Statistically and Dynamically Downscaled, Calibrated, Probabilistic 10-m Wind Vector Forecasts Using Ensemble Model Output Statistics

BRYAN P. HOLMAN

Department of Ocean Engineering and Sciences, Florida Institute of Technology, Melbourne, Florida

STEVEN M. LAZARUS

Department of Physics and Space Sciences, Florida Institute of Technology, Melbourne, Florida

MICHAEL E. SPLITT

College of Aeronautics, Florida Institute of Technology, Melbourne, Florida

ABSTRACT

A computationally efficient method is developed that performs gridded postprocessing of ensemble 10-m wind vector forecasts. An expansive set of idealized WRF Model simulations are generated to provide physically consistent, high-resolution winds over a coastal domain characterized by an intricate land/water mask. The ensemble model output statistics (EMOS) technique is used to calibrate the ensemble wind vector forecasts at observation locations. The local EMOS predictive parameters (mean and variance) are then spread throughout the grid utilizing flow-dependent statistical relationships extracted from the downscaled WRF winds. In a yearlong study, the method is applied to 24-h wind forecasts from the Global Ensemble Forecast System (GEFS) at 28 east-central Florida stations. Compared to the raw GEFS, the approach improves both the deterministic and probabilistic forecast skill. Analysis of multivariate rank histograms indicates that the postprocessed forecasts are calibrated. A downscaling case study illustrates the method as applied to a quiescent easterly flow event. Strengths and weaknesses of the approach are presented and discussed.

1. Introduction

As numerical weather prediction (NWP) models continue to improve, increasing value is being recognized in quantifying not only the best guess of a future atmospheric state (deterministic forecast) but also the uncertainty in that forecast (Gneiting and Raftery 2005). To that end, ensemble systems have been developed at many weather prediction centers. Ensembles account for uncertainty by varying dynamical cores, initial and/or boundary conditions, or model physics (Mass 2003). Because the spread between ensemble solutions has been shown to be related to the forecast error, a number of spread–skill relationships have been developed (e.g., Whitaker and Loughe 1998; Grimit and Mass 2007). However, in order to realize the full potential of

ensemble output, the forecasts must be reliable and free of systematic error. Unfortunately, this is not generally the case as current ensemble systems are inherently biased and suffer from underdispersion (i.e., observations frequently fall outside the range of individual ensemble member solutions) (Toth et al. 1997; Buizza et al. 2005).

These issues can be addressed by statistically postprocessing ensemble output (Gneiting and Raftery 2005). The goal of ensemble postprocessing is to create probabilistic forecasts that are sharp (i.e., reduced spread) and calibrated, meaning forecast probabilities are consistent with observations (Gneiting 2014). Two state-of-the-art ensemble calibration methods are Bayesian model averaging (BMA; Raftery et al. 2005), and ensemble model output statistics (EMOS; Gneiting et al. 2005). These approaches convert ensemble forecasts into predictive probability density functions (PDF) using training data comprising recent forecast errors. The predictive PDF parameters, its mean, and variance, represent bias- and

Corresponding author: Bryan P. Holman, bholman2013@my.fit.edu

dispersion-corrected functions of the ensemble mean and ensemble variance, respectively (Gneiting 2014).

In this study we focus on calibration methods for ensemble wind forecasts. BMA has been used for wind direction (Bao et al. 2010) and both BMA and EMOS have been adapted for wind speed (Sloughter et al. 2010; Thorarinsdottir and Gneiting 2010). Yet wind speed and direction are not independent, and preserving the relationship between the two is essential for many applications (Schuhen et al. 2012). Given that forecasting surface winds is a multidimensional problem, bivariate BMA (Sloughter et al. 2013) and EMOS (Schuhen et al. 2012) methods have been developed to calibrate ensemble wind vector forecasts. Alternatively, Pinson (2012) introduced a recursive and adaptive method for wind vectors that preserves the rank dependence structure of the original ensemble, resulting in a discrete ensemble as opposed to a predictive density.

Regardless of methodology, forecasts calibrated using point observations are only available at station locations, while postprocessed forecasts are needed everywhere (Gneiting and Raftery 2005; Mass et al. 2009). A number of different approaches have been developed that attempt to address this. For example, in lieu of observations, a gridded product (analysis) can be used to develop the calibration model, as done for temperature by Kann et al. (2009) and for wind vectors by Pinson (2012). However, high-resolution analyses are not generally available everywhere (Mass et al. 2008), and resulting postprocessed forecasts can inherit systematic errors present in the analysis product (Gneiting 2014). As an alternative, gridded training datasets can be constructed using forecast errors from point stations that share related characteristics (Mass et al. 2008) to which a calibration model is then fit (Kleiber et al. 2011). Mass et al. (2009) utilize this latter approach to provide operational, calibrated, probabilistic forecasts of temperature and precipitation over the Pacific Northwest. One of the more popular approaches performs a local calibration (i.e., at observation locations) and then spreads the predictive PDF parameters onto a grid using geostatistical methods. This technique has been applied extensively to surface temperature forecasts (Berrocal et al. 2007; Kleiber et al. 2011; Scheuerer and Büermann 2014; Scheuerer and König 2014; Feldmann et al. 2015), and recently to forecasts of wind speed (Scheuerer and Möller 2015).

We introduce a unique approach to the ensemble wind forecast problem that focuses on an area with a relatively complex land/water mask. The study area comprises a coastal estuary in east-central Florida, the Indian River Lagoon (IRL), whose intricate coastline and narrow width requires downscaling in order to be resolved by current ensemble systems (Holman et al. 2017). The technique we present seeks to combine the benefits of dynamical downscaling (i.e., physically consistent flow dependencies) and statistical postprocessing (i.e., calibration using recent forecast errors) using a three-step process that includes 1) an antecedent application of the Weather Research and Forecasting (WRF) Model (Skamarock et al. 2008) to downscale an expansive set of idealized wind simulations to high resolution (333-m horizontal grid spacing), 2) using bivariate EMOS to calibrate ensemble wind vector forecasts at observing stations, and 3) the spreading of EMOS predictive parameters (i.e., mean and variance) from station locations to nearby observation-free locations using directionally dependent statistical relationships extracted from the downscaled WRF wind fields.

The remainder of this paper is as follows. Section 2 describes the ensemble wind vector forecasts and the verifying observations. Section 3 motivates the use of bivariate EMOS to calibrate these forecasts and briefly outlines its methodology. Section 4 details the idealized WRF downscaling framework and discusses how the statistical relationships are determined and then used to spread the EMOS parameters. A data withdrawal experiment is performed in section 5 to test the procedure and the results are presented in section 6. Section 7 presents a summary and discussion of the results and offers suggestions for future work.

2. Data

a. Observations

In this study we verify wind vector forecasts at both land- and water-based stations. As will be shown in section 4b, one objective of the method introduced here is the ability to estimate winds over the IRL using nearby observations, regardless of station type (i.e., land, shoreline, or open water). A total of 28 quality-controlled observing stations were selected from 4 observation networks for this purpose (Fig. 1b): 11 from WeatherFlow Inc. (http://www.weatherflow.com), 10 Automated Surface Observing Stations (ASOS), 4 from the South Florida Water Management District (SFWMD), and 3 stations from the National Data Buoy Center (NDBC). In total, 12 of the stations are on land, 8 are located near shorelines, and 8 are over water. In general these stations border the east coast of Florida, with the exception of the four Lake Okeechobee stations, Okeechobee airport, and buoy 41009 in the Atlantic Ocean (Fig. 1b).

For forecast verification purposes the reported wind speed and wind direction at these stations are broken into vector components. Only the most recent observations (within a 10-min window) to the 6-hourly GEFS output
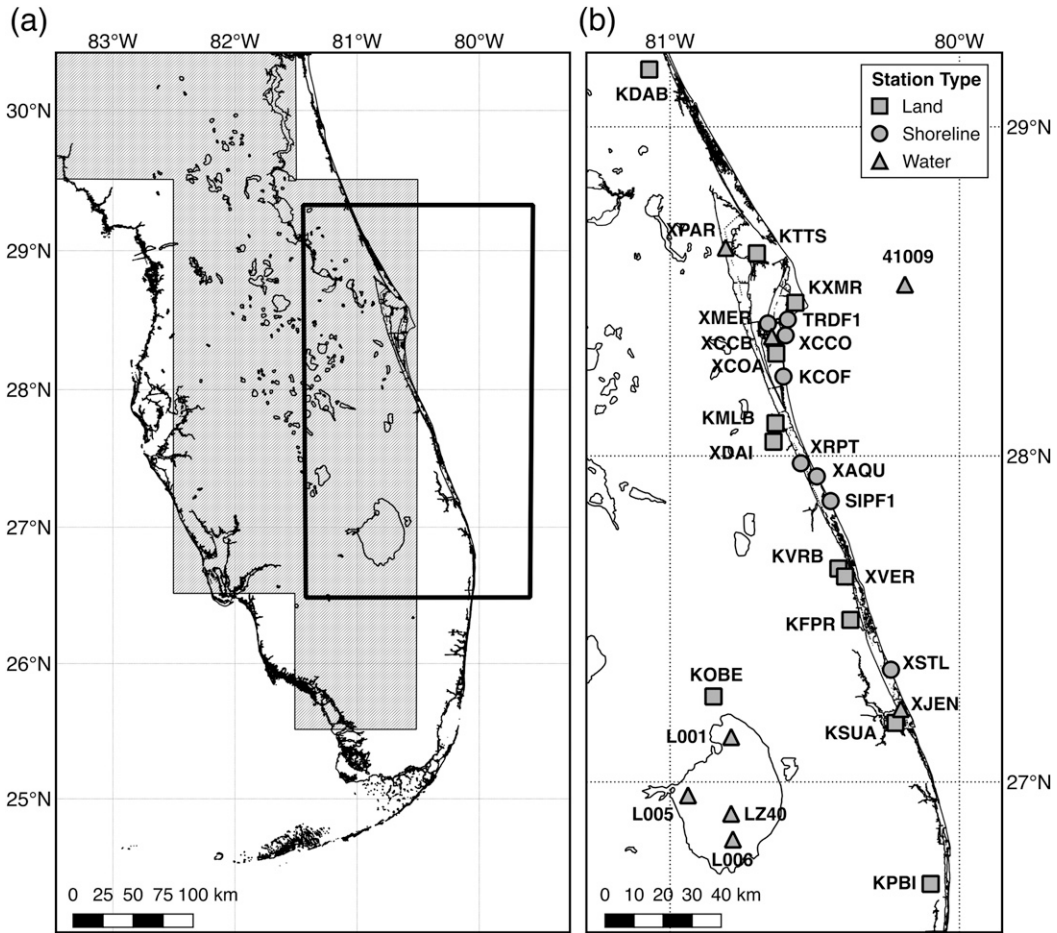
FIG. 1. The Florida Peninsula with (a) land mask for the GEFS ensemble system (gray shading), and the domain used by the WRF Model (thick black box), and (b) locations of the 28 stations used in this study. The type of station (land, shoreline, open water) is depicted by the symbol shape as portrayed in the legend (top right).

are used. Wind reporting practices vary across the four networks used here. For example, ASOS stations, sourced here from hourly METARs, report 2-min-averaged wind speed and direction rounded to the nearest knot (1 kt = 0.5144 m s^{-1}) and 10°, respectively, and calm winds (<3 kt) are reported as zero (Nadolski 1998). In contrast, WeatherFlow Inc. stations report 5-min-averaged winds to the nearest 0.01 m s^{-1} and degree. Despite these differences, no conversion between observations with competing averaging periods was made since each represents an unbiased estimate of the true mean wind (Harper et al. 2010). Anemometer heights are not 10 m at two land, four shoreline, and all eight open water stations. At these stations 10-m equivalent winds are calculated following Hsu et al. (1994).

b. Model output

Ensemble surface (10 m) wind vector forecasts for this study were obtained from the National Centers for Environmental Prediction (NCEP) Global Ensemble Forecast System (GEFS). The GEFS cycles every 6 hours (0000, 0600, 1200, and 1800 UTC), producing global output on a 1.0° latitude–longitude grid. The GEFS grid spacing over Florida is roughly 100 km, resulting in a coarse-resolution land mask over the region (Fig. 1a). We mine the GEFS 24-h wind vector forecasts during a 13-month period ending 31 December 2015. This allows for verification of postprocessed forecasts (which use a 30-day training window) across the 2015 calendar year. The GEFS forecasts are bilinearly interpolated to the location of each observing station.

3. Bivariate EMOS for wind vectors

The aim of this study is to calibrate 24-h ensemble wind vector forecasts at observing stations and spread the model parameters (mean and variance) to nearby, observation-free locations. We employ ensemble model

output statistics (EMOS; Gneiting et al. 2005) for wind vectors for local calibration as implemented by Junk et al. (2014). The performance of EMOS is competitive with Bayesian model averaging (BMA; Raftery et al. 2005; Sloughter et al. 2013), but is more computationally efficient (Thorarinsdottir and Gneiting 2010; Junk et al. 2014; Feldmann et al. 2015).

