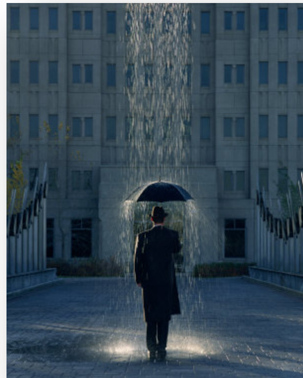
Predicting Events Using Diverse Ensemble Models

2016 SIAM Annual Meeting
Mini-Symposium on Forecasting from Big, Noisy Data: Challenges and Techniques

Amy Sliva
Charles River Analytics
asliva@cra.com

12 July 2016

Prediction is essential for decision making



Will it rain today? Do I need my umbrella?

What is the remaining useful life of an engine turbine for an aircraft? When should it be replaced?

What will be the impact of Brexit? How should I adjust my investment strategy?

…but prediction is hard, especially about the future

- Accurate prediction requires

  1. The right data
  - Free of noise (or with well-behaved noise distributions)
  - Complete coverage over all possible cases
  - Contains orthogonal features/independent variables
  - Annotated with classes or cases of the dependent variables

  2. The right models
  - Quantify confidence or uncertainty
  - Account for complex relationships
  - Computationally tractable on “big data”

- Back to the real world
  - Data is noisy, incomplete, unlabeled, statistically troublesome (e.g., autocorrelated), or unavailable
  - Countless models and analytics—which to choose and why?

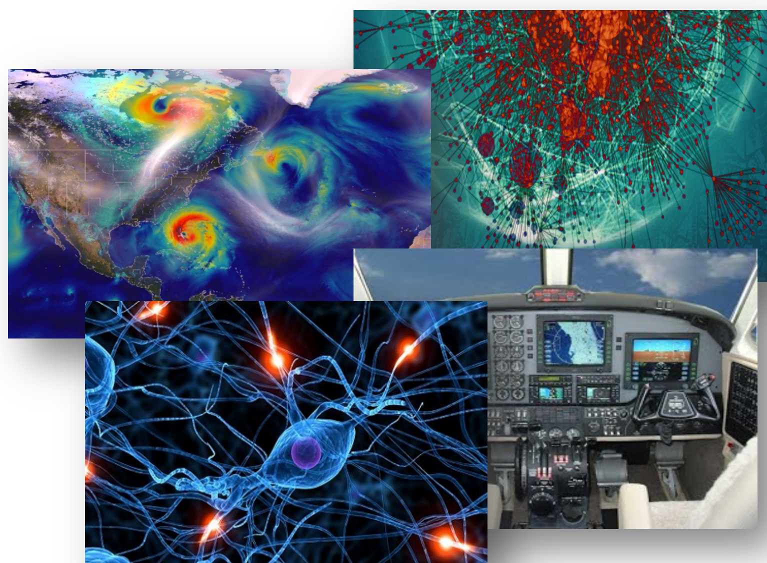
Complexity is a driver of these challenges

- Data is not just noisy due to measurement error or bad sensors

- Complexity of the underlying systems makes it hard to capture perfect data
  - Interactions between individual components produce emergent patterns in the larger system over time

- Attempts to predict behavior in complex systems contribute to both the huge volume of data and proliferation of modeling approaches

- Examples
  - Climate
  - Aircraft
  - Human brain
  - Social systems
  - …



Example: predicting behaviors in sociocultural systems

- Human behavior is a confluence of interconnected and interacting factors
  - Cultural, economic, political, social, historical, etc…
  - Real situations contain hundreds or thousands of possible actions
  - Huge variability in human behavior at any given time point

- Study of political violence—what are the likely future actions of a militant organization (e.g., ISIS)?
  - What data is available?
  - How can we use it effectively to make predictions?

