter and spring. Regarding the latter, five of the seven models predict precipitation increases in the winter and spring for both scenarios in the latter two thirty-year periods considered here. The model-average increases in precipitation (±1σ) by 2070-2099 for winter and spring, are, respectively, 8 ± 7% and 9 ± 8% for A2 and 7 ± 7% and 5 ± 4% for B2.

A2 and B2 are only two among six greenhouse gas scenarios considered “equally sound” by Nakićenović and Swart (2000). Though we do not have access to the output of all models for all scenarios (in part because it may not exist), we can estimate the impact of other scenarios in the region by scaling the model output to globally averaged results. To estimate the temperature change ∆Tjm for a given scenario m in watershed j we use

\[ \Delta T_{jm} = r_m \Delta T_{jA2}, \]  

where \( r_m \) is the ratio of the scenario-m global temperature change to the A2 global temperature change and \( \Delta T_{jA2} \) is the temperature change in watershed j according to the A2 scenario. To evaluate the approach, we compared the model-average B2 annual
temperature change for the three future time periods in the CBW with that predicted from
Equation 2. Differences were less than 4%. Projected seven-model-average temperature
change for the CBW varies from 2.8° C for the B1 scenario to 5.9° C for the A1FI
scenario (Figure 9a), a range of 3.1° C, which is more than twice the difference between
model-averages of A2 and B2 (1.2° C).

If we restrict the projections to those models that performed best (CCCM, ECHM,
HADC and NCAR; see Section 3.1.4), then the model-averaged projected climate
becomes cooler and wetter than the climate projected by the seven-model average (Figure
9), and the consensus (as given by the standard deviation or coefficient of variation)
improves. For these four models, the 100-year average projections over the CBW for
temperature and precipitation change are, respectively, 3.9 ± 1.1° C and 9 ± 12% for the
A2 scenario and 2.9 ± 0.6° C and 8 ± 4% for the B2 scenario (Table 6). To establish the
significance of this shift towards a smaller temperature increase and a greater
precipitation increase, we computed the average temperature and precipitation changes of
all possible four-model combinations (a total of 7×6×5×4 = 840 combinations). The
chances of the warming being less than 3.9° C and precipitation being greater than 9%
are only 13% and 6%, respectively, suggesting a significant shift.

The temperature and precipitation increases in the Hudson River and the
Delaware Bay watersheds are projected to be slightly higher than those of the CBW and
have slightly less spread among the models (Tables 5 and 6). This northward increase in
temperature and precipitation change is a general feature of mid-latitude climate change,
as we will discuss below.
4. Discussion

4.1. Comparison with other studies and differences among models

4.1.1. Model evaluation

Our finding of an overall Mid-Atlantic cool bias of the models is a consistent feature among numerous modeling studies. The three models analyzed by McCabe and Ayers (1989) were too cool (by 0-4° C, mean of 2° C) as was the regional climate model of Jenkins and Barron (1997) in the SRB (by 2.7° C, analyzed in (Najjar, 1999)). These two studies found most of the cool bias to be outside of winter, in contrast to our findings. Similar to our results (Figure 3), the GFDL model used by McCabe and Ayers (1989) had excessive summer warmth. As part of the AR4, Randall et al. (2007) presented global maps of various metrics of model skill for 23 GCMs, from which we roughly estimated the skill in the Mid-Atlantic region; we also used tabulated averages over Eastern North America for the same set of models (Christensen et al., 2007). For annual mean surface air temperature, the overall bias was 1-2° C too cool and the RMSE was 2-3° C, which suggests somewhat less skill than the simulations presented here (Table 2). We found little spatial variation in the temperature error in the Mid-Atlantic, at least at the watershed scale (Figure 4), which agrees with the lack of spatial gradients over the Eastern U.S. in the error maps of Randall et al. (2007). Over Eastern North America, the cool bias is 1.2° C (Christensen et al., 2007). The cool bias is a large-scale feature of many GCMs, particularly in the Northern Hemisphere (Covey et al., 2000; Randall et al., 2007).

Covey et al. (2000) noted that cool biases are worse in regions of topography, which suggests that improved resolution might reduce this bias. To evaluate this, we
regressed the error in CBW annual mean temperature against mean horizontal grid spacing (average of latitude and longitude spacing, see Table 1). We found a correlation coefficient of -0.04 with 80% confidence intervals of -0.52 to 0.43; this suggests that the cool bias is not influenced by resolution for this group of models. Covey et al. (2000) also noted that flux adjustment reduces model surface temperature errors over the ocean. The Mid-Atlantic region’s proximity to the sea prompted us to compare the CBW’s absolute temperature error of the five models with flux adjustment to the two models without it. The average absolute error of the flux-adjusted models was 1°C smaller than that of the non-flux adjusted models. But the difference is not very significant, because less than 80% (7 of 10) of the possible paired differences showed larger errors in the flux adjusted models.

The early mid-Atlantic studies of McCabe and Ayers (1989) and Jenkins and Barron (1997) found a dampened annual surface air temperature cycle in models, whereas numerous (and mostly more recent) multi-model comparative studies (including ours, Section 3.1.1) find an exaggerated one in the Eastern or Northeastern U.S. (Christensen et al., 2007; Gates et al., 1996; Nigam and Ruiz-Barradas, 2006; Randall et al., 2007). Soils that are too dry are expected to lead to an enhanced annual cycle in surface air temperature. Bucket hydrological models tend to have too much evaporation (Sellers, 1992) and therefore are expected to have soils that are too dry, so we compared the CBW’s summer-minus-winter temperature difference of the three models that have bucket hydrology with the four models that have more complex land surface schemes (Table 1). The average seasonal temperature range of the models with bucket hydrology is 2.9°C greater than that of the models with complex land surface hydrology, consistent
with our hypothesis. However, the significance of the difference is marginal because less than 80% (9 of 12) of the possible paired differences show that the models with bucket hydrology have greater seasonal temperature ranges than those that do not.

