J4.4 HOW WELL CAN WE MODEL THE HOUSTON AREA FOR AIR POLLUTION STUDIES?

Wayne M. Angevine[1,2], Mark Žagar[3], Jerome Brioude[1,2] , Robert Banta[2],
Christoph Senff[1,2], HyunCheol Kim[4] , and Daewon Byun[4]
[1]Cooperative Institute for Research in Environmental Sciences (CIRES), University of Colorado, Boulder, Colorado, USA
[2]NOAA Earth System Research Laboratory, Boulder, Colorado USA
[3]Meteorological Service of Slovenia, Ljubljana
[4]Institute for Multi-dimensional Air Quality Studies, University of Houston, USA

1. INTRODUCTION

The evaluation of model results is vital to the improvement of models. Statistics are one common approach, case studies another. A third possibility is to examine specific features of interest for a particular application. Here, we are interested in evaluating the results of mesoscale Eulerian numerical model runs to decide whether changes to the modeling system improve results for air quality applications. We will show statistics, but finding them not entirely satisfactory, will also show several new metrics related to specific features, in particular, the sea breeze, stagnation, and recirculation. We will also show preliminary results from a Lagrangian particle dispersion model, driven by our Eulerian results, which may allow us to evaluate the model by examining the placement of point-source pollutant plumes.

Many groups have run models over the Houston area and reported results (Angevine et al. 2008; Bao et al. 2005; Cheng and Byun 2008; Cheng et al. 2008; Zhang et al. 2007). Our goal here is *not* to claim that the results we present are superior. In fact, our goal is to demonstrate the difficulty of supporting such a claim, and contribute to the discussion about how such claims can be evaluated.

The second Texas Air Quality Study (TexAQS II) was held in 2006. Many groups contributed resources, including several aircraft, ground chemistry sites, and a heavily-instrumented ship. These resources augmented the operational chemistry and meteorology monitoring network operated by state and local authorities.

We present results from runs of the Advanced Research core of the Weather Research and Forecasting model (WRF-ARW) model system for 75 days at 5 km grid spacing. Results from shorter runs with a finer grid have been shown by Angevine et al. (2008). That presentation included comparisons with mixing heights measured by Doppler lidar, and with surface fluxes, over the waters of the Bay and Gulf. Three WRF configurations are presented here: With four-dimensional data assimilation (FDDA) of three wind profilers; FDDA plus 1-hour water surface temperatures; and no FDDA. All three use soil moisture values reduced over land to produce approximately correct near-surface temperatures, since the default soil moisture tables produced temperatures that were substantially too cool. All runs were initialized at 0000 UTC every day from the ECMWF analyses, and used analysis nudging to the 6-hourly ECMWF analyses on the outer (15-km spacing) grid. All the WRF configurations used the MYJ PBL scheme, 5-layer thermal diffusion ("slab") land surface, RRTM longwave and Dudhia shortwave radiation.

For comparison, we primarily use surface meteorology measurements from the operational air-quality monitoring network in the Houston area. These measurements have the advantage of continuous availability, and

---

Corresponding author: Wayne M. Angevine, NOAA ESRL R/CSD04, 325 Broadway, Boulder, CO 80305-3337, email: Wayne.M.Angevine@noaa.gov, tel: (303)497-3747

are representative of the types of measurements available in many locations when no major field campaign is in progress. We also use wind measurements from the radar wind profiler at LaPorte. Figure 1 shows the locations of the sites used.

2. TRADITIONAL STATISTICS

The basic figures of merit for model – measurement comparisons are systematic and random error, also known as accuracy and precision, bias and scatter, and many other terms. Of the many possible statistics that could be shown, we present a small selection. Figure 2 shows the mean difference (bias) and standard deviation of the difference (random error) of the wind speed and direction every 3 hours at 7 sites over the entire 75-day run. One challenge is immediately apparent: The ECMWF analysis used to initialize the WRF model performs better than any of the WRF configurations for some sites and measures. However, there is a general tendency for FDDA to improve the random error of wind direction.

Figure 3 shows the random error of wind speed and direction by time of day for a single site, C45. This is the closest site to the Bay and Gulf, so it would be expected to have the most sea breeze influence. Only data for the 17 days with ozone levels of regulatory concern are included. Here, FDDA improves random error in both speed and direction. The use of 1-h SST (rather than default daily values from ECMWF) improves random error in the afternoon, but makes it worse at night. ECMWF has different but comparable errors, but WRF with FDDA is better at hours 18 and 21in the critical afternoon period.

3. DISTRIBUTION OF ERRORS

Means and standard deviations can mask important information that can be revealed by looking at the actual distribution. If we compare the 2-m temperature between the model run with FDDA and the measurements (figure 4), we find that 10 days have at least one hour with temperature difference > 5K at site C35 (28 hours total). All differences > 5K have model > measurement (model too warm). All 10 days have convection or a cold front in reality. The model also has clouds and fronts but different amount, timing, or location.

4. NEW METRICS

Having found that traditional statistics do not clearly demonstrate the improvements we expected to find by running a better modeling system, we devised other metrics. Analysis of air quality in Houston has shown that the sea breeze is implicated in the most severe episodes. Stagnation is also a factor in Houston and many other areas.

