LLNL-TR-499162

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Observing System Simulation Experiments (OSSEs) for the Mid-Columbia Basin

J. Zack, E. J. Natenberg, G. V. Knowe, K. Waight,
J. Manobianco, D. Hanley, C. Kamath

September 14, 2011

Disclaimer

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

Observing System Simulation Experiments (OSSEs) for the Mid-Columbia Basin

John Zack, Eddie Natenberg, Glenn Van Knowe,
Ken Waight, John Manobianco and Deborah Hanley
AWS Truepower, LLC

Chandrika Kamath
Lawrence Livermore National Laboratory

6 September 2011

1. Introduction

The overall goal of this multi-phased research project known as WindSENSE is to develop an observation system deployment strategy that would improve wind power generation forecasts. The objective of the deployment strategy is to produce the maximum benefit for 1- to 6-hour ahead forecasts of wind speed at hub-height (~80 m). In this phase of the project the focus is on the Mid-Columbia Basin region, which encompasses the Bonneville Power Administration (BPA) wind generation area (Figure 1) that includes the Klondike, Stateline, and Hopkins Ridge wind plants.

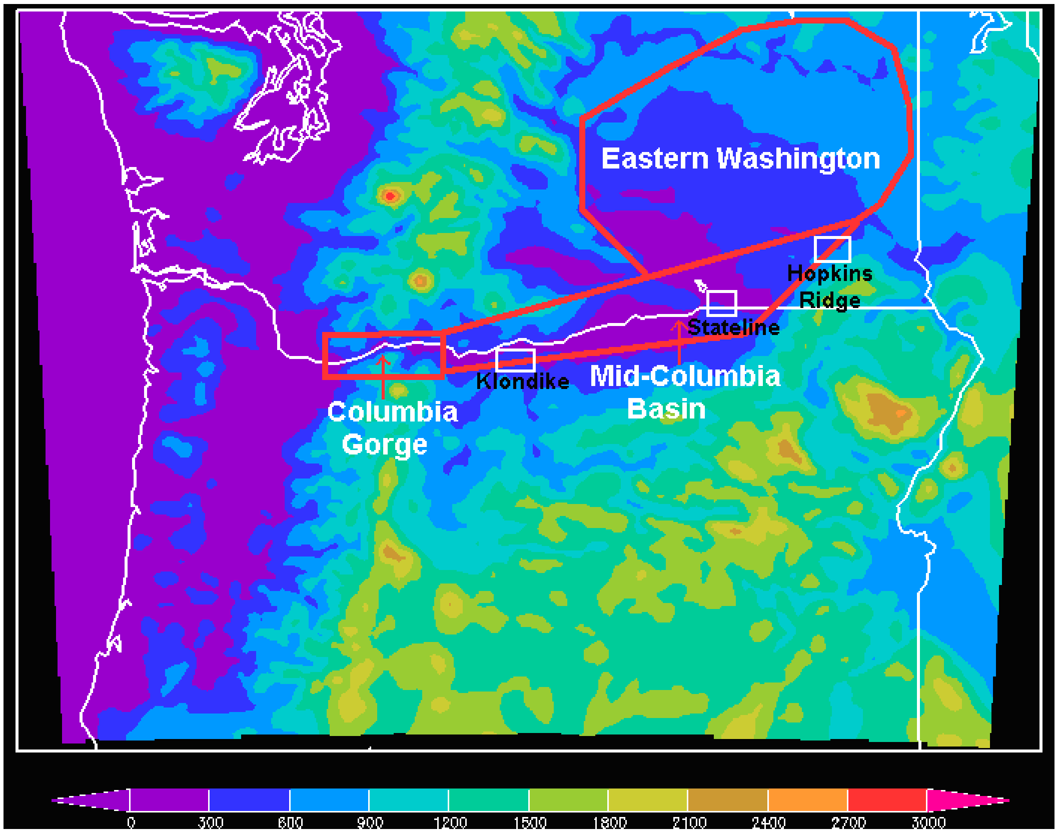

Figure 1. *The approximate geographical area used in the OSSE simulations. The color shading depicts the terrain elevation (m) on the scale of the model grid. The white boxes denote the forecast target areas for which the average forecast metric (80-m wind speed) was calculated.*

There are two tasks in the current project effort designed to validate the Ensemble Sensitivity Analysis (ESA) observational system deployment approach in order to move closer to the overall goal:

(1) Perform an Observing System Experiment (OSE) using a data denial approach. The results of this task are presented in a separate report.

(2) Conduct a set of Observing System Simulation Experiments (OSSE) for the Mid-Colombia basin region.

This report presents the results of the OSSE task. The specific objective is to test strategies for future deployment of observing systems in order to suggest the best and most efficient ways to improve wind forecasting at BPA wind farm locations. OSSEs have been used for many years in meteorology to evaluate the potential impact of proposed observing systems, determine trade-offs in instrument design, and study the most effective data assimilation methodologies to incorporate the new observations into numerical weather prediction (NWP) models (Atlas 1997; Lord 1997). For this project, a series of OSSEs will allow consideration of the impact of new observing systems of various types and in various locations.

2. Methods and Configuration

An OSSE typically consists of the following components:

(1) A very high resolution model in both the vertical and horizontal is used to create a surrogate (simulated) numerical atmosphere. This is called the ‘nature run’.

(2) The data set that represents the surrogate atmosphere is then sampled according to the characteristics of proposed sensors and existing sensors in the area.

(3) The data from the sampling process is then used to initialize a lower resolution forecast simulation.

(4) Simulations with and without the “proposed and/or existing simulated observations” are then executed, usually with a different atmospheric model than was used in the nature runs. These are called the ‘forecast simulations’.

(5) The effect of the simulated sensor data on forecast performance is quantified by comparing the forecast runs to the nature runs at desired sites.

The following sections describe these five components of the OSSEs that were performed in this project. The results from OSSEs are generally considered more robust if different models are used for the nature runs and forecast runs – the so-called ‘‘identical twin’’ problem (Atlas 1997). Therefore, the Advanced Regional Prediction System (ARPS) was used for the nature runs and the Weather Research and Forecasting (WRF) system was used for the forecast runs. Figure 2 shows a timeline of the nature and forecast runs.

