Ensemble Visualization and Verification in Python

Tyler Wixtrom

Texas Tech University

tyler.wixtrom@ttu.edu

Unidata Users Workshop

25-28 June 2018

Boulder, CO

Overview

1. Plots
   - Spaghetti Plots and Postage Stamps
   - Paintball Plots
   - Probability Plots
2. Verification
   - RMSE
3. Examples

Examples and Data Repository

https://github.com/tjwixtrom/workshop2018

Why Ensembles and Fancy Plots?

Ensembles are more than a collection of deterministic forecasts

- Probabilistic forecasts
    - Most likely and range of solutions
- Information regarding forecast uncertainty
- Variability in time, space, and intensity

Spaghetti Plots and Postage Stamps

Spaghetti plots and postage stamp plots are both concise, simple methods of viewing all members in a ensemble on the same plot.

Spaghetti plots are generally used for contoured fields (e.g. 500 hPa geopotential height) while postage stamps are more commonly reserved for shaded fields (e.g. simulated reflectivity).
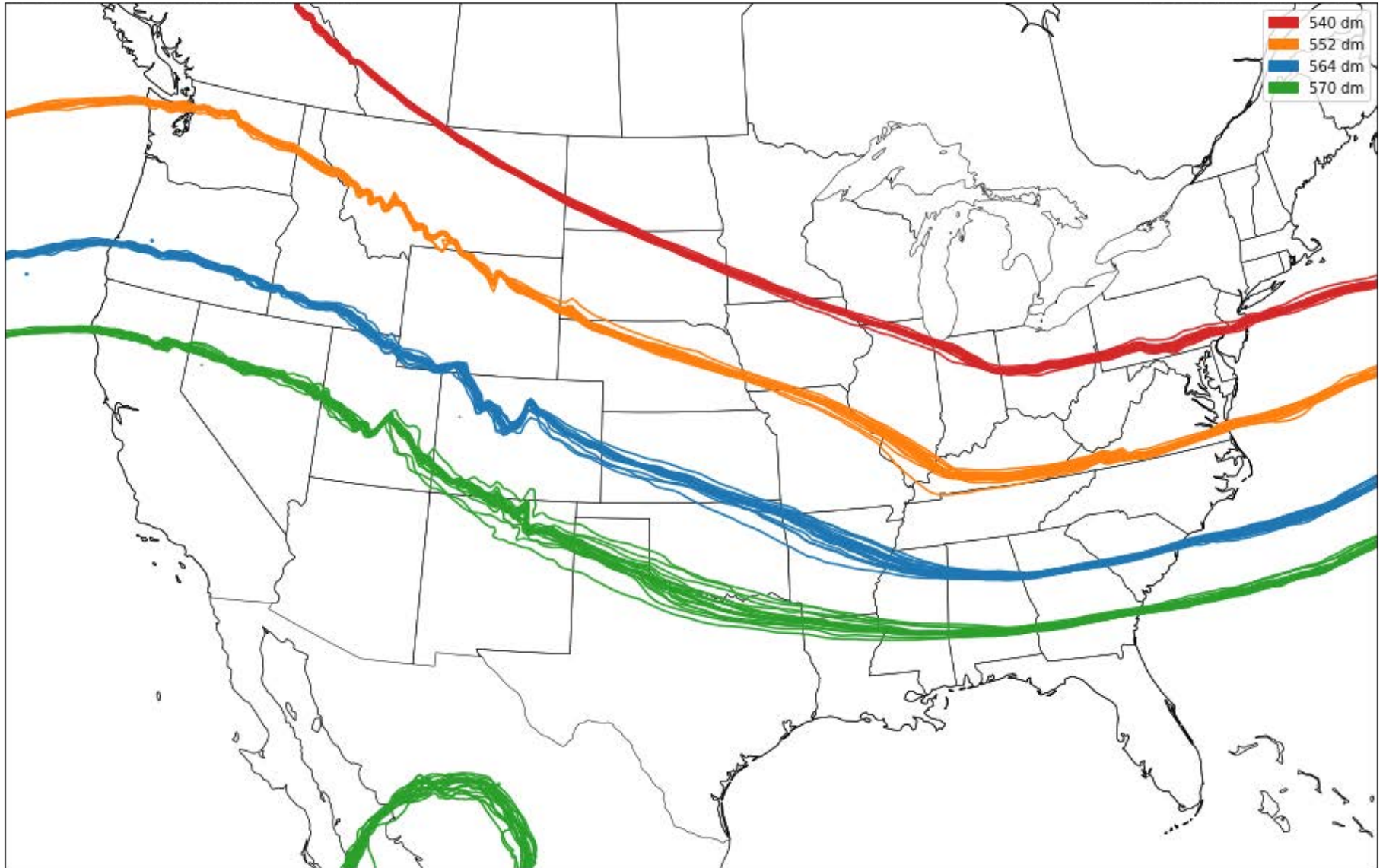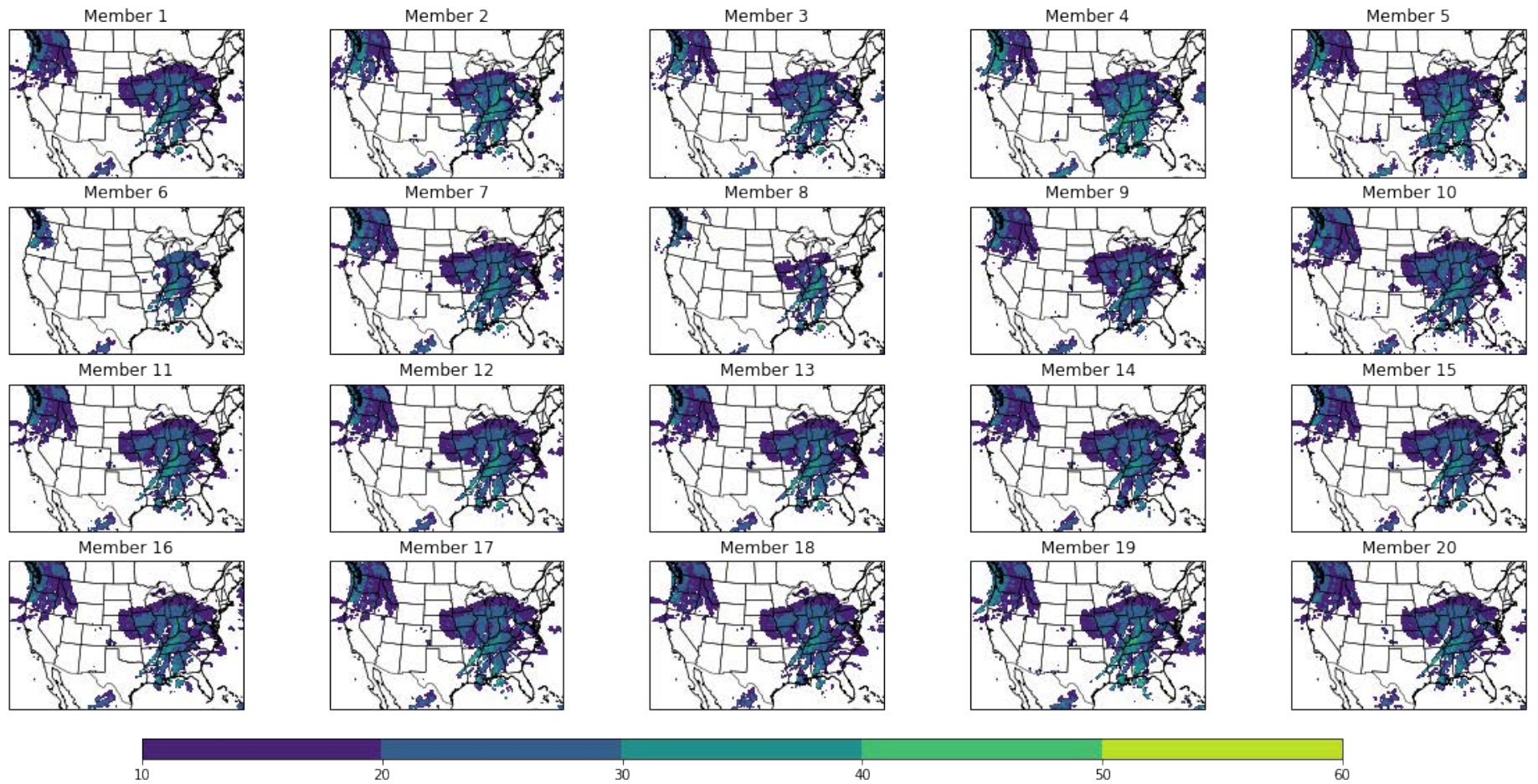
Example 500 hPa Spaghetti Plot

