Evaluating WRF Meteorological Downscaling Performance for Use in Air Quality Dispersion Modeling Studies

Gary Moore ▪ William Leatham ▪ Jeffery Connors ▪ Robert Paine (AECOM, Chelmsford, MA)

Introduction

Traditional model performance evaluation statistics used to evaluate prognostic meteorological models generally are not geared towards local near source air quality modeling.

Air quality modeling exercise protocols require local scale meteorology (100's of meters). Dynamical downscaling using a multi-nest application of a prognostic model like the Weather Research Forecast (WRF) model is handed off to a diagnostic or even a CFD model.

Additional information about the suitability of the WRF model output for the surface meteorology in the local modeling domain needs to be generated.

Several additional diagnostic statistical measures and graphics were developed to examine a WRF model output's suitability for input to the diagnostic meteorological model. These added performance tools are described and discussed in this presentation.

Motivation and Approach

Motivation: Improve air quality met model evaluation products
Approach: Add diagnostics to traditional WRF evaluation exercises

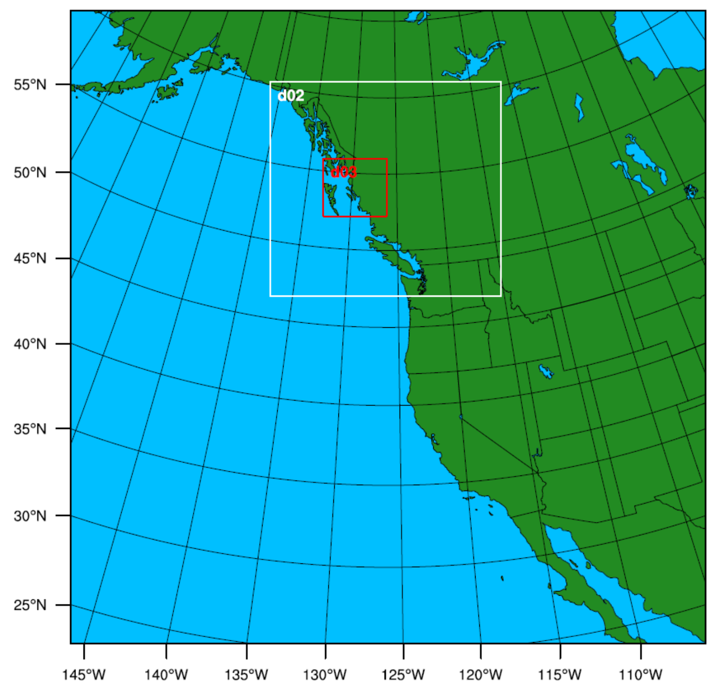
TRADITIONAL STATISTICS

- Mean bias
- Mean square error or its square root (MSE or RMSE)
- (absolute) Gross Error
- Index of Agreement (IOA)
- Normalization (e.g. differences over averages or division by observed)
- A priori performance targets

VALUE ADDED EVALUATION ATTRIBUTES

- Confidence intervals – robust, flexible, and meaningful
- Stratification of residuals -focus on conditions of interest
- Directionality - bias in trajectory headings is important
- Persistence - degree of hourly independence of residuals
- Accumulation metrics – duration and buildup of biases
- Episodic summaries – frequency of failures
- Classification of residuals – are there useful principle components?

WRF v3.51 ARW Prince Rupert, British Columbia Lambert Conformal Domain



EXERCISE SUMMARY

- Nested 36, 12, and 4 km domains with 37 vertical layers.
- Two-way nesting
- National Center for Environmental Protection (NCEP) Global Forecast System (GFS) for initialization conditions, boundary conditions and sea surface temperature (SST)
- Four Dimensional Data Assimilation (FDDA) analysis nudging is performed on the 36 and 12 km domains, while FDDA observational nudging (OBSGRID) is performed on the 4 km domain (u|v|T|q)
- Based on the University of Alaska Fairbanks (UAF) weather forecast model configuration

WRF Use: Dynamical Downscaling of Hourly Meteorology into CALMET Centered on Alder Met Tower

4 KM DOMAIN CHALLENGES

- Complex/rugged terrain
- Coastal environment
- Large seasonal surface variations (snow/ice)
- Limited observations (24 surface, 1 rawinsonde)
- Large changes in Aleutian low intensity and location for modeling years 2009,2010, 2012-2013

