High-Resolution Rapid Refresh Model Analytics in a High-*Throughput* Computing Environment

Brian K. Blaylock and John D. Horel
University of Utah Department of Atmospheric Sciences

Introduction

The National Centers for Environmental Prediction operational High Resolution Rapid Refresh (HRRR) model is primarily used for short-term forecasting and situational awareness

- Hourly analyses and 18-hr forecasts
- 3 km grid over contiguous United States (1.9 million grid points)
- Advanced data assimilation system

3 years of select HRRR output is archived at Utah's Center for High Performance Computing

http://hrrr.chpc.utah.edu/

Goal: Synthesize HRRR statistics from a 3-year archive via high-throughput computing

Options to compute basic model statistics from a large model data set:

1. Use small, dedicated multicore processors—*requires lengthy run times*
2. Use high performance computing facilities with large number of interconnected nodes—*not necessary*
3. Use high-throughput computing, like the Open Science Grid (OSG)—*an effective solution*

OSG farms out individual jobs to opportunistic compute cycles at computer facilities around the country

https://www.opensciencegrid.org/

Method:

Divide task into embarrassingly parallel jobs: A single Python script downloads and calculates percentiles from HRRR analyses (f00) within a 30-day window for every hour of the year from the three year archive

- 8,784 unique jobs (one job for every hour of the year)
- 90 analysis samples in each calculation (rolling 30 day window, times 3 years)
  - *Example: Percentiles for 1 June at 0000 UTC are calculated using the 0000 UTC analyses of the 15 days before and 14 days after 1 June from all the available years (2015-2017)*

Percentiles Calculated

00	01	02	03	04	05	10	25	33	50
66	75	90	95	96	97	98	99	100	mean

Jobs on OSG are managed and submitted with the Directed Acyclic Graph Manager (DAGMan) and HTCondor

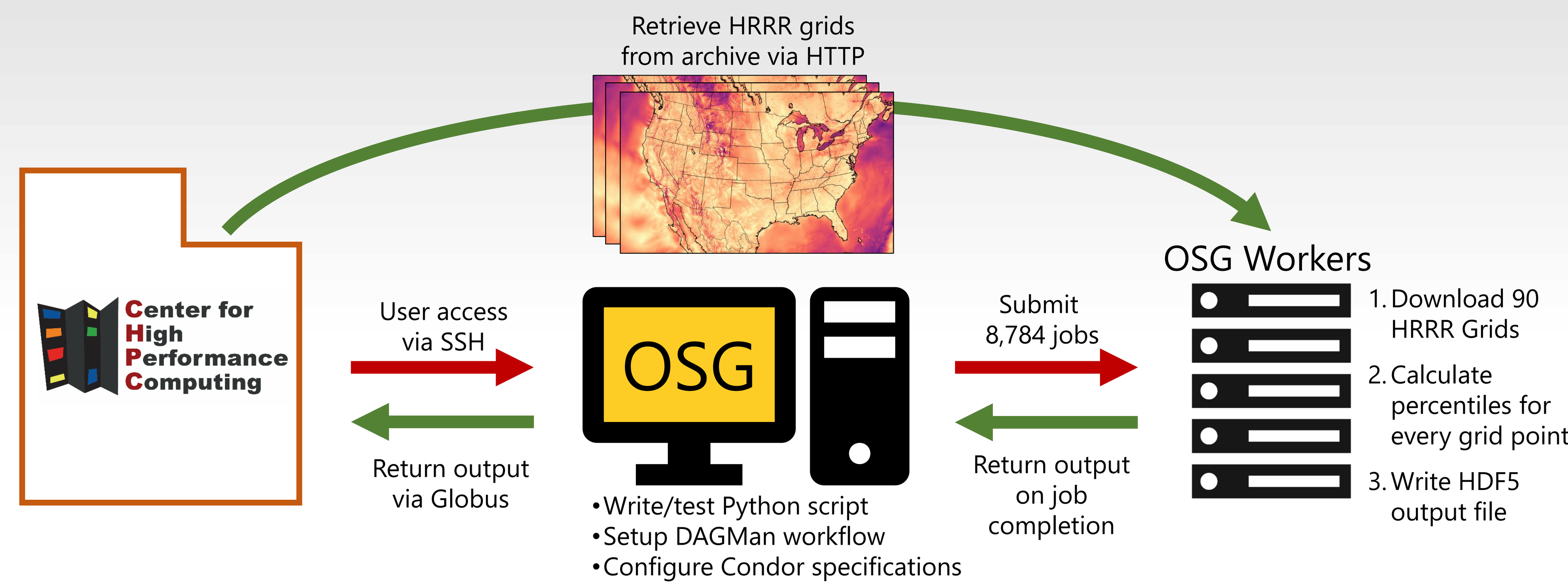
Each job runs on a remote OSG worker within a Singularity container

