Forecasting the Probability of Tropical Cyclone Formation: the reliability of NHC forecasts from the 2012 hurricane season

A. S. Jarman[1] and L. A. Smith[1,2]

[1] Centre for the Analysis of Time Series, Dept. of Statistics, London School of Economics, U.K.
[2] Pembroke College, Oxford, U.K.

NHC Short-term Tropical Cyclone Forecasts

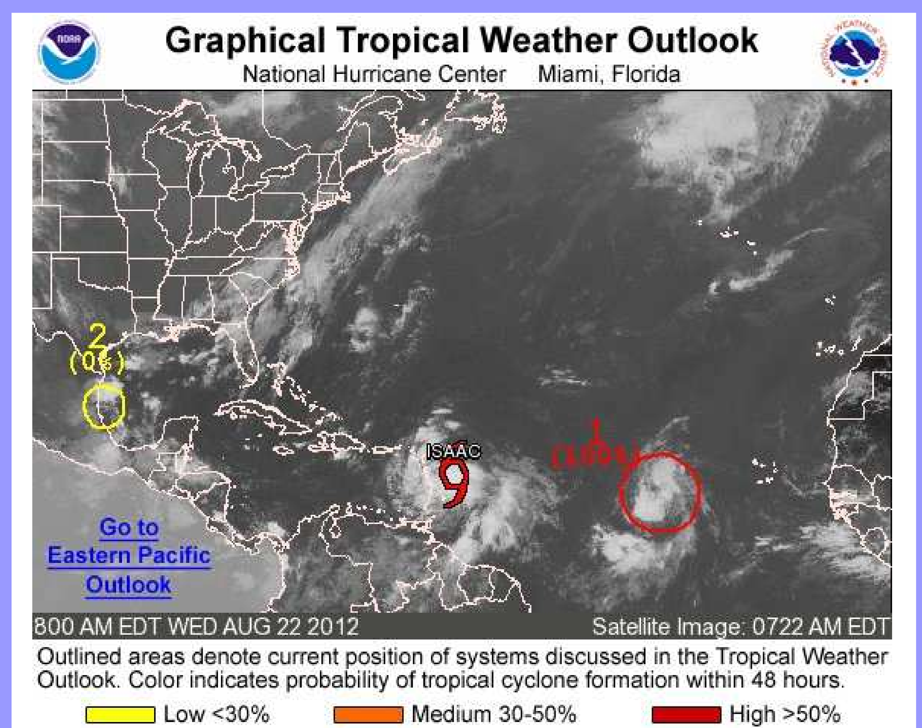

Fig 1: NHC Graphical Tropical Weather Outlook 2nd October 2012

The US National Hurricane Center (NHC) issues 6-hourly binary probability forecasts of tropical cyclone (TC) formation in a specific area of weather disturbance in the north Atlantic Basin within the next 48 hours during the hurricane season (June 1st to November 30th). These subjective probability forecasts called Tropical Weather Outlooks (TWO) can be found at *http://www.nhc.noaa.gov/pastall.shtml*. Probabilities of a TC occurrence are typically assigned in 10% increments by the NHC forecaster but those stated as 0% (and 100%) are called near zero (near 100) in the supporting text (see Fig 1).

Reliability diagrams for each season plot the observed relative frequencies of a TC event, given by

$$f_k = \frac{\sum_{i \in I_k} Y_i}{\# I_k}, \qquad (1)$$

($Y_i = 1$ for an occurrence and $Y_i = 0$ for a non-occurrence over $i = 1, ..., N$ forecast/verification pairs) against the averaged forecast probabilities,

$$r_k = \frac{\sum_{i \in I_k} X_i}{\# I_k}, \qquad (2)$$

assigned to the outcome (where $\# I_k$ is the number of elements in a collection of indices I_k for which forecast X_i falls into bin B_k, $k = 1, ..., K$).

A perfectly reliable forecast system is indicated if all the points line up near to the diagonal, $f_k = r_k$. The diagrams in their traditional format do not convey a truly representative measure of reliability, however, because they do not account for sampling error. Even perfectly reliable forecasts would be expected to deviate from the diagonal line.

