Distinguishing between Skill and Value in Hurricane Forecasting

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] Oxford Centre for Industrial and Applied Mathematics, Oxford University, U.K.

Identifying skill with small datasets

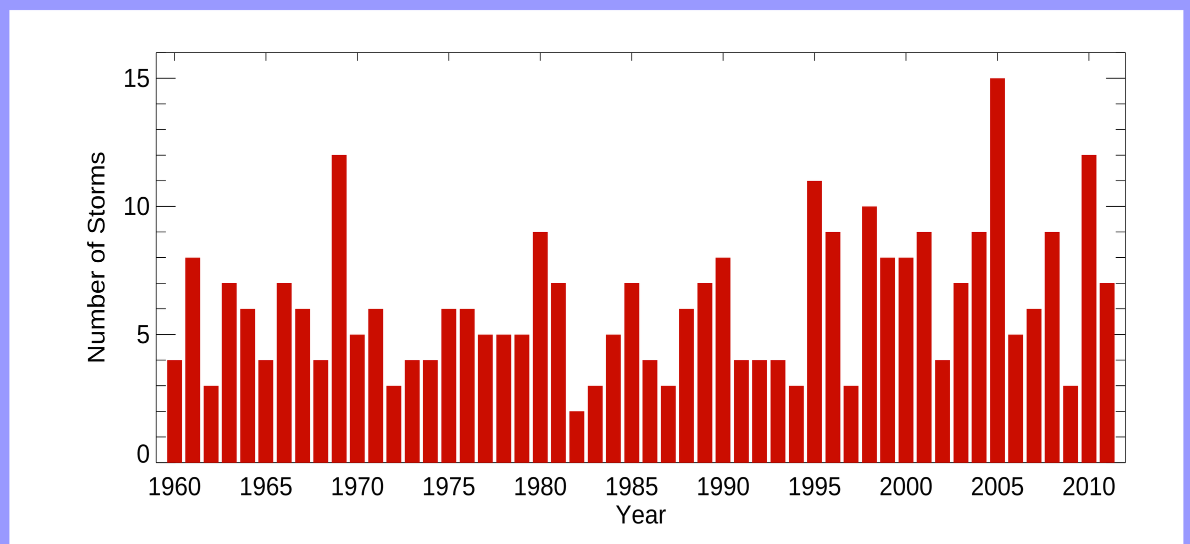

Fig 1: HURDAT data: Number of Atlantic Basin hurricanes from 1960 to 2011 [1]

Establishing out-of-sample skill in an annual hurricane forecasting system is unrealistic on less than decadal timescales because of the slow rate new verification information is gathered. The range of uncertainty in a sample of forecast model evaluations increases with decreasing time duration, and thus genuine skill cannot be reliably ascribed to a forecast model which is verified with few data.

If demonstrating genuine skill with limited datasets is not possible, then should a decision maker wait for proof of skill in a model before using it? Might they be forgoing the opportunity to benefit from forecast information whilst seeking statistical reassurance? The expectation may be that forecast value is similarly indemonstrable in the short term so here we address the following questions: 1) what would be the cost of waiting to prove skill for a decision-maker? And 2) how long does it actually take to prove skill?

Profit and Proof: What is the cost of waiting?

If a decision-maker believes in the skill of their model, they might rationally choose to begin implementing it and will begin to receive value before those who choose to delay. The chance to profit before proving can be conceptualised in the context of what is called the "Swindled Statistician Scam" (see Box 1).

Consider a toy hurricane system in which the mean number of storms follows a 24 year cycle, while the number of storms in any given year is determined at random.

To illustrate that structural model error does not preclude value to a decision-maker, consider an imperfect model of that system with the same cycle period, but where the probability distribution function used is incorrect in shape, not merely in parameter.