Our finding of little bias in the model-average annual precipitation in the Mid-Atlantic is similar to results from the AR4. The multi-model mean precipitation in the maps presented by Randall et al. (2007) is in excellent agreement with the observations, though a slight wet bias may exist; over Eastern North America it is only 9% (Christensen et al., 2007). In our Mid-Atlantic analysis, the model-mean precipitation is closer to observations than any individual model, similar to results for the global scale distribution of annual precipitation (Nohara et al., 2006). Even early studies in the Mid-Atlantic found error cancellation among models in their precipitation simulations. McCabe and Ayers (1989) found annual precipitation errors ranging from -60% to 80% in the three models they analyzed, but the average bias was only 25%. Comparison with this older study also illustrates the dramatic improvements in the simulation of Mid-Atlantic precipitation since the 1980s: the largest absolute error found among the seven models analyzed here is 14%, less than a fifth of the largest absolute error in the three models presented in McCabe and Ayers (1989).

Improvements in simulated precipitation have been less dramatic on seasonal time scales. On average, the models of McCabe and Ayers (1989) were too wet in winter (30%) and too dry in summer (-20%). The regional model of Jenkins and Barron (1997) simulated precipitation that was too seasonal, similar to our findings, though in their study the biases were greatest in summer (+87%) and fall (-30%). Our biases are small in summer and winter but considerable in the spring and fall (~25%, Table 2). Boyle (1998)
found that the mean of 30 models overestimated spring-summer rainfall in the Eastern U.S., a result that is essentially unchanged after several generations of intercomparison projects (Christensen et al., 2007; Phillips and Gleckler, 2005). The too-strong annual cycle in the Eastern U.S. is even a feature of the National Center for Environment Prediction (NCEP) reanalysis (Boyle, 1998; Nigam and Ruiz-Barradas, 2006), which rules out climate model drift as the cause of the discrepancy. Boyle (1998) noted that model performance is lowest when and where precipitation is dominated by convection. This suggests that increased resolution might help, and so we regressed the CBW’s normalized precipitation range (seasonal maximum minus minimum, divided by the annual mean) versus horizontal grid spacing. We found a correlation coefficient of 0.55, which, being positive, is consistent with the hypothesis, but the 80% confidence bounds of -0.22 to 0.86 make the result insignificant. Improved resolution to much finer levels (~0.5°) using nested regional models also does not cure the anomalously large seasonality of precipitation in the Northeast U.S. (Liang et al., 2004), though the problems here could also be related to lateral boundary conditions and choice of convection scheme. The more recent, higher resolution North American Regional Reanalysis (NARR) has a much improved annual precipitation cycle (Nigam and Ruiz-Barradas, 2006), but this is presumably because the NARR directly assimilates precipitation, unlike the NCEP reanalysis.

Our finding of simulated interannual variability in temperature that exceeds the observations has been noted previously in model intercomparison studies for land regions in general (Bell et al., 2000; Räisänen, 2002). Räisänen (2002) showed that more than 80% of the 19 models analyzed had interannual variability in excess of the observations.
Bell et al. (2000) showed that excessive variability over land was related to land surface hydrology, with the simpler (bucket-type) schemes producing greater interannual variability than the observations and more complex land surface models because the simpler schemes underestimate soil moisture. We therefore compared the CBW’s interannual variability in annual-mean and summer-mean temperature of the three models that have bucket hydrology with the four models that have more complex land surface schemes. The average interannual variability of the models with bucket hydrology for the annual and summer time frames was 41% and 96% greater, respectively, than that of the models with complex land surface hydrology, consistent with the hypothesis. The result can be considered marginally significant because 2 of the 12 (for the annual time frame) and 3 of the 12 (for the summer time frame) possible paired differences show that the models with bucket hydrology have smaller interannual variability than those that do not.

Räisänen (2002) suggested that interannual variability in winter temperature is driven by ice and snow physics, which was supported by his finding of a negative correlation between mean temperature and temperature variability in winter. We found a negative correlation between these quantities in the CBW ($r = -0.62$) with bootstrapped 80% confidence bounds of -0.21 to -0.95, consistent with the hypothesis that the models’ cold bias in winter causes the excessive interannual temperature variability in winter. Räisänen (2002) also found a positive correlation between summer mean temperature and summer interannual variability in temperature, arguing that excessive drying related to high temperature would increase interannual variability. We also found a positive
correlation between these quantities in the CBW ($r = 0.82$), with the 80% confidence intervals (0.43 to 0.95) supporting the hypothesis.

Our finding of too little interannual variability in precipitation is consistent with Boyle’s (1998) analysis of 30 AMIP models for the U.S., which showed too much variation at the seasonal time frame at the expense of interannual variability. Räisänen (2002), however, found no bias among 19 models in their simulation of the variability of precipitation at interannual time scales.

Our results regarding climate change over the 20th century (Figures 6 and 7) are in good agreement with those of Hayhoe et al. (2007), who found that climate model simulations of the Northeast U.S. were able to capture 20th-century trends in temperature but not precipitation.