Example: predicting behaviors in sociocultural systems

- "Simple" event dataset categorizes violent behavior into 41 individual actions (i.e., kidnap, suicide bomb, etc.)
  - Predicted behavior consists of any possible combination of actions
  - **2^{41} (about 10^{12}) possibilities**
- Actions in this dataset can have an intensity in the range 0-7
  - **$2^{41 \times 8} = 2^{328}$ possible behaviors**
- Does not account for location, etc.
  - Include geospatial data for only 100 locations in a region
  - **$2^{32,800} \approx 10^{9,900}$ possible behaviors!**
- Struggle to even write down a model that captures all this, let alone solve the prediction problem

Current approaches to complex systems predictions

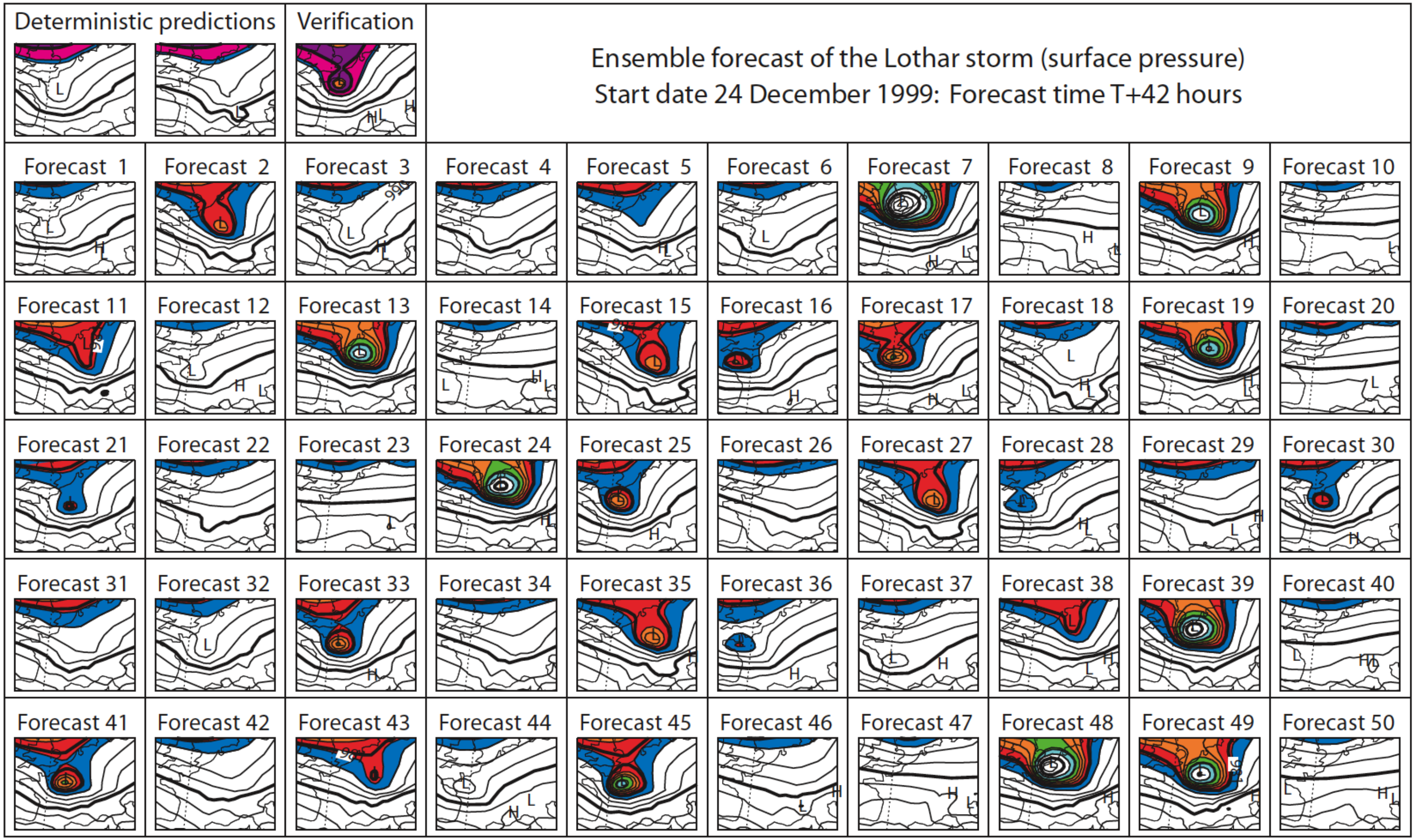
- Rely on human intuition
  - Human experts who can “see the matrix” often have good understanding, but cannot articulate a formal model
  - Just get a few of these experts to do the analysis and make a prediction
  - **Information overload!**
  - **Turns out individual experts are no better than simple lag-models…**