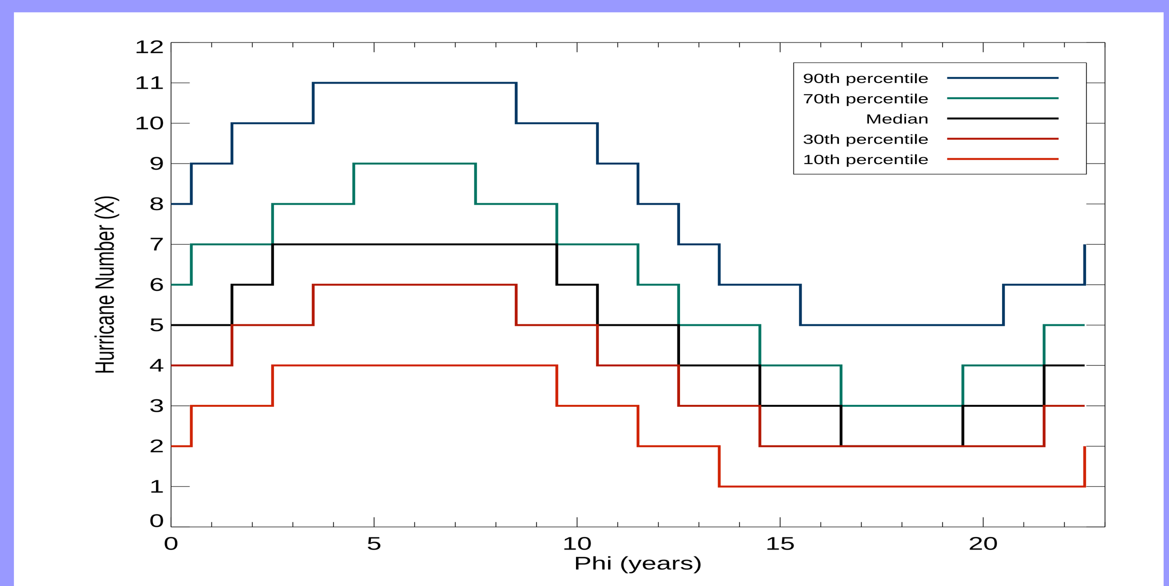

Fig 2: Contours of the cumulative distribution function showing how the system probability distribution changes as a function of the phase of the 24 year cycle. Note that occurrences of higher hurricane numbers are more likely in years with phase 5, 6 or 7 than those with phase 17, 18 or 19.

This model will then be used in games of Hurricane Roulette, where the imperfect (but time dependent) model probabilities are used to place bets against odds set by a house using the correct (but not time dependent) climatological probability distribution. The results can be reported in either bits of information or as an expected annual return (see [2]).

The system is defined as a Poisson process, $X \sim Pois(\lambda(t))$, where X is the number of hurricanes in a given year and has a sinusoidal time-dependent mean determined by the equation $\lambda(t) = 2.5sin(2\pi t/T + \phi) + 5.0$. The probability distribution of the system over each of the 24 phases is illustrated in Figs. 2 and 3 which describe the cumulative distribution function for all phases, and the probability density function corresponding to phase year 12.

Box 1. The Swindled Statistician Scam: A wily underwriter approaches a non-Floridian statistician with a business deal: the statistician will produce a probability forecast of the number of destructive events in the coming year, the underwriter will use her market contacts to bet on the forecast. As soon as the statistician can prove the forecast really does have skill, the underwriter will pay royalties. Will this leave the statistician swindled out of a small fortune?

