Object-Based Analog Forecasts for Surface Wind Speed

MARIA E. B. FREDIANI^a

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

THOMAS M. HOPSON AND JOSHUA P. HACKER^b

National Center for Atmospheric Research, Boulder, Colorado

EMMANOUIL N. ANAGNOSTOU

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

LUCA DELLE MONACHE AND FRANCOIS VANDENBERGHE

National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 18 January 2017, in final form 17 October 2017)

ABSTRACT

Analogs are used as a forecast postprocessing technique, in which a statistical forecast is derived from past prognostic states. This study proposes a method to identify analogs through spatial objects, which are then used to create forecast ensembles. The object-analog technique preserves the field's spatial relationships, reduces spatial dimensionality, and consequently facilitates the use of artificial intelligence algorithms to improve analog selection. Forecast objects are created with a three-step object selection, combining standard image processing algorithms. The resulting objects are used to find similar forecasts in a training set with a similarity measure based on object area intersection and magnitude. Storm-induced power outages in the Northeast United States motivated the method's validation for 10-m AGL wind speed forecasts. The training set comprises reforecasts and reanalyses of events that caused damages to the utility infrastructure. The corresponding reanalyses of the best reforecast analogs are used to produce the object-analog ensemble forecasts. The forecasts are compared with other analog forecast methods. Analogs representing lower and upper predictability limits provide references to distinguish the method's ability (to find good analogs) from the training set's ability (to provide good analogs) to generate skillful ensemble forecasts. The object-analog forecasts are competitively skillful compared to simpler analog techniques with an advantage of lower spatial dimensionality, while generating reliable ensemble forecasts, with reduced systematic and random errors, maintaining correlation, and improving Brier scores.

1. Introduction

Analog forecasts are predictions based on the idea of finding events from the past that are analogous (i.e., similar) to a current prediction. From these past events, we may be able to infer the errors of the current forecast. The methods used to produce analog forecasts vary in

the data source used to select analogous events, in the similarity criteria, and in the ground-truth data used to build the final statistical prediction.

Since the early advances in deterministic chaos, researchers have been investigating nearest neighbors in the atmosphere's phase space to better understand the limits of predictability (Lorenz 1963). Two similar atmospheric states are expected to remain close for a relatively short time interval (Van den Dool 1989), potentially allowing for using one known phase space orbit to predict another. The first successful use of analog forecasts for numerical weather prediction (NWP) appeared when researchers stopped seeking similarities over an entire hemisphere (Namias 1951) and shifted the

^a NCAR Visiting Scientist, Boulder, Colorado.

^b Current affiliation: Jupiter Technology Systems, Boulder, Colorado.

Corresponding author: Maria E. B. Frediani, maria.frediani@uconn.edu

focus to smaller regions (Van den Dool 1989; Vislocky and Young 1989; Kruizinga and Murphy 1983).¹ The positive results were a consequence of fewer spatial degrees of freedom, which increased the chances of finding good analogs, and was justified because short-term tendencies are determined by local or quasi-local processes.

Four decades later, research on using past forecast similarities continues, and the goal has evolved from improving over persistence to extending predictability and uncertainty quantification. Increased computational power has allowed for fast-paced advances in data assimilation and enhanced forecast resolution (spatial, temporal, and spectral). Current efforts in numerical weather and climate prediction focus on best representing localized processes (e.g., clouds and convection), on improving predictability of nonlinear quantities (e.g., variables associated with turbulent transport), and on modeling complex flows (e.g., alpine meteorology, large-eddy simulation). As scientists gain the ability to model very localized phenomena and explicitly resolve processes, the number of degrees of freedom in the system escalates. Consequently, the primary challenge to finding good analogs in high-dimensional systems reemerges.

Analog forecasts are often used to create statistical ensembles and estimate uncertainties of a deterministic prediction. The method has been validated by many applications, such as for NWP of precipitation (Hamill and Whitaker 2006; Hamill et al. 2015; Keller et al. 2017), wind and temperature (Delle Monache et al. 2011, 2013), renewable energy (Martín et al. 2014; Alessandrini et al. 2015; Vanvyve et al. 2015; Zhang et al. 2015), air quality (Djalalova et al. 2015), flood forecasting (Hopson and Webster 2010), orographic precipitation (Panziera et al. 2011), climate downscaling (Pinto et al. 2014), tropical cyclone track prediction (Sievers et al. 2000; Fraedrich et al. 2003), precipitation downscaling based on weather regime analogs (Rostkier-Edelstein et al. 2016), and flows in complex terrain (Manor and Berkovic 2015).

Commonly, analog sets are selected based on model quantities interpolated to an observation location or at a grid point, over a small time window (Delle Monache et al. 2013). When applied to individual points independently, spatial relationships are not retained. Alternatively, selecting analogs that cover a spatial domain can potentially preserve spatial covariances, such as matching a grid tile centered over a point of interest

(Hamill and Whitaker 2006). We propose a new method to find analogs based on spatial objects, that is, features with distinguishable shapes in any NWP 2D field, including forecasts and analysis, obtained from any NWP variable. In this study, analog forecasts are based on similarities between a deterministic forecast and previous forecasts from the same NWP system, and an ensemble is produced with the corresponding analogs' observed states.

An object in a forecast field is a unit enclosed by defined boundaries described by a set of attributes, such that each attribute is a unitary quantity representative of the object's entire spatial extent. General object attributes are speed, propagation direction, orientation, position, area, magnitude, etc. Searching for analogs over object spaces, rather than grids, reduces the spatial degrees of freedom required to find spatially similar fields. By grouping spatially similar information into an object unit, spatial relationships are retained and the local dynamical coherence is preserved. The method's reduced spatial dimensionality consequently facilitates implementation of advanced algorithms to build the analog set (e.g., artificial intelligence), which are currently challenged by the algorithms' computational costs incurred over the high spatial dimension of traditional forecast grids.

This study is motivated by the need of high-resolution probabilistic forecasts over the Northeast United States (NE). In this region, frequent storm-induced power outages are caused by the constant interaction between trees and the adjacent overhead distribution lines, which in turn, are driven by moderate wind speeds and gusts. With foreknowledge of the number and location of such power outages, restoration crews could more optimally be allocated within a service territory, with the potential of significantly diminishing the time between power failure and restoration.

To help utilities prepare and manage restoration resources, a power outage prediction model was developed at the University of Connecticut (Wanik et al. 2015; He et al. 2017). In predicting power outages, this model accounts for a number of sources of variability impacting outage occurrence, such as variations in vegetation cover and power line density across the region. However, one of the largest sources of uncertainty is in the weather itself, predominantly the wind speed's impact on outages, such as when areas of high wind speed are predicted at the incorrect location. The outage prediction model currently operates with regional deterministic NWP forecasts, and the potential to obtain ensemble wind speed predictions (thus accounting for weather variability) to drive the power outage prediction model at no additional computational cost motivated the development of this analog technique.

¹ Studies using analogs began in the 1940s; however, predictions were not numerical.

With wind speed being the most important nonstatic variable in outage prediction model, the object-based analog technique is applied to predict 10-m above ground level (AGL) wind speeds. The wind speed objects selected in this study represent features of similar wind speed structure in the lower levels of the planetary boundary layer, generated by large-scale dynamics combined with local forcings (terrain, landscape, thermals, etc.). Object attributes may be used to distinguish location errors from magnitude errors (and any other errors assigned to an attribute: orientation, size, etc.), which cannot be directly accomplished with a fixed gridpoint analog approach.

Forecast objects are used in forecast verification techniques to emulate an evaluation performed by the human eye (Davis et al. 2006a,b; Brown et al. 2007; Gilleland et al. 2009). The approach to verifying forecasts using objects also aims at reducing the penalty to the overall forecast skill from spatial displacements between model and observation (or among models). Object-based (or feature based) verification approaches belong to a larger group of spatial verification methods (Casati et al. 2008; Ahijevych et al. 2009; Gilleland et al. 2009), which include scale-dependent approaches (Briggs and Levine 1997; Casati et al. 2004; Lack et al. 2010), field deformation approaches (Hoffman et al. 1995; Gilleland et al. 2010), and fuzzy approaches (Ebert 2008).

Some of these object-based verification methods use image smoothing algorithms, and invariably, they all require thresholds to identify objects in the model field. The verification is made between two gridded fields, so objects in the forecast must be compared to objects in the reference field. Often, objects in two fields are compared after matching all corresponding objects, using criteria specific to each technique.² When matching is required, object attributes are used to find matchings between forecast and reference fields. The matching is successful when similar objects are created in both fields.

However, objects created with thresholding may not always be matched, particularly when the magnitude differences between fields, or within the same field, are high. Threshold issues are more apparent in continuous fields, such as wind speed. In fields of continuous variables, nonzero values are present over the entire domain, as opposed to dichotomous variable fields, such as

precipitation, for which objects may be created directly with any threshold above zero. To overcome limitations associated with object selection and matching, we develop a new object selection and matching technique that is the basis for the object-based analog forecasting method (Obj-An) and is also a forecast verification method on its own. The object selection procedure combines three standard image processing algorithms (three-step object selection) and the matching technique may be applied to any pair of fields, regardless of their 2D spatial pattern and magnitude differences.

