

# The effects of climate downscaling technique and observational data set on modeled ecological responses

AFSHIN POURMOKHTARIAN,<sup>1,4</sup> CHARLES T. DRISCOLL,<sup>1</sup> JOHN L. CAMPBELL,<sup>2</sup> KATHARINE HAYHOE,<sup>3</sup> AND ANNE M. K. STONER<sup>3</sup>

<sup>1</sup>*Department of Civil and Environmental Engineering, Syracuse University, Syracuse, New York 13244 USA*

<sup>2</sup>*US Forest Service, Northern Research Station, Durham, New Hampshire 03824 USA*

<sup>3</sup>*Climate Science Center, Texas Tech University, Lubbock, Texas 79409 USA*

**Abstract.** Assessments of future climate change impacts on ecosystems typically rely on multiple climate model projections, but often utilize only one downscaling approach trained on one set of observations. Here, we explore the extent to which modeled biogeochemical responses to changing climate are affected by the selection of the climate downscaling method and training observations used at the montane landscape of the Hubbard Brook Experimental Forest, New Hampshire, USA. We evaluated three downscaling methods: the delta method (or the change factor method), monthly quantile mapping (Bias Correction-Spatial Disaggregation, or BCSD), and daily quantile regression (Asynchronous Regional Regression Model, or ARRM). Additionally, we trained outputs from four atmosphere–ocean general circulation models (AOGCMs) (CCSM3, HadCM3, PCM, and GFDL-CM2.1) driven by higher (A1fi) and lower (B1) future emissions scenarios on two sets of observations (1/8° resolution grid vs. individual weather station) to generate the high-resolution climate input for the forest biogeochemical model PnET-BGC (eight ensembles of six runs).

The choice of downscaling approach and spatial resolution of the observations used to train the downscaling model impacted modeled soil moisture and streamflow, which in turn affected forest growth, net N mineralization, net soil nitrification, and stream chemistry. All three downscaling methods were highly sensitive to the observations used, resulting in projections that were significantly different between station-based and grid-based observations. The choice of downscaling method also slightly affected the results, however not as much as the choice of observations. Using spatially smoothed gridded observations and/or methods that do not resolve sub-monthly shifts in the distribution of temperature and/or precipitation can produce biased results in model applications run at greater temporal and/or spatial resolutions. These results underscore the importance of carefully considering field observations used for training, as well as the downscaling method used to generate climate change projections, for smaller-scale modeling studies. Different sources of variability including selection of AOGCM, emissions scenario, downscaling technique, and data used for training downscaling models, result in a wide range of projected forest ecosystem responses to future climate change.

**Key words:** *climate change; ecological modeling; forested watershed; Hubbard Brook Experimental Forest (HBEF), New Hampshire, USA; hydrology; LTER; net primary productivity; statistical downscaling; stream nitrate; uncertainty analysis; variability.*

## INTRODUCTION

Dynamic forest biogeochemical models are useful tools to understand, evaluate, and predict the interactive effects of climate change, atmospheric CO<sub>2</sub> increases, and atmospheric deposition of anthropogenic pollutants on the hydrology and water quality of watersheds. Models have the ability to simulate the dynamics of energy, water, and element cycles in terrestrial ecosystems over spatio-temporal scales that are difficult to achieve through direct observation and experimentation.

Any model analysis is subject to multiple sources of uncertainty. For biogeochemical models, uncertainties stem from simplifications and assumptions regarding the hydrological, biological, and geochemical processes depicted in the model (structural uncertainty) (Gupta et al. 2012), as well as inaccurate parameterizations due to lack of data or uncertain observations (parametric uncertainty) (Wellen et al. 2015). When models are used to assess the potential impacts of climate change on a terrestrial ecosystem, uncertainties in climate projections enter the analysis. These include uncertainties in estimates of future emissions due to human activities (human or scenario uncertainty) (Stott and Kettleborough 2002), the ability of the atmosphere–ocean general circulation models

Manuscript received 27 April 2015; revised 19 October 2015; accepted 1 December 2015. Corresponding Editor: D. S. Schimel.

<sup>4</sup>E-mail: afshin.pourmokhtarian@gmail.com

(AOGCMs), or global climate models, to simulate the response of the climate system to human forcings (structural and parametric uncertainty) (Andrews et al. 2012), and uncertainties in the methods and/or models used to translate large-scale change into high-resolution projections (downscaling uncertainty) (Gutiérrez et al. 2012, Gutmann et al. 2014). The latter is the focus of this study.

Application of biogeochemical models to assess climate change impacts on forest ecosystems necessitates inputs of high-resolution simulated climatic variables covering a historical and future time period. Due to their coarse resolution (typically  $\sim 1\text{--}2^\circ$  in latitude and longitude) and recognized biases in the absolute values of temperature and precipitation, direct AOGCM projections are usually inadequate to assess climate change impacts on small watersheds, particularly in complex mountainous terrain that may be affected by highly localized topography and weather patterns. Often, statistical techniques have been applied to downscale coarse resolution AOGCM output to a finer spatial resolution, matching long-term observations (Hayhoe et al. 2004, 2007, 2008, Stoner et al. 2012). Numerous studies have evaluated the effect of different statistical downscaling techniques on hydrological model output (e.g., Chen et al. 2011, 2013, Liu et al. 2011, Gutmann et al. 2014), but we are not aware of previous studies that have considered the influence of different downscaling techniques on productivity and biogeochemical processes of forested watershed ecosystems or the effects of the choice of observations for training downscaling models.

We assessed the effects of three different downscaling methods and two sets of observations, which were used to train those techniques, on simulations of hydrology, water quality, and forest growth under potential future changing climate using a forest biogeochemical watershed model (PnET-BGC). Previous work has shown that PnET-BGC simulations are relatively sensitive to changes in meteorological inputs (temperature, precipitation, and photosynthetically active radiation [PAR]; (Pourmokhtarian et al. 2012). The goals of this study were to: (1) evaluate variability in climate change projections associated with the application of different downscaling techniques and sets of observations, which were used to train those techniques; (2) assess this variability in the context of other sources of variability in climate projections (i.e., AOGCM variability, scenario variability); and (3) quantify the implications of this variability in simulations of the response of a forest ecosystem in montane terrain to future climate change. PnET-BGC was applied at an intensively studied northern hardwood forest watershed at the Hubbard Brook Experimental Forest (HBEF) in New Hampshire, USA to characterize the sensitivity of projected climate change impacts resulting from the downscaling approach selected and the observations used for training, and to compare the magnitude of this variability to other sources of climate projection variability. This analysis improves understanding of the strengths and limitations of common statistical downscaling techniques

and selected data set for training, and helps guide the application of biogeochemical watershed models in climate change research.

## METHODS

### *Site description*

This study was conducted using long-term data from a relatively undisturbed reference watershed (watershed 6) at the Hubbard Brook Experimental Forest (HBEF) in the White Mountains of New Hampshire, USA ( $43^\circ 56' \text{ N}$ ,  $71^\circ 45' \text{ W}$ ) (Likens and Bormann 1995). The HBEF was established as a center for hydrological research in 1955 by the U.S. Forest Service, and joined the National Science Foundation Long-Term Ecological Research (LTER) network in 1987. The HBEF ([http://www.hubbardbrook.org/data/dataset\\_search.php](http://www.hubbardbrook.org/data/dataset_search.php)) has one of the longest and most extensive records on meteorology, hydrology, and biogeochemistry in the U.S. (Likens and Bormann 1995, Campbell et al. 2011). The climate is cool-temperate, humid continental, with an annual mean precipitation of  $\sim 1400$  mm that is distributed evenly over the year. Watershed 6 is  $0.13 \text{ km}^2$  and has complex terrain with an elevation range of 549–792 m. Soils are largely well-drained Spodosols, with bedrock at an average depth of 1–2 m. Vegetation is mostly northern hardwoods (Johnson et al. 2000). A detailed description of the site and long-term monitoring program is provided by Likens and Bormann (1995).

### *Ecosystem model PnET-BGC*

PnET-BGC is a deterministic forest-soil-water model that simulates energy, water, and element fluxes at the small watershed scale. PnET-BGC has been used to assess climate change impacts on northern hardwood forests (Campbell et al. 2009, 2011, Wu and Driscoll 2009, Pourmokhtarian et al. 2012) as well as atmospheric deposition and land disturbance on soil and surface waters across northern forest ecosystems (Chen and Driscoll 2005). The model was developed by linking a water, carbon, and nitrogen model (PnET-CN) (Aber and Driscoll 1997, Aber et al. 1997) to a biogeochemical (BGC; Gbondo-Tugbawa et al. 2001) sub-model, which enables the simultaneous simulation of major biotic and abiotic processes of major elements ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ ,  $\text{Na}^+$ , C, N, P, S, Si,  $\text{Al}^{3+}$ ,  $\text{Cl}^-$ , and  $\text{F}^-$ ) (Gbondo-Tugbawa et al. 2001). Climatic inputs to the model include meteorological data (PAR, precipitation, maximum and minimum temperature), atmospheric  $\text{CO}_2$  concentration, and atmospheric deposition (wet and dry). PAR was derived from solar radiation with the method described by Aber and Freuder (2000). A detailed description of model inputs and parameters are provided by Aber and Driscoll (1997), Aber et al. (1997), and Gbondo-Tugbawa et al. (2001), including a sensitivity analysis of parameters. A monthly time-step was used for model simulations. The model spin-up period started

at the year 1000 in order for the soil and vegetation pools to reach steady state. Methods for reconstruction of historical (hindcast) meteorological values are provided by Aber and Federer (1992), and Driscoll et al. (2001). The atmospheric deposition for 2012–2100 was assumed to not change from current conditions (business as usual). The dry-to-wet deposition ratios were assumed to be constant during the entire simulation period (Yanai et al. 2013) and wet deposition inputs are from the National Atmospheric Deposition Program (NADP) station (NH02) at the HBEF. Note that for these simulations we did not consider the effects of potential CO<sub>2</sub> on forest vegetation (Curtis et al. 1995, Lewis et al. 1996, Saxe et al. 1998, Ellsworth 1999, Ainsworth and Long 2005). Pourmokhtarian et al. 2012 used PnET-BGC to model the effects of atmospheric CO<sub>2</sub> on forest productivity, hydrology, and water quality. Although invoking CO<sub>2</sub> effects in this study would change the absolute forest ecosystem response under each climate change scenario, the relative differences within the ensembles of simulations would be similar; therefore, the findings of this study are independent of CO<sub>2</sub> effects.