4.1.2. Model Projections

As several past studies of Mid-Atlantic climate change have been conducted for a CO$_2$ doubling (McCabe and Ayers, 1989; Moore et al., 1997; Najjar, 1999) we selected transient results from the literature and from our study that corresponded most closely to these conditions: 2090-2099 for the 1% yr$^{-1}$ simulations of Polsky et al. (2000) and 2070-2099 for the A2 simulations presented by Hayhoe et al. (2007) and us. Figure 10 summarizes the results. Multi-model means from the studies all indicate a warmer (2.5° to 4.7° C) and wetter (3-17%) Mid-Atlantic. Our seven-model means tend toward the drier and warmer part of the range, whereas the “best”-four-model means sit closer to the center. As part of the AR4, Christensen et al. (2007) presented maps of North American climate change under the A1B scenario (2080-2099 with respect to 1980-1999) averaged over 21 GCMs, from which we estimated the change over the Mid-Atlantic. Annual
temperature and precipitation change was approximately 3-4° C and 5-10% over the Mid-Atlantic, which is consistent with prior estimates, even after correcting for the ~20% cooler predictions of A1B with respect to A2 (Christensen et al., 2007). We found that three of the seven models analyzed here had precipitation increases in the Mid-Atlantic (Figure 8), whereas of the 21 more recent models analyzed by Christensen et al. (2007) more than 17-20 showed precipitation increases in the Mid-Atlantic; this indicates that model consensus is improving with time.

The range among the various models analyzed here tends to be greater than previous studies, which is partly due to the large number of models considered, though Hayhoe et al. (2007) found smaller temperature and precipitation ranges among nine models. The wider spread found here can be explained by the CCSR model, which produced the largest warming (8.6° C) and the greatest precipitation decrease (-17%). The strong warming and drying in the CCSR simulation fits in with the tendency among climate simulations of the Mid-Atlantic to have a negative correlation between temperature change and precipitation change. The correlation coefficient between temperature change and precipitation change among the models we analyzed for the CBW in 2070-2099 under the A2 scenario is -0.76 (80% confidence interval of -0.27 to -0.92). A negative correlation also exists between the mean values of the studies presented in Figure 10a (-0.78, with 80% confidence intervals of -0.65 to -0.98, excluding the “best 4” data point to make the results independent of the previous correlations).

Though not noted by the authors, this pattern was also seen in the two models analyzed by Polsky et al. (2000) and the three models analyzed by McCabe and Ayers (1989). A negative correlation is expected on physical grounds: a wetter climate will result in a
greater fraction of radiative heating to evaporate water at the expense of heating the
ground and hence the air above it.

Our finding of a larger precipitation increase and greater consensus in winter
compared to summer is a general feature of GCMs in the Mid-Atlantic (Figure 10b). Of
the 21 GCMs analyzed by Christensen et al. (2007) as part of the AR4, precipitation
increases within the Mid-Atlantic occurred in 17-21 of the models in winter and 8-18 in
summer; mean precipitation increases in these seasons were 10-15% and 0-5%,
respectively. Seasonality of temperature change is more equivocal among prior studies.
Whereas we and Christensen et al. (2007) find no indication of seasonality in the Mid-
Atlantic, Hayhoe et al. (2007) project greater warming in the summer and Moore et al.
(1997) report greater warming in the winter. Many GCMs do consistently show greater
warming in winter due to ice-albedo feedback, but this is generally limited to regions
north of the Mid-Atlantic.

Our seven- and four-model average A2 projections for 2070-2099 are 1.6° and
0.8° C (25-50%) greater than the corresponding global mean temperature projections in
Figure 2c (3.0° C). This follows from the fact that warming is greater over land and in
the Northern Hemisphere, particularly at higher latitudes. More specifically, at 40° N
over land, temperature increases reported in the AR4 are, averaged over latitude and
multiple models, about 40% greater than the global mean for the A2 scenario by the end
of the 21st century (Meehl et al., 2007).

Models tend to predict a precipitation increase in the Mid-Atlantic because the
region lies poleward of the boundary between subtropical moisture divergence and
subpolar moisture convergence. Global warming causes an overall increase in humidity
and intensification of the hydrological cycle, which results in greater transport of moisture from the subtropics to subpolar regions (Christensen et al., 2007). This then leads to a subtropical decrease and subpolar increase in precipitation. There is a large range among models in the projected Mid-Atlantic precipitation change because the divergence/convergence boundary lies just south of the Mid-Atlantic; hence small errors in the location of this boundary result in large errors in projected precipitation. The model errors tend to decrease from summer to winter as the convergence/divergence boundary moves southward. This also explains why the model spread in precipitation projections tends to be smaller in the Delaware and Hudson River Basins compared to the more southward CBW (Tables 5 and 6).

4.2. Implications for water temperature and streamflow

A treatment of the impacts of projected climate change on Mid-Atlantic estuaries is beyond the scope of this study. The reader is referred to the numerous studies that have considered the potential impacts of climate change on coastal systems (Harley et al., 2006; Kennedy et al., 2002; Scavia et al., 2002), including the Mid-Atlantic coastal region (Najjar et al., 2000). However, to make the climate projections presented here and elsewhere relevant for estuarine science, we discuss how water temperature and streamflow may change. Precipitation itself has little direct influence on estuaries but is extremely important through its influence on streamflow and nutrient and sediment loading. In the Mid-Atlantic, annual streamflow changes are 1.5-2.0 times those of annual precipitation changes on a fractional basis (Sankarasubramanian et al., 2001). In other words, a year with precipitation that is 10% above normal will typically have streamflow that is 15-20% above normal. Furthermore, in the U.S. Northeast, a greater
fraction of nitrogen inputs to a watershed are exported to the ocean during wet years (Howarth et al., 2006). These two observations illustrate the high sensitivity of streamflow and nutrient loading to precipitation in the Mid-Atlantic region.

4.2.1. Water temperature

Estuaries are too small to be resolved by GCMs, so estimates of water temperature change must be inferred from changes in the regional climate. A clear result from numerous studies in the Mid-Atlantic region is that estuarine and coastal water temperature is highly correlated to regional atmospheric and oceanic temperature at time scales ranging from months to decades (Cronin et al., 2003b; Hayhoe et al., 2007; Preston, 2004; Secor and Wingate, 2008). Thus it appears highly likely that as the Mid-Atlantic region warms, estuaries of the Mid-Atlantic will warm as well.