In the following section, we present the data, model setup, and metrics used to verify the analog forecasts and validate the method by comparing verification metrics to other analog forecasts. In section 3 we describe the three-step object selection algorithm and the procedure to match objects between model fields. Section 4 describes the Obj-An technique, including object selection and object matching. Results are shown in section 5, followed by the conclusions in section 6.

2. Data and models

In this study, the Obj-An method is evaluated using a set of storms comprising 89 events (Table 1), with variable duration and moderate wind speeds that affected the electric power distribution in the NE. The events consist of storms from all seasons, including tropical storms and nor'easters. The method is verified over a subregion of the NE, including the state of Connecticut and part of Massachusetts (Fig. 1).

The training set used in this study is restricted to verified positives and does not include false forecast positives. The training set is based on power outage utility data, and reports consist of outage counts by date. So, records of dates in which events were expected but did not occur are not available. Nevertheless, verified negatives exist within the events forecast duration, which is not limited to times of high wind speed, rather, the forecast period encloses the event, and therefore, include calm conditions preceding and succeeding the event.

In this study, the Obj-An method is evaluated using reforecasts (for the forecast fields required by the method) and reanalyses (representing the forecast verification fields). These NWP models are run over the subregion using an 81×74 gridpoint subdomain, with 3.3 km of grid spacing (Fig. 1). Forecasts are verified over land only, so the forecast effective grid area is reduced from 5994 (81×74) to 4635 grid points. The reforecasts are generated hourly, between 24- and 90-h lead times, for a total of 5773 forecast fields.

Analog forecasts comprising 15 ensemble members are created for every forecast lead hour of each storm. To

² Some methods, such as (structure, amplitude, and location) SAL (Wernli et al. 2008), do not match individual objects, instead, they use a metric to quantify the entire field, where the location component is given by the objects' relative distance to the field's center of mass.

TABLE 1. Storm dates used in this study. The columns contain the initialization date and duration in hours considered for each storm. The reforecast simulations are initialized at 0000 UTC (except on dates indicated below) and the initialization hour is not included in the Obj-An method's validation. The reanalyses used for verification are initialized 12 h prior to the reforecast start time.

28 Sep 2004 66 h	27 Jul 2006 54 h	21 Jun 2008 66 h	7 May 2010 84 h
1 Apr 2005 84 h	31 Jul 2006 54 h	17 Jul 2008 54 h	25 May 2010 78 h
28 Sep 2005 54 h	14 Jul 2007 48 h	8 May 2009 84 h	28 Oct 2011 78 h
24 May 2005 66 h	1 Sep 2006 72 h	22 Jul 2008 78 h	1 Jun 2010 72 h
28 Jun 2005 60 h	27 Oct 2006 78 h	26 Jul 2008 48 h	4 Jun 2010 73 h ^a
18 Jul 2005 54 h	30 Nov 2006 78 h	5 Aug 2008 54 h	5 Jun 2010 78 h
21 Jul 2005 54 h	1 Mar 2007 54 h	5 Sep 2008 72 h	5 Jul 2010 48 h
26 Jul 2005 78 h	14 Apr 2007 84 h	24 Oct 2008 72 h	20 Jul 2010 72 h
4 Aug 2005 54 h	14 May 2007 66 h	10 Dec 2008 72 h	31 May 2011 84 h
11 Aug 2005 78 h	31 May 2007 54 h	29 Dec 2008 54 h	8 Jun 2011 60 h
16 Sep 2005 48 h	8 Jul 2007 54 h	6 Jan 2009 90 h	27 Aug 2011 66 h
7 Oct 2005 72 h	29 Jul 2007 24 h	25 Jun 2009 54 h	22 Jun 2012 36 h
15 Oct 2005 72 h	16 Aug 2007 78 h	6 Jul 2009 60 h	28 Oct 2012 66 h
24 Oct 2005 78 h	2 Dec 2007 78 h	23 Jul 2009 54 h	7 Nov 2012 66 h
2 Jan 2006 72 h	13 Jan 2008 54 h	29 Jul 2009 78 h	28 Dec 2012 71 h
13 Jan 2006 90 h	12 Feb 2008 72 h	30 Jul 2009 54 h	30 Jan 2013 66 h
16 Jan 2006 24 h	8 Mar 2008 60 h	20 Aug 2009 60 h	7 Feb 2013 84 h
16 Feb 2006 84 h	19 Mar 2008 90 h	6 Oct 2009 72 h	10 May 2013 54 h
31 May 2006 54 h	26 May 2008 78 h	27 Nov 2009 72 h	23 May 2013 61 h ^b
6 Jun 2006 72 h	7 Jun 2008 60 h	22 Feb 2010 72 h	24 May 2013 60 h
10 Jul 2006 54 h	9 Jun 2008 60 h	28 Apr 2010 54 h	28 May 2013 60 h
16 Jul 2006 54 h	13 Jun 2008 54 h	3 May 2010 54 h	7 Jun 2013 54 h
			13 Jun 2013 48 h

^a Storm start hour is at 1200 UTC.

^b Storm start hour is at 1800 UTC.

obtain a fixed number of members in the ensembles, a threshold cannot be used to enforce a maximum similarity. Thus, the most similar 15 analogs candidates compose the ensembles.

a. Model: Forecast and verification fields

The reforecast initial and boundary conditions consist of the second-generation NOAA Global Ensemble Forecast System Reforecast control member (GEFS/R; Hamill et al. 2013). It is based on the 2012 version of the NCEP Global Forecast System run at T254L42 (approximately 40 km). The ensemble control member (member 0) is dynamically downscaled using the WRF Model, version 3.4.1 (Skamarock and Klemp 2008), with three nested domains of 29.7-, 9.9-, and 3.3-km grid spacing, 30 arc s (~ 1000 m) of terrain resolution, a 30-s time step, two-way feedback between nests, and 28 vertical levels. Parameterizations include the following: Thompson microphysics (Thompson et al. 2008); implicit cumulus parameterization using Grell 3D convection (Grell and Dévényi 2002) in the outer and intermediate domains, and explicitly integrated in the inner nest; RRTM longwave radiation (Mlawer et al. 1997); Goddard shortwave radiation (Chou and Suarez 1994); Unified Noah LSM (Tewari et al. 2004); Yonsei University PBL (Hong et al. 2006); MM5 similarity for surface layer (Zhang and Anthes 1982); and a

topographic correction for surface wind (Jiménez and Dudhia 2012). We label these deterministic forecasts as “Raw” in later sections.

The verification field consists of high-resolution reanalysis generated using dynamical initialization (Newtonian relaxation or nudging) to assimilate observations

FIG. 1. Inner model domain with 3.3-km grid intervals and study area enclosed in box showing the terrain elevation. The red markers indicate the three locations used in the gridpoint evaluation presented in section 5.

into the NWP model. The dynamical initialization consists of a preforecast integration step during which the time-dependent variables—mass, momentum, water vapor (i.e., u , v , θ , and q)—are gradually nudged toward observations (Auroux and Blum 2008). The four-dimensional data assimilation (Stauffer et al. 1991) system was configured according to the Army Test and Evaluation Command Real-Time Four-Dimensional Data Assimilation (ATEC RTFDDA) system (Liu et al. 2005, 2008) incorporated into the WRF Model.

The WRF Model V3.5.1 is used to generate a reanalysis with a domain identical to the reforecast, using initial and boundary conditions from the Climate Forecast System Reanalysis (CFSR; Saha et al. 2010), at a T382 grid (approximately 38 km). The simulations are initialized 12 h ahead of the start time of each event. The parameterizations are the same for the downscaled reforecast, except for the topographic parameterization (Topo-wind; Jiménez and Dudhia 2012), which was not applied.³ The observations provided to the NWP model comprise quality controlled surface data from the Meteorological Assimilation Data Ingest System (MADIS, <http://madis.noaa.gov>), and 10-m wind over ocean from scatterometers.

b. Obj-An forecast verification

1) ANALOG ENSEMBLE METHODS

The Obj-An method is validated with a leave-one-out cross validation. All forecast lead hours from the held-out event are removed from the 89-event set, such that the training set comprises the remaining 88 events. The forecasts obtained with the Obj-An method are verified and compared with forecasts from other analog methods, which we discuss here. The comparison among multiple analog techniques characterizes the method's performance relative to simpler analog techniques.

The forecast ensembles generated from the Obj-An are compared with the unprocessed deterministic forecast (Raw), and ensembles generated from four other analog ensemble methods. In any analog ensemble, the two common steps consist of 1) finding the best analog forecasts, and 2) taking the corresponding “truth” to build the ensemble. The techniques differ in the similarity measure and constraints used to select the best analogs, and at times, the use of weights to calibrate the ensemble (weights are not used in this study). The four analog ensembles (all generating the same number, 15,

of ensemble members) included with the Obj-An verification results are as follows:

- 1) Grid-An: A single-gridpoint analog ensemble, created with a similarity measure calculated individually at each grid point, given by the wind speed difference between Raw and forecasts in the training set. The Grid-An method does not require physical coherence in the 2D space; that is, the ensemble members of neighboring grid points are independent of each other. Consequently, the sample of high-ranked gridpoint candidates increases because their similarity measure is not constrained by any further spatial or temporal relationship.
- 2) Field-An: A field-based analog ensemble, created from 2D fields using the root-mean-square (RMS) difference between Raw and forecasts in the training set as the similarity measure. Forecasts derived with the Field-An method are physically coherent fields in the 2D space, in that all grid points of each of the ensemble members belong to the same numerical model output.
- 3) Random: An ensemble created with randomly sampled 2D fields, based on a uniform distribution.
- 4) Best-An: A verification-based selection representing the best analog ensemble derivable from the training set. For a given sample, the Best-An ensemble comprises 2D fields with the smallest RMS difference between the “actually occurring” verification field and the verification fields available in the reforecast–verification pair training set. The Best-An ensemble requires advanced knowledge of the forecast verification, so it is not feasible operationally. Yet, in a validation exercise, it reveals the best possible statistical forecast provided by the training set and serves as an assessment of the sufficiency of the training set size and extent. Together, the Best-An and Random ensembles provide the higher and lower boundaries, respectively, of any analog prediction using the same training set.