#### *Climate scenarios and models*

The magnitude of future climate change depends on human emissions of CO<sub>2</sub> and other greenhouse gases. These in turn depend on a broad range of social, technological, and economic factors that affect energy use. To capture the range of plausible future scenarios, we compared projected climate change under the Intergovernmental Panel on Climate Change (IPCC-AR4) Special Report on Emissions Scenarios (SRES) higher (A1fi, or fossil-intensive) vs. the lower (B1) emissions scenarios (Nakicenovic et al. 2000). Under the higher scenario, atmospheric CO<sub>2</sub> levels reach nearly 970 ppm by 2100, while under the lower scenario, CO<sub>2</sub> concentrations stabilize around 550 ppm by 2100. Since the time of the analyses in this study, the more recent IPCC-AR5 scenarios have become available; however, use of these data would not affect the outcome since this work evaluates the influence of the downscaling techniques on forest ecosystem responses, rather than the scenarios used.

These scenarios were used as inputs to four AOGCMs: the Community Climate System Model version 3 (CCSM3) from the National Center for Atmospheric Research (NCAR) (Collins et al. 2006), the U.K. Meteorological Office Hadley Centre Coupled Model, version 3 (HadCM3) (Pope et al. 2000), the U.S. Department of Energy/National Center for Atmospheric Research Parallel Climate Model (PCM) (Washington et al. 2000), and the U.S. National Oceanographic and Atmospheric Administration/Geophysical Fluid Dynamics Laboratory (GFDL) model CM2.1 (Delworth et al. 2006). These AOGCMs were selected to provide a plausible range across AOGCMs with different climate sensitivities. We used output from each of the four AOGCMs corresponding to the fossil fuel-intensive, high-end A1fi scenario, as well as the

resource-efficient, low-end B1 scenario, as described in the Special Report on Emissions Scenarios (SRES) (Nakicenovic et al. 2000). In total, eight scenarios were developed for this application (two emissions scenarios applied to four AOGCMs), which cover the upper and lower range of values for future climate projections.

#### *Downscaling approaches*

Atmosphere–ocean general circulation models output for four daily variables (maximum and minimum temperature, precipitation, and solar radiation converted to PAR) were statistically downscaled using three methods in order to translate coarse-scale projections down to the small watershed-scale. Statistical downscaling accomplishes this translation by combining past observations with AOGCM simulations of historical conditions to calibrate a statistical model at the local scale. Historical observations can consist of individual station-based time series, specific to a certain weather station, or processed observations that have been interpolated onto a regular grid (e.g., Maurer et al. 2002). The statistical model “trains” a relationship between observed climate and model output to correct for biases in the climate model (Hayhoe et al. 2008). The relationship between AOGCM output and measured climate variables is then used to downscale future AOGCM outputs to the same scale (grid or weather station). The resulting time series is intended to match the statistics for the simulated historical output with observed climate over a relatively long period (at least 20 yr). Individual days or years do not match observations, as downscaled projections are based on AOGCM simulations, which are not constrained to match observed natural variability, but rather are allowed to develop their own patterns of variability based on initial conditions when the model simulation begins, typically prior to 1900 (NECIA, 2006, Hayhoe et al. 2008). Statistical downscaling is based on the assumption that the relationships between large and small scale processes remain constant over time. Although this assumption of constant relationship over time has been shown to vary by downscaling method and by quantile (Hayhoe et al. 2012), the statistical technique has a substantial time and cost benefit over dynamical downscaling and therefore is often more practical (Hayhoe et al. 2008). In this study we focused on statistical downscaling methods, rather than dynamical, because of their widespread use in ecological applications. Here, we compared results of forest biogeochemical model simulations using three different statistical downscaling approaches: the delta method (or the change factor method); monthly quantile mapping (Bias Correction-Spatial Disaggregation, or BCSD); and daily quantile regression (Asynchronous Regional Regression Model, or ARRM). Each method was trained on two sets of observations: interpolated gridded observations at a scale of 1/8° and individual weather station records. In total, we developed 48 runs (four AOGCMs, two

emissions scenarios, three downscaling methods, and two sets of observations).

The delta method is relatively simple and therefore widely used (e.g., in the U.S. Global Change Research Program National Assessment; USGCRP, 2001). The two main underlying assumptions for the delta method are: changes in climate occur only over large areas, comparable to the size of the AOGCM grid cell or larger, and existing relationships between climate variables will remain constant in the future (Hay et al. 2000). Therefore AOGCM accuracy in simulating changes in climate variables over the grid cell dictate the “delta” method output (Hay et al. 2000). The difference of the means between a future period and a historical period is calculated for an AOGCM simulation. This change or “delta” is added to observed mean monthly values to create future projections (Hay et al. 2000). Therefore the exact shape of the monthly distributions is retained, and the values are shifted by “delta” (for precipitation, the delta value is multiplicative rather than additive). Consequently, although the delta method is easy to compute, it is more limiting than other approaches because it does not account for changes in the frequency of extreme events (Hay et al. 2000) and assumes that AOGCMs are more reliable in simulating relative change than absolute change (Hay et al. 2000).

The BCSD is a relatively simple and commonly used method based on an empirical statistical technique known as quantile mapping, whereby probability density functions (PDFs) for modeled monthly temperature, precipitation, and solar radiation for a period of time (in this application, 1960–1999 for the ensemble members using gridded observations and 1964–2011 for ensemble members using station observations) are mapped by quantile onto historical observations (Maurer et al. 2002) (e.g., 90th percentile of historical AOGCMs are mapped on 90th percentile of historical observations) to create simulated monthly temperature and precipitation projections (Maurer and Hidalgo 2008). Monthly projections are disaggregated to the daily scale by sampling from observed months and scaling those daily values by the difference between the observed and simulated monthly mean (Maurer and Hidalgo 2008). The BCSD technique was originally developed to downscale ensemble climate model forecasts as input to a macro-scale hydrologic model, the Variable Infiltration Capacity (VIC) model, to simulate streamflow at spatial and temporal scales appropriate for large-scale water management (Wood et al. 2002). Following successful applications of the VIC model to forecast hydrological responses to climate change (e.g., Liang et al. 1994, Hamlet and Lettenmaier 1999, Wood et al. 2002, VanRheenen et al. 2004), the 1/8° grid, known as VIC grid, has become popular among hydrologists and climate scientists. A more detailed description of the BCSD downscaling method is provided by Hayhoe et al. (2004, 2007, 2008); Northeast Climate Impact Assessment (NECIA) (2006); and Campbell et al. (2011).

The ARRM is a more complex statistical method that uses quantile regression to determine relationships

between two quantities which have approximately normal distributions. These quantiles do not have temporal correspondence, but are expected to have similar statistical properties such as mean and variance (O'Brien et al. 2001). Although daily observations and daily AOGCM simulations do not have temporal correspondence, over climatological timescales (a minimum of 20 yr) they should theoretically have similar probability distributions for the historical period if the AOGCM simulates the climate effectively. However since this is not always the case, a downscaling model can be used to correct major dissimilarities. Assuming stationarity between observations and the AOGCM, simulations for the future require the same amount of correction for each quantile as historical periods. In ARRM, daily historical observations and AOGCM simulations are sorted by month, reordered by rank, and transformed into a near-normal distribution if necessary (e.g., wet-day precipitation is transformed using a log function). An independent piecewise linear regression is then derived for each month (Stoner et al. 2012) and used to correct future AOGCM output to match site-specific conditions in the future. The relationship between measured values and AOGCM simulations is improved further by additional steps such as pre-filtering the AOGCM output by principal component analysis (PCA) to remove low-level noise (meaning very small portions of the variability that do not contribute to the signal or trend in the data; Jackson 2004), spatially interpolating the AOGCM observations to the scale of the observations, and including information generated by the climate models for convective and large-scale precipitation (Stoner et al. 2012). A more detailed description of the ARRM method is provided by Stoner et al. (2012). The ARRM model has been used to generate gridded projections for the entire U.S. as used in The Third U.S. National Climate Assessment (Melillo et al. 2014), as well as station-based projections for individual climate impact studies for Chicago (Illinois, USA), Mobile (Alabama, USA), the state of Delaware (USA), and other locations (e.g., Hayhoe et al. 2010, 2014, U.S. DOT, 2012).

#### *Observations used in training the downscaling method*

The training component of the downscaling approach using station data was performed with observations obtained directly from the Hubbard Brook Experimental Forest (HBEF) weather stations: meteorological station 1 for maximum and minimum temperature, watershed 6 for precipitation (areal weighting of three precipitation collectors located in the vicinity of the watershed), and the HBEF headquarters building for solar radiation. The training period is based on available data for the historical period, which was 1964–2011 for temperature and precipitation variables, and 1959–2011 for solar radiation. A detailed description of the instrumentation and methods used to collect these data are presented in Bailey et al. (2003). Temperature and precipitation data for

1950–1999 from Maurer et al. 2002, which is observational data that has been interpolated to a 1/8° rectangular grid, were used to train the downscaling model for the gridded versions of the downscaled output. Solar radiation (<http://maps.nrel.gov/prospector>) grid-scale downscaling was trained on daily averaged hourly solar radiation from 1998–2009, for the center of a 10-km grid cell from the National Renewable Energy Laboratory.

*Statistical tests*

A Fisher’s Least Significant Difference (LSD) test was used to determine statistically significant differences ( $\alpha$  level of 0.05) in modeled streamflow, net primary productivity (NPP), and nitrate for each downscaling method/emissions scenario (e.g., ARRM-A1fi) over the three future periods (2011–2040, 2041–2070, and 2071–2100). The same statistical method was used to determine statistically significant differences among all six combinations of downscaling approaches/emissions scenario (ARRM-A1fi, ARRM-B1, Delta-A1fi, Delta-B1, BCSD-A1fi, and BCSD-B1) within each time period (e.g., 2011–2040). For each time period, an analysis of variance (ANOVA;  $\alpha$  level of 0.05) and eta squared ( $\eta^2$ ) effect size were performed for each downscaling method and emissions scenario (e.g., BCSD-A1fi and BCSD-B1 over 2071–2100). Similarly, for each time period, ANOVA ( $\alpha = 0.05$ ) and eta squared ( $\eta^2$ ) effect size were also applied to each downscaling technique for different sets of observations (e.g., ARRM-A1fi grid based [VIC] vs. ARRM-A1fi station based over 2011–2040).