Output is transferred back to our local compute facility with a Globus bulk data transfer

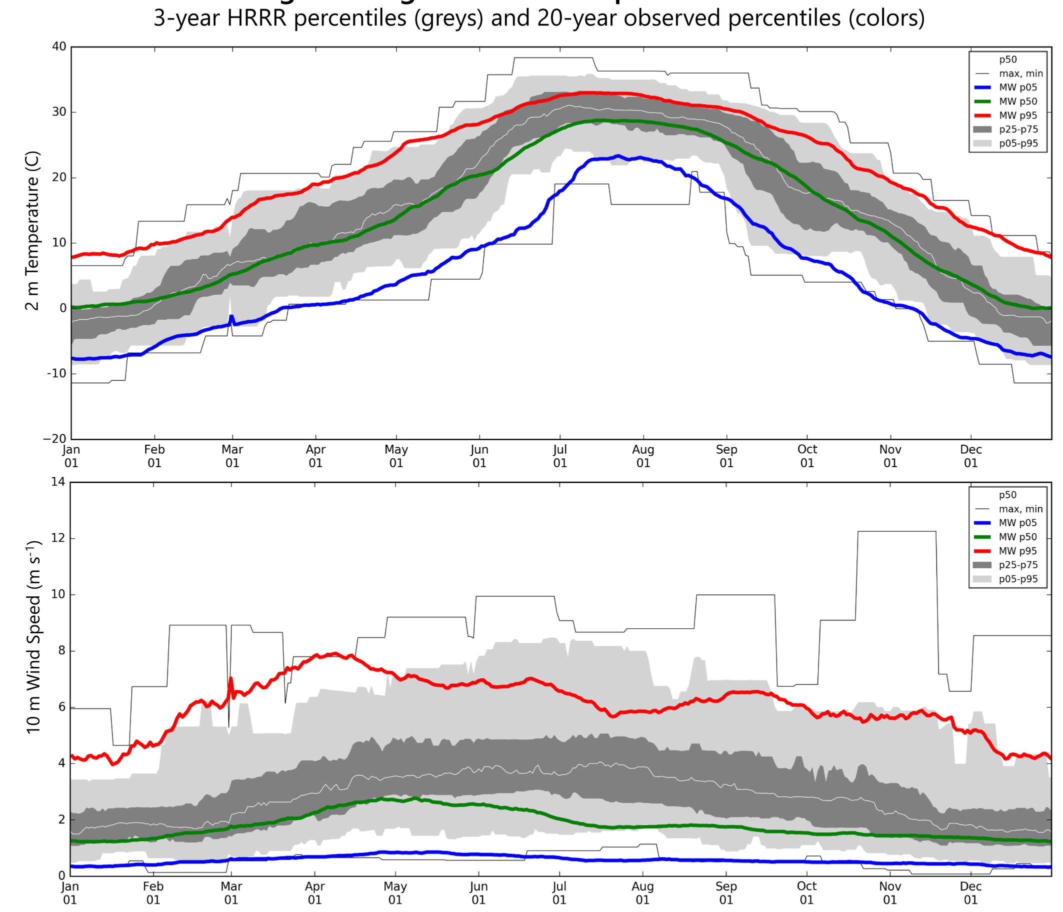
	Computation Time*	Data Transfer Time*	Size of Data Created*
OSG	2-3 Hours	3-4 hours	500-800 GB
Local Node	7 days	n/a	

