Key Contacts

Tom Bolt Director, Performance Management
Telephone: 020 7327 6700 tom.bolt@lloyds.com

Trevor Maynard Head, Exposure Management
Telephone: 020 7327 6141 trevor.maynard@lloyds.com

Neil Smith Manager, Emerging Risks & Research
Telephone: 020 7327 5605 neil.j.smith@lloyds.com

Contents

Acknowledgements

Emerging risks team

The Emerging Risks team is part of the Performance Management Directorate at Lloyd's. We define an emerging risk as an issue that is perceived to be potentially significant, but which may not be fully understood or allowed for in insurance terms and conditions, pricing, reserving or capital setting. Our objective is to ensure that the Lloyd's market is aware of potentially significant emerging risks so that it can decide on an appropriate response to them. The Lloyd's Emerging Risks team maintains a database of emerging risks that is updated regularly through conversations with the Lloyd's emerging risks Special Interests Group, which consists of experts within the Lloyd's market put together with help from the Lloyd's Market Association. The team also maintains contact with the academic community, the wider business community and government. Contact with academics is often facilitated through the Lighthill Risk Network, an organisation that is run as not-for-profit funded by Aon Benfield, Catlin, Guy Carpenter and Lloyd's.
More details can be found at www.lloyds.com/emergingrisks.

About the author

Dr Matt Huddleston is Principal Consultant on Climate Change at the Met Office where he leads the development of climate services for the financial industry. He has worked on a number of insurance related projects including an assessment of climate change for the Association of British Insurers. As a physicist, he previously led the development of long-range forecasts in the Met Office Hadley Centre and has published research on various topics including tropical storm forecasting and ocean observations. In producing this report, Dr Huddleston was assisted by a number of members of the Met Office Hadley Centre, including Dr Jeff Knight and Dr Anca Brookshaw. Andres Drew also contributed whilst on placement from the London School of Economics.

About the Met Office

The Met Office is the UK's national weather service and a world leader in weather and climate science. It provides global weather services for public, governmental, military and commercial use. The Met Office Hadley Centre undertakes research on both natural and man-made climate change. Uniquely, the Met Office undertakes global research and forecasting on timescales from minutes to decades ahead. It is a World Meteorological Organization designated Global Producing Centre for long-range forecasts.

Workshop participants

Lloyd's would like to thank the following people for attending workshops at Lloyd's that debated the issues around longer-range forecasting and helped develop the thinking presented within this report:

Grant Baxter, Liberty; Robert Caton, Hiscox; Alex Jarman, London School of Economics; Paul Kershaw, Catlin; Gero Michel, Willis; Nicola Ranger, London School of Economics; James Slaughter, Liberty; Lenny Smith, Oxford University; Swenja Surminski, London School of Economics; Dickie Whittaker, Lighthill Risk Network; Andrew Voysey, ClimateWise.

ClimateWise
REDUCING THE RISK FOR TOMORROW

Lloyd's is a member of ClimateWise, the global collaboration of leading insurers focused on reducing the risks of climate change. Launched in 2007 by HRH The Prince of Wales, and facilitated by the University of Cambridge Programme for Sustainability Leadership, ClimateWise brings together over 40 international members from Europe, North America, Asia and Southern Africa.

All members publicly commit to abide by the ClimateWise Principles, which cover climate risk analysis, public policy, climate awareness amongst customers, investment strategies and the impact of their business operations. Members also commit to independent public reporting against all of these commitments.

1 Foreword

Catastrophe modelling plays a vital role in helping insurers manage their exposures to natural catastrophes, such as hurricanes, flooding and earthquakes. While statistical models using climate data continue to be used by the majority of insurers, emerging forecasting methods and techniques provide risk experts with additional options to further develop and refine their modelling practices.

The climate is clearly changing, with increasing evidence this climate change is leading to more frequent severe weather events. This results, together with increasing property values and concentrations of population in catastrophe-exposed areas, in increased insurance losses. Insurers need to find better ways of predicting these extreme events and the techniques of long-range forecasting provide a welcome addition to the debate on how to do so.

As our awareness and understanding of climate change has increased, there have been corresponding developments in forecasting technology and modelling techniques. Forecasting scientists are developing models to predict weather events and patterns over a longer timeframe – from the seasonal out to five years. These increasingly sophisticated models take into account ocean and atmosphere conditions, as well as seasonal and regional climate trends, such as the El Niño-La Niña cycle. Equally importantly, forecasters are measuring and refining the skill of these models and making them more relevant to the premium setting and capital decisions that insurers have to make.

We believe that such longer-range forecasting techniques will have an increasingly important role to play in the insurance market and help to significantly improve and develop existing practices – particularly as the impacts of climate change are increasingly felt. That is why we held a series of workshops with modelling and forecasting experts from within and outside the Lloyd's market. It is important that long-range forecasting and related modelling techniques are properly debated and evaluated and the findings shared for the wider benefit of the industry.

Throughout this process we have been working closely with the Met Office, who has undertaken much work in this area. This paper, prepared in co-operation with the Met Office, draws out the key points from these workshops and highlights both the challenges and potential benefits of introducing long-term forecasting into our industry.

Trevor Maynard
Head, Exposure Management
Lloyd's

2 Executive summary

Forecasts should be useful to the insurance industry.

Traditionally, the industry has established the probability of a weather event by the use of past weather data (climatology) because forecasts, particularly over the longer term, were not always reliable. Weather forecasts were generally informative over a period of days, not the months between the issuing of a policy and the start of a storm season. However, recent research suggests that some phenomena can be forecast months in advance more accurately than by using climatology – for example, sea surface temperatures in the Pacific and Atlantic tropical storm numbers. While it is impossible to forecast with certainty, the range of outcomes is becoming more accurate and decision-relevant for the insurance industry.

Is past data still a useful guide to the future?

Our climate is changing, and we appear to be seeing more frequent extreme weather events, so the wisdom of relying solely on historical observations is being debated. In addition, the recent past does not capture the impact of rare catastrophic events, such as from large volcanic eruptions, and this is an area where the forecasting sciences may help.

Ocean conditions can be used to produce long-range forecasts.

Recent research has shown that current ocean conditions can be used to inform likely risk levels of future extreme loss-making events over one to seven years ahead. The ocean and atmosphere evolve partly in response to large-scale natural variations, as happens with El Niño, and understanding and using such causal physical processes will enhance statistical approaches.

Premium and capital decisions should be based on prospective levels of risk rather than relying only on long-term averages.

Sea surface temperature-conditioned models are already recommended best practice for analysing catastrophe risk for insurance[1]. This report suggests methods and techniques which the industry could use to further inform pricing and capital decisions. In terms of Solvency II, risk forecasts could be seen as a useful component of a wider risk management framework.

The concept of skill is critical when assessing the value of forecasting information. An individual forecast – whether it is seen as accurate or not – is not a good guide for overall utility of a forecasting system.

It is not easy to assess the accuracy or the skill of forecasts - and skill scores rarely give the full picture. However, industry risk modelling teams should educate themselves on the ways in which the skill of forecasts can be measured. Further development is needed so that skill measures are based on their usefulness to end users and so test how useful a forecast is in guiding practical business decisions.

More detailed global earth observations will improve forecasting.

More in depth observations are needed to improve the models and our understanding of the climate. Present-day forecast accuracy estimates may also be understated simply because we did not have the benefit of data from earth observation systems in the past. Earth observation data will need to be both maintained and enhanced in the future to deliver the maximum benefits from advances in forecast systems.

Will first adopters of forecast models be at a competitive advantage?

Models which bring together ocean and atmospheric interactions are still new. However, the development of forecasting technologies – in particular in relation to increased computing power – is likely to improve forecasting of perils useful for insurers. Additionally, academics are increasingly keen to tackle the issues and questions most relevant to insurers.

Research should be carried out to establish the value of long-range forecasts.

Whilst use of a new technology represents considerable investment for insurers, we believe that market participants should investigate how long-range forecasting methods might improve current modelling work, and also how it might mitigate uncertainties created due to climate change.

[1] Lloyd's ICA guidance notes: "…there is strong evidence that hurricane risk in the North Atlantic is raised above long term averages."

3 Introduction: the insurance market and forecasting natural hazard risk

The insurance industry has undertaken extensive research to understand the risk from extreme loss-making events – perhaps more than any other industry. Analysis of the observed past, sometimes with statistical adjustment for trends, has been used to assess future risk of perils. However, operational decisions in the industry are focused on the next contractual period; predominantly the next 12 months. As such, while strategically relevant, climate change research remains largely focused beyond contractual timescales.

Lloyd's convened an expert working group of underwriters, catastrophe modellers and research scientists to explore the potential of using long-range weather and climate forecasts in insurance to predict events for the coming months to five years ahead. The group met at two workshops and the key themes of the discussion are presented below. This is followed by an examination of insurance industry requirements of long-range forecasts.

