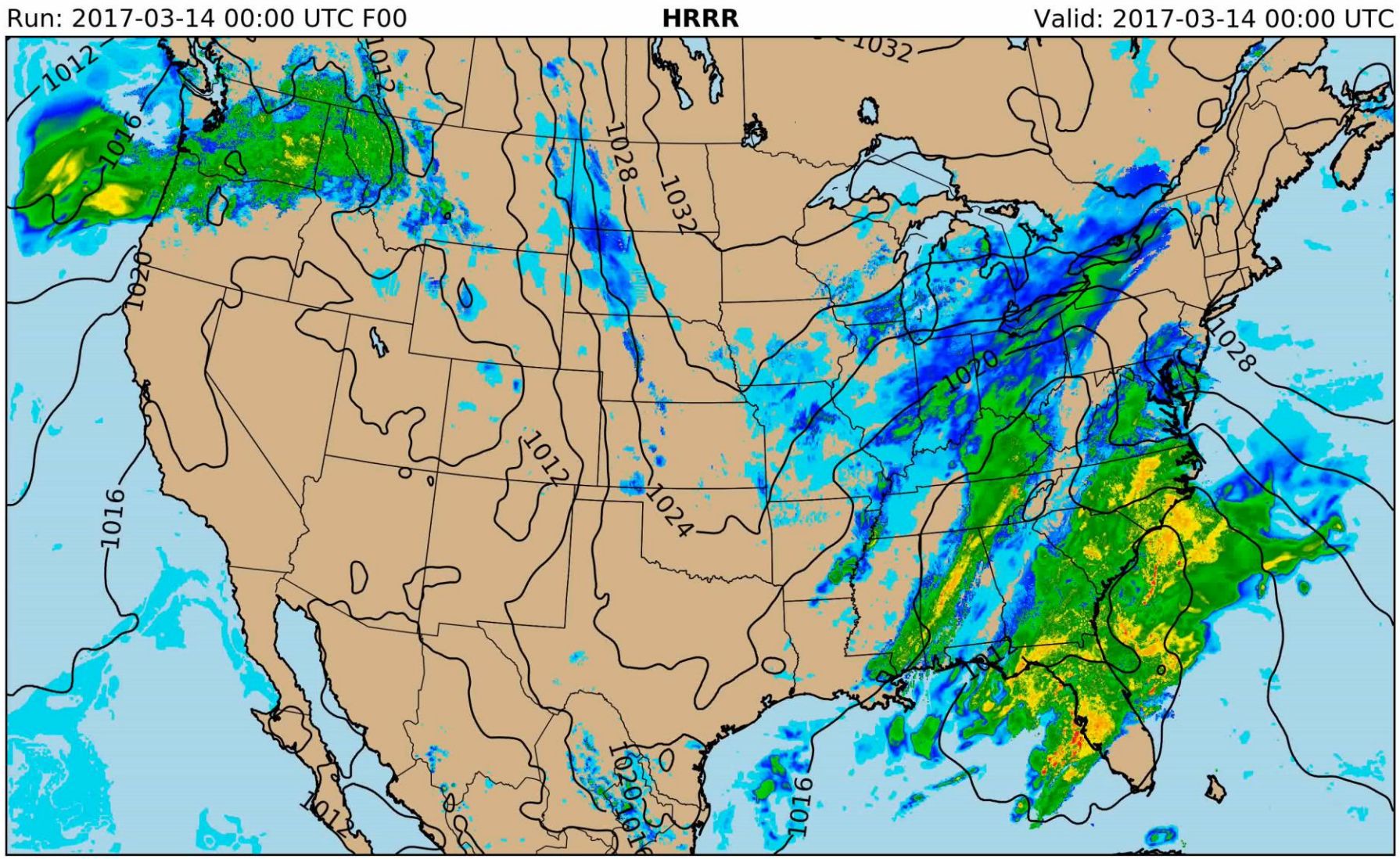
Multi-year Analytics of NOAA's High Resolution Rapid Refresh Model

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

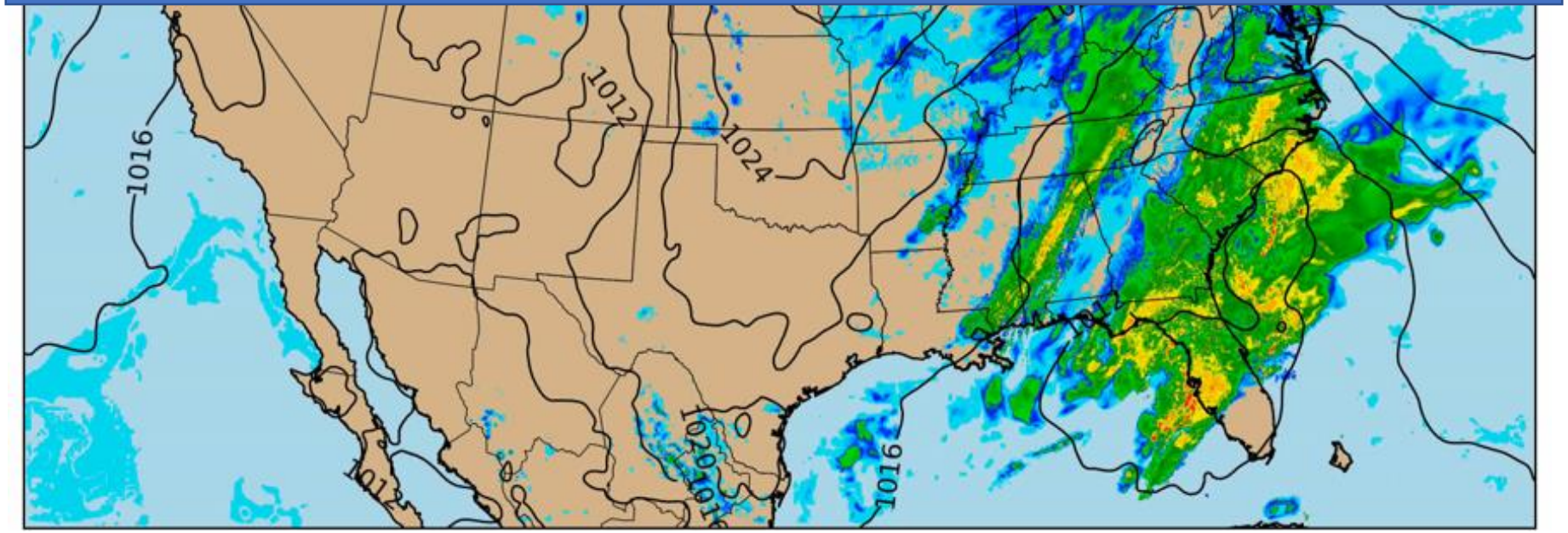
March 21, 2018
Salt Lake City, Utah
Open Science Grid All-hands Meeting

What is the **H**igh **R**esolution **R**apid **R**efresh Model?



What is the **H**igh **R**esolution **R**apid **R**efresh Model?

- 3 km grid spacing (1.9 million points)
- Updated every hour
- Produces 18 hour forecasts
- Advanced data assimilation



What is the **H**igh **R**esolution **R**apid **R**efresh Model?

- 3 km grid spacing (1.9 million points)
- Updated every hour
- Produces 18 hour forecasts
- Advanced data assimilation

The highest resolution weather model run operationally by NOAA's National Centers for Environmental Prediction

What is the **H**igh **R**esolution **R**apid **R**efresh Model?

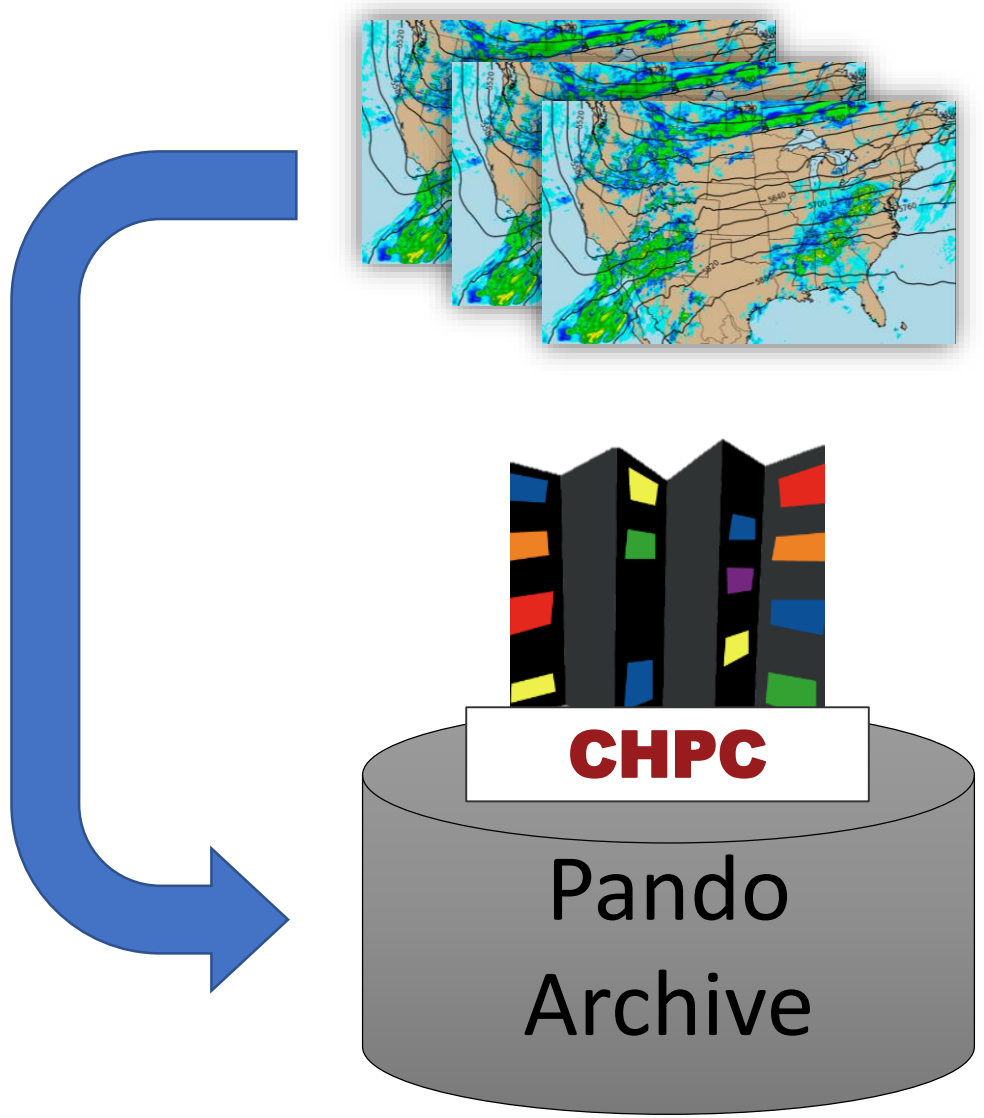
Applications:

- Aviation
- Fighting Wildfires
- Water management
- Solar and wind energy
- Agriculture
- Severe weather

We archive raw HRRR output

- HRRR data is in GRIB2 format (Gridded Binary Version 2)
  - Highly compressed data
  - Data Size: ~20 TB/year
  - Data Source: *NOAA Operational Model Archive and Distribution System*

- Pando Archive at CHPC
  - Object storage like Amazon S3
  - Access: http://hrrr.chpc.utah.edu



OSG Acknowledgments

Wim Cardon

- Introduced me to OSG at CHPC workshop

Bala Desinghu

- Got me started and rapidly replied to my questions

Mats Rynge

- Prepared a Singularity image for me with `pygrib` and other dependencies

Benedikt Riedel

- Wrote my first DAGMan file for me

Sam Liston

- Helped me with Globus transfers at CHPC

Science Question

Structured our science question to be answered with parallel computing

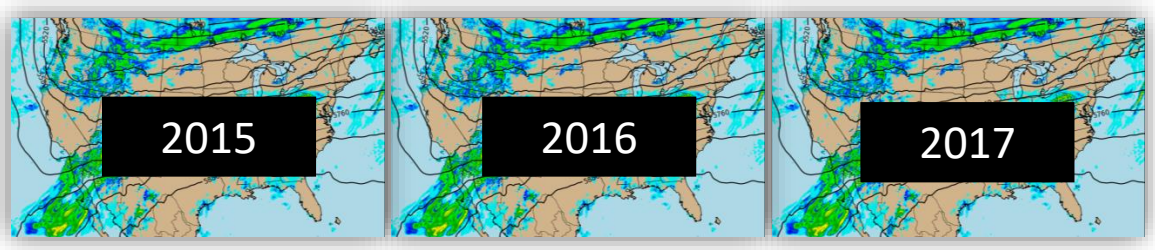
What is the range of weather conditions
at every grid point
for every hour of the year
in the last three years?

