# Probabilistic Forecasting

Alberto Ruiz, Stephen Loew
Mentor: George Corliss
GasDay Lab

## Introduction

Probabilistic forecasting is increasingly becoming a useful tool for forecasting when taking account of uncertainty is of great concern. Much work has been done on probabilistic forecasting on various fields [2]. Therefore it is of paramount importance to have the proper tools to distinguish between good and bad probabilistic forecasts. Furthermore, much confusion can arise when working with probabilistic forecasts communicating the results. Hence the need for a clear unambiguous definition of what a probabilistic forecast is. In addition, it is helpful to compare different scoring metrics for probabilistic forecasts.

## Properties of Probabilistic Forecasts

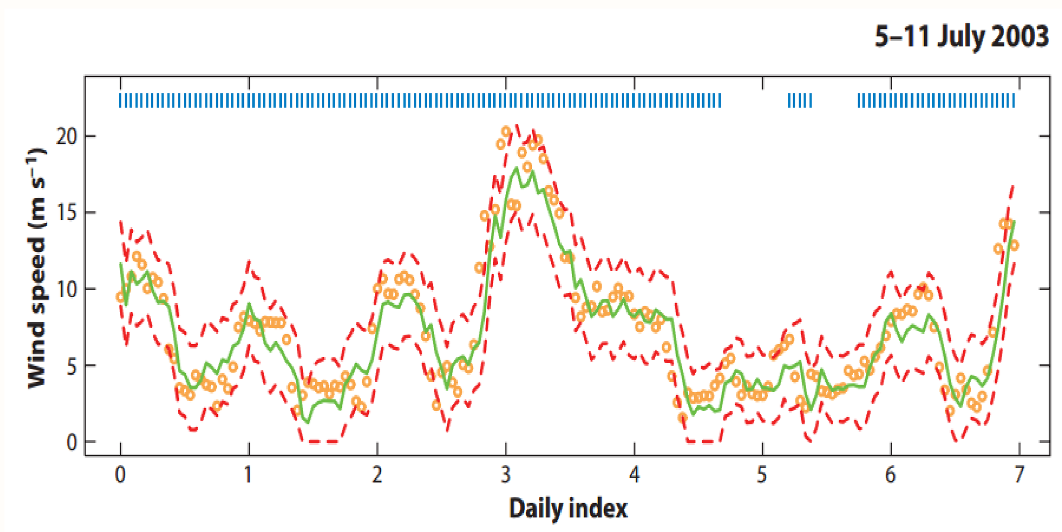


Figure 1: Two-hour-ahead regime-switching space-time forecasts of hourly average wind speed at the Stateline wind energy center for the 7-day period beginning July 5, 2003. The predictive mean is shown in green, the 90% central prediction interval in red, and the realized wind speed in orange. The blue marks at the top indicate forecasts in the prevalent westerly regime.

The restriction to density forecasts can be impractical, and we now discuss proper scoring rules that are specified directly in terms of the *predictive cumulative distribution* (CDF) function. There are various properties a predictive CDF possesses. Among the most important is *sharpness* of the predictive CDF subject to *calibration*.

### *CDF Properties*

One way to assess calibration is with the probability integral transform (PIT) of the CDF.

