Ensemble forecast verification

Craig Schwartz
The National Center for Atmospheric Research

schwartz@ucar.edu

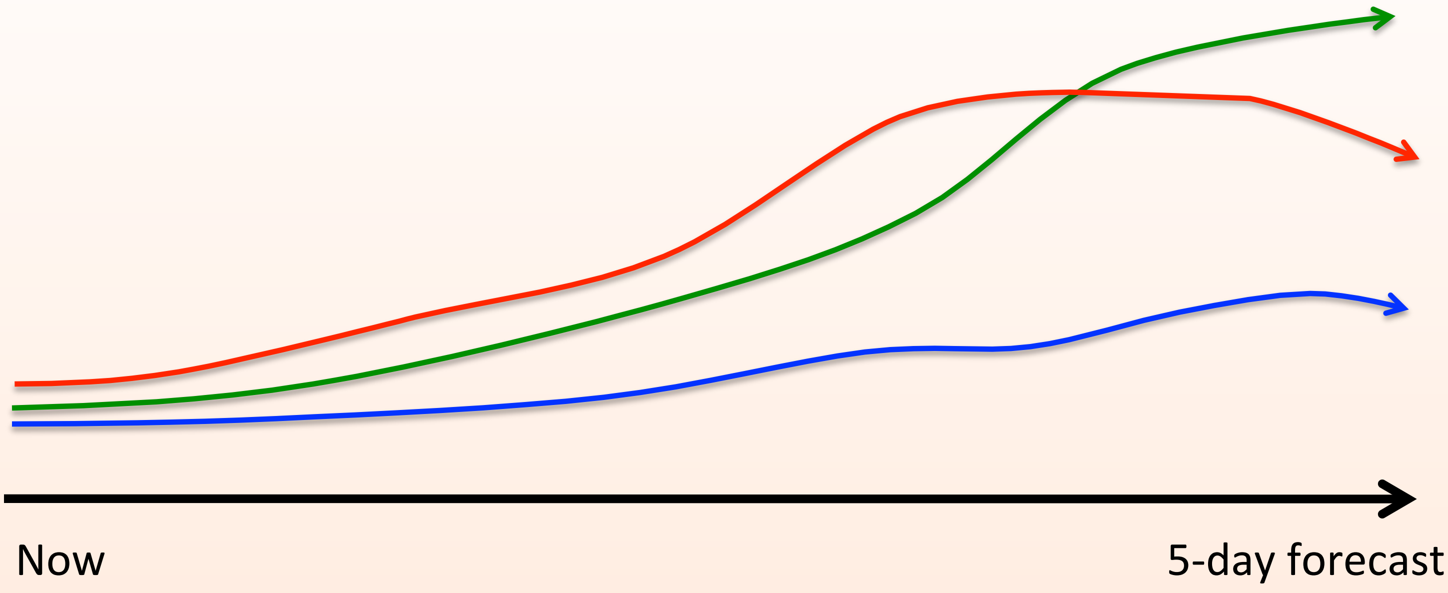
NCAR is sponsored by the National Science Foundation

Ensemble verification

- Many concepts are important for ensemble verification
- The literature is rich with ideas and metrics
  - Can only scratch the surface in a 15 minute talk
- I'll discuss some aspects of ensemble verification that can be confusing

Ensemble predictions

- Each individual member of an ensemble forecast is a deterministic forecast
- Tempting to verify each member individually with deterministic approaches



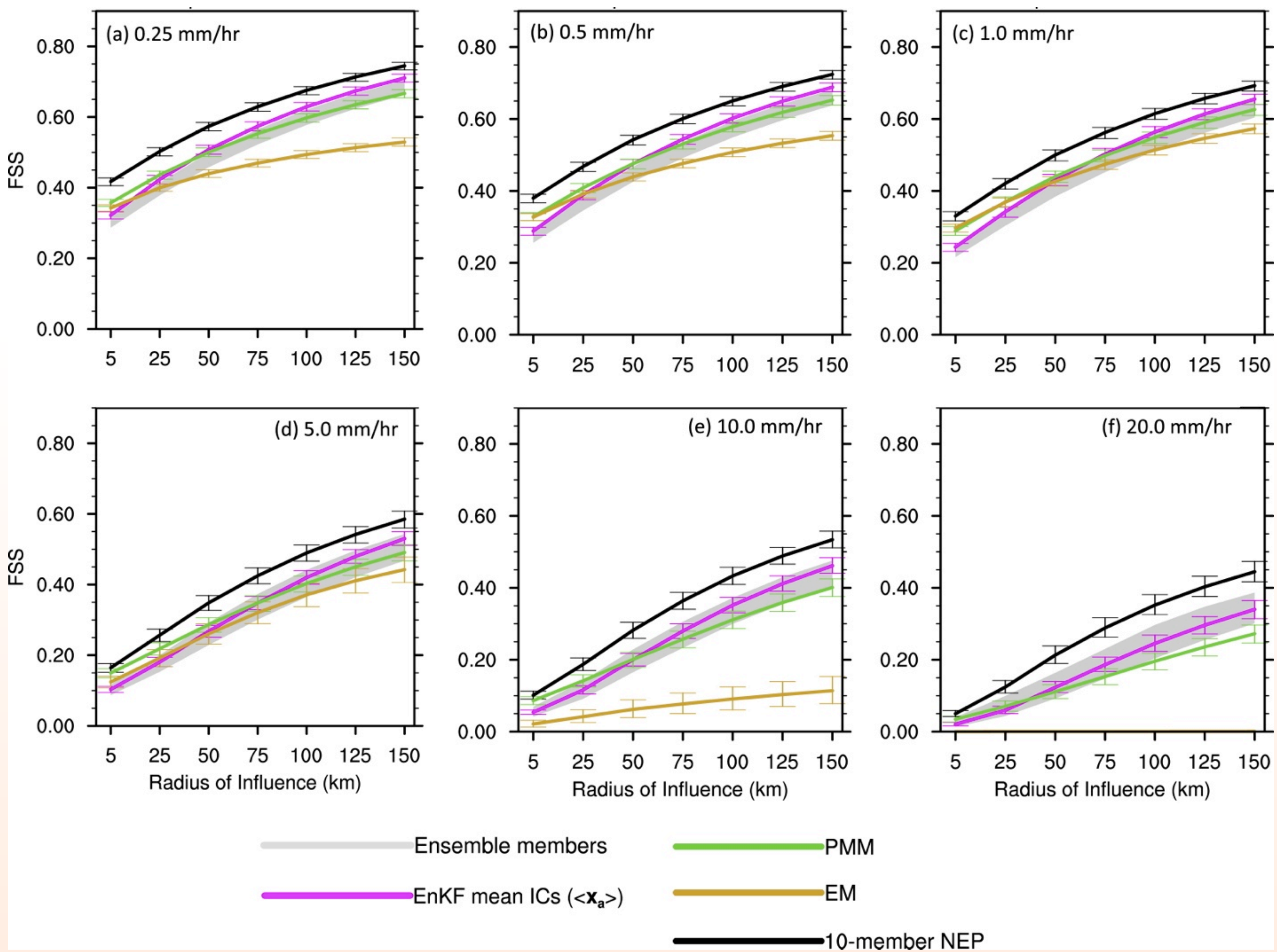
Value of ensembles

- Ensemble value comes primarily from probabilities
  - Ensembles should be verified probabilistically

