Applied Energy 209 (2018) 516–528

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Simulated building energy demand biases resulting from the use of representative weather stations

Casey D. Burleyson*, Nathalie Voisin, Z. Todd Taylor, Yulong Xie, Ian Kraucunas

Pacific Northwest National Laboratory, Richland, WA, United States

HIGHLIGHTS

- Building energy model biases in the WECC depend on the location/number of representative cities.
- Using 1 station per IECC climate zone results in a mean absolute summer temperature bias of 4.0 °C.
- Using 1 station per IECC zone can lead to a 20–40% overestimate of peak loads during summer/winter.
- Using all available stations reduces the mean absolute load bias by a factor of 2.5.
- Using 4 stations per IECC zone reduces both temperature/load biases and computational burden.

ARTICLE INFO

Keywords:
Building energy demand
Representative cities
BEND
EnergyPlus
Weather

ABSTRACT

Numerical building models are typically forced with weather data from a limited number of "representative cities" or weather stations representing different climate regions. The use of representative weather stations reduces computational costs, but often fails to capture spatial heterogeneity in weather that may be important for simulations aimed at understanding how building stocks respond to a changing climate. We quantify the potential reduction in temperature and load biases from using an increasing number of weather stations over the western U.S. Our novel approach is based on deriving temperature and load time series using incrementally more weather stations, ranging from 8 to roughly 150, to evaluate the ability to capture weather patterns across different seasons. Using 8 stations across the western U.S., one from each IECC climate zone, results in an average absolute summertime temperature bias of ~4.0 °C with respect to a high-resolution gridded dataset. The mean absolute bias drops to ~1.5 °C using all available weather stations. Temperature biases of this magnitude could translate to absolute summertime mean simulated load biases as high as 13.5%. Increasing the size of the domain over which biases are calculated reduces their magnitude as positive and negative biases may cancel out. Using 8 representative weather stations can lead to a 20–40% bias of peak building loads during both summer and winter, a significant error for capacity expansion planners who may use these types of simulations. Using weather stations close to population centers reduces both mean and peak load biases. This approach could be used by others designing aggregate building simulations to understand the sensitivity to their choice of weather stations used to drive the models.

1. Introduction

A large direct societal cost of climate change could come from the need to build an energy system capable of meeting spikes in energy demand under heat wave conditions that are changing in frequency in response to warmer temperatures [1–7]. It is important to energy system planners and other stakeholders that we understand the detailed regional, seasonal, and diurnal characteristics of building energy demand under current and future climate conditions. There are efforts underway across multiple research disciplines to understand and quantify this potential impact. Much of the literature on this topic is based on empirical studies which often utilize static representations of building stock and therefore do not capture dynamic responses to extreme events or evolving building technologies [8–15]. Climate and weather impacts on individual buildings have also been explored using numerical building models such as eQUEST [http://www.doe2.com/equest/], TRACE 700 [http://www.trane.com/commercial/north-america/us/en/products-systems/design-and-analysis-tools/analysis-tools/trace-700.html], and the Department of Energy's (DOE) EnergyPlus [16]. A typical approach with these models is to use weather

* Corresponding author at: Pacific Northwest National Laboratory, PO Box 999/MS K9-24, Richland, WA 99352, United States.
E-mail address: casey.burleyson@pnnl.gov (C.D. Burleyson).

http://dx.doi.org/10.1016/j.apenergy.2017.08.244
Received 24 April 2017; Received in revised form 23 August 2017; Accepted 24 August 2017
Available online 06 November 2017

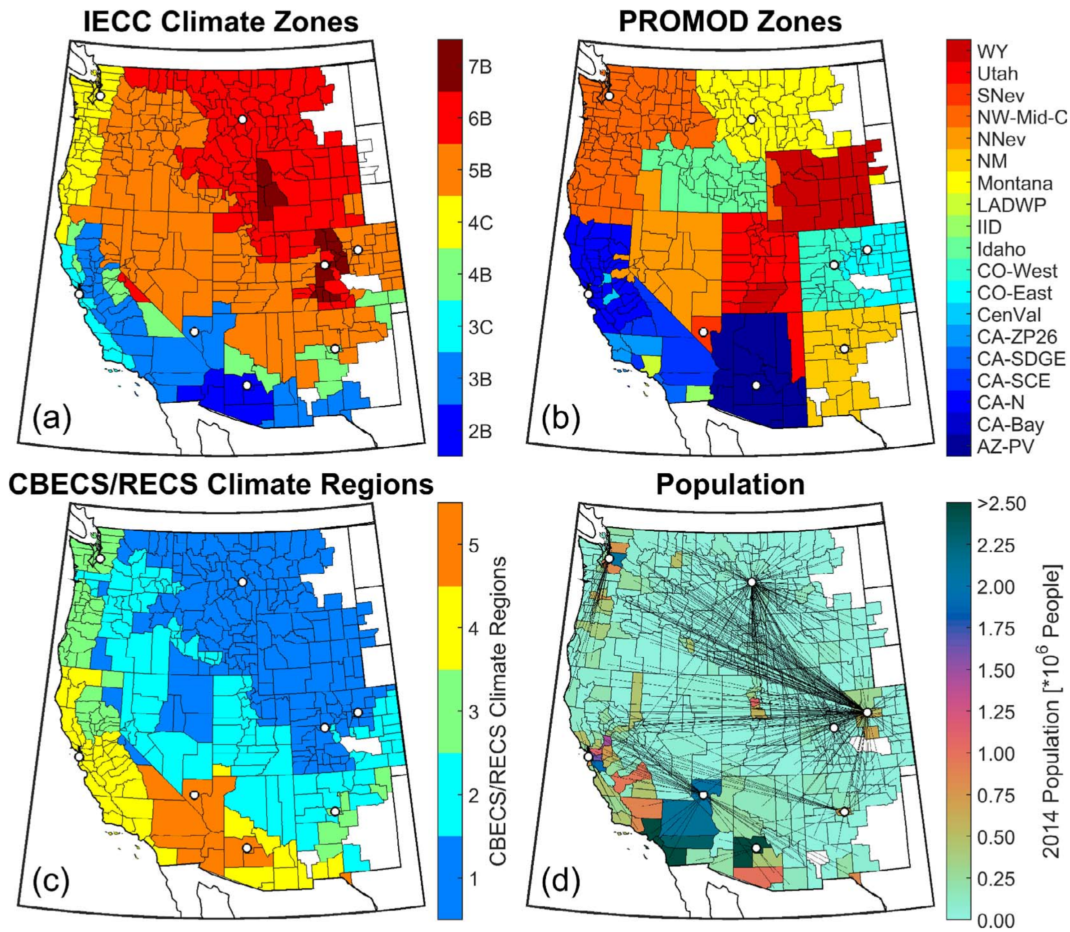

Fig. 1. (a) The 8 IECC climate zones that are present in the Western Electricity Coordinating Council (WECC); (b) the 19 PROMOD zones used in the calibration of our BEND simulation; (c) the 5 climate regions in the CBECS/RECS building databases; and (d) the population within each county in 2014. In all panels the white dots indicate the original 8 Class A representative weather stations, one for each climate zone, that are used to force BEND. Panel (d) includes a mapping from each county to its associated Class A representative weather station. The two blank counties in Colorado and New Mexico are not part of the WECC.

data at a single location as input to the model and then analyze the building energy demand response to changing weather conditions at that location [5,17–19].

One advancement facilitating new insights in this space is that composite models of building energy demand using tens of thousands of individual building simulations, based on different combinations of building technologies and characteristics or climate scenarios, are now possible due to increased access to high-performance computing. In these composite models, individual building simulations under current or future climate scenarios are aggregated on city [20,21], state [7,11], or national scales [5,22]. The composite model approach delivers the detailed, physically-based aspects derived from modeling individual buildings as well as information on the aggregate effect on larger spatial or longer temporal scales. Two key uses of these aggregate building energy models are to quantitatively evaluate the impact of proposed energy efficiency measures for energy efficient building designs or to assess the impact of climate or economic changes on aggregate building energy demand.

The DOE's Pacific Northwest National Laboratory (PNNL) has developed the aggregate Building ENergy Demand (BEND) model to explore the interaction between weather conditions and building energy demand (alternatively referred to as building loads in this paper). The BEND model was first described and utilized in Dirks et al. [7], which explored climate change impacts on peak and annual building energy demand in the Eastern Interconnection (EIC). At its root, the BEND model is a mechanism to aggregate EnergyPlus simulations for a representative sample of building types in a given geographical area. BEND uses the Commercial Buildings Energy Consumption Survey [CBECS; https://www.eia.gov/consumption/commercial/about.php] and Residential Energy Consumption Survey [RECS; https://www.eia.gov/consumption/residential/] datasets to generate a population of buildings that span the range of residential and commercial building sizes and vintages in a climate similar area of a census region. Energy usage in each building in the sample population is then simulated using forcing from Energy Plus Weather files that contain an hourly time series of observed or predicted meteorological variables (e.g., temperature, humidity, solar radiation). These forcing files are the primary mechanism by which the model responds to changes in weather and they represent the physical linkage between climate and building energy demand. The selection of which weather datasets or locations are used to force the model is a key component of the simulation design.