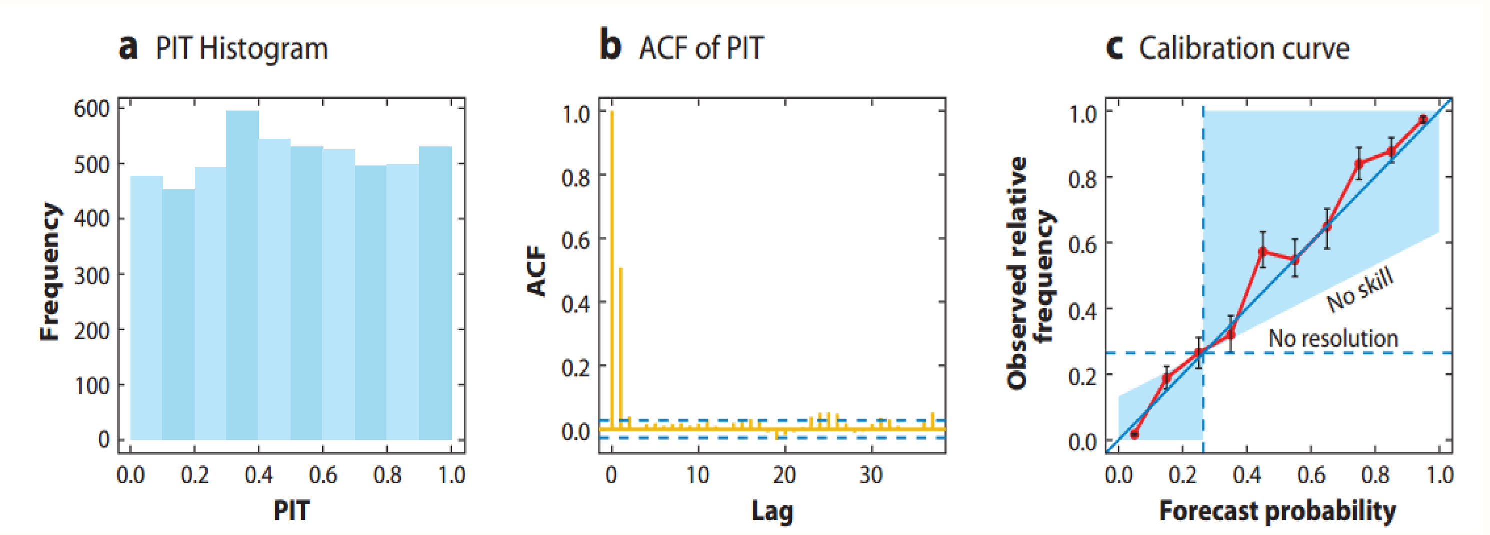


Figure 2: (a) Probability integral transform (PIT) histogram, (b) sample autocorrelation function (ACF) for the PIT values, and (c) calibration curve for exceedance of 10 m/s with bootstrap confidence intervals for two-hour-ahead regime-switching space-time forecasts of hourly average wind speed at the Stateline wind energy center.

## Scoring Metrics:

### *Continuous Ranked Probability Score*

The continuous ranked probability score (CRPS) is defined as:

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}\{y \le x\})^2 \, dx$$

where $Y$ and $Y'$ are independent random variables with CDF $F$ and finite first moment (Gneiting & Raftery 2007 , Matheson & Winkler 1976).

### *Dawid–Sebastiani Score*

The CRPS has many attractive properties, but it can be hard to compute for complex forecast distributions. A viable alternative that depends on the probabilistic forecast, $F$, through only its first two central moments, $\mu_F$ and $\sigma_F^2$, is given by the proper Dawid–Sebastiani score (DSS) (Dawid & Sebastiani 1999),

$$\mathrm{DSS}(F, y) = \frac{(y - \mu_F)^2}{\sigma_F^2} + 2 \log \sigma_F.c$$

### *Pinball Score*

For a quantile forecast $q_a$ at the $\frac{a}{100}$ target quantile and having $y$ as the realized value the *Pinball score* $L(q_a, y)$ is defined as:

$$L(q_a, y) = \begin{cases} (1 - \frac{a}{100})(q_a - y) & y < q_a \\ (\frac{a}{100})(y - q_a) & y \ge q_a \end{cases}$$