Ensemble Postage Stamps of Reflectivity Initialized 2015-02-01 12:00:00 VALID: 2015-02-01 18:00:00

Advantages

- Displays information from each individual member
- Full location and magnitude of each field is plotted
- Simple to interpret

Disadvantages

- Can be very busy
- Becomes increasingly hard to interpret as spread increases
- Easy to miss small differences among members (esp. postage stamps)

Paintball Plots

Paintball plots are plots where individual points greater than a specified threshold (e.g. simulated reflectivity $\geq$ 25 dBZ) are plotted, color-coded by ensemble member.

Ensemble Member Paintballs Reflectivity > 25 dBz

VALID: 2015-02-01 18:00:00
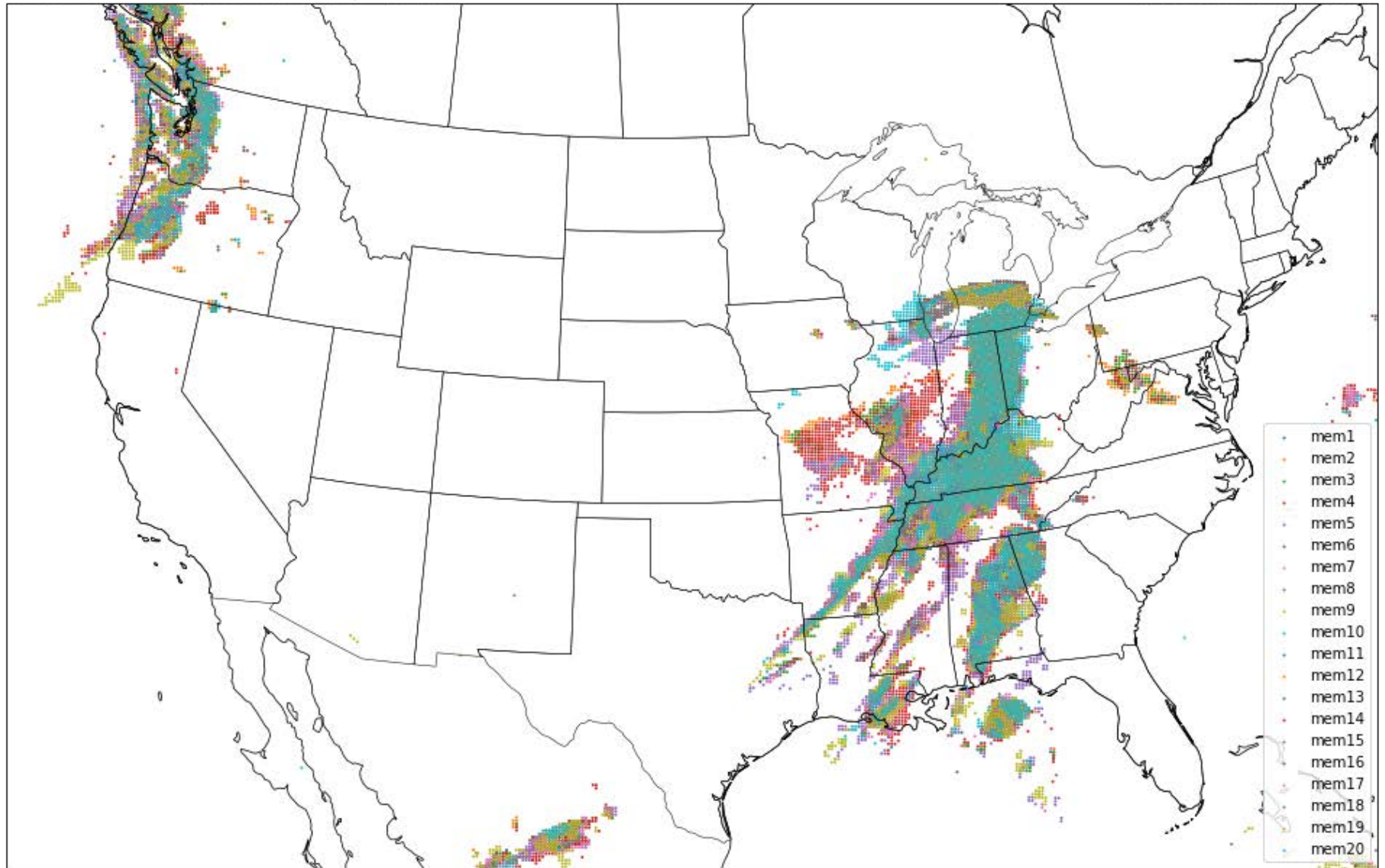
Advantages

- Easy to quickly see spatial variations among individual member solutions
