Dear Dr. Garrett,

We have addressed the comments from Reviewers 1 and 2 and incorporated changes into a revised manuscript. Please find the referees' comments, our point-by-point responses, and a marked up version of the manuscript below. The text in black are the replies uploaded to ACPD. During incorporation of the revisions, we change our minds on a few points. These changes are noted in red text, as are indications for where to find the related changes in the manuscript.

Thank you for considering this manuscript for publication in ACP.

Thanks,

Natalie Wagenbrenner

Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2015-761-RC1, 2016

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja" *by* N. S. Wagenbrenner et al.

Anonymous Referee #1

Received and published: 21 February 2016

Review

This is very well-written and well-constructed, scientifically sound paper that is of benefit to the scientific community. I don't have any major issues with the paper as-is. However, some discussion of the following issues would strengthen the paper in my opinion:

1. For Figures 1, and 6-8 it would be really helpful to have the Big Southern Butte area zoomed in more. I have a hard time seeing what is going on in that region and then the rest of the plots are mostly empty with a few observations scattered downstream. What I would recommend would be to have a zoomed-in plot of the Big Southern Butte area.

Printer-friendly version

Discussion paper

2. On the same note, it would be helpful if sensor R2, R26 discussed in Fig. 4 were indicated on the map in Fig. 1.

3. Why is the only meteorological parameter discussed in this paper wind? Is there no parallel for downscaling for other critical fire parameter: relative humidity and temperature? For the general reader some discussion of this in the intro would be helpful for the more general audience of ACP.

4. I am under the impression that strongly-forced wind events are the most important factor for high fire spread, and therefore the strengths of this study play to that need. I think this point should be made somewhere, even if it is fairly obvious.

5. WRF-HRRR fields are now available to drive the 1.33 km model. Were any tests conducted to determine the improvement in WRF using HRRR analyses to drive the 1.33 km model instead of NARR? Could this be included as potential future work?

6. I would separate the discussion more clearly into 'externally forced flows' (large-scale winds) and locally-forced flows (upslope and downslope). The current text discusses weak external forcing in many places; my opinion is that it would be better to distinguish the two cases by referring to 'locally-forced' more when the external forcing is weak.

7. Why was a 'quasi-large-eddy simulation' (200-400 m horizontal grid spacing) not conducted with WRF and compared to the other runs? This simply requires turning off the PBL scheme and a few other parameters (WRF-LES settings). A lack of numerical resources is a good argument. While the argument is made that these simulations are not feasible in real-time, the 1.33 km runs are also not feasible in real-time. If the LES run was conducted, then a comparison could be made such as 'the downscaling is equivalently good as the quasi-LES run over the area where the terrain structure is explicitly simulated. There may be issues with CFL criteria with high resolution topography in the simulations, but many investigators are starting to use WRF as a LES model with the PBL turned off. . .would be curious how these results would compare

to the downscaled. In any case, a sentence or two in future work discussing how this could be done in the future would be good.

8. The potential limitations of the slope flow parameterization and discussion of a companion paper is looking at this is discussed but in a rather scattered manner. If a summary of the strengths and weaknesses of WindNinja were included in the introduction this would help make the findings in the body of the paper flow better.

9. The weakness with windninja simulating lee-side flows is an important point, as many fires occur on the 'downslope' side of mountain ranges. Does this weakness also happen during strongly forced conditions? It was not clear to me.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2015-761, 2016.

Printer-friendly version

Discussion paper

Atmos. Chem. Phys. Discuss.,
doi:10.5194/acp-2015-761-RC2, 2016

Atmospheric Chemistry and Physics Discussions

Interactive comment on "Downscaling surface wind predictions from numerical weather prediction models in complex terrain with WindNinja" *by* N. S. Wagenbrenner et al.

Anonymous Referee #2

Received and published: 23 February 2016

The authors compare different WRF simulations and a diagnostic model against wind observations over an isolated mountain. Spacial emphasis is given to quantify the added value of downscaling the WRF simulations with the diagnostic model (WindNinja). The topic is relevant to progress in our understanding of the ability of mesoscale simulations, and diagnostic models, to reproduce the wind over complex terrain. The observational dataset is quite unique. The article is well written and the conclusions are supported by the results. Perhaps it could be highlighted that none of the WRF simulations is able to represent the mountain. Only the simulation at 1.33 km of horizontal resolution starts to "see" something (fig. 1). More specific comments, all of them of minor character, are provided below.

Printer-friendly version

Discussion paper

I therefore recommend acceptance of the manuscript pending to minor revisions

SPECIFIC COMMENTS

1. Page 5, line 9: Please define the FDM acronym.

2. Page 5, line 25: I believe you are using the Kain-Fritsch scheme to represent the cumulus in WRF. If so, please update the reference.

3. Page 6, line 4: The first model level of WRF-UWis 40m a.g.l. Later in the manuscript it is mentioned that these winds are interpolated to 3 m using a log profile. Is this interpolation taking into account the atmospheric stability? Or the atmosphere is assumed to be neutral? I say this because the target height, 3 m a.g.l, is far from 40 m a.g.l. and the interpolation method could play a major role in the evaluation.

4. Page 7, line 7: The NAM model has the first level at 200 m and the horizontal resolution is 12 km. NAM is obviously not representing the mountain at all. RAP provides information of the winds in the valley. It would be informative to have some discussion of this here pointing out why NAM is used even though the mountain is not represented at all.

5. Page 13: liens 11-15: None of the WRF simulations have enough horizontal resolution to represent the mountain. Only at 1.3 km WRF starts to see something. The WRF predictability probably comes from the simulation of the flow in the Snake River Plain. It may be interesting to compute anomalies with respect to the mean flow and compare them with the equivalent observations to see if there is any benefit as a result of increasing the horizontal resolution.

6. Page 14: It will be good to show a spatial plot showing the under prediction of windward and ridge top locations.

7. Page 18, first paragraph Summary: You should mention that the horizontal resolutions are too coarse to represent the mountain in the WRF simulations.

Interactive comment on Atmos. Chem. Phys. Discuss., doi:10.5194/acp-2015-761, 2016.

Printer-friendly version

Discussion paper

Reply to Reviewer 1

1. Figures 1 and 6-8 will be updated to show a zoomed in version of the butte. Zooming in on the butte would be nice for the left panel of Figure 1, however, it's not really possible for the other three panels, since the butte is represented by just one or two pixels; therefore, we chose to leave these figures as is. Additionally, these figures depict the domain extent used in our downscaling simulations and so there is value in leaving extents in the figures as is.

2. R2 and R26 will be added to Figure 1. Added in Figure 1, p. 34.

3. The diagnostic model evaluated in this paper, WindNinja, is only designed to downscale the flow. WindNinja includes physics for modeling the mechanical and thermal effects of the terrain on the flow field. WindNinja is capable of interpolating other parameters (e.g., temperature and relative humidity) to a finer grid, but does not provide any additional physics (e.g., conservation of energy) or parameterizations to simulate terrain effects on these parameters. For these reasons, WindNinja does not output additional downscaled weather parameters. Additionally, wind varies more spatially than temperature and RH, so is more important to predict at a high resolution. Wind is known to often be the driving environmental variable for wildfire spread and behavior. We will clarify these points in the paper. Some discussion on this was added in lines 101-104.

4. Yes, it is correct that high winds are often the most important factor for wildfire spread. This point will be incorporated into the paper. Added in lines 93-94, 384-385, 473-475.

5. HRRR-initialized 1.33 km WRF runs were not considered in this study, but could be considered in the future.

6. The discussion will be adjusted accordingly to more clearly separate the externally-forced flow and locally-forced flow discussion. After reviewing this section, we decided to leave the organization as is. We currently have sections formally separated into wind speed vs. wind direction and all data vs. diurnal/externally-forced flows. The discussion is organized by paragraph (no mixed discussion of externally-forced/externally-weak flows in a paragraph), but we didn't feel it was necessary to add another formal section heading to separate these.

7. LES was not considered for a couple of reasons. Most importantly, LES is too computationally intensive to be used in an operational context in an emergency response situation such as wildland fire. Additionally, there appear to still be many issues regarding LES in complex terrain. For example, as we understand it, WRF-LES cannot be run in complex terrain with the typical meshing algorithm employed by WRF; instead some other method, such as IBM must be used. Because of these issues, LES was not considered. However, we are working with colleagues who have substantial experience with LES that are investigating LES simulations at Big Southern Butte. We plan to make comparisons between WindNinja, the next generation WindNinja with a RANS-based solver added, and these LES simulations in the future.

8. The discussion of the slope flow parameterization will be re-worked. We will also include some background information in the introduction to set the stage for this discussion. More discussion was added in the introduction in lines 93-98.

9. Yes, the weakness in simulating lee-side recirculation occurs under high wind speeds as well. We will re-work this discussion to clarify the lee-side flow behavior and difficulty in simulating that behavior.

Reply to Reviewer 2

1. FDM will be defined. Added in line 117.

2. Yes, this reference will be added. Added in line 131.

3. The interpolation assumes neutral atmospheric stability. This information will be added in the methods. Added in lines 202-203.

4. We will include additional discussion of the terrain representation NAM and its inability to resolve the butte. Added in lines 163-166.

5. Yes, looking at the perturbations to the mean flow could be an interesting addition to our analysis. We will consider adding this in the revised manuscript. We decided not to add this at this time, but will consider this method in future evaluation work we have planned.

6. We will consider adding a spatial plot of the bias at the windward and ridgetop locations. We decided not to add this, but will consider this type of plot in our future evaluation work.

7. We will include discussion of the horizontal resolution and terrain representation in the summary. Added in lines 462-463.

Downscaling Surface Wind Predictions from Numerical Weather Prediction Models in Complex Terrain with WindNinja

N.S. Wagenbrenner[1], J.M. Forthofer[1], B.K. Lamb[2], K.S. Shannon[1], B.W. Butler[1]

[1]{US Forest Service, Rocky Mountain Research Station, Missoula Fire Sciences Laboratory, 5775 W Highway 10, Missoula, MT 59808, USA}

[2]{Laboratory for Atmospheric Research, Department of Civil and Environmental Engineering, Washington State University, Pullman, WA 99164, USA}

Correspondence to: N.S. Wagenbrenner (nwagenbrenner@fs.fed.us)

66

Abstract

Wind predictions in complex terrain are important for a number of applications. Dynamic downscaling of numerical weather prediction (NWP) model winds with a high resolution wind model is one way to obtain a wind forecast that accounts for local terrain effects, such as wind speed-up over ridges, flow channeling in valleys, flow separation around terrain obstacles, and flows induced by local surface heating and cooling. In this paper we investigate the ability of a mass-consistent wind model for downscaling near-surface wind predictions from four NWP models in complex terrain. Model predictions are compared with surface observations from a tall, isolated mountain. Downscaling improved near-surface wind forecasts under high-wind (near-neutral atmospheric stability) conditions. Results were mixed during upslope and downslope (non-neutral atmospheric stability) flow periods, although wind direction predictions generally improved with downscaling. This work constitutes evaluation of a diagnostic wind model at unprecedented high spatial resolution in terrain with topographical ruggedness approaching that of typical landscapes in the western US susceptible to wildland fire.

