

1 Dear Dr. Garrett,

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

6

7 Thank you for considering this manuscript for publication in ACP.

8

9 Thanks,

10 Natalie Wagenbrenner

11

12 Reply to Reviewer 1

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

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

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

28 4. Yes, it is correct that high winds are often the most important factor for wildfire spread. This point will
29 be incorporated into the paper. **Added in lines 93-94, 384-398, 473-475.**

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

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

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

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

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

51

52 Reply to Reviewer 2

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

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

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

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

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

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

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

66

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

69

70 **N.S. Wagenbrenner¹, J.M. Forthofer¹, B.K. Lamb², K.S. Shannon¹, B.W. Butler¹**

71 [1]{US Forest Service, Rocky Mountain Research Station, Missoula Fire Sciences Laboratory,
72 5775 W Highway 10, Missoula, MT 59808, USA}

73 [2]{Laboratory for Atmospheric Research, Department of Civil and Environmental Engineering,
74 Washington State University, Pullman, WA 99164, USA}

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

76

77

78 **Abstract**

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

93

94

95 **1. Introduction**

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

104

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

112

113 The model grid horizontal resolution in operational NWP models is limited due, in part, to the
114 high computational demands of NWP. Routine gridded forecast products are typically provided
115 at grid resolutions of 3 km or larger. The High Resolution Rapid Refresh (HRRR) model produces
116 3-km output grids and is currently the highest-resolution operational forecast in the U.S.

117

118 NWP models have been run successfully with grid resolutions of less than 1 km in complex
119 terrain for specific cases when modifications were made to the meshing (Lundquist et al. 2010)
120 or PBL schemes (Ching et al., 2014; Seaman et al., 2012) or when large-eddy simulation (LES)
121 was used (Chow and Street, 2008). While successful for specific test cases, these efforts
122 employ specialized model configurations that have not been incorporated into routine
123 forecasting frameworks, either because they are not sufficiently robust, have not been
124 thoroughly tested, or are too computationally intense for routine forecasting. For example, the
125 configuration used in Seaman et al. (2012) is applicable for stable nocturnal conditions only.

126

127 Additionally, these modifications require technical expertise in NWP and access to substantial
128 computing resources, which many consumers of NWP output do not have. Perhaps, the biggest
129 limitation to running NWP models on grids with fine horizontal resolution is the computational
130 demand. Time-sensitive applications, such as operational wildland fire support, require fast
131 solution times (e.g., less than 1 hr) on simple hardware (e.g., laptop computers with 1-2
132 processors). Thus, there remains a practical need for fast-running tools that can be used to
133 downscale coarse NWP model winds in complex terrain.

134

135 Dynamic downscaling with a steady-state (diagnostic) wind model is one option for obtaining
136 near-surface high-resolution winds from routine NWP model output (e.g., Beaucage et al.,
137 2014). The NWP model provides an initial wind field that accounts for mesoscale dynamics
138 which is then downscaled by a higher resolution wind model to enforce conservation of mass

139 and, in some cases, momentum and energy on the flow field on a higher resolution grid that
140 better resolves individual terrain features. Dynamic downscaling can be done in a steady-state
141 fashion for each time step of the NWP model output. One advantage of using a steady-state
142 downscaling approach is that the spatial resolution can be increased with no additional
143 computational cost associated with an increase in temporal resolution.

144
145 Diagnostic wind models have primarily been evaluated with observations collected over
146 relatively simple, low elevation hills. Askervein Hill (Taylor and Teunissen, 1987) and Bolund Hill
147 (Berg et al., 2011) are the two mostly commonly used datasets for evaluating diagnostic wind
148 models. These are both geometrically simple, low-elevation hills compared to the complex
149 terrain exhibited in many regions of the western U.S. susceptible to wildland fire. Lack of
150 evaluations under more complex terrain is due in part to the lack of high-resolution datasets
151 available in complex terrain. Recently, Butler et al. (2015) reported high-resolution wind
152 observations from a tall, isolated mountain (Big Southern Butte) in the western U.S. Big
153 Southern Butte is substantially taller and more geometrically complex than both Askervein and
154 Bolund hills.

155
156 In this work, we investigate the ability of a mass-conserving wind model, WindNinja (Forthofer
157 et al., 2014a), for dynamically downscaling NWP model winds over Big Southern Butte.
158 WindNinja is a diagnostic wind model developed for operational wildland fire support. It is
159 primarily designed to simulated mechanical effects of terrain on the flow, which are most
160 important under high-wind conditions; however, WindNinja also contains parameterizations for

161 local thermal effects, which are more important under periods of weak external forcing.
162 WindNinja has primarily been evaluated under high-wind conditions, which are thought to be
163 most important for wildland fire behavior, and so these the thermal parameterizations have not
164 been thoroughly tested. WindNinja has previously been evaluated against the Askervein Hill
165 data (Forthofer et al., 2014a) and found to capture important terrain-induced flow features,
166 such as ridgetop speed-up, and it has been shown to improve wildfire spread predictions in
167 complex terrain (Forthofer et al., 2014b). We focus on downscaling wind in this work because it
168 is typically more spatially and temporally variable than temperature or relative humidity, and
169 thus, more important to predict at high spatial resolution. Wind is also often the driving
170 environmental variable for wildfire behavior.

171
172 The goals of this work were to (1) investigate the accuracy of NWP model near-surface wind
173 predictions in complex terrain on spatial scales relevant for processes driven by local surface
174 winds, such as wildland fire behavior and (2) assess the ability of a mass-consistent wind model
175 to improve these predictions through dynamic downscaling. Wind predictions are investigated
176 from four NWP models operated on different horizontal grid resolutions. This work constitutes
177 one of the first evaluations of a diagnostic wind model with data collected over terrain with a
178 topographical ruggedness approaching that of western U.S. landscapes susceptible to wildland
179 fire.

180

181 **2. Model descriptions and configurations**

182 WRF is a NWP model that solves the non-hydrostatic, fully compressible Navier-Stokes
183 equations using finite difference method (FDM) discretization techniques (Skamarock et al.,
184 2008). All of the NWP models investigated in this work use either the Advanced Research WRF
185 (ARW) or the non-hydrostatic multi-scale model (NMM) core of the WRF model (Table 1).