2) FORECAST VERIFICATION METRICS

The forecast verification includes deterministic and probabilistic metrics. Bias, mean absolute error (MAE), centered root-mean-square error (CRMSE), and correlation are used to compare the verification results of the four analog ensemble means and Raw. The metrics are calculated for forecast–verification pairs at individual land grid points in the forecast domain (see the appendix for details). Results are presented in boxplots, where boxes enclose the inner quartile around the median (McGill et al. 1978). The values enclosed by the boxes represent the spatial distribution of verification

³ The topographic parameterization introduced biases into the nudged simulated wind speeds, possibly overcorrecting tendencies (Frediani et al. 2016) after observation nudging.

metrics aggregated over time. The median is denoted by the line separating lower and upper quartiles in two color shades, and the mean is denoted by a star. Vertical bars (whiskers) enclose the minimum and maximum values. Notches indicating 95% confidence intervals are narrow and not visible in the figures.

The comparison of verification metrics is supported by direct forecast comparisons. Verification metrics show errors between a forecast and its corresponding verification, whereas direct forecast comparisons reveal the significance of performance differences among forecast methods. A direct comparison between each of the four analog ensemble means and the Raw forecast indicates whether the analog forecast metrics are significantly different from the Raw forecast metric. The direct forecast comparison includes MAE and CRMSE, bootstrapped with the percentile interval method using 1000-element samples resampled 1000 times for confidence intervals enclosing the 10th and 90th percentiles.

MAE and CRMSE skill scores are calculated using the random ensemble mean as the climatological reference. The skill score boxplots are presented for metrics aggregated in time and space. When metrics are aggregated in time, the boxplots represent the distribution of grid points over land (i.e., the MAE and CRMSE correspond to averages over all times at single grid points). For metrics aggregated in space, the distributions represented by the boxplots enclose multiple events at varying forecasts hours (i.e., the MAE and CRMSE correspond to spatial errors of the entire forecast area).

Probabilistic metrics include a rank histogram and the Brier score. In the rank histogram, bars indicate the verification wind speed frequency at each position within a rank given by the ensemble forecast members. The position zero indicates the verifying wind speed was lower than all ensemble members. Thus, higher frequency at lower positions indicates the verifying field is often underestimated by the ensemble. The histograms were created using the ensemble forecast for all data points: 4635 points in space, and 5773 points in time, yielding $\sim 26\,900\,000$ points.

Forecast probabilities are verified using the Brier score, which corresponds to the mean squared error of probabilistic forecasts, and its components, reliability, resolution, and uncertainty (the equations are presented in the [appendix](#)). Probability bins of size 0.2 are used to aggregate probabilities of exceeding 4, 6, 8, and 10 m s^{-1} thresholds. In the Brier score, probabilities similar to the binary outcome (true or false) yield values close to zero, whereas larger discrepancies between the probability of exceeding a threshold result in values closer to one ([Wilks 2006](#)).

3. Three-step object selection and object matching technique

The method developed for object selection is an essential component of the Obj-An forecast, in that it enables object matching between two fields. Continuous fields, such as wind speed, exhibit spatial patterns influenced by a combination of atmospheric state and physical boundaries (e.g., topography). It is common practice to use object-based techniques in forecast verification, which are applied to identify these patterns and delineate regions of interest within the field. Objects within two gridded fields are identified and must be matched so that verification statistics can quantify the differences between corresponding objects. Current object selection techniques (including their respective matching algorithm) rely on thresholds to identify objects and are suited for discontinuous quantities (e.g., precipitation), but may yield poor matchings or poor object distinction when applied to continuous variables. Direct threshold application (sometimes preceded by image smoothing) often yields fewer objects than interest areas, or oversized objects enclosing distinct patterns.

The major challenge of matching objects in continuous fields is associated with magnitude differences when matching a forecast to its verifying analysis. Magnitude differences in forecast fields matched against respective analysis lead to numerous unmatched objects. When magnitudes differ substantially, the threshold applied during the object selection step is able to define objects in the higher magnitude field, but is not able to depict the corresponding object in the field of lower magnitude. Lowering the threshold altogether creates oversized objects in the overestimating field. Applying different thresholds in each field may work in some cases, but often reveals another challenge to object selection of continuous variables: thresholds limit the depiction of multiple objects of varying magnitude in the same field. As thresholds are lowered, objects of higher magnitude enclose larger areas and embed lower magnitude objects in distinct regions of the domain. In addition, varying thresholds to enable object matching would require a complex automation algorithm, which would unlikely be portable to other domains and applications, and consequently, is not a feasible solution for verifying a large forecast set.

a. Object selection

The technique we propose to select objects has three steps that ensure patterns are preserved and regions of different magnitudes are distinctly depicted. Essentially, three computationally efficient algorithms, standard in image processing, are combined as shown in [Fig. 2](#): 1) a median order-statistics filter, 2) an image histogram

FIG. 2. Steps used in the object selection for the 25th forecast lead hour initialized at 0000 UTC 13 Jun 2013. (from left to right) Raw forecast, forecast after the order statistics median filter (step 1), filtered forecast after the histogram equalization (step 2), segmentation created with the watershed transform (step 3), and resulting forecast objects.

equalization, and 3) a watershed transform.⁴ Refer to Acharya and Ray (2005) and Dougherty (1992) for a full description of the three algorithms.

Order-statistics filters are nonlinear filters based on robust statistics, used to suppress noise while enhancing local contrast and preserving edges in a signal. These characteristics are useful to prepare a forecast field for object selection, in that it separates distinct patterns without introducing blur. The median filter analyzes a square neighborhood of a grid point, sorts it in ascending order, and replaces it with the middle value.

Histogram equalization is applied to adjust the field magnitude and further enhance contrast. The equalization is such that the resulting cumulative distribution is transformed to a linear ramp, yielding equal probability bins. The equalization allows for depiction of low and high magnitude objects in the same field.

The watershed transform is used to identify the equivalent of a river catchment on a topographic surface (or image). Each pixel in an image is assigned to a “basin” defined by the influence zone of the local minima. Watersheds defined over actual topographic relief delineate the surface over which fluid would flow to the same end point. In this study, the algorithm result is a segmented field corresponding to the objects selected from a wind speed topography field.

The three-step object selection allows for a complete match between objects in the forecast fields and their verifying analysis because individual model fields are entirely segmented such that all grid points belong to an object. Objects produced with the three-step technique consist of model fields enclosed in the segmented regions. It is possible to apply thresholds to each individual region and refine the resulting objects according to specific thresholds, but this approach is not needed in this study. Applications with fixed thresholds may use the three-step technique to select objects and to guarantee objects in different images are matched. The complete field segmentation is the key to more accurate object matching.

b. Matching forecast objects

Let A and B be any two different 2D model gridded fields from which objects must be matched. Let $\mathbb{A}$ and $\mathbb{B}$ be the object set from fields A and B with n_A and n_B indicating the total number of objects in each set. The objects a_i and b_j denote any object element in sets $\mathbb{A}$ and $\mathbb{B}$, obtained with the three-step object selection of fields A and B (section 3a):

$$\begin{aligned}\mathbb{A} &= \{a_1, \dots, a_{n_A}\}, & a_i &\in \mathbb{A} \\ \mathbb{B} &= \{b_1, \dots, b_{n_B}\}, & b_j &\in \mathbb{B}.\end{aligned}\quad (1)$$

The object matching consists of assigning each object a_i to a corresponding object b_j , and the matching is complete when all objects in the sets $\mathbb{A}$ and $\mathbb{B}$ are matched. An object a_i match b_j (and vice versa) when their relative area intersection (RAI; given in grid-point units) is maximum. The simplest matching rule given by Eqs. (2) and (3) returns a positive match when the area intersection between two objects a_i and

⁴ Algorithms from IDL, a product of Exelis Visual Information Solutions, Inc., a subsidiary of Harris Corporation (Exelis VIS), <http://www.harrisgeospatial.com>. The functions and parameters in each step were: 1) estimator_filter, applied twice, using a two-dimensional neighborhood of width 5 (grid units); 2) hist_equal, using a 25% stretch; and 3) watershed, using a byte type inverted input (255b-input).

b_j is at least half the size of the largest object area. This means that a large object a_i perfectly intersecting with two smaller objects in $\mathbb{B}$ is matched to the largest b_j object:

$$\text{RAI}_{(a_i, b_j)} = \frac{\text{Area}_{a_i} \cap \text{Area}_{b_j}}{\max\{\text{Area}_{a_i}, \text{Area}_{b_j}\}}, \quad (2)$$

$$a_i \equiv b_j \quad \forall (a_i, b_j): \text{RAI}_{(a_i, b_j)} > 0.5. \quad (3)$$