- Trying to find the "best" deterministic forecast from an ensemble is not a great idea
  - Ignores probabilistic value from ensembles

- Ensemble mean and probability matched mean are sometimes useful

Probabilistic forecasts are best



From Schwartz et al. (2014); *Weather and Forecasting*

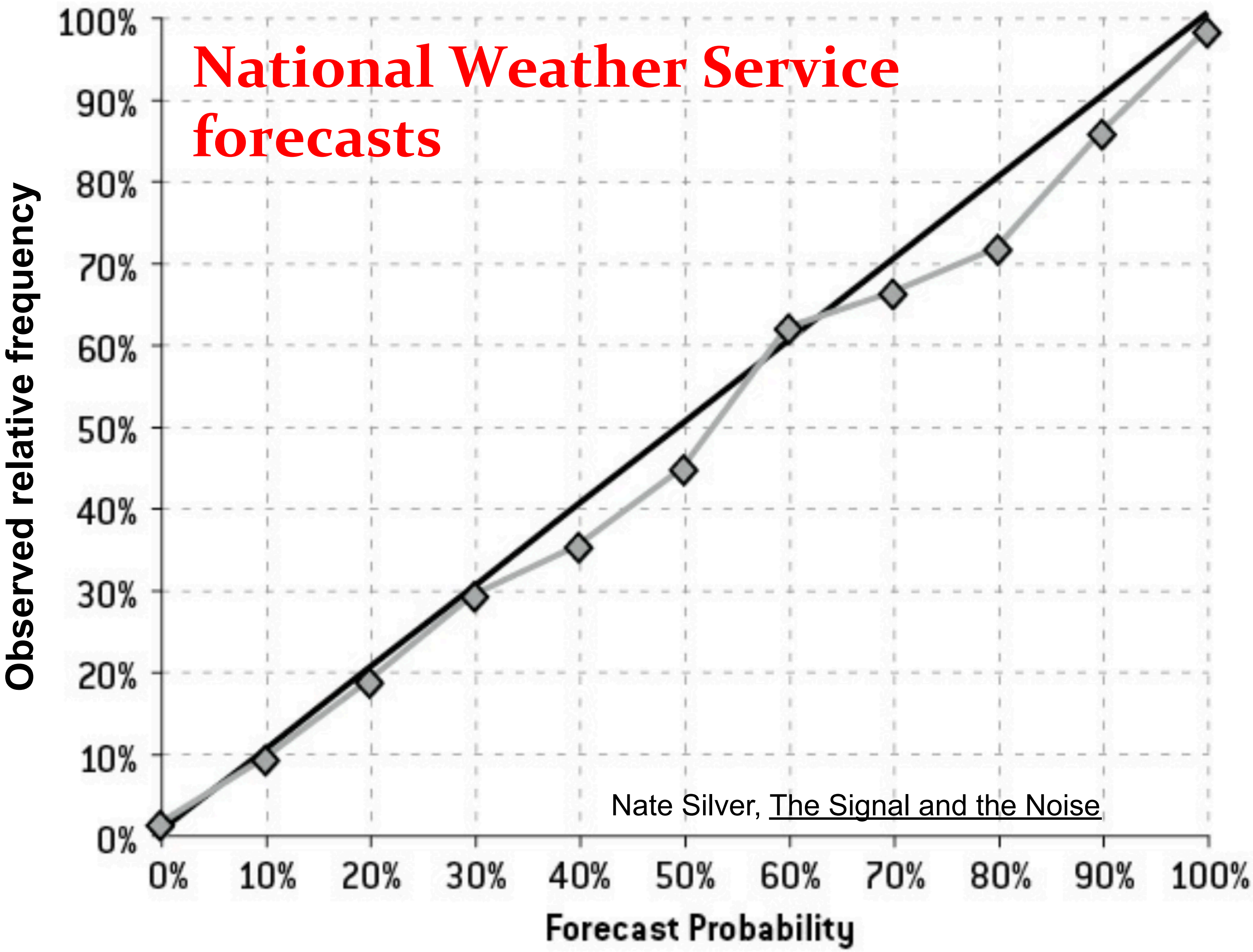
Deterministic vs. probabilistic forecasts

- Deterministic forecasts:
  - Only 0% or 100% probabilities
- Probabilistic forecasts:
  - Convey uncertainty on a *continuum* between 0% and 100%
- Probabilistic forecast quality cannot be verified with a single event