The EMOS postprocessing method used here utilizes the ensemble mean and spread from GEFS wind vector forecasts to output a bivariate normal predictive PDF. The bivariate normal PDF for a wind vector (u, v) is given [Eq. (4.33) in Wilks (2011)] as follows:

$$f(u, v) = \frac{1}{2\pi\sigma_u\sigma_v\sqrt{1-\rho_{uv}^2}} \exp\left\{-\frac{1}{2(1-\rho_{uv}^2)}\left[\frac{(u-\mu_u)^2}{\sigma_u^2}\right.\right.$$

$$\left.\left. -2\rho_{uv}\frac{(u-\mu_u)(v-\mu_v)}{\sigma_u\sigma_v} + \frac{(v-\mu_v)^2}{\sigma_v^2}\right]\right\}. \quad (1)$$

Equation (1) contains five parameters that require estimation: the u and v wind component predictive means (μ_u and μ_v), variances (σ_u^2 and σ_v^2), and correlation (ρ_{uv}). In the EMOS framework these parameters represent modified versions of ensemble forecasts that have been statistically optimized given verifying observations. Specifically, μ_u and μ_v represent bias-corrected components of the ensemble mean wind vector, σ_u^2 and σ_v^2 represent the ensemble spread (variance) that has been scaled to correct for dispersion errors, and ρ_{uv} represents the estimated correlation between the ensemble mean wind components.

In their implementations of bivariate EMOS, Schuhen et al. (2012) and Junk et al. (2014) begin by first estimating ρ_{uv} offline (i.e., using an independent dataset) at each observation location. The correlation is expressed as a trigonometric function of the ensemble mean wind direction whose parameters are determined via weighted nonlinear least squares. Here, ρ_{uv} is set to zero because, for the stations investigated in this study, estimations via its explicit modeling yielded only minimal differences in calibration metrics (not shown).

Following Junk et al. (2014), the predictive means μ_u and μ_v of the bivariate density are defined as functions of the ensemble mean wind components ($\overline{u}$ and $\overline{v}$):

$$\mu_u = b_{0u} + b_{1u}\overline{u} + b_{2u}\overline{v} \quad \text{and} \quad \mu_v = b_{0v} + b_{1v}\overline{u} + b_{2v}\overline{v}. \quad (2)$$

The three coefficients for each wind component are first estimated by regressing $\overline{u}$ and $\overline{v}$ against observed wind components using a running training window of 30 days, with the constraint that forecast–observation pairs are valid at the same hour as the current forecast. Thus, the

predictive means represent bias-corrected versions of the ensemble mean wind components.

The variance parameters σ_u^2 and σ_v^2 are estimated following Schuhen et al. (2012):

$$\sigma_u^2 = c_u + d_u s_u^2 \quad \text{and} \quad \sigma_v^2 = c_v + d_v s_v^2, \quad (3)$$

where s_u^2 and s_v^2 are the variances in the u and v components across m ensemble members defined by

$$s_u^2 = \frac{1}{m}\sum_{i=1}^{m}(u_i - \overline{u})^2 \quad \text{and} \quad s_v^2 = \frac{1}{m}\sum_{i=1}^{m}(v_i - \overline{v})^2. \quad (4)$$

For the GEFS ensemble, $m = 21$. The coefficients c_u, d_u, c_v, and d_v in Eq. (3), referred to as dispersion parameters, are constrained to be nonnegative and are estimated following the maximum likelihood framework outlined by Schuhen et al. (2012). In general, the σ_u^2 and σ_v^2 parameters that result are much greater than the individual ensemble variances themselves, sometimes by two orders of magnitude or more. The parameter estimates are then applied [in Eq. (1)] to produce a bivariate normal predictive density function that is optimized based on recent forecast errors.

4. Interpolated EMOS

a. Idealized WRF simulations

The WRF simulations are designed to provide a physically consistent, high-resolution surface (10 m) wind field that can be used to downscale a coarse-resolution ensemble and spread locally calibrated estimates of its mean and variance to nearby observation-free locations. The functionality of the WRF Model in this process is manifest in its surface physics and resolution, which provide a means of extracting flow-dependent statistical relationships associated with the complex land–water interface of a coastal estuary.

A total of 180 simulations were run using the Advanced Research version of WRF (WRF-ARW) version 3.8 (Skamarock et al. 2008). Model setup follows that of the convective-radiation ideal case (*em_convrad*) added in WRF, version 3.7 (see http://www2.mmm.ucar.edu/wrf/users/wrfv3.7/updates-3.7.html), with modifications described as follows. To encompass the entirety of the IRL and nearby observing stations we use a 550 (*nx*) by 950 (*ny*) lambert conic conformal grid with horizontal spacing equal to 333 m. The resulting horizontal domain is roughly $180 \times 315 \text{ km}^2$ (thick black box in Fig. 1a) with open boundaries on all sides. Land-use and terrain information are determined using the WRF preprocessing system and the ~30-m 2011 NLCD land-use dataset (Homer et al. 2015). In the vertical, model top is

set at 30 km with 35 eta levels, left unchanged from the *em_convrad* default settings.

The purpose for conducting these simulations is to resolve mean wind fields in quasi-equilibrium with the heterogeneous land surface under neutral atmospheric conditions. To this end, we neglect diurnal and radiative effects by turning off the longwave and shortwave physics options. As a result, sea and lake breezes do not appear in the model output. Cloud and microphysics options are also turned off to limit boundary interactions that result from such features.

WRF surface winds are sensitive to PBL scheme selection (Akylas et al. 2007; Cheng et al. 2013). For this study we parameterize PBL and surface layer processes with the Yonsei University (YSU; Hong et al. 2006) and accompanying MM5 similarity (Beljaars 1995) schemes, respectively. The YSU scheme was selected over the Mellor–Yamada–Janjić (MYJ; Janjić 1994) because preliminary analysis showed a larger wind speed contrast between land and water grid cells (not shown). Simulations are initialized from an idealized sounding, with a land and sea surface pressure of 1000 hPa and a temperature of 300.65 K (the annually averaged high temperature for Melbourne, Florida). The thermodynamic profile is identical for all simulations with neutral stability ($d\theta/dz = 0$) in the lowest kilometer and stable conditions above. The 180 simulations are defined by systematically varying the initial wind direction (18 total, every 20° starting at due north) and speed (10 total, including 1.0, 2.5, 3.75, 5.0, 6.25, 7.5, 8.75, 10.0, 12.5, and 15.0 m s^{-1}). The initial wind profiles have no directional or speed shear and thus are represented as a single point on a hodograph. The only difference in the simulations is the initial values of the u and v components.

The simulations were run until surface winds reached a quasi–steady state (i.e., when near-surface wind speed and direction essentially ceased changing in response to the surface characteristics). The adjustment process, which takes on the order of 6 h, results in a decrease in magnitude and a subsequent cross-isobaric down-gradient shift in direction that is more pronounced for land-based grid cells. An example of this process is shown for the WRF simulation initialized with WNW (300°) winds at 8.75 m s^{-1} (Fig. 2). At 15 min into the simulation, the initial wind direction has changed little but speeds have decreased to 7.5 m s^{-1} over the ocean and 3.5 m s^{-1} over portions of the peninsula. By 6 h, 10-m wind speeds have decreased substantially (to approximately 2 m s^{-1}) over the peninsula and have backed to a more westerly flow. Downwind of the peninsula, over the open ocean, the winds accelerate in the nearshore and gradually veer to the WNW. To mitigate the impact of turbulent structures on the mean flow

(of primary interest here), each simulation is extended an additional hour (to 7 h) over which the average wind components are calculated. Thus, the mean wind field is estimated as 1-h average wind components after equilibrium.

The result of these simulations compose a large multidimensional database populated by high-resolution (333 m) idealized surface wind fields. These data are used to estimate directionally dependent relationships between WRF grid cells with and without observations as described in the next section.

b. Using WRF output to spread EMOS parameters

This section outlines the process of utilizing the WRF winds to spread PDF parameters. This process is performed on the fly, after applying bivariate EMOS locally, using a rolling 30-day training period. Local EMOS (hereafter LEMOS) is carried out at the 28 stations introduced previously with ρ_{uv} assumed to be zero (section 3). The remaining predictive distribution parameters (μ_u, μ_v, σ_u^2, and σ_v^2) are spread as follows.

1) SPREADING THE MEAN PARAMETERS

The directionally dependent relationships used to spread mean parameters are developed by first identifying the WRF simulation most analogous to an observing station's current LEMOS mean forecast. This is accomplished by minimizing the Euclidean norm in $\Re^2$ between this forecast vector and the 180 WRF vectors from the grid cell containing this station. The regression (linear model) is then populated by selecting the WRF simulations with the same initial direction as that obtained via the distance metric. Given the inherent directional variability, the winds from the two adjacent direction bins are also included. The 30 simulations (i.e., 10 wind speeds per direction, 3 direction bins) are then used to separately regress the u and v components at the observation-containing grid cell (predictor) against those from the grid cell of interest (predictand). The observing station's predictive mean wind components (μ_u and μ_v) are then inserted into the directionally dependent regression equations, resulting in estimates of the predictive means at the grid cell of interest.

To provide an example, we detail the process of spreading the 0000 UTC 7 February 2015 LEMOS mean forecast at XRPT, a station on the IRL's western shore, to the grid cell containing the station KMLB. The choice of KMLB here is arbitrary, as information can be spread anywhere on the WRF grid. The WRF wind vectors for the grid cell that contains XRPT are shown in Fig. 3a. The LEMOS mean forecast at XRPT is depicted by the white filled triangle. This forecast is most consistent with simulations initialized with NE winds (40°, filled light
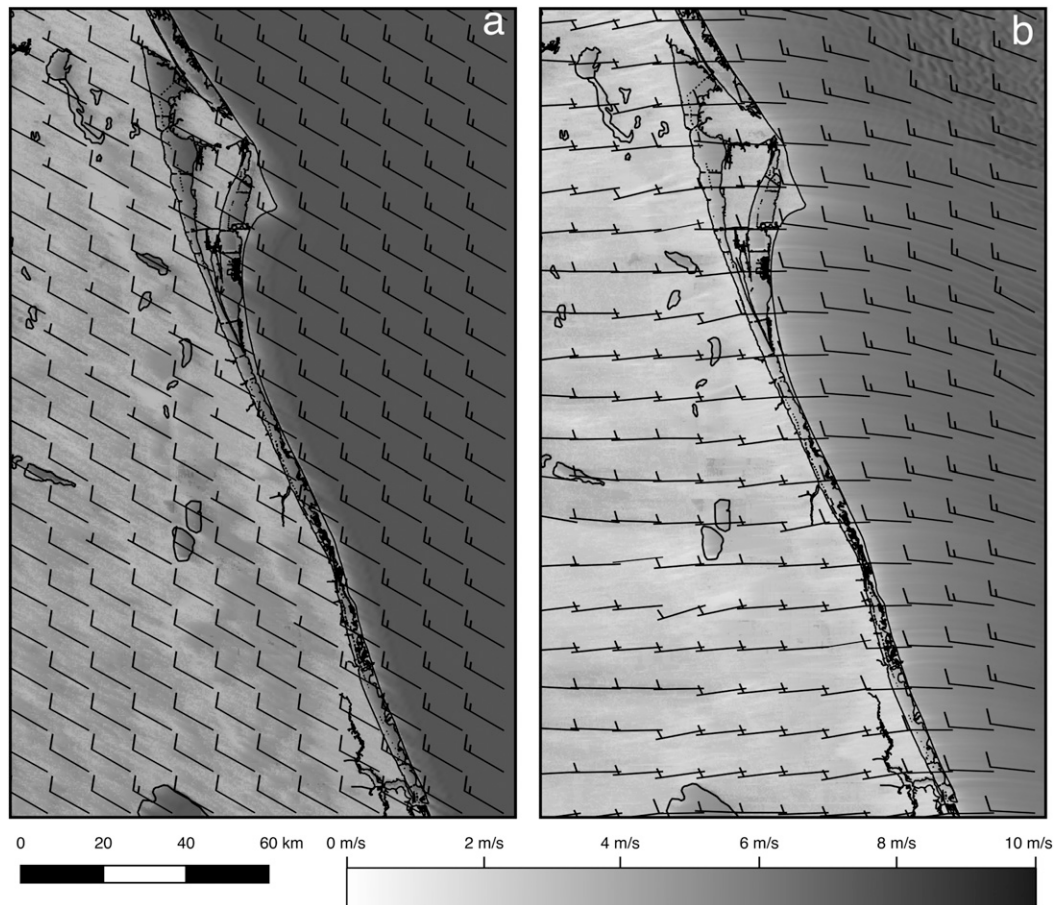
FIG. 2. WRF domain subset with wind speed (m s^{-1}, shaded contours) and associated wind barbs (full barb equals 5 m s^{-1}) after (a) 15 min and (b) 6 h of the simulation initialized with WNW winds (300°) at 8.75 m s^{-1}.

gray circles). The dark gray squares are the vectors initialized within the adjacent directional bins (i.e., 20° and 60°). Given the location of XRPT, the magnitude of the simulated onshore flow (easterly, u component less than zero) is nearly double that of the offshore (westerly, u component greater than zero). The u components of the simulated vectors at XRPT (x axis) and KMLB (y axis) are shown in Fig. 3b. The least squares fit (solid black line) yields a regression equation of $u_{\text{KMLB}} = 0.58 + 0.68u_{\text{XRPT}}$. This regressed fit is key, representing the directionally dependent relationship between the two station's WRF grid cells conditional on the XRPT LEMOS forecast. While determined using WRF data, this relationship is used to spread the u component from the LEMOS predictive mean forecast at XRPT to KMLB by inserting the current XRPT value ($\mu_u = -2.95\,\text{m s}^{-1}$, white filled triangle in Fig. 3a), yielding a μ_u estimate at KMLB of $-1.41\,\text{m s}^{-1}$ (white filled diamond). A similar regression using the v components yields a KMLB μ_v estimate of $-3.22\,\text{m s}^{-1}$ (not shown).