In an ideal scenario, simulations using BEND or other building energy models would be run using weather information that exactly corresponds to the physical location of each simulated building. However, computational and data constraints make this impractical and unwarranted: The CBECS and RECS databases only provide a statistical representation of the nation's building stock, rather than a complete geospatially explicit inventory, and there are only ~1000 surface weather stations in the U.S. with sufficient data density and quality to drive the underlying EnergyPlus models [23]. The key questions to address when developing an aggregated building energy demand model are thus (1) how many representative buildings are needed to adequately represent the diversity of building types and corresponding energy demand profiles, and (2) how many weather

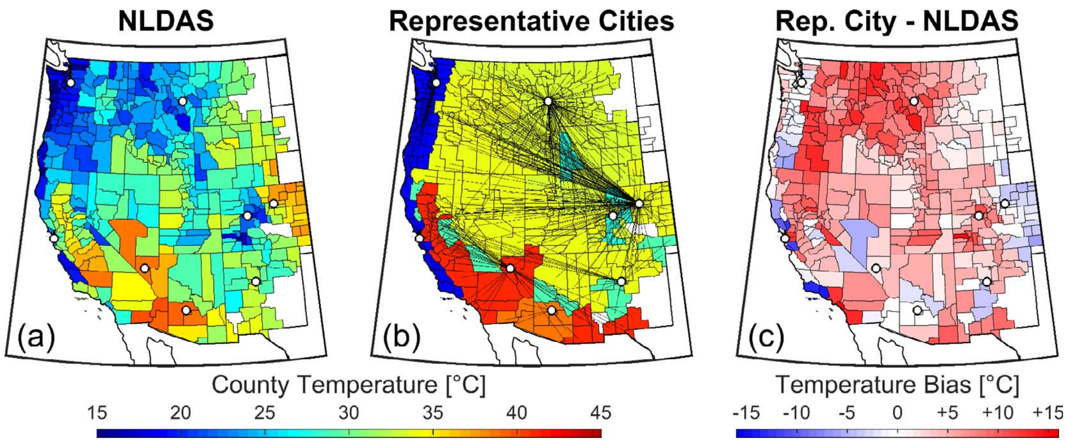

Fig. 2. An example of the bias in temperature that results from the use of 8 representative weather stations: (a) the surface temperature from the NLDAS reanalysis mapped to each county in the WECC on 5-July 2008 at 2000 UTC; (b) the temperature from the 8 representative stations at the same time mapped to all of the counties within a given IECC climate zone; (c) the difference between (a) and (b).

stations are needed to characterize the environmental factors that influence energy demand across that population of buildings? This paper focuses on the latter question; a forthcoming publication will address the more complex question of calibrating the BEND model for a given region, including the balance between computational efficiency and number of different building types needed to accurately represent energy demand.

Previous studies have used a number of different techniques to select representative weather sites to force building energy demand models. For example, in EnergyPlus simulations meant to provide a baseline for energy efficiency research in commercial buildings [24], the team used weather data from 15 cities to capture weather variability across the U.S. These cities, one from each of the International Energy Conservation Code (IECC) climate zones, were subjectively selected from the available sites to best capture the mean weather in each IECC climate zone. The 15 IECC climate zones in the U.S. were designed to identify areas with similar climates while retaining an association with important political boundaries (Fig. 1a). The IECC climate zones are similar to, but do not necessarily map directly to, the 5 climate regions used to distinguish building characteristics in the CBECS/RECS datasets (Fig. 1c). Hong et al. [25] also used one representative city from each of the IECC climate zones to examine errors associated with using Typical Meteorological Years (TMY) instead of Actual Meteorological Years (AMY) to force a set of building models over the U.S. Shen [20] used future climate projections for 4 U.S. cities to quantify changes in simulated residential building energy demand. Similarly, Degelman [26] used weather data from 6 cities to represent the climate conditions impacting a majority of the world's population, Spandagos [15] used projected weather conditions in 3 large Asian cities, Crawley [27] used weather data from 25 sites across different Köppen climate regions, Wan et al. [28] used weather data from 5 cities to represent different climates in China, and Dirks et al. [7] used 100 representative weather stations in the EIC. For work analyzing residential building codes and standards, PNNL developed a set of 120 representative cities, identified by overlaying the 15 IECC climate zones onto state boundaries so that there is coverage of all code zones in every state [29].

The issue of selecting appropriately representative weather stations is relevant to multiple other fields as well. For example, wind energy researchers trying to predict the spatial variability of wind resources often use statistical methods to go from a limited number of point measurements to spatially-distributed estimates of available energy [30]. Air pollution researchers try to identify the best weather stations for understanding pollution observations at surface monitoring sites [31]. Weather observed at locations close to agricultural fields can be used to predict climate change impacts on crop yields [32,33]. Representative weather stations have even been used in ecological studies such as predicting climate impacts on the hatch rate of leatherback turtles based on observed weather conditions close to nesting beaches [34]. All of these examples illustrate that the selection of weather stations for forcing building energy demand simulations, among multiple other applications, can be influenced by a number of factors including jurisdictional boundaries, population distributions, data availability, and computational burden.

The use of representative buildings and weather stations necessarily introduces biases, or differences between simulated and actual energy demand [19]. These differences may be amplified in the future if climate change impacts different locations in different ways, effectively increasing the spatial variability and decreasing the "representativeness" of stations selected when designing the simulation. The use of representative weather stations also creates challenges when it comes to calibrating aggregate building models [35]. Load biases can be introduced in areas where the actual temperature is drastically different than the temperature at the assigned representative weather station. For example, San Diego, Los Angeles, and Las Vegas are all in IECC climate zone 3B, which may utilize weather observations at any of the three cities as representative data. On any given day the actual temperature can be significantly different among these locations (hereafter temperature biases). An example of these differences for one summer day is shown in Fig. 2.

Using biased temperatures to force the building models potentially leads to simulated loads that are significantly different from the observed values (hereafter load biases). Given that many of the analyses made possible by aggregate building modeling revolve around understanding the sensitivity of the current or future energy system to changes in meteorology, it is critical to understand the potential magnitude of these biases. To date there are no evaluations of the sensitivity of aggregate building simulations to the number or location of weather sites used to force the models. The purpose of this paper is to characterize and quantify the magnitude of the potential biases that result from using a limited number of representative weather stations and to develop a protocol for selecting an optimal number of stations for a particular analysis. This is the first step of a robust error characterization for the BEND model.

While our primary objective can be considered mechanistic (i.e., what are the magnitudes of potential temperature and load biases and how many weather stations are necessary to capture temperature patterns across a given region?), there are a pair of additional science questions that can be addressed along the way that will be of interest to a broader audience:

(1) How does the spatial scale over which temperature and load biases are computed impact the magnitude of the biases?

Table 1
A summary of the total number of stations used and the number of stations per IECC climate zone in each of our 6 Classes of data.

	Class A	Class B	Class C	Class D	Class E	Class F
Total stations	8	8	16	32	64	~150
Stations per IECC Climate Zone	1	1	2	4	8	All available
Stations/approach	Phoenix, Las Vegas, San Francisco, Albuquerque, Seattle, Denver, Helena, Aspen	Random Selection	Class B + Random Selection	Class C + Random Selection	Class D + Random Selection	All available

(2) Can the spatial density of weather sites used to force an aggregate model tell us anything about the requisite spatial scales that are necessary to capture heat wave impacts on building energy demand?

2. Data and methods

Our current study builds on ongoing work by our team trying to calibrate BEND in the Western Electricity Coordinating Council (WECC). In this WECC study, we use weather data from 8 representative weather stations, one from each of the IECC climate zones present in the WECC (Fig. 1a). The 8 stations and their associated climate zone were Phoenix, AZ (2B), Las Vegas, NV (3B), San Francisco, CA (3C), Albuquerque, NM (4B), Seattle, WA (4C), Denver, CO (5B), Helena, MT (6B), and Aspen, CO (7B). We may refer to the weather stations as representative cities, although not every station lies within a city. We used those 8 stations to force the EnergyPlus simulations in BEND. The choice to use only 8 stations was based largely on computational constraints.

The aggregate building simulations in this BEND run using 8 representative cities were calibrated using load profiles from the 19 default spatial zones spanning the WECC in the PROMOD production cost model for the year 2010 (Fig. 1b). PROMOD is a zonal energy dispatch model used to balance energy supply and demand for a given region. To avoid confusion, we will use "climate zones" to reference the IECC climate zones and "PROMOD zones" to reference the PROMOD zones. Our calibration method involves scaling and bias correcting the BEND output to minimize the difference between the simulated loads and those from PROMOD. This bias corrected simulation is the base case scenario for BEND upon which our analysis is built.

The meteorological data we use was collected at Automated Surface Observing Systems (ASOS) sites across the country and was quality-controlled and packaged as the National Solar Radiation Database (NSRDB; [23]). We ran the NSRDB dataset through standard processing scripts to generate the Energy Plus Weather files used to drive the model. We know *a priori* that 8 weather stations are insufficient to capture all details of the spatially heterogeneous meteorological fields (e.g., Fig. 2). Choosing the appropriate number of weather stations is a balance between computational expense and simulation fidelity. Based on a hypothetical building sample of 10,000 unique buildings, using 8 weather stations requires a total of roughly 220,000 EnergyPlus runs whereas 1,620,000 runs would be required if we utilized all available weather stations in the WECC. An expansion of this calculation is given in the Supplemental Material Table S1. The computational savings are obvious, but are they worth the degradation in simulation quality? The fundamental design of this analysis is to quantify how much information is gained if we were to use more than 8 stations to represent weather across the WECC. Rather than rerun all of the possible building models using all of the possible weather stations (which would be very computationally expensive), this study employs an analytical approach whereby we calculate how much the temperature and simulated building loads would change if we used incrementally increasing numbers of weather stations.

We begin by mapping temperatures from the original 8 representative weather stations, 1 from each climate zone, to the county level. Each county is assigned to the weather station from the same climate zone. This creates an hourly time series of temperature for every county in the WECC. We call this data our Class A temperature time series. We then derive additional time series of county level temperatures by increasing the number of weather stations available in each climate zone. Each combination of weather stations is referred to as a Class and we use 6 total Classes in our analysis.

Class A: The original 8 weather stations - 1 from each IECC climate zone (Fig. 1).
Class B: This Class also uses 1 weather station from each climate zone, but the station is chosen at random (excluding the Class A stations) as opposed to being subjectively selected like the Class A stations. There are a total of 8 weather stations in Class B, 1 from each climate zone.
Class C: The Class B stations plus 1 additional station from each climate zone. The additional station is chosen at random (excluding the Class A stations) for a total of 16 stations, 2 from each climate zone.
Class D: The Class C stations plus 2 additional stations from each climate zone. The additional stations are chosen at random (excluding the Class A stations) for a total of 32 stations, 4 from each climate zone.
Class E: The Class D stations plus 4 additional stations from each climate zone. The additional stations are chosen at random (excluding the Class A stations) for a total of 64 stations, 8 from each climate zone.
Class F: All of the possible stations from each climate zone. While the exact number varies by year, there are roughly 150 weather stations available in the WECC.