RESULTS

*Observations used in training the downscaling method*

We had two *a priori* expectations for these analyses. First, since both the individual station and gridded data are based on observations, we expected a relatively high correspondence between values at the daily time scale. Secondly, we anticipated that interpolation over a grid would smooth topographical features that might produce more extreme conditions (warmer or colder temperatures, higher rainfall amounts) at any given location within a grid cell as compared to a given station that falls inside that cell. Hence, the mean values of the station-based observations could be offset relative to the gridded values, and furthermore, the variance or extremes of the station-based observations would likely be greater than those for the entire grid cell.

Monthly and seasonal means and standard deviations (SDs) are shown in Table 1 for both data sets over the same historical period (1964–2000 for maximum and minimum temperature and precipitation and 1998–2009 for solar radiation). The maximum temperature for gridded observations compared to station-based values is higher over both monthly and seasonal time scales, while the minimum temperature is higher for station-based observations. The SDs for both maximum and minimum

TABLE 1. Comparison of two sets of observations (station, abbreviated as S, and grid, G) used for training the three downscaling techniques over monthly and seasonal scales.

	1964–2000 (1998–2009 for PAR)											
	Latitude, longitude		Monthly		Spring		Summer		Fall		Winter	
	HBEF stations	center of grid	S	G	S	G	S	G	S	G	S	G
Temp. max. (°C)	43.9523 N, 71.7254 W	43.9375 N, 71.8125 W	10.4 (9.9)	11.8 (10.3)	15.9 (5.5)	17.7 (5.3)	21.5 (2.9)	23.3 (2.9)	5.1 (5.8)	6.2 (6.0)	−1.0 (3.4)	−0.1 (3.5)
Temp. min. (°C)	43.9523 N, 71.7254 W	43.9375 N, 71.8125 W	0.9 (9.4)	−1.1 (9.2)	5.5 (4.9)	3.5 (4.6)	11.6 (2.7)	9.3 (2.7)	−3.2 (5.3)	−5.0 (5.1)	−10.2 (3.8)	−12.3 (3.9)
Precip. (cm)	43.9535 N, 71.7396 W	43.9375 N, 71.8125 W	118.4 (53.4)	90.5 (42.5)	116.9 (58.6)	89.8 (44.2)	123.3 (55.7)	99.9 (40.8)	124.4 (48.7)	95.9 (42.6)	108.8 (49.4)	76.5 (38.9)
PAR (mmol·m <sup>−2</sup> ·s <sup>−1</sup> )	43.9439 N, 71.7019 W	43.9375 N, 71.8125 W	493 (173)	585 (164)	637 (93)	729 (83)	635 (66)	712 (63)	302 (87)	382 (65)	397 (112)	518 (97)

*Notes:* HBEF stations and grid center are shown as latitude (°N), longitude (°W). Mean of measured values is shown for the reference period (1964–2000), with the exception of solar radiation (1998–2009). Standard deviation (SD) is shown in parentheses. Seasons are defined as spring (April–June), summer (July–September), fall (October–December), and winter (January–March).

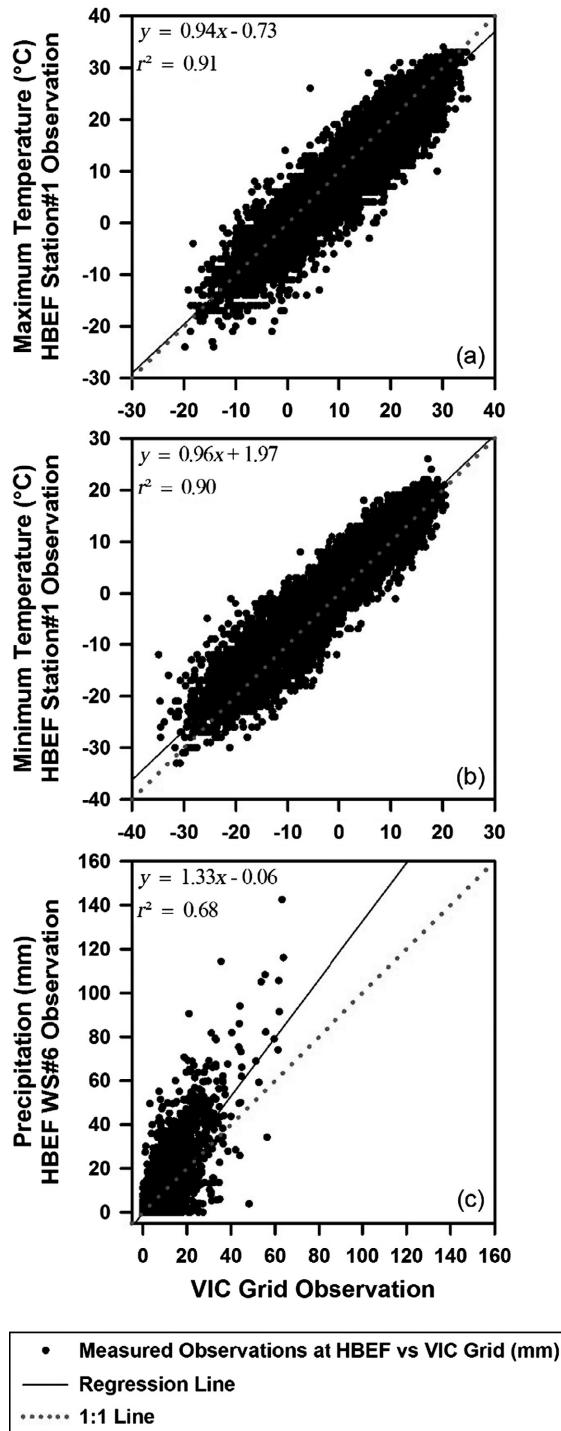


FIG. 1. Regression analysis between measured (a) maximum temperature ( $^{\circ}\text{C}$ ), (b) minimum temperature ( $^{\circ}\text{C}$ ) at HBEF station #1, and (c) measured precipitation (mm) at the Hubbard Brook Experimental Forest (HBEF; New Hampshire, USA) watershed (WS) #6 and values for the variable infiltration capacity (VIC) grid over the period 1964–2000.

temperature are similar, indicating that both data sets have similar variability across months and seasons. The mean and SD of precipitation values for station-based observations are higher compared to gridded values. The observed gridded solar radiation data are greater compared to station-based, although the variability in both sets of observations is similar.

To further test these comparisons, we performed a regression analysis between observations from the  $\sim 10 \text{ km}^2$  grid cell located over the HBEF stations and the station observations (Fig. 1). Observations from the grid cell on the  $x$ -axis were compared with station observations on the  $y$ -axis for (Fig. 1a) maximum temperature ( $^{\circ}\text{C}$ ), (Fig. 1b) minimum temperature ( $^{\circ}\text{C}$ ), and (Fig. 1c) precipitation (mm). There was a strong correlation for both maximum and minimum temperature ( $r = 0.91$ ;  $\beta = 0.94$ ;  $\text{SE} = 0.003$ ;  $P < 0.001$  and  $r = 0.90$ ;  $\beta = 0.96$ ;  $\text{SE} = 0.003$ ;  $P < 0.001$ , respectively). Maximum temperature shows a slight offset from the 1:1 line, which indicates that the station values are slightly cooler than the value for the entire grid cell. For minimum temperature, the values for the station are consistently slightly warmer than the grid cell values, suggesting that a local topographical influence (e.g., south-facing aspect) causes the station to be warmer than its surroundings within the  $1/8^{\circ}$  grid cell. The relationship for precipitation was not as strong ( $r = 0.68$ ;  $\beta = 1.33$ ;  $\text{SE} = 0.008$ ;  $P < 0.001$ ), suggesting that the relationship between precipitation amounts over the entire grid is not consistently correlated with the amount of precipitation falling at the station. This result is not surprising in an area with complex topography, but it does have important implications for the observations used to represent the hydrology of the small watershed. In addition, station-based precipitation amounts were generally greater than the grid-based values, especially for extreme events (Fig. 1c). Again, this pattern was expected, as spatial averaging has a tendency to smooth out local extremes. This result is illustrated by comparing time series of precipitation from the station-based observations (red) and the grid cell (blue) (Fig. 2).

It appears that using gridded vs. station-based observations results in little effect on mean maximum temperatures, slightly cooler mean minimum temperatures, and slightly damped daily extreme values of temperature. However, it has a profound effect on precipitation by altering the day-to-day distribution of precipitation as well as quantity. Therefore, there may be substantial differences in precipitation between the spatial average and an individual watershed location.

#### Future climate projections

Statistically downscaled AOGCM temperature projections for the HBEF, using all combinations (i.e., AOGCM, emissions scenario, downscaling method, training data set), indicated increases above the mean

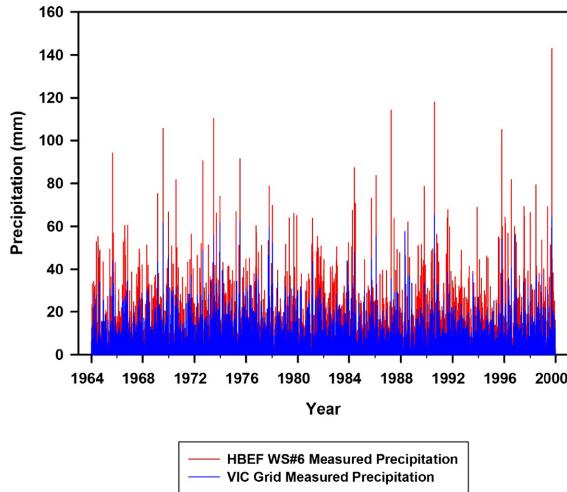


FIG. 2. Comparison of time series for measured monthly precipitation at the HBEF WS#6 and VIC grid (mm) over the period 1964–2000.

long-term annual measured air temperature of  $5.7^{\circ}\text{C}$ , ranging from  $+0.6^{\circ}\text{C}$  to  $+4.9^{\circ}\text{C}$  by the end of the century under the BCSD-VIC-B1 and Delta-A1fi, respectively (Table 2; VIC refers to gridded data). The choice of emissions scenario (higher A1fi vs. lower B1) and set of observations used for training the downscaling model had nearly equivalent impact on projected changes in temperature. The projected changes in temperature using station-based observations were higher compared to grid-based projections across all downscaling methods and emissions scenarios. The higher emissions scenario (A1fi) showed greater increases in temperature compared to the lower emissions scenario (B1). Differences due to downscaling technique were not significant (i.e., ARRMA1fi, Delta-A1fi, and BCSD-A1fi projected changes are similar). These results are consistent with Hayhoe et al. (2012), who found that different statistical downscaling models produce nearly identical changes in mean temperature.