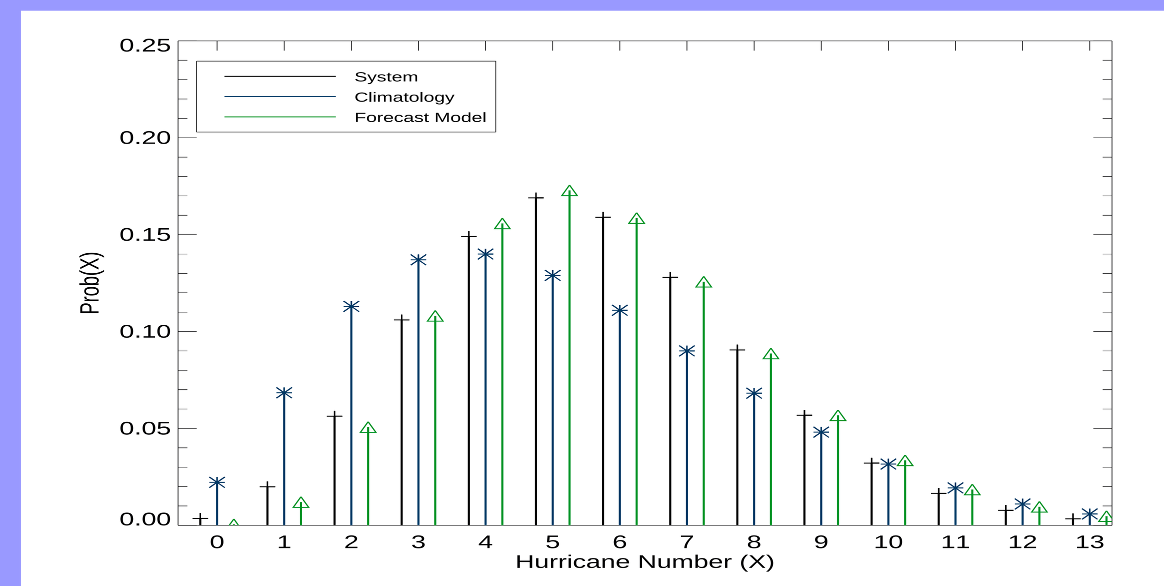

Fig 3: This figure shows the probability distributions for the system (black), and an imperfect model (green) for phase year 12 of the 24 year cycle. The climatological distribution (computed over all values of phase) is also shown in blue.

Hurricane Roulette proceeds as follows: at the start of each annual hurricane season a decision-maker is offered odds defined by the climatology PDF (equally-weighted sum of the 24 system phase PDFs). She then places her bet by distributing all of her current wealth (based on the Kelly betting strategy [3]) according to the forecast probabilities assigned to each possible hurricane number outcome (X). The actual outcome determines the pay-off on each annual bet.

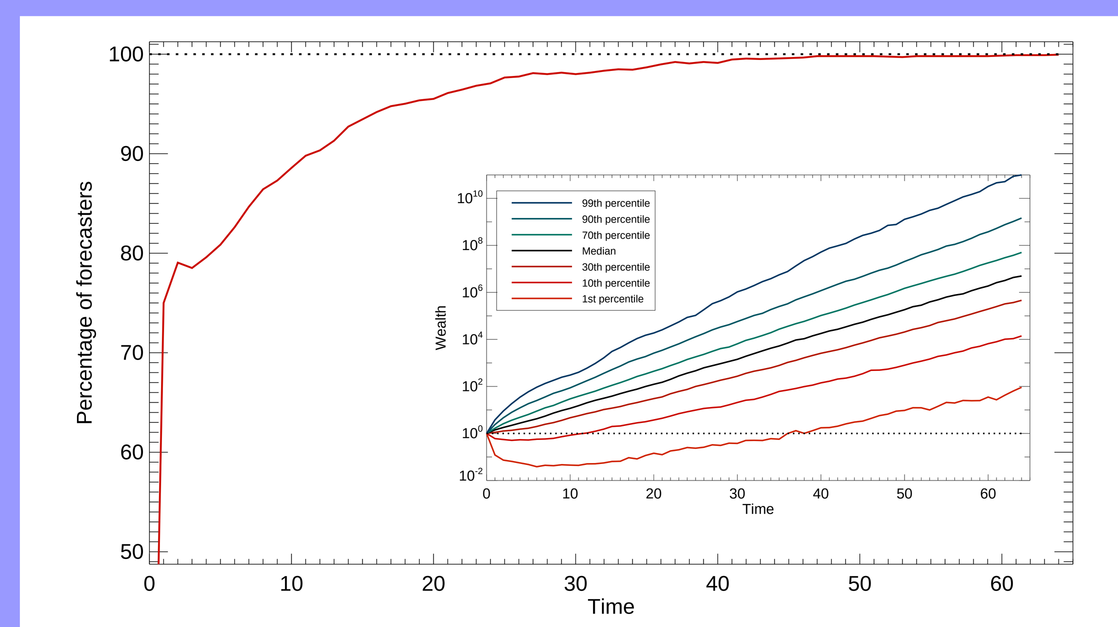

Fig 4: Percentage of 2048 decision-makers expected to make a profit with time when betting against climatology using the imperfect model in a game of hurricane roulette (main plot), and frequency distribution of decision-makers' wealth with time (inset plot).