Each of our six Classes of data are summarized in Table 1. The purpose of the Class B dataset is to test the sensitivity to our initial subjective choice of which 8 stations to use. The remaining Classes are used to quantify how much information is added, and potentially how much the simulated load biases may be reduced, by forcing BEND with weather data from more and more weather stations. By intentionally including sites from the previous Classes in our Class C-E datasets we ensure that any differences in error characteristics from Class to Class are primarily due to changes in the number of sites utilized as opposed to changes in the exact sites selected. In the Supplemental Material we show that our error estimates are similar compared to when the sampling between Classes B-E is independent. In the event that there are not enough weather stations in a given climate zone to satisfy the requested number of stations for a given Class, all of the stations in that zone are used – effectively going straight to Class F. This is only an issue in the smaller climate zones (i.e., 2A and 7C) and for Class E. In each Class, every county gets mapped to the closest station located in the same climate zone. An example of this mapping for the year 2008 is shown in Fig. 3. Because the number and location of available stations with a robust data record varies by year, the exact stations in each Class and the mapping of stations to counties are redefined each year. We use temperature time series from 2007–2010 for each of these Classes in our analysis.

The stations in each Class are chosen at random in order to

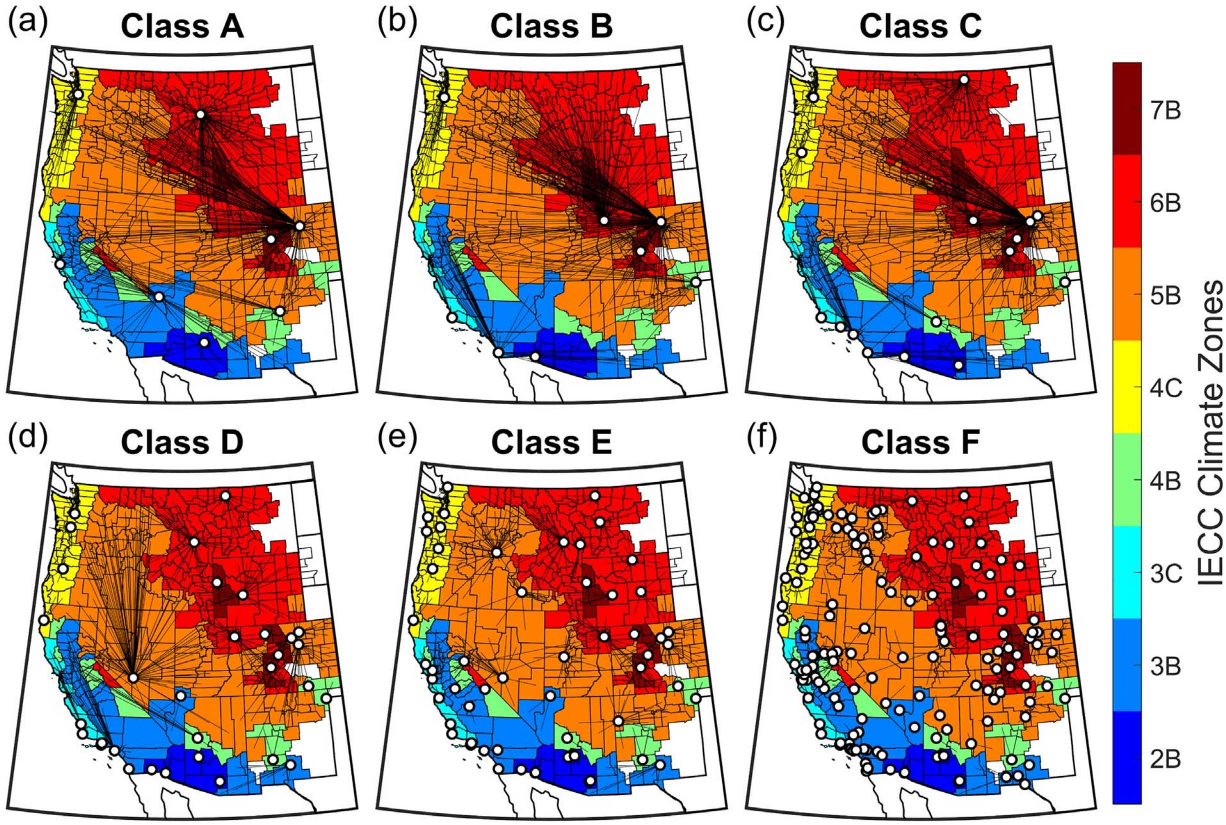

Fig. 3. An example of the mapping between weather stations and counties for each of our 6 Classes of representative weather stations in the year 2008.

minimize the chance that the information gained by using additional stations is dependent on the exact stations chosen. If an ideal number of stations were selected from our analysis then, in theory, the temperature and load biases using that number of stations could be further reduced by objectively selecting the stations that minimize the temperature and load biases in each climate zone. Even in Class F there are still a relatively small number of stations used to represent the weather across the entire western U.S. To evaluate the impact of using a limited number of stations to capture spatially heterogeneous meteorology, we use an additional dataset of surface temperature that has homogenous sampling across the WECC. Specifically, we use the 2-m air temperature estimates from the North American Land Data Assimilation System (NLDAS-2; [36]). Air temperature in NLDAS is taken from the National Center for Environmental Prediction North American Regional Reanalysis. Reanalyses are estimates of true conditions based on a blend of observations and model predictions. NLDAS has an hourly temporal resolution and a 1/8° spatial resolution across the U.S. Because the 1/8° NLDAS dataset largely captures the spatial heterogeneity in temperature across the WECC, we use it as "truth" to evaluate the inhomogenously sampled representative weather station temperatures. We map the gridded NLDAS temperatures to each county by matching the population-weighted centroid of every county to the closest 1/8° NLDAS cell. We use NLDAS data from 2007–2010 to match the time series from our 6 Classes of representative weather stations.

The base dataset used in our analysis is an hourly time series from 2007–2010 of temperature for each county in the WECC from the Class A-F weather stations as well as the NLDAS reanalysis. Having these temperatures at the county level also allows us to calculate population-weighted temperatures on larger spatial scales. Population-weighted temperatures reflect the temperature impacting the most people, a metric important to the estimation of regional electric loads from buildings [6]. Our analysis of this dataset proceeds in three sequential steps:

(1) For each PROMOD zone, we calculate the population-weighted mean temperatures for every hour in the time series using data from the Class A-F datasets as well as NLDAS. Using population to upscale from the county level ensures that the PROMOD zone mean temperature reflects the temperature impacting the most people within each PROMOD zone. We also calculate a time series of temperature biases compared to NLDAS for each Class. Temperature biases are defined as the representative weather station temperature minus the NLDAS temperature. The root-mean-squared (RMS) biases are analyzed to quantify how much temperature biases are reduced when increasing from the original 8 stations (Class A) to all available stations (Class F).
(2) The results of the temperature analysis are insightful, but the key question is how much do these temperature biases matter to the simulation of building energy demand in the WECC? It is impractical to run BEND for all of our Classes of representative cities, so to answer this we use the calibrated BEND load time series for each PROMOD zone from our previous BEND simulation to derive a regression relationship between the zone mean Class A temperatures and the simulated BEND loads. The regressions are second-order polynomial fits between the hourly population-weighted mean temperature and the simulated hourly building energy demand. We tested a variety of different regression models, including the more standard piecewise linear model, and found that the 2nd order polynomial model consistently gave higher correlations with simulated BEND loads in almost every PROMOD zone we examined. Since the BEND run we utilize was forced using the 8 Class A weather stations, these regression relationships reflect the model's intrinsic response to changes in temperature. Examples of these regression relationships for 4 PROMOD zones are shown in Fig. 4 and the relationships for all zones are included in the Supplemental Material Fig. S1.
Once these regression relationships are established, a time series of building energy demand (the predictand) is generated using the

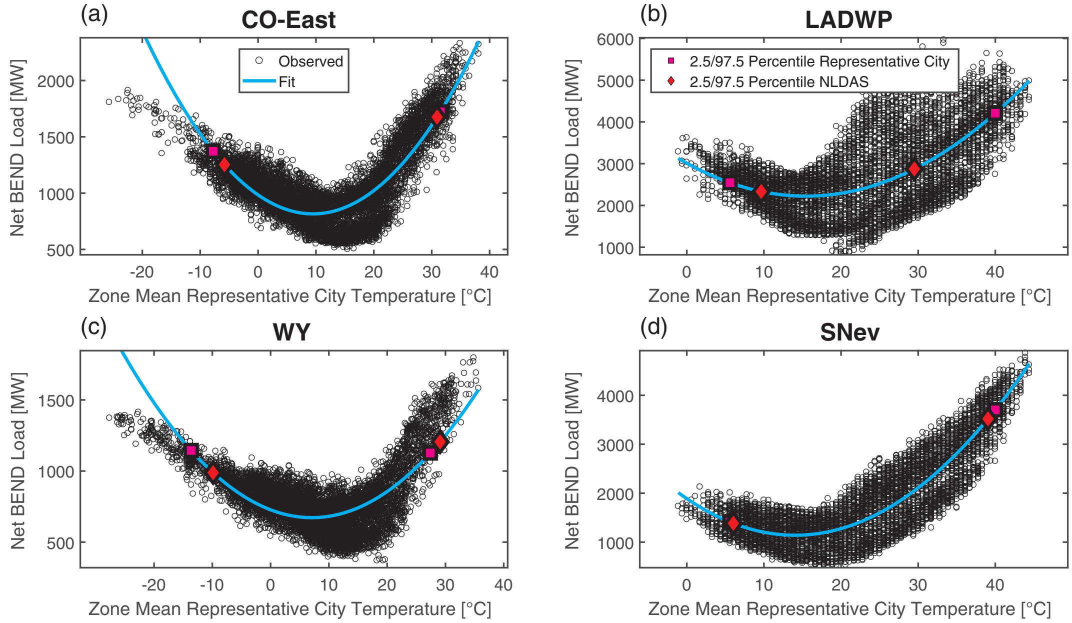

Fig. 4. Examples of the relationship between the zone mean Class A temperatures in 2010 (x-axes) and simulated building energy demand (y-axes) using the calibrated BEND model in 4 different PROMOD zones. In each panel the black dots indicate hourly values and the gray line is the second-order polynomial regression between temperature and building energy demand. The 2.5 and 97.5 percentiles of zone mean temperature and building load from the polynomial regression are shown for the representative weather station (white squares) and NLDAS (white diamonds) datasets.

time series of mean temperatures (the predictor) for the Class A-F and NLDAS datasets in each PROMOD zone. The basic premise is that, if we make the simplifying assumptions that the variability in building energy demand in a given PROMOD zone is due only to changes in temperature and that the relationship between temperature and load is deterministic, then the difference in predicted building energy demand between each Class of representative weather stations and the NLDAS data reflects how much the simulated building loads would change by increasing the number of weather stations used in the simulation. While recognizing that even the calibrated BEND simulations are biased with respect to observations and that the regressions are imperfect representations of the coupling between temperature and building energy demand, we minimize the impact of these uncertainties by applying the same regression relationship to both the representative city temperature time series and the NLDAS temperature time series and basing our interpretation of load biases on the relative rather than absolute differences in the predicted building loads.

