



A comparison of statistical and dynamical downscaling for surface temperature in North America

Scott Spak,¹ Tracey Holloway,¹ Barry Lynn,² and Richard Goldberg³

Received 26 September 2005; revised 11 October 2006; accepted 14 October 2006; published 20 April 2007.

[1] Projections from general circulation model (GCM) simulations must be downscaled to the high spatial resolution needed for assessing local and regional impacts of climate change, but uncertainties in the downscaling process are difficult to quantify. We employed a multiple linear regression model and the MM5 dynamical model to downscale June, July, and August monthly mean surface temperature over eastern North America under greenhouse gas-driven climate change simulation by the NASA GISS GCM. Here we examine potential sources of apparent agreement between the two classes of models and show that arbitrary parameters in a statistical model contribute significantly to the level of agreement with dynamical downscaling. We found that the two methods and all permutations of regression parameters generally exhibited comparable skill at simulating observations, although spatial patterns in temperature across the region differed. While the two methods projected similar regional mean warming over the period 2000–2087, they developed different spatial patterns of temperature across the region, which diverged further from historical differences. We found that predictor domain size was a negligible factor for current conditions, but had a much greater influence on future surface temperature change than any other factor, including the data sources. The relative importance of SD model inputs to downscaled skill and domain-wide agreement with MM5 for summertime surface temperature over North America in descending order is Predictor Domain; Training Data/Predictor Model; Predictor Variables; and Predictor Grid Resolution. Our results illustrate how statistical downscaling may be used as a proxy for dynamical models in sensitivity analysis.

Citation: Spak, S., T. Holloway, B. Lynn, and R. Goldberg (2007), A comparison of statistical and dynamical downscaling for surface temperature in North America, *J. Geophys. Res.*, 112, D08101, doi:10.1029/2005JD006712.

1. Introduction

[2] While general circulation models (GCMs) allow for projection of future climate under a range of scenarios, understanding the local and regional impacts of climate change often necessitates data at a higher resolution than that of the global GCM. The problem of translating coarse-scale GCM and reanalysis output to the finer spatial scales required for local climate change projection and regional impacts analysis relies on two general classes of methods to estimate climate variables at a higher resolution: dynamical and statistical downscaling. Dynamical downscaling takes GCM boundary conditions to drive a regional climate model (RCM), such as those used in operational weather forecasting, in which atmospheric properties are calculated on a finer grid by solving equations of motion and thermodynamics. These models are able to generate a dynamically

consistent suite of climate variables, but there is significant uncertainty in parameterization of sub-grid-scale processes, and the computational costs of RCMs are high. Statistical downscaling (SD) evaluates observed spatial and temporal relationships between large-scale (predictors) and local climate variables (predictands) over a specified training period, and extends these relationships to project the time series of predictands from the predictors.

[3] This experiment complements the New York Climate and Health Project (NYCHP), a funded health impacts assessment study based at Columbia University and funded by the U.S. Environmental Protection Agency. The NYCHP employs a multiscale modeling framework for assessing the changes in heat- and ozone-related mortality in the 31-county New York City metropolitan area resulting from projected climate and land use change over the next 80 years. The NYCHP framework incorporates a global climate model, a regional climate model (B. H. Lynn et al., The GISS-MM5 regional climate modeling system: Sensitivity of simulated current and future climate to model physics configuration and grid-resolution, submitted to *Journal of Climate*, 2005; hereinafter referred to as Lynn et al., submitted manuscript, 2005), regional land use modeling [Solecki and Oliveri, 2004], and a regional air

¹Center for Sustainability and the Global Environment, University of Wisconsin-Madison, Madison, Wisconsin, USA.

²Center for Climate Systems Research, Columbia University, New York, New York, USA.

³NASA Goddard Institute for Space Sciences, New York, New York, USA.

quality model [Hogrefe *et al.*, 2004] to generate temperature and ozone projections for human health risk analysis [Knowlton *et al.*, 2004].

[4] Publications from the NYCHP [Lynn *et al.*, 2004; Hogrefe *et al.*, 2004] have highlighted uncertainty within the RCM regional temperature scenario and emphasized the importance of the regional temperature scenario as a primary influence on projected changes in regional ozone and corresponding health impacts. A more thorough understanding of uncertainty in regional temperature resulting from the choice of climate downscaling procedure is a critical requirement if integrated assessment results are to be applied to inform policy decisions.

[5] Performance of dynamical and statistical methods in simulating contemporary climate has been formally compared in studies by Kidson and Thompson [1998], Mearns *et al.* [1999], Murphy [1999], and Oshima *et al.* [2002], with analysis limited to temperature and precipitation fields and confined to North America, Europe, and Japan. Similar levels of skill for present-day climate for the dynamical and statistical methods are a common finding, independent of region, RCM, GCM, statistical technique, temporal scale, and even performance metric.

[6] Mearns *et al.* [1999], Murphy [2000], Oshima *et al.* [2002], and Hanssen-Bauer *et al.* [2003] compared RCM and SD methods directly under projected climate change. These comparative studies all found divergence between the two downscaling methods for temperature projections under climate change forcings, but without systematic explanations for the magnitude of divergence. Murphy noted a change in the strength of predictor/predictand relationships, and Mearns *et al.* found that SD produced an amplified seasonal cycle, while the RCM generated greater variability in the spatial patterns of regional temperature change.

[7] In light of the comparable skill exhibited by RCMs and SD methods at daily and monthly timescales under present conditions, and the consistent suggestion of even limited agreement between the methods for future projections, both classes of downscaling techniques may be used to generate plausible regional climate scenarios. Used together, the two methods may contribute to improved qualitative and quantitative metrics of structural [Thorne *et al.*, 2005] uncertainty in the downscaled results.

[8] While a combination of these methods is a best practice [Wilby *et al.*, 2004], it is currently unreasonable to expect that every impact assessment will have the budget and personnel to include RCM downscaling in the assessment framework, particularly in nations with limited capacity in climate modeling. It is also unlikely that every assessment will have timely access to statistical climatologists who can incorporate expert knowledge of SD algorithms and regional climate.

[9] Instead, SD methods known to yield results consistent with RCMs are a tempting choice as the sole source of downscaled scenarios in impact assessment when RCM simulation is unavailable, or where the time and expense of RCM simulation are otherwise prohibitive. Statistical downscaling represents regional- and local-scale phenomena better than GCM change factors, whereby future changes in climate projected by GCMs are applied directly to a local baseline climatology [Diaz-Nieto and Wilby, 2005].

[10] This study provides a basis for identifying the elements of an SD model that contribute most to potential

agreement with RCMs, and attempts to understand the physical reasons behind any apparent agreement. The objective is twofold: to provide more comprehensive understanding of confidence and uncertainty in the NYCHP downscaled scenarios in accordance with current best practices in integrated assessment, and to provide a global template for SD model selection by the impacts analyst. We have established a flexible, globally applicable protocol for statistical downscaling of surface temperature (T_{SFC}), and identified a framework for selecting appropriate downscaling parameters.

[11] We projected summer monthly mean T_{SFC} in order to describe chronic, rather than peak, effects of high temperature for analysis of potential human health impacts. Although our methodology is not timescale dependent, the choice of the monthly mean is reinforced by the scarcity of archived GCM and standardized, gridded observation data at higher temporal resolution. Geographically specific daily and sub-daily time series projections can also be generated from the monthly mean through weather generators.