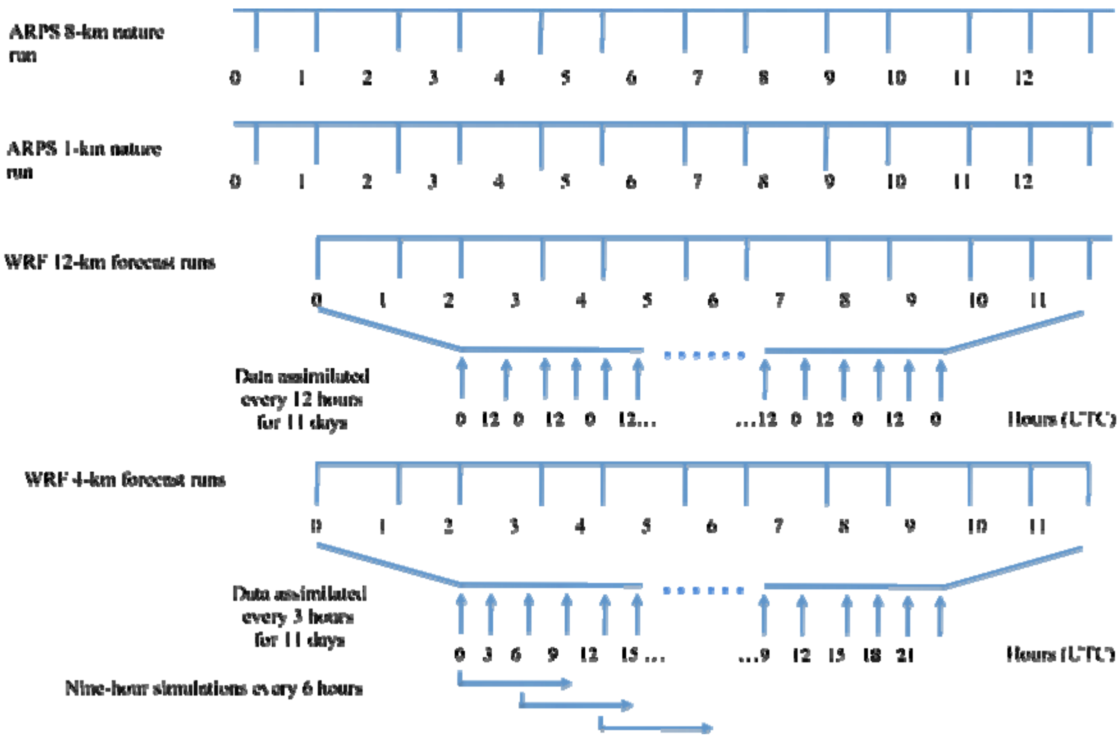

Figure 2. *Timeline of the ARPS nature runs and WRF forecast runs, showing the times of data assimilation.*

2.1 Configuration of the ARPS Nature Runs

The NWP model used for the nature run was version 5.2.11 of ARPS (Xue et al. 2000, Xue et al. 2001). The model domain included an inner and outer grid using 50 vertical levels and one way nesting. The horizontal grid spacing was 8 km (1 km) on the outer (inner) grid with 199 x 199 (795 x 687) points in the north-south and east-west directions, respectively. The average vertical spacing between levels for both the outer and inner grids was 350 m with higher resolution in the boundary layer to better resolve turbine level winds. Both grids were centered in the Mid-Columbia Basin at 45.56°N and 120.98°W.

Initial and lateral boundary conditions were provided by a set of North American Mesoscale model (NAM) initial fields. The outer and inner grid runs consisted of single simulations to avoid any discontinuities caused by reinitializing the model periodically. The warm season simulations began at 0000 UTC 29 April 2007 and ended at 0000 UTC 11 May 2007 (288 hours). The cool season simulations began at 0000 UTC 1 January 2010 and ended at 0000 UTC 13 January 2010. Model output was saved at 10-minute intervals.

2.2 Extraction of Synthetic Observations from the Nature Runs

To extract the simulated (synthetic) observations from ARPS output, software was modified for each type of simulated data: rawinsonde, surface, meteorological (met) tower and SOnic Detection And Ranging (SODAR) observations. To emulate rawinsonde data, approximately 60-70 levels of temperature, dew point temperature, wind direction, and wind speed values were interpolated to vertical levels spaced 250 m apart, with an additional level at the surface. For this experiment, "perfect" observations were assumed, meaning that estimated representativeness or instrumental errors were not added to the extracted data. Data were extracted at actual rawinsonde and surface observation locations. Synthetic sodar data were created by interpolating wind speed data from the nature runs to vertical levels from 30 to 200 m above ground level, every 10 m, typical of an actual sodar instrument. Simulated met tower data were created by interpolating temperature, dew point, wind speed and wind direction to the 80-m level.

2.3 Assimilation of Synthetic Data

Version 3.3 of the WRF Data Assimilation system (WRFDA) was used to assimilate synthetic data of various types extracted from the ARPS nature runs into the WRF forecasting runs. Default parameters of the scheme were used in winter but the horizontal scale of the unbalanced potential velocity (one of the tuning parameters) was reduced from 1.0 and 0.3 for the warm season runs, after finding that the default setting resulted in analyses that were too smooth.

2.4 WRF Forecast Configuration

Components of the WRF modeling system were used to make the forecast simulations that use data from the nature runs similar to the way that models ordinarily use actual observed data. These runs used versions 3.3 of the WRF pre-processing system, the WRFDA, and the WRF model. Both the warm and cool season runs used a domain that included an inner and an outer grid with 31 vertical levels and one way nesting. The horizontal grid spacing was 12 km (4 km) on the outer (inner) grid with 225 x 225 (181 x 181) points in the north-south and east-west directions, respectively.

Two 12-hour outer grid simulations were executed each day, beginning at 0000 UTC and 1200 UTC. The runs used initial and boundary conditions from the same NAM dataset used by the nature runs. In an OSSE, it is important that the nature and forecast runs differ sufficiently from each other, to mimic the usual differences between a model forecast and the actual atmosphere. Therefore, the forecast runs were staggered from the nature runs (the forecast runs start one day later), so that the nature runs would have enough time to diverge from the NAM initial conditions, adding to the differences generated by using two different atmospheric models (Figure 2). To increase the separation further, the first day of forecast runs were omitted from the set of experiments that were validated, so that the experiments consisted of nine days of forecast simulations after the first "spin-up" day.

The first warm season outer grid simulation began at 0000 UTC 30 April 2007 and the last one ended at 0000 UTC 11 May 2007. The first simulation was a "cold start", in which data from a larger scale model (NAM) provided the first guess fields and lateral boundary conditions, and rawinsonde and surface data were assimilated to create the initial state. Each simulation after the first one was a “warm start”, using the 12-hour forecast from the previous simulation as a first guess, and assimilating synthetic surface data and rawinsonde data extracted from the nature runs at the usual observation locations. The first cool season outer grid simulation began at 0000 UTC 2 January 2010 and the last one ended at 0000 UTC 14 January 2010.

