Using a Bayesian Regression Approach on Dual-Model Windstorm Simulations to Improve Wind Speed Prediction

JAEMO YANG, MARINA ASTITHA, AND EMMANOUIL N. ANAGNOSTOU

Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

BRIAN M. HARTMAN

Department of Statistics, Brigham Young University, Provo, Utah

(Manuscript received 3 June 2016, in final form 1 February 2017)

ABSTRACT

Weather prediction accuracy is very important given the devastating effects of extreme-weather events in recent years. Numerical weather prediction systems are used to build strategies to prevent catastrophic losses of human lives and the environment and have evolved with the use of multimodel or single-model ensembles and data-assimilation techniques in an attempt to improve the forecast skill. These techniques require increased computational power (thousands of CPUs) because of the number of model simulations and ingestion of observational data from a wide variety of sources. In this study, the combination of predictions from two state-of-the-science atmospheric models [WRF and RAMS/Integrated Community Limited Area Modeling System (ICLAMS)] using Bayesian and simple linear regression techniques is examined, and wind speed prediction for the northeastern United States is improved using regression techniques. Retrospective simulations of 17 storms that affected the northeastern United States during the period 2004–13 are performed and utilized. Optimal variances are estimated for the 13 training storms by minimizing the root-mean-square error and are applied to four out-of-sample storms [Hurricane Irene (2011), Hurricane Sandy (2012), a November 2012 winter storm, and a February 2013 blizzard]. The results show a 20%–30% improvement in the systematic and random error of 10-m wind speed over all stations and storms, using various storm combinations for the training dataset. This study indicates that 10–13 storms in the training dataset are sufficient to reduce the errors in the prediction, and a selection that is based on occurrence (chronological sequence) is also considered to be efficient.

1. Introduction

Weather forecasting, applied to global and regional scales, has evolved with the use of multimodel or single-model ensembles (Doblas-Reyes et al. 2005; Palmer et al. 2008; Weigel et al. 2009; Kirtman et al. 2014), data-assimilation techniques (Barker et al. 2012; Wang et al. 2013; Ancell et al. 2015), and high-resolution grid spacing (Roberts 2003; Speer et al. 2003; Steppeler et al. 2003; Gego et al. 2005; Schwartz et al. 2009) in an attempt to improve the forecast skill. Despite the noted

improvements, inaccuracies caused by random and systematic errors are a continuous topic for research (Krishnamurti et al. 2004; Mass et al. 2008; Ancell et al. 2011; Ancell 2012; Delle Monache et al. 2011). The ability of numerical weather prediction (NWP) models to accurately describe atmospheric conditions under various dynamic states is influenced by errors stemming from the implemented physical parameterizations, initial state, boundary conditions, and data availability. Atmospheric complexity and inability to handle subgrid-scale phenomena also cause errors in the predicted meteorological variables (Libonati et al. 2008; Louka et al. 2008; Idowu and Rautenbach 2009). Restrictions in the resolution cause an imperfect representation of the actual surface properties (e.g., topography, vegetation, soil types and moisture, and sea surface temperature) that can result in significant model error along the sharp gradients. In addition, inaccurate

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JAMC-D-16-0206.s1>.

Corresponding author e-mail: Marina Astitha, astitha@engr.uconn.edu

DOI: 10.1175/JAMC-D-16-0206.1

© 2017 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

prediction of land surface interactions can be disadvantageous to the NWP (Koster and Suarez 2001; Drusch and Viterbo 2007; Papadopoulos et al. 2008; Serpetzoglou et al. 2010), and approximation of the planetary boundary layer representation can be an error source for the prediction of surface variables (Arakawa 2004; Pleim 2007; Hu et al. 2010; Nielsen-Gammon et al. 2010; Frediani et al. 2016).

Using high spatial resolution and/or data assimilation does not always assure high forecast accuracy because of the important role of the input fields, initial conditions, and inherent model uncertainties that influence the prediction. Statistical postprocessing approaches play a useful role to address this issue and contribute to the reduction of prediction errors. Various techniques on statistical postprocessing for error/bias correction have been suggested in the literature. These techniques are based on 1) running-mean bias removal (Stensrud and Yussouf 2003, 2005; Eckel and Mass 2005; Hacker and Rife 2007; Wilczak et al. 2006; Müller 2011; Homleid 1995; Roeger et al. 2003; McCollor and Stull 2008; Rincon et al. 2010; Delle Monache et al. 2006, 2008, 2011; Djalalova et al. 2010; Kang et al. 2010), and 3) model output statistics (Glahn and Lowry 1972; Carter et al. 1989; Jacks et al. 1990; Mao et al. 1999; Wilson and Vallée 2002, 2003; Hart et al. 2004; Wilks and Hamill 2007; Glahn et al. 2009).

Combining statistical postprocessing techniques with NWP ensemble simulations is of particular interest because of the ability to characterize model uncertainty and improve the predicted variables (wind speed, temperature, humidity, etc.). Even though there is no consensus on the adequate number of ensemble members or on the best way to combine them (Weigel et al. 2009), the computational cost of ensemble simulations can be a deterrent factor. The motivation for this work has its basis in the use of a computationally efficient scheme that uses only two NWP models and statistical postprocessing techniques over a set of meteorological storms that have common characteristics. One typical method for optimum weighting of the ensemble members is Bayesian model averaging (BMA), which estimates each member's contributing weight (Raftery et al. 2005; Wilson et al. 2007; Fraley et al. 2010; Erickson et al. 2012; Sloughter et al. 2007, 2010). In the BMA approach, the calibrated weights reflect the forecast skill of each ensemble member over a training period (Fraley et al. 2010). Fraley et al. (2010) implemented BMA with 86 members, which total represents a relatively large number of ensemble members, to show how BMA can be adapted to handle exchangeable ensemble members. Erickson et al. (2012) ran BMA for specific weather storms, including

fire weather, that caused poor air quality. This test was conducted using sequential (the most recent days) and conditional (the most recent similar days) training periods and showed that the correction of the conditional training period was better than that of the sequential training.

The similarity or difference between training and out-of-sample conditions can affect the results of statistical postprocessing methods that accompany training algorithms. Although statistical postprocessing methods correct the errors over general cases, specific storms may not be improved if the training period does not consider patterns that are similar to those of the target storms. In other words, if the training dataset reflects the characteristics of the target storm, the modeled field may be improved more efficiently. Especially for high-wind speed storms, distinguishing the mean atmospheric conditions and using a training scheme with a dataset fitted to similar weather conditions can be a critical factor for the success of the error correction. Our reference to extreme storms includes tropical storms, heavy precipitation associated with floods, blizzards with strong sustained winds, and seasonal thunderstorms.

The main objective of this study is to improve surface wind speed prediction under extreme weather conditions, because it is strongly correlated with negative effects in civil infrastructure, power grid, and the environment. To this end, the combination of wind speed predictions from two atmospheric models using a Bayesian linear regression (BLR) approach is explored, and the potential to improve wind speed prediction against single-model simulations and simple linear regression (SLR) techniques is demonstrated. The combination of two atmospheric modeling systems with simple bias-correction techniques serves two purposes: it minimizes computational cost, since only two model members are being employed, and it determines the value added by Bayesian regression in a deterministic framework. An additional goal of this work is to assess the efficient length for the training period in chronological and nonchronological sequences, which will be important in the operational application of the described method. The work presented here will support the operational prediction of power outages in the northeastern United States that are strongly influenced by wind severity (Wanik et al. 2015; He et al. 2017). The power-outage modeling system is currently operating with meteorological inputs from the WRF Model (Wanik et al. 2015). Section 2 describes the model configuration and data used, section 3 presents the methods for SLR and BLR, and section 4 includes discussion of the results.

FIG. 1. Model domains: (a) WRF and (b) RAMS/ICLAMS. (c) NOAA/NCEP/National Weather Service stations over the northeastern United States (black circles), and elevation (m).

Conclusions and future work are summarized in [section 5](#).

2. Models and data

a. Atmospheric modeling systems

Two mesoscale meteorological modeling systems are implemented to simulate the selected storms. The Advanced Research version of the Weather Research and Forecasting (WRF) Model (version 3.4.1; [Skamarock et al. 2008](#)) and the Regional Atmospheric Modeling System/Integrated Community Limited Area Modeling System (RAMS/ICLAMS, referred to hereinafter as ICLAMS; [Cotton et al. 2003](#); [Solomos et al. 2011](#); [Kushta et al. 2014](#)). ICLAMS is an integrated air-quality and chemical weather modeling system that is based on

RAMS, version 6 ([Pielke et al. 1992](#); [Cotton et al. 2003](#)). It directly couples meteorological fields with air-quality components and includes gaseous, aqueous, aerosol-phase chemistry and partitioning of cloud condensation nuclei, giant cloud condensation nuclei, and ice nuclei as predictive quantities (atmospheric chemistry and feedback processes are not included in the ICLAMS simulations for this work).