- One model to rule them all
  - If we can find the “best” way to represent the system and relationships, then we can make accurate predictions
  - **Given the diversity and scale of data and the complexity of the problems, there is no single best solution…**

Addressing complexity with ensemble models

- Variety of data sources and modeling approaches can be an asset rather than a challenge
- **Ensemble models** combine diverse predictive models into a single forecast
- Ensembles can provide more predictive power and accuracy than any single model alone
  - Hurricane forecasting at the National Hurricane Center
  - Political instability prediction under DARPA ICEWS combines statistical and agent-based models

Example: forecasting storms using ensemble models

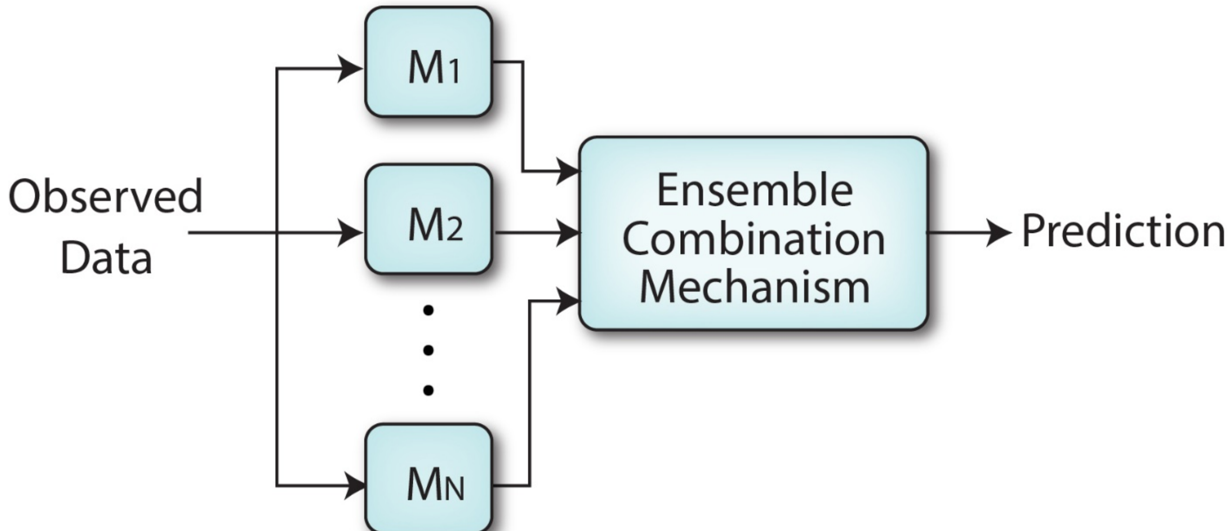


Citation: Palmer, T. N. "Predictability of weather and climate: From theory to practice. Predictability of Weather and Climate, TN Palmer and R. Hagedorn, Eds." (2006).

How do ensembles work?

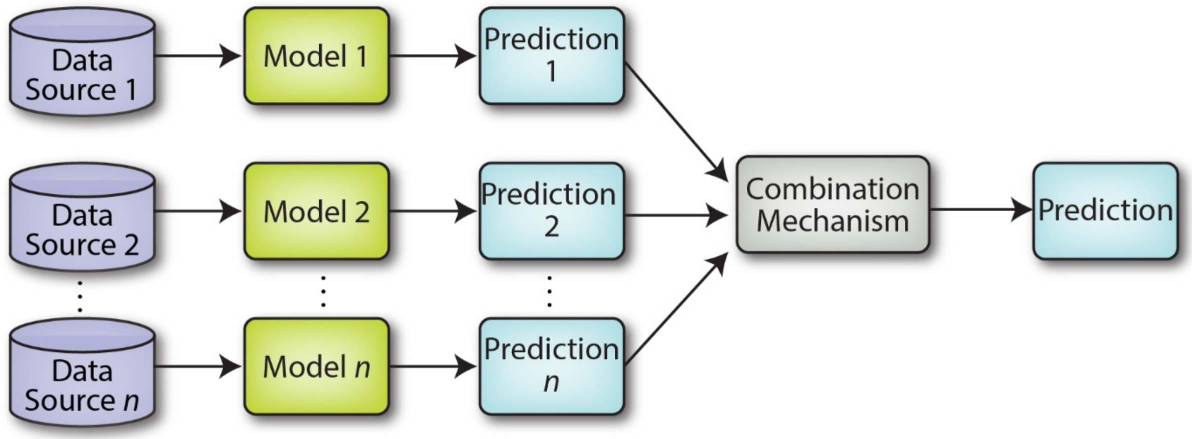
- Address uncertainty introduced by
  - Chaotic, dynamic nature of complex systems
  - Error in individual models