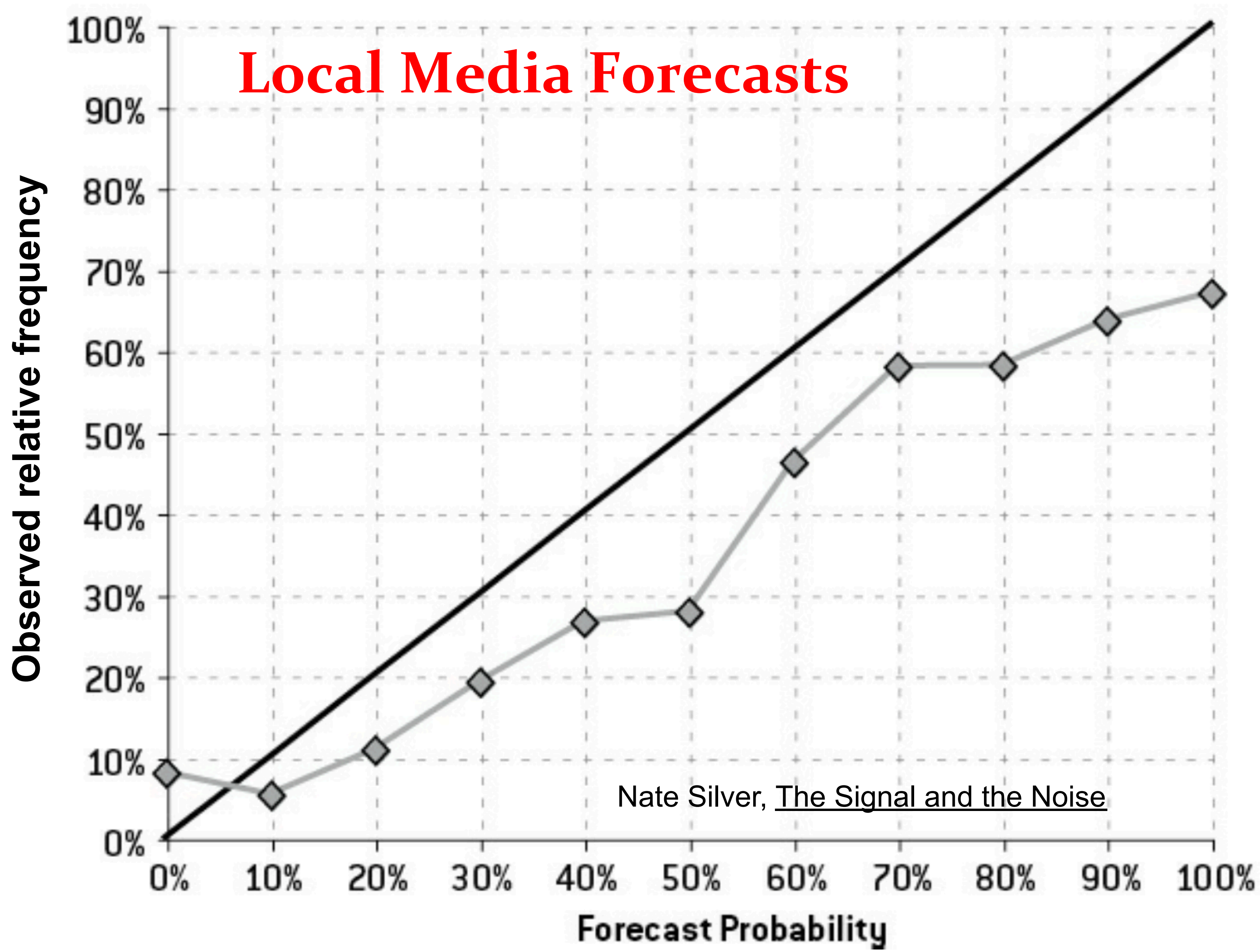
Reliability

- Reliability
  - Given a probabilistic forecast of an event, how often does the event actually occur?
- An important part of an ensemble system
  - Post-processing can improve ensembles that have poor reliability

Reliability



Reliability



Brier score (BS)

- The Brier score (BS) is commonly used to verify probabilistic forecasts:

$$BS = \frac{1}{N}\sum_{i=1}^{N}(p_i - o_i)^2$$

- p_i: Probabilistic forecast at point i
- o_i: Observations at i. $o_i = 1$ if the event occurred at i and $o_i = 0$ otherwise

Brier score decomposition (Murphy 1973)

- Brier score can be decomposed into 3 terms:

$$BS = \frac{1}{N}\sum_{k=1}^{K} n_k (p_k - \overline{o}_k)^2 - \frac{1}{N}\sum_{k=1}^{K} n_k (\overline{o}_k - \overline{o})^2 + \overline{o}(1-\overline{o})$$

reliability　　　　resolution　　　　uncertainty

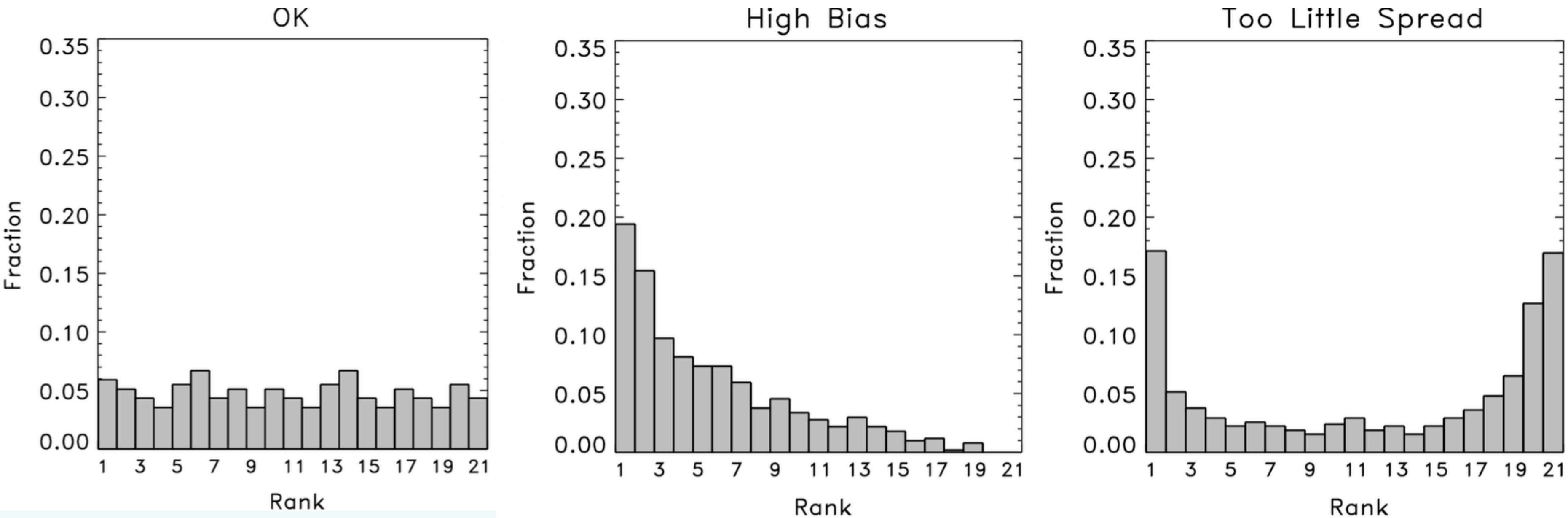
- Uncertainty term depends only on *observations*
  - Therefore, BS should not be used to compare forecasts from different samples
  - Use a Brier skill score (BSS) to circumvent this issue
  - BSS compares a BS to a reference BS

Rank histogram

- Examines ensemble spread
  - Do observations fall within range of the ensemble?
- Sort ensemble members in increasing order and determine where the observation lies with respect to the ensemble members

