Nested Mesoscale Large-Eddy Simulations with WRF: Performance in Real Test Cases

CHARLES TALBOT, ELIE BOU-ZEID, AND JIM SMITH
Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey

(Manuscript received 5 May 2011, in final form 4 April 2012)

ABSTRACT

This paper assesses the performance of the Weather Research and Forecasting Model (WRF) as a tool for multiscale atmospheric simulations. Tests are performed in real and idealized cases with multiple configurations and with resolutions ranging from the mesoscale (gridcell size ~10 km) for the real cases to local scales (gridcell size ~50 m) for both real and idealized cases. All idealized simulations and the finest real-case simulations use the turbulence-resolving large-eddy simulation mode of WRF (WRF-LES). Tests in neutral conditions and with idealized forcing are first performed to assess the model’s sensitivity to grid resolutions and subgrid-scale parameterizations and to optimize the setup of the real cases. An increase in horizontal model resolution is found to be more beneficial than an increase in vertical resolution. WRF-LES is then tested, using extensive observational data, in real-world cases over complex terrain through nested simulations in which the mesoscale domains drive the LES domains. Analysis of the mesoscale simulations indicates that the data needed to force the largest simulated domain and to initialize surface parameters have the strongest influence on the results. Similarly, LES model fields are primarily influenced by their mesoscale meteorological forcing. As a result, the nesting of LES models down to a 50-m resolution does not improve all aspects of hydrometeorological predictions. Advantages of using fine-resolution LES are noted at nighttime (under stable conditions) and over heterogeneous surfaces when local properties are required or when resolving small-scale surface features is desirable.

1. Introduction

Mesoscale atmospheric models, limited by their grid resolution (>1 km), cannot capture the small-scale spatial variability of meteorological dynamics and land–atmosphere exchanges, especially over complex terrain. Local-scale atmospheric simulations (resolution ~50 m), focusing on smaller domains, should ideally offer a more realistic representation of near-surface flows and surface fluxes owing to their ability to capture surface heterogeneity and resolve the large turbulent eddies in the flow. However, the ability of these local models to simulate real-world cases is often hindered by a lack of realistic lateral boundary conditions needed to drive the flow in the fine domain. Nested atmospheric simulations, where mesoscale models are driven by reanalysis data fields and in turn drive local-scale simulations, offer an appealing approach that can combine the strengths of the different modeling scales and enable high-resolution simulations of real-world cases. Such simulations would allow the downscaling of climatic or meteorological forecasts to the local scale using grid-nesting techniques and would be of significant importance for applications in hydrometeorology, urban micrometeorology, and wind power mapping, especially when the meteorological dynamics are expected to be influenced by surface heterogeneities and soil conditions (Bou-Zeid et al. 2007; Wu et al. 2009; Lemone et al. 2010; Rihani et al. 2010).

Examples of atmospheric problems that can be better investigated with multiscale simulations include 1) the dynamics of convection in the atmospheric boundary layer (ABL) and convective clouds and rainfall (Siebesma et al. 2003; Moeng et al. 1996; Bryan et al. 2008; Stevens and Bretherton 1999), 2) the impact of urban areas on the microclimate (Oke 1988) and thunderstorm dynamics over built terrain (Ntelekos et al. 2008), and 3) the influence of abrupt changes in the local landscape and land cover on atmospheric flows and land–atmosphere exchanges (Bou-Zeid et al. 2004; Talbot et al. 2007a; Courault et al. 2007; Chan 2009; Rotunno et al. 2009; Stoll and Porté-Agel 2009; Bou-Zeid et al. 2009), as well as their influence on the transport of pollutants (Talbot et al. 2007b).
One model that allows multiscale nested simulations is the Weather Research and Forecasting Model (WRF), developed mainly at the National Center for Atmospheric Research and the National Oceanic and Atmospheric Administration. The WRF is a community model that is being increasingly used to study atmospheric dynamics and land–atmosphere interaction at various scales (Zhang et al. 2009; Rotunno et al. 2009; Catalano and Moeng 2010). Some of the appealing features of the model for this type of simulation, in addition to its nesting capabilities, is the design of the code to run on massively parallel computers, as well as the availability of real-world land-use and topography data and of regional-scale meteorological forcing data (used to drive WRF) that are easily imported into the model. Recent studies have assessed the sensitivity of mesoscale WRF simulations to grid resolution (Alifò et al. 2009; Hill and Lackmann 2009) and to the various choices of physical models (Borge et al. 2008; Misenis and Zhang 2010; Shin and Hong 2011). Other studies investigated the turbulence-resolving large-eddy simulation mode of WRF (WRF-LES) in idealized simulations with no coupling to mesoscale simulations or real-world meteorological forcing. For example, Moeng et al. (2007) tested a new subgrid-scale (SGS) turbulence model for LES to reduce a bias in temperature and vertical velocity outputs of a nested model in a two-way nesting run. Mirocha et al. (2010) and Kirkil et al. (2012) implemented the nonlinear backscatter and anisotropy model (Kosovic 1997), the dynamic reconstruction model (Chow et al. 2005), and the Lagrangian scale-dependent dynamic model (Bou-Zeid et al. 2005) to improve the representation of SGS turbulence in WRF-LES and showed significant improvement in idealized simulations with the new SGS models compared to the basic Smagorinsky formulation with a constant coefficient (only the first of these models is now available in the public releases of WRF but was not included at the time the simulations of this study were run).

However, tests of the skill of WRF-LES in nested real-world simulations are still rare and are critically needed in view of the increasing and broadening use of the model. The challenges in performing such simulations are often different from the challenges of idealized cases. Recently, Liu et al. (2011) tested WRF-LES, with data assimilation, over real terrain in a nested mesoscale–LES configuration for wind power applications. Their comparison of modeled fields to observations showed a discrepancy and they recommended further tests to elucidate whether inaccurate synoptic forcing or coarse resolution (100 m in their study) are the sources of the model errors. This paper aims to bridge this gap in real tests of WRF-LES and to assess its performance, nested in the WRF mesoscale model, in simulating real cases. We use an a posteriori validation approach where we assess the ability of the model to reproduce observed parameters, with a particular focus on the role of the reanalysis forcing data and mesoscale simulation setup that drive the finescale LES and on the effect of grid resolution. Preliminary tests in idealized cases are also performed to test specific aspects of WRF-LES that are relevant to real-case simulations, mainly independent horizontal and vertical grid refinements and the sensitivity of the model to SGS turbulence parameterization schemes in idealized cases, which we then contrast to the sensitivity to SGS schemes in real cases.

The paper is organized as follows. Section 2 is a presentation of the idealized LES tests for different configurations under neutral conditions. The setup of the nested real-case simulations, the experimental data used for validation, and the sensitivity of the mesoscale-modeled fields to various options are presented in section 3. In section 4, we validate the nested mesoscale-LES runs for various configurations and options, using measurements of ABL depth, surface fluxes, and near-surface meteorological parameters. A summary and discussion follow in section 5.

