Evaluation of PBL Parameterizations for Modeling Surface Wind Speed during Storms in the Northeast United States

MARIA E. B. FREDIANI
Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

JOSHUA P. HACKER
National Center for Atmospheric Research, Boulder, Colorado

EMMANOUIL N. ANAGNOSTOU
Department of Civil and Environmental Engineering, University of Connecticut, Storrs, Connecticut

THOMAS HOPSON
National Center for Atmospheric Research, Boulder, Colorado

(Manuscript received 14 October 2015, in final form 25 July 2016)

ABSTRACT

This study identifies conditions that determine errors in numerical simulations of 10-m wind speed over moderately complex terrain, emphasizing winds that lead to overhead power-line damage over a subregion of the northeast United States. Simulations with the Mellor–Yamada–Janjic’ (MYJ) scheme, the Yonsei University (YSU) scheme, and a subgrid-scale topographic drag correction (Topo) applied to YSU are used to investigate error components. The wind speed distribution is dominated by low speeds, which are well depicted by Topo, but are underestimated by the MYJ and YSU schemes. Conversely, moderate and high speeds are underestimated by Topo, and MYJ and YSU perform better across specific ranges. Verification samples are conditioned by season, diurnal cycle, topography, and spatial patterns obtained with a clustering analysis. The systematic error is characterized by a positive bias in low speeds, and as speed increases the biases become more negative. Quantile comparisons, along with systematic and random errors, indicate that beyond the dependence on wind speed itself, errors also depend on seasonal characteristics, indirectly defined by scheme stability profiles. The positive relationship between absolute bias and speed originates in the friction velocity parameterization, and the correction for drag in the Topo scheme exacerbates the effect. The Topo scheme adjusts the total bias and sharpens the bias spread but penalizes moderate and high winds. Clusters reveal that in Topo the bias is primarily driven by wind direction. Excessive correction occurs on terrain-interacting flows, and oceanic flow modulates the adjustment, enhancing the scheme’s performance.

1. Introduction

This study analyzes near-surface (10 m) wind speeds of numerical simulations using criteria that elucidate the specific conditions determining the wind speed and wind component bias for damaging winds. Biases from different planetary boundary layer (PBL) schemes, conditioned on flow regime, topography, season, and diurnal cycle, inform model implementation decisions and forecaster guidance for the damage scale and impact areas. Motivated by a necessity for computationally efficient predictions of potential wind damage to the local electric utility power grid, we evaluate three schemes: two PBL models and a topography-based wind speed correction, available in the Advanced Research version of the Weather Research and Forecasting (ARW) Model.

In the extensive forest territory of the U.S. Northeast (McCaffrey 2006), storms often uproot trees and break branches that fall on power lines, directly affecting the electricity distribution network. The region’s economy
and community are vulnerable to weather-driven blackouts that have lasted several days or weeks, compromising communications, heating, and the water supply (McGee et al. 2012). The Northeast includes the greatest percentage of houses immersed in wildland forests in the country (Radeloff et al. 2005). The power distribution network is predominantly composed of overhead lines, aligned with primary and local roads, frequently found in proximity to forest stand edges susceptible to wind damage. When storms affect the region, even moderate wind speeds are sufficient to damage branches or trees that fall onto the adjacent power lines (Wanik et al. 2015). The greatest cause of damage to the power grid is wind-damaged trees (Maliszewski and Perrings 2012; Connecticut Light and Power 2014). Ultimately, we seek improved surface wind simulations over the northeast United States in damaging wind scenarios and, consequently, improved prediction of tree-induced damage to the electricity distribution network.

Small and short-time-scale features such as wind gusts and flow interactions with topography are important for predicting wind damage. We expect surface wind speed accuracy from numerical weather prediction (NWP) models to improve with high atmospheric and topographic resolution (e.g., Santos-Alamillos et al. 2013), and deterministic forecasts driven by larger analyses and forecasts from operational centers remain a viable alternative. As an immediate and computationally low-cost solution, we seek optimal parameterizations for deterministic NWP forecasts, focusing on refining the regional prediction of wind magnitude and location.

Surface wind, temperature, and mixing ratio simulations are strongly influenced by the choice of surface and boundary layer parameterization (Burk and Thompson 1989; Pleim 2007). The surface-layer scheme and the land surface model (LSM) combine to calculate surface momentum, heat, and moisture fluxes into the atmosphere. These provide lower boundary conditions for PBL parameterizations. The PBL schemes determine vertical mixing from turbulent eddies and thermals, determining boundary layer evolution. Several studies have evaluated model sensitivities to different PBL schemes (Zhang and Zheng 2004; Nolan et al. 2009; Shin and Hong 2011; Santos-Alamillos et al. 2013; Zhang et al. 2013). Among wind speed investigations, Zhang and Zheng (2004) indicated that PBL schemes tend to underestimate wind speeds during the daytime and overestimate them at night. Nolan et al. (2009) found that hurricane intensities were low because of uncapped drag coefficients for high speeds, and the typical neutral boundary layer makes surface winds strongly dependent on the surface roughness. Shin and Hong (2011) found larger biases at night. Zhang et al. (2013) found that error characteristics change with terrain complexity, and emphasize flow-dependent errors under strong synoptic forcing. Santos-Alamillos et al. (2013) showed the importance of terrain representation on surface wind speed and direction.

Errors in surface wind predictions originate primarily in the approximate physics representing the PBL (Arakawa 2004; Pleim 2007; Hu et al. 2010; Nielsen-Gammon et al. 2010) and the LSM. Regions strongly influenced by mesoscale processes and variable land surface features as is the Northeast are prone to large errors. Sea-breeze circulations (Miller and Keim 2003), convection (Murray and Colle 2011), and terrain-forced flows (Wood 2000) are examples of processes that could be misrepresented by choosing inappropriate PBL and LSM parameterizations.

The Northeast’s land surface includes moderately complex features, such as lakes, rivers, low mountains, urban centers immersed into tall vegetation, and irregular coastlines (Colle et al. 2003a,b; Jones et al. 2007). These contribute to unresolved complex topography, impacting NWPs by introducing errors into boundary layer processes and transport (Hanna and Yang 2001; Zhong and Fast 2003; Zhang and Zheng 2004; Hart et al. 2005; Liu et al. 2008; Zhang et al. 2013). Typically, the 10-m wind speed in WRF is too high over plains and too low over unresolved orographic features (Cheng and Steenburgh 2005; Bernardet et al. 2008; Roux et al. 2009).

Jiménez and Dudhia (2012) suggest that the WRF Model’s positive surface wind biases can be attributed to an underrepresented drag coefficient over smoothed topography and developed a parameterization (Topo-wind scheme) that accounts for subgrid-scale topographic variability. The scheme showed successful results when tested over the Iberian Peninsula (Jiménez and Dudhia 2012, 2013; Jiménez et al. 2013; Lorente-Plazas 2014), a region of much higher terrain complexity than the Northeast. The scheme has not been widely applied outside the Iberian Peninsula. Giovannini et al. (2014) used it to validate the accuracy of wind speed simulations over the Italian Alps. Hari Prasad et al. (2015) employed the scheme in dispersion simulations in India. But neither of these two studies validated the parameterization itself. An evaluation was performed by Lee et al. (2014), and they found that the scheme reduced the wind speed bias over flat regions in East Asia. In our study, the scheme is applied to a region with moderate subgrid-scale variability with relatively low hilltops. The regional topography is less complex than those in previous studies, providing a new testing ground.
for the Topo-wind scheme. Within the goals of this study, we investigate whether the wind parameterization can also reduce surface wind speed biases over the Northeast.