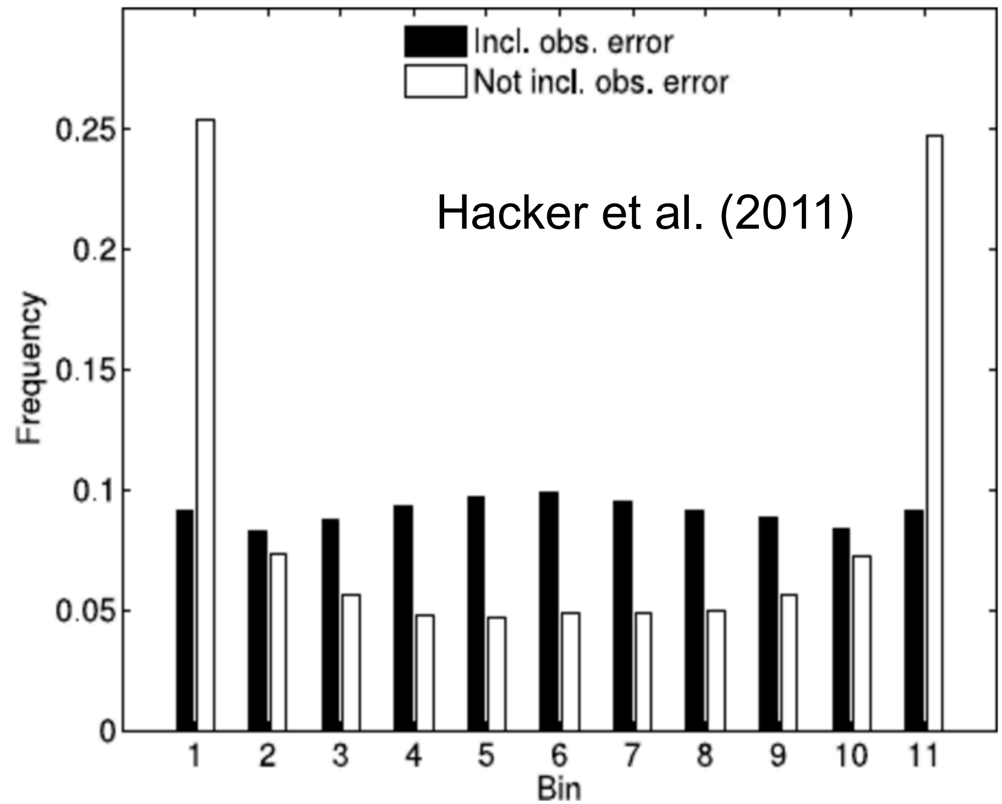
Rank histogram

- Examines ensemble spread
  - Do observations fall within range of the ensemble?



Rank histogram

- Should consider observation errors when producing rank histograms
  - Observation error = measurement error + representativeness error
  - Add noise to each ensemble member

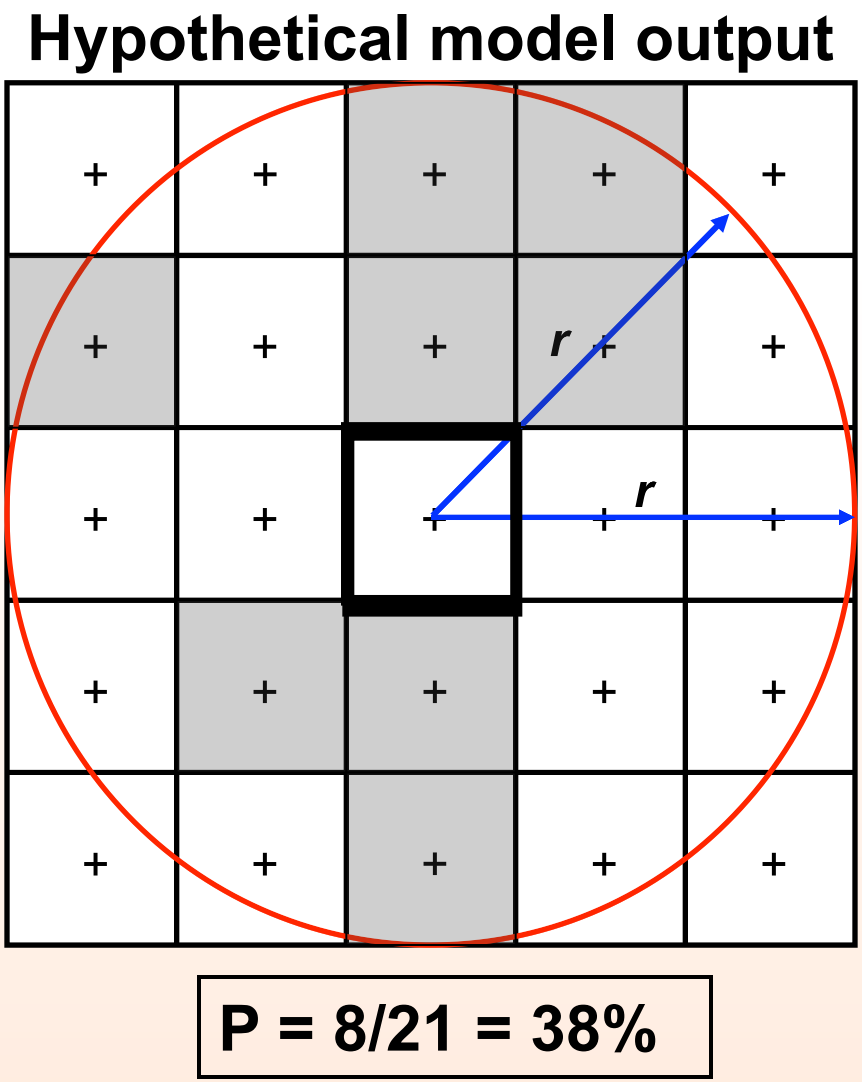


"It's a beautiful day in the neighborhood"

- High-res models are inaccurate at the grid scale
  - Verification methods requiring forecast and observed events match at the grid scale are inappropriate
- Instead, use a "neighborhood approach"
  - Specify a "radius of influence" (*r*) that defines the tolerance for error
  - Can use either square or circular geometry