The second section of the report provides an overview of the current state of forecasting science, focusing on dynamical climate models being used to forecast one month to several years ahead (long-range forecasting). This section explains why dynamical forecasts[2] are possible and the components of the technology used for long-range forecasting. Dynamical forecasts of loss-making events offer the possibility of going beyond the statistical use of historical data as the sole method for loss estimation: indeed, in a changing climate, past data may no longer be a useful guide to the future.

3.1 Current market practice regarding the use of forecasts

The use of unadjusted long-term observed climate as a baseline for pricing risk is already viewed as the lowest common denominator by many underwriters. Such an approach will inevitably fail to allow for multi-year trends or cycles. For instance, a five year expectation of Atlantic tropical storm activity derived from expert views and statistics has been available for some time[i] and is widely used in the market. Financial risk management professionals, including actuaries and underwriters, are in principle able to adjust hazard activity rates within the catastrophe modelling approach currently used in the insurance market.

There is a perceived lack of comparison between forecast models and observations for verification. For long-range forecasts however, a single high-profile seasonal forecast that gives bad advice can significantly impact people's views on the overall value of forecasting. This is despite widespread understanding that an individual poor forecast is not a good indicator of the overall value of long-range forecasting. Forecasts of probabilities and likelihood or risk-based forecasts need to be evaluated over many decisions. This presents a significant challenge in communicating the value of risk-based long-range hazard forecasts. Forecasts can also be challenged by current perceived wisdom when they give signals at odds with past experience or standard methodologies i.e. a forecast that differs significantly from climatology or a near-term view of climate may have significant implications for current business, for example requiring increased capital allocations, and is therefore more likely to be challenged.

That said, the Lloyd's Working Group believed that a number of the leading underwriters and risk managers already have a good understanding of how forecast information could be incorporated into their decision making, but are limited due to inconsistencies in how the information is communicated. There are cost implications in the use of new methods that also need to be considered, both in their implementation and in the communication to stakeholders regarding changes in risk management practices.

Across the market and in the wider forecasting science community differing terms and concepts are used and common standards and terminology would help people use forecasts already available.

3.2 Reputational issues

It can be difficult for an individual company to be the first adopter in the market place. Some comfort, in terms of information, education and market debate, may be needed before individuals can challenge current market practices for pricing and capital decisions. This would likely change rapidly were forecasts shown to be more accurate than existing methods. A solution to this would to be to apply new longer-range forecasting methods alongside current practices.

3.3 Pricing

Prior to the annual insurance contract cycle, long-range forecasts could be used in a number of different ways within insurance markets. Wholesale price is subject to market constraints, location and class of property insured; however key pricing decisions can be adjusted within certain limits. The use of forecasts in this way is thought to be possible with the current level of forecast information that is already available in the market place.

---

[2] By a *dynamical* forecast we mean a numerical forecast from a computer model of the earth's weather and climate system that takes into account initial atmospheric and ocean conditions and evolves them into the future using the laws of dynamics and thermodynamics.

After contracts are in place, alternative risk transfer products, such as catastrophe bonds, can also be bought or sold to offset forecast risks or to take advantage of new opportunities from decreasing risks. Taking the decision to leave or re-enter a market is an option of last resort due to significant associated costs and impacts on long-standing relationships.

Using skilful[3] long-range forecasts would likely mean a certain amount of asymmetry in the response to a high hazard activity forecast compared to a low one, which may enhance profit. For a low hazard activity forecast, it is likely there would be a lower limit to how far capital requirements or prices could be reduced, as even a quiet year could have a significant loss making event, such as Hurricane Andrew in 1992.

To integrate forecasts into existing pricing mechanisms, more transparency may be required around current processes and pricing solutions. Volatile pricing however may be resisted by insurance customers and result in calls for long-term pricing of policies. This could be at odds with efforts to improve capital adequacy as long-term policies may weaken market responsiveness.

As forecasts generally improve as we approach an event, we may see the development of secondary markets to enable the transfer of risk to capital and other markets on shorter timescales. Opportunities for mid-year portfolio re-adjustment would generally be welcome. In theory, if forecasts become too accurate it could be difficult to find insurance counterparties, although in practice there will always be an aspect of chaos or randomness in the weather ensuring the loss remains unexpected and other markets - for example, capital markets, may be encouraged to participate in wider risk diversification.

3.4 Markets

As forecasts are enhanced and the understanding of forecasts improves, and if they are recognised as useful in the marketplace, it is likely wider market pressure would be placed on companies to use them in risk management. The timing of such a significant change to the market could be critical. A low investment-return environment, such as we have at present, may be an ideal time to focus on the quality of underwriting risk information and establish new levels of confidence in market innovation and practices. Previous experience has shown that there are tipping points in market behaviours, where new ideas move from being idiosyncratic to normal market practice – for example, the adoption of catastrophe models in the 1990s following the rapid post-event loss estimation for Hurricane Andrew in 1992.

In the future, Solvency II and other market regulatory reforms may require forecasts to be judged within overall enterprise risk management practices. There would be risks to the market, however, if too few forecast sources feed into a small number of risk modelling frameworks.

Overall, members of the two workshops felt that current long-range weather and climate forecasts should be of value to the market. However, they also felt that it would be hard to demonstrate how the benefits of the forecasts can be fully realised under current market conditions (i.e. a generally soft market) and practices. Even if they are used conservatively, forecasts are likely to perform better than the long-term baseline risk / climatological approach, but significant innovation will be needed to demonstrate the benefit and foster adoption. As such, there was agreement that it would be a good idea to establish a virtual re/insurance business for research on risk management and pricing to establish the value of innovations, such as longer-range forecasts.

[3] *Skill* is a term used to define the expected performance of a forecasting system for a particular type of forecast. Skill measures help users assess how reliable a particular forecast is. It is often derived by running a near-identical forecasting system over a historical period. These historical forecasts are called hindcasts or re-forecasts.

4 The science – forecasting beyond the next week out to years ahead

4.1 Requirements of a forecast

Longer-range forecasts should be of use to the insurance industry, where relevant timescales (see below) and perils are considered and the forecasts are skilful enough. For pricing the impact of meteorological hazards, longer-range forecasts are needed that fit into the annual cycle of the insurance industry and fulfil four basic criteria:

1 Timing

The highest proportion of reinsurance business is written on 1 January, although other dates (including 1 April, May, June and July) are also used. Almost all contracts are 12 months in length. This means that a forecast for the peril concerned is needed ahead of contract execution (plus preparation time.) Assuming that catastrophe risk calculations can be updated in a timely manner, this suggests that the required forecast lead time is 4-12 months for most risk pricing decisions.

For example, a forecast of European wind risk for the coming winter period (January to April) of an annual contract could be used in contracts executed on 1 January if provided latest on 1 December. For USA wind risk from Atlantic hurricanes, a forecast of the next June-November season's risk would ideally be needed by the previous 1 December although updated forecasts ahead of subsequent key contract dates will also be relevant.

Other types of alternative risk transfer, such as insurance linked securities and CAT bonds, can be purchased at any time if contractual counter-parties can be found. This suggests that forecasts on timescales of days to several years may also be useful.

Even longer-range forecasts of hazard activity over one to five years, may be useful in some insurance markets for pricing and also especially for the strategic management of portfolios of risk within an enterprise risk management framework.

For those that are concerned with market strategy and reducing damage done by hazards, even longer-range forecasts – out to ten or more years – may be useful, for example, for informing building codes and design standards.

2 Relevance

The forecasts must address key impacts from the perils concerned. A typical example is tropical storms, which can cause significant losses when they strike land. As many storms do not strike land, a landfalling forecast is highly desirable as opposed to overall activity forecasts. An individual major storm making landfall in a quiet year can cause significant losses, for example Hurricane Andrew in an otherwise quiet 1992.

Extreme events, such as one-in-200 year floods or storms, are of most interest generally. While these may be much harder to predict (not least because of the lack of observed data for verification), extreme events or clusters of events may be linked to some large scale driving forces or cycles, such as El Niño, which have more predictable impacts.

Indeed, large-scale patterns such as an El Niño have been shown to make extreme climatic conditions more predictable[ii,iii]. This means, when something significant is happening in the global climate system, local impacts are more predictable.

3 Usability

Users need to be able to integrate forecasts into their risk management systems and pricing decisions. Current cutting-edge catastrophe modelling methods use large hazard event sets describing, for example the equivalent of 10,000 years of hurricanes. These are generated by random (stochastic) re-sampling of the parameters that describe the observed hazards from the last ~50 years, such as maximum wind speed in a storm. These are now increasingly enhanced by simulations from high-resolution numerical weather models. Such hazard event sets can be re-sampled to reflect forecasts derived from statistical or expert elicitation exercises.

For dynamical climate forecasts to be used in the current framework of risk management, new and longer event sets could be produced or the current event sets re-sampled so that they match the expected distribution of hazards in the forecast, as has been done previously for climate change studies[iv].

Underwriters and catastrophic risk modellers already have the ability to update the event frequency of hazards within the current catastrophe modelling process and this offers one approach to the use of long-range forecast information that could be used today.

4 Confidence

Users must have confidence in the methods, the systems and underlying science. Communicating complex systems to decision makers is a challenge, but efforts must be made to explain the benefits and limitations of these methods. Central to this is the idea of skill and the key question is: how accurate are the forecasts?