Model "climatology"

Hourly Percentiles from 3 years of Data

- For each hour in a year:
  - Retrieve model analysis grid from 2015, 2016, 2017
  - Compute statistics for each grid point

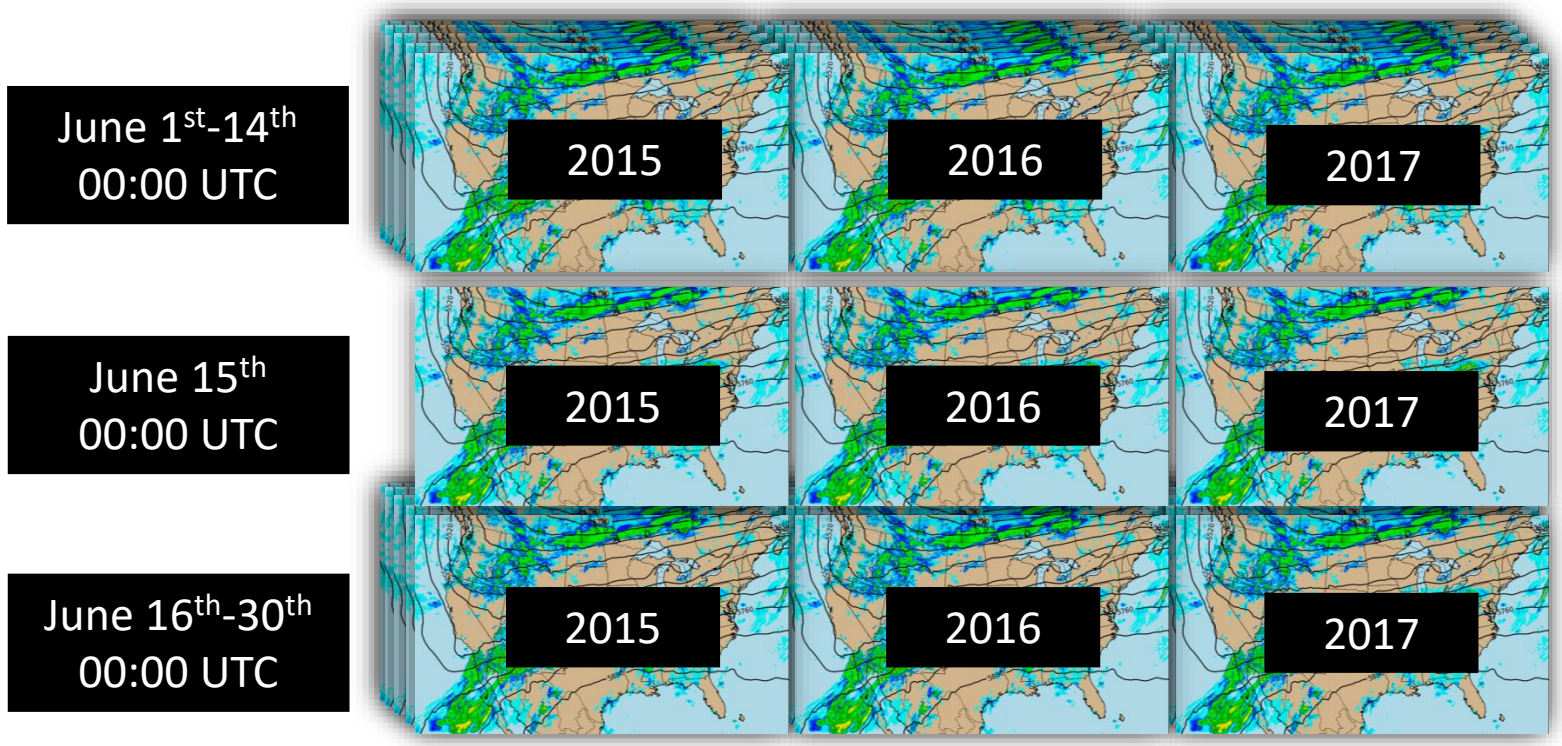
June 15th
00:00 UTC



Samples:
3

Hourly Percentiles from 3 years of Data

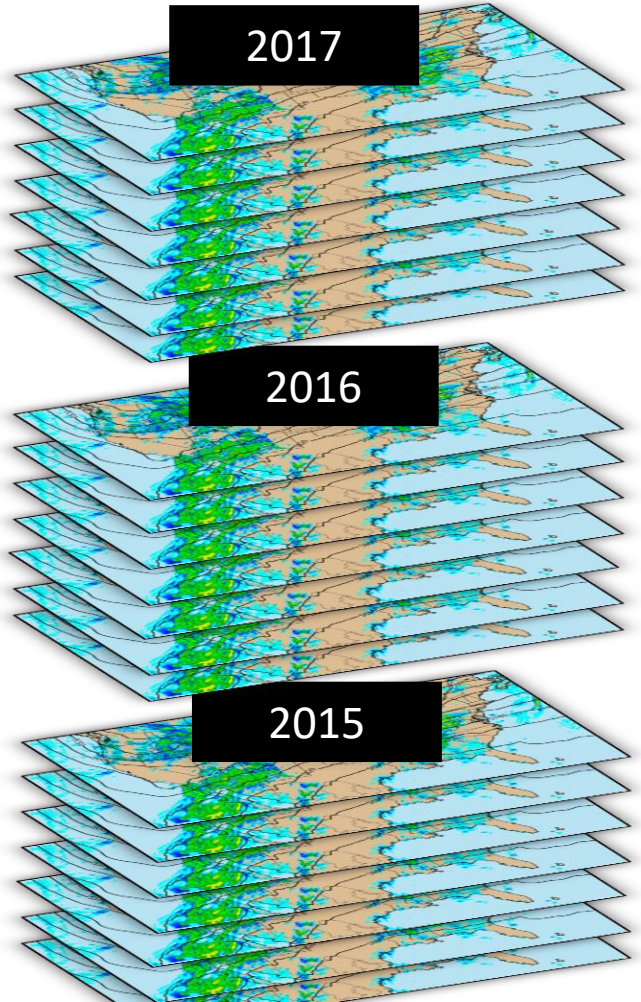
- For each hour in a year:
  - Retrieve model analysis grid from 2015, 2016, 2017
  - Compute statistics for each grid point
  - **Increase sample size by including +/- 15 days**



Samples:
90

Hourly Percentiles from 3 years of Data

Download 90 Grids



Calculate Statistics

numpy.
percentile()

Percentiles Calculated

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

mean

Output

HDF5

1 unique OSG job for every hour of the year

8,784 unique jobs

366 days x 24 hours

For a single variable, this work takes 2-3 hours on OSG

This same work takes ~7 days on our local node

Sacrifice data download efficiency for high-throughput computing

Each HRRR file is downloaded 30 times in 30 different jobs, but downloads are quick!

1 unique OSG job for every hour of the year

8,784 hours × **8 Variables** = **70,272 jobs**

1. 2 m Temperature
2. 2 m Dew Point
3. 10 m Wind Speed
4. Max 10 m Wind Speed
5. 80 m Wind Speed
6. Surface Gusts
7. Simulated Composite Reflectivity
8. 500 mb Height

Tools and Workflow

Singularity Container – *Needed* `pygrib` *module*

Python/2.7 – *Main program*

DAGMan – *Manage jobs*

Globus – *Transfer files to CHPC*

New Data Created

20 statistics calculated for each variable at each model grid point

Output in **HDF5** format → more bloated than **GRIB2**, but easier to work with

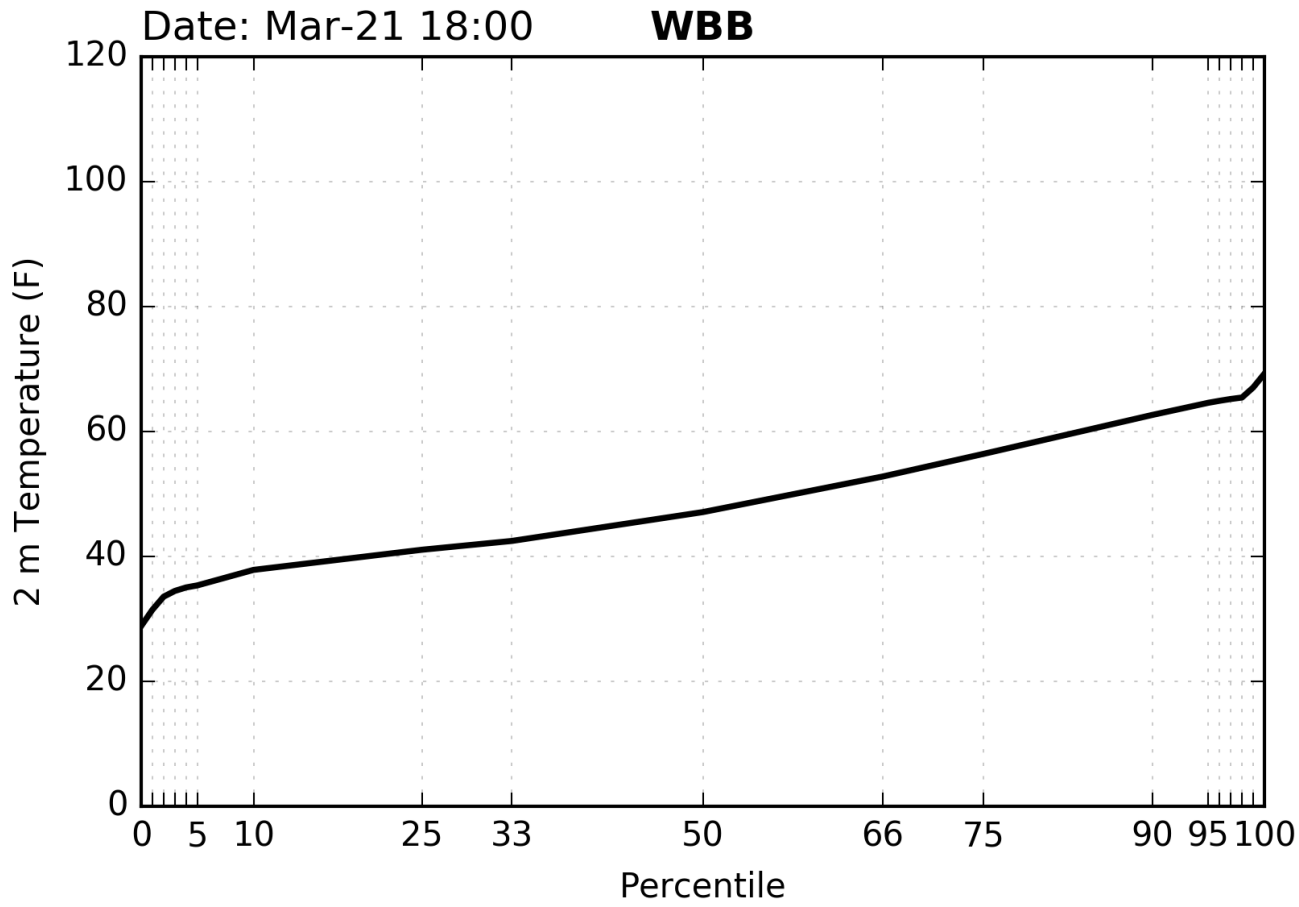
Science

- Wind and temperature climatology
  - Full year
  - By month/season
  - Single hour