- Combine predictions to provide a **probabilistic** forecast



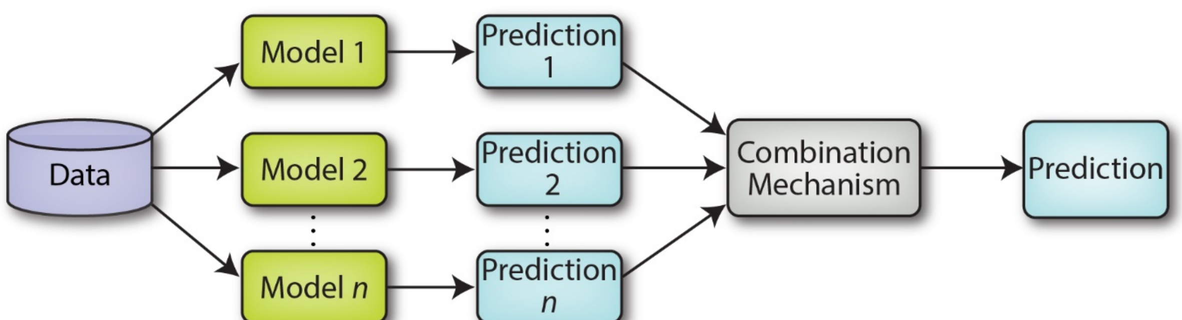
- **Challenge: what models to include, when, and with what combination mechanism?**