2. Idealized large-eddy simulations

Idealized simulations are performed with a periodic flow domain (the flow exiting the domain is reentering on the other side) in the horizontal directions. A homogeneous and steady forcing is used since it allows comparison of the results to theoretical laws or to normalized observational findings. The aim of these tests is to investigate specific aspects of the WRF-LES capabilities that are important for real cases. For detailed tests of WRF-LES in idealized simulations the reader is referred to Moeng et al. (2007), Mirocha et al. (2010), and Kirkil et al. (2012). First, we investigate the ability of the LES code, under neutral conditions (zero surface buoyancy flux), to yield the logarithmic velocity profile (log law) predicted by the law-of-the-wall and given by (Pope 2000):

$$\frac{u}{u_*} = \frac{1}{\kappa} \ln \left( \frac{z - d}{z_0} \right),$$

(1)

where $\kappa$ is the von Kármán constant (=0.4), $z$ is the height above the ground, $d$ is the displacement height of the surface, and $z_0$ is the surface-roughness length. The friction velocity, $u_*$ = ($\tau_s/\rho$)$^{1/2}$, is computed by the model from the surface stress $\tau_s$ and the density $\rho$. We also analyze the profiles of velocity variances and energy spectra produced by WRF-LES.

The idealized simulation domain is 6.72 km in the streamwise ($x$) and the cross-stream ($y$) directions, and
1.2 km in the vertical (z). The horizontal domain size is more than five times the atmospheric boundary layer height to allow the largest turbulent eddies to go through a few turnover cycles before exiting (and reentering) the domain in order to mimic the large horizontal extent of atmospheric flows and to allow each eddy to evolve without interacting with itself (Bou-Zeid et al. 2004). It is also necessary in nonperiodic WRF simulations to have a representative ensemble average of large eddies and avoid nonzero mean vertical velocities when such values are not physical (Moeng et al. 2007). The flow is driven by a mean horizontal pressure gradient (i.e., a geostrophic wind), and the Coriolis acceleration is included.

The different horizontal and vertical resolutions used are all summarized in Table 1 along with the friction velocities that they produce. The default WRF numerical discretization options were used—consisting of a third-order Runge–Kutta scheme for time advancement, a fifth-order finite-difference advection scheme in the horizontal directions, and a third-order advection scheme in the vertical. We note that these odd-order finite difference schemes will generate numerical diffusion terms that will dissipate resolved energy in a manner similar to what the subgrid-scale (SGS) model does: we will return to this point later in this section. We also note that WRF includes other filters, some of which might also dissipate kinetic energy. The “polar filter” is not relevant in our study. The “sixth-order spatial filter” and the “vertical velocity damping filter” were turned off, as recommended in the user’s manual (Skamarock et al. 2008). These are also the filters that are most likely to produce additional numerical diffusion. However, the “three-dimensional divergence damping filter,” the “external-mode filter,” and “the semi-implicit acoustic step off centering,” which are useful in real-world simulations and which act on the acoustic modes (and hence are less likely to create artificial diffusion and dissipate turbulent kinetic energy), are kept on and their default parameters were used (see Skamarock et al. 2008 and references therein for details).

These numerical configurations were also used in the real-world runs. A damping layer near the domain top is not used in the idealized cases (which only go up to 1.2 km) since they appear to produce very high shear and result in high variances near the domain top. A damping layer, however, will be used later in the much deeper real-case domains to avoid abrupt changes in the stress and to prevent reflection of waves.

The initial vertical wind profile, for the idealized simulations, is obtained from an Ekman profile with a 10 m s⁻¹ geostrophic wind aligned with the x axis at the top of the domain. The simulations were run for a warm-up period of 40 h to ensure that the results reach statistical stationarity regardless of the initialization, for example, no more significant change is observed in the spatially averaged vertical profiles. The runs were then pursued for 10 more hours over which data were collected for analysis. Reynolds ensemble averages were approximated using spatial averaging over each horizontal plane as well as time averaging over the 10 h. The surface pressure value is 1000 hPa. All idealized simulations have zero surface buoyancy flux in order to compare the wind vertical profile to the log law. The surface roughness is z₀ = 0.1 m. Simulation time steps decreased from 1 to 0.33 s as the resolution was increased to maintain numerical stability.

Our control simulation uses the turbulent kinetic energy (TKE) 1.5 order of closure (TKE1.5) SGS model where a prognostic equation is solved for the subgrid-scale turbulent kinetic energy (SGS-TKE). Subgrid-scale stresses are then modeled using an eddyviscosity approach where the eddy viscosity is proportional to the square root of the SGS-TKE (Sagaut 2006). In this study, all coefficients for this SGS-TKE equation are kept at the default WRF values (Cₑ = 0.18, Cₓ = 0.1, smdiv = 0.1, emdiv = 0.01, and epssm = 0.1; see Skamarock et al. 2008). Although some studies (Moeng et al. 2007) have tested different coefficients, we did not want to focus on the effect of these model coefficients in this study.

### a. Vertical profiles in the neutral atmospheric boundary layer: Sensitivity to grid resolution

Since nesting in WRF is only done in the horizontal directions, the vertical resolution remains constant as the horizontal resolution improves (the vertical resolution

---

**Table 1. Horizontal and vertical resolutions and grid sizes for the idealized simulations.**

<table>
<thead>
<tr>
<th>Domain size</th>
<th>Simulation identifier</th>
<th>Grid points</th>
<th>Horizontal resolution Δx (m)</th>
<th>Height of first grid point (z₁)</th>
<th>Vertical grid resolution (dz = 2z₁)</th>
<th>u₀ (m s⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lₓ = 6.72 km</td>
<td>1-1</td>
<td>140 x 140 x 80</td>
<td>48</td>
<td>z₁ = 7.25 m, dz = 14.5 m</td>
<td>0.363</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-2</td>
<td>140 x 140 x 40</td>
<td>48</td>
<td>z₁ = 14.5 m, dz = 29 m</td>
<td>0.357</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-3</td>
<td>140 x 140 x 20</td>
<td>48</td>
<td>z₁ = 30 m, dz = 60 m</td>
<td>0.366</td>
<td></td>
</tr>
<tr>
<td>Lᵧ = 6.72 km</td>
<td>1-4</td>
<td>140 x 140 x 40 + stretching</td>
<td>48</td>
<td>z₁ = 4.8 m, 9.6 &lt; dz &lt; 84 m</td>
<td>0.369</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-5</td>
<td>70 x 70 x 40</td>
<td>96</td>
<td>z₁ = 14.5 m, dz = 29 m</td>
<td>0.370</td>
<td></td>
</tr>
<tr>
<td>Lz = 1.2 km</td>
<td>1-6</td>
<td>47 x 47 x 40</td>
<td>144</td>
<td>z₁ = 14.5 m, dz = 29 m</td>
<td>0.373</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1-7</td>
<td>35 x 35 x 40</td>
<td>192</td>
<td>z₁ = 14.5 m, dz = 29 m</td>
<td>0.379</td>
<td></td>
</tr>
</tbody>
</table>
can be increased if each nest is run independently but that introduces some disadvantages). Therefore, 1) nested simulations will improve upon coarser simulations results only if a grid refinement in the horizontal directions is beneficial, 2) the identical vertical resolution for all nested runs must be selected to optimize the results of the finer domains even if such a high vertical resolution is not required in the coarser domains, and 3) the grid aspect ratio, which can have a significant impact on the results (Scotti et al. 1997; Brasseur and Wei 2010; Mirocha et al. 2010), will change as the grid is refined.

We address these issues by performing simulations with different horizontal (vertical) resolutions where the vertical (horizontal) resolution is kept constant (Table 1). Our aim is to optimize the setup of the nested grids for the real cases by assessing the improvement in the results that can be expected from grid refinement in the different directions—within a range of resolutions that can be realistically attained. Ideally, we would observe convergence of the results (no further change as the resolution is increased) for the mean profiles in our simulations, but we do not expect this convergence to apply for variances since we only report the resolved part of these statistics, which will keep increasing in large-eddy simulations as the resolution is increased (an estimate of the subgrid-scale component of the variances can be made as in Moeng et al. 2007 but introduces uncertainties that we prefer to avoid here).

