Spatiotemporal post-calibration in a Numerical Weather Prediction model for quantifying building energy consumption

Youngchan Jang, Eunshin Byon, *Member, IEEE*, Soham Vanage, Kristen Cetin, David E. Jahn, William Gallus, Jr, Lance Manuel

***Abstract*—Characterizing localized climate conditions is becoming important in many aspects of modern society. The Weather Research and Forecasting (WRF) models have been used to predict localized environmental variations. Further, the recently developed Urban Canopy Model (UCM), derived from energy balance equations, represents more detailed urban characteristics, when it is coupled with the WRF model. However, such physics-based numerical models can exhibit a spatially and temporarily heterogeneous discrepancy pattern compared to actual climate conditions possibly due to inappropriate model specifications and/or incorrect choices of model parameters. This study devises a new method that post-calibrates geographically- and temporally-varying discrepancy in an integrative framework. Tested on urban temperature data collected in the central Texas region during heat wave events, our case study demonstrates that the proposed method substantially reduces prediction errors over the original WRF/UCM projection and other alternative approaches. Based on the results, we quantify the building energy consumption at spatially dispersed locations.**

***Note to Practitioners*– As electricity is a key component of modern society, the economy, and human well-being, the ability to assess the consumption of electricity is increasingly important. This assessment is challenging due to heterogeneous meteorological patterns. While the numerical weather prediction model, such as WRF/UCM, provides localized weather conditions, its output may not be accurate when its parameters and boundary conditions are not properly specified. In particular, we observe the prediction error from WRF/UCM is spatially- and temporally-varying due to the urban heat island effect during heat wave events. This study develops an integrative framework which provides a methodology to post-correct such heterogeneous prediction error. Our case study using data collected in the central Texas region suggests that the energy consumption at highly urbanized areas can be 0.4 kWh-0.5 kWh larger than its surrounding areas during the heat wave periods. The results demonstrate how the proposed approach can benefit a real application use case.**

Youngchan Jang is with the Precedent Study Team for C4ISR System, Korea Research Institute for defense Technology planning and advancement, Jinju, Korea (mapsossa@krit.re.kr). Eunshin Byon is with the Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109 (corresponding author; ebyon@umich.edu). Soham Vanage and Kristen Cetin are with the Department of Civil and Environmental Engineering at Michigan State University, East Lansing, MI 48824 (vanageso@msu.edu, cetinkri@msu.edu). David E. Jahn is with the Cooperative Institute for Severe and High-Impact Weather Research and Operations at the University of Oklahoma and the Strom Prediction Center of the NOAA/National Weather Service (david.jahn@noaa.gov). William Gallus, Jr is with the Department of Geological and Atmospheric Sciences at Iowa State University, Ames, IA 50011 (wgallus@iastate.edu). Lance Manuel is with the Department of Civil, Architectural and Environmental Engineering at The University of Texas at Austin, Austin, TX 78712 (lmanuel@mail.utexas.edu). This work was supported by the National Science Foundation (Grant No. CMMI-1662553, CMMI-1662691, CMMI-2013161, CMMI-1663044).

***Index Terms*—Heat wave events, Post-processing, Urban Canopy Model, Urban island effect, Weather Research and Forecasting Model**

Notation

S	Number of weather stations
T	Number of time points
s	Location index
t	Time index
s^*	Target location
t^*	Target time point
$y(s,t)$	Measured temperature at site s and time t
$x(s,t)$	WRF/UCM-predicted temperature at site s and time t
$\beta_i^{(F)}$	i^{th} fixed effect for $i=0,1$
$\beta_i^{(R)}(s,t)$	i^{th} random effect for $i=0,1$
ϵ_y	Random noise in $y(s,t)$
σ_y^2	Variance of ϵ_y
$\mathbf{Y}$	Vector of measured temperature
$\mathbf{X}^{(F)}$	Design matrix for the fixed effect
$\mathbf{1}$	Vector of ones
$\mathbf{x}$	Vector of WRF/UCM outputs
$\mathbf{I}_{ST}$	$ST \times ST$ Identity matrix
$\mathbf{D_x}$	Diagonal matrix whose diagonal elements are WRF/UCM temperature projections
$\epsilon_\mathbf{y}$	Vector of random noise in $y(s,t)$
$\mathcal{B}^{(F)}$	Vector of fixed effects
$\mathcal{B}_i^{(R)}$	Vector of i^{th} random effects for $i=0,1$
$\beta_i^{(T)}(t)$	i^{th} temporal random effect for $i=0,1$
$\beta_i^{(S)}(s)$	i^{th} spatial random effect for $i=0,1$
ρ_i	ith AR parameter for $i=0,1$
$\xi_i(t)$	Random noise in $\beta_i^{(T)}(t)$ for $i=0,1$
δ_i^2	Variance of $\xi_i(t)$ for $i=0,1$
$\mathcal{B}_i^{(T)}$	Vector of i^{th} temporal random effects
$\Sigma_i^{(T)}$	Covariance matrix of $\mathcal{B}_i^{(T)}$ for $i=0,1$
$\mathcal{B}_i^{(S)}$	Vector of i^{th} spatial random effects for $i=0,1$
$\Sigma_i^{(S)}$	Covariance matrix of $\mathcal{B}_i^{(S)}$ for $i=0,1$
L_s	Location of site s
L_{s_1}	Horizontal coordinate of site s
L_{s_2}	Vertical coordinate of site s
$\|\|\cdot\|\|_2$	2-norm for the Euclidean distance
τ_i^2	Variance of $\Sigma_i(S)$ for $i=0,1$
ℓ_i	Scale parameter of $\Sigma_i(S)$ for $i=0,1$
$\Sigma_{\mathbf{Y}\|\mathbf{x}}$	Covariance matrix of $\mathbf{Y}\|\mathbf{x}$
$\Sigma_\mathbf{x}$	Covariance matrix of $\mathbf{x}$
L	Likelihood of parameters
$\theta_\mathbf{x}$	Set of hyperparameters in the density $\mathbf{x}$
$\theta_{\mathbf{Y}\|\mathbf{x}}$	Set of hyperparameters in the density $\mathbf{Y}\|\mathbf{x}$
$\mathbf{d}_*$	Covariance vector at all pairs of the train and test points in $\mathbf{x}$

d_{**}	Variance at the test points in the WRF/UCM projection
$\mathbf{c}_*$	Covariance vector at all pairs of the train and test points in $\mathbf{Y}\|\mathbf{x}$
c_{**}	Variance at test points in $\mathbf{Y}\|\mathbf{x}$
$\mathbf{c}_{i,*}$	Covariance vector constructed from the training and test points for $\beta_i^{(R)}$ for $i = 0, 1$
$c_{i,**}$	Variance constructed from the test points for $\beta_i^{(R)}$ for $i = 0, 1$

I. Introduction

Ambient environmental condition affects many aspects of human life. The energy consumption from residential and commercial buildings is largely affected by weather conditions [1], [2]. To understand and forecast environmental conditions, physics-based models such as the Weather Research and Forecasting (WRF) model are often employed [3]. The WRF is a mesoscale numerical weather prediction (NWP) model that has been used to study climate change with regionally downscaled versions. However, these simulations do not completely account for local climate characteristics [4]. To better estimate localized weather conditions in urban areas, recent studies have coupled WRF with an Urban Canopy Model (UCM) that parameterizes urban canopy features such as building and street properties [3], [5]. The WRF model coupled with UCM is referred to as WRF/UCM in this study

This study is concerned with the spatiotemporal temperature characterization with the WRF/UCM model during heat wave events. The excessive electricity consumption during heat waves can result in power blackouts and brownouts, which can put occupants of the buildings at risk. Even worse, the so-called urban heat island effect produces additional heat stresses due to interactions between the built-environment and the atmosphere, causing higher temperatures in highly urbanized areas as compared to their surrounding areas [4].

Although WRF/UCM can provide more accurate localized weather profiles, we observe a substantial discrepancy between actual and WRF/UCM-predicted temperatures. Such discrepancies likely arise due to inappropriate settings in the WRF and UCM models, e.g, inaccurate initial and/or boundary conditions and model formulation in WRF, and/or incorrect parameterization in UCM [6], [7]. In particular, UCM employs several parameters such as land use, land cover information, percentage of impervious surface, building dimensions, and surface albedo. When information on these parameters is not available or is unknown, *educated guesses* from domain experts are often employed, but inappropriate assumptions can cause substantial deviations in model prediction outputs from actual climate conditions [7]. One remedy is to adjust the parameter values iteratively, but doing so can be challenging and cumbersome when a large number of such parameters is involved. Furthermore, running WRF/UCM with each new parameter setting is computationally demanding.

Interestingly, WRF/UCM outputs exhibit both *temporally- and spatially-varying discrepancy patterns*. To illustrate, Fig. 1 shows a network of ground-based weather stations around Austin, Texas, where each black circle represents the location of a weather station. Fig. 2 compares actual temperatures (black curves) at these locations with their corresponding WRF/UCM projections (red curves) during three heat wave events. A large discrepancy is observed during the daytime on these heat wave days. In particular, WRF/UCM tends to underestimate the temperature from the late afternoon to the evening hours. This is mainly because heat is stored in the thermal mass of buildings and asphalt paved roads, which causes the ambient temperatures to decrease more slowly. It appears that WRF/UCM cannot fully account for such local climate characteristics and thus, its projections exhibit the *temporally-varying* discrepancy.

Fig. 3 further shows discrepancy patterns on 8/8/2013 at three locations: ASTSL, ASTPK, and BDDHL identified in Fig. 1. As observed in Fig. 2, WRF/UCM outputs suggest a clear discrepancy pattern over the day at each location. The discrepancy patterns at two nearby locations, ASTSL and ASTPK, are more similar to each other but somewhat different from the pattern seen at BDDHL; this suggests a possible spatial correlation. In summary, WRF/UCM's prediction errors are not independent and identically distributed. Rather, they are unevenly distributed both spatially and temporally.

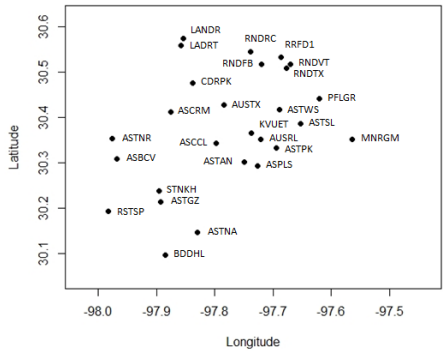


Fig. 1: Ground-based weather stations in Central Texas

We propose a new method that adjusts the unevenly distributed prediction error patterns both spatially and temporally for more accurate and reliable post-calibration of WRF/UCM. By utilizing information collected at geographically dispersed locations in both WRF/UCM projections and actual measurements, we aim to learn heterogeneous prediction error patterns in a collaborative manner. The main contribution of this study is three-fold: (a) our approach captures heterogeneous prediction error characteristics which are uneven across a day-long period and over space; (b) the proposed approach fully harnesses uncertainties in both WRF/UCM projections and spatiotemporal prediction errors in an integrative framework; (c) and we provide a post-calibrated predictive density in a closed form.

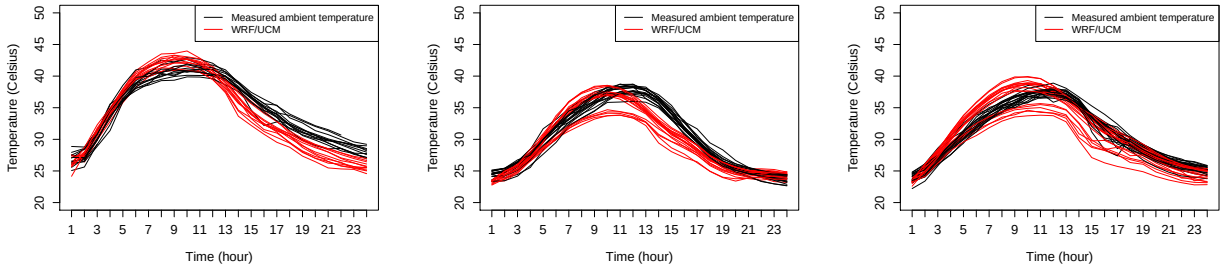
We implement the proposed method using numerical examples and conduct a case study. Our implementation results demonstrate that the proposed approach is capable of characterizing time- and space-varying discrepancy patterns. The advantage of the proposed approach is reflected by better point and probabilistic correction performance, compared to the WRF/UCM projections and alternative approaches. Based on the results, we estimate the building energy consumption over dispersed locations in the central Texas. We believe these results can be effectively used for building operations, such as demand response program during heat wave events [1], [8].