The projected changes in precipitation were much more variable, ranging from a 243 mm decrease to a 327 mm increase above the long-term annual measured mean of 1440 mm (Table 1). The choice of training data for downscaling had a large impact on future projections of precipitation. Here, station-based projected changes in annual precipitation for all combinations of ARRMA, delta, and BCSD were significantly greater than grid-based projections. Station-based projections for all downscaling methods projected increases in precipitation, in contrast to the consistent decreases for grid-based downscaling. The station-based projections also showed higher mean precipitation, and more frequent extreme rainfall events during summer and fall compared to grid-based values. This pattern might be expected, considering downscaling trained on the station observations has higher mean precipitation and higher

precipitation extremes compared with the gridded observations. The impact of emissions scenarios on future projections of precipitation was not pronounced with the exception of ARRMA.

Projections of PAR were highly variable and indicated both increases and decreases ranging from  $-52$  to  $41 \text{ mmol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  compared to the long-term annual mean PAR of  $566 \text{ mmol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  (Table 2). The changes in PAR were small though (9% decrease to 7% increase) and there was no pattern of change associated with the choice of downscaling method, set of observations, and emissions scenario.

### Hydrology

When projected changes in temperature, precipitation, and solar radiation were used as input to PnET-BGC, simulations for all combinations (i.e., AOGCM, emissions scenario, downscaling method, training data set) indicate future changes in the timing of streamflow compared to the historical period (1981–2010), consistent with previous analyses (Campbell et al. 2011, Pourmokhtarian et al. 2012). All projections show that the current discharge regime of snowmelt-driven spring-flows that peak in April will likely shift to larger fall/winter streamflows under changing climate. In contrast to changes in the timing of streamflow, future projections of the overall amount of annual streamflow vary significantly depending on the AOGCM, downscaling method and training data used.

The 30-yr means ( $\pm$  SD) of projected annual streamflow for four AOGCMs with two emissions scenarios (A1fi and B1) using three downscaling methods and both sets of training observations are shown in Fig. 3.

TABLE 2. Summary of projected changes in mean annual temperature, precipitation, and PAR for three statistical downscaling techniques, two emissions scenarios, and two sets of training observations.

Downscaling/ emissions	Temperature ( $^{\circ}\text{C}$ )	Precipitation (mm)	PAR ( $\text{mmol}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ )
ARRM-A1fi	4.7 (5.0)	327 (135)	8.3 (7.3)
ARRM-B1	2.3 (4.9)	168 (114)	25.2 (10.4)
ARRM-VIC-A1fi	3.1 (6.1)	-163 (70)	-32.1 (13.8)
ARRM-VIC-B1	0.7 (6.3)	-236 (75)	-2.5 (12.1)
Delta-A1fi	4.9 (4.6)	233 (234)	-62.7 (49.7)
Delta-B1	2.5 (4.7)	192 (231)	-37.6 (52.2)
Delta-VIC-A1fi	3.3 (6.1)	-218 (173)	40.0 (21.4)
Delta-VIC-B1	0.8 (6.2)	-243 (169)	41.0 (21.5)
BCSD-A1fi	4.8 (4.4)	173 (109)	-51.8 (9.6)
BCSD-B1	2.3 (4.7)	198 (102)	-19.6 (12.7)
BCSD-VIC-A1fi	3.0 (6.1)	-218 (92)	-51.3 (11.9)
BCSD-VIC-B1	0.6 (6.3)	-195 (84)	-20.4 (15.3)

Notes: Each value shows the difference between the average of four AOGCMs for the period of 2070–2100 and the mean of measured values for the reference period of 1970–2000. Values in parentheses represent SD.

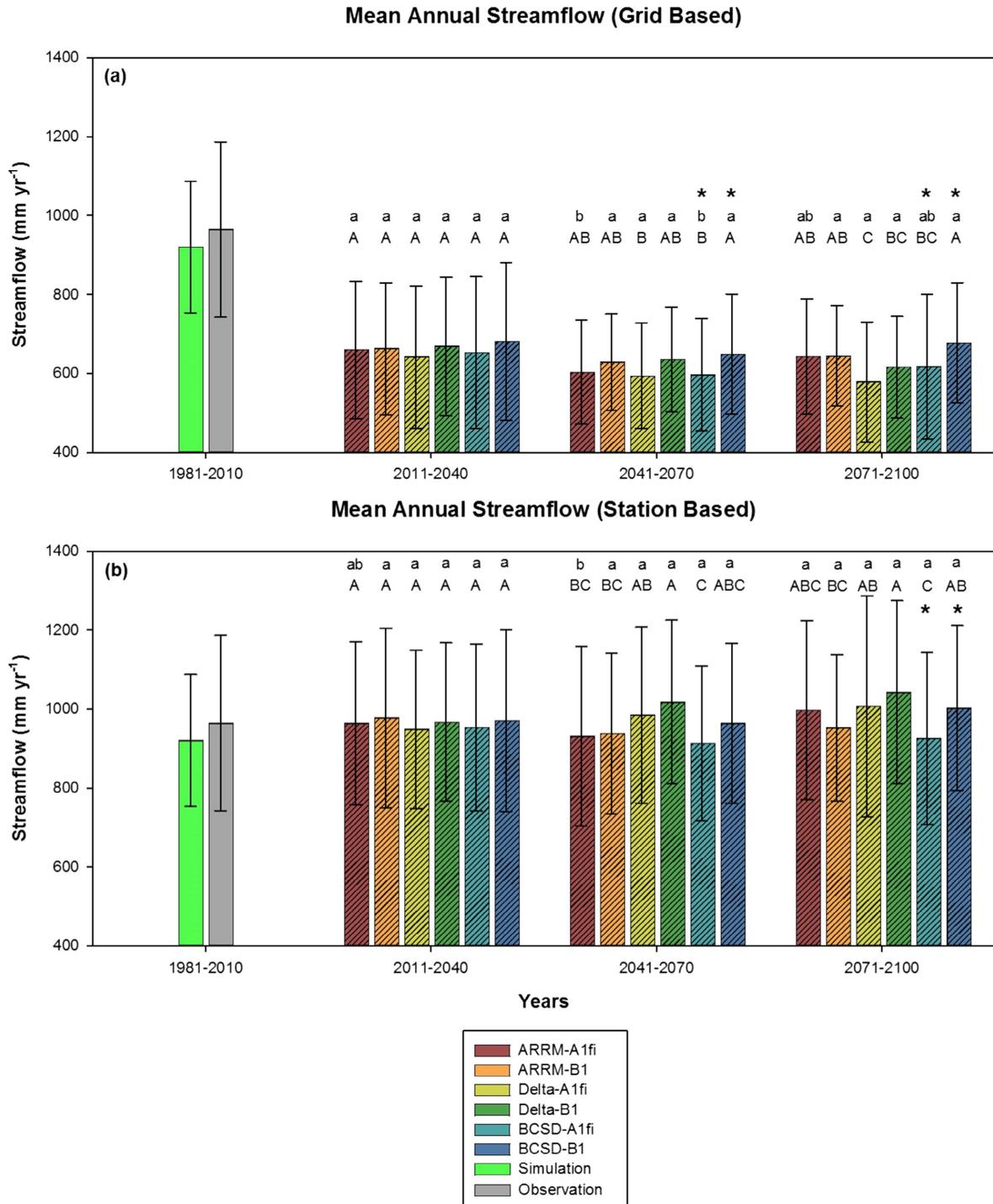


FIG. 3. Future projections of mean ( $\pm$  standard deviation; SD) annual streamflow over three 30-yr periods for three downscaling techniques and two sets of observations used for downscaling training. Lowercase letters shows the result of the Fisher's least significance difference (LSD) test with 95% confidence intervals for each downscaling method/emissions scenario (e.g., ARRM-A1fi) over the three projected time periods (2011–2040, 2041–2070, and 2071–2100). Uppercase letters indicates the Fisher's LSD results with 95% confidence interval among all six combinations (ARRM-A1fi, ARRM-B1, Delta-A1fi, Delta-B1, BCSD-A1fi, and BCSD-B1) within each time period (e.g., 2011–2040). The asterisk indicates significant differences ( $\alpha = 0.05$ ) determined with analysis of variance (ANOVA) between two emissions scenarios (A1fi and B1) at each time period for each downscaling method (e.g., BCSD-A1fi and BCSD-B1 over 2071–2100) (Table 2). Hatching indicates significant ANOVA test ( $\alpha = 0.05$ ) between different sets of observations (grid vs. station) for each downscaling technique at similar time step (e.g., ARRM-A1fi grid based vs. ARRM-A1fi station based over 2011–2040) (Table 3). The notations and symbols are the same for Figs. 4 and 5; net primary production (NPP) and streamwater nitrate, respectively.

TABLE 3. Summary of effect size for analysis of variance (ANOVA) between two emissions scenarios (A1fi and B1) for combinations of three downscaling techniques and two sets of observations over three 30-yr periods and throughout the century for stream discharge, net primary production (NPP), and stream water  $\text{NO}_3^-$  concentrations.

Time, by state variable	ARRM	ARRM-VIC	Delta	Delta-VIC	BCSD	BCSD-VIC
Stream discharge						
2011–2040	0.00	0.00	0.00	0.01	0.01	0.01
2041–2070	0.00	0.04	0.01	0.03	0.05	0.09
2071–2100	0.04	0.00	0.01	0.02	0.15	0.12
2011–2100	0.00	0.00	0.01	0.01	0.05	0.05
NPP						
2011–2040	0.00	0.01	0.00	0.00	0.03	0.00
2041–2070	0.06	0.05	0.00	0.00	0.31	0.08
2071–2100	0.02	0.00	0.10	0.01	0.01	0.12
2011–2100	0.00	0.01	0.01	0.00	0.03	0.00
Stream nitrate						
2011–2040	0.12	0.04	0.02	0.00	0.13	0.01
2041–2070	0.65	0.53	0.12	0.56	0.78	0.52
2071–2100	0.61	0.70	0.14	0.61	0.78	0.82
2011–2100	0.42	0.22	0.07	0.18	0.55	0.24

Notes: Values less than 0.01 are considered as small effect, values above 0.06 are medium effect, and values greater than 0.13 are large effect size. Values above 0.13 mean choice of emissions scenarios has a significant effect on the results for a selected state variable over the same time period.