A new inner grid simulation began every three hours, when synthetic observations were assimilated. After an initial cold start from NAM data, each simulation was a warm start using the 3-hour forecast from the previous run as a first guess, and lateral boundary conditions from the outer grid runs. As shown in Figure 2, four simulations were run each day (beginning at 0000, 0600, 1200 and 1800 UTC) for a 9-hour forecast period to evaluate data impacts. The runs at intermediate times (beginning at 0300, 0900, 1500 and 2100 UTC) were 3-hour simulations necessary to provide a first guess for the following runs.

Synthetic rawinsonde data at the usual observation locations were assimilated every 12 hours and synthetic surface data at the usual observation locations were assimilated every three hours. The assimilation of other types of simulated observations forms the basis for the set of OSSEs listed in Section 3. The first inner grid warm season run began at 0000 UTC 30 April 2007 and the last one ended at 0300 UTC 10 May 2007. The first inner grid cool season run began at 0000 UTC 2 January 2010 and the last one ended at 0300 UTC 12 January 2010. Model output was saved to a file every hour.

2.5 Description and Configuration of Evaluation Methods

The evaluation of the OSSEs centered on comparing errors at the three BPA wind farm locations: Klondike, Stateline and Hopkins Ridge (Fig. 1). For the warm and cool season simulations, 80-m wind speed values were extracted from the grid points nearest to these locations and compared to values from the nature runs at the same locations. Mean absolute errors for each forecast hour were calculated over the entire set of 9-hour simulations (those initialized at 0000, 0600, 1200, and 1800 UTC). Because there were usually small differences in the forecasts after six hours, results from only the first six forecast hours are shown.

3. Description of the OSSEs

Table 1 gives a brief description of each experiment along with the variables assimilated into the forecast runs. For some of the following experiments, simulated observations were extracted at high or low sensitivity points for Klondike and other BPA wind farms, as determined by the ESA work described in a previous report (Zack et al. 2010). The ESA approach uses data generated by a set (ensemble) of perturbed NWP simulations for a sample time period to diagnose the sensitivity of a specified forecast variable (metric) for a target location to initial condition state variable(s) at other locations and prior times. The result is a spatial map of sensitivity values, where it is expected that observations in locations with higher values are more likely to impact the short-term wind forecast at the target site.

Table 1. List of Experiments

Experiment	Variables Assimilated (beyond Control)	Description
Control		Simulated rawinsonde and surface data only
1-Tower-Klondike	T, p, T_d, u, v at 80 m	Simulated rawinsonde and surface data, plus met tower data at Klondike
4-Tower-Klondike-High	T, p, T_d, u, v at 80 m	Simulated rawinsonde and surface data, plus met tower data at Klondike and three high sensitivity locations for Klondike
81-Tower-High	T, p, T_d, u, v at 80 m	Simulated rawinsonde and surface data, plus met tower data at 81 high sensitivity locations, a 3x3 matrix of points at three high sensitivity locations for each of the three observing sites
1-Sodar-Klondike	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at Klondike
1-Sodar-High	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at a high sensitivity location for Klondike
1-Sodar-Low	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at a low sensitivity location for Klondike
2-Sodar-Klondike-High	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at Klondike and a high sensitivity location for Klondike
4-Sodar-Klondike-High	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at Klondike and three high sensitivity locations for Klondike
10-Sodar-Klondike-High	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at Klondike and 9 high sensitivity locations, three for each of the three observing sites
81-Sodar-High	u, v from 30-200 m	Simulated rawinsonde and surface data, plus sodar data at 81 high sensitivity locations, a 3x3 matrix of points at three high sensitivity locations for each of the three observing sites

The two most likely instruments for future deployment are met towers and sodars, so the OSSEs focused on them. The 1-Tower-Klondike and 1-Sodar-Klondike experiments explored the value of a single new observation at the target location. The 1-Sodar-High experiment used an observation at the location with the highest forecast sensitivity for 80-m wind speed for Klondike instead of Klondike itself, and 1-Sodar-Low used a single observation at approximately the same distance from Klondike but in an area where the ESA did not have high sensitivity values (Figure 3).

A series of experiments assessed the impact of increasing the number of local observations. The 2-Sodar-Klondike-High experiment used observations at both Klondike and the same high sensitivity location in Figure 3. The 4-Tower-Klondike-High and 4-Sodar-Klondike-High experiments continued the tower-sodar comparison with observations at Klondike, the previous high sensitivity location, along with the locations with the highest forecast sensitivity to 80- to 500-m wind shear and 10- to 80-m wind shear ("High Sensitivity 1", "High Sensitivity 2" and "High Sensitivity 3" in Figure 4, respectively).

The 10-Sodar-Klondike-High experiment included observations at Klondike, the previous three high sensitivity locations for Klondike, and the high sensitivity locations for the same variables (80-m wind speed, 80- to 500-m wind shear, and 10- to 80-m wind shear) for Stateline and Hopkins Ridge (Figure 5). For the 81-Sodar-High and 81-Tower-High experiments, synthetic observations were used from a 3x3 matrix of points, spaced 0.3 deg (~30 km) apart, centered on the nine high sensitivity locations (Figure 5), but not at Klondike.