We first test vertical resolution refinement by maintaining a fixed horizontal resolution of 48 m in all simulations, while the vertical resolution is varied. The results, obtained with the TKE1.5 SGS model, and corresponding to simulations I-1 to I-4 in Table 1, are reported in Fig. 1. The mean velocity profiles, depicted in Fig. 1a, are not in very good agreement with the logarithmic profile [Eq. (1), solid line]. The lowest two resolutions depart from the log law at the second grid point, while the highest vertical resolution run and the run with the stretched grid produce somewhat better results near the surface. The Smagorinsky SGS model (Smagorinsky 1963) (not shown here) results in very similar profiles in the surface layer. The mismatch with the log law is consistent with previous applications of many SGS models and previous tests of WRF-LES (Andren et al. 1994; Mason and Thomson 1992; Moeng et al. 2007; Mirocha et al. 2010). It is linked to an overestimation of the SGS dissipation (extraction of the TKE from the resolved scales by the SGS model) and can be fixed by new-generation dynamic SGS models (Porte-Agel et al. 2000; Bou-Zeid et al. 2005; Chow et al. 2005; Stoll and Porte-Agel 2006) as demonstrated for WRF in Kirkil et al. (2012).

The horizontal velocity variance profiles (Fig. 1b) are in good agreement with the typical ranges observed in field measurements and reported for neutral ABLs (Stull 1988); increased vertical resolution slightly increases the resolved variance and results in peak values closer to the ones observed in other LES studies of the neutral ABL (e.g., Bou-Zeid et al. 2005). The vertical velocity variance (Fig. 1c), on the other hand, peaks too far from the wall, although the value at the peak is in agreement with...
observations and other simulations that suggest this variance (normalized by the squared friction velocity) should peak at a value of about 1.5–2.5 somewhere around the top of the atmospheric surface layer (Stull 1988; Bou-Zeid et al. 2005). The velocity variance profiles are similar to results reported in Moeng et al. (2007, their Fig. 2b) at comparable resolutions, although they added the modeled SGS variances and used a different SGS model. Refining the grid vertically moves the profile in the right direction since it produces higher variance near the surface and peaks exceeding 1.5. This is expected since these finer simulations resolve a larger fraction of the variances.

In general, apart from the very low vertical resolution simulation with 20 levels, the results are not highly sensitive to the increased vertical resolution. The effect of the grid stretching is noticeable: the 40-level stretched grid simulation matches the 80-level results near the ground but is closer to the 20-level results farther aloft. Based on these results, we selected a stretched vertical grid, with 109 levels, for the real cases presented later (with a much deeper domain) such that the vertical resolution near the surface will fall between the vertical resolutions of I-2 and I-4 of the idealized tests since Fig. 1 suggests very moderate improvement beyond such vertical resolutions.

We then tested the sensitivity of the model results to horizontal resolution with 40 vertical levels without vertical stretching, again using the TKE1.5 SGS model. These tests represent the differences between nested LES domains that share the same vertical resolution but where the horizontal resolution increases; their details are presented in Table 1 (simulations I-2, I-5, I-6, and I-7). The mean wind profiles do not show a clear improvement trend (better log-law prediction), with an increasing horizontal resolution (Fig. 2a). Nevertheless, the finest simulation clearly produces a more linear profile suggesting at least that a log law is observed, albeit with a different von Kármán constant than the 0.4 value used in the WRF wall model and in the log profile plotted in Fig. 2a. The horizontal resolution had a very large impact on the variances profiles. The lowest resolution (192 m) produces very high streamwise and very low vertical velocity variances, pointing to significant errors in the representation of the turbulence. As the resolution is increased, the peaks of the variances occur closer to the surface, decrease in magnitude for the streamwise velocity variance, and increase for the vertical velocity variance. In both cases, the higher horizontal resolution profiles are in better agreement with literature data (Stull 1988) and other LES results (Bou-Zeid et al. 2005). Based on these results that show continued improvement with increasing horizontal resolution, we refined our horizontal grid spacing down to ~ 50 m in the real cases presented later to match the finest idealized simulation resolution that we tested here (I-2). Additional refinement is not realistic with current computational resources.

Overall, the grid resolution tests indicate that numerical convergence is not observed (as expected for the grid resolution used here, see Mirocha et al. 2010), but higher-resolution simulations tended to have better results compared to lower resolution simulations. More importantly, these test show that the results are significantly more

![Fig. 2. Sensitivity test to horizontal resolution for simulations with the TKE1.5 subgrid-scale model and 40 vertical levels and horizontal resolutions of 48 m (I-2), 96 m (I-5), 144 m (I-6), and 192 m (I-7): (a) wind velocity profiles, (b) resolved variance of the streamwise wind velocity, and (c) resolved variance of the vertical velocity, all normalized by the squared friction velocity.](image-url)
sensitive to horizontal than vertical resolution increases, which is an encouraging result for nested simulations where the refinement of the grid is in the horizontal directions only.

b. Sensitivity to the subgrid-scale model

We also perform a brief analysis of the sensitivity of the idealized cases to the two SGS models available in the version of WRF used here. Our aim is to contrast the sensitivity of WRF to SGS closure schemes under idealized and real cases; the reader is referred to other references (Moeng et al. 2007; Mirocha et al. 2010; Kirkil et al. 2012) for an in-depth analysis of SGS closure schemes in WRF. We compare the results obtained with the TKE1.5 parameterization (used in the previous subsection) and the Smagorinsky parameterization (Smagorinsky 1963) with a constant coefficient. The constant Smagorinsky coefficient used in our simulations has a value of $C_s = 0.18$ as implemented in WRF although a reduction of the value close to the wall is common and, indeed, needed in ABL applications (Mason and Thomson 1992); our aim of contrasting the SGS closure role under idealized and real cases should however not be very sensitive to model coefficient tuning.

The two SGS models gave very similar results for the mean and variance profiles (not shown here). Another useful comparative test is to verify the model’s performance regarding SGS dissipation—that is, the cascade of TKE from the resolved to the SGS scales produced by the two SGS models, which is linked to the poor match with the log law. This is often regarded as the most critical function of an SGS model and as an imperative for the correct evolution of turbulent kinetic energy during the simulation (Meneveau and Katz 2000). The energy spectra from time series of the velocity, normalized by $u^*_2$ and the height over the surface $z$, are plotted in Fig. 3 versus the nondimensional angular frequency $n = k z / \langle u \rangle$ (as in Kaimal and Finnigan 1994, for example), where $k$ is the angular frequency in radians per second. In the inertial subrange, the Kolmogorov theory (Kolmogorov 1941) suggests a decrease of the TKE with increasing frequency with a slope of $-5/3$ (on a log–log plot) while in the production range empirical and theoretical data for wall-bounded flows suggests that the decrease follows a slope of $-1$ (Perry et al. 1987; Katul et al. 1995). Time series were collected during the simulations at various model levels, ranging from the first level (14.5 m) to the middle of the ABL at 600 m and at various locations with horizontal separation distances of 2 km to minimize the correlation of the time series. The time series are collected from the last 4 h of the simulations and then divided into eight periods of 30 min each and the spectra of each period and for the various horizontal locations at the same height are then averaged.