The correspondence between sea breeze occurrence in each model run and in the surface observations is shown in figure 5a. This is a measure of how often a sea breeze occurs simultaneously in the simulation and the measurement. The figure shows results when a sea breeze is defined as a northerly component >1 m/s between 0600 and 1200 UTC and a southerly component >1 m/s after 1200 UTC. Many other definitions are possible, but the results are not particularly sensitive to the threshold speed. The runs with FDDA alone or with FDDA and 1-h SST are better (closer to the measurement, that is, to 1) at all 7 sites, although at two sites the improvement is marginal.

A measure of stagnation, defined as wind speed < 1 m/s at any hour between 1500 and 2300 UTC, was not helpful in distinguishing the performance of the different model runs (figure 5b).

5. NET TRAJECTORY DISTANCE

Banta et al. (2009a, b) found that the meteorological variable that correlated best with daily peak ozone in Houston was the net distance traveled by a trajectory. Here we show how the equivalent distance computed from the WRF simulations behaves. Figure 6a shows the relationship between the net distance and the total distance. Trajectories were calculated from 10-minute WRF winds starting midway along the Ship Channel at 1400 UTC each day, extending for 10 hours, at a constant model level of approximately 190 m AGL. The net distance is the distance between the start and end of the 10-h trajectory. The total distance is the distance traveled along the trajectory. The difference is thus a rough measure of recirculation. The lower left portion of the diagram is of most interest since this is where winds are light. There are a number of days on which the trajectories have considerable curvature (difference between net and total distance). Figure 6b is a scatter plot of peak ozone vs. net trajectory distance. The correlation is rather good, with correlation coefficient r = -0.85 and r^2 = 0.72. The correlation with total distance is worse (r = -0.57). The results shown in figure 6 are from the run with FDDA but without 1-h SST. The run with 1-h SST produces similar results.

6. AVERAGE WIND

Can a simpler representation of the wind field perform as well as net trajectory distance? We show in figure 7a the relationship between vector average wind and average wind speed for each of the 75 days. The average wind speed is simply the speed for each 10 minutes, averaged over 10 hours (1400-2400 CST). The vector average wind is found by averaging the u and v components over 10 hours and computing the speed from those. It thus accounts for recirculation. In an extreme example, if the wind blows from the south at 6 m/s for 5 hours and then switches to the north at the same speed for 5 hours, the average speed is 6 m/s but the vector average wind is zero. The figure shows no such extreme departures, but some significant differences between the two averages (points off the 1:1 line). As for the trajectories above, the winds used here are those simulated at 190 m AGL.

There is a good relationship (not shown) between (either) average wind in the column above LaPorte from the model runs with FDDA and the wind profiler measurement. This is not too surprising since the profiler data was assimilated, so the comparison is not independent.

Figure 7b shows the relationship between the vector average wind and the airborne maximum ozone (Banta et al. 2009). This is a fully independent comparison. The correlation coefficient is r = -0.91 or r^2 = 0.83. The correlation with the average wind speed is almost as good , r = -0.88. Both are a little better than found above for the net trajectory distance. The results shown in figure 7 are from the WRF run with FDDA and 1-h SST. In this case, the run without 1-h SST performs about the same. The run without FDDA is clearly worse (not shown).

7. FLEXPART RESULTS

Figure 8 shows results of running the FlexPART Lagrangian particle diffusion model (Stohl et al. 2005) driven by the WRF output with FDDA. Real emissions of SO_2 are used to drive the model, and the resulting concentrations are plotted along the flight track of the NOAA P3 on 26 September 2006. A similar run driven by ECMWF model output with horizontal grid spacing of 0.25x0.25 degree is also shown. The WRF fields produce much better agreement in the position and strength of the point-source plumes.

8. SUMMARY

The search for metrics that clearly indicate whether the addition of complexity to the model system in the form of data assimilation or improved water surface temperatures is at best partially successful. We begin with the handicap that the ECMWF analysis used for initialization is already quite good, as good as any results in the literature. However, we expect that simulations with finer resolution, such as those shown here, should resolve fine-scale features driven by the complex coastline effects better. Even this is difficult to conclusively demonstrate. Part of the explanation for that difficulty may be related to the unique nature of the sea breeze at 30 degrees latitude (see sidebar in (Banta et al. 2005). The large scale of the coastal oscillation near that critical latitude may make finer resolution less useful than at other locations. Another factor may be the lack of terrain and the relatively small contrasts in temperature between land and water in the Houston/Galveston area. Finer resolution is likely to be more important in areas with coastal mountains and/or cold water.

Traditional statistics (bias and standard deviation), whether computed over the full period of time or only on episode days, by individual sites or time of day, do not crisply display differences between runs. They do, however, generally indicate improvement with FDDA of wind profiler data.

Looking at the distribution of errors is clearly useful in diagnosing when and why major problems occur. In these simulations, large errors in temperature (>5K) occurred when moist convection was present in reality.

The new metric of sea breeze correspondence shows improved model performance at all 7 surface sites with FDDA. The stagnation metric does not demonstrate improvement.

Net trajectory distance correlates better with observed ozone than total distance. It does not clearly distinguish between the runs with and without 1-h SST.

The average wind over the middle of the day is a very good predictor of maximum ozone. The vector average is slightly better than the scalar average, but the difference may not be significant. Runs with FDDA are clearly better than without by this metric.

The Lagrangian plume model provides clear information about directly relevant performance of the model, that is, the simulated wind field. It is not obvious how to encapsulate that information in an easily reportable form.

Uncertainty analysis is needed to establish the significance of both traditional and new metrics.