Both models have three nested domains covering the northeastern United States with horizontal grid spacing of 18 km (outer domain), 6 km (inner intermediate domain), and 2 km. The third gridded domain is the focus area in this work ([Figs. 1a,b](#)). To initialize the two models, the National Centers for Environmental Prediction (NCEP) Global Forecast System ($1^{\circ} \times 1^{\circ}$; 6-hourly intervals) analyses ([NOAA/National Centers for Environmental Prediction 2007](#)) and the Final

TABLE 1. WRF and ICLAMS configuration (here, P_{top} indicates the pressure at the highest altitude level).

	WRF	ICLAMS
Grid structure	Three horizontal grids: 18, 6, and 2 km; 27 vertical levels ($P_{\text{top}} = 50$ hPa)	Three horizontal grids: 18, 6, and 2 km; 50 vertical levels ($P_{\text{top}} = 60$ hPa)
Horizontal grid scheme	Arakawa C grid	Arakawa C grid
Nesting	Two-way nesting	Two-way nesting
Initial conditions	NCEP GFS ($1^\circ \times 1^\circ$; 6 h)	NCEP FNL ($1^\circ \times 1^\circ$; 6 h)
Cumulus scheme (per grid)	Grell 3D scheme (Grell and Devenyi 2002) on the parent and second grids; no parameterization on the third grid	Kain–Fritsch cumulus parameterization on the parent and second grids; no parameterization on the third grid
Cloud microphysics	Thompson et al. (2008) scheme	Warm-rain processes; five ice condensate species; two-moment bulk scheme (Walko et al. 1995; Meyers et al. 1997); explicit cloud droplet activation scheme (Nenes and Seinfeld 2003; Fountoukis and Nenes 2005) with prescribed aerosols
Planetary boundary layer	Yonsei scheme (Hong et al. 2006)	Mellor–Yamada scheme (Mellor and Yamada 1982)
Boundary conditions	SST: NCEP GFS; topography: USGS GTOPO30 at $30''$; land cover: USGS at $30''$; soil texture: FAO at $5'$ and North America State Soil Geographic (STATSGO) at $30''$	SST daily; normalized difference vegetation index: USGS at $30''$; topography: NASA SRTM90, v4.1, at $3''$; land cover: USGS Olson Global Ecosystem at $30''$; soil texture: FAO at $2'$
Radiation	Goddard for shortwave radiation (Chou and Suarez 1994); RRTM for longwave radiation (Mlawer et al. 1997)	RRTM for shortwave and longwave radiation (Mlawer et al. 1997)
Land surface	WRF Noah (Tewari et al. 2004)	Land Ecosystem–Atmosphere Feedback model (LEAF-3; Walko et al. 2000)

Analysis (FNL) ($1^\circ \times 1^\circ$; 6-hourly intervals) data (NOAA/National Centers for Environmental Prediction 2000) are used for WRF and ICLAMS, respectively. Configuration details for both WRF and ICLAMS are summarized in Table 1.

The storms that compose the training and validation datasets are selected after a k -means cluster analysis of sea level pressure, 2-m temperature, and 10-m wind speed from the European Centre for Medium-Range Weather Forecasts interim reanalysis (ERA-Interim; Simmons et al. 2007) for 80 storms that affected the power network in the northeastern United States (from 20 outages to >15 000 outages) and span the period 2004–13 (M. Frediani 2015, personal communication). A subset of 17 storms that belong to two clusters representing winter and late-summer/autumn storms with strong winds and intense pressure gradients is selected. The selected storms include three major storms for the northeastern United States: Hurricane Irene (2011), Hurricane Sandy (2012), and the 8–9 February 2013 blizzard. General information about the storms is included in Table 2.

b. Observations

The Automated Surface Observing System (ASOS) observation datasets at NCEP are used for model evaluation and also for the implementation of error

optimization (SLR and BLR). ASOS generally provides minute-by-minute observations and generates the Meteorological Terminal Aviation Routine Weather Report (METAR) and Aviation Selected Special Weather (SPECI) report. ASOS is installed at more than 900 airports across the United States, and data from 80 stations over the northeastern United States are used in this study (Fig. 1c). The wind speed at observational locations is matched with the modeled wind speed using bilinear interpolation (nested grid with a grid spacing of $2 \text{ km} \times 2 \text{ km}$).

3. Method

Two statistical postprocessing methods, SLR and BLR, are applied for error correction of the modeled wind speed. Thirteen storms are used for the training dataset, and four storms are used for the out-of-sample application (validation). The first application uses a chronological sequence to select the storms for the training dataset. The second application uses all possible combinations of the 13 storms to compose the training dataset, regardless of the date of occurrence. The two regression methods are described in sections 3a and 3b, section 3c presents details on the training scheme, and section 3d includes information about data processing and statistical metrics.

TABLE 2. Storm type and date (the duration of all simulations was 61 h; source: NOAA Significant Weather Events Archive).

Storms	Model start time and date	Storm type
5 Nov 2004	1800 UTC 4 Nov 2004	Windstorm
1 Dec 2004	1200 UTC 30 Nov 2004	Windstorm
1–3 Apr 2005	1800 UTC 1 Apr 2005	Rain/windstorm
16 Oct 2005	1800 UTC 15 Oct 2005	Windstorm
24–25 Oct 2005	1800 UTC 24 Oct 2005	Northeaster
14–15 Jan 2006	0000 UTC 14 Jan 2006	Rain/snow/windstorm
18 Jan 2006	0000 UTC 17 Jan 2006	High windstorm
17 Feb 2006	1800 UTC 16 Feb 2006	High windstorm
15–16 Apr 2007	1800 UTC 14 Apr 2007	Northeaster
6–8 Jan 2009	0000 UTC 7 Jan 2009	Ice storm/windstorm
12–15 Mar 2010	0000 UTC 13 Mar 2010	Heavy rain/high windstorm
26–27 Dec 2010	1800 UTC 25 Dec 2010	Blizzard
11–12 Jan 2011	1800 UTC 11 Jan 2011	Heavy snowstorm
28 Aug 2011	0000 UTC 28 Aug 2011	Hurricane (Irene)
29 Oct 2012	1800 UTC 28 Oct 2012	Hurricane (Sandy)
7 Nov 2012	0600 UTC 7 Nov 2012	Northeaster
8–9 Feb 2013	0000 UTC 8 Feb 2013	Blizzard

a. Simple linear regression

The SLR model consists of the mean function (Weisberg 2005), defined as follows:

$$E(\mathbf{Y}|\mathbf{X}=x) = \beta_0 + \beta_1 x. \quad (1)$$

Intercept β_0 is the value of $E(\mathbf{Y}|\mathbf{X}=x)$ when x is 0, and the slope β_1 is the rate of change in $E(\mathbf{Y}|\mathbf{X}=x)$ for a unit change in x , respectively. The unknown parameters β_0 and β_1 are estimated from the modeled–observed wind speed pairs given the independent and dependent variable vectors $\mathbf{X}$ and $\mathbf{Y}$, respectively.

In this study, the SLR model with a single predictor (wind speed at 10 m) is developed for WRF and ICLAMS separately. To evaluate the estimators, storms are selected to arrange the training datasets first, and each estimator of β_0 and β_1 is calculated from the training storms by the ordinary least squares (OLS) method as follows:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{and} \quad (2)$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \quad (3)$$

where $\bar{x}$ and $\bar{y}$ are the averages of modeled values x and observed values y in the training datasets. Since a linear relationship between observed and modeled wind speed exists, this relationship points toward the possibility of model prediction correction (Sweeney et al. 2013). The SLR method is developed for each station because the linear relationship is spatially variable (from station to

station) and the spatial error heterogeneity must be preserved in the results. Therefore, the SLR analysis is implemented for each station (a total of 80 stations) by the OLS method, and overall the final SLR model for WRF and ICLAMS is given as

$$\hat{\mathbf{Y}}_{\text{station}_m} = \hat{\beta}_{0,\text{station}_m} + \hat{\beta}_{1,\text{station}_m} \mathbf{X}_{\text{station}_m}, \quad (4)$$

where $\hat{\beta}_{0,\text{station}_m}$ and $\hat{\beta}_{1,\text{station}_m}$ are the estimators of station m evaluated from the training dataset and $\mathbf{X}_{\text{station}_m}$ is the predictor of the station m from the WRF or ICLAMS out-of-sample storms. The $\hat{\mathbf{Y}}_{\text{station}_m}$ is the final product of the SLR model for station m .

b. Bayesian linear regression

BLR is implemented as a new approach to improve WRF and ICLAMS 10-m wind speed fields. Bayesian statistics are based on Bayes's theorem that deals with uncertainty of unknown parameters, and the parameters are basically inferred to probabilistic forms from the observed data under the Bayesian framework (Chu and Zhao 2011). The Bayesian formula to infer the unknown parameter vector $\boldsymbol{\theta}$ is thus governed by