The scientific community and forecasters need to create tools which risk managers and insurance professionals can understand and use. Specifically new skill measures based on usefulness for end users are needed rather than science-focused verifications and this may mean integrating these measures into current risk management and pricing methodologies. In other words a model is only useful if it helps with business-relevant decisions – and skill measures should be designed to test this. Plausible scenarios and explanations of current climate signals may also assist in confident decision making, as opposed to relying on pure statistical analysis – for example, the ability to issue more confident forecasts of extreme weather events during certain events, such as an El Niño.

All levels of users from analysts to underwriters and board-level executives would benefit from a deeper understanding of forecasting. Drivers, such as Solvency II and Enterprise Risk Management, suggest that increased education and broader skill sets will be needed in the future.

4.2 How are long-range forecasts possible?

The short answer is that the oceans store an enormous amount of heat. The top three metres of the ocean hold more heat than the entire atmosphere. The heat moves around in the various currents and circulations relatively slowly and interacts with the atmosphere at the ocean surface.

The net balance of energy on the planet needs to be near-zero for a stable climate. Within the climate system, solar heating in the tropics flows naturally to the cold poles. Half of this heat is moved towards the poles by the global oceans and half by the atmosphere.

In addition to this, the weather-bearing atmosphere is affected by moisture and heat from the land surface and vegetation; snow and ice cover; radiative properties of atmospheric aerosols and ash; and even high-level winds in the stratosphere. Many of these phenomena have a long-term footprint; they can influence the atmosphere over weeks, months and years.

Like the vibrations of a drum or violin, there are also a number of natural cycles in the climate system that characterise patterns in global weather over long timescales. These are sometimes coupled interactions between the oceans and the atmosphere - features storing heat in the ocean interact with the atmosphere and vice versa. Beyond the day and night cycle and the seasons, the most dominant and predictable pattern of natural variability is the El Niño Southern Oscillation, which varies between a period of three and eight years. No two El Niño events are identical, but the type of impacts are well documented and relatively well understood.

The El Niño – La Niña cycle is globally important as its influence is felt worldwide. It has a significant impact on Atlantic tropical storm season activity and its influence is also directly felt in Australia, where the El Niño phase causes a significant increased risk of bush fires, and the wet La Nina phase increases the risk of flooding. The intense rainfall in Australia in late 2010 and early 2011, for example, was predicted several months in advance and led to devastating impacts on individuals, property and a range of industries.

El Niño also modulates European winters[v], as does the stratosphere and its Quasi-Biennial Oscillation. A more detailed description of these natural modes of variation is described in Appendix 1.

4.3 Traditional long-range forecasts

The key to weather and climate forecasts is an understanding of large-scale natural modes and variations as they influence day-to-day (synoptic) weather patterns and allow a hierarchy of predictable phenomena. Beyond a purely black-box statistical analysis of past weather and forecast data, an understanding of the causal physical reasons why certain events are happening allows models to be improved, forecasts to be verified and an overall increase in confidence in forecasting systems.

Traditional forecasts using statistical or empirical forecasting methods use patterns of past climate and current observations as predictors of the future (for example, Bill Gray's original tropical storm forecasts from CSU - Colorado State University[vi]). This approach to forecasting requires the historical period used to derive the patterns to be representative of current climate and is unlikely to pick up the complex (non-linear) interactions of different modes of natural climate variation, such as El Niño. This is because local climate conditions in a particular season can significantly modify the impact of large-scale phenomena like El Niño. Statistical approaches are very useful to set the baseline level of skill performance for dynamical models. They can also offer additional skill where biases or errors in the numerical models are high and a clear physical mechanism has been identified, such as the link between ocean surface temperature and hurricane formation.

The provision of long-range forecast advice requires the use of a range of methods, including statistical and numerical/dynamical methods, to allow forecasters to explore and communicate the full range of possibilities to end users.

4.4 Certainty in forecasts and the role of chaos

Weather forecasts start with the initial state of the atmosphere and may be described as initial value problems. Forecasts assume that the future state of the atmosphere can be predicted by using the physical laws of nature on the current global weather conditions, for example the laws of motion, thermodynamics etc. For weather forecasts going out to a few days, the boundaries of the atmosphere include the ocean and land surface. The constituents and radiative properties of the atmosphere and incoming solar irradiance are also considered boundary conditions.

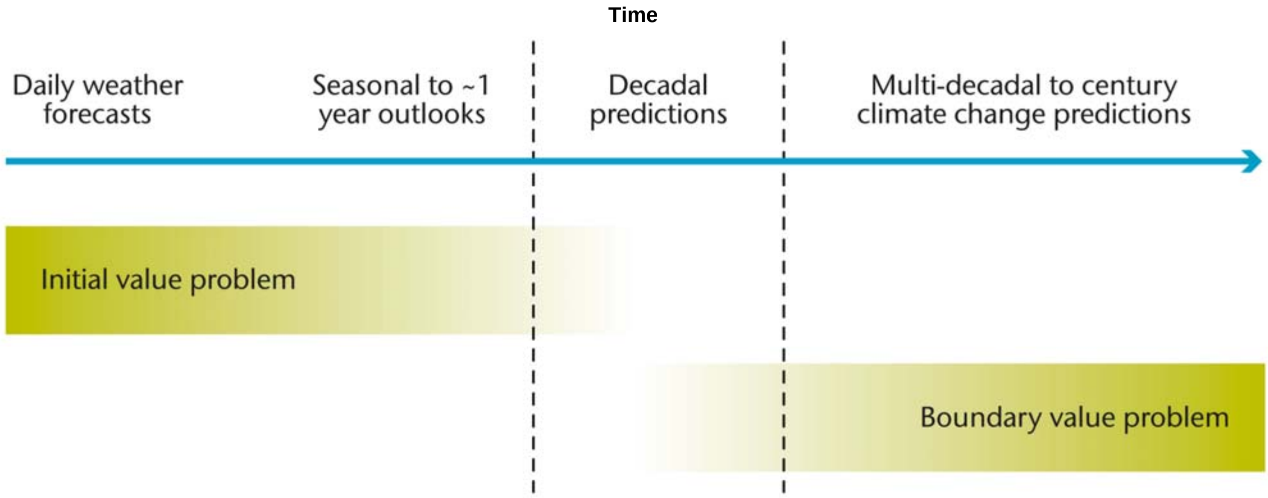

Figure 1: Schematic illustrating the progression from initial-value problems, with daily weather forecasts at one end, to boundary value problems with century-scale climate projections at the other, with seasonal and decadal predictions in between[vii].

These boundaries of the atmosphere do change over time, but these changes are not large enough to affect short timescale weather forecasts (a few days for most phenomena). However, with longer-range forecasts they become increasingly important until, with multi-decadal forecasts, the boundaries have a more significant impact than that of the initial conditions (see Figure 1).

One of the reasons for this is that even the smallest of errors in the initial analysed state of the atmosphere can have very significant impacts on the outcome of a forecast due to the highly complex and interconnected nature of the climate system. The climate system is highly non-linear; not only are there tipping points and thresholds within it, but the dynamics of the system are inherently chaotic. This means it is impossible to forecast with certainty. However the range of the possible outcomes may still be predictable and this is far more relevant to the insurance industry.

For all timescales beyond a few days, probabilistic methods are required and these aim to sample the range of possible outcomes in a rigorous manner. This is usually done by running a forecasting system many times into the future to sample the range of different future scenarios – an ensemble forecast (see Figure 2). The uncertainties sampled may include those in the initial conditions, the boundaries, the physical laws and approximations made to encode them. Additionally, several forecasting systems may be combined to sample their different formulations (so called structural uncertainty), forming a multi-model ensemble[viii].