The remainder of this article is organized as follows. Section II reviews relevant studies. Section III introduces the proposed approach. Section IV presents numerical studies for evaluating the performance of the proposed approach, and Section V discusses the case study for predicting the building energy consumption. Finally, Section VI summarizes the paper and outlines future research plans.

II. Literature Review

In the literature, several approaches have been proposed to reduce the uncertainty of computer models. Broadly speaking, they can be grouped into two major categories: post-calibration and parameter calibration.

While this study focuses on the post-calibration of NWP models, it is worthwhile to review relevant studies for the parameter calibration of numerical computer models. Studies in this category aim to improve the accuracy of numerical models by identifying correct parameter values employed in computer models. In this approach, building upon the seminal study by Kennedy and O'Hagan [9], Bayesian calibration has been extensively studied and applied to various applications. Extending the basic framework of the calibration model presented in [9], Higdon *et al.* [10] devised a new approach for high-dimensional outputs. Chang *et al.* [11] developed a dimension-reduced climate model calibration method to handle high dimensional spatial data. Xie and Xu [12] proposed the Bayesian projected calibration to solve the calibration problem when a computer model is



(a) 8/28/2011 (b) 8/8/2013 (c) 7/24/2017

Fig. 2: Comparison between actual temperature and WRF/UCM prediction

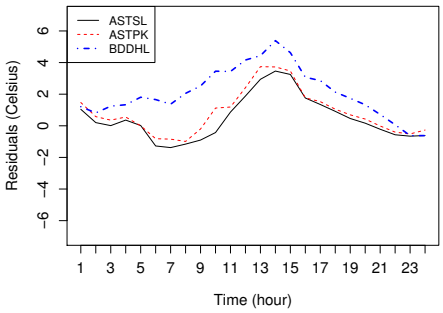


Fig. 3: WRF/UCM's prediction error patterns at three stations, ASTSL, ASTPK, and BDDHL on 8/8/2013

imperfect. Recently, focusing on minimizing the discrepancy between computer model outputs and physical system outputs, Tuo and Wu [13] presented the L_2 calibration approach in a frequentist framework. Liu *et al.* [14] and Jain *et al.* [15] present a stochastic optimization approach to calibrate computer model parameters in a big data setting.

While significant improvement has been made in the parameter calibration in the literature, most studies are limited to cases where the number of calibration parameters is small. When a large number of parameters are involved, as is the case of WRF/UCM model, and the computer model is expensive to run, parameter calibration is still challenging. Therefore, this study focuses on post-correcting the WRF/UCM projections. The post-calibration approach adjusts the physics-based forecasting model outputs by comparing them with observational data. This approach is also called post-processing or forecast calibration [16].

The most common model in the post-calibration approach is the linkage model that characterizes a relationship between actual weather observations and the corresponding NWP projections. Liu *et al.* [17] used multiple linear regression to establish the relationship between historical NWP and corresponding measured wind speed and correct the future NWP wind speed based on this relationship. In their study, the climate model prediction error is assumed to be a white noise process, implying that they do not make any spatiotemporal corrections to NWP outputs but simply make fixed location and scale adjustments. Similarly, Dupré *et al.* [18] used the linkage model to calibrate the prediction error in wind speed forecasts from NWP and predicted hourly and sub-hourly wind speed and wind power outputs. In these studies, the climate model's prediction error is assumed to be stationary, implying that the probabilistic relationship between numerical model outputs and measurements is assumed not to change over time and space. In other words, they adjust the prediction error in NWP projections by the same amount, regardless of the time of day and location.

Recently, Heredia *et al.* [19] extended the linkage model and formulated the additive discrepancy term with Gaussian process (GP) to accommodate a spatially-varying prediction error. Further, Tang *et al.* [20] proposed a numerical forecast correction method for the temperature and wind using a single-station and single-time spatial lightGBM approach. They decoupled the spatiotemporal correlation of the numerical forecast by modeling the forecasting model for each time interval and location, while the two nearest stations' observed data is added as input data for the lightGBM to reflect the spatial correlation. Du [21] proposed the ensemble-based model for the wind power forecasts by combining the WRF output and meteorological data from weather stations. However, these studies do not account for temporally nonstationary adjustments of the numerical forecast error. Clearly, this assumption appears to be unjustified, given what is observed in Fig. 2 and Fig. 3.

Another approach, called the delta change method, computes the average discrepancy in historical projections from NWP models and adds this to projections from NWP models [22], [23]. Berg *et al.* [22] employed the delta method to calibrate the temperature prediction error from a regional climate model (RCM). Similarly, Teutschbein and Seibert [23] corrected the prediction error in monthly mean temperature and precipitation predictions from an RCM. These studies do not address spatially- and temporally-varying discrepancy trends of NWP models. Recently, Jahani *et al.* [7] computed the average prediction error at each hour to accommodate a time-varying pattern, but they do not consider temporal correlation.

In summary, existing studies in the post-calibration of NWP models do not fully account for spatially and temporally varying prediction error patterns. Moreover, they mainly focus on predicting the response variable (weather conditions) at pre-specified locations where the NWP outputs are available. This study fills the gap in the literature in a few critical aspects. First, we present a new linkage model that adjusts nonstationary spatiotemporal NWP projection errors. Next, we provide predictions at spatially dispersed locations where NWP outputs are unavailable. Lastly, the prediction uncertainties, arising from both NWP output estimation and post-calibration models, are fully quantified through post-calibrated predictive densities in a closed-form.

III. Methodology

This section presents a post-calibration method to provide a post-corrected predicted density in a closed form.

A. Model Formulation

Let $y(s,t)$ and $x(s,t)$, $s = 1, \cdots, S$ and $t = 1, \cdots, T$, denote the measured and WRF/UCM-predicted temperature at site s and time t, respectively. In our analysis, we use hourly data in each day (thus, $T = 24$), and S denotes the number of weather stations.

The most common linkage model post-processes the NWP projection by assuming a linear relationship between $y(s,t)$ and its WRF/UCM prediction $x(s,t)$, modeled as

$$y(s,t) = \beta_0 + \beta_1 x(s,t) + \epsilon, \tag{1}$$

where β_0 and β_1 are the intercept and slope coefficients, respectively, representing additive (or location) and multiplicative (or scale) corrections, and ϵ is a random noise. The main issue with this model is that it assumes stationary prediction errors across time and space.

To reflect heterogeneous error patterns and capture spatially and temporally varying trends, we formulate the problem with a varying-coefficient model that allows the model coefficients to change [24], [25]. Specifically, we decompose each model coefficient into fixed as well as spatially- and temporally-varying components as follows.

$$\begin{aligned} & y(s,t) \\ & = \beta_0(s,t) + \beta_1(s,t)x(s,t) + \epsilon_y \\ & = \Big[\beta_0^{(F)} + \beta_0^{(R)}(s,t)\Big] + \Big[\beta_1^{(F)} + \beta_1^{(R)}(s,t)\Big]x(s,t) + \epsilon_y \\ & = \Big[\beta_0^{(F)} + \beta_1^{(F)}x(s,t)\Big] + \Big[\beta_0^{(R)}(s,t) + \beta_1^{(R)}(s,t)x(s,t)\Big] + \\ & \epsilon_y, \end{aligned} \tag{2}$$

where $x(s,t)$, $s = 1, \cdots, S$ and $t = 1, \cdots, T$, denotes the WRF/UCM-predicted temperature at site s and time t and ϵ_y represents random noise which is assumed to obey a Gaussian distribution with zero mean and variance σ_y^2.

The overall framework is illustrated in Fig. 4. Specifically, the intercept and slope coefficients are decomposed into fixed and random effects; $\beta_0^{(F)}$ and $\beta_1^{(F)}$ represent fixed effects, whereas $\beta_0^{(R)}(s,t)$ and $\beta_1^{(R)}(s,t)$ are random effects. The fixed effects are space- and time-invariant, representing the global (or common) adjustments of WRF/UCM projections in all sites and hours, whereas the random effects represent the space- and time-variant adjustments. Further, each random effect breaks down into two components: spatial and temporal random effects (this will be detailed in the subsequent discussion). The use of spatiotemporal random effects in both intercept and slope coefficients provides enhanced flexibility when the relationship between WRF/UCM and actual measurements is nonstationary.

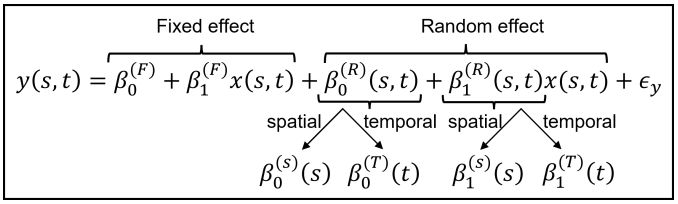


Fig. 4: Overall framework of the proposed model

Let us express the proposed model in a matrix form. Let $\mathbf{Y}$ denote the vector of measured temperature, $\mathbf{X}^{(F)} = [\mathbf{1}, \mathbf{x}] \in R^{ST\times 2}$ denote a design matrix for the fixed effect terms with $\mathbf{1} \in R^{ST\times 1}$ being a vector of ones and $\mathbf{x}$ representing a vector of WRF/UCM outputs, $\mathbf{I}_{ST}$ denote a $ST \times ST$ identity matrix, $\mathbf{D_x}$ denote the diagonal matrix whose diagonal elements are WRF/UCM projections, and $\epsilon_\mathbf{y}$ denote the vector of random noise.

The WRF provides the temperature field over the spatial grid in the study region. The UCM further introduces effects on the urban environment, including the albedo of roads and roof surfaces, the thermal conductivity of buildings, and vegetation fraction [26]. These UCM values are determined, depending on the urban category at each location, such as low-intensity residential, high-intensity residential, or commercial/industrial. In practice, it is difficult to set all UCM parameters in all sites over the entire region. In this study, assumptions are made with the WRF/UCM model applied in conjunction with environmental data only at discrete weather station sites so as to demonstrate the methodology. Then, to estimate the WRF/UCM outputs at other locations and account for the estimation uncertainty, we formulate the WRF/UCM output as a GP.

Combining the GP model for WRF/UCM projections with the post-calibration model in (2), we have

$$\mathbf{x} \sim \mathcal{GP} \tag{3}$$

$$\mathbf{Y}|\mathbf{x} = \mathbf{X}^{(F)}\mathcal{B}^{(F)} + \mathbf{I}_{ST}\mathcal{B}_0^{(R)} + \mathbf{D_x}\mathcal{B}_1^{(R)} + \epsilon_\mathbf{y}. \tag{4}$$

Here $\mathcal{B}^{(F)} = [\beta_0^{(F)}, \beta_1^{(F)}]^T \in R^{2\times 1}$ is a vector of the fixed effects, and $\mathcal{B}_0^{(R)}$ and $\mathcal{B}_1^{(R)}$ are $(ST \times 1)$-dimensional vectors of intercept random effects and slope random effects, respectively.

Next, we discuss how to formulate the spatially- and temporally-varying coefficients. We note that the discrepancy patterns at sites located close to each other and temporally adjacent points exhibit greater similarity, compared to those for widely separated locations and/or times. This implies that the model coefficient $\beta_i^{(R)}(s,t)$ should be similar to $\beta_i^{(R)}(s',t')$ for s close to s' and t close to t', for $i = 0, 1$. To reflect such temporal and spatial correlation, we formulate the site- and time-specific effects $\beta_i^{(R)}(s,t)$ as latent random processes and further decompose them into temporal and spatial random effects, denoted by $\beta_i^{(T)}(t)$ and $\beta_i^{(S)}(s)$, respectively (Fig. 4).