1. Introduction

Researchers from multiple disciplines rely on routine forecasts from numerical weather prediction (NWP) models to drive transport and dispersion models, conduct wind assessments for wind energy projects, and predict the spread of wildfires. These applications require fine-scale, near-surface wind predictions in regions where rugged terrain and vegetation have a significant effect on the local flow field. Terrain effects such as wind speed-up over ridges, flow channeling in valleys, flow separation around terrain obstacles, and enhanced surface roughness alter the flow field over spatial scales finer than those used for routine, operational NWP forecasting.

Numerous operational mesoscale NWP model forecast products are available in real-time, such as those provided by National Centers for Environmental Prediction (NCEP). Access to these output products is facilitated by automated archiving and distribution systems such as the National Operational Model Archive and Distribution System (NOMADS). These routine forecast products are highly valuable to researchers and forecasters, for example, as inputs to drive other models. In many cases, however, the spatial resolution of the system of interest (e.g., wildland fire spread) is much finer than that of the NWP model output.

The model grid horizontal resolution in operational NWP models is limited due, in part, to the high computational demands of NWP. Routine gridded forecast products are typically provided

104 at grid resolutions of 3 km or larger. The High Resolution Rapid Refresh (HRRR) model produces
105 3-km output grids and is currently the highest-resolution operational forecast in the U.S.
106
107 NWP models have been run successfully with grid resolutions of less than 1 km in complex
108 terrain for specific cases when modifications were made to the meshing (Lundquist et al. 2010)
109 or PBL schemes (Ching et al., 2014; Seaman et al., 2012) or when large-eddy simulation (LES)
110 was used (Chow and Street, 2008). While successful for specific test cases, these efforts
111 employ specialized model configurations that have not been incorporated into routine
112 forecasting frameworks, either because they are not sufficiently robust, have not been
113 thoroughly tested, or are too computationally intense for routine forecasting. For example, the
114 configuration used in Seaman et al. (2012) is applicable for stable nocturnal conditions only.
115
116 Additionally, these modifications require technical expertise in NWP and access to substantial
117 computing resources, which many consumers of NWP output do not have. Perhaps, the biggest
118 limitation to running NWP models on grids with fine horizontal resolution is the computational
119 demand. Time-sensitive applications, such as operational wildland fire support, require fast
120 solution times (e.g., less than 1 hr) on simple hardware (e.g., laptop computers with 1-2
121 processors). Thus, there remains a practical need for fast-running tools that can be used to
122 downscale coarse NWP model winds in complex terrain.
123

Dynamic downscaling with a steady-state (diagnostic) wind model is one option for obtaining near-surface high-resolution winds from routine NWP model output (e.g., Beaucage et al., 2014). The NWP model provides an initial wind field that accounts for mesoscale dynamics which is then downscaled by a higher resolution wind model to enforce conservation of mass and, in some cases, momentum and energy on the flow field on a higher resolution grid that better resolves individual terrain features. Dynamic downscaling can be done in a steady-state fashion for each time step of the NWP model output. One advantage of using a steady-state downscaling approach is that the spatial resolution can be increased with no additional computational cost associated with an increase in temporal resolution.

Diagnostic wind models have primarily been evaluated with observations collected over relatively simple, low elevation hills. Askervein Hill (Taylor and Teunissen, 1987) and Bolund Hill (Berg et al., 2011) are the two mostly commonly used datasets for evaluating diagnostic wind models. These are both geometrically simple, low-elevation hills compared to the complex terrain exhibited in many regions of the western U.S. susceptible to wildland fire. Lack of evaluations under more complex terrain is due in part to the lack of high-resolution datasets available in complex terrain. Recently, Butler et al. (2015) reported high-resolution wind observations from a tall, isolated mountain (Big Southern Butte) in the western U.S. Big Southern Butte is substantially taller and more geometrically complex than both Askervein and Bolund hills.

In this work, we investigate the ability of a mass-conserving wind model, WindNinja (Forthofer et al., 2014a), for dynamically downscaling NWP model winds over Big Southern Butte. WindNinja is a diagnostic wind model developed for operational wildland fire support. It is primarily designed to simulated mechanical effects of terrain on the flow, which are most important under high-wind conditions; however, WindNinja also contains parameterizations for local thermal effects, which are more important under periods of weak external forcing. WindNinja has primarily been evaluated under high-wind conditions, which are thought to be most important for wildland fire behavior, and so these the thermal parameterizations have not been thoroughly tested. WindNinja~~It~~ has previously been evaluated against the Askervein Hill data (Forthofer et al., 2014a) and found to capture important terrain-induced flow features, such as ridgetop speed-up, and it has been shown to improve wildfire spread predictions in complex terrain (Forthofer et al., 2014b). We focus on downscaling wind in this work because it is typically more spatially and temporally variable than temperature or relative humidity, and thus, more important to predict at high spatial resolution. Wind is also often the driving environmental variable for wildfire behavior.

The goals of this work were to (1) investigate the accuracy of NWP model near-surface wind predictions in complex terrain on spatial scales relevant for processes driven by local surface winds, such as wildland fire behavior and (2) assess the ability of a mass-consistent wind model to improve these predictions through dynamic downscaling. Wind predictions are investigated from four NWP models operated on different horizontal grid resolutions. This work constitutes

166 one of the first evaluations of a diagnostic wind model with data collected over terrain with a
167 topographical ruggedness approaching that of western U.S. landscapes susceptible to wildland
168 fire.
169
170 **2. Model descriptions and configurations**
171 WRF is a NWP model that solves the non-hydrostatic, fully compressible Navier-Stokes
172 equations using finite difference method (FDM) ~~FDM~~ discretization techniques (Skamarock et
173 al., 2008). All of the NWP models investigated in this work use either the Advanced Research
174 WRF (ARW) or the non-hydrostatic multi-scale model (NMM) core of the WRF model (Table 1).
175 *2.1. Routine Weather Research and Forecasting (WRF-UW)*
176 Routine WRF-ARW forecasts with 4 km horizontal grid resolution were acquired from the
177 University of Washington Atmospheric Sciences forecast system
178 (www.atmos.washington.edu/mm5rt/info.html). These forecasts are referred to as WRF-UW.
179 The outer domain of WRF-UW has a horizontal grid resolution of 36 km and covers most of the
180 western US and northeastern Pacific Ocean. This outer domain is initialized with NCEP Global
181 Forecast System (GFS) 1-degree runs. The 36 km grid is nested down to 12 km, 4 km, and an
182 experimental 1.33 km grid which covers a limited portion of the Pacific Northwest. The 4 km
183 grid investigated in this study covers the Pacific Northwest, including Washington, Oregon,
184 Idaho, and portions of California, Nevada, Utah, Wyoming, and Montana. Physical
185 parameterizations employed by WRF-UW include the Noah Land Surface Model (Chen et al.,

186 1996), Thompson microphysics (Thompson et al., 2004), Kain-Fritsch convective scheme ~~Kain~~
187 (2004), Rapid Radiative Transfer Model (RRTM) for longwave radiation (Mlawer et al., 1997),
188 Duhdia (1989) for shortwave radiation, and the Yonsei University (YSU) boundary layer scheme
189 (Hong et al., 2006). WRF-UW is run at 00z and 12z and generates hourly forecasts out to 84
190 hours. The computational domain consists of 38 vertical layers. The first grid layer is
191 approximately 40 m AGL and the average model top height is approximately 16000 m AGL.

192 *2.2. Weather Research and Forecasting Reanalysis (WRF-NARR)*

193 WRF-ARW reanalysis runs were performed using the NCEP North American Regional Reanalysis
194 (NARR) data (Mesinger et al., 2006). The reanalysis runs are referred to as WRF-NARR. The
195 same parameterizations and grid nesting structures used in WRF-UW were also used for the
196 WRF-NARR simulations, except that the WRF-NARR inner domain had 33 vertical layers and a
197 horizontal grid resolution of 1.33 km (Table 1). Analysis nudging (e.g., Stauffer and Seaman,
198 1994) was used above the boundary layer in the outer domain (36 km horizontal grid
199 resolution). Hourly WRF-NARR simulations were run for 15 day periods with 12 hours of model
200 spin up prior to each simulation. The first grid layer was approximately 38 m AGL and the
201 average model top height was approximately 15000 m AGL. WRF-NARR differs from the other
202 models used in this study in that it is not a routinely run model. These were custom simulations
203 conducted by our group to provide a best-case scenario for the NWP models. Routine forecasts
204 are already available for limited domains (e.g., UW provides WRF simulations on a 1.33 km grid

for a small domain in the Pacific Northwest of the US) and are likely to become more widely available at this grid resolution in the near future.

2.3. North American Mesoscale Model (NAM)

The North American Mesoscale (NAM) model is an operational forecast model run by NCEP for North America (http://www.emc.ncep.noaa.gov/index.php?branch=NAM). The NAM model uses the NMM core of the WRF model. The NAM CONUS domain investigated in this study has a horizontal grid resolution of 12 km. NAM employs the Noah Land Surface model (Chen et al., 1996), Ferrier et al. (2003) for microphysics, Kain (2004) for convection, GFDL (Lacis and Hansen, 1974) for longwave and shortwave radiation, and the Mellor-Yamada-Janjic (MJF) boundary layer scheme (Janjic, 2002). The NAM model is initialized with 12-hr runs of the NAM Data Assimilation System. It is run four times daily at 00z, 06z, 12z, and 18z and generates hourly forecasts out to 84 hours. The computational domain consists of 26 vertical layers. The first grid layer is approximately 200 m AGL and the average model top height is approximately 15000 m AGL. NAM forecasts are publicly available in real time from NCEP. Although the 12-km horizontal resolution used in NAM is not sufficient to resolve the butte, this resolution is sufficient for resolving the surrounding Snake River Plain and therefore can be used to generate a domain-average flow for input to WindNinja.