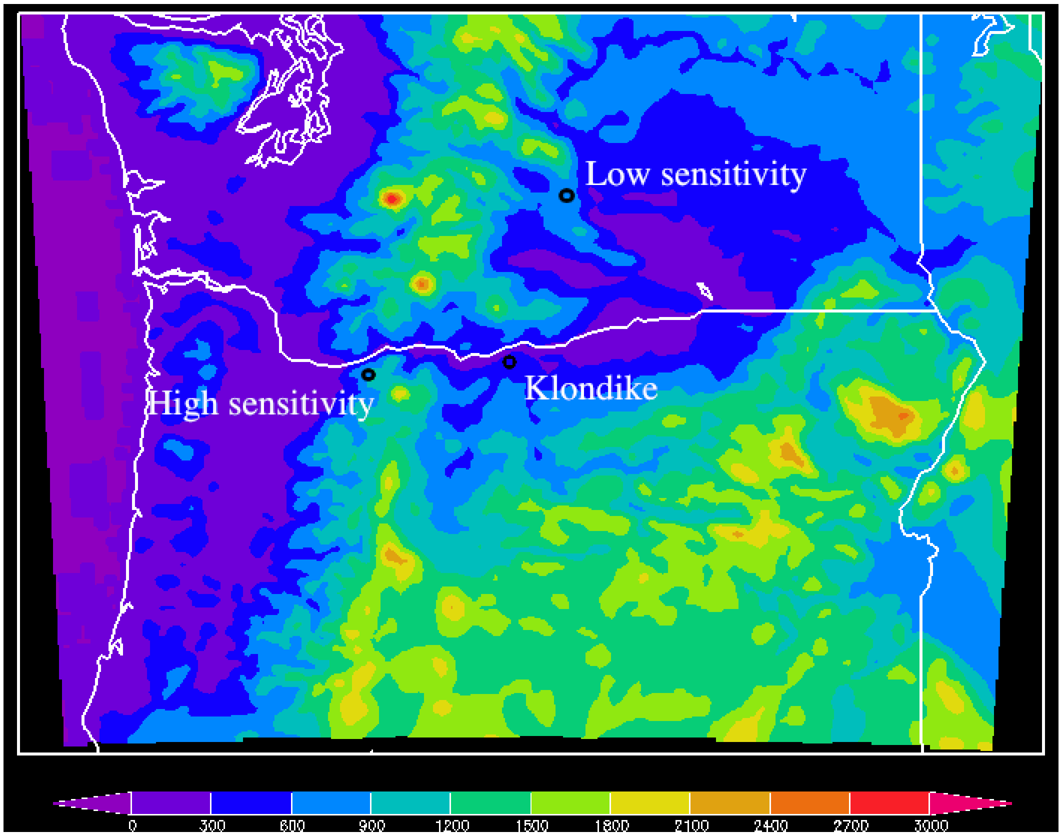

Figure 3. *The low and high sensitivity locations used in the 1-Sodar-High and 1-Sodar-Low experiments. The color shading depicts the terrain elevation (m) on the scale of the model grid.*

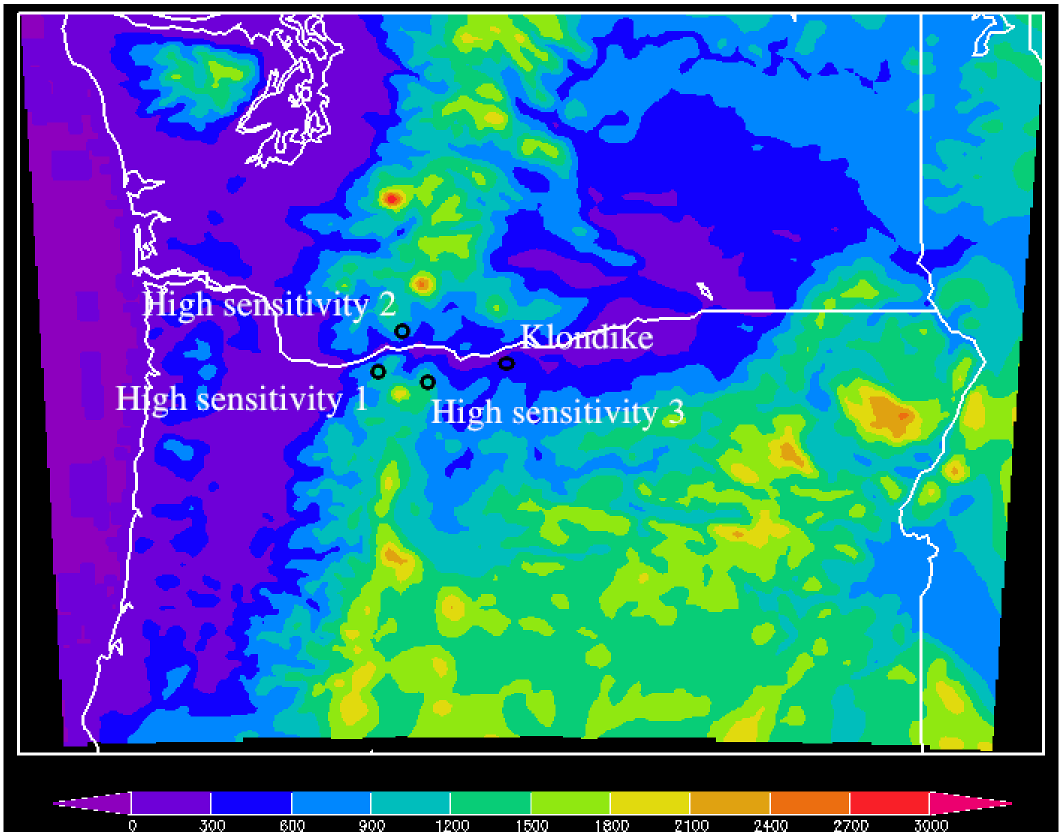

Figure 4. *The three high sensitivity locations for Klondike used in the 4-Tower-Klondike-High and 4-Sodar-Klondike-High experiments. The color shading depicts the terrain elevation (m) on the scale of the model grid.*