The degree of warming, however, is somewhat less certain. Projected sea surface temperatures in coastal grid boxes of GCMs show smaller increases than the surface air temperatures above adjacent land areas. In the A2 scenario by 2070-2099, for example, warming in the Gulf of Maine was 3.3° C in contrast to a 4.5° C warming over the U.S. Northeast (Hayhoe et al., 2007). This is consistent with the historical warming of the Gulf of Maine over the 20th century (0.05° C decade\(^{-1}\)), which was less than the rate of warming over the U.S. Northeast over the same time period (0.08° C decade\(^{-1}\)) (Hayhoe et al., 2007). However, the Gulf of Maine (especially as represented in GCMs), is in greater contact with the open ocean than any of the estuaries considered here, so we might expect less of a difference between changes in watershed air temperature and changes in estuarine temperature. Furthermore, Mid-Atlantic estuaries are very shallow, with mean depths of the Chesapeake Bay, the Delaware Bay and the Hudson River
Estuary of 6.5, 8 and 10 m, respectively; this indicates a modest thermal inertia and thus a close thermal coupling between water and air. Water temperature change in the upper Hudson River occurred at a rate of 0.12°C decade\(^{-1}\) from 1918 to 1990 (Ashizawa and Cole, 1994), which is greater than the warming rate of the watershed reported here (0.08°C decade\(^{-1}\)) over a slightly different time period (Figure 7). Secor and Wingate (2008) analyzed surface water and air temperatures from 1960 to 2006 in the lower Patuxent River Estuary (located roughly halfway down the mainstem Chesapeake) and found similar trends in air and water. Overall, these studies suggest that Mid-Atlantic estuaries will warm at a rate similar to or perhaps slightly less than the warming rate of the Mid-Atlantic region.

The implications of warming by the end of the 20\(^{th}\) century, such as those presented in Figure 9 (increases of ~2-5°C) are difficult to quantify because the historical record shows interannual variability that is far smaller (~0.5°C, Figure 10a). In other words, there are few (if any) years in the recent past that can be used as analogues for the future. Due to the complexity and nonlinearity of biogeochemical and ecological systems, attempts to quantify their responses to climate change based on first principles in the absence of empirical, whole-system observations, are fraught with uncertainty. Exceptions to this rule are some straightforward physical and chemical metrics, such as dissolved oxygen solubility, the temperature dependence of which is well known.

### 4.2.2. Streamflow

Table 7 shows a summary of previous estimates of the impact of CO\(_2\) increase on streamflow. Results vary dramatically among the studies and even within them, with an overall range of nearly ±40%. The range within a given study typically reflects the
different climate predictions of GCMs, whereas the variation among studies includes this in addition to differences in hydrological models. Milly et al. (2005), Nohara et al. (2006), and Christensen et al. (2007) presented maps of multi-model averages (19-35 members) of runoff change in response to increasing CO$_2$ with different weighting schemes. All three studies, which had a different number of ensemble members (19-35) and different weighting schemes, showed a weak tendency for models to predict future runoff increases in the Mid-Atlantic.

Projections of changes in streamflow due to climate change in the Mid-Atlantic region reflect the combined uncertainties in the projected climate change (Figure 10) and the hydrological models used to translate climate into flow (Sankarasubramanian et al., 2001). Problems in the Mid-Atlantic are particularly vexing because projected temperature and precipitation increases have opposing effects, making the projected streamflow the difference of two large, uncertain numbers. Another difficulty in projecting streamflow in the Mid-Atlantic is the fact that interannual variability in temperature is far less than the projected climate change (Figure 10a), leaving hydrological models an inadequate range for calibrating the streamflow sensitivity to temperature at interannual timescales. It is for this reason, we suggest, that Wolock and McCabe (1999), Neff et al. (2000), and Frei et al. (2002) reported such different streamflow projections using the same GCMs as input (Table 7).

Nevertheless, it is clear that precipitation has controlled annual streamflow in the past in the Mid-Atlantic region and we can use this knowledge to anticipate what the maximum annual streamflow increases might be. This can be done very well because projected precipitation changes are generally in the range of interannual variability
We acquired annual streamflow records for the four watersheds considered here, divided by the upstream area of the watershed to get flow in the same units as precipitation, and created annual averages for different calendar years (January-December, February-January, etc.). We then linearly regressed these against matching calendar-year average precipitation. We also estimated the fractional change in annual streamflow divided by the fractional change in annual precipitation, which is known as the precipitation elasticity of streamflow, the sensitivity factor, or (as called here) the amplification factor (Sankarasubramanian et al., 2001, and references therein). Table 8 shows the results for the calendar year with the highest correlation coefficients. Similar to previous analyses, a very large fraction of annual streamflow in the Mid-Atlantic is accounted for by annual precipitation (~80%) and the amplification factor is between 1.5 and 2.0 (Najjar, 1999; Sankarasubramanian et al., 2001). As an example, we take A2 2070-2099 precipitation changes within the CBW of the best four models (mean ± 1σ = 9 ± 12%, Table 6) and multiply by an amplification factor of 1.72 (Table 7) to get a maximum streamflow change of 15 ± 20%.