2.4. High Resolution Rapid Refresh (HRRR)

223 The High Resolution Rapid Refresh (HRRR) system is a nest inside of the NCEP-Rapid Refresh
224 (RAP) model (13 km horizontal grid resolution; http://ruc.noaa.gov/hrrr/). HRRR has a
225 horizontal grid resolution of 3 km and is updated hourly. HRRR uses the WRF model with the
226 ARW core and employs the RUC-Smirnova Land Surface Model (Smirnova et al., 1997; Smirnova
227 et al., 2000), Thompson et al. (2004) microphysics, RRTM longwave radiation (Mlawer et al.,
228 1997), Goddard shortwave radiation (Chou and Suarez, 1994), the MYJ boundary layer scheme
229 (Janjic, 2002). HRRR is initialized from 3-km grids with 3-km radar assimilation over a 1-hr
230 period. HRRR is currently the highest resolution operational forecast available in real time. The
231 computational domain consists of 51 vertical layers. The first grid layer is approximately 8 m
232 AGL and the average model top height is approximately 16000 m AGL.

233 *2.5. WindNinja*

234 WindNinja is a mass-conserving diagnostic wind model developed and maintained by the USFS
235 Missoula Fire Sciences Laboratory (Forthofer et al., 2014a). The theoretical formulation is
236 described in detail in Forthofer et al. (2014a). Here we provide a brief overview of the
237 modeling framework. WindNinja uses a variational calculus technique to minimize the change
238 in an initial wind field while conserving mass locally (within each cell) and globally over the
239 computational domain. The numerical solution is obtained using finite element method (FEM)
240 techniques on a terrain-following mesh consisting of layers of hexahedral cells that grow
241 vertically with height.

242

243 WindNinja includes a diurnal slope flow parameterization (Forthofer et al., 2009). The diurnal
244 slope flow model used in WindNinja is the shooting flow model in Mahrt (1982). It is a one-
245 dimensional model of buoyancy-driven flow along a slope. A micrometeorological model
246 similar to the one used in CALMET (Scire et al., 2000; Scire and Robe, 1997) is used to compute
247 surface heat flux, Monin-Obukhov length, and boundary layer height. The slope flow is then
248 calculated as a function of sensible heat flux, distance to ridgetop or valley bottom, slope
249 steepness, and surface and entrainment drag parameters. The slope flow is computed for each
250 grid cell and added to the initial wind in that grid cell. Additional details can be found in
251 Forthofer et al. (2009).

252

253 WindNinja was used to dynamically downscale hourly 10-m wind predictions from the above
254 NWP models. The WindNinja computational domain was constructed from 30-m resolution
255 Shuttle Radar Topography Mission (SRTM) data (Farr et al., 2007). The 10-m NWP winds were
256 bilinearly interpolated to the WindNinja computational domain and used as the initial wind
257 field. Layers above and below the 10-m height were fit to a logarithmic profile (neutral
258 atmospheric stability) based on the micrometeorological model. The computational domain
259 consisted of 20 vertical layers. The first grid layer is 1.92 m AGL and the average model top
260 height is 931 m AGL.

261 *2.6. Terrain representation*

The four NWP models used in this study employ an implementation of the WRF model. They use different initial and boundary conditions, incorporate different parameterizations for sub-grid processes, such as land surface fluxes, convection, and PBL evolution, but in terms of surface wind predictions under the conditions investigated in this study (inland, dry summertime conditions), the horizontal grid resolution is arguably the most important difference among the models. The horizontal grid resolution affects the numerical solution since fewer terrain features are resolved by coarser grids. Coarser grids essentially impart a smoothing effect which distorts the actual geometry of the underlying terrain (Fig. 1). As horizontal cell size and terrain complexity increase, the accuracy of the terrain representation and thus, the accuracy of the near-surface flow solution deteriorate.

3. Evaluations with field observations

3.1. Observations at Big Southern Butte

Surface wind data (Butler et al., 2015) collected from an isolated mountain (Big Southern Butte, hereafter 'BSB'; 43.395958, -113.02257) in southeast Idaho were used to evaluate surface wind predictions (Fig. 1). BSB is a predominantly grass-covered volcanic cinder cone with a horizontal scale of 5 km and a vertical scale of 800 m and surrounded in all directions by the relatively flat Snake River Plain. The portion of the Snake River Plain surrounding BSB slopes downward gently from the northeast to the southwest.

Three-meter wind speeds and directions were measured with cup-and-vane anemometers at 53 locations on and around BSB. The anemometers have a measurement range of 0 to 44 m s^{-1}, a resolution of 0.19 m s^{-1} and 1.4°, and are accurate to within ±0.5 m s^{-1} and ±5°. The anemometers measured wind speed and direction every second and logged 30-s averages. We averaged these 30-s winds over a 10-min period at the top of each hour (five minutes before and 5 minutes after the hour). The 10-min averaging period was chosen to correspond roughly with the time scale of wind predictions from the NWP forecasts. The NWP output is valid at a particular instant in time, but there is always some inherent temporal averaging in the predictions. The temporal averaging associated with a given prediction depends on the time-step used in the NWP model and is typically on the order of minutes. The 10-min averaged observed data are referred to in the text as 'hourly' observations (since they are averaged at the top of each hour) and are compared directly with the hourly model predictions.

Butler et al. (2015) observed the following general flow features at BSB. During periods of weak synoptic and mesoscale forcing (hereafter, referred to collectively as 'external forcing'), the observed surface winds at BSB were decoupled from the large-scale atmospheric flows, except for at high-elevation ridgetop locations. Diurnal slope flows dominated the local surface winds under periods of weak external forcing. There were frequent periods of strong external forcing, during which the diurnal slope winds on BSB were completely overtaken by the larger-scale winds. These periods of strong external forcing at BSB were typically characterized by large-

scale southwesterly flow aligned with the Snake River Plain, although occasionally there were also strong early morning winds from the northeast. Under periods of strong external forcing wind speeds commonly varied by as much as 15 m s^{-1} across the domain due to mechanical effects of the terrain (e.g., speed-up over ridges and lower speeds on leeward slopes). Additional details regarding the BSB field campaign can be found in Butler et al. (2015).

3.2. Evaluation methods

Hourly observations were compared against corresponding hourly predictions from the most recent model run. Modeled and observed winds were compared by interpolating the modeled surface wind variables to the observed surface sensor locations at each site. The 10-m winds from the NWP forecasts were interpolated to sensor locations, using bilinear interpolation in the horizontal dimension and a log profile in the vertical dimension. A 3-D interpolation scheme was used to interpolate WindNinja winds to the sensor locations. This 3-D interpolation was possible because the WindNinja domain had layers above and below the surface sensor height (3.0 m AGL). A 3-D interpolation scheme was not possible for the NWP domains since there were not any layers below the three meter surface sensor height.

Model performance was quantified in terms of the mean bias, root-mean-square error (RMSE), and standard deviation of the error (SDE):

(3)

320 $$\overline{\varphi'} = \frac{1}{N}\sum_{i=1}^{N} \varphi'$$

321 $$\mathrm{RMSE} = \left[\frac{1}{N}\sum_{i=1}^{N} (\varphi_i')^2\right]^{1/2}$$

322 $$\mathrm{SDE} = \left[\frac{1}{N-1}\sum_{i=1}^{N} (\varphi_i' - \overline{\varphi'})^2\right]^{1/2}$$

323 where φ' is the difference between simulated and observed variables and N is the number of
324 observations.

325 *3.3. Case selection*

326 We selected a five-day period from July 15-19 2010 for model evaluations. This specific period
327 was chosen because it included periods of both strong and weak external forcing, conditions
328 were consistently dry and sunny, and was a period for which we were able to acquire forecasts
329 from all NWP models selected for investigation in this study.

330

331 The observed data from the five-day period were broken into periods of upslope, downslope,
332 and externally-driven flow conditions to further investigate model performance under these
333 particular types of flow regimes. We used the partitioning schemes described in Butler et al.
334 (2015). Externally-driven events were partitioned out by screening for hours during which wind
335 speeds at a designated sensor (R2, located 5 km southwest of the butte in flat terrain) exceeded

a predetermined threshold wind speed of 6 m s^{-1}. This sensor was chosen because it was located in flat terrain far from the butte and therefore was representative of near-surface winds that were largely unaffected by the butte itself. Hours of upslope and downslope flows (i.e., observations under weak external forcing) were then partitioned out of the remaining data. Additional details regarding the partitioning scheme can be found in Butler et al. (2015). Statistical metrics were computed for these five-day periods.

We also chose one specific hour representative of each flow regime within the 5-day period to qualitatively investigate model performance for single flow events under the three flow regimes. This directly comparison of NWP model predictions, downscaled predictions, and observations for single events in order to get a visual sense for how the models performed spatially while avoiding any inadvertent complicating issues that may have arose from temporal averaging over the flow regimes.

4. Results and discussion

4.1. Overview of the five-day simulations

Fig. 2 shows observed vs. forecasted wind speeds during the five-day period. The following generalizations can be made. The NWP models predicted wind speeds below 5 m s^{-1} reasonably well on average, although HRRR tended to over predict at speeds below 3 m s^{-1} (Fig. 2). There is a lot of scatter about the regression lines, but the regressions follow the line of agreement fairly well up to observed speeds around 5 m s^{-1}. Downscaling did not improve wind

speed predictions much in this range. NWP forecast accuracy declined for observed speeds between 5 and 10 m s^{-1}, and accuracy sharply dropped off for observed speeds above 10 m s^{-1}. This is indicated by the rapid departure of the NWP model regression lines from the line of agreement (Fig 2). Downscaling improved wind speed predictions for all NWP forecasts for observed speeds greater than around 5 m s^{-1} and the biggest improvements were for observed speeds greater than 10 m s^{-1} (Fig. 2). This is indicated by the relative proximity of the downscaled regression lines to the line of agreement (Fig. 2).

Poor model accuracy at higher speeds is largely due to the models under predicting windward slope and ridgetop wind speeds. Observed speeds at these locations were often three or four times higher than speeds in other locations in the study area (e.g., note the spatial variability in Fig 3). Butler et al. (2015) showed that the highest observed speeds occurred on upper elevation windward slopes and ridgetops and the lowest observed speeds occurred on the leeward side of the butte and in sheltered side drainages on the butte itself. Downscaling with WindNinja offers improved predictions at these locations as indicated by Fig. 2 (regression lines in closer proximity to the line of agreement) and Fig. 3 (spatial variability in predictions more closely matches that of the observations).