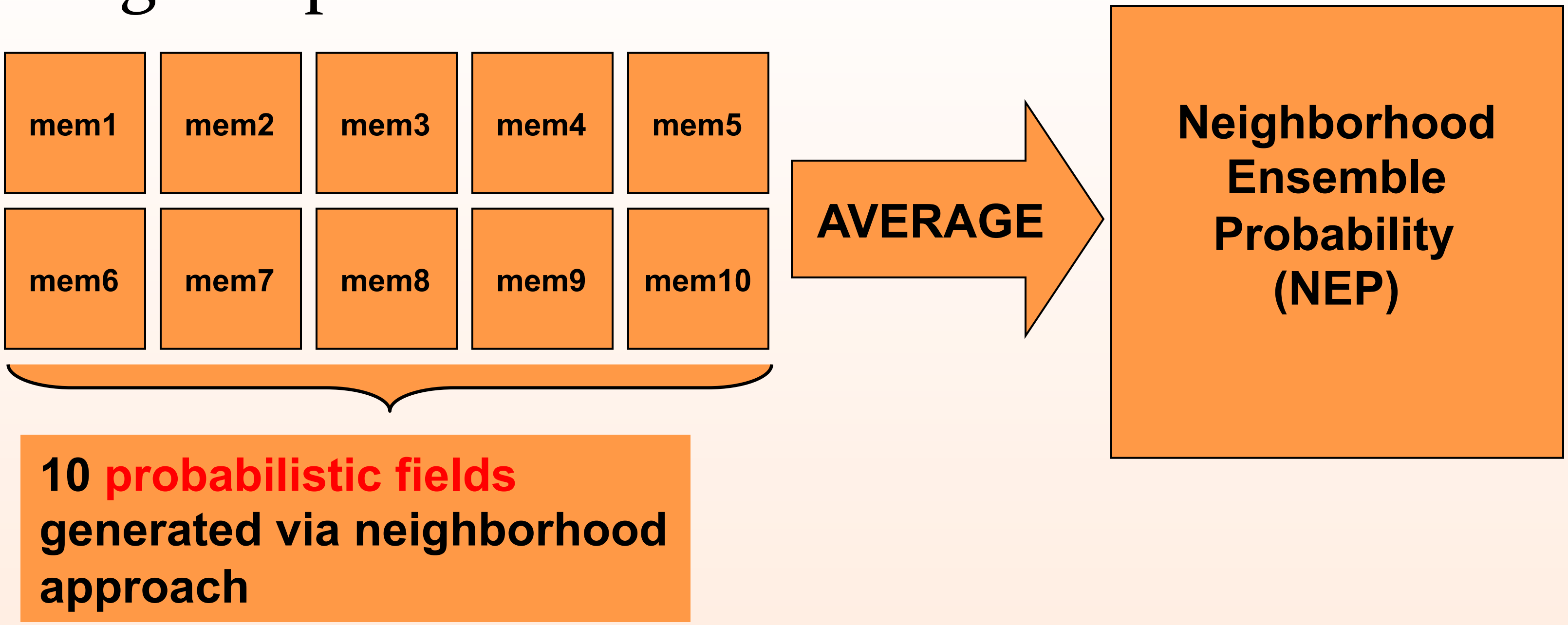
How to create probabilities from deterministic guidance?

- Pick an event
- The event has occurred in the shaded boxes
- In this schematic, r = 2.5 times the grid spacing



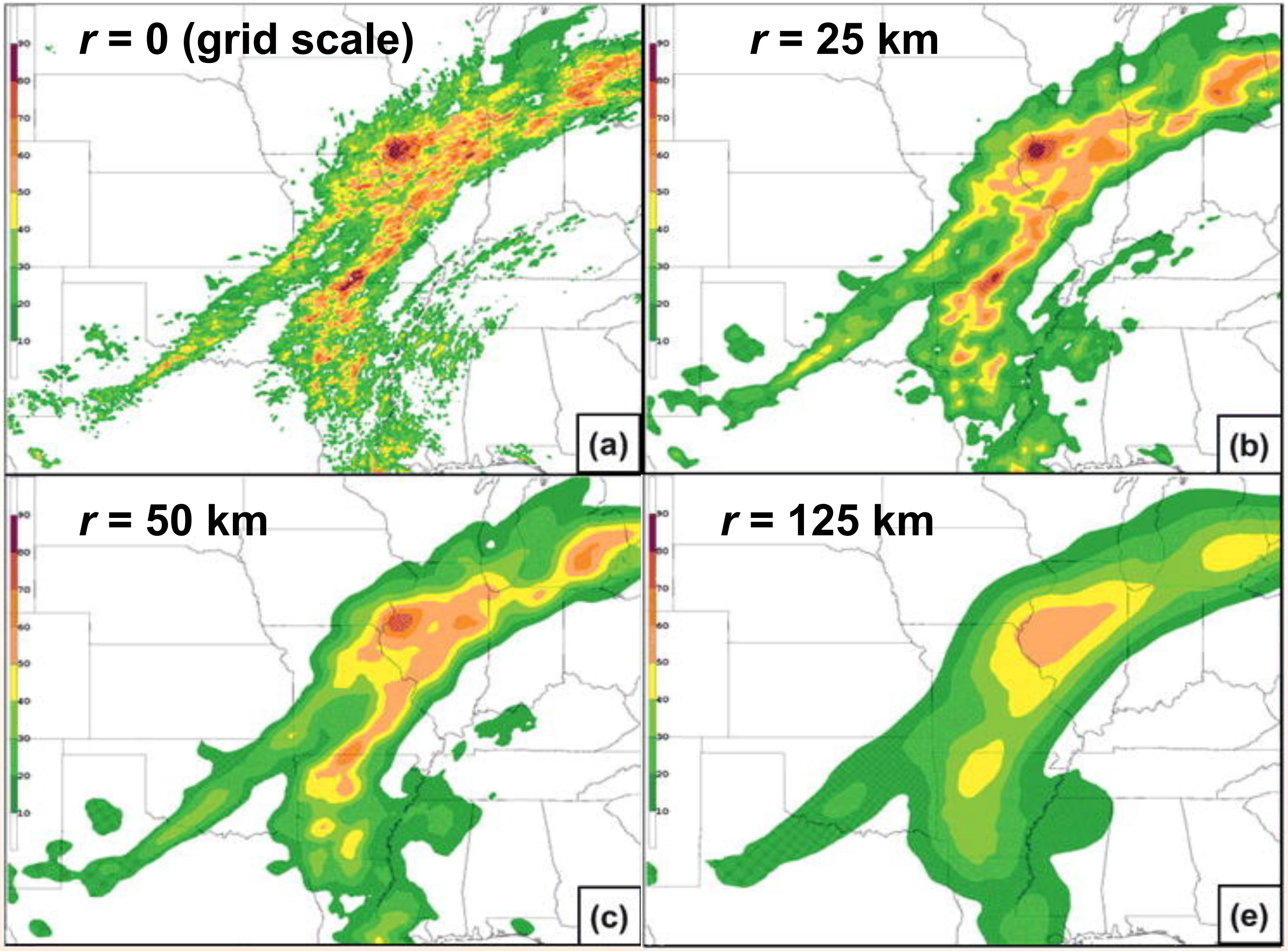
Neighborhood approach applied to ensembles option 1

- Apply neighborhood approach as described on previous slide to each member separately
  - For each member, get a value between 0 and 1
  - Average all probabilistic fields