The spectra plots in Fig. 3 indicate an overdissipation of turbulent kinetic energy by both SGS models (and probably by the odd-order finite-difference numerical schemes of WRF) that directly influences the inertial range $[k z / u(z) > 1]$, yielding slopes much steeper than

![Fig. 3. Streamwise velocity spectra for the two subgrid-scale models: (a) TKE1.5 and (b) Smagorinsky. The different curves represent different heights with the spectra at the first grid level (farthest to the left) and the spectra at the middle of the ABL (farthest to the right). The straight solid lines on each figure represent the expected decay of energy with a slope of $-1$ for the production range and $-5/3$ for the inertial range.](image-url)
The reduced TKE in the inertial subrange then increases the cascade from the production range \( [kz/\ u(z) < 1] \) also leading to slopes steeper than \( -1 \) in that range. This overdissipation is more pronounced at the first model level (far left curve on the figure), where the flow is poorly resolved and strongly affected by the SGS and wall models. In general, the two SGS models yield very similar results in idealized cases; one would then also expect them to be similar in real tests.

3. Real cases: Forcing data, nesting, and mesoscale simulation setup

The tests of WRF-LES in idealized simulations illustrated limitations of the model in reproducing the physical features of a neutral ABL. The turbulent eddies in the neutral ABL are smaller than those in a convective unstable ABL but larger than those in a stable ABL. Thus we expect the model performance to be better for idealized convective cases and worse for idealized stable cases. However, higher resolutions, especially in the horizontal direction, improved the model performance as illustrated above and better (dynamic) SGS models can further improve the WRF-LES performance, as illustrated in other studies (e.g., Kirkil et al. 2012). Furthermore, LES has been able to significantly improve our understanding of idealized ABL dynamics and land–atmosphere interaction despite shortcomings in SGS closures, wall models, and numerical discretization. With the additional ability in WRF to force the flow in the LES domain using real mesoscale meteorological fields, WRF could potentially further our understanding of real-world ABLs in a similar way. Therefore, the rest of this paper focuses on testing the performance of WRF-LES in real cases. In such tests, various inputs and physical features of the simulation domain, such as the land surface model, surface boundary conditions, meteorological forcing, and nesting scheme, play an important role.

First we assess the potential errors related to mesoscale forcings: can mesoscale simulations in WRF provide adequate boundary and initial conditions for LES? Then we proceed to nest WRF down to LES scales and compare its predictions to various field measurements including 1) surface fluxes at an eddy-covariance (EC) station in complex terrain, 2) vertical profiles measured using soundings, and 3) ABL height determined using a light detection and ranging (lidar) system with three wavelengths. These measurements are mainly taken over New Jersey (NJ) and the Princeton University campus in particular; therefore, our nested simulations are centered over Princeton. These real-case simulations focus on one diurnal cycle during a cloud-free day to avoid errors related to cloud and radiative parameterizations (e.g., development of stable conditions under a convective cloud: Aligo et al. 2009).

a. Meteorological inputs and numerical setup

In real-world simulations, the Monin–Obukhov similarity theory (Monin and Obukhov 1954) is used as a wall model that provides friction velocity to the WRF and couples it to the land surface model via surface fluxes of heat and water vapor. The evolution of parameters of relevance to the surface energy balance and surface water balance such as soil moisture, surface skin temperature, and heat fluxes are computed during the simulation by the Noah land surface model (LSM) (Chen and Dudhia 2001). Surface fluxes calculated by the Noah LSM are used in the mesoscale model, via the planetary boundary layer (PBL) scheme, to predict the ABL height and dynamics, and in the prognostic turbulent kinetic energy equation to provide estimates of buoyant TKE production or destruction near the surface. The PBL dynamics in mesoscale runs are simulated using the Yonsei University PBL scheme (Noh et al. 2003; Hong et al. 2006)—referred to as YSU PBL. This scheme was one of the best performers in a recent comparative study (Shin and Hong 2011) in which it was concluded that there are serious shortcomings with all PBL schemes especially under stable (nighttime) conditions. Over urban areas, Noah is coupled to the urban canopy model (UCM) of Kusaka et al. (2001) to compute fluxes for a given urban density, street alignment, and urban cover fraction. This urban canopy model simulates thermal processes in urban areas with very good detail, but does not resolve building-scale flow dynamics; it simply represents the urban areas as separate land-use types with higher momentum roughness length that depends on the urban density (see, e.g., Wang et al. 2011a,b). As such, the first model level needs to be higher than the sum of roughness length \( (\ z_0) \) and the displacement height \( (d) \) in the UCM [to keep the logarithmic term in Eq. (1) positive], which limits the vertical resolution near the surface. The implementation of building-resolving capabilities in WRF has however been previously demonstrated (Lundquist et al. 2010), though it is not included in the WRF version used in this paper and thus is not used here.

Meteorological initial and boundary conditions are provided to the coarsest mesoscale simulations from the National Centers for Environmental Prediction (NCEP) Eta model reanalysis products. Two such products are used; the first one is the North American Regional Reanalysis (NARR; details can be found on the NARR homepage http://www.emc.ncep.noaa.gov/mmb/reanalysis/ and on http://dss.ucar.edu/datasets/ds608.0/) with outputs every 3 h. This dataset has a resolution of 32 km horizontally and 30 vertical levels up to 100 hPa. The second
A product that could provide the needed input to the mesoscale simulations is the Global Energy and Water Cycle Experiment (GEWEX) Continental-Scale International Project (GCIP) output data, which has a frequency of 3 h and a resolution of 42 km horizontally, with 27 vertical levels up to 50 hPa (details of the WRF input from this dataset can be found at http://dss.ucar.edu/datasets/ds609.2/).

The chosen radiative scheme is the Rapid Radiative Transfer Model (RRTM) (Mlawer et al. 1997), which computes shortwave and longwave radiation at fine time scales (e.g., every 2 min). The cumulus parameterization is present only in the coarsest mesoscale simulation of 12.15-km resolution and relies on the Grell scheme (Grell and Devenyi 2002). As we mentioned earlier however, we have selected a clear day when the scheme is not relevant.

Six domains were nested for the simulations; the largest and coarsest three are used for mesoscale simulations (Fig. 4a), while in the smallest and finest three (Fig. 4b) the large-eddy simulation technique is used. The coarsest LES is forced by the finest mesoscale simulation with...
boundary conditions interpolated linearly in time and space from the mesoscale simulations fields, which are saved every 30 min. The nesting is one way, so no information is passed from the LES to the mesoscale simulations: previous comparisons by Zhu et al. (2010) did not find significant improvement when two-way nesting was used in WRF simulations of stratocumulus clouds. The mesoscale simulations and the LES are run separately (mesoscale inputs were saved every 30 min and later used to force the LES). As depicted in Fig. 4a and Fig. 4b, the largest domain (d1) extends over the entire northeastern United States, while the smallest mesoscale simulation (d3) covers the northern part of New Jersey and southern part of New York. The finest LES simulation (d6) covers an area of approximately $5 \times 5$ km$^2$ centered over Princeton at $40.34^\circ$N, $74.65^\circ$W (Fig. 4). The details of the six domains are provided in Table 2. The vertical grid is stretched in all simulations up to the top boundary of the domain at 100 hPa.

As discussed by Wyngaard (2004), both the finest mesoscale simulation and the coarsest LES might fall in what he termed “terra incognita”—a region where Reynolds-averaged closures can be problematic, since some turbulent fluxes might be resolved, and where LES is also challenging since the grid resolution does not allow the simulation to capture enough resolved fluxes. In addition to these potential problems with d3 and d4, at the transition from a coarser to a finer nest, some distance is needed for the small-scale features to develop. For example, the turbulence at the transition from d3 to d4 will require some distance to develop since d3—a Reynolds-averaged mesoscale simulation—does not have any resolved turbulence to feed into d4 (except for large mesoscale convective motions that may be captured by d3 and that would help create regions of high shear in domain d4 that would enhance turbulence production).