The results of a sample of 2048 realisations (or worlds) of roulette demonstrate that the decision-maker would be very likely to have made a nontrivial profit even before two system cycles have completed (NB: the phase, ϕ, is selected at random for each realisation to avoid bias). This is evident in Fig. 4 which shows the percentage of decision-makers who are likely to profit and frequency distribution of their wealth over time.

How long does it take to prove skill?

To complete the assessment of the cost of waiting for the statistician the forecasts of the 2048 decision-makers are evaluated with the ignorance skill score [4]. Figure 5 shows the distribution of the decision-makers' forecast skill - measured by the ignorance of their forecasts relative to climatology. In this case, the minimum time required for over 99% of the decision-makers to prove skill is at least 2^7 (128) years; much longer than the time required to profit by betting on the forecasts.

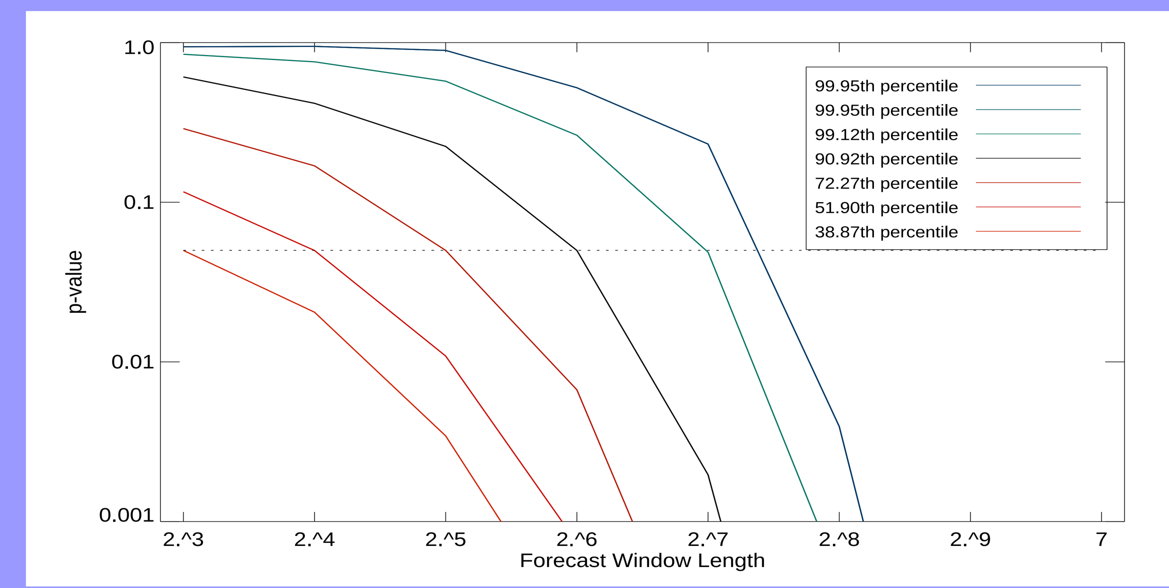

Fig 5: Distribution of forecast skill p-values (H1: $IGN < 0$) of 2048 decision-makers; 99.12% have established statistically significant skill ($p \leq 0.05$) by 128 years. NB: this percentage of d-ms had already made a nontrivial profit by 40 years (see Fig. 4).

Box 2. Discussion Points

- Establishing forecast skill on systems with long time scales poses a challenge
- A decision-maker may accept risk in order to gain profit, rather than first wait to establish statistical confidence in the forecast
- Ways of benefiting from an imperfect model can be demonstrated through the use of games like Hurricane Roulette

References

[1] National Oceanic and Atmospheric Administration. Atlantic hurricane database re-analysis project. http://www.aoml.noaa.gov/hrd/data_sub/re_anal.html.

[2] R. Hagedorn and L. A. Smith. Communicating the value of probabilistic forecasts with weather roulette. Meteorological Applications, 16(2):143155, October 2009.

[3] J. L. Kelly. A new interpretation of information rate. Bell Systems Tech., 35:917-926, 1956.

[4] Mark S. Roulston and L. A. Smith. Evaluating Probabilistic Forecasts Using Information Theory. Monthly Weather Review, 130(6):1653-1660, 2002.