The value of the WRF here is providing gridded data across the study domain, and the process of first identifying relevant WRF data (given a postprocessed forecast) allows for directionally dependent relationships to be established. The process outlined here can estimate predictive means at any location on the WRF grid (521 001 cells) using a forecast from any observing station.

2) SPREADING THE VARIANCE PARAMETERS

The variance parameters are scaled according to the ratio of the average magnitude of the same 30 wind vectors at the grid cell of interest (KMLB) to that of the station-containing grid cell (XRPT). In this example the mean wind speeds for the WRF grid cells containing KMLB and XRPT are 2.16 and 4.05 m s^{-1}, respectively (a ratio of 0.53). Multiplying the ratio and the current estimate of the XRPT LEMOS variance parameters (3.25 and 5.00 m^2 s^{-2}) results in scaled estimates of σ_u^2 and σ_v^2 at KMLB of 1.73 and 2.66 m^2 s^{-2}, respectively.
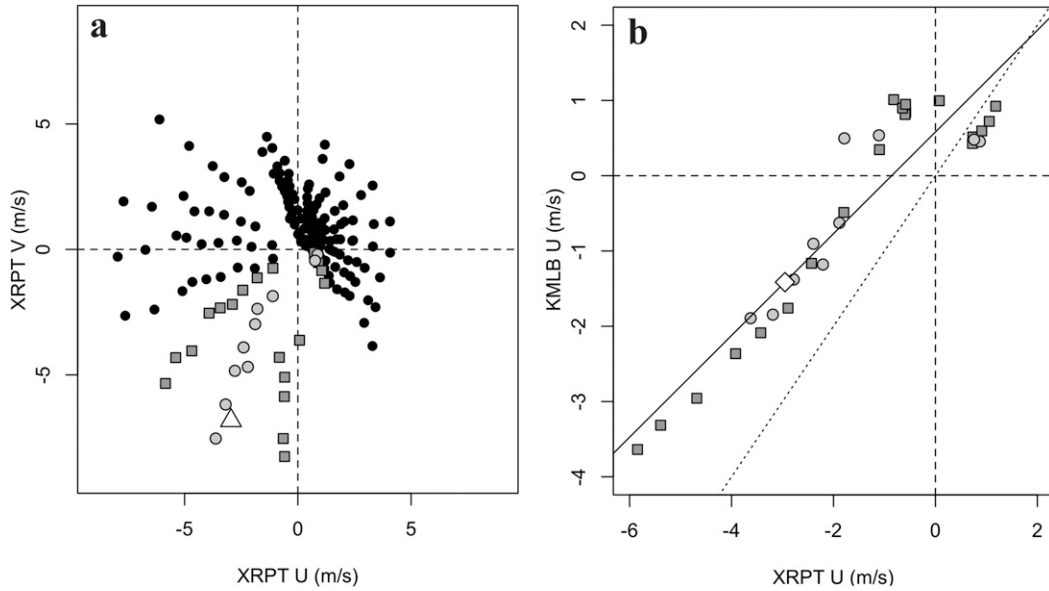
FIG. 3. (a) WRF wind vectors for the gridcell-containing station XRPT (filled circles). The 30 wind vectors (filled dark gray squares and light gray circles) determined using the direction of the 0000 UTC 7 Feb 2015 mean LEMOS forecast at XRPT (filled white triangle, see text for details). (b) The u components of the 30 wind vectors for the WRF KMLB and XRPT grid cells. The predictive μ_u estimate at KMLB (white filled diamond, $-1.41\,\mathrm{m\,s}^{-1}$) conditional on the corresponding predictive mean at XRPT ($-2.95\,\mathrm{m\,s}^{-1}$) lies on the least squares fit line (solid black). The dotted line represents a one-to-one fit.

This approach is motivated by the fact that, at the 28 stations investigated here, LEMOS variances are positively correlated with WRF wind speed. The median LEMOS σ_u^2 and σ_v^2 values (calendar year 2015, y axis) versus the median WRF wind speeds (all simulations, x axis) at the locations of the 28 observing stations are shown in Fig. 4. The variance roughly follows the one-to-one line (dashed), although more so for the median σ_u^2 values (black circles) than σ_v^2 (gray triangles). While the median values shown here encompass all wind directions, in practice the ratios are directionally dependent.

3) WEIGHING PARAMETER ESTIMATES

The relationships outlined in the two previous subsections are used to map the LEMOS parameters from the 28 stations to each cell (521 001 total) of the 333-m WRF grid. For each forecast cycle and grid cell, the LEMOS quantities (μ_u, μ_v, σ_u^2, and σ_v^2) are weighted to combine the 28 stations into a single estimate of each parameter. Following traditional data assimilation, the weights are defined to be inversely proportional to estimates of their precision. Three proxy measures were considered: the distance between a station's location and the grid cell of interest, a station's recent deterministic forecast error [as defined by the bivariate mean absolute error bMAE (see section 5b)

associated with LEMOS mean forecasts over the previous 30 days], and the standard error of the regressions used to spread the predictive means. Thus, for example, increased weighting would be assigned as a result of a station's 1) proximity, 2) higher LEMOS forecast skill (reduced bMAE), or 3) similarity to other stations (as defined by WRF-determined linear relationships).

Weights are determined by minimizing deterministic error (bMAE) across all 28 stations during a yearlong data withdrawal study (section 5a). The optimization process can be visualized by simultaneously considering the impact of the three precision quantities on the bMAE (Fig. 5a). The vertices represent full (no) weight for the respective (remaining two) schemes. The interpolated EMOS (hereafter TEMOS) forecasts are least skillful (bMAE 2.38 m s^{-1}, lower left) when parameter estimates are weighed solely according to recent LEMOS skill. The bMAE decreases as distance and WRF similarity receive progressively more weight, with maximum skill (bMAE 2.33 m s^{-1}) for weights of 85% and 15% (distance and WRF similarity, respectively). While the range of bMAE values is small (2.33–2.38 m s^{-1}), the difference between the lowest and highest values is significant with 95% confidence (not shown). Hence, for the data withdrawal experiment, the weights given to the zonal
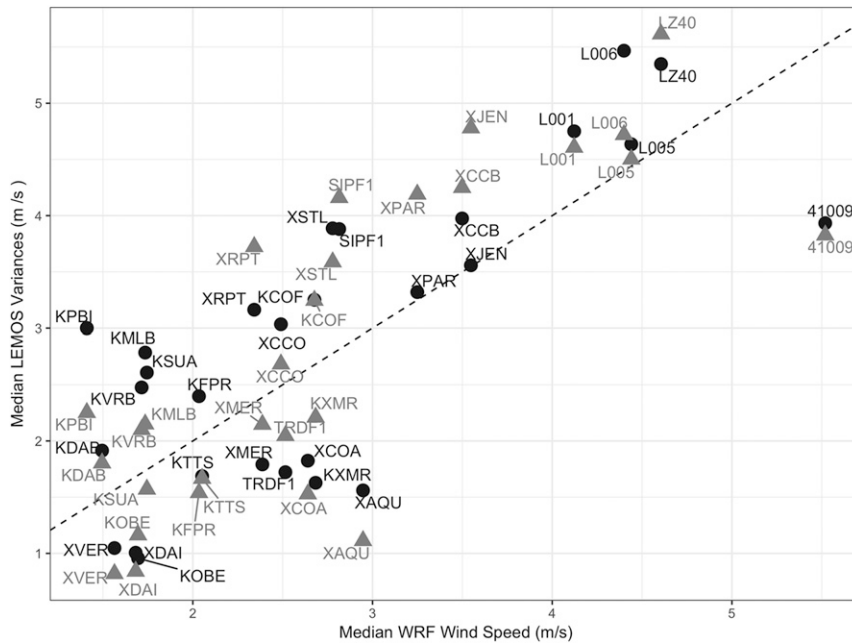
FIG. 4. Comparison of median LEMOS σ_u^2 (black circles) and σ_v^2 (gray triangles) values at each of the 28 stations for calendar year 2015 vs the median WRF wind speed at the corresponding station locations. The WRF median wind speeds were obtained by sampling from the full suite (180) of simulations at each of the station locations. The dashed line represents a one-to-one fit.

(w_u) and meridional (w_v) mean and variance estimates for station i are

$$w_{ui} = 0.85 \frac{\frac{1}{d_i}}{\sum_{j=1}^{k} \frac{1}{d_j}} + 0.15 \frac{\frac{1}{e_{ui}}}{\sum_{j=1}^{k} \frac{1}{e_{uj}}} \quad \text{and}$$

$$w_{vi} = 0.85 \frac{\frac{1}{d_i}}{\sum_{j=1}^{k} \frac{1}{d_j}} + 0.15 \frac{\frac{1}{e_{vi}}}{\sum_{j=1}^{k} \frac{1}{e_{vj}}}, \quad (5)$$

where k is the number of stations (i.e., 28); d_i is the distance between station i and the grid cell of interest; and e_{ui} and e_{vi} are the standard errors of regression models used to spread the μ_u and μ_v parameters, respectively [section 4b(1)].

While the bulk approach to optimizing the weights favors a combination of both distance and similarity, this is not necessarily the case for individual stations. For example, the bMAE at the suburban/forested station XDAI (near Melbourne, Florida; center of Fig. 1b) is lowest when estimates are weighted according to LEMOS skill and station similarity, prioritizing information from comparably protected stations (Fig. 5b). In contrast, the bMAE is lowest at Lake Okeechobee's LZ40 when

distance weighted only as this places more emphasis on the three remaining lake stations (Fig. 5c). The sensitivity is relatively large as the bMAE increases by almost 0.35 m s^{-1} depending on the choice of weighting.

5. Experimental setup

a. Design

The TEMOS approach is assessed using the 28 stations in a yearlong data withdrawal experiment in which an individual station's parameters are estimated from the other 27. TEMOS deterministic and probabilistic metrics are compared with those of the raw GEFS and the LEMOS estimated directly from the extracted station. Forecasts are verified four times per day (corresponding with GEFS cycle times) over calendar year 2015.

b. Verification

The skill of deterministic forecasts is gauged using the bMAE, which represents the magnitude of the average error vector between forecasts (f_i) and observations (o_i) across N cases:

$$\text{bMAE} = \frac{1}{N} \sum_{i=1}^{N} ||f_i - o_i||, \quad (6)$$

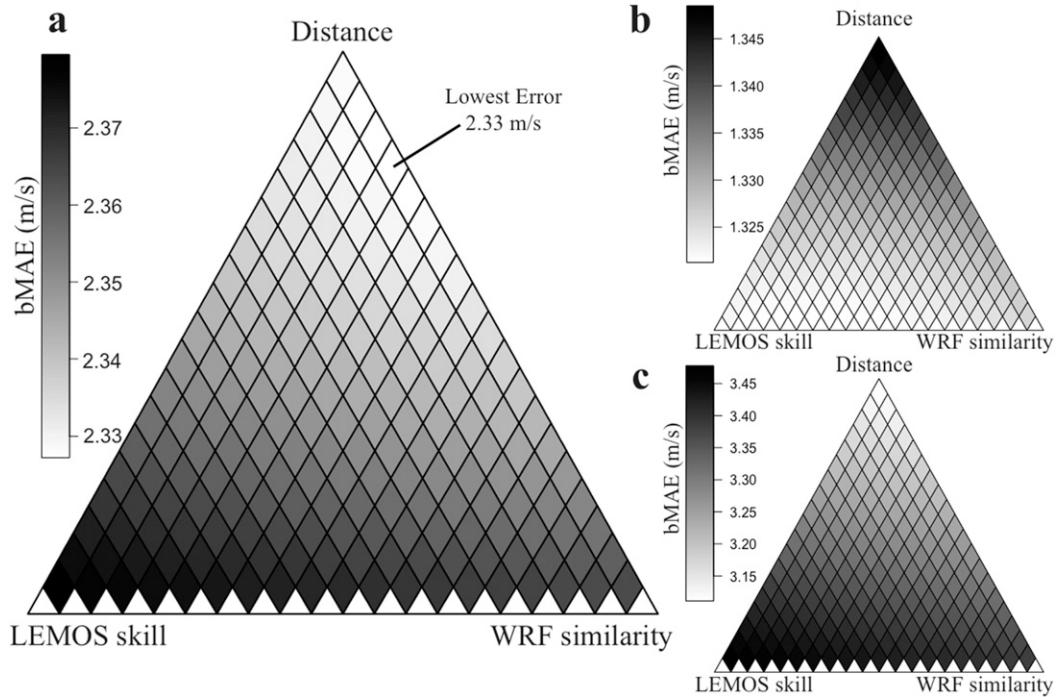where $||\cdot||$ denotes the Euclidean norm in $\Re^2$.

FIG. 5. TEMOS deterministic error as a function of the three weighting schemes (a) averaged across all stations and at (b) XDAI and (c) LZ40. The lower left (top, lower right) rhombus indicates the bMAE (m s^{-1}) that results when TEMOS parameters are exclusively weighed by LEMOS skill (distance, WRF similarity).