Similarly, small-scale turbulence will need to develop as we transition from d4 to d5 or from d5 to d6. We performed detailed analysis of the TKE levels in the various domains to ensure that a realistic level of turbulence is reached. While we do not present the details of this analysis here, we observed that the turbulence in domains d4 and d5 develops relatively quickly (specific TKE values exceeding $1 \text{ m}^2 \text{s}^{-2}$) after the transition from a coarser to a finer grid, even when the transition is from a mesoscale to an LES. We attribute this to the topography and to the unsteadiness and spatial variability in the mean incoming flow that will promote TKE generation. This aspect of nested WRF simulations however warrants more in-depth analysis to validate and improve model performance at grid transitions (adding, for example, a synthetic turbulence generator between d3 and d4 in WRF). We note however that the design of the simulations and nests aimed to minimize such transition effects: 1) we always use a nesting ratio ($dx_{coarse}/dx_{fine}$) of 3, as recommended for WRF (Skamarock 2004), to ensure that the small-scale information is gradually generated as the grid is refined and 2) we center our domains in their parents (coarser forcing domain) and ensure that nests are separated from the boundary by at least one ABL depth. We also mainly focus on results at a distance of one ABL depth from the inflow boundary. These elements of the analysis are guided by the results of Vanella et al. (2008), who found that the small-scale turbulent eddies were well developed after a distance from the transition of about one integral scale (which is on the order of the ABL depth) in their LES with sudden mesh refinement.

The large number of vertical grid points (the same for all runs) and the vertical stretching allow a fine resolution near the ground, with the first grid point at about 12 m above ground level on average. All simulations have 44 vertical levels in the lowest 1200 m, yielding a vertical resolution that falls between that of idealized simulations I-2 and I-4. One can notice in Table 2 that the time steps for the LES runs are smaller [even when reported as (wind speed) $\times dt/dx$] than for neutral idealized cases. Larger time steps in the real cases led to numerical instabilities that could be related to the higher spatial variability of the maximum wind speeds, to significantly higher vertical velocities relative to the small $dz$ used, or to the transition regions in the nested real-world simulations.

For the mesoscale simulations, the geographical data for the land use and topography are obtained from the standard U.S. Geological Survey dataset (http://www.usgs.gov/) and have resolutions of 2 arc minute for the first domain d1 and 30 arc second for d2 and d3 (respectively 4.05 and 1.35 km). For the LES simulations, the topography is obtained from the 1-arc-second resolution (30 m) Shuttle Radar Topography Mission dataset (SRTM) (Rabus et al. 2003; Farr et al. 2007; http://www2.jpl.nasa.gov/srtm/; Fig. 4c). The land-use dataset, used by the Noah land surface model for the LES simulations, is adapted from the National Land Cover Dataset (NLCD 2001)

### Table 2. The six nested models used for the multiscale atmospheric simulations.

<table>
<thead>
<tr>
<th>Simulation type</th>
<th>PBL treatment</th>
<th>$dx$, $dy$ (m)</th>
<th>Grid points</th>
<th>Time step (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>d1 mesoscale</td>
<td>YSU PBL</td>
<td>12 150</td>
<td>91 $\times$ 82 $\times$ 109</td>
<td>72</td>
</tr>
<tr>
<td>d2 mesoscale</td>
<td>YSU PBL</td>
<td>4050</td>
<td>91 $\times$ 82 $\times$ 109</td>
<td>24</td>
</tr>
<tr>
<td>d3 mesoscale</td>
<td>YSU PBL</td>
<td>1350</td>
<td>91 $\times$ 82 $\times$ 109</td>
<td>8</td>
</tr>
<tr>
<td>d4 microscale</td>
<td>LES</td>
<td>450</td>
<td>91 $\times$ 82 $\times$ 109</td>
<td>2/3</td>
</tr>
<tr>
<td>d5 microscale</td>
<td>LES</td>
<td>150</td>
<td>91 $\times$ 82 $\times$ 109</td>
<td>2/9</td>
</tr>
<tr>
<td>d6 microscale</td>
<td>LES</td>
<td>50</td>
<td>100 $\times$ 100 $\times$ 109</td>
<td>2/27</td>
</tr>
</tbody>
</table>
Observational datasets

To validate the real-case simulations, we use a combination of field-measured data over the Princeton University campus and other available measurements in the simulation domains. WRF mesoscale outputs are compared to radio soundings from Upton, New York, located 150 km northeast of Princeton (Fig. 4a). The finest domain where this measurement point is included is d2. These soundings provide vertical profiles of potential temperature, water vapor mixing ratio, and wind speed and direction every 12 h.

Closer to the center of the domain, within 30 km from Princeton, the model output near the ground is tested by comparing the simulation results to measurements from two local meteorological stations of the New Jersey Weather and Climate Network (NJ Mesonet) (http://climate.rutgers.edu/njwxnet/). One station is in New Brunswick, New Jersey, and the other is at the Trenton Mercer County Airport [part of the Automated Surface Observing System (ASOS) of the National Weather Service]. Both setups follow the World Meteorological Organization guidelines and are depicted with yellow dots on the map in Fig. 4b. These measurements include air temperature and humidity and wind speed and direction.

Local measurements made near the ground level at Princeton University are the most challenging in terms of numerical modeling since that meteorological station is intentionally located in a heterogeneous terrain and, hence, measures a local pattern of meteorological parameters at a height of 3 m. The station is above a grassland area, which is not representative of the dominant land cover in domain d6 (although we ensured that the land use at the grid point where the station is located is set as grassland). Its position is indicated by a black circle in Fig. 4d. Measurements at the Princeton station include eddy covariance (EC) surface fluxes of heat, water vapor and momentum, air humidity and temperature, soil moisture and temperature, surface temperature, the four radiative components, and wind speed and direction.

To verify the WRF skill in reproducing the bulk dynamics of the atmospheric boundary layer, data from a backscatter lidar system are used. The lidar measures the backscatter of light in three wavelengths (near infrared, visible, and ultraviolet). Backscatter is from particulate matter in the atmosphere and hence a sharp reduction in the return signal indicates the top of the ABL, which we detect using the local maxima of the first derivative of the backscattered infrared signal averaged over 2 min. As a result of restrictions on the laser operation for aviation safety, lidar measurements were done only during daytime from 1100 to 1744 local time on the simulated day. The simulated and measured data correspond to a cloud-free full diurnal cycle (24 September 2007) as indicated by shadowband radiometer measurements; these conditions were selected to ensure that the cloud parameterizations of WRF would not affect the local surface energy budget.

c. Mesoscale sensitivity to turbulence closure and meteorological inputs

To investigate the effect of mesoscale model dynamics and forcing on the finer large-eddy simulations, a series of three mesoscale simulations were realized. The aim is to assess the sensitivity of the nested simulations to GCIP and NARR meteorological forcing and horizontal turbulence models [Horizontal Smagorinsky (HorSmag) and TKE1.5]; the details of the three simulations are given in Table 3. GCIP and NARR reanalysis datasets, with a 3-h resolution, are used as initial and boundary conditions. Once the model runs start, these data are linearly interpolated in time and space to update the outer domain (d1), at every time step and boundary grid point, from the coarser reanalysis data. Figure 5 depicts vertical profiles of wind speed and direction, potential temperature, and water vapor mixing ratio from the mesoscale simulations’ outputs located in domain d2 and from the NARR and GCIP datasets at the same location as the radio-sounding measurements at Upton (see map Fig. 4a), which are also reported in the figure for comparison. Initial data profiles extracted from NARR and GCIP outputs are close to radiosonde measurements, as expected, since