Additionally, the downscaled NAM wind speeds were as accurate as the downscaled HRRR and WRF-UW wind speeds (Fig. 2). This indicates that the NAM forecast was able to capture the important large-scale flow features around BSB such that the additional resolution provided by HRRR and WRF-UW was not essential to resolve additional flow features in the large scale flow around BSB.

The accuracy of the NAM forecast at BSB is likely due to the fact that Snake River Plain which surrounds BSB is relatively flat and extends more than 50 km in all directions from the butte. Even a 12 km grid resolution would be capable of resolving the Snake River Plain and diurnal flow patterns within this large, gentle-relief drainage. Coarse-resolution models would not be expected to offer this same level of accuracy in areas of more extensive complex terrain, however. In areas surrounded by highly complex terrain it may be necessary to acquire NWP model output on finer grids in order to resolve the regional flow features.

The NWP forecasts predicted the overall temporal trend in wind speed (Fig. 3), but underestimated peak wind speeds due to under predictions on ridgetops and windward slopes as previously discussed, and also occasionally in the flat terrain on the Snake River Plain surrounding the butte (Fig. 4).

NWP models with coarser resolution grids predicted less spatial variability in wind speed (Fig. 3). This is because there were fewer grid cells covering the domain, and thus fewer prediction points around the butte. The spatial variability in the downscaled wind speed predictions more closely matched that of the observed data, although the highest speeds were still under predicted (Fig. 3). Although downscaling generally improved the spatial variability of the predictions, there were cases where NWP errors clearly propagated into the downscaled simulations. For example, HRRR frequently over predicted morning wind speeds associated with down-drainage flow on the Snake River Plain; this error was amplified in the downscaled simulations, especially at the ridgetop locations (e.g., Fig. 3-4, 15-17 July).

The mean bias, RMSE, and SDE for wind speed and wind direction were smaller in nearly all cases for the downscaled simulations than for the NWP forecasts during the five-day period (Table 2). Mean biases in wind speed were all slightly negative and NAM and WRF-UW had the largest mean biases. The RMSE and SDE in wind speed were largest for HRRR. Although mean bias, RMSE, and SDE in wind direction for the downscaled forecasts were smaller or equal to those for the NWP forecasts, the differences were small, with a maximum reduction in mean bias in wind direction of just 4°.

410 It is difficult to draw too many conclusions from the spatially and temporally averaged 5-day
411 statistics, however, since this period included a range of meteorological conditions (e.g., high-
412 wind events from different directions, upslope flow, downslope flow) each of which could have
413 been predicted with a different level of skill by the models. Qualitatively, however, the 5-day
414 results demonstrate that the spatial variability in the downscaled winds better matches that of
415 the observed winds at BSB (Fig. 3) and, although the reductions were small in some cases,
416 nearly all statistical metrics also improved with downscaling. The analysis is broken down by
417 flow regime in the next section for more insight into model performance.

418 *4.2. Performance under Upslope, downslope, and externally-forced flows*

419 Local solar heating and cooling was a primary driver of the flow during the slope flow regime at
420 BSB (Butler et al. 2015), with local thermal effects equal to or exceeding the local mechanical
421 effects of the terrain on the flow. Because there is weak external forcing (i.e., input wind
422 speeds to WindNinja are low), the downscaling is largely driven by the diurnal slope flow
423 parameterization in WindNinja during the slope flow regimes.

424

425 During upslope flow, the diurnal slope flow parameterization increases speeds on the windward
426 slopes and reduces speeds (or reverses flow and increases speeds, depending on the strength
427 of the slope flow relative to the prevailing flow) on lee slopes due to the opposing effects of the
428 prevailing wind and the thermal slope flow. The parameterization has the opposite effect

during downslope flow; windward slope speeds are reduced (or possibly increased if downslope flow is strong enough to reverse the prevailing flow) and lee side speeds are enhanced.

4.2.1 Wind speed

The biggest improvements in wind speed predictions from downscaling occurred during externally-driven flow events (Fig. 5). This is not surprising since the highest spatial variability in the observed wind speeds occurred during high-wind events due to mechanically-induced effects of the terrain, with higher speeds on ridges and windward slopes and lower speeds in sheltered side drainages and on the lee side of the butte (Fig. 6-8). Since WindNinja is designed primarily to simulate the mechanical effects of the terrain on the flow, it is during these high-wind events that the downscaling has the most opportunity to improve predictions across the domain. This has important implications for wildfire applications since high-wind events are often associated with increased fire behavior.

The NWP models tended to under predict wind speeds on the windward slopes, ridgetops, and surrounding flat terrain, and over predict on the lee side of the butte during high wind events (e.g., Fig. 6). The largest NWP errors in wind speed during high wind events were on the ridgetops, where speed-up occurred and the NWP under predicted speeds. These largest wind speed errors were reduced by downscaling (e.g., Fig. 6). Downscaling reduced NWP wind speed

447 errors in most regions on the butte, although the general trend of under predicting wind speeds
448 on the windward side and over predicting on the lee side did not change (e.g., Fig. 6).

449

450 There were consistent improvements in predicted wind speeds from downscaling during the
451 upslope regime, although the improvements were smaller than for the externally-driven regime
452 (Fig. 5). Wind speeds were lower during the slope flow regimes than during the externally-
453 forced regime (Fig. 6-8), and thus, smaller improvements were possible with downscaling.
454 There was some speed-up predicted on the windward side of the butte during the
455 representative upslope case which appeared to match the observed wind field (Fig. 8).

456

457 Results were mixed for the downslope regime, as wind speeds improved with downscaling for
458 WRF-UW and NAM, but not for WRF-NARR or HRRR (Fig. 5). The poor wind speed predictions
459 from HRRR during the downslope regime is partly due to the fact that HRRR tended to over
460 predict early morning winds associated with down drainage flows on the Snake River Plain.
461 These errors were amplified by the downscaling, especially at ridgetop locations (Fig. 4). In
462 reality, the high-elevation ridgetop locations tended to be decoupled from lower-level surface
463 winds during the slope flow regimes due to flow stratification. WindNinja assumes neutral
464 atmospheric stability, however, so this stratification is not handled. A parameterization for
465 non-neutral atmospheric conditions is currently being tested in Windninja.

The diurnal slope flow parameterization in WindNinja resulted in lower speeds on the windward side and higher speeds on the lee side of the butte for the representative downslope case (Fig. 7). These downscaled speeds better matched those of the observed wind field, although speeds were still under predicted for ridgetops and a few other locations around the butte (Fig. 7). The high observed speeds at the ridgetop locations are not likely due to thermal slope flow effects, but could be from the influence of gradient-level winds above the nocturnal boundary layer. These ridgetop locations are high enough in elevation (800 m above the surrounding plain) that they likely protruded out of the nocturnal boundary layer and were exposed to the decoupled gradient-level winds. Butler et al. (2015) noted that ridgetop winds did not exhibit a diurnal pattern and tended to be decoupled from winds at other locations on and around the butte. Lack of diurnal winds at the summit of the butte is also confirmed by National Oceanic and Atmospheric Administration Field Research Division (NOAA-FRD) mesonet station data collected at the top of BSB (described in Butler et al., 2015; http://www.noaa.inel.gov/projects/INLMet/INLMet.htm).

Under predictions on the lower slopes and on the plain surrounding the butte could be due to overly weak slope flows being generated by the slope flow parameterization in WindNinja (Fig. 7-8). Overly weak slope flows could be caused by a number of things: improper parameterization of surface or entrainment drag parameters, poor estimation of the depth of

the slope flow, or deficiencies in the micrometeorological model used. The slope flow parameterization is being evaluated in a companion paper.

4.2.2 Wind direction

The biggest improvement in wind direction predictions from downscaling occurred during the downslope regime (Fig. 5). Wind direction improved with downscaling for all NWP models during periods of downslope flow. This indicates that the diurnal slope flow model helped to orient winds downslope. This is confirmed by inspection of the vector plots for the representative downslope case which show the downscaled winds oriented downslope on the southwest and northeast faces of the butte (Fig. 7). Downscaling reduced speeds on the northwest (windward) side of the butte, but did not predict strong enough downslope flow in this region to reverse the flow from the prevailing northwest direction (Fig. 7). This again suggests that perhaps the diurnal slope flow algorithm is predicting overly weak slope flows.

Wind direction predictions during the upslope regime also improved with downscaling for all NWP models except HRRR (Fig. 5). Downscaled winds for the representative upslope case were oriented upslope on the southwest (lee side) of the butte and matched the observed winds in this region well (Fig. 8). This is an improvement over the NWP wind directions on the lee side of the butte.

There was no improvement in wind direction predictions with downscaling during the externally-driven regime (Fig. 5). Looking at the vector plots during the representative externally-driven event (Fig. 6), it is clear why this would be. The representative event was a high-wind event from the southwest. Wind directions are well predicted on the windward side of the butte, but not on the leeward side, where the observed field indicates some recirculation in the flow field (Fig. 6). The prevailing southwesterly flow is captured by the NWP model, but the lee side recirculation is not. WindNinja does not predict the lee side recirculation, and thus, the downscaling does not improve directions on the lee side of the butte (Fig. 7). This is an expected result, as WindNinja has been shown to have difficulties simulating flows on the lee side of terrain features due to the fact that it does not account for conservation of momentum in the flow solution (Forthofer et al., 2014a).

5. Summary

The horizontal grid resolutions of NWP models investigated in this study were too coarse to resolve the BSB terrain. Results showed that the NWP models captured the important large-scale flow features around BSB under most conditions, but were not capable of predicting the high spatial variability (scale of 100s of meters) in the observed winds on and around the butte induced by mechanical effects of the terrain and local surface heating and cooling. Thus, surface winds from the NWP models investigated in this study would not be sufficient for forecasting wind speeds on and around the butte at the spatial scales relevant for processes driven by local surface winds, such as wildland fire spread, for example.

525

526 Wind predictions generally improved for all NWP models by downscaling with WindNinja. The
527 biggest improvements occurred under high-wind events (near-neutral atmospheric stability)
528 when observed wind speeds were greater than 10 m s^{-1}. This finding has important
529 implications for fire applications since increased wildfire behavior is often associated with high
530 winds. Downscaled NAM wind speeds were as accurate as downscaled WRF-UW and HRRR
531 wind speeds, indicating that a NWP model with 12 km grid resolution was sufficient for
532 capturing the large-scale flow features around BSB.