To quantify the skill of probabilistic forecasts it is customary to employ the energy score (ES; Gneiting and Raftery 2007), which jointly evaluates the calibration and sharpness of wind vector forecasts. For an ensemble forecast $F = F_{\mathrm{ens}}$ consisting of M members $f_1, \ldots, f_M \in \Re^2$ and verifying observation o, the ES is defined as

$$\mathrm{ES}(F_{\mathrm{ens}}, o) = \frac{1}{M} \sum_{m=1}^{M} ||f_m - o|| - \frac{1}{2M^2} \sum_{n=1}^{M} \sum_{m=1}^{M} ||f_n - f_m||.$$

$$(7)$$

Equation (7) is used to calculate the ES for raw individual GEFS forecasts. Since the TEMOS and LEMOS methods produce predictive PDFs, following Schuhen et al. (2012) we pull $K = 10\,000$ random samples from their respective bivariate densities and approximate the ES with

$$\mathrm{ES}(F_{\mathrm{ens}}, o) = \frac{1}{K} \sum_{k=1}^{K} ||f_k - o|| - \frac{1}{2(K - 1)} \sum_{k=1}^{K-1} ||f_k - f_{k+1}||.$$

$$(8)$$

Similar to bMAE, across N cases the mean energy score (hereafter ES) is

$$\mathrm{ES} = \frac{1}{N} \sum_{i=1}^{N} \mathrm{ES}(F_{\mathrm{ens},i}, o_i).$$

$$(9)$$

In addition, the reliability of the probabilistic forecasts is evaluated using multivariate rank histograms (Gneiting et al. 2008). As explained by Hamill (2001), a uniform rank histogram implies a calibrated ensemble, while sloped, U, and bell-shaped rank histograms indicate biased, underdispersive, and overdispersive ensembles, respectively. Examples of multivariate rank histograms are shown in section 6b.

6. Results

a. Deterministic verification

The following deterministic results are separated into two subsections, the first of which identifies the general strengths and weaknesses of the TEMOS approach by examining results averaged over all 28 stations. A more detailed analysis at select stations then follows.

1) GENERAL RESULTS

The bMAE associated with the raw GEFS ensemble mean (dark gray), TEMOS mean (gray), and LEMOS mean (light gray) forecasts is shown in Fig. 6a. Values represent averages across all stations as well as by station characterization. Error bars represent 95% confidence intervals determined by bootstrapping. Raw GEFS error ("none" category) is lowest at land stations
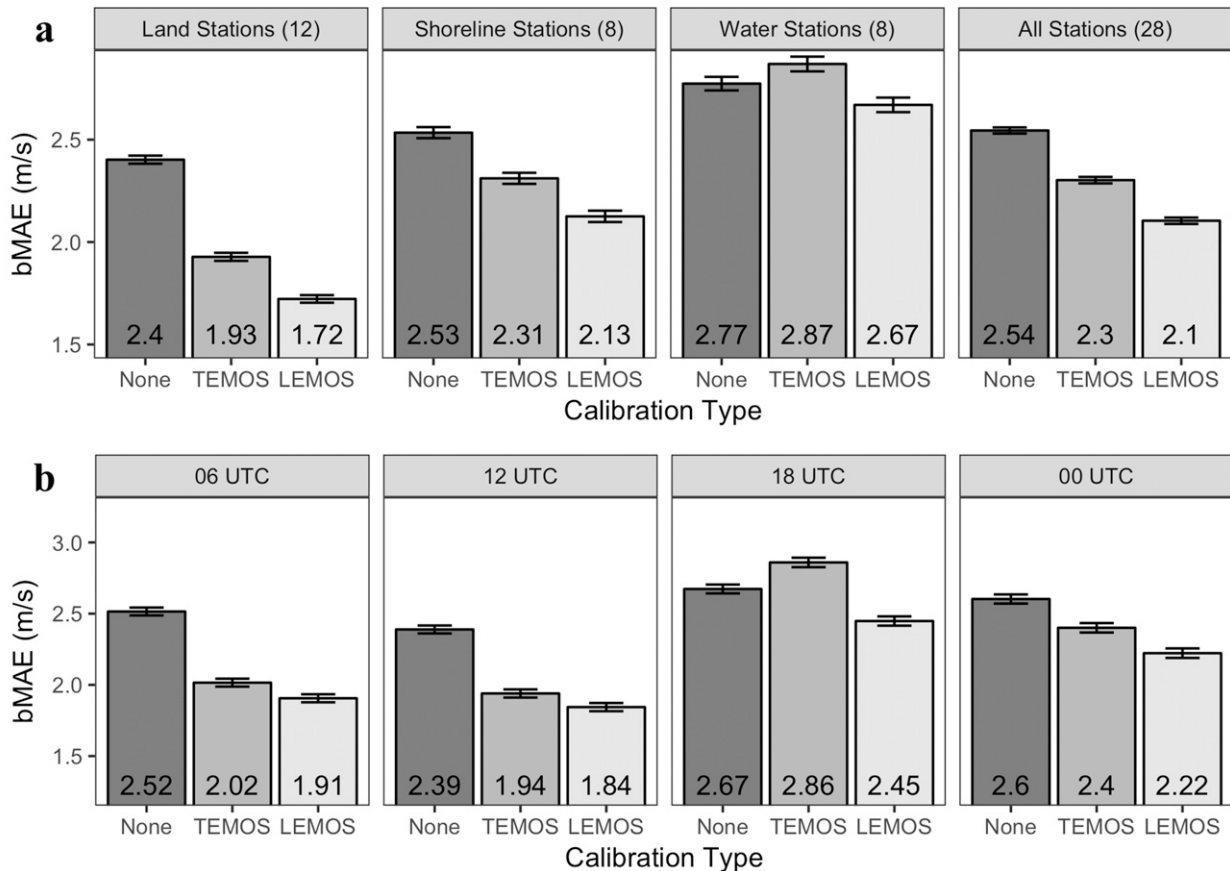
FIG. 6. (a) Bivariate mean absolute error (bMAE; m s^{-1}) from GEFS 24-h ensemble mean forecasts (dark gray), LEMOS (light gray), and TEMOS (gray) averaged by station type (land, shoreline, water) and for all stations during calendar year 2015. (b) The bMAE for all stations but separated by forecast cycle. The columns are arranged to correspond with the local diurnal cycle. (from left to right) middle of the night (0600 UTC), morning (1200 UTC), early afternoon (1800 UTC), and early evening (0000 UTC). Error bars represent 95% bootstrapped confidence intervals for the respective categorical means.

(2.40 m s^{-1}), and highest at water stations (2.77 m s^{-1}). LEMOS reduces the bMAE by 17.3%, an amount that is comparable to that of Schuhen et al. (2012). LEMOS appears to be more effective for land stations (28.3% bMAE reduction) compared to water stations (3.6% bMAE reduction)—a finding consistent with Junk et al. (2014). This suggests that, in general, water (land) stations exhibit less (greater) wind vector bias in the GEFS forecasts. LEMOS reduces the bMAE at all 28 stations, 17 of which are statistically significant (i.e., 95% confidence) including 3 shoreline and all 8 water stations.

The raw GEFS and LEMOS provide a direct means by which to gauge TEMOS performance as the former (latter) yields a lower (upper) limit on forecast skill improvement. Ultimately, postprocessing is useful only if it improves forecast skill over the raw model while the best-case scenario is that obtained directly from LEMOS. Overall, TEMOS reduces the bMAE by 9.4%, roughly half that of LEMOS. Analogous to LEMOS,

TEMOS is most effective at reducing bMAE at land stations (20%), but increases the bMAE at water stations by 3.6%. When results are stratified by forecast cycle (Fig. 6b), the TEMOS deterministic skill is close to that of LEMOS during the middle of the night (0600 UTC), morning (1200 UTC), and, to a lesser extent, during early evening (0000 UTC). During the afternoon (1800 UTC) the TEMOS bMAE is 7.1% higher than the raw GEFS. With the exception of the 1800 UTC cycle (discussed shortly), TEMOS appears to be a reasonable approach for spreading LEMOS mean forecasts.

In general TEMOS performs best at locations well represented by the WRF (i.e., where modeled upwind surface elements and neutral stability represent the actual conditions). For example, XDAI is well characterized by the WRF, that is, the observed and modeled magnitudes are similar (note its location near the one-to-one line in the lower left of all four panels of Fig. 7), and TEMOS is quite close to LEMOS across all forecast
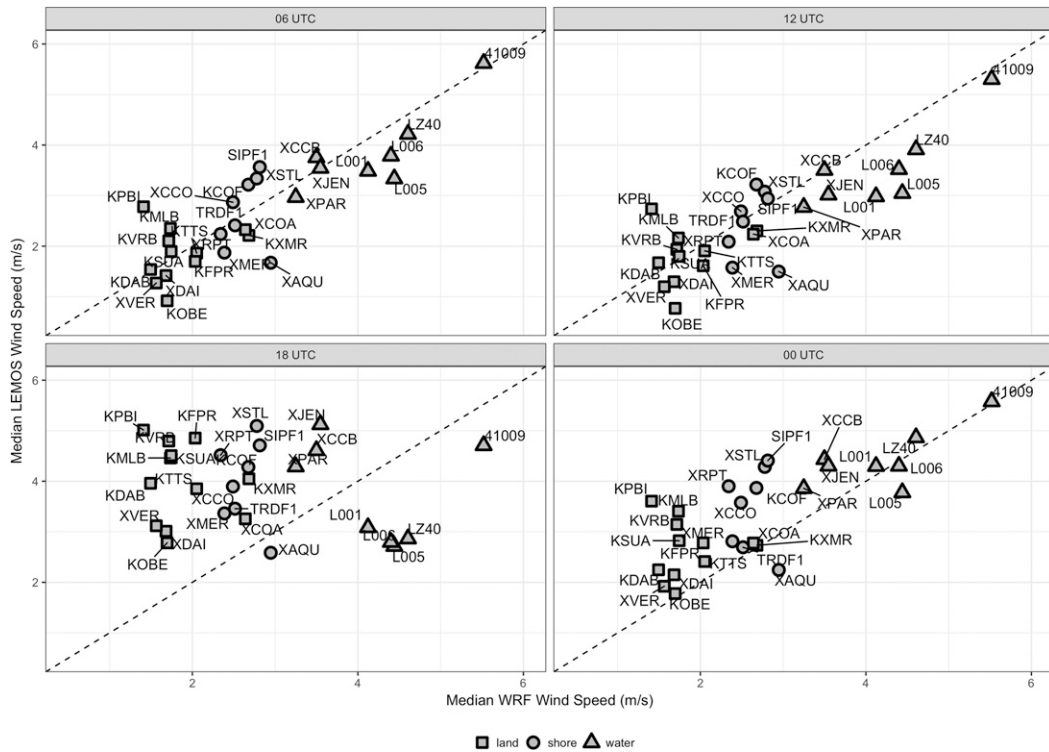
FIG. 7. As in Fig. 4 but for median LEMOS wind speed (*y* axis) vs median WRF wind speed (*x* axis), separated by forecast cycle and segregated by station type. Squares, circles, and triangles denote land-, shoreline-, and water-based stations, respectively. Dashed lines represent a one-to-one fit.

cycles (Fig. 8), reducing the bMAE (cf. the raw GEFS) by 45.5% on average. This underscores the capabilities of TEMOS at stations that are consistent with their WRF grid characterization. In contrast, the WRF KPBI (West Palm Beach, Florida) grid cell is classified as urban (surface roughness of 0.8 m), while KPBI is actually a large, relatively smooth, airfield (surface roughness on the order of 0.01 m). As a result, TEMOS underforecasts KPBI wind speeds (not shown), leading to increased error (cf. the raw GEFS) for all cycles (Fig. 8). Conversely, the poor TEMOS performance at XAQU (shoreline station near Melbourne Beach, Florida; near the center of Fig. 1b) is an artifact of overforecasting the wind speed. Despite its ostensibly fetch favorable shoreline location, XAQU experiences local obstructions (e.g., groves of trees, buildings) uncaptured by the WRF.

2) 1800 UTC PERFORMANCE

To explain the poor TEMOS performance at 1800 UTC, the agreement between observed and modeled winds is assessed by comparing median LEMOS wind speeds (over the study period), which are unbiased estimates of future observations, to median WRF wind speeds (from all 180 simulations) at each station, separated by forecast cycle (Fig. 7). At 0600 and 1200 UTC the stations are relatively evenly distributed about the one-to-one

lines (dashed) and, to a lesser extent, at 0000 UTC. However, at 1800 UTC median LEMOS wind speeds are essentially independent of, and are greater than (with the exception of some of the water stations), those from the WRF. This is an artifact of various physical processes, present in the observations, which were explicitly not accounted for in the WRF simulations—including radiation, microphysics, and variations in static stability. As a result, the increased mixing associated with an unstable afternoon surface layer, sea and lake breezes, and thunderstorm outflows are not present in the simulations. The absence of such features is particularly problematic in the mid- to late afternoon, and it limits the ability of TEMOS to effectively spread information between locations, particularly those that are dissimilar (e.g., from land to water and vice versa). Thus, the TEMOS mean forecasts exhibit larger error at 1800 UTC because the WRF relationships on which TEMOS relies are not consistent with the observations.