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\mathbf{y})}, \quad (5)$$

where $p(\boldsymbol{\theta}|\mathbf{y})$ is the posterior probability density function (PDF) of $\boldsymbol{\theta}$ given the observed data information $\mathbf{y}$, $p(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood function, $p(\boldsymbol{\theta})$ is the prior PDF for the unknown parameter vector of $\boldsymbol{\theta}$, and $p(\mathbf{y})$ is the PDF of the observation vector $\mathbf{y}$. For the continuous case for the Bayes theorem, Eq. (5) is formulated as follows:

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) d\boldsymbol{\theta}}. \quad (6)$$

In this study, two controlled variables produced from WRF and ICLAMS are used in the BLR, and therefore the normal linear model is regarded as a normal multiple regression form defined by two predictor variables. In vector form, the normal multiple regression equation is defined by

$$\mathbf{y} = \boldsymbol{\beta}\mathbf{X} + \boldsymbol{\varepsilon}, \quad (7)$$

where $\mathbf{y}$ is an $n \times 1$ vector of observations, $\mathbf{X}$ is an $n \times p$ matrix of independent variables incorporating the unit matrix of the first column for the intercepts β_0 , $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients $[\beta_0 \beta_1 \beta_2]$, and the error term is $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$ with the unknown dispersion parameter σ^2 . After consideration of the elements of the parameter vector $\boldsymbol{\theta} = [\beta_0 \beta_1 \beta_2 \sigma^2]$, Eq. (6) becomes

$$p(\beta_0, \beta_1, \beta_2, \sigma^2 | \mathbf{y}) = \frac{p(\mathbf{y} | \beta_0, \beta_1, \beta_2, \sigma^2) p(\beta_0, \beta_1, \beta_2, \sigma^2)}{\iiint p(\mathbf{y} | \beta_0, \beta_1, \beta_2, \sigma^2) p(\beta_0, \beta_1, \beta_2, \sigma^2) d\beta_0 d\beta_1 d\beta_2 d\sigma^2}. \quad (8)$$

It is assumed that the posterior probability distribution is in the same family as the prior probability distribution and that prior information of σ^2 can be inferred. The posterior mean of $\boldsymbol{\beta}$ can be calculated by Eq. (9) (Cattin et al. 1983; Zellner 1996; O'Hagan and Forster 2004; Sorensen and Gianola 2002; Walter et al. 2007; Walter and Augustin 2010):

$$\bar{\boldsymbol{\beta}} = \begin{bmatrix} \bar{\beta}_0 \\ \bar{\beta}_1 \\ \bar{\beta}_2 \end{bmatrix} = (\mathbf{X}^T \mathbf{X} + \mathbf{V}_{\boldsymbol{\beta}}^{-1})^{-1} (\mathbf{X}^T \mathbf{y} + \mathbf{V}_{\boldsymbol{\beta}}^{-1} \boldsymbol{\mu}_{\boldsymbol{\beta}}), \quad \text{with}$$

$$\mathbf{y} = \begin{bmatrix} \text{Obs}_1 \\ \text{Obs}_2 \\ \vdots \\ \text{Obs}_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1\text{WRF}} & x_{1\text{ICLAMS}} \\ 1 & x_{2\text{WRF}} & x_{2\text{ICLAMS}} \\ \vdots & \vdots & \vdots \\ 1 & x_{n\text{WRF}} & x_{n\text{ICLAMS}} \end{bmatrix},$$

$$\mathbf{V}_{\boldsymbol{\beta}} = \begin{bmatrix} \sigma_0^2 & 0 & 0 \\ 0 & \sigma_1^2 & 0 \\ 0 & 0 & \sigma_2^2 \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\mu}_{\boldsymbol{\beta}} = \begin{bmatrix} \mu_0 \\ \mu_1 \\ \mu_2 \end{bmatrix}, \quad (9)$$

where $\mathbf{y}$ is the column vector of 10-m observed wind speed for n time steps, $\mathbf{X}$ is the matrix of WRF and ICLAMS wind speed for n time steps, $\mathbf{V}_{\boldsymbol{\beta}}$ is a diagonal matrix including three prior variances that correspond to each element of $\bar{\boldsymbol{\beta}}$, and $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ is a prior mean vector. It is assumed a priori that the best model will be a simple unbiased average of the two simulations, implying a mean vector of $\boldsymbol{\mu}_{\boldsymbol{\beta}} = [0 \ 0.5 \ 0.5]$. Certainty about that assumption is defined by the size of the prior variances (σ_i^2). To rely more on the data to inform the final model, the prior variances are made much larger. In an extreme case, as the prior variances go to infinity, Bayesian posterior estimates will match the OLS estimates. As the prior variances get smaller, the results will shrink toward the a priori assumptions. Shrinkage of this type will allow the model to be more robust in the presence of outliers and other strange and influential data. To develop BLR using Eq. (9), optimal prior variances are searched with the vector representing the prior mean ($\boldsymbol{\mu}_{\boldsymbol{\beta}}$). The final BLR model that is based on WRF and ICLAMS is formulated as follows:

$$\hat{\mathbf{Y}}_{\text{station}_m} = \bar{\beta}_{0,\text{station}_m} + \bar{\beta}_{1,\text{station}_m} \mathbf{X}_{\text{WRF},\text{station}_m} + \bar{\beta}_{2,\text{station}_m} \mathbf{X}_{\text{ICLAMS},\text{station}_m}, \quad (10)$$

where $\bar{\beta}_{0,\text{station}_m}$ is the intercept of the BLR equation and $\bar{\beta}_{1,\text{station}_m}$ and $\bar{\beta}_{2,\text{station}_m}$ denote the regression coefficients for the two column-vector predictor variables $\mathbf{X}_{\text{WRF},\text{station}_m}$ and $\mathbf{X}_{\text{ICLAMS},\text{station}_m}$ for station m . The column vector $\hat{\mathbf{Y}}_{\text{station}_m}$ is the adjusted 10-m wind speed field for station m using the BLR method.

c. Training scheme

The regression coefficients for SLR and BLR can be estimated using a variety of training datasets. To investigate the sensitivity of the results to training-period length, a variation in the number of storms as well as a change in the chronological order of the training dataset is examined. For instance, in the case of using one storm for the training dataset, SLR is implemented for each station, and then the number of storms is gradually increased to make different combinations. Among the 17 storms, 13 storms from 2004 to 2011 are selected as training storms and the other four storms are used for out-of-sample applications/validations. The four storms represent significant storms over the northeastern United States during 2011–13: Hurricane Irene (2011), Hurricane Sandy (2012), a November 2012 storm (affected by Hurricane Sandy), and a February 2013 blizzard (maximum 1-h wind speed from 80 inland stations: 21, 25, 22, and 24 m s^{-1} for Irene, Sandy, the November 2012 storm, and the February 2013 blizzard, respectively).

In the first approach, an increasing number of storms in chronological order are employed. The second approach consists of training datasets composed of all possible combinations of the 13 storms to analyze the behavior of coefficient of determination R^2 , root-mean-square error (RMSE), mean bias, and centered root-mean-square error (CRMSE) in conformity with changes in combinations. Each quantity of combination using different number of storms can be calculated by $n!/[r!(n-r)!]$, which represents the number of “ r combinations” from a given set of n elements (r is an integer and $1 \leq r \leq 13$ with $n = 13$). To be specific, for the case of using a single storm 13 individual storms constitute the training dataset and for the case of using two storms 78 training datasets are required (Table 3). The experiments for all possible combinations (with a total of 8191 training datasets; Table 3) are implemented for each observation station for SLR and BLR.

The BLR approach for the first application has three phases (Fig. 2): 1) A random selection of 10 000 prior variance sets $\mathbf{V}_{\boldsymbol{\beta}}$ [Eq. (9)] within the interval $[10^{-10}, 1]$ is

TABLE 3. Number of possible storm-sequence combinations for the training datasets.

No. of storms r	No. of r combinations
1	13
2	78
3	286
4	715
5	1287
6	1716
7	1716
8	1287
9	715
10	286
11	78
12	13
13	1
All possible combinations	8191

made. Each variance set (of the 10 000) is used to estimate $\bar{\beta}$ [Eq. (9)] for each station using all training storms. The estimated $\bar{\beta}$ is applied to individual storms of the training dataset to compute the global RMSE for each storm and each station (phase 1 in Fig. 2). 2) The

RMSE that corresponds to each variance set is summed over all k storms (phase 2 in Fig. 2). 3) The optimal prior variances that corresponded to the minimum summation of the RMSE from phase 2 are used for the calculation of the final $\bar{\beta}$ for each station, which is then applied to out-of-sample storms (phase 3 in Fig. 2).