- Percentiles at point or an area

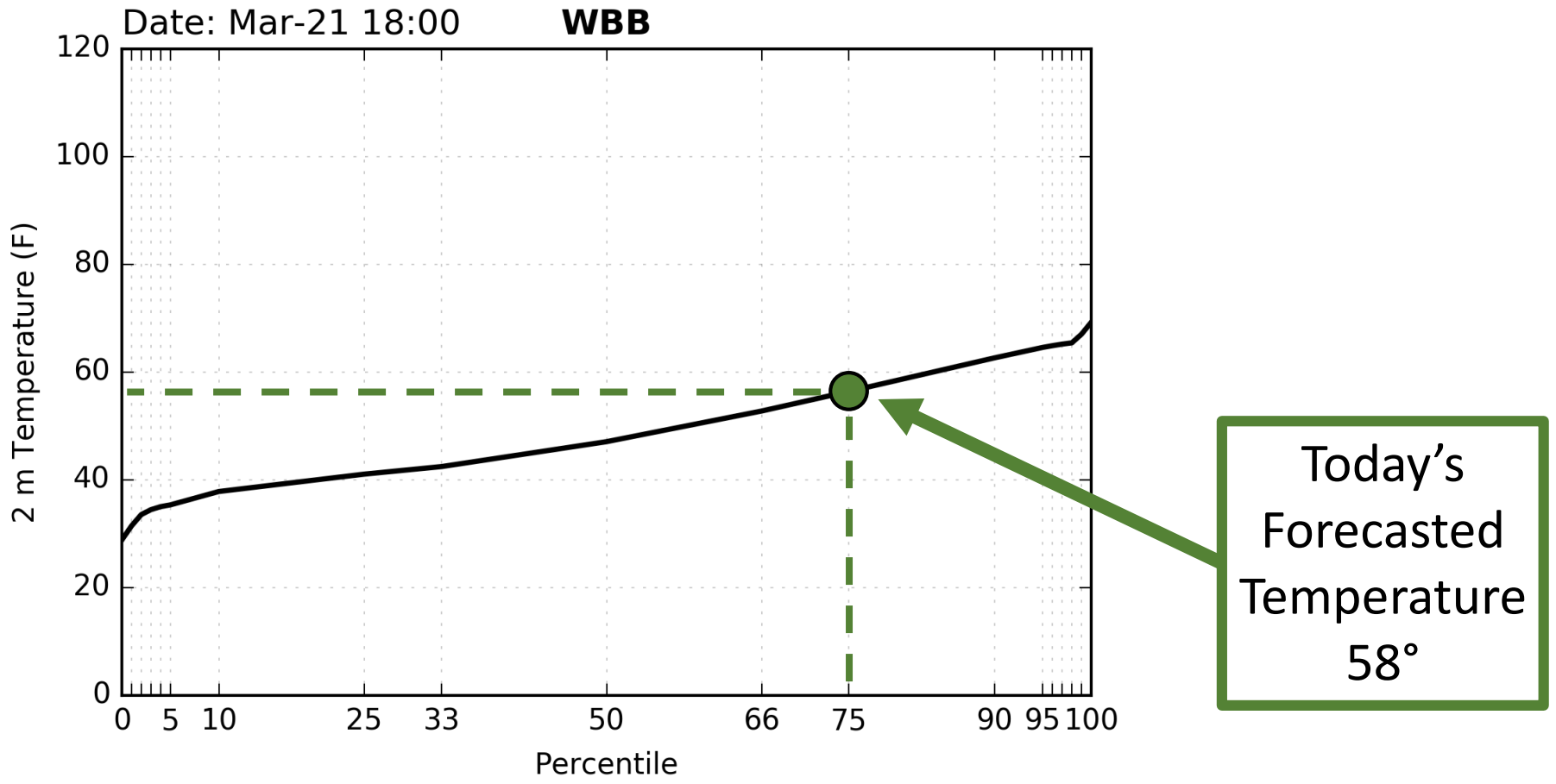
Temperature Percentiles at a Point

Single Hour



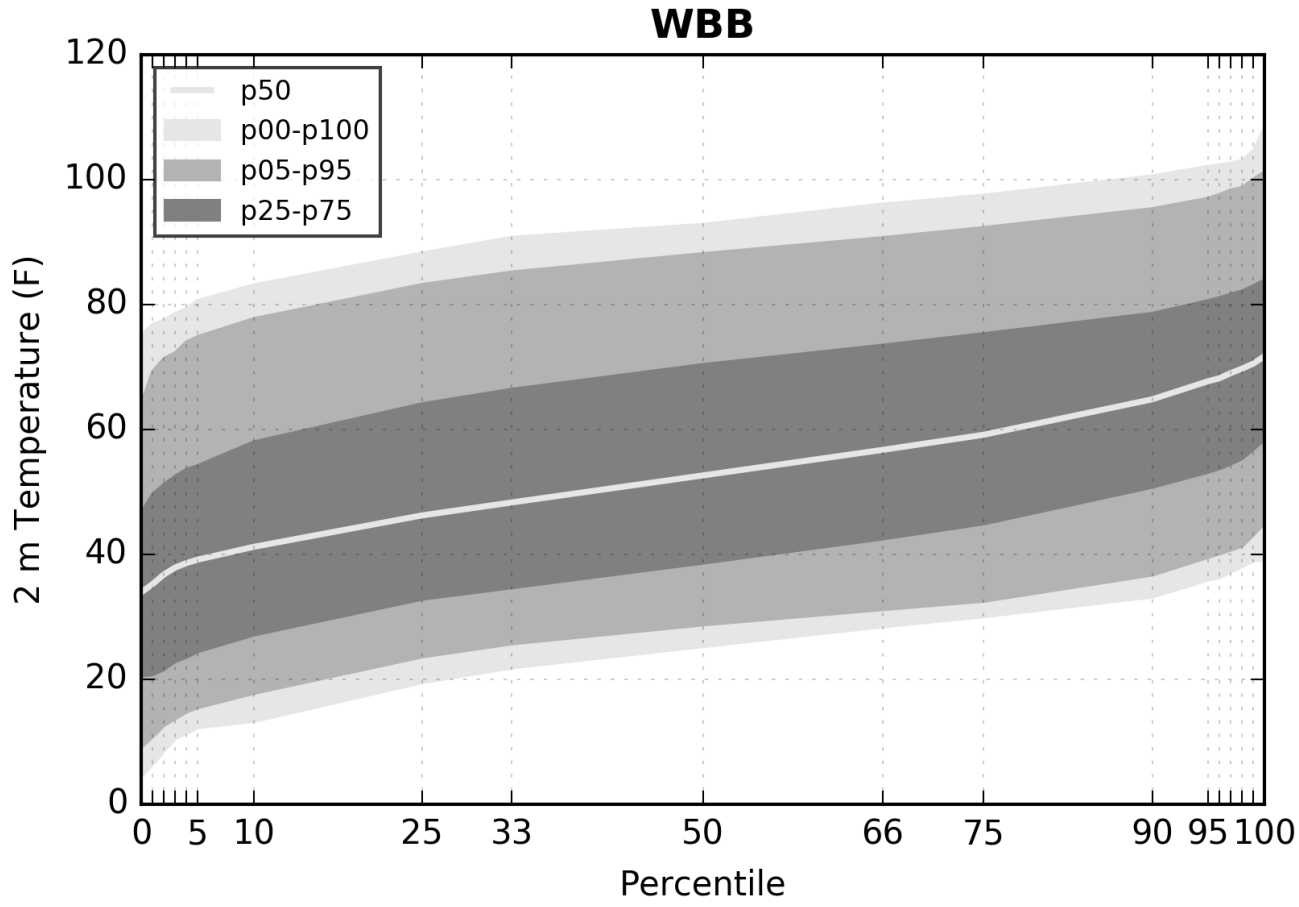
Temperature Percentiles at a Point

Single Hour



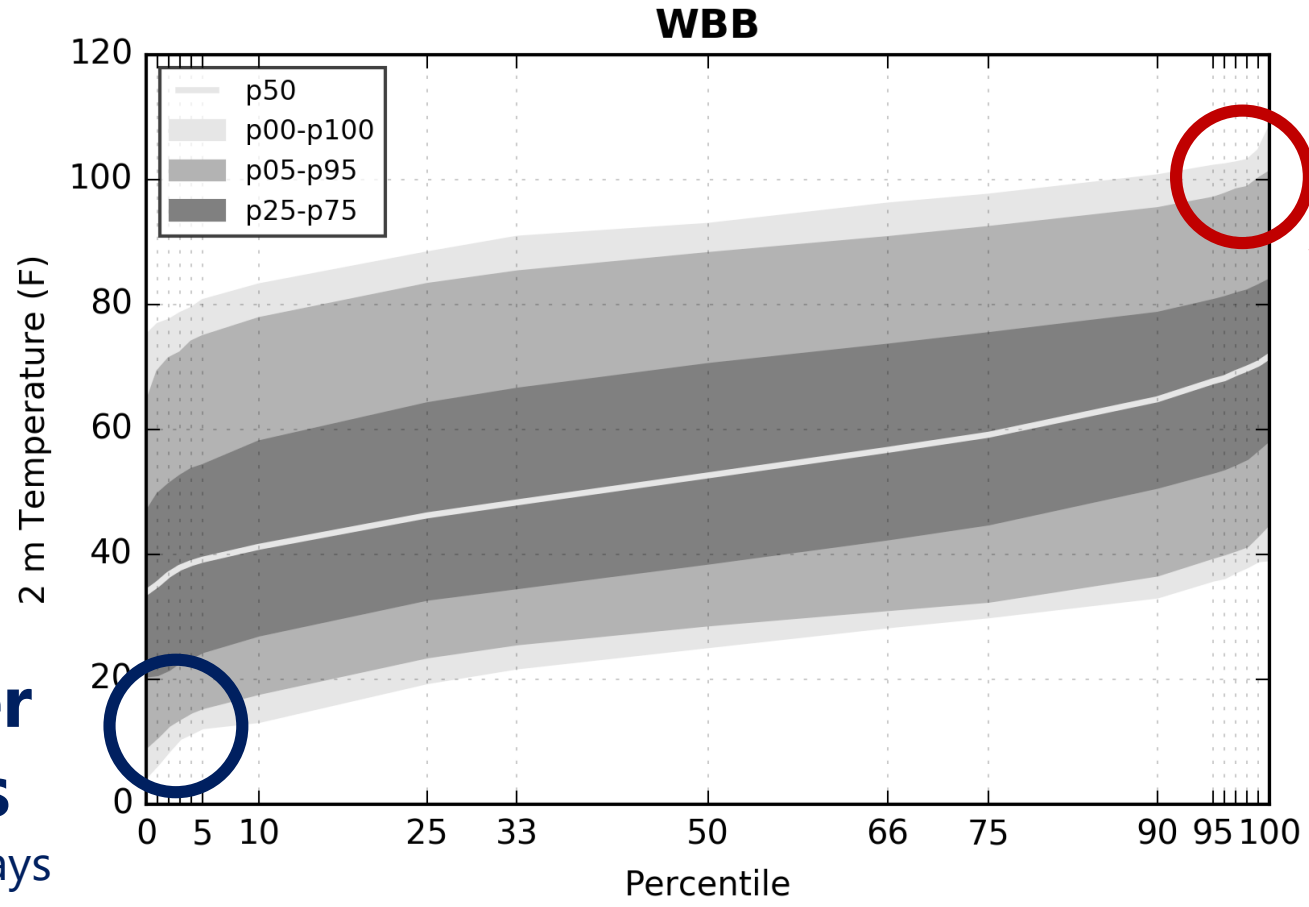
Temperature Percentiles at a Point

8,784 Hours



Temperature Percentiles at a Point

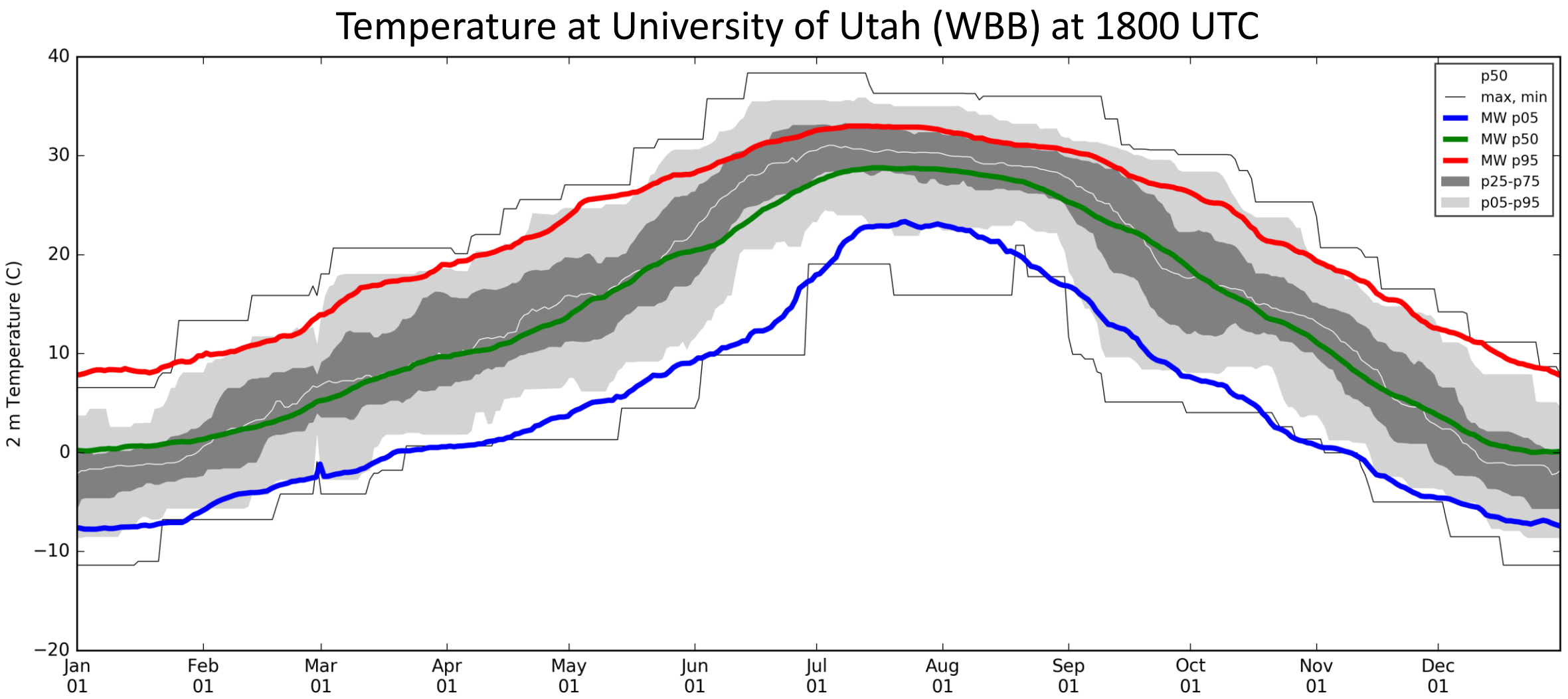
8,784 Hours



Hottest Summer Temperatures
warmest of the warmest days

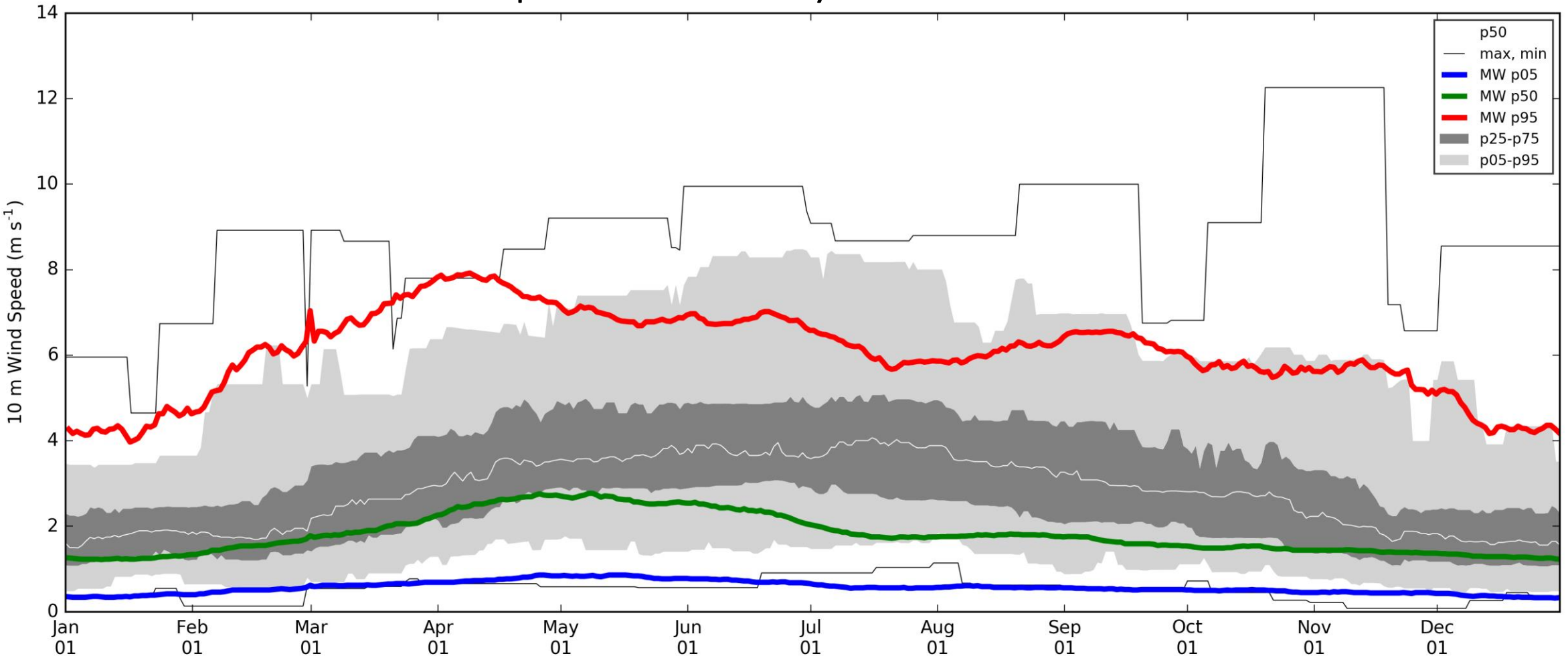
Coldest Winter Temperatures
coldest of the coldest days