533

534 WindNinja did not predict the observed lee-side flow recirculation at BSB that occurred during
535 externally-forced high wind events. Previous work has shown that WindNinja has difficulties
536 simulating lee-side flows (Forthofer et al., 2014a). This is partly due to lack of a momentum
537 equation in the WindNinja flow solution as discussed in Forthofer et al. (2014a). Work is
538 currently underway to incorporate an optional momentum solver in WindNinja which is
539 anticipated to improve flow predictions on the lee-side of terrain obstacles.

540

541 Results indicated that WindNinja predicted overly weak slope flows compared to observations.
542 Weak slope flow could be caused by several different issues within the diurnal slope flow
543 parameterization in WindNinja: improper parameterization of surface or entrainment drag

544 parameters, poor estimation of the depth of the slope flow, or deficiencies in the
545 micrometeorological model. These issues will be explored in future work.

546

547 This work constitutes evaluation of a diagnostic wind model at unprecedented high spatial
548 resolution and terrain complexity. While extensive evaluations have been performed with data
549 collected in less rugged terrain (e.g., Askervein Hill and Bolund Hill, relatively low elevation hills
550 with simple geometry), to our knowledge, this study is the first to evaluate a diagnostic wind
551 model with data collected in terrain with topographical ruggedness approaching that of typical
552 landscapes in the western US susceptible to wildland fire. This work demonstrates that NWP
553 model wind forecasts can be improved in complex terrain, at least in some cases, through
554 dynamic downscaling via a mass-conserving wind model. These improvements should
555 propagate on to more realistic predictions from other model applications which are sensitive to
556 surface wind fields, such as wildland fire behavior, local-scale transport and dispersion, and
557 wind energy applications.

558 **Acknowledgements**

559 Thanks to Dave Ovens and Cliff Mass of the University of Washington for providing access to
560 the WRF-UW simulations and Eric James of NOAA–GSD Earth System Research Laboratory for
561 access to the HRRR simulations. Thanks to Serena Chung of the Laboratory for Atmospheric
562 Research, Washington State University, for guidance on the WRF-NARR simulations. We also

563 thank the participants in the BSB and SRC field campaigns, including Dennis Finn, Dan Jimenez,
564 Paul Sopko, Mark Vosburgh, Larry Bradshaw, Cyle Wold, Jack Kautz, and Randy Pryhorocky.

565

References

Beaucage, P., Brower, M.C., and Tensen, J.: Evaluation of four numerical wind flow models for wind resource mapping, Wind Energy, 17, 197–208, 2014.

Berg, J., Mann, J., Bechmann, A., Courtney, M.S., and Jorgensen, H.E.: The Bolund experiment, Part I: flow over a steep, three-dimensional hill, Bound.-Lay. Meteorol., 141, 219–243, 2011.

Butler, B.W., Wagenbrenner, N.S., Forthofer, J.M., Lamb, B.K., Shannon, K.S., Finn, D., Eckman, R.M., Clawson, K., Bradshaw, L., Sopko, P., Beard, S., Jimenez, D., Wold, C., and Vosburgh, M.: High-resolution observations of the near-surface wind field over an isolated mountain and in a steep river canyon, Atmos. Chem. Phys., 15, 3785–3801, 2015.

Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H., Koren, V., Duan, Y., Ek, M., and Betts, A.: Modeling of land-surface evaporation by four schemes and comparison with FIFE observations, J. Geophys. Res., 101, 7251–7268, 1996.

Ching, J., Rotunno, R., LeMone, M., Martilli, A., Kosovic, B., Jimenez, P.A., and Dudhia, J.: Convectively induced secondary circulations in fine-grid mesoscale numerical weather prediction models, Mon. Wea. Rev., 142, 3284–3302, 2014.

Chou, M., and Suarez, M.J.: An efficient thermal infrared radiation parameterization for use in general circulation models, Technical Report Series on Global Modeling and Data Assimilation, NASA Tech. Memo. 104606, 3, 85 pp, 1994.

584 Chow, F.K., and Street, R.L.: Evaluation of turbulence closure models for large-eddy simulation
585 over complex terrain: Flow over Askervein Hill, J. Appl. Meteor. Climatol., 48, 1050–1065,
586 2008.

587 Dudhia, J.: Numerical study of convection observed during the winter monsoon experiment
588 using a mesoscale two-dimensional model, J. Atmos. Sci., 46, 3077–3107, 1989.

589 Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M.,
590 Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M.,
591 Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, Reviews of Geophysics,
592 45, RG2004, doi: 10.1029/2005RG000183, 2007.

593 Ferrier, B., Lin, Y., Parrish, D., Pondeca, M., Rogers, E., Manikin, G., Ek., M., Hart, M., DiMego,
594 G., Mitchell, K., Chuang, H.-Y.: Changes to the NCEP Meso Eta analysis and forecast system:
595 Modified cloud microphysics, assimilation of GOES cloud-top pressure, assimilation of
596 NEXRAD 88D radial wind velocity data, NWS Technical Procedures Bulletin, 2003.

597 Forthofer, J., Shannon, K., and Butler, B.: Simulating diurnally driven slope winds with
598 WindNinja, Eighth Symposium on Fire and Forest Meteorology. Oct 13-15. Kalispell, MT,
599 2009.

Forthofer, J.M., Butler, B.W., and Wagenbrenner, N.S.: A comparison of three approaches for simulating fine-scale winds in support of wildland fire management: Part I. Model formulation and accuracy, Int. J. Wildland Fire, 23, 969–981, 2014a.

Forthofer, J.M., Butler, B.W., McHugh, C.W., Finney, M.A., Bradshaw, L.S., Stratton, R.D, Shannon, K.S., and Wagenbrenner, N.S.: A comparison of three approaches for simulating fine-scale surface winds in support of wildland fire management. Part II. An exploratory study of the effect of simulated winds on fire growth simulations, Int. J. Wildland Fire, 23, 982–994, 2014b.

Hong, S.-Y., Noh, Y., and Dudhia, J.: A new vertical diffusion package with an explicit treatment of entrainment, Mon. Wea. Rev., 134, 2318–2341, 2006.

Janjic, Z.: Nonsingular implementation of the Mellor-Yamada level 2.5 scheme in the NCEP meso model, NCEP Office Note No. 437, 60, 2002.

Kain, J.: The Kain-Fritsch convective parameterization: An update, J. Meteor. Climatol., 43, 170–181, 2004.

Lacis, A.A., and Hansen, J.E., 1974: A parameterization for the absorption of solar radiation in the earth's atmosphere, J. Atmos. Sci., 31, 118–133, 1974.

Lundquist, K.A., Chow, F.K., and Lundquist, J.K.: An immersed boundary method for the Weather Research and Forecasting Model, Mon. Wea. Rev., 138, 796–817, 2010.

618 Mahrt, L.: Momentum balance of gravity flows, J. Atmos. Sci., 39, 2701–2711, 1982.

619 Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jovic, D., Woollen,
620 J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y.,
621 Manikin, G., Parrish, D., and Shi, W.: North American regional reanalysis, Bull. Amer.
622 Meteor. Soc., 87, 343–360, 2006.

623 Mlawer, E.J., Taubman, S.J., Brown, P.D., Iacono, M.J., and Clough, S.A.: Radiative transfer for
624 inhomogenous atmospheres: RRTM, a validated correlated-k model for the longwave, J.
625 Geophys. Res., 102, 16663–16682, 1997.

626 Scire, J.S., Robe, F.R., Fernau, M.E., and Yamartino, R.J.: A user's guide for the CALMET
627 meteorological model, Earth Tech, Inc., Concord. MA. Available online at
628 src.com/calpuff/download/CALMET_UsersGuide.pdf, 2000.

629 Scire, J.S., and Robe, F.R.: Fine-scale application of the CALMET meteorological model to a
630 complex terrain site, Air & Waste Management Associations's 90th Annual Meeting &
631 Exhibition, Toronto, Ontario, Canada, 16 pp, 1997.

632 Seaman, N.L., Gaudet, B.J., Stauffer, D.R., Mahrt, L., Richardson, S.J., Zielonka, J.R., and
633 Wyngaard, J.C.: Numerical prediction of submesoscale flow in the nocturnal stable
634 boundary layer over complex terrain, Month. Wea. Rev., 140, 956–977, 2012.

635 Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Duda, M.G., Huang, X., Wang,
636 W., and Powers, J.G.: A description of the advanced research WRF version 3. NCAR Tech.
637 Note NCAR/TN-475STR, 2008.

638 Smirnova, T.G., Brown, J.M, and Benjamin, S.J.: Performance of different soil model
639 configurations in simulating ground surface temperature and surface fluxes, Mon. Wea.
640 Rev., 125, 1870–1884, 1997.

641 Smirnova, T.G, Brown, J.M., Benjamin, S.G., and Kim, D.: Parameterization of cold season
642 processes in the MAPS land-surface scheme, J. Geophys. Res., 105, 4077–4086, 2000.

643 Stauffer, D.R., and Seaman, N.L.: Multiscale four-dimensional data assimilation, J. Appl. Meteor.
644 33, 416–434, 1994.

645 Taylor, P.A., and Teunissen, H.W.: The Askervein Hill project: Overview and background data,
646 Bound.-Lay. Meteorol., 39, 15–39. 1987.

647 Thompson, G.R., Rasmussen, R.M., and Manning, K.: Explicit forecasts of winter precipitation
648 using an improved bulk microphysics scheme. Part I: description and sensitivity analysis,
649 Mon. Wea. Rev., 132, 519–542, 2004.
650

651 **Tables**

652 Table 1. Model specifications.

Model	Horizontal grid resolution	Number vertical layers	First layer height[a] (m AGL)	Top height[a] (m AGL)	Numerical core	Run frequency
NAM	12 km	26	200	15000	NMM	00z, 06z, 12z, 18z
WRF-UW	4 km	38	40	16000	ARW	00z, 12z
HRRR	3 km	51	8	16000	ARW	hourly
WRF-NARR	1.33 km	33	38	15000	ARW	NA
WindNinja	138 m	20	1.92	931	NA	NA

653 [a]Approximate average height AGL.