* per HRRR variable

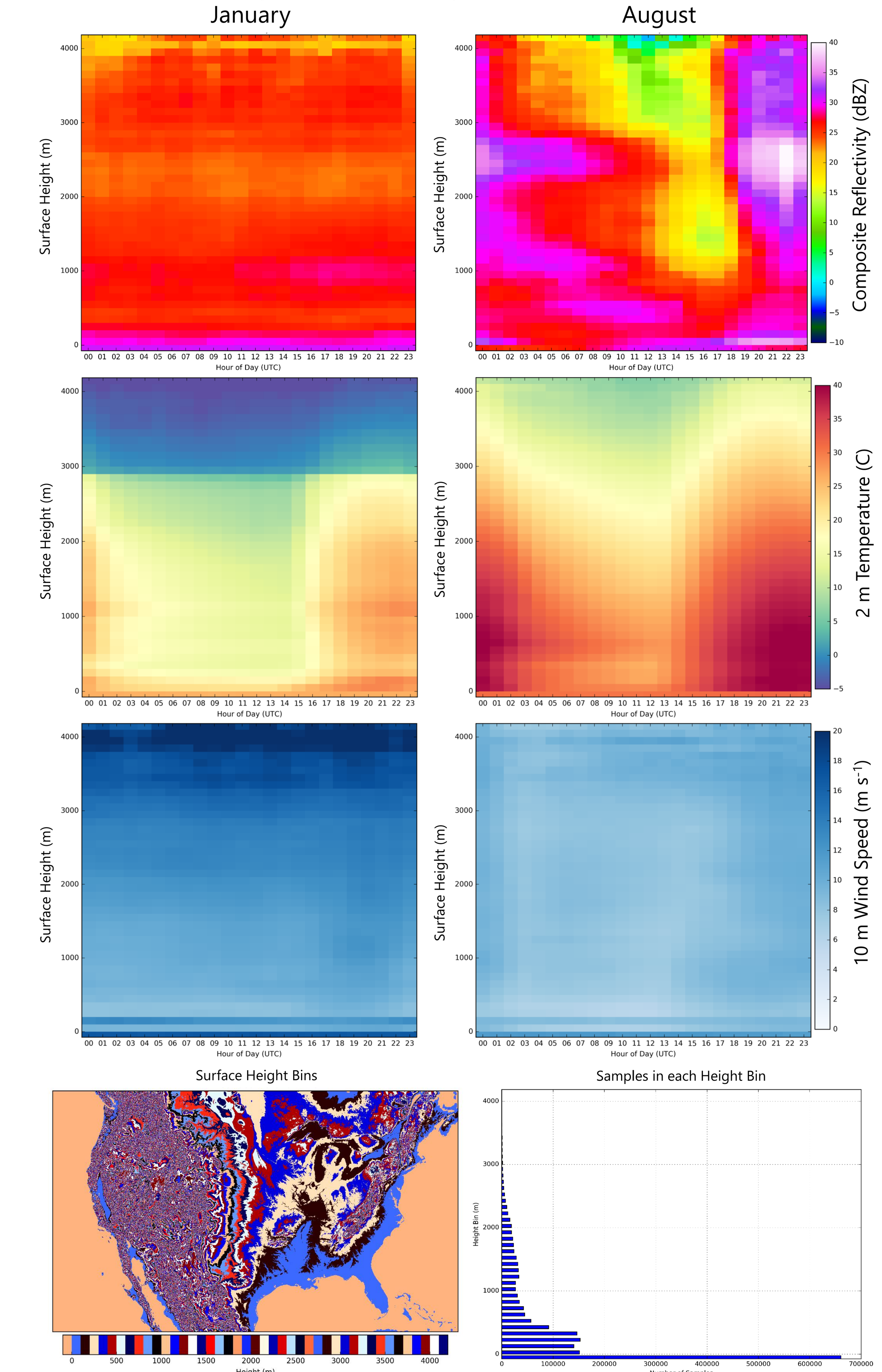
HRRR Percentile Calculations on the Open Science Grid