186 *2.1. Routine Weather Research and Forecasting (WRF-UW)*

187 Routine WRF-ARW forecasts with 4 km horizontal grid resolution were acquired from the
188 University of Washington Atmospheric Sciences forecast system
189 (www.atmos.washington.edu/mm5rt/info.html). These forecasts are referred to as WRF-UW.
190 The outer domain of WRF-UW has a horizontal grid resolution of 36 km and covers most of the
191 western US and northeastern Pacific Ocean. This outer domain is initialized with NCEP Global
192 Forecast System (GFS) 1-degree runs. The 36 km grid is nested down to 12 km, 4 km, and an
193 experimental 1.33 km grid which covers a limited portion of the Pacific Northwest. The 4 km
194 grid investigated in this study covers the Pacific Northwest, including Washington, Oregon,
195 Idaho, and portions of California, Nevada, Utah, Wyoming, and Montana. Physical
196 parameterizations employed by WRF-UW include the Noah Land Surface Model (Chen et al.,
197 1996), Thompson microphysics (Thompson et al., 2004), Kain-Fritsch convective scheme (Kain,
198 2004), Rapid Radiative Transfer Model (RRTM) for longwave radiation (Mlawer et al., 1997),
199 Duhdia (1989) for shortwave radiation, and the Yonsei University (YSU) boundary layer scheme
200 (Hong et al., 2006). WRF-UW is run at 00z and 12z and generates hourly forecasts out to 84
201 hours. The computational domain consists of 38 vertical layers. The first grid layer is
202 approximately 40 m AGL and the average model top height is approximately 16000 m AGL.

203 *2.2. Weather Research and Forecasting Reanalysis (WRF-NARR)*

204 WRF-ARW reanalysis runs were performed using the NCEP North American Regional Reanalysis
205 (NARR) data (Mesinger et al., 2006). The reanalysis runs are referred to as WRF-NARR. The
206 same parameterizations and grid nesting structures used in WRF-UW were also used for the
207 WRF-NARR simulations, except that the WRF-NARR inner domain had 33 vertical layers and a
208 horizontal grid resolution of 1.33 km (Table 1). Analysis nudging (e.g., Stauffer and Seaman,
209 1994) was used above the boundary layer in the outer domain (36 km horizontal grid
210 resolution). Hourly WRF-NARR simulations were run for 15 day periods with 12 hours of model
211 spin up prior to each simulation. The first grid layer was approximately 38 m AGL and the
212 average model top height was approximately 15000 m AGL. WRF-NARR differs from the other
213 models used in this study in that it is not a routinely run model. These were custom simulations
214 conducted by our group to provide a best-case scenario for the NWP models. Routine forecasts
215 are already available for limited domains (e.g., UW provides WRF simulations on a 1.33 km grid
216 for a small domain in the Pacific Northwest of the US) and are likely to become more widely
217 available at this grid resolution in the near future.

218 *2.3. North American Mesoscale Model (NAM)*

219 The North American Mesoscale (NAM) model is an operational forecast model run by NCEP for
220 North America (<http://www.emc.ncep.noaa.gov/index.php?branch=NAM>). The NAM model
221 uses the NMM core of the WRF model. The NAM CONUS domain investigated in this study has
222 a horizontal grid resolution of 12 km. NAM employs the Noah Land Surface model (Chen et al.,
223 1996), Ferrier et al. (2003) for microphysics, Kain (2004) for convection, GFDL (Lacis and

224 Hansen, 1974) for longwave and shortwave radiation, and the Mellor-Yamada-Janjic (MJF)
225 boundary layer scheme (Janjic, 2002). The NAM model is initialized with 12-hr runs of the NAM
226 Data Assimilation System. It is run four times daily at 00z, 06z, 12z, and 18z and generates
227 hourly forecasts out to 84 hours. The computational domain consists of 26 vertical layers. The
228 first grid layer is approximately 200 m AGL and the average model top height is approximately
229 15000 m AGL. NAM forecasts are publicly available in real time from NCEP. Although the 12-
230 km horizontal resolution used in NAM is not sufficient to resolve the butte, this resolution is
231 sufficient for resolving the surrounding Snake River Plain and therefore can be used to generate
232 a domain-average flow for input to WindNinja.

233 *2.4. High Resolution Rapid Refresh (HRRR)*

234 The High Resolution Rapid Refresh (HRRR) system is a nest inside of the NCEP-Rapid Refresh
235 (RAP) model (13 km horizontal grid resolution; <http://ruc.noaa.gov/hrrr/>). HRRR has a
236 horizontal grid resolution of 3 km and is updated hourly. HRRR uses the WRF model with the
237 ARW core and employs the RUC-Smirnova Land Surface Model (Smirnova et al., 1997; Smirnova
238 et al., 2000), Thompson et al. (2004) microphysics, RRTM longwave radiation (Mlawer et al.,
239 1997), Goddard shortwave radiation (Chou and Suarez, 1994), the MYJ boundary layer scheme
240 (Janjic, 2002). HRRR is initialized from 3-km grids with 3-km radar assimilation over a 1-hr
241 period. HRRR is currently the highest resolution operational forecast available in real time. The
242 computational domain consists of 51 vertical layers. The first grid layer is approximately 8 m
243 AGL and the average model top height is approximately 16000 m AGL.

244 *2.5. WindNinja*

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

253

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

263

264 WindNinja was used to dynamically downscale hourly 10-m wind predictions from the above
265 NWP models. The WindNinja computational domain was constructed from 30-m resolution

266 Shuttle Radar Topography Mission (SRTM) data (Farr et al., 2007). The 10-m NWP winds were
267 bilinearly interpolated to the WindNinja computational domain and used as the initial wind
268 field. Layers above and below the 10-m height were fit to a logarithmic profile (neutral
269 atmospheric stability) based on the micrometeorological model. The computational domain
270 consisted of 20 vertical layers. The first grid layer is 1.92 m AGL and the average model top
271 height is 931 m AGL.

272 *2.6. Terrain representation*

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

283 284 **3. Evaluations with field observations**

285 *3.1. Observations at Big Southern Butte*

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

292

293 Three-meter wind speeds and directions were measured with cup-and-vane anemometers at
294 53 locations on and around BSB. The anemometers have a measurement range of 0 to 44 m s⁻¹,
295 a resolution of 0.19 m s⁻¹ and 1.4°, and are accurate to within ±0.5 m s⁻¹ and ±5°. The
296 anemometers measured wind speed and direction every second and logged 30-s averages. We
297 averaged these 30-s winds over a 10-min period at the top of each hour (five minutes before
298 and 5 minutes after the hour). The 10-min averaging period was chosen to correspond roughly
299 with the time scale of wind predictions from the NWP forecasts. The NWP output is valid at a
300 particular instant in time, but there is always some inherent temporal averaging in the
301 predictions. The temporal averaging associated with a given prediction depends on the time-
302 step used in the NWP model and is typically on the order of minutes. The 10-min averaged
303 observed data are referred to in the text as 'hourly' observations (since they are averaged at
304 the top of each hour) and are compared directly with the hourly model predictions.