The BLR procedure demands a relatively longer computation time than SLR since it incorporates a phase to sample 10 000 prior variance sets. To reduce the computational time for BLR experiments related to all possible storm combinations, it is necessary to use fewer random samples of variance sets, instead of the previous 10 000. The variation of the RMSE from all cases, employing from 1 up to 13 storms for the training dataset, is analyzed to determine the reasonable number of prior variance samples that would reduce the computational cost and succeed in minimizing the RMSE to a level that is similar to that of the 10 000 variance sets.

RMSE variability with increasing sample size is quantified by calculating the normalized difference of RMSE (NDiff) from the final minimized RMSE of all 10 000 samples. NDiff is calculated as

$$\text{NDiff} = \frac{\left[\min_{i \in j} \left(\sum_{n=1}^k \text{RMSE}_{\text{storm},i} \right), 1 \leq j \leq 10\,000 \right] - \min_{i \in 10\,000} \left(\sum_{n=1}^k \text{RMSE}_{\text{storm},i} \right)}{\min_{i \in 10\,000} \left(\sum_{n=1}^k \text{RMSE}_{\text{storm},i} \right)} \times 100\%, \quad (11)$$

where i is the number of variance sets in the range from 1 to 10 000, j is the number of variance sets to be used for reduction of the computational cost in the range from 1 to 10 000, and k is the number of storms. For example, if 2 storms and 20 variance sets are used to calculate NDiff, then

$$\text{NDiff} = \frac{\left[\min_{i \in 20} \left(\sum_{n=1}^2 \text{RMSE}_{\text{storm},i} \right) \right] - \min_{i \in 10\,000} \left(\sum_{n=1}^2 \text{RMSE}_{\text{storm},i} \right)}{\min_{i \in 10\,000} \left(\sum_{n=1}^2 \text{RMSE}_{\text{storm},i} \right)} \times 100\%.$$

All cases commonly display that the NDiff values are decreased near the 20 variance sets (Fig. 3). Thus, 20 samples are identified as a proper sample size for prior variance sets and are used to implement BLR for the all-storm combinations.

d. Data processing and statistical metrics

The first 6 h are regarded as the model spinup time and are discarded from the analysis. The missing and zero values for 10-m wind speed observations are not

included in the modeled–observed pairs. Five statistical metrics that offer complementary views on the model and regression performances are used. To evaluate the impact of regression techniques on the 10-m modeled wind speed, the metrics are calculated for raw WRF, raw ICLAMS, WRF_{SLR}, ICLAMS_{SLR}, and BLR. The five statistical metrics used in this study are R^2 , RMSE, mean bias, CRMSE (Murphy 1988; Taylor 2001; Delle Monache et al. 2011), and skill score (SS). The first four are determined as follows:

FIG. 2. A schematic diagram of the BLR approach.

$$R^2 = \left(\frac{N \sum_N (XY) - \left(\sum_N X \right) \left(\sum_N Y \right)}{\left\{ \left[N \sum_N X^2 - \left(\sum_N X \right)^2 \right] \left[N \sum_N Y^2 - \left(\sum_N Y \right)^2 \right] \right\}^{1/2}} \right)^2, \quad (12)$$

$$\text{RMSE} = \left[N^{-1} \sum_N (X - Y)^2 \right]^{1/2}, \quad (13)$$

$$\text{bias} = N^{-1} \sum_N (X - Y), \quad \text{and} \quad (14)$$

$$\text{CRMSE} = \left\{ N^{-1} \sum_N [(X - \bar{X}) - (Y - \bar{Y})]^2 \right\}^{1/2}, \quad (15)$$

where the modeled value is represented by X , the observed wind speed is represented by Y , N is the total number of data points, and $\bar{X}$ and $\bar{Y}$ are the modeled and observed wind speed averages over the N values used in the calculations. RMSE is used to evaluate model performances and the crucial objective function aiming to mitigate errors for the BLR approach.

CRMSE is a measure of the random component of RMSE, and the systematic component is represented by the bias.

To measure the relative improvement of the regression techniques, the SS with regard to RMSE and R^2 (e.g., Wilks 1995; Libonati et al. 2008; Idowu and Rautenbach 2009; Delle Monache et al. 2011) is calculated. An example of the SS calculation is shown in Eqs. (16) and (17):

$$\text{SS}_{\text{RMSE}} = \frac{\text{RMSE}_{\text{raw}} - \text{RMSE}_{\text{SLR/BLR}}}{\text{RMSE}_{\text{raw}}} \times 100\% \quad \text{and} \quad (16)$$

$$\text{SS}_{R^2} = \frac{R^2_{\text{SLR/BLR}} - R^2_{\text{raw}}}{R^2_{\text{raw}}} \times 100\%. \quad (17)$$

FIG. 3. RMSE normalized difference (NDiff) using different sample sizes for the BLR training datasets.

Equations (16) and (17) estimate the relative improvement of the SLR and BLR approaches versus raw-WRF and raw-ICLAMS predictions. Positive values of SS_{RMSE} and SS_{R^2} indicate that the suggested regression method improves the raw model outputs.

4. Results and discussion

a. Chronologically ordered storm combinations for the training dataset

The variation of R^2 , RMSE, bias, and CRMSE (Figs. 4 and 5) for each out-of-sample storm shows an increase in R^2 and a decrease in RMSE, bias, and CRMSE when the number of storms of the training dataset increases. All three models [WRF_{SLR} (triangles), ICLAMS_{SLR} (squares), and BLR (circles)] exhibit poor performance as indicated by low R^2 and high RMSE, bias, and CRMSE when employing one storm for training, which denotes that one historical storm is not sufficient to improve wind speed predictions of future storms. The statistical metrics progressively improve with increasing number of storms in the training dataset, and the trend reaches a plateau after 8–10 storms to an almost constant value. This is indicative of the number of storms that will be efficient and effective for correcting the modeled wind speed error. BLR is consistently performing better across all storm cases with the only exception of the November 2012 storm, for which BLR and ICLAMS_{SLR} share comparable performances. We include 95% bootstrapped confidence intervals for all statistical metrics

and out-of-sample storms (shown in Figs. 4 and 5). Nonoverlapping bootstrapped intervals show that results are significantly different when looking at the RMSE for Irene and Sandy, for the maximum number of storms in the training dataset. For the two most recent out-of-sample storms, ICLAMS_{SLR} and BLR are not significantly different in terms of the RMSE.

The mean bias is almost entirely removed for most out-of-sample storms, with all models being successful (Fig. 5). The mean bias of raw model outputs for the four storms is in the range from -1.0 to 0.5 m s^{-1} . At least five storms are required for the success of bias removal by SLR and BLR (Fig. 5). Hurricane Sandy wind speed exhibits a positive bias even with the inclusion of 13 storms in the training dataset, and BLR has higher bias than WRF or ICLAMS. This is attributed to the fact that model predictions of Sandy exhibit error characteristics that are distinctly different from those of the other storms in the database. To explore this behavior further, Hurricane Irene was included in the training dataset (14 storms instead of 13) as the storm closest in character to Hurricane Sandy. Sandy, the November 2012 storm, and the February 2013 blizzard are kept as the out-of-sample storms, and the results did not show significant differences (not shown here). The average RMSE for Sandy changed by only 0.01 m s^{-1} , and the spatial distribution was not significantly affected. The case of Sandy shows that the mean bias can be reduced when the available number of training storms increases. Bias removal is consistent with the systematic error removal that is an expected outcome of regression techniques. The random component of RMSE, denoted by CRMSE, has a decreasing trend for all storms as the number of training storms increase (Fig. 5). For both RMSE and CRMSE, BLR results are more successful than those of SLR, with the exception of the November 2012 storm (as previously noted).

So far, results from the regression techniques are discussed without mentioning the “raw” atmospheric model performance. The correlation increases to 0.6–0.8 and the RMSE decreases to 1.7 – 2.0 m s^{-1} , for the different out-of-sample storms. The distribution of weights [beta coefficients from Eq. (9)] given for NWP models in the BLR approach (including 13 storms in the training dataset) shows a slight “preference” toward the ICLAMS model (Fig. 6). These beta values give the optimal RMSE in the training and are subsequently applied to all out-of-sample storms but vary for each station. To put things in perspective, statistical metrics employing the raw model outputs are presented using the SS.

The SS (the relative improvement in percent for a given metric), grouped by storm and raw model output,

FIG. 4. Chronological storm-sequence experiment: (left) R^2 and (right) RMSE (m s^{-1}) variation by increasing the number of training storms for WRF SLR (triangles), ICLAMS SLR (squares), and BLR (circles). The 95% bootstrap confidence intervals are indicated by the error bars for WRF SLR (blue), ICLAMS SLR (red), and BLR (purple).

FIG. 5. As in Fig. 4, but for (left) bias (m s^{-1}) and (right) CRMSE (m s^{-1}).