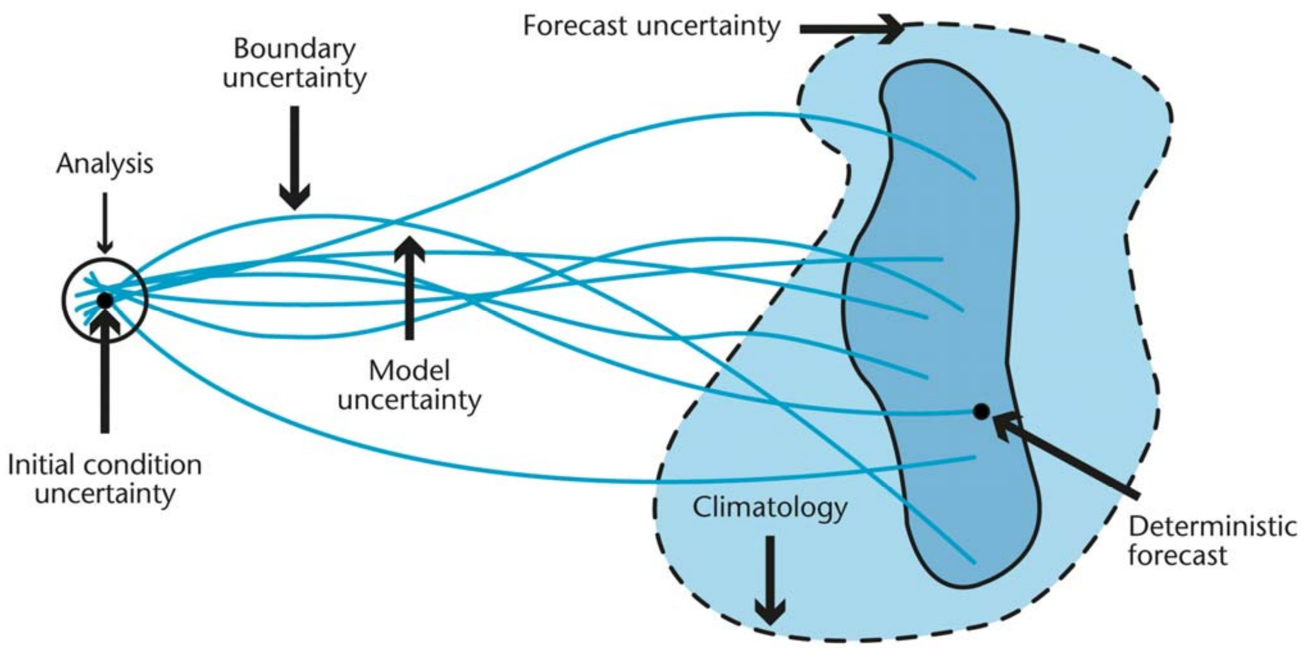

Figure 2: A schematic representation of an ensemble forecast system and the uncertainties being sampled. The ensemble of forecasts is run forward in time to sample the range of possible outcomes. Each ensemble member is physically consistent and provides an equally likely realisation of the future. This is somewhat similar to the stochastic modelling carried out by insurers when assessing capital, but with the addition of a consistent physical basis to model. Note that the climatology (the observed past) is also uncertain due to measurement errors and missing data.

The spread in the ensemble forecasts should in theory represent the certainty to which one can make a particular decision. This requires a forecast system to be well calibrated, so that the spread does in fact match the error growth caused by the various sources of error and the non-linear system. The spread may be useful in decision making if it is reliable, allowing efficient hedging strategies to be developed (this is often done, for example, in medium-range trading strategies for energy commodity trading).

4.5 The forecasting science and technology – how is it done?

A long-range forecasting system comprises of a number of components (see Figure 3). These include the observational analysis and initialisation procedure, the forecast model and ensemble generation system, and the analysis of the forecast data to turn it into decision-relevant aids and tools.

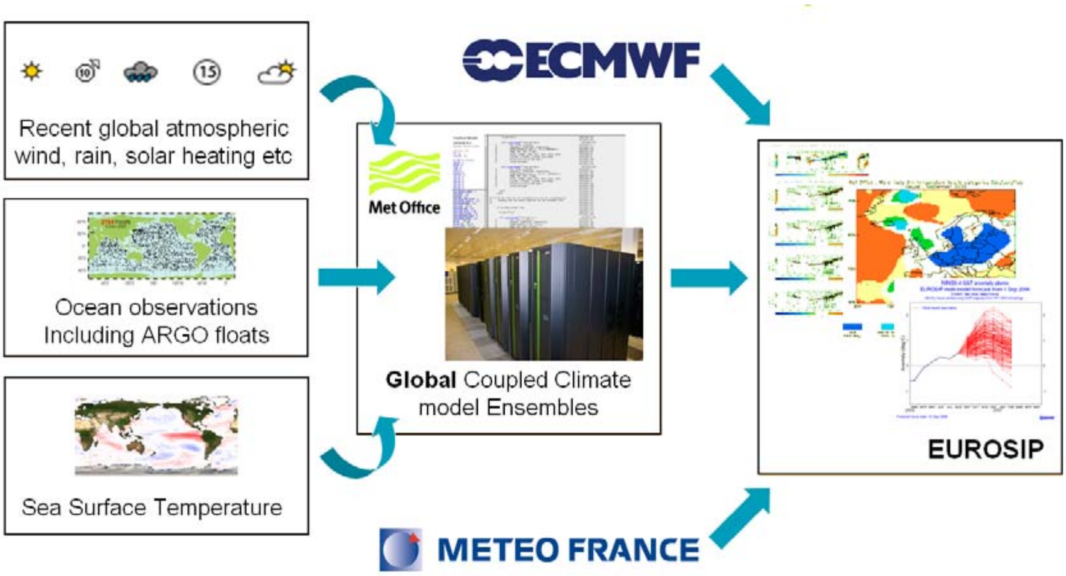

Figure 3: Schematic of a cutting edge long-range forecasting system – in this instance the European multi-model ensemble[ix]. The observations are combined to form a global analysis or nowcast, which initialises the forecast models. The outputs from three separate forecasting centres are combined to produce the forecasts to reduce uncertainties. Aside from the Met Office, the other forecasters are ECMWF (European Centre for Medium Range Weather Forecasts), and Météo France.

Climate models are numerical models of the fluid dynamics of the atmosphere and oceans, including various physical process and approximations (these are called parameterisations and include unresolved sub-grid processes such as clouds) coupled together with land-surface models, ice models, atmospheric chemistry and various other components. As mentioned above, the various physical laws that govern the movement and conservation of heat, moisture and momentum are encoded together and represented on three dimensional grids reaching from the ocean floor to high in the atmosphere.

Figure 4: Global Producing Centres making operational long-range forecasts using global circulation models as designated by the World Meteorological Organization[4].

Currently, the World Meteorological Organization has 11 designated Global Producing Centres[5] which are developing and operating global long-range dynamical forecasting systems of varying complexity (Figure 4). The Climate Prediction Centre (NOAA), Met Office and the European Centre for Medium Range Weather Forecasts (ECMWF) operate three of the most cutting-edge systems. Europe, among others, has historically funded and led several large international collaborative research programmes (such as DEMETER and ENSEMBLES) to develop the technology behind these systems and explore the application of the science [viii, x].

The design of the forecasting system depends on the overall computing power available – five factors contribute to the overall calculation: resolution; complexity; number of forecasts in the ensemble; re-forecasts and calibration; and forecast initialisation. Computing power is very much the limiting factor and often means that difficult decisions and trade-offs are required. These trade-offs mean that, for example, the range of the forecast or the geographic detail are necessarily limited. As an example of a global system, the Met Office GloSea4 long-range forecasting system is illustrated in Appendix 2.

4.6 Skill – how good are the forecasts?

'All models are wrong, some are useful'

George Box, Industrial statistician

One year with significant losses can drive market sentiment for some years after and occasional long-range forecasts that have failed to capture an event, for instance the statistical forecasts of the Atlantic hurricane seasons of 2005 and 2006, have led to some scepticism in the markets. Conversely, dynamical forecasts in 2005 and 2006 captured the swings of activity well, but again an individual year's forecast is not an indicator of the long-term reliability of the system.

The truth is that single-value forecasts are prone to over-interpretation by users giving the figure a high level of confidence and, as with any uncertain system, the use of probabilities is the only rational approach. Beyond the use of the statistics of overall skill and long-term performance, human expert physical understanding is needed to interpret whether the model results are credible.

An important aspect of live or operational long-range forecasting, as opposed to research studies, is that the human forecaster is a crucial component of the forecast process. There are numerous sources of observational and forecast data, including research studies and other forecasting systems - all with their own benefits and biases. Here, the role of the expert is to assist the user in understanding and using the right information for the decision in question.

In any one live case, there may be options for combining, weighting or excluding sources of information. This is not dissimilar to operational weather forecasting where forecasters are still required to decide whether to trust the model on a given occasion or not. A

[4] Original figure courtesy of the Korean Meteorological Administration
[5] http://www.wmo.int/pages/prog/wcp/wcasp/clips/producers_forecasts.html

human-free forecast is also available from these systems for transparency and robustness (see Met Office web site), but experience has shown that live analysis is needed to understand and apply climate forecast data to specific decisions[xi,xii].

The concept of skill is similar to form in horse racing. Previous forecasts are analysed to see how well they predict certain events, and these statistics are used to show the potential relevance of the latest unverified forecast (see Figure 5). Skill can be also be used in itself to calibrate or weight/combine forecasts. Skill measures can also be misleading at times as they rarely give a complete picture of performance for a particular forecast and a variety of different measures may be needed.

One specific limiting factor in measuring skill in long-range forecasts is that the coverage of subsurface ocean observations has improved so dramatically over the last few years that, due to the ARGO robotic float program, forecasts now benefiting from new data are potentially more skilful than hindcasts based on sparse observations. In short, present-day forecast skill may be underestimated by hindcasts. Additionally, current day observations of the earth system are absolutely core to our ability to predict the future.

As such, a mechanistic or process-based view of the forecast – seeing if the physical reasons for an outcome occurring are coherent - is also a valid method to explore whether a particular scenario is rational and trustworthy. This also allows real-time observations to be used to verify the unfolding forecast.