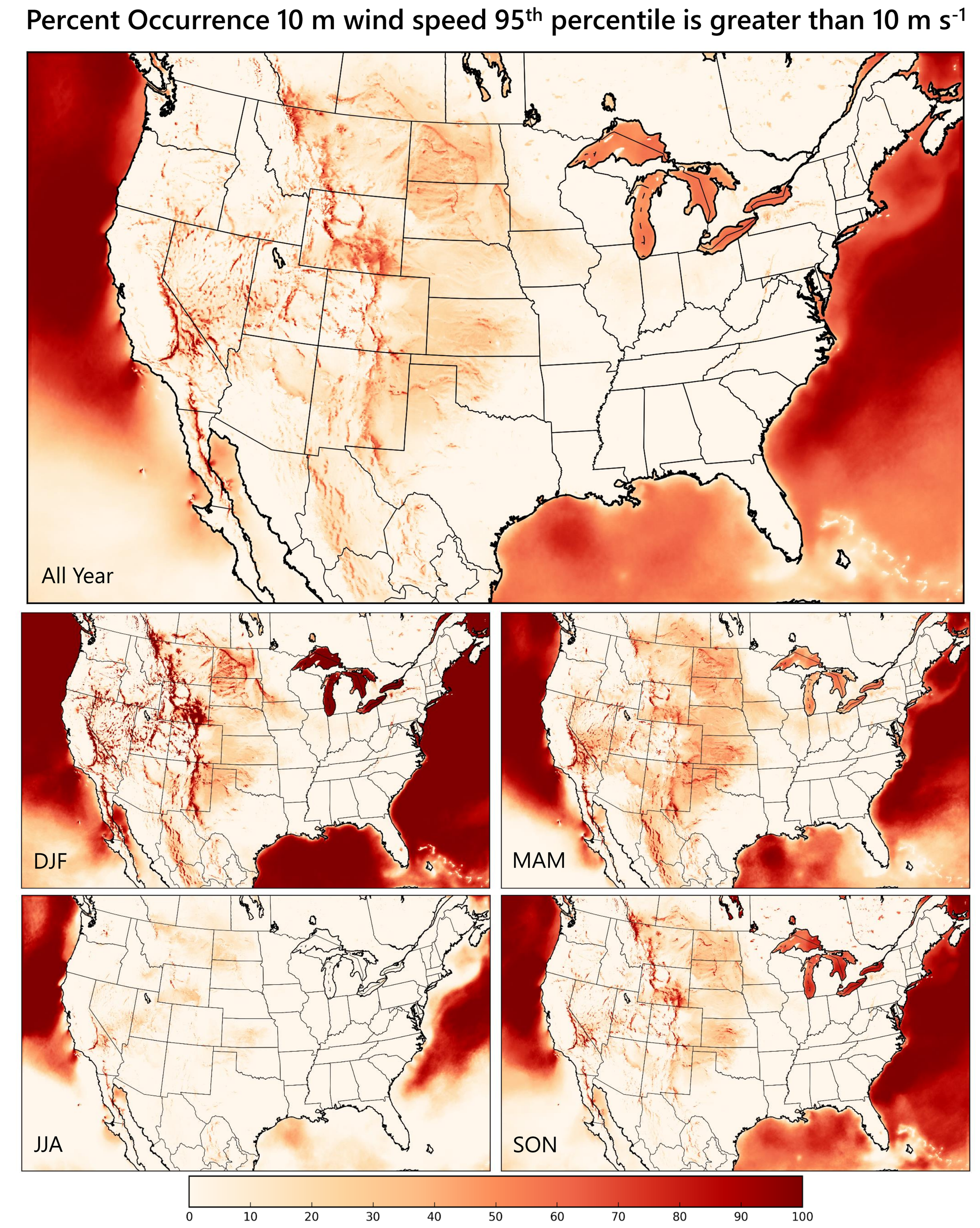
William Browning Building 1800 UTC Temperature and Wind Percentiles
3-year HRRR percentiles (greys) and 20-year observed percentiles (colors)