Consider KOBE, which is approximately 10 km north of Lake Okeechobee (Fig. 1). Similar to XDAI, where TEMOS performs well (Fig. 8), KOBE is approximately consistent with its WRF characterization (Fig. 7). However, because of the proximity-dominated weighting of the individual LEMOS estimates, KOBE's TEMOS
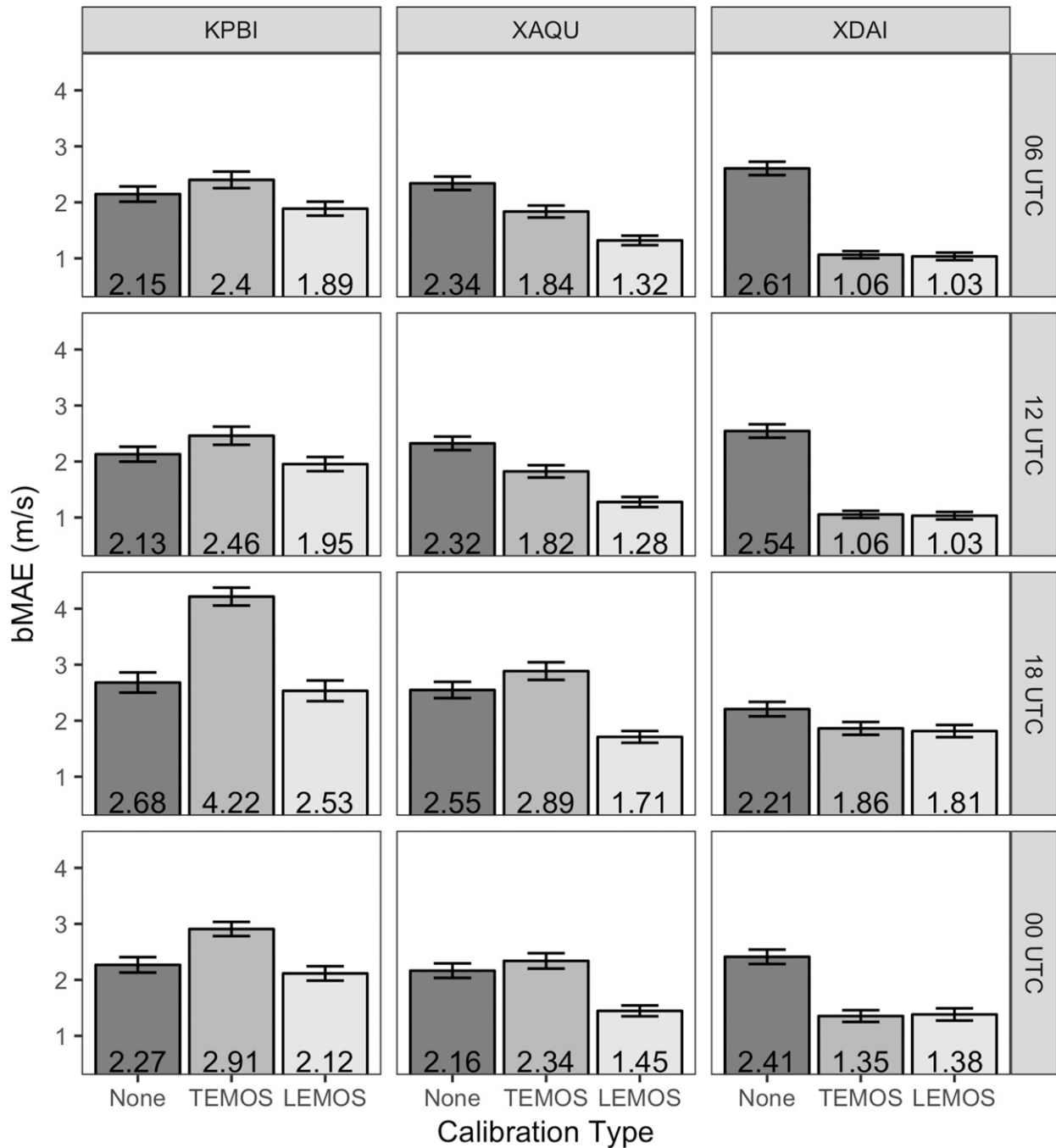
FIG. 8. As in Fig. 6b, but for stations KPBI, XAQU, and XDAI.

forecasts are heavily influenced by the four nearby Lake Okeechobee stations (L001, L005, L006, and LZ40). The WRF scale factor between these locations is approximately 2-to-1 from the lake stations to KOBE. However, lake surface divergence in tandem with the increased mixing associated with an unstable surface layer at KOBE results in comparable 1800 UTC observed median wind speeds at these locations (Fig. 7). As a result, TEMOS spreads winds from the lake to KOBE that are approximately half the magnitude of those observed at KOBE, and thus there is no improvement in the bMAE compared to the raw GEFS (Fig. 9). Similar problems are manifest at other locations, such as the isolated oceanic station 41009, where
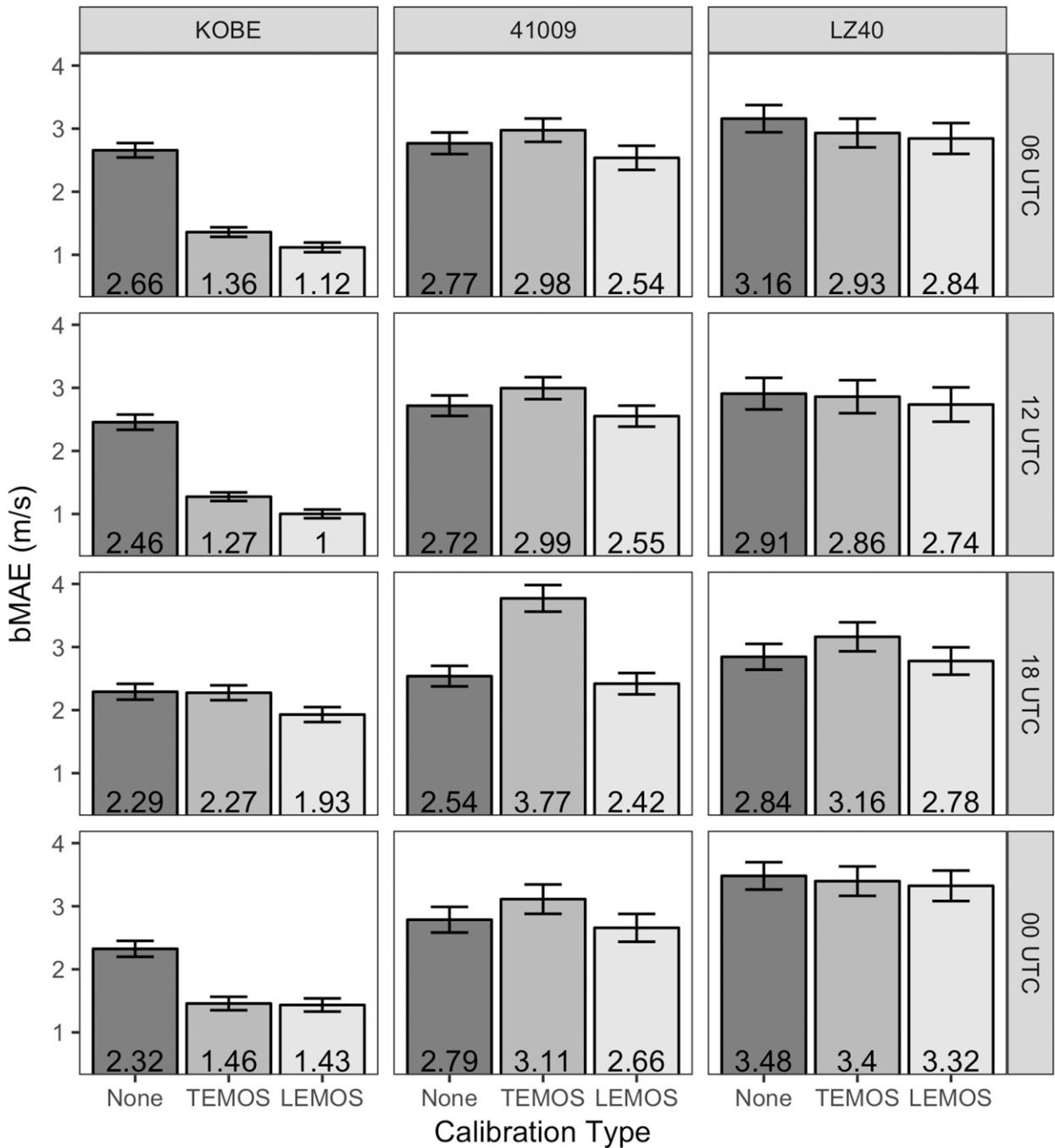
FIG. 9. As in Fig. 8, but for stations KOBE, 41009, and LZ40.

elevated wind speeds (related to the sea breeze and increased instability) at nearby land and shoreline stations lead to overforecasted afternoon TEMOS winds, degrading the deterministic skill by 48% (Fig. 9).

Despite the absence of a model lake breeze, the effects of using WRF gridcell relationships that are inconsistent with observed afternoon wind fields are less

of a problem at Lake Okeechobee stations. In this case, the observed 1800 UTC wind speeds associated with the lake breeze are reduced at all four lake stations (Fig. 7), and the distance weighting yields TEMOS parameter estimates that are primarily determined from the remaining (similar) three lake stations. Hence, in contrast to 41009, the availability of similar nearby wind
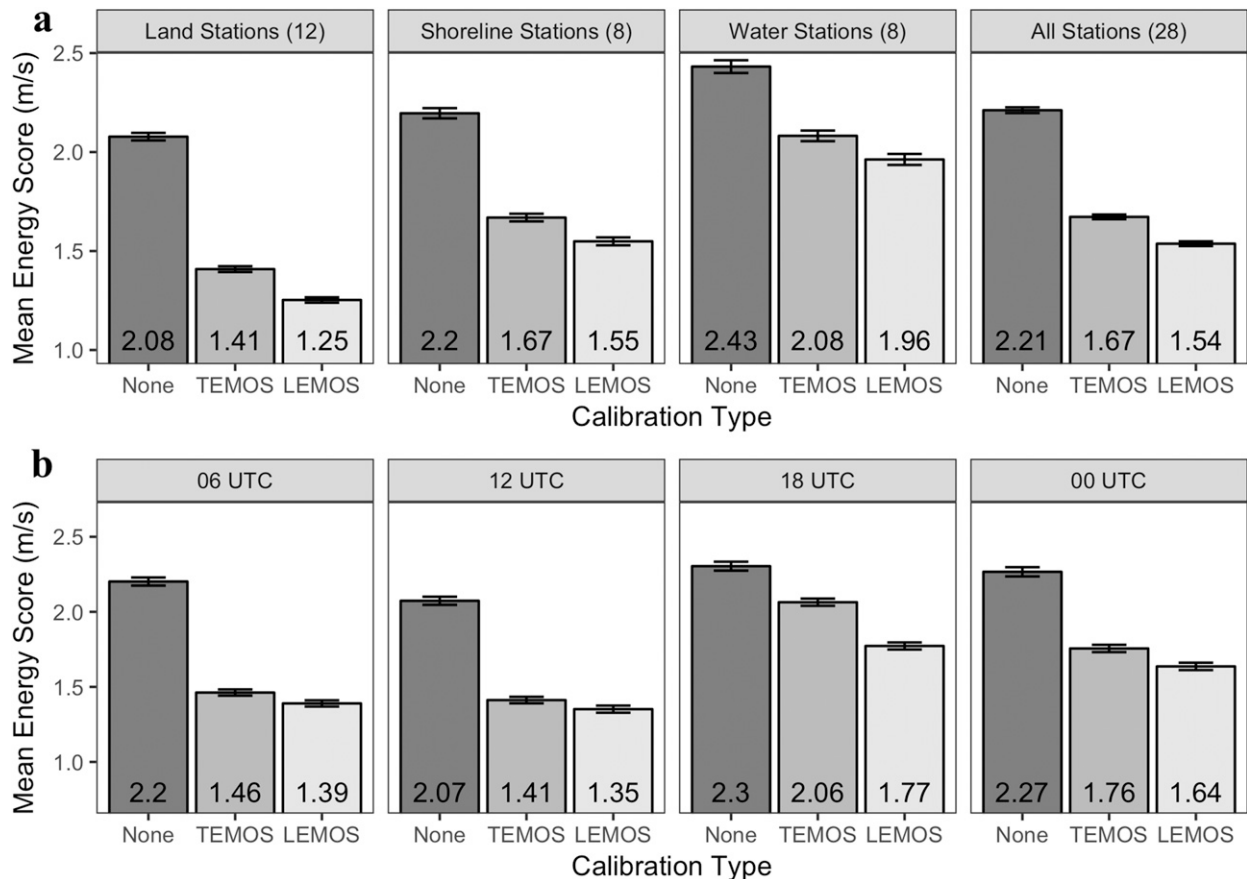
FIG. 10. As in Fig. 6, but for mean energy scores (m s^{-1}).

information results in overlapping TEMOS and LEMOS confidence intervals across all cycles at LZ40, as desired (Fig. 9). This indicates that TEMOS skill can be improved with increased observational density—on the condition of proximity and similarity in both the observational and WRF frameworks.

b. Probabilistic verification

The negatively oriented (i.e., smaller is better) energy score metric [Eq. (7)] simultaneously quantifies calibration and sharpness. It is improved by reducing the average deterministic error of the ensemble, increasing the average Euclidean distance between each member (i.e., increasing the ensemble spread toward the observation), and/or enhancing the representation of the actual correlation structure. The latter does not apply here since we set $\rho_{uv} = 0$ in the EMOS model (section 3), but its relation to the deterministic results (bMAE) is important. The raw GEFS produces larger mean energy scores (ES) at water stations (2.43 m s^{-1}), where deterministic error is the largest of the three station types (Figs. 6a and 10a). The raw ensemble clearly possesses

dispersion errors, as LEMOS and TEMOS improve the ES by 30.3% and 24.4%, respectively. Comparing these values with their deterministic counterparts (17.3% and 9.4% improvement in the bMAE for LEMOS and TEMOS, respectively) suggests that TEMOS is more effective at gridded calibration of probabilistic forecasts than deterministic bias correction alone. Although TEMOS increases the deterministic error at 1800 UTC, the ES is improved over the raw ensemble by 10.4% (Fig. 10b), indicating that the probabilistic benefits outweigh the diminished deterministic skill.

