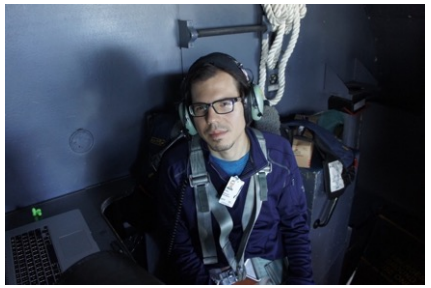
Interpretation of Probabilistic Surface Ozone Forecasts: A Case Study for Philadelphia

10th International Workshop on Air Quality Forecasting Research
20-22 October 2021

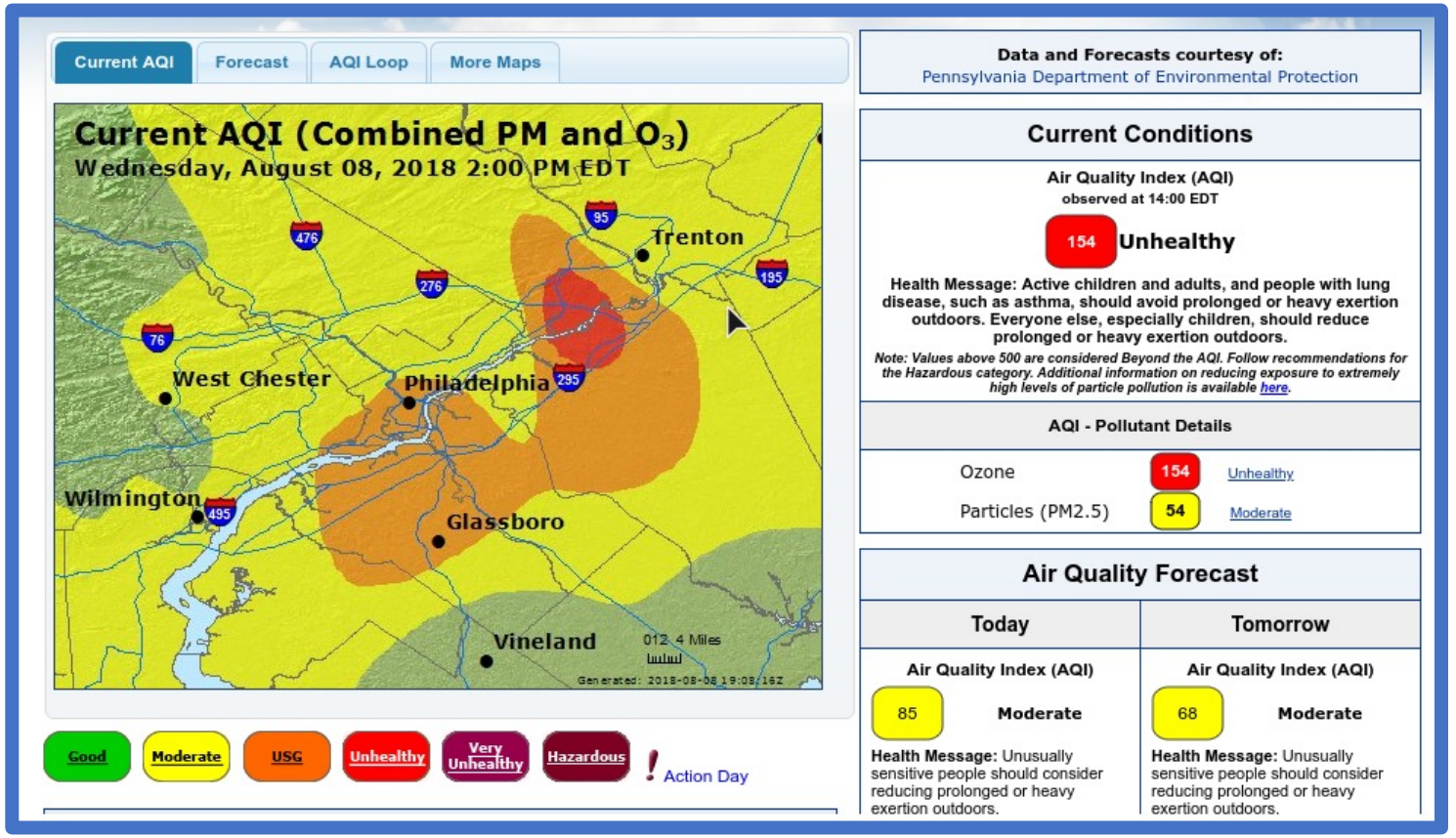
Nikolay V. Balashov[1,2], Amy K. Huff[3], Anne M. Thompson[4,5]

[1]NASA Global Modeling and Assimilation Office (GMAO), NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA
[2]Visiting Assistant Research Scientist, ESSIC, University of Maryland, 5825 University Research Court, College Park, MD, 20740, USA
[3]I. M. Systems Group, College Park, Maryland
[4]Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD, 20771, USA
[5]UMBC/JCET, Baltimore, MD 21250