Types of ensembles



Data Ensembles

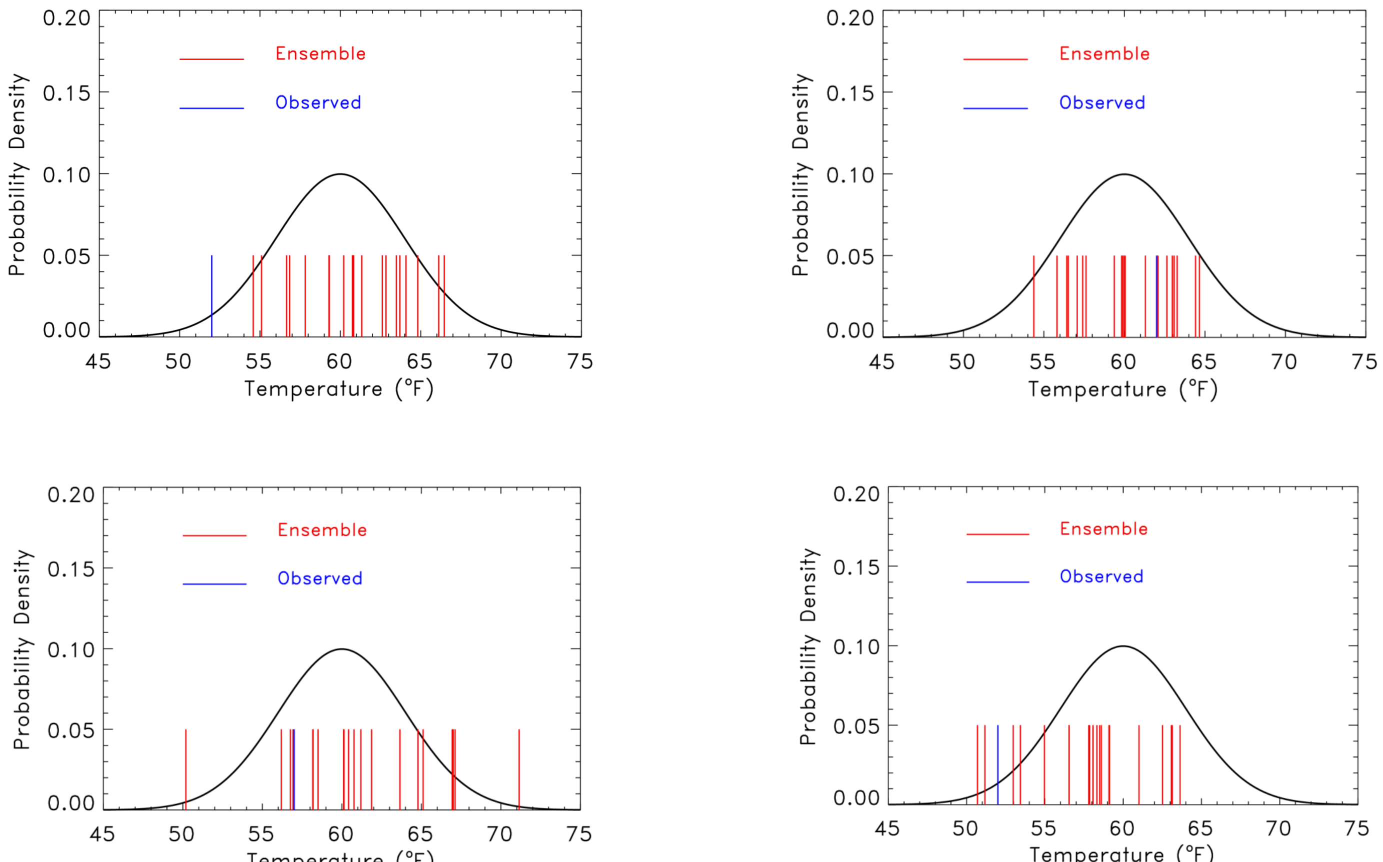
Chain Ensembles



Multi-Model Ensembles

What makes a “good” ensemble?

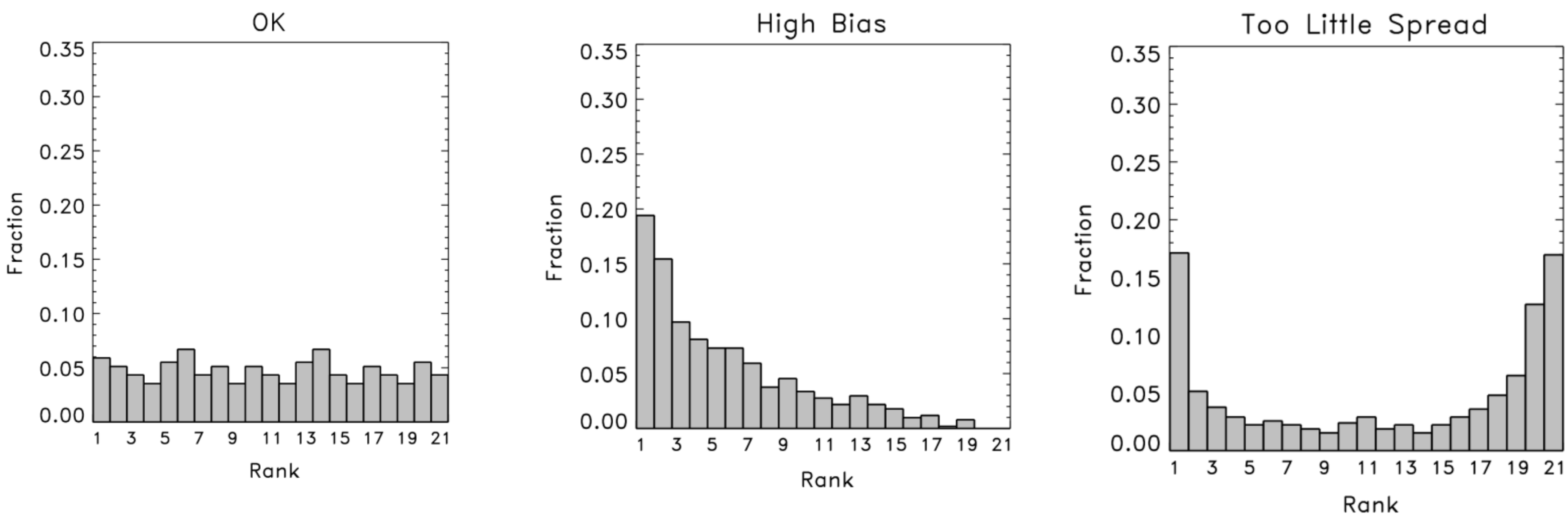
- Reliability
  - Ensemble and ground truth draw from the same distribution
  - Events occur with the same relative frequency as the prediction



Citation: Hamill, Thomas M. "Interpretation of rank histograms for verifying ensemble forecasts." *Monthly Weather Review* 129.3 (2001): 550-560.