Multivariate rank histograms are examined to further assess the degree to which TEMOS improves calibration (Fig. 11). The U-shaped rank histograms for the raw GEFS indicate that it is underdispersive, regardless of station type. Both TEMOS and LEMOS correct for this, as evidenced by their nearly uniform rank histograms. The reliability index (Δ; Delle Monache et al. 2006), which quantifies a distribution's deviation from uniformity, is also shown. The Δs for TEMOS and LEMOS histograms are comparable ($\Delta = 0$ indicates a perfectly uniform distribution)—indicating that the former appears
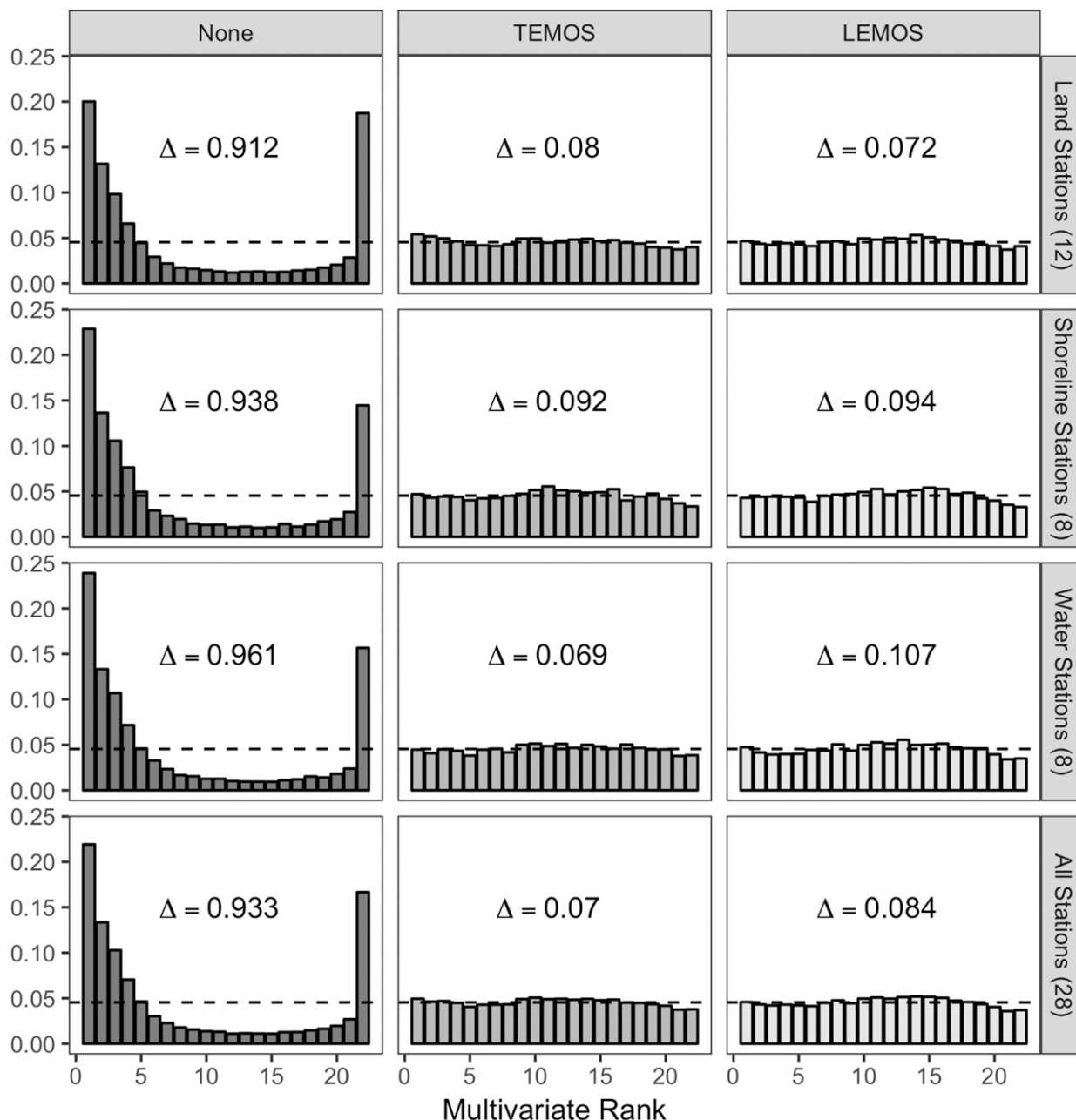
FIG. 11. Multivariate rank histograms and reliability indices (Δ; Delle Monache et al. 2006) of (left) raw GEFS, (middle) TEMOS, and (right) LEMOS 24-h wind vector forecasts for (bottom row) all stations and station type. Dashed lines indicate uniformity.

to be as effective as local calibration at correcting for dispersion errors.

The rank histograms provide insight into the reliability of the calibrated forecast probabilities. As explained previously, the ES is improved by increasing deterministic skill (e.g., bias correction only). The bias correction portion of LEMOS [Eq. (2)] accounts for more than half (55%) of the ES reduction when compared to the raw GEFS (not shown). However, the rank histograms retain their U shape and exhibit only marginal improvement (i.e., Δ decreases from 0.933 to 0.924, not shown). The extent that bias correction alone fails to improve probabilistic reliability is an artifact of the underdispersive nature of the GEFS ensemble (as is the case for many current ensemble systems). This underscores the added value in dispersion correction (i.e., scaling the ensemble spread) during calibration, which yields rank histograms that are essentially uniform for both the TEMOS and LEMOS approaches.
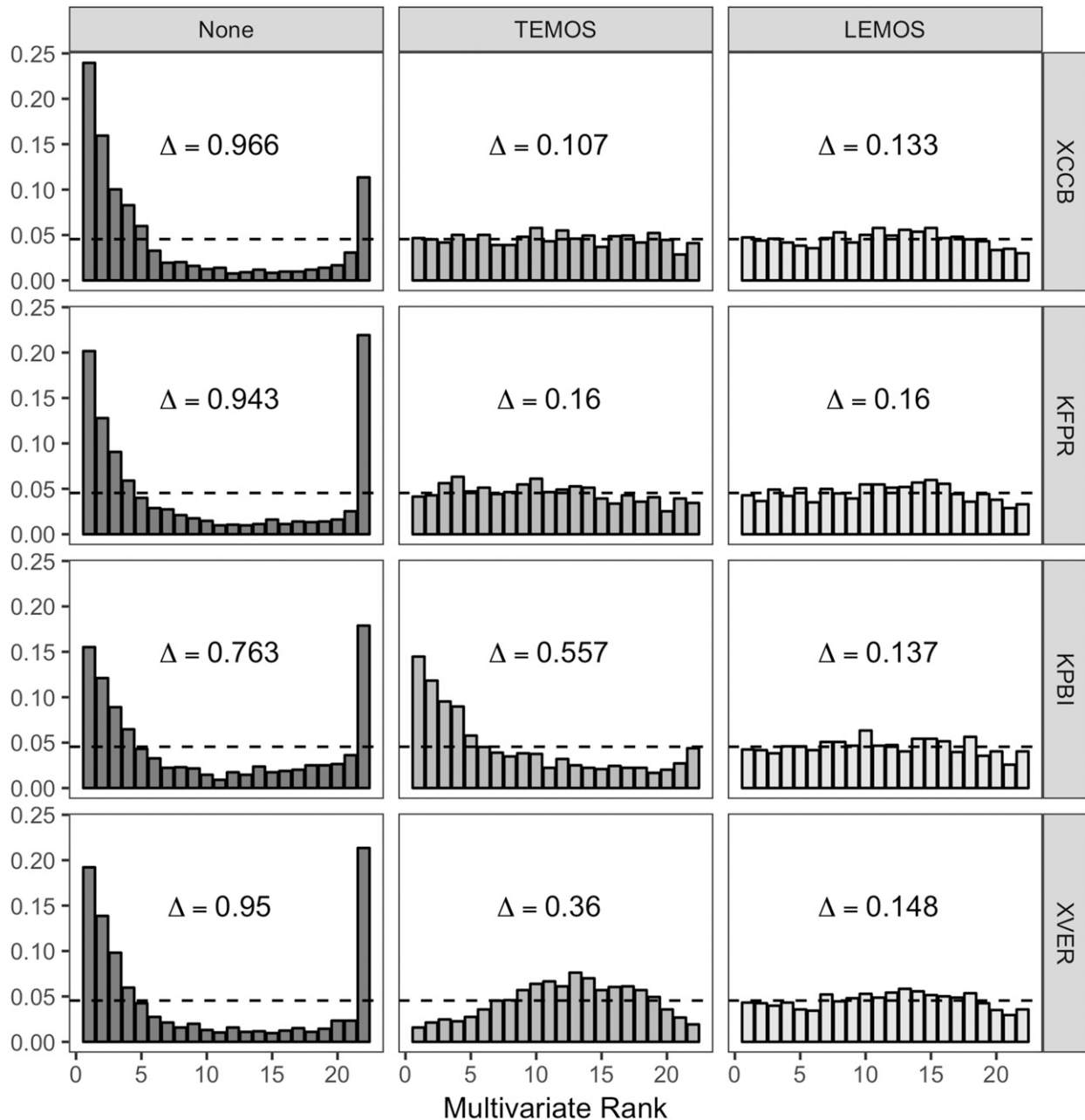
FIG. 12. As in Fig. 11, but at stations XCCB, KFPR, KPBI, and XVER.

At the station level, dispersion errors, and thus calibration, are impacted by representativeness issues. For example, with the exception of 1800 UTC, both XCCB and KFPR appear to be consistent with their respective WRF grid cells as both lie close to best fit lines (Fig. 7). Consequently, the TEMOS multivariate rank histograms at these stations are essentially uniform (Fig. 12, top two rows). However, as discussed previously, the wind speeds at KPBI are faster than resolved by the WRF simulations. As a result, the TEMOS forecasts are

underdispersive and biased (i.e., left-skewed histogram). At the urban station XVER (Vero Beach, Florida), where the variance parameters are overestimated (not shown), the bell-shaped rank histogram indicates overdispersion.

c. Case study, easterly flow, 0000 UTC 5 May 2015

To illustrate the TEMOS method, it is applied to a case study and compared with raw GEFS ensemble mean 24-h wind vector forecasts. This is accomplished by applying LEMOS using GEFS forecasts and observed
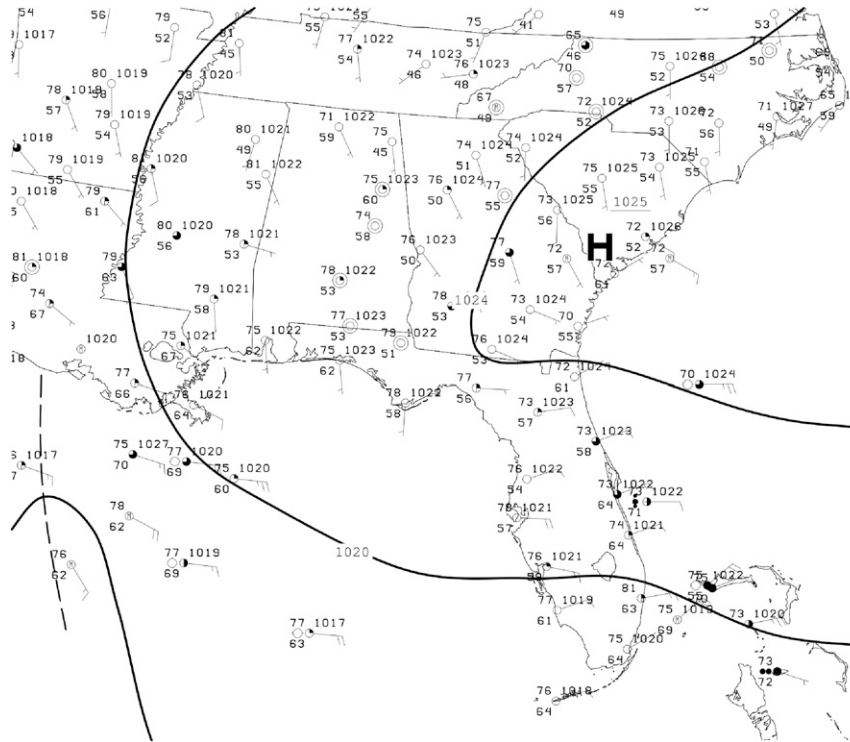
FIG. 13. Surface analysis valid at 0000 UTC 5 May 2015.

winds at the 28 stations to produce estimates of the bivariate predictive distribution parameters (mean and variance, section 3). These parameters are then spread onto a downscaled 333-m grid (more than 500 000 grid cells), using directionally dependent statistical relationships determined from the idealized WRF simulations.