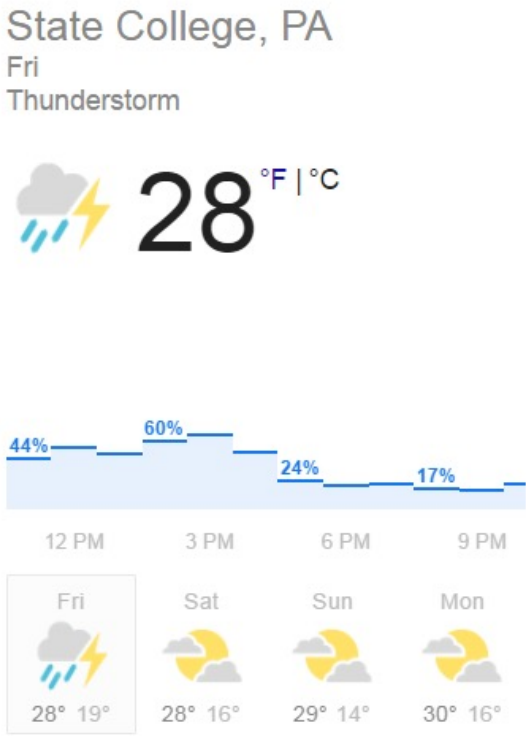
Operational Forecasts

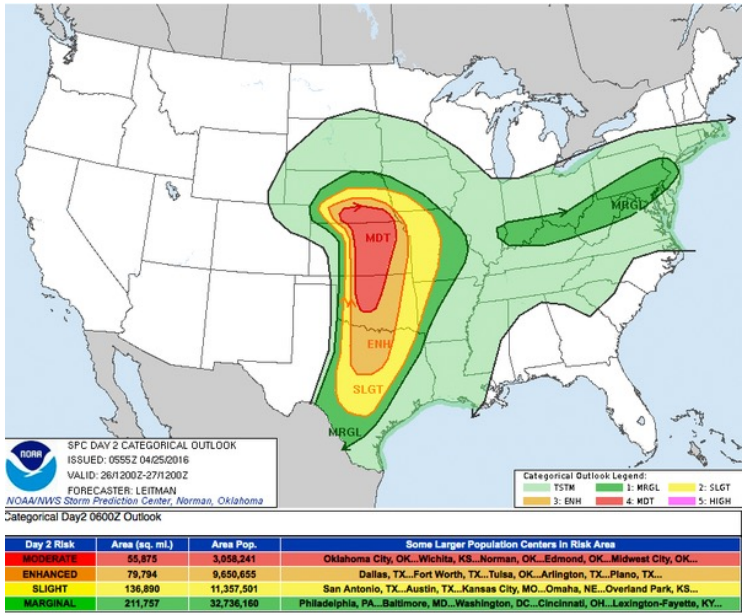
- Daily operational forecast is issued in the US (by AirNow system) to warn the public of potential unhealthy air
- Forecasters who issue these forecasts use a variety of prediction tools
- These tools include NOAA National Air Quality Forecasting Capability (NAQFC) modeling system
- Recently there has been movement toward probabilistic prediction of ozone (*e.g. Pinder et al., 2009*)

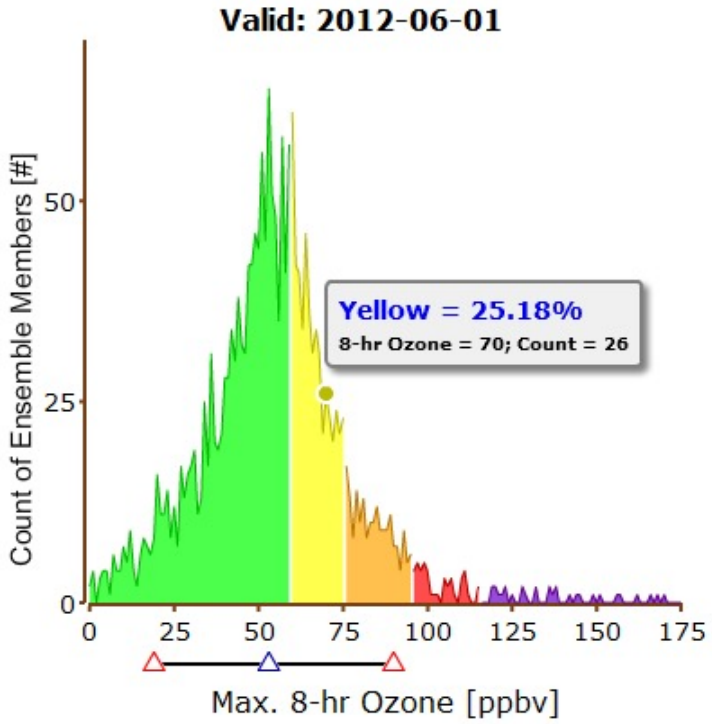


Probabilistic prediction

Probabilistic forecasts are vague!
-Is it going to rain?
-Will there be a tornado?
-Will there be an exceedance?
-Decision makers want “yes” or “no” answer