2. Materials and Methods

2.1. Observed Data

2.1.1. Predictands

[12] The primary surface temperature record employed for SD training was the University of Delaware Air Temperature and Precipitation $0.5^\circ \times 0.5^\circ$ monthly mean gridded data set (Willmott and Matsuura (DE) [2003], hereinafter referred to as DE) Although SD studies typically analyze time series from individual locations, we choose to use interpolated climate data, with global coverage over all landmasses for 1951–1999. This choice of training data facilitates the transferability of our results to other areas. Over the contiguous United States, DE is based primarily on interpolation of station observations in the U.S. Historical Climate Network [Karl *et al.*, 1990a], by way of their inclusion in the Global Historical Climate Network version 2 [Vose *et al.*, 1992], using the spherical version of Shepard's distance-weighting method [Willmott *et al.*, 1985], as well as digital elevation model height interpolation and spatial adjustment by Climatologically Aided Interpolation [Willmott and Robeson, 1995]. The U.S. Historical Climate Network, compiled by the National Climatic Data Center, has been adjusted to remove bias introduced by station moves, instrument changes, time-of-observation differences, and urbanization effects. Vose and Menne [2004] found the network exceeded the density required over the period 1971–2000 to capture changes in the spatial mean of climate parameters at a regional scale. Station averages were interpolated to a $0.5^\circ \times 0.5^\circ$ grid, with grid nodes centered on the 0.25 degree. An average of 20 nearby stations influence each grid node, and over the United States the Climatologically Aided Interpolation was conducted using a 5685 station high-resolution network.

[13] The University of East Anglia Climate Research Unit (CRU) TS 2.0 global 0.5° monthly transient climate grids for 1901–2000 [New *et al.*, 2000; Mitchell and Jones, 2004] served as an alternative training record. Since CRU is on the same grid as DE, and was compiled from similar source data by other methods, the inclusion of CRU data allowed us to examine SD model sensitivity to small differences in inter-

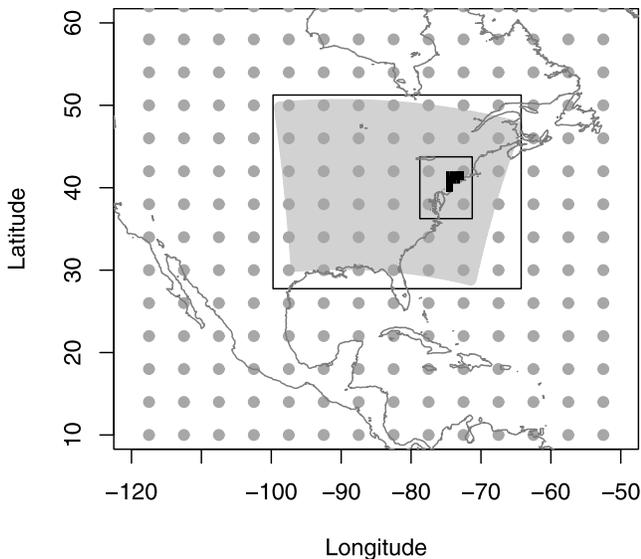


Figure 1. Predictor field and downscaling study areas. The regional modeling domain is shaded in light gray. Boxes represent boundaries of the three scales of predictor domain. *Wilmott and Matsuura (DE)* [2003] grid cells representing the three state, 31-county New York City metropolitan area (NYC) are shown in black. General circulation model (GCM) grid cell centers are shown as gray points.

polated observation data, which may arise because of variations in the number and distribution of stations included, and the choice of interpolation techniques.

2.1.2. Predictors

[14] The National Centers for Environmental Prediction (NCEP) $2.5^\circ \times 2.5^\circ$ Reanalyses of 1990–2004 monthly mean T_{SFC} and sea level pressure (*Kalnay et al.* [1996], NCEP/NCAR Reanalysis Project, 2004, through July 2004; NCEP/NCAR Reanalysis 1, NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado) were used as large-scale predictors to provide performance benchmarks against which GCM-downscaled results for the 1990s could be compared.

2.2. Climate Simulation Methods and Data

2.2.1. Global Climate Models

[15] We downscaled June, July, and August monthly mean temperature for the eastern half of the United States and southern Canada from monthly mean climate variables in simulations with the NASA Goddard Institute for Space Studies $4^\circ \times 5^\circ$ resolution Global Atmosphere-Ocean Model GCM (GISS-GCM) [*Russell et al.*, 1995]. The GCM simulated conditions for model years 1990–2087, with projections based on climate forcing from the IPCC 'A2' scenario [*IPCC*, 2000], as described by *Lynn et al.* [2004] (see also *Lynn et al.*, submitted manuscript, 2005).

[16] The Canadian Centre for Climate Modelling and Analysis Coupled Global Climate Model (CGCM2) [*Flato and Boer*, 2001] simulation for 1990–2087 with the A2 forcings was employed as an alternative predictor source.

2.2.2. Regional Climate Models

[17] Dynamically downscaled current and future regional climate fields were obtained by coupling the Pennsylvania State University/National Center for Atmospheric Research

mesoscale regional climate model (MM5) to GISS-GCM in a one-way mode through initial conditions and lateral boundaries (*Lynn et al.*, submitted manuscript, 2005). Simulations were performed for five consecutive summer seasons (June–August) in the 1990s and three future decades, namely 1993–1997, 2023–2027, 2053–2057, and 2083–2087. Following the NYCHP, we chose model runs by *Lynn et al.* [2004], with the Betts-Miller cumulus parameterization (MIBR) as the primary RCM for comparison. In order to directly compare the relative sensitivity and uncertainty stemming from arbitrary choices in the RCM to those in SD, we employed MM5 with the Grell cumulus parameterization (MIGR). MM5 results were interpolated from the original 36×36 km grid to the $0.5^\circ \times 0.5^\circ$ DE grid for downscaling model intercomparison.

2.2.3. Statistical Downscaling: Downscaling Methods

[18] In order to compare SD directly with the RCM employed by the NYCHP, we downscaled GISS-GCM over the same domain as MM5. We assessed performance across the entire domain and over the three state, 31-county New York City metropolitan area (NYC) used for the NYCHP health impacts assessment. The SD domain and NYC subdomain are depicted in Figure 1.

2.2.3.1. Downscaling Model

[19] Surface temperature at each DE grid point was estimated by multiple linear regression statistical downscaling, a method well suited to the normally distributed anomalies in DE, and for monthly T_{SFC} in general. Spatial patterns of anomalies in the predictor fields, with the annual cycle removed, were decomposed into empirical orthogonal functions (EOFs). These functions are orthogonal eigenvectors aligned so the leading EOF describes the spatial pattern that maximizes variance [*Preisendorfer*, 1988]. The twenty leading EOFs (or for the small-scale predictor field, the spatially constrained maximum of four) were calculated, with the first EOF of T_{SFC} and mean sea level pressure (MSLP) explaining 85–95% of the GISS-GCM and NCEP variance in these variables at all scales, and the eight leading EOFs always capturing more than 99% of the variance. We chose to include only the leading eight EOFs to avoid inflating the warming response, as *Huth* [2004] found that downscaled climate change estimates for local daily T_{SFC} in central and western Europe from multiple linear regression of EOFs was most dependent on the inclusion of monopolar or imbalanced multipolar EOFs as predictors. It is critical to limit the inclusion of EOFs that contain monopoles or unbalanced dipoles to avoid erroneously compounding the climate change signal.

[20] Downscaling was conducted by multiple linear regressions on the eight leading EOFs of monthly anomalies in predictor fields and the time series of temperature at each of the 1875 DE grid points in the region, with the annual cycle and linear trend removed (as in *Murphy* [1999]). The algorithms of *Benestad* [2004a] were used to fit the predictor-predictand relationship to a fifth-order polynomial in time. Full-year time series maximized training data density, and full-year monthly projections were generated for the years 1997–2087. Since previous comparative studies highlighted changes in the strength of predictor/predictand relationships, seasonal cycles, and spatial patterns of regional temperature change, we did not perform variance inflation [*Karl et al.*, 1990b] to adjust variance.