That said, skill scores are used to evaluate a system and determine what information to trust. Here a common deterministic approach is for the forecast ensemble mean to be used as a proxy for the forecast outcome. This can then be verified against observations by using simple statistical comparisons, such as correlations or standard error measures.

A correlation score is good for detecting changes in direction or phase – for example, more or fewer hurricanes. Anomaly correlation coefficients (ACC) then measure the variation of the forecast against the observed outcome over many cases, rather than measuring bias in, for example, whether a system forecasts consistently too many or too few storms. Note a system may be skilful at forecasting change and variations even with significant biases.

For forecast biases the root mean square error (RMS), the difference between the forecast and the observed outcome over many cases, is commonly used. Other, often highly complex, measures are used to characterise skill in a probabilistic way by examining how often the various possible outcomes in the forecast are realised in reality.

In the media, individual forecasts are often assessed as being right or wrong. Where a likelihood or probability forecast is produced however, many forecasts are necessarily needed for verification. For example, given a forecast of a 70% chance of an event, this event should happen 70% of the time in a well calibrated forecasting system over many similar forecasts. This means the converse should also happen 30% of the time. When the most likely event does occur, it does not imply forecast success or failure as this cannot be assessed from that single event alone. Success or failure is thus measured over many forecasts and not one.

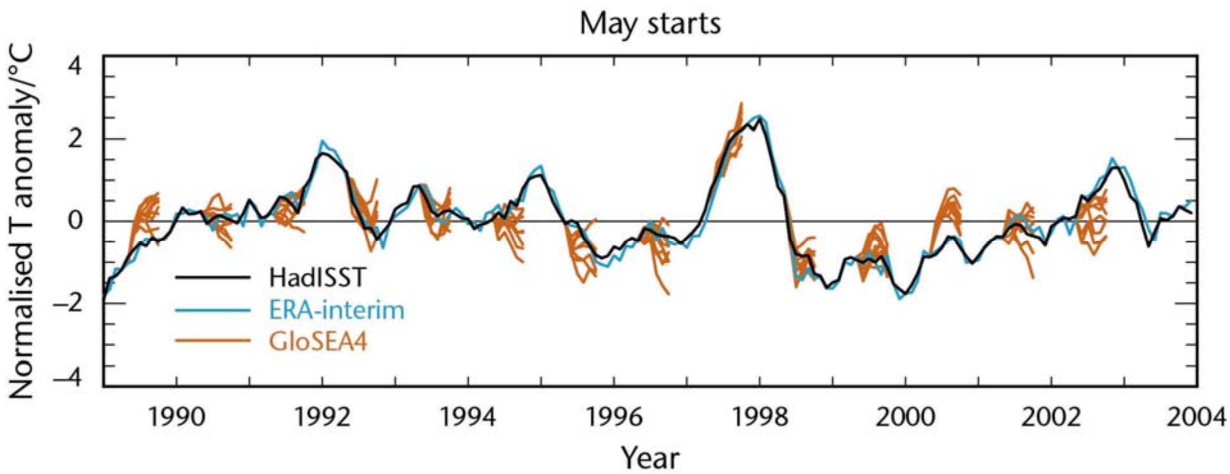

Figure 5: Forecast plumes showing ensemble members for central Pacific Sea Surface Temperatures (central Pacific Niño 3.4 region) compared to observations (HadISST and ECMWF ERA-Interim climatological analysis) relative to the long-term mean temperature. Individual forecasts are shown as orange lines for six month forecasts starting 1st May each year ii). The graph shows that the Pacific Ocean temperature can be forecast with some accuracy although there are occasional forecasts where the ensemble of individual forecasts doesn't span the observed outcome.

Skill varies depending on (1) start date of the forecast within the annual seasonal cycle; (2) forecast length (lead time); (3) location (some regions are easier to forecast than others) and (4) the phenomena being examined. Published deterministic forecast skill in state-of-the-art seasonal forecasting systems, excluding any live expert intervention, includes:

- **Pacific sea surface temperatures are highly[6] predictable to at least 6 months ahead:**
- Correlation (ACC) of 0.86 for central Pacific sea surface temperatures at a forecast lead time of six months[x].
- **Numbers of Atlantic tropical storms are highly predictable from before the beginning of the storm season (1 June):**
- Correlation (ACC) of 0.81 for tropical storm numbers in the Atlantic hurricane season (June to November) from 1st June start dates[ix] with an RMS error of 3.26 storms using a dynamical multi-model. Note CSU statistical forecasts for the same period scored ACC=0.39 with an RMS error of 4.76 storms (note CSU forecasts have now been updated). For forecasts from 1st May, an ACC of 0.62 was found using a single model[xiii].
- Numbers of Atlantic tropical storms have moderate predictability from the preceding November:
- Correlation (ACC) of 0.4 of predictions of numbers of tropical storms in the next Atlantic hurricane season from previous November[xiii].
- **Average numbers of Atlantic tropical storms over the next five-years are highly predictable:**
- Correlation (ACC) of 0.75 of predictions of 5-year mean numbers of tropical storms in the Atlantic hurricane season from previous November, compared to ACC=0.4 for using the last 5 years' mean activity[xiii]. These forecasts are statistically different at the 95% significance level.

All of these systems are more skilful than using past climatology as the forecast (significantly so in some cases) or persisting with current observed climate anomalies for the phenomena described. Several studies also point to the fact that skill is high in years with a significant event, such as El Niño or La Niña[iii]. As such, the above skill scores could be over-estimated for years where there is no such event, or underestimated for years when something significant is happening. This suggests the use of natural climate variations in addition to the more common statistical presentation, would enhance verification and communication to end users.

A layer of complexity in the analysis of forecast performance comes from forecasting probabilities (risk) rather than individual deterministic events. Probabilistic methods require that some likelihood of a category (e.g. cooler or warmer than average) or a threshold being transgressed (eg warmer than 26.5°C) is calculated. Probabilistic skill measures are inherently difficult to communicate and need many forecasts to populate the skill statistics. That said, they are arguably more useful to insurers as a more sophisticated audience.

Typically, probabilistic skill scores are unfamiliar to those not working in the probabilistic sciences and include, for example, the Brier Skill Score and the Relative Operating Characteristic. These scores also allow one to value the forecast in comparison to using, for example, climatology or another forecasting system. Detailed analysis exists for some phenomena using these scores adding a much more nuanced understanding of the skill of forecasting systems[x,ii].

Attempts have also been made to introduce more intuitive measures that relate information from a forecast and forecast value to more widely understood concepts, for example the ignorance measure corresponds to the expected returns of a gambler placing proportional bets on outcomes[xiv]. This measure asks whether you would have made more money using the forecast or not. This is a much more business-focussed statistical measure. Ideally, these skill scores approximate the overall value of the forecast system to something more familiar e.g. an interest rate used in banking[xv].

How these utility scores can be applied to risk-based forecasts where you only make a few decisions per year, rather than hedging over many decisions in, for example, energy commodity trading, remains a key question that needs exploration.

4.7 Hurricane focus: latest Atlantic tropical storm activity forecasts

Climate models forced by observed sea surface temperatures faithfully simulate the inter-annual variability of Atlantic tropical storm frequency[xvi], but actual predictions using dynamical models have previously been limited to a season ahead[ix].

Climate models are now yielding skilful regional forecasts beyond seasonal timescales. Skilful climate model predictions have been demonstrated well beyond the seasonal timescale for the first time by the Met Office Hadley Centre DePreSys system[xvii,xiii] – which won the Lloyd's Science of Risk Prize in 2010. Figure 6 shows that there is significant potential skill at forecasting Atlantic hurricane season activity rates at timescales out to seven years. This is beyond what can be achieved by using, for example, the average of the last five years as a forecast.

The study also shows that the increase in Atlantic hurricane activity since the 1970s and the decrease in the 1960s appear not to be caused by internal variability alone, but were at least partly externally forced by a combination of man-made changes in greenhouse gas, ozone and aerosol concentrations, and natural variations in solar irradiance and volcanic aerosol.

[6] Here, calibrated language is used to express long-range forecast skill in terms of correlation: high is greater or equal to r = 0.6, moderate is between 0.4 and 0.6, and low is less than 0.4

Initialisation of the Hadley Centre model with the observed state of the climate improves forecast skill, mainly through improved predictions of ocean conditions in the tropical Pacific and north Atlantic consistent with established modes of natural variability (see appendix 1).