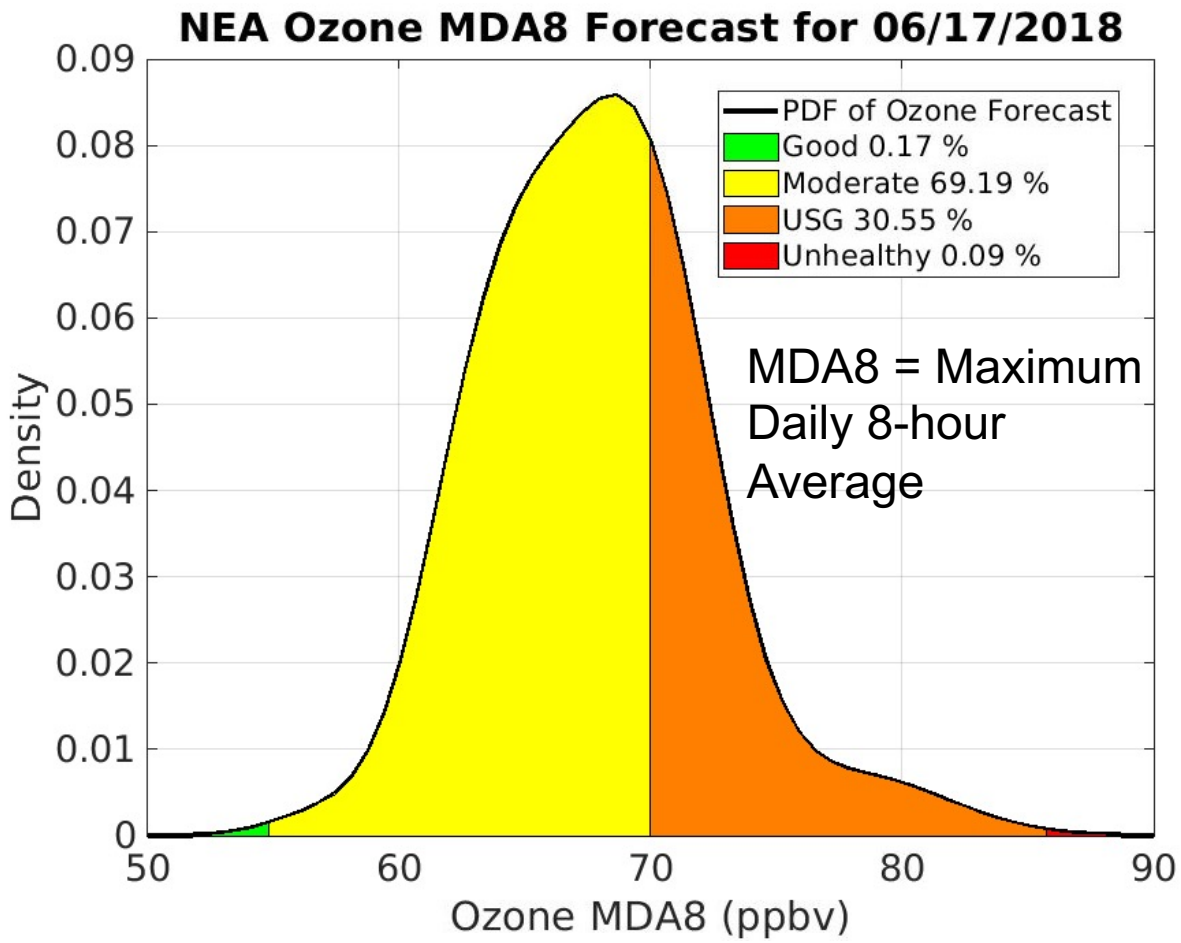






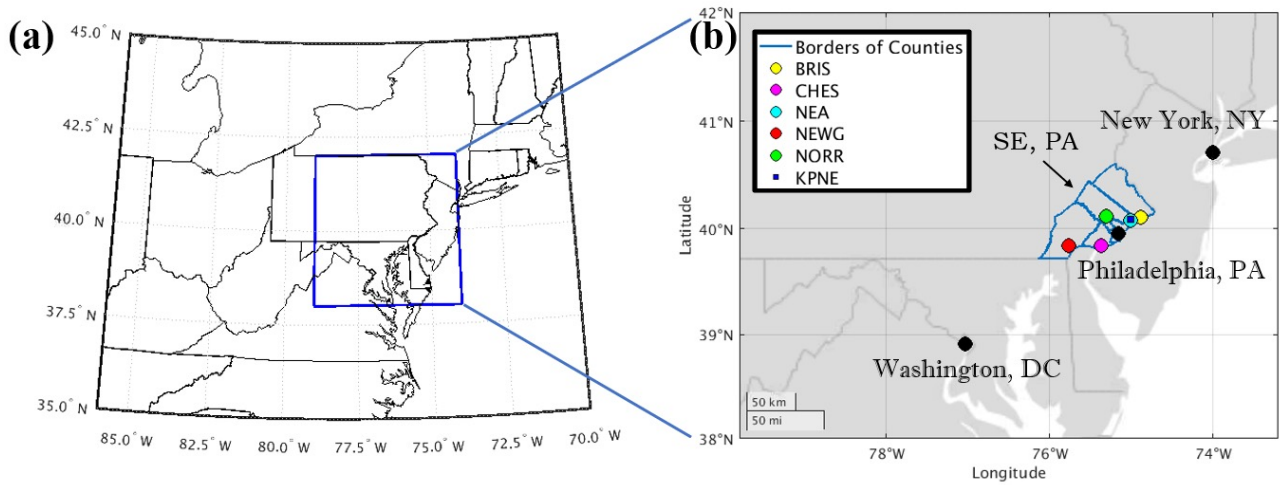
Exceedance (yes) or no exceedance (no)?

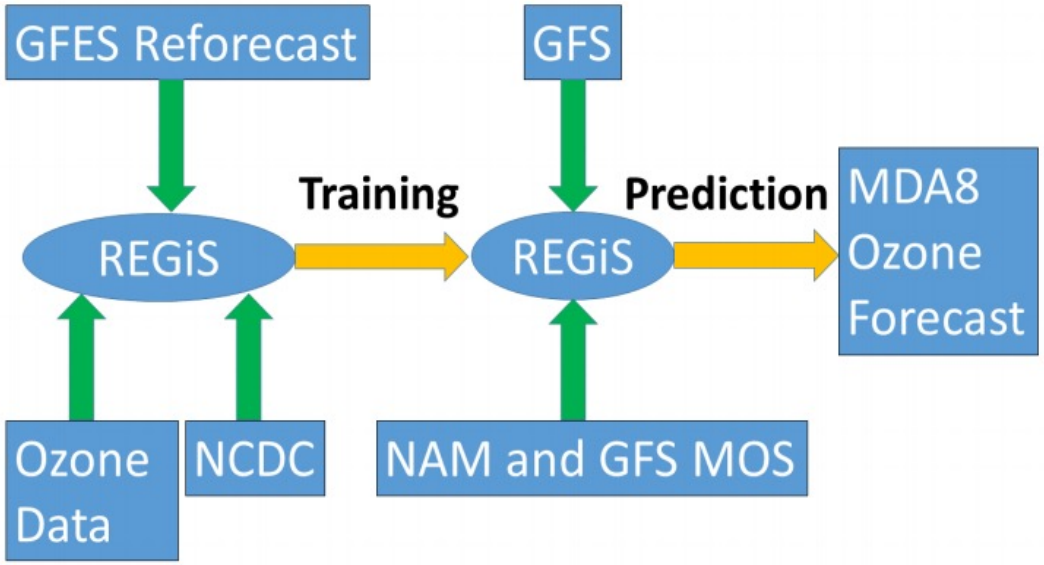
- Probabilistic forecasts contain more information than deterministic ones and decision makers would like to make use of these probabilistic forecasts
- But their interpretation can be challenging
- One approach is to consider ozone exceedance: current EPA ozone standard is 70 ppbv maximum daily 8–hour average (MDA8)
- For instance, should we forecast exceedance or not in the example forecast shown on the right?
- The question is not trivial and depends on the model, location, and stake holders' objectives
- Here we will perform a case study to examine this problem in more detail