In conclusion, we find that assimilation of wind profiler data improves the model results overall, primarily by reducing the random error in wind direction. The improvement is most easily seen in the wind profile away from the surface, and confirmed by a tight correlation with measured ozone.

Some philosophical issues must be considered. How good is good enough? What if we know we have improved the model, but can’t show that we have improved the results? Are there fundamental or practical limits to predictability of the phenomena most relevant to air quality, such as stagnation?

ACKNOWLEDGEMENTS:

We are grateful to Bryan Lambeth and Ben Coughran from the Texas Commision on Environmental Quality, who oversaw the LaPorte wind profiler and provided its data. Tom Ryerson provided the NOAA P3 ozone data, and John Holloway provided the P3 SO_2 data. WMA acknowledges partial funding from Vaisala Inc.

REFERENCES

Angevine, W. M., M. Zagar, S. C. Tucker, C. W. Fairall, L. Bariteau, D. E. Wolfe, and W. A. Brewer, 2008: Modeling the boundary layer over Galveston Bay and the Gulf of Mexico for air pollution studies. Preprints/*Proc.*/*Extended Abstract 18th Symposium on Boundary Layers and Turbulence*, Stockholm, Sweden, American Meteorological Society.

Banta, R. M. and coauthors, 2009a: Dependence of peak daily ozone concentrations in Houston, Texas on the sea breeze and meteorological variables. Preprints/*Proc.*/*Extended Abstract 8th Symposium on the Urban Environment*, Phoenix, AZ, American Meteorological Society.

Banta, R. M., C. J. Senff, J. W. Nielsen-Gammon, L. S. Darby, T. B. Ryerson, R. J. Alvarez, S. P. Sandberg, E. J. Williams, and M. Trainer, 2005: A bad air day in Houston. *Bull. Amer. Meteor. Soc.*, **86,** 657-669.

Banta, R. M., C. J. Senff, D. D. Parrish, R. J. Alvarez, L. S. Darby, R. M. Hardesty, B. Lambeth, A. B. White, M. Trainer, T. B.

Ryerson, J. A. Neuman, W. M. Angevine, and S. J. Sandberg, 2009b: Dependence of daily peak O3 concentrations near Houston, Texas on environmental factors: Wind speed, temperature, and boundary-layer depth. *In preparation*.

Bao, J.-W., S. A. Michelson, S. McKeen, and G. A. Grell, 2005: Meteorological evaluation of a weather-chemistry forecasting model using observations from the Texas AQS 2000 field experiment. *J. Geophys. Res.*, **110,** D21105, doi:10.1029/2004JD005024.

Cheng, F.-Y. and D. W. Byun, 2008: Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area: Part I Meteorological simulation results. *Atmos. Environ.*, **In press**.

Cheng, F.-Y., S. Kim, and D. W. Byun, 2008: Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area: Part II Air quality simulation results. *Atmos. Environ.*, **42,** 4853-4869.

Stohl, A., C. Forster, A. Frank, P. Seibert, and G. Wotawa, 2005: Technical note: The Lagrangian particle dispersion model FLEXPART version 6.2. *Atmospheric Chemistry and Physics*, **5,** 2461-2474.

Zhang, F., N. Bei, J. W. Nielsen-Gammon, G. Li, R. Zhang, A. Stuart, and A. Aksoy, 2007: Impacts of meteorological uncertainties on ozone pollution predictability estimated through meteorological and photochemical ensemble forecasts. *J. Geophys. Res.*, **112,** D04304, doi:10.1029/2006JD007429.

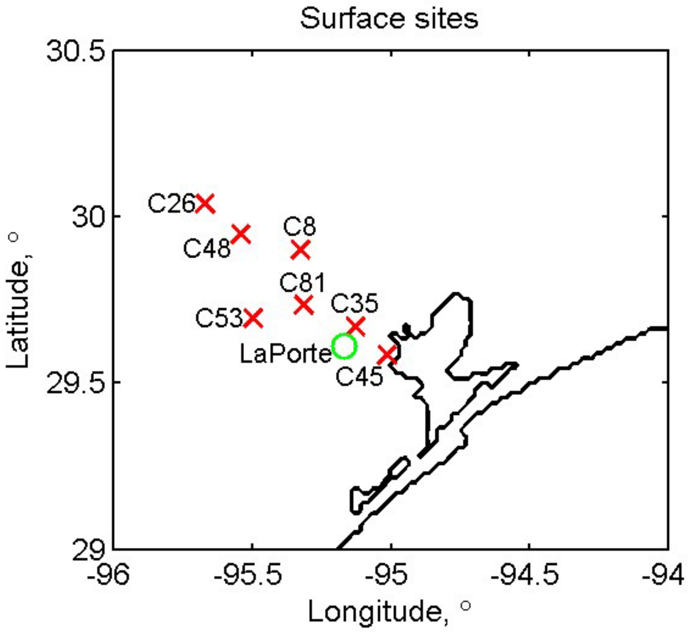

Figure 1: Map of surface sites used in comparisons. The sites are all within the Houston metropolitan area. Site C45 is near the western edge of Galveston Bay. The Gulf coast runs diagonally across the lower right quadrant.