Percentiles at a Point

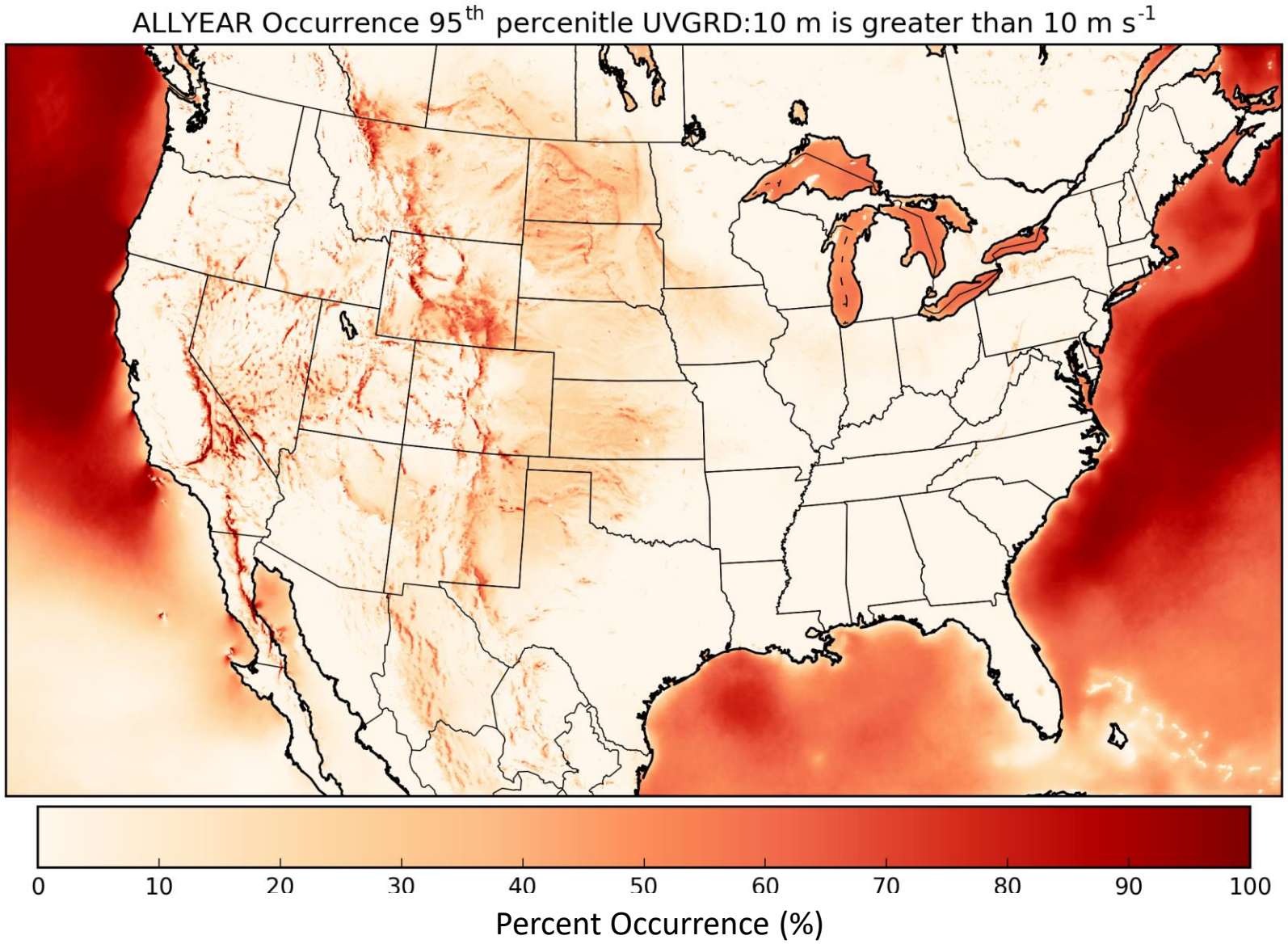


Percentiles at a Point

Wind Speed at University of Utah at 1800 UTC

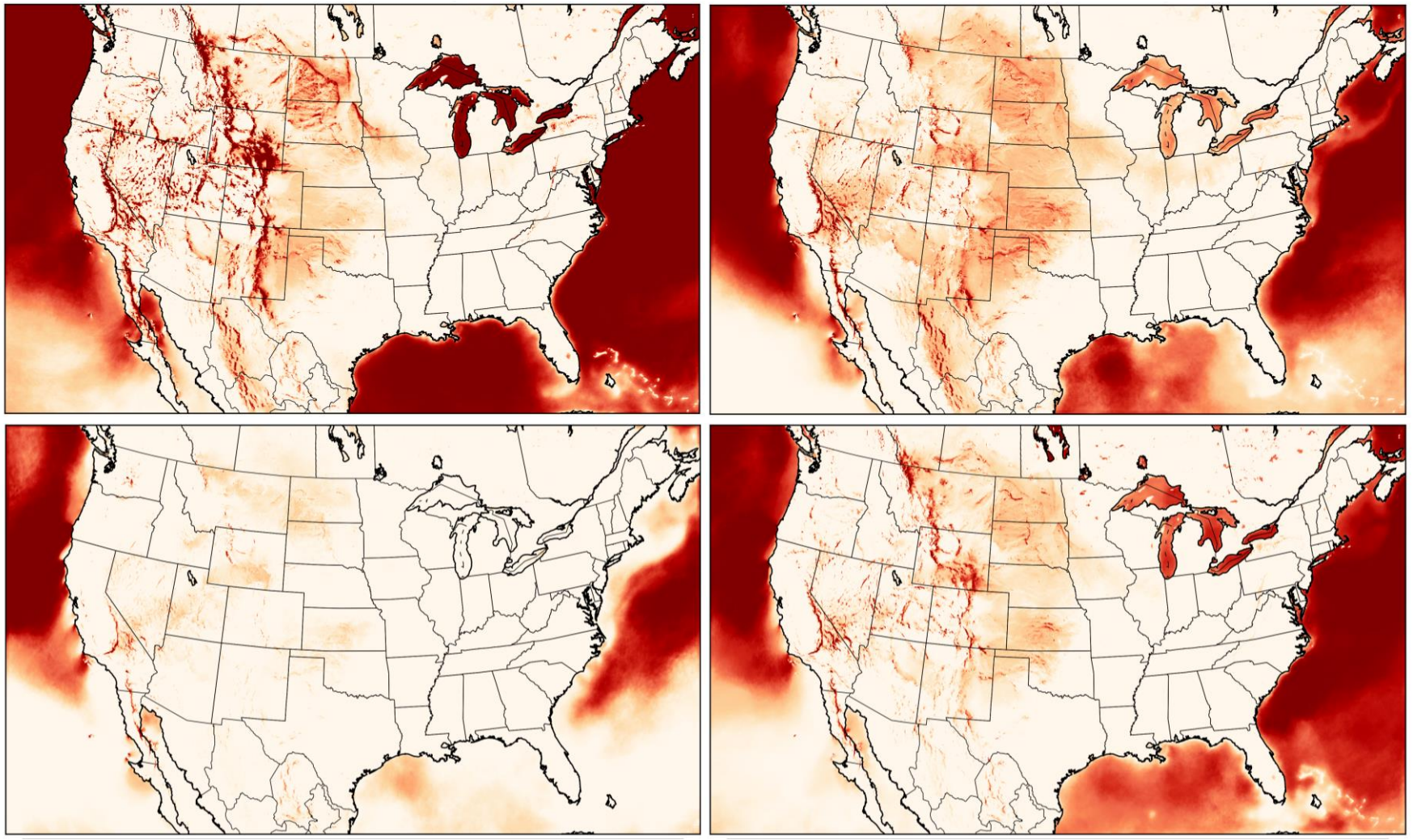


Wind Climatology



Wind Climatology

December
January
February

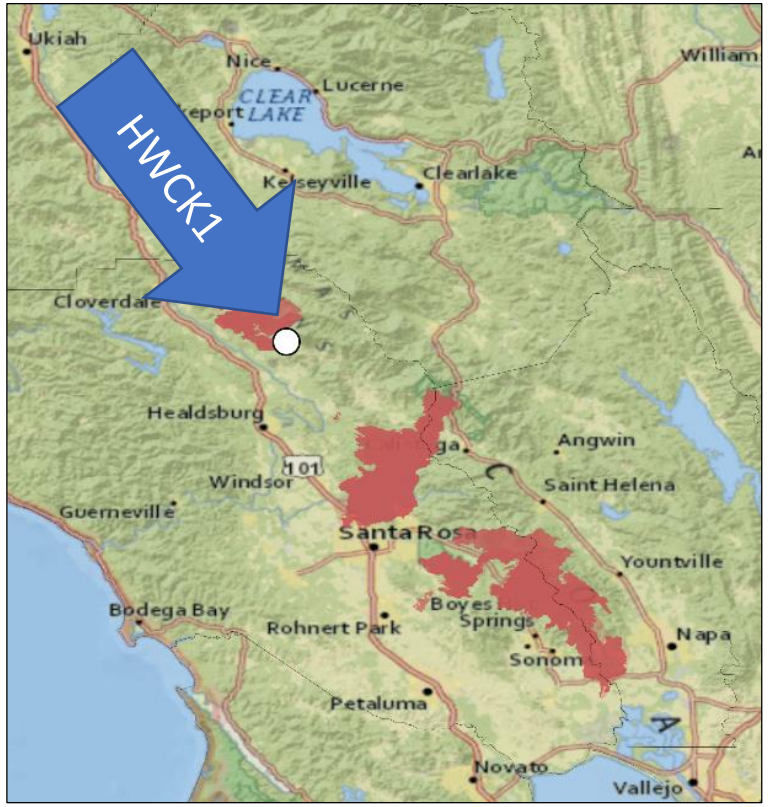


March
April
May

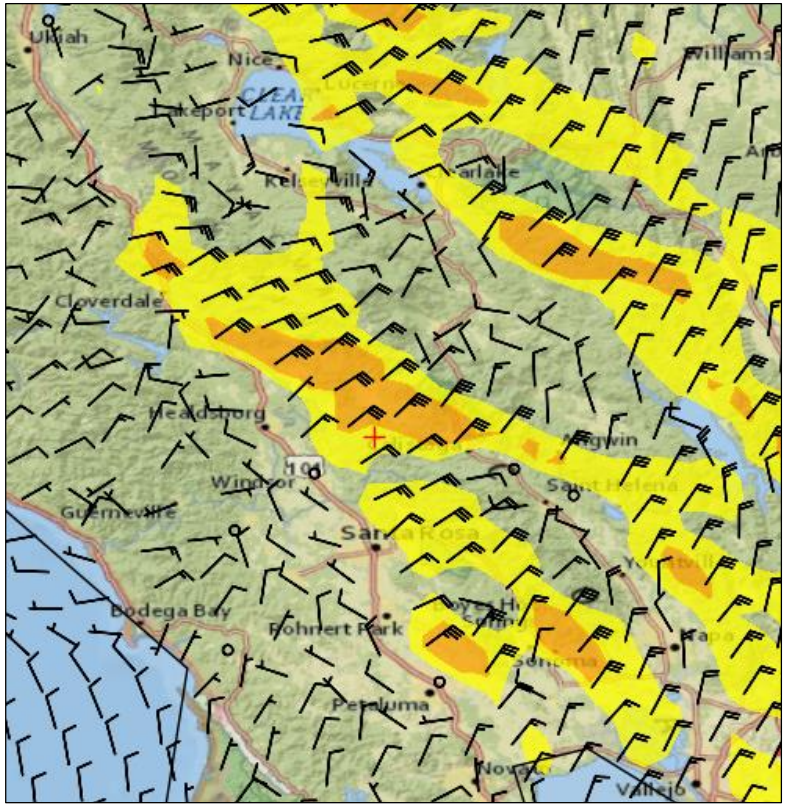
June
July
August

September
October
November

Pocket, Tubbs, and Nuns Fires
California, October 2017

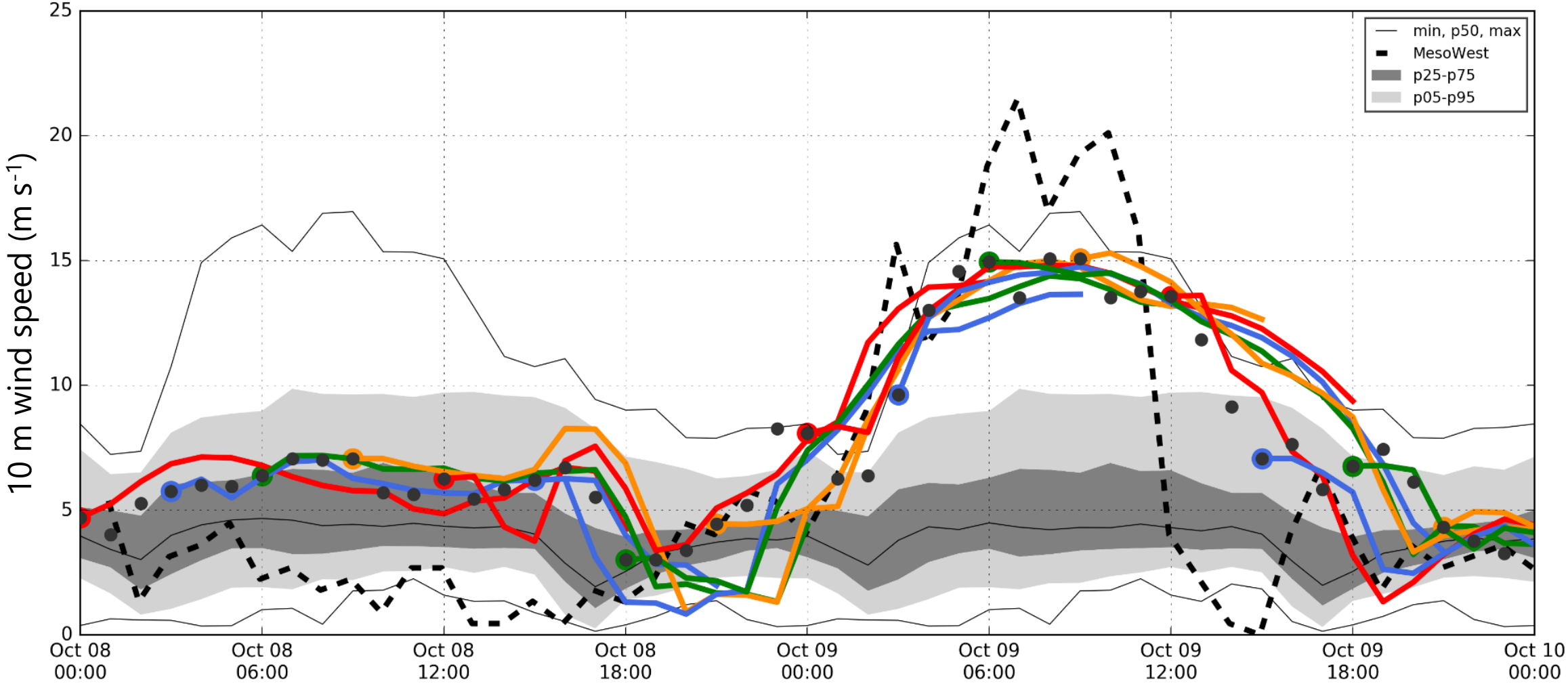


Fire Perimeters



HRRR 10 m wind
09 Oct 2017 0700 UTC