A typical synoptic-scale easterly surface flow case is presented here (Fig. 13). The GEFS ensemble mean 24-h forecast wind speeds gradually diminish from near $9\,m\,s^{-1}$ over the ocean to $4\,m\,s^{-1}$ over the Florida Peninsula (Fig. 14a). While the coastline and lakes (including Lake Okeechobee and the IRL), are not resolved in the raw GEFS, these features are clearly visible in the TEMOS product (Fig. 14b). Furthermore, the spurious oceanic gradient in wind speed in the raw GEFS (in part an artifact of the interpolation of the coarse data onto the refined grid) is no longer present in the TEMOS (calibrated) forecast. In contrast, TEMOS exhibits marked changes in magnitude as the flow transitions across the land/water mask. Only subtle changes in wind direction exist between the two products. Increased cross-isobaric flow is evident in association with the rougher surface elements (on the order of 1 m; Fig. 14d) in the Orlando metropolitan area (dashed rectangle in Fig. 14b), and the winds back to a more easterly direction on the downwind (western) side of

Lake Okeechobee (roughness on order of $10^{-4}\,m$). The relatively small impact of TEMOS on the calibrated wind direction is, in general, consistent with previous studies that indicate little in the way of wind direction bias in current operational NWP forecast systems (e.g., Bao et al. 2010).

The wind speed impact of the calibration is examined by differencing the two products (TEMOS minus the raw GEFS; Fig. 14c). As expected, there is considerable correlation between the wind speed differences and WRF surface roughness. The largest impacts (on the order of $3\,m\,s^{-1}$) tend to occur in association with regions populated by rougher surface elements (urban, forest; Fig. 14d) for which the TEMOS winds are reduced, but are also manifest (as increased wind speed) over portions of Lake Okeechobee, the coastal ocean, and some estuaries. The TEMOS sensitivity to flow-dependent changes in the surface roughness is also apparent as winds increase from east to west over inland water bodies. In general, TEMOS results in an increase in wind speeds over the IRL, except for the downwind regions west of Cape Canaveral and Merritt Island (Fig. 14c, inset) where the resolved upwind land fetch decreases the wind speed (by $1–2\,m\,s^{-1}$) relative to the GEFS.

Because TEMOS parameter estimates are primarily distance weighted, potentially undesirable results can
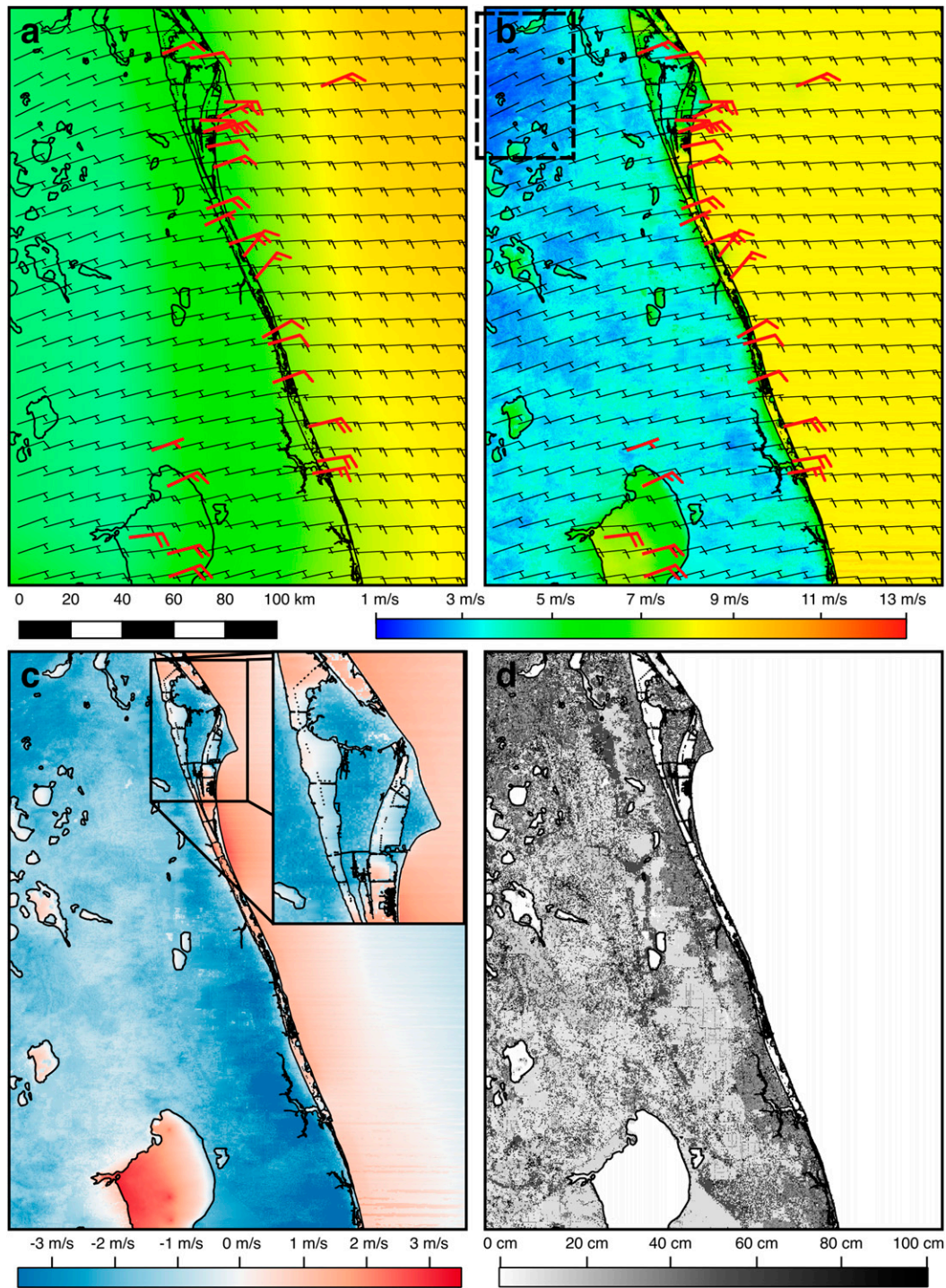
FIG. 14. Gridded (333 m) (a) GEFS ensemble mean and (b) mean TEMOS 24-h wind speed forecasts (m s^{-1}) and associated wind barbs (black) valid at 0000 UTC 5 May 2015. Verifying wind observations are colored red. Dashed rectangle indicates the location of the Orlando metropolitan area, mentioned in the text. (c) Wind speed difference (TEMOS minus GEFS ensemble mean, m s^{-1}), with red (blue) shading indicating that TEMOS wind speeds are faster (slower) than the GEFS ensemble mean. Inset provides a zoomed-in view of the northern IRL, Cape Canaveral, and Merritt Island. (d) WRF Model surface roughness lengths (cm).

occur in the wind field near observing stations. This is evident with respect to two stations in the southern portion of Lake Okeechobee, identifiable as dark red bull's-eyes since TEMOS wind speeds at these stations exceed those of the surrounding water (Fig. 14c). Similarly, an enhanced region of increased wind speeds associated with nearby (low roughness) airports bleeds into the coastal ocean east of KDAB (top of Fig. 14a). Additionally, despite averaging the flow over the last hour of the WRF simulations, some still contain bands of increased wind speed over the oceanic portions of the model domain (e.g., SE corner of Figs. 14b and 14c). These features appear to be a result of the open boundary conditions in the WRF Model. While time averaging mitigates the impact of these features, their persistence suggests that additional smoothing may be needed to extract the mean wind field only. Regardless, their existence does not significantly impact TEMOS results.

While the data-withdrawal experiments indicate that TEMOS, in general, reduces the 0000 UTC deterministic error (Fig. 6b), that is not the case here. TEMOS improves (degrades) deterministic skill at 11 (17) of the 28 stations resulting in an overall small ($0.05 \, \mathrm{m \, s^{-1}}$) increase in the average bMAE compared to the raw GEFS. Nonetheless, TEMOS reduces the ES by 20.5% (from 2.34 to $1.86 \, \mathrm{m \, s^{-1}}$) as the number of observations falling outside the ensemble spread decreases from 15 to 5—highlighting its probabilistic capabilities. Moreover, the calibrated predictive densities at each cell can be sampled to produce fields of wind speed representing the tails of the distribution. For example, the lower and upper 10% likely wind speed thresholds can be estimated by drawing 1000 samples from each grid cell's bivariate density, sorting, and then extracting the 100th and 900th values (Fig. 15). Water bodies are easily identifiable in both of the wind fields. As a result of the positive correlation between forecast wind speed and uncertainty, the range of likely wind speeds is larger for water-based grid cells (6–$13 \, \mathrm{m \, s^{-1}}$) than those over land ($1$–$5 \, \mathrm{m \, s^{-1}}$).

7. Summary and discussion

A combined dynamical/statistical approach is introduced that transforms wind forecasts from a coarse, biased, underdispersive ensemble into high-resolution, gridded, calibrated wind fields. This method, referred to here as interpolated ensemble model output statistics (TEMOS), utilizes flow-dependent statistical relationships extracted from idealized high-resolution WRF simulations to spread locally calibrated winds from locations with observations to those without. Although there is considerable overhead with respect to generating the suite of WRF simulations, the TEMOS implementation is relatively straightforward, adaptable to any domain and resolution with reasonable computational expense.

TEMOS was applied to calibrate and downscale one year of 24-h wind forecasts from the 1° GEFS ensemble over east-central Florida. The domain features, which include a coastal estuary with a complex land/water mask, are an ideal test bed for examining the proposed methodology. The raw GEFS forecasts are biased, especially at land stations, and underdispersive. Using 28 Florida stations and data withdrawal experiments, TEMOS improved the deterministic (bMAE) and probabilistic (ES) forecast skill by 9.4% and 24.4%, respectively. Compared to LEMOS, which provides an upper limit of the method's utility (conditional on WRF representativeness), TEMOS corrects for 54.3% and 80.5% of bias and dispersion errors in the GEFS 24-h forecasts, respectively. Additionally, uniform multivariate rank histograms indicate that TEMOS forecasts are calibrated. The improved probabilistic forecast skill is a key benefit of the TEMOS approach.

TEMOS is applied to an easterly flow event. In contrast to the raw GEFS, the downscaled (and calibrated) wind fields resolve the water bodies in the region, producing winds that better reflect the region's heterogeneous surface characteristics. While TEMOS slightly increases the deterministic error, its enhancement of the variance increases the number of observations encompassed by the ensemble spread by 77%.

Situations where the meteorological conditions are more dynamic can be problematic. In a separate analysis of a frontal passage event, TEMOS fails to adequately resolve the wind shift (not shown). These as well as high-impact events, in general, require more tailored solutions such as implementing a radius of influence (e.g., Yussouf and Stensrud 2006) when spreading parameter estimates and/or limiting training data to analog cases (e.g., Delle Monache et al. 2011). The former places additional emphasis on station proximity and thus requires increased observational density. The analog approach distinguishes between pre- and postfrontal errors in the training data, theoretically improving local mean forecasts and the subsequent dispersal of that information through TEMOS if a sufficient number of representative training cases are available.

The TEMOS method essentially relies on two basic assumptions. The first is that the WRF spatial relationships represent those of the actual atmosphere. The other assumes that each observing station is representative of its corresponding WRF grid cell. TEMOS performs well only when these assumptions are generally valid. Stations poorly characterized by the WRF, such as
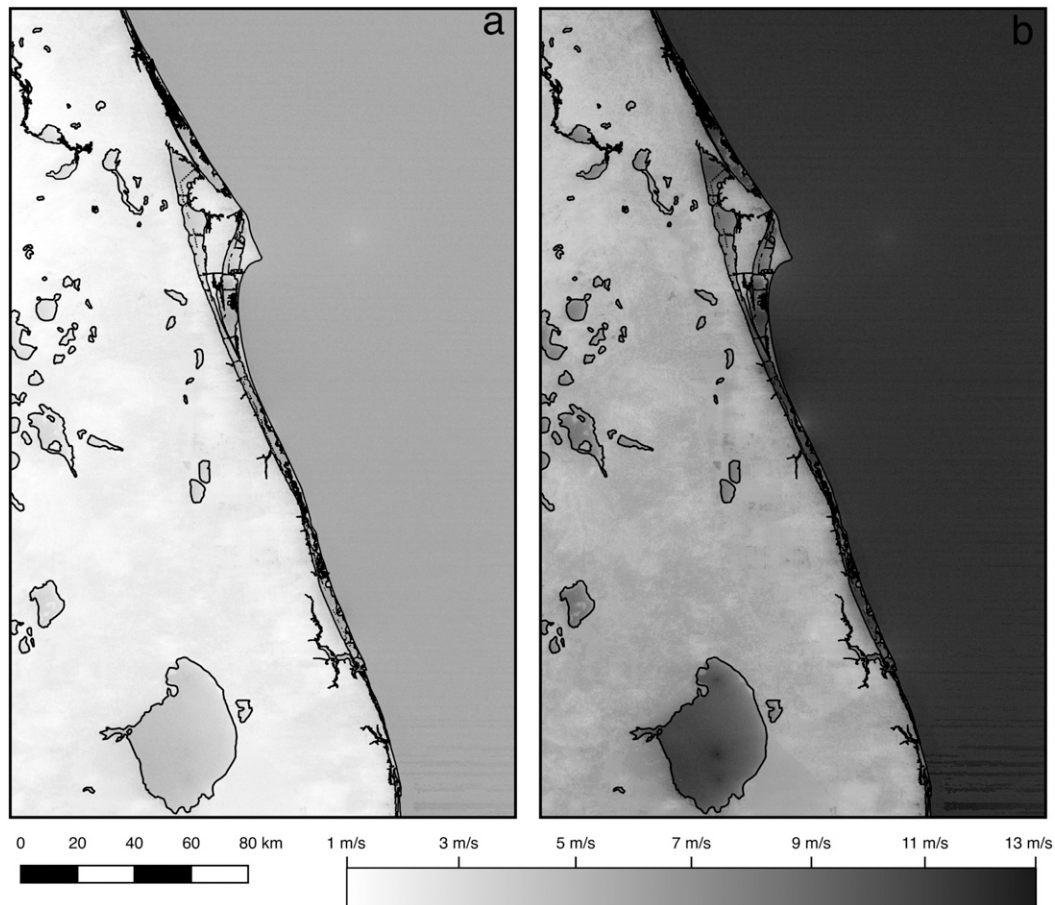
Fig. 15. Gridded estimates of (a) lower and (b) upper 10% likely TEMOS 24-h wind speed forecasts (m s^{-1}) valid at 0000 UTC 5 May 2015. See text for details.

