Munich Re programme Symposium

Hurricane Forecasting: Skill and Value[1]

15 May 2012, LSE

Minutes by Emma Suckling, Alex Jarman, Lyn Grove

INTRODUCTION

The London School of Economics (LSE) hosted an academic symposium on Hurricane Forecasting as part of its Munich Re programme, with a focus on questions relating to the construction, evaluation and use of hurricane forecasts on seasonal scales.

Participants from the climate research and forecast user communities included representatives from the UK Met Office, UCL, Imperial College, KNMI, the Bermuda Institute of Biological Sciences, the William J Clinton Foundation, Christian Aid, as well as members of the Munich Re programme at LSE and Munich Re itself.[2]

The aim of the symposium was to open up discussion between the climate research and forecast user communities to share knowledge and better understand the needs of decision-makers when communicating uncertainty, and distinguishing between forecast skill and value. Various aspects of the end-to-end process of forecast production and implementation such as lead time, forecast targets, and good practice in forecast verification were covered in detail. Findings from the LSE's work in this area, funded by the CCCEP Munich Re programme, were presented to stakeholders and other users.

The principle focus was on the skill and value of Atlantic basin hurricane forecasts but other areas of extreme weather risk such as floods, landslides, and drought were also reviewed. LSE research in this area continues to be shaped by the requirements of forecast users and the knowledge and capabilities of forecast providers. Participants were invited to share knowledge to assist LSE in their understanding of the needs of decision-makers when communicating uncertainty, and distinguishing between forecast skill and value.

---

[1] This symposium formed part of the programme *'Evaluating the economics of climate risks and opportunities in the insurance sector'* at the London School of Economics, funded by Munich Re.

[2] A full list of participants is provided at the end.

MINUTES

The minutes have been organized under the following five themes that came out of the discussions:

- Understanding user needs
- Communication
- Forecast skill, value and verification
- Decision making and Uncertainty
- Evaluation

i) Understanding user needs

The focus of the meeting was on the months+ timescale for a focus on longer term planning. For **Humanitarian Aid** organisations, shorter lead time information - forecast information provided days or weeks in advance - is the most useful in terms of providing aid, resources and preparation (training, early warning alarms) to regions that are most vulnerable. But all information that can be provided is useful for deciding where to focus resources, including both seasonal forecasts and shorter term information about the risks in specific regions that may be vulnerable. Currently accessible information for vulnerable regions such as Central/South America is not available to NGOs, although that would be most useful. Regional decision-making is what **NGO**s are most interested in for disaster risk development, etc. Their key priority is a better use of the science available.

In the **insurance sector** seasonal hurricane forecasts could be used for risk management purposes. Landfall probabilities, in particular for major hurricanes (cat 3-5) that account for the largest fraction of historical losses, would be useful information for insurers (underwriters are trained in statistics and are able to cope with probabilities rather than just a number).

Statistical hurricane properties (expected frequency-intensity distribution) would be of interest to insurers not only for the coming season, but also at timescales of over 5+ years. However, underwriters' and brokers' memory in the wider insurance industry is often dominated by recent (non-)damage experience that may contradict forecast risk. In addition, there might be distrust of and reluctance to use forecasts: e.g., if one seasonal forecast is unskilful then the following year it will not likely be utilised. In a more general perspective, insurance business is about risk (probability times consequence) usually conceptualized as a dateless probabilistic attribute of an insured object rather than about forecasts for specific periods. Hence, there is also some fundamental tension between the idea of forecasts and the traditional concept of risk and insurance. Integer hurricane forecasts, e.g. not 2.2 landfalls, would be more useful for users within the insurance sector. Seasonal insurance risk management practices (treaties are renewed January and July) should take into account some degree of stochastic behaviour (volatility) in addition to forecast information.

Probability forecasts could be helpful for **reinsurance** where risk assessment is important. It is good to incorporate natural climate variability into predictions, e.g. El Niño, but one must be careful about interference or "masking" caused by the presence of other phenomena when analysing trends. Precipitable water content and short-term wind gusts are also of interest for damages rather than just number of storms or intensity, etc.