Case study set-up

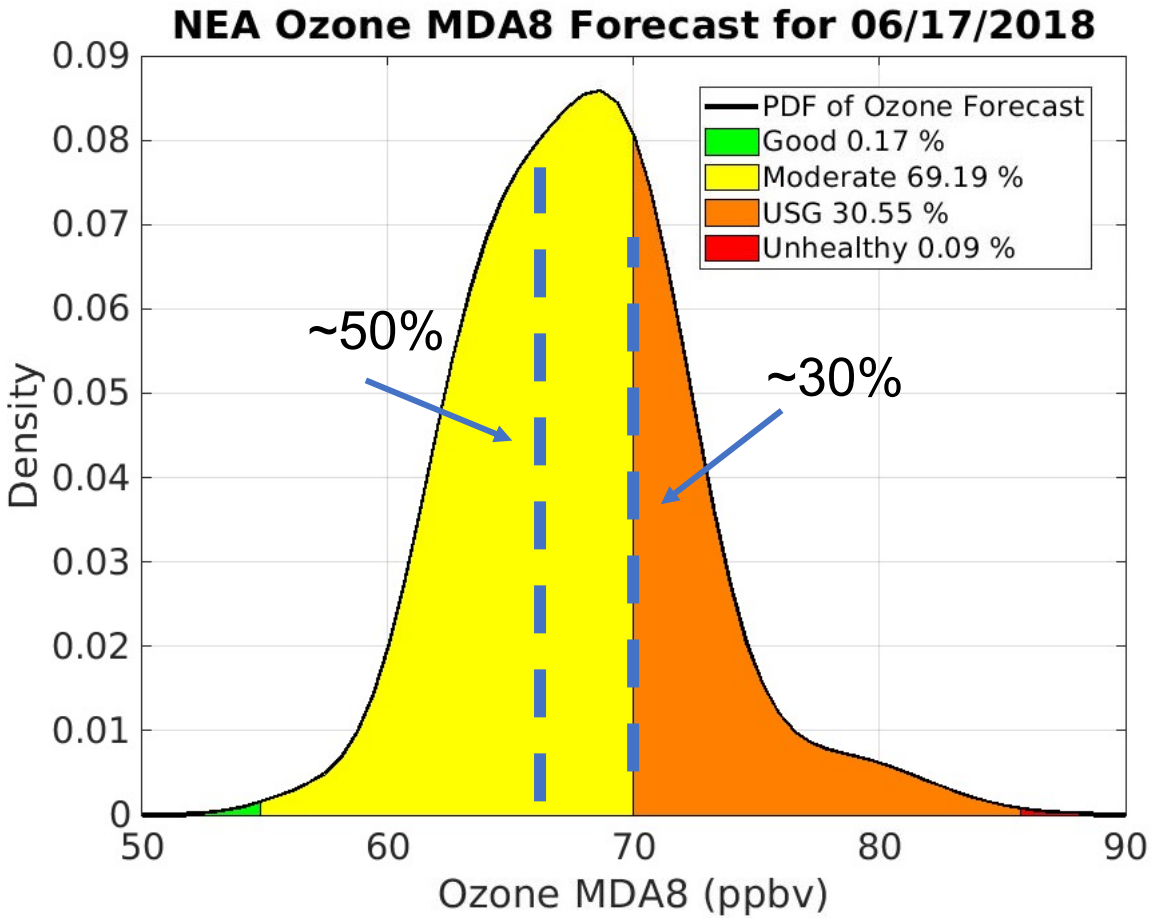
- We perform our experiment in Southeastern PA – Philadelphia
- We use experimental statistical probabilistic model REGiS developed by *Balashov et al., 2017*
- The probabilistic forecast is the result of ozone prediction based on a variety of synoptic patterns
- REGiS operational model schematic is shown below the maps
- Training data for REGiS: 2000-2011
- Calibration data: 2012-2014
- Evaluation data: 2015-2018





Calibration 1

- Now that training is done (2000-2011), we come back to our initial example shown earlier
- We will use calibration data (2012-2014) to decide at what **threshold** we should declare exceedance
- REGiS **exceedance threshold** decides what is “yes” and what is “no” (Is it 50%? Is it 30%?)
- *Wilks 2012* recommends 4 different methods to determine such **exceedance threshold**



Calibration 2

Depending on what **exceedance threshold** we pick we are going to have different combinations of metrics based on the 2 by 2 contingency table shown below:

EVENT		OBSERVATIONS YES	OBSERVATIONS NO	Total
MODEL	YES	a (hit)	b (false alarm)	a+b Yes (Model)
MODEL	NO	c (miss)	d (correct rejection)	c+d No (Model)
	Total	a+c Yes (OBS)	b+d No (OBS)	N=a+b+c+d Total of events

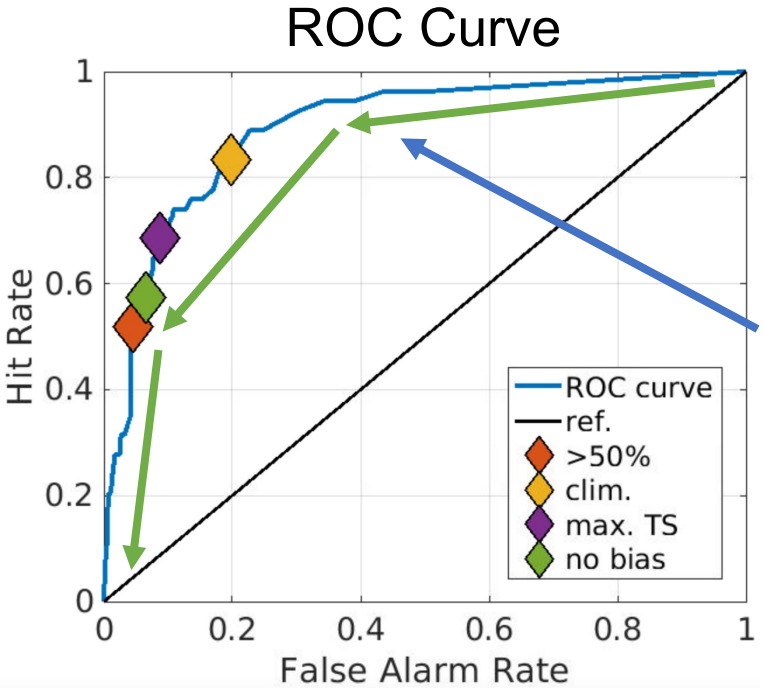
Determining exceedance threshold

1) More likely event (>50%)