Streamflow results using three downscaling methods were not statistically different over each time period, although toward the end of 21<sup>st</sup> century some statistical differences emerged without any specific pattern.

Compared to gridded training observations, use of station-based observations led to higher annual streamflow for all downscaling techniques under all emissions scenarios as well as higher SDs across all AOGCMs. Results indicated that the choice of emissions scenario (A1fi or B1) had the smallest effect size on future projected annual streamflow and differences were only significant for BCSD downscaling (Fig. 3, Tables 3 and Appendix S1: Table S1). Sets of training observations used for downscaling had the greatest effect size on projected annual streamflow compared to emissions scenarios and downscaling method (Tables 4 and Appendix S1: Table S2). The combination of different downscaling techniques, sets of training observations, and emissions scenarios showed a wide range of projected future annual water yield, ranging from a decrease of 342 mm per year to an increase of 122 mm, averaged across all AOGCMs. Projected changes in mean annual streamflow under BCSD showed greater variability, expressed as normalized projected percentage changes (SD divided by mean), when using grid-based observations compared to station-based (Appendix S1: Table S3). Patterns for the delta method and ARRM differed, in that projected variabilities were higher under the high emissions scenario (A1fi) compared to the low emissions scenario (B1) across all downscaling methods over the later two periods.

Model simulations using grid-based observations, projected lower flows during the summer (July–September) and higher flows in winter (January–March) compared to the historical period. The station-based simulations

projected higher flows compared to grid-based during the early winter and spring snowmelt, as well as summer. Model projections under both sets of observations showed an increase in future streamflow in late fall (October–December) and early winter, consistent with warmer air temperatures, less snow pack accumulation, and a decrease in the ratio of snow to rain. The use of station-based training observations led to a deeper snowpack and associated greater snowmelt-derived streamflow compared to grid-based observations.

Future model projections (i.e., 2011–2040, 2041–2070, 2071–2100) of soil moisture indicated a decline compared to 1981–2010 values (Appendix S1: Figure S1A). Soil moisture projections using station-based observations were higher than those derived with gridded observations (Appendix S1: Figure S1B). Model projections showed that under all simulations, decreases in soil moisture started earlier in the spring (April–June) due to earlier loss of snowpack, and wet-up occurred later into the fall. This phenomenon was more pronounced under the grid-based simulations.

#### *Net primary productivity (NPP)*

Future model projections of NPP using station-based observations were significantly greater than grid-based for all downscaling techniques (Fig. 4, Table 4). Model simulations using grid-based observations showed a decline in annual NPP for all downscaling methods. Simulations for ARRM-B1 and BCSD-B1 using grid-based observations remained relatively constant and there was no statistically significant difference over time. Results under ARRM-A1fi for grid-based showed statistically significant differences only for the 2011–2040

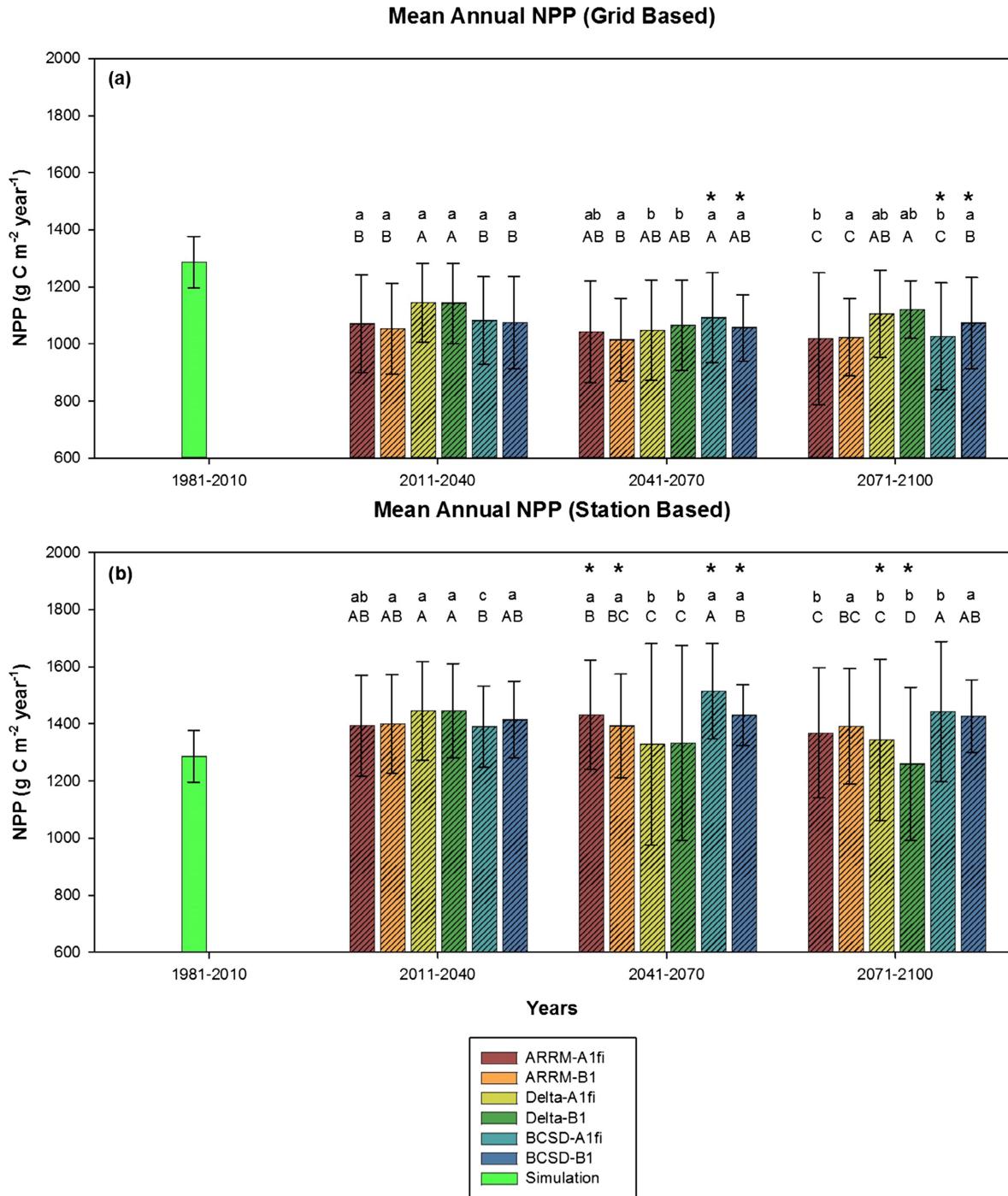


FIG. 4. Future projections of mean ( $\pm$  SD) net primary production (NPP) over three 30-yr periods for three downscaling techniques and two sets of observations used for downscaling training.

and 2071–2100 periods and BCSD-A1fi only showed a significant difference over the later future period (2071–2100) compared to the two earlier future periods (2011–2040 and 2041–2070). The delta method under both emissions scenarios showed a statistically significant difference only for the 2011–2040 and 2041–2070 periods.

The Fisher’s LSD test results indicated that the most significant difference under different grid-based downscaling techniques and emissions scenarios occurs over 2071–2100 (Fig. 4a). The future projected NPP using station observations were significantly greater than grid-based projections for all downscaling methods, and the

TABLE 4. Summary of effect size for analysis of variance (ANOVA) for two sets of observations used for three downscaling techniques over three 30-yr periods and throughout the century for stream discharge, NPP, and stream water nitrate concentrations.

Time, by state variable	ARRM-A1fi	ARRM-B1	Delta-A1fi	Delta-B1	BCSD-A1fi	BCSD-B1
Stream discharge						
2011–2040	0.64	0.63	0.41	0.40	0.60	0.57
2041–2070	0.78	0.80	0.60	0.59	0.76	0.68
2071–2100	0.77	0.80	0.60	0.60	0.73	0.79
2011–2100	0.71	0.73	0.54	0.53	0.69	0.67
NPP						
2011–2040	0.72	0.79	0.53	0.55	0.78	0.79
2041–2070	0.92	0.86	0.35	0.33	0.91	0.91
2071–2100	0.84	0.90	0.47	0.31	0.87	0.91
2011–2100	0.82	0.85	0.41	0.35	0.83	0.87
Stream nitrate						
2011–2040	0.03	0.05	0.14	0.11	0.00	0.15
2041–2070	0.54	0.63	0.18	0.00	0.47	0.68
2071–2100	0.79	0.87	0.70	0.34	0.83	0.91
2011–2100	0.29	0.43	0.11	0.02	0.26	0.54

Notes: Values less than 0.01 are considered as small effect, values above 0.06 are medium effect, and values greater than 0.13 are large effect size. Values above 0.13 indicates using different sets of observations has a significant effect on the results for a selected state variable over the same time period.

effect size of station-based training data was large ( $>0.5$ ) (Fig. 4b, Table 4, and Appendix S1: Table S2). The results indicated an increase in projected means and SDs of annual NPP for each time period compared to the historical period of 1981–2010 under station-based downscaling. The Fisher's LSD test showed that under ARRM-B1 and BCSD-B1, changes in NPP throughout the 21<sup>st</sup> century did not differ significantly among the three periods. The remaining four station-based downscaling simulations indicated statistically significant, but inconsistent, changes among these three periods, with BCSD-A1fi exhibiting the greatest change over time (Fig. 4b). The choice of emissions scenario did not have a large effect size during 2011–2040 for all downscaling methods/training observations used, but showed some significant differences for the later time periods, especially for the BCSD method (Table 3). Model projections of water use efficiency (WUE) indicated a significant decline using the grid-based observation (data not shown). In contrast, the station-based downscaling simulations showed higher WUE compared to grid-based projections for all scenarios. Overall, the set of observation used for training of the downscaling technique had the most profound effect on future projections of NPP, compared to emissions scenarios and downscaling technique. Projected changes in mean NPP under BCSD and ARRM showed higher variability (with the exception of ARRM-B1 over the 2071–2100 time period), when using grid-based observations compared to station-based (Appendix S1: Table S3). The delta method exhibited the opposite trend. Across all downscaling methods, projected variabilities were generally higher (with a few exceptions under delta method) under the high emissions scenario (A1fi) compared to low emissions scenario (B1).