The Topo-wind scheme’s enhanced drag is expected to decelerate the wind speed over topographically complex grid cells and reduce the momentum advection across the entire domain. The impact of including the scheme may vary depending on the flow pattern (field speed, direction, and flow characteristics conditioned by orography features). To investigate this hypothesis, we identify the main flow structures by applying a clustering technique to capture the various wind patterns. We investigate whether improvements obtained with the Topo-wind scheme are connected to specific flow patterns, and identify the advantages of using a clustering method to define error components, instead of a more general classification defined by season or region.

Here, we carefully verify and compare the WRF Model performance using the topographic adjustment (Topo experiment), against two other designs: 1) an experiment using the Mellor–Yamada–Janjić PBL and its corresponding Monin–Obukhov-based scheme for the surface layer (MYJ experiment) and 2) an experiment using the Yonsei University PBL, with a Monin–Obukhov-based scheme for the surface layer (YSU experiment); these are the same PBL and surface schemes used in the Topo experiment. Focusing on speeds that can damage trees, we assess error characteristics associated with the terrain, season, diurnal cycle, and flow pattern, with the goal of improving forecast guidance and decision-making.

The study is conducted using simulations of storms that caused significant damage to the electric distribution grid infrastructure of utilities in Connecticut and Massachusetts. The impact of these storms can be inferred by the number of customers with service interrupted. The lower-impact storms affected tens of thousands of customers and the highest-impact systems, such as Tropical Storm Irene and the October 2011 nor’easter, affected 671 000 and 831 000 customers, respectively.

This study is innovative in that it analyzes the schemes (MYJ, YSU, and Topo-wind) under different atmospheric forcing and stability regimes (seasonally and diurnally), under variable topographic influences (spatially), and the main synoptic patterns (clusters). We assess the Topo-wind scheme more extensively than has been done in previous work. The focus on damaging surface wind filters the data period to specific events and contributes to an innovative scheme assessment.

The following section describes the data and the three model configurations. The methods used to verify the experiments and the clustering characterization are presented in section 3. Section 4 contains the analyses of the models according to season, diurnal cycle, spatial location, and clustered wind field. Section 5 summarizes this study and primary conclusions.

2. Data and models

This analysis is developed upon a database of 103 storm simulations between 2001 and 2013. Among them, 27 storms are during spring, 36 are associated with deep convection during summer, 19 are in the fall, and 21 are during winter. The database includes three hurricanes/tropical storms (Hanna, Irene, and Sandy), six nor’easters, and three blizzards (Table 1).

The simulations composing the storm database are generated using ARW version 3.4.1 (Skamarock et al. 2008) to dynamically downscale NCEP’s Global Forecast System (at 6-hourly intervals with 1.0° of resolution). The model is set up with three nested domains using 18-, 6-, and 2-km grid spacing. All domains use 30-arcsecond (~1000 m) terrain as input. A subregion in the inner-most domain provides the modeled atmospheric conditions evaluated in this study (the subregion is illustrated by the red box in Fig. 1).

The WRF Model is configured to use a 30-s time step, two-way feedback loop between nests, as well as 28 default vertical levels. Parameterizations used in all three experiments include Thompson microphysics (Thompson et al. 2008); Grell 3D convection (Grell and Dévényi 2002) (except for the 2-km inner domain); RRTM longwave radiation (Mlawer et al. 1997) computed every 18, 6, and 2 min, respectively, on each domain; Goddard shortwave radiation (Chou and Suarez 1996); and the Unified Noah LSM (Tewari et al. 2004).

The WRF PBL is configured in three ways, which define the experiments evaluated. In the MYJ experiment we use Janjić’s Eta Monin–Obukhov scheme for the surface layer (Janjić 1996, 2002) and the Eta Mellor–Yamada–Janjić approach for the PBL (Janjić 1994). In the YSU experiment and the Topo experiment we use the MM5 similarity for the surface layer (Zhang and Anthes 1982) and the Yonsei University model for the PBL (Hong et al. 2006) above it. In the Topo experiment exclusively, we add the topographic correction for surface wind (Topo-wind) to the YSU experiment to represent data from the subgrid-scale topography (Jiménez and Dudhia 2012). The 10-m wind speed is interpolated from the surface to the lowest model layer using Monin–Obukhov similarity theory.

For each event, the WRF Model is initialized 6 h prior to the time of the first recorded damage (according to local utility records), producing hourly outputs for a
maximum lead time of 60 h to accommodate longer events. An event period is assessed using wind speed observations obtained from 16 METAR stations. The event duration is defined as the period of time when the subregion mean wind speed remains persistently above 2.5 m s\(^{-1}\). An additional filtering, presented in section 3c, was applied to remove fields with very low wind speeds overall. The 16 METAR stations within the analysis domain provide the wind speed observations at 10 m AGL for evaluating the model.\(^1\)

3. Methods

a. Topographic parameterization

The topographic correction is of primary interest because it offers a potentially simple way to improve wind speed predictions in the Northeast. The full description can be found in Jiménez and Dudhia (2012). The scheme distinguishes mountains from plains and valleys and applies a specific adjustment to each. In the region we are analyzing, the scheme classifies all grid cells as valley/plains. So, for the scheme, the mountains in the region are not particularly steep. Consequently, this analysis does not enclose all conditions on which the Topo-wind scheme operates.

The adjustment for valleys and plains consist of a factor \(c_t\) introduced to the surface drag term in the momentum equation integrated by the YSU PBL parameterization.

The surface drag term in the momentum equation is given by

\[
\frac{\partial u}{\partial t} = f + \frac{u^2}{\Delta z} \nabla V, \tag{1}
\]

where \(u\) is the zonal wind component at the first model level, \(V\) is the wind speed, and \(\Delta z\) is the thickness of the lowest model layer. The friction velocity \(u_\text{f}\) is computed as described in Zhang and Anthes (1982):

\[
u_f = \kappa \frac{V}{\ln(z/z_0)} - \psi. \tag{2}
\]

where \(\kappa\) is the Von Kármán constant; \(z\) and \(z_0\) are the roughness length at the layer above ground and at the surface, respectively; and \(\psi\) is the integrated flux-profile relationship.

The coefficient \(c_t\) is the only factor differentiating the YSU and Topo experiments. In the YSU scheme, \(c_t\) is constant and equal to 1. In the Topo-wind scheme, \(c_t\) is related to the orography variability on the subgrid scale. Specifically, this relationship is given by the logarithm of

\[^1\] Observation uncertainty is 0.5 m s\(^{-1}\) for speeds lower or equal to 5 m s\(^{-1}\) and 10% for speeds greater than 5 m s\(^{-1}\) (Jarraud 2008).