FIG. 6. Cumulative distribution function (CDF) of the beta coefficients at 80 stations when including 13 storms in the training dataset ($\text{BLR} = \bar{\beta}_0 + \bar{\beta}_1 \times \text{WRF} + \bar{\beta}_2 \times \text{ICLAMS}$).

indicates improvement by BLR relative to SLR (Table 4), marked by increased SS values for BLR versus raw model outputs for all storms. This result indicates that the BLR approach has been successful in improving the RMSE and R^2 statistical metrics for wind speed relative to raw model outputs. In addition, the RMSE and R^2 are analyzed for each model normalized by BLR to identify how BLR performs over the other methods. Normalized RMSE values of greater than 1 indicate that BLR performs better (all normalized RMSE values are greater than 1, with the exception of the November 2012 storm and ICLAMS_{SLR}; Table 5). Inversely, normalized R^2 of smaller than 1 demonstrates that BLR results outperform the other model results in the out-of-sample storms. These normalized RMSE and R^2 values listed in Table 5 show the same patterns in terms of the BLR performance. Normalized metrics indicate that BLR improves the wind speed statistical metrics for Irene, Sandy, and

the February 2013 storm. ICLAMS_{SLR} performs as well as, if not better than, BLR for the November 2012 storm. The results from the normalized metrics are consistent with the conclusions from the confidence intervals discussed previously.

The spatial distribution of RMSE was also analyzed, with 13 storms as the training dataset for all out-of-sample storms (Fig. 7). In each plot, colored circles represent the RMSE value calculated using observations at each station location. All suggested regression methods in this study successfully reduce the RMSE of the raw WRF and raw ICLAMS for almost all stations and storms. The RMSE in raw model outputs ranges between 1.6 and 3.5 ms^{-1} , whereas when regression techniques were used a large number of stations show decreased values within a range from 1.0 to 2.5 ms^{-1} (more abundance of lower-range RMSE values). Overall, BLR is shown to be an effective method to

TABLE 4. Skill score (%) evaluated by RMSE and R^2 of WRF_{SLR}, ICLAMS_{SLR}, and BLR, with the number of storms in the training dataset equal to 13.

Storm	SS _{RMSE} (%)				SS _{R²} (%)			
	WRF _{SLR} vs	BLR vs	ICLAMS _{SLR} vs	BLR vs	WRF _{SLR} vs	BLR vs	ICLAMS _{SLR} vs	BLR vs
	WRF _{raw}	WRF _{raw}	ICLAMS _{raw}	ICLAMS _{raw}	WRF _{raw}	WRF _{raw}	ICLAMS _{raw}	ICLAMS _{raw}
Irene	20.4	24.6	9.2	17.1	12.0	15.7	8.0	14.5
Sandy	16.6	22.0	15.1	18.3	21.0	32.7	16.7	23.7
7 Nov 2012	24.4	31.9	21.9	21.7	27.0	47.3	25.5	25.5
8–9 Feb 2013	22.9	31.3	15.0	18.0	20.0	33.5	18.8	22.8

reduce RMSE for the stations of our case study, with highest reductions in the range of 17%–32% when compared with raw model outputs. More details on the time series and RMSE values for individual stations are provided in Table S1 and Figs. S1–S4 of the online supplemental material. In addition, the spatial distribution of bias and CRMSE is provided in supplemental Figs. S5 and S6 for a more detailed view of the BLR efficiency at the station level.

b. All-storm combinations for the training dataset

In this section, the training dataset comprises all possible storm combinations while increasing the number of storms. The intention of this test is to define the sensitivity of BLR and SLR results to a random combination of storm sequences and to denote the confidence that can be placed in the BLR method if a convergence in the results is achieved. The results showing 25th, 50th, and 75th percentiles (horizontal bars) and minimum and maximum (error bars) (Figs. 8 and 9) are similar to the chronologically ordered selection of training storms in the sense that the bias is almost entirely removed in most cases and RMSE and CRMSE are decreased with the addition of storms in the training dataset. The variability of all metrics is clearly reduced by adding more storms in the training dataset (box-and-whisker plots in Figs. 8 and 9). Even at the combination of six or seven storms (largest number of combinations = 1716; Table 3), the distribution is narrow. BLR starts with a relatively narrow width distribution in comparison with the other models, having higher R^2 values and lower RMSE and CRMSE. For example, with use of a

single training storm in the case of Irene (Fig. 8), statistical metrics for BLR exhibit the following ranges: $R^2 = [0.73, 0.78]$, RMSE = $[1.91, 2.17] \text{ (m s}^{-1}\text{)}$ and CRMSE = $[1.91, 2.10] \text{ (m s}^{-1}\text{)}$. These ranges are narrower than those of WRF_{SLR} ($R^2 = [0.59, 0.74]$, RMSE = $[2.07, 2.59] \text{ (m s}^{-1}\text{)}$, and CRMSE = $[2.07, 2.52] \text{ (m s}^{-1}\text{)}$) and ICLAMS_{SLR} ($R^2 = [0.65, 0.72]$, RMSE = $[2.19, 2.92] \text{ (m s}^{-1}\text{)}$, and CRMSE = $[2.16, 2.79] \text{ (m s}^{-1}\text{)}$). In addition, the median values corresponding to BLR (R^2 : 0.76, RMSE: 2.05 m s^{-1} , and CRMSE: 1.98 m s^{-1}) indicate statistically significant improvements when compared with WRF_{SLR} (R^2 : 0.68, RMSE: 2.28 m s^{-1} , and CRMSE: 2.26 m s^{-1}) and ICLAMS_{SLR} (R^2 : 0.68, RMSE: 2.32 m s^{-1} , and CRMSE: 2.31 m s^{-1}).

When the lowest possible RMSE is selected (minimum RMSE in the 12 storm combinations; Figs. 8 and 9) and used for the calculation of BLR weighting factors for the out-of-sample application, there is no significant change in the average RMSE over all stations or in the spatial distribution of RMSE (not shown). The results from combining all possible storm sequences denote a convergence in the wind prediction improvements by both the chronological and all-combinations approaches, giving confidence in the performance of the proposed BLR technique.

5. Conclusions

In this study, a simple linear regression and a Bayesian linear regression are introduced as postprediction error-correction techniques to improve modeled 10-m wind speed of storms that exhibit high-wind speed occurrences. Both simple and Bayesian linear regressions rely

TABLE 5. Normalized RMSE and R^2 by the relevant metrics for BLR, with the number of storms in the training dataset equal to 13.

Storm	RMSE normalized by BLR				R^2 normalized by BLR			
	WRF _{raw} /	ICLAMS _{raw} /	WRF _{SLR} /	ICLAMS _{SLR} /	WRF _{raw} /	ICLAMS _{raw} /	WRF _{SLR} /	ICLAMS _{SLR} /
	BLR	BLR	BLR	BLR	BLR	BLR	BLR	BLR
Irene	1.33	1.21	1.06	1.10	0.86	0.87	0.97	0.94
Sandy	1.28	1.22	1.07	1.04	0.75	0.81	0.91	0.94
7 Nov 2012	1.47	1.28	1.11	1.00	0.68	0.80	0.86	1.00
8–9 Feb 2013	1.46	1.22	1.12	1.04	0.75	0.81	0.90	0.97

FIG. 7. Spatial distribution of RMSE for the chronologically ordered training-dataset application.

on the training dataset and the appropriate selection of storms with similar weather characteristics. A selection of 17 storms in total is used to study the efficiency of the two methods in reducing the wind speed systematic and random errors for station locations in the northeastern United States. Thirteen storms constitute the training dataset, and four high-impact storms (two hurricanes, one blizzard, and one northeaster) are used for the out-of-sample applications.

Both SLR and BLR reduce systematic and random errors for most out-of-sample storms. The statistical metrics and spatial distribution of root-mean-square error indicate that BLR is more successful in the surface wind speed error correction because it takes into account wind predictions from two atmospheric modeling systems. Such a result is promising because the two-model application reduces the computational cost associated with multimodel or single-model ensemble

FIG. 8. Randomized storm-sequence experiment: (left) R^2 and (right) RMSE (m s^{-1}) spread behavior related to the number of storms in the training dataset for WRF SLR (blue), ICLAMS SLR (red), and BLR (purple).

forecasts without compromising the accuracy of the wind speed error reduction.

The selection of storms in the training dataset does not depend on the chronological sequence of storm occurrence but rather depends mostly on their abundance.

The randomized experiment shows a good convergence of wind speed forecast improvements for all possible storm combinations, increasing the confidence in the proposed BLR technique. A suggestion that applies to the specific type of weather storms included in this work

FIG. 9. As in Fig. 8, but for (left) bias (m s^{-1}) and (right) CRMSE (m s^{-1}).

is that 10–13 storms in the training dataset are sufficient to reduce the errors in the prediction by 20%–30% for all stations relative to raw model outputs (Table 4) and by up to 60% for individual stations (online supplemental Fig. S7). A selection based on occurrence (chronological sequence) is also considered to be

sufficient. This conclusion allows for planning of real-time operational wind speed error correction using the BLR technique.