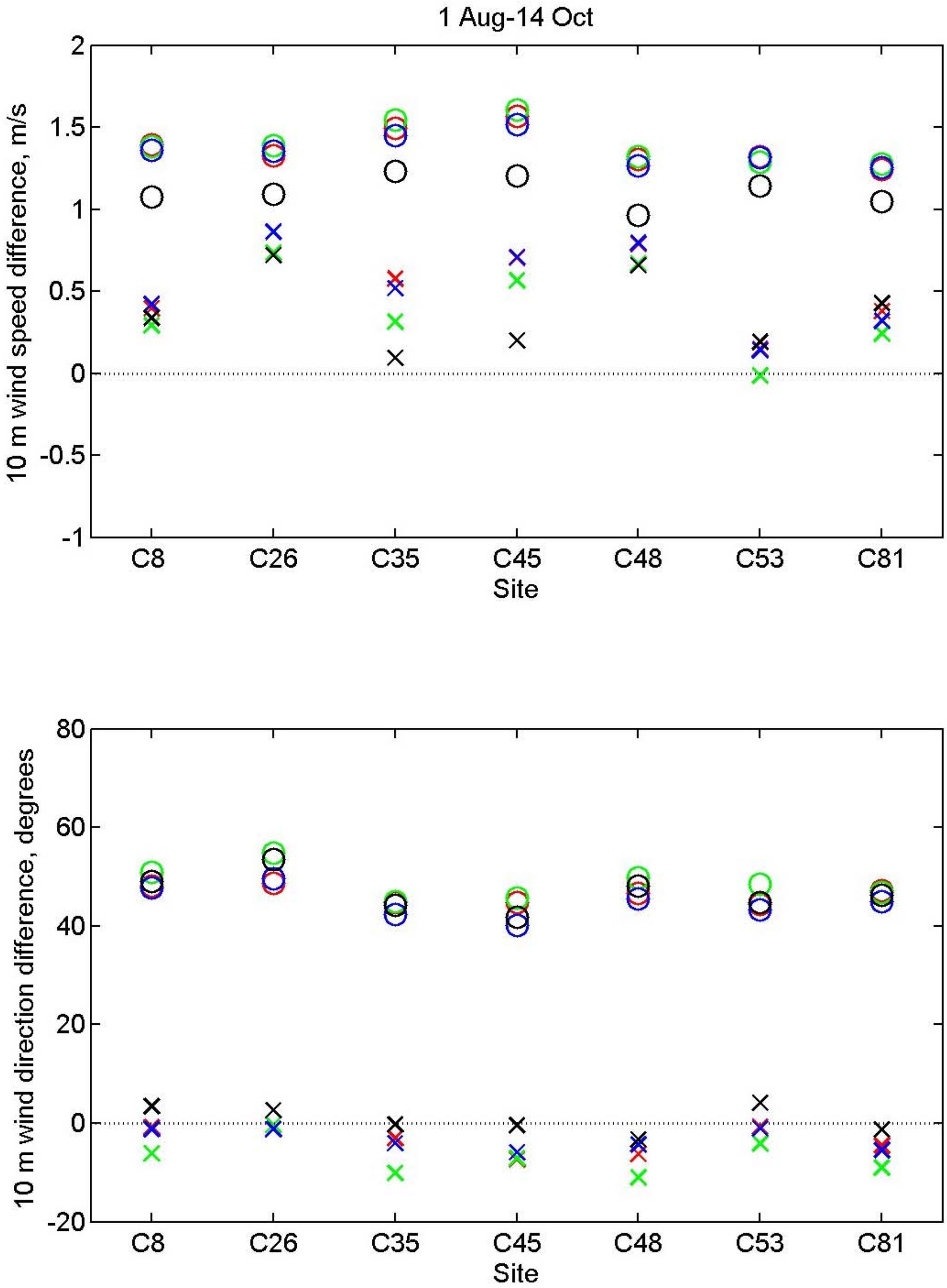

Figure 2: Mean (X) and standard deviation (O) of wind speed and direction difference between model and observation at 7 surface sites every 3 hours for the entire 75-day run. The runs are color coded: Black = ECMWF analysis, green = WRF without FDDA, red = WRF with FDDA, blue = WRF with FDDA and 1-h SST.