HRRR wind forecasts and observations at HWKC1



Future Work

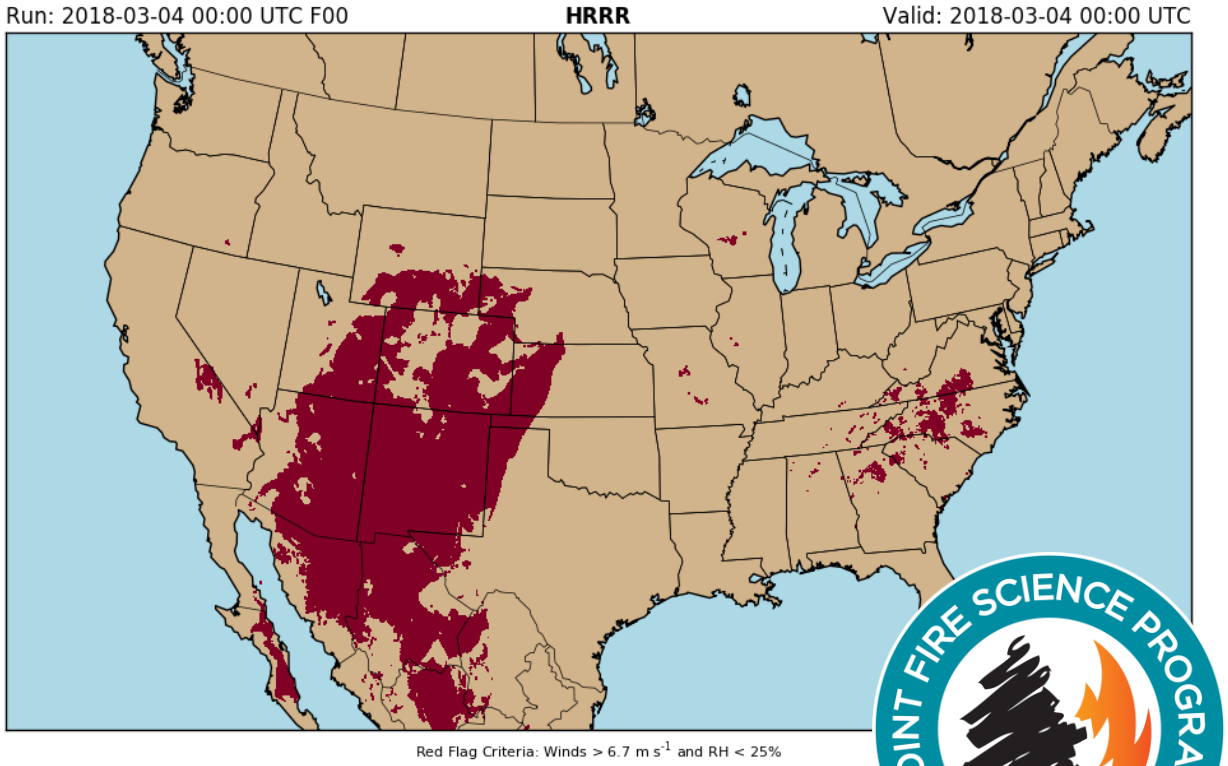
Multivariate Percentiles

Occurrence of **strong wind** and **low relative humidity**

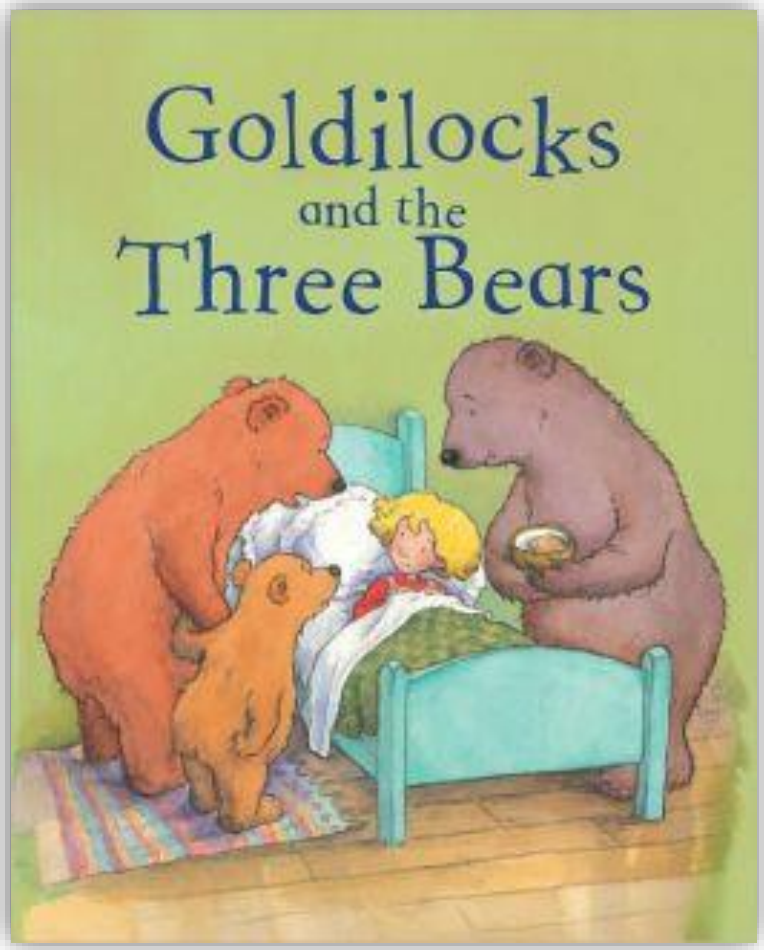
Red flag conditions for wildfire forecasting

Identify Model Bias

Compute percentiles for forecast hours and comparing with analyses and observations. Identify bias by variable, location, time of day and year, etc.



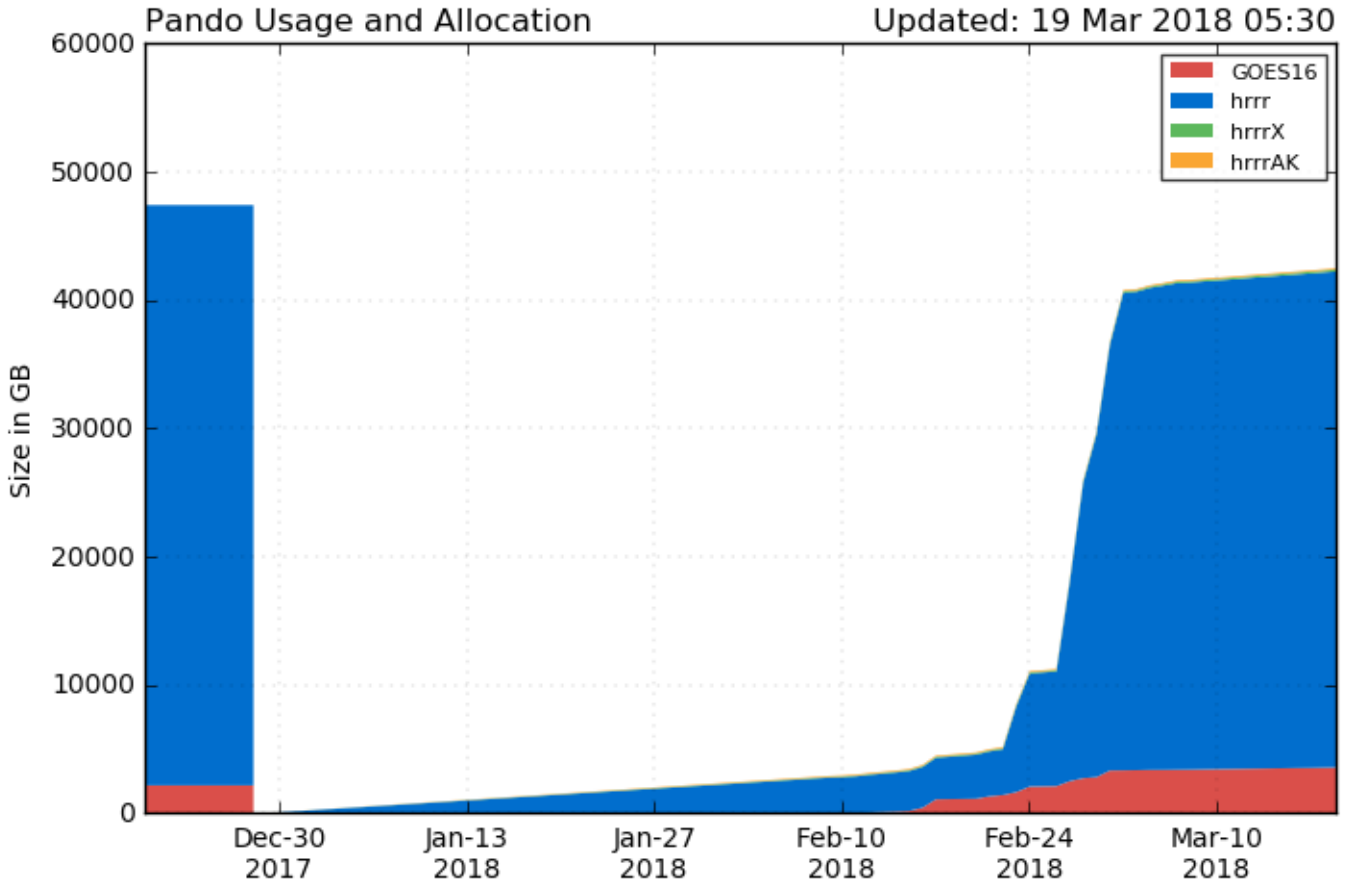
Why the OSG was good for this work?