#### *Stream nitrate*

All future simulations using grid-based observations showed significant increases in annual volume-weighted  $\text{NO}_3^-$  concentrations over the next century, although the magnitude and variability depend on the emissions scenario and downscaling technique (Fig. 5a). The use of station-based data in projections of mean annual volume-weighted  $\text{NO}_3^-$  concentrations had a large effect size compared to those of the grid-based downscaled simulations for all, with the exceptions of ARRM-A1fi and B1, BCSD-B1 over the 2011–2040 period and the Delta-B1 over the 2041–2070 period (Table 4 and Appendix S1: Table S2). Nitrate simulated with the station-based observations was lower than grid-based observations (Fig. 5b). The Fisher's LSD test revealed that future projections of streamwater  $\text{NO}_3^-$  concentrations for every downscaling method and each training observation data set used, changed significantly in at least one of the 30-yr periods compared to the other two periods throughout the 21<sup>st</sup> century, with the exception of ARRM-B1 under station-based observation downscaling. The choice of emissions scenario had a more pronounced impact on streamwater  $\text{NO}_3^-$  compared to stream discharge and NPP (Table 3, and Appendix S1: Table S1). In most cases, the B1 scenarios led to lower stream  $\text{NO}_3^-$  concentrations for all downscaling methods and observation training data set combinations. The exceptions were all three methods during the 2011–2040 time period using grid-based observations and the delta method using station-based observations, which did not show any significant differences over the three periods (Fig. 5, Table 3 and Appendix S1: Table S1). Similar to water yield and NPP, the choice of training observation data set had the most

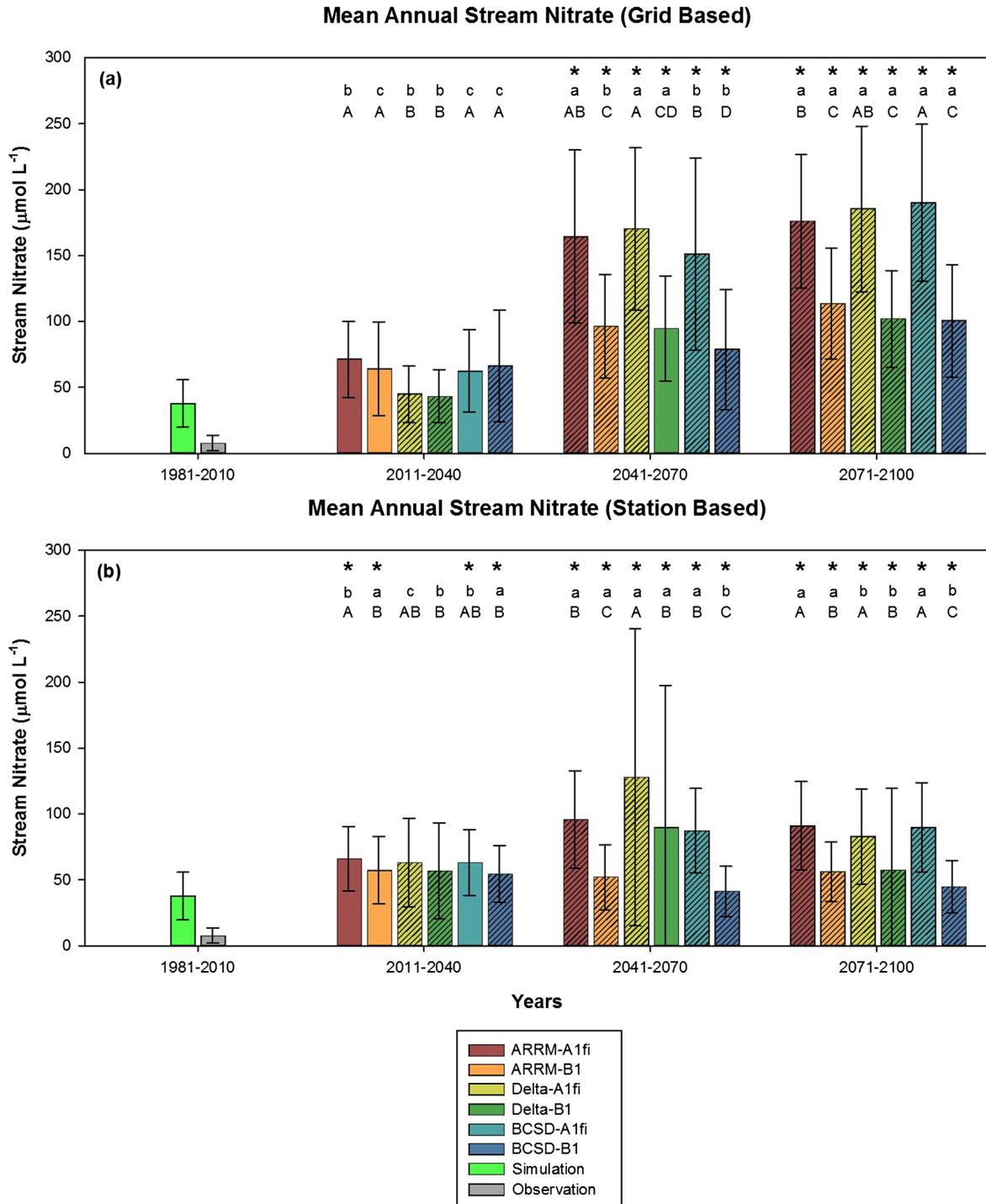


FIG. 5. Future projections of mean ( $\pm$  SD) annual streamwater  $\text{NO}_3^-$  concentrations over three 30-yr periods for three downscaling techniques and two sets of observations used for downscaling training.

significant impact on future projections of streamwater  $\text{NO}_3^-$  concentrations.

The selection of observation data set used for downscale training had the smallest impact on stream  $\text{NO}_3^-$  during the 2011–2040 period, and became more significant

over the last two periods. The Fisher’s LSD test showed that over each period the choice of training observation had a more significant impact on stream  $\text{NO}_3^-$  concentrations compared to streamflow and NPP (Fig. 5). Projected changes in streamwater  $\text{NO}_3^-$  concentrations

under the delta method showed lower variability using grid-based observations compared to station-based (Appendix S1: Table S3). The BCSD and ARRM methods had an opposite pattern, with the exception of the 2071–2100 time period and for ARRM-B1 over 2041–2070. Across all downscaling methods, projected variabilities were higher (with the exception of Delta-VIC over 2011–2040) under the low emissions scenario (B1) compared to high emissions scenario (A1fi).

## DISCUSSION

### *Future climate projections*

Overall, average temperature projections under the three downscaling techniques were similar for a given emissions scenario. This pattern was expected, as there is generally a good match between the gridded observations and HBEF station measurements of temperature. However, projected increases in temperature using station-based observations had a tendency to be higher than grid-based. This difference is likely due to not only the mean, but also the *shape* of the daily distribution of temperature, which is projected to shift toward higher temperatures in the future, with extreme high temperature days becoming relatively more frequent compared to the same quantile in the past, regardless of the mean value. Since there is generally a good match between sets of grid observations and HBEF station measurements of temperature, overall, temperature projections under the three downscaling techniques were relatively similar compared to precipitation projections. The station-based projections of precipitation were significantly higher than grid-based values, which affected modeled streamflow.

### *Streamflow forecasting*

There is a strong relation between precipitation and streamflow at the HBEF (Campbell et al. 2011) therefore, not surprisingly the modeled streamflow was sensitive to differences in precipitation associated with the downscaling method and observations used to train that method. Streamflow modeled with station-based training data was significantly higher than that generated using grid-based data for BCSD and ARRM. This difference is attributed to differences in the precipitation observations that were used for training the three downscaling techniques. Daily precipitation for watershed 6 (0.13 km<sup>2</sup>) at the HBEF is calculated with the Thiessen weighting method based on three rain gauges located along an elevational transect in the vicinity of the watershed. In contrast, grid-based downscaling uses a regional retrospective gridded observational database to train the downscaling model. Measured temperature and precipitation for a 1/8° grid cell (~10 km<sup>2</sup>) are not from a single station, but rather are statistical interpolations among multiple stations, not all of which are necessarily in the grid cell of interest. Therefore, in a relatively small

watershed in a mountainous region like watershed 6 at the HBEF, measurements of precipitation from on-site stations would be more representative of actual precipitation compared to values for the entire 1/8° grid. Note that a comparison of observations (graph not shown) by rank showed that the biases are rank- and not time-dependent. High precipitation amounts at the HBEF are always underestimated by the VIC grid regardless of the day they occur, presumably because precipitation is influenced by elevation, and the elevation for the entire grid is lower than the elevation for the site.

For the gridded projections, the annual discharge decreased under all simulations, due to the projected higher temperature and associated increase in evapotranspiration, coupled with the fact that the gridded observations underestimated historical (and therefore future) local precipitation patterns and extreme events. In contrast, with station-based downscaling, model simulations showed an increase in annual streamflow. This projected increase in streamflow occurs because the station, on which station-based downscaling is trained, better captures the magnitude of extreme events and local precipitation. As a result, the use of station-based observations more accurately depicts changes in the daily distribution of average and extreme events, resulting in significantly greater streamflow than grid-based observations. Extreme events are important in assessment of climate change impacts. There is a growing body of literature that focuses on climate variability, changes in return period, and the intensity of extreme events rather than “soft” extremes (Klein Tank and Können 2003), which are typically in the 90–95th percentile (Fowler et al. 2007). Changes in precipitation variability and extreme events have a strong impact on the hydrological cycle. Model simulations of future stream discharge indicate the increased importance of individual storms during summer and fall due to more frequent and intense extreme rainfall. Therefore, use of a station-based downscaling approach that is more capable of resolving daily extremes (e.g., ARRM), had a profound effect on modeled forest hydrological responses. ARRM is able to capture simulated changes in large precipitation events on a daily basis, by accurately resolving the relationship at the tails of the distribution (Stoner et al. 2012). Although this method is not useful if trained on grid-based observations, in which case it produces results that are similar to BCSD and the delta method. These results highlight the need to correctly characterize the quantity and distribution of future precipitation for accurate streamflow forecasting.