<table>
<thead>
<tr>
<th>Case</th>
<th>Meteorological input data</th>
<th>Horizontal turbulence model</th>
<th>Horizontal resolutions for d1, d2, d3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>NARR</td>
<td>TKE1.5</td>
<td>12.15 km, 4.05 km, 1.35 km</td>
</tr>
<tr>
<td>B</td>
<td>NARR</td>
<td>HorSMAG</td>
<td>12.15 km, 4.05 km, 1.35 km</td>
</tr>
<tr>
<td>C</td>
<td>GCIP</td>
<td>TKE1.5</td>
<td>12.15 km, 4.05 km, 1.35 km</td>
</tr>
</tbody>
</table>
these soundings were assimilated in both products. Nevertheless, significant discrepancies can be observed near the surface, especially with the wind speed profiles where NARR and GCIP give significantly higher winds compared to the radiosonde, and with the potential temperature, which is overestimated by NARR, GCIP, and the WRF simulations. An interesting result to note is that WRF seems to be able to correct some of the bias in the mixing ratio; the results from all cases are closer to the sounding near the ground than either GCIP or NARR profiles.

Both turbulence closure models—the horizontal Smagorinsky and the TKE1.5 (case A versus case B)—follow the trend of the reanalysis data used to force the model and overpredict the wind speed (Fig. 5c). The two turbulence models perform similarly due to the fact that both use the same vertical turbulent mixing coefficients (the most important turbulent flux coefficient) produced by the PBL parameterization scheme. Some differences can be noted between the NARR and GCIP simulations using the same closure model (case A versus case C), especially in the wind speed profiles, suggesting a more important role for the meteorological forcing than for the horizontal turbulence closure model. The effect of different PBL schemes was not investigated here, but the reader is referred to Shin and Hong (2011) where such a study was performed.

Another aspect of the forcing of the simulations, in addition to the synoptic scale forcing, is the initialization of surface conditions. In contrast to synoptic scale forcing, these surface conditions are only provided to the model at the first time step (with the exception of the sea surface temperatures, which are provided at every step or kept constant). The evolution of the surface properties is then modeled by the coupled WRF–land surface model. Nevertheless, the initialization, and possible biases it involves, can have an impact on high-resolution LES, which are typically ran at most for a full diurnal cycle and thus are still affected by these initial surface conditions, especially the initial soil moisture. At the model initialization for the present simulations, GCIP data at the Princeton meteorological (EC) station indicated a value of the 10-cm-deep soil moisture of $0.220 \text{ m}^3\text{ m}^{-3}$, while the NARR initial value is $0.185 \text{ m}^3\text{ m}^{-3}$; measurements at the EC station indicated a drier surface with a soil moisture of $0.14 \text{ m}^3\text{ m}^{-3}$.

---

**FIG. 5.** Vertical profiles at 1200 UTC of (a) potential temperature, (b) mixing ratio, (c) wind speed, and (d) wind direction from WRF mesoscale runs compared with the Upton sounding (40.868°N, 72.883°W) and profiles from NARR and GCIP.
This difference between the soil moisture contents at the initialization of the model may lead to differences in ABL growth during daytime and in its peak depth during the simulated day; this is due to the higher latent heat fluxes and the associated lower sensible heat fluxes that the higher soil moisture values produce. To confirm whether this indeed occurs, we present in Fig. 6 the evolution of the daytime PBL as measured over Princeton campus from the backscattered lidar signal. The normalized and range-corrected backscatter is depicted as a grayscale color map in the figure, the data markers and lines overlaid on the backscatter plot are the computed ABL heights from the lidar signal (maximum gradient of the signal) and from simulations A, B, and C for the finest mesoscale simulation (d3). WRF values are computed by the YSU PBL scheme and are interpolated for the lidar location from the closest grid points. The height of the ABL deduced from the lidar increases from 500 to 1200 m during the daytime period from 1100 to 1744 eastern daylight time (EDT) (i.e., local time). As anticipated, the GCIP simulation, which had the highest soil moisture, yields the lowest PBL height: 300 m below the measured values. NARR simulations give slightly better results potentially due to the fact that they had a soil moisture initialization that was closer to the measured value, but the peak modeled PBL height remains about 200 m below the measured peak height. In addition, all simulations predict the collapse of the mixed convective ABL earlier than suggested by the measurement, though the measurements by the lidar could be detecting the height of the residual layer rather than the turbulent layer after 1700.

The growth rate is similar for both HorSMAG and TKE1.5 with NARR forcing (simulations A and B), again highlighting the fact that, in this case study, the input data from reanalysis products is significantly more important to the model performance than the horizontal turbulence closure model.

4. Real cases: Nested large-eddy simulation results

After assessing the mesoscale flow fields that WRF generates, we proceeded to study the large-eddy simulated flow fields from the inner three domains. Several simulations with different configurations were performed as summarized in Table 4; they are run separately from mesoscale simulations. Cases BLT and BLS were both forced by outputs of the same mesoscale simulation (case B in Table 3) using horizontal Smagorinsky and NARR, but BLT uses the TKE1.5 SGS model while BLS uses the Smagorinsky 3D SGS model. Their comparison allows the effect of the SGS model (if any, considering the results in section 2) to be determined. LES run CLT uses TKE1.5 as both the SGS and mesoscale turbulence model, but is forced by mesoscale simulations using GCIP data (case C in Table 3); its comparison with BLT allows the effect of the forcing (GCIP versus NARR) to be determined since the effect of the mesoscale horizontal turbulence model is negligible, as shown in the previous subsection.

a. The effect of increasing LES resolution

As deduced from the previous section, mesoscale case B and the resulting LES cases BLT and BLS are expected to match the experimental data better than CLT owing to the more accurate forcing, especially surface moisture initialization, that NARR provides for this test case. We have also verified that cases BLT and BLS give very similar results, and hence we will present results only from BLS in this section, where we focus on the differences between the three LES domains with different resolutions. In the next section, we will revert back to using all three LES runs to compare the different LES setups with the different SGS and forcing options listed in Table 4.

The evolution of the ABL height for the BLS simulation, from different domains, is presented in Fig. 7; we verified (not shown here) that cases BLS and CLT indeed

<table>
<thead>
<tr>
<th>Case</th>
<th>Meteorological input data</th>
<th>Parent mesoscale turbulence model</th>
<th>SGS turbulence model</th>
<th>Horizontal resolutions for d4, d5, d6</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLT</td>
<td>NARR</td>
<td>HorSMAG</td>
<td>TKE1.5</td>
<td>450 m, 150 m, 50 m</td>
</tr>
<tr>
<td>BLS</td>
<td>NARR</td>
<td>HorSMAG</td>
<td>SMAG</td>
<td>450 m, 150 m, 50 m</td>
</tr>
<tr>
<td>CLT</td>
<td>GCIP</td>
<td>TKE1.5</td>
<td>TKE1.5</td>
<td>450 m, 150 m, 50 m</td>
</tr>
</tbody>
</table>
yield very similar ABL heights. The figure displays the vertical time–height profiles of specific humidity on a grayscale map, on which the ABL height measured every 2 min by the lidar is overlaid. The other overlaid lines represent the ABL heights determined for mesoscale simulation d3 and for LES d4 and d6. The method of determination of the ABL heights \( (Z_i) \) is consistent for the LES and mesoscale simulations in this figure and relies on the detection of the highest vertical gradient of both temperature and water vapor mixing ratio (this method is referred to as the hybrid method) following

\[
Z_i = \max \left( \frac{dT}{dz} \times 10^3, \left| \frac{dq}{dz} \times 10^5 \right| \right)
\]

with \( \frac{dT}{dz} > 0, \frac{dq}{dz} < 0 \). (2)