Detecting (In)Consistent Points on a Reliability Diagram

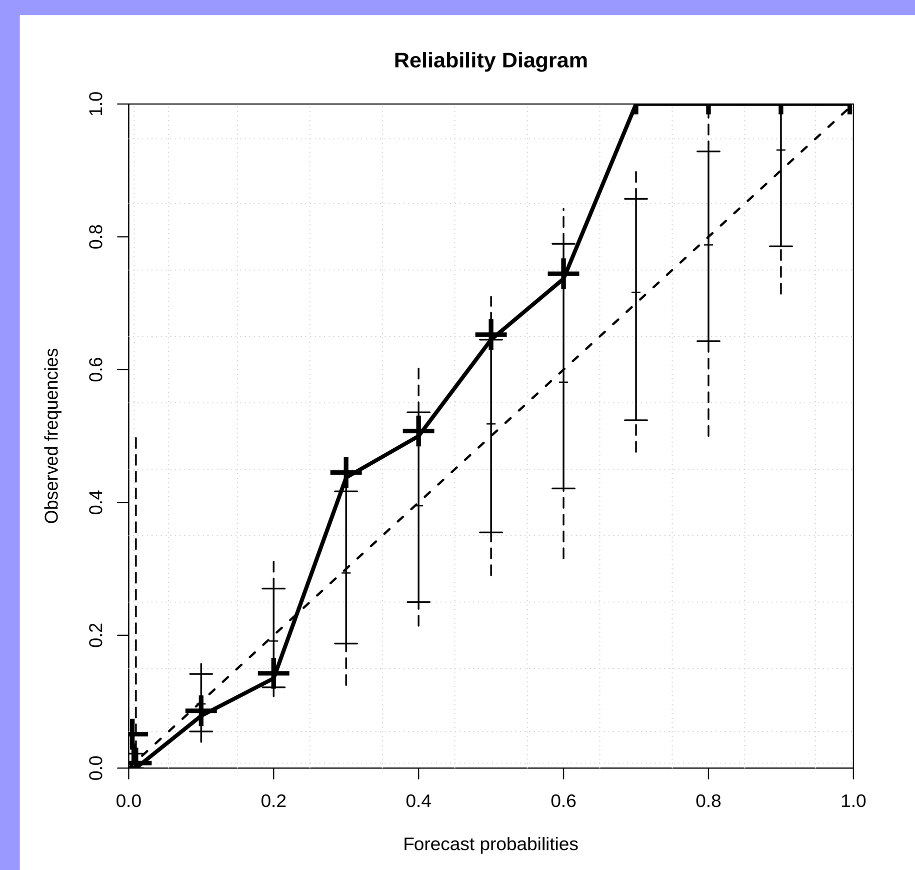

Fig 2: Reliability diagram for NHC 2012 hurricane forecasts with 5% - 95% (1% - 99% dashed line) consistency bars. The bin boundaries (light grey vertical lines) are positioned at [0.0075, 0.055, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95].

To indicate the expected sampling error 5% to 95% consistency bars [1] are included in the reliability diagram at each forecast probability bin (Fig 2).