### *Forest growth and biomass*

Projected increases in precipitation and associated higher soil moisture and WUE under station-based downscaling resulted in increased tree growth compared to grid-based. In PnET-BGC, the length of growing season is determined by the minimum temperature, while

the maximum temperature affects photosynthesis and respiration. The difference between maximum and minimum temperatures determines vapor pressure deficit (VPD), which affects the WUE. In the absence of water stress, projected future forest growth should be higher than 20<sup>th</sup> century values due to the warmer, wetter climate and a longer growing season. An increase in both maximum and minimum temperature results in a longer growing season and higher VPD, causing water stress that offsets the enhancement of tree growth to some extent. Repeated water stress and drought during the growing season, results in decreases in projected maximum leaf area index, decreasing NPP. Increasing temperature can increase forest growth in two ways: by increasing the number of days with optimum photosynthetic temperature, or by alleviating N limitation through higher rates of soil N mineralization. However, the extent of these effects is limited by precipitation quantity and the seasonal pattern in soil moisture. This example illustrates the important interplay between projections of changing temperature and precipitation and their effects on the growth of the Northern Forest.

#### *Nutrient export from watersheds*

The export of elements from forested watersheds is strongly influenced by stream discharge (Likens and Bormann 1995), therefore future changes in the hydrological cycle, especially the seasonality and quantity of discharge, will likely affect water quality and nutrient loss. Soils at the HBEF have low base saturation and are sensitive to inputs of strong acid anions (Driscoll et al. 2001). Therefore, projections of elevated leaching of  $\text{NO}_3^-$  could re-acidify soil and stream water in acid-sensitive regions that have been impacted by acid deposition like the HBEF (Driscoll et al. 2003). Moreover, elevated export of  $\text{NO}_3^-$  from forest lands could alter the nutrient status of adjacent N-growth limited coastal waters (Driscoll et al. 2003). The simulated decline in soil moisture induced mid-summer drought stress on vegetation, which could decouple the linkage between soil and vegetation. Midsummer droughts and water stress decreases N uptake by trees and increases N availability, which leads to elevated loss of  $\text{NO}_3^-$  (Pourmokhtarian et al. 2012). For model runs using station-based downscaling, plant WUE increased and midsummer drought did not occur to the same extent as with grid-based simulations. Under station-based simulations, higher precipitation and associated increases in soil moisture and WUE offset the effect of higher temperatures, thereby minimizing future  $\text{NO}_3^-$  loss. In PnET-BGC, the decomposition rate of soil organic matter increases exponentially with increases in temperature, but increases linearly with increasing soil moisture. Therefore, under station-based downscaled scenarios of higher projected soil moisture and temperature, it might be anticipated that  $\text{NO}_3^-$  leaching would exceed projections compared to the grid-based approach. However, the absence of midsummer

drought and water stress due to higher precipitation, allowed plant demand for N to keep pace with the rate of soil N mineralization. As a result, the assimilation of the additional N produced under warmer temperatures, limited  $\text{NO}_3^-$  leaching. Nevertheless under the A1fi scenario during the second and third periods evaluated, the optimum temperature for photosynthesis is exceeded under all downscaling techniques, resulting in similar  $\text{NO}_3^-$  leaching. Seasonal patterns of annual N export varied with the downscaling technique and sets of observations used; with grid-based training data, elevated concentrations were projected in fall and with station-based,  $\text{NO}_3^-$  peaked during winter and spring snowmelt.

The ARRM method uses all the information provided by AOGCMs regarding projected changes in day-to-day variability and allows the shape of the probability distribution to change over time, shifting the mean, variance, and even the skewness (symmetry) of the distribution (Stoner et al. 2012). In contrast, the delta method derives the shape of the daily distribution from historical observations and therefore does not simulate projected changes in the shape of the distribution that affect the variance, compared with the historical period. We expected that these differences for delta compared to BCSD and ARRM would result in different projected biogeochemical responses, but interestingly, the differences were not as profound as expected and the selection of the set of observations used to train the downscaling method had a greater influence on the results.

#### *Quantifying variabilities*

Our study quantified different sources of variability in climate change projections, including AOGCMs, emissions scenarios, downscaling techniques, data used for training downscaling models, and quantified the manifestation of these variabilities in projected responses of a forest ecosystem in a montane terrain (Figs. 3–5, Table 2, and Appendix S1: Table S2). Our findings indicate that there is large variability across different AOGCMs. This variability demonstrates the importance of using multiple AOGCMs in climate impact assessments in order to capture a plausible range of responses. Additionally, a comparison of variability between two different emissions scenarios showed the profound effects of how policy choices on controlling future greenhouse gases influences forested ecosystems. Surprisingly, our analysis showed limited differences among the three downscaling approaches. The similarity in results from different downscaling techniques is likely due to the monthly time step considered (Hayhoe 2010). If the time step was at a finer interval (e.g., daily), differences in downscaling methods would have likely been more evident, especially for extreme events. This work did illustrate that the set of observations used for downscaling training can significantly impact variability in projected ecosystem responses to future climate change, particularly for precipitation in mountainous landscapes.

### Implications

This study provides new insights into the selection of statistical downscaling techniques and appropriate observations for “training” those techniques. It also introduces new sets of uncertainties beyond those generally associated with models used for climate change impact assessments in small forested watersheds. The choice of observational data compared to downscaling method had a much more profound effect on hydrology, which in turn affected forest growth and stream chemistry. These projected changes were directly related to the ability of the downscaling technique to mimic observed precipitation, emphasizing the need for careful selection of observations for “training” the downscaling technique at a scale that appropriately suits the scale and topography of watershed, as well as selection of a downscaling method that appropriately captures aspects of the distribution that contribute to observed impacts. In developing climate projections for small watersheds, particularly in areas with complex, mountainous terrain, it is important to use a downscaling technique that relies on measurements from the station within the watershed boundaries that can resolve projected changes at the sub-monthly scale. These measurements capture the actual variability of meteorological conditions for that watershed which improves the ability of the downscaling process (including both model and observations) to mimic the local climate patterns.

The findings of this study may be applicable to other ecosystems and modeling studies that use downscaled climatic input to project future climate change impacts. However, results may differ for larger basins because climate data from a single station may not adequately represent broad regions. At larger spatial scales, grid-based climate data may be preferable to station-based data. Grid-based data may also be suitable in flatter terrains, where climate is more uniform. When using grid-based data at the small basin scale, if weather stations are not located within or in close proximity to the study area, care should be taken to ensure that the gridded data represent the study area reasonably well. This is particularly important for precipitation, because smoothing within the grid cell tends to misrepresent extremes and underestimate volume.

### ACKNOWLEDGMENTS

The authors would like to thank the two anonymous reviewers for providing helpful comments that improved this manuscript. Funding for this study was provided by the Environmental Protection Agency through the STAR program, the USDA Northeastern States Research Cooperative, and the National Science Foundation (NSF) through the Long-Term Ecological Research (LTER) program. This manuscript is a contribution of the Hubbard Brook Ecosystem Study. Hubbard Brook is part of the LTER network, which is supported by the NSF. The Hubbard Brook Experimental Forest is operated and maintained by the USDA Forest Service, Northern Research Station, Newtown Square, Pennsylvania, USA.

### LITERATURE CITED

- Aber, J. D., and C. T. Driscoll. 1997. Effects of land use, climate variation, and N deposition on N cycling and C storage in northern hardwood forests. *Global Biogeochemical Cycles* 11:639–648.
- Aber, J. D., and C. A. Federer. 1992. A generalized, lumped-parameter model of photosynthesis, evapotranspiration and net primary production in temperate and boreal forest ecosystems. *Oecologia* 92:463–474.
- Aber, J. D., and R. Freuder. 2000. Variation among solar radiation data sets for the Eastern US and its effects on predictions of forest production and water yield. *Climate Research* 15:33–43.
- Aber, J. D., S. V. Ollinger, and C. T. Driscoll. 1997. Modeling nitrogen saturation in forest ecosystems in response to land use and atmospheric deposition. *Ecological Modelling* 101:61–78.
- Ainsworth, E. A., and S. P. Long. 2005. What have we learned from 15 years of free-air CO<sub>2</sub> enrichment (FACE)? A meta-analytic review of the responses of photosynthesis, canopy properties and plant production to rising CO<sub>2</sub>. *New Phytologist* 165:351–371.
- Andrews, T., J. M. Gregory, M. J. Webb, and K. E. Taylor. 2012. Forcing, feedbacks and climate sensitivity in CMIP5 coupled atmosphere-ocean climate models. *Geophysical Research Letters* 39:L09712.
- Bailey, A. S., J. W. Hornbeck, J. L. Campbell, and C. Eagar. 2003. Hydrometeorological database for Hubbard Brook Experimental Forest: 1955–2000. Gen. Tech. Rep. NE-305. Newtown Square, PA: US Department of Agriculture, Forest Service, Northeastern Research Station. 36 p.
- Campbell, J. L., et al. 2009. Consequences of climate change for biogeochemical cycling in forests of northeastern North America. *Canadian Journal of Forest Research* 39:264–284.
- Campbell, J. L., C. T. Driscoll, A. Pourmokhtarian, and K. Hayhoe. 2011. Streamflow responses to past and projected future changes in climate at the Hubbard Brook Experimental Forest, New Hampshire, United States. *Water Resources Research* 47:15 pp.
- Chen, L., and C. T. Driscoll. 2005. A two-layer model to simulate variations in surface water chemistry draining a northern forest watershed. *Water Resources Research* (W09425) 41:8 pp.
- Chen, J., F. P. Brissette, and R. Leconte. 2011. Uncertainty of downscaling method in quantifying the impact of climate change on hydrology. *Journal of Hydrology* 401:190–202.
- Chen, J., F. P. Brissette, D. Chaumont, and M. Braun. 2013. Performance and uncertainty evaluation of empirical downscaling methods in quantifying the climate change impacts on hydrology over two North American river basins. *Journal of Hydrology* 479:200–214.
- Collins, W. D., et al. 2006. The Community Climate System Model Version 3 (CCSM3). *Journal of Climate* 19:2122–2143.
- Curtis, P. S., C. S. Vogel, K. S. Pregitzer, D. R. Zak, and J. A. Teeri. 1995. Interacting effects of soil fertility and atmospheric CO<sub>2</sub> on leaf area growth and carbon gain physiology in *Populus × euramericana* (Dode) Guinier. *New Phytologist* 129:253–263.
- Delworth, T. L., et al. 2006. GFDL’s CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *Journal of Climate* 19:643–674.
- Driscoll, C. T., G. B. Lawrence, A. J. Bulger, T. J. Butler, C. S. Cronan, C. Eagar, K. F. Lambert, G. E. Likens, J. L. Stoddard, and K. C. Weathers. 2001. Acidic deposition in the northeastern United States: sources and inputs, ecosystem effects, and management strategies. *BioScience* 51:180–198.