(3) The final part of our analysis involves two shorter investigations. First, we examine how the spatial scale over which the temperature and load biases are calculated impacts their distribution. Our current study uses building simulation results which were aggregated to the spatial scale of PROMOD zones to reflect the data available for calibration, but in theory they could be aggregated to the utility or balancing authority level, the IECC climate zone level, or the state level. Each of these spatial scales represents distinct stakeholders in the energy system. To convert the estimated building loads across scales we start with the time series of loads at the PROMOD zone level and disaggregate those values downscale to the county level using populations as weighting factors. This gives us both temperature and estimated building loads at the county scale for each of our Classes of data, which we then aggregate to incrementally larger spatial scales to examine scale dependencies. Temperatures are population-weighted averages when converting to larger spatial scales while loads are simply added (they were previously population-weighted when downscaling from the PROMOD zones). This methodology is summarized in Fig. 5. The approach in Fig. 5 was repeated for each of our 6 Classes of data, for each season, and the final spatial aggregations were performed on multiple spatial scales (Section 3.4).

We examine the mean and extreme biases on each spatial scale. Our final analysis looks at how extremes in the temperature and load distributions vary across the different Classes of representative weather stations. This analysis of extremes reflects the community's growing interest in how extreme temperatures and peak loads may change in a warmer climate (e.g., [14]). Each of these analyses is done seasonally to allow us to distinguish the impact of hot or cold temperature biases. We use meteorological seasons: winter (December-January-February; DJF), spring (March-April-May; MAM), summer (June-July-August; JJA), and fall (September-October-November; SON).

3. Results

3.1. Temperature bias characteristics

As outlined in the Data and Methods section, our Class A dataset utilizes 1 weather station from each of the 8 climate zones in the WECC. As a starting point to understanding how "representative" these stations may be, one might postulate that their representativeness is inversely proportional to the variability of temperatures within each climate zone. Fig. 6 shows the seasonal mean population-weighted temperature along with the standard deviation of county temperatures within each climate zone. The range of temperatures within a given climate zone is a function of the size of the zone and the variability of weather within it. For example, zone 2B is the most concentrated in area and correspondingly has the smallest variability. Larger zones such as 5B have considerably more spread. The largest standard deviation occurs during meteorological summer in every zone except 4B and the largest of these summertime standard deviations occur in zones 3B (represented by Las Vegas) and 3C (represented by San Francisco). These patterns are reflected throughout the remainder of our analysis.

Fig. 7 shows the seasonal mean temperature bias from 2007–2010 for each county in the WECC calculated using the Class A and Class F datasets. As a reminder, biases are computed with respect to the NLDAS reanalysis. The spatial and seasonal patterns in Fig. 7 reflect the

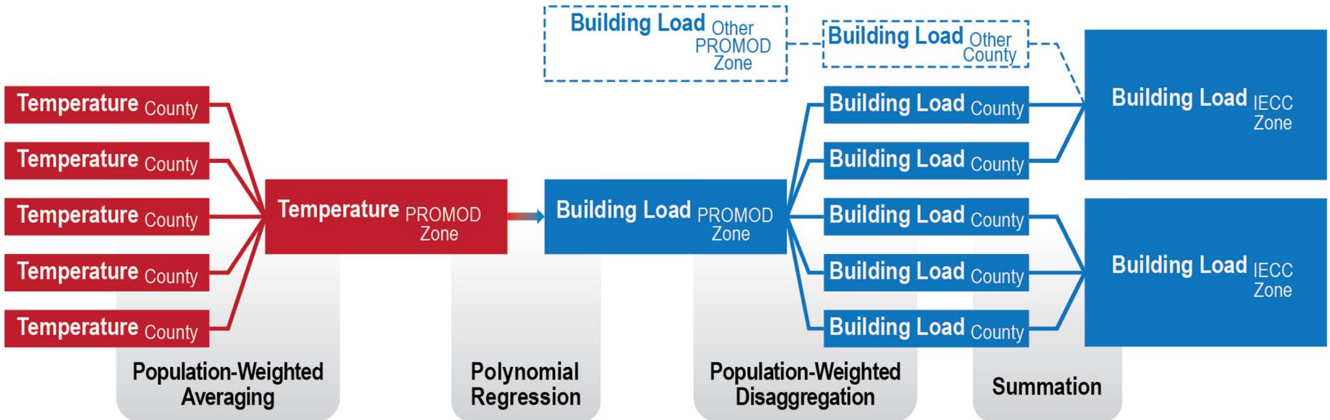

Fig. 5. A flow-chart summarizing our approach to going from temperatures at the county level (far left) to building loads at the IECC climate zone level (far right) using multiple steps to upscale and downscale the data. The dashed boxes indicate data-streams from other PROMOD zones that contribute to building loads at the IECC climate zone spatial scale.

variability in Fig. 6. The largest biases tend to occur during the summer and in southern and central California (climate zone 3B). While the biases are not quite as large, temperature biases during MAM are sizeable in the same regions. Fig. 7 also shows the degree to which temperature biases are reduced in the Class F dataset. The positive temperature biases in southern California apparent in Class A largely disappear as more weather stations are included that are closer to the major population centers in southern California. Bias maps for the other Classes are included in the Supplemental Material Fig. S2-5.

The distribution of population-weighted RMS temperature biases for the entire WECC for each of our six Classes are shown in Fig. 8. Data in this figure is based on 10 random samples of the Class B-E weather stations for the full 2007–2010 time series. Because the exact stations selected varies by year, each year is taken as a separate sample. The variability in the Class A and Class F data in Fig. 8 reflects the inter-annual variability in representativeness (i.e., the same Class A/F sites may be more representative in 2007 compared to 2010). Consistent with the Figs. 6 and 7, the largest biases occur during summer. The large Class A warm biases in the heavily-populated areas of southern California are reflected in Fig. 8. The Class A dataset in JJA has an average RMS bias of ~4.0 °C, the largest of any Class or season. The summertime temperature bias decreases as more and more sites are made available – a signal that is consistent across all seasons. The RMS bias for the Class F JJA dataset is only ~1.5 °C – roughly 2.5× smaller than in Class A. This ratio is 1.5×, 2.2×, and 1.5× during DJF, MAM, and SON, respectively. The decrease in temperature biases from one Class to another is not always linear. For example, the bias during JJA is reduced by 1.1 °C when going from 1 (Class B) to 4 (Class D) representative weather stations, but only falls by an additional 0.5 °C when all available stations are utilized (Class F). A similar pattern is observed during the other seasons. For our purposes, this signal of diminishing returns means we may be able to significantly improve the performance of BEND, or another similar aggregate building model, by marginally increasing the number of stations used and total simulations required, but that there may not be much to gain by using all of the possible stations. This is not altogether surprising because weather systems tend to similarly affect fairly large geographic regions; once the spatial resolution of the representative weather stations set is within the typical bounds of weather systems' reach, temperatures from additional stations are likely highly correlated with other stations already included. Computationally this is an important result because the number of individual building simulations required scales non-linearly with the number of stations used to force the model (Table S1).

One other interesting aspect of the temperature biases can be seen by comparing the distributions of the biases Class A and Class B datasets for 1 random sample of the full 2007–2010 time series (Supplemental Material Fig. S6). Recall that the Class B dataset also uses only 1 representative station per climate zone, but the stations were randomly

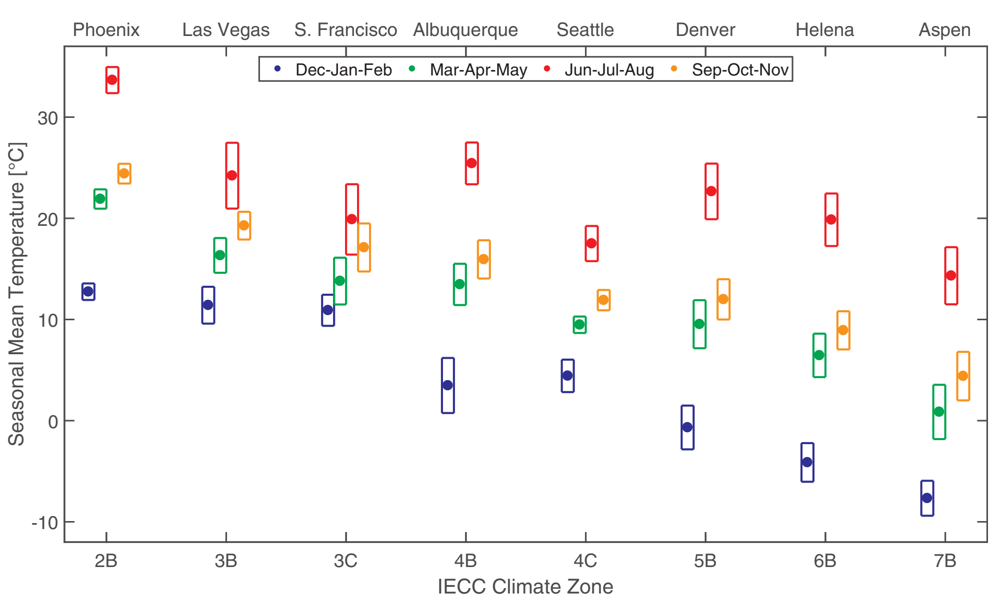

Fig. 6. A summary of the 8 IECC climate zones, their representative weather stations, and their seasonal climate characteristics derived from the NLDAS dataset using all available data from 2007–2010. Seasonal mean temperatures are indicated by the solid markers and the boxes bound ±1 standard deviation. Mean temperatures are population-weighted by county to reflect the temperature impacting the most people. The temperature standard deviation is meant to convey spatial variability and is calculated as the standard deviation of the seasonal mean temperature across all counties within a given climate zone.