2) Forecast probability is greater than a given climatology

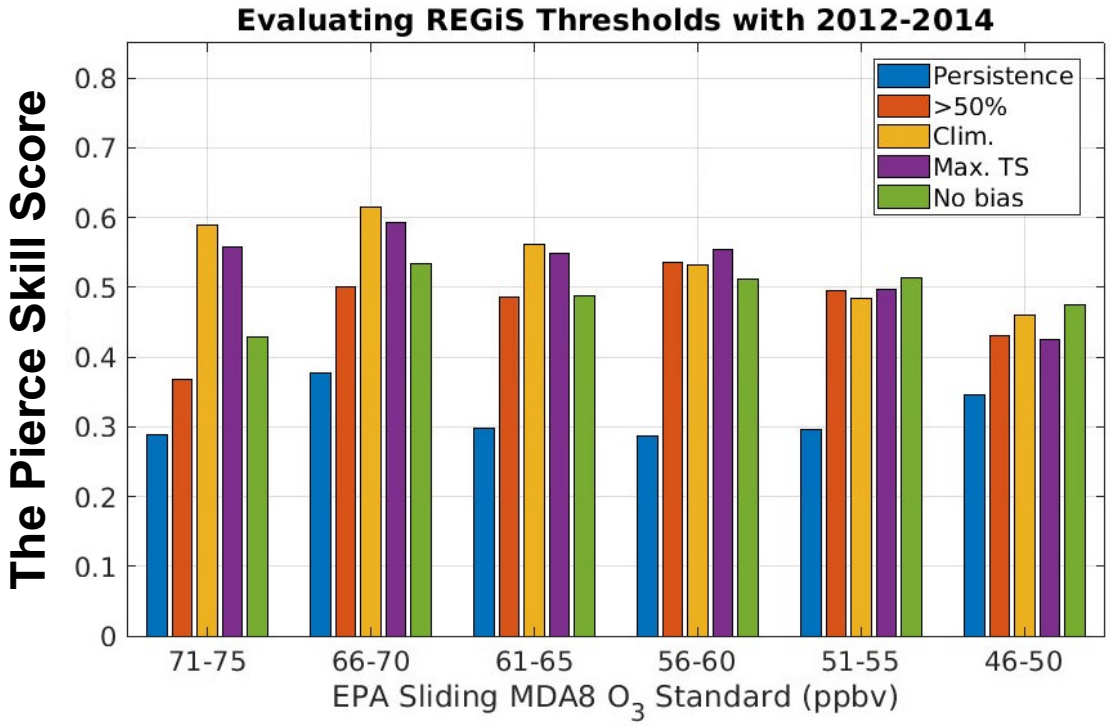
3) Maximize $the\ threat\ score\ (TS) = \frac{hits}{hits+misses+false\ alarms}$

4) Minimize $the\ bias\ ratio = \frac{hits+false\ alarm}{hits+misses}$



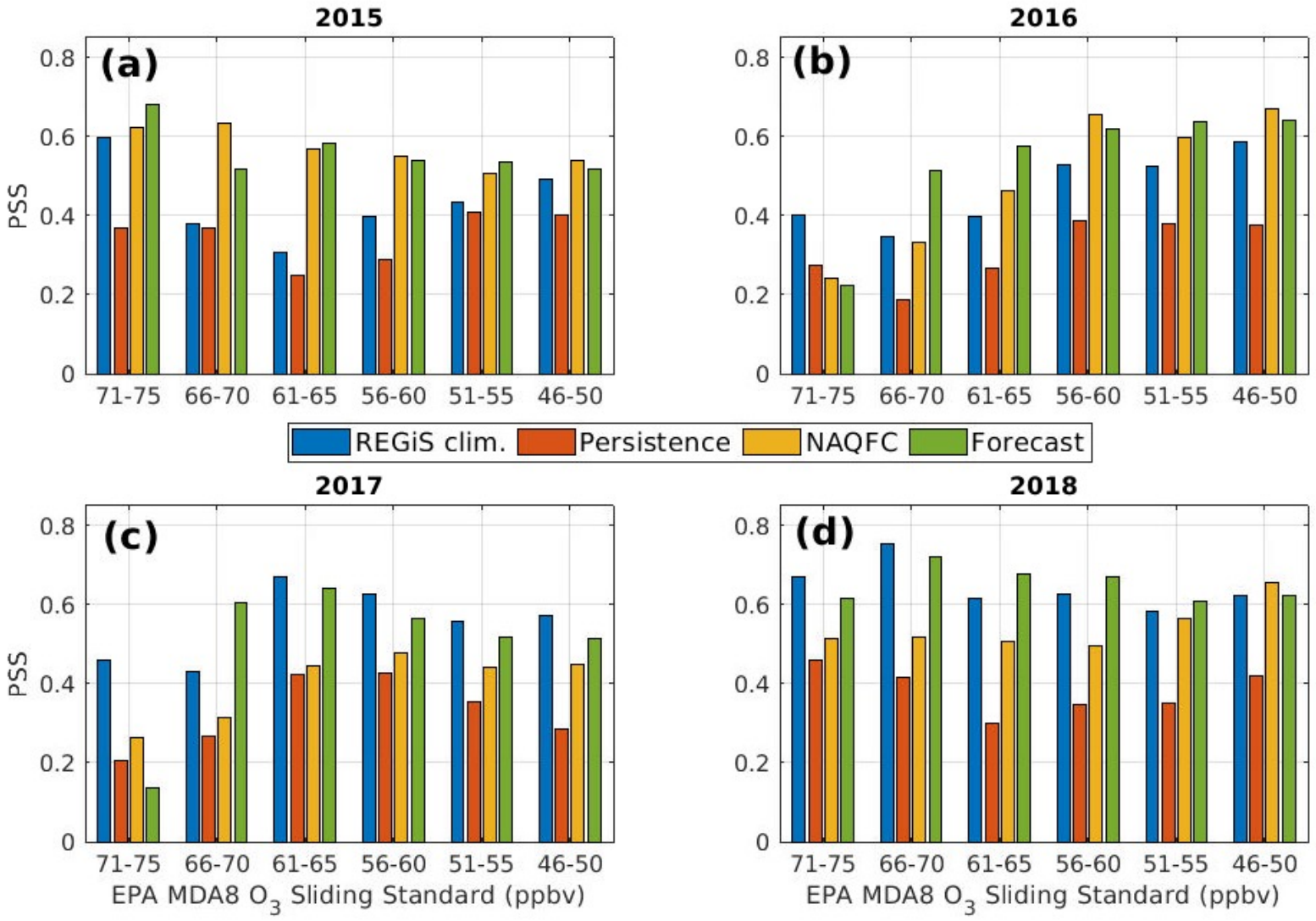
For MDA8 O_3 = 70 ppbv

As we change threshold from 0 to 100 hit rate decreases along with false alarm rate



Evaluation (2015-2018)

- We evaluating calibrated REGiS against persistence, NAQFC (NOAA ozone model), and operational forecast
- Operational forecasters tend to outperform other predictions
- REGiS calibrated by climatology and NAQFC are comparable, and even occasionally REGiS outperforms NAQFC
- As expected, persistence shows lowest skill



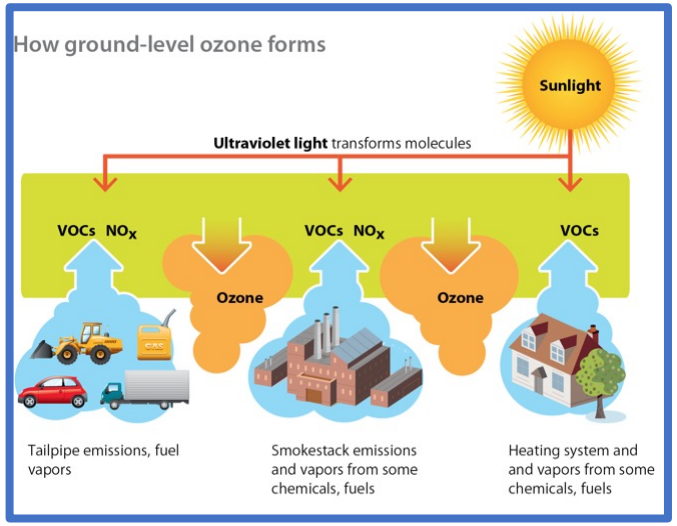
Summary

- Reducing probabilistic forecast to ”yes” and “no” is an important and relevant problem
- In this work we test probabilistic statistical ozone model called REGiS at Philadelphia
- We find that calibrating REGiS exceedance threshold using climatology produces the most skillful forecast based on the PSS
- It is possible that for other probabilistic models different thresholds need to be used, but the process of calibration nonetheless is recommended

Thank you!