First, to describe the temporal correlation at consecutive time points, $\beta_i^{(T)}(t)$ should depend on $\beta_i^{(T)}(t-1)$. We employ an autoregressive (AR) process to capture temporal dependency as follows.

$$\beta_i^{(T)}(t) = \rho_i\beta_i^{(T)}(t-1) + \xi_i(t), \tag{5}$$

for $i = 0, 1$, where $|\rho_i| \leq 1$ represents the AR parameter, quantifying the dependence intensity between two consecutive time points, and $\xi_i(t) \sim N(0, \delta_i^2)$ is assumed to be Gaussian random noise. A higher-order AR process can be employed, however, we adopt AR(1) in this study due to its simplicity. This AR(1) formulation implies that the collection of temporal random effects, $\mathcal{B}_i^{(T)} = [\beta_i^{(T)}(1), \beta_i^{(T)}(2), \cdots, \beta_i^{(T)}(T)]^T$ for $i = 0, 1$, follows a multivariate Gaussian distribution as

$$\mathcal{B}_i^{(T)} \sim N(0, \Sigma_i^{(T)}),$$

where $\Sigma_i^{(T)}$ represents a $T \times T$ temporal covariance matrix [27] whose $(t,t')^{th}$ component is defined as

$$\Sigma_i^{(T)}(t,t') = \delta_i^2 \frac{\rho_i^{|t-t'|}}{1-\rho_i^2},$$

for $t, t' = 1, \cdots, T$. It implies that the correlation intensity gets stronger, leading to a more similar pattern when t and t' are adjacent.

Next, the temperature at closely spaced locations should be similar to one another, as observed in Fig. 3. In other words, each site-specific random effect $\beta_i^{(R)}(s)$ at location s should be similar (less similar) to $\beta_i^{(R)}(s')$ for a close (remote) location s' for $i = 0, 1$ [28], [29]. To address such spatial dependency across sites, we employ GP [19]. Due to its nonparametric nature, a GP can flexibly represent the spatial variation. For the collection of site-specific random effects, $\mathcal{B}_i^{(S)} = [\beta_i^{(S)}(1), \beta_i^{(S)}(2), \cdots, \beta_i^{(S)}(S)]^T$ for $i = 0, 1$, we have

$$\mathcal{B}_i^{(S)} \sim N(0, \Sigma_i^{(S)}),$$

where $\Sigma_i^{(S)}$ is an $S \times S$ covariance matrix. Among several possible functions, we use the exponential covariance function in this study. Thus, the $(s,s')^{th}$ element of $\Sigma_i^{(S)}$ is defined as

$$\Sigma_i^{(S)}(s,s') = \tau_i^2 \exp\Big\{-\frac{||L_s - L_{s'}||_2^2}{\ell_i}\Big\}, \tag{6}$$

for $i = 0, 1$ and $s, s' = 1, \cdots, S$, where L_s is the location of site s, $||\cdot||_2$ denotes the 2-norm for the Euclidean distance, and τ_i^2 and ℓ_i, respectively, represent the variance term and scale parameter.

Finally, we combine the temporal and spatial random effects to represent $\beta_i^{(R)}(s,t)$ in (2) for $i = 0, 1$. We consider the separable covariance structure in time and space with the additive structure $\beta_i^{(R)}(s,t) = \beta_i^{(S)}(s) + \beta_i^{(T)}(t)$, and its covariance function becomes

$$\Sigma_i^{(R)}((s,t),(s',t')) = \tau_i^2 \exp\Big\{-\frac{||L_s - L_{s'}||_2^2}{\ell_i}\Big\} + \delta_i^2 \frac{\rho_i^{|t-t'|}}{1-\rho_i^2}. \tag{7}$$

B. Parameter Estimation

To estimate the parameters in the proposed post-calibration model with latent spatial and temporal processes, we derive the joint density of $\mathbf{x}$ and $\mathbf{Y}|\mathbf{x}$ in (3) and (4). Following the property of GP, $\mathbf{x}$ obeys the multivariate normal desnity. Next, we derive the density of $\mathbf{Y}|\mathbf{x}$. Because the mean of the random effects is zero, the mean of $\mathbf{Y}$, given $\mathbf{x}$, becomes $X^{(F)}\mathcal{B}^{(F)}$. Recall that the covariance matrices of $\mathcal{B}_0^{(R)}$ and $\mathcal{B}_1^{(R)}$, denoted by $\Sigma_0^{(R)}$ and $\Sigma_1^{(R)}$, respectively, can be obtained from (7). Assuming the independence among $\mathcal{B}_0^{(R)}$, $\mathcal{B}_1^{(R)}$, and the random noise, the covariance of $\mathbf{Y}$, given $\mathbf{x}$, becomes

$$\Sigma_{Y|\mathbf{x}} = \Sigma_0^{(R)} + \mathbf{D}_\mathbf{x}\Sigma_1^{(R)}\mathbf{D}_\mathbf{x} + \sigma_y^2 I_{ST}. \tag{8}$$

As such, $\mathbf{Y}|\mathbf{x}$ follows a multivariate Gaussian distribution as

$$\mathbf{Y}|\mathbf{x} \sim N(\mathbf{X}^{(F)}\mathcal{B}^{(F)}, \Sigma_Y|\mathbf{x}). \tag{9}$$

With the covariance structure in (7), we introduce slightly modified forms to facilitate parameter estimation as follows.

$$K_0^{(R)}((s,t),(s',t')) = \exp\Big\{-\frac{||L_s - L_{s'}||_2^2}{\ell_0}\Big\} + g_0\frac{\rho_0^{|t-t'|}}{1-\rho_0^2},$$

with $g_0 = \delta_0^2/\tau_0^2$, and

$$K_1^{(R)}((s,t),(s',t')) = g_1\exp\Big\{-\frac{||L_s - L_{s'}||_2^2}{\ell_1}\Big\} + g_2\frac{\rho_1^{|t-t'|}}{1-\rho_1^2},$$

with $g_1 = \tau_1^2/\tau_0^2$ and $g_2 = \delta_1^2/\tau_0^2$. Then, $\Sigma_{Y|\mathbf{x}}$ in (8) becomes

$$\begin{aligned}\Sigma_{Y|\mathbf{x}} &= \Sigma_0^{(R)} + \mathbf{D}_\mathbf{x}\Sigma_1^{(R)}\mathbf{D}_\mathbf{x} + \sigma_y^2 I_{ST}\\ &= \tau_0^2(K_0^{(R)} + \mathbf{D}_\mathbf{x}K_1^{(R)}\mathbf{D}_\mathbf{x} + g_3 I_{ST})\\ &= \tau_0^2 K_Y,\end{aligned}$$

with $g_3 = \sigma_y^2/\tau_0^2$ and $K_Y = K_0^{(R)} + \mathbf{D}_\mathbf{x}K_1^{(R)}\mathbf{D}_\mathbf{x} + g_3 I_{ST}$.

Finally, we estimate the parameters and hyperparameters in the proposed model by maximizing the likelihood. The likelihood of parameters, given $(\mathbf{x},\mathbf{Y})$, is $L = L_\mathbf{x}\cdot L_{\mathbf{Y}|\mathbf{x}}$, with

$$L_\mathbf{x} = (2\pi)^{-ST/2}|\Sigma_\mathbf{x}|^{-1/2}\exp\Big\{-\frac{1}{2}\mathbf{x}^T\Sigma_\mathbf{x}^{-1}\mathbf{x}\Big\}, \tag{10}$$

and

$$\begin{aligned}L_{\mathbf{Y}|\mathbf{x}} &= (2\pi\tau_0^2)^{-ST/2}|K_Y|^{-1/2} \\ &\exp\Big\{-\frac{1}{2\tau_0^2}(\mathbf{Y}-\mathbf{X}^{(F)}\mathcal{B}^{(F)})^T K_Y^{-1}(\mathbf{Y}-\mathbf{X}^{(F)}\mathcal{B}^{(F)})\Big\}.\end{aligned} \tag{11}$$

Here, $\Sigma_\mathbf{x}$ is the spatial and temporal covariance matrix, similar to (7) but with different hyperparameters.

Let $\boldsymbol{\theta}_\mathbf{x}$ denote a set of hyperparameters in (10). Its log-likelihood is

$$\ell_\mathbf{x}(\boldsymbol{\theta}_\mathbf{x}) = -\frac{ST}{2}(2\pi) - \frac{1}{2}|\Sigma_\mathbf{x}| - \frac{1}{2}\mathbf{x}^T\Sigma_\mathbf{x}^{-1}\mathbf{x}.$$

Next, let $\boldsymbol{\theta}_y = \{l_0, l_1, \rho_0, \rho_1, g_0, g_1, g_2, g_3\}$ denote the set of hyperparameters in the density of $\mathbf{Y}|\mathbf{x}$. Using (11), we obtain its log-likelihood $\ell_{\mathbf{Y}|\mathbf{x}}(\mathcal{B}^{(F)}, \tau_0^2, \boldsymbol{\theta}_y)$ as

$$\begin{aligned}\ell_{\mathbf{Y}|\mathbf{x}}(\mathcal{B}^{(F)},\tau_0^2,\boldsymbol{\theta}_y) &= -\frac{ST}{2}\log(2\pi) - \frac{ST}{2}\log(\tau_0^2) - \\ &-\frac{1}{2}\log|K_Y| - \frac{1}{2\tau_0^2}(\mathbf{Y}-\mathbf{X}^{(F)}\mathcal{B}^{(F)})^T K_Y^{-1}(\mathbf{Y}-\mathbf{X}^{(F)}\mathcal{B}^{(F)}).\end{aligned}$$

Then with $L = L_\mathbf{x}\cdot L_{\mathbf{Y}|\mathbf{x}}$ the log-likelihood of all parameters is

$$\ell(\mathcal{B}^{(F)},\tau_0^2,\boldsymbol{\theta}_\mathbf{x},\boldsymbol{\theta}_y) = \ell_\mathbf{x}(\boldsymbol{\theta}_\mathbf{x}) + \ell_{\mathbf{Y}|\mathbf{x}}(\mathcal{B}^{(F)},\tau_0^2,\boldsymbol{\theta}_y)).$$

Therefore, we can estimate the parameters in $\boldsymbol{\theta}_\mathbf{x}$ separately from $\{\mathcal{B}^{(F)},\tau_0^2,\boldsymbol{\theta}_y\}$. We get the maximum likelihood estimator (MLE) of $\boldsymbol{\theta}_\mathbf{x}$ by maximizing $\ell_\mathbf{x}(\boldsymbol{\theta}_\mathbf{x})$. To get the MLEs of $\{\mathcal{B}^{(F)},\tau_0^2,\boldsymbol{\theta}_y\}$, we first take partial derivatives of $\ell_{\mathbf{Y}|\mathbf{x}}(\mathcal{B}^{(F)},\tau_0^2,\boldsymbol{\theta}_y)$ with respective to $\mathcal{B}^{(F)}$ and τ_0^2 and setting them to 0 to obtain

$$\hat{\mathcal{B}}^{(F)} = (\mathbf{X}^{(F)T}K_Y^{-1}\mathbf{X}^{(F)})^{-1}\mathbf{X}^{(F)T}K_Y^{-1}\mathbf{Y}, \tag{12}$$

and

$$\hat{\tau}_0^2 = \frac{1}{ST}(\mathbf{Y}-\mathbf{X}^{(F)}\hat{\mathcal{B}}^{(F)})^T K_Y^{-1}(\mathbf{Y}-\mathbf{X}^{(F)}\hat{\mathcal{B}}^{(F)}). \tag{13}$$

Plugging $\hat{\mathcal{B}}^{(F)}$ and $\hat{\tau}_0^2$ into (11), the log-likelihood function in (11) becomes a function of $\boldsymbol{\theta}_y$ as follows.

$$\begin{aligned}&\ell_{\mathbf{Y}|\mathbf{x}}(\boldsymbol{\theta}_y)\\ &= -\frac{ST}{2} - \frac{ST}{2}\log(2\pi) - \frac{1}{2}\log|K_Y| - \frac{ST}{2}\log\Big\{\frac{1}{ST}\\ &(\mathbf{Y}-\mathbf{X}^{(F)}(\mathbf{X}^{(F)T}K_Y^{-1}\mathbf{X}^{(F)})^{-1}\mathbf{X}^{(F)T}K_Y^{-1}\mathbf{Y})^T K_Y^{-1}\\ &(\mathbf{Y}-\mathbf{X}^{(F)}(\mathbf{X}^{(F)T}K_Y^{-1}\mathbf{X}^{(F)})^{-1}\mathbf{X}^{(F)T}K_Y^{-1}\mathbf{Y})\Big\}.\end{aligned} \tag{14}$$

Once we obtain the MLE of $\boldsymbol{\theta}_y$ that maximizes (14), we plug it into (12) and (13) and attain the MLEs of $\mathcal{B}^{(F)}$ and τ_0^2.

In our implementation, we use a numerical optimization, specifically, L-BFGS-B algorithm. We employ the 'optim' function in 'stats' package in the statistical software R [30].