How much will warming-induced evapotranspiration offset this flow change? Unfortunately, direct estimates of evapotranspiration changes were not available for the TAR models, and so we used other estimates of the sensitivity of evapotranspiration to temperature. McKenney and Rosenberg (1993) analyzed the sensitivity of potential evapotranspiration formulae to temperature and found a range of 4-10% °C⁻¹. In a completely independent approach, Huntington (2003) inferred evapotranspiration from annual flow and precipitation data in U.S. East Coast watersheds. Regression analysis yielded a range of 2-4 cm °C⁻¹, which, given that annual evapotranspiration is about 50
cm in the Mid-Atlantic, is in excellent agreement with McKenney and Rosenberg (1993). Given a warming of about 4° C and annual evapotranspiration of about 50 cm in the Mid-Atlantic, projected warming-induced streamflow decreases are roughly 15-40%, easily offsetting the flow increases due to precipitation change.

5. Conclusions

The results presented here show that models vary widely in their ability to simulate the climate of the Mid-Atlantic region. For the mean climate, the errors are random, so that the output averaged over many models tends to look more like the real climate than the output of any individual model. The same can be said for 20th-century temperature trends, except for the Lower Chesapeake Watershed, which models consistently exaggerate the warming of. The observed long-term increase in precipitation throughout Mid-Atlantic is also missed by the models. Finally, interannual variability in the models is overestimated for temperature and underestimated for precipitation. To summarize the overall model evaluation, we find that the multi-model mean is superior to any individual model, a finding similar to numerous climate model intercomparison studies conducted at global and continental scales (Gleckler et al., 2008; Lambert and Boer, 2001; Phillips and Gleckler, 2006; Reichler and Kim, 2008).

Under projected changes in atmospheric composition, the models consistently predict warming but show little consensus in annual precipitation change. Winter and spring precipitation, however, are shown to increase in most models. By selecting models that are least prone to error in simulating the past and present climate, future projections shift significantly towards less warming and increased precipitation.
The projected temperature changes by the end of the 21\textsuperscript{st} century are outside the variability of the 20\textsuperscript{th} century. On the other hand, the precipitation projections tend to be in the neighborhood of past interannual variability. Thus, we are in the unfortunate position in the Mid-Atlantic—and probably in many other mid-latitude estuarine watersheds—of on the one hand being rather certain that it will warm but uncertain about the hydrological and estuarine response, and on the other hand of being uncertain about precipitation change but certain about what the response of any plausible change might be. Thus, efforts are needed in the Mid-Atlantic region to better constrain hydrological and estuarine responses to temperature change and to improve the quality of precipitation forecasts.

Acknowledgments. This work was supported by the Global Change Research Program, Office of Research and Development, US EPA (Cooperative Agreement R-83053301). We thank Ann Fisher and Jim Shortle for guidance and leadership of the Consortium for Atlantic Regional Assessment. We also thank the anonymous reviewer who provided us with useful feedback on this paper.

References


Miller WD, Harding LW (2007) Climate forcing of the spring bloom in Chesapeake Bay. Marine Ecology Progress Series 331:11-22


Neff R, Chang HJ, Knight CG, Najjar RG, Yarnal B, Walker HA (2000) Impact of climate variation and change on Mid-Atlantic Region hydrology and water resources. Climate Research 14:207-218


Roeckner E, Oberhuber JM, Bacher A, Christoph M, Kirchner I (1996) ENSO variability and atmospheric response in a global coupled atmosphere-ocean GCM. Climate Dynamics 12:737-754


Schubel JR, Pritchard DW (1986) Responses of Upper Chesapeake Bay to variations in discharge of the Susquehanna River. Estuaries 9:236-249

Secor DH, Gunderson TE, Karlsson K (2000) Effect of temperature and salinity on growth performance in anadromous (Chesapeake Bay) and nonanadromous (Santee-Cooper) strains of striped bass Morone saxatilis. Copeia 1:291-296


UNEP/GPA, The Hague


Figure Captions

Figure 1. Study area showing the four main watersheds (shaded): Hudson River Basin, (HRB), Delaware River Basin (DRB), Susquehanna River Basin (SRB), and Lower Chesapeake Watershed (LCW). Also shown are locations of U. S. Historical Climate Network Stations (circles) and streamgages for some major rivers (green triangles). From south to north, the streamgages are located on the Potomac River, the James River, the Susquehanna River, the Delaware River, and the Hudson River. States are outlined with thick gray lines and climate divisions by thin black lines.

Figure 2. Projections from the IPCC Third Assessment Report (TAR). Six scenarios are shown, described in Nakićenović and Swart (2000). (a) CO₂ emissions; (b) modeled levels of carbon dioxide according to the Bern carbon cycle model; (c) global mean surface air temperature from the average of nine TAR models. Data for these figures was taken from Appendix II of Houghton et al. (2001).

Figure 3. Simulated and observed average temperature (left) and precipitation (right) across the Chesapeake Bay Watershed for each season and the annual average. AVG = seven-model average, OBS = observations; other abbreviations are individual models (see Table 1).

Figure 4. Annual- and watershed-averaged temperature (a) and precipitation (b) for the period 1971-2000 given by the observations (OBS) and the seven-model average (AVG)
for each watershed (LCW = Lower Chesapeake Watershed, DRB = Delaware River Basin, SRB = Susquehanna River Basin, HRB = Hudson River Basin).

Figure 5. Modeled and observed interannual variability in spatially averaged temperature (left) and precipitation (right) over the Chesapeake Bay Watershed for the period 1971-2000 for each season and the annual average. AVG = seven-model average, OBS = observations; other abbreviations are individual models (see Table 1). Model-average standard deviations are averages of the individual model standard deviations as opposed to the standard deviation of the model averages, which are very small due to models being uncorrelated at interannual time scales.

Figure 6. Changes in Chesapeake Watershed-average temperature (left) and precipitation (right) from 1911-1940 to 1971-2000 by season and annual average. AVG = five-model average, OBS = observations; other abbreviations are individual models (see Table 1).