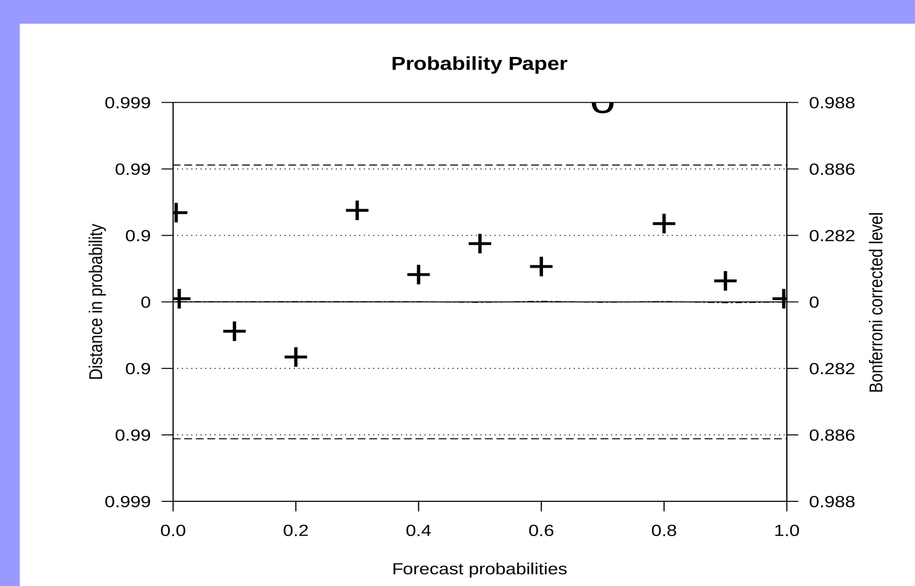

Fig 3: Reliability diagram on probability paper for NHC 2012 hurricane forecasts. The dash–dotted line denotes the exact position of the diagonal, and a semi-circle indicates an observed frequency outside the range of the y axis. The right-hand axis indicates the equivalent Bonferroni corrected levels i.e. for a reliable forecast, all of the points (12 bins) would be expected to fall within the 0.99 probability distance band with an 88.6% chance. In addition, the dashed lines indicate where the entire diagram would be expected to fall within with a 90% chance.

Forecast reliability is indicated if the observed frequencies lie within the consistency bars.

Plotting the reliability diagram on probability paper (see Fig 3) provides a useful companion to the main diagram by displaying the observed frequencies by their distance in probability from the 50% quantile of the consistency bars. This should lie close to or on the diagonal. Comparisons of forecast reliability between probability bins and an assessment of the calibration of a forecast system can be made using this diagram.

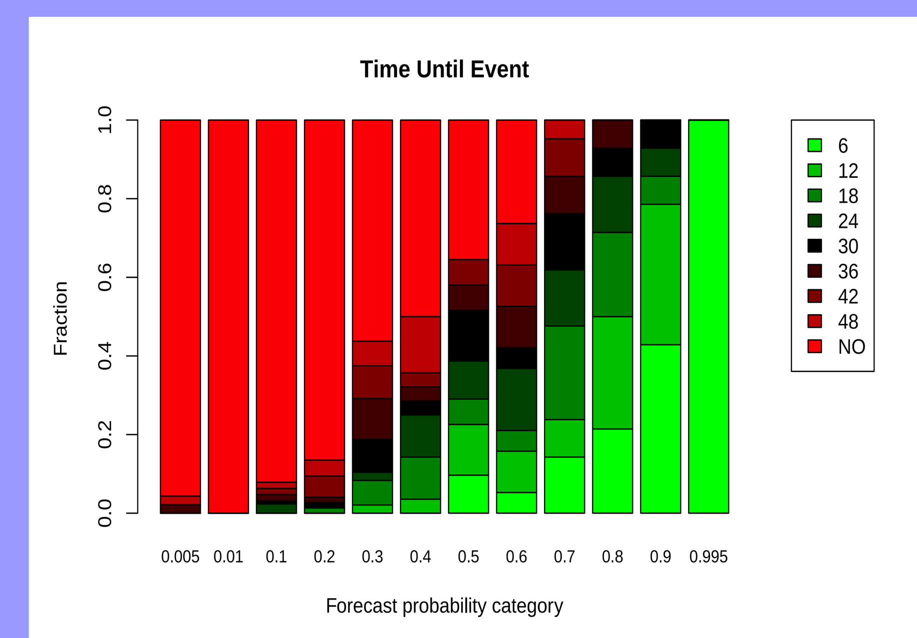

Fig 4: Relative frequencies of time until formation (in hours) for each NHC 2012 hurricane forecast probability bin. The categories indicate the occurrence of TC formation between the time given and 6 hours previous to it. A "NO" indicates a non-occurrence of TC formation within 48 hours.

We note that the time until formation is not uniformly distributed across the 48 hour window. Indeed all of the highest probability forecasts correspond to near immediate formation, as shown in Fig 4.

Strengths and weaknesses of 2012 NHC short-term TC forecasts

Figure 2 indicates overforecasting for the lowest forecast probabilities (except for 0.5% probability forecasts), and more significant underforecasting for higher forecast probabilities. This demonstrates a degree of underconfidence but good resolution in the 2012 forecasts [2]. Most of the observed frequencies lie within the 5% to 95% consistency bars (Fig 3) and the 90% reliability band (Fig 4), demonstrating a skilful forecast system on the whole. Underforecasting at the highest probabilities may suggest less predictability of tropical cyclone formation at shorter times until formation.