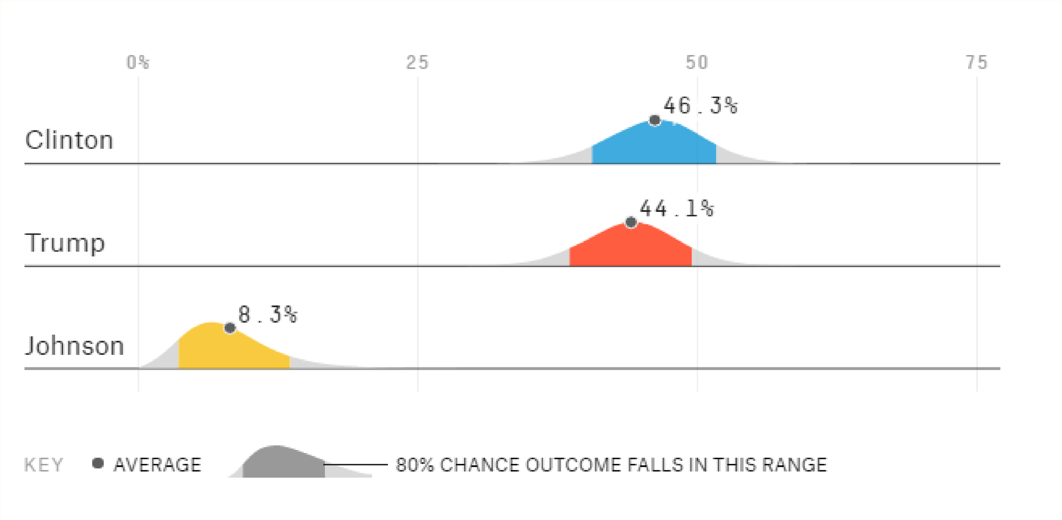


Figure 3: This model produces a distribution of outcomes for the national popular vote. The curves will get narrower as the election gets closer and the forecasts become more confident. [1]

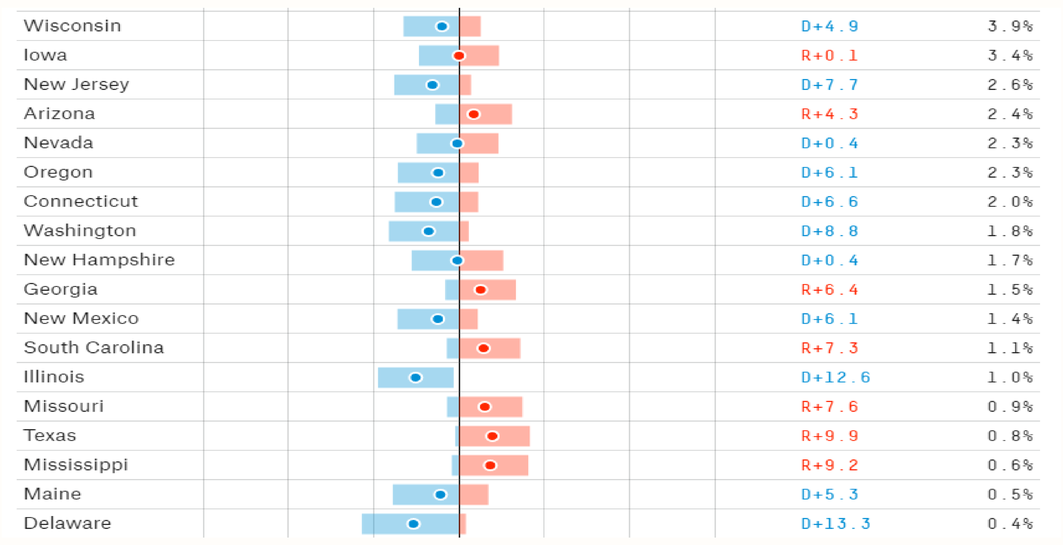


Figure 4: These win probabilities come from simulating the election 20,000 times, which produces a distribution of possible outcomes for each state. The closer the dot is to the center line, the tighter the race. And the wider the bar, the less certain the model is about the outcome. [1]

## Conclusions

Probabilistic forecasting is imperative for making decisions in real life scenarios. Right now there is no consensus on what should be the best scoring metric to use in different scenarios. There is also little research done in the area of probabilistic energy forecasting [2]. Due to this we need a better understanding of the strength and weakness of this metrics and a better understanding of what a probabilistic forecast is. This way will be more prepared for making decisions taking account of uncertainty more effectively.

## Future Work

- Try to find weaknesses in the scoring metrics that could be exploited.
- Create a probabilistic forecast engine to test different scoring metrics against each other.
- Create a machine learning model that tries to improve the score of a probabilistic forecast engine.
- Further research into probabilistic forecast ensembles might be helpful [3].

## Acknowledgements

This work was made possible by the National Science Foundation REU Site grant, #ACI-1461264, 'Computation Across the Disciplines', at Marquette University. We would like to acknowledge Marquette University, MSCS, and mentor Dr. George Corliss as well as Dr. Dennis Brylow and Dr. Kim Factor for their work in organizing the REU.

### *References*

[1] 2016 election forecast | fivethirtyeight.
http://projects.fivethirtyeight.com/2016-election-forecast/.
(Accessed on 07/18/2016).

[2] Tilmann Gneiting and Matthias Katzfuss.
Probabilistic forecasting.
*Annual Review of Statistics and Its Application*, 1:125–151, 2014.

[3] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski.
Using bayesian model averaging to calibrate forecast ensembles.
*Monthly Weather Review*, 133(5):1155–1174, 2005.