Rather than probabilistic forecasts, information about whether the season would be above or below average activity would be useful to decision-makers and insurers. For example, below average, average, or above average forecasts of rainfall as warning for landslides; or how much rainfall will occur over the critical period for sugar cane growth. Inter-seasonal oscillations that can relate to

tropical activity/storms at month-season timescales would also be useful. Information is needed at certain times for business, and it is always important to know the range of possible outcomes.

ii) Communication

Forecast producers need to give careful consideration to how they distribute/provide information, depending on whether it is important on a global/national/international scale. Communicating information to local people is also important.

Being transparent about forecast procedures - post-processing, evaluation, interpretation, uncertainties and limitations, etc. - is fundamental to building up trust amongst users, as well as good practice. Indeed both forecasters and scientists should work more with users, i.e. maintain a 2-way stream of information. Establishing a dialogue can help communicate the information and limitations of forecasts and the underlying science, so that users can understand how best to use the information they are given.

When talking about skill and value, it is important to be able to explain what different skill measures mean and how they can be interpreted. We need to be careful not to examine several skill scores and then select those that make our own forecasts look good. A better alternative is to have a justification for using certain measures, or follow a best practice approach. Forecasters should also consider whether to issue forecasts from dynamical models when they know the model has low skill. In these cases should they blend a model output with climatology? If this were done, people might tend to have more confidence in forecasts when they are issued.

It is also important to communicate how the physical phenomena contribute to the uncertainty in probabilistic predictions, to lend clarity and credibility to forecasts. Delivering forecasts without information about the uncertainties can lead to much bigger impacts if the forecast doesn't verify.

Communication should be targeted at different users: e.g. communication of skill through interest rates or something similar could be useful and would be easy for some users to understand, whereas a packaging of forecast information would be useful for others.

iii) Forecast skill, value and verification

Forecasts tend to be bound by information on specific timescales, e.g. dynamical models tend to be launched on fixed dates, incorporating specific observations. However, a useful addition to the standard dynamical model forecasts would be to go back and update the forecast given new information, or launching at intermediate start dates if there is information that could improve predictability. The UK Met Office produce hurricane forecasts a few times during the season with updates from seasonal models in between. WMO are carrying out comparison/verification of seasonal forecasts.

It would be useful to run stress tests on the means coming out of models and think about approaches to get an understanding of uncertainties on probability distributions. Guidance on numerical output is still commonly used and deliberation should be exercised when deciding to produce a forecast purely from model output or when blending with climatology if expecting low skill to avoid negative public perception.

Skill-value feedback (*chicken and egg?*) was suggested as a dynamic for forecast demand. To improve a forecast and leading to an increase in skill, its value must be proven to users first. However, given the historical lack of evaluations of operational forecasting systems (Gray et al at CSU are the only ones to have produced forecasts for more than 10 years) and modifications to systems, we cannot prove forecast skill as yet. Transparency with respect to these modifications is important to reliably establish confidence in skill. Furthermore, clarity in communication of what skill means is important in a score.

A factor which may influence users' perception of forecast skill is the 'guaranteed winner scam': if there is a large diversity of hurricane predictions, e.g. hurricane landfall predictions made by an expert elicitation panel, then there is a chance of 'lucky strikes' I.e. forecasting correctly by chance, which may mislead users as to the genuine skill of a forecasting system.

Another challenge for proving forecast skill arises when there is a presence of serial dependence in observational time series – this can result in over-precise confidence intervals and hence, misleading estimates of forecast skill.

Establishing meaningful skill/value measures for seasonal forecasts or one-a-year event type forecasts is much harder than for a weather forecast evaluation, due to the lack of statistical significance and the 'noise' associated with proper skill measures. We need to think more about what evaluation or skill means in such cases. Ideally we would arrive at an optimised PDF (maximum skill). Analogue, constructed analogue or regression-based forecasts could be useful to users as benchmarks or simply cheaper options. However, a lack of transparency can also be a problem with such forecasts and there are some products that appear to be well marketed, perhaps without any solid foundation/justification, and therefore no evaluation of them has been considered. Careful design of the value/loss function needs to be considered and tailored to specific users. The focus on high impact events could be important.

iv) Decision-making and uncertainty

Results from a survey conducted by the Institute of Psychology showed that sharper forecasts and higher resolution are preferred by users, even if on average a less sharp forecast performs to the same quality. The preference for forecasts that are too narrow tends to lead people to make worse decisions. We should make efforts to engage in dialogue and communicate uncertainties (first or second order). Better decisions will be made if we are honest about limitations.