Extra Slides

National Ambient Air Quality Standard (NAAQS) for Ozone



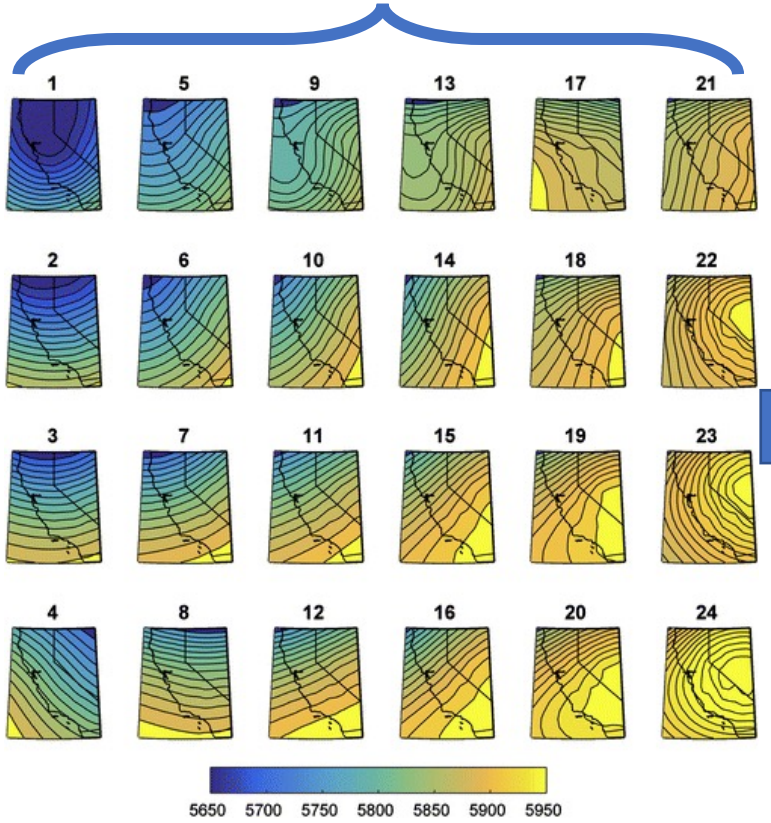
- Prolonged exposure is harmful for humans, animals, plants, etc.
- Ozone daily maximum 8-hour average (MDA8) – is regulated (running average)
- Exceedance threshold – MDA8 of 70 ppbv (set by EPA – lowered it periodically)

O_3 (ppbv)	Category
0-54 (8-hr)	Good
55-70 (8-hr)	Moderate
71-85 (8-hr)	Unhealthy for Sensitive Groups
86-105 (8-hr)	Unhealthy
106-200 (8-hr)	Very Unhealthy
201-500 (8-hr)	Hazardous

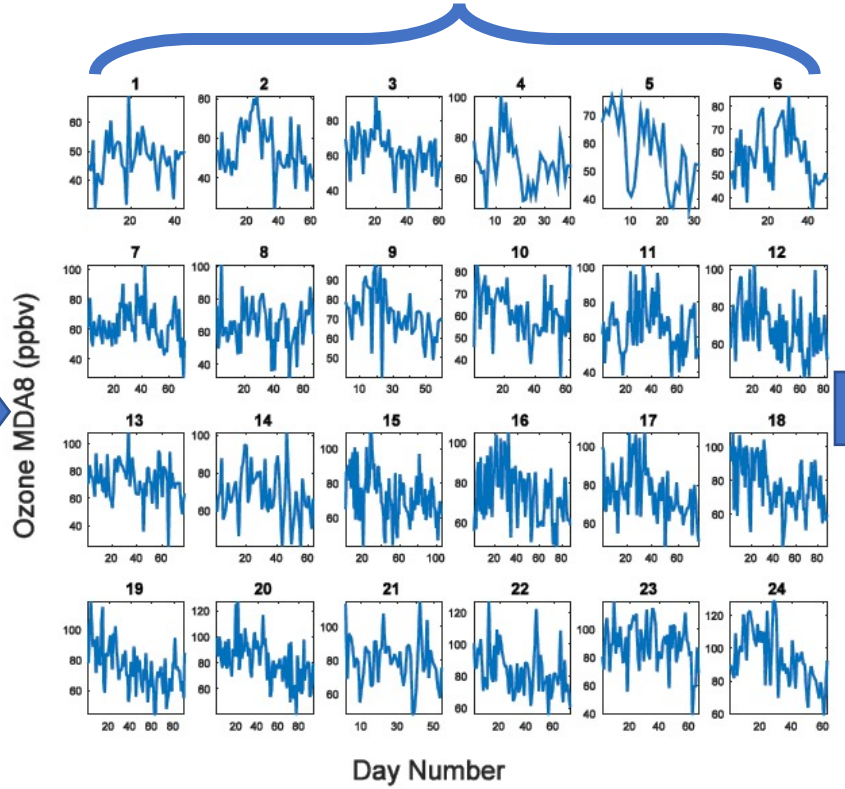
What is REGiS?