When an object a_i partially intersects with two or more objects b_j , Eq. (3) does not yield a match for all objects in sets $\mathbb{A}$ and $\mathbb{B}$ because object areas may not intersect by at least half the area of the largest object. Unmatched objects are then matched using the RAI' given by Eq. (4), defined as the sum of the area intersection relative to each object area. In this case, the matching rule returns a positive match between objects that yield the maximum RAI' [Eq. (5)]. The maximum RAI' is calculated between any possible object pair in which at least one object is unmatched, as many times until all objects are matched. Objects merge when multiple objects in set $\mathbb{A}$ correspond to a single object in set $\mathbb{B}$ (and vice versa). Often, n_A and n_B are different, so merging must occur to result in a complete match of all elements in the sets:

$$\text{RAI}'_{(a_i, b_j)} = \left(\frac{\text{Area}_{a_i} \cap \text{Area}_{b_j}}{\text{Area}_{a_i}} + \frac{\text{Area}_{a_i} \cap \text{Area}_{b_j}}{\text{Area}_{b_j}} \right), \quad (4)$$

$$a_i \equiv b_j \quad \forall (a_i, b_j): \text{RAI}'_{(a_i, b_j)} = \max\{\text{RAI}'_{(a_i, b_j)}\}. \quad (5)$$

The example illustrated in Fig. 3a shows an object a_1 with 16 grid points that intersect with all 12 grid points of b_1 and all 4 grid points of b_2 . The $\text{RAI}_{(a_1, b_1)}$ is 0.75, therefore, a_1 and b_1 match, but $\text{RAI}_{(a_1, b_2)}$ is 0.25 and b_2 is not yet matched. Here b_2 is entirely contained in the area corresponding to a_1 , so they should also be a match. So, for b_2 , the total area intersection relative to a_1 is 0.25 and relative to itself, 1, yielding an $\text{RAI}'_{(a_1, b_2)}$ of 1.25. The term b_2 does not intersect with any other object, so the total relative area intersection between a_1 and b_2 is the maximum RAI' (i.e., 1).

Two or more objects in $\mathbb{B}$ may be equivalent to a single object in $\mathbb{A}$, and vice versa, as long as the rules in Eqs. (3) and (5) are satisfied for any unmatched object. In the example, b_2 only intersects with a_1 , so it is the only possible match. Figure 3b shows a more elaborate example in which object a_1 matches b_1 , and a_2 matches b_3 by their RAI [Eq. (3)]. The unmatched objects a_3 and b_2 are matched by their RAI' [Eq. (5)]. Object a_2 matches b_2

and object a_3 matches b_3 . Finally, objects a_2 and a_3 merge, and objects b_2 and b_3 merge because two objects are matched to one. That is, a_2 merges with a_3 because both match b_3 , and b_2 merges with b_3 because b_2 matches a_2 .

The result of the matching algorithm is shown in Fig. 4. The objects in the forecast used to illustrate the three-step object selection in Fig. 2 are matched with objects from two forecast-analog candidates. The matching algorithm allows the same forecast field matched to different forecast-analog candidates to merge differently according to fields being matched (Figs. 4a,b).

4. The Obj-An forecast method

The Obj-An forecast method builds upon the objects obtained with the three-step object selection and object matching described in sections 3a and 3b. In an operational forecast setting, the objects in the training set are selected in advance and only the real-time forecast objects must be created at run time. The analog candidates are ranked once the training set and real-time forecast objects are matched (and possibly merged). A flowchart illustrates the process in Fig. 5.

The best analogs are selected using a similarity measure [Eq. (6)] with two components: the object intersection ratio and object magnitude ratio. The measure is used to quantify the similarity between forecasts in the training set (analog candidates) and the current forecast. The object intersection component accounts for location and size, and the inclusion of a magnitude component completes a metric that accounts for the main object attributes used in this study:

$$\text{Similarity} = \text{intersect ratio} + \text{magnitude ratio}. \quad (6)$$

The intersection ratio of each analog candidate is given by the area-weighted sum of RAI [Eq. (2)] calculated for positive matches identified between objects in the forecast and candidate fields. The RAI for analog classification is calculated only over matched objects after objects are merged, following the procedure described in section 3b. Let A be the current forecast field and B an analog candidate, the intersection ratio is given by

$$\text{Intersect ratio} = \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} \frac{\text{Area}_{a_i}}{\text{Area}_A} \text{RAI}_{(a_i, b_j)}, \quad \forall a_i \equiv b_j. \quad (7)$$

A hypothetical perfect spatial match between forecast and analog candidate yields an intersection ratio equal to one.

Similarly, the magnitude ratio is the sum of the area-weighted wind speed ratio between forecast and candidate analog. The wind speed ratio of any two matched objects is given by the mean of ratios over the objects'

(a)

Objects	Intersection Area	Largest Object Area	Smallest Object Area	RAI	RAI'	Match
A ₁ , B ₁	12	16	12	12/16 = 0.75	-	true
A ₁ , B ₂	4	16	4	4/16 = 0.25	4/16+4/4 = 1.25	true

(b)

Objects	Intersection Area	Largest Object Area	Smallest Object Area	RAI	RAI'	Match
A ₁ , B ₁	12	16	12	12/16 = 0.75	-	true
A ₁ , B ₂	4	16	8	4/16 = 0.25	4/16+4/8 = 0.75	false
A ₂ , B ₁	-	-	-	-	-	-
A ₂ , B ₂	3	8	6	3/8 = 0.375	3/8+3/6 = 0.875	true
A ₃ , B ₁	-	-	-	-	-	-
A ₃ , B ₂	1	8	2	1/8 = 0.125	1/8+2/8 = 0.375	false
A ₁ , B ₃	-	-	-	-	-	-
A ₂ , B ₃	3	6	4	0.5	-	true
A ₃ , B ₃	1	4	2	1/4 = 0.25	1/4+1/2 = 0.75	true

FIG. 3. Schematics for the matching procedure with relative area intersection [Eqs. (3) and (5)]. (a) Objects B₁ and B₂ match A₁. (b) According to Eq. (3), objects A₁ matches B₁, and A₂ matches B₃. The unmatched objects A₃ and B₂ are matched by Eq. (5). Object A₂ matches B₂ and object A₃ matches B₃. Objects merge when two or more objects are matched to one. Thus, A₂ merges with A₃ because both match B₃, and B₂ merges with B₃ because B₂ matches A₂.

intersecting grid points, represented by f_{a_i} and f_{b_j} , respectively. When the mean wind speed is equal between forecast and analog objects, the magnitude ratio has a value of 1. To limit the range of magnitude ratios between $[0, 1]$, the object magnitude is given by the minimum between f_{a_i}/f_{b_j} and f_{b_j}/f_{a_i} as follows:

$$\text{Magnitude ratio} = \sum_{i=1}^{n_A} \min \left\{ \frac{f_{a_i}}{f_{b_j}}, \frac{f_{b_j}}{f_{a_i}} \right\} \frac{\text{Area}_{a_i}}{\text{Area}_A}, \quad \forall a_i \equiv b_j. \quad (8)$$

The sum of intersection and magnitude ratios corresponds to the similarity measure used to classify the

analog candidates, such that identical object fields measure 1 and fields less similar approach zero.

In Figs. 4a and 4b, two analog candidates selected using the similarity measure in Eq. (6) are shown. The analog candidate in Fig. 4a corresponds to the highest ranked candidate, and the one in Fig. 4b corresponds to the fifth best candidate.

5. Results

The Obj-An method is verified using deterministic and probabilistic verification metrics. The Obj-An method performance is compared to the performance of four other analog ensemble methods and Raw.

FIG. 4. Matching procedure applied to forecast and analog candidate. (a),(b) Matchings between the forecast objects shown in Fig. 2 and objects from different analog candidates are illustrated. The merged objects in a field change according to the objects in the field to which it is matched. The merged objects improve the object intersection area between any pair of forecast and analog candidate. In (a), the forecast object 1 is matched to six small objects in the candidate field, which merge to become the corresponding candidate object 1; the forecast object 2 is a merge itself and matches the candidate object 2, also given by a merge of two objects; and similarly for object 3. In (b), the forecast objects merge differently than in (a). In (b), the forecast object 2 is not a merged object itself, and it matches two objects corresponding to candidate object 2; similarly, the forecast objects 3 and 8 are result of different mergings than those in (a).

The mean field from the analog ensembles is shown for the forecast field used to illustrate the three-step object selection (Fig. 2) and the Obj-An candidate selection (Fig. 4) for the 13 June 2013 event's 25th lead hour. Figure 6 shows the Raw forecast field; the mean of the Obj-An, Grid-An, Field-An, Random, and Best-An ensembles; and the corresponding verification field. The Grid-An mean field shows higher systematic errors and higher spatial variability because the analog selection technique does not consider neighboring grid points. The Random is created from randomly selected fields, not grid points, and consequently exhibit a certain degree of spatial consistency when compared to Grid-An.

a. Forecast verification over study area

The deterministic verification metrics described in section 2b—Bias, MAE, CRMSE, and correlation of forecasts (Raw and ensemble means)—are summarized in Fig. 7a. The analog forecasts improve the Raw forecast systematic and random errors and differences among the analog forecasts indicate that Obj-An and Field-An outperform Grid-An and Random.