Hourly 95th Percentile by Elevation (95th of 95th)



Percent Occurrence 10 m wind speed 95th percentile is greater than 10 m s⁻¹



Top Left:

3-year HRRR (greys) and 20-year observed (colors) temperature and wind percentiles for William Browning Building at the University of Utah at 1800 UTC

Bottom Left:

Percent occurrence that 95th percentile 10 m wind speed is greater than 10 m s⁻¹

Top Right:

95th percentile in a 100 m height bin of the 95th percentile for each hour of January (left) and August (right) for simulated reflectivity (1st row), 2 m temperature (2nd row), and 10 m wind speed (3rd row)

The bottom two figures show the spatial grid points within each height bin (left) and the number of samples in each bin (right)

What is a 95th of a 95th? This statistic is calculated along two axes. The first axis is time, where the 95th percentile is calculated for each grid point from the 3 years of data (calculated on OSG). The second axis is the height axis, where the 95th is calculated from all grid points within a specific height bin (calculated in postprocessing).

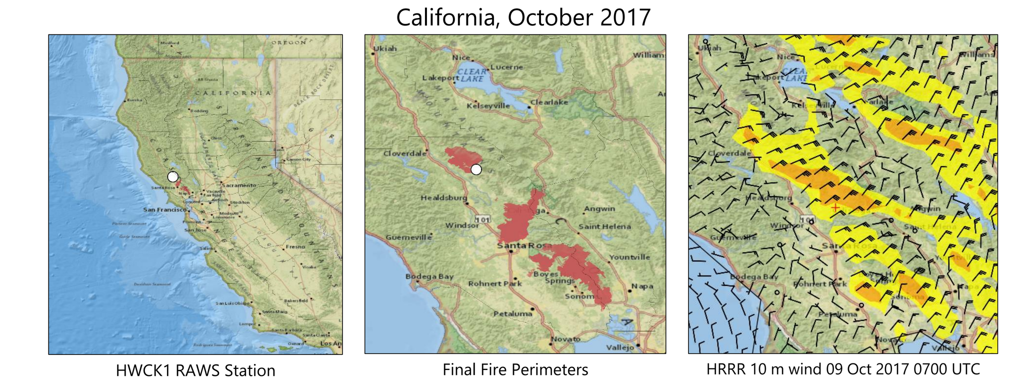
Applications

Statistics of HRRR model analyses may be useful to the renewable energy and agriculture sector and benefit forecasters unfamiliar with an area

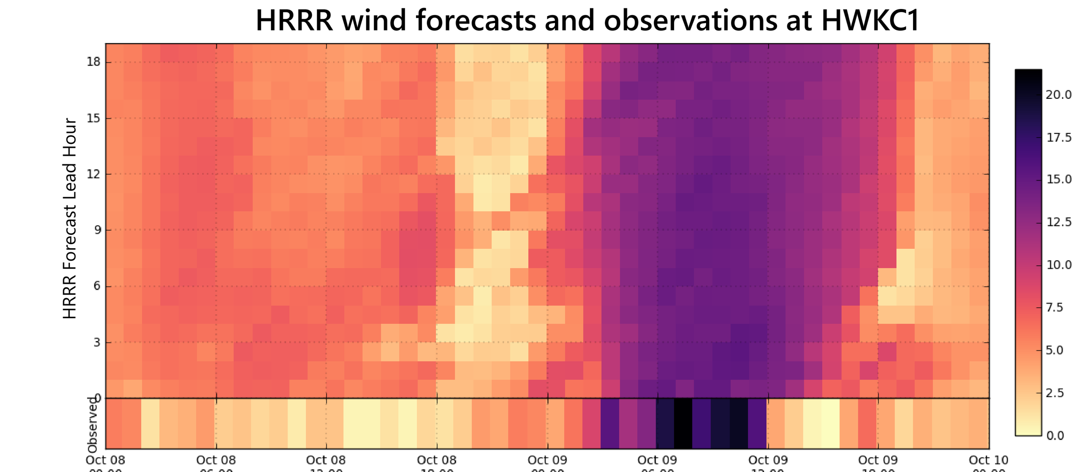
The continued focus of this work is to assist incident meteorologists assigned to major wildfires