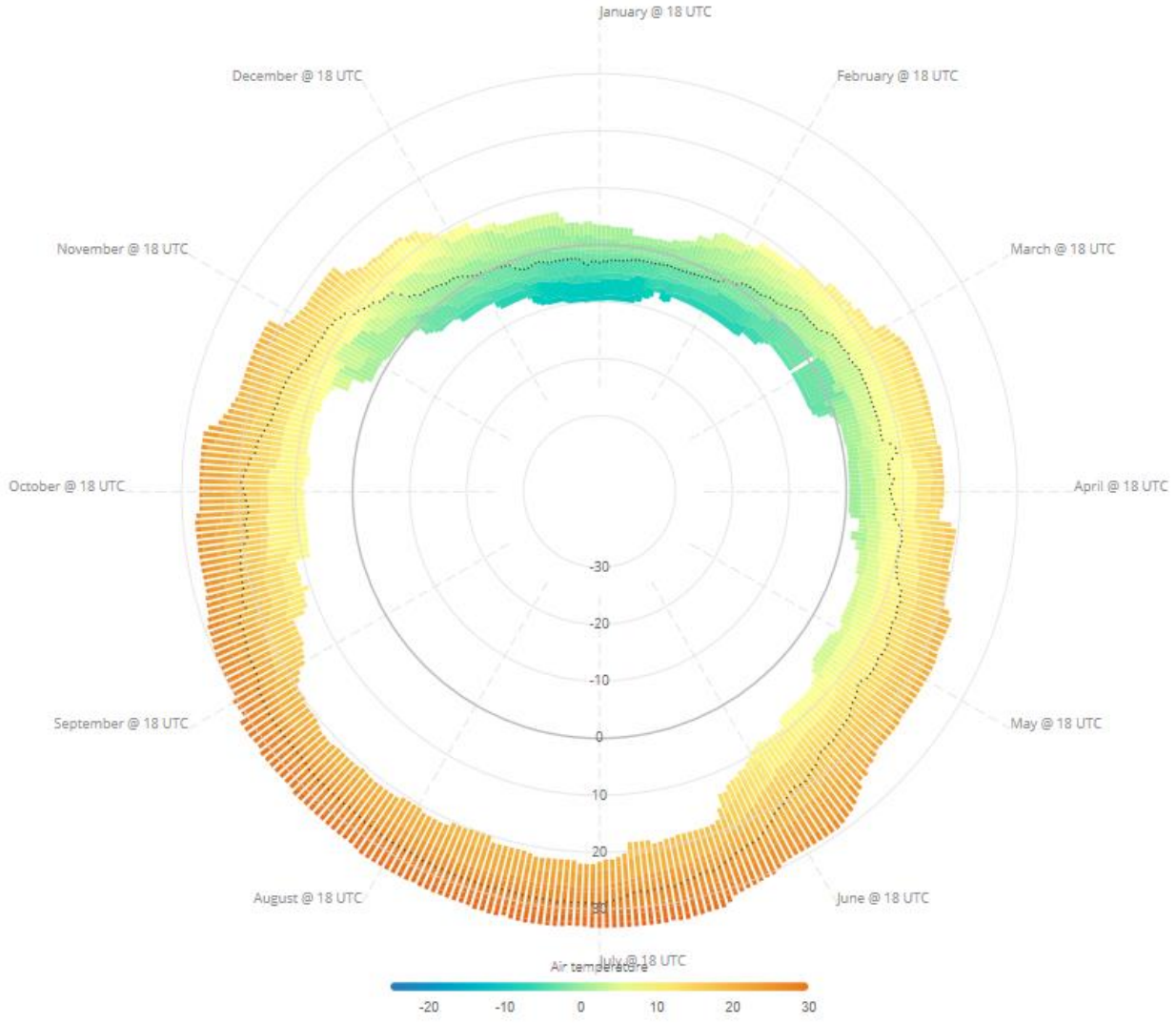
1. Small, dedicated node—*lengthy run time*
2. HPC with interconnected nodes—*not necessary*
3. **Open Science Grid high-throughput computing—** ***just right!***

Future Work
(Ugh…Technology)

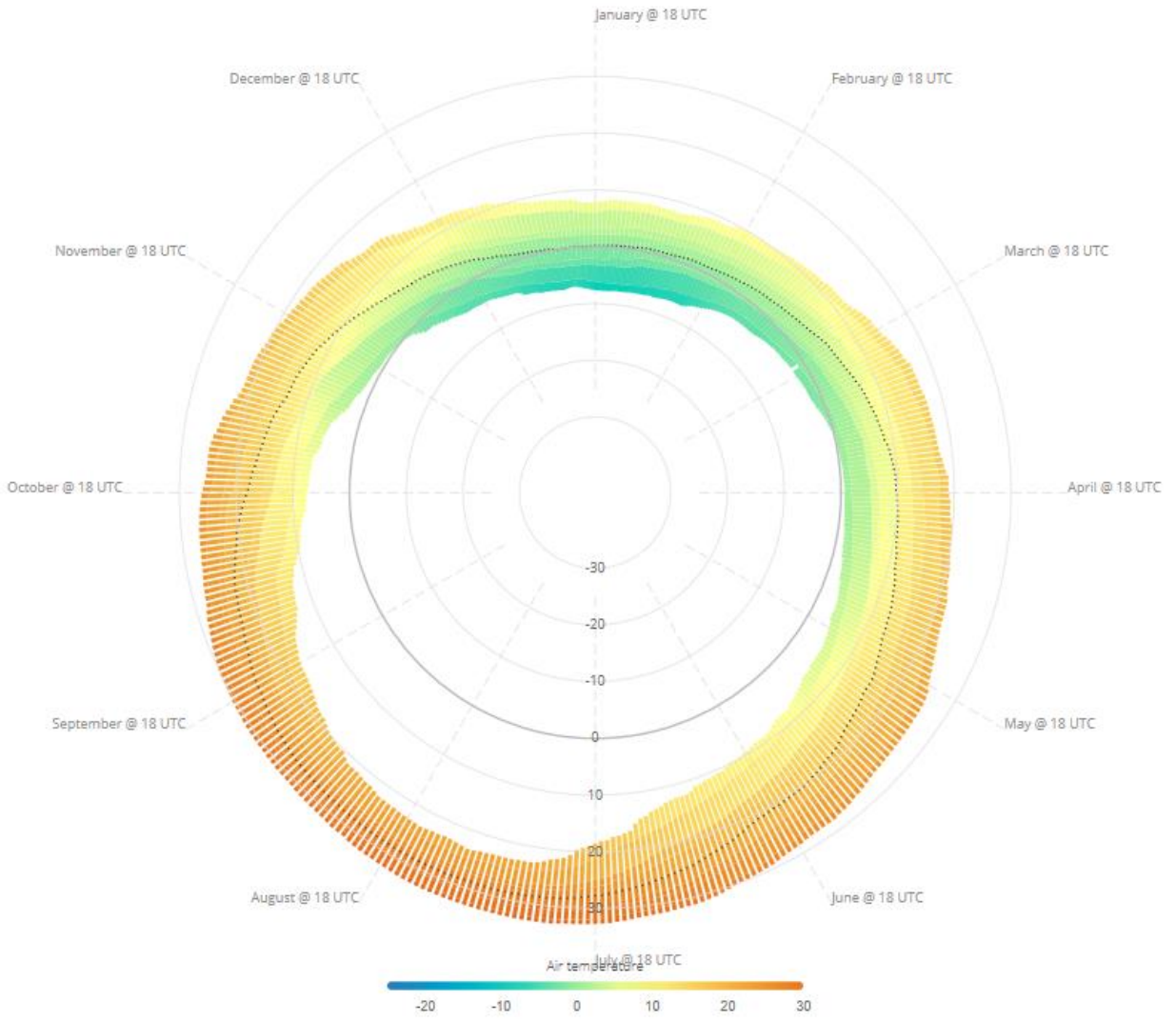
- The Pando archive system failed and we lost all the data.
- We were able to recover data after July 2016
- After July this year, we can recompute percentiles for 2 years of HRRR data.