- Plot remains relatively uncluttered, even for ensembles with many members

Disadvantages

- Very little information regarding variations in intensity
- Thresholds are arbitrary and may not be best for all cases

Probability Plots

If an ensemble contains a sufficient number of members and sufficient spread, probabilities of event occurence can be calculated based on the individual members. This is done at a single point with the following method (Schwartz and Sobash, 2017):

Ensemble Probability

1. Define the Binary Probability (BR) for each member at each point for field f and threshold q
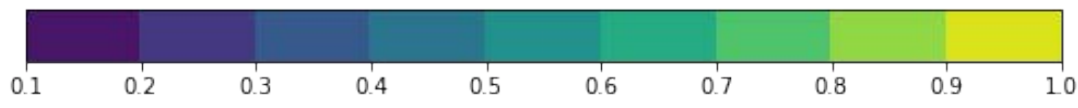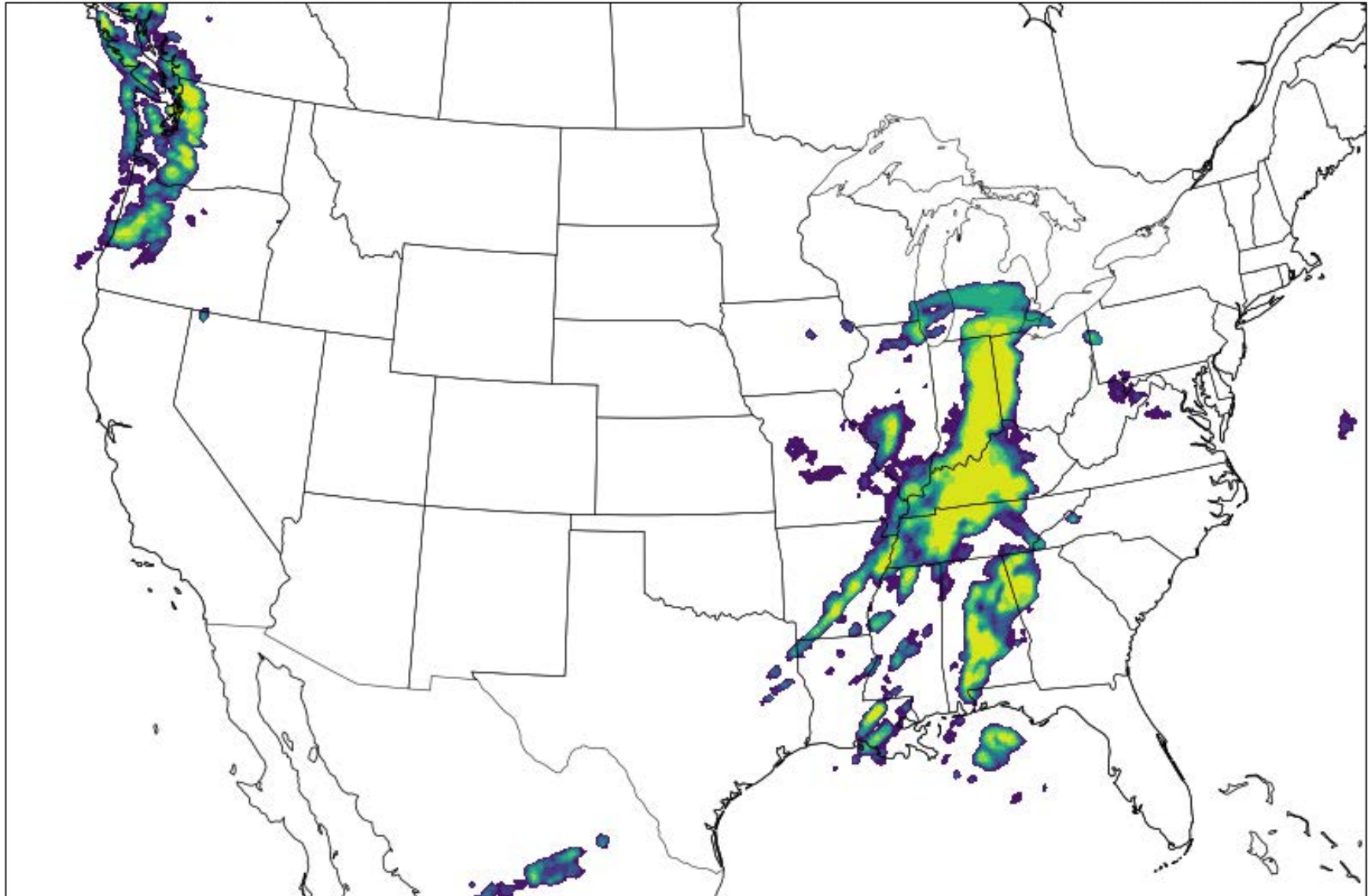
\begin{equation} BP(q)_{ij} = \left{

```
\begin{array}{ll}
    1 & \quad f_{ij} \geq q \\
    0 & \quad f_{ij} < q
\end{array}
```