[21] In order to quantify SD sensitivity to model parameters, we selected predictor variables and predictor domains for study and downscaled using every combination of parameters. We then selectively evaluated predictor resolution and predictor/predictand data sources.

2.2.3.2. Predictor Variables

[22] In order to yield the most direct comparison between statistical and dynamical downscaling, we downscaled variables from the same GISS-GCM simulation used to drive the RCM in *Lynn et al.* [2004] (see also *Lynn et al.*, submitted manuscript, 2005). Most statistical downscaling studies to date have generated predictions based in part on free atmospheric variables, with the rationale that these are better simulated in GCMs than surface variables. As a result, a primary source of uncertainty in downscaling under climate change scenarios is the extent to which observed relationships between large-scale predictor variables and downscaled predictands will remain consistent under altered climate regimes. *Murphy* [2000] notes that both dynamical and statistical estimates of downscaled surface warming may be misleading because the link between T_{SFC} and 850 mb temperature (T_{850}) in GCMs are often stronger than the corresponding link found in observations, and it may be that GCMs do not simulate T_{850} substantially better than they simulate T_{SFC} . The GISS-GCM appears to fit this description, as spatial patterns in JJA T_{SFC} and T_{850} are strongly correlated over the study and predictor domains, and their biases (-2 to $+3^{\circ}\text{C}$) are of the same magnitude and spatial distribution.

[23] Some early studies showed a strong agreement between daily GCM-simulated series of T_{SFC} and observed values at single locations [*Portman et al.*, 1992; *Robinson et al.*, 1993] or averaged over only a few (three or fewer) stations [*Rind et al.*, 1989]. Furthermore, *Benestad* [2003] found that large-scale surface predictor fields from GCMs, including MSLP, explained observed regional surface-level variance over northern Europe better than free atmospheric circulation indices, and *Benestad* [2004b] then used GCM 2 m surface temperature as a sole predictor for downscaled T_{SFC} . Since local temperature under increased atmospheric CO_2 may be dominated by changes in the radiative properties of the atmosphere rather than changes in upper atmosphere circulation, employing the large-scale T_{SFC} field from the GCM as a predictor may be an effective means of capturing the climate change signal [*Dehn and Buma*, 1999].

[24] Following *Hanssen-Bauer et al.* [2003], we selected T_{SFC} and MSLP a priori as predictor variables to create a parsimonious “base case” predictor set for flexible, global application independent of season or location. Two predictors were assessed: T_{SFC} alone, and a mixed field (TP) of T_{SFC} and MSLP concatenated, with the combined field decomposed into principal components [*Benestad et al.*, 2002]. These two fields meet the criteria for successful predictors [*Wilby et al.*, 2004], consistently explain significant variance in global land surface temperature, capture climate change signals, and are the most frequently chosen predictors for T_{SFC} when stepwise regression is conducted using all available GCM output and derived variables [*Murphy*, 1999]. As a result, they represent an effective minimal set to which additional geographically specific predictors may be added. Finally, these variables are common to climate models, reanalyzed and gridded climate data sets; and have reason-

ably long and dense global in situ and remote observational records.

2.2.3.3. Predictor Domain

[25] If a predictor domain larger than the single proximate grid cell is considered, the domain that best describes the relationship between large-scale influences and local effects for a particular climate variable at any point may change dramatically if climatic changes alter synoptic patterns. While *Huth* [2002] found that the size of the predictor domain had a negligible influence on downscaled daily T_{SFC} in six nations in central and western Europe under contemporary conditions, *Benestad* [2001] found that the choice of predictor domain impacted climate change trend estimates.

[26] As there is no optimal a priori predictor domain for any region or downscaled climate variable, we addressed these concerns by employing the same downscaling methodology using three vastly different scales of predictor domain and compared their performance over the study area as well as over the NYC area only. The three predictor scales were (1) continental: the majority of the North American continent and the Caribbean; (2) regional: the eastern U.S. study area itself, as in the work of *Murphy* [2000], at approximately the planetary Rossby wavelength scale in longitude and latitude; and (3) the northeast (NE), an area representing the two GCM grid cells covering the NYC metropolitan area, centered on 42 N, 72.5W, and 38N, 72.5W, and the two to their west, upstream under typical conditions. This domain includes the Appalachian mountain ranges and Atlantic Ocean, capturing the orographic and maritime influences that we expected were the primary local forcings for the area. Because of the GCM grid layout, we found this to be the most appropriate local predictor scale, rather than downscaling the NYC area from a single grid cell centered on western Massachusetts or from nine proximate GCM grid cells, an area 40 times larger than the NYC area. Predictor domains are shown in Figure 1.

2.2.3.4. Training/Predictor Model

[27] Archived data from GCM simulations of the SRES scenarios typically begin at 1990, and we set out to work within this common constraint. Since SD performance coterminous with the 1993–1997 MM5 downscaling (*Lynn et al.*, submitted manuscript, 2005) was an essential metric, the period 1990–1996 was used as training, allowing 1997 for cross comparison among the RCM, SD, and observations to highlight the relative skill of the two downscaling approaches. This 7-year calibration interval is at the lower bound of the preferred 10–30 year range for climate downscaling, but comparable in span to the 8-year period examined by *Murphy* [1999] in assessing regional dynamical and statistical downscaling of T_{SFC} . Statistical relationships between the predictand and predictors may vary in time [*Wilby*, 1997], and we reasoned that this training period represents the recent warming of the 1990s, ensuring any calibration bias will be toward similar warming in the future, which could be better training for future relationships in this explicit warming scenario than the extended climatological trend. Figure 2 illustrates that the training period contains anomalously warm as well as cool summers, according to all predictor and observation data sets, and is representative of the range of regional climatology of T_{SFC} . Further justification for using only the

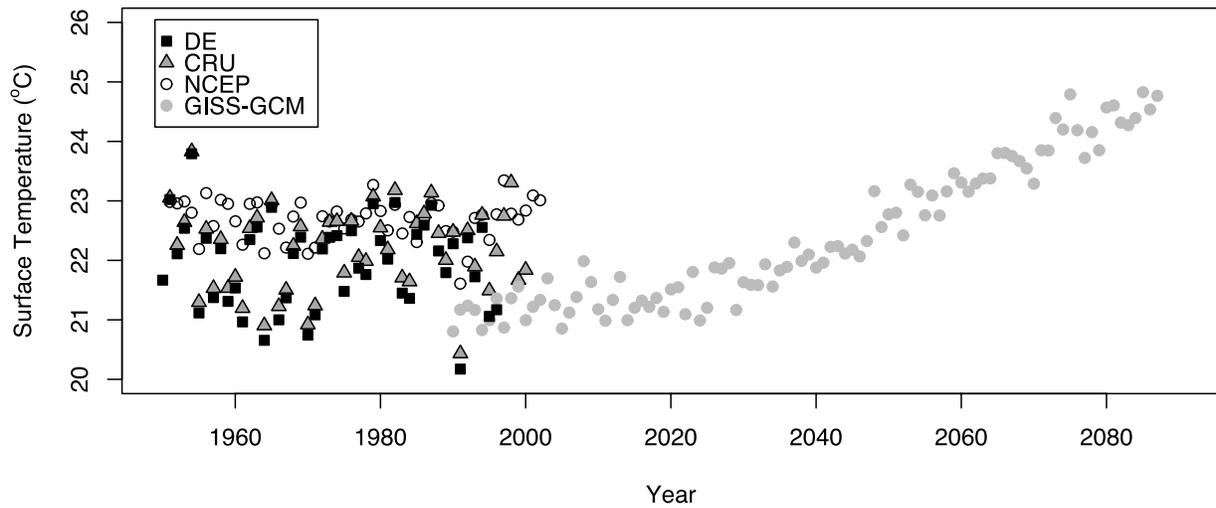


Figure 2. 1951–2087 JJA regional mean T_{SFC} ($^{\circ}C$) from observations, National Centers for Environmental Prediction (NCEP) reanalysis, and GISS-GCM.

most recent years as training came from the consistency between 1990s observed trends and RCM-projected future changes in mean summer precipitation and large-scale circulations in western Europe found by *Pal et al.* [2004]. In fact, we found that NCEP reanalysis 1990–1996 training performed much better than 1951–1996 in predicting summer 1997–1999 temperatures. Mean error for the regional and NE predictor domains increased by more than 100% when the 45-year climatological record was used for training. We performed leave-one-out cross validation [*Michaelsen, 1987*] on the NYC area for 5-year periods in 1990–1996 with the GISS-GCM, and found little variation (less than $0.2^{\circ}C$) in 1997–1999 JJA mean temperatures.