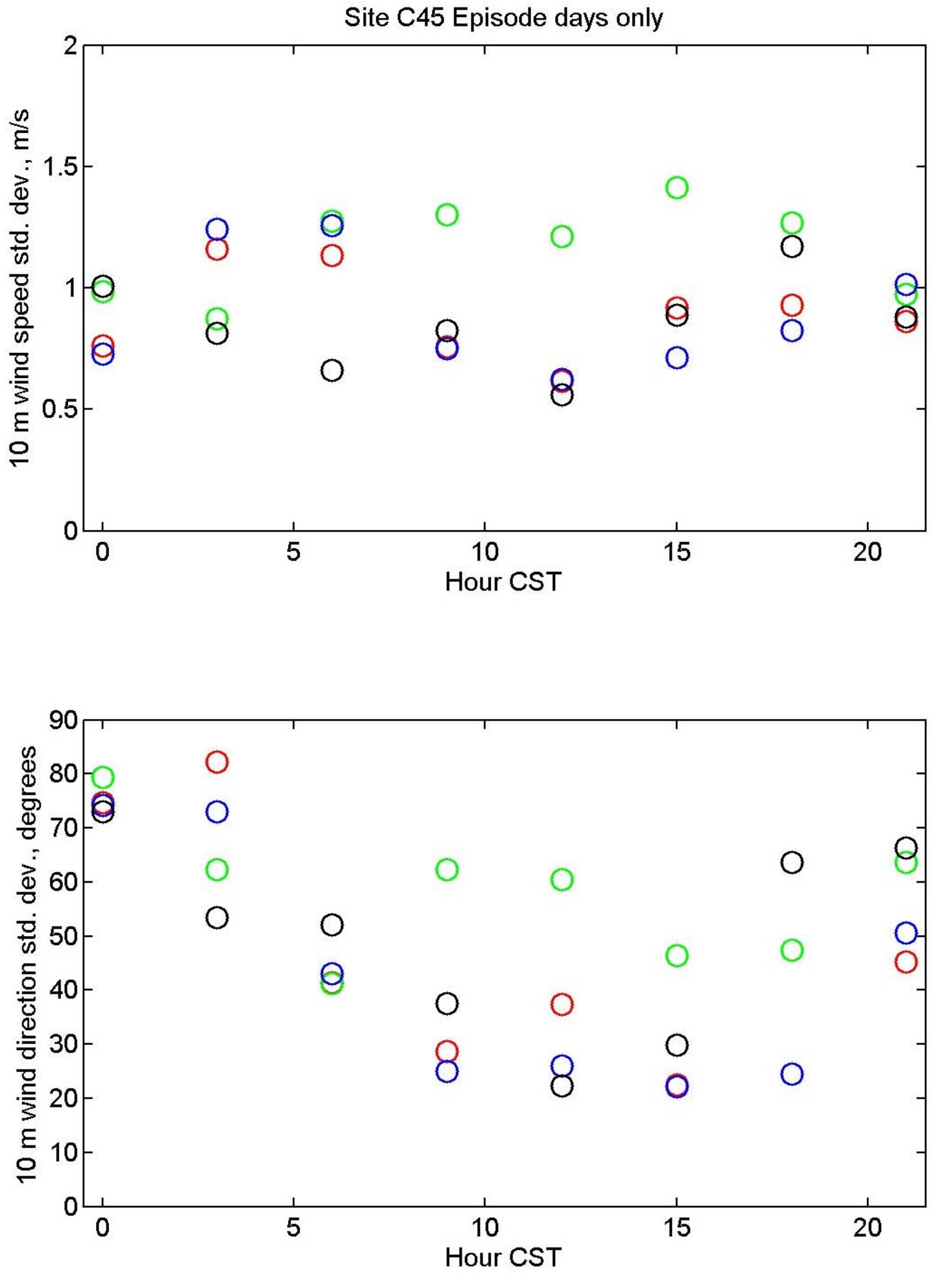

Figure 3: Standard deviation of wind speed and direction difference between model and observation at site C45 by hour of day (every 3 hours) for the 17 ozone episode days. The runs are color coded: Black = ECMWF analysis, green = WRF without FDDA, red = WRF with FDDA, blue = WRF with FDDA and 1-h SST.

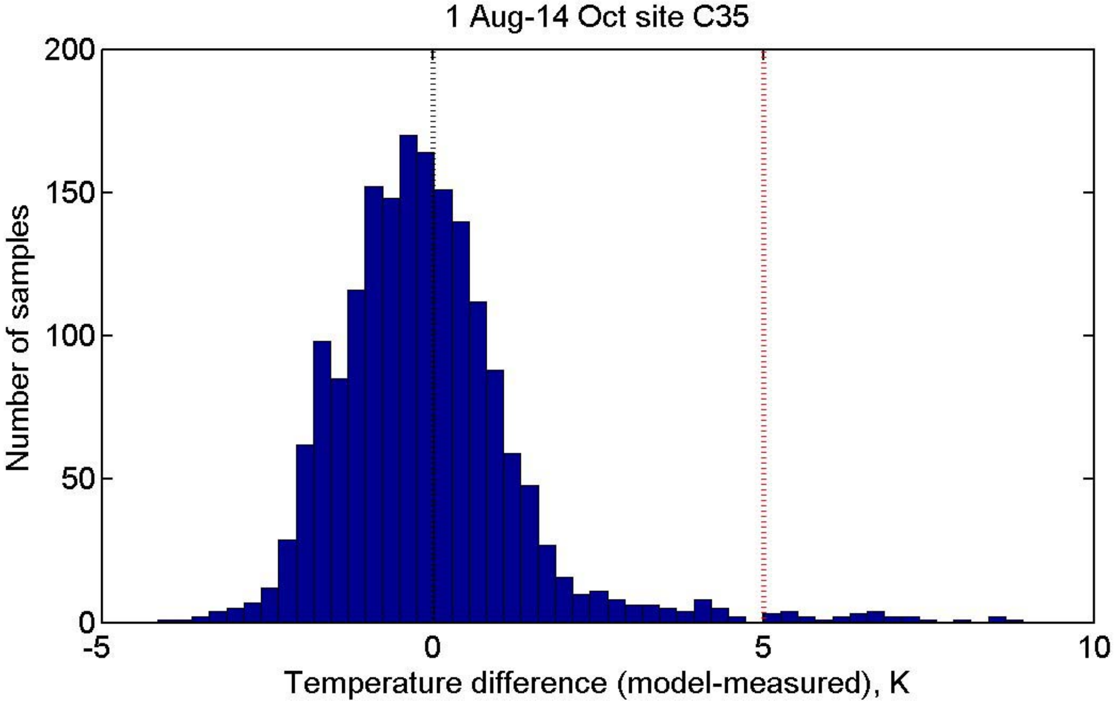

Figure 4: Distribution of hourly temperature differences at site C35 for all 75 days.