C. Post-calibrated Predictive Distribution

We now derive the predictive distribution of temperature at location s^* and time t^*. We consider the uncertainties in both the WRF/UCM temperature field model and the spatiotemporal discrepancy adjustment model in obtaining the predictive density. First, following the properties in the GP, the predictive distribution of the WRF/UCM temperature projection x^* (or $x(s^*,t^*)$) is given by

$$x^*|\mathbf{x} \sim N(\mu_{x^*}, \sigma_{x^*}^2), \tag{15}$$

with

$$\mu_{x^*} = \mathbf{d}_*^T\Sigma_\mathbf{x}^{-1}\mathbf{x}, \tag{16}$$

$$\sigma_{x^*}^2 = d_{**} - \mathbf{d}_*^T\Sigma_\mathbf{x}^{-1}\mathbf{d}_* + \sigma_x^2, \tag{17}$$

where $\mathbf{d}_*$ is a covariance vector constructed by all pairs of the training and test points in the WRF/UCM projection and d_{**} is constructed by test points as

$$\begin{pmatrix}\mathbf{x}\\ x^*\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma_\mathbf{x} & \mathbf{d}_*\\ \mathbf{d}_*^T & d_{**}\end{pmatrix}\right). \tag{18}$$

Next, we obtain the predictive density of temperature y^* at x^*. It follows a Gaussian distribution as

$$y^*|x^* \sim N(\mu_{y^*|x^*}, \sigma_{y^*|x^*}^2), \tag{19}$$

with

$$\mu_{y^*|x^*} = \mathbf{X}^*\mathcal{B}^{(F)} + \mathbf{c}_*^T\Sigma_Y^{-1}(\mathbf{Y}-\mathbf{X}^{(F)}\mathcal{B}^{(F)}) \tag{20}$$

and

$$\sigma_{y^*|x^*}^2 = c_{**} - \mathbf{c}_*^T\Sigma_Y^{-1}\mathbf{c}_* + \sigma_y^2, \tag{21}$$

where $\mathbf{X}^* = [1, x^*]$ is the design matrix and $\mathbf{c}_*$ is a covariance vector whose components provide the covariance between the training and test points (i.e., between $y|x(s,t)$ and $y|x(s^*,t^*)$), defined as

$$\mathbf{c}_* = \mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}x^*, \tag{22}$$

and c_{**} is the variance at the test point as

$$c_{**} = c_{0,**} + (x^*)^2 c_{1,**}. \tag{23}$$

Here, $\mathbf{c}_{i,*}$ is the covariance vector constructed from the training and test points for $\beta_i^{(R)}$ for $i = 0, 1$, and $c_{i,**}$ is the variance term of test points as

$$\begin{pmatrix}\mathcal{B}_i^{(R)}\\ \beta_i^{(R)}(s^*,t^*)\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\Sigma_i^{(R)} & \mathbf{c}_{i,*}\\ \mathbf{c}_{i,*}^T & c_{i,**}\end{pmatrix}\right). \tag{24}$$

Combining the predictive density of x^* in (15)-(18) with the predictive density of $y^*|x^*$ in (19)-(24), the predictive distribution of y^* becomes

$$p(y^*) = \int p(y^*|x^*)p(x^*|\mathbf{x})dx^*.$$

Because both $x^*|\mathbf{x}$ and $y^*|x^*$ follow a Gaussian distribution, y^* also obeys a Gaussian distribution as

$$y^* = y(s^*, t^*) \sim N(\mu_{y^*}, \sigma^2_{y^*}),$$

where μ_{y^*} and $\sigma^2_{y^*}$, respectively, denote the mean and variance of y^* [31].

Next, we derive μ_{y^*} and $\sigma^2_{y^*}$ using the law of total expectation and the law of total variance, respectively, as follows.

$$\begin{aligned}
\mu_{y^*} &= E\Big[E[y^*|x^*]\Big] \\
&= E\Big[\beta_0^{(F)} + \beta_1^{(F)}x^* + \mathbf{c}_*^T\Sigma_Y^{-1}(\mathbf{Y} - \mathbf{X}^{(F)}\mathcal{B}^{(F)})\Big] \quad (25)\\
&= \beta_0^{(F)} + \beta_1^{(F)}\mu_{x^*} + (\mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}\mu_{x^*})^T\Sigma_Y^{-1} \\
&\quad (\mathbf{Y} - \mathbf{X}^{(F)}\mathcal{B}^{(F)}), \quad (26)
\end{aligned}$$

where (25) is obtained using (20), and (26) is due to (22). Further, using (20) and (21), we get

$$\begin{aligned}
\sigma^2_{y^*} &= E\Big[Var[y^*|x^*]\Big] + Var\Big[E[y^*|x^*]\Big] \\
&= E\Big[c_{**} + \sigma_y^2 - \mathbf{c}_*^T\Sigma_Y^{-1}\mathbf{c}_*\Big] + \quad (27)\\
&\quad Var\Big[\beta_0^{(F)} + \beta_1^{(F)}x^* + \mathbf{c}_*^T\Sigma_Y^{-1}(\mathbf{Y} - \mathbf{X}^{(F)}\mathcal{B}^{(F)})\Big].
\end{aligned}$$

Below we derive each term in (27). First, $E\Big[c_{**} + \sigma_y^2\Big]$ becomes

$$\begin{aligned}
E\Big[c_{**} + \sigma_y^2\Big] &= E\Big[c_{0,**} + (x^*)^2c_{1,**} + \sigma_y^2\Big] \quad (28)\\
&= c_{0,**} + E\Big[(x^*)^2\Big]c_{1,**} + \sigma_y^2 \\
&= c_{0,**} + (\mu_{x^*}^2 + \sigma_{x^*}^2)c_{1,**} + \sigma_y^2.
\end{aligned}$$

Here, (28) is due to (23). We also attain $E\Big[\mathbf{c}_*^T\Sigma_Y^{-1}\mathbf{c}_*\Big]$ as

$$\begin{aligned}
E\Big[\mathbf{c}_*^T\Sigma_Y^{-1}\mathbf{c}_*\Big] &= E\Big[(\mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}x^*)^T\Sigma_Y^{-1}(\mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}x^*)\Big] \\
&= (\mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}\mu_{x^*})^T\Sigma_Y^{-1}(\mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}\mu_{x^*}) \\
&\quad +\sigma_{x^*}^2 tr(\mathbf{c}_{1,*}^T\mathbf{D}_\mathbf{x}\Sigma_Y^{-1}\mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}),
\end{aligned}$$

where the last equation is obtained using the mean of a quadratic form. Finally, the last term of (27) becomes

$$\begin{aligned}
&Var\Big[\beta_0^{(F)} + \beta_1^{(F)}x^* + \mathbf{c}_*^T\Sigma_Y^{-1}(\mathbf{Y} - \mathbf{X}^{(F)}\mathcal{B}^{(F)})\Big] \\
&= Var\Big[\beta_0^{(F)} + \beta_1^{(F)}x^* + (\mathbf{c}_{0,*} + \mathbf{D}_\mathbf{x}\mathbf{c}_{1,*}x^*)^T \\
&\quad \Sigma_Y^{-1}(\mathbf{Y} - \mathbf{X}^{(F)}\mathcal{B}^{(F)})\Big] \\
&= \sigma_{x^*}^2\Big(\beta_1^{(F)} + \mathbf{c}_{1,*}^T\mathbf{D}_\mathbf{x}\Sigma_Y^{-1}(\mathbf{Y} - \mathbf{X}^{(F)}\mathcal{B}^{(F)})\Big)^2.
\end{aligned}$$

Note that in this integrative model, the uncertainty in predicting $x(s^*, t^*)$ is seamlessly incorporated into the post-calibration model. When $x(s^*, t^*)$ is available from the WRF/UCM model, the prediction of $x(s^*, t^*)$ becomes deterministic and thus, the predictive density of y^* is reduced to $p(y^*|x^*)$ defined in (19)-(24). When the WRF/UCM projection is not available at the target location s^*, the overall prediction uncertainty gets increased due to the inflated variance in (27) (mainly the second term of (27)).

IV. Numerical Examples

A. Alternative Approaches

We compare the performance of the proposed approach presented in Section III-A with the WRF/UCM projection and four other alternative approaches. The first alternative model is the original linkage model that assumes a homogeneous linear relationship in (1). This model does not account for spatially and temporally heterogeneous discrepancy patterns. Second, we consider the model presented in [19] that incorporates the spatial GP into the linkage model, which formulates the problem as

$$y(s,t) = \beta_1 x(s,t) + f(s) + \epsilon, \quad (29)$$

where $f(s)$ represents the spatial correction term modeled by GP. Thus, $f(s)$ is formulated as $f(s) \sim N(0, \Sigma)$, where Σ is the spatial covariance function similar to (6). The model in [19] does not include the slope parameter β_1, however, we add it here to enhance this model's prediction performance. While this model adjusts the spatially correlated error, it does not handle the temporal correlation.

The third alternative approach is the linkage model with temporal correction, similar to the delta change approach presented in [7]. To adjust the temporally changing discrepancies, it contains the time-varying intercept term as

$$y(s,t) = \beta_0(t) + x(s,t) + \epsilon,$$

and estimates $\beta_0(t)$ at each time point. The least-squares estimate of the intercept $\beta_0(t)$ becomes equivalent to the average of the difference between $x(s,t)$ and $y(s,t)$ at each time $t = 1, \cdots, T$. Because this model does not correct the multiplicative correction term (i.e., slope parameter), it is less flexible than the original linkage model in some sense. However, it adjusts the discrepancy of $x(s,t)$ at each hour to account for the temporally-varying prediction error. Still, the temporal correlation of prediction error is not taken into consideration, because it computes the temporal discrepancy separately in each hour. Thus, this approach addresses the temporal variation only partially. Moreover, it does not address spatial correlation.

Lastly, we consider the spatiotemporal error process model, which is one of the common approaches in handling the spatiotemproal data. This model lets the error term ϵ follow a spatiotemporal process [32] as

$$y(s,t) = \beta_0^{(F)} + \beta_1^{(F)}x(s,t) + \epsilon(s,t). \quad (30)$$

Suppose we use an additive structure of the spatial and temporal correlation for $\epsilon(s,t)$. The covariance function of $\epsilon(s,t)$ becomes $\Sigma_0 + \sigma^2 I$, where Σ_0 is similar to (7).

We would like to highlight the difference between the proposed post-calibration model and the error process model in (30). Without the random effect in the slope parameter (scale adjustment parameter) by setting $\beta_1^{(R)}(s,t) = 0$, the proposed model in (2) is reduced to

$$\begin{aligned}
y(s,t) &= \beta_0^{(F)} + \beta_1^{(F)}x(s,t) + \beta_0^{(R)}(s,t) + \epsilon \quad (31)\\
&= \beta_0^{(F)} + \beta_1^{(F)}x(s,t) + \epsilon'(s,t)
\end{aligned}$$

where $\epsilon'(s,t) = \beta_0^{(R)}(s,t) + \epsilon$. Here the covariance matrix of $\epsilon'(s,t)$ is $\Sigma_0^{(R)} + \sigma^2 I$, which is the same as $Cov(\epsilon(s,t))$ in the spatiotemporal error process model in (30). As such, the spatiotemporal error process is a special case of our proposed modeling approach without considering random effects in the scale adjustment parameter. It implies that the proposed model presents a more flexible representation than the spatiotemporal error process model.

Then, when does such added flexibility is useful? The spatiotemporal error process model assumes that the correlation pattern of discrepancy (more specifically, $\epsilon(s,t) = y(s,t) - \beta_0^{(F)} - \beta_1^{(F)}x(s,t)$) is stationary over space and time. However, when the pattern changes, it cannot capture nonstationarity. In contrast, the use of spatiotemporal processes for both location and scale adjustments in the proposed approach provides enhanced flexibility and allows us to characterize nonstationary correlation. Table I describes whether or not each

method accounts for spatially- and temporally-varying discrepancy patterns. Our approach provides a collaborative learning framework to seamlessly identify spatiotemporal nonstationary relationships between two data sources.

Additionally, we would like to emphasize that the spatiotemporal error process model and other aforementioned alternatives mainly focuses on the spatiotemporal prediction at the pre-specified locations only where $x(s,t)$ is available. In our problem context, it means that the WRF/UCM output should be provided at any site. Conversely, by modeling $x(s,t)$ with GP, our approach can provide predictions for locations where WRF/UCM outputs are not available. For the comparison purpose, in implementing the alternative approaches, we use the GP model for estimating $x(s,t)$ at testing sites.