### Table 1. Model initialization dates.

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Date</th>
<th>Time</th>
<th>Date</th>
<th>Time</th>
<th>Date</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1200 UTC 5 Mar 2001</td>
<td>1800 UTC 31 May 2006</td>
<td>0000 UTC 22 Jun 2008</td>
<td>0000 UTC 2 Jun 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 UTC 1 Jul 2004</td>
<td>1200 UTC 10 Jul 2006</td>
<td>1800 UTC 22 Jul 2008</td>
<td>1800 UTC 5 Jun 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0000 UTC 20 Aug 2004</td>
<td>1800 UTC 16 Jul 2006</td>
<td>1800 UTC 26 Jul 2008</td>
<td>1200 UTC 5 Jul 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 UTC 8 Sep 2004</td>
<td>1800 UTC 27 Jul 2006</td>
<td>1200 UTC 5 Aug 2008</td>
<td>0000 UTC 21 Jul 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 UTC 17 Sep 2004</td>
<td>0000 UTC 1 Aug 2006</td>
<td>0000 UTC 6 Sep 2008</td>
<td>1800 UTC 29 Sep 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 UTC 25 Dec 2004</td>
<td>1800 UTC 1 Mar 2007</td>
<td>0000 UTC 7 Jan 2009</td>
<td>1800 UTC 1 Feb 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1800 UTC 1 Apr 2005</td>
<td>1800 UTC 14 Apr 2007</td>
<td>0000 UTC 9 May 2009</td>
<td>0000 UTC 1 Jun 2011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0000 UTC 29 Sep 2005</td>
<td>1200 UTC 13 Jan 2008</td>
<td>1800 UTC 27 Nov 2009</td>
<td>0600 UTC 30 Jan 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1800 UTC 7 Oct 2005</td>
<td>1200 UTC 12 Feb 2008</td>
<td>1200 UTC 23 Feb 2010</td>
<td>0000 UTC 8 Feb 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1800 UTC 15 Oct 2005</td>
<td>0600 UTC 8 Mar 2008</td>
<td>0000 UTC 13 Mar 2010</td>
<td>0000 UTC 11 May 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1200 UTC 2 Jan 2006</td>
<td>1200 UTC 26 May 2008</td>
<td>1200 UTC 3 May 2010</td>
<td>0600 UTC 25 May 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0000 UTC 14 Jan 2006</td>
<td>0000 UTC 8 Jun 2008</td>
<td>0000 UTC 4 May 2010</td>
<td>0000 UTC 29 May 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0000 UTC 17 Jan 2006</td>
<td>0000 UTC 10 Jun 2008</td>
<td>1200 UTC 7 May 2010</td>
<td>0600 UTC 13 Jun 2013</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1800 UTC 16 Feb 2006</td>
<td>1800 UTC 13 Jun 2008</td>
<td>1800 UTC 25 May 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
the standard deviation of the subgrid-scale orography \( \sigma_{\text{orog}} \), \( c_t = \ln(\sigma_{\text{orog}}) \). The subgrid-scale orography (and consequently \( \sigma_{\text{orog}} \)) depends solely on the computational grid specification and varies by grid cell.

In complex terrain, high biases originate in grid cells with unresolved topography, which underrepresent the subgrid surface roughness. The Topo-wind correction scales the acceleration from drag, and reduces the biases in these specific grid cells (\( c_t \) cannot be less than 1).

b. Assessment approach

We verify the experiment results based on season, diurnal cycle, terrain, and through the use of a two-step clustering technique for grouping the most similar wind fields. We apply a set of statistical error metrics based on comparisons of model outputs to surface observations from METAR stations. The location of the stations and the standard deviation of orography representing the terrain variability in the subdomain are shown in Fig. 1.

We examine how the model performs against observations by looking into the wind speed frequency distributions, biases, random errors, and distribution quantiles. The bias is given by the mean difference of the forecast from the observations. The random error is given by the standard deviation of the difference between the forecast and the observations. These two quantities are the components of the mean square error (MSE). We use the MSE components separately to isolate systematic errors from errors related to differences in the model structure, phase, and/or amplitude of the variations. The latter can be treated as the random error, even if it includes errors at a range of temporal and spatial scales, because it is distinct from the mean error that estimates the bias. The MSE is represented by two components as

\[
\text{MSE} = \text{Bias}^2 + \sigma_{f-o}^2. \tag{3}
\]

Bias and \( \sigma_{f-o} \) are given by the following:

\[
\text{Bias} = \frac{1}{n} \sum_{i=1}^{n} (f_i - o_i) = \bar{f} - \bar{o} \quad \text{and} \tag{4}
\]

\[
\sigma_{f-o}^2 = \frac{1}{n} \sum_{i=1}^{n} [(f_i - \bar{f}) - (o_i - \bar{o})]^2. \tag{5}
\]

This simplifies to

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (f_i - o_i)^2. \tag{6}
\]

The simulated wind speed at each observation site is obtained using bilinear interpolation considering the four closest grid points. The interpolation is used to reduce large differences caused by topography representativeness errors in the model. The sample size \( n \) is given by the number of observations from 16 surface stations matching the hourly model outputs in the set of simulations.

The quantification of errors is based on comparisons between experiments and observations using the 16 METAR stations and corresponding interpolated model grid cells. Direct comparisons among experiments use all grid cells in the evaluation subregion. Throughout the investigation, we look for biases associated with the season, diurnal cycle, topography, and flow patterns.

c. Two-step clustering

The two-step clustering analysis was devised to find common attributes in the wind fields without any a priori classification, such as storm type. Through this analysis we expect to find particular circumstances of atmosphere–surface interaction that may transcend a spatial, seasonal, or diurnal classification.

Storm wind fields are clustered according to hourly wind components (\( u \) and \( v \)). Each hourly field is an independent element in the clustering dataset, implying that one coherent storm event can have hourly fields distributed in different clusters. Taking hourly wind fields as independent elements allows us to cluster the wind patterns interacting directly with the surface layer,
regardless of its triggering mechanisms or its temporal continuity. Subsequent analysis is based on observations valid at times matching the hourly fields assigned to each cluster.

Previous studies applied clustering analysis to wind fields with different goals. Davis and Walker (1992) and Weber and Kaufmann (1995) applied hierarchical clustering techniques to find the initial number of clusters and use them as seeds in a subsequent clustering analysis. Weber and Kaufmann (1995) compared several hierarchical algorithms for a wind field classification finding advantages in the complete linkage algorithm. Because the method does not vary under monotonic transformations, they found it appropriate for identifying wind patterns. In a following study, Kaufmann and Weber (1996) applied the two-step cluster analysis using the complete linkage as the first step and $k$-means as the second step.

In all these studies the classification was based on observations. This study is different, in that we perform the two-step classification based on the MYJ simulation wind fields. The model fields are used instead of observations because it allows the inclusion of all grid cells in the domain, resulting in clusters that best depict spatial patterns. We follow the method described by Kaufmann and Whiteman (1999).

In the two-step method, the first step determines the adequate number of clusters using hierarchical clustering, and the second step assigns elements (wind fields) to the clusters according to their similarities. The $u$- and $v$-wind components provide the Euclidean distance, the dissimilarity metric used to build the clusters.