Summary of Traditional Evaluation Statistics

- Passing "grade" for all variables except wind speeds – always an over prediction
- 2012-2013 appears to be overall best predicted of the 3 years (had most observed data)
- Not much physical information to judge adequacy for air quality modeling
- 6 different WRF configurations were exercised as sensitivity runs with none showing significantly better performance

Parameter	Statistic	Statistical Benchmark		Average Values		
		Tesche	Complex Terrain	2009 Annual	2010 Annual	2012-2013 Annual
Wind Speed (m/s)	RMSE	≤ 2.0	≤ 2.5	3.16	3.26	2.85
	Bias	± 0.5		0.99	1.02	0.79
	IOA	≥ 0.6		0.71	0.70	0.72
Wind Direction (degrees)	Bias	± 10.0		4.02	6.30	2.53
	Gross Error	≤ 30.0	≤ 55.0	39.26	37.30	38.94
Temperature (K)	Bias	± 0.5	± 2.0	-0.03	-0.19	0.05
	Gross Error	≤ 2.0	≤ 3.5	1.38	1.32	1.21
	IOA	≥ 0.8		0.81	0.79	0.81
Humidity (g/kg)	Bias	± 1.0	± 1.0	0.28	0.34	0.23
	Gross Error	≤ 2.0	≤ 2.0	0.63	0.61	0.61
	IOA	≥ 0.6		0.68	0.71	0.71

Cross Correlation/ Factor Analysis

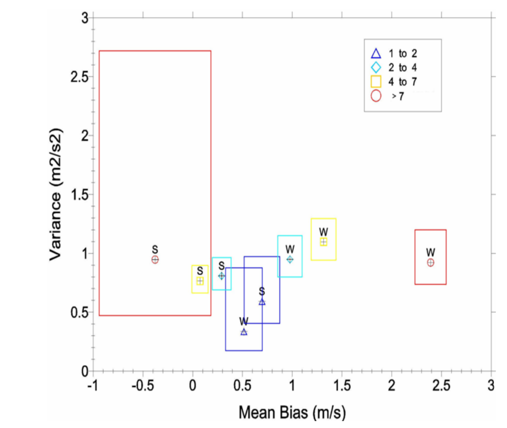
4km	9-Jul	10-Jul	13-Jul
ospd vs pspd	0.838	0.891	0.858
bdir vs pspd	0.332	0.332	0.449
odir vs pdir	0.978	0.960	0.953
otmp vs ptmp	0.874	0.946	0.851
ohum vs phum	0.788	0.799	0.792
bspd vs pspd	0.342	0.387	0.512
bhum vs phum	0.701	0.475	0.663
rmspd vs ospd	0.405	0.459	0.443
rmtmp vs otmp	0.463	0.643	0.477
rmhum vs ohum	0.333	0.270	0.215
otmp vs btmp	-0.177	-0.385	-0.162
otmp vs bhum	0.523	0.478	0.566
odir vs rmspd	0.137	0.148	0.256
odir vs rmhum	0.363	0.081	0.167

- Correlations of residuals with observations or predictions or each other are generally rather small
- Residual-variable factor analysis does not find one or two really dominant components (only 20-30% of residual variance each)

Footnotes

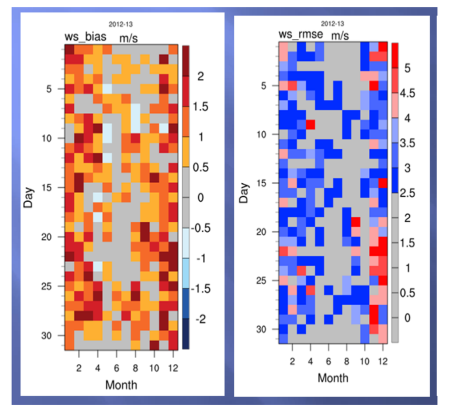
- Only correlations .ne. 0 passing Fisher Z at 95% confidence level are included
- (b) = bias
- (rm) = root mean
- (o) = observed
- (p) = predicted