NEP

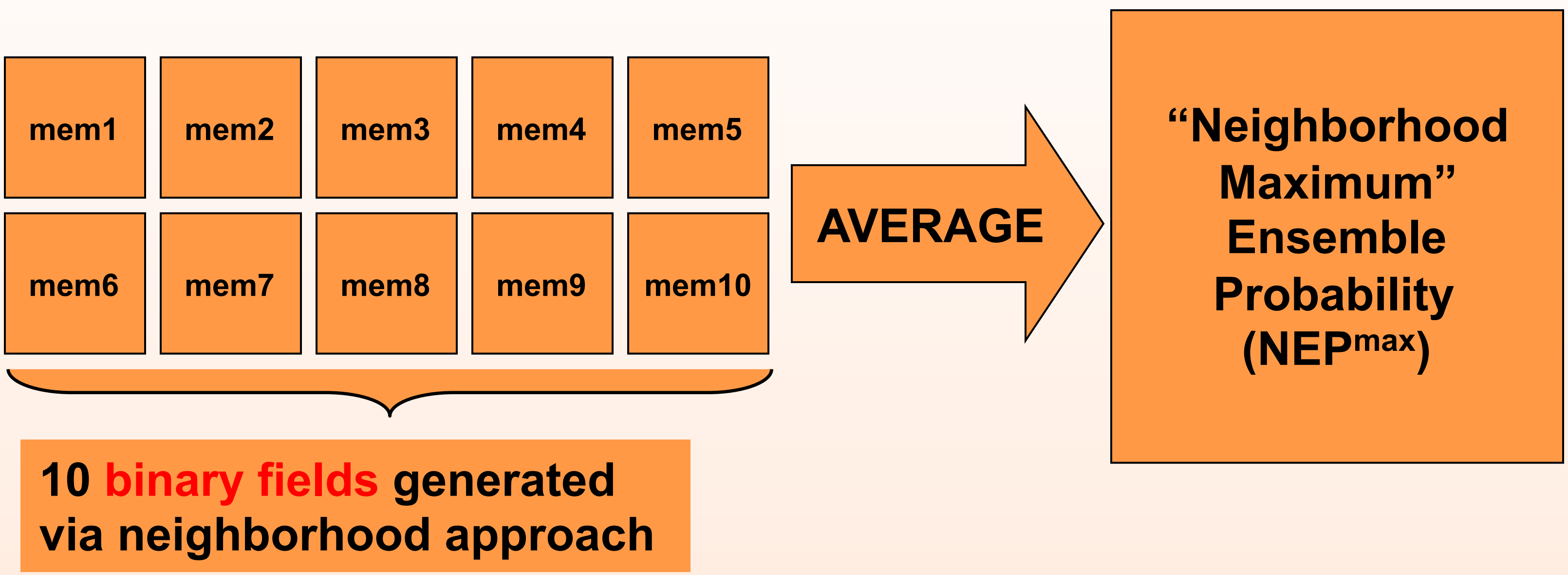
- NEPs of 1-hr precipitation ≥ 1.0 mm/hr



From Schwartz et al. (2010)

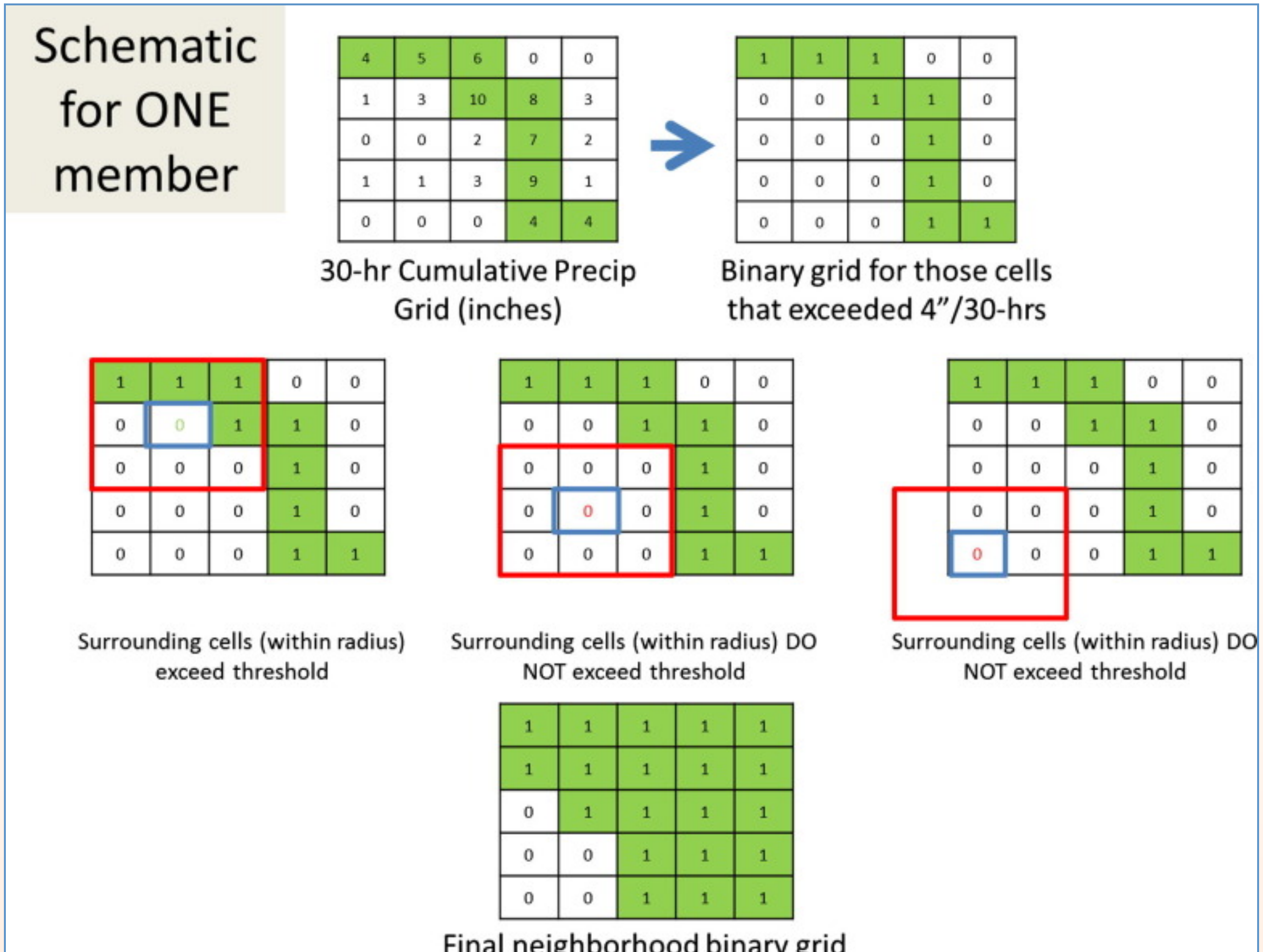
Neighborhood approach applied to ensembles option 2

- If, at a point, an event occurs anywhere within the neighborhood, give the point a value of 1, otherwise 0
  - Do this for all ensemble members individually
  - Average across the ensemble at each point to get a probability between 0 and 1



Schematic of NEP^{max}

- From Hardy et al. (2016)



- Do this for *N* members, then average the fields

Interpretations

- NEP
  - "The probability of an event occurring at grid point *i* (considering a radius of influence *r*)
  - Grid-scale probability
- NEP^{max}
  - "The probability of an event occurring within *r*-km of *i*"
  - Non-grid-scale probability

- Two methods are identical at grid scale ($r = 0$)

References

Hacker, J., and Coauthors, 2011: The U.S. Air Force Weather Agency's mesoscale ensemble: Scientific description and performance results. Tellus, 63A, 625–641, doi:10.1111/j.1600-0870.2010.00497.x.