To filter very low-speed wind fields and assure we maintain localized gusts, only wind fields (from the MYJ experiment) of mean speed above 5 m s$^{-1}$, or with at least five grid points of speed greater than 8 m s$^{-1}$, are kept in the analysis, resulting in 3517 valid elements in the database (elements correspond to simulated wind fields). The complete linkage hierarchical clustering method (first step) was applied, as described in Kaufmann and Whiteman (1999) and indicates 18 is an adequate number of clusters to classify the wind fields. The centers of each of the 18 clusters obtained with the complete linkage algorithm are used as the starting point for the next step. In the second step we apply the $k$-means method, commonly used for unsupervised data clustering. The $k$ means starts with $k$ seeds ($k = 18$) corresponding to the centroids of the clusters; we then group the data by minimizing the intracluster variability and maximizing the intercluster variability.

d. Cluster characteristics

The resulting 18 clusters (Fig. 2) are subjectively binned into three groups according to cluster-center wind speeds. Table 2 and Fig. 3 summarize the clusters within three ranges: low speed (2–5 m s$^{-1}$), intermediate or moderate speed (5–8.5 m s$^{-1}$), and high speed (8.5–12 m s$^{-1}$). Clusters 0–4 (in increasing center speed) are in the low-speed group. Clusters 5–10 are in the intermediate group, and clusters 11–17 are in the high-speed group. Table 2 shows the number of elements in each cluster, along with the cluster-center wind speed and direction. Figure 3 presents the count of wind fields according to month and local time.

Clusters with center wind speed below 5 m s$^{-1}$ include mostly storms in May–July concentrated during the afternoon (after 1400 LST) and early evening. These characteristics dominate clusters 1 (southwesterly winds), 2 (northwesterly winds), and 4 (southerly winds). During May–July we expect mostly convective storms that under normal conditions initiate early in the day, feed on excessive heat and moisture throughout the afternoon, and dissipate at sunset if no other source of energy is available. Given that these fields are mostly convectively driven, the low wind speed can be explained by one or more factors: 1) insufficient potential energy to develop

2 In Kaufmann and Whiteman (1999) the wind components are normalized spatially and temporally. In this study, to preserve variability, we do not normalize the fields.

3 These speed ranges are also used as qualifiers throughout the analysis.
into severe storms; 2) lack of kinetic energy to maintain the storms, that is, vertical development occurring in conditions of low wind shear; and/or 3) decaying stage of storm development.

Clusters with center wind speeds between 5 and 8.5 m s\(^{-1}\) are characterized by fields in two distinct periods: during winter and between May and July. Most of the winter fields occur in January and February. Because these are larger-scale storms, elements within the same 24-h period tend to be in the same cluster. These winter characteristics are found in clusters 5 and 6 (both northeasterly winds) and 10 (northwesterly winds).

During the May–July period the fields accompany the development of the boundary layer, between 9 and 15 h. Such characteristics dominate clusters 8 and 9 (southwesterly and south winds). Unlike the low speed group, these May–July fields use the heat available for convection at times when it is most intense, developing stronger pressure gradients at the surface and producing stronger bursts from downdrafts.

Clusters with speeds above 8.5 m s\(^{-1}\) contain mostly the September–December elements, with an additional set of elements from March and a smaller set from June to July between 14 and 19 h. Clusters 12 (northerly winds), 14 (westerly winds), 15 (northeasterly winds), and 16 (northeasterly winds) contain elements from March and the fall season. Similar to the winter elements in the intermediate group, 24-h periods tend to fall in the same cluster. Again the mesoscale (or larger) dynamics dominate over the surface–atmosphere heat exchange and other diurnal characteristics.

4. Results

a. Bulk statistics

The distribution of wind speeds from the storms used in this study (Fig. 4) shows that among the three parameterizations, the Topo experiment most accurately represents the distribution median, located at the lower speed range, and its associated high frequency. However, it cannot reproduce the high-wind speed frequencies. According to observations, the most frequent speed is \(~3.5\) m s\(^{-1}\), agreeing with results from Topo-wind simulations. In the moderate and high-wind speed range the Topo-wind scheme leads to frequencies lower than observed.

The MYJ and YSU experiments have shorter and wider distributions, with the median shifted toward higher wind speeds (6 and 4.5 m s\(^{-1}\), respectively). The

### Table 2. Results from the two-step cluster analysis based on MYJ experiment wind fields grouped into three speed categories: low, medium, and high. Shown are the cluster number, the number of elements (hourly wind fields), the cluster-center wind speed, and the center wind direction.

<table>
<thead>
<tr>
<th>Cluster No.</th>
<th>No. of elements</th>
<th>Center speed (m s(^{-1}))</th>
<th>Center direction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>163</td>
<td>2.1</td>
<td>SE</td>
</tr>
<tr>
<td>1</td>
<td>376</td>
<td>3.6</td>
<td>SW</td>
</tr>
<tr>
<td>2</td>
<td>496</td>
<td>4.3</td>
<td>NW</td>
</tr>
<tr>
<td>3</td>
<td>159</td>
<td>4.8</td>
<td>SE</td>
</tr>
<tr>
<td>4</td>
<td>390</td>
<td>4.9</td>
<td>S</td>
</tr>
<tr>
<td>Medium speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>248</td>
<td>5.4</td>
<td>NE</td>
</tr>
<tr>
<td>6</td>
<td>193</td>
<td>6.6</td>
<td>NE</td>
</tr>
<tr>
<td>7</td>
<td>75</td>
<td>7.3</td>
<td>E</td>
</tr>
<tr>
<td>8</td>
<td>131</td>
<td>7.6</td>
<td>SW</td>
</tr>
<tr>
<td>9</td>
<td>262</td>
<td>8.0</td>
<td>S</td>
</tr>
<tr>
<td>10</td>
<td>189</td>
<td>8.4</td>
<td>NW</td>
</tr>
<tr>
<td>High speed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>107</td>
<td>8.5</td>
<td>S</td>
</tr>
<tr>
<td>12</td>
<td>121</td>
<td>9.2</td>
<td>N</td>
</tr>
<tr>
<td>13</td>
<td>118</td>
<td>9.2</td>
<td>W</td>
</tr>
<tr>
<td>14</td>
<td>246</td>
<td>9.5</td>
<td>W</td>
</tr>
<tr>
<td>15</td>
<td>136</td>
<td>10.0</td>
<td>NE</td>
</tr>
<tr>
<td>16</td>
<td>81</td>
<td>11.4</td>
<td>E</td>
</tr>
<tr>
<td>17</td>
<td>26</td>
<td>11.8</td>
<td>SE</td>
</tr>
</tbody>
</table>

FIG. 3. Summary of the two-step cluster analysis based on the MYJ experiment wind fields. Clusters are grouped into three speed categories determined by the cluster-center speed: (left) low, (center) medium, and (right) high. Shown are the counts of wind fields contained in each speed category according to month and local time. Note that the count scale changes between panels. The low-speed range contains clusters 0–4 (in increasing center speed order), the medium-speed range clusters 5–10, and the high-speed range clusters 11–17.
low- to moderate-speed frequencies are largely underestimated, and high-speed frequencies are overestimated. The distributions obtained in the Topo and YSU experiments agree with the results of Lorente-Plazas (2014), who show that the Topo-wind scheme enhances the low-speed distribution mode and reduces the high-speed frequency given by the YSU scheme.