We measure the prediction performance in terms of both point and probabilistic estimation capability using several metrics. For the point estimation performance, we use the root mean square error (RMSE),

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}\{y(s,t)-\hat{y}(s,t)\}^2}{T}},$$

for $s = 1,\cdots,S$ and $t = 1,\cdots,T$, where $\hat{y}(s,t)$ denotes the estimated temperature at site s at hour t. For evaluating the probabilistic performance, the continuous ranked probability score (CRPS) is employed [33], defined as

$$\begin{aligned} CRPS &= \frac{1}{T}\sum_{t=1}^{T}\hat{\sigma}(s,t)\Big[z(s,t)\Big\{2\Phi\Big(z(s,t)\Big)+1\Big\}+ \\ &\quad 2\phi\Big(z(s,t)\Big)-\frac{1}{\sqrt{\pi}}\Big], \end{aligned}$$

where $z(s,t) = (y(s,t)-\hat{y}(s,t))/\hat{\sigma}(s,t)$ is the standardized temperature at site s and time t and ϕ and Φ, respectively, denote the standard Gaussian probability density function and cumulative density function. A smaller value of CRPS indicates better performance. In addition, other probabilistic performance metrics, including the width of the prediction interval (PI) and PI coverage rate are also examined for assessing probabilistic prediction performance.

B. Problem Setup

We consider the following four numerical examples.

- Ex 1

$$\begin{aligned} x(s,t) &= \frac{2}{3}\exp(L_{s_1}+L_{s_2})(0.01+0.001t+0.0005t^2) \\ &\quad -4\sin(50t) \\ y(s,t) &= \Big(1.2+\frac{\cos(t)}{10}+2L_{s_1}^2-L_{s_2}^2\Big)x(s,t) \\ &\quad -\Big(1+\frac{\sin(t)}{10}+L_{s_1}+L_{s_2}\Big)+\epsilon_1, \end{aligned}$$

for $t = 1,2,\cdots,100$.

- Ex 2

$$\begin{aligned} x(s,t) &= L_{s_1}+L_{s_2}+\cos(50t) \\ y(s,t) &= \frac{t(L_{s_1}+L_{s_2})}{100}x(s,t)+\frac{t(L_{s_1}^2-2L_{s_2}^2)}{10}+\epsilon_2, \end{aligned}$$

for $t = 1,2,\cdots,30$.

- Ex 3

$$\begin{aligned} x(s,t) &= \frac{2}{3}\exp(L_{s_1}+L_{s_2})-\sin(t) \\ y(s,t) &= (L_{s_1}+L_{s_2})\exp\Big((L_{s_1}+L_{s_2}-0.3)x(s,t)\Big)-1 \\ &\quad +\epsilon_3, \end{aligned}$$

for $t = 1,2,\cdots,25$.

- Ex 4

$$\begin{aligned} x(s,t) &= \frac{2}{3}\exp(L_{s_1}+L_{s_2})-\sin(t) \\ y(s,t) &= 1.2x(s,t)-1+\epsilon_4, \end{aligned}$$

for $t = 1,2,\cdots,25$.

In these examples, we assume $\epsilon_i \sim N(0,0.1)$, $i = 1,2,\cdots,4$, and L_{s_1} and L_{s_2}, respectively, denote the vertical and horizontal coordinates of location s. In our implementation, we draw 10 random locations from $U(0,1)$ for Ex 1 to Ex 3 and 15 locations for Ex 4.

C. Implementation Results

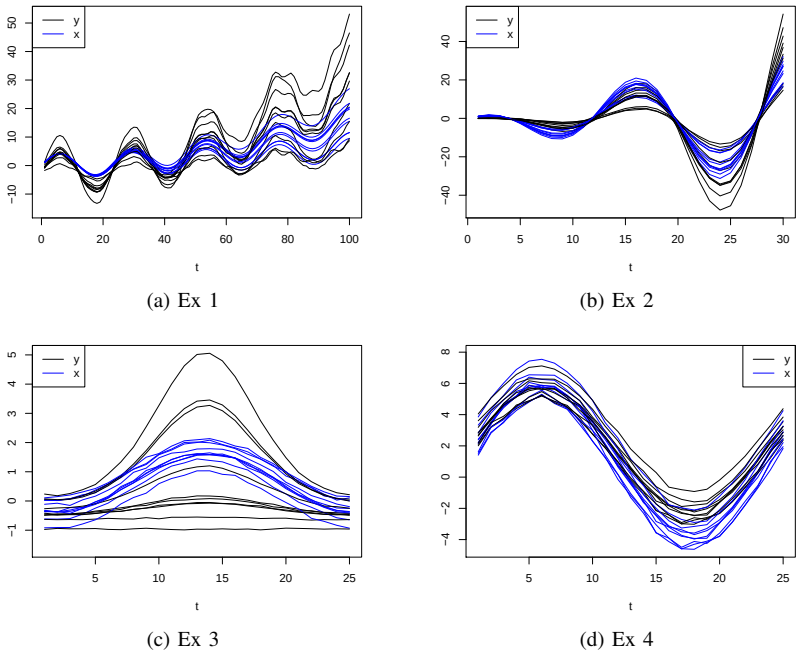
Fig. 5 depicts the generated data for each case, where each black and blue curve, respectively, represent real measurements $y(s,t)$ and WRF/UCM outputs $x(s,t)$ at their corresponding location s. Fig. 5(a)-Fig. 5(c) show that the spatial pattern of the discrepancy between $y(s,t)$ and $x(s,t)$ changes over time. In Ex 1, the location-to-location spatial variation in actual measurements gets more substantial as t increases. While the WRF/UCM outputs also exhibit growing spatial variations, it does not fully characterize the variation for large t and thus, the deviation becomes nonstationary. In Ex 2 and Ex 3, large spatial variations are observed at some time intervals. These three examples represent the nonstationary spatial and temporal correlation structure in the deviation between $y(s,t)$ and $x(s,t)$. In contrast, Fig. 5(d) does not exhibit any distinct nonstationary patterns.

The post-calibrated outputs from the alternative and proposed approaches in Ex 1 and Ex 4 are depicted in Fig. 6 and Fig. 7, respectively. In each subfigure, the red curve represents the calibrated prediction at each location. The proposed approach captures the spatiotemporal nonstationary pattern successfully, whereas all the alternative approaches fail to do so in Ex 1. In general, the alternative approaches tend to generate relatively parallel predictions over time, so they do not characterize the changing spatial variations. The linkage model with spatial GP seems to capture the spatiotemporal nonstationary pattern successfully (see Fig. 6(b)), but there are large deviations when t is small, resulting in the large value of RMSE (see Table II). On the other hand, for Ex 4 where the spatiotemporal pattern is stationary in both $x(s,t)$ and $y(s,t)$, Fig. 7 depicts similar prediction results from the original linkage model, the spatiotemporal error process model, and the proposed approach.

Table II summarizes the average RMSE and CRPS from the spatial leave-one-out tests, where each site is considered a testing site. In the first three examples, the proposed approach provides substantially smaller RMSE and CRPS over the original WRF/UCM and other alternative approaches, demonstrating that the proposed approach

TABLE I: Comparison of alternative methods (X: not addressed, △: partially addressed, ◯: fully addressed)

	Spatial prediction error pattern	Temporal prediction error pattern	Spatiotemporal stationary error pattern	Spatiotemporal nonstationary error pattern
Original linkage model	X	X	X	X
Linkage model with spatial GP	◯	X	X	X
Linkage model with temporal correction	X	△	X	X
Spatiotemporal error process model	◯	◯	◯	X
Proposed approach	◯	◯	◯	◯



(a) Ex 1 (b) Ex 2

(c) Ex 3 (d) Ex 4

Fig. 5: Simulated data in numerical examples

TABLE II: Average RMSE and CRPS over the testing sites from spatial out-of-sample testing using the simulated dataset (the values in parentheses are the standard deviation values at the testing sites)

Measure	Model	Ex 1	Ex 2	Ex 3	Ex 4
RMSE	WRF/UCM	4.82 (3.11)	4.33 (1.68)	0.94 (0.22)	0.43 (0.02)
	Original linkage model	5.04 (2.91)	6.50 (2.41)	0.94 (0.43)	**0.11 (0.02)**
	Linkage model with spatial GP	4.47 (1.31)	4.55 (1.11)	0.75 (0.34)	0.57 (0.13)
	Linkage model with temporal correction	4.94 (3.20)	5.03 (2.85)	0.85 (0.51)	0.16 (0.06)
	Spatiotemporal error process model	4.26 (2.60)	4.90 (2.94)	0.65 (0.36)	**0.11 (0.02)**
	Proposed approach	**2.60 (1.80)**	**2.69 (1.57)**	**0.40 (0.35)**	**0.11 (0.02)**
CRPS	WRF/UCM	-	-	-	-
	Original linkage model	3.09 (1.52)	3.70 (1.23)	0.57 (0.27)	**0.06 (0.01)**
	Linkage model with spatial GP	2.73 (0.82)	2.58 (0.58)	0.45 (0.23)	0.57 (0.13)
	Linkage model with temporal correction	2.87 (1.67)	2.67 (1.69)	0.44 (0.30)	0.10 (0.05)
	Spatiotemporal error process model	2.68 (1.57)	2.92 (1.83)	0.39 (0.22)	0.08 (0.02)
	Proposed approach	**1.82 (0.86)**	**1.81 (0.59)**	**0.21 (0.19)**	**0.06 (0.01)**

can successfully capture spatially and temporally nonstationary relationships between two data sources, as shown in Fig. 6. Among the alternative approaches, the spatiotemporal error process model generally improves the prediction accuracy over the simple adoption of WRF/UCM output (except Ex 2) and other alternatives, but the proposed approach presents a more flexible extension. Interestingly, in Ex 2, the RMSEs from all of the alternative approaches are even higher than that from WRF/UCM, demonstrating that the attempt to correct the discrepancy does not provide any improvement. However, the proposed approach leads to a considerable reduction in both RMSE and CRPS in Ex 2. In Ex 4, the original linkage model, the spatiotemporal error process model, and the proposed approach provide comparable performance; still, the proposed approach achieves a smaller CRPS value than the spatiotemporal error process model.

Overall, our proposed approach can flexibly characterize both stationary and nonstationary spatiotemporal discrepancy patterns and its merit becomes more obvious in nonstationary cases.

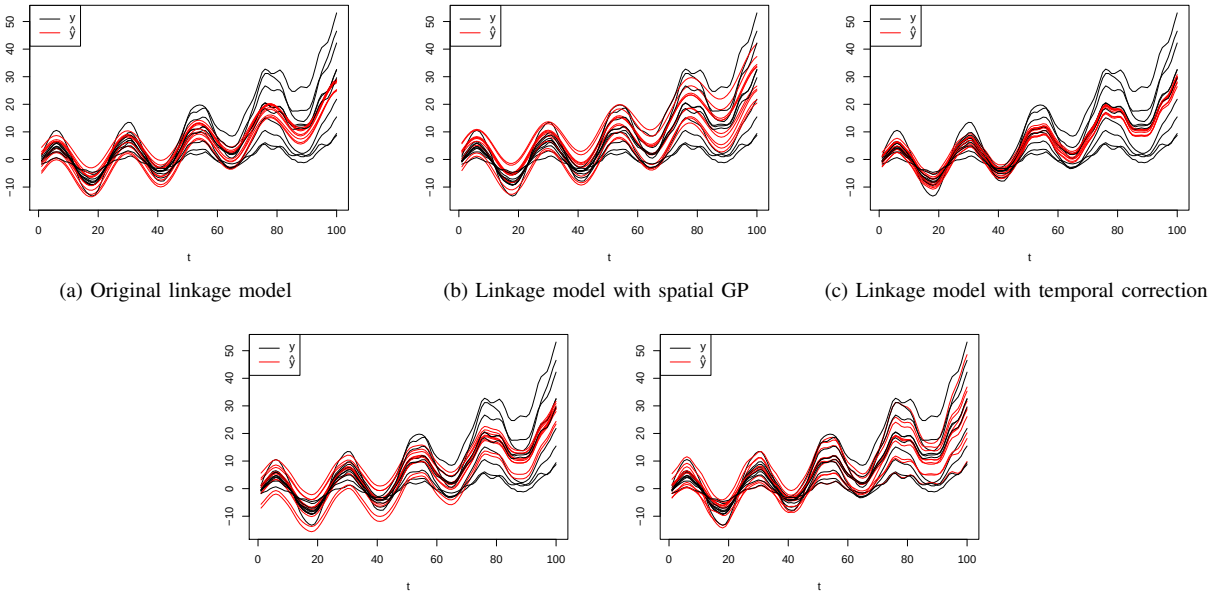
V. Case Study

We conduct the case study using real temperature measurements and examine how the temperature variation affects the building energy consumption in the central Texas region during the heat wave events.