305

306 Butler et al. (2015) observed the following general flow features at BSB. During periods of weak
307 synoptic and mesoscale forcing (hereafter, referred to collectively as ‘external forcing’), the
308 observed surface winds at BSB were decoupled from the large-scale atmospheric flows, except
309 for at high-elevation ridgetop locations. Diurnal slope flows dominated the local surface winds
310 under periods of weak external forcing. There were frequent periods of strong external forcing,
311 during which the diurnal slope winds on BSB were completely overtaken by the larger-scale
312 winds. These periods of strong external forcing at BSB were typically characterized by large-
313 scale southwesterly flow aligned with the Snake River Plain, although occasionally there were
314 also strong early morning winds from the northeast. Under periods of strong external forcing
315 wind speeds commonly varied by as much as 15 m s^{-1} across the domain due to mechanical
316 effects of the terrain (e.g., speed-up over ridges and lower speeds on leeward slopes).
317 Additional details regarding the BSB field campaign can be found in Butler et al. (2015).

318 *3.2. Evaluation methods*

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

328

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

$$331 \quad \overline{\varphi'} = \frac{1}{N} \sum_{i=1}^N \varphi' \quad (1)$$

$$332 \quad \text{RMSE} = \left[\frac{1}{N} \sum_{i=1}^N (\varphi'_i)^2 \right]^{1/2} \quad (2)$$

$$333 \quad \text{SDE} = \left[\frac{1}{N-1} \sum_{i=1}^N (\varphi'_i - \overline{\varphi'})^2 \right]^{1/2} \quad (3)$$

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

336 *3.3. Case selection*

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

341

342 The observed data from the five-day period were broken into periods of upslope, downslope,
343 and externally-driven flow conditions to further investigate model performance under these
344 particular types of flow regimes. We used the partitioning schemes described in Butler et al.

345 (2015). Externally-driven events were partitioned out by screening for hours during which wind
346 speeds at a designated sensor (R2, located 5 km southwest of the butte in flat terrain) exceeded
347 a predetermined threshold wind speed of 6 m s^{-1} . This sensor was chosen because it was
348 located in flat terrain far from the butte and therefore was representative of near-surface
349 winds that were largely unaffected by the butte itself. Hours of upslope and downslope flows
350 (i.e., observations under weak external forcing) were then partitioned out of the remaining
351 data. Additional details regarding the partitioning scheme can be found in Butler et al. (2015).
352 Statistical metrics were computed for these five-day periods.

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

359 **4. Results and discussion**

360 *4.1. Overview of the five-day simulations*

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

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

373

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

383

384 Additionally, the downscaled NAM wind speeds were as accurate as the downscaled HRRR and
385 WRF-UW wind speeds (Fig. 2). This indicates that the NAM forecast was able to capture the

386 important large-scale flow features around BSB such that the additional resolution provided by
387 HRRR and WRF-UW was not essential to resolve additional flow features in the large scale flow
388 around BSB.

389

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

397

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

402

403 NWP models with coarser resolution grids predicted less spatial variability in wind speed (Fig.
404 3). This is because there were fewer grid cells covering the domain, and thus fewer prediction
405 points around the butte. The spatial variability in the downscaled wind speed predictions more

406 closely matched that of the observed data, although the highest speeds were still under
407 predicted (Fig. 3). Although downscaling generally improved the spatial variability of the
408 predictions, there were cases where NWP errors clearly propagated into the downscaled
409 simulations. For example, HRRR frequently over predicted morning wind speeds associated
410 with down-drainage flow on the Snake River Plain; this error was amplified in the downscaled
411 simulations, especially at the ridgetop locations (e.g., Fig. 3-4, 15-17 July).

412

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

420

421 It is difficult to draw too many conclusions from the spatially and temporally averaged 5-day
422 statistics, however, since this period included a range of meteorological conditions (e.g., high-
423 wind events from different directions, upslope flow, downslope flow) each of which could have
424 been predicted with a different level of skill by the models. Qualitatively, however, the 5-day
425 results demonstrate that the spatial variability in the downscaled winds better matches that of

426 the observed winds at BSB (Fig. 3) and, although the reductions were small in some cases,
427 nearly all statistical metrics also improved with downscaling. The analysis is broken down by
428 flow regime in the next section for more insight into model performance.

429 *4.2. Performance under Upslope, downslope, and externally-forced flows*

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

435

436 During upslope flow, the diurnal slope flow parameterization increases speeds on the windward
437 slopes and reduces speeds (or reverses flow and increases speeds, depending on the strength
438 of the slope flow relative to the prevailing flow) on lee slopes due to the opposing effects of the
439 prevailing wind and the thermal slope flow. The parameterization has the opposite effect
440 during downslope flow; windward slope speeds are reduced (or possibly increased if downslope
441 flow is strong enough to reverse the prevailing flow) and lee side speeds are enhanced.

442 *4.2.1 Wind speed*

443 The biggest improvements in wind speed predictions from downscaling occurred during
444 externally-driven flow events (Fig. 5). This is not surprising since the highest spatial variability in
445 the observed wind speeds occurred during high-wind events due to mechanically-induced

446 effects of the terrain, with higher speeds on ridges and windward slopes and lower speeds in
447 sheltered side drainages and on the lee side of the butte (Fig. 6-8). Since WindNinja is designed
448 primarily to simulate the mechanical effects of the terrain on the flow, it is during these high-
449 wind events that the downscaling has the most opportunity to improve predictions across the
450 domain. This has important implications for wildfire applications since high-wind events are
451 often associated with increased fire behavior.

452

453 The NWP models tended to under predict wind speeds on the windward slopes, ridgetops, and
454 surrounding flat terrain, and over predict on the lee side of the butte during high wind events
455 (e.g., Fig. 6). The largest NWP errors in wind speed during high wind events were on the
456 ridgetops, where speed-up occurred and the NWP under predicted speeds. These largest wind
457 speed errors were reduced by downscaling (e.g., Fig. 6). Downscaling reduced NWP wind speed
458 errors in most regions on the butte, although the general trend of under predicting wind speeds
459 on the windward side and over predicting on the lee side did not change (e.g., Fig. 6).

460

461 There were consistent improvements in predicted wind speeds from downscaling during the
462 upslope regime, although the improvements were smaller than for the externally-driven regime
463 (Fig. 5). Wind speeds were lower during the slope flow regimes than during the externally-
464 forced regime (Fig. 6-8), and thus, smaller improvements were possible with downscaling.

465 There was some speed-up predicted on the windward side of the butte during the
466 representative upslope case which appeared to match the observed wind field (Fig. 8).