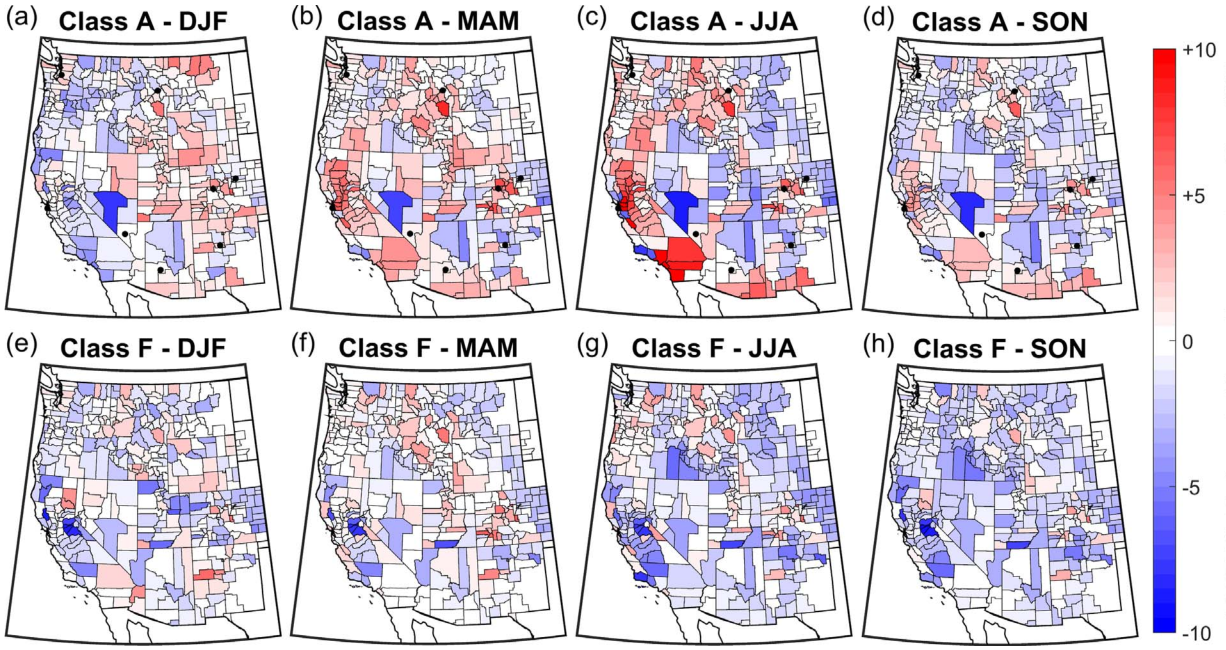

Fig. 7. The seasonal mean temperature bias for each county in the WECC calculated using the (top row) Class A representative weather stations and (bottom row) all available weather stations. The Class A stations are indicated by black dots in the top row maps. The biases are based on all data from 2007–2010.

rather than objectively selected. While the JJA RMS temperature bias in Class B is slightly smaller compared to Class A, the more interesting signal is how the distribution shifts from being skewed towards positive values to being skewed toward negative values (Fig. S6c). This implies that if we had picked an alternate 8 stations to drive the model our simulation could have changed from having a significantly positive temperature and load bias to significantly negative biases. For the purposes of investigating how many stations should be used for calibrating and forcing BEND or other aggregate models, 1 station per climate zone is not enough to consistently represent the variance in temperature. This result is also supported by Huang et al. [13], who found that only using 1 station to represent the weather in an IECC climate zone can lead to systematically underestimated variability in building loads.

3.2. Load bias characteristics

The underlying purpose of this paper is to better understand how temperature biases resulting from using representative weather stations to force aggregate building models are likely to impact the load characteristics in the simulations. To this end, we converted the hourly temperature time series from each of our Classes as well as the NLDAS dataset to a load time series using the regression approach described in the Data and Methods section. This allows us to analyze load biases that are more relevant to future studies of aggregate building energy demand. Because we know that our simulation produces building loads that are flawed compared to observed values, we chose to base our analysis on *relative* load biases rather than *absolute* load biases. The relative bias is computed by calculating a total building load time series across the WECC using the NLDAS-derived loads and comparing it to the total load time series from each of our 6 Classes of representative weather stations. The seasonal distributions of these relative biases are shown in Fig. 9.

The distributions in Fig. 9 reflect many of the features in Fig. 8. The absolute reduction in bias from Class A to Class F ranges from −7.5% in

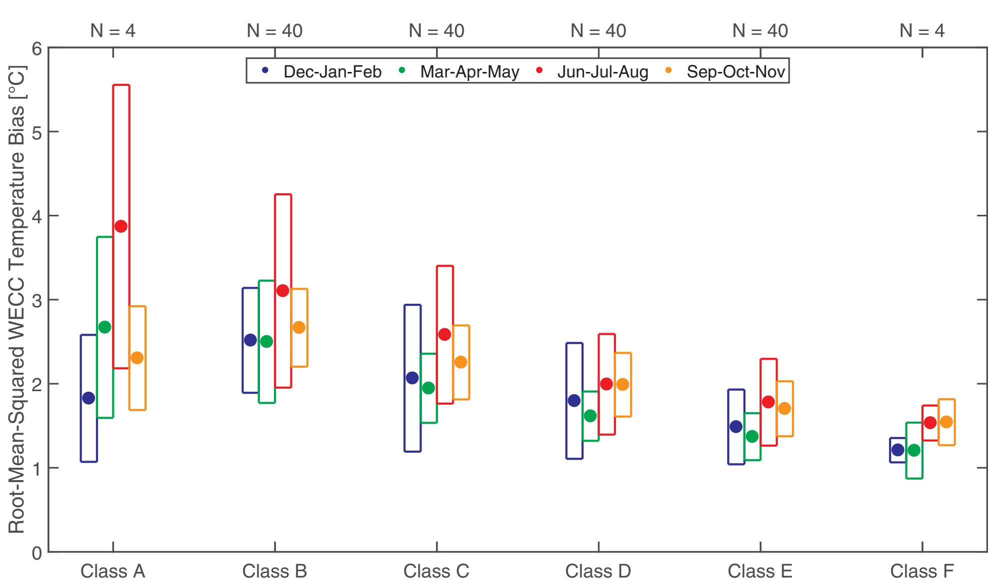

Fig. 8. Seasonal distributions of the population-weighted root-mean-squared (RMS) temperature bias across the WECC for each of the six Classes of representative weather stations. The solid markers indicate the average RMS bias for each season and the boxes bound ±1 standard deviation. The distributions are based on 8 random samples of the Class B-E representative weather stations. Each year (2007–2010) is taken as a separate sample and the sample size for each Class of data is given on the top axis. Biases are computed at the county level and then averaged across the WECC using county populations as weighting.

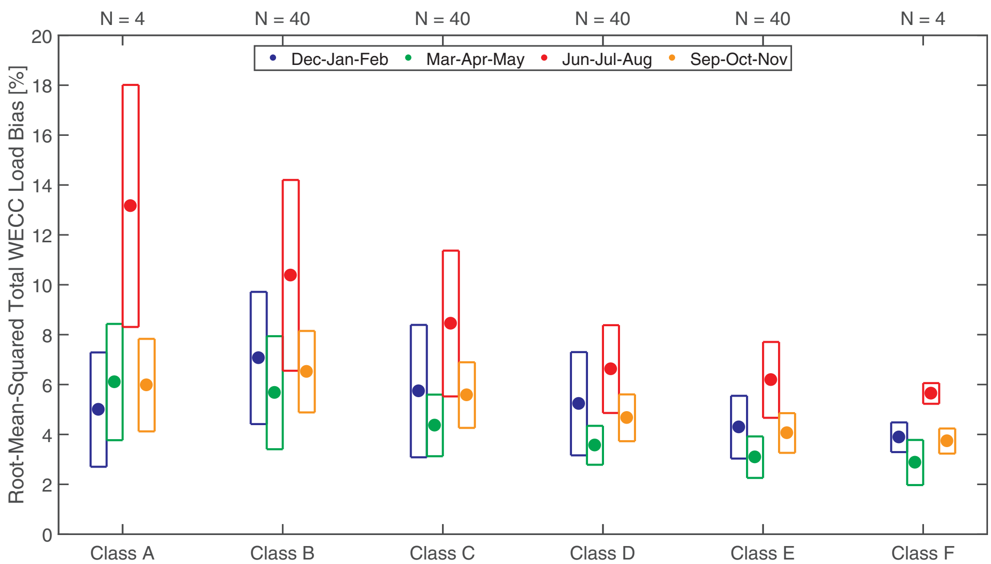

Fig. 9. Seasonal distributions of the total building energy demand bias for the entire WECC for each of the six Classes of representative weather stations. The solid markers indicate the average RMS bias for each season and the boxes bound ±1 standard deviation. The distributions are based on 8 random samples of the Class B-E representative weather stations. Each year (2007–2010) is taken as a separate sample and the sample size for each Class of data is given on the top axis.

summer to −1.1% in winter. As with the temperature biases, the reduction is not always linear. During JJA, the mean RMS bias of 13.2% using the Class A representative weather stations is reduced to 5.6% in the Class F dataset. A bias of this magnitude has implications for both utilities and electricity grid managers. Much of the long-term capacity expansion planning for these entities is based on the idea of ensuring that adequate reserve margins are maintained even as the climate shifts towards larger loads due to increased cooling demand in the summer. If these groups used aggregate building simulations as a tool in their planning, a 13% bias would lead to a significant overestimation of the capacity required to meet future demand.