A. Implementation Setting

We use hourly temperature measurements collected at multiple locations in Austin, Texas on 8/28/2011, 8/8/2013, and 7/24/2017, when heat wave events occurred as illustrated in Fig. 2. Due to missing data at some weather stations, we have a different number of measurements available for each heat wave event: temperature measurements from 12, 16, and 18 stations are available in 2011, 2013, and 2017, respectively. The distance between adjacent weather stations ranges from 1.16 km to 61.86 km. We also obtain WRF/UCM temperature projections on those heat wave days. The WRF model ver. 3.9 [34] coupled with an UCM [7] was used to simulate heat wave events along with urban island effects in Austin, Texas region. The initial and boundary conditions of the outer 36 km WRF domain were taken from Global Forecast System data provided by the NOAA National Centers for Environmental Information [35]. For a detailed description of the WRF/UCM model considered in this study, we refer the readers to the study in [7]. Finally, we use the predicted temperature to quantify the heterogeneous building energy consumption at multiple sites during the heat wave events.

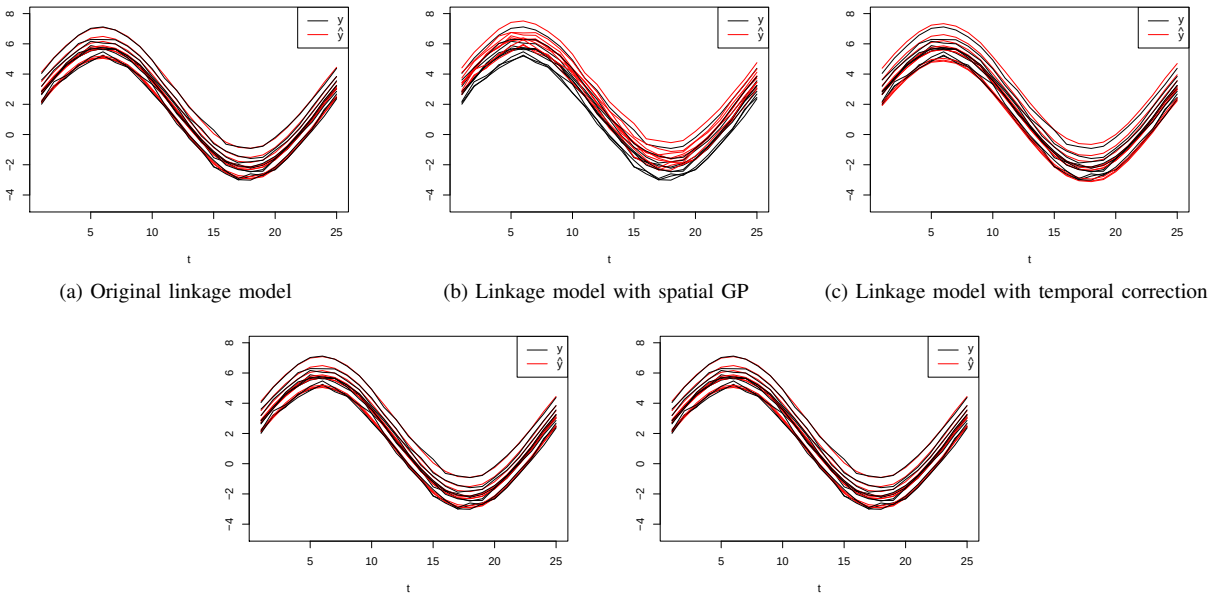
As the WRF/UCM's prediction error patterns are different in the three heat wave events as shown in Fig. 2, we apply the proposed procedure to each event separately. For evaluating the prediction



(a) Original linkage model (b) Linkage model with spatial GP (c) Linkage model with temporal correction

(d) Spatiotemporal error process model (e) Proposed approach

Fig. 6: Comparison of the prediction results results from the proposed and alternative approaches in Ex 1



(a) Original linkage model (b) Linkage model with spatial GP (c) Linkage model with temporal correction

(d) Spatiotemporal error process model (e) Proposed approach

Fig. 7: Comparison of the prediction results results from the proposed and alternative approaches in Ex 4

performance, we perform leave-one-out test. We divide the multiple sites into training and testing sites. The training set (in-sample) includes the temperature measurements at $S-1$ locations, whereas the testing set (out-of-sample) includes the measurements for 24 hours at the remaining one location. With the training set, we estimate the model parameters. Then the resulting post-calibration model is applied to predict the temperature at the testing site. This procedure is repeated S times so every site becomes the testing site.

B. Implementation Results

The proposed approach reduces the average RMSE of the WRF/UCM projection by about 60% from 1.87 to 0.77. In the next section, we provide additional details on these results. One notable advantage of the proposed model is that it can quantify different types of variability. The overall variability (i.e., the variance of the data) of actual temperature at multiple sites can be explained by temporal, spatial, and unstructured contributions. Table III shows that temporal variability explains more than 89% of the total variability. This is due to the diurnal temperature pattern over a day. While spatial variability appears to be less substantial, it still exceeds the unstructured variability from the random noise in two cases. This implies that accounting for spatial correlation can help improve estimation performance.

We further study the role of spatiotemporal random effects in the

intercept and slope parameters of our model. Table IV shows the average ratio of variability in $\beta_0^{(R)}$ and $\beta_1^{(R)}$, and the random noise. In the heat wave event cases of 2011 and 2017, most of the variability is explained by the random effect in the intercept coefficient $\beta_0^{(R)}$, whereas the random effect $\beta_1^{(R)}$ captures a greater portion of the variability in 2013. It demonstrates that including random effects in both intercept and slope coefficients provides greater flexibility in the model. In all cases, the two random effects explain greater than 90% of the variability in total, greatly exceeding the unstructured variability resulting from the random noise.

C. Comparison with Alternative Methods

Table V summarizes the average RMSE, CRPS, 95% PI width, and their coverage rate from the spatial leave-one-out tests, where each site is considered as a testing site. Clearly, our proposed approach provides noteworthy improvement over the original WRF/UCM and alternatives in terms of RMSE. The RMSE values for our proposed approach are substantially lower than those of the original WRF/UCM in all three heat wave events. The proposed approach also results in the lowest RMSE value, compared to other alternatives, except for 2017. The spatiotemporal error process model outperforms our approach in the 2017 case. We believe this is because most of the variability in 2017 is explained by $\beta_0^{(R)}$, as summarized in the last column of Table IV. The linkage model with temporal correction appears to adjust the original WRF/UCM projections relatively well, albeit its simple treatment for handling the heterogeneous discrepancy pattern. But it does not account for the temporal variation fully because it estimates the temporal discrepancy separately at each hour without considering its temporal correlation. Interestingly, the linkage model with spatial GP performs worse than the linkage model with temporal correction. We believe this is because the temporal variability accounts for a significant portion of the entire variability, ranging 89%-97%, as shown in the third row of Table III.

Fig. 8 compares the residuals at three selected stations, ASTSL, ASTPK, and BDDHL, from each approach. As we already observed in Fig. 3, the residuals from the original linkage model exhibit temporal and spatial patterns. The residual trends from the linkage model with spatial GP are similar to those of the original linkage model. This is because, as we discussed earlier, the temporal variation is much larger than the spatial variation. Another possible reason is that it does not account for the spatial correlation in the multiplicative adjustment β_1 in (29), but our result in Table IV indicates that the multiplicative term explains a significant portion of variability during the heat event day in 2013.

On the other hand, the linkage model with temporal correction generates smaller residuals than the aforementioned two alternative approaches, but the spatial correlation in the residuals is still present; see the similar patterns at ASTSL and ASTPK. Our approach does not generate any clear residual patterns and the magnitudes of residuals are small, demonstrating that much of the spatially and temporally correlated discrepancy is adjusted well.

In evaluating the probabilistic performance with CRPS, the proposed model performs better than the alternative approaches in most cases, eliciting strong probabilistic prediction capability. Note that the WRF/UCM provides deterministic projections, so its CRPS cannot be computed. Moreover, the proposed model provides the narrow PIs on the average, which indicates a greatly reduced prediction uncertainty. The ideal coverage rate for 95% PI is 95%, implying that 95% of actual measurements should be located within their 95% PIs. Our coverage rates are slightly lower than the linkage model with spatial GP; however, they are close to the desired rate of 95%. While the spatial GP model achieves the highest coverage rate on average, it does so with much greater PIs; its average PI width (9.47) is much wider than the proposed approach's (3.15).

To elaborate further, Fig. 9 compares actual and predicted temperatures and their prediction intervals at the site ASPLS in 2013 when ASPLS is a testing site. The black circle denotes the measured temperature and the blue curve represents the WRF/UCM temperature projections. The red solid and dashed curves represent the predicted temperature and their corresponding 95% PIs from each method, respectively. Fig. 9(a) shows that the original linkage model vertically shifts the WRF/UCM output by the nearly same amount, failing to capture large discrepancy between 11:00 and 16:00. This is because it does not take the temporally heterogeneous discrepancy pattern into consideration. In Fig. 9(b), the linkage model with spatial GP shows similar prediction patterns. Fig. 9(c) shows that the linkage model with temporal correction provides a better point correction performance, compared to the original linkage model, because it calibrates the WRF/UCM projection in each time point. However, there is still a discrepancy between the real and predicted temperature between 9:00 to noon. We believe that this is because it does not account for spatial variation but adjusts the WRF/UCM output by the same amount over multiple locations at each time.

In contrast, our approach provides better calibration performance as is evident in Fig. 9(e). Its point estimates align well with actual measurements. We also observe that the PI is narrower, exhibiting sharper estimation. Despite its narrow interval, our approach includes most measurements within the 95% PIs.

D. Application to Quantification of Building Energy Consumption

Using the results obtained from the proposed approach, we investigate how the building energy consumption varies at different places due to heterogeneous temperature caused by urban heat island effects. We use the building energy simulator, EnergyPlus (version 9.3.0) developed by the U.S. Department of Energy's National Renewable Energy Laboratory [36], [37]. EnergyPlus can simulate hourly energy consumptions of a building throughout a pre-specified time period. We employ a residential building obtained from a dataport of Pecan street research institute [38]. This building has a rectangular floor plan with a dimension of 14 m $\times$ 9 m and 10 m $\times$ 9 m on the first and second floors, respectively. For more detailed information about the simulation setting and building profile, please refer to [39].

Fig. 10 depicts heat maps of predicted building energy consumption during the three heat wave events in the south Texas region, which shows spatially heterogeneous building energy consumption patterns. From our model, ambient temperatures in the region are estimated to vary by $1^\circ C \sim 2^\circ C$ in the afternoon, which can be explained by the urban heat island effect. This difference is translated into energy consumption difference by 0.4 kWh $\sim$ 0.5 kWh in the

TABLE III: The average ratio (expressed in percent) of the spatial, temporal, and unstructured variabilities (the values in parentheses are standard deviations at the testing sites)

Type of variability	Formula	8/28/2011	8/8/2013	7/24/2017
Spatial	$\frac{\tau_0^2+[x(s,t)]^2\tau_1^2}{Var(y(s,t))}$	4.69 (1.58)	1.87 (0.73)	3.07 (1.12)
Temporal	$\frac{\frac{\delta_0^2}{1-\rho_0^2}+[x(s,t)]^2\frac{\delta_1^2}{1-\rho_1^2}}{Var(y(s,t))}$	89.21 (3.90)	96.68 (0.92)	95.75 (1.15)
Unstructured	$\frac{\sigma^2}{Var(y(s,t))}$	6.08 (3.00)	1.45 (0.43)	1.18 (0.19)

TABLE IV: The average ratio (expressed in percent) of the variability in random effects $\beta_0^{(R)}$ and $\beta_1^{(R)}$ and unstructured variability (the values in parentheses are the standard deviations at the testing sites)

Variability type	Formula	8/28/2011	8/8/2013	7/24/2017
$\beta_0^{(R)}$	$\frac{\tau_0^2+\frac{\delta_0^2}{1-\rho_0^2}}{Var(y(s,t))}$	63.61 (35.56)	44.65 (20.68)	96.84 (1.63)
$\beta_1^{(R)}$	$\frac{[x(s,t)]^2\left(\tau_1^2+\frac{\delta_0^2}{1-\rho_0^2}\right)}{Var(y(s,t))}$	30.30 (36.77)	53.90 (20.56)	1.98 (1.57)
Unstructured	$\frac{\sigma^2}{Var(y(s,t))}$	6.09 (3.00)	1.45 (0.43)	1.18 (0.19)