Figure 7. Annual mean temperature (a) and precipitation (b) change from 1911-1940 to 1971-2000 given by the observations (OBS) and the five-model average (AVG) for each watershed (LCW = Lower Chesapeake Watershed, DRB = Delaware River Basin, SRB = Susquehanna River Basin, HRB = Hudson River Basin).

Figure 8. Annual and seasonal temperature (top) and precipitation (bottom) changes averaged over the Chesapeake Bay Watershed with respect to 1971-2000 predicted under the A2 scenario for the periods 2010-2039, 2040-2069 and 2070-2099.
Figure 9. Projected change in annual mean temperature (a and b) and precipitation (c and d) of the Chesapeake Bay Watershed for six IPCC scenarios averaged over all seven climate models (a and c) and the four ranked highest (b and d).

Figure 10. Comparison of available studies of CO$_2$-induced climate change in the Mid-Atlantic: (a) annual precipitation change vs. annual temperature change and (b) winter and summer precipitation change. For each study, the mean results are shown by the symbols and the overall range of individual models by the lines. Domain and CO$_2$ levels differed among the studies, as follows. McCabe and Ayers (1989), Delaware River Basin, CO$_2$ doubling; Moore et al. (1997), average of New England and Mid-Atlantic, CO$_2$ doubling; Najjar (1999), Susquehanna River Basin, CO$_2$ doubling; Polsky et al. (2000), Susquehanna River Basin, 2090-2099 with respect to 1983-1994, CO$_2$ increase of 1% yr$^{-1}$; Hayhoe et al. (2007), U.S. Northeast, A2 scenario, 2070-2099 with respect to 1961-1990; This study, the Chesapeake Bay Watershed, A2 scenario, 2070-2099 with respect to 1971-2000. Dashed lines show observed interannual variability in temperature and precipitation as given by ±1 standard deviation of temperature and precipitation for the Chesapeake Bay Watershed. Annual-mean results were not available for Moore et al. (1997).
Table 1. Characteristics of climate models whose output is used in this study. Adapted from McAvaney et al. (2001)

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Model Name</th>
<th>Modeling Center</th>
<th>Time period analyzed</th>
<th>Atmospheric model resolutiona (Latitude, longitude, # levels)</th>
<th>Land Surface Model</th>
<th>Flux Adjustment?</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCCM</td>
<td>CGCM2</td>
<td>Canadian Centre for Climate Modelling and Analysis</td>
<td>1911-2099</td>
<td>3.8(^\circ), 3.8(^\circ), 10</td>
<td>Bucket</td>
<td>Yes</td>
<td>Flato and Boer (2001)</td>
</tr>
<tr>
<td>CCSR</td>
<td>CCSR/ NIES2</td>
<td>Center for Climate System Research</td>
<td>1911-2099</td>
<td>5.6(^\circ), 5.6(^\circ), 20</td>
<td>Bucket</td>
<td>Yes</td>
<td>Nozawa et al. (2000)</td>
</tr>
<tr>
<td>CSIR</td>
<td>CSIRO-Mk2</td>
<td>Commonwealth Scientific and Industrial Research Organization</td>
<td>1911-2099</td>
<td>3.2(^\circ), 5.6(^\circ), 9</td>
<td>Complex</td>
<td>Yes</td>
<td>Gordon and O'Farrell (1997)</td>
</tr>
<tr>
<td>ECHM</td>
<td>ECHM4/ OPYC3</td>
<td>Max Planck Institute for Meteorology</td>
<td>1911-2099</td>
<td>2.8(^\circ), 2.8(^\circ), 19</td>
<td>Complex</td>
<td>Yes</td>
<td>Roeckner et al. (1996)</td>
</tr>
<tr>
<td>GFDL</td>
<td>GFDL-R30</td>
<td>Geophysical Fluid Dynamics Laboratory</td>
<td>1970-2099</td>
<td>2.25(^\circ), 3.75(^\circ), 14</td>
<td>Bucket</td>
<td>Yes</td>
<td>Knutson et al. (1999)</td>
</tr>
<tr>
<td>HADC</td>
<td>HadCM3</td>
<td>Hadley Centre for Climate Prediction and Research</td>
<td>1911-2099</td>
<td>2.5(^\circ), 3.75(^\circ), 19</td>
<td>Complex</td>
<td>No</td>
<td>Gordon et al. (2000)</td>
</tr>
<tr>
<td>NCAR</td>
<td>DOE PCM</td>
<td>National Center for Atmospheric Research</td>
<td>1970-2099</td>
<td>2.8(^\circ), 2.8(^\circ), 18</td>
<td>Complex</td>
<td>No</td>
<td>Washington et al. (2000)</td>
</tr>
</tbody>
</table>

aAll models except HADC are spectral, so horizontal resolution is approximate.
Table 2. Seasonal and annual model errors in temperature and precipitation averaged over each watershed and over the 1971-2000 period. Shown are the mean and standard deviation of the seven models minus the observations.

<table>
<thead>
<tr>
<th>Temperature (°C)</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chesapeake</td>
<td>-3.2 ± 3.8</td>
<td>-1.6 ± 2.1</td>
<td>1.1 ± 3.6</td>
<td>0.2 ± 1.9</td>
<td>-0.8 ± 2.0</td>
</tr>
<tr>
<td>Delaware</td>
<td>-3.0 ± 3.8</td>
<td>-1.7 ± 2.2</td>
<td>1.0 ± 3.4</td>
<td>0.3 ± 1.9</td>
<td>-0.9 ± 2.0</td>
</tr>
<tr>
<td>Hudson</td>
<td>-2.4 ± 3.0</td>
<td>-1.9 ± 2.5</td>
<td>1.0 ± 3.2</td>
<td>0.5 ± 1.8</td>
<td>-0.7 ± 1.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Precipitation (%)</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chesapeake</td>
<td>4 ± 21</td>
<td>24 ± 29</td>
<td>5 ± 14</td>
<td>-28 ± 36</td>
<td>0 ± 10</td>
</tr>
<tr>
<td>Delaware</td>
<td>21 ± 27</td>
<td>13 ± 19</td>
<td>-4 ± 15</td>
<td>-27 ± 34</td>
<td>-3 ± 12</td>
</tr>
<tr>
<td>Hudson</td>
<td>3 ± 15</td>
<td>21 ± 27</td>
<td>5 ± 20</td>
<td>-26 ± 32</td>
<td>-2 ± 12</td>
</tr>
</tbody>
</table>