Another notable component of these relative load biases is the decoupling between temperature biases and load biases during spring and fall. As seen in Figs. 7 and 8 and discussed in the previous section, there are fairly large Class A temperature biases during MAM. However, these positive temperature biases do not translate to large load biases during MAM (Fig. 9). The likely cause of this is the relationship between temperature and load. Simulated building loads in almost every PROMOD zone respond sharply to small changes in temperature under very warm or very cold conditions, but have a flatter response for temperatures closer to the balance-point temperature (~13–18 °C in most zones) where neither heating nor cooling is required (Fig. 4). In terms of biases, this means that temperature biases during MAM or SON, which have seasonal temperatures closer to the balance-point (Fig. 6), translate to smaller building energy demand biases compared to similar magnitude temperature biases that occur during JJA or DJF.

3.3. Repeated sampling for a single season

As demonstrated in Figs. 6–9, the largest temperature and load biases tend to occur during the summer months. For this reason, the remainder of our analysis will focus on June-July-August. To increase the robustness of our results we increased our sample size by randomly reselecting the weather stations in each Class B-E 100 times for the summer of 2010. In addition to providing statistical clarity, focusing on a single season and year also removes seasonal and interannual variability in representativeness for each of our Classes. Any differences in representativeness for these 100 random samples are purely due to changes in the number of weather stations utilized. The trends in temperature and load biases for these 100 random samples are shown in Fig. 10. The trends in JJA 2010 are similar to those noted using the full 2007–2010 data. Notably, the median RMS temperature bias drops by 1.0 °C between Classes B and D, but only an additional 0.5 °C when going all the way to Class F. Similarly, the median RMS load bias drops by 3.4% from Class B to D, but only 1.1% between Class D and F.

3.4. Scale dependencies and heat wave impacts

The previous two sections have focused on mechanistic questions related to designing weather forcing datasets for building energy demand models. For example, how much could the simulated building energy demand change if we used more representative weather stations to force the model? The final two components of our analyses are centered on more decision-maker-centric questions: How do the magnitudes of the temperature and load biases depend on the scale over which they are calculated and to what extent are extreme temperatures and loads captured when we use a limited number of point measurements to represent temperature over an area as large as the WECC. Each of these questions can be framed in terms of how they might impact a decision maker who utilizes aggregate building models like BEND. In terms of the scale dependencies, a rational assumption is that the larger the area over which the biases are calculated the more likely there are positive and negative biases within the domain that can potentially offset each other. For planning purposes of an individual utility, which typically cover only a few contiguous counties, the impact of these biases in the aggregate simulations will be larger because there is a reduced likelihood of offsets. However, on the scale of an entire state, which may use simulations similar to ours to evaluate the impact of potential changes to building codes, the potential for cancelling effects is larger. The issue of extreme temperatures and loads will impact capacity expansion planners who design their future systems to maintain reliability during peak loads – for example, heat waves. For this reason it makes sense to examine how well heat waves and extreme loads are captured in a simulation using a limited number of representative weather stations.

To address the scale dependency issue, we take the JJA 2010 time series of temperature and building load for each of our Classes and the NLDAS dataset and aggregate them across different spatial scales ranging from balancing authorities (typically the smallest) to IECC climate zones (typically the largest). The maximum load biases on each spatial scale are shown in Fig. 11 and the maximum temperature biases are shown in Fig. S8 in the Supplemental Material. In the Class A data, the largest mean temperature bias for any balancing authority in the WECC is 14.1 °C while for states the largest bias is only be 3.9 °C. Maximum load biases shrink from 44.1% for balancing authorities to only 21.6% for states (Fig. 11). Similar patterns are observed in the PROMOD zones, which are typically smaller than most states in the western U.S. Even using the Class F datasets, the largest PROMOD zone or balancing

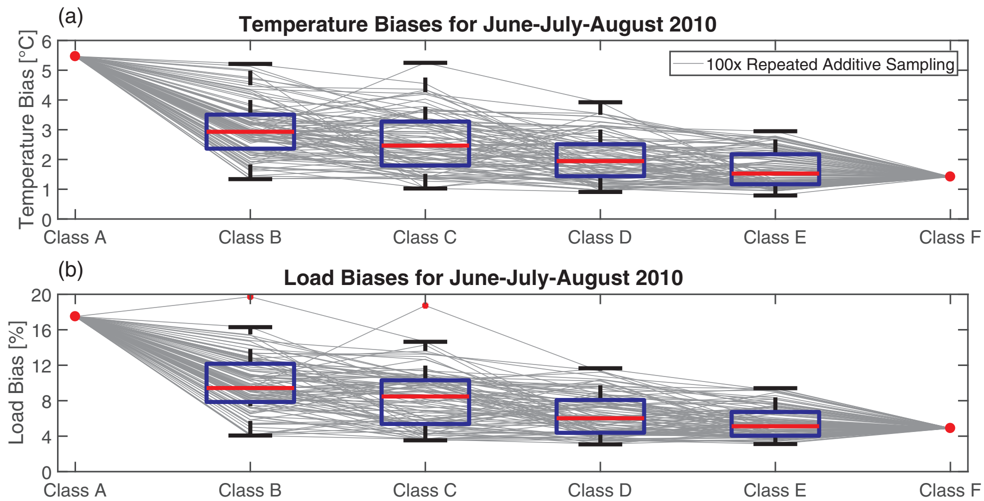

Fig. 10. The trend in root-mean-squared (a) temperature and (b) load biases during June-July-August 2010 for 100 random samples of weather stations for Classes B through E. In the box plots the horizontal lines indicate the median value, the boxes indicate the area between the 25th and 75th percentiles, and the whiskers extend to the 5th and 95th percentiles. The thin gray lines shown the trends for individual samples.

authority temperature or load biases are still 200–400% of the maximum bias for states. This means that even if we used all available weather stations to force the aggregate building model we could still end up with significant biases at the balancing authority or PROMOD zone scale.

Fig. 12 shows one way to visualize the spatial structure of heat waves in the various datasets we use. Here we define heat wave following the American Meteorological Society convention – "A period of abnormally and uncomfortably hot and usually humid weather". To create this figure, we used the 2008 time series of temperature from the NLDAS dataset as well as each Class of representative weather stations to look for extended periods of hot weather impacting the WECC. We take each hourly time series and compute a 5-day running mean of population-weighted temperature for the entire WECC. The maximum value from that population-weighted temperature time series is used as the center time for the peak heat wave in each dataset. Fig. 12 shows the 5-day average temperature around this center point identified in the NLDAS, Class A, and Class F datasets. Several features are apparent. As expected, the Class A dataset lacks the spatial variability that is present in either the NLDAS or Class F datasets. The Class A data are actually colder in many areas across the WECC (e.g., Utah, Arizona, and New Mexico), but are anomalously hot over Las Vegas and Phoenix. Because these cities map to densely populated areas, including southern California, they are the driving force behind the overall heat wave signal in the population-weighted temperatures. Additionally, the Class A dataset has a peak population-weighted temperature that is ~3 °C warmer than the other datasets. In terms of understanding aggregate building models and their potential biases, this shows that unusually hot temperatures in two relatively small metropolitan areas could actually manifest in the model as a heat wave impacting a much larger area. The spatial variability of the heat wave in the NLDAS dataset suggests that we may actually need weather data from individual counties to resolve peak demand during a heat wave using representative weather stations.

Another way to look at heat waves and extreme temperatures is to examine how the tails of the temperature or load distributions could change by including more and more weather stations. We examine tails in the distributions by calculating the 97.5 percentile of temperature in each PROMOD zone to look for heat waves during spring and summer and the 2.5 percentile to look for cold snaps during fall and winter. We also calculate the 97.5 percentiles of the corresponding building loads calculated using the regression relationships. The 2nd order polynomial regressions we utilize predict these extreme load values reasonably well compared to the calibrated BEND model for most PROMOD zones (see Figs. 4 and S1). Fig. 13 shows the difference in extreme temperatures compared to NLDAS on the x-axis for the Class A (closed circles) and Class F (open circles) datasets. The y-axis shows the ratio of the load extremes. As an example of how to read this figure, a closed circle value of +5 °C on the x-axis and 1.2 on the y-axis during JJA would indicate that the Class A data for that PROMOD zone has a 97.5 percentile of temperature that is 5 °C warmer than NLDAS and a 97.5 percentile of building load that is 1.2x that from NLDAS.

The limitations of the Class A data discussed throughout this paper

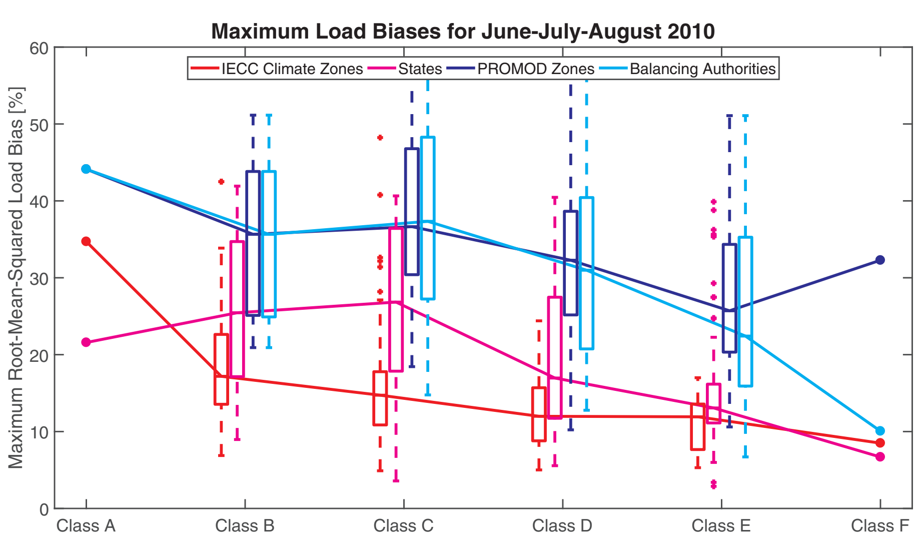

Fig. 11. Distributions of the maximum June-July-August 2010 load bias for each Class of data calculated over different spatial scales: IECC climate zones (red), states (magenta), PROMOD zones (blue), and balancing authorities (cyan). Distributions are based on 100 random samples of weather stations for Classes B through E. In the box plots the horizontal lines indicate the median value, the boxes indicate the area between the 25th and 75th percentiles, and the whiskers extend to the 5th and 95th percentiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