[28] We examined the effects of the choice of GCM and training data on the downscaled result using CGCM2 and GISS-GCM at the Domain predictor scale with both DE and CRU 1990–1996 surface data. Table 1 lists the GCM, surface record, predictor domain, and predictor variables for each SD model.

3. Results and Discussion

[29] Analysis of downscaled scenarios was confined to areas common to the interpolated MM5 results and the DE data set, excluding parts the Atlantic Ocean, Gulf of Mexico, and a few coastal areas. Since MM5 simulates

Table 1. Downscaling Performance Statistics^a

Model	Predictor Domain	Predictor Variables	90% Confidence Level, $\pm 2\ sd, T_{SFC}, ^{\circ}C$		1997 JJA RMS Error in $T_{SFC}, ^{\circ}C$		JJA Mean Warming, 2080s–1990s $T_{SFC}, ^{\circ}C$		Regional Average Correlation Between JJA Downscaled T_{SFC} Field and GCM T_{SFC} Predictor			
			Region	NYC	Region	NYC	Region	NYC	1990s	2020s	2050s	2080s
<i>Dynamical</i>												
GISS-GCM: MM5 MIBR			0.69	0.80	2.99	1.42	3.05	2.38	44	34	34	21
GISS-GCM: MM5 MIGR			0.31	0.32	2.20	1.80	2.22	1.32	97	60	69	51
<i>Statistical</i>												
GISS-GCM: DE	Continental	T_{SFC}	2.13	1.57	(1.26)	0.83	4.85	5.19	70	52	48	42
		TP	2.49	2.33	1.95	1.12	5.96	6.80	82	51	38	46
	Regional	T_{SFC}	(0.80)	(0.52)	1.56	1.48	2.24	2.73	32	52	40	–3
		TP	1.43	1.13	1.86	1.80	(2.73)	3.13	77	57	47	55
NE	T_{SFC}				(0.80)		(2.21)	77	91	88	82	
	TP				3.08		3.25	95	90	96	95	
GISS-GCM: CRU	Regional	T_{SFC}	0.86	(0.53)	1.32	1.26	2.11	2.54	60	69	52	29
CGCM2: DE	Regional	T_{SFC}	2.08	1.80	1.34	0.70	1.75	1.92	42	90	85	74
CGCM2: CRU	Regional	T_{SFC}	1.64	1.61	1.43	1.44	1.72	2.10	83	85	82	49

^aModel column indicates the combination of general circulation model (GCM) and downscaling model; cumulus parameterization for MM5, training data set for SD. Values for SD models best matching *Wilmott and Matsuura (DE)* [2003] observations (RMS Error) or closest in performance to MM5 MIBR in each column are in parentheses. The 90% confidence interval for each model was estimated as twice the standard deviation of 1993–1997 JJA seasonal means. Regional average correlations between downscaled field and GCM predictor were limited to the 95% confidence level. NYC is the three state, 31-county New York City metropolitan area.

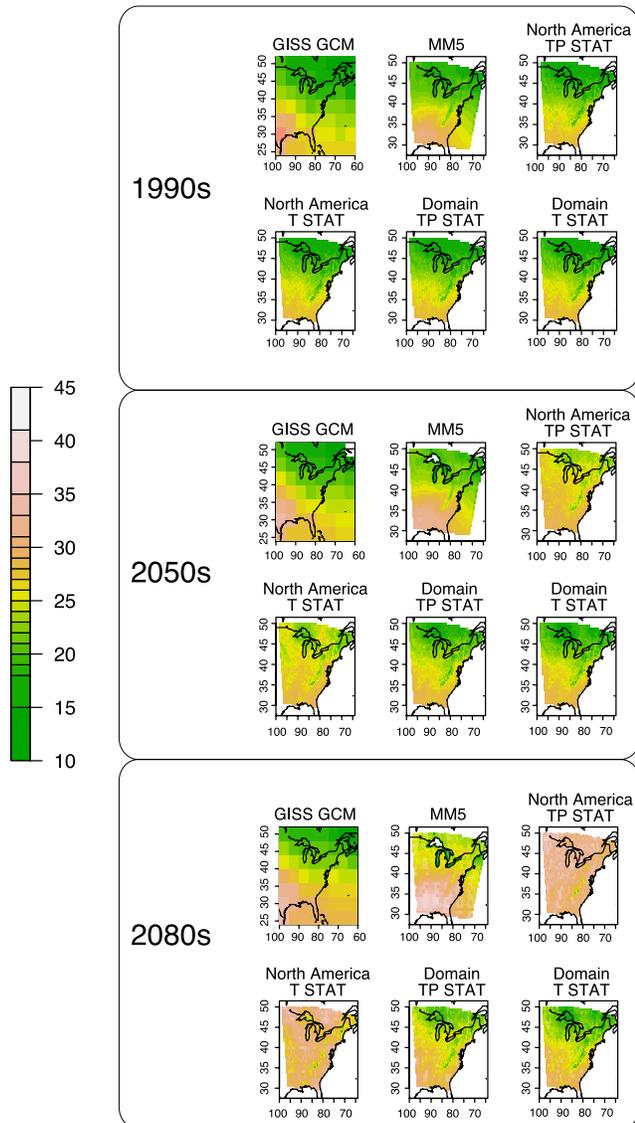


Figure 3. Projected JJA mean T_{SFC} ($^{\circ}\text{C}$).

lake surface temperatures directly, while the DE data set included higher temperatures interpolated from land station observations near the lakes, grid cells representing the large lakes resolved by MM5 were also excluded from analysis. The DE 1993–1999 T_{SFC} record was used as verification for all downscaling models, and skill was assessed through areally averaged root mean squared (RMS) error.

3.1. Skill Under Current Conditions

[30] NCEP Reanalysis was a skillful predictor of local T_{SFC} over eastern North America, and an effective performance benchmark. For summers 1997–1999, RMS error on monthly mean T_{SFC} downscaled from 1990–1996 NCEP Reanalysis was between 0.66 and 0.84 $^{\circ}\text{C}$ across all predictor domains and variables. Neither the T_{SFC} nor the TP predictor variable set was consistently superior across all analyzed scales, although the regional-scale T_{SFC} predictor was marginally best across both the domain and NYC. Predictor domain played a negligible role, as the difference in RMS error for NYC between the continental North America and NE predictor domains was less than 0.04 $^{\circ}\text{C}$.

This confirmed *Huth's* [1999] finding that predictor domain has a negligible effect on downscaled T_{SFC} for present conditions.