To measure the accuracy of the general speed magnitude in the experiments, we divide the observed speed range into 1 m s\(^{-1}\) intervals and calculate the bias of each experiment for that interval. As a result of reduced sample sizes, we restrict the bias and error calculations to speeds below 15 m s\(^{-1}\) (Figs. 5 and 6).

All model configurations are positively biased at wind speeds lower than 3 m s\(^{-1}\) (Fig. 5). The lower biases occur at low wind speeds, and the Topo experiment (also indicated by Fig. 4) shows the lowest bias and best matches the observed distribution. The low-speed range is the most frequent observed speed, so it practically controls the experiment total bias. During high-wind intervals the sample size is considerably smaller, and the contribution to the total bias is smaller than the contribution from the weaker winds. Consequently, the Topo experiment total bias appears to be the smallest among the three experiments, when in fact it hides large negative biases that result at higher wind speeds.

A similar relationship between the bias and wind speed is also shown in Jiménez and Dudhia (2012) for YSU. Their simulations over the Iberian Peninsula also show a positive bias when speeds are low, as well as a diminishing bias value, resulting in negative biases at greater wind speeds.

Comparing the bias versus wind speed in the YSU and MYJ experiments shows similar downward slopes, leading to negative biases at wind speeds greater than approximately 8 m s\(^{-1}\) (for YSU) and 10 m s\(^{-1}\) (for MYJ). At greater wind speeds the drag formulation based on Monin–Obukhov similarity theory is invalid. Although other factors may be relevant, as described below, this drag formulation results in a drag that is too strong at large wind speeds.

At greater wind speeds, the drag adjustment from the Topo-wind scheme exacerbates the negative biases already resulting from the YSU and MYJ schemes. At moderate wind speeds the Topo scheme appears as an overcorrection, making the positive YSU and MYJ biases negative. Consequently, both the MYJ and YSU provide more accurate winds than the Topo experiment in intervals above 7 m s\(^{-1}\).

The biases indicate that the Topo-wind scheme correction works well when wind speeds are low, but the drag adjustment is excessive for higher wind speeds. The Topo coefficient \(c_t\) on the drag term in the momentum equation [Eq. (1)] depends only on the subgrid-scale orographic variability. The drag also depends on the friction velocity \(u_*,\) which is parameterized proportional to the wind speed \(V\) by Eq. (2). Assuming all other variables are constant, the drag is then proportional to the product of \(V\) and the individual wind component (i.e., \(V^2\) in a direction aligned with the mean wind). From the bias results, it appears that a superlinear drag is too strong at greater speeds. When \(\sigma_{\text{orog}} > e (c_t > 1),\) the drag is an even steeper function of wind speed. The effect can be seen in the steeper slope of the Topo bias curve in Fig. 5, compared to the YSU and MYJ.

The random error component (Fig. 5) of the MSE (hereafter standard error) increases with speed in the three models, and it is persistently smaller in the Topo experiment. The lower standard error from the Topo experiment may result directly from the slower speeds in the Topo-wind scheme. That is, the amplitude of the wind speed variability is related to the speed magnitude, as it is also shown by the increasing standard error with increasing speed in the three experiments. Hence, the lower speeds in the Topo experiment lead to lower overall standard error from the Topo-wind scheme.

b. Seasonal analysis

Analyzing the results by season indicates that biases have ties to both the synoptic and mesoscale dynamics. During the summer, damaging winds result from local convective storms, and the overall performance of the models depends more on the land–atmosphere interaction than on the large-scale circulation. Gusty winds are localized, and the average wind over the
domain is weak. During winter, the large-scale flow controls the surface wind, which is considerably stronger than in summer.

Figure 6 shows a panel with quantiles (top row), biases (middle row), and standard errors (bottom row) for each season. Biases and standard errors are binned by speed intervals, and biases are accompanied by the wind speed distribution on the secondary y axis. Quantiles of the observed distribution are plotted against the equivalent simulated quantiles (q–q plots). Each circle in the q–q plots represents a 10% mark in the model and observations distribution, going from 10% to 100%. Identical distributions should have the quantiles on a 45\(^\circ\) diagonal. The bias plots only show speeds below 15 m s\(^{-1}\) due to their small sample size, but higher speeds are included in the q–q plots. In quantile analysis, observations and model simulations are not matched in space and time. Temporal and spatial shifts among distributions do not penalize the q–q comparisons as they penalize biases, so for quantile analysis we include all the speeds.

The q–q plots show the most noticeable differences among the three experiments in summer. The quantile distribution in the Topo experiment is more accurate in summer and poorer in spring. The bias curves indicate that none of the experiments show a seasonal pattern of behavior that is different from the average biases shown in section 4a. The magnitude and the rate over which the biases increase (absolute values) with speed vary slightly across the seasons. In spring the absolute bias rate shows the sharpest increase with speed. In winter, the MYJ and YSU experiment biases are the most similar.

During summer, the observation probability distribution in the bias chart (middle and bottom row, solid black line) shows a narrow distribution dominated by low-wind speeds (\(-3\) m s\(^{-1}\)), with minimal occurrence of speeds above 8 m s\(^{-1}\), indicating the lowest frequency of the moderate and high speeds among seasons. The Topo experiment is the best model fitting the observation quantiles. Lorente-Plazas (2014) also reported the most noticeable improvements of the Topo-wind scheme during summer. The MYJ and YSU experiments present the most overestimated model distributions (on observed speeds up to \(-17\) m s\(^{-1}\)), suggesting that corrections concerning surface interactions for typical summer storms are more important during this season. In the 90th percentile, the experiments change subtly, as a result of wind fields in the sample from tropical storms (mostly Tropical Storms Irene and Hanna). At this scale, storms are less influenced by subgrid variability, and part of the highest quantile in the summer distribution is best represented by the YSU scheme. The low observed frequency of high-wind speeds and their greater variability related to the presence of tropical storms in the sample result in wider confidence intervals in all experiment biases.

In the fall, the observed probability distribution is wider, with a higher median wind speed (\(-4\) m s\(^{-1}\)), and the Topo experiment quantiles are overall the most consistent with observations, including the high-speed quantile. The slight accentuation of the quantile curve below the 45\(^\circ\) line in summer and fall shows that underestimation of the wind speed distribution exists, and yet the Topo experiment distributions match best...
the observations compared with the MYJ and YSU experiments.

The Topo experiment underestimates most of the spring and winter quantiles. On average, the YSU scheme is the best choice for both, in that its quantiles show the best match to the observations. In spring the biases are the sharpest, indicating that all models have deficiencies when simulating high-speed storms in the season.

The seasonal analysis shows that errors depend more on the meteorology than on the seasonal distributions of the observed wind speed. That is, the distributions of the observed wind speeds in winter and fall are most similar, and the summer distribution resembles most that of the spring. In turn, the q–q plots indicate that the errors in fall are most similar to those in summer, and the errors in spring are more similar to those in winter. These correspondences indicate skill that most reflects the different meteorological conditions driving events across seasons, rather than skill depending on wind speed alone.