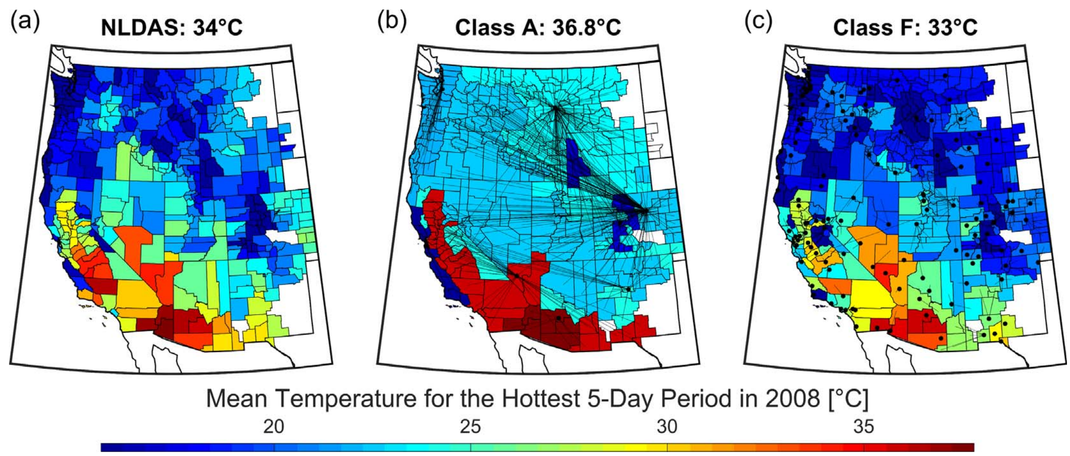

Fig. 12. The mean temperature over the hottest 5-day period in 2008 identified using (a) NLDAS [06/19/08 to 06/23/08], (b) the Class A representative weather stations [06/28/08 to 07/02/08], and (c) all available weather stations [06/19/08 to 06/23/08]. Panels (b) and (c) include a mapping between each county to its associated representative station. The peak population-weighted temperature over the 5-day period for each dataset is given in the panel titles.

are readily apparent in Fig. 13. Focusing on the summer and winter seasons, the Class A data in many PROMOD zones overestimates the extreme load values by a factor ranging from 1.1x all the way up to 1.6x. For the PROMOD zone containing Los Angeles (LADWP), the use of Las Vegas weather in the Class A data leads to a summertime extreme load that is 50% higher than NLDAS. In the winter months many PROMOD zones have extreme cold temperatures that are 5–10 °C colder than NLDAS and corresponding load values that are 20–40% too high. While there are some examples of even the Class F data performing poorly in this particular analysis, for the most part the use of all available weather stations would decrease the aforementioned biases to more acceptable ranges (e.g., ±5 °C and ±10%). In terms of capacity expansion planning, the Class A data could give the false impression that more capacity is required than is actually necessary to accommodate heat waves and cold snaps.

4. Discussion and conclusions

Numerical building simulations are used in many sectors to explore how current and future building stocks may respond to climate change or changes in building technologies or codes. Recent advances in computational power have facilitated the use of aggregate building models in which many thousands of buildings of different sizes and vintages are simulated in a given geographical area. Aggregate models are attractive because they maintain the detailed, physically-based, and potentially non-linear responses of individual buildings while at the same time yielding a more holistic view of building energy demand on larger spatial or longer temporal scales. Observed or projected weather at select locations are used to force the models – providing the physical linkage between building energy demand and changes in climate or weather. A common approach to reduce the computational demand of these simulations is to use a subset of “representative” weather stations to drive the models. While reducing computational costs, this simplification potentially leads to significant biases in the simulated building loads [13,19]. By quantifying the relationship between the number of sites and the bias in aggregate building loads we bounded one common source of errors in aggregate building models.

This analysis began as a way to understand biases in PNNL’s aggregate Building ENergy Demand (BEND) model, which we are utilizing in a separate project to examine climate change impacts on building energy demand in the WECC. In that simulation 8 representative weather stations, 1 for each IECC climate zone in the WECC, were used

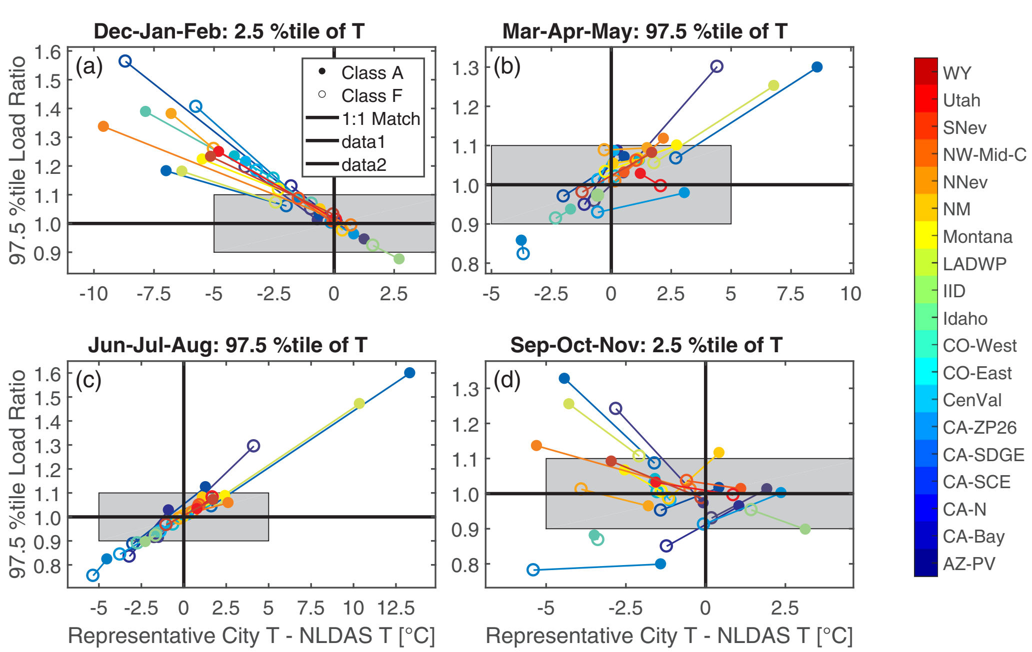

Fig. 13. The dots indicate the difference in extreme seasonal temperatures (x-axes) and the ratio of the 97.5 percentile of building energy demand (y-axes) between the (closed circles) Class A and (open circles) Class F representative weather stations values compared to the values in the NLDAS dataset. Extremes during SON and DJF are defined using the 2.5 percentile of temperature to demonstrate sensitivity to extreme cold temperatures while data from MAM and JJA use the 97.5 percentile of temperature to demonstrate sensitivity to extreme warm temperatures. The lines indicate the change in ratios observed when we increase the number of weather stations from 8 (Class A) to all available stations (Class F). Differences are calculated for each season and each PROMOD zone in the WECC for the full 2007–2010 time series. Note that the axis limits vary in each panel. The 1:1 lines and shaded areas (±5 °C and ±10%) are included for reference.

to force the building models. This paper quantifies how much the temperature and aggregate building load biases may be reduced by using increasing numbers of stations. Rather than rerunning the simulations, we employed an analytical approach using county level temperature time series based on the 8 representative stations (our Class A dataset) as a base and then increasing the number of stations until we eventually use all stations in the WECC (Class F). These county level temperature time series were converted to building loads using polynomial regressions based on calibrated BEND output. Temperature and load biases were computed with respect to the 1/8° NLDAS weather reanalysis dataset that was also mapped to county level.

The primary results of our analysis are as follows:

(1) During meteorological summer, when temperature variability peaks in many climate zones, the Class A dataset had an average absolute temperature bias of ~4.0 °C compared to NLDAS. The distribution of this bias is skewed toward anomalously warm values. This is primarily a result of using weather data from Las Vegas to represent the densely populated areas in southern California. The mean bias drops to only ~ 1.5 °C when using all of the available weather stations. Class A and Class B temperature and load biases were similar – indicating that the biases using only 1 weather station per climate zone were mostly independent of the exact stations used. For all but the largest aggregate spatial scale, one station per climate zone is insufficient to calibrate and force aggregate building energy demand models (e.g., [13]).
(2) The mean aggregate building load bias when using only 1 weather station per climate zone was 13.2% in the summer. This bias drops to 5.6% when using all available stations. While still non-zero, reducing load biases by a factor of 2.5 could be important to many stakeholders. Similar magnitude temperature biases in the spring and fall had a smaller overall impact on aggregate building loads because temperatures in these seasons are closer to the balance-point temperature, around which building loads respond more slowly to variations in temperature compared to biases occurring farther from the balance-point.
(3) Increasing the size of the domain over which biases are calculated reduces their magnitude as positive and negative biases may cancel each other out. For utilities and small balancing authorities, a maximum load bias of 44.1% using the Class A data would likely be a cause for concern whereas a maximum Class A load bias of 21.6% at the state level may be more tenable. At the spatial scale of IECC climate zones, the largest used in our analysis, one station may be sufficient as the maximum temperature and load errors are insensitive to the number of stations used. This result also suggests that simulations aimed at predicting annual energy demand, which would be subject to similar cancelling biases as the larger spatial scales in our analysis, may be less sensitive to the number of representative cities used to force the model.
(4) Using a small number of representative weather stations results in the overestimation of peak loads during both summer and winter. Capturing heat waves, which have distinct spatial patterns, may require using weather data from individual counties. During the winter, peak loads ranged from 20–40% too high when using only 1 station per climate zone. Biases of this magnitude would obviously be of concern to utilities, grid managers, and capacity expansion planning groups who often engineer their systems to meet peak demand under current and future climate conditions. Simulations aimed at predicting peak loads under heat wave conditions, which vary sharply in space and time, are more sensitive to the number of weather stations used.