1st order uncertainty only appears to be communicated to decision-makers in Economics forecasting – although economists do attempt to arrive at a PDF of the PDFs that are constructed from model output. Second order uncertainty is important and needs to be communicated to decision-makers, but many economists don't think about this type of uncertainty. The Bayesian framework is not suited to thinking about second order uncertainty as their assumptions involve the idea that in a set of models there is one that is truth.

Point forecasts may assume that the user cannot handle uncertainty which is probably detrimental. Users are generally happy to take on information about uncertainty, but deciding how to deal with it

is tricky. However, although there is a lack of capacity to deal with uncertain information in some organisations, efforts are being made to rectify this.
Poisson uncertainty……Dealing with uncertainty is much trickier when numbers are small.
We have to consider uncertainties not just on numbers that come out of forecasts, but on impacts too.

v) Evaluation
Models provide output and forecasts for all regions, but more work is needed to evaluate/verify and then distribute information to more of the regions of interest (Central/South America).

Ouija boards: how do we go about distinguishing between good/poor methodology when more spurious forecasting systems are incorporated into a forecast – e.g. expert elicitation that includes a range of models. A telescoping approach - or necessary conditions to proceed to the next round - ought to weed these out, i.e. starting with predicting SSTs over a large region, then zooming in on hurricane number, or something like that. But there is a question of how to do that properly.

(Reference was made to Steve Sparks who suggested a credibility rating or weighting in volcanology expert elicitations as a way of forming forecast uncertainty PDFs with multiple input from experts).

One [science] participant has a forecast comparison website that has been running for several years now, providing an independent evaluation of seasonal forecasts from nearly all past models and versions. He tends to show correlation coefficients as they are easy to understand and because proper skill score measures tend to be noisy. He is aware of the need to perhaps use more robust evaluations, in particular adding confidence intervals or error bars on the measures on his evaluation site.

Another participant is running a hurricane forecast competition to encourage forecasters from different centres and with different approaches to enter their forecasts and engage in the evaluation exercise. By asking forecasters for several regions the more sensible forecasts should do better.
Projects like these to independently evaluate forecasts from different models/centres would be beneficial, and would help in both establishing best practice and transparency amongst groups.

Stress testing of forecast means should be carried out.
Evaluating forecasts with statistical measures may be missing the point – the loss function can often be more useful to a user.
Small sample sizes are a big problem for seasonal+ forecasts and we need to think much harder about how we can provide robust evaluations of our forecasts at these timescales.
There is a question of how you evaluate forecasts given few verifications and a need to carefully consider what categories should be judged to weed out the Ouija board type forecasters.

PARTICIPANTS

D. James Baker	William J Clinton Foundation – Clinton Climate Initiative
Simon Dietz	GRI/CCCEP, LSE
Walter Distaso	Imperial College
Eberhard Faust	Munich re
Nigel Harvey	UCL
Leon Hermanson	UK Met Office
Matt Huddleston	UK Met Office
Alex Jarman	CATS & CCCEP, LSE
Adam Lea	UCL
Rick Murnane	Bermuda Institute of Ocean Science (BIOS)
Geert Jan van Oldenborough	KNMI, the Netherlands
Jose Luis Penya	Christian Aid
Nicola Ranger	CCCEP, LSE
Leonard Smith	CATS & CCCEP, LSE
Emma Suckling	CATS & CCCEP, LSE
Swenja Surminski	CCCEP, LSE
Erica Thompson	Imperial College, London
Emma Visman	Humanitarian Futures, KCL
Oliver Walker	CCCEP, LSE