The seasonal variations demonstrate the effects of stability on the drag term. The MYJ scheme performs better in the high-speed quantiles during winter and spring than it does at other times. The MYJ scheme performance is related to how the atmospheric stability is parameterized in more stable or nearly neutral conditions. According to Hu et al. (2010), the practical difference between the YSU and MYJ schemes is mainly in the vertical mixing strength and entrainment of air above the PBL. The YSU scheme shows stronger vertical mixing being more appropriate in convective boundary layer profiles. Local schemes such as the MYJ show less entrainment, and a cooler, moister, and shallower PBL (Dudhia 2014). In stable conditions (e.g., during winter), local closure schemes have weaker
vertical mixing, and overall perform better than the nonlocal options (Shin and Hong 2011; Hu et al. 2013). Then, in the MYJ, \( \nu_a \), and therefore the drag, are smaller than in the YSU.

At low wind speeds, the systematic error is the largest error component in the MYJ and YSU experiments. As the biases shift from positive to negative with increasing speed, the standard error (Fig. 6, bottom row) begins to grow and becomes larger than the bias. In the Topo experiment, even though the random component also increases with wind speed, the bias is the dominant component in all seasons.

The standard error also indicates that the relationship of random errors increasing with speed, described in section 4a, is partially due to seasonal differences. Errors increasing with speed are more pronounced in summer and fall, which are the seasons when the model quantiles in the Topo experiment best match the observations. Wider variability amplitudes, as well as frequent spatial and phase shifts, are expected in unstable atmospheres favoring thermals and localized convective cells. In spring and winter, when the weather is mostly controlled by large-scale features, the flow over flat areas is less perturbed, and systems develop over a longer temporal scale.

c. Diurnal cycle analysis

The surface heat exchange process is a function of the available energy, which in turn depends on the time of the day. Potential underestimation of the frictional stress has more impact when the heat transfer between the surface and atmosphere is greatest, because the effects of the surface momentum sink are more efficiently transported upward via convective turbulence. The two combinations of surface layer/PBL schemes model different boundary layer profiles, which are reflected in the bias time series (Fig. 7).

The diurnal range of bias for any individual model implementation can be large. The MYJ in particular shows greater diurnal variability than either the YSU or the Topo-wind schemes in the spring and winter (~2 and \( \sim 1.2 \text{ m s}^{-1} \), respectively). The MYJ experiment during spring and winter exhibits a sharp bias increase associated with the development of the boundary layer in the early morning, indicating that the scheme underestimates the vertical mixing (Hu et al. 2010). Because in the winter the mixing within the PBL is weaker, the bias peaks at 1000 LST, and the diurnal amplitude is shallower than in springtime. During the spring it increases until 1500 LST, when the near-ground temperature is at its maximum and abundant energy is available for turbulence. As night approaches and the atmosphere becomes more stable, the bias decreases sharply.

reduced biases in the MYJ experiment late at night in spring indicate that the overestimation is minimized in a stable atmosphere when the heating exchange decreases. In summer and fall, the MYJ experiment bias is characterized by a diurnal evolution of lower amplitude, with a persistent high bias throughout the daily cycle.

In the YSU experiment the subtle day–night contrast indicates that the experiment better captures the PBL evolution, and performs more evenly throughout the 24-h daily cycle. In spring and summer, the YSU experiment shows an increase in bias after 1600 LST, which is opposite to the decrease in the MYJ experiment. As discussed above, the MYJ is expected to perform well in more stable higher-wind environments.

At night the increased drag from the Topo-wind scheme brings the biases in YSU closer to zero. This is more evident in the summer, but also appears in spring and fall. Because of the stable environment, the Topo adjustment to the drag is slightly smaller during the night than during the day. During the day, heating at the surface increases \( \nu_a \) and therefore the drag, resulting in a large negative wind speed bias that is most obvious during summer but that is also clear in spring. Overall, during the day (0800–1600 LST), the YSU and Topo experiments have comparable bias magnitudes (of opposite sign), between 1 and 1.5 m s\(^{-1}\) (~0.5 and 1 m s\(^{-1}\), respectively).

During winter, the YSU and Topo experiment biases show less diurnal variability than during spring and summer. The speeds in these seasons are overall higher, so the drag is amplified by the Topo-wind scheme adjustment coefficient. Even though the MYJ experiment shows lower biases during winter compared to the other seasons, it still shows greater absolute biases than the YSU experiments.

d. Spatial analysis

In this section we analyze the bias and standard error for individual meteorological stations, and how each behaves according to terrain variability. For each meteorological station, the \( \sigma_{orog} \) is taken from the closest grid-cell center point (Fig. 1). The domain includes 16 surface meteorological stations. Nine are located in a region close to the coast and are influenced by the ocean (stations 2, 4, 6, 7, 8, 9, 10, 13, and 15) from which three of these are also in a grid cell with greater terrain complexity (stations 2, 10, and 13). Four stations are in a region with greater terrain complexity that is distant from the coastal and valley influence (5, 11, 12, and 14); three are in the Connecticut River valley (0, 1, and 3). Jiménez and Dudhia (2012) point to a relationship between wind speed biases and orography variability, and to reduce the bias, they include in the Topo-wind scheme a
correction for smoothed topography effects. In this section we analyze spatial biases related to higher $\sigma_{\text{orog}}$.

Figure 8 is arranged with increasing bias in the YSU experiment (the PBL scheme that runs with the Topo-wind scheme), and $\sigma_{\text{orog}}$ is shown by the black curve on the secondary y axis. There is no indication that $\sigma_{\text{orog}}$ controls the bias at individual stations, but the number of stations is not sufficient to state that a trend does not exist. Yet, Fig. 8 shows additional factors leading to intrastation variability that surpasses the effect of the subgrid topography.

Although there is no indication of an explicit effect from $\sigma_{\text{orog}}$, there is evidence of a geographical influence. The majority of the stations with low bias in the YSU experiment are in the river valley (diamonds; stations 0, 1, and 3) and have low $\sigma_{\text{orog}}$. The Topo-wind parameterization leads to an absolute bias higher than YSU at stations 0, 1, 3, and 12. These four stations show the lowest biases in YSU, and the use of the Topo-wind scheme overcorrects the biases at these locations.

In the Topo experiment, all stations categorized as coastal-influenced (circle) show the smallest bias among the experiments (stations 2, 4, 6, 7, 8, 9, 10, 13, and 15), as well as three out of the four complex terrain stations (stations 5, 11, and 14). The Topo-wind scheme has no effect over ocean, but it shows improvements on circulations that may enter the continent along the coast. Urban development and forest patches are features present near the coast that can enhance drag, especially during flows entering the continent and facing abrupt increases in roughness. With this analysis it is not possible to identify the flow pattern and verify the hypothesis, but in the cluster analysis we show that lower absolute biases are associated with oceanic wind flow in the Topo experiment.