654

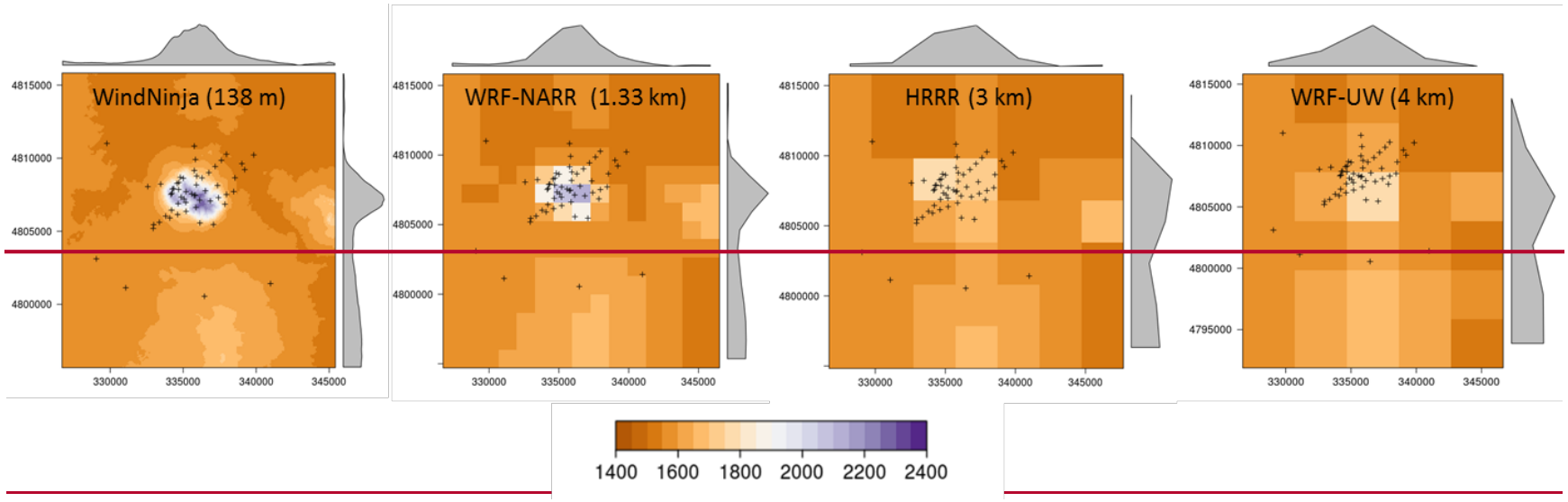
Table 2. Model mean bias, root-mean-square error (RMSE), and standard deviation of errors (SDE) for surface wind speeds and directions during the 5-day evaluation period at Big Southern Butte. Downscaled values are in parentheses. Smaller values are in bold. The 5-day period includes the Downslope, Upslope, and Externally-driven time periods.

Time period	Statistic	NAM	WRF-UW	HRRR	WRF-NARR
		Wind Speed (m s^{-1})			
5-day	Bias	-0.84 (**-0.67**)	-1.17 (**-0.95**)	-0.40 (**-0.14**)	-0.31 (**-0.08**)
	RMSE	2.31 (**2.04**)	2.39 (**2.07**)	2.52 (**2.47**)	2.33(**2.21**)
	SDE	2.15 (**1.92**)	2.08 (**1.83**)	2.49 (**2.47**)	2.31 (**2.21**)
Downslope	Bias	-1.07 (**-0.76**)	-1.15 (**-0.74**)	**-0.09** (0.48)	-0.48 (**0.12**)
	RMSE	2.08 (**1.92**)	2.03 (**1.83**)	**2.36** (2.66)	**2.19** (2.28)
	SDE	1.79 (**1.77**)	**1.67** (1.68)	**2.36** (2.62)	**2.14** (2.28)
Upslope	Bias	-0.81 (**-0.74**)	-1.11 (**-0.98**)	-0.81 (**-0.75**)	0.06 (**0.05**)
	RMSE	1.73 (**1.62**)	2.02 (**1.86**)	1.93 (**1.81**)	1.86 (1.86)
	SDE	1.52 (**1.44**)	1.69 (**1.58**)	1.76 (**1.64**)	1.86 (1.86)
Externally-driven	Bias	**-0.57** (-0.62)	**-1.28** (-1.32)	**-0.94** (-1.03)	**-0.22** (-0.33)
	RMSE	3.06 (**2.48**)	3.21 (**2.58**)	3.17 (**2.59**)	2.92 (**2.39**)
	SDE	3.00 (**2.40**)	2.94 (**2.22**)	3.02 (**2.38**)	2.92 (**2.37**)
		Wind Direction (°)			
5-day	Bias	59 (**56**)	57 (**53**)	64 (**60**)	57 (**54**)
	RMSE	76 (**72**)	74 (**71**)	80 (**76**)	73 (**71**)
	SDE	47 (**46**)	47 (**46**)	47 (**46**)	46 (46)
Downslope	Bias	67 (**60**)	61 (**56**)	76 (**67**)	66 (**61**)
	RMSE	83 (**77**)	78 (**72**)	88 (**81**)	81 (**75**)
	SDE	49 (**47**)	48 (**46**)	46 (46)	47 (**45**)
Upslope	Bias	55 (**52**)	58 (**54**)	56 (56)	52 (**49**)
	RMSE	70 (**67**)	74 (**71**)	72 (72)	68 (**65**)
	SDE	44 (**42**)	46 (**45**)	45 (46)	44 (**42**)
Externally-driven	Bias	**48** (49)	**45** (46)	51 (**50**)	44 (**46**)

	RMSE	**64** (65)	**63** (65)	68 (**67**)	**62** (65)
	SDE	**43** (44)	**44** (47)	45 (**44**)	**43** (46)

658

Figures


WindNinja (138 m)
WRF-NARR (1.33 km)
HRRR (3 km)
WRF-UW (4 km)
4815000
4810000
4805000
4800000
4795000
330000
335000
340000
345000
1400
1600
1800
2000
2200
2400

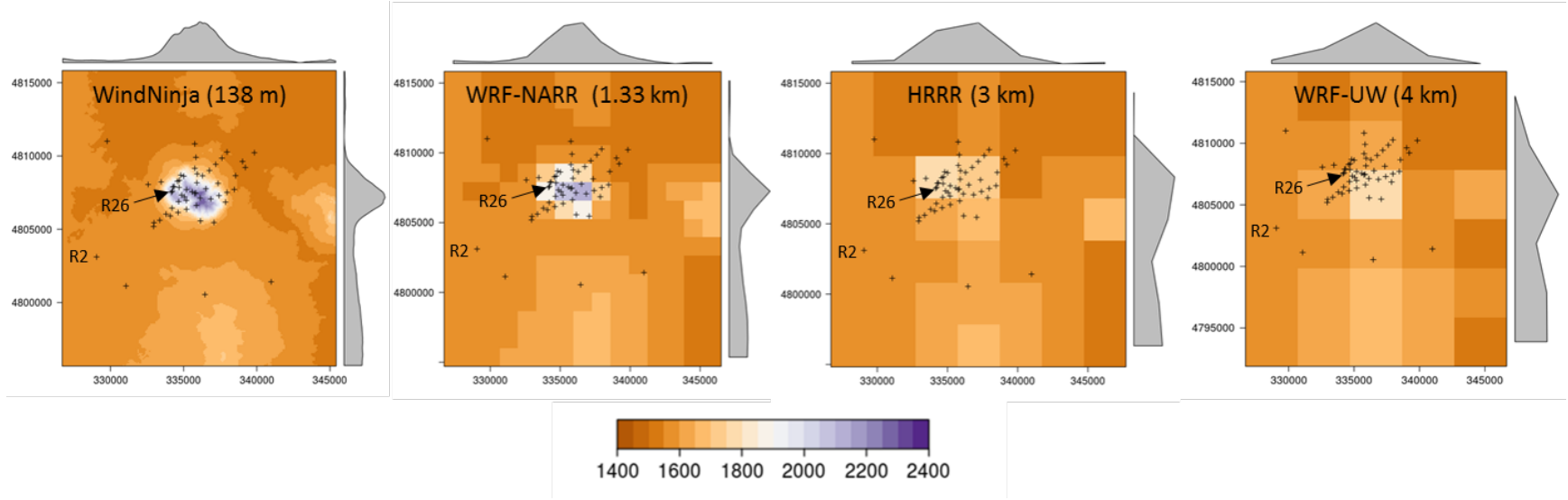
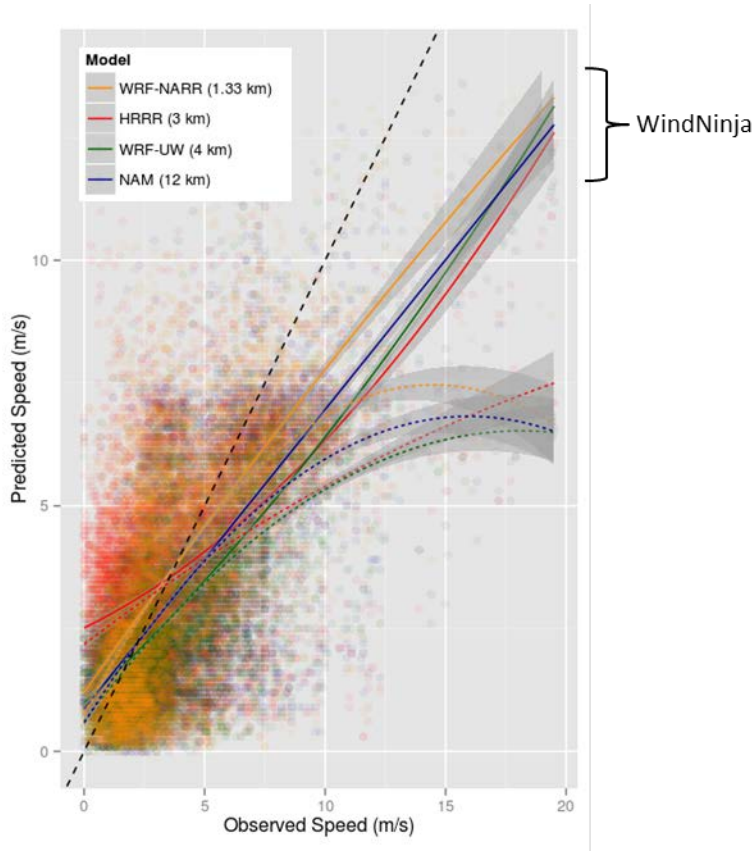

WindNinja (138 m)
WRF-NARR (1.33 km)
HRRR (3 km)
WRF-UW (4 km)
R26
R2
4815000
4810000
4805000
4800000
4795000
330000
335000
340000
345000
1400
1600
1800
2000
2200
2400

Figure 1. Terrain representation (m ASL) in WindNinja, WRF-NARR, HRRR, and WRF-UW for the Big Southern Butte. Crosses indicate surface sensor locations. Maps are projected in the Universal Transverse Mercator (UTM) zone 12 coordinate system. Axis labels are eastings and northings in m. Profiles in gray are the average elevations for rows and columns in the panel. NAM (12 km) terrain is represented by just four cells and is not shown here.