- Driscoll, C. T., et al. 2003. Nitrogen pollution in the northeastern United States: sources, effects, and management options. *BioScience* 53:357–374.
- Ellsworth, D. S. 1999. CO<sub>2</sub> enrichment in a maturing pine forest: are CO<sub>2</sub> exchange and water status in the canopy affected? *Plant, Cell & Environment* 22:461–472.
- Fowler, H. J., S. Blenkinsop, and C. Tebaldi. 2007. Linking climate change modelling to impacts studies: recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology* 27:1547–1578.
- Gbondo-Tugbawa, S. S., C. T. Driscoll, J. D. Aber, and G. E. Likens. 2001. Evaluation of an integrated biogeochemical model (PnET-BGC) at a northern hardwood forest ecosystem. *Water Resources Research* 37:1057–1070.
- Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye. 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resources Research* 48:W08301.
- Gutiérrez, J. M., D. San-Martín, S. Brands, R. Manzanos, and S. Herrera. 2012. Reassessing statistical downscaling techniques for their robust application under climate change conditions. *Journal of Climate* 26:171–188.
- Gutmann, E., T. Pruitt, M. P. Clark, L. Brekke, J. R. Arnold, D. A. Raff, and R. M. Rasmussen. 2014. An intercomparison of statistical downscaling methods used for water resource assessments in the United States. *Water Resources Research* 50:7167–7186.
- Hamlet, A. F., and D. P. Lettenmaier. 1999. Effects of climate change on hydrology and water resources in the Columbia River basin. *JAWRA Journal of the American Water Resources Association* 35:1597–1623.
- Hay, L. E., R. L. Wilby, and G. H. Leavesley. 2000. A comparison of delta change and downscaled GCM scenarios for three mountainous basins in the United States. *JAWRA Journal of the American Water Resources Association* 36:387–397.
- Hayhoe, K. A. 2010. A standardized framework for evaluating the skill of regional climate downscaling techniques. University of Illinois, Urbana-Champaign, Illinois, USA.
- Hayhoe, K., et al. 2004. Emissions pathways, climate change, and impacts on California. *Proceedings of the National Academy of Sciences of the USA* 101:12422–12427.
- Hayhoe, K., et al. 2007. Past and future changes in climate and hydrological indicators in the US Northeast. *Climate Dynamics* 28:381–407.
- Hayhoe, K., C. Wake, B. Anderson, X.-Z. Liang, E. Maurer, J. Zhu, J. Bradbury, A. DeGaetano, A. Stoner, and D. Wuebbles. 2008. Regional climate change projections for the northeast USA. *Mitigation and Adaptation Strategies for Global Change* 13:425–436.
- Hayhoe, K., J. VanDorn, T. Croley II, N. Schlegal, and D. Wuebbles. 2010. Regional climate change projections for Chicago and the US Great Lakes. *Journal of Great Lakes Research* 36:7–21.
- Hayhoe, K., K. W. Dixon, A. M. K. Stoner, J. R. Lanzante, and A. Radhakrishnan(2012), Is the past a guide to the future? Evaluating the assumption of climate stationarity in statistical downscaling, *in* AGU Fall Meeting, Abstract GC41D-04, San Francisco, California, USA.
- Hayhoe, K., A. Stoner, and R. Gelca (2014), Climate Change Projections and Indicators for Delaware. *ATMOS Research & Consulting*. [http://www.dnrec.delaware.gov/energy/Documents/Climate%20Change%202013-2014/ARC\\_Final\\_Climate\\_Report\\_Dec2013.pdf](http://www.dnrec.delaware.gov/energy/Documents/Climate%20Change%202013-2014/ARC_Final_Climate_Report_Dec2013.pdf)
- Intergovernmental Panel on Climate Change (IPCC) (2007), *Climate Change 2007: The Physical Science Basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K.B. Averyt, M. Tignor, and H.L. Miller, editors. Cambridge University Press, Cambridge, UK and New York, New York, USA.
- Jackson, J. E. 2004. A user's guide to principal components, p. 592, John Wiley & Sons, Inc., New York, New York, USA, doi: 10.1002/0471725331.
- Johnson, C. E., C. T. Driscoll, T. G. Siccama, and G. E. Likens. 2000. Element fluxes and landscape position in a northern hardwood forest watershed ecosystem. *Ecosystems* 3:159–184.
- Klein Tank, A., and G. Können. 2003. Trends in indices of daily temperature and precipitation extremes in Europe, 1946–99. *Journal of Climate* 16:3665–3680.
- Lewis, J. D., D. T. Tissue, and B. R. Strain. 1996. Seasonal response of photosynthesis to elevated CO<sub>2</sub> in loblolly pine (*Pinus taeda* L.) over two growing seasons. *Global Change Biology* 2:103–114.
- Liang, X., D. P. Lettenmaier, E. F. Wood, and S. J. Burges. 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research* 99:14415–14428.
- Likens, G. E., and F. H. Bormann. 1995. *Biogeochemistry: of a forested ecosystem*. Springer-Verlag, New York, New York, USA.
- Liu, L., Z. Liu, X. Ren, T. Fischer, and Y. Xu. 2011. Hydrological impacts of climate change in the Yellow River Basin for the 21st century using hydrological model and statistical downscaling model. *Quaternary International* 244:211–220.
- Maurer, E. P., and H. G. Hidalgo. 2008. Utility of daily vs. monthly large-scale climate data: an intercomparison of two statistical downscaling methods. *Hydrology and Earth System Sciences* 12:551–563.
- Maurer, E. P., A. W. Wood, J. C. Adam, D. P. Lettenmaier, and B. Nijssen. 2002. A long-term hydrologically based dataset of land surface fluxes and states for the conterminous United States\*. *Journal of Climate* 15:3237–3251.
- Melillo, J. M., T. C. Richmond, and G. W. Yohe (2014), *Climate Change Impacts in the United States: The Third National Climate Assessment*. U.S. Global Change Research Program, 841 pp. doi:10.7930/J0Z31WJ2.
- Nakicenovic, N., J. Alcamo, G. Davis, B. de Vries, J. Fenhann, S. Gaffin, K. Gregory, A. Grübler, T. Y. Jung, and T. Kram (2000), *Special Report on Emissions Scenarios: A Special Report of Working Group III of the Intergovernmental Panel on Climate Change*, First edition, Cambridge University Press, Cambridge, UK and New York, New York, USA.
- Northeast Climate Impact Assessment (NECIA) (2006). *Climate Change in the U.S. Northeast. A Report of the Northeast Climate Impacts Assessment*. Union of Concerned Scientists (UCS), UCS Publications, Cambridge, Massachusetts, USA.
- O'Brien, T. P., D. Sornette, and R. L. McPherron. 2001. Statistical asynchronous regression: determining the relationship between two quantities that are not measured simultaneously. *Journal of Geophysical Research* 106:13247–13259.
- Pope, V. D., M. L. Gallani, P. R. Rowntree, and R. A. Stratton. 2000. The Impact of new physical parametrizations in the Hadley centre climate model: HadAM3. *Climate Dynamics* 16:123–146.
- Pourmokhtarian, A., C. T. Driscoll, J. L. Campbell, and K. Hayhoe. 2012. Modeling potential hydrochemical responses to climate change and increasing CO<sub>2</sub> at the Hubbard Brook Experimental Forest using a dynamic biogeochemical model (PnET-BGC). *Water Resources Research* 48:W07514.
- Saxe, H., D. S. Ellsworth, and J. Heath. 1998. Tree and forest functioning in an enriched CO<sub>2</sub> atmosphere. *New Phytologist* 139:395–436.

- Stoner, A. M. K., K. Hayhoe, X. Yang, and D. J. Wuebbles. 2012. An asynchronous regional regression model for statistical downscaling of daily climate variables. *International Journal of Climatology* 33:2473–2494. doi:10.1002/joc.3603.
- Stott, P. A., and J. A. Kettleborough. 2002. Origins and estimates of uncertainty in predictions of twenty-first century temperature rise. *Nature* 416:723–726.
- U.S. DOT (2012), Climate Variability and Change in Mobile, Alabama: Task 2 Final Report. Impacts of Climate Change and Variability on Transportation Systems and Infrastructure: The Gulf Coast Study, Phase 2. Report # FHWA-HEP-12-053. The U.S. Department of Transportation Center for Climate Change and Environmental Forecasting, Washington, D.C., USA.
- USGCRP: National Assessment Synthesis Team. 2001. Climate change impacts on the United States: the potential consequences of climate variability and change. Cambridge University Press, Cambridge, UK.
- VanRheenen, N. T., A. W. Wood, R. N. Palmer, and D. P. Lettenmaier. 2004. Potential implications of PCM climate change scenarios for Sacramento-San Joaquin River Basin hydrology and water resources. *Climatic Change* 62:257–281.
- Washington, W. M., et al. 2000. Parallel climate model (PCM) control and transient simulations. *Climate Dynamics* 16:755–774.
- Wellen, C., A.-R. Kamran-Disfani, and G. B. Arhonditsis. 2015. Evaluation of the current state of distributed watershed nutrient water quality modeling. *Environmental Science and Technology* 49(6):3278–3290. doi:10.1021/es5049557.
- Wood, A. W., E. P. Maurer, A. Kumar, and D. P. Lettenmaier. 2002. Long-range experimental hydrologic forecasting for the eastern United States. *Journal of Geophysical Research: Atmospheres* 107:ACL 6-1–ACL 6-15.
- Wu, W., and C. T. Driscoll. 2009. Impact of climate change on three-dimensional dynamic critical load functions. *Environmental Science and Technology* 44:720–726.
- Yanai, R. D., M. A. Vadeboncoeur, S. P. Hamburg, M. A. Arthur, C. B. Fuss, P. M. Groffman, T. G. Siccamo, and C. T. Driscoll. 2013. From missing source to missing sink: long-term changes in the nitrogen budget of a northern hardwood forest. *Environmental Science and Technology* 47(20):11440–11448. doi:10.1021/es4025723.

#### SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at <http://onlinelibrary.wiley.com/doi/10.1890/15-0745/suppinfo>

#### DATA AVAILABILITY

Data associated with this paper are available at LTER: <https://portal.lternet.edu/nis/mapbrowse?packageid=knb-lter-hbr.173.2>