Qualitatively, the MAE and CRMSE medians are equivalent among Obj-An, Field-An, and Grid-An, yet the metrics' whiskers are narrower in the Obj-An and Field-An forecasts. In Grid-An and Random, biases are negligibly lower, and the bias whiskers are narrower than Obj-An and Field-An (up to $\sim \pm 0.5 \text{ m s}^{-1}$). The Grid-An and Random bias boxplots indicate the forecast biases converge toward zero but the higher values of MAE and CRMSE (up to $\sim 1 \text{ m s}^{-1}$) reveal the lower biases do not represent lower systematic error, rather,

errors oscillate and on the average, the forecasts converge to the sample climatological mean. The narrower whiskers in the MAE and CRMSE boxplots of Obj-An and Field-An indicate the metrics' spatial variability is lower. That is, errors are generally lower over the forecast domain in Obj-An and Field-An.

The Raw and Obj-An forecast correlation coefficient distributions are very similar. The Field-An forecast correlation whiskers indicate slightly higher coefficients (medians within 0.55–0.6) than Raw and Obj-An. Grid-An and Random show considerably lower correlation coefficients (0.4 and 0, respectively), which indicates that, although the error metrics suggest these forecasts reduce the overall forecast error, their relationship to the verification is weaker.

The forecasts from Obj-An and Field-An show comparable metrics. This result is expected in that these forecasts are conceptually similar, that is, all grid points in each analog ensemble member belong to a single wind speed field from the training set. The analogs generated from 2D fields produce similar metrics in terms of magnitude and spread, and some of these forecast errors originate from the enforced spatial dependence when creating analog forecasts from 2D fields. For example, it is possible that in some areas over the domain a local similarity measure would yield poorer similarity, but the overall similarity measure is high, so the field is selected as an analog. Assuming that a perfect similarity measure yields the best forecast analog, a high but imperfect similarity measure where subregions of lower similarity exist degrades the overall forecast skill.

In the Random ensemble, a similar number of positive and negative forecast errors is expected such that the

FIG. 5. The flowchart shows the sequence of processes required to produce the Obj-An ensemble forecast.

mean yields zero bias and fields show nearly zero spatial correlation. In Grid-An, each grid point has its own set of ensemble members, such that errors are spatially independent, so biases converges to zero. Yet, because the Grid-An ensemble members are selected with a criterion, as opposed to Random that has no similarity measure, the Grid-An fields show higher spatial correlation than Random.

The Best-An forecast metrics show that overall, the training set contains analogs that could consistently improve the metrics over the Raw forecast. The MAE and CRMSE interquartiles are lower than the other forecasts ($\sim 0.7 \text{ ms}^{-1}$, $\sim 1 \text{ ms}^{-1}$) and the correlation is higher (0.86). Bias and correlation show a skewed spatial distribution. The bias tail stretches toward negative values

(forecast lower than verification) and the correlation tail indicate that some forecast regions are less correlated to the verification. The Best-An is created directly from verification fields, so the analog ensemble errors are not intrinsically forecasting errors, rather, they reveal the sufficiency of the training set. The Best-An biases are mostly a consequence of the more severe events not having sufficiently similar fields in the training set. Ensembles are created with 15 members, regardless of the degree of similarity between fields, and members are required to be from different events to prevent temporal dependence among members. However, the most severe storms in the training set are limited in number. So, the less similar members in the resulting ensemble are storms of lower magnitude, which in turn lead to negative bias and lower correlation.

Direct forecast comparisons (Fig. 7b) between each analog ensemble mean and Raw indicate whether the absolute and random error metrics are statistically different. The results in Fig. 7a show the metric distributions interquartile range often overlap and some of the metrics overall distributions are only minimally shifted. The forecast comparisons in Fig. 7b provide additional information regarding the significance of the metric differences, and ultimately whether one forecast shows consistently lower errors than the other. The confidence intervals enclosing the 10th and 90th percentiles show that forecasts errors (MAE and CRMSE) from Obj-An, Field-An, and Best-An are consistently lower than Raw. Conversely, the statistical difference between the Grid-An and Raw errors is not significant. The Random absolute error is significantly higher than the error from Raw, and the CRMSE shows no statistical difference.

The forecast correlations are tested with a Steiger's Z test (Steiger 1980) to identify the statistical significance of the differences between two correlation coefficients. The test results comparing the Raw correlation coefficient to each of the four analog ensemble correlations (all tested coefficients refer to a forecast-verification set) do not indicate correlated correlations (plot not shown).

Figure 7c shows the MAE and CRMSE skill scores for metrics aggregates in time and in space. The top row shows the skill scores for land grid points, with metrics calculated over forecast hours. Some of the grid points in Raw and Grid-An forecasts exhibit negative MAE skill, indicating that at some grid points, a forecast based on climatology is overall better. The MAE and CRMSE skill scores of Obj-An and Field-An are similar to each other, being the difference between mean values of ~ 0.03 . The Best-An skill score is higher than the scores of all the other forecasts (mean of ~ 0.5 , at least twice as skillful than Obj-An and Field-An).

FIG. 6. Forecasts for the 25th forecast lead hour initialized at 0000 UTC 13 Jun 2013. Raw, analog ensemble means (Obj-An, Field-An, Grid-An, Random, and Best-An), and corresponding verification.

The bottom row of Fig. 7c shows the skill scores for the events' forecast times. The wider distribution represented by the boxplots show that there is more variability in skill when the metrics represent different forecast times and distinct events. The MAE skill score indicate that some forecasts may be almost perfect (MAE equals zero and skill score equals one), whereas others may show no skill. The MAE and CRMSE skill scores of the Obj-An, Field-An, and Grid-An ensembles show minimal differences.

This difference in skill between the Best-An and the other analog forecasts provide important insights into what may be accomplished with analog ensembles. When skill is segregated in space, it indicates that the training set may potentially provide more skillful analogs over specific regions, given that an adequate similarity measure is found. When forecasts are segregated in time, the training set skill limits revealed by the Best-An indicate that the potential for improvements in the analog methods is lower. That is, some forecasts have little to gain with an analog method, which may be a result of an insufficient training set for specific storm types and/or a consequence of the skill-lead time dependence.

b. Point forecast evaluation

The training set skill is further explored with forecasts presented at grid points to confirm results indicated by metrics averaged over space and time. In Fig. 8, three grid points locations are selected and the point-based forecasts are compared to the corresponding verification reanalysis. At point location we compare the forecast ensemble members with their corresponding mean. The point locations are placed over a diagonal line across the domain to verify forecasts at locations characterized by different terrain heights and to contrast coastal versus inland influence. The geographical location of these points is indicated in Fig. 1. Each panel in Fig. 8 shows the Raw wind speed forecast (in m s^{-1}) on top of 2 columns containing the forecast ensemble mean (left) and the 15 ensemble members (right) of Obj-An, Grid-An, Field-An, and Best-An (plots for Random are not shown).

Higher wind speeds are observed at point 3, followed by point 1. The Raw forecast overestimates wind speeds higher than the verification's 99th wind speed percentile of $\sim 10 \text{ m s}^{-1}$ on point 3 more frequently, whereas it underestimates them at points 1 and 2.

The ensemble mean given by the Best-An is a performance baseline for other analog ensembles created from the same training set. Events in which the Best-An forecast verifies with a higher absolute error, are events for which analog techniques are not expected to exhibit forecast skill. When the verification indicates high wind speed (above $\sim 10 \text{ m s}^{-1}$), the Best-An members are

FIG. 7. Wind speed verification metrics. (a) Metrics show bias (m s^{-1}), mean absolute error (MAE, m s^{-1}), centered root-mean-square error (CRMSE, m s^{-1}), and correlation coefficient of Raw and the four analog ensemble means: Obj-An, Grid-An, Field-An, Random, and Best-An. (b) MAE and CRMSE of forecasts (ensemble means and Raw) are directly compared and the confidence intervals indicate the 10th and 90th percentiles of forecast differences. (c) MAE and CRMSE skill scores calculated using Random as climatological reference.

mostly clustered under the main diagonal, with only a few analogs indicating similar wind speeds. This effect is caused by a low frequency of high wind speed events in the training set and reflects directly on the Best-An ensemble, which is populated with less similar members

(in this case, with lower wind speed). Consequently, the Obj-An, Grid-An, and Field-An forecasts of higher wind speeds are penalized, not due to the methods themselves, but because the training set is limited and does not provide good analogs.

FIG. 8. Scatterplots of forecasts vs reanalysis verification fields. Panels show the Raw wind speed forecast (m s^{-1}) on top of 2 columns containing (left) the wind speed forecast ensemble mean and (right) the 15 ensemble members of Obj-An, Grid-An, Field-An, and Best-An. The color bar indicates the density of points in the corresponding subplot columns. Panels correspond to the three locations indicated in Fig. 1.