[31] Predictions for 1993–1999 showed the extent to which both MM5 and the SD models could debias GCM output, and 1997 provided a contemporaneous performance metric against observations. We found that all SD models tested were more skillful than either of the MM5 models for 1997, the only historical summer of comparison. Downscaling from GISS-GCM was not as skillful as downscaling from 1990–1996 NCEP reanalysis. We estimated the 90% confidence interval for regional RMS error on all models as twice the standard deviation of 1993–1997 JJA seasonal means. Results are shown in Table 1. None of the SD models demonstrated a statistically significant improvement in RMS error over either MM5 simulation. Overall, the regional T_{SFC} and NE T_{SFC} were the only SD models to closely match MM5 seasonal mean sensitivity to GISS-GCM, and all SD models demonstrated statistically significant RMS differences from both MM5 models over the region and NYC.

3.2. Statistical and Dynamical Downscaling Under Climate Change

[32] Regional mean temperature increases from the 1990s to 2080s predicted by the GISS-CGM (3.31 $^{\circ}\text{C}$) and CGCM2 (4.14 $^{\circ}\text{C}$), as well as those in the downscaled results produced by SD and MM5, are within the range of temperature changes predicted by other global climate models for the A2 scenario for the same region [IPCC, 2001]. MM5 MIBR and MIGR models performed comparably to all SD models using the regional and NE predictor domains in predicting less warming over the region than the host GCM. For SD from CGCM2, downscaled regional warming was less than half of what the GCM predicted.

[33] Figure 3 illustrates the progression of projected quasi-decadal JJA mean T_{SFC} , in GISS-GCM and downscaled regional scenarios, in the 1990s, 2050s, and 2080s. While all SD predictor domains yielded reasonably consistent spatial patterns for the 1990s, different predictor domains caused the regression estimates to diverge steadily for future projections. Differences between the T_{SFC} and TP predictor sets are slight. The inclusion of high latitudes up to 60 $^{\circ}\text{N}$ in the continental-scale predictors forecasted warming greatly increased beyond the GCM, especially over northern parts of the region.

[34] Although there is agreement among models that the Arctic warms more than subpolar regions when subject to increasing levels of greenhouse gases in the atmosphere, GISS-GCM showed the greatest polar amplification of climate change among 14 models in the Coupled Model Intercomparison Project 2 [Holland and Bitz, 2003]. While the model's climate sensitivity of 2.7 $^{\circ}\text{C}$ for doubled CO_2 is well within the empirical range of $3 \pm 1^{\circ}\text{C}$ [Hansen et al., 2006], GISS-GCM demonstrated anomalous increases in both poleward ocean heat transport and winter polar cloud cover at doubled CO_2 conditions, which may influence polar amplification. Statistical downscaling was sensitive to this pattern, as DE surface temperature observations in the NE region were highly correlated with the Canadian circumpolar north (>90%) and had low correlations with the rest of the continental predictor domain. This may be why

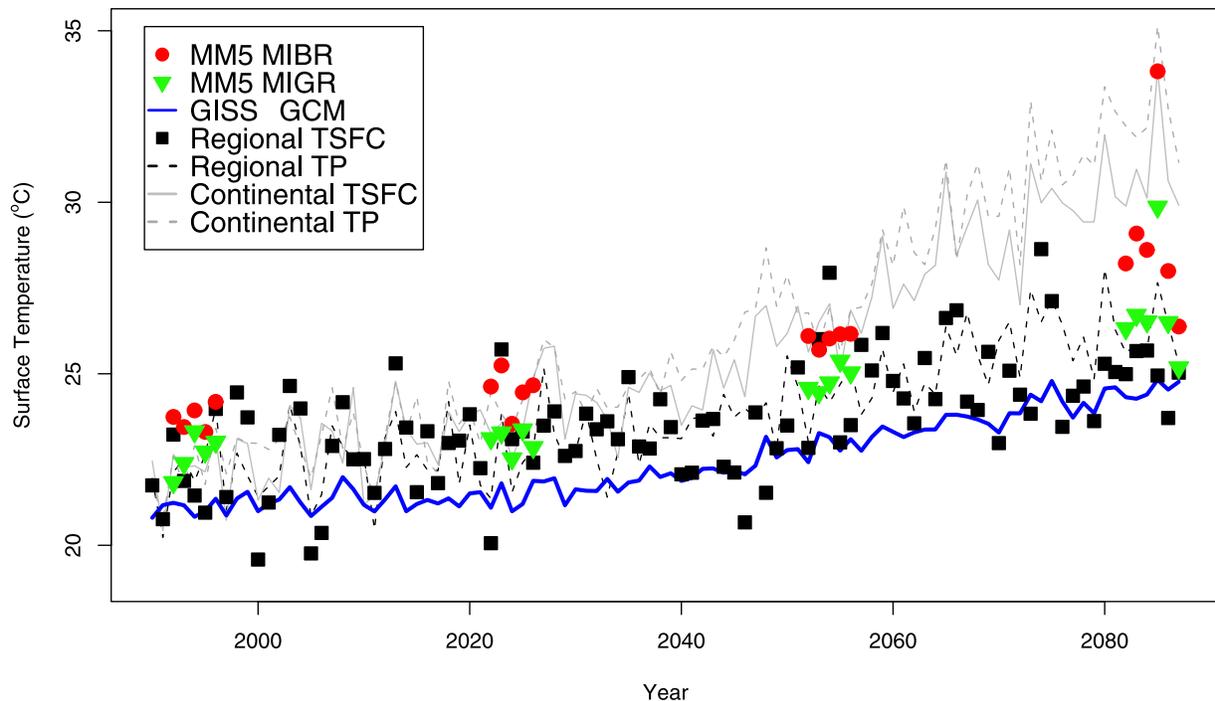


Figure 4. 1993–2087 JJA regional mean T_{SFC} projections ($^{\circ}\text{C}$).

the predictor domain had relatively little influence on SD under observed conditions, but mattered greatly under climate change scenarios.

[35] Increased warming from the continental predictors is also attributable to the fact that leading EOFs of both T_{SFC} and MSLP are multipolar over this large domain. The first EOF of MSLP and second and third EOFs of T_{SFC} are imbalanced dipoles, which *Benestad* [2002] and *Huth* [2004] found to increase warming and yield spurious trend estimates.

[36] Although the MM5 simulations covered limited time intervals, the continuous monthly time series of regional mean T_{SFC} from SD puts them into perspective (Figure 4) and illustrates that, with the exception of the last 4 years of the simulation (2084–2087), all MM5 MIBR and MIGR projections fell within the envelope of the SD simulations specified with our a priori assumptions. This indicates that until 2084, the RCM regional response to the GCM increase in T_{SFC} remained linear, and is sufficiently captured by linear regression SD models. For 2084–2087, the RCM response was highly nonlinear, and covariant across the region and NYC. The consistent difference in T_{SFC} between MM5 MIBR and MIGR, which *Lynn et al.* [2004] found to be a result of differences in diurnal cycles of precipitation, illustrates the magnitude of uncertainty in just one of the many parameterizations employed by RCMs. We find that both MM5 and SD produce much greater interannual variance in regional mean T_{SFC} than the host GCM or the historical records of DE, CRU, and NCEP Reanalysis, shown in Figure 2. This variation, sometimes as much as 5°C between concurrent years, suggests that both methods added a potentially implausible level of interannual noise to the climate change signal in the process of downscaling.

[37] With significant uncertainties in both SD and MM5, we cannot say which projection is “the best.” While

downscaling model performance at present climate conditions does not imply applicability to climate change studies [*Huth*, 2004], similar performance of the two classes of downscaling models both at present and under climate change suggests that for studies focused on defining projection uncertainty, or for studies with computation resource limitations, SD may be appropriate to apply for high-resolution modeling in lieu of an RCM, and to supplement the RCM as a proxy in sensitivity analysis.