Thus, the d3 heights are slightly different from Fig. 6 where they were computed by the YSU PBL scheme. The profiles of temperature and humidity used to compute the maximum gradients are averaged over an area centered at the lidar location and covering a number of adjacent grid points (2 × 2 grid points for d3, 4 × 4 for d4, and 8 × 8 for d6); this choice is a compromise to keep our averaging area close to the small physical averaging area of the lidar, while including enough grid points to avoid gradients with very high variability. The ABL depth computed from these profiles with different spatial-averaging extents in the different domains should not be very different since the ABL depth at a given location is controlled by fluxes over a large footprint. However, some effects of the averaging can be seen: for example, the height determined from d6 exhibits the highest variability owing to the fact that it was computed with the profiles with the least spatial averaging. Two sudden dips in the height determined from d6 coincide with entrainment of dry air, as can be seen in the grayscale map of the mixing ratio.

The results indicate that, overall, the ABL depth determined from the mesoscale simulation (d3) matches the lidar measurements better than the LES simulations (d4 and d6). ABL heights from LES domains are consistently lower than the measurements by about 300 m, while results from d3 are within about 150 m. Simulation results converge later in the day (after ~1500) but remain about 200 m lower than lidar measurements. One possible reason for this discrepancy is the difference in soil moisture between the simulations (high) and the measurements (low). However, one would need to measure surface conditions and compute the surface fluxes, which control the ABL height over the footprint area affecting the ABL height at the lidar location, to directly verify this hypothesis. The lack of surface parameter measurements covering the whole area prevents us from performing such an analysis.

Nevertheless, we can compute some surface parameters and fluxes to verify whether they vary with grid resolution in a manner consistent with our conclusions about the ABL depth. This allows us to directly validate modeled surface fluxes against measurements and elucidate the benefits of increasing the horizontal resolution and resolving more of the heterogeneities of the landscape on the simulation dynamics. This is especially relevant with the Noah land surface model since it only takes into account the dominant category of land cover in each grid cell. We compare latent and sensible heat fluxes from the different domains (again only for simulation BLS) to the eddy-covariance fluxes measured at the EC station. Figure 8a illustrates large biases in the modeled sensible heat flux during nighttime for the mesoscale simulation d3 and for the LES simulation d4. The finest simulation d6 produces better results at night and until about 1200 noon. After 1200 the LES simulation results from d4 and d6 collapse and yield lower sensible heat fluxes than the measurements; the mesoscale simulation agreement with the measurements during daytime improves beyond that of the LES runs. Results from different model resolutions for latent heat flux are depicted in Fig. 8b. The nighttime fluxes are insensitive to resolution, but this is probably related to their very small magnitude. However, during the day the differences in evaporation are significant with the mesoscale run d3 matching the measurements at the eddy-covariance station better than the LES runs, which produce higher latent heat fluxes. These results are in agreement with the ABL height analysis where the LES was found to produce a shallower ABL than the mesoscale simulation or the lidar measurements.

On closer inspection, however, the better results provided by domain d3 do not seem to be linked to better model performance. In the grid cell in d3 where the EC station is located, the dominant land-use type is the “low
intensity residential” and the fluxes from d3 do not reflect the correct land-use type of the EC station (grassland). Simulation domain d6, on the other hand, resolves the small grassland patches at the EC station and has the correct land-use type at this location. Because LES d6 is initialized with higher soil moisture, higher evaporation rates and lower sensible heat fluxes are predicted compared to the measurements, as one should expect.

The variability of the surface fluxes over d6 is captured in Fig. 9. In comparison, for example, d3 resolves this area with $5 \times 5$ grid cells. The figure depicts all the surface-flux variability that coarser simulations cannot capture. It also illustrates that, over landscapes like the one modeled here, the fluxes from the different land-use types are very different and hence the WRF–Noah coupled model with coarse resolutions can have significant errors in estimating fluxes because of the use of the dominant surface type in each individual grid cell. This also confirms that the good match between the latent heat flux ($LE$) modeled by d3 and measured at the EC station in the previous figure is a coincidence since low-intensity residential surfaces have slightly lower evaporative fluxes than grassland surfaces at noon.

The area of low evaporation in the center of Fig. 9 is the built-up core of Princeton where the land-use types according to Fig. 4d are medium- and high-intensity residential, while the elongated and curved area (to the bottom right) with both low $LE$ and low sensible heat flux ($H$) is Lake Carnegie. The lake evaporation dominated $LE$ during the night (not shown here). During the day, however, the air temperature rises but the lake surface temperature remains constant, creating statically stable conditions over the lake and leading to lower lake evaporation in WRF [see Mahrt and Ek (1984) for a detailed analysis of the effect of atmospheric stability on potential evaporation]. Note that Noah does not model water bodies, which are assumed to keep a constant temperature, as initialized.

Figure 10 depicts the wind speed, wind direction, temperature, and mixing ratio measured by the Princeton EC station and simulated over that location at a height of 3 m. The striking result in this figure is the discrepancy between measured and modeled wind fields, especially during nighttime, rather than the differences between the simulations. During the day, the numerical–experimental agreement for the wind speed and direction is better but without any clear advantage for the higher resolution simulations. During nighttime, d6 seems to produce lower wind speeds than the coarser simulations; it is thus closer to the measurements but remains significantly higher. Simulation d4 has significant errors in wind speed during the morning transition and in the mixing ratio during the day.

In these simulations, it is clear that the wind pattern for each simulation follows from the “parent” simulation that drives it. As such, the bias in wind speed and direction in the mesoscale simulations (Fig. 5) is passed on to all LES simulations, which are unable to fully correct this
bias in the forcing. The results for temperature on the other hand are good. The water vapor mixing ratio is captured on average, but the trends of its variation do not match the measurements. Overall, the finest LES simulation d6 gives somewhat better results during nighttime: lower wind speeds and higher mixing ratios at the end of the simulations, although the errors remain quite large for all domains. Also note that the effect of the LES resolution (d4 versus d6) seems to be especially important at night under stable conditions.

b. The effect of meteorological forcing and the SGS model on nested LES

In this section we will perform further tests focusing on near-surface wind and meteorological parameters to investigate if other settings (cases BLT and CLT) can improve the results and to contrast the model sensitivity to the SGS model to the (in)sensitivity we noted for idealized simulations. At ground level, comparisons are done between model outputs from LES domain d6 and the measurements from the Princeton EC stations, as well as between results from LES d4 with a resolution of 450 m and measurements from the New Jersey Weather and Climate Network, ASOS, and Mesonet stations. These stations are near the boundary of domain d4 and boundary effects can be important when the wind is blowing from the direction where the station is located.

Figure 11 depicts the comparison of the d6 simulation with data over Princeton. The differences between the BLT and BLS simulations, which differ by the SGS model used, are relatively small, but more significant than the differences between the two models under idealized simulations. Larger differences can be noted however between the results of the BLT/BLS simulations using NARR and the CLT results using GCIP, but there is no clear advantage of one reanalysis product over the other. Overall, temperature variations are captured well; however, considerable errors exist in wind speed and direction, especially during nighttime. Specific humidity variations are not well captured.