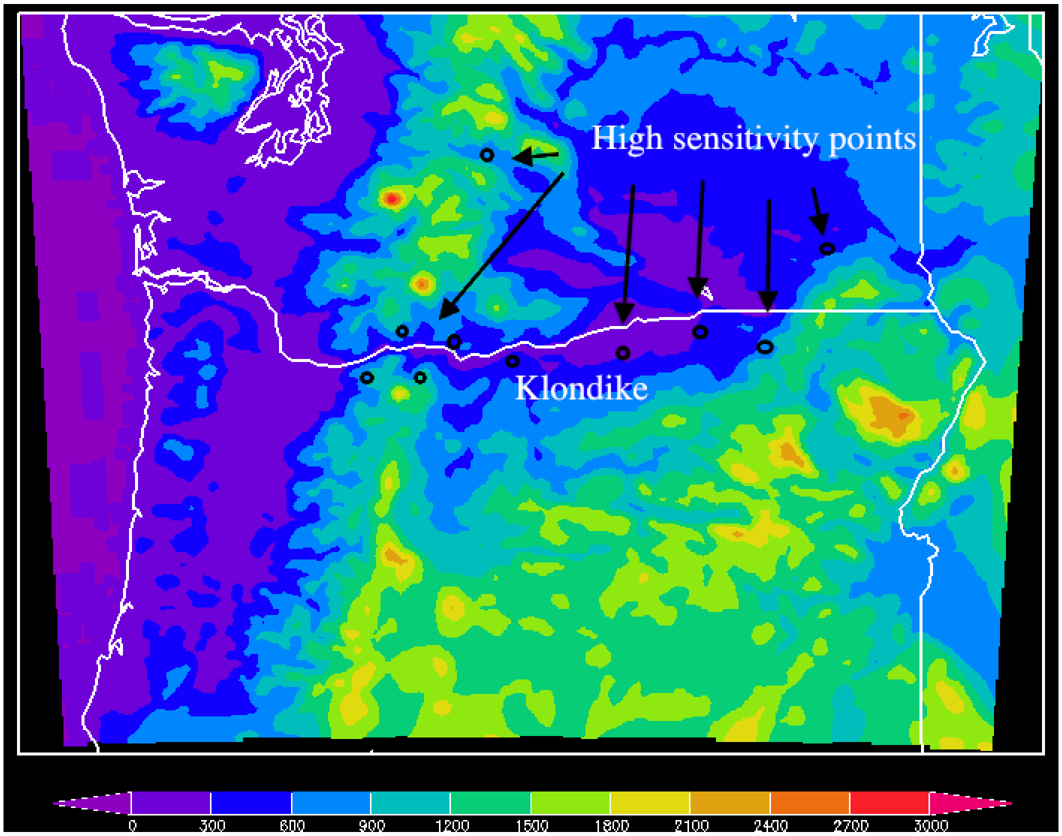

Figure 5. *The nine high sensitivity locations used in the 10-Sodar-Low experiments. For the 81-Tower-High and 81-Sodar-High experiments, a 3x3 matrix of points separated by 0.3 deg (~30 km) was centered on each of the nine high sensitivity points. The color shading depicts the terrain elevation (m) on the scale of the model grid.*

4. Results

Results of the warm and cool season OSSEs are discussed in this section. The experiments focused on the Klondike location, and all validation results shown here are for that site. The results from similar experiments at Stateline and Hopkins Ridge had some differences, but the main conclusions drawn from Klondike are generally applicable to all three sites.

The overall results suggest that assimilation of an observation at the target location would improve the initial value of 80-m wind speed, but the forecast was only improved for one hour. A single observation in a different location, even in one recommended by the ESA to impact the target location, did not improve the forecast. The combination of observations at the target location and at a second location suggested by the ESA produced an improvement in the forecast that persisted for a few hours more. The best enhanced observation scenarios improved the forecast for about four to five hours from the initial time. The improvement was much more evident in the summer than in the winter. The results also indicated that the introduction of one or more sodars improved the forecast significantly more than the equivalent number of meteorological towers.

4.1 *Warm Season 2007*

The mean absolute error (MAE) of 80-m wind speed at Klondike as a function of forecast hour for the warm season experiments described in Table 1 is shown in Figure 6 (all of the experiments are shown here to provide an overview; they are separated into subsets in the following sections). The Control run (the thick red line) assimilated only rawinsonde and surface data, and the error at the initial time (forecast hour zero) was about 3.2 ms^{-1}. The MAE generally decreased with time for this run, which is not typical.

The size of the initial error is probably due to the relatively large distance from the nearest rawinsonde locations (Quillayute and Spokane, Washington; Salem, Oregon and Boise, Idaho). There were a few surface observations relatively close to Klondike, but those 10-m wind speed observations were not sufficient to properly determine the 80-m wind speed at a location in complex terrain such as Klondike. Over the first few hours of the simulation, the errors decreased as the model better captured the terrain-induced flow along the Columbia River Valley.

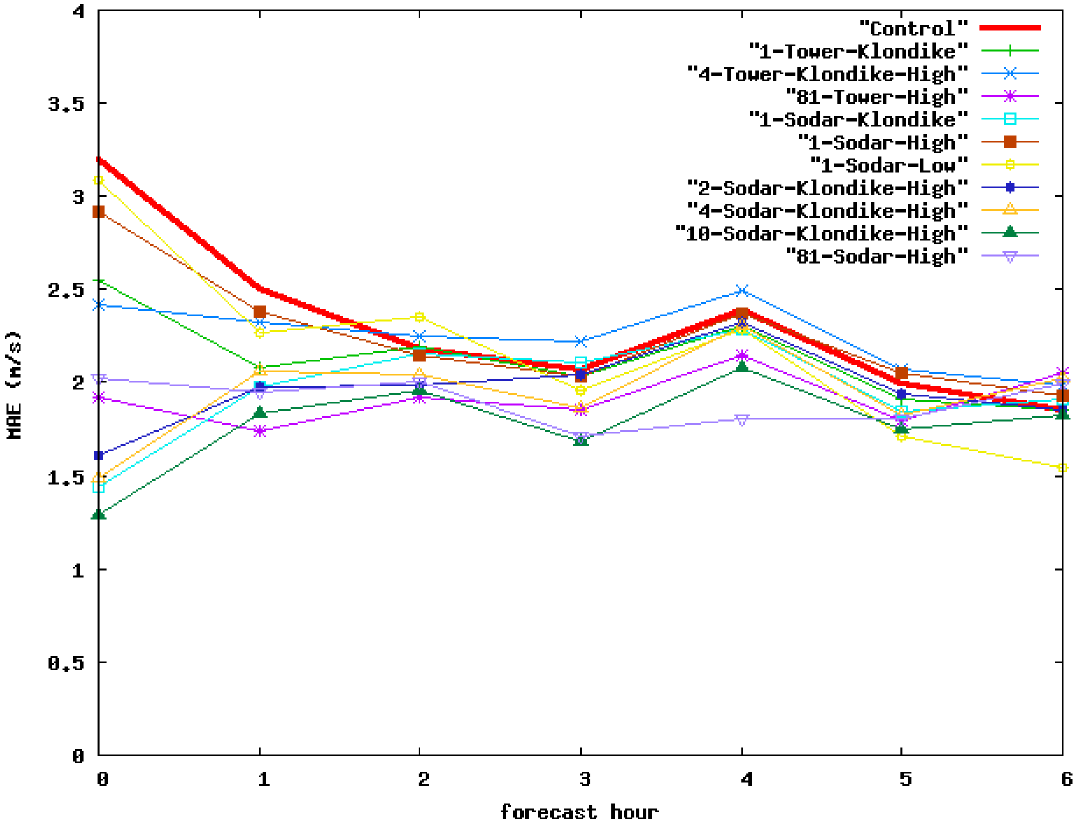

Figure 6. *Mean absolute error (MAE) of 80-m wind speed (ms^{-1}) at Klondike for all of the warm season (May 2007) experiments.*

4.1.1 Comparison of Tower and Sodar Experiments

Figure 7 shows errors for a subset of the experiments, in order to highlight the comparison of sodar and tower data. In the 1-Tower-Klondike experiment, 80-m wind data was assimilated only at the Klondike location itself every three hours, as if a meteorological tower were present there. The MAE at the initial time was reduced from about 3.2 to 2.4 ms^{-1}, but the improvement in the simulation lasted for only one hour after that, mostly because the better initial value at Klondike persisted for a while. The corresponding 1-Sodar-Klondike experiment resulted in an even better initial value at Klondike (MAE of 1.5 ms^{-1} at the initial time), but again, the improvement of the short-term forecast lasted for only one hour.