What makes a "good" ensemble?

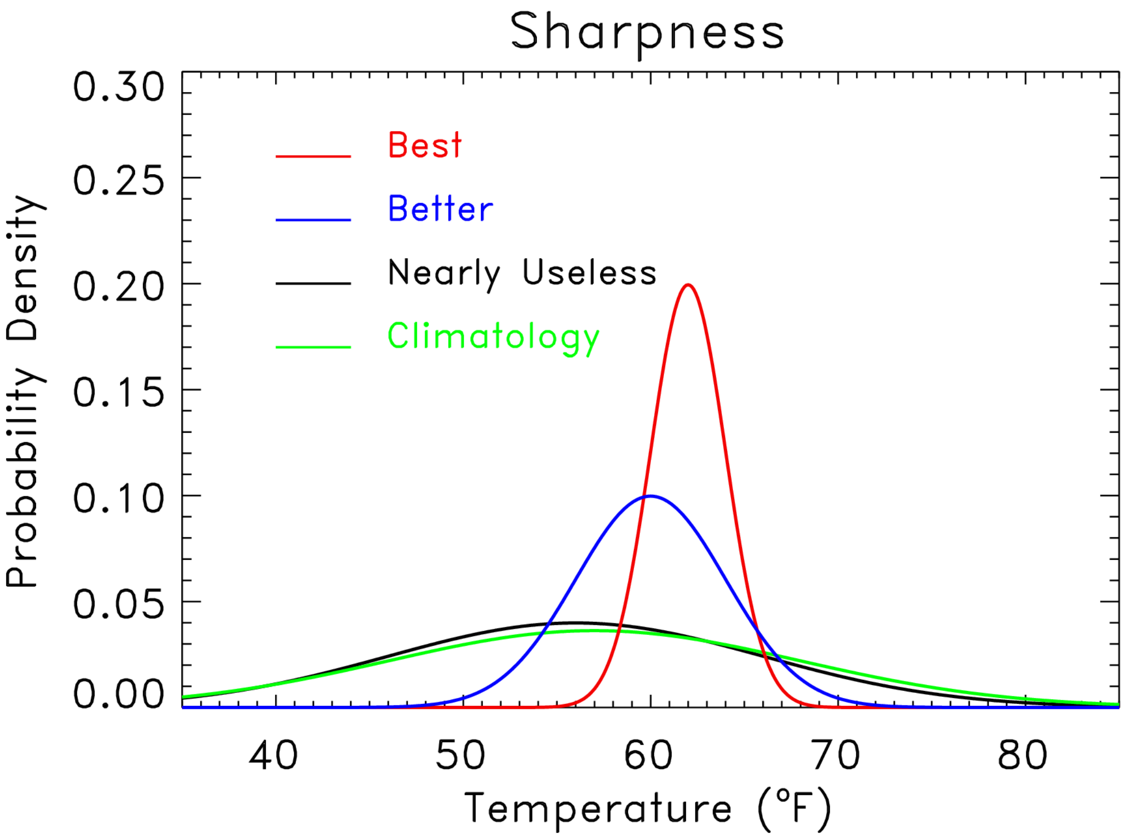
- Reliability
  - Need a lot of data and trials to actually evaluate an ensemble
  - Benefit of the big data world!



Citation: Hamill, Thomas M. "Interpretation of rank histograms for verifying ensemble forecasts." *Monthly Weather Review* 129.3 (2001): 550-560.

What makes a “good” ensemble?