Figure 12 presents the time series of air temperature, specific humidity, and wind for the stations at Trenton and New Brunswick (see locations in Fig. 4b) versus LES outputs. Similar to the findings over Princeton with d6, the diurnal cycle of temperature is well reproduced by the model for both sites to within a few degrees. For the mixing ratio, the trends are not well reproduced by the model, although daytime measured and modeled values are in good agreement. Both of these results are similar to the observed model performance for the Princeton station. The wind speed simulated at New Brunswick is significantly higher than the measurements and there is a marked error in wind directions, as was the case over Princeton. On the other hand, the wind direction and

![Fig. 9. Maps of surface fluxes (W m⁻²) at 1200 noon local time over d6 from simulation BLS: (left) sensible heat flux and (right) latent heat flux.](image-url)
wind speed are well reproduced by the model for Trenton, indicating strong spatial variability in the ability of WRF to reproduce observed meteorological variables.

Differences between the simulations are mostly related to differences in reanalysis forcing (NARR versus GCIP) rather than to turbulence parameterization models. As observed in the idealized LES simulations, the currently implemented SGS models for LES give similar results. Some differences can be noted however in the magnitude of specific humidity and wind speeds and directions between different SGS models (both with NARR forcing) especially at the Trenton station during daytime. These differences in real cases are higher than what the idealized case comparison would suggest and could be related to the wider range of stability conditions in the real-case simulations resulting in a more important role for the SGS model. We note that data with westerly winds at Trenton should be analyzed with caution since the station is at the western edge of d4; while the measurements have no westerly winds (−270°), all models seem to produce westerly winds in the afternoon. For the New Brunswick station, northeasterly winds (−0°−90°) are also unreliable owing to the station location in d4. Thus, the data comparison for New Brunswick is more reliable in the afternoon. The results however show no clear improvements in the agreement with measurements for periods when the winds are blowing from the center of d4 toward the stations.

5. Summary and discussion

The WRF large-eddy simulation capability has been tested with horizontal grid resolutions down to 50 m over a flat and homogeneous (infinite) terrain in idealized neutral conditions and then compared to local meteorological
measurements in real-world conditions during a cloud-free diurnal cycle over a heterogeneous and complex terrain.

In idealized simulations, the 1.5-order TKE and the Smagorinsky SGS models gave relatively poor agreement with the expected log law of the vertical wind profile in the surface layer. Sensitivity tests to grid resolution showed that the model results are much more sensitive to changes in horizontal resolution than vertical resolution and tend to improve with increasing resolution. The vertical profiles of horizontal wind variances for the higher-resolution runs were similar to experimental and simulated profiles reported in the literature, both in vertical variability and magnitude, while the profiles of vertical wind variance were of lesser quality. Minor differences were found between the energy spectra and the profiles produced by the two SGS models tested here.

The WRF sensitivity in real-case mesoscale simulations to the choice of horizontal turbulence model and input data source (NARR and GCIP NCEP Eta products) was then tested. This is important since in nested simulations the output of the mesoscale domains is used to force real-case LES. Although the horizontal turbulence closure scheme for the mesoscale model was found to have little effect on the results, the forcing input data yielded significant differences that propagated to the smallest domains in the simulations.

The performance of WRF-LES was then analyzed in real cases. The choice of the synoptic forcing was found to be more important to the LES results than the choice of LES subgrid-scale or mesoscale horizontal turbulence closures. Different meteorological inputs from GCIP and NARR gave different forcing wind profiles and different surface properties at initialization. When the LES and mesoscale real-world simulations were compared, it was found that for bulk ABL dynamics (e.g., ABL height) the mesoscale simulations gave better agreement with

FIG. 11. Measurements at the Princeton eddy-covariance station vs LES model outputs from BLS, BLT, and CLT, all from domain d6, at a height of 3 m.
measurements than the LES with increasing resolutions. We observed that the basis for this better agreement of the mesoscale simulations does not seem to be improved model performance since all model resolutions were initialized with higher soil moisture than the actual measurements. Benefits of increasing model resolution were however noticeable for surface fluxes and near-surface meteorological parameters, particularly during stable nighttime conditions. In real cases, the differences between the results obtained with different SGS models were still minor but were more significant than in idealized simulations where the results were almost identical.

**FIG. 12.** Comparison of LES d4 outputs to measurements at (left) Trenton and (right) New Brunswick, New Jersey.
This higher sensitivity could be due to the wider range of stability conditions in the real-case simulations, which could be more challenging for the SGS model.

The lessons learned from this study that are useful for future application of WRF as a multiscale model for atmospheric simulations down to LES resolutions are the following.

1) Increased resolution improves the ability of WRF to capture surface variability, which facilitates comparison with measurements and thus improves model validation. It also allows the model to capture heterogeneous surface fluxes of importance in hydrometeorological applications. But, increased resolution does not seem to result in improvements in regionally averaged simulation outputs or in bulk ABL properties.

2) The mesoscale horizontal turbulence closure has little effect on the results analyzed and presented in this study. WRF recommends the 2D Smagorinsky, and we found in results not shown here that the 2D Smagorinsky seems to generate mesoscale features that appear more realistic than the TKE1.5 model. This insensitivity is related to the fact that the most important turbulent fluxes, in the vertical direction, are controlled by the diffusivities provided by the ABL scheme. Here we used the YSU PBL scheme and refer the reader to Shin and Hong (2011) for a comparison of different schemes.

3) The LES SGS model had a relatively small impact on our results: this is due to the fact the two models implemented in the WRF version that we tested are both eddy-viscosity type models (although the determination of the eddy viscosity is significantly different). Overall, the performance of these two models is poor when assessed based on the simulated idealized wind profiles and energy spectra, and similar to their performance in other codes. New-generation models being implemented in new versions of WRF (Mirocha et al. 2010; Kirkil et al. 2012) have a significantly improved performance, especially in matching the expect logarithmic wind profile.

4) The most important controls on model results were the meteorological forcing data. We used two products, NARR and GCIP, that use the same model (the Eta model from NCEP) to assimilate data and produce reanalysis products. The two datasets, however, produced differences in the forcing that were non-negligible. Moreover, the forcings they provided did not match soundings or ground measurements of surface conditions. Some of these forcing biases are slightly reduced by the nested WRF simulations (profile of specific humidity or near-surface wind speed in domain d6), but the biases have been largely downscaled by WRF and their impact on the smaller domains remained significant. Tests can be performed where mesoscale simulations are started a long time before the LES. This will result in initial surface conditions for the LES dictated by the WRF–land surface model dynamics, rather than the meteorological forcings. But, there is no guarantee that such an approach will yield better initial surface conditions. Ensemble averages of reanalysis products can also be potentially tested.

Acknowledgments. This work is supported by NSF under Grant CBET-1058027 by the High Meadows Sustainability Fund, and by the Mid-Infrared Technology for Health and the Environment (MIRTHE) NSF center at Princeton University, Grant EEC-0540832.

We are especially grateful to New Jersey climatologist Professor David Robinson from Rutgers University and the other members of the NJ weather and climate network for providing us access to and assistance in analyzing the data from the Trenton and New Brunswick stations and to the lidar development team at the City College of New York, especially Professors Fred Moshary and Barry Gross.

The simulations were performed on the supercomputers of the National Center for Atmospheric Research (NCAR) under Project 36861019 and the TIGRESS High Performance Computing Center of Princeton University. The eddy-covariance data collection and analysis from the Princeton campus were performed by June Yeung, Yan Zang, and Mary-Lynn Baek.

REFERENCES