\right. \end{equation}

1. Calculate the average of ensemble member binary probabilities at each point to define the Ensemble Probability (EP)

$$EP(q)_i = \frac{1}{N} \sum_{j=1}^{N} BP_{ij}$$

Point Probability of Reflectivity Greater Than 25 dBz   VALID: 2015-02-01 18:00:00

Example Ensemble Probability Plot

Neighborhood Ensemble Probability

Since the ensemble probability is valid for only a single point and does not account for small spatial differences in member solutions, the Neighborhood Ensemble Probability (NEP) can be defined as the probability of event occurence within a specified radius of any point (Schwartz and Sobash, 2017).

To calculate the NEP for a given field and threshold q:

- Begin by calculating the EP at each point
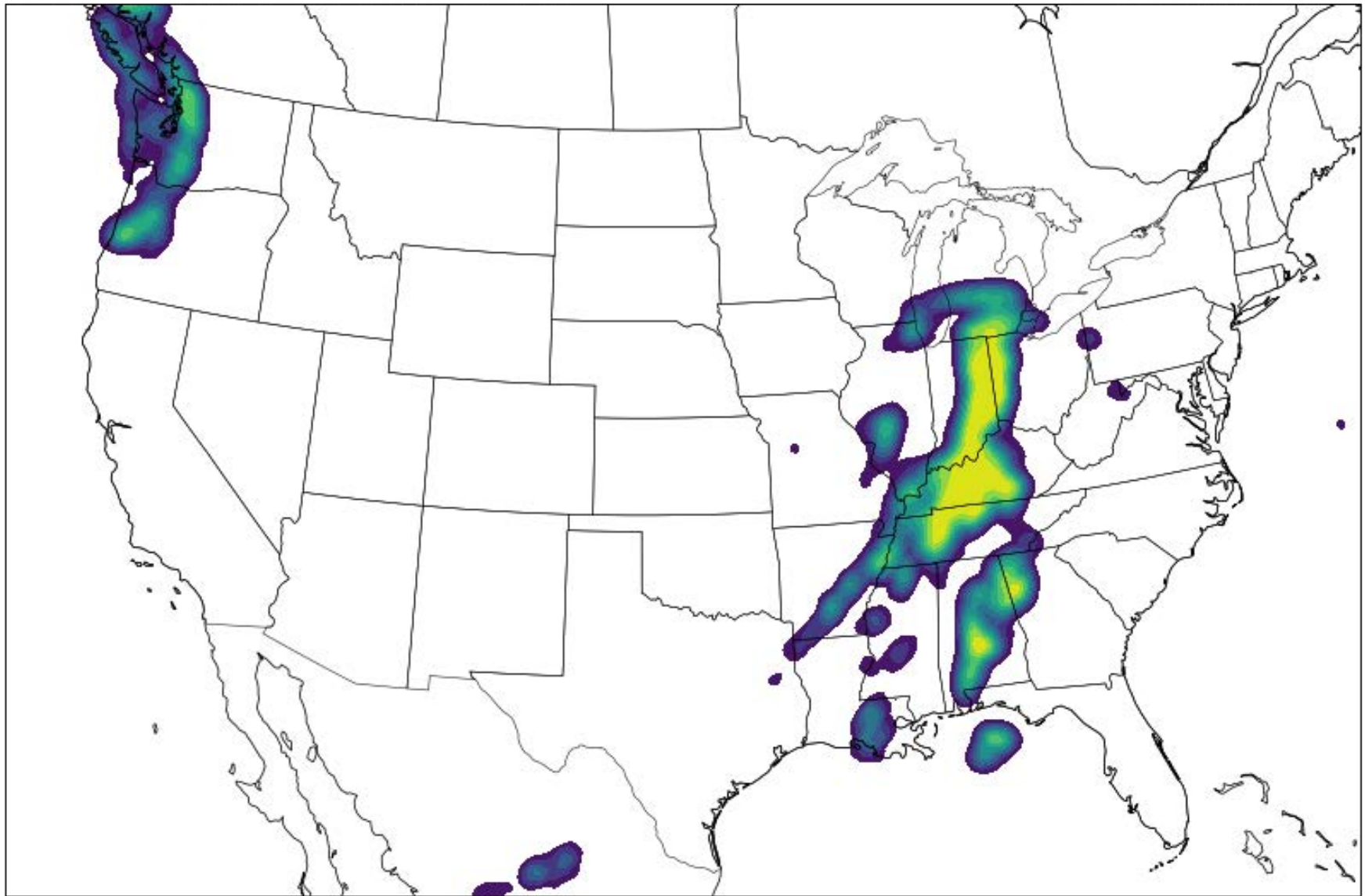- Find all points within specified radius and calculate mean of EP