Closer Look at Wind Speed ('Brickplot') Using 96% Confidence Intervals for 2009



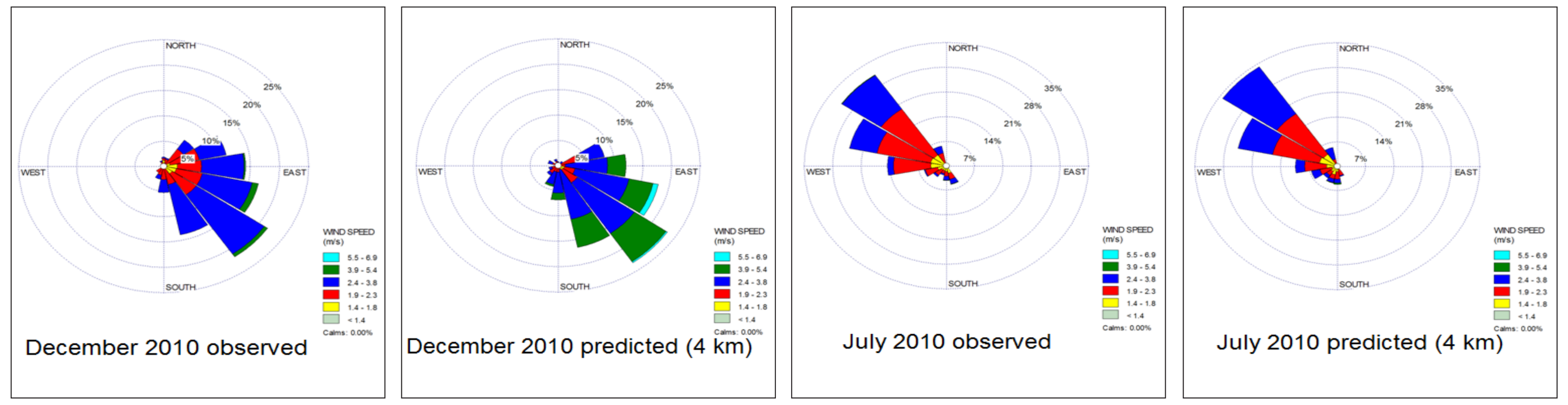
- Winter does better than summer for low wind speeds – opposite is true for higher wind speeds
- Summer residual biases are often not significant from zero for higher wind speeds
- Winter wind residuals show the greatest variability and bias

Bakergrams for Daily Wind Speed Statistics for 2012-2013

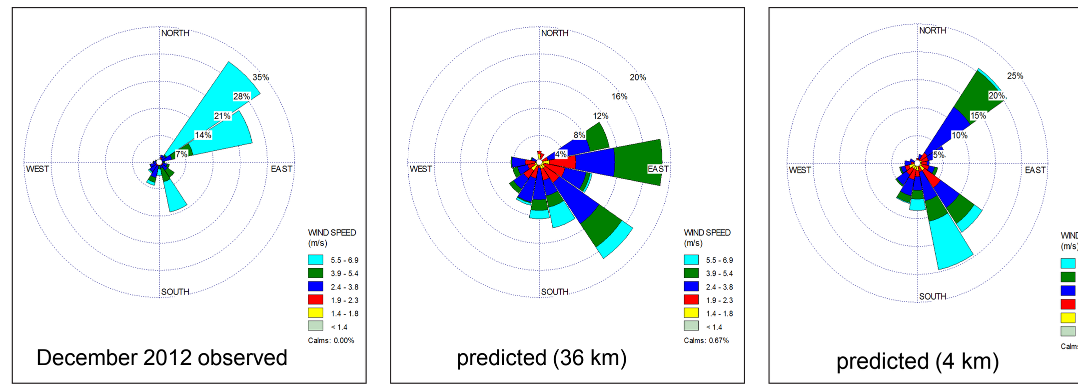


- Summer month wind speeds are better predicted with long runs of days of "passing"
- Bias has 29% days "passing"
- RMSE has 38% days "passing"

Monthly Station Ensemble Wind Rose Comparisons (2010) — High Degree of Agreement (Some Over Prediction)