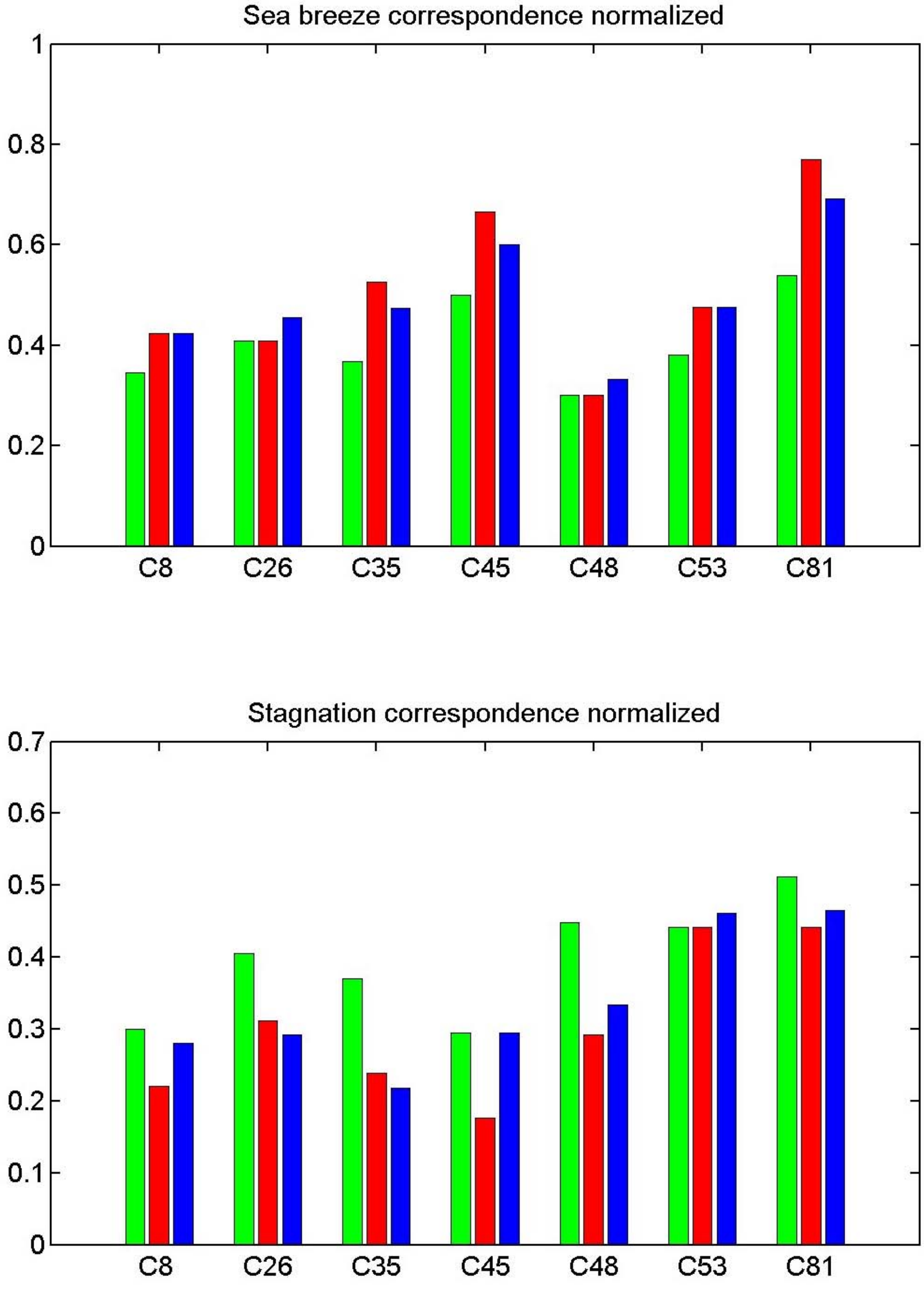

Figure 5: Correspondence between sea breeze (top) or stagnation (bottom) in each of the model runs and the measurements at 7 surface sites. Perfect agreement would be indicated by a value of 1. See text for definitions. The runs are color coded: Green = WRF without FDDA, red = WRF with FDDA, blue = WRF with FDDA and 1-h SST.