One signal that was clear in our analysis was the diminishing returns in bias reduction when using more than 4 stations per climate zone. For all but the largest spatial scale (i.e., IECC climate zones), most seasons and regions had a significant drop in mean temperature and load biases when increasing from 1 to 4 weather stations per climate zone, but showed diminishing returns with more than 4 stations. This suggest that using 4 stations per climate zone, which would drastically reduce computational costs compared to using all available weather stations, may be good enough for some applications. We did not attempt to identify the specific stations that minimized the overall temperature or load biases in our analysis. However, using the example described above where 4 randomly selected stations per climate zone provided decent results, the temperature and load biases could be reduced even further by objectively identifying the 4 stations per zone that minimized the absolute temperature or load biases. Moving forward, our calibration tests for BEND utilize 4 weather stations per climate zone.

By quantifying the sensitivity of temperature and load biases to the number of representative weather stations we created a baseline for future studies over the western U.S. Our novel approach, which required no new simulations, could also be useful to other groups designing or calibrating aggregate building model simulations – particularly those looking at the impact of future climate change scenarios. Future temperature and load biases may be enhanced if climate change impacts different locations in different ways, increasing the spatial variability and decreasing the “representativeness” of the stations used to drive the model. Our work also highlighted the importance of using stations close to population centers which dominate energy usage in the large but mostly rural WECC. This paper should be of interest to utilities and grid managers attempting to forecast how much power may be required to serve future instantiations of the grid. Our analysis demonstrates that while aggregate building simulations are useful, one needs to carefully evaluate how the models are forced to reduce the likelihood of significant biases in required energy. Our analysis is a first step toward a longer term vision of developing a robust, spatially distributed, aggregate building energy demand model whose potential biases are well characterized.

Acknowledgments

This work was supported by the Office of Science of the U.S. Department of Energy, Office of Biological and Environmental Research as part of the Integrated Assessment Research Program. The initial development of the BEND model was supported by the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory, a multi-program national laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830. The authors wish to thank J. Hathaway and Trenton Pulsipher for some relevant work on this topic during the initial development of the BEND model.

Conflict of interest

The authors declare no conflicts of interest involving this work.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.apenergy.2017.08.244.

References

[1] Franco G, Sanstad AH. Climate change and electricity demand in California. Clim Change 2008;87:139–51. http://dx.doi.org/10.1007/s10584-007-9364-y.
[2] Lu N, et al. Climate change impacts on residential and commercial loads in the Western U.S. grid. IEEE Transact Power Syst 2010;25:480–8. http://dx.doi.org/10.1109/TDC.2010.5484537.
[3] Li DHW, Yang L, Lam JC. Impact of climate change on energy use in the built environment in different climate zones – a review. Energy 2012;42:103–12. http://dx.doi.org/10.1016/j.energy.2012.03.044.
[4] Rose SK, Turner D, Blanford G, Bistline J, de la Chesnaye F, Wilson T. Understanding the social cost of carbon: a technical assessment. Washington, D.C.:

Electric Power Research Institute; 2014. Report. Technical Report, 3002004657.

[5] Wang H, Chen Q. Impact of climate change heating and cooling energy use in buildings in the United States. Energy Build 2014;82:428–36. http://dx.doi.org/10.1016/j.enbuild.2014.07.034.

[6] Zhou Y, et al. Modeling the effect of climate change on U.S. state-level buildings energy demands in an integrated assessment framework. Appl Energy 2014;113:1077–88. http://dx.doi.org/10.1016/j.apenergy.2013.08.034.

[7] Dirks JA, et al. Impacts of climate change on energy consumption and peak demand in buildings: a detailed regional approach. Energy 2015;79:20–32. http://dx.doi.org/10.1016/j.energy.2014.08.081.

[8] Sailor DJ. Relating residential and commercial sector electricity loads to climate – evaluating state level sensitivities and vulnerabilities. Energy 2001;26:645–57. http://dx.doi.org/10.1016/S0360-5442(01)00023-8.

[9] Mirasgedis S, Sarafidis Y, Georgopoulou E, Lalas DP, Moschovits M, Karagiannis F, et al. Models for mid-term electricity demand forecasting incorporating weather influences. Energy 2006;31:208–27. http://dx.doi.org/10.1016/j.energy.2005.02.016.

[10] Wan KKW, Li DHW, Pan W, Lam JC. Impact of climate change on building energy use in different climate zones and mitigation and adaptation implications. Appl Energy 2012;97:274–82. http://dx.doi.org/10.1016/j.apenergy.2011.11.048.

[11] Xu P, Huang YJ, Miller N, Schlegel N, Shen P. Impacts of climate change on building heating and cooling energy patterns in California. Energy 2012;44:792–804. http://dx.doi.org/10.1016/j.energy.2012.05.013.

[12] Brown MA, Cox M, Staver B, Baer P. Modeling climate-driven changes in U.S. buildings energy demand. Clim Change 2016;134:29–44. http://dx.doi.org/10.1007/s10584-015-1527-7.

[13] Huang J, Gurney KR. Impact of climate change on U.S. building energy demand: sensitivity to spatiotemporal scales, balance point temperature, and population distribution. Clim Change 2016;137:171–85. http://dx.doi.org/10.1007/s10584-016-1681-6.

[14] Auffhammer M, Baylis P, Hausman CH. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the United States. Proc Natl Acad Sci USA 2017;114:1886–91. http://dx.doi.org/10.1073/pnas.1613193114.

[15] Spandagos C, Ng TL. Equivalent full-load hours for assessing climate change impact on building cooling and heating energy consumption in large Asian cities. Appl Energy 2017;189:352–68. http://dx.doi.org/10.1016/j.apenergy.2016.12.039.

[16] Crawley DB, et al. EnergyPlus: creating a new-generation building energy simulation program. Energy Build 2001;33:319–31. http://dx.doi.org/10.1016/S0378-7788(00)00114-6.

[17] Jentsch MF, Bahaj AS, James PAB. Climate change future proofing of buildings – generation and assessment of building simulation weather files. Energy Build 2008;40:2148–68. http://dx.doi.org/10.1016/j.enbuild.2008.06.005.

[18] Chan ALS. Developing future hourly weather files for studying the impact of climate change on building energy performance in Hong Kong. Energy Build 2011;43:2860–8. http://dx.doi.org/10.1016/j.enbuild.2011.07.003.

[19] Wang L, Mathew P, Pang X. Uncertainties in energy consumption introduced by building operations and weather for a medium-size office building. Energy Build 2012;53:152–8. http://dx.doi.org/10.1016/j.enbuild.2012.06.017.

[20] Shen P. Impacts of climate change on U.S. building energy use by using downscaled hourly future weather data. Energy Build 2017;134:61–70. http://dx.doi.org/10.1016/j.enbuild.2016.09.028.

[21] Shimoda Y, Asahi T, Taniguchi A, Mizuno M. Evaluation of city-scale impact of residential energy conservation measures using the detailed end-use simulation model. Energy 2007;32:1617–33. http://dx.doi.org/10.1016/j.energy.2007.01.007.

[22] Chung M, Park H-C. Development of a software package for community energy system assessment – Part I: Building a load estimator. Energy 2010;35:2767–76. http://dx.doi.org/10.1016/j.energy.2010.02.027.

[23] National Renewable Energy Laboratory. National solar radiation database 1991–2010 update: user's manual. Golden CO: National Renewable Energy Laboratory; 2012. Technical Report, NREL/TP-5500-54824.

[24] National Renewable Energy Laboratory. U.S. Department of Energy Commercial Reference Building Models of the National Building Stock. Golden CO: National Renewable Energy Laboratory; 2011. Technical Report, NREL/TP-5500-46861.

[25] Hong T, Chang W-K, Lin H-W. A fresh look at weather impact on peak electricity demand and energy use of buildings using 30-year actual weather data. Appl Energy 2013;11:333–50. http://dx.doi.org/10.1016/j.apenergy.2013.05.019.

[26] Degelman LO. Which came first – building cooling loads or global warming? – A cause and effect examination. Building Serv Eng Res Technol 2002;23:259–67. http://dx.doi.org/10.1191/0143624402bt049oa.

[27] Crawley DB. Estimating the impacts of climate change and urbanization on building performance. J Build Perform Simulat 2008;1:91–115. http://dx.doi.org/10.1080/19401490802182079.

[28] Wan KKW, Li DHW, Liu D, Lam JC. Future trends of building heating and cooling loads and energy consumption in different climates. Build Environ 2011;46:223–34. http://dx.doi.org/10.1016/j.buildenv.2010.07.016.

[29] Pacific Northwest National Laboratory. Methodology for Evaluating Cost-Effectiveness of Residential Energy Code Changes. Richland, WA: Pacific Northwest National Laboratory; 2015. Technical Report, PNNL-21294 Rev 1.

[30] Sheridan B, Baker SD, Pearre NS, Firestone J, Kempton W. Calculating the offshore wind power resource: Robust assessment methods applied to the U.S. Atlantic Coast. Renew Energy 2012;43:1–10. http://dx.doi.org/10.1016/j.renene.2011.11.029.

[31] Piersanti A, Vitali L, Righini G, Cremona G, Ciancarella L. Spatial representativeness of air quality monitoring stations: a grid model based approach. J Atmos Poll Res 2015;6:953–60. http://dx.doi.org/10.1016/j.apr.2015.04.005.

[32] van Wart J, Grassini P, Cassman KG. Impact of derived global weather data on simulated crop yields. Glob Change Biol 2013;19:3822–34. http://dx.doi.org/10.1111/gcb.12302.

[33] Tomasek BJ, Williams II MM, Davis AS. Changes in field workability and drought risk from projected climate change drive spatially variable risks in Illinois cropping systems. PLoS ONE 2017;12:e0172301. http://dx.doi.org/10.1371/journal.pone.0172301.

[34] Tomillo PS, et al. Global analysis of the effect of local climate on the hatchling output of leatherback turtles. Sci Rep 2015;5:16789. http://dx.doi.org/10.1038/srep16789.

[35] Wang N, Makhmalbaf A, Srivastava V, Hathaway JE. Simulation-based coefficients for adjusting climate impact on energy consumption of commercial buildings. Build Simul 2017;10:309–22. http://dx.doi.org/10.1007/s12273-016-0332-1.

[36] Mitchell KE, et al. The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J Geophys Res 2011;109:D07S90. http://dx.doi.org/10.1029/2003JD003823.