[38] RMS differences between the different statistical models and MM5 diverged by the 2020s; the regional difference always represents more than the 100% of the mean warming in MM5 MIGR by the 2080s, and is higher in absolute magnitude and percentage of mean regional warming than the RMS differences found by *Murphy* [2000] between a different set of statistical and dynamical models over northern Europe under a similar warming scenario. Differences between SD and MM5 were highly temperature dependent, and neither method predicted a consistently different downscaled regional mean temperature (Figure 5).

3.3. Choice of Statistical Downscaling Variables

[39] The relative importance of SD model inputs to SD downscaled skill (RMS error) and domain-wide spatial and statistical agreement with MM5 for JJA T_{SFC} over North America in descending order is (1) predictor domain; (2) Surface Record/Predictor Model; (3) Predictor Variables; and (4) Predictor Grid Resolution.

3.3.1. SD Predictor Domain

[40] The size of the large-scale region from which EOFs are calculated critically influenced SD results under climate change, although this finding may be specific to the region, SD model, and predictor set. In this experiment, skill against observations, stationarity, and agreement with MM5 over NYC were highest for the local NE predictor

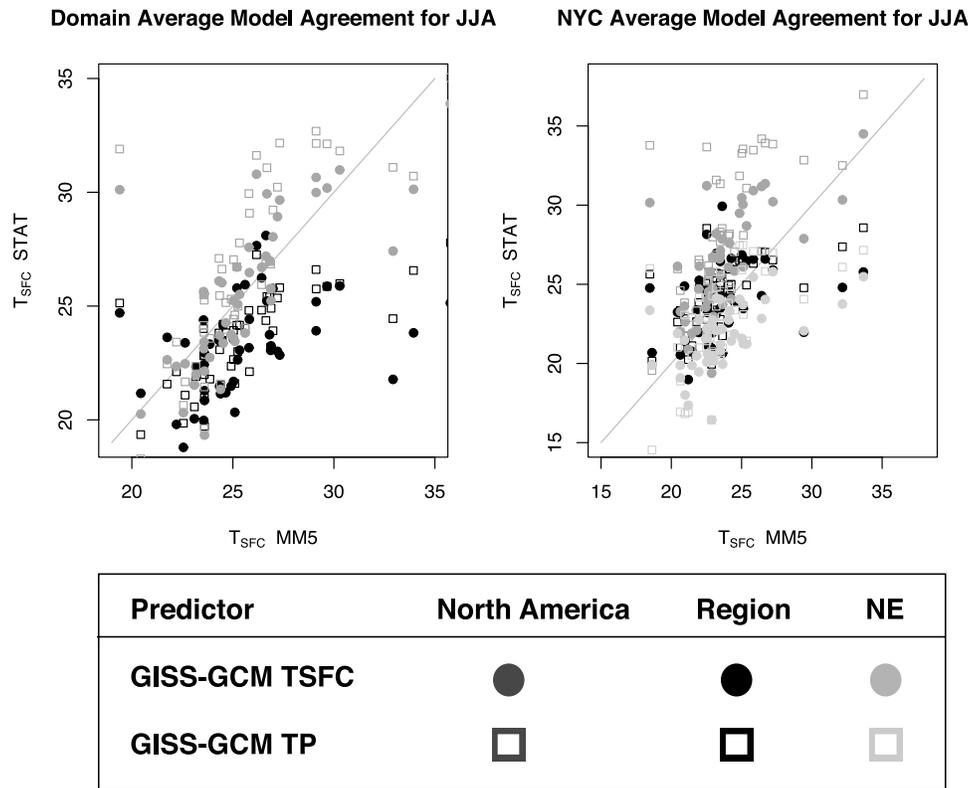


Figure 5. Average model agreement between SD and MM5 (MIBR) over the study domain and NYC, 1997–2087 ($^{\circ}\text{C}$). Contemporaneous predicted monthly mean T_{SFC} from GISS-GCM/MM5 and GISS-GCM statistically downscaled to DE by predictor domain and predictor variables.

domain. Over the entire region (for which only the regional and continental predictor domains could be evaluated), spatial patterns of SD and MM5 were closest for the regional predictor over the entire region and for the NE predictor over NYC. However, predictor region is not necessarily the best predictor domain, as the continental-scale predictor performed better against observations and, when other parameters were changed, was less variable than the regional predictor for regional downscaling.

3.3.2. Input Data for SD (Predictor GCM and Surface Temperature Record)

[41] The choice of surface record was slightly more important than the choice of GCM, though for the options tested here, all input choices were less important than the choice of predictor domain. Differences in regional mean warming between models using DE and CRU as training were of the same magnitude (less than 0.3°C) as the 1951–1999 differences, regardless of the input GCM.

3.3.3. Predictor Variable (T_{SFC} , MSLP, Combination)

[42] The MSLP predictor explained variance in observed temperature, but diminished predictive performance when downscaling from GISS-GCM. Regional mean MSLP in GISS-GCM decreased linearly at a rate of 0.12 mb/decade from 1990–2087, and the influence of this trend in MSLP in the combined predictor (TP) tended to slightly increase warming in downscaling from GISS-GCM.

3.3.4. Predictor Resolution

[43] Across all three predictor domains, differences in downscaled results between coarse ($4^{\circ} \times 5^{\circ}$) and finer

($2.5^{\circ} \times 2.5^{\circ}$) scale predictor fields from NCEP 1990–1996 reanalysis are negligible, with a mean RMS difference of 0.01°C . This apparent insensitivity to resolution is likely due to the absence of important smaller-scale features in both predictor fields. While predictor resolution probably has minimal effect on downscaled T_{SFC} globally, we caution that this result is not necessarily transferable to precipitation and other climate variables of interest for regional downscaling.

3.4. Spatial Relationships as Indicators of Agreement

[44] Stationarity implies that the mean, variance and autocorrelation structure do not change over time. Some statistical relationships are stationary while others vary, and it is uncertain which of these relationships will be important to the downscaling. In particular, the relationship between the time series at each observation station or surface grid point and the time series of a large-scale predictor pattern may remain constant even while the aggregate relationship between a regional map of such surface observations and the large-scale pattern varies. The SD models used here assume a constant local predictor-predictand relationship, but do not incorporate the relationships between the predictor and the regional map of observations. Changes to this regional aggregate relationship may be an indicator of the relative contributions of the strength of the predictor forcing and the codified relationship between large-scale and local features built into the RCM or SD algorithms.

[45] In this study, the regional mean correlation between DE T_{SFC} and NCEP reanalyzed T_{SFC} was 0.92 for each

decade in the period 1950–1996 and varied by less than 1%, indicating stationarity. Table 1 shows that all of the regional correlations at the 95% significance level between downscaled results and associated GCM predictors are variable over the quasi-decades chosen for comparison. In fact, neither MM5 for 1993–1997 nor any of the SD predictions from GISS-GCM for 1997–1999 even maintained the same relationship with GISS-GCM as the DE training.

[46] Spatial scale of the predictors was the most important contributor to stationarity during both training with reanalysis and projection with GCMs. Two SD models with the NE predictor maintained their historical correlation, and one strengthened, but predictor/predictand field relationships in projections from GCMs generally weakened from the 1990s to 2080s, and were weaker than historical patterns from DE and NCEP Reanalysis. However, Figure 3 shows that MM5 MIBR and SD with regional-scale predictors each retained the spatial patterns they developed in downscaling the 1990s from GISS GCM. This suggests that for these particular domain sizes and locations, the strength of the local training relationship in SD and model formulation in MM5 was more important in the downscaling model than the input GCM's spatial variability, so that the GCM serves more as a mean warming trend upon which to overlay statistical momentum or dynamical calculations than as the driving force behind spatial patterns of variance from predictor EOFs. This similar response to GCM input, despite different downscaling methods, may be a reason for apparent agreement between the downscaled scenarios. However, it is possible that the weakening of predictor-predictand relationships in projections may come from the fact that predictor temperatures in future climates fall beyond the interval on which the SD models were trained.