Table 3. Seasonal and annual model errors in interannual variability as measured using standard deviations for the 1971-2000 period of watershed-averaged temperature and precipitation. The mean and standard deviation of the seven models minus the observations was computed. Shown are these figures divided by the observed interannual variability.

<table>
<thead>
<tr>
<th>Temperature (%)</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chesapeake</td>
<td>16 ± 24</td>
<td>55 ± 60</td>
<td>57 ± 95</td>
<td>47 ± 61</td>
<td>49 ± 65</td>
</tr>
<tr>
<td>Delaware</td>
<td>16 ± 24</td>
<td>53 ± 58</td>
<td>59 ± 98</td>
<td>52 ± 65</td>
<td>47 ± 59</td>
</tr>
<tr>
<td>Hudson</td>
<td>16 ± 25</td>
<td>40 ± 49</td>
<td>55 ± 94</td>
<td>37 ± 52</td>
<td>34 ± 47</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Precipitation (%)</th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Fall</th>
<th>Annual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chesapeake</td>
<td>-31 ± 26</td>
<td>-32 ± 33</td>
<td>-6 ± 26</td>
<td>-33 ± 43</td>
<td>-17 ± 21</td>
</tr>
<tr>
<td>Delaware</td>
<td>-42 ± 43</td>
<td>-21 ± 27</td>
<td>3 ± 36</td>
<td>-22 ± 42</td>
<td>-16 ± 29</td>
</tr>
<tr>
<td>Hudson</td>
<td>-12 ± 16</td>
<td>-21 ± 26</td>
<td>3 ± 37</td>
<td>-24 ± 39</td>
<td>-2 ± 29</td>
</tr>
</tbody>
</table>
Table 4. Normalized error index characterizing the ability of models (Table 1) to simulate temperature and precipitation averaged over the Chesapeake Bay Watershed (CBW), the Delaware River Basin (DRB), the Hudson River Basin (HRB), and the nine states from the Consortium for Atlantic Regional Assessment (CARA). Categories are for 1971-2000 means, 1971-2000 standard deviations (variability), and change from 1911-1940 to 1971-2000 (trend). The lowest normalized error index (i.e., the highest rank) is bold and the second lowest is in bold italic.

<table>
<thead>
<tr>
<th></th>
<th>Mean &amp; variability</th>
<th>Mean, variability &amp; trend</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBW</td>
<td>0.71</td>
<td>0.80</td>
</tr>
<tr>
<td>DRB</td>
<td>0.93</td>
<td>1.09</td>
</tr>
<tr>
<td>HRB</td>
<td>0.67</td>
<td>1.09</td>
</tr>
<tr>
<td>CARA</td>
<td>0.84</td>
<td>0.79</td>
</tr>
<tr>
<td>CBW</td>
<td>0.80</td>
<td>0.99</td>
</tr>
<tr>
<td>DRB</td>
<td>1.02</td>
<td>1.40</td>
</tr>
<tr>
<td>HRB</td>
<td>0.84</td>
<td>1.24</td>
</tr>
<tr>
<td>CARA</td>
<td>1.29</td>
<td>1.18</td>
</tr>
<tr>
<td>CCCM</td>
<td>1.03</td>
<td>1.11</td>
</tr>
<tr>
<td>CCSR</td>
<td>1.28</td>
<td>1.06</td>
</tr>
<tr>
<td>CSIR</td>
<td>1.31</td>
<td>1.31</td>
</tr>
<tr>
<td>ECHM</td>
<td>0.71</td>
<td>0.66</td>
</tr>
<tr>
<td>GFDL</td>
<td>0.89</td>
<td>1.14</td>
</tr>
<tr>
<td>HADC</td>
<td>0.77</td>
<td>0.83</td>
</tr>
<tr>
<td>NCAR</td>
<td>0.81</td>
<td>0.95</td>
</tr>
<tr>
<td>AVG</td>
<td>0.44</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.61</td>
</tr>
</tbody>
</table>

Table 5. All-model mean temperature and precipitation change (± one standard deviation) by watershed for three future periods. Results are for the A2 scenario with respect to the 1971-2000 period.

<table>
<thead>
<tr>
<th></th>
<th>2010-2039</th>
<th>2040-2069</th>
<th>2070-2099</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature (°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chesapeake</td>
<td>1.2 ± 0.3</td>
<td>2.7 ± 1.0</td>
<td>4.7 ± 2.0</td>
</tr>
<tr>
<td>Delaware</td>
<td>1.2 ± 0.3</td>
<td>2.9 ± 1.0</td>
<td>4.9 ± 1.8</td>
</tr>
<tr>
<td>Hudson</td>
<td>1.2 ± 0.4</td>
<td>2.9 ± 0.9</td>
<td>5.0 ± 1.7</td>
</tr>
<tr>
<td>Precipitation (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chesapeake</td>
<td>1 ± 2</td>
<td>2 ± 7</td>
<td>3 ± 12</td>
</tr>
<tr>
<td>Delaware</td>
<td>0 ± 2</td>
<td>3 ± 4</td>
<td>4 ± 7</td>
</tr>
<tr>
<td>Hudson</td>
<td>1 ± 4</td>
<td>5 ± 3</td>
<td>5 ± 4</td>
</tr>
</tbody>
</table>
Table 6. Four-model mean of temperature and precipitation change (± one standard deviation) by watershed for three future periods. Results are for the A2 scenario with respect to the 1971-2000 period. The four models are CCCM, ECHM, HADC and NCAR.