It is somewhat surprising that sodar data at one location produced a much better initial wind value than one tower observation. The tower observations included temperature and other variables in addition to wind, but perhaps the multiple levels of sodar data forced the variational scheme to give more weight to the sodar data. The errors for the 4-Tower-Klondike-High experiment were also much higher than those for the 4-Sodar-Klondike-High experiment. These results suggest that if only a limited number of observations sites are possible, a small set of sodars may outperform an equivalent set of towers.

It is possible that this difference relates more to details of the analysis scheme than it reflects the comparative value of sodar and tower data. This issue should be resolved by further research. The WRF variational scheme is complex, with many adjustable parameters. Further testing could determine optimal settings for tower and sodar data (maybe different settings for each) that would allow the observations to make the maximum positive impact, and possibly affect the conclusion that sodar data are more helpful at these target locations.

To test the robustness of this finding, 81-Tower-High and 81-Sodar-High experiments were made where synthetic observations were extracted from a 3x3 matrix of points centered on the nine high sensitivity locations in Figure 4. These two runs produced the lowest errors of all of the warm season experiments, showing that a much larger set of observations is better, and that the tower observations were better for the first few hours and then not quite as good for the last few hours.

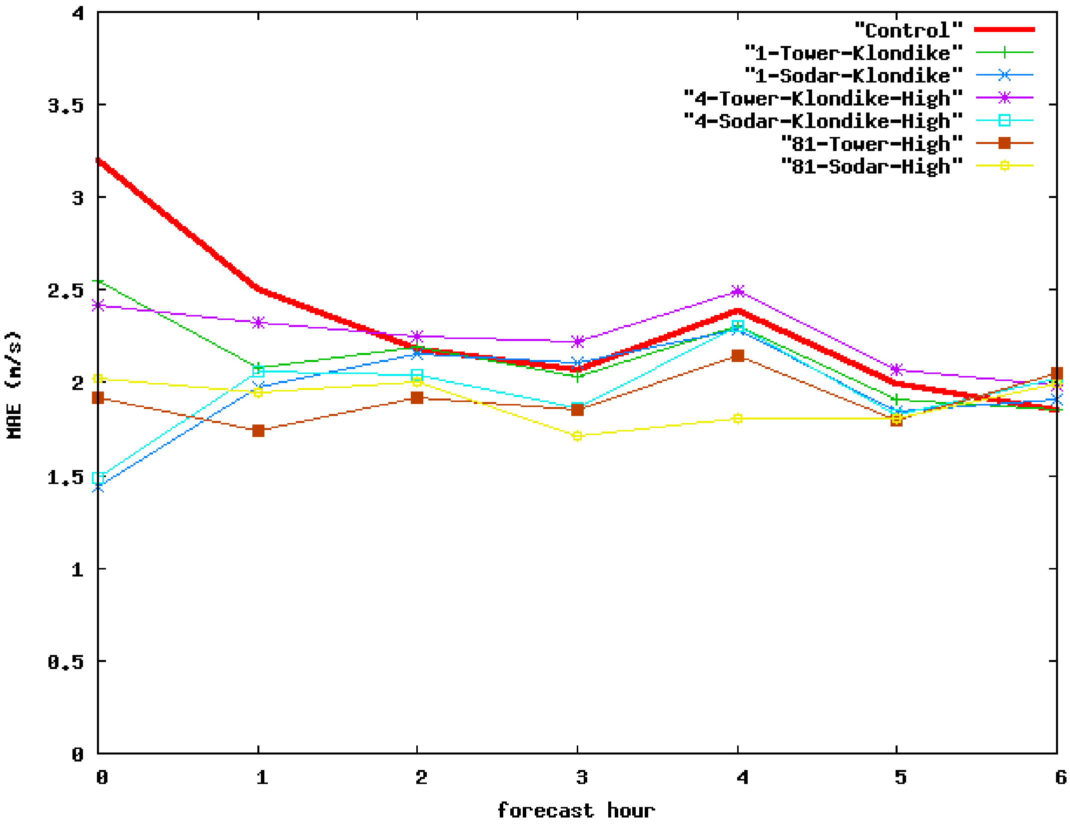

Figure 7. *Mean absolute error (MAE) of 80-m wind speed (m/s) at Klondike for sodar and tower warm season experiments.*

4.1.2 Single Sodar Experiments

Figure 8 shows the results of experiments when data from a single sodar were assimilated at various locations. The 1-Sodar-Klondike experiment was described in the previous section. The ESA results from Zack et al. (2010) indicated, for each BPA wind farm location, a set of other locations that had the greatest potential impact on the short-term (one to three hour) wind forecast. The 1-Sodar-High experiment shows the effect of introducing sodar data only at the highest sensitivity point for Klondike. The errors for that experiment were almost identical to the Control run, so a single observation at the sensitive point did not help the short-term forecast, even for one hour.

In the 1-Sodar-Low experiment, a single sodar was placed at a location found to have low sensitivity for Klondike in the ESA findings (north of the Columbia River). Those sodar observations also did not improve the forecast in the first few hours. These results suggest that a single new observation was unable to improve the short-term forecast, except for about one hour and only when data were available at the forecasting site itself (1-Sodar-Klondike).