467

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

477

478 The diurnal slope flow parameterization in WindNinja resulted in lower speeds on the
479 windward side and higher speeds on the lee side of the butte for the representative downslope
480 case (Fig. 7). These downscaled speeds better matched those of the observed wind field,
481 although speeds were still under predicted for ridgetops and a few other locations around the
482 butte (Fig. 7). The high observed speeds at the ridgetop locations are not likely due to thermal
483 slope flow effects, but could be from the influence of gradient-level winds above the nocturnal
484 boundary layer. These ridgetop locations are high enough in elevation (800 m above the
485 surrounding plain) that they likely protruded out of the nocturnal boundary layer and were
486 exposed to the decoupled gradient-level winds. Butler et al. (2015) noted that ridgetop winds
487 did not exhibit a diurnal pattern and tended to be decoupled from winds at other locations on

488 and around the butte. Lack of diurnal winds at the summit of the butte is also confirmed by
489 National Oceanic and Atmospheric Administration Field Research Division (NOAA-FRD) mesonet
490 station data collected at the top of BSB (described in Butler et al., 2015;
491 <http://www.noaa.inel.gov/projects/INLMet/INLMet.htm>).

492
493 Under predictions on the lower slopes and on the plain surrounding the butte could be due to
494 overly weak slope flows being generated by the slope flow parameterization in WindNinja (Fig.
495 7-8). Overly weak slope flows could be caused by a number of things: improper
496 parameterization of surface or entrainment drag parameters, poor estimation of the depth of
497 the slope flow, or deficiencies in the micrometeorological model used. The slope flow
498 parameterization is being evaluated in a companion paper.

499 *4.2.2 Wind direction*

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

509

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

515

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

527 **5. Summary**

528 The horizontal grid resolutions of NWP models investigated in this study were too coarse to
529 resolve the BSB terrain. Results showed that the NWP models captured the important large-

530 scale flow features around BSB under most conditions, but were not capable of predicting the
531 high spatial variability (scale of 100s of meters) in the observed winds on and around the butte
532 induced by mechanical effects of the terrain and local surface heating and cooling. Thus,
533 surface winds from the NWP models investigated in this study would not be sufficient for
534 forecasting wind speeds on and around the butte at the spatial scales relevant for processes
535 driven by local surface winds, such as wildland fire spread.

536

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

544

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

551

552 Results indicated that WindNinja predicted overly weak slope flows compared to observations.

553 Weak slope flow could be caused by several different issues within the diurnal slope flow

554 parameterization in WindNinja: improper parameterization of surface or entrainment drag

555 parameters, poor estimation of the depth of the slope flow, or deficiencies in the

556 micrometeorological model. These issues will be explored in future work.

557

558 This work constitutes evaluation of a diagnostic wind model at unprecedented high spatial

559 resolution and terrain complexity. While extensive evaluations have been performed with data

560 collected in less rugged terrain (e.g., Askervein Hill and Bolund Hill, relatively low elevation hills

561 with simple geometry), to our knowledge, this study is the first to evaluate a diagnostic wind

562 model with data collected in terrain with topographical ruggedness approaching that of typical

563 landscapes in the western US susceptible to wildland fire. This work demonstrates that NWP

564 model wind forecasts can be improved in complex terrain, especially under high-wind events,

565 through dynamic downscaling via a mass-conserving wind model. These improvements should

566 propagate on to more realistic predictions from other model applications which are sensitive to

567 surface wind fields, such as wildland fire behavior, local-scale transport and dispersion, and

568 wind energy applications.

569 **Acknowledgements**

570 Thanks to Dave Ovens and Cliff Mass of the University of Washington for providing access to

571 the WRF-UW simulations and Eric James of NOAA–GSD Earth System Research Laboratory for

572 access to the HRRR simulations. Thanks to Serena Chung of the Laboratory for Atmospheric
573 Research, Washington State University, for guidance on the WRF-NARR simulations. We also
574 thank the participants in the BSB and SRC field campaigns, including Dennis Finn, Dan Jimenez,
575 Paul Sopko, Mark Vosburgh, Larry Bradshaw, Cyle Wold, Jack Kautz, and Randy Pryhorocky.
576

577 **References**

- 578 Beaucage, P., Brower, M.C., and Tensen, J.: Evaluation of four numerical wind flow models for
579 wind resource mapping, *Wind Energy*, 17, 197–208, 2014.
- 580 Berg, J., Mann, J., Bechmann, A., Courtney, M.S., and Jorgensen, H.E.: The Bolund experiment,
581 Part I: flow over a steep, three-dimensional hill, *Bound.-Lay. Meteorol.*, 141, 219–243, 2011.
- 582 Butler, B.W., Wagenbrenner, N.S., Forthofer, J.M., Lamb, B.K., Shannon, K.S., Finn, D., Eckman,
583 R.M., Clawson, K., Bradshaw, L., Sopko, P., Beard, S., Jimenez, D., Wold, C., and Vosburgh,
584 M.: High-resolution observations of the near-surface wind field over an isolated mountain
585 and in a steep river canyon, *Atmos. Chem. Phys.*, 15, 3785–3801, 2015.
- 586 Chen, F., Mitchell, K., Schaake, J., Xue, Y., Pan, H., Koren, V., Duan, Y., Ek, M., and Betts, A.:
587 Modeling of land-surface evaporation by four schemes and comparison with FIFE
588 observations, *J. Geophys. Res.*, 101, 7251–7268, 1996.
- 589 Ching, J., Rotunno, R., LeMone, M., Martilli, A., Kosovic, B., Jimenez, P.A., and Dudhia, J.:
590 Convectively induced secondary circulations in fine-grid mesoscale numerical weather
591 prediction models, *Mon. Wea. Rev.*, 142, 3284–3302, 2014.
- 592 Chou, M., and Suarez, M.J.: An efficient thermal infrared radiation parameterization for use in
593 general circulation models, *Technical Report Series on Global Modeling and Data*
594 *Assimilation*, NASA Tech. Memo. 104606, 3, 85 pp, 1994.

595 Chow, F.K., and Street, R.L.: Evaluation of turbulence closure models for large-eddy simulation
596 over complex terrain: Flow over Askervein Hill, *J. Appl. Meteor. Climatol.*, 48, 1050–1065,
597 2008.

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

600 Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M.,
601 Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M.,
602 Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography Mission, *Reviews of Geophysics*,
603 45, RG2004, doi: 10.1029/2005RG000183, 2007.

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

608 Forthofer, J., Shannon, K., and Butler, B.: Simulating diurnally driven slope winds with
609 WindNinja, Eighth Symposium on Fire and Forest Meteorology. Oct 13-15. Kalispell, MT,
610 2009.

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

614 Forthofer, J.M., Butler, B.W., McHugh, C.W., Finney, M.A., Bradshaw, L.S., Stratton, R.D,
615 Shannon, K.S., and Wagenbrenner, N.S.: A comparison of three approaches for simulating
616 fine-scale surface winds in support of wildland fire management. Part II. An exploratory
617 study of the effect of simulated winds on fire growth simulations, *Int. J. Wildland Fire*, 23,
618 982–994, 2014b.