- Sharpness
  - Specificity of the forecast



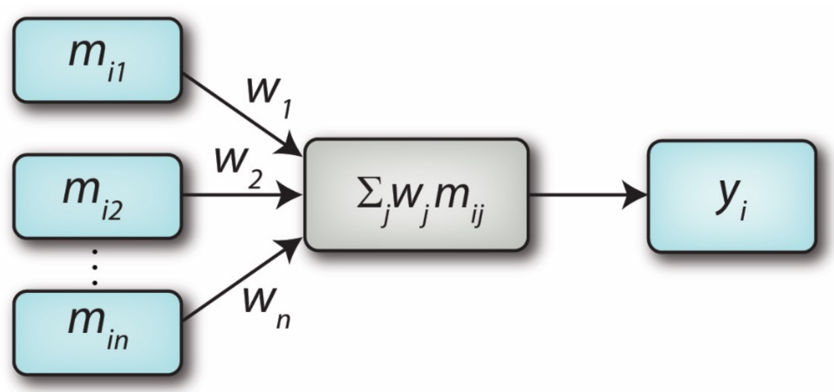
Ensemble success depends on the combination mechanism

- How we combine the models is just as important as what models to include in the ensemble

- Statistical combination mechanisms
  - Voting (linear combination)
  - Bayesian model averaging
  - Stacking

Voting-based linear combination mechanisms

$$y_i = \sum_j m_{ij} w_j \text{ where } w_j \geq 0, \sum_j w_j = 1$$



- Common Variations:
  - Ensemble Mean—all w_j are equal
  - Most probable
    $y_i = argmax_{y_i} n_{y_i} / n_t$ where n_t is the ensemble size and n_{y_i} is the number of ensembles that forecasted y_i

Ensemble Bayesian model averaging

- Specialized form of voting that uses Bayesian priors to inform the weighted average
  - Finite mixture of *K* ensemble components
  - Priors come from model performance in a training and validation period

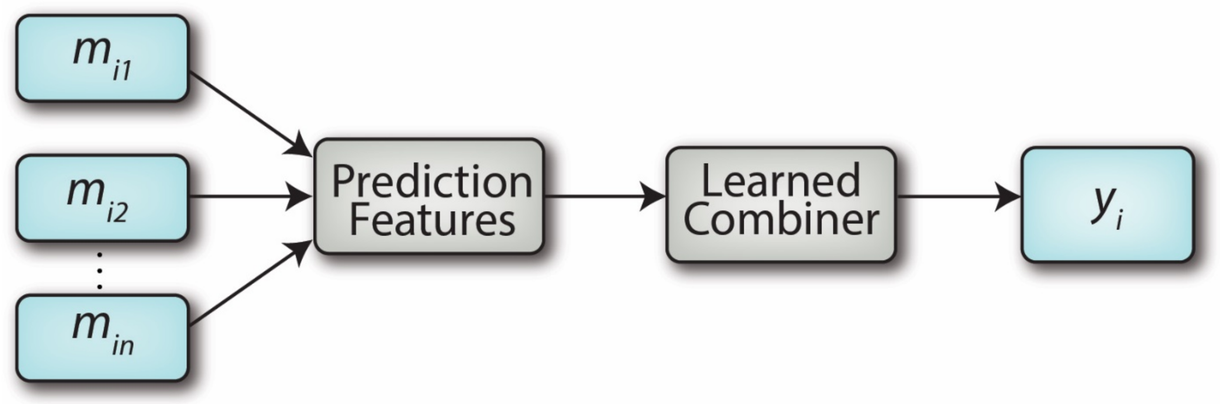
$$p(y \mid f_1^{s|t*}, \ldots, f_K^{s|t*}) = \sum_{k=1}^{K} w_k \, g_k(y \mid f_k^{s|t*})$$

Probability of event *y* given forecasts from *K* models

Weighted sum over the component probability density functions *g* for each model

Stacking to learn the best combination mechanism

- Machine learning used to find the combination approach
  - Not limited to linear combinations
  - Learns how the component models of the ensemble make errors
  - Estimate and correct for the biases of the component models



Experimental Results: stacking causal models

- Additive Noise Models (ANM) are a recent approach to determining causal/predictive relationships from observational data
  - Assume $y = f(x) + N$ where N is some noise distribution
  - Enables analysis of causal direction in cases with two variables

- CauseEffectPairs (CEP) Benchmark data set
    - 99 cause-effect pairs from a variety of domains
    - Altitude vs. Temperature
    - Weight vs. Bone Density
    - Age vs. Blood Pressure