\begin{equation}

`NEP(q)_{i} = \frac{1}{N_b}\sum^{N_b}_{j=1}EP_{ij}`

\end{equation}

NMEP of Reflectivity Greater Than 25 dBz within 40 km　　　　VALID: 2015-02-01 18:00:00

EP and NEP Differences

- NEP will have smoothed gradients when compared to EP
- EP is valid at each point while NEP is valid within a radius of each point
- NEP better for forecasts with high spatial variability
  - e.g. mesoscale convection

Ensemble Probability

Neighborhood Ensemble Probability

Point Probability of Reflectivity Greater Than 25 dBz   VALID: 2015-02-01 18:00:00

NMEP of Reflectivity Greater Than 25 dBz within 40 km   VALID: 2015-02-01 18:00:00
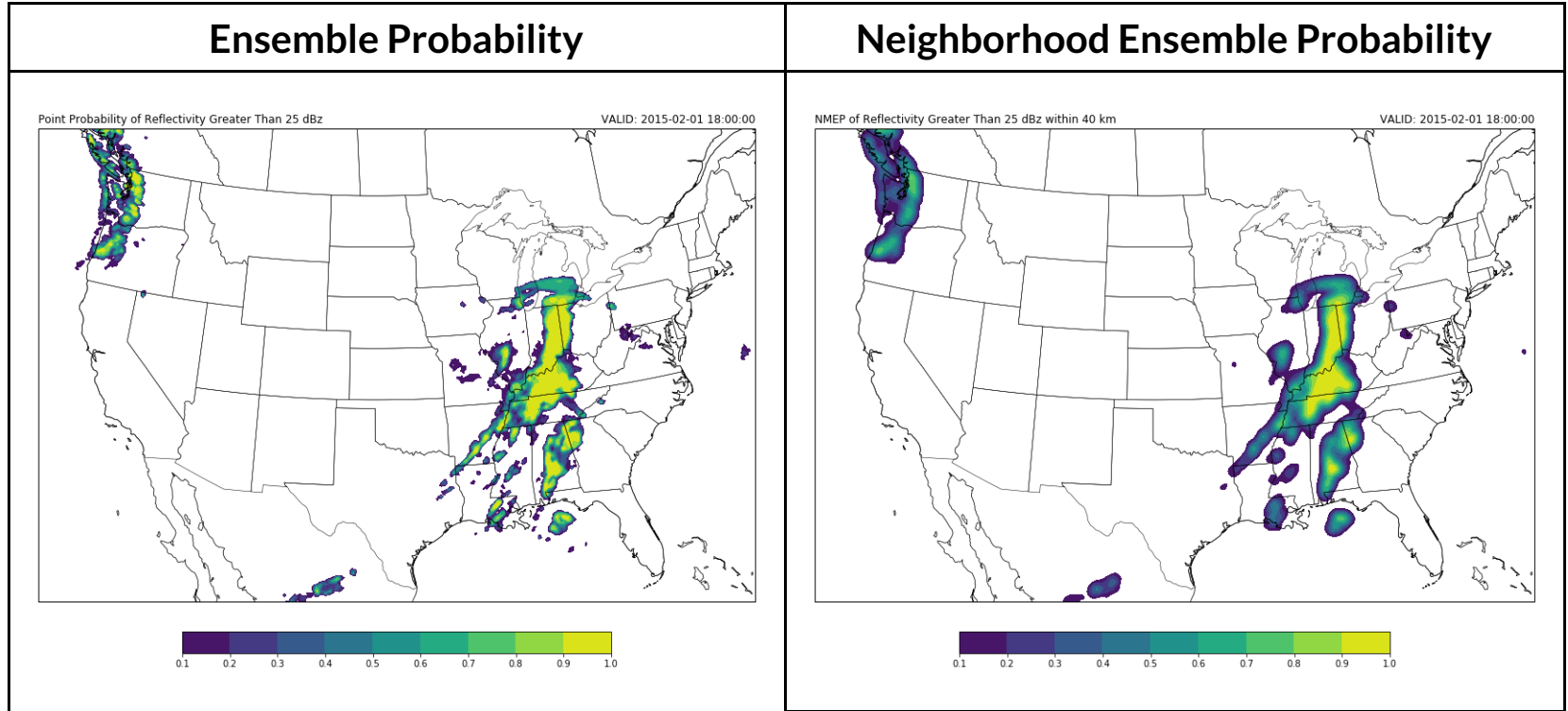
Advantages

- Probabilistic forecasts of specific events
- Quickly shows ensemble confidence
- Relatively simple to interpret

Disadvantages

- Probabilities are often too high due to low model spread
- Assumes that all member solutions are equally likely
- Limited information regarding member differences in intensity and location

Plots Summary

1. Spaghetti Plots and Postage Stamps
   - Best for viewing individual members, ensemble spread, range of solutions
2. Postage Stamp Plots
   - Best for quickly viewing spatial differences among member solutions
3. Probability Plots
   - Best for probability of event occurence and ensemble confidence

Verification

There are many verification metrics available which can be placed into three categories:

- Grid-point (Wolff et al., 2014)
- Object-based (Davis et al., 2006)
- Neighborhood approaches (Schwartz and Sobash, 2017)

One very simple approach that is often used in the Root Mean Square Error (RMSE).

Root Mean Square Error

The Root Mean Square Error is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(F_i - O_i)^2}$$

- Measure of mean magnitude of forecast error
- Same units as forecast value
- Simple to compute and interpret
- Useful for comparing two models, multiple members, etc.
- Holmes (2000)

Examples

1. Open terminal
2. Navigate to `workshop2018` folder
3. `source activate workshop` for mac users, `activate workshop` for windows users
4. Type `jupyter notebook`

References

Davis, C., B. Brown, and R. Bullock, 2006: Object-Based Verification of Precipitation Forecasts. Part I: Methodology and Application to Mesoscale Rain Areas. *Mon. Weather Rev.*, **134 (7)**, 1772-1784.

Holmes, S., 2000: RMS Error. URL statweb.stanford.edu/~susan/courses/s60/split/node60.html.

Schwartz, C. S., and R. A. Sobash, 2017: Generating Probabilistic Forecasts from Convection-Allowing Ensembles Using Neighborhood Approaches: A Review and Recommendations. *Mon. Weather Rev.*, **145 (9)**, 3397-3418.

Wolff, J. K., M. Harrold, T. Fowler, J. H. Gotway, L. Nance, and B. G. Brown, 2014: Beyond the Basics: Evaluating Model-Based Precipitation Forecasts Using Traditional, Spatial, and Object-Based Methods. *Weather Forecast.*, **29 (6)**, 1451-1472.