The point forecast evaluation suggests that the best analog method performance varies between Field-An and Obj-An according to region. The ensemble mean indicates that at point 3, Obj-An is able to forecast higher wind speeds; at point 1, Field-An provides better analogs, and in all three points Grid-An is the less skillful with respect to the 95th verification percentile of 7 m s^{-1} . The cluster of dots formed by the ensemble members suggests that Grid-An retrieves analogs that overpredict lower wind speeds (in the cluster of dots, the ellipsoid main axis formed by the graph's main diagonal is shorter and the perpendicular axis is longer in Grid-An).

c. Probabilistic forecast verification

The verification rank histogram in Fig. 9 shows whether the ensemble members are equally likely. The ensembles produced with the Grid-An, and Random methods are statistically consistent, that is, the probability of outcomes at any position in the rank within the ensemble distribution is generally equal, and consequently, the histogram is approximately flat. The Obj-An

rank shows the verification tends to occur slightly more often in the first four positions and slightly less often in the last four (frequency departures are less than 0.5% from the uniform histogram). Ensembles from the Field-An show a similar trend, but the positive forecast bias (verification frequently at lower positions in the rank) is more frequent, as previously identified in Fig. 7, with departures of up to 2%.

The Best-An ensemble shows the verification wind speed may be lower than any of the ensemble members', but more often, the verification is at the three higher rank positions (indicated by the higher frequency of these positions in the histogram). However, the Best-An is not an actual forecast, it is an ensemble with the most similar verification fields that could be drawn from the training set. Hence, its histogram shape also reveals that the best 15 verification fields of analog candidates are not as similar to the forecast verification field. Higher bars appear at the highest rank positions showing that the ensemble somewhat underestimates the verification wind speed. It reveals that some events are not well represented by the magnitudes contained in the training

FIG. 9. Verification rank histogram for the ensemble members of Obj-An, Grid-An, Field-An, Random, and Best-An.

set, and in such events, any analog method is expected to fail.

In Random, the uniform histogram is expected, in that samples are drawn from the climatological distribution. Yet, it is not expected to find resolution on Random because forecasts have the same probability on any occasion.

Figure 10 shows the Brier score and Brier score components: reliability, resolution, and uncertainty. The scores are calculated for the three locations indicated in Fig. 1.

The forecast accuracy given by the reliability term varies according to region and wind speed threshold. In points 1 and 2, Obj-An and Field-An are the most reliable in the 4 and 6 m s^{-1} thresholds. In point 1, the Best-An reliability term worsens at the 6 m s^{-1} threshold. Both, Grid-An and Random are the least reliable. In point 3, the sample reliability indicates the Obj-An is more reliable; however, the confidence intervals of the 10th and 90th percentiles indicate that the ensemble reliabilities almost always overlap.

The resolution term shows the Best-An resolution is the best in all three points, Obj-An and Field-An are statistically better resolved than Grid-An in points 1 and 2, and Random shows no resolution, as expected. The Grid-An resolution indicates these analogs are weak to distinguish events, and the scatterplot of ensemble members discussed in Fig. 8 reveal a poor ability to find high-speed analogs. Yet, the one-dimensional ensembles from Grid-An (no spatial dependence required among grid points) provide a higher probability of finding more similar analogs (given its lower number of degrees of freedom in the analog search), which does not translate into better forecasts. On the contrary, analogs become indistinguishably similar, yielding low forecast resolution and poorer overall skill.

The climatological uncertainty is the largest component of the Brier score. In this analysis, the uncertainties of lower thresholds correspond to high probability events (>0.8), whereas those of higher thresholds correspond to rare occurrences (<0.1). These uncertainties represent common and uncommon events (4 and 10 m s^{-1} thresholds, respectively) and constitute the largest terms in the Brier score.

The Brier score terms referring to the 8 and 10 m s^{-1} thresholds indicate forecasts are reliable with poor resolution, and consequently, the value expressed by the Brier score corresponds mostly to the uncertainty term.

The Brier score of Obj-An and Field-An forecasts indicates these are more accurate and precise than the Grid-An forecasts. The result originates mostly in the resolution component, with a lower contribution from reliability.

6. Conclusions

This study describes a new analog technique to produce statistical forecasts based on spatial characteristics of flow. Image processing algorithms are used to create objects of a forecast field, segmenting the forecast domain according to instantaneous spatial patterns provided by an NWP model. The proposed Object Analog (Obj-An) technique searches for similar forecasts in a forecast training set using an object-based similarity measure: object area intersection and object magnitude. A statistical ensemble is then created using the verification fields of the corresponding best analogs. We apply the Obj-An technique to generate 15-member ensembles of wind speed forecasts, over a subregion of

FIG. 10. Brier score and components (reliability, resolution, and uncertainty) for four wind speed thresholds: 4, 6, 8, and 10 m s^{-1} . Scores are given for the three locations indicated in Fig. 1. Confidence intervals indicate the bootstrap 10th and 90th percentiles.

the NE. The training set comprises events that caused storm-induced damages to the power distribution lines in the states of Connecticut and Massachusetts. The training set used in the cross-validation and method's forecast verification is restricted to verified positive outcomes (occurrence of power outages) and does not include false forecast positives or negative outcomes. Ensemble forecasts created with the Obj-An are compared with other analog techniques: 1) a single gridpoint analog (Grid-An), 2) a 2D field analog (Field-An), 3) a random ensemble (Random), and 4) the best possible ensemble (based solely on the verifying field, Best-An). Random and Best-An are used as lower and upper limit references of skill, respectively. The comparison among the analog techniques provides a relative measure of performance for the Obj-An to other analog forecasts created from the same training set, with the five ensemble verification results compared to the uncalibrated deterministic forecast (Raw).

Overall, the most skillful analog forecasts are the Obj-An and Field-An approaches. In comparison, the Grid-An overestimates low wind speeds more often, has lower reliability, and lower resolution than the Obj-An and Field-An. The gridpoint independence in the similarity measure of the Grid-An favors excessive similarities among all analog candidates because poor similarities in neighboring points cannot penalize the

single gridpoint classification. Because samples from Grid-An are less constrained, forecast analogs of common events may be indistinguishably similar. A potential solution to improve skill in the Grid-An would be to include other criteria, as done in Delle Monache et al. (2011, 2013), who used temporally constrained analogs and weights based on similarity. In Field-An and Obj-An, a spatial constraint is imposed implicitly by having all grid points (i.e., the entire 2D field) from the same forecast for each analog member.

The Best-An ensemble is important to identify limitations associated with the training set that otherwise may be attributed to the analog forecast technique (e.g., insufficient high-wind-speed analogs). The “best possible analog forecast” is an important baseline against which analog forecasts can be referenced to assess training set size sufficiency.

The advantages of the Obj-An method over other methods in the literature include lower computational costs due to an analog selection algorithm utilizing a reduced spatial dimensionality. The method implicitly has skill in identifying flow-dependent analogs by searching for similar spatial patterns, and also provides benefits in generating ensembles with realistic spatial relationships.

The Obj-An is a 2D spatial analog method and is different from analogs obtained according to similarities

in initial conditions due to the source of uncertainty they characterize. Analogs based on initial conditions express uncertainties associated with the evolution of physical processes or model error saturation with time, whereas the Obj-An estimate spatial uncertainties based on model errors at a given time. The Obj-An does not attempt to predict the atmospheric state evolution, instead, it creates a statistical forecast for an instantaneous atmospheric state, based only on spatial differences between model and reference. In this framework, the Obj-An probabilities inform the past model error given by the analog set and their corresponding verification, for a forecast with similar spatial patterns. Spatial patterns in the wind field may be used to detect dynamical features reflected on the flow. Patterns obtained from flows can distinguish among states induced by different dynamical phenomena occurring at different spatial scales (e.g., a wind field from a tropical storm is different than one generated by convective cells or terrain-induced flows) and identify features produced by specific flow–terrain interaction patterns (e.g., terrain shadow effect). Frediani et al. (2016) show that forecasts classified according to wind fields are naturally grouped into physically meaningful categories implicitly classifying model states according to event scale, season, and diurnal cycle. Consequently, spatial patterns may be sufficient to aggregate similar meteorological phenomena into categories that better characterize the associated model error and uncertainty.

The Obj-An method may be further improved with the assignment of weights or adjustment coefficients to analog members according to their degree of similarity. Future research goals include investigating individual object selection, in that a selection is made separately for each object in the field, which is then reassembled into a single probability field.

Acknowledgments. This research was carried out with partial funding from Eversource Energy and data provided by its subsidiary, Connecticut Light and Power. The views expressed are those of the authors and do not necessarily represent the official policy or position of the utility company. This study was enabled by a long-term visit by M. Frediani to NCAR supported by NCAR's Advanced Study Program (ASP). T. Hopson's involvement was partially funded by the U.S. Army Test and Evaluation Command (ATEC) through an interagency agreement with the National Science Foundation. Computational aspects of this research was supported by the high-performance computational resources by Yellowstone (ark:/85065/d7wd3xhc) maintained by NCAR's Computational and Information Systems Laboratory. The National Center for Atmospheric

Research is sponsored by the National Science Foundation. I would like to express my deepest gratitude to my advisors, Manos, Tom, and Josh, for all the support I received along my path.