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

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

623 Kain, J.: The Kain-Fritsch convective parameterization: An update, *J. Meteor. Climatol.*, 43, 170–
624 181, 2004.

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

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

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

630 Mesinger, F., DiMego, G., Kalnay, E., Mitchell, K., Shafran, P.C., Ebisuzaki, W., Jovic, D., Woollen,
631 J., Rogers, E., Berbery, E.H., Ek, M.B., Fan, Y., Grumbine, R., Higgins, W., Li, H., Lin, Y.,

632 Manikin, G., Parrish, D., and Shi, W.: North American regional reanalysis, Bull. Amer.
633 Meteor. Soc., 87, 343–360, 2006.

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

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

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

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

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

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

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

654 Stauffer, D.R., and Seaman, N.L.: Multiscale four-dimensional data assimilation, *J. Appl. Meteor.*
655 33, 416–434, 1994.

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

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

661

662 **Tables**

663 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

664 ^aApproximate average height AGL.

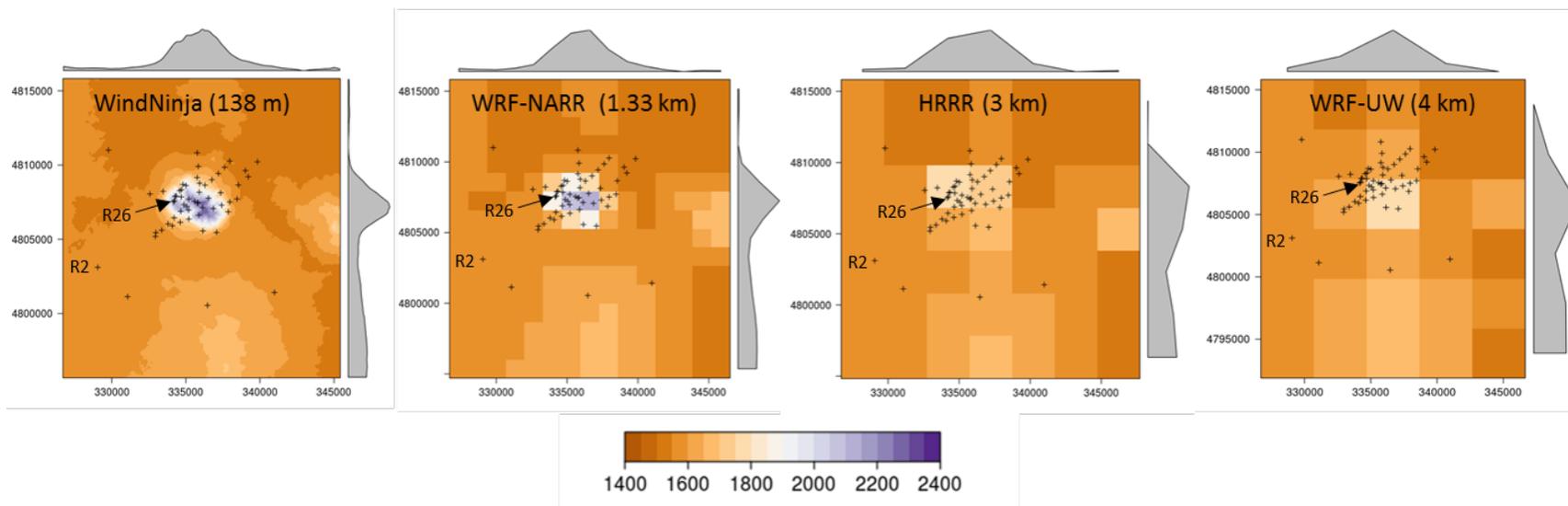
665

666 Table 2. Model mean bias, root-mean-square error (RMSE), and standard deviation of errors (SDE) for surface wind speeds and
 667 directions during the 5-day evaluation period at Big Southern Butte. Downscaled values are in parentheses. Smaller values are in
 668 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 ($^{\circ}$)					
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)

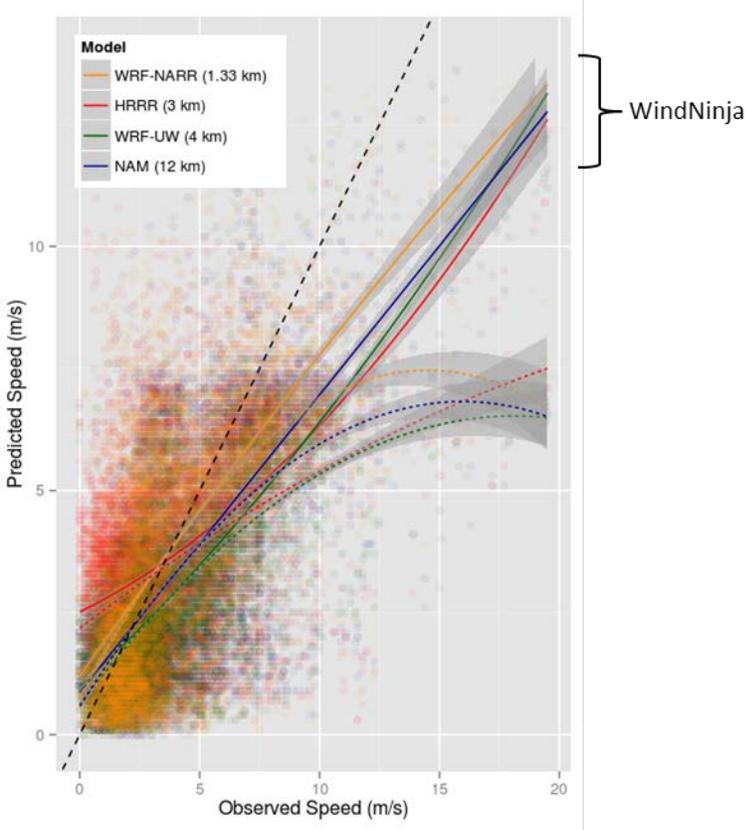
669

670 **Figures**



671
672
673 Figure 1. Terrain representation (m ASL) in WindNinja, WRF-NARR, HRRR, and WRF-UW for the Big Southern Butte. Crosses indicate
674 surface sensor locations. Maps are projected in the Universal Transverse Mercator (UTM) zone 12 coordinate system. Axis labels are
675 eastings and northings in m. Profiles in gray are the average elevations for rows and columns in the panel. NAM (12 km) terrain is
676 represented by just four cells and is not shown here.