TABLE V: Average RMSE, CRPS, 95% PI width, and their coverage rate (%) over the testing sites from spatial out-of-sample testing (the values in parentheses are the standard deviation values at the testing sites, unit: celsius)

Measure	Model	8/28/2011	8/8/2013	7/24/2017	Average
RMSE	WRF/UCM	2.12 (0.47)	1.81 (0.55)	1.69 (0.55)	1.87 (0.52)
	Original linkage model	1.51 (0.21)	1.63 (0.43)	1.59 (0.51)	1.58 (0.38)
	Linkage model with spatial GP	1.69 (0.31)	1.61 (0.43)	1.71 (0.45)	1.67 (0.44)
	Linkage model with temporal correction	0.98 (0.21)	0.90 (0.52)	1.34 (0.66)	1.09 (0.55)
	Spatiotemporal error process model	0.91 (0.21)	0.69 (0.31)	**0.76 (0.32)**	0.79 (0.28)
	Proposed approach	**0.90 (0.13)**	**0.64 (0.33)**	0.80 (0.35)	**0.77 (0.33)**
CRPS	WRF/UCM	-	-	-	-
	Original linkage model	0.87 (0.18)	0.93 (0.25)	0.93 (0.32)	0.91 (0.25)
	Linkage model with spatial GP	0.98 (0.19)	0.92 (0.25)	1.00 (0.30)	0.97 (0.27)
	Linkage model with temporal correction	0.98 (0.21)	0.90 (0.52)	1.34 (0.66)	1.09 (0.55)
	Spatiotemporal error process model	0.54 (0.15)	0.43 (0.21)	**0.46 (0.20)**	0.47 (0.19)
	Proposed approach	**0.52 (0.09)**	**0.39 (0.22)**	0.49 (0.20)	**0.46 (0.21)**
95% PI width	WRF/UCM	-	-	-	-
	Original linkage model	5.94 (0.08)	6.52 (0.02)	6.41 (0.02)	6.29 (0.04)
	Linkage model with spatial GP	8.14 (0.23)	10.64 (0.48)	9.63 (1.38)	9.47 (0.70)
	Linkage model with temporal correction	4.35 (0.75)	4.13 (1.57)	5.79 (2.53)	4.83 (2.01)
	Spatiotemporal error process model	4.15 (0.48)	**2.68 (0.22)**	**3.00 (0.29)**	3.19 (0.33)
	Proposed approach	**3.49 (0.21)**	2.73 (0.26)	3.29 (0.61)	**3.15 (0.36)**
95% PI coverage rate	WRF/UCM	-	-	-	-
	Original linkage model	98.21 (2.69)	80.73 (6.78)	89.58 (17.11)	89.51 (8.86)
	Linkage model with spatial GP	**96.88 (4.40)**	**99.22 (3.12)**	**99.54 (1.35)**	**98.54 (1.45)**
	Linkage model with temporal correction	96.15 (6.24)	94.79 (10.15)	94.44 (17.68)	95.04 (12.64)
	Spatiotemporal error process model	96.53 (4.29)	91.67 (15.21)	90.74 (14.75)	92.57 (11.42)
	Proposed approach	95.14 (5.57)	92.97 (16.30)	92.59 (13.30)	93.39 (11.72)

considered building over geographically dispersed locations in the study region.

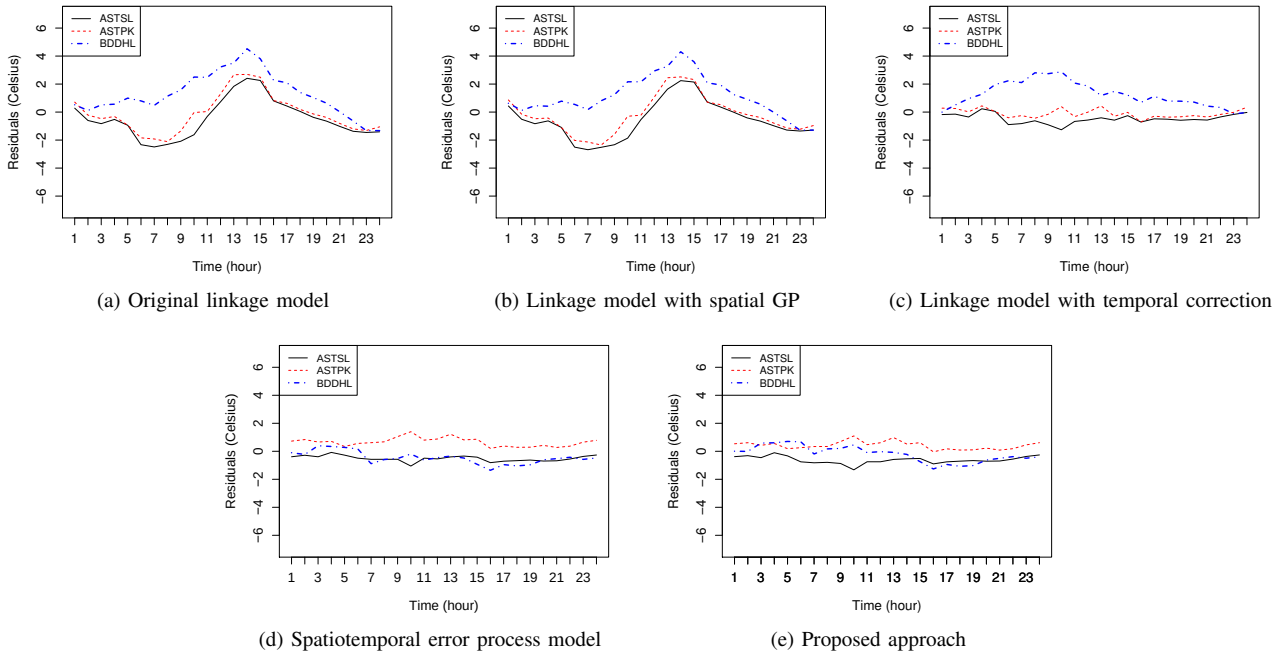
Interestingly, the spatial variation is different in three heat wave event days. During the heat wave events that occurred in 2011 and 2013, the energy consumption in the south-eastern part of the study region was greater than in other areas, whereas, in 2017, the central part showed more consumption. We believe that this is because the regional characteristics (e.g., population, building density) have rapidly changed over the years in this region. The population in the city of Austin grew by 30% over the 2010-2019 period with rapid urbanization [40], intensifying the urban heat island effects when the heat wave event occurred. This explains the higher energy consumption pattern in the central part of Fig. 10(c) and Fig. 10(f).

We believe these results offer useful insights into understanding of energy consumption and electricity demand variations, depending on urbanization intensity. Such studies can be useful for predicting electricity demand and effectively managing power system operations such as demand response programs [8].

VI. Conclusions and Future Plans

The NWP models can provide useful climate information, however, they inevitably exhibit spatiotemporal discrepancy patterns because of incomplete characterization of local or regional variations. This study presents a new probabilistic post-calibration approach that takes the spatially and temporally heterogeneous discrepancy into consideration. We formulate the post-calibration model with random effect coefficients that vary over space and time. A case study using actual temperature data and associated temperature predictions from WRF/UCM demonstrates that the proposed framework can successfully yield reliable solutions for limiting the spatially- and temporarily-varying prediction errors presented in WRF/UCM, while substantially reducing the average RMSE by about 60% from $1.87^\circ C$ to $0.77^\circ C$. Our approach also provides accurate and reliable probabilistic prediction performance, outperforming the alternative approaches in most performance criteria evaluated in this study, including CRPS, PI width, and coverage rates. We present how the spatially- and temporally-varying temperature affects the building energy consumption using the building energy simulator.

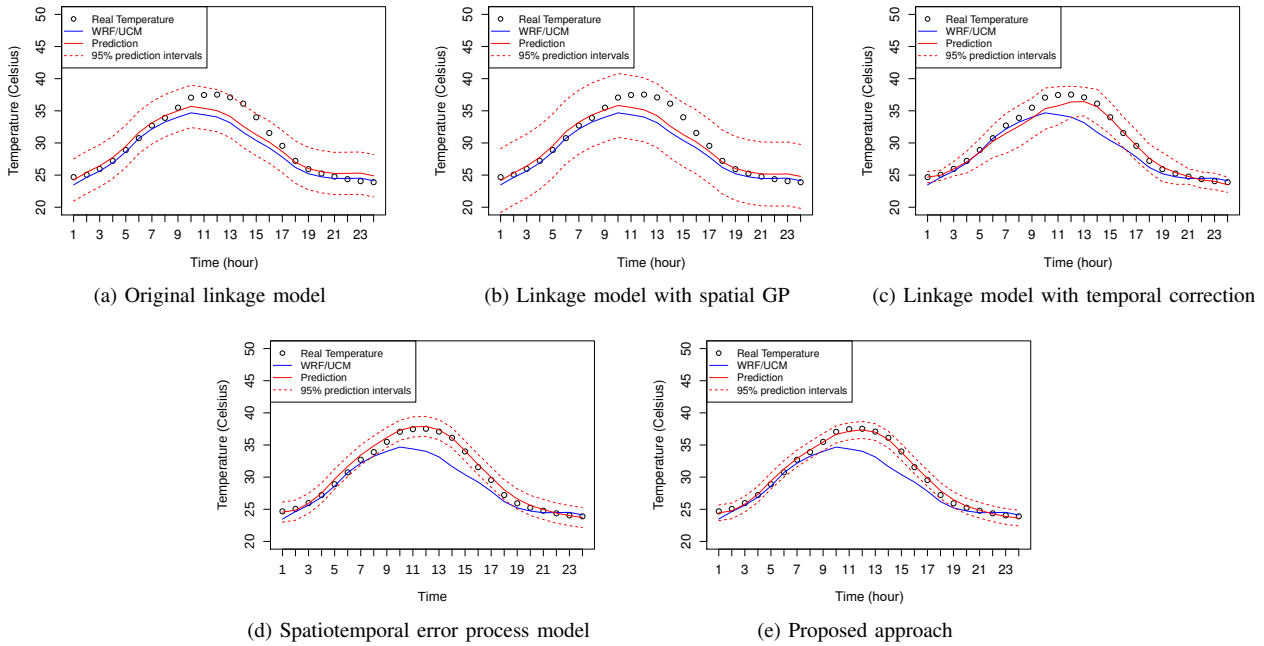
In the future, we plan to evaluate temporal prediction performance with larger datasets. In this study, due to the limited WRF/UCM data available to us (i.e., 1 day of data for each heat wave event), we can only apply our approach to the spatial out-of-sample evaluation, but we intend to extend the procedure to provide day(s)-ahead prediction by adaptively learning discrepancy patterns [41]. Additionally, we plan to investigate an alternative method for modeling WRF/UCM using a multi-fidelity approach with WRF being a low fidelity model and UCM being a high fidelity model and to integrate the resulting multi-fidelity model into the proposed post-calibration model. We will also incorporate other factors, such as seasonal effects and socio-economic factors, in our future work. Lastly, the extension of this work for electricity usage prediction and decision-making in building



(a) Original linkage model (b) Linkage model with spatial GP (c) Linkage model with temporal correction

(d) Spatiotemporal error process model (e) Proposed approach

Fig. 8: Residuals in the alternative and proposed models at three stations, ASTSL, ASTPK, and BDDHL on 8/8/2013



(a) Original linkage model (b) Linkage model with spatial GP (c) Linkage model with temporal correction

(d) Spatiotemporal error process model (e) Proposed approach

Fig. 9: Comparison of measured and predicted temperatures and their prediction intervals at the testing site ASPLS in 2013 from spatial out-of-sample estimation

energy systems and power system operations remains a subject for future study [42], [43].