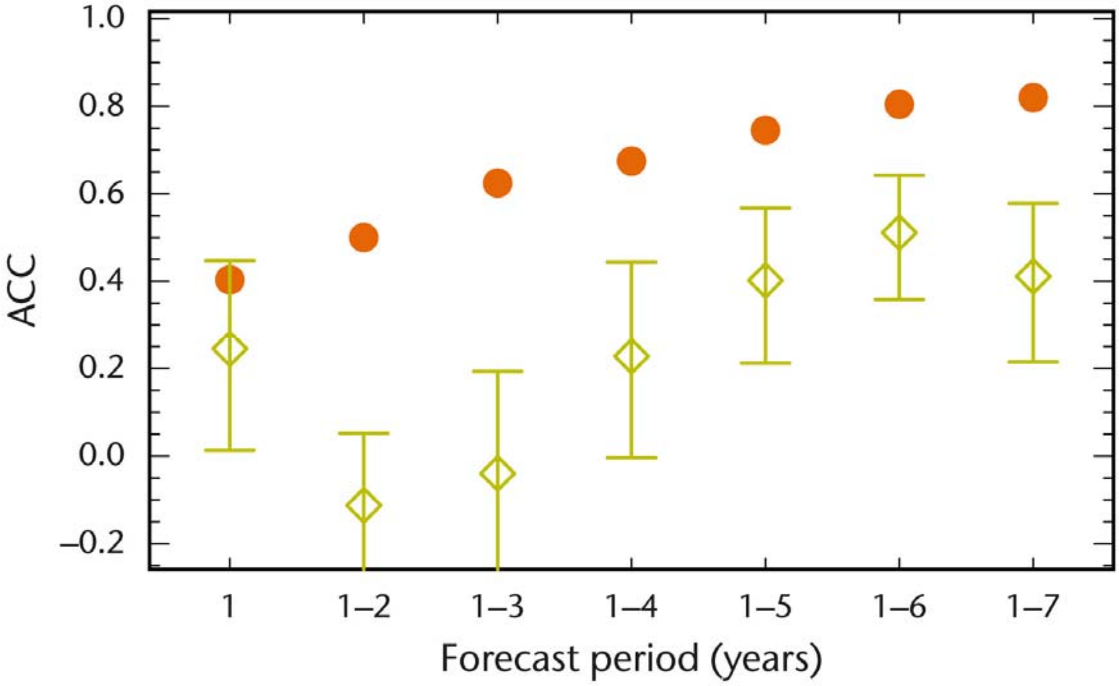

Figure 6: Skill of multi-year Atlantic tropical storm number forecasts showing the moderate to significant ability of dynamical models to predict tropical storm activity years ahead. Orange dots show the correlation of forecasts and the observed outcomes (Anomaly Correlation Coefficients where ACC=1.0 is perfect skill) plotted against the length of the forecast for re-forecasts for 1962-2005. Period 1 represents forecasts of the following June-November Atlantic tropical storm numbers from previous 1 November (one-year forecast). Period 1-2 represents the following two years' hurricane seasons storm numbers from 1 November (forecast covering following two years), and so on. Green lines show persistence forecast skill for comparison ie forecast period 1-5 is compared with the previous five years' mean storm numbers[xiii]. For interpretation, one can see that the numerical forecasts out-perform forecasts that say the near future is the same as the near past at all forecast ranges, and the overall skill improves the longer the forecast range.

4.8 Future potential and insurance relevance

Active research at global forecast centres around the world to develop improved near-term climate predictions and the next Intergovernmental Panel on Climate Change assessment means that there will be rapid developments in the science of inter-annual and decadal prediction in the coming years and increased forecast data availability.

Supercomputing is a significant limiting factor due to the competing needs for granularity, quantified uncertainty, physical process complexity and observation data assimilation. Due to the vast scope of science needed, it is unlikely that any one commercial company or governmental organisation will be able to implement such a system without significant collaboration.

One relevant question is: how good will forecasts be in five or ten years' time? This is not easy to answer and is dependant on increased scientific understanding of fundamental processes in the climate system, supercomputing resource and the development of user-relevant impact models and tools to aid decision making. The science objectives needed to fulfil the aim of forecasting catastrophic hazards at long range include:

- Understanding of the key physical process and modes of natural variability leading to more skilful predictions on 1-to-10 year timescales.
- Use of the latest high resolution (~50km to ~1km) coupled climate models that will resolve more physical processes directly - for example air-sea interaction, hail generated by thunderstorms and the physical processes related to tropical storm intensity.
- Increased understanding of connections between different perils and hazards, from large-scale phenomena, such as El Niño, down to small-scale hazards, such as tornadoes.
- Exploration of more relevant forecast parameters, such as landfalling of tropical storms rather than activity rates.
- Inclusion of impact models that link flood and crop models to underpinning meteorology.
- Exploration of the performance skill of hazard predictions using measures more relevant to the end users, including the insurance industry.
- Exploration of the impact of major volcanic eruptions and changes in the solar cycle on meteorological hazards.
- Clearer understanding of the impact and interaction of both natural and man-made climate change on damaging hazards and extreme climatic events.

With these developments in mind, the operational capability to forecast changes in financial loss-rates of major peril classes (including hurricanes, winter storms, floods and fire) is likely to increase and will lead to changes in pricing methods and market strategies in terms of volume and territories covered. Links between different peril classes will also become more apparent as different hazards in different locations are linked by large-scale climatic drivers and may allow reinsurances to manage risk across peril classes and within wider portfolios of risk (enterprise risk management).

5 Conclusions

Forecasts of loss-making hazards in the near term climate are relevant to the insurance industry. To a large degree the industry does not currently integrate long-range forecasts from statistical or dynamical forecast models into core risk pricing decisions. It does appear Atlantic tropical storm forecasts are at least monitored for general market context. Should they now be proven to be skilful and provide more relevant information (for example, the likelihood of landfalling tropical storms), in a way that better meets end-users' needs, long-range forecasts should be useful to the insurance industry for managing risk exposure and ensuring profitable business. New opportunities await those that exploit this developing science.

The science itself is predicated on the ability of dynamical climate models to represent near-term climate and especially modes of natural variation, such as El Niño. The studies now appearing suggest that there is an increasing ability to capture these natural shifts and provide insurance-relevant forecasts of some extreme phenomena. Further development in the models, their initialisation and the exploration of uncertainty will likely demand significantly larger computing resources. The relative benefit of this investment to the insurance industry, as well as disaster risk reduction and other markets, could also be significant if the forecasts are proven to be reliable.

Development in long-range forecasting technology is probably twenty years behind weather forecasting. However, it offers unique potential in catastrophic risk management. This uniqueness could be keenly felt in at least two scenarios where the past does not represent the future; should we experience significant cooling for a few years as a result of a massive volcanic eruption, or significant warming due to man-made climate change. Under such scenarios, it may be the only technology available to understand future risk.

6 Appendix 1 - Natural climate variability modes

The potential for skilful seasonal to decadal predictions depends largely on whether models simulate sufficient inter-annual and decadal variability, both in terms of magnitude and structure (i.e. where, when and how much). Climate records reveal several large-scale modes of variability intrinsic to the natural climate system with timescales of months, years and decades which, if captured in prediction systems, would contribute to the enhanced prediction of climate impacts around the globe.

In tropical regions the Madden-Julian Oscillation (MJO), also known as the '40-day wave' and 'intra-seasonal wave' is principally a west-east propagating (zonal) large-scale atmospheric wave. Its passage through the Indo-Pacific region disrupts regional rainfall, therefore influencing break periods in the Asian and Australian monsoons. Although irregular in behaviour, it is sufficiently predictable at a monthly range for forecasts to be of practical use[xviii].

Outside tropical regions inter-annual variability is characterised by large-scale atmospheric pressure patterns. The North Atlantic Oscillation (NAO[xix]) has connections to European and eastern North America climate variability on seasonal and longer timescales, and has some seasonal predictability that is associated with variations in North Atlantic sea surface temperature. The NAO is the strongest mode of inter-annual variability in the Northern Hemisphere mid-latitudes in both winter and summer. The magnitude of the NAO patterns in each winter is known as the NAO index, which is most often expressed as the difference in average sea level pressure between the Azores and Iceland. A positive NAO results in low pressure over Iceland and high pressure over the Azores and is associated with stronger westerly winds. Opposite impacts tend to occur during the negative phase of the NAO.

The NAO is a very important driver of European climate, for example, for the trajectory and intensity of European winter storms and has also been linked to the steering of Atlantic tropical storms. The North Atlantic extra-tropical storm track tends to shift north when the NAO is positive, giving more northern European wind storms. The storm track shifts south during negative NAO conditions. There are more extreme North Atlantic cyclones during positive NAO and fewer during negative NAO and most central European storms occur during positive NAO.

The most important mode of climate variability on seasonal timescales is the El Niño-Southern Oscillation (ENSO), associated with the warming and cooling of the surface of the tropical Pacific Ocean over a 3-8 year period. The term El Niño event is used for the widespread anomalous warming of equatorial Pacific sea surface temperature, and ENSO is an irregular cycle of warm (El Niño) and cold events (La Nina) that last several months. Through atmospheric dynamical effects (teleconnections), ENSO events have globally widespread effects on temperature and rainfall[xx], and substantial socio-economic impacts.

Within the last ten years it has been recognised that an east-west contrast in Indian Ocean sea surface temperature anomalies (the Indian Ocean Dipole) is another influential and potentially predictable climate variability feature[xxi].

The Pacific exhibits a pattern of longer time-scale variability known as either the Pacific Decadal Oscillation (PDO) or the Inter-decadal Pacific Oscillation (IPO), which appears to be related to the frequency of ENSO phases and seems to modulate ENSO impacts, such as flood and drought. Likewise in the North Atlantic Ocean strong variability with a phase of three to four decades has been observed, and is referred to as the Atlantic Multi-decadal Oscillation (AMO). Decadal variations in hurricane activity have been linked to the multi-decadal sea surface temperature variability in the Atlantic[xxii]; something which has been simulated well by numerous models[xxiii,xxiv,xxv].