The relationship between biases and $\sigma_{\text{orog}}$ given by Jiménez and Dudhia (2012) is not evident; yet, other indications of a geographic relationship are present and noticeable in the spatial correlations among the experiments. Additional controls on the bias might include the model characterization of vegetation and land use (roughness length), the soil type, or observation siting. An analysis of other features such as high-resolution vegetation cover and terrain height did not reveal any immediate insights.

e. Performance by cluster

The clusters reveal aspects not apparent when stratifying the data spatially, by season, or by diurnal cycle. In the same season, the winds can be driven by a number of mesoscale processes and varying microscale conditions in the surface layer. To understand which

![Fig. 7. Biases in the diurnal cycle for different seasons in the three model experiments (MYJ, red; Topo, blue; and YSU, green). The x axis marks the local time (LST), and the secondary y axis indicates the sample size (solid black line). Confidence intervals reflect a significance level of 0.05.]

![Fig. 8. Biases at the surface meteorological stations displayed with the YSU experiment’s (green) increasing bias. The stations are identified by the numbers 0–15, which refer to stations shown in Fig. 1. The symbols indicate the topography in the model grid cell corresponding to the station location: triangles represent terrain of greater complexity, diamonds represent the river valley, and circles represent the coastal area. The colors indicate the three model experiments (MYJ, red; Topo, blue; and YSU, green). The standard deviation of the subgrid-scale orography is displayed (solid black line) along the secondary y axis.]

microscale characteristics are most influential to the simulation errors, we apply the clustering technique (section 3c) to classify data into similar wind field groups. Each cluster can contain a range of seasons and times of day. Even though the storms strongly depend on the season, this analysis shows that direction, and consequently terrain interaction, have important effects.

Ranked according to the Topo experiment biases, the biases computed from samples defined by individual clusters illustrate a lack of correlation between Topo and both the YSU and MYJ experiments. (Fig. 9). The MYJ and YSU experiment biases show a strong relationship with the cluster elements mean speed, consistent with the clusters definition based on the MYJ experiment database. This correlation between bias and cluster speed is known and has been discussed in section 4a. Yet, when clustering separates wind fields by direction and speed, the biases from the Topo experiment do not show the same relationship. In the Topo experiment the biases are not tied to the cluster-center speed. The Topo experiment bias presents a relationship between the cluster’s $u$ and $v$ components. The lowest absolute biases, in increasing order, occur in clusters 4, 11, 9, 7, 3, 1, 8, and 17. The $u$ and $v$ components of these clusters (Fig. 2) suggest that the Topo experiment performance is best when $v$ is positive (clusters 4, 11, 9, 7, and 3 all have positive $v$). In the local geography, a positive $v$ component corresponds to flow entering along the coast. Conversely, its largest biases occur in clusters 14, 10, 15, 13, 5, 12, 2, 16, 0, and 6, all with a negative $v$ component (except for cluster 13). Clusters 10 and 14 have the highest absolute bias, both characterized by northwesterly high-speed winds. Along the northwesterly trajectory, the winds interact with the highest and most complex terrain in the model’s inner domain.

These errors indicate that areas of more complex terrain have poorer simulated winds than smoother topography areas. This evidence agrees with results shown by Jiménez et al. (2013), who refer to errors in direction angle, whereas we use speed components to infer the direction influencing the errors in speed.

Winds with a southerly component have more interactions with the ocean and face less friction. The Topo-wind scheme does not modulate the wind over ocean, but the upwind interactions over land in cyclonic circulations, and the sudden increased roughness along Long Island peninsula and at the parallel urbanized coastline, reflect on the fields entering the region from the ocean. The favorable lower biases from southerly winds are not present in the MYJ and YSU experiments, so the accuracy in positive $v$ flows does not originate in the boundary conditions or the PBL scheme.

The clear difference between the Topo experiment biases when the flow interacts more intensively with terrain (negative $v$) than when adjustments propagate from less complex areas (positive $v$) is evidence that terrain variability influences the wind speed bias, but the drag correction applied requires modulation. The flow over ocean prevents excessive drag adjustments, limiting the correction to the upwind speeds over land and Long Island, resulting in a balanced amount of drag between the enhanced Topo-wind scheme drag and the default model drag. In turn, the negative $v$ flow carried over the continent removes momentum excessively, resulting in higher absolute biases.

The increasing absolute bias with increasing speed (section 4a) is not different in the clustering analysis. The total bias is controlled by the bias on speeds corresponding to the distribution median, and biases increase wind speeds at a superlinear rate. However, the spread of cluster biases from the Topo experiment is narrower than the bias spread from YSU (Fig. 10). In both positive and negative $v$ components the Topo-wind scheme sharpens the systematic error distribution and reduces the total bias. When the Topo-wind scheme adjustment is modulated by the ocean, it produces wind speed distributions matching low- and high-speed observations. The bias is still present at high speeds, once the issue does not depend on how the dataset is analyzed. The errors remain until the parameterization for $u_a$ is improved and the drag correction in the Topo-wind scheme is revised.

The standard error (Fig. 9) shows lower errors and narrower confidence intervals at moderate and high speeds than the seasonal classification (Fig. 6). The k-means clustering method groups the data elements by minimizing the intracluster variability. The result is that elements are grouped by speed and flow direction, which reduces the variability and, consequently, the standard deviation of the errors and confidence intervals. Figure 9 shows the clusters ranked by mean cluster speed. The trend indicates a slight increase in error with increasing speed, but the relationship is not as evident as in Figs. 4 and 6.

5. Conclusions

Three experiments were designed to evaluate and identify environmental drivers of surface wind speed
errors related to PBL parameterization in moderately complex terrain. A database with storm simulations was compared to surface station observations using multiple classification criteria to investigate conditional model biases. The events were categorized by season and diurnal cycle, as well as spatially, and then clustered using a two-step cluster analysis to identify spatial patterns associated with speed and direction.

The Topo-wind scheme, which adds a drag term to account for subgrid-scale orography, provides the most accurate wind speed distribution in that it best represents the speed and frequency of the observed distribution median. When the analysis is based solely on experiment total biases, low wind speeds dominate the sample, and the accuracy levels of the different model configurations at moderate and high wind speeds are not clear. Consequently, the scheme with lower bias in the speed intervals enclosing the distribution median may erroneously be interpreted as being the best scheme. It is possible that averaging biases hide such effects in Jiménez and Dudhia (2012) and Lee et al. (2014).

Looking at frequency distributions clarifies that the Topo experiment underpredicts the observed frequencies of higher speeds. Conversely, in the MYJ and YSU experiments the occurrence of low and moderate wind speeds is largely underestimated, and the frequencies of higher speeds are overestimated.

Biases binned by observation speed intervals also reveal that the Topo-wind scheme accurately adjusts low wind speeds, but when the same correction is applied to higher speeds, negative biases result and intensify with increasing speed. The MYJ and YSU experiments provide the best simulations at speeds above 7 m s⁻¹. Biases also indicate underestimation of speeds above 10 m s⁻¹ in all configurations.

The negative bias in all experiments at high wind speeds is caused by a superlinear relationship between drag and wind speed, given through the parameterization of friction velocity. The result is a downward bias–speed slope from positive toward large negative biases. When the adjustment coefficient from the Topo-wind scheme is applied to the drag term, it reduces the low-wind-speed bias, but emphasizes the downward slope. Consequently, biases in the Topo experiment become more negative more rapidly than biases in the YSU and MYJ experiments. The practical result is that the Topo-wind scheme performs poorly when simulating high-speed events.