677



678

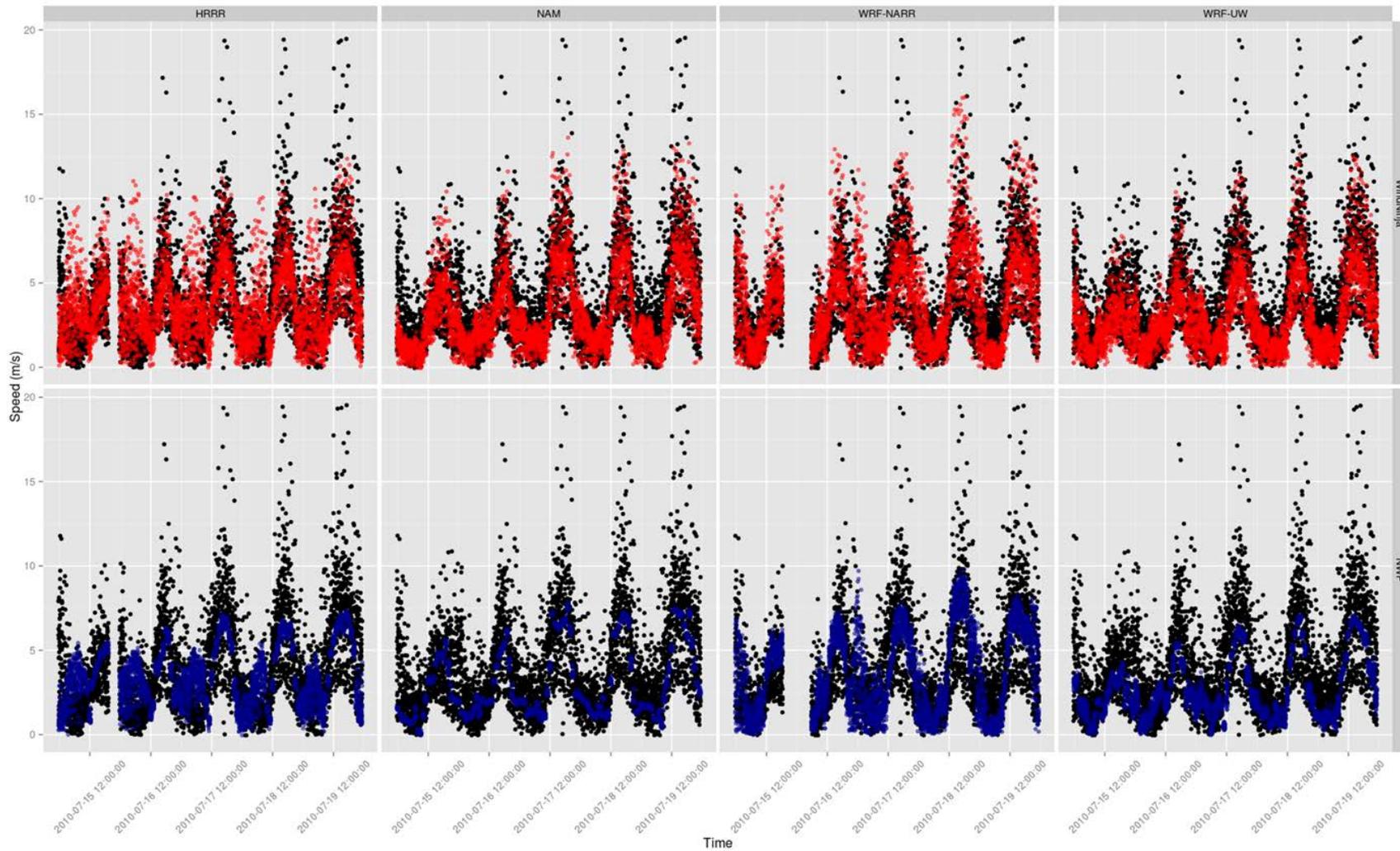
679 Figure 2. Observed vs. predicted wind speeds for the 5-day evaluation period at Big Southern

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

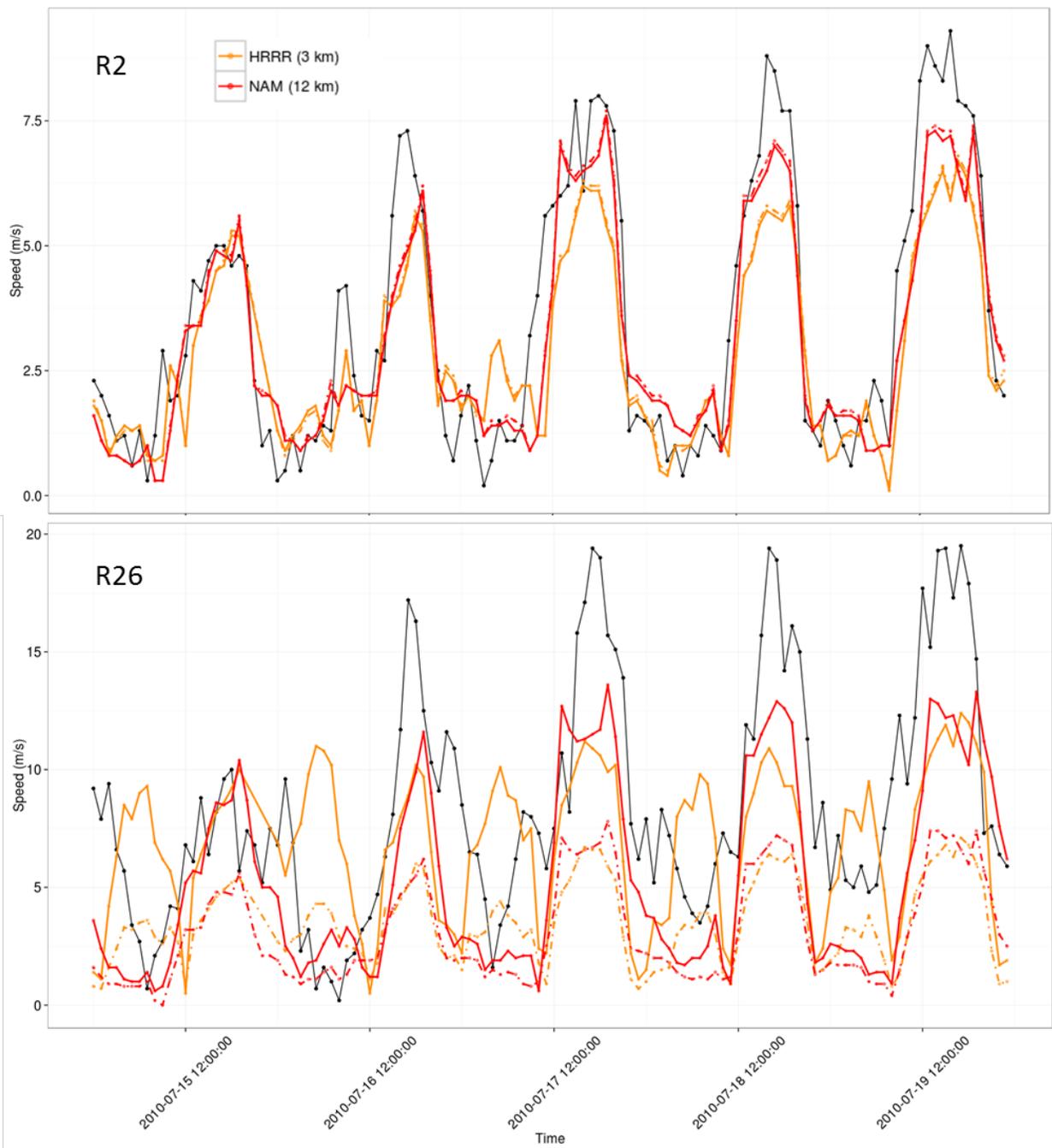
681 (quadratic fit); dashed lines are NWP models and solid lines are NWP forecasts downscaled with