4. Summary and Conclusion

[47] We employed regression techniques and a dynamical regional climate model to downscale June, July, and August monthly mean surface temperature over eastern North America under greenhouse gas-driven climate change simulation by the NASA GISS GCM. We found that the two methods and all permutations of regression parameters generally exhibited comparable skill at simulating historical observations, although spatial patterns in temperature across the region differed. While the two methods projected similar regional mean warming over the period 2000–2087, they again developed vastly different spatial patterns of temperature across the region, which diverged greatly from their historical differences. We found that for statistical downscaling with multiple linear regressions, predictor domain size was a negligible factor for current conditions, but had a much greater influence on future surface temperature change than any other factor, including the source of predictor and training data sets. We found that employing a smaller predictor domain maintained stationarity and led to better agreement with the RCM, while continental-scale predictors simulated much greater warming than regional and local predictors.

[48] These results illustrate the broad range of potentially plausible local scenarios that can be generated from a single GCM run using the same methodology, and highlight the importance of evaluating each variable in the process of

statistical downscaling. The location and size of the predictor domain demand special attention, since this variable is responsible for the most variation in downscaled results, is inherently specific to each application, and is currently chosen in a more arbitrary manner than other factors. A combination of expert knowledge, objective analysis, and sensitivity testing is necessary to reduce uncertainty in this area.

[49] Downscaled projections provide an estimate of specific, localized response to climate change that raw GCM output cannot yet provide, but substitute the known limitations of GCMs, such as inadequate spatial resolution, with a different set of local uncertainties. The unique value of local projections in integrated assessment is contingent upon the bias and noise added in downscaling, as well as the transparency of the downscaling process. This study highlights the advantages and relative ease for integrated assessments to take into account multiple sources of information, at all available scales, in order to quantify uncertainty and reduce the assessment's reliance on a few linkages and arbitrary settings. Regional surface temperature scenarios, and the assessments to which they contribute, can be improved by assessing multiple downscaling methods for the same GCM, ranging from state-of-the-science dynamical models to relatively simple statistical predictions; and by using multiple downscaling methods with an ensemble of GCMs and surface data sets to yield the most plausible projections and develop a comprehensive understanding of the physical and mathematical reasons behind apparent agreement among distinct regional downscaling techniques. Conversely, the divergence in projections in this study shows how inappropriate it may be to pick just one SD analysis for comparison with RCM results, and especially when using SD as a standalone tool.

[50] This qualitative and quantitative comparison and sensitivity analysis is a step toward the development of a comprehensive methodology for estimating the uncertainty added in the process of downscaling climate change scenarios. While this experiment focused on a single variable, temperature, that is well-suited to linear regression downscaling, future studies of other statistical downscaling methods and climate variables with greater spatial and temporal inhomogeneity will illustrate the transferability of these results to the generalized problem of downscaling climate change.

[51] **Acknowledgments.** The authors thank Daniel J. Vimont for advice on EOF analysis and combination, Christian Hogrefe for comments, and Patrick L. Kinney for coordination of the NYCHP. The reviews of three anonymous referees led to considerable improvements to the original manuscript. Results were obtained using clim.pact by Benestad (the Norwegian Meteorological Institute and the Norwegian Research Council's RegClim programme). NCEP Reanalysis and University of Delaware Air Temperature and Precipitation data provided by the NOAA-CIRES Climate Diagnostics Center, Boulder, Colorado, United States, from their Web site at <http://www.cdc.noaa.gov>.

References

- Benestad, R. E. (2001), A comparison between two empirical downscaling strategies, *Int. J. Climatol.*, 21, 1645–1668, doi:10.1002/joc.703.
- Benestad, R. E. (2002), Empirically downscaled temperature scenarios for northern Europe based on a multi-model ensemble, *Clim. Res.*, 21(2), 105–125.
- Benestad, R. E. (2003), What can present climate models tell us about climate change?, *Clim. Change*, 59(3), 311–331, doi:10.1023/A:1024876611259.