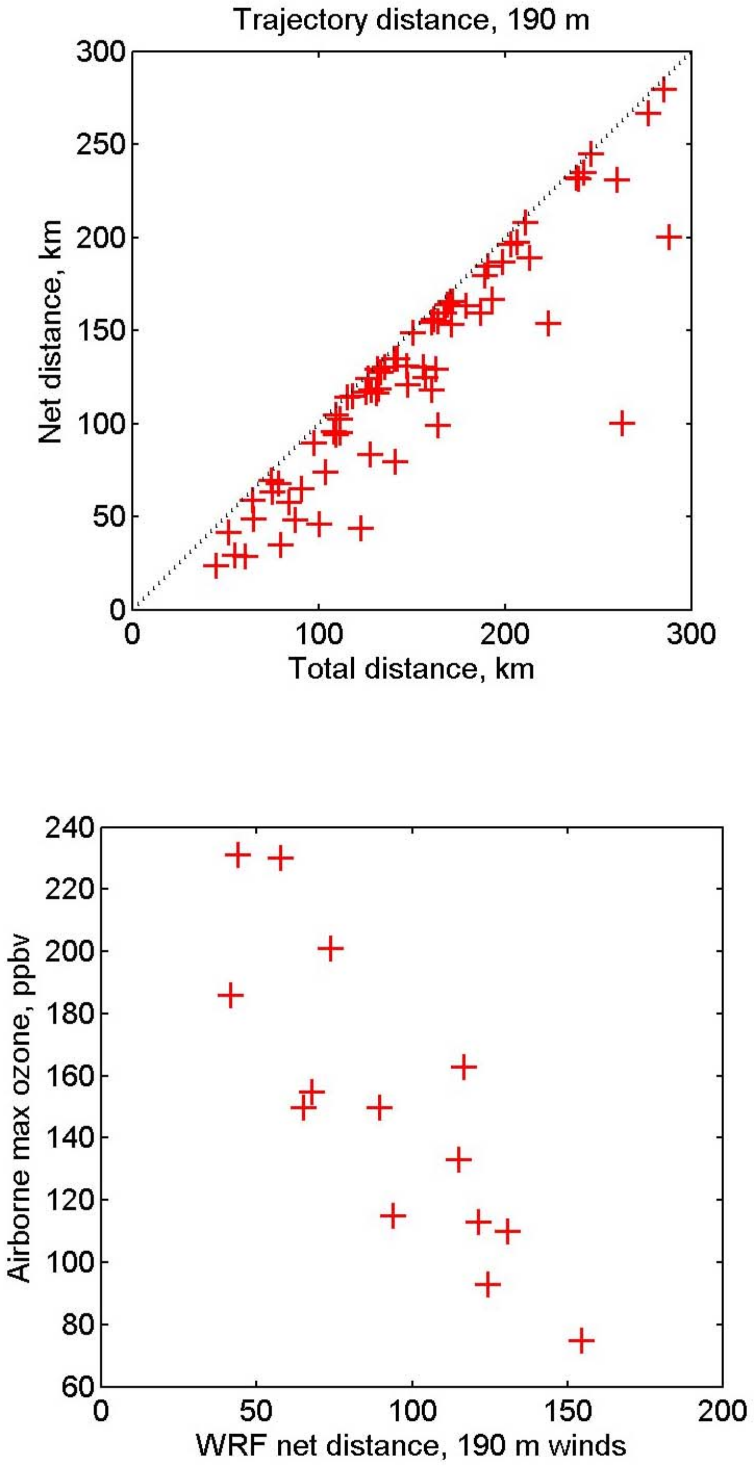

Figure 6: Total vs. net distance along 10-h trajectory for all 75 days (top). Maximum ozone observed by airborne platforms on 14 days vs. net trajectory distance (bottom). See text for details of trajectory calculation.