The physical mechanisms of some of the multi-decadal modes of variability (PDO/IPO and AMO) are still being clarified by research and as such their long-term predictability is not certain. Atlantic processes related to the oceanic overturning circulation might provide the necessary AMO feedback mechanisms, although Pacific processes tend to suggest the PDO/IPO is an integration of shorter-period variability with lower predictability. Regardless of the longer-term predictability of these modes, the persistence of their anomalies is sufficiently long that the correct initialisation of AMO and PDO/IPO phases in a climate model at the outset is vital for predictions up to a decade ahead.

While all these modes have multi-year 'memory' due to the large heat capacity of the ocean, there is one purely atmospheric mode that can provide multi-annual predictability. The Quasi-Biennial Oscillation (QBO) is an oscillation with a period of approximately 28 months in the equatorial stratospheric winds. It is highly predictable for up to a few years ahead and has been shown to affect Northern Hemisphere winter climate[xxvi].

In addition to variability with solely internal origins within the climate system, decadal predictability can also arise from forced changes[7]. The solar cycle, with a period of around 11 years, is a clear example. Although the magnitude of solar cycle effects is still uncertain even at the global scale, first steps are now being taken towards identifying regional responses.

Step changes in climate, such as those from volcanic eruptions can lead to climatic perturbations for several years and can result in the failure of statistical long-range forecasts. Fortunately, it appears that if the size and characteristics of an eruption can be identified, then it is possible to simulate much of its climatic impact. This suggests that existing decadal forecasts could be revised soon after an eruption.

Finally, while the underlying long-term trend from anthropogenic climate change is currently small compared to year-to-year variability, it begins to add predictive value for periods approaching a decade or more and may already have had significant impact on the relative

[7] See http://www.lloyds.com/News-and-Insight/360-Risk-Insight/Research-and-Reports/Space/Space-Weather for discussion

frequency of extreme events[xxvii]. The impacts of anthropogenic climate change on natural modes of climate, such as El Niño, is an area of active research, but no clear consensus regarding any natural mode of variability has been reached.

Finally, as a cautionary note, geological records do show extremely rare examples of abrupt natural climate change. An example is the Younger Dryas de-glaciation period, which terminated abruptly ~11,500 years ago with a warming of ~15°C in Greenland, half of which is believed to have occurred over as little as 15 years (essentially bringing global temperatures in line with those experienced today.) Accompanying this and other abrupt jumps in temperature were rapid shifts within diverse components of the earth system, ranging from East African aridity to Western European storminess[xxviii].

7 Appendix 2 – Global forecasting systems

Below is a description of the major aspects that describe a global forecasting system with examples from the Met Office GloSea4 system.

1 Resolution

State-of-the art coupled general circulation climate models that include weather forecasting atmospheric models are now being used operationally for seasonal forecasting, with the ideal of capturing realistic weather patterns within long-range forecasts. The latest UK Met Office seasonal forecasts system (GloSea4[ii]) uses:

- An atmospheric grid of approximately 120km horizontally (in mid-latitudes) with 85 vertical levels up to 85km altitude.
- An oceanic grid of 1 degree in latitude and longitude, and 1/3 degree at the equator with 75 vertical levels from the sea surface to the ocean floor.
- Interactive sea ice and land surface schemes with associated resolution constraints.

Operational five day global weather forecasts are currently run deterministically at 25km resolution and regional weather forecasts at resolutions up to 1km and finer. The long forecasts and large number of ensemble members required in long-range forecasting however limits the resolution of seasonal and decadal forecasts. For a doubling of horizontal resolution in both latitude and longitude, an increase of a factor of about 8 in computing power is required. This means that some important physical processes, such as the north Atlantic drift in the ocean and the structure of hurricanes, can only be partially resolved under current computational resources.

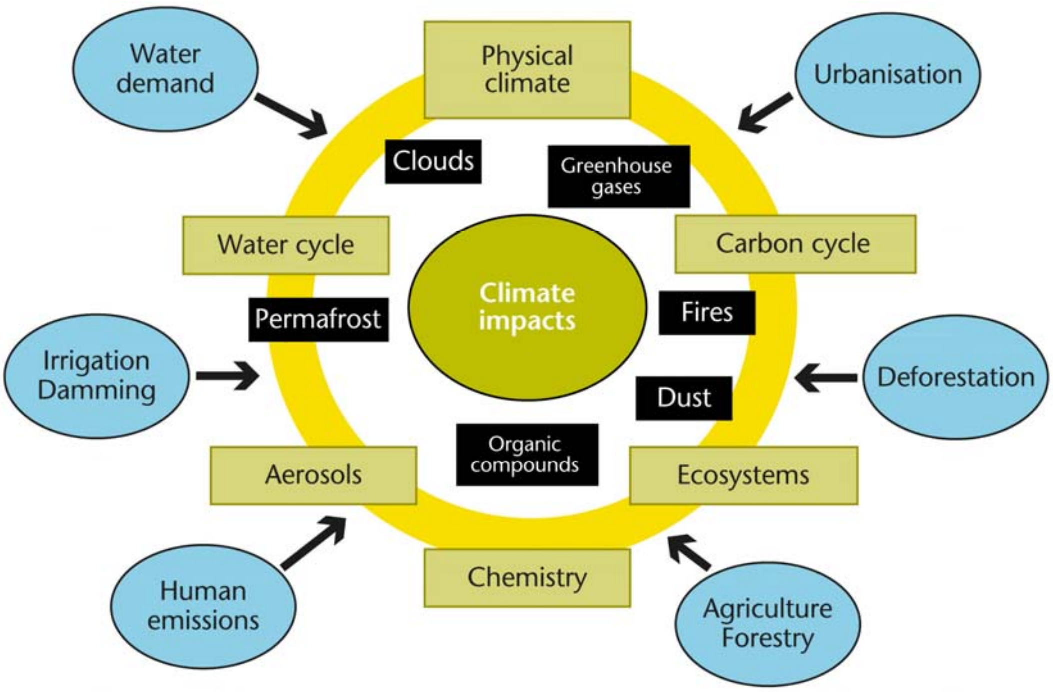

Figure 7: Complexity of the climate system. The physical ocean-atmospheric system is influenced by a wide range of other interactions operating over a range of timescales.

2 Complexity

Additional sub-system complexity increases computational demand. Excluding processes known to impact the weather and climate means that significant risks may not be captured (similar to externalities in economics). Additional costs come from factors such as: the inclusion of the stratosphere, ocean/land surface biology, interactive aerosols, including dust, and the exchange of fluxes of heat and momentum between the sub-systems at more frequent intervals to allow more complex physical interactions. Within the computer resources available, forecasters include as many of the processes as possible. See Figure 7 above.

GloSea4 is among the most complex forecasting systems used for long-range forecasting globally and includes sea ice, ozone and aerosol parameterisations. The model now includes the stratosphere (which has been shown to be important for European winter predictability[xxvi]) and an increased ocean surface layer resolution to improve ocean-atmospheric heat exchange.

3 Number of forecasts in the ensemble

To detect a change in the likelihood of an event many realisations of the future are needed. The size of the ensemble needed depends on how visible the events that one is looking for are - for example, a global scale El Niño compared to a rare clustering of winter storms. This is the so-called signal to noise ratio problem where one needs much more forecast information to detect the likelihood of rarer events. A larger ensemble would be needed to resolve the relatively noisy North Atlantic (extra-tropics) in comparison to changes in the temperature of the Pacific for El Niño forecasting for instance. Currently, GloSea4 uses forecasts initialised every day, allowing collective monthly forecasts for the next six months using 42 forecast ensemble members.

The ensemble is carefully structured to account for as many sources of uncertainty in the forecast as possible (see figure 2). These include: initial condition uncertainty (errors in our knowledge of today's weather and climatic state), boundary conditions and uncertainty and errors in the model formulation.

For GloSea4, the ensemble members sample uncertainties in initial conditions by using different start dates and uncertainties in model physics.

Large-scale international projects to produce multi-model ensembles have also been used to explore risk based forecasting. These combine forecasts produced at different forecasting centres using different models to sample a wider range of uncertainties and give a larger ensemble size and increase forecast confidence[x].

4 Re-forecasts and calibration

All climate models have their own preferred climate. This means the models can have biases compared to our current observed climate that the model prefers to maintain. To understand their performance, calibrate biases and benchmark the forecasts, retrospective re-forecasts or hindcasts are used. These are run over extensive historical periods as if they were run in real time i.e. with no prior knowledge of the future.

The GloSea4 system uses 14 years of hindcasts run currently from 1996 to 2009 with a reduced ensemble size (see section on skill.)

5 Forecast initialisation

As with weather forecasts, the entire earth system represented in the model needs initial conditions to start from. Whilst those in the lower atmosphere will soon be lost to chaos, the deep ocean and land surface will maintain a memory of the initial conditions for a longer period into the forecast.