Overall, this study has demonstrated that the application of two regression methods can improve the surface wind speed prediction from single- and dual-model

simulations. The dual-model combination of the BLR approach is more skillful and merits further investigation. Future extensions of this work include distribution of optimized BLR coefficients to each grid point of the model domain to improve the modeled wind speed for all locations. Furthermore, beta testing will be expanded to an operational setup in the northeastern U.S. region, where the real-time wind speed prediction of a storm using the two-modeling system will be corrected on the basis of historical storms included in the training dataset. This will be accomplished by operationally running both numerical weather prediction models (WRF and RAMS/ICLAMS) daily with a 5-day forecast window (WRF is currently operational). The current practice of identifying a potential future storm by consulting the in-house NWP as well as other operational forecasts (e.g., National Weather Service and NCEP) will be implemented, in which event BLR will be applied to provide optimal dual-model wind speed predictions.

Acknowledgments. The work was supported by Eversource Energy through a research grant awarded by the Eversource Energy Center at the University of Connecticut. WRF is developed and maintained by the National Center for Atmospheric Research, which is sponsored by the National Science Foundation.

REFERENCES

- Ancell, B. C., 2012: Examination of analysis and forecast errors of high-resolution assimilation, bias removal, and digital filter initialization with an ensemble Kalman filter. *Mon. Wea. Rev.*, **140**, 3992–4004, doi:[10.1175/MWR-D-11-00319.1](https://doi.org/10.1175/MWR-D-11-00319.1).
- , C. F. Mass, and G. J. Hakim, 2011: Evaluation of surface analyses and forecasts with a multiscale ensemble Kalman filter in regions of complex terrain. *Mon. Wea. Rev.*, **139**, 2008–2024, doi:[10.1175/2010MWR3612.1](https://doi.org/10.1175/2010MWR3612.1).
- , E. Kashawlic, and J. L. Schroeder, 2015: Evaluation of wind forecasts and observation impacts from variational and ensemble data assimilation for wind energy applications. *Mon. Wea. Rev.*, **143**, 3230–3245, doi:[10.1175/MWR-D-15-0001.1](https://doi.org/10.1175/MWR-D-15-0001.1).
- Arakawa, A., 2004: The cumulus parameterization problem: Past, present, and future. *J. Climate*, **17**, 2493–2525, doi:[10.1175/1520-0442\(2004\)017<2493:RATCP>2.0.CO;2](https://doi.org/10.1175/1520-0442(2004)017<2493:RATCP>2.0.CO;2).
- Barker, D., and Coauthors, 2012: The Weather Research and Forecasting Model's Community Variational/Ensemble Data Assimilation System: WRFDA. *Bull. Amer. Meteor. Soc.*, **93**, 831–843, doi:[10.1175/BAMS-D-11-00167.1](https://doi.org/10.1175/BAMS-D-11-00167.1).
- Carter, G. M., J. P. Dallavalle, and H. R. Glahn, 1989: Statistical forecasts based on the National Meteorological Center's numerical weather prediction system. *Wea. Forecasting*, **4**, 401–412, doi:[10.1175/1520-0434\(1989\)004<0401:SFBOTN>2.0.CO;2](https://doi.org/10.1175/1520-0434(1989)004<0401:SFBOTN>2.0.CO;2).
- Cattin, P., A. E. Gelfand, and J. Danes, 1983: A simple Bayesian procedure for estimation in a conjoint model. *J. Mark. Res.*, **20**, 29–35, doi:[10.2307/3151409](https://doi.org/10.2307/3151409).
- Chou, M.-D., and M. J. Suarez, 1994: An efficient thermal infrared radiation parameterization for use in general circulation models. NASA Tech. Memo. 104606, Vol. 3, 85 pp. [Available online at <https://permanent.access.gpo.gov/gpo60401/19950009331.pdf>.]
- Chu, P.-S., and X. Zhao, 2011: Bayesian analysis for extreme climatic events: A review. *Atmos. Res.*, **102**, 243–262, doi:[10.1016/j.atmosres.2011.07.001](https://doi.org/10.1016/j.atmosres.2011.07.001).
- Cotton, W. R., and Coauthors, 2003: RAMS 2001: Current status and future directions. *Meteor. Atmos. Phys.*, **82**, 5–29, doi:[10.1007/s00703-001-0584-9](https://doi.org/10.1007/s00703-001-0584-9).
- Delle Monache, L., T. Nipen, X. Deng, Y. Zhou, and R. Stull, 2006: Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. *J. Geophys. Res.*, **111**, D05308, doi:[10.1029/2005JD006311](https://doi.org/10.1029/2005JD006311).
- , and Coauthors, 2008: A Kalman-filter bias correction method applied to deterministic, ensemble averaged and probabilistic forecasts of surface ozone. *Tellus*, **60B**, 238–249, doi:[10.1111/j.1600-0889.2007.00332.x](https://doi.org/10.1111/j.1600-0889.2007.00332.x).
- , T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. *Mon. Wea. Rev.*, **139**, 3554–3570, doi:[10.1175/2011MWR3653.1](https://doi.org/10.1175/2011MWR3653.1).
- Djalalova, I., and Coauthors, 2010: Ensemble and bias-correction techniques for air quality model forecasts of surface O₃ and PM_{2.5} during the TEXAQs-II experiment of 2006. *Atmos. Environ.*, **44**, 455–467, doi:[10.1016/j.atmosenv.2009.11.007](https://doi.org/10.1016/j.atmosenv.2009.11.007).
- Doblas-Reyes, F. J., R. Hagedorn, and T. N. Palmer, 2005: The rationale behind the success of multi-model ensembles in seasonal forecasting—II. Calibration and combination. *Tellus*, **57A**, 234–252, doi:[10.3402/tellusa.v57i3.14658](https://doi.org/10.3402/tellusa.v57i3.14658).
- Drusch, M., and P. Viterbo, 2007: Assimilation of screen-level variables in ECMWF's Integrated Forecast System: A study on the impact on the forecast quality and analyzed soil moisture. *Mon. Wea. Rev.*, **135**, 300–314, doi:[10.1175/MWR3309.1](https://doi.org/10.1175/MWR3309.1).
- Eckel, F. A., and C. F. Mass, 2005: Aspects of effective mesoscale, short-range ensemble forecasting. *Wea. Forecasting*, **20**, 328–350, doi:[10.1175/WAF843.1](https://doi.org/10.1175/WAF843.1).
- Erickson, M. J., B. A. Colle, and J. J. Charney, 2012: Impact of bias-correction type and conditional training on Bayesian model averaging over the northeast United States. *Wea. Forecasting*, **27**, 1449–1469, doi:[10.1175/WAF-D-11-00149.1](https://doi.org/10.1175/WAF-D-11-00149.1).
- Fountoukis, C., and A. Nenes, 2005: Continued development of a cloud droplet formation parameterization for global climate models. *J. Geophys. Res.*, **110**, D11212, doi:[10.1029/2004JD005591](https://doi.org/10.1029/2004JD005591).
- Fraley, C., A. E. Raftery, and T. Gneiting, 2010: Calibrating multimodel forecast ensembles with exchangeable and missing members using Bayesian model averaging. *Mon. Wea. Rev.*, **138**, 190–202, doi:[10.1175/2009MWR3046.1](https://doi.org/10.1175/2009MWR3046.1).
- Frediani, M. E., J. P. Hacker, E. N. Anagnostou, and T. Hopson, 2016: Evaluation of PBL parameterizations for modeling surface wind speed during storms in the northeast United States. *Wea. Forecasting*, **31**, 1511–1528, doi:[10.1175/WAF-D-15-0139.1](https://doi.org/10.1175/WAF-D-15-0139.1).
- Gego, E., C. Hogrefe, G. Kallos, A. Voudouri, J. S. Irwin, and S. T. Rao, 2005: Examination of model predictions at different horizontal grid resolutions. *Environ. Fluid Mech.*, **5**, 63–85, doi:[10.1007/s10652-005-0486-3](https://doi.org/10.1007/s10652-005-0486-3).
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, doi:[10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2).
- , M. Peroutka, J. Wiedenfeld, J. Wagner, G. Zylstra, B. Schuknecht, and B. Jackson, 2009: MOS uncertainty estimates