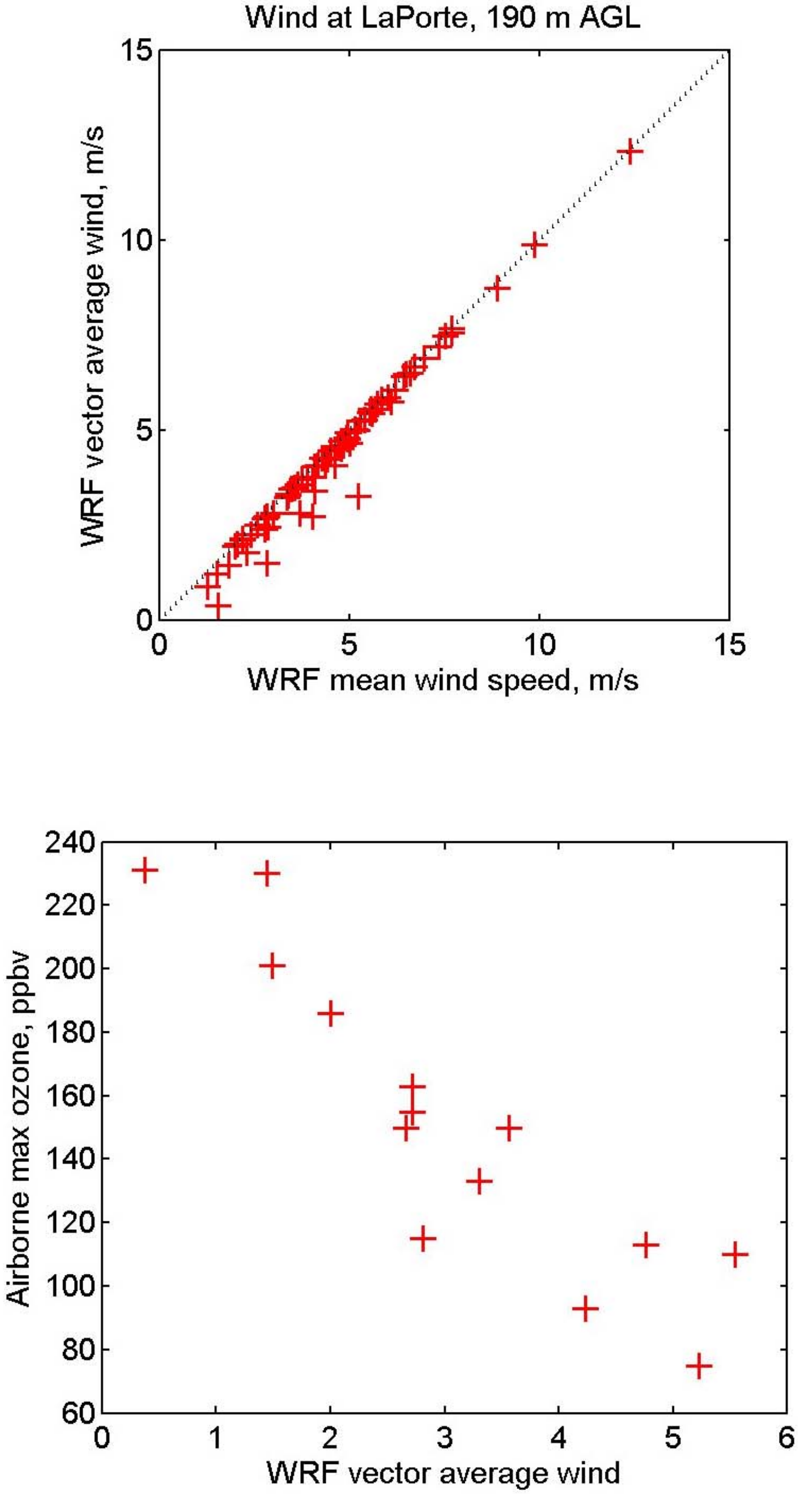

Figure 7: Vector average wind versus average wind speed at LaPorte (top). Averages are over 10 hours, 1400-2400 UTC. Maximum ozone observed by airborne platforms on 14 days vs. vector average wind from WRF (bottom).

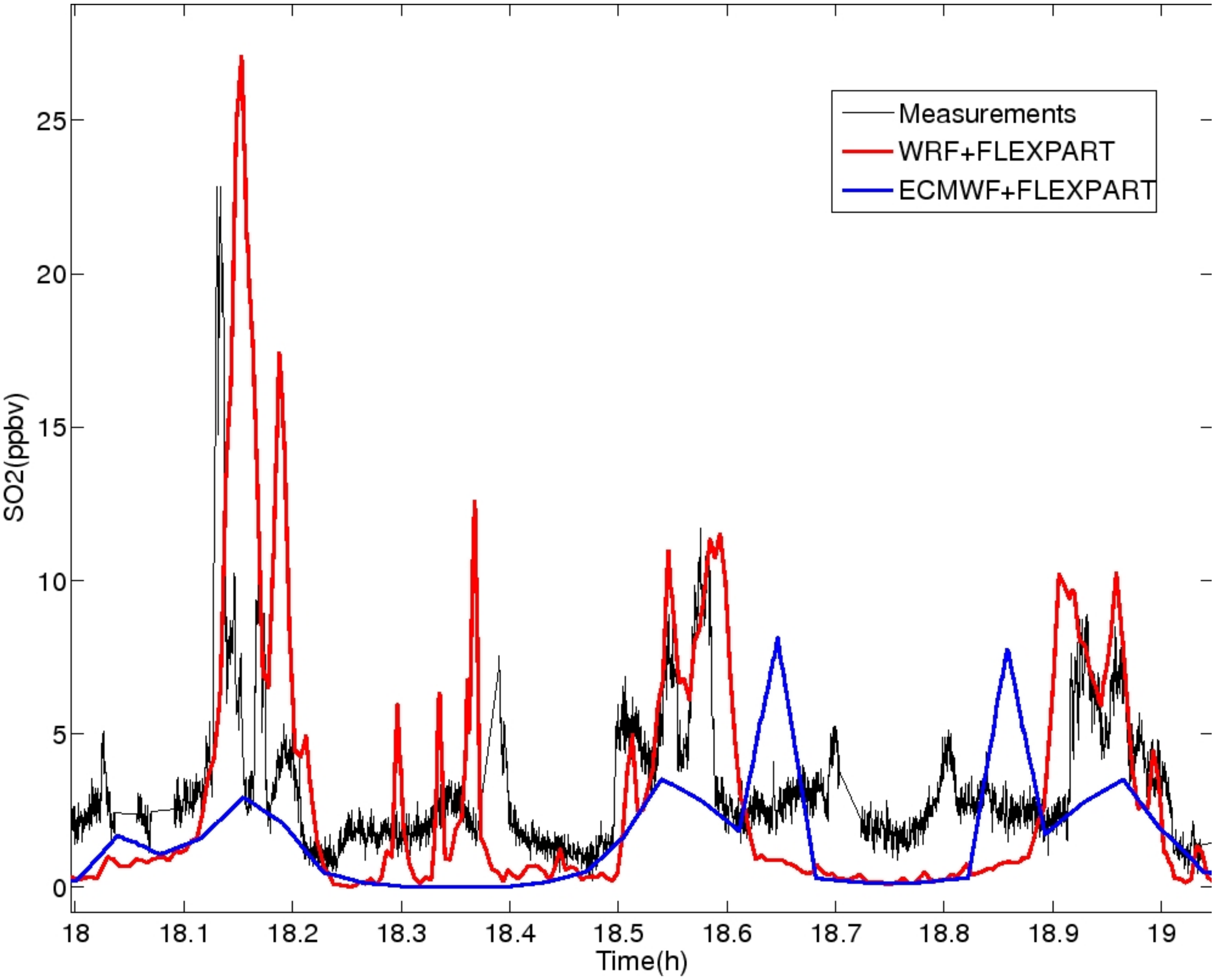

Figure 8: FlexPART model results when driven by WRF run with FDDA (red) and by ECMWF (blue). The model is given real emissions. A 1-h segment of a flight by the NOAA P3 is shown (measurements in black).