668



669

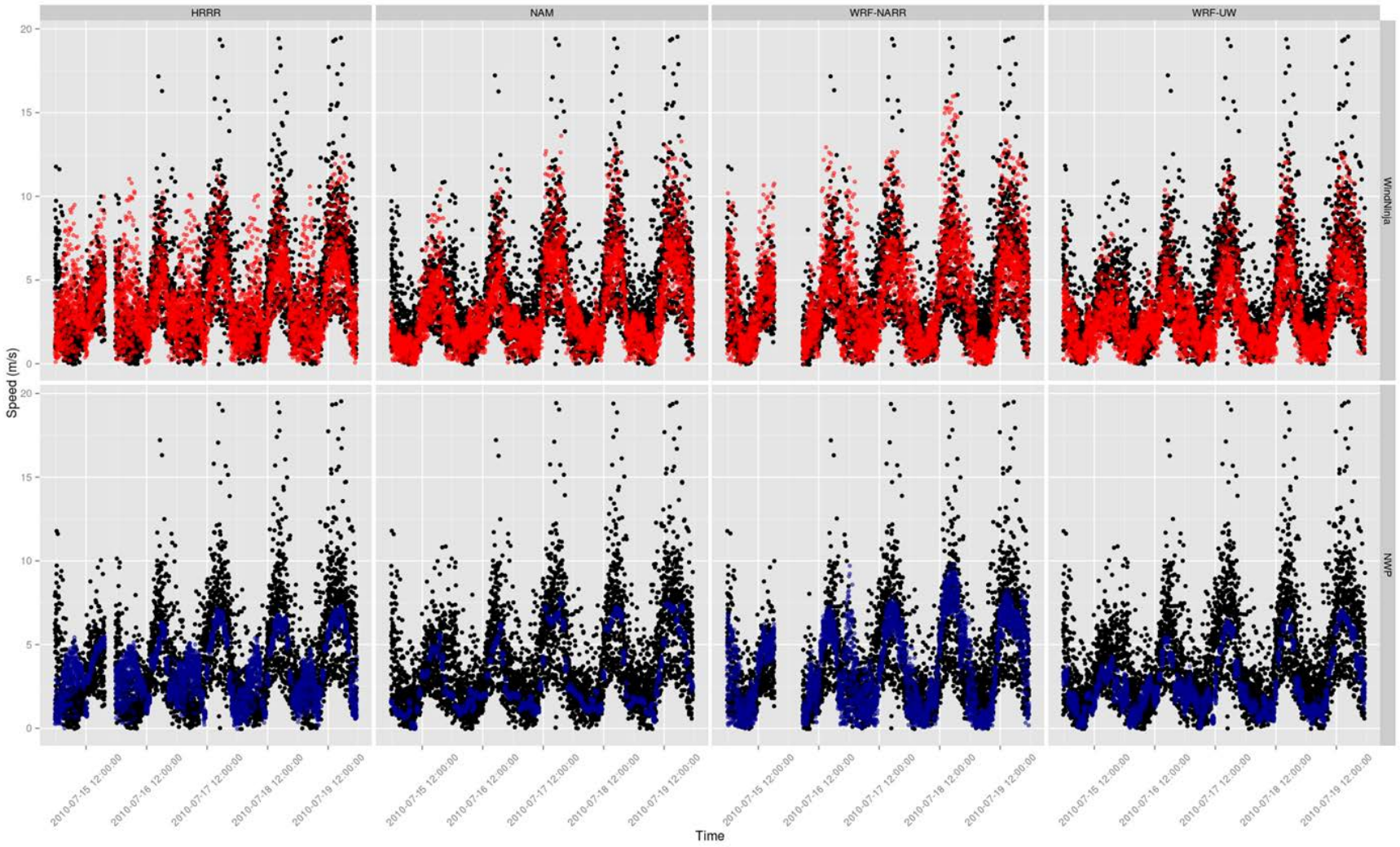
670 Figure 2. Observed vs. predicted wind speeds for the 5-day evaluation period at Big Southern

671 Butte. Dashed black line is the line of agreement. Colored lines are linear regressions

672 (quadratic fit); dashed lines are NWP models and solid lines are NWP forecasts downscaled with

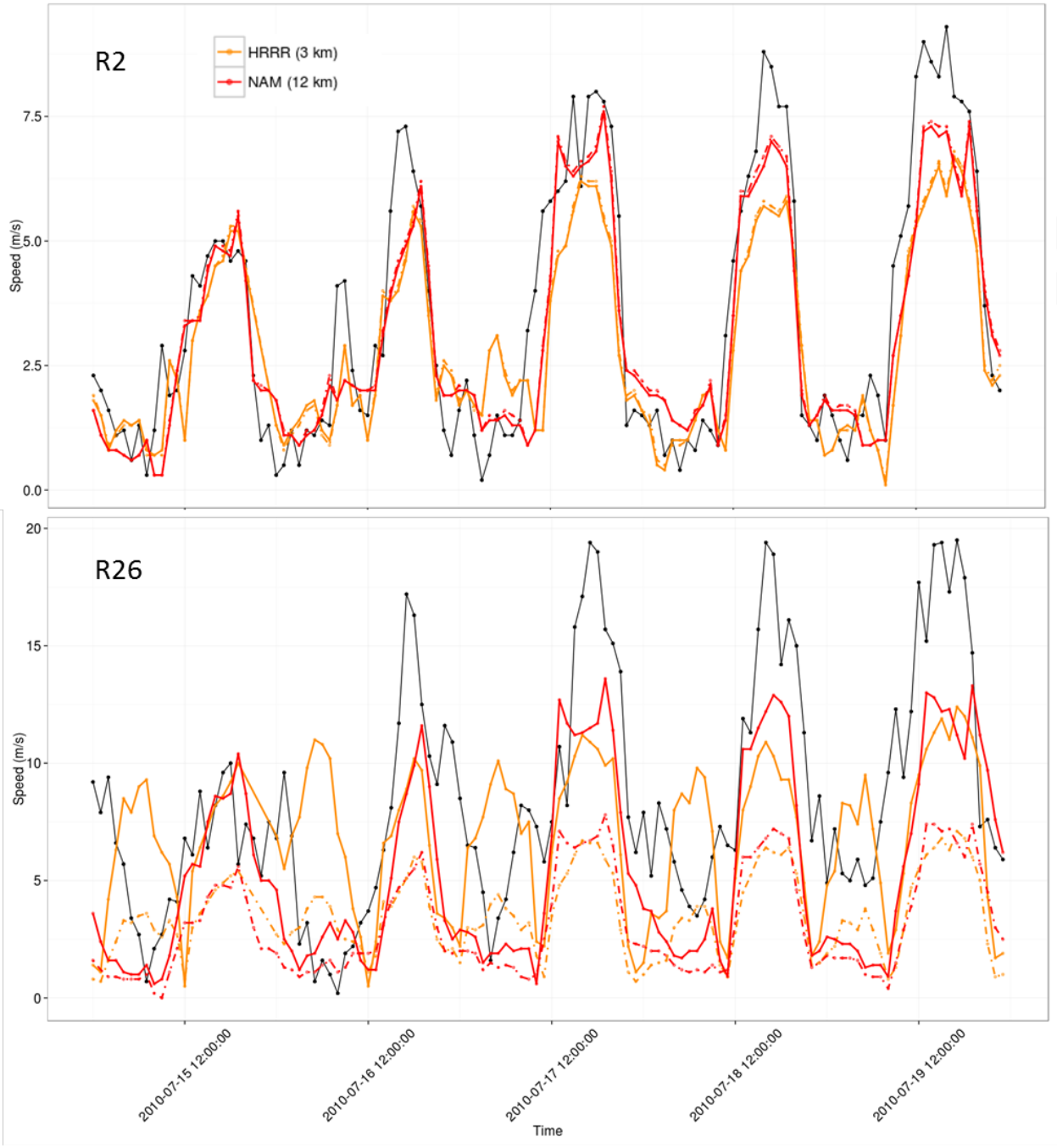
673 WindNinja. Shading indicates 95% confidence intervals.

674



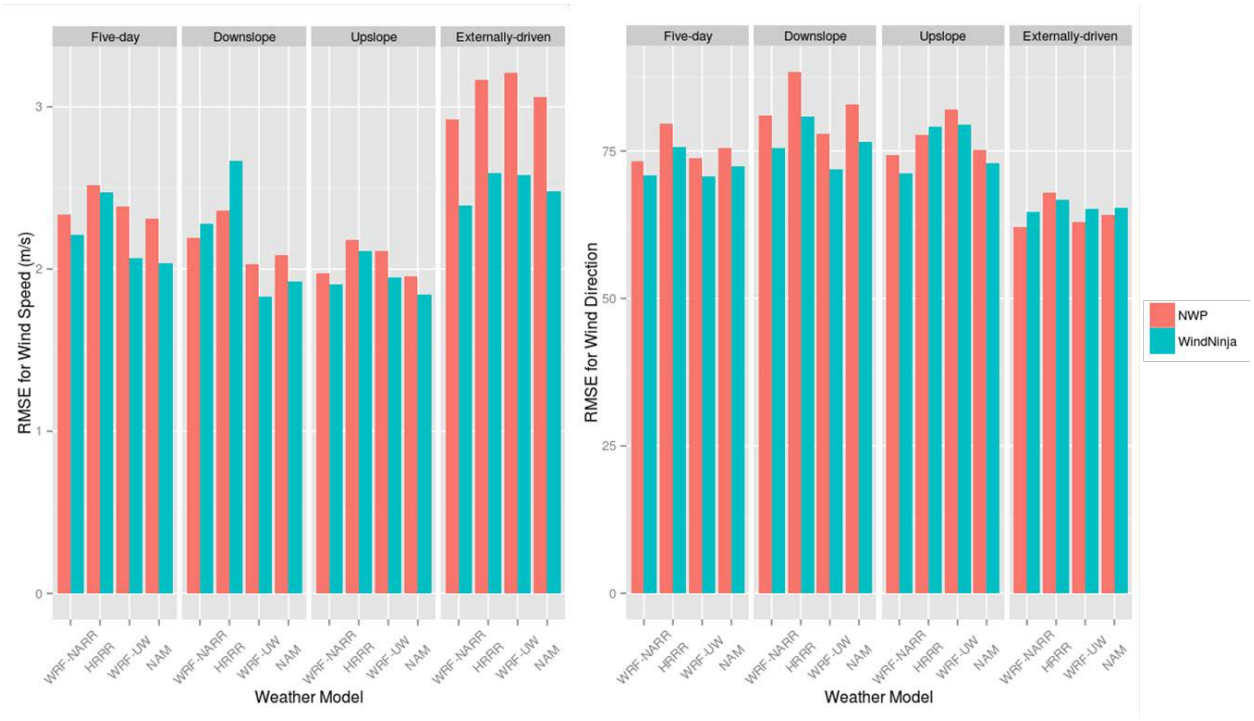
675
676 Figure 3. Observed (black) and predicted (colored) winds speeds at all sensors for 15 July 2010–19 July 2010 at Big Southern Butte.