- Benestad, R. E. (2004a), Empirical-statistical downscaling in climate modeling, *Eos Trans. AGU*, 85(42), 417, doi:10.1029/2004EO420002.
- Benestad, R. E. (2004b), Tentative probabilistic temperature scenarios for northern Europe, *Tellus, Ser. A.*, 56(2), 89, doi:10.1111/j.1600-0870.2004.00039.x.
- Benestad, R. E., E. J. Førland, and I. Hanssen-Bauer (2002), Empirically downscaled temperature scenarios for Svalbard, *Atmos. Sci. Lett.*, 3(2–4), 71–93, doi:10.1006/asle.2002.0050.
- Dehn, M., and J. Buma (1999), Modelling future landslide activity based on general circulation models, *Geomorphology*, 30, 175–187, doi:10.1016/S0169-555X(99)00053-7.
- Diaz-Nieto, J., and R. L. Wilby (2005), A comparison of statistical downscaling and climate change factor methods: Impacts on low flows in the River Thames, United Kingdom, *Clim. Change*, 69(2–3), 245–268, doi:10.1007/s10584-005-1157-6.
- Flato, G. M., and G. J. Boer (2001), Warming asymmetry in climate change simulations, *Geophys. Res. Lett.*, 28, 195–198, doi:10.1029/2000GL012121.
- Hansen, J., et al. (2006), Dangerous human-made interference with climate: A GISS modelE study, *Atmos. Chem. Phys. Discuss.*, 6, 12,549–12,610.
- Hanssen-Bauer, I., E. J. Førland, J. E. Haugen, and O. E. Tveito (2003), Temperature and precipitation scenarios for Norway: Comparison of results from dynamical and empirical downscaling, *Clim. Res.*, 25, 15–27.
- Hogrefe, C., B. Lynn, K. Civerolo, J.-Y. Ku, J. Rosenthal, C. Rosenzweig, R. Goldberg, S. Gaffin, K. Knowlton, and P. L. Kinney (2004), Simulating changes in regional air pollution over the eastern United States due to changes in global and regional climate and emissions, *J. Geophys. Res.*, 109, D22301, doi:10.1029/2004JD004690.
- Holland, M. M., and C. M. Bitz (2003), Polar amplification of climate change in coupled models, *Clim. Dyn.*, 21, 221–232, doi:10.1007/s00382-003-0332-6.
- Huth, R. (1999), Statistical downscaling in central Europe: Evaluation of methods and potential predictors, *Clim. Res.*, 13, 91–101.
- Huth, R. (2002), Statistical downscaling of daily temperature in Central Europe, *J. Clim.*, 15, 1731–1742, doi:10.1175/1520-0442(2002)015<1731:SDODTI>2.0.CO;2.
- Huth, R. (2004), Sensitivity of local daily temperature change estimates to the selection of downscaling models and predictors, *J. Clim.*, 17, 640–652, doi:10.1175/1520-0442(2004)017<0640:SOLDTC>2.0.CO;2.
- Intergovernmental Panel on Climate Change (IPCC) (2000), *Special Report on Emissions Scenarios*, edited by N. Nacenov and R. Swart, 612 pp., Cambridge Univ. Press, New York.
- Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate Change 2001: The Scientific Basis*, edited by J. T. Houghton et al., 944 pp., Cambridge Univ. Press, New York.
- Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, 77, 437–471, doi:10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.
- Karl, T. R., C. N. Williams Jr., F. T. Quinlan, and T. A. Boden (1990a), *United States Historical Climatology Network (HCN) Serial Temperature and Precipitation Data*, *Environ. Sci. Div. Publ. 3404*, 389 pp., Carbon Dioxide Inf. and Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Karl, T. R., W. C. Wang, M. E. Schlesinger, R. W. Knight, and D. Portman (1990b), A method of relating general circulation model simulated climate to the observed local climate. part I: Seasonal statistics, *J. Clim.*, 3(10), 1053–1079, doi:10.1175/1520-0442(1990)003<1053:AMORGC>2.0.CO;2.
- Kidson, J. W., and C. S. Thompson (1998), A comparison of statistical and model-based downscaling techniques for estimating local climate variations, *J. Clim.*, 11(4), 735–753, doi:10.1175/1520-0442(1998)011<0735:ACOSAM>2.0.CO;2.
- Knowlton, K., J. E. Rosenthal, C. Hogrefe, B. H. Lynn, S. Gaffin, R. Goldberg, C. Rosenzweig, K. Civerolo, J. Y. Ku, and P. L. Kinney (2004), Assessing ozone-related health impacts under a changing climate, *Environ. Health Perspect.*, 112(15), 1557–1563, doi:10.1289/ehp.7163.
- Lynn, B. H., L. Druyan, C. Hogrefe, J. Dudhia, C. Rosenzweig, R. Goldberg, D. Rind, R. Healy, J. Rosenthal, and P. Kinney (2004), Sensitivity of present and future surface temperatures to precipitation characteristics, *Clim. Res.*, 28(1), 53–65.
- Mearns, L. O., I. Bogardi, F. Giorgi, I. Matyasovszky, and M. Palecki (1999), Comparison of climate change scenarios generated from regional climate model experiments and statistical downscaling, *J. Geophys. Res.*, 104(D6), 6603–6622, doi:10.1029/1998JD200042.
- Michaelsen, J. (1987), Cross-validation in statistical climate forecast models, *J. Appl. Meteorol.*, 26, 1589–1600, doi:10.1175/1520-0450(1987)026<1589:CVISCF>2.0.CO;2.
- Mitchell, T. D., and P. D. Jones (2004), An improved method of constructing a database of monthly climate observations and associated high-resolution grids, *Int. J. Climatol.*, 25(6), 693–712, doi:10.1002/joc.1181.
- Murphy, J. M. (1999), An evaluation of statistical and dynamical techniques for downscaling local climate, *J. Clim.*, 12, 2256–2284, doi:10.1175/1520-0442(1999)012<2256:AEOSAD>2.0.CO;2.
- Murphy, J. M. (2000), Predictions of climate change over Europe using statistical and dynamical downscaling techniques, *Int. J. Climatol.*, 20(6), 489–501, doi:10.1002/(SICI)1097-0088(200004)20:5<489::AID-JOC484>3.0.CO;2-6.
- New, M., M. Hulme, and P. D. Jones (2000), Representing twentieth century space-time climate variability. part II: Development of 1901–96 monthly grids of terrestrial surface climate, *J. Clim.*, 13, 2217–2238, doi:10.1175/1520-0442(2000)013<2217:RTCSTC>2.0.CO;2.
- Oshima, N., H. Kato, and S. Kadokura (2002), An application of statistical downscaling to estimate surface air temperature in Japan, *J. Geophys. Res.*, 107(D10), 4095, doi:10.1029/2001JD000762.
- Pal, J. S., F. Giorgi, and X. Q. Bi (2004), Consistency of recent European summer precipitation trends and extremes with future regional climate projections, *Geophys. Res. Lett.*, 31, L13202, doi:10.1029/2004GL019836.
- Portman, D. A., W.-C. Wang, and T. R. Karl (1992), Comparison of general circulation model and observed regional climates: Daily and seasonal variability, *J. Clim.*, 5, 343–353, doi:10.1175/1520-0442(1992)005<0343:COGCMA>2.0.CO;2.
- Preisendorfer, R. W. (1988), *Principal Component Analysis in Meteorology and Oceanography*, Elsevier, New York.
- Rind, D., R. Goldberg, and R. Ruedy (1989), Change in climate variability in the 21st century, *Clim. Change*, 14, 5–37, doi:10.1007/BF00140173.
- Robinson, P. J., A. N. Samel, and G. Madden (1993), Comparison of modelled and observed climate for impacts assessments, *Theor. Appl. Climatol.*, 48, 75–87, doi:10.1007/BF00864915.
- Russell, G. L., J. R. Miller, and D. Rind (1995), A coupled atmosphere-ocean model for transient climate change studies, *Atmos. Ocean*, 33, 683–730, doi:0705-5900/95/0000-0359\$01.25/0.
- Solecki, W. D., and C. Oliveri (2004), Downscaling climate change scenarios in an urban land use change model, *J. Environ. Manage.*, 72, 105–115, doi:10.1016/j.jenvman.2004.03.014.
- Thorne, P. W., D. E. Parker, J. R. Christy, and C. A. Mears (2005), Uncertainties in climate trends: Lessons from upper-air temperature records, *Bull. Am. Meteorol. Soc.*, 86, 1437–1442, doi:10.1175/Bull.Am.Meteorol.Soc.-86-10-1437.
- Vose, R. S., and M. J. Menne (2004), A method to determine station density requirements for climate observing networks, *J. Clim.*, 17, 2961–2971, doi:10.1175/1520-0442(2004)017<2961:AMTDSO>2.0.CO;2.
- Vose, R. S., R. L. Schmoyer, P. M. Steurer, T. C. Peterson, R. Heim, T. R. Karl, and J. Eischeid (1992), *The Global Historical Climatology Network: Long-Term Monthly Temperature, Precipitation, Sea Level Pressure, and Station Pressure Data*, ORNL/CDIAC-53, NDP-041, Carbon Dioxide Inf. Anal. Cent., Oak Ridge Natl. Lab., Oak Ridge, Tenn.
- Wilby, R. L. (1997), Non-stationarity in daily precipitation series: Implications for GCM down-scaling using atmospheric circulation indices, *Int. J. Climatol.*, 17, 439–454, doi:10.1002/(SICI)1097-0088(19970330)17:4<439::AID-JOC145>3.0.CO;2-U.
- Wilby, R. L., S. P. Charles, E. Zorita, B. Timbal, P. Whetton, and L. O. Mearns (2004), Guidelines for use of climate scenarios developed from statistical downscaling methods, IPCC Data Distribution Centre report, 27 pp., Univ. of East Anglia, Norwich, U.K.
- Wilmott, C. J., and K. Matsuura (DE) (2003), *Univ. of Delaware Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950–1999)*, version 1.02, NOAA-CIRES Clim. Diagn. Cent., Boulder, Colo.
- Wilmott, C. J., and S. M. Robeson (1995), Climatologically aided interpolation of terrestrial air temperature, *Int. J. Climatol.*, 15, 221–229.
- Wilmott, C. J., C. M. Rowe, and W. D. Philpot (1985), Small-scale climate maps: A sensitivity analysis of some common assumptions associated with grid-point interpolation and contouring, *Am. Cartogr.*, 12, 5–16.

R. Goldberg, NASA Goddard Institute for Space Sciences, 2880 Broadway, New York, NY 10025, USA.

T. Holloway, Center for Sustainability and the Global Environment, University of Wisconsin-Madison, 1710 University Avenue, Room 201A, Madison, WI 53726, USA.

B. Lynn, Center for Climate Systems Research, Columbia University, Armstrong Hall, 2880 Broadway, New York, NY 10025, USA.

S. Spak, Center for Sustainability and the Global Environment, University of Wisconsin-Madison, 1710 University Avenue, Room 207, Madison, WI 53726, USA.