682 WindNinja. Shading indicates 95% confidence intervals.

683

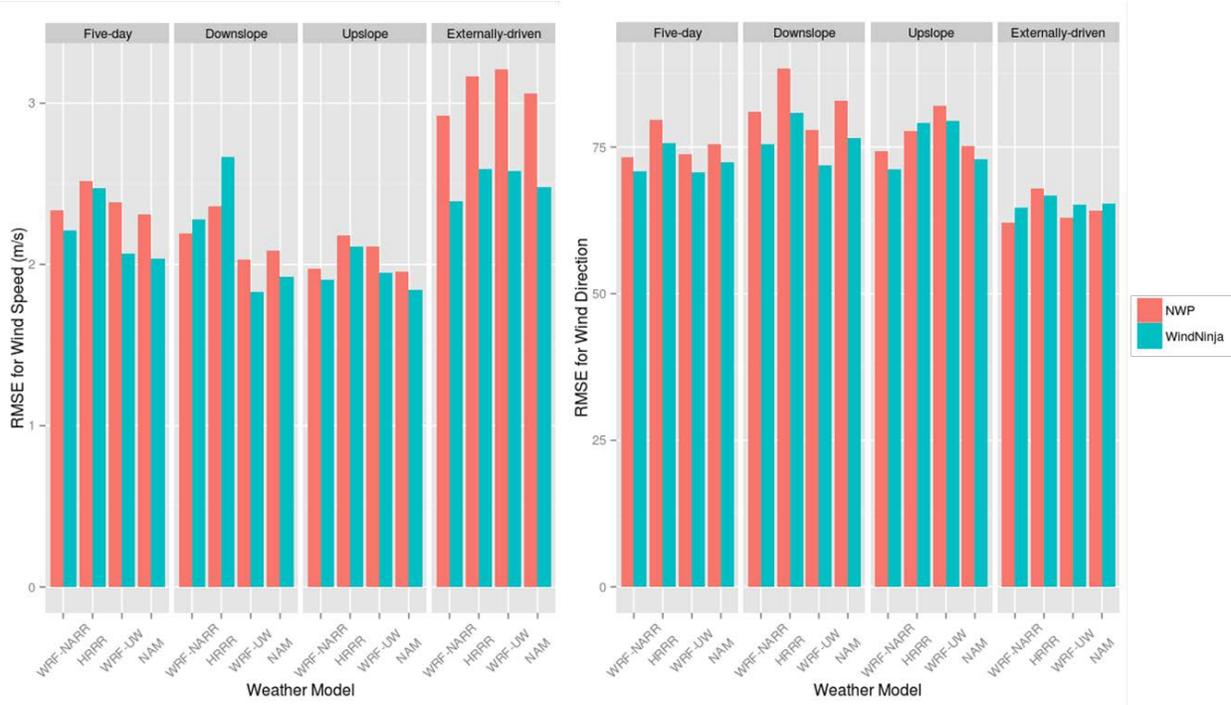


684
685 Figure 3. Observed (black) and predicted (colored) winds speeds at all sensors for 15 July 2010–19 July 2010 at Big Southern Butte.
686 Top panels are WindNinja predictions. Bottom panels are NWP predictions.



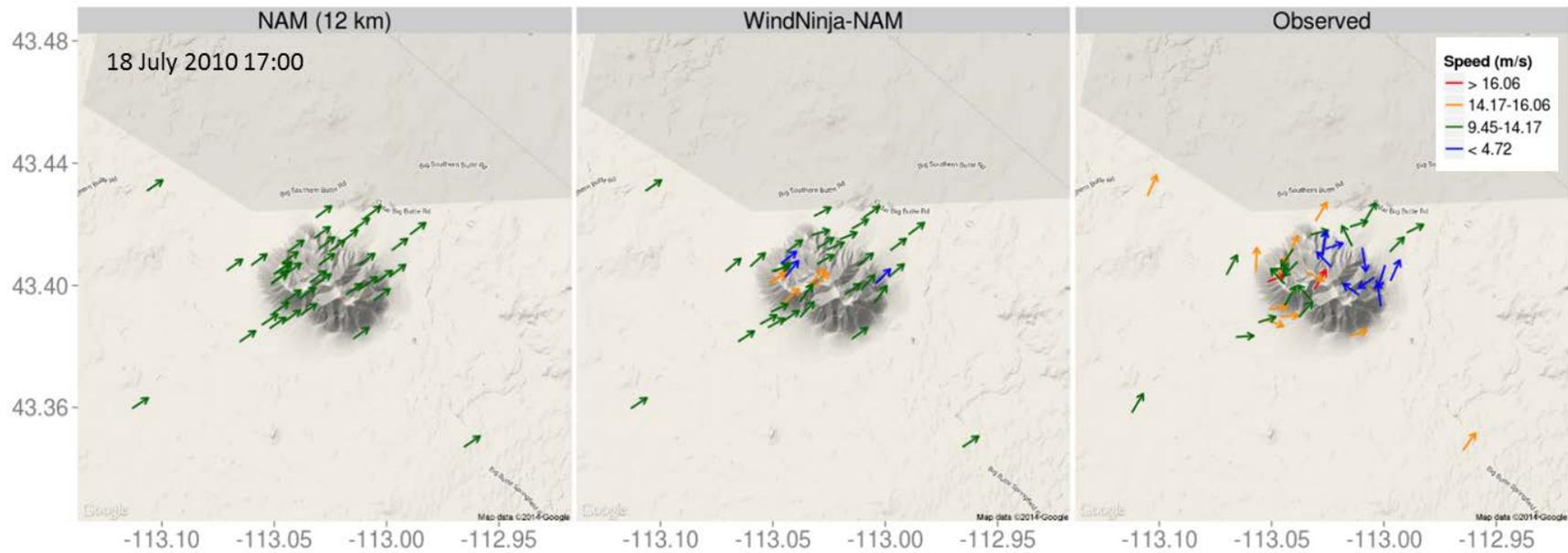
687

688 Figure 4. Observed (black line) and predicted (colored lines) wind speeds for sensor R2 located
 689 5 km southwest of Big Southern Butte on the Snake River Plain and sensor R26 located on a
 690 ridgetop. Dashed colored lines are NWP models and solid colored lines are WindNinja.