APPENDIX

Verification Metrics

a. Bias, MAE, CRMSE, and correlation

The forecasts are verified using point verification metrics. The mean squared error (MSE) is represented through its two components, bias and CRMSE:

$$\text{Bias} = \frac{1}{n} \sum_{i=1}^n (f_i - v_i) = \bar{f} - \bar{v}, \quad (\text{A1})$$

$$\text{CRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n [(f_i - \bar{f}) - (v_i - \bar{v})]^2}, \quad (\text{A2})$$

where f_i and v_i are the forecast and corresponding verification, respectively; and n is the total number of forecast–verification pairs at each grid point. Bias and CRMSE represent the systematic and random error components, which combined yield the MSE:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (f_i - v_i)^2 = \text{Bias}^2 + \text{CRMSE}^2. \quad (\text{A3})$$

The MAE and Pearson correlation coefficient are calculated as follows:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |f_i - v_i|, \quad (\text{A4})$$

$$\text{Correlation} = \frac{\sum_{i=1}^n (f_i - \bar{f})(v_i - \bar{v})}{\sqrt{\sum_{i=1}^n (f_i - \bar{f})^2} \sqrt{\sum_{i=1}^n (v_i - \bar{v})^2}}. \quad (\text{A5})$$

b. Brier score and Brier score components

The Brier score (BS) is a proper score to verify probabilistic forecasts given a binary outcome. The Brier score formulation used in this study is

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2, \quad (\text{A6})$$

where i denotes the N forecast–outcome pairs, o_i is the binary outcome ($o_i = 1$ if outcome is true and $o_i = 0$ if

outcome is false), and f_i is the forecast outcome probability given by the ensemble. The Brier score components with respect to m mutually exclusive probability bins with index $k = 1, 2, \dots, m$ are given by

$$\text{BS} = \text{Rel} - \text{Res} + \text{Uncertainty}$$

$$\text{Rel} = \frac{1}{N} \sum_{k=1}^m n_k (\bar{f}_k - \bar{o}_k)^2$$

$$\text{Res} = \frac{1}{N} \sum_{k=1}^m n_k (\bar{o}_k - \bar{o})^2 + \text{WBV} - \text{WBC}$$

$$\text{Uncertainty} = \bar{o}(1 - \bar{o}) \quad (\text{A7})$$

$$\text{WBV} = \frac{1}{N} \sum_{k=1}^m \sum_{j=1}^{n_k} (f_{kj} - \bar{f}_k)^2$$

$$\text{WBC} = \frac{2}{N} \sum_{k=1}^m \sum_{j=1}^{n_k} (o_{kj} - \bar{o}_k)(f_{kj} - \bar{f}_k),$$

such that n_k is the number of probability forecasts in the k th bin, denoted by f_{kj} for $j = 1, 2, \dots, n_k$; and $\sum_{k=1}^m n_k = N$. Here $\bar{f}_k$ is the average forecast bin probability, $\bar{o}_k$ is the bin frequency of positive outcomes, and $\bar{o}$ is the climatological mean probability. The terms representing within-bin variance (WBV) and covariance (WBC) are included in the resolution term according to [Stephenson et al. \(2008\)](#) so that the sum of components yields the exact Brier score.

The reliability term (or calibration) refers to the bias of forecast probabilities conditioned on outcomes, the resolution term (or refinement) refers to the variance of forecast probabilities conditioned on outcomes, and the uncertainty term refers to the climatological uncertainty. The reliability term provides a measure of distance between conditioned binary outcomes and the forecast probability of each conditioning bin. The Brier score and the reliability term are negatively oriented, that is values close to zero indicate more accuracy and reliability. High probabilities forecasting an event with a *true* outcome, and low probabilities forecasting a *false* outcome yield low values of the reliability term. Likewise, higher values of the reliability term indicate large discrepancies between forecast and outcomes.

High resolution (or refinement) indicate that forecasts are able to discern among different events. The outcomes conditioned to their corresponding forecast probability are compared to the climatological uncertainty of the event. Accurate predictions of common events result in lower resolution, whereas accurate predictions of rarer events result in higher resolution.

In the climatological uncertainty term, outcomes with probability around 0.5 have higher uncertainty (uncertainty equals 0.25 for events with a probability of 0.5) than outcomes with higher or lower probability.

REFERENCES

- Acharya, T., and A. K. Ray, 2005: *Image Processing: Principles and Applications*. John Wiley & Sons, 425 pp.
- Ahijevych, D., E. Gilleland, B. G. Brown, and E. E. Ebert, 2009: Application of spatial verification methods to idealized and NWP-gridded precipitation forecasts. *Wea. Forecasting*, **24**, 1485–1497, <https://doi.org/10.1175/2009WAF2222298.1>.
- Alessandrini, S., L. Delle Monache, S. Sperati, and J. Nissen, 2015: A novel application of an analog ensemble for short-term wind power forecasting. *Renewable Energy*, **76**, 768–781, <https://doi.org/10.1016/j.renene.2014.11.061>.
- Auroux, D., and J. Blum, 2008: A nudging-based data assimilation method: The Back and Forth Nudging (BFN) algorithm. *Nonlinear Processes Geophys.*, **15**, 305–319, <https://doi.org/10.5194/npg-15-305-2008>.
- Briggs, W. M., and R. A. Levine, 1997: Wavelets and field forecast verification. *Mon. Wea. Rev.*, **125**, 1329–1341, [https://doi.org/10.1175/1520-0493\(1997\)125<1329:WAFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(1997)125<1329:WAFV>2.0.CO;2).
- Brown, B. G., R. Bullock, J. H. Gotway, D. Ahijevych, C. Davis, E. Gilleland, and L. Holland, 2007: Application of the MODE object-based verification tool for the evaluation of model precipitation fields. *22nd Conf. on Weather Analysis and Forecasting/18th Conf. on Numerical Weather Prediction*, Park City, UT, Amer. Meteor. Soc., 10A.2, https://ams.confex.com/ams/22WAF18NWP/techprogram/paper_124856.htm.
- Casati, B., G. Ross, and D. B. Stephenson, 2004: A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteor. Appl.*, **11**, 141–154, <https://doi.org/10.1017/S1350482704001239>.
- , and Coauthors, 2008: Forecast verification: Current status and future directions. *Meteor. Appl.*, **15**, 3–18, <https://doi.org/10.1002/met.52>.
- 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., <https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/19950009331.pdf>.
- Davis, C., B. Brown, and R. Bullock, 2006a: Object-based verification of precipitation forecasts. Part I: Methodology and application to mesoscale rain areas. *Mon. Wea. Rev.*, **134**, 1772–1784, <https://doi.org/10.1175/MWR3145.1>.
- , —, and —, 2006b: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1796, <https://doi.org/10.1175/MWR3146.1>.
- Delle Monache, L., 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>.
- , F. A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. *Mon. Wea. Rev.*, **141**, 3498–3516, <https://doi.org/10.1175/MWR-D-12-00281.1>.
- Djalalova, I., L. Delle Monache, and J. Wilczak, 2015: PM_{2.5} analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. *Atmos. Environ.*, **108**, 76–87, <https://doi.org/10.1016/j.atmosenv.2015.02.021>.