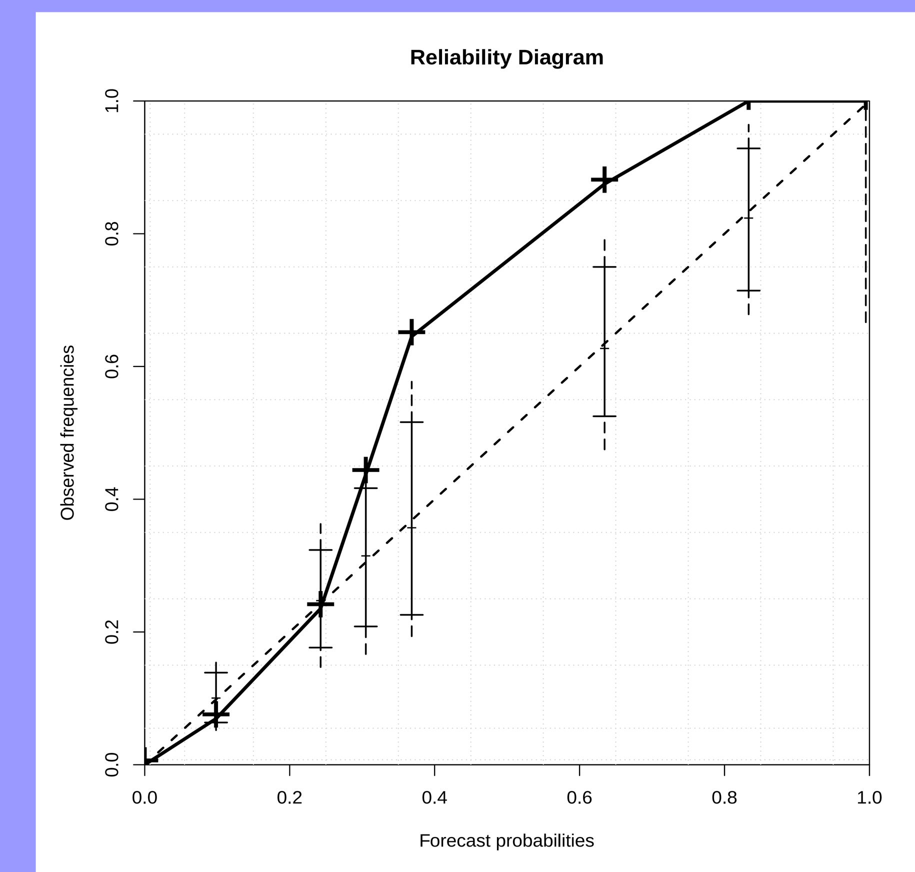

Fig 5: Recalibration reliability diagram (using 2011 forecasts as training set) for the NHC 2012 hurricane forecasts with 5% - 95% (1% - 99% - dashed line) consistency bars. The bin boundaries (light grey vertical lines) are positioned at [0.0075, 0.055, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95].

Attempts to recalibrate forecast systems by reassigning the probabilities, r_k, with observed relative frequencies, f_k are often suggested. Figure 5 shows how the forecast system would have performed for the 2012 hurricane season using the 2011 forecasts as a training set (see Fig 3) - it is significantly less skilful.

Discussion Points

- Reliability diagrams provide a quick and simple view of forecast performance
- The NHC 2012 forecasts are skilful
- Times until formation vary between the 2012 NHC forecasts suggesting additional analysis to identify information beyond reliability of the 48 hour forecast
- Recalibration of the 2012 forecasts using the 2011 forecasts leads to a poorer forecast performance

References [1] J. Broecker and L. A. Smith. Increasing the reliability of reliability diagrams. Weather and Forecasting, 22(3):651-661, 2007.

[2] D. S. Wilks. Statistical Methods in the Atmospheric Sciences, volume 91 of International Geophysics. Academic Press, 3rd edition, 2011.

[3] A. S. Jarman and L. A. Smith. On the reliability of 2012 Hurricane forecasts, 2013 (in preparation).

Acknowledgements: This research was funded as part of and the ESRC Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council and Munich Re.