Observations are the basis of this step and their location and technology varies significantly with time – not least of all with the introduction of satellites in the 1970s and with the robotic ARGO[8] array in the ocean since 2000[xxix].

The technology to generate global gridded nowcast analyses ranges from simple statistical formulations to fantastically complicated systems that iterate the numerical forecast model back and forward in time; the whole process is called data-assimilation. The cost of these systems can vary considerably and they also demonstrate varying benefits on forecast skill. Some forecast systems also use anomaly initialisation – a process of providing the forecast system with initial differences from normal climate rather than absolute observations[xvii]. This is used to overcome forecast drift as the model tries to get back to its own inherent climate. Coupled data assimilation (where one analyses both atmospheric and oceanic data at the same time) is a nascent but much-needed technology.

The GloSea4 system uses the high-order (4DVAR) analysed state of the atmosphere from operational global 25km weather forecasts to initialise the atmosphere and an optimal-interpolation initialisation of the ocean data using temperature and salinity. Climate forcings (CO2) follow observations and then IPCC A1B scenarios from 2000 and ozone and atmospheric aerosols are fixed at climatological values including a seasonal cycle. Observed solar variability is specified in the hindcasts and an 11-year cycle is included in the forecasts.

[8] ARGO is an army of ~3200 robotic ocean buoys that sample the vertical profile of temperature and salinity as they float around the world's oceans. See http://argo.jcommops.org/

8 references

[i] Lonfat, M, A. Boissonnade and R. Muir-Wood, 2007. Atlantic basin, U.S. and Caribbean landfall activity rates over the 2006–2010 period: an insurance industry perspective. Tellus A Volume 59, Issue 4, pages 499–510, August 2007.

[ii] Arribas, A., M. Glover, A. Maidens, K. Peterson, M. Gordon, C., MacLachlan, R. Graham, D. Fereday, J. Camp, A.A. Scaife, P. Xavier, P. McLean, A. Colman, S. Cusack 2011. The GloSea4 ensemble prediction system for seasonal forecasting. Mon. Wea. Rev., accepted for publication.

[iii] Jin E. K., James L. Kinter III, B. Wang, C.-K. Park, I.-S. Kang, B. P. Kirtman, J.-S. Kug, A. Kumar, J.-J. Luo, J. Schemm, J. Shukla and T. Yamagata, 2008: Current status of ENSO prediction skill in coupled ocean–atmosphere models. Clim. Dyn., 31, 647–666.

[iv] Dailey, P., M.R. Huddleston, S. Brown and D. Fasking, 2009. The Financial Risks of Climate Change, ABI Research Paper No 19: Examining the financial implications of climate change using climate models and insurance catastrophe risk models. Publisher, ABI, 108 pp.

[v] Ineson, S., A. A. Scaife, 2009. The role of the stratosphere in the European climate response to El Niño. Nature Geoscience, 2, 32-36, doi:10.1038/ngeo381

[vi] Klotzbach, P. J., W. M. Gray, 2009. Twenty-five years of Atlantic basin seasonal hurricane forecasts. Geophys. Res. Lett., 36, L09711, doi:10.1029/2009GL037580.

[vii] Meehl G.A., L. Goddard, J. Murphy, R.J. Stouffer, G. Boer, G. Danabasoglu, K. Dixon, M.A. Giorgetta, A.M. Greene, E. Hawkins, G. Hegerl, D. Karoly, N. Keenlyside, M. Kimoto, B. Kirtman, A. Navarra, R. Pulwarty, D. Smith, D. Stammer and T. Stockdale, 2009. Decadal Prediction, Can It Be Skillful? Am Met Soc, 90, ISS 10, Pg 1467-1484, DOI:10.1175/2009BAM2778.1

[viii] Palmer, T.N., A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, M. Déqué, E. Díez, F.J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, M. C. Thomson, 2004. Development of a European multi-model ensemble system for seasonal to inter-annual prediction (DEMETER). Bulletin of the American Meteorological Society, 85, 853-872.

[ix] Vitart, F. M.R. Huddleston, M. Déqué, D. Peake, T.N. Palmer, T.N. Stockdale, M.K. Davey, S. Ineson and A. Weisheimer, 2007. Dynamically-based seasonal forecasts of Atlantic tropical storm activity issued in June by EUROSIP. *Geophys. Res. Letts,* **34***,* L16815, doi: 10.1029/2007GLO30740

[x] Weisheimer, A., F.J. Doblas-Reyes, T.N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, P. Rogel, 2009. ENSEMBLES - a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs *Geophys. Res. Letters,* Vol. 36, No. 21.

[xi] Graham R. J., C. Gordon, M. R. Huddleston, M. Davey, W. Norton, A. Colman, A. A. Scaife, A. Brookshaw, B. Ingleby, P. McLean, S. Cusack, E. McCallum, W. Elliott, K. Groves, D. Cotgrove, D. Robinson, 2006. The 2005/06 winter in Europe and the United Kingdom: Part 1 - How the Met Office forecast was produced and communicated. *Weather*, 61, 326 – 362.

[xii] Folland C. K., D. E. Parker, A. A. Scaife, J. J. Kennedy, A. W. Colman, A. Brookshaw, S. Cusack, M. R. Huddleston, 2006. The 2005/06 winter in Europe and the United Kingdom: Part 2 - Prediction techniques and their assessment against observations. Weather, 61, 337-346.

[xiii] Smith, D. M., Eade, R., Dunstone, N. J., Fereday, D., Murphy, J. M., Pohlmann, H., and Scaife, A. A., 2010. Skilful multi-year predictions of Atlantic hurricane frequency. - Nature Geoscience, 3, 846-849, doi:10.1038/ngeo1004.

[xiv] Roulston, M. S., & Smith, L. A. 2002. Evaluating probabilistic forecasts using information theory. *Monthly Weather Review, 130*(6), 1653--1660.

[xv] Hagedorn, Renate and Smith, Leonard A., 2009. Communicating the value of probabilistic forecasts with Weather Roulette. *Meteorological applications*, 16 (2). pp. 143-155. ISSN 1350-4827

[xvi] Zhao, M., Held, I. M., Lin, S.-J. & Vecchi, G. A, 2009. Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50km resolution GCM. *J. Climate* **22**, 6653.6678.

[xvii] Smith D., S. Cusack, A. Colman, C. Folland, G.Harris and J. Murphy, 2007. Improved surface temperature prediction for the coming decade from a global climate model. *Science,* **317**, 796-799

[xviii] Lau WKM, Waliser DE. 2005. Intraseasonal Variability in the Atmosphere–Ocean Climate System. Springer: Chichester, UK

[xix] Hurrell, J. W., and C. Deser, 2009. North Atlantic climate variability: The role of the North Atlantic Oscillation. *J. Mar. Syst.*, 78, No. 1, 28-41, DOI:10.1016/j.jmarsys.2008.11.026.

[xx] Ropelewski, C.F. and M.S. Halpert, 1987. Global and regional scale precipitation patterns associated with El Niño/Southern Oscillation, Mon. Wea. Rev. 115:1606-1626.

[xxi] Luo, J.-J., S. Behera, Y. Masumoto, H. Sakuma, and T. Yamagata, 2008. Successful

prediction of the consecutive IOD in 2006 and 2007. *Geophys. Res. Lett.*, 35, L14S02, doi:10.1029/2007GL032793.

[xxii] Goldenberg, S. B., et al., 2001. The recent increase in Atlantic hurricane activity: Causes and implications. *Science*, **293**, 474– 479

[xxiii] Delworth, T., S. Manabe, and R. J. Stouffer 1993. Interdecadal variations of the thermohaline circulation in a coupled ocean-atmosphere model, J. Clim., 6, 1993– 2011.

[xxiv] Latif, M., 1998. Dynamics of interdecadal variability in coupled ocean–atmosphere models. *J. Climate,* **11,** 602–624.

[xxv] Knight, J. R., R. J. Allan, C. K. Folland, M. Vellinga, and M. E. Mann 2005. A signature of persistent natural thermohaline circulation cycles in observed climate, *Geophys. Res. Lett.*, *32*, L20708, doi:10.1029/2005GL024233.

[xxvi] Marshall, A. G. and A. A. Scaife, 2009. Impact of the QBO on surface winter climate. Journal of Geophysical Research (Atmospheres), 114, 18 110–+, doi:10.1029/2009JD011737.

[xxvii] Stott, P. A., D. A. Stone, and M. R. Allen, 2004. Human contribution to the European heatwave of 2003, *Nature*, *432*, 610–614

[xxviii] McNeal, D.J, P.R. Halloran, P.Good and R.A.Betts, 2010. Analysing Abrupt and Nonlinear Climate Changes and their Impacts. Submitted to Wiley Interdisciplinary Reviews: Climate Change.

[xxix] Ingleby, B., M. Huddleston, 2007. Quality control of ocean temperature and salinity profiles - historical and real-time data. Journal of Marine Systems, 65, 158-175 10.1016/j.jmarsys.2005.11.019