- in an ensemble framework. *Mon. Wea. Rev.*, **137**, 246–268, doi:[10.1175/2008MWR2569.1](https://doi.org/10.1175/2008MWR2569.1).
- Grell, G. A., and D. Devenyi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, 1693, doi:[10.1029/2002GL015311](https://doi.org/10.1029/2002GL015311).
- Hacker, J. P., and D. L. Rife, 2007: A practical approach to sequential estimation of systematic error on near-surface mesoscale grids. *Wea. Forecasting*, **22**, 1257–1273, doi:[10.1175/2007WAF2006102.1](https://doi.org/10.1175/2007WAF2006102.1).
- Hart, K. A., W. J. Steenburgh, D. J. Onton, and A. J. Siffert, 2004: An evaluation of mesoscale-model-based model output statistics (MOS) during the 2002 Olympic and Paralympic Winter Games. *Wea. Forecasting*, **19**, 200–218, doi:[10.1175/1520-0434\(2004\)019<0200:AEOMMO>2.0.CO;2](https://doi.org/10.1175/1520-0434(2004)019<0200:AEOMMO>2.0.CO;2).
- He, J., D. W. Wanik, B. M. Hartman, E. N. Anagnostou, and M. Astitha, 2017: Nonparametric tree-based predictive modeling of storm damage to power distribution network. *Risk Anal.*, doi:[10.1111/risa.12652](https://doi.org/10.1111/risa.12652), in press.
- Homleid, M., 1995: Diurnal corrections of short-term surface temperature forecasts using the Kalman filter. *Wea. Forecasting*, **10**, 689–707, doi:[10.1175/1520-0434\(1995\)010<0689:DCOSTS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0689:DCOSTS>2.0.CO;2).
- Hong, S.-Y., Y. Noh, and J. Dudhia, 2006: A new vertical diffusion package with an explicit treatment of entrainment processes. *Mon. Wea. Rev.*, **134**, 2318–2341, doi:[10.1175/MWR3199.1](https://doi.org/10.1175/MWR3199.1).
- Hu, X. M., J. W. Nielsen-Gammon, and F. Zhang, 2010: Evaluation of three planetary boundary layer schemes in the WRF Model. *J. Appl. Meteor. Climatol.*, **49**, 1831–1844, doi:[10.1175/2010JAMC2432.1](https://doi.org/10.1175/2010JAMC2432.1).
- Idowu, O. S., and C. J. deW. Rautenbach, 2009: Model output statistics to improve severe storms prediction over western Sahel. *Atmos. Res.*, **93**, 419–425, doi:[10.1016/j.atmosres.2008.10.035](https://doi.org/10.1016/j.atmosres.2008.10.035).
- Jacks, E., J. B. Bower, V. J. Dagostaro, J. P. Dallavalle, M. C. Erickson, and J. C. Su, 1990: New NGM-based MOS guidance for maxima and minima temperature, probability of precipitation, cloud amount, and surface wind. *Wea. Forecasting*, **5**, 128–138, doi:[10.1175/1520-0434\(1990\)005<0128:NNBMGF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1990)005<0128:NNBMGF>2.0.CO;2).
- Kang, D., R. Mathur, and S. T. Rao, 2010: Assessment of bias-adjusted PM_{2.5} air quality forecasts over the continental United States during 2007. *Geosci. Model Dev.*, **3**, 309–320, doi:[10.5194/gmd-3-309-2010](https://doi.org/10.5194/gmd-3-309-2010).
- Kirtman, B. P., and Coauthors, 2014: The North American Multimodel Ensemble: Phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, doi:[10.1175/BAMS-D-12-00050.1](https://doi.org/10.1175/BAMS-D-12-00050.1).
- Koster, R. D., and M. J. Suarez, 2001: Soil moisture memory in climate models. *J. Hydrometeor.*, **2**, 558–570, doi:[10.1175/1525-7541\(2001\)002<0558:SMMICM>2.0.CO;2](https://doi.org/10.1175/1525-7541(2001)002<0558:SMMICM>2.0.CO;2).
- Krishnamurti, T. N., J. Sanjay, A. K. Mitra, and T. S. V. V. Kumar, 2004: Determination of forecast errors arising from different components of model physics and dynamics. *Mon. Wea. Rev.*, **132**, 2570–2594, doi:[10.1175/MWR2785.1](https://doi.org/10.1175/MWR2785.1).
- Kushta, J., G. Kallos, M. Astitha, S. Solomos, C. Spyrou, C. Mitsakou, and J. Lelieveld, 2014: Impact of natural aerosols on atmospheric radiation and consequent feedbacks with the meteorological and photochemical state of the atmosphere. *J. Geophys. Res. Atmos.*, **119**, 1463–1491, doi:[10.1002/2013JD020714](https://doi.org/10.1002/2013JD020714).
- Libonati, R., I. Trigo, and C. C. Dacamará, 2008: Correction of 2 m-temperature forecasts using Kalman filtering technique. *Atmos. Res.*, **87**, 183–197, doi:[10.1016/j.atmosres.2007.08.006](https://doi.org/10.1016/j.atmosres.2007.08.006).
- Louka, P., G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, 2008: Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering. *J. Wind Eng. Ind. Aerodyn.*, **96**, 2348–2362, doi:[10.1016/j.jweia.2008.03.013](https://doi.org/10.1016/j.jweia.2008.03.013).
- Mao, Q., R. T. McNider, S. F. Mueller, and H.-M. H. Juang, 1999: An optimal model output calibration algorithm suitable for objective temperature forecasting. *Wea. Forecasting*, **14**, 190–202, doi:[10.1175/1520-0434\(1999\)014<0190:AOMOCA>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<0190:AOMOCA>2.0.CO;2).
- Mass, C. F., J. Baars, G. Wedam, E. Grimit, and R. Steed, 2008: Removal of systematic model bias on a model grid. *Wea. Forecasting*, **23**, 438–459, doi:[10.1175/2007WAF2006117.1](https://doi.org/10.1175/2007WAF2006117.1).
- McCollor, D., and R. Stull, 2008: Hydrometeorological accuracy enhancement via post-processing of numerical weather forecasts in complex terrain. *Wea. Forecasting*, **23**, 131–144, doi:[10.1175/2007WAF2006107.1](https://doi.org/10.1175/2007WAF2006107.1).
- Mellor, G. L., and T. Yamada, 1982: Development of a turbulence closure model for geophysical fluid problems. *Rev. Geophys.*, **20**, 851–875, doi:[10.1029/RG020i004p00851](https://doi.org/10.1029/RG020i004p00851).
- Meyers, M. P., R. L. Walko, J. Y. Harrington, and W. R. Cotton, 1997: New RAMS cloud microphysics parameterization. Part II: The two-moment scheme. *Atmos. Res.*, **45**, 3–39, doi:[10.1016/S0169-8095\(97\)00018-5](https://doi.org/10.1016/S0169-8095(97)00018-5).
- Mlawer, E. J., S. J. Taubman, P. D. Brown, M. J. Iacono, and S. A. Clough, 1997: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-*k* model for the longwave. *J. Geophys. Res.*, **102**, 16 663–16 682, doi:[10.1029/97JD00237](https://doi.org/10.1029/97JD00237).
- Müller, M. D., 2011: Effects of model resolution and statistical postprocessing on shelter temperature and wind forecasts. *J. Appl. Meteor. Climatol.*, **50**, 1627–1636, doi:[10.1175/2011JAMC2615.1](https://doi.org/10.1175/2011JAMC2615.1).
- Murphy, A. H., 1988: Skill score based on the mean square error and their relationship to the correlation coefficient. *Mon. Wea. Rev.*, **116**, 2417–2424, doi:[10.1175/1520-0493\(1988\)116<2417:SSBOTM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2).
- Nenes, A., and J. H. Seinfeld, 2003: Parameterization of cloud droplet formation in global climate models. *J. Geophys. Res.*, **108**, 4415, doi:[10.1029/2002JD002911](https://doi.org/10.1029/2002JD002911).
- Nielsen-Gammon, J. W., X.-M. Hu, F. Zhang, and J. E. Pleim, 2010: Evaluation of planetary boundary layer scheme sensitivities for the purpose of parameter estimation. *Mon. Wea. Rev.*, **138**, 3400–3417, doi:[10.1175/2010MWR3292.1](https://doi.org/10.1175/2010MWR3292.1).
- NOAA/National Centers for Environmental Prediction, 2000: NCEP FNL Operational Model Global Tropospheric Analyses, continuing from July 1999 (updated daily). National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 15 August 2015, doi:[10.5065/D6M043C6](https://doi.org/10.5065/D6M043C6).
- , 2007: NCEP Global Forecast System (GFS) Analyses and Forecasts. National Center for Atmospheric Research Computational and Information Systems Laboratory Research Data Archive, accessed 15 August 2015. [Available online at <http://rda.ucar.edu/datasets/ds084.6/>].
- O'Hagan, A., and J. J. Forster, 2004: *Bayesian Inference*. Vol. 2B, *Kendall's Advanced Theory of Statistics*, Arnold, 496 pp.
- Palmer, T. N., F. J. Doblas-Reyes, A. Weisheimer, and M. J. Rodwell, 2008: Toward seamless prediction: Calibration of climate change projections using seasonal forecasts. *Bull. Amer. Meteor. Soc.*, **89**, 459–470, doi:[10.1175/BAMS-89-4-459](https://doi.org/10.1175/BAMS-89-4-459).