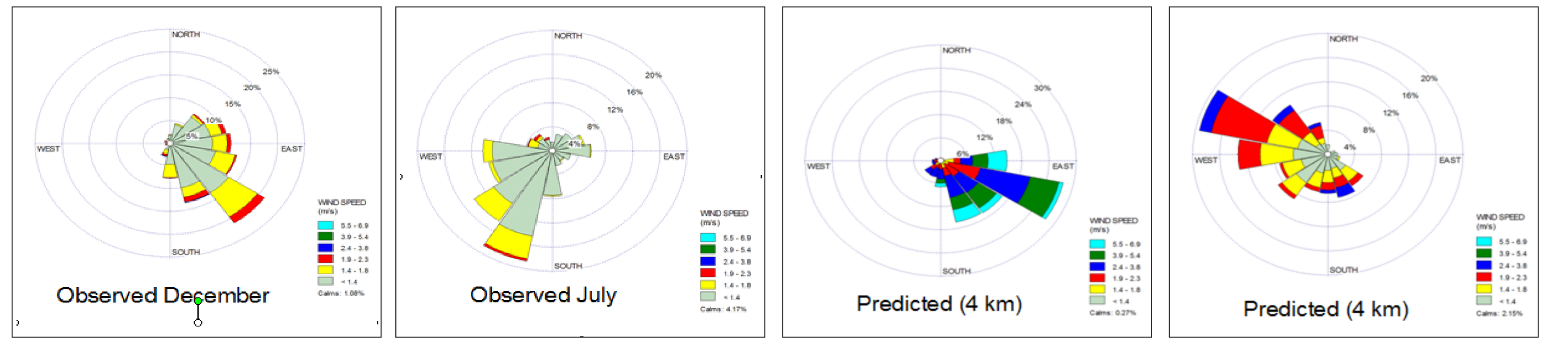


Grey Islet Wind Rose Comparison

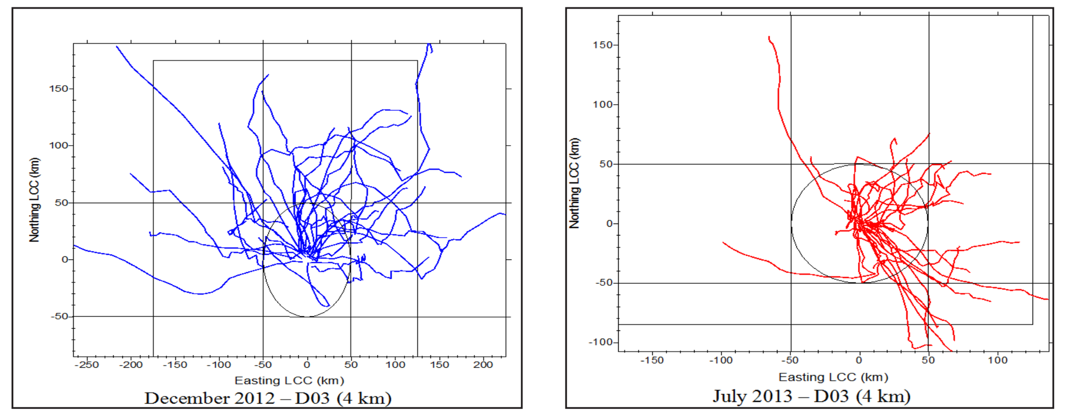


- Strong drainage present in observations does not appear in 36 km cell WRF wind rose
- Drainage is present in 4 km cell WRF wind rose but speeds are under predicted
- 4 km cell WRF winds retain an along coast peak that is not observed

Alder Station Wind Rose Comparison (2012-2013) — Topography Driven Biases in Speed and Direction