Percentiles at a Point



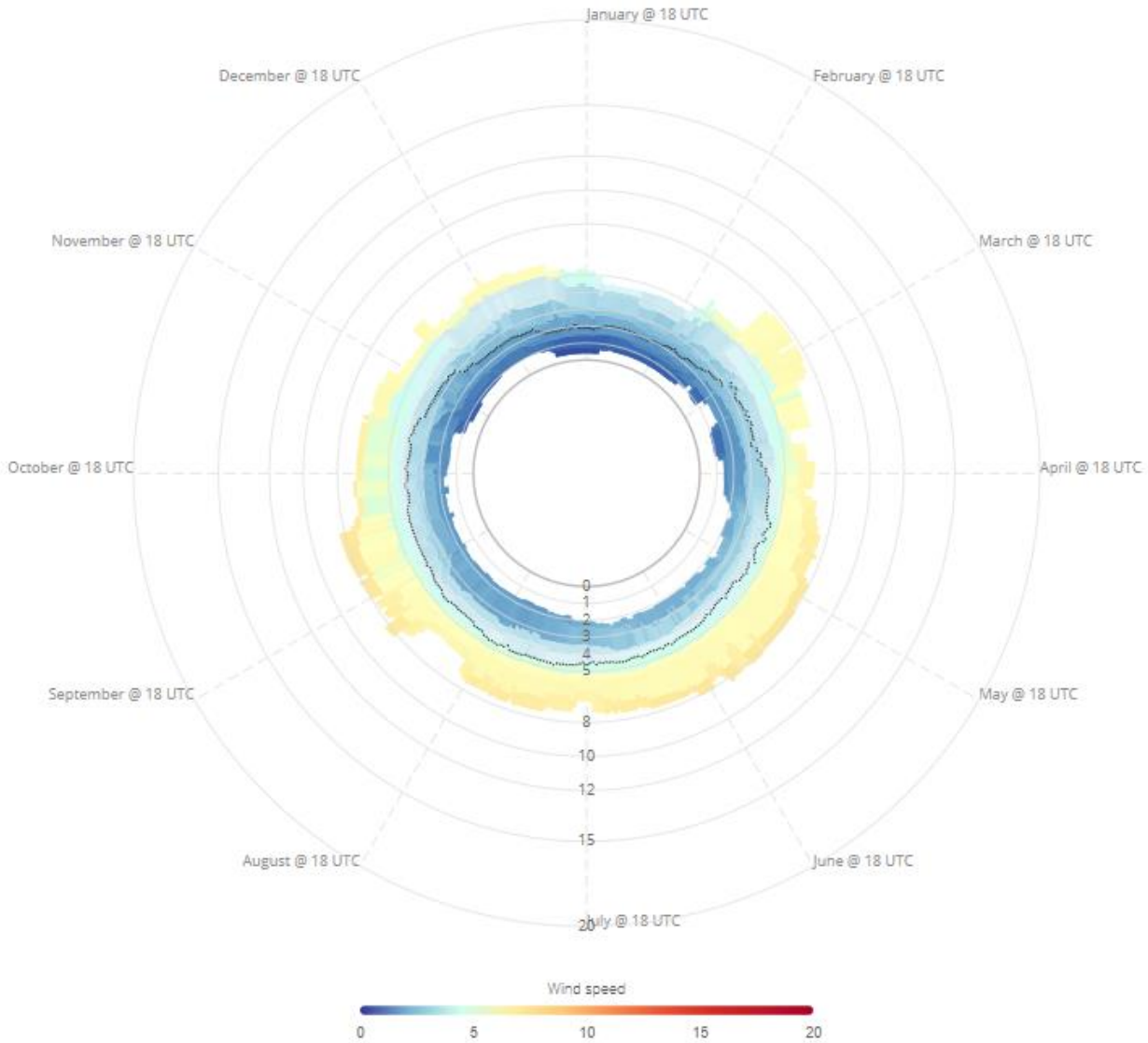
3-year HRRR 5th-95th Percentiles



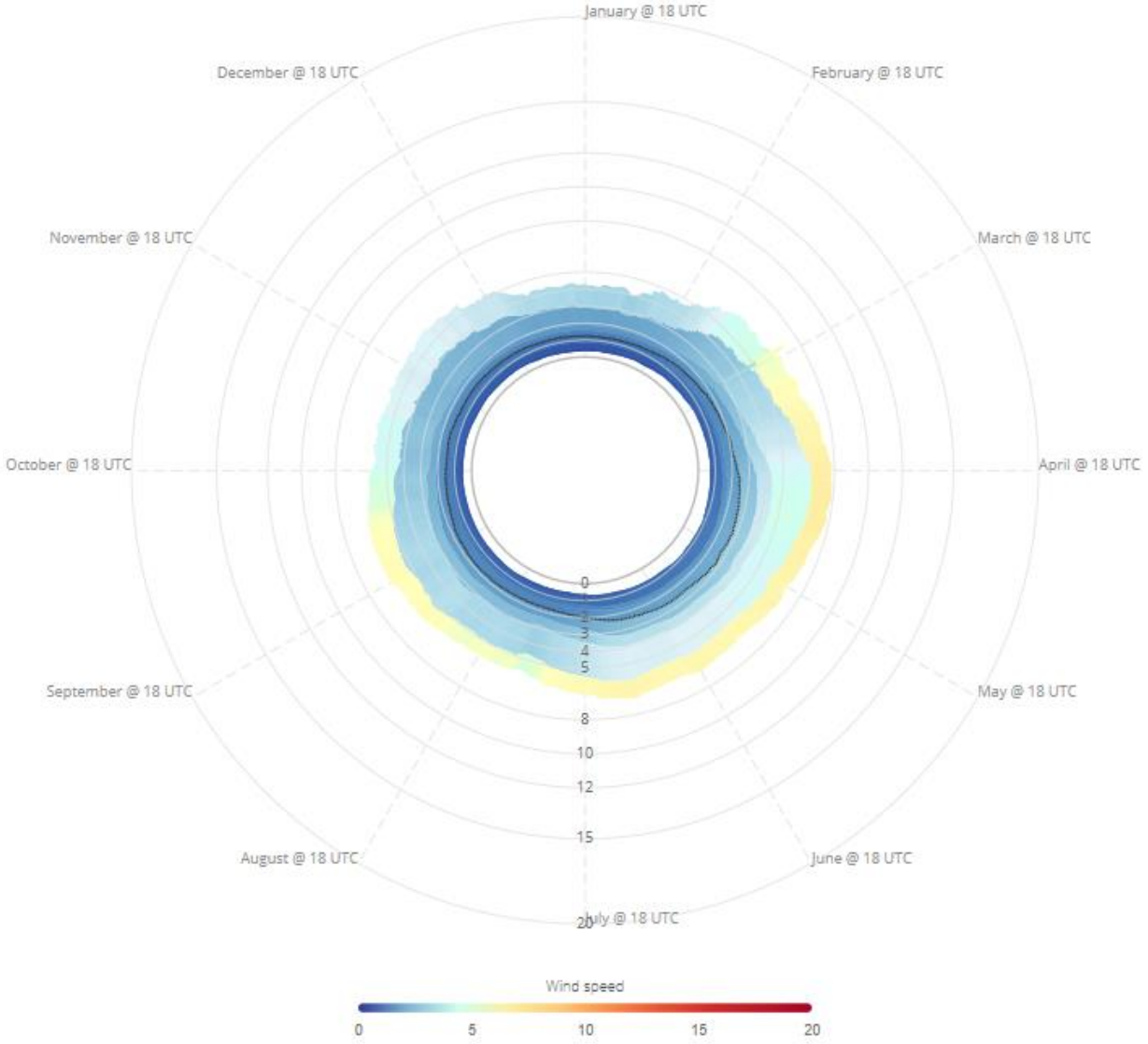
20-year Observed 5th-95th Percentiles

Credit: Chris Galli

Percentiles at a Point



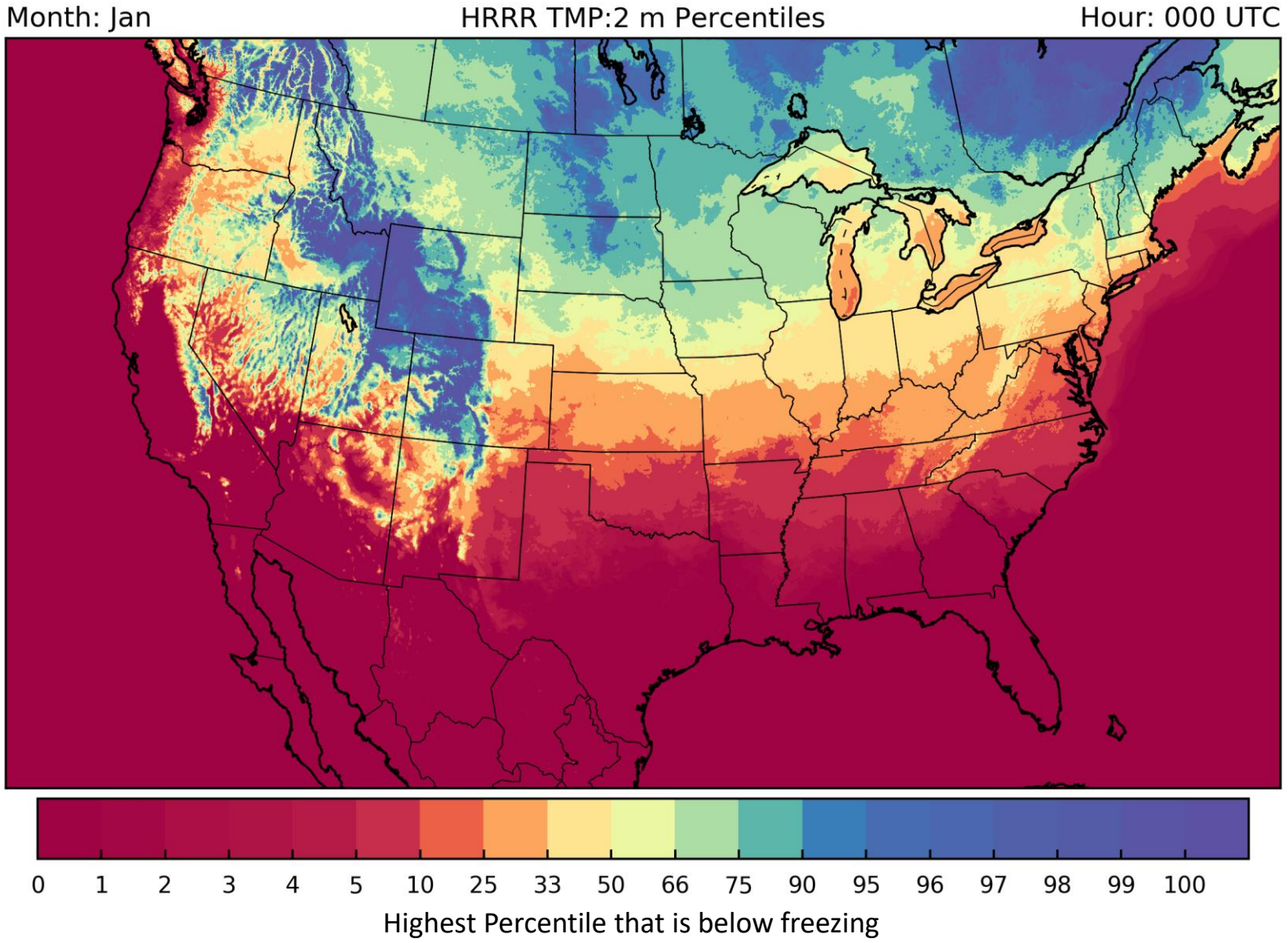
3-year HRRR 5th-95th Percentiles



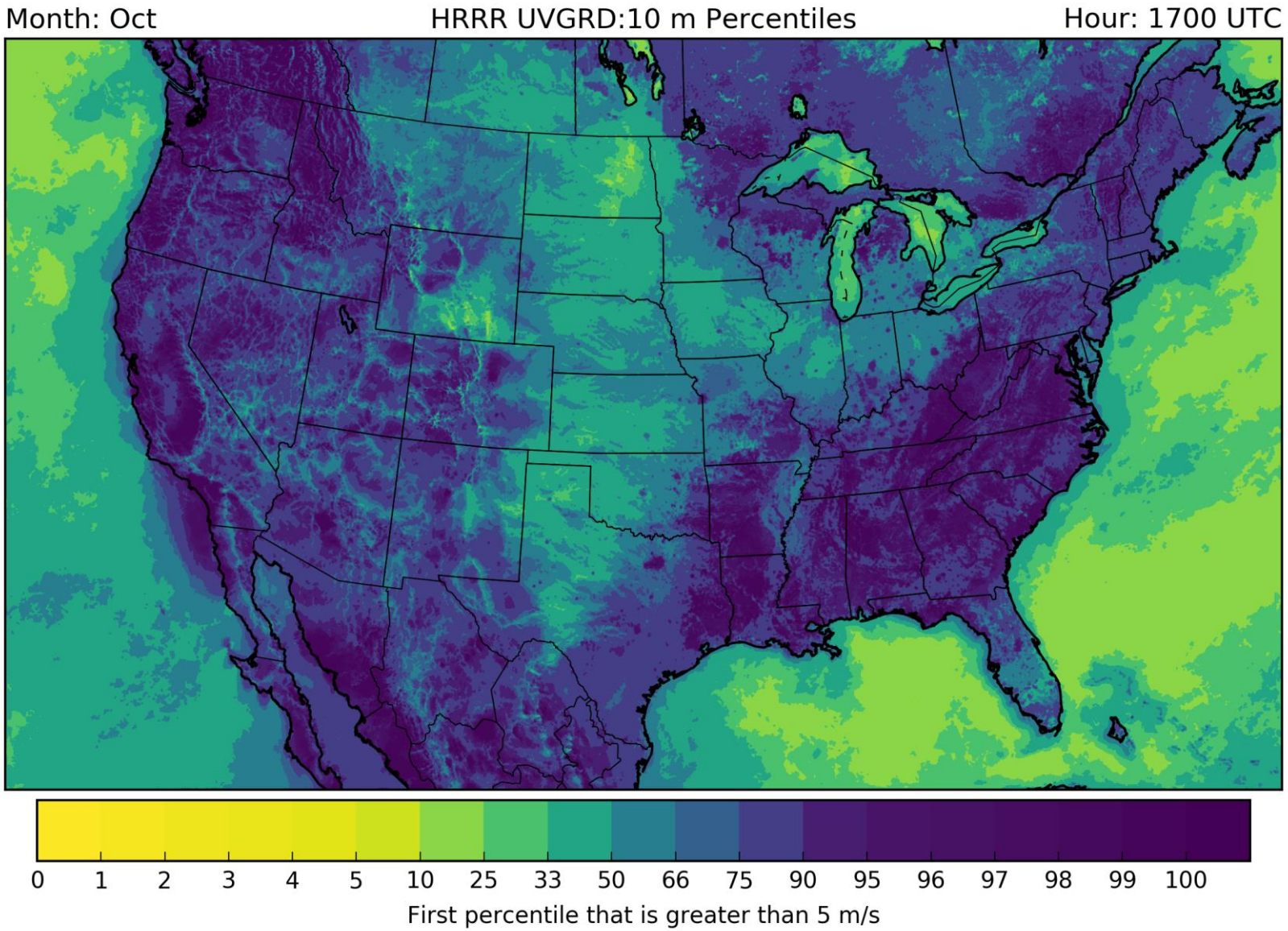
20-year Observed 5th-95th Percentiles

Credit: Chris Galli

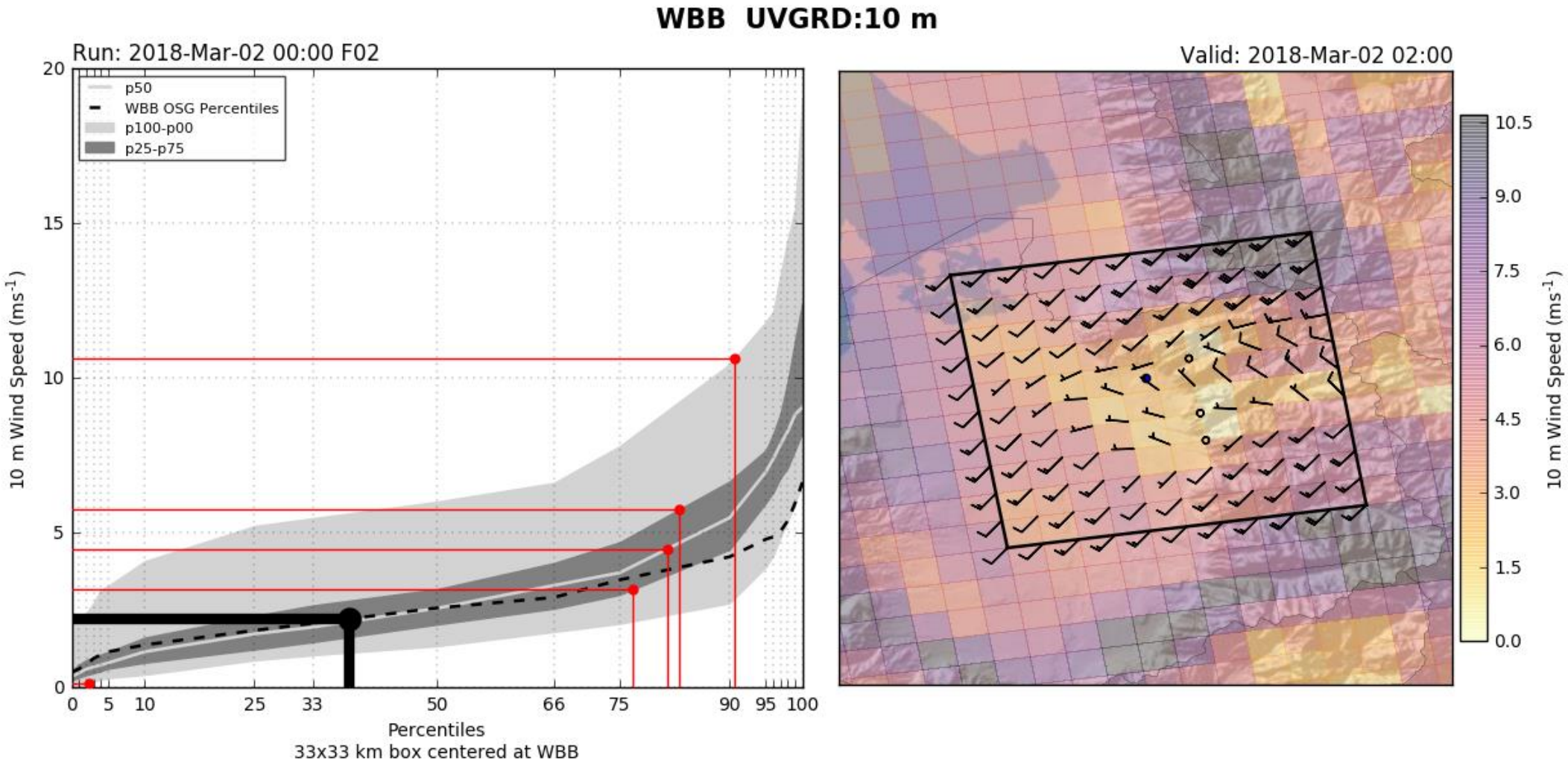
January Freezing Temperature Climatology