- Best ANMs have an accuracy score in high 60s, worst is in 30s, average mid 50s

Experimental Results: stacking causal models

- **Methodology**
  1. Run a variety of ANMs over the CEP data—collect the results
  2. Create a feature vector for each pair of data points
     - Class = ground truth (we know the true cause and effect)
     - Features = Each model's outcome for that pair
  3. Build a classifier using these features to learn the best way to combine ANMs

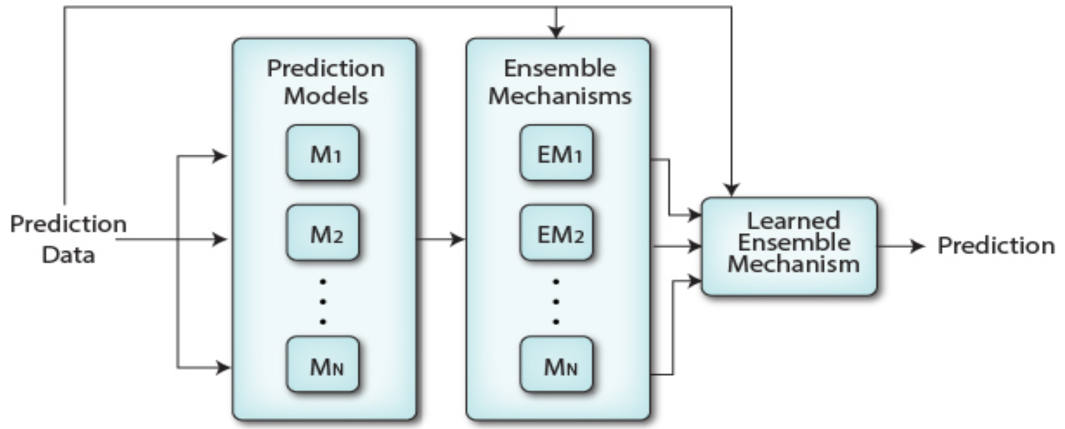
- **Results**
  - Ensemble improved drastically over any individual model (~90% accuracy)

Method	CA	Sens	Spec	AUC	F1	Prec	Recall
SVM	.7694	.9000	.1000	.5569	.8672	.8367	.9000
Naïve Bayes	.8306	.9561	.1875	.5718	.9043	.8578	.9561
Class Tree	.8367	1.000	.000	.5000	.9111	.8367	1.000

Ongoing Research: Meta-Stacking

- Most ensemble combination mechanisms focus on the "how" and not the "when"
  - Some models are just better in certain situations
  - Not captured by standard statistical approaches

- Meta-stacking—uses context features to learn the most effective model combination mechanism in a particular instance
  - Two-level stacks using support vector machines, neural networks, classifier trees, logistic regression, k-nearest neighbor, rules, random forests, and genetic programs



Ongoing Research: Semantic Ensemble Combinations

- Statistical combinations assume the component models are in competition doing roughly the same thing
- What about models that are functionally distinct?

- Multi-Formalism Modeling framework describes meta-features of models
  - Modeling formalisms
  - Time
  - Scope
  - Entities being modeled
- Guides **semantic combination** of complementary models to generate more complete predictions

Citation: Levis, Alexander H., and Ahmed Abu Jbara. "Multi-Modeling, Meta-Modeling, and Workflow Languages." *Theory and Application of Multi-Formalism Modeling* (2013)

Conclusions

- One of the biggest challenges in predictive analytics is complexity of the problem space
  - Introduces uncertainty
  - Magnified by big, noisy data

- No single predictive model that can provide the “best” solution
- Ensembles to the rescue!

- Challenges of what models to combine and when
  - Popular statistical combination mechanisms cannot fully address the problem

- Ongoing and future work
  - Meta-stacking of ensembles to consider context
  - Enhancing combinations with semantics

Charles River Analytics Points of Contact

Amy Sliva
Senior Scientist

617.491.3474 Ext. 751
asliva@cra.com

charles river analytics

Charles River Analytics Inc.
625 Mount Auburn St.
Cambridge, MA 02138
p: 617.491.3474
f: 617.868.0780
www.cra.com

Questions?