Hardy, J., J. J. Gourley, P.-E. Kirstetter, Y. Hong, F. Kong, and Z. L. Flamig, 2016: A method for probabilistic flash flood forecasting. *Journal of Hydrology*. In press, doi:10.1016/j.jhydrol.2016.04.007.

Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Meteor.*, **12**, 595–600, doi: 10.1175/1520-0450(1973)012<0595:ANVPOT>2.0.CO;2.

Murphy, A. H., 1993: What is a good forecast? An essay on the nature of goodness in weather forecasting. *Wea. Forecasting*, 8, 281–293, doi:10.1175/1520-0434(1993)008<0281:WIAGFA>2.0.CO;2.

Schwartz, C. S., and Coauthors, 2010: Toward improved convection-allowing ensembles: Model physics sensitivities and optimizing probabilistic guidance with small ensemble membership. *Wea. Forecasting*, **25**, 263–280, doi:10.1175/2009WAF2222267.1.

Schwartz, C. S., G. S. Romine, K. R. Smith, and M. L. Weisman, 2014: Characterizing and optimizing precipitation forecasts from a convection-permitting ensemble initialized by a mesoscale ensemble Kalman filter. *Wea. Forecasting*, **29**, 1295–1318, doi:10.1175/WAF-D-13-00145.1.

Wilks, D., 1995: Statistical Methods in the Atmospheric Sciences. Academic Press, 464pp.

References

- A useful powerpoint: http://www.atmos.umd.edu/~ekalnay/syllabi/AOSC630/ensemble101.pdf

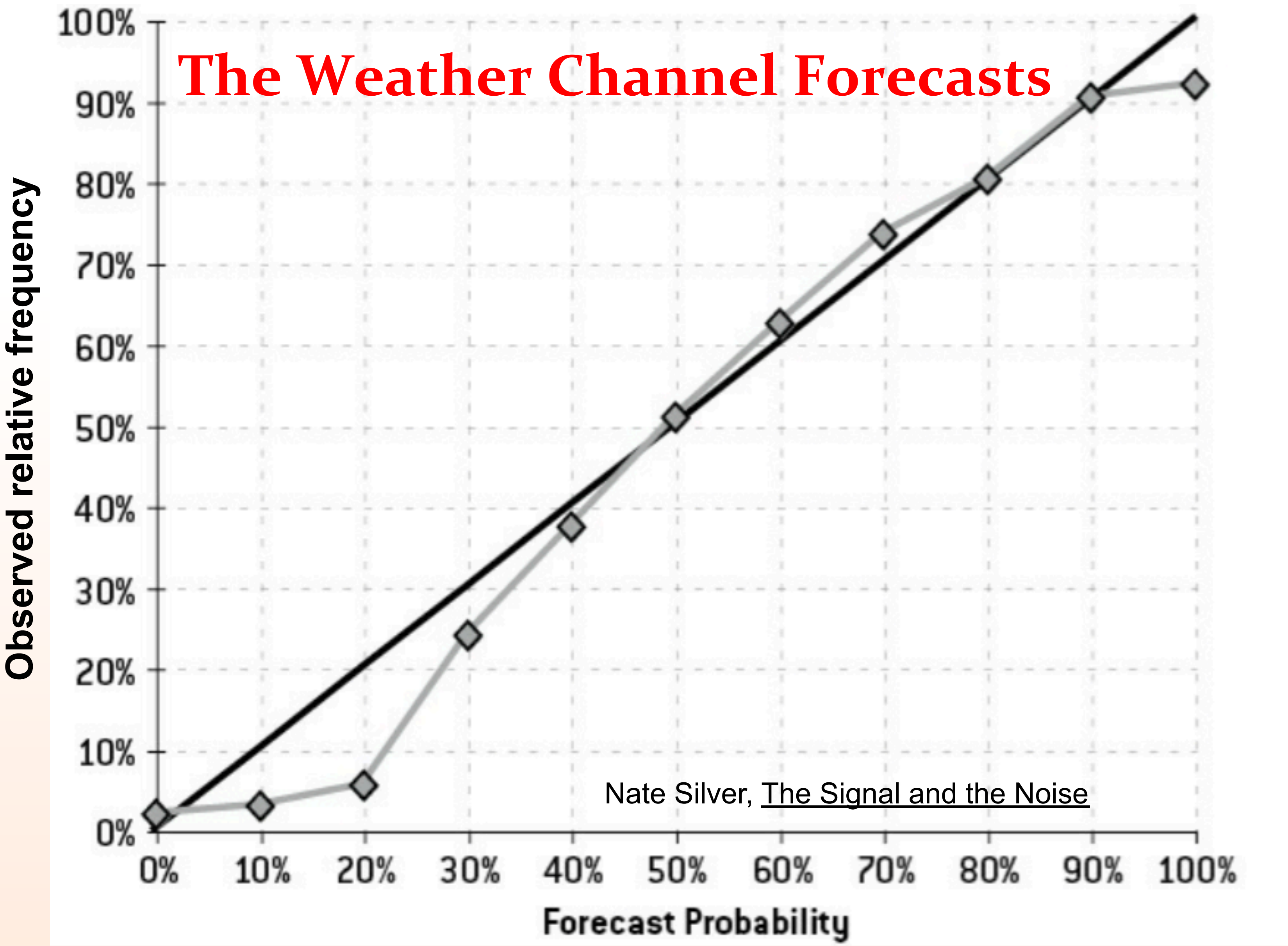
Conclusions

- Verify ensembles probabilistically
  - Do not treat an ensemble as a collection of deterministic forecasts!

- Be careful dealing with observations in the Brier score and rank histogram

- If using a neighborhood approach, explicitly state your methods and interpretations of resulting probabilistic fields

Reliability



Why ensemble forecasts are desirable

- Quantification of uncertainty
  - Naturally produces probabilities!
  - Allows forecasters to forecast their "true beliefs"
  - Allows users to make decisions based on expected value and cost-loss scenarios

- Errors of different members cancel when combining forecasts across members
  - Forecasts combining information across all members are better than single deterministic forecasts