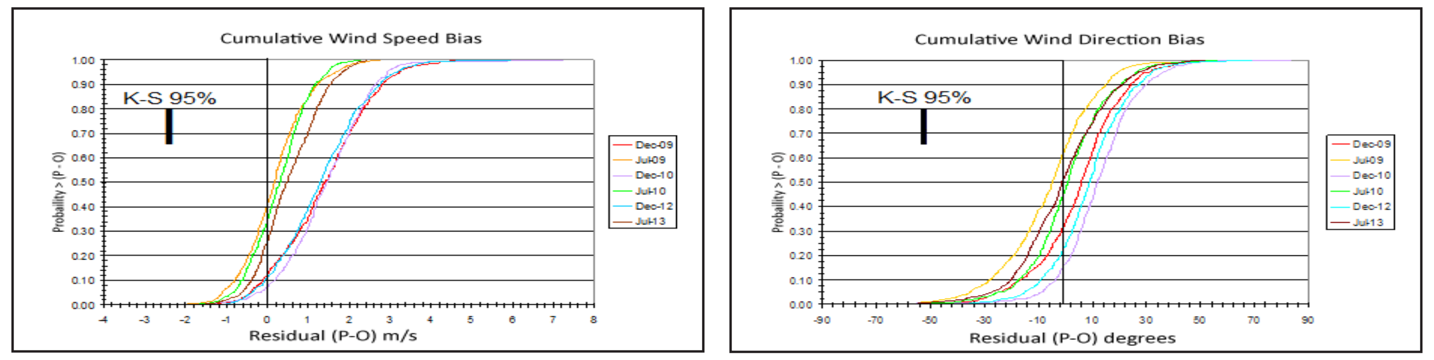


Trajectory Drift Plots (2012-2013)



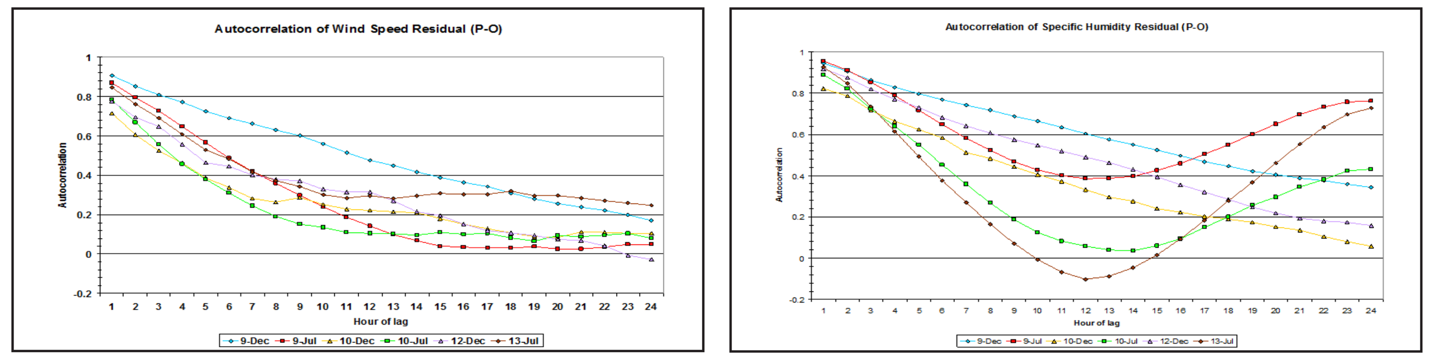
- Winter drift range due to stronger and more persistent winds is greater than summer
- Drift is biased northwards during winter and southwards during summer
- Probability of 50 km 'escape' is 77% during winter and 65% during summer

Summer-Winter Cumulative Distribution Comparisons



- K-S test indicates summer and winter wind speed residuals are significantly different
- Only December 2009 and 2012 wind speed residuals are similar and pass K-S test
- Largest failures occur near the average residual point for both speed and direction

Ensemble Station Time Persistence of Residuals



- Wind speed residuals decay monotonically
- Shortest speed residuals e-fold times - 6 hrs occur during summer (2010) and longest -17 hrs during the winter (2009)
- Summer humidity residuals have a diurnal cycle and in winter are monotonic
- Winter monotonic e-folding times for humidity range from 12 to 24 hrs between years suggesting a variation in duration of weather systems between observations and predictions

Conclusions

Traditional statistics tell us that only wind speed over predictions are a problem in the current application, with winter worse than summer. Diagnostic statistics as illustrated by some of the previous examples provide us far more information about residuals including.

- If modeled results require further downscaling and are suitable for direct use in CALMET
- When low wind speed and direction residuals are most likely to create persistent biases in transport
- How rapidly residuals of each of the variables become independent of each other
- The degree to which multi-day episodes of similar residuals can be related to specific weather conditions.
- What stations/locations may be most difficult to apply further downscaling modifications

These and other findings provide an indication of how the further downscaling of the meteorology for air quality worst case conditions should be conducted. In the current example the 4 km WRF output appears suitable to drive high resolution diagnostic modeling during the years 2009, 2010, and 2012-2013 when more refined terrain and roughness are considered.