XAQU (Melbourne Beach, Florida) and KPBI (West Palm Beach, Florida) are easily identified after TEMOS application and thus can be excluded when implementing the approach operationally.

The first assumption appears to be an issue during typical Florida afternoons (1800 UTC), where the median LEMOS wind speeds are essentially independent of those from the WRF. As a result, the afternoon TEMOS forecasts degrade the deterministic skill by 7.1%. Strong afternoon winds observed at many land stations are likely a result from increased turbulent mixing when the surface layer is most unstable. To isolate the response of the 10-m wind to the heterogeneous land surface, mesoscale features such as sea breezes, lake breezes, and thunderstorm outflows are not represented in the idealized WRF simulations. The assumption that, to a first order, the flow is strongly driven by bulk surface inhomogeneities (i.e., surface roughness) appears to be reasonable except at 1800 UTC. These mesoscale features could be incorporated, for example,

by performing additional WRF simulations initialized with a seasonally averaged afternoon air–sea temperature difference, resolving sea- and lake-breeze circulations under various steering flow regimes. Relationships could then be extracted from these data and applied to the forecasts valid at 1800 UTC following the approach used here. Regardless, the mean energy scores are improved for all forecast cycles, and the nearly uniform rank histograms indicate that variance parameters are effectively spread irrespective of the time of day and thus the approach effectively spreads calibrated information.

Given the initial success of the method used here, it might be possible to improve the work by incorporating other techniques. For example, Scheuerer and Möller (2015) introduce a kriging model that leverages annual mean wind speed information using a 200 m^2 gridded wind speed climatology. Devoid of such a high-resolution wind product over this study's domain of interest, a kriging approach could instead utilize wind speed information from the high-resolution WRF simulations. Additionally,

Lerch and Baran (2017) introduce a semilocal, clustering-based approach of estimating EMOS parameters for probabilistic wind speed forecasts. Building off of their work, the method introduced here could be extended by identifying similar grid cells in the WRF output, estimating EMOS parameters with pooled training data from stations within each cluster, then spreading these parameters between clusters (as opposed to individual cells).

As presented here we optimize the station weights to spread the predictive PDF parameters by minimizing the bMAE [section 4b(3)]. As an alternative metric, the ES could also be leveraged for this purpose. As the ES accounts for both bias and dispersion, it is likely to offer some additional improvement over that of the bMAE alone. Future implementations of this approach should make this consideration.

REFERENCES

Akylas, E., V. Kotroni, and K. Lagouvardos, 2007: Sensitivity of high-resolution operational weather forecasts to the choice of the planetary boundary layer scheme. *Atmos. Res.*, **84**, 49–57, https://doi.org/10.1016/j.atmosres.2006.06.001.

Bao, L., T. Gneiting, E. P. Grimit, P. Guttorp, and A. E. Raftery, 2010: Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Mon. Wea. Rev.*, **138**, 1811–1821, https://doi.org/10.1175/2009MWR3138.1.

Beljaars, A., 1995: The parametrization of surface fluxes in large-scale models under free convection. *Quart. J. Roy. Meteor. Soc.*, **121**, 255–270, https://doi.org/10.1002/qj.49712152203.

Berrocal, V. J., A. E. Raftery, and T. Gneiting, 2007: Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Mon. Wea. Rev.*, **135**, 1386–1402, https://doi.org/10.1175/MWR3341.1.

Buizza, R., P. Houtekamer, G. Pellerin, Z. Toth, Y. Zhu, and M. Wei, 2005: A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. *Mon. Wea. Rev.*, **133**, 1076–1097, https://doi.org/10.1175/MWR2905.1.

Cheng, W. Y. Y., Y. Liu, Y. Liu, Y. Zhang, W. P. Mahoney, and T. T. Warner, 2013: The impact of model physics on numerical wind forecasts. *Renewable Energy*, **55**, 347–356, https://doi.org/10.1016/j.renene.2012.12.041.

Delle Monache, L., J. P. Hacker, Y. Zhou, X. Deng, and R. B. Stull, 2006: Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *J. Geophys. Res.*, **111**, D24307, https://doi.org/10.1029/2005JD006917.

——, 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, https://doi.org/10.1175/2011MWR3653.1.

Feldmann, K., M. Scheuerer, and T. L. Thorarinsdottir, 2015: Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Mon. Wea. Rev.*, **143**, 955–971, https://doi.org/10.1175/MWR-D-14-00210.1.

Gneiting, T., 2014: Calibration of medium-range weather forecasts. Tech. Memo. 719, European Centre for Medium-Range Weather Forecasts, 30 pp., https://www.ecmwf.int/sites/default/files/elibrary/2014/9607-calibration-medium-range-weather-forecasts.pdf.

——, and A. E. Raftery, 2005: Weather forecasting with ensemble methods. *Science*, **310**, 248–249, https://doi.org/10.1126/science.1115255.

——, and ——, 2007: Strictly proper scoring rules, prediction, and estimation. *J. Amer. Stat. Assoc.*, **102**, 359–378, https://doi.org/10.1198/016214506000001437.

——, ——, A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Wea. Rev.*, **133**, 1098–1118, https://doi.org/10.1175/MWR2904.1.

——, L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds. *TEST*, **17**, 211–235, https://doi.org/10.1007/s11749-008-0114-x.

Grimit, E. P., and C. F. Mass, 2007: Measuring the ensemble spread–error relationship with a probabilistic approach: Stochastic ensemble results. *Mon. Wea. Rev.*, **135**, 203–221, https://doi.org/10.1175/MWR3262.1.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. *Mon. Wea. Rev.*, **129**, 550–560, https://doi.org/10.1175/1520-0493(2001)129<0550:IORHFV>2.0.CO;2.

Harper, B. A., J. D. Kepert, and J. D. Ginger, 2010: Guidelines for converting between various wind averaging periods in tropical cyclone conditions. Tech. Doc. WMO/TD-1555, World Meteorological Organization, 64 pp., https://www.wmo.int/pages/prog/www/tcp/documents/WMO_TD_1555_en.pdf.

Holman, B. P., S. M. Lazarus, and M. E. Splitt, 2017: A fetch-based statistical method to bias correct and downscale wind speed over unresolved water bodies. *Wea. Forecasting*, **32**, 1637–1657, https://doi.org/10.1175/WAF-D-17-0016.1.

Homer, C. G., and Coauthors, 2015: Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.*, **81**, 345–354.

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, https://doi.org/10.1175/MWR3199.1.

Hsu, S. A., E. A. Meindl, and D. B. Gilhousen, 1994: Determining the power-law wind-profile exponent under near-neutral stability conditions at sea. *J. Appl. Meteor.*, **33**, 757–765, https://doi.org/10.1175/1520-0450(1994)033<0757:DTPLWP>2.0.CO;2.

Janjić, Z. I., 1994: The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. *Mon. Wea. Rev.*, **122**, 927–945, https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2.

Junk, C., L. von Bremen, M. Kühn, S. Späth, and D. Heinemann, 2014: Comparison of postprocessing methods for the calibration of 100-m wind ensemble forecasts at off-and onshore sites. *J. Appl. Meteor. Climatol.*, **53**, 950–969, https://doi.org/10.1175/JAMC-D-13-0162.1.

Kann, A., C. Wittmann, Y. Wang, and X. Ma, 2009: Calibrating 2-m temperature of limited-area ensemble forecasts using high-resolution analysis. *Mon. Wea. Rev.*, **137**, 3373–3387, https://doi.org/10.1175/2009MWR2793.1.

Kleiber, W., A. E. Raftery, J. Baars, T. Gneiting, C. F. Mass, and E. Grimit, 2011: Locally calibrated probabilistic temperature forecasting using geostatistical model averaging and local Bayesian model averaging. *Mon. Wea. Rev.*, **139**, 2630–2649, https://doi.org/10.1175/2010MWR3511.1.

Lerch, S., and S. Baran, 2017: Similarity-based semilocal estimation of post-processing models. *J. Roy. Stat. Soc.*, **66**, 29–51, https://doi.org/10.1111/rssc.12153.

Mass, C. F., 2003: IFPS and the future of the national weather service. *Wea. Forecasting*, **18**, 75–79, https://doi.org/10.1175/1520-0434(2003)018<0075:IATFOT>2.0.CO;2.

——, J. Baars, G. Wedam, E. Grimit, and R. Steed, 2008: Removal of systematic model bias on a model grid. *Wea. Forecasting*, **23**, 438–459, https://doi.org/10.1175/2007WAF2006117.1.

——, and Coauthors, 2009: PROBCAST: A Web-based portal to mesoscale probabilistic forecasts. *Bull. Amer. Meteor. Soc.*, **90**, 1009–1014, https://doi.org/10.1175/2009BAMS2775.1.

Nadolski, V. L., 1998: Automated Surface Observing System (ASOS) user's guide. National Oceanic and Atmospheric Administration, Department of Defense, Federal Aviation Administration, U. S. Navy, 74 pp., https://www.weather.gov/media/asos/aum-toc.pdf.

Pinson, P., 2012: Adaptive calibration of (*u, v*)-wind ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, **138**, 1273–1284, https://doi.org/10.1002/qj.1873.

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, https://doi.org/10.1175/MWR2906.1.

Scheuerer, M., and L. Büermann, 2014: Spatially adaptive post-processing of ensemble forecasts for temperature. *J. Roy. Stat. Soc.*, **63**, 405–422, https://doi.org/10.1111/rssc.12040.

——, and G. König, 2014: Gridded, locally calibrated, probabilistic temperature forecasts based on ensemble model output statistics. *Quart. J. Roy. Meteor. Soc.*, **140**, 2582–2590, https://doi.org/10.1002/qj.2323.

——, and D. Möller, 2015: Probabilistic wind speed forecasting on a grid based on ensemble model output statistics. *Ann. Appl. Stat.*, **9**, 1328–1349, https://doi.org/10.1214/15-AOAS843.

Schuhen, N., T. L. Thorarinsdottir, and T. Gneiting, 2012: Ensemble model output statistics for wind vectors. *Mon. Wea. Rev.*, **140**, 3204–3219, https://doi.org/10.1175/MWR-D-12-00028.1.

Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., https://doi.org/10.5065/D68S4MVH.

Sloughter, J. M., T. Gneiting, and A. E. Raftery, 2010: Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *J. Amer. Stat. Assoc.*, **105**, 25–35, https://doi.org/10.1198/jasa.2009.ap08615.

——, ——, and ——, 2013: Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Wea. Rev.*, **141**, 2107–2119, https://doi.org/10.1175/MWR-D-12-00002.1.

Thorarinsdottir, T. L., and T. Gneiting, 2010: Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression. *J. Roy. Stat. Soc.*, **173**, 371–388, https://doi.org/10.1111/j.1467-985X.2009.00616.x.

Toth, Z., E. Kalnay, S. M. Tracton, R. Wobus, and J. Irwin, 1997: A synoptic evaluation of the NCEP ensemble. *Wea. Forecasting*, **12**, 140–153, https://doi.org/10.1175/1520-0434(1997)012<0140:ASEOTN>2.0.CO;2.

Whitaker, J. S., and A. F. Loughe, 1998: The relationship between ensemble spread and ensemble mean skill. *Mon. Wea. Rev.*, **126**, 3292–3302, https://doi.org/10.1175/1520-0493(1998)126<3292:TRBESA>2.0.CO;2.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences.* Academic Press, 676 pp.

Yussouf, N., and D. J. Stensrud, 2006: Prediction of near-surface variables at independent locations from a bias-corrected ensemble forecasting system. *Mon. Wea. Rev.*, **134**, 3415–3424, https://doi.org/10.1175/MWR3258.1.