677 Top panels are WindNinja predictions. Bottom panels are NWP predictions.



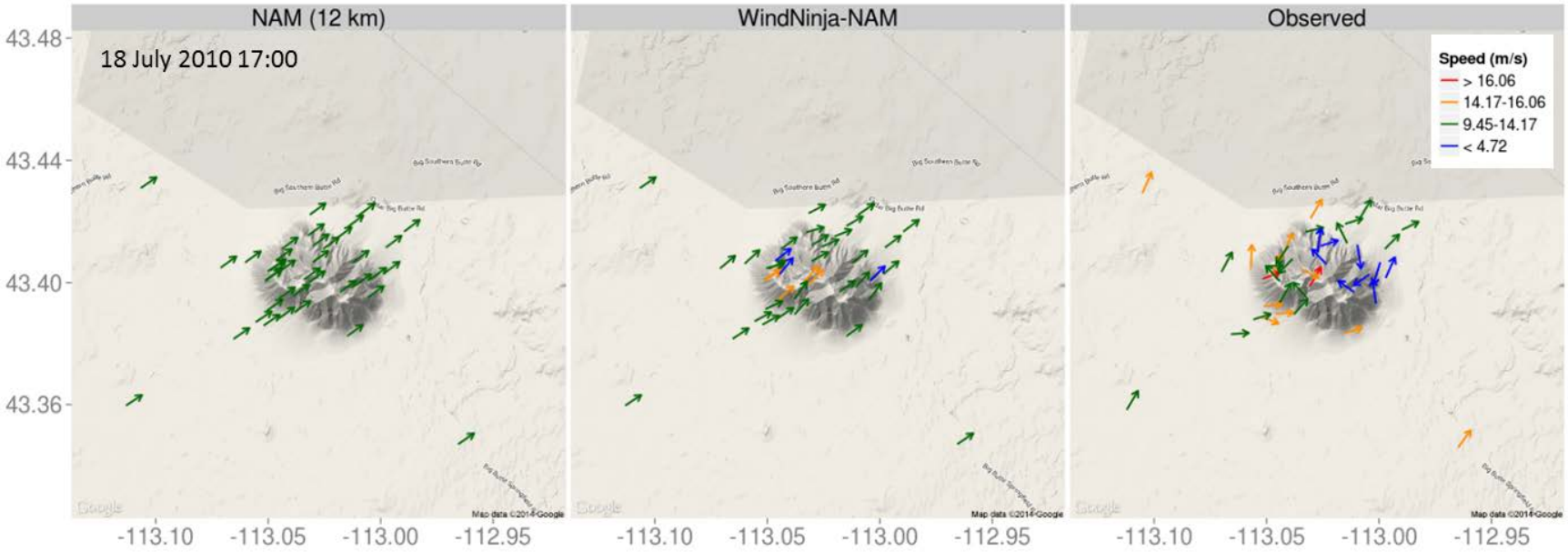
678

679 Figure 4. Observed (black line) and predicted (colored lines) wind speeds for sensor R2 located
680 5 km southwest of Big Southern Butte on the Snake River Plain and sensor R26 located on a
681 ridgetop. Dashed colored lines are NWP models and solid colored lines are WindNinja.



682

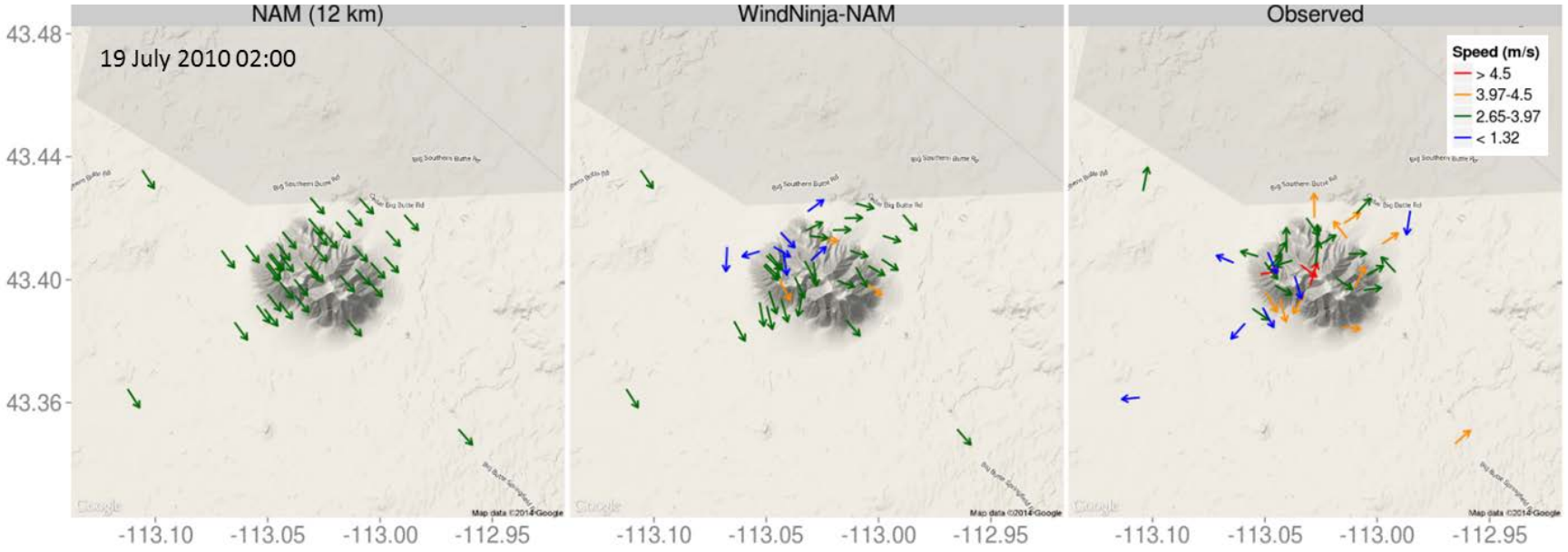
683 Figure 5. Root-mean-square error in wind speed (left) and wind direction (right) at Big Southern
684 Butte for the five-day evaluation period (*N* = 4149), and downslope (*N* = 1593), upslope (*N* =
685 717), and externally -driven (*N* = 966) periods within the five-day period. Sample size, *N* =
686 number of hours x number of sensor locations.



687

688 Figure 6. Predicted and observed winds for an externally-forced flow event at Big Southern Butte.

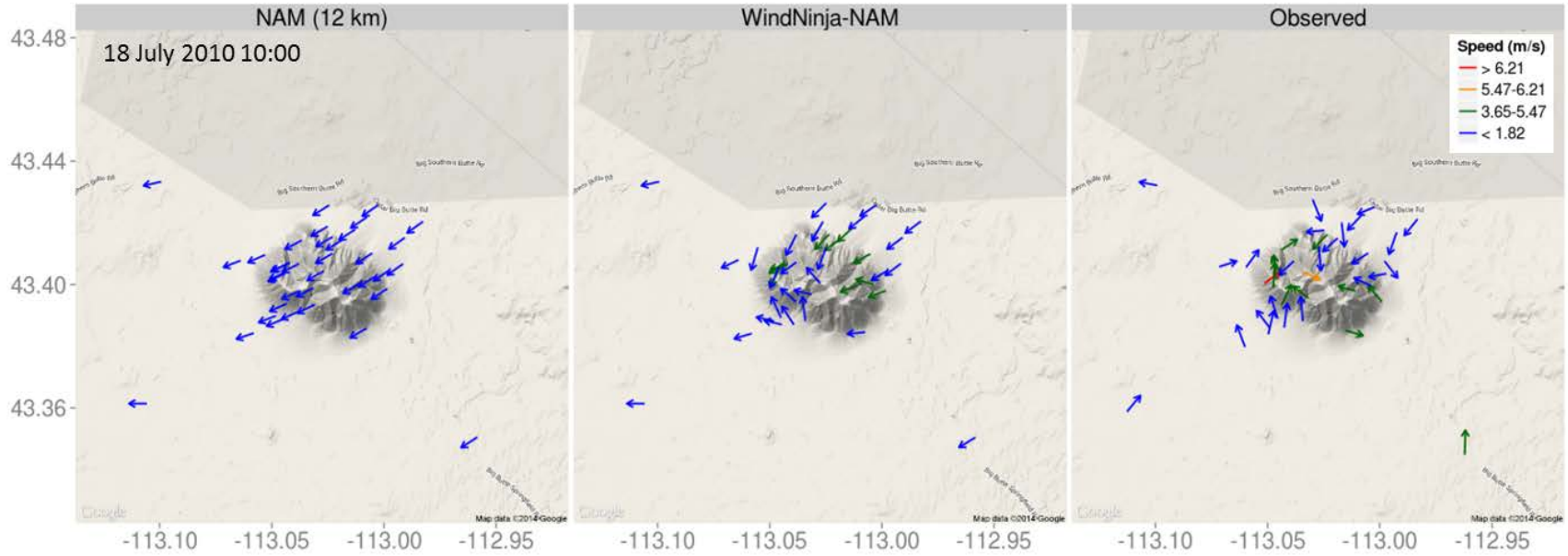
689



690

691 Figure 7. Predicted and observed winds for a downslope flow event at Big Southern Butte.

692



693

694 Figure 8. Predicted and observed winds for an upslope flow event at Big Southern Butte.

695