Systematic and random errors can be controlled by large-scale and mesoscale dynamics. The Topo experiment underpredicts the observed occurrences of higher speeds. Conversely, in the MYJ and YSU experiments the occurrence of low and moderate wind speeds is largely underestimated, and the frequencies of higher speeds are overestimated.

Biases binned by observation speed intervals also reveal that the Topo-wind scheme accurately adjusts low wind speeds, but when the same correction is applied to higher speeds, negative biases result and intensify with increasing speed. The MYJ and YSU experiments provide the best simulations at speeds above 7 m s⁻¹. Biases also indicate underestimation of speeds above 10 m s⁻¹ in all configurations.

The negative bias in all experiments at high wind speeds is caused by a superlinear relationship between drag and wind speed, given through the parameterization of friction velocity. The result is a downward bias–speed slope from positive toward large negative biases. When the adjustment coefficient from the Topo-wind scheme is applied to the drag term, it reduces the low-wind-speed bias, but emphasizes the downward slope. Consequently, biases in the Topo experiment become more negative more rapidly than biases in the YSU and MYJ experiments. The practical result is that the Topo-wind scheme performs poorly when simulating high-speed events.

Systematic and random errors can be controlled by large-scale and mesoscale dynamics. The Topo experiment underpredicts the observed occurrences of higher speeds. Conversely, in the MYJ and YSU experiments the occurrence of low and moderate wind speeds is largely underestimated, and the frequencies of higher speeds are overestimated.
Differences between MYJ and YSU experiments, mainly in spring and winter, reflect the stability arising from the different parameterizations. The YSU scheme accounts for countergradients and entrainment at the PBL top, which translates into stronger vertical mixing (Hu et al. 2010; Zhang and Zheng 2004). The known weak vertical mixing in the MYJ scheme is shown by the increasing bias during the PBL diurnal evolution. Conversely, the YSU scheme bias decreases as the PBL develops. In a more stable boundary layer (during the nighttime), the YSU and MYJ schemes perform similarly. The weaker winds at night during the spring, summer, and fall are accurately adjusted by the Topo-wind scheme, but the stronger winds in the convective boundary layer during the day are excessively reduced. That is, the effect of the constant subgrid-scale adjustment in the increasing wind speed is also represented in the diurnal cycle. In winter the atmosphere is typically stable, systems are mesoscale and larger, and most importantly the wind speeds are higher than during other seasons. In such conditions the excessive Topo-wind-adjusted drag affects the entire daily cycle.

The random component of the MSE (standard deviation of the error, referred to as the standard error) shows that the error variability is related to the wind speed magnitude, increasing with increasing speed and overestimation in any particular scheme. The random error component also depends on the season and associated storm types. The summer, characterized by greater convective activity, elicits greater random errors from errors in the intensity and location of the convective storms.

A dependence of bias on station location indicates that there is a geographical influence on the biases found in all of the experiments. However, the limited number of surface stations is not sufficient to verify the relationship between the wind speed bias and the orography variability shown by Jiménez and Dudhia (2012). Nevertheless, the Topo experiment improves all nine stations in the region influenced by the ocean, and three of the four stations in grid cells containing more variable subgrid-scale terrain. The YSU experiment shows smaller biases at three stations in the river valley, for which the Topo experiment correction is excessive.

The cluster analysis also shows the correlation between increasing biases and increasing wind speed, but only in the MYJ and YSU experiments. The Topo experiment exhibits rather different behavior, lacking a

---

**FIG. 10.** Cluster biases for (top) positive and (bottom) negative meridional wind component in the (left) YSU and (right) Topo experiments. Biases are binned by observation speed intervals. The gray lines show the bias of each cluster, and the colored line shows the total bias. The mean sample size is shown by the solid black line along the secondary y axis.
clear relationship between the bias and cluster speed. Ranked biases show that the Topo experiment biases are related to the clusters’ \( u \) and \( v \) components, and suggest that its performance is best when \( v \) is positive. The two highest biases occur in northwesterly (negative \( v \)) high-speed clusters, where the circulation pattern promotes more interactions of the systems with the highest and most complex terrain in the domain. The differences among the biases between terrain-interacting flows and flows entering from the ocean, which have a reduced drag adjustment, reveal that a modulated adjustment can improve the simulations. Analysis of clusters with ocean flow shows wind speed distributions closely matching the observations, a sharper systematic error distribution, and reduced total bias. The negative bias is still present in high wind speeds, but the issue originates in the friction velocity parameterization, and is only exacerbated by the constant coefficient adjusting the drag in the Topo-wind scheme.

In general, NWP model analysis on complex terrain must consider that wind speed and direction are strongly associated with terrain features and, therefore, with model errors. To prevent averaging out significant results, the conditions upon which errors may originate should be investigated. The seasonal analysis in this study indicated that errors are associated with synoptic conditions, followed by geography. Diurnal cycle variations confirmed known issues in the parameterizations. Finally, the clustering analysis provided evidence that errors are conditioned on flow and offered a starting point for further work on minimizing errors.

Results indicate that although the Topo-wind scheme can improve simulations in many cases and regimes, it needs tuning for optimal performance in various conditions. Forecasts and simulations can be constrained to use the scheme for typical summer storms, events of low wind speeds, or for minimizing total biases. Improvements to the scheme could be achieved with a conditional activation determined by atmospheric stability, by modulating the drag adjustment according to wind speed, and possibly by including an ad hoc correction to the superlinear relationship between \( u_* \) and high wind speed. We emphasize that the scheme evaluation presented here is partial, in that we only analyzed adjustments applied in regions identified as valleys and plains. The Topo-wind scheme is also able to correct overestimated winds in more extreme terrain not present in this study.

The results here also indicated that forecast verification can benefit from different approaches that take into account other conditioning factors such as flow over terrain. We conducted a clustering analysis capable of separating patterns within a season and elucidating a dependency on speed and direction. Our results revealed flow peculiarities that may inform postprocessing methods and further developments of PBL parameterizations. An ensemble of models using weights adjusted to prefer models of specific abilities can be developed using clusters without any a priori information relative to the mesoscale dynamics. Operationally, clusters offer an opportunity for conditional bias correction, as well as for enhancing forecasts of specific, problematic flows. Parameterization developers benefit from cluster analyses designed to identify specific conditions within which a scheme thrives or fails, providing additional feedback that augments traditional case or seasonal focused analyses.

**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. This study was enabled by a long-term visit supported by NCAR’s Advanced Study Program (ASP), and high-performance computational resources by Yellowstone (ark:/85065/d7wd3xhc) maintained by NCAR’s Computational and Information Systems Laboratory and sponsored by the National Science Foundation. MEBF is in residency as a long-term visitor at the National Center for Atmospheric Research, Boulder, Colorado.

**REFERENCES**


——, ——, and ——, 2003b: Multiseason verification of the MM5. Part II: Evaluation of high-resolution precipitation forecasts.