REGiS is a machine learning model that generates probabilistic ozone forecasts (Balashov et al., 2017)

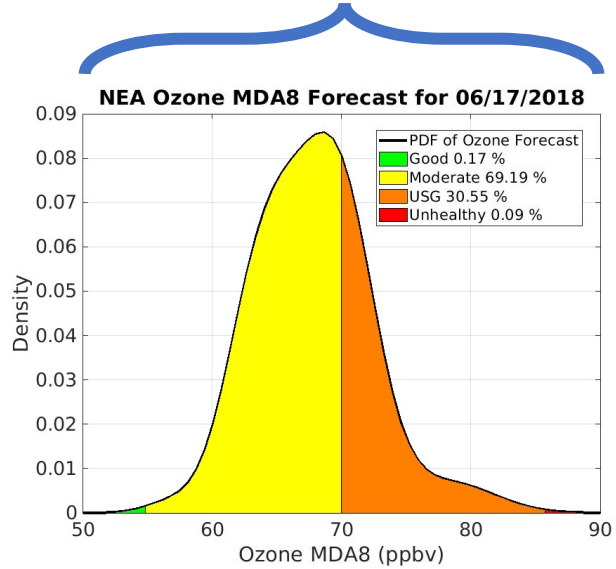
Use Self-Organizing Maps (SOMs) to classify multiple synoptic patterns



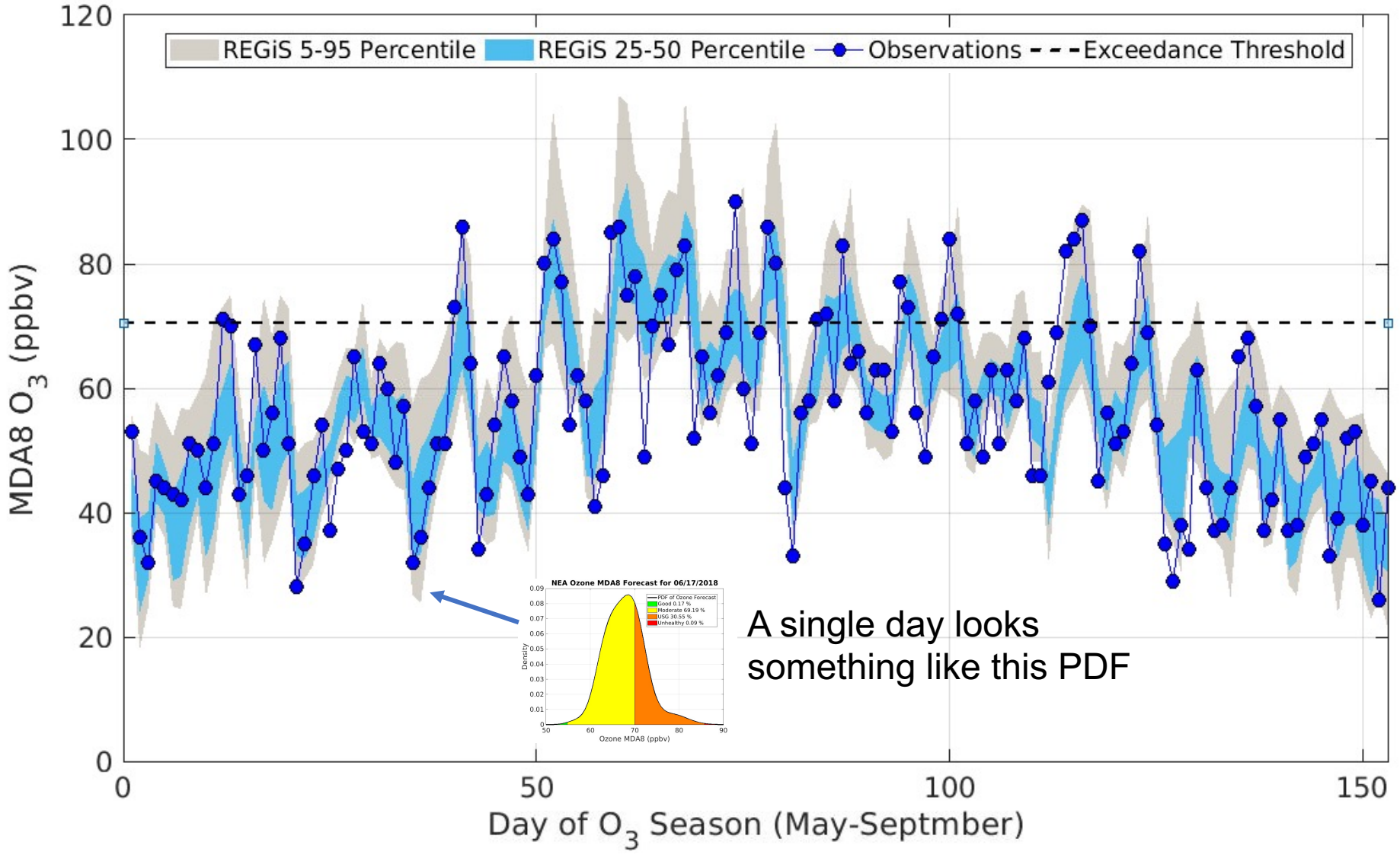
Create regressions using meteorological predictors as a function of synoptic pattern