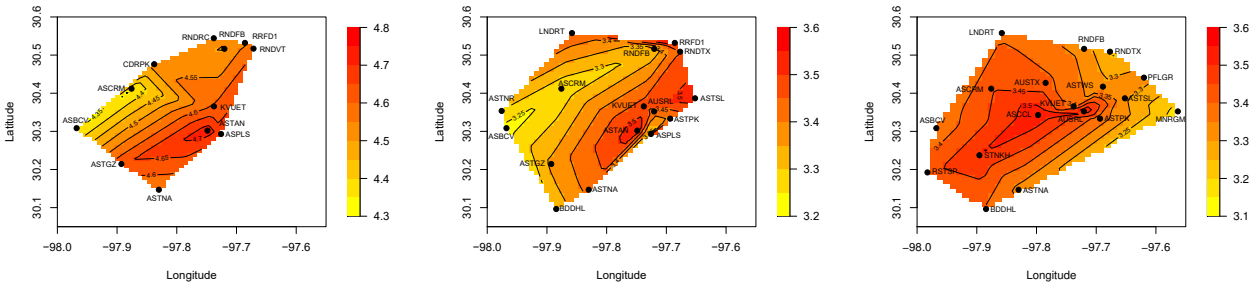
REFERENCES

[1] M. P. Fanti, A. M. Mangini, M. Roccotelli, and W. Ukovich, "A district energy management based on thermal comfort satisfaction and real-time power balancing," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 4, pp. 1271-1284, 2015.

[2] S. Ning, E. Byon, T. Wu, and J. Li, "A sparse partitioned-regression model for nonlinear system environment interactions," *IISE Trans.*, vol. 49, no. 8, pp. 814-826, 2017.

[3] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, "A description of the advanced research WRF version 2," National Center For Atmospheric Research, Tech. Rep., 2005.

[4] D. E. Jahn, W. A. Gallus, P. T. T. Nguyen, Q. Pan, K. Cetin, E. Byon, L. Manuel, Y. Zhou, and E. Jahani, "Projecting the most likely annual urban heat extremes in the central United States," *Atmosphere*, vol. 10, no. 12, 2019.



(a) 12:00, 2011 (b) 12:00, 2013 (c) 12:00, 2017

(d) 17:00, 2011 (e) 17:00, 2013 (f) 17:00, 2017

Fig. 10: Heat maps for the building energy consumption (kWh) simulated by EnergyPlus using the temperatures predicted by the proposed model at 12:00 and 17:00 for three heat wave events in Austin, Texas region (the darker color is, the higher energy consumption is; each curve shows the contour plot of the same energy consumption.)

[5] B. Bueno, M. Roth, L. Norford, and R. Li, "Computationally efficient prediction of canopy level urban air temperature at the neighbourhood scale," *Urban Clim.*, vol. 9, pp. 35-53, 2014.

[6] T. N. Krishnamurti, J. Sanjay, A. K. Mitra, and T. S. V. Vijaya Kumar, "Determination of forecast errors arising from different components of model physics and dynamics," *Mon. Weather Rev.*, vol. 132, no. 11, pp. 2570-2594, 2004.

[7] E. Jahani, S. Vanage, D. Jahn, W. Gallus, and K. S. Cetin, "City-scale energy modeling to assess impacts of extreme heat on electricity consumption and production using WRF-UCM modeling with bias correction," in *Canadian Society of Civil Engineers Annual Conference*, 2019.

[8] Y. Jang, E. Byon, E. Jahani, and K. Cetin, "On the long-term density prediction of peak electricity load with demand side management in buildings," *Energy Build.*, vol. 228, p. 110450, 2020.

[9] M. C. Kennedy and A. OHagan, "Bayesian calibration of computer models," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 63, no. 3, pp. 425-464, 2001.

[10] D. Higdon, J. Gattiker, B. Williams, and M. Rightley, "Computer model calibration using high-dimensional output," *J. Am. Stat. Assoc.*, vol. 103, no. 482, pp. 570-583, 2008.

[11] W. Chang, M. Haran, R. Olson, and K. Keller, "Fast dimension-reduced climate model calibration and the effect of data aggregation," *Ann. Appl. Stat.*, vol. 8, no. 2, pp. 649-673, 2014.

[12] F. Xie and Y. Xu, "Bayesian projected calibration of computer models," *J. Am. Stat. Assoc.*, vol. 116, no. 536, pp. 1965-1982, 2021.

[13] R. Tuo and J. C. Wu, "A theoretical framework for calibration in computer models: parameterization, estimation and convergence properties," *SIAM-ASA J. Uncertain.*, vol. 4, no. 1, pp. 767-795, 2016.

[14] B. Liu, X. Yue, E. Byon, and R. A. Kontar, "Parameter calibration in wake effect simulation model with stochastic gradient descent and stratified sampling," *Ann. Appl. Stat.*, vol. 16, no. 3, pp. 1795-1821, 2022.

[15] P. Jain, S. Shashaani, and E. Byon, "Wake effect calibration in wind power systems with adaptive sampling based optimization" In *Proceedings of the IIE Annual Conference*, pp. 43-48, 2021.

[16] T. Nipen and R. Stull, "Calibrating probabilistic forecasts from an NWP ensemble," *Tellus A: Dyn. Meteorol. Oceanogr.*, vol. 63, no. 5, pp. 858-875, 2011.

[17] Y. Liu, Y. Wang, L. Li, S. Han, and D. Infield, "Numerical weather prediction wind correction methods and its impact on computational fluid dynamics based wind power forecasting," *J. Renew. Sustain. Energy*, vol. 8, no. 3, pp. 770-778, 2016.

[18] A. Dupré, P. Drobinski, B. Alonzo, J. Badosa, C. Briard, and R. Plougonven, "Sub-hourly forecasting of wind speed and wind energy," *Renew. Energy*, vol. 145, pp. 2373-2379, 2020.

[19] M. B. Heredia, C. Junquas, C. Prieur, and T. Condom, "New statistical methods for precipitation bias correction applied to WRF model simulations in the Antisana region, Ecuador," *J. Hydrometeorol.*, vol. 19, no. 12, pp. 2021-2040, 2018.

[20] R. Tang, Y. Ning, C. Li, W. Feng, Y. Chen, and X. Xie, "Numerical forecast correction of temperature and wind using a single-station single-time spatial lightGBM method," *Sensors*, vol. 22, no. 1, p. 193, 2022.

[21] P. Du, "Ensemble machine learning-based wind forecasting to combine NWP output with data from weather station," *IEEE Trans. Sustain. Energy*, vol. 10, no. 4, pp. 2133-2141, 2018.

[22] P. Berg, H. Feldmann, and H.-J. Panitz, "Bias correction of high resolution regional climate model data," *J. Hydrol.*, vol. 448, pp. 80-92, 2012.

[23] C. Teutschbein and J. Seibert, "Bias correction of regional climate model simulations for hydrological climate-change impact studies: Review and evaluation of different methods," *J. Hydrol.*, vol. 456, pp. 12-29, 2012.

[24] D. Wang, K. Liu, and X. Zhang, "Spatiotemporal thermal field modeling using partial differential equations with time-varying parameters," *IEEE Trans. Autom. Sci. Eng.*, vol. 17, no. 2, pp. 646-657, 2020.

[25] Y. Jang and E. Byon, "Probabilistic characterization of wind diurnal variability for wind resource assessment," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2535-2544, 2020.

[26] P. Vahmani and G. A. Ban-Weiss, "Impact of remotely sensed albedo and vegetation fraction on simulation of urban climate in WRF-urban canopy model: A case study of the urban heat island in Los Angeles," *J. Geophys. Res. Atmos.*, vol. 121, no. 4, pp. 1511-1531, 2016.

[27] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control.* John Wiley & Sons, 2015.

[28] M. You, E. Byon, J. J. Jin, and G. Lee, "When wind travels through turbines: A new statistical approach for characterizing heterogeneous wake effects in multi-turbine wind farms," *IISE Trans.*, vol. 49, no. 1, pp. 84-95, 2017.

[29] M. You, B. Liu, E. Byon, S. Huang, and J. Jin, "Direction-dependent power curve modeling for multiple interacting wind turbines," *IEEE Trans. Power Syst.*, vol. 33, no. 2, pp. 1725-1733, 2018.

[30] R Core Team, *R: A Language and Environment for Statistical Com-*

puting, R Foundation for Statistical Computing, Vienna, Austria, 2014. [Online]. Available: http://www.R-project.org/
[31] C. M. Bishop, *Pattern recognition and machine learning*. Springer, 2006.
[32] N. Cressie and C. K. Wikle, *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.
[33] T. Gneiting, A. E. Raftery, A. H. Westveld III, and T. Goldman, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation," *Mon. Weather Rev.*, vol. 133, no. 5, pp. 1098-1118, 2005.
[34] W. C. Skamarock, J. B. Klemp, J. Dudhia, D. O. Gill, D. M. Barker, W. Wang, and J. G. Powers, "A description of the Advanced Research WRF version 3. NCAR Technical note -475+STR," 2008.
[35] National Oceanic and Atmospheric Administration, "Global Forecast System," https://www.ncei.noaa.gov/products/weather-climate-models/global-forecast, 2021, accessed: 2021-01-30.
[36] U.S. Department Energy, National Renewable Energy Laboratory, "EneryPlus essentials," Tech. Rep., 2019.
[37] S. Li, K. Deng, and M. Zhou, "Sensitivity analysis for building energy simulation model calibration via algorithmic differentiation," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 2, pp. 905-914, 2016.
[38] Pecan Street, "Dataport," https://www.pecanstreet.org/dataport/, 2021, accessed: 2021-01-30.
[39] E. Jahani, S. Vanage, and K. Cetin, "City-scale high-resolution WRF-UCM urban weather predictions compared to a dense network of ground-based weather station data for assessment of urban building energy consumption," in *Proc. Annu. Conf. - Can. Soc. Civ. Eng.*, 2019.
[40] United States Census Bureau, "Population and housing unit estimates datasets," https://www-census-gov.proxy.lib.umich.edu/ programs-surveys/popest/data/data-sets.html, 2021.
[41] E. Byon, Y. Choe, and N. Yampikulsakul, "Adaptive learning in time-variant processes with application to wind power systems," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 2, pp. 997-1007, 2016.
[42] D. Li, C. Menassa, V. R. Kamat, E. Byon, "HEAT–Human Embodied Autonomous Thermostat," *Build. Environ.*, vol. 178, p. 106879, 2020.
[43] J. Wang, S. Chung, A. AlShelahi, R. Kontar, E. Byon, and R. Saigal, "Look-ahead decision making for renewable energy: A dynamic predict and store approach," *Appl. Energy*, vol. 296, p. 117068, 2021.

Youngchan Jang received a Ph.D. degree in Industrial and Operations Engineering from the University of Michigan in 2021. He is a researcher with the Precedent Study Team for C4ISR System at the Korea Research Institute for defense Technology planning and advancement. His research interests include spatio-temporal data analytics, and quality and reliability engineering.

Eunshin Byon received the Ph.D. degree in Industrial and Systems Engineering from Texas A&M University in 2010. She is an Associate Professor with the Industrial and Operations Engineering Department at the University of Michigan. Her research interests include optimizing operations and management, data science, and simulation optimization. She is a member of IIE, INFORMS, and IEEE.

Soham Vanage is a Ph.D. student in the Civil Engineering department at Michigan State University. He completed his masters in Energy System Engineering from Iowa State University in Fall 2019 with a focus on building energy systems. His areas of interest are energy and daylighting modeling and validation and integration of buildings with the grid using demand response technologies.

Kristen Cetin is an Associate Professor in the Department of Civil and Environmental Engineering at Michigan State University. She is also the Director of the MSU Industrial Assessment Center. Her research focuses on the investigation of the performance of buildings, building systems, and other infrastructure and the development of solutions to improve the energy performance, efficiency, and energy demand contributions of the built environment.

David Jahn is a Research Scientist with the Cooperative Institute for Severe and High-Impact Weather Research and Operations at the University of Oklahoma and the Storm Prediction Center of the NOAA/National Weather Service. His Ph.D. from Iowa State University in 2016 was in Meteorology and Wind Energy Science Engineering and Policy. His current research interests include numerical weather prediction, severe weather calibrated forecast guidance, and forecast verification.

William Gallus received a Ph.D. in Atmospheric Sciences from Colorado State University in 1993. He is a professor of meteorology in the Department of Geological and Atmospheric Sciences at Iowa State University, and has published over 110 refereed articles primarily related to numerical weather prediction of thunderstorm systems, winds at turbine height, and climate change impacts on mesoscale phenomena.

Lance Manuel earned a Ph.D. degree in Civil Engineering from Stanford University in 1993. He is the Texas Atomic Energy Research Foundation Professor of Engineering in the Department of Civil, Architectural and Environmental Engineering at The University of Texas at Austin. His research deals with uncertainty quantification and with atmospheric/ocean climate extremes and associated loads on wind turbines and wave energy converters. He is a Fellow of AAAS, ASCE, and ASME.