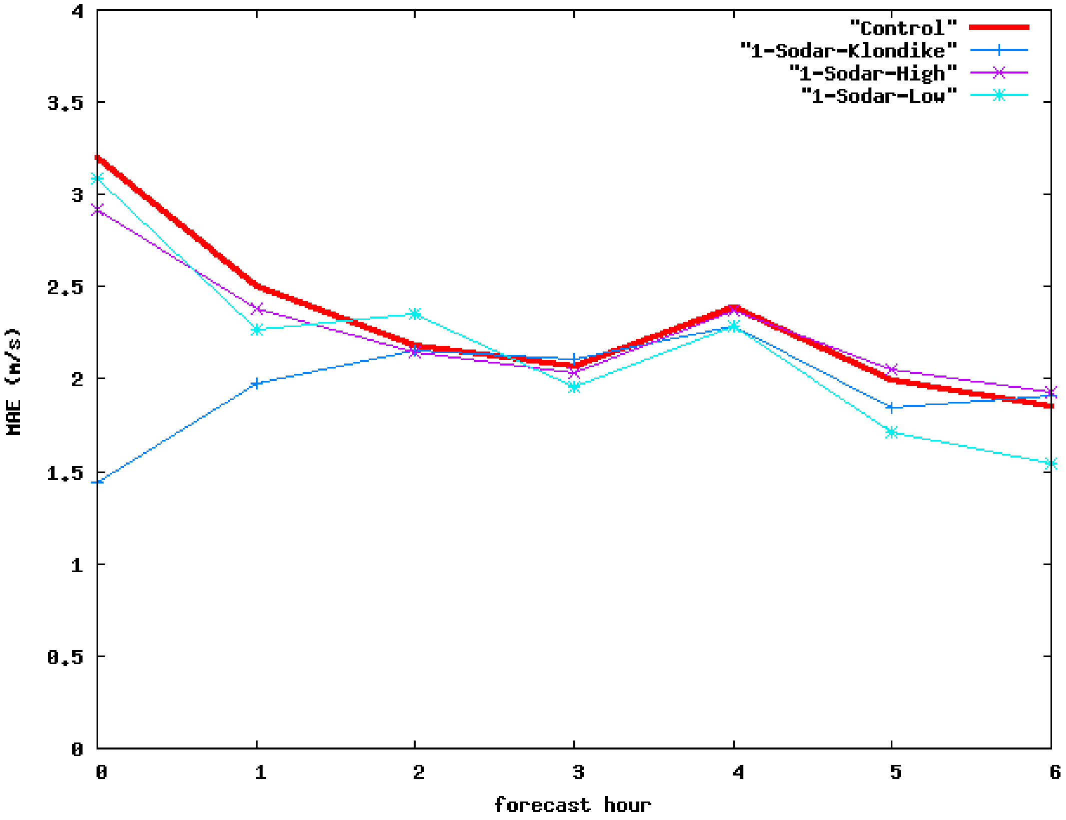

Figure 8. *Mean absolute error (MAE) of 80-m wind speed (m/s) at Klondike for single sodar warm season experiments.*

4.1.3 Multiple Sodar Experiments

Figure 9 shows the results of experiments in which sodar data were assimilated at multiple locations. The 2-Sodar-Klondike-High experiment shows that placement of one sodar at Klondike and another at a high sensitivity point for Klondike produced a forecast improvement for the first two hours. In the 4-Sodar-Klondike-High experiment, sodars were placed at Klondike and three high sensitivity locations for Klondike. The improvements now stretched beyond two hours, with some benefits at three and five hours into the simulations.

The 10-Sodar-Klondike-High experiment tested the scenario of adding a sodar at Klondike, along with nine sodars scattered around the sensitive locations for all three BPA wind farms. These nine locations essentially cover the areas along the Columbia River and along the eastern side of the Cascade Mountains that were found by the ESA work to most significantly impact wind forecasts at Klondike, Stateline, and Hopkins Ridge. These additional observations resulted in further improvement, especially during the first four hours. Results from the 81-Sodar-High experiment were not as good early in the simulations because of the lack of an observation at Klondike but were otherwise similar to 10-Sodar-Klondike-High experiment.

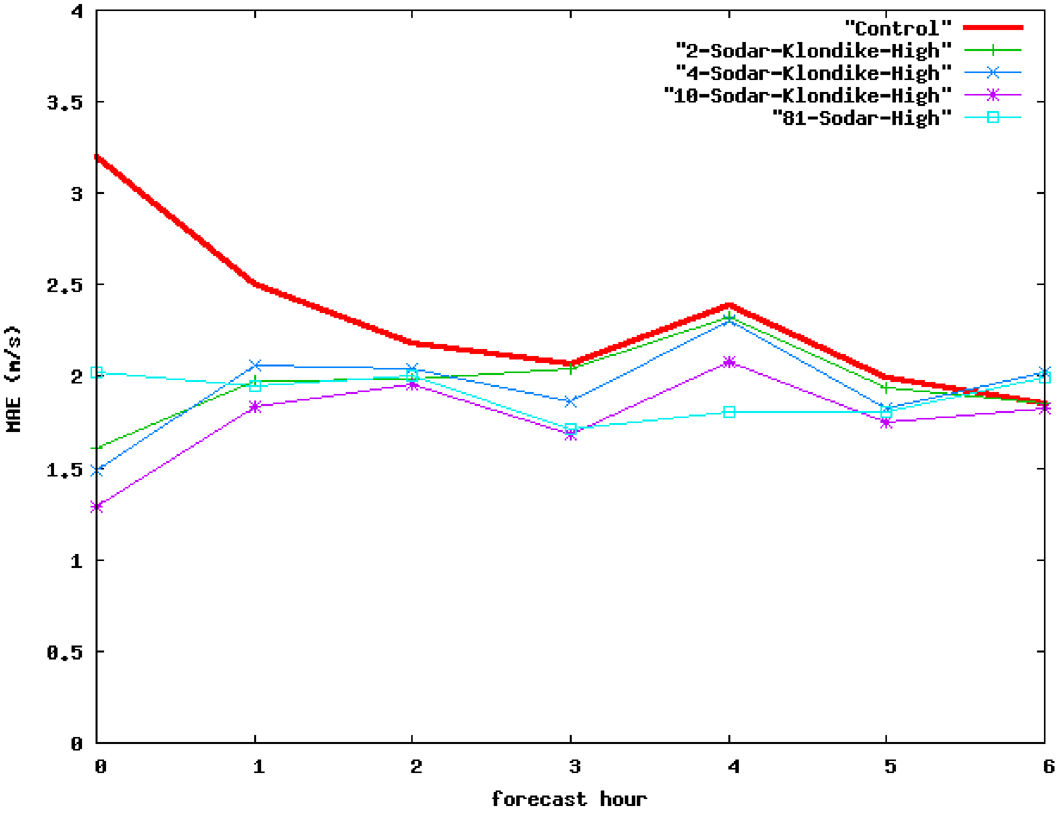

Figure 9. *Mean absolute error (MAE) of 80-m wind speed at Klondike for multiple sodar warm season experiments.*

4.2 Cool Season 2010

Figure 10 shows the results from the cool season experiments. The 80-m wind speed MAEs for the cool season Control experiment were significantly smaller than for the warm season. In general, atmospheric models perform better when the large-scale influences are greater, so that conventional rawinsonde and surface observations are better able to capture the most significant features of the initial state of the atmosphere.

Errors for the 1-Tower-Klondike experiment were almost identical to the Control experiment, and 1-Sodar-Klondike was only slightly better. The introduction of sodar data only at Klondike again improved the initial value there, but it did not improve the 1-h forecast, unlike the warm season experiment. Using a large number of sodars (81-Sodar-High) produced only a modest improvement at the initial time, no improvement at all for the first three hours, and a modest improvement at four and five hours.