Combine into probabilistic forecast

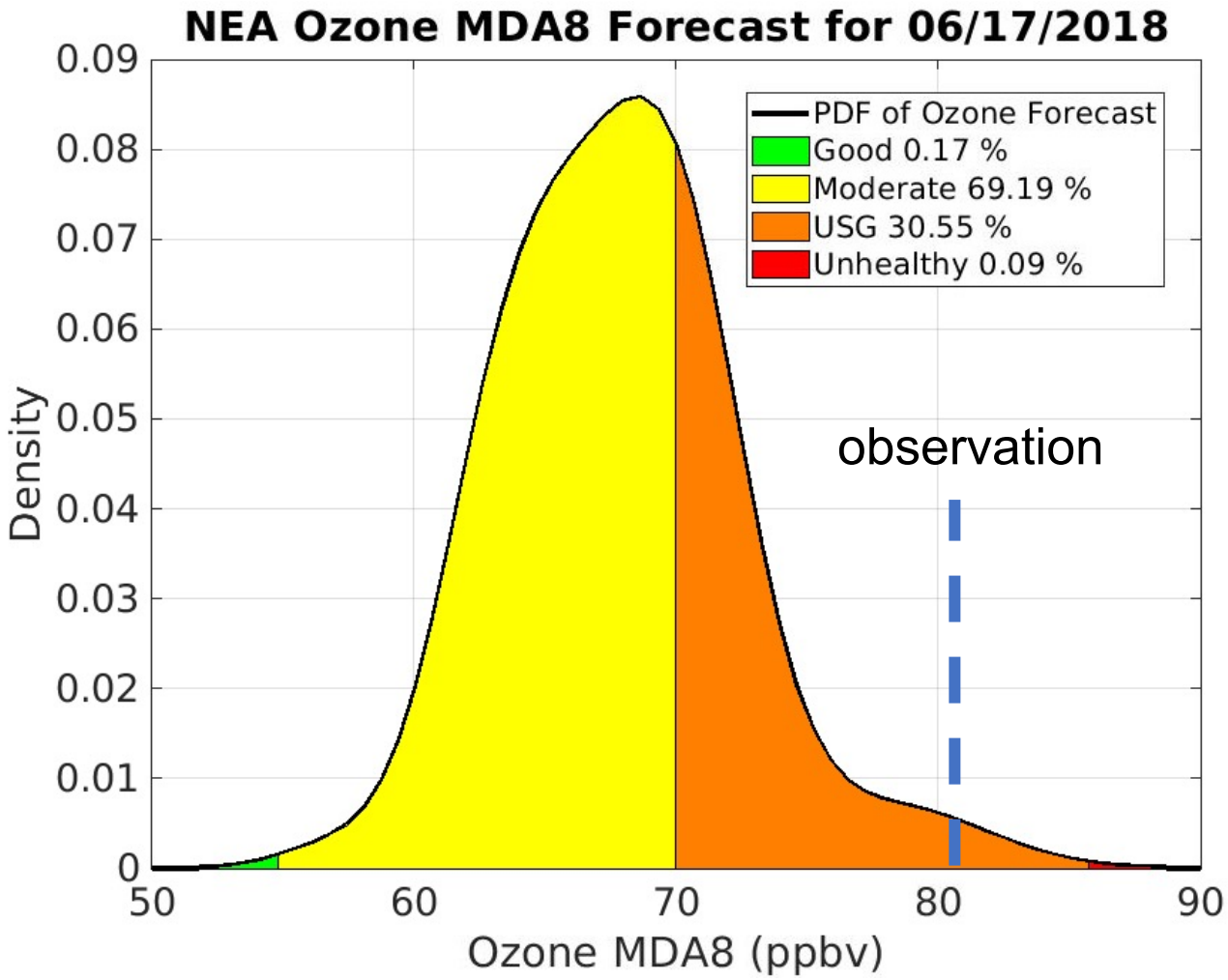


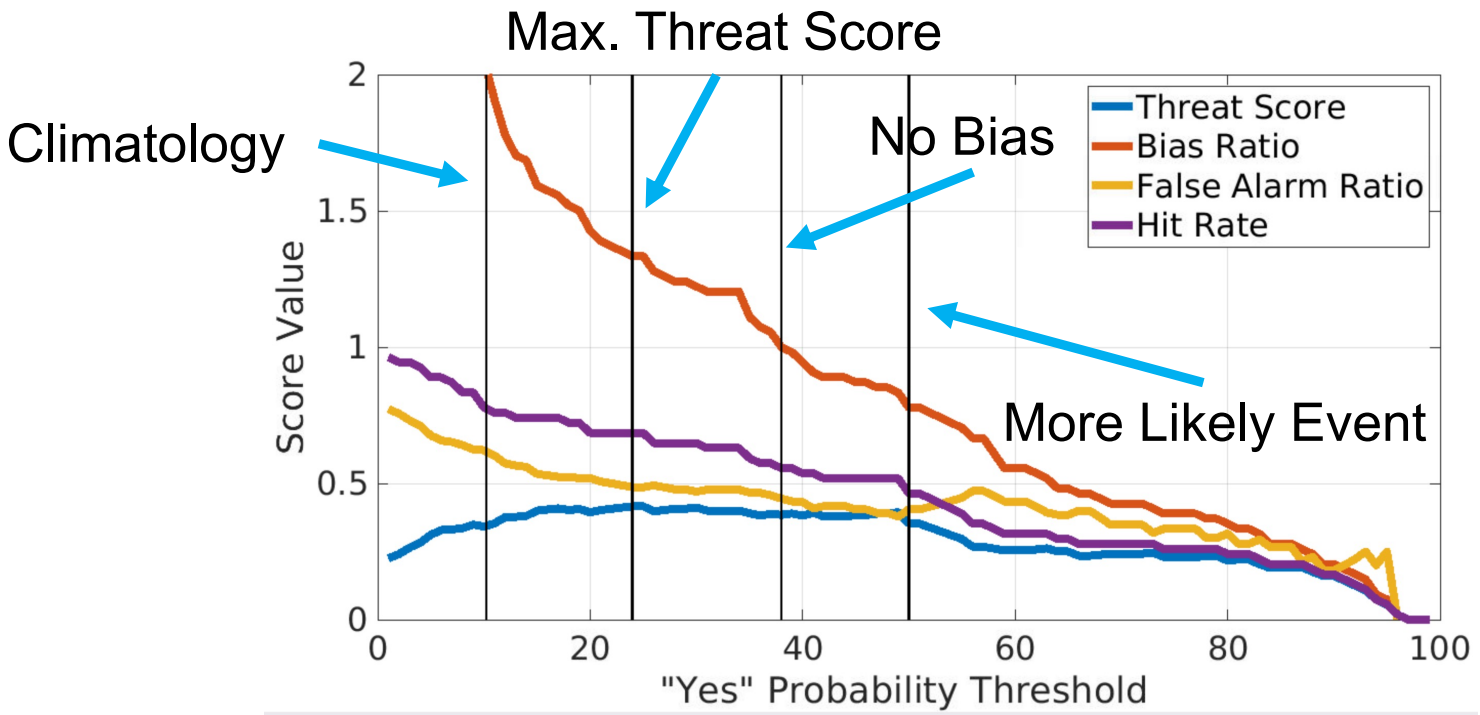
MDA8 Ozone 2012, NEA (ozone station in Philadelphia) Obs. (blue) vs. REGiS (light blue/gray)



So, is it exceedance (yes) or no exceedance (no)?

- Here we come back to the the probabilistic forecast shown earlier (forecast is derived from REGiS)
- Using climatology and maximum TS will give us correct answer here
- Other thresholds that are above 31% would yield a wrong answer here




Max. Threat Score
Climatology
No Bias
More Likely Event
Threat Score
Bias Ratio
False Alarm Ratio
Hit Rate
Score Value
"Yes" Probability Threshold
0
0.5
1
1.5
2
0
20
40
60
80
100

Evaluation 1

- We evaluate REGiS using Pierce Skill Score (PSS) – an equitable Score
- X-axis shows sliding EPA MDA8 exceedance scale
- Sliding scale allows for simulation of a variety of exceedances
- Evaluating calibrated REGiS using independent 2015-2018 data indicates that in the given case **climatology** gives highest PSS score (especially when there are few exceedances), while using **more likely event** produces lowest PSS score