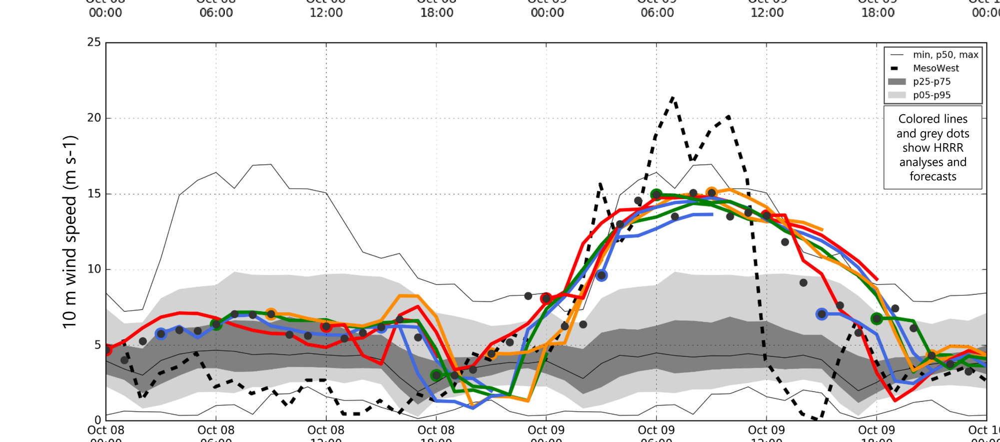
Forecasters assigned to an incident may use these statistics to become familiar with numerical model performance for the area or identify exceptional weather events

Pocket, Tubbs, and Nuns Fires
California, October 2017



HRRR wind forecasts and observations at HWKC1





Above illustrates several ways to view wind forecasts from hourly HRRR output. This high wind event caused the Pocket, Tubbs, and Nunn fires in California to spread quickly. Strong winds were forecasted by the HRRR model in all runs prior to the event, though not as strong as observed.

The bottom figure shows HRRR wind forecasts (colored and grey dots) in relation to observed wind speed at HWKC1 (black dash) and the percentile statistics (thin grey lines and shading).

HRRR statistics are also being used to flag potentially erroneous station data archived by MesoWest

Future work may ultimately perform these same statistics for each forecast hour to help identify potential forecast biases in the HRRR model

Contact: Brian Blaylock
brian.blaylock@utah.edu

https://github.com/blaylockbk/pyBKB_v2/tree/master/OpenScienceGrid

Benjamin, S. G., and Coauthors, 2016: A North American Hourly Assimilation and Model Forecast Cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669-1694, doi: 10.1175/MWR-D-15-0242.1.

Blaylock, B., J. Horel, and S. Liston, 2017: Cloud Archiving and Data Mining of High Resolution Rapid Refresh Forecast Model Output. *Computers and Geosciences.* **109**, 43-50, doi: 10.1016/j.cageo.2017.08.005.

This research was done using resources provided by the Open Science Grid, which is supported by the National Science Foundation award 1148698, and the U.S. Department of Energy's Office of Science.

The support and resources from the Center for High Performance Computing at the University of Utah and NOAA ESRL are gratefully acknowledged. This research has been supported by the Joint Fire Science Program L17AC00225.