- Papadopoulos, A., E. Serpetzoglou, and E. N. Anagnostou, 2008: Improving NWP through radar rainfall-driven land surface parameters: A case study on convective precipitation forecasting. *Adv. Water Resour.*, **31**, 1456–1469, doi:10.1016/j.advwatres.2008.02.001.
- Pielke, R. A., and Coauthors, 1992: A comprehensive meteorological modeling system—RAMS. *Meteor. Atmos. Phys.*, **49**, 69–91, doi:10.1007/BF01025401.
- Pleim, J. E., 2007: A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. *J. Appl. Meteor. Climatol.*, **46**, 1396–1409, doi:10.1175/JAM2534.1.
- Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, **133**, 1155–1174, doi:10.1175/MWR2906.1.
- Rincon, A., O. Jorba, and J. M. Baldasano, 2010: Development of a short-term irradiance prediction system using post-processing tools on WRF-ARW meteorological forecasts in Spain. *Extended Abstracts, 10th European Conf. on Applied Meteorology*, Zurich, Switzerland, European Meteorological Society, EMS2010-406. [Available online at <http://meetingorganizer.copernicus.org/EMS2010/EMS2010-406-1.pdf>.]
- Roberts, N. M., 2003: Results from high-resolution simulations of convective events. Met Office Tech. Rep. 402, 47 pp.
- Roeger, C., R. Stull, D. McClung, J. Hacker, X. Deng, and H. Modzelewski, 2003: Verification of mesoscale numerical weather forecast in mountainous terrain for application to avalanche prediction. *Wea. Forecasting*, **18**, 1140–1160, doi:10.1175/1520-0434(2003)018<1140:VOMNWF>2.0.CO;2.
- Schwartz, C. S., and Coauthors, 2009: Next-day convection-allowing WRF Model guidance: A second look at 2-km versus 4-km grid spacing. *Mon. Wea. Rev.*, **137**, 3351–3372, doi:10.1175/2009MWR2924.1.
- Serpetzoglou, E., E. N. Anagnostou, A. Papadopoulos, E. I. Nikolopoulos, and V. Maggioni, 2010: Error propagation of remote sensing rainfall estimates in soil moisture prediction from a land surface model. *J. Hydrometeor.*, **11**, 705–720, doi:10.1175/2009JHM1166.1.
- Simmons, A., S. Uppala, D. Dee, and S. Kobayashi, 2007: ERA-Interim: New ECMWF reanalysis products from 1989 onwards. *ECMWF Newsletter*, No. 110, ECMWF, Reading, United Kingdom, 25–35. [Available online at <http://www.ecmwf.int/sites/default/files/elibrary/2006/14615-newsletter-no110-winter-200607.pdf>.]
- Skamarock, W. C., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., doi:10.5065/D68S4MVH.
- Sloughter, J. McL., A. E. Raftery, T. Gneiting, and C. Fraley, 2007: Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 3209–3220, doi:10.1175/MWR3441.1.
- , T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, **105**, 25–35, doi:10.1198/jasa.2009.ap08615.
- Solomos, S., G. Kallos, J. Kushta, M. Astitha, C. Tremback, A. Nenes, and Z. Levin, 2011: An integrated modeling study on the effects of mineral dust and sea salt particles on clouds and precipitation. *Atmos. Chem. Phys.*, **11**, 873–892, doi:10.5194/acp-11-873-2011.
- Sorensen, D., and D. Gianola, 2002: *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics*. Springer, 740 pp.
- Speer, M. S., L. M. Leslie, and L. Qi, 2003: Numerical prediction of severe convection: Comparison with operational forecasts. *Meteor. Appl.*, **10**, 11–19, doi:10.1017/S1350482703005024.
- Stensrud, D. J., and N. Yussouf, 2003: Short-range ensemble predictions of 2-m temperature and dewpoint temperature over New England. *Mon. Wea. Rev.*, **131**, 2510–2524, doi:10.1175/1520-0493(2003)131<2510:SEPOMT>2.0.CO;2.
- , and —, 2005: Bias-corrected short-range ensemble forecasts of near surface variables. *Meteor. Appl.*, **12**, 217, doi:10.1017/S135048270500174X.
- Steppele, J., G. Doms, U. Schättler, H. W. Bitzer, A. Gassmann, U. Damrath, and G. Gregoric, 2003: Meso-gamma scale forecasts using the nonhydrostatic model LM. *Meteor. Atmos. Phys.*, **82**, 75–96, doi:10.1007/s00703-001-0592-9.
- Sweeney, C. P., P. Lynch, and P. Nolan, 2013: Reducing errors of wind speed forecasts by an optimal combination of post-processing methods. *Meteor. Appl.*, **20**, 32–40, doi:10.1002/met.294.
- Taylor, K. E., 2001: Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, **106**, 7183–7192, doi:10.1029/2000JD900719.
- Tewari, M., and Coauthors, 2004: Implementation and verification of the unified Noah land surface model in the WRF Model. *20th Conf. on Weather Analysis and Forecasting/16th Conf. on Numerical Weather Prediction*, Seattle, WA, Amer. Meteor. Soc., 14.2A. [Available online at <https://ams.confex.com/ams/pdfpapers/69061.pdf>.]
- Thompson, G., P. R. Field, R. M. Rasmussen, and W. D. Hall, 2008: Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, **136**, 5095–5115, doi:10.1175/2008MWR2387.1.
- Walko, R. L., W. Cotton, M. Meyers, and J. Harrington, 1995: New RAMS cloud microphysics parameterization. Part I: The single-moment scheme. *Atmos. Res.*, **38**, 29–62, doi:10.1016/0169-8095(94)00087-T.
- , and Coauthors, 2000: Coupled atmosphere–biophysics–hydrology models for environmental modeling. *J. Appl. Meteor.*, **39**, 931–944, doi:10.1175/1520-0450(2000)039<0931:CABHMF>2.0.CO;2.
- Walter, G., and T. Augustin, 2010: Bayesian linear regression—Different conjugate models and their (in)sensitivity to prior-data conflict. *Statistical Modelling and Regression Structures*, T. Kneib and G. Tutz, Eds., Springer Physica-Verlag, 59–78.
- , —, and A. Peters, 2007: Linear regression analysis under sets of conjugate priors. *Proc. Fifth Int. Symp. on Imprecise Probabilities and Their Applications*, Prague, Czech Republic, Society for Imprecise Probability: Theories and Applications, 445–455. [Available online at <http://www.sipta.org/isipta07/proceedings/proceedings-optimised.pdf>.]
- Wang, X., D. Parrish, D. Kleist, and J. Whitaker, 2013: GSI 3DVar-based ensemble–variational hybrid data assimilation for NCEP Global Forecast System: Single-resolution experiments. *Mon. Wea. Rev.*, **141**, 4098–4117, doi:10.1175/MWR-D-12-00141.1.
- Wanik, D. W., E. Anagnostou, B. M. Hartman, M. E. Frediani, and M. Astitha, 2015: Storm outage modeling for an electric distribution network in northeastern USA. *Nat. Hazards*, **79**, 1359–1384, doi:10.1007/s11069-015-1908-2.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2009: Seasonal ensemble forecasts: Are recalibrated single models better than multimodels? *Mon. Wea. Rev.*, **137**, 1460–1479, doi:10.1175/2008MWR2773.1.

- Weisberg, S., 2005: *Applied Linear Regression*. John Wiley and Sons, 310 pp.
- Wilczak, J. M., S. A. McKeen, I. Djalalova, and G. Grell, 2006: Bias-corrected ensemble and probabilistic forecasts of surface ozone over eastern North America during summer of 2004. *J. Geophys. Res.*, **111**, D23S28, doi:[10.1029/2006JD007598](https://doi.org/10.1029/2006JD007598).
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Sciences*. Academic Press, 467 pp.
- , and T. M. Hamill, 2007: Comparison of ensemble-MOS methods using GFS reforecasts. *Mon. Wea. Rev.*, **135**, 2379–2390, doi:[10.1175/MWR3402.1](https://doi.org/10.1175/MWR3402.1).
- Wilson, L. J., and M. Vallée, 2002: The Canadian updateable model output statistics (UMOS) system: Design and development tests. *Wea. Forecasting*, **17**, 206–222, doi:[10.1175/1520-0434\(2002\)017<0206:TCUMOS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2002)017<0206:TCUMOS>2.0.CO;2).
- , and —, 2003: The Canadian updateable model output statistics (UMOS) system: Validation against perfect prog. *Wea. Forecasting*, **18**, 288–302, doi:[10.1175/1520-0434\(2003\)018<0288:TCUMOS>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0288:TCUMOS>2.0.CO;2).
- , S. Beauregard, A. E. Raftery, and R. Verret, 2007: Calibrated surface temperature forecasts from the Canadian Ensemble Prediction System using Bayesian model averaging. *Mon. Wea. Rev.*, **135**, 1364–1385, doi:[10.1175/MWR3347.1](https://doi.org/10.1175/MWR3347.1).
- Zellner, A., 1996: *An Introduction to Bayesian Inference in Econometrics*. John Wiley and Sons, 431 pp.