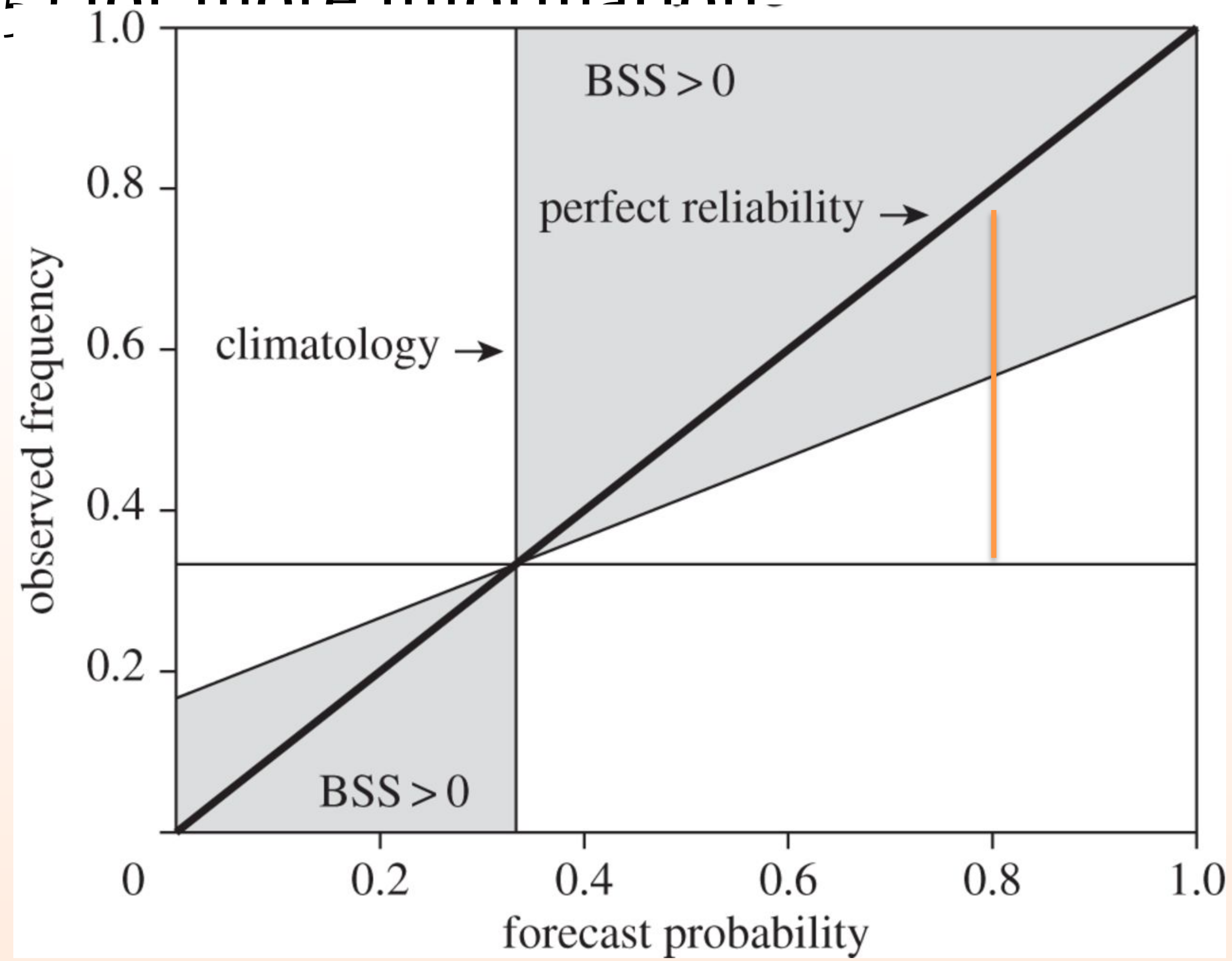
Resolution

- Resolution refers to the ability of the ensemble to distinguish between various events
  - Typically requires having a training period

- Unlike reliability, resolution cannot be easily fixed!
  - Accordingly, some people believe resolution is the most important aspect of an ensemble

Attributes diagram

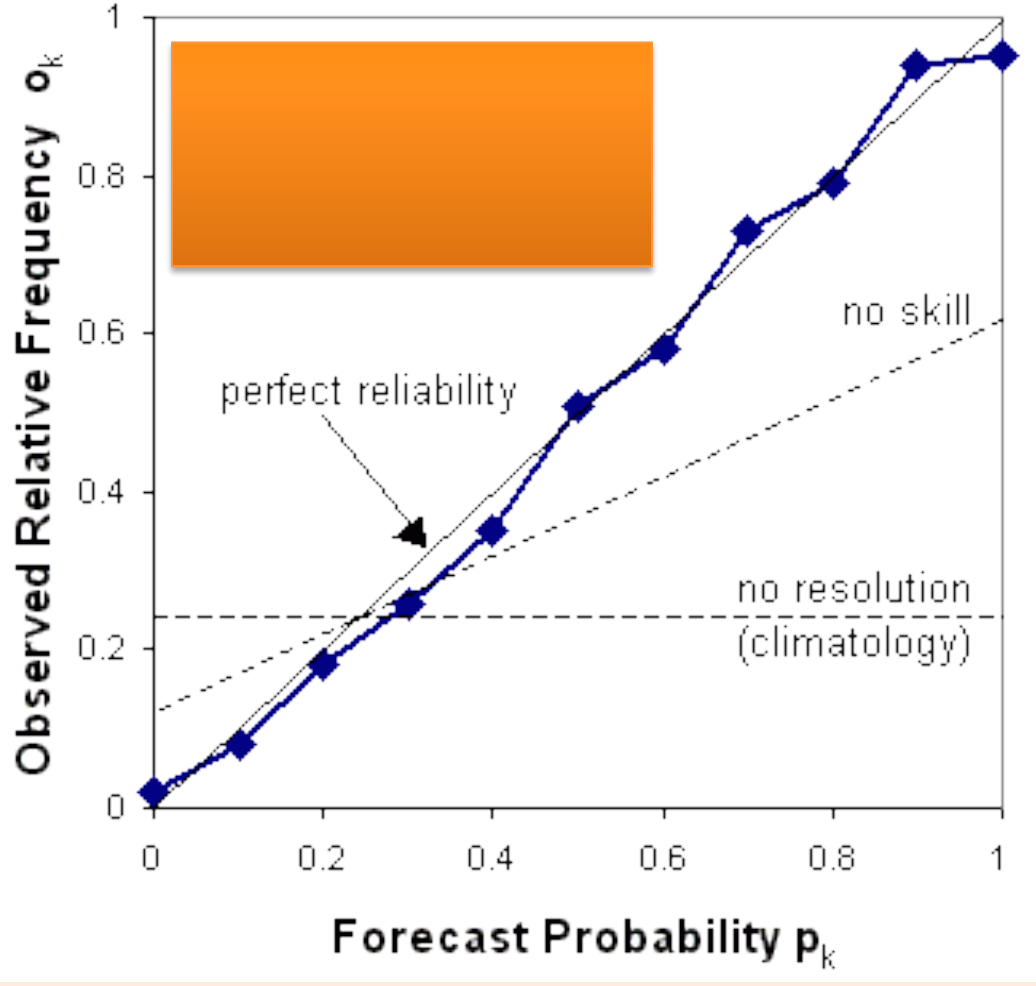
- Synthesize information about reliability, resolution, and Brier skill score
  - See Wilks (1995) for more information



*

Attributes diagram

- Synthesize information about reliability, resolution, and Brier skill score
  - See Wilks (1995) for m



*