691

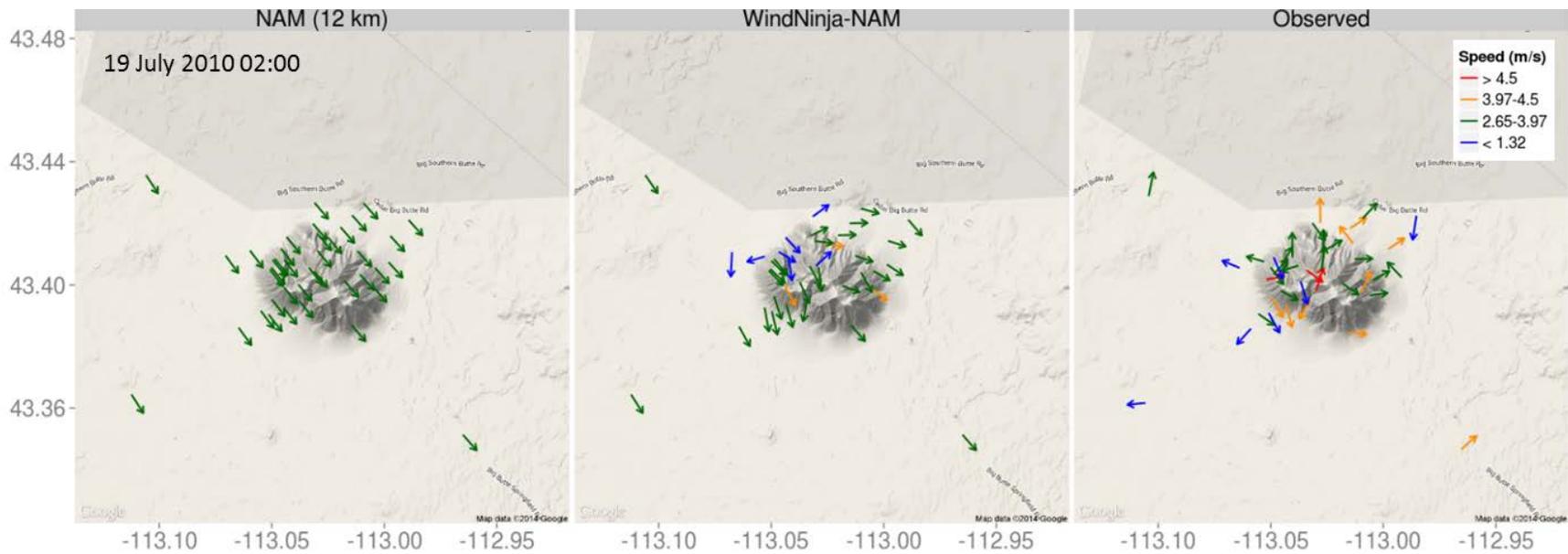
692 Figure 5. Root-mean-square error in wind speed (left) and wind direction (right) at Big Southern
 693 Butte for the five-day evaluation period ($N = 4149$), and downslope ($N = 1593$), upslope ($N =$
 694 717), and externally -driven ($N = 966$) periods within the five-day period. Sample size, $N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



696

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

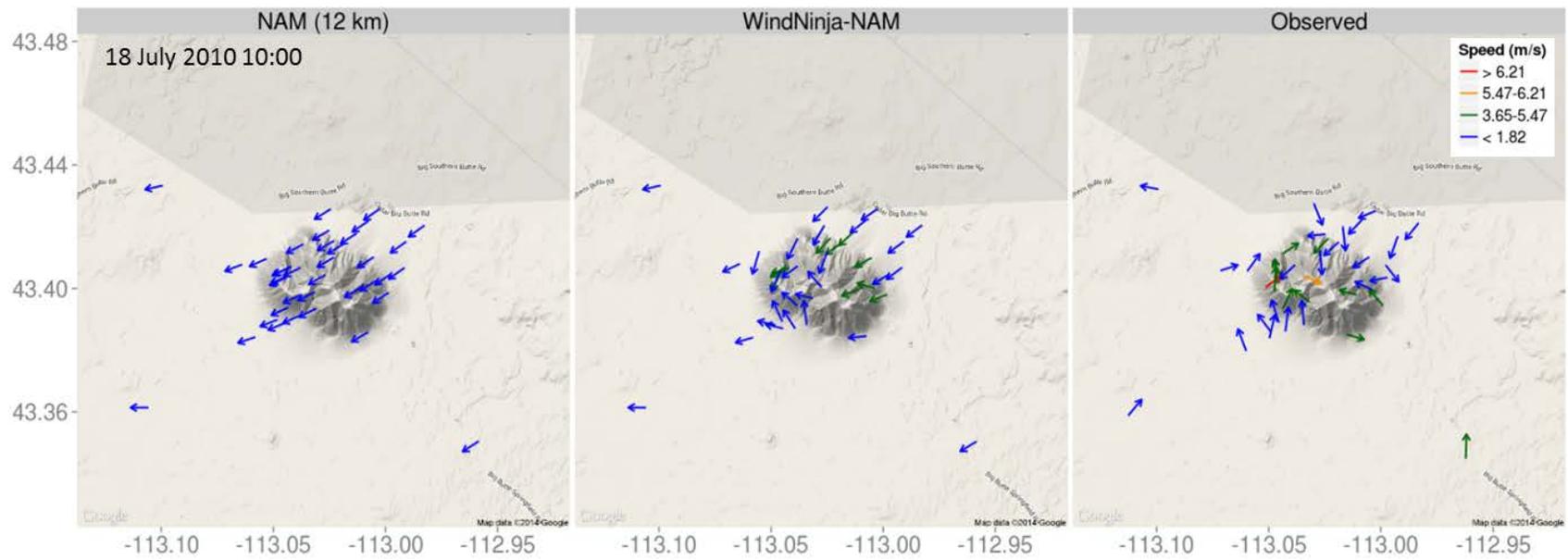
698



699

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

701



702

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