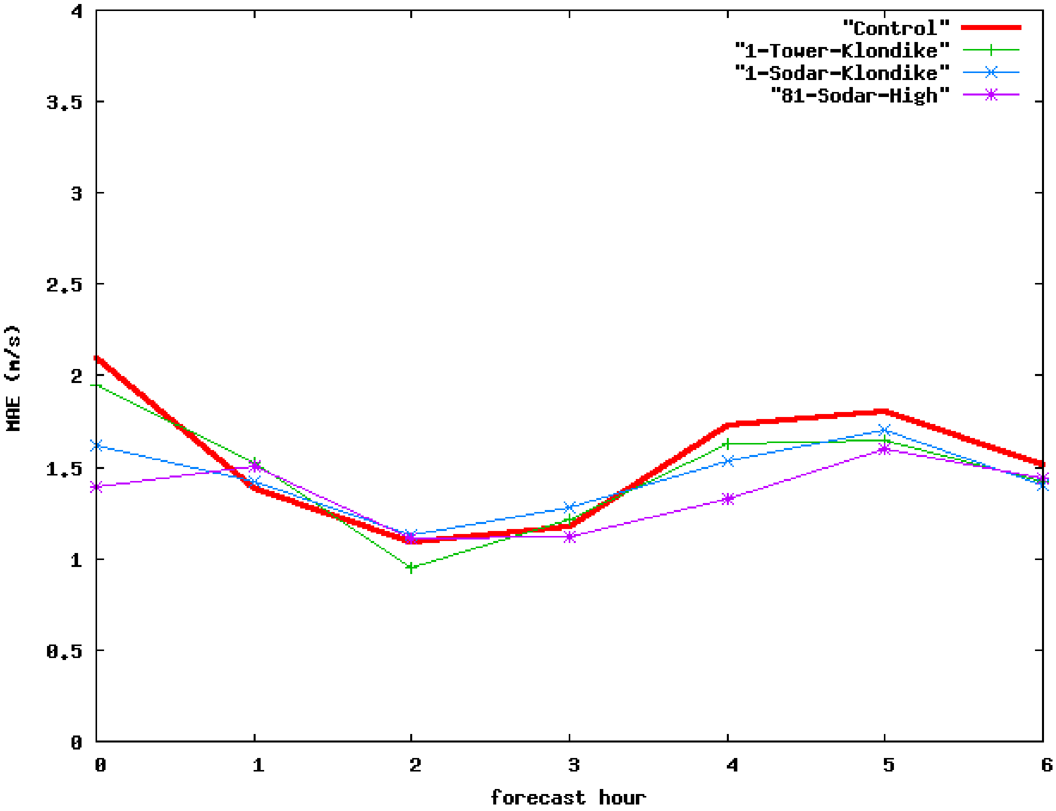

Figure 10. *Mean absolute errors (MAE) of 80-m wind speed at Klondike for all of the cool season (January 2010) experiments.*

5. Summary

In order to test ESA strategies for future instrumentation deployment, a set of OSSEs were performed for a 9-day period for both cool season and warm season periods. The control experiments assimilated synthetic rawinsonde and surface data from the nature runs. The remaining experiments tested various possible deployment scenarios involving 80-m met towers that measure temperature, pressure, dew point temperature, wind speed, and wind direction only at one level, and sodars that measure only wind between about 30 and 200 m above ground level. Further research is warranted to determine the best deployment strategy on a cost-benefit basis considering the number, types, and locations of future instrumentation.

The warm season results suggest that assimilating a single observation at the target location improved the initial conditions, but the forecast was only improved for the first hour. The addition of a modest number of sodars (ten or less) in locations suggested by the ESA analyses reduced the MAE of the 80-m wind speed from 10-20% through the first four to five hours of the forecasts. The use of sodar data, which has wind information over multiple vertical levels surrounding wind turbine hub heights, had a greater positive impact than meteorological tower data at a single level, despite the use of thermodynamic variables from tower data.

For the cool season, none of the OSSEs produced a significant improvement for the first three forecast hours, although there was some improvement in the fourth and fifth hours. It seems likely that it is more difficult to improve forecasts in the cool season when synoptic influences are greater and conventional data captures a greater portion of the atmospheric circulations that impact wind speeds in the Mid-Columbia Basin.

Task 2 results (in a companion report) indicated that the use of met tower data in locations with a high sensitivity for BPA sites could improve the short-term forecast, if the data assimilation system is properly configured to maximize the value provided by the observations. This is consistent with these Task 3 results – improvements in the forecast for the first few hours may be possible with a modest number of observations in high sensitivity areas, but more work is needed to find and tune the best assimilation procedure for this particular geographic area and set of observation types. Also, the finding that a small number of sodars performed better than an equivalent number of towers needs to be evaluated carefully.

6. Acknowledgements

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DEAC52-07NA27344. The work of the authors from AWS Truepower was performed under subcontract B591225 from Lawrence Livermore National Laboratory.

7. References

Atlas R., 1997: Atmospheric observations and experiments to assess their usefulness in data assimilation. *J. Royal Meteor Soc Japan*, **75**, 111–30.

Lord, S. J., E. Kalnay, R. Daley, G. D. Emmitt, and R. Atlas, 1997: Using OSSEs in the design of the future generation of integrated observing systems. Preprints, First Symposium on Integrated Observing Systems, Amer. Meteor. Soc., Long Beach, CA, 45-47.

Xue, M., K. K. Droegemeier, V. Wong, 2000: The Advanced Regional Prediction System (ARPS)—A multiscale nonhydrostatic atmospheric simulation and prediction tool. Part I: Model dynamics and verification. *Meteor. Atmos. Phys*., **75**, 161–193.

Xue, M., K. K. Droegemeier, and V. Wong, A. Shapiro, K. Brewster, F. Carr, D. Weber, Y. Liu, and D.-H. Wang, 2001: The Advanced Regional Prediction System (ARPS)—A multi-scale nonhydrostatic atmospheric simulation and prediction tool. Part II: Model physics and applications. *Meteor. Atmos. Phys*., **76**, 143–165.

Zack, J., E. Natenberg, S. Young, G. Van Knowe, K. Waight, J. Manobianco, and C. Kamath, 2010b: Application of ensemble sensitivity analysis to observation targeting for short term wind speed forecasting in the Washington - Oregon Region. Technical Report LLNL-TR-458086, 65 pp., Lawrence Livermore National Laboratory, Livermore, CA.