<table>
<thead>
<tr>
<th></th>
<th>2010-2039</th>
<th>2040-2069</th>
<th>2070-2099</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Temperature (° C)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chesapeake</em></td>
<td>1.1 ± 0.4</td>
<td>2.3 ± 0.6</td>
<td>3.9 ± 1.1</td>
</tr>
<tr>
<td><em>Delaware</em></td>
<td>1.1 ± 0.4</td>
<td>2.5 ± 0.7</td>
<td>4.1 ± 1.1</td>
</tr>
<tr>
<td><em>Hudson</em></td>
<td>1.1 ± 0.5</td>
<td>2.5 ± 0.7</td>
<td>4.2 ± 1.0</td>
</tr>
<tr>
<td><strong>Precipitation (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Chesapeake</em></td>
<td>3 ± 1</td>
<td>7 ± 4</td>
<td>9 ± 12</td>
</tr>
<tr>
<td><em>Delaware</em></td>
<td>2 ± 2</td>
<td>5 ± 3</td>
<td>7 ± 8</td>
</tr>
<tr>
<td><em>Hudson</em></td>
<td>2 ± 3</td>
<td>6 ± 2</td>
<td>7 ± 5</td>
</tr>
</tbody>
</table>

Table 7. Summary of modeling studies on the influence of climate change on streamflow in the Mid-Atlantic region.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Region</th>
<th>CO₂ scenario</th>
<th>Time period</th>
<th>Number of GCMs</th>
<th>Annual streamflow change (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moore et al. (1997)</td>
<td>Mid-Atlantic/New England</td>
<td>doubling</td>
<td>--</td>
<td>4</td>
<td>-32 to 6</td>
</tr>
<tr>
<td>Wolock and McCabe (1999)</td>
<td>Mid-Atlantic</td>
<td>1% yr⁻¹ increase</td>
<td>2090-2099</td>
<td>2</td>
<td>-25 to 33</td>
</tr>
<tr>
<td>Neff et al. (2000)</td>
<td>Susquehanna River Basin</td>
<td>1% yr⁻¹ increase</td>
<td>2090-2099</td>
<td>2</td>
<td>-4 to 24</td>
</tr>
<tr>
<td>Frei et al. (2002)</td>
<td>Southeastern New York</td>
<td>1% yr⁻¹ increase</td>
<td>2080-2089</td>
<td>2</td>
<td>-28 to 10</td>
</tr>
<tr>
<td>Hayhoe et al. (2007)</td>
<td>Pennsylvania and New Jersey</td>
<td>A1FI and B1</td>
<td>2070-2099</td>
<td>2</td>
<td>9 to 18</td>
</tr>
</tbody>
</table>
Table 8. Relationship between annual streamflow ($Q$) and precipitation ($P$) in Mid-Atlantic watersheds. See Figure 1 for streamgage locations.

<table>
<thead>
<tr>
<th>Watershed</th>
<th>USGS streamgage (station #)</th>
<th>Period analyzed</th>
<th>Calendar year</th>
<th>Relationship</th>
<th>$r^2$</th>
<th>Amplification factor$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chesapeake Bay$^b$</td>
<td>Conowingo (01578310) Potomac (01646500) James (02035000)</td>
<td>1930-2000</td>
<td>May-April</td>
<td>$Q = 0.698P - 2.636 \text{ cm mon}^{-1}$</td>
<td>0.78</td>
<td>1.72</td>
</tr>
<tr>
<td>Delaware Bay</td>
<td>Trenton (01463500)</td>
<td>1912-2000</td>
<td>June-May</td>
<td>$Q = 0.896P - 3.574 \text{ cm mon}^{-1}$</td>
<td>0.85</td>
<td>1.73</td>
</tr>
<tr>
<td>Susquehanna River</td>
<td>Conowingo (01578310)</td>
<td>1911-2000</td>
<td>July-June</td>
<td>$Q = 0.964P - 3.777 \text{ cm mon}^{-1}$</td>
<td>0.84</td>
<td>1.76</td>
</tr>
<tr>
<td>Hudson River</td>
<td>Green Island (01358000)</td>
<td>1946-2000</td>
<td>August-July</td>
<td>$Q = 0.868P - 2.930 \text{ cm mon}^{-1}$</td>
<td>0.83</td>
<td>1.61</td>
</tr>
</tbody>
</table>

$^a$Computed from $m \langle P \rangle / \langle Q \rangle$, where $m$ is the slope, and $\langle P \rangle$ and $\langle Q \rangle$ are long-term mean precipitation and streamflow, respectively (Najjar, 1999).

$^b$These three gages were summed.
<table>
<thead>
<tr>
<th>Time period</th>
<th>Precipitation change (%)</th>
<th>Temperature change (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2010−39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2040−69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2070−99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Average of all models

Average of CCCM, ECHM, HADC & NCAR
The figure compares temperature change (°C) with precipitation change (%) for different time periods and studies. The data points represent various studies as follows:

- McCabe & Ayers (1989)
- Moore et al. (1997)
- Najjar (1999)
- Polsky et al. (2000)
- Hayhoe et al. (2007)
- This study—-all models
- This study—best 4

The analysis indicates that there is a general trend of increased temperature change and decreased precipitation change, with a more pronounced effect during the summer season (SUMMER) compared to the winter season (WINTER).