- Dougherty, E., 1992: *An Introduction to Morphological Image Processing*. Tutorial Texts in Optical Engineering, Vol. 9, SPIE Optical Engineering Press, 161 pp.
- Ebert, E. E., 2008: Fuzzy verification of high-resolution gridded forecasts: A review and proposed framework. *Meteor. Appl.*, **15**, 51–64, <https://doi.org/10.1002/met.25>.
- Fraedrich, K., C. C. Raible, and F. Sielmann, 2003: Analog ensemble forecasts of tropical cyclone tracks in the Australian region. *Wea. Forecasting*, **18**, 3–11, [https://doi.org/10.1175/1520-0434\(2003\)018<0003:AEFOTC>2.0.CO;2](https://doi.org/10.1175/1520-0434(2003)018<0003:AEFOTC>2.0.CO;2).
- Frediani, M. E. B., 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, <https://doi.org/10.1175/WAF-D-15-0139.1>.
- Gilleland, E., D. Ahijevych, B. G. Brown, B. Casati, and E. E. Ebert, 2009: Intercomparison of spatial forecast verification methods. *Wea. Forecasting*, **24**, 1416–1430, <https://doi.org/10.1175/2009WAF2222269.1>.
- , J. Lindström, and F. Lindgren, 2010: Analyzing the image warp forecast verification method on precipitation fields from the ICP. *Wea. Forecasting*, **25**, 1249–1262, <https://doi.org/10.1175/2010WAF2222365.1>.
- Grell, G. A., and D. Dévényi, 2002: A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophys. Res. Lett.*, **29**, <https://doi.org/10.1029/2002GL015311>.
- Hamill, T. M., and J. S. Whitaker, 2006: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Mon. Wea. Rev.*, **134**, 3209–3229, <https://doi.org/10.1175/MWR3237.1>.
- , G. T. Bates, J. S. Whitaker, D. R. Murray, M. Fiorino, T. J. Galarneau, Y. Zhu, and W. Lapenta, 2013: NOAA's second-generation global medium-range ensemble reforecast dataset. *Bull. Amer. Meteor. Soc.*, **94**, 1553–1565, <https://doi.org/10.1175/BAMS-D-12-00014.1>.
- , M. Scheuerer, and G. T. Bates, 2015: Analog probabilistic precipitation forecasts using GEFS reforecasts and climatology-calibrated precipitation analyses. *Mon. Wea. Rev.*, **143**, 3300–3309, <https://doi.org/10.1175/MWR-D-15-0004.1>.
- He, J., D. W. Wanik, B. M. Hartman, E. N. Anagnostou, M. Astitha, and M. E. B. Frediani, 2017: Nonparametric tree-based predictive modeling of storm outages on an electric distribution network. *Risk Anal.*, **37**, 441–458, <https://doi.org/10.1111/risa.12652>.
- Hoffman, R. N., Z. Liu, J.-F. Louis, and C. Grassoti, 1995: Distortion representation of forecast errors. *Mon. Wea. Rev.*, **123**, 2758–2770, [https://doi.org/10.1175/1520-0493\(1995\)123<2758:DROFE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1995)123<2758:DROFE>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, <https://doi.org/10.1175/MWR3199.1>.
- Hopson, T. M., and P. J. Webster, 2010: A 1–10-day ensemble forecasting scheme for the major river basins of Bangladesh: Forecasting severe floods of 2003–07. *J. Hydrometeorol.*, **11**, 618–641, <https://doi.org/10.1175/2009JHM1006.1>.
- Jiménez, P. A., and J. Dudhia, 2012: Improving the representation of resolved and unresolved topographic effects on surface wind in the WRF Model. *J. Appl. Meteor. Climatol.*, **51**, 300–316, <https://doi.org/10.1175/JAMC-D-11-084.1>.
- Keller, J. D., L. Delle Monache, and S. Alessandrini, 2017: Statistical downscaling of a high-resolution precipitation reanalysis using the analog ensemble method. *J. Appl. Meteor. Climatol.*, **56**, 2081–2095, <https://doi.org/10.1175/JAMC-D-16-0380.1>.
- Kruizinga, S., and A. H. Murphy, 1983: Use of an analogue procedure to formulate objective probabilistic temperature forecasts in the Netherlands. *Mon. Wea. Rev.*, **111**, 2244–2254, [https://doi.org/10.1175/1520-0493\(1983\)111<2244:UOAAPT>2.0.CO;2](https://doi.org/10.1175/1520-0493(1983)111<2244:UOAAPT>2.0.CO;2).
- Lack, S. A., G. L. Limpert, and N. I. Fox, 2010: An object-oriented multiscale verification scheme. *Wea. Forecasting*, **25**, 79–92, <https://doi.org/10.1175/2009WAF2222245.1>.
- Liu, Y., A. Bourgeois, T. Warner, S. Swerdlin, and J. Hacker, 2005: Implementation of observation-nudging based FDDA into WRF for supporting ATEC test operations. *2005 WRF Users' Workshop*, Boulder, CO, 4 pp., https://www.researchgate.net/publication/228942774_Implementation_of_observation-nudging_based_on_FDDA_into_WRF_for_supporting_AFEC_test_operations.
- , and Coauthors, 2008: The operational mesogamma-scale analysis and forecast system of the U.S. Army Test and Evaluation Command. Part I: Overview of the modeling system, the forecast products, and how the products are used. *J. Appl. Meteor. Climatol.*, **47**, 1077–1092, <https://doi.org/10.1175/2007JAMC1653.1>.
- Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, [https://doi.org/10.1175/1520-0469\(1963\)020%3C0130:DNF%3E2.0.CO;2](https://doi.org/10.1175/1520-0469(1963)020%3C0130:DNF%3E2.0.CO;2).
- Manor, A., and S. Berkovic, 2015: Bayesian inference aided analog downscaling for near-surface winds in complex terrain. *Atmos. Res.*, **164**–**165**, 27–36, <https://doi.org/10.1016/j.atmosres.2015.04.014>.
- Martín, M. L., F. Valero, A. Pascual, J. Sanz, and L. Frias, 2014: Analysis of wind power productions by means of an analog model. *Atmos. Res.*, **143**, 238–249, <https://doi.org/10.1016/j.atmosres.2014.02.012>.
- McGill, R., J. W. Tukey, and W. A. Larsen, 1978: Variations of box plots. *Amer. Stat.*, **32**, 12–16, <https://doi.org/10.2307/2683468>.
- 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, <https://doi.org/10.1029/97JD00237>.
- Namias, J., 1951: General aspects of extended-range forecasting. *Compendium of Meteorology*, T. F. Malone, Ed., Amer. Meteor. Soc., 802–813.
- Panziera, L., U. Germann, M. Gabella, and P. V. Mandapaka, 2011: NORA—Nowcasting of Orographic Rainfall by means of Analogues. *Quart. J. Roy. Meteor. Soc.*, **137**, 2106–2123, <https://doi.org/10.1002/qj.878>.
- Pinto, J. O., A. J. Monaghan, L. Delle Monache, E. Vanvyve, and D. L. Rife, 2014: Regional assessment of sampling techniques for more efficient dynamical climate downscaling. *J. Climate*, **27**, 1524–1538, <https://doi.org/10.1175/JCLI-D-13-00291.1>.
- Rostkier-Edelstein, D., P. Kunin, T. M. Hopson, Y. Liu, and A. Givati, 2016: Statistical downscaling of seasonal precipitation in Israel. *Int. J. Climatol.*, **36**, 590–606, <https://doi.org/10.1002/joc.4368>.
- Saha, S., and Coauthors, 2010: The NCEP Climate Forecast System Reanalysis. *Bull. Amer. Meteor. Soc.*, **91**, 1015–1058, <https://doi.org/10.1175/2010BAMS3001.1>.
- Sievers, O., K. Fraedrich, and C. C. Raible, 2000: Self-adapting analog ensemble predictions of tropical cyclone tracks. *Wea. Forecasting*, **15**, 623–629, [https://doi.org/10.1175/1520-0434\(2000\)015<0623:SAAEPO>2.0.CO;2](https://doi.org/10.1175/1520-0434(2000)015<0623:SAAEPO>2.0.CO;2).
- Skamarock, W. C., and J. B. Klemp, 2008: A time-split non-hydrostatic atmospheric model for weather research and

- forecasting applications. *J. Comput. Phys.*, **227**, 3465–3485, <https://doi.org/10.1016/j.jcp.2007.01.037>.
- Stauffer, D. R., N. L. Seaman, and F. S. Binkowski, 1991: Use of four-dimensional data assimilation in a limited-area mesoscale model. Part II: Effects of data assimilation within the planetary boundary layer. *Mon. Wea. Rev.*, **119**, 734–754, [https://doi.org/10.1175/1520-0493\(1991\)119<0734:UOFDDA>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<0734:UOFDDA>2.0.CO;2).
- Steiger, J. H., 1980: Tests for comparing elements of a correlation matrix. *Psychol. Bull.*, **87**, 245–251, <https://doi.org/10.1037/0033-2909.87.2.245>.
- Stephenson, D. B., C. S. Coelho, and I. T. Jolliffe, 2008: Two extra components in the Brier score decomposition. *Wea. Forecasting*, **23**, 752–757, <https://doi.org/10.1175/2007WAF2006116.1>.
- 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, https://ams.confex.com/ams/84Annual/techprogram/paper_69061.htm.
- 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, <https://doi.org/10.1175/2008MWR2387.1>.
- Van den Dool, H., 1989: A new look at weather forecasting through analogues. *Mon. Wea. Rev.*, **117**, 2230–2247, [https://doi.org/10.1175/1520-0493\(1989\)117<2230:ANLAWF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<2230:ANLAWF>2.0.CO;2).
- Vanvyve, E., L. Delle Monache, A. J. Monaghan, and J. O. Pinto, 2015: Wind resource estimates with an analog ensemble approach. *Renewable Energy*, **74**, 761–773, <https://doi.org/10.1016/j.renene.2014.08.060>.
- Vislocky, R. L., and G. S. Young, 1989: The use of perfect prog forecasts to improve model output statistics forecasts of precipitation probability. *Wea. Forecasting*, **4**, 202–209, [https://doi.org/10.1175/1520-0434\(1989\)004<0202:TUOPPF>2.0.CO;2](https://doi.org/10.1175/1520-0434(1989)004<0202:TUOPPF>2.0.CO;2).
- Wanik, D. W., E. N. Anagnostou, B. M. Hartman, M. E. B. Frediani, and M. Astitha, 2015: Storm outage modeling for an electric distribution network in northeastern USA. *Nat. Hazards*, **79**, 1359–1384, <https://doi.org/10.1007/s11069-015-1908-2>.
- Wernli, H., M. Paulat, M. Hagen, and C. Frei, 2008: SAL—A novel quality measure for the verification of quantitative precipitation forecasts. *Mon. Wea. Rev.*, **136**, 4470–4487, <https://doi.org/10.1175/2008MWR2415.1>.
- Wilks, D. S., 2006: *Statistical Methods in the Atmospheric Sciences*. 2nd ed. International Geophysics Series, Vol. 100, Academic Press, 648 pp.
- Zhang, D., and R. A. Anthes, 1982: A high-resolution model of the planetary boundary layer—Sensitivity tests and comparisons with SESAME-79 data. *J. Appl. Meteor.*, **21**, 1594–1609, [https://doi.org/10.1175/1520-0450\(1982\)021<1594:AHMOT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1982)021<1594:AHMOT>2.0.CO;2).
- Zhang, J., C. Draxl, T. Hopson, L. Delle Monache, E. Vanvyve, and B.-M. Hodge, 2015: Comparison of numerical weather prediction based deterministic and probabilistic wind resource assessment methods. *Appl. Energy*, **156**, 528–541, <https://doi.org/10.1016/j.apenergy.2015.07.059>.