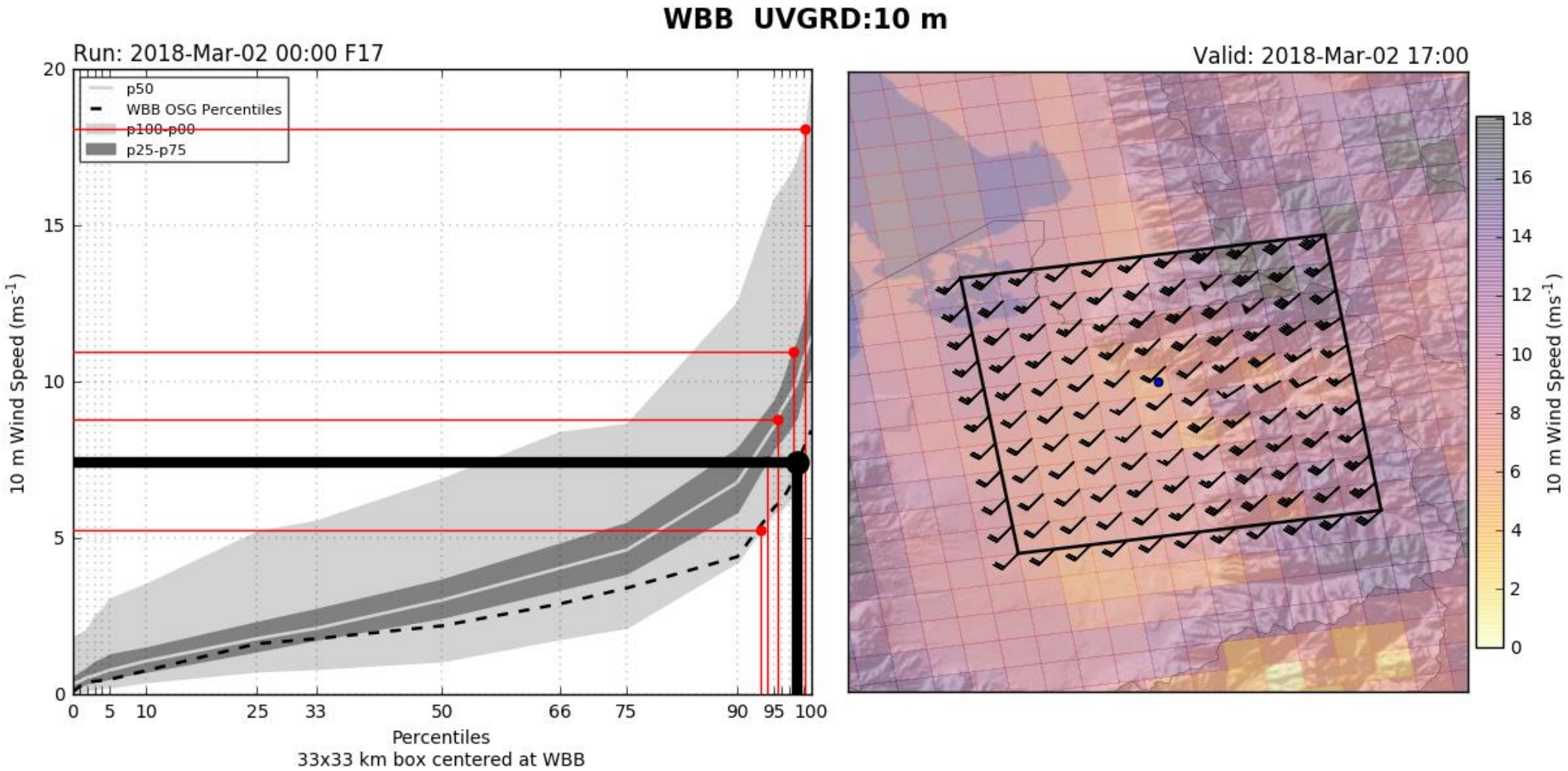
October Wind Climatology



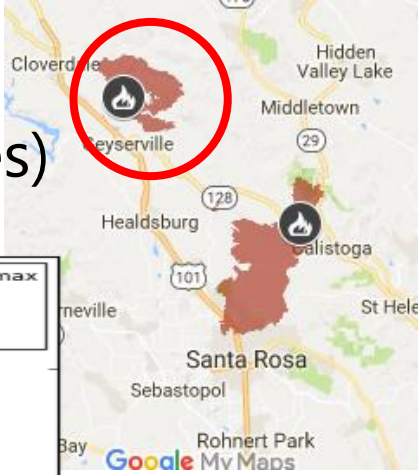
HRRR Climatology vs Single HRRR Run

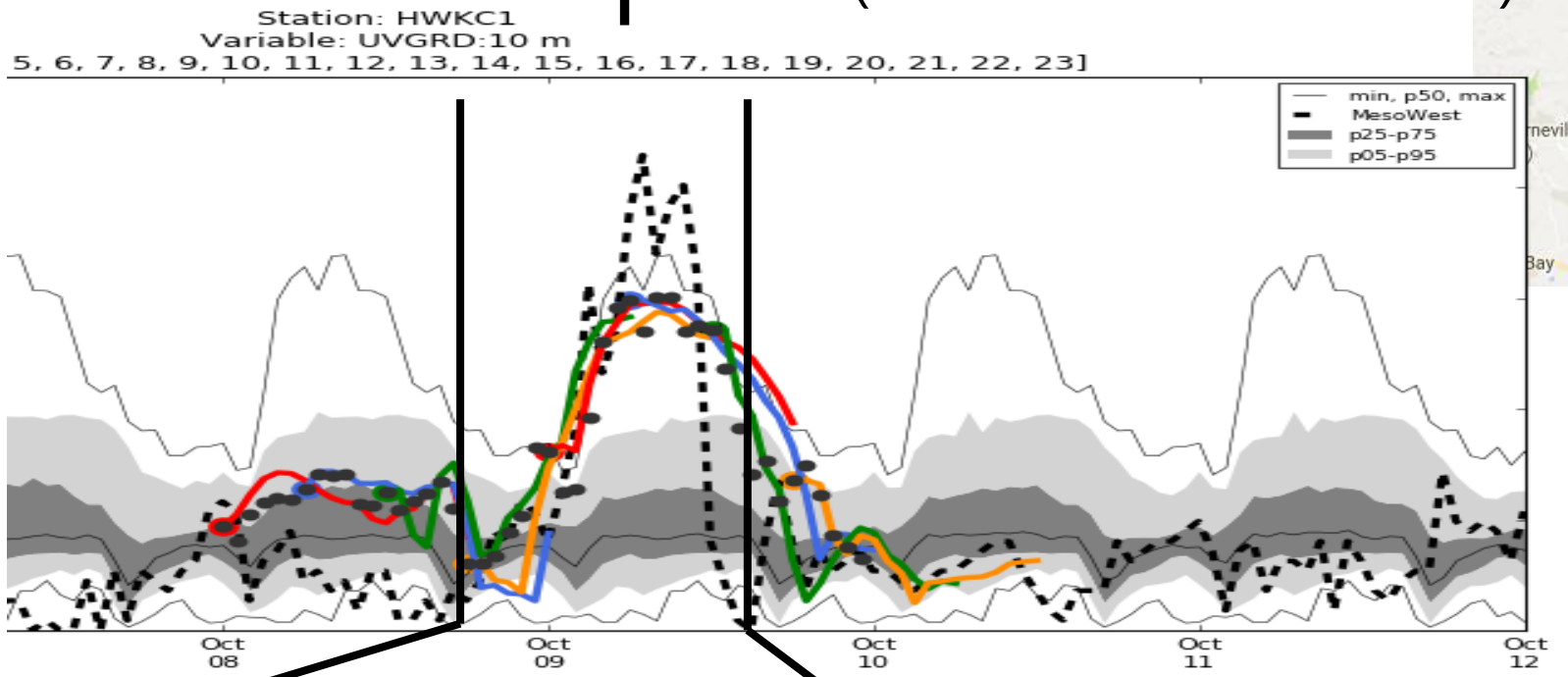


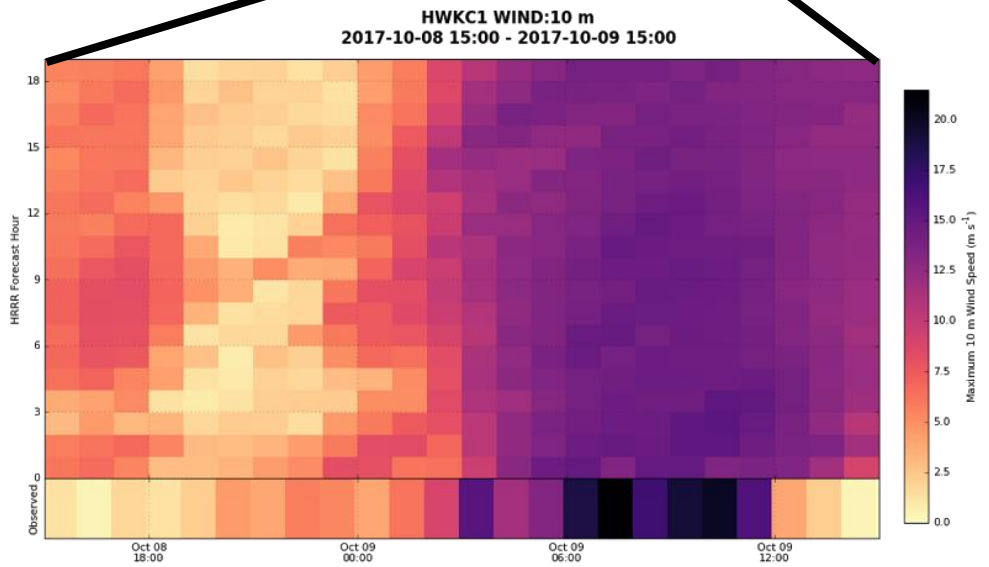
HRRR Climatology vs Single HRRR Run



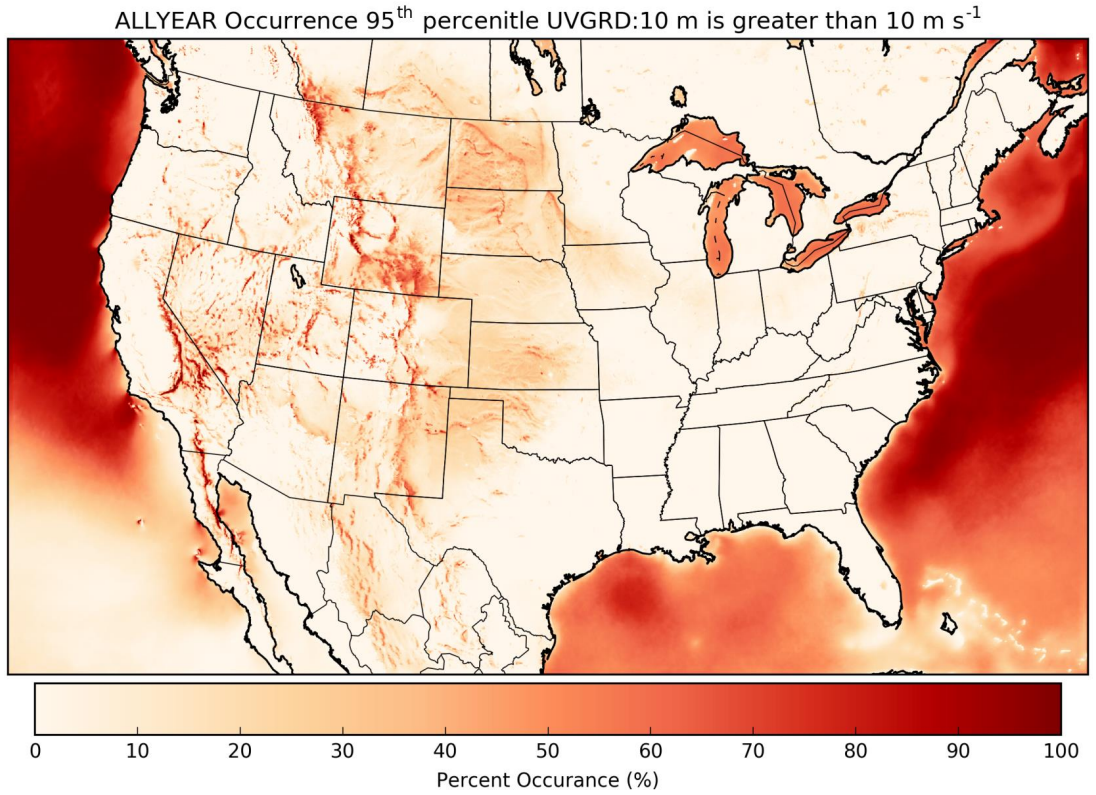
Central LNU Complex (Tubbs and Pocket Fires)

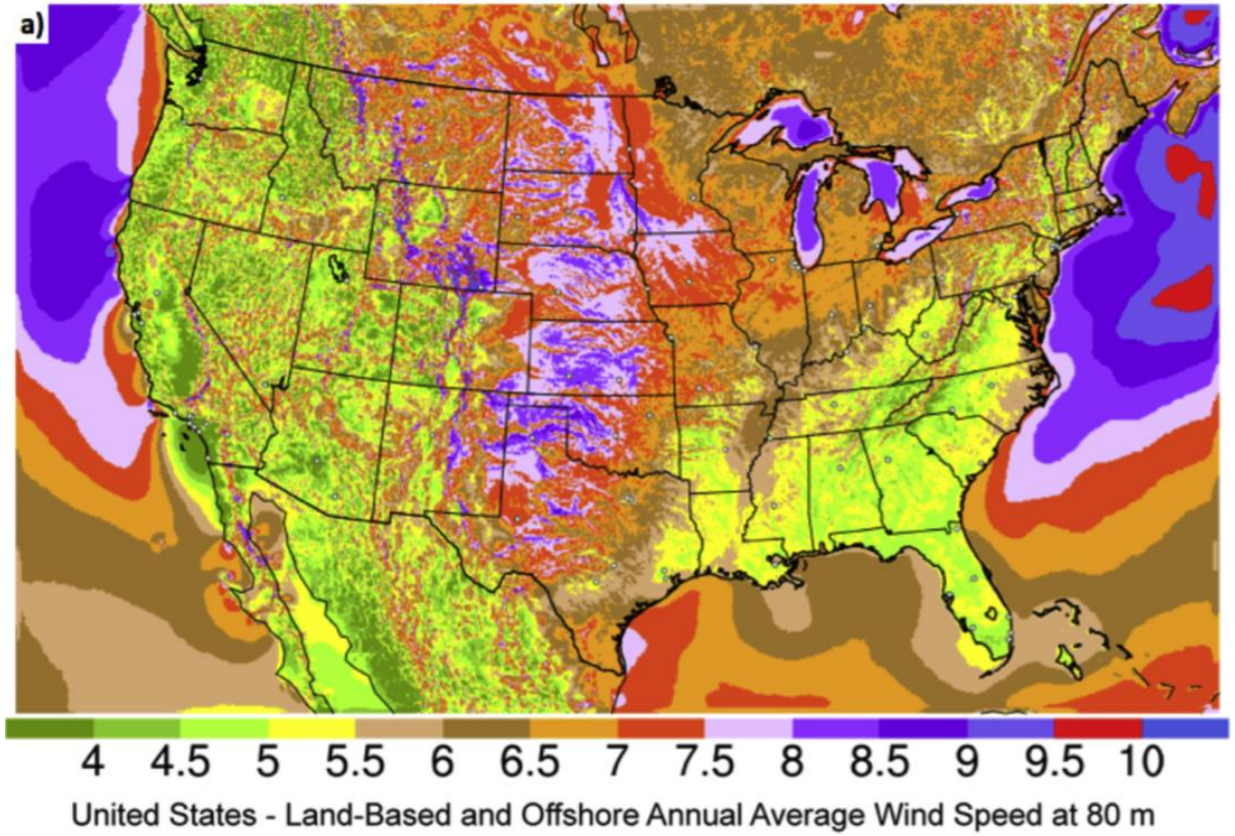






Wind Climatology





James et al. 2017