

??

# WEATHER AND CLIMATE FORECASTING

Mark S. Roulston and Leonard A. Smith

University of Oxford and London School of Economics

## History of Forecasting: Statistics vs Dynamics

*"When it is evening ye say, it will be fair weather for the sky is red. And in the morning, it will be foul weather today for the sky is red and lowering." (Matthew Ch. 16 v. 2)*

The Biblical quotation above is an ancient example of attempting to make meteorological predictions based on empirical observations. The idea it contains tends to work as a forecasting scheme, and it has been expressed in many different ways in the subsequent 2,000 years. It is one of many sayings and rhymes of folklore that provide a qualitative description of the weather. In the 17th century, the introduction of instruments to measure atmospheric variables meant that meteorology became a quantitative science.

By the 19th century meteorological data was being collected all over Europe. It was during the 1800s that many scientists began lamenting the fact that the collection of meteorological data had far outpaced attempts to analyse and understand this data (Lempfert 1932). Attempts had been to find patterns in tables of meteorological, but these had usually ended in failure. One example is the efforts to link weather to celestial motions made by the Palatine Meteorological Society of Mannheim (the World's first such society) in the late 18th century.

The breakthrough came when scientists began plotting the growing meteorological database on maps; the tendency of mid-latitude low-pressure systems to advance eastward was discovered: clouds in the west provide a red sky at dawn before bringing bad weather, clouds in the east redden the sky at sunset before making way for clear skies. The discovery of mid-latitude eastward flow not only explained the success of ancient folk wisdom, combined with the telegraph it also provided a means to extend predictability beyond the horizon. The late 19th and early 20th century saw many efforts to identify cycles in time series of weather data. One the most prolific cycle seekers, and ultimately one of the few successful ones, was Sir Gilbert Walker. Walker collated global weather data and spent years seeking correlations (Walker 1923, 1924, 1928, 1937, Walker and Bliss 1932). During this work he discovered that the pressure in Tahiti and in Darwin were anti-correlated. This discovery withstood the subsequent tests and is called the "Southern Oscillation". It is now known to be the atmospheric component of El Nino-Southern Oscillation climate phenomenon.

Many other claims for the existence of weather cycles were made, but few withstood the scrutiny of statistical tests of robustness. By the 1930s, mainstream meteorology had largely given up attempting to forecast the state of the atmosphere using statistical approaches based solely on data (Nebeker 1995). Since then, estimates of the *recurrence* time of the atmosphere have implied that globally

"similar" atmospheric states can only be expected to repeat on time scales far greater than the age of the Earth (van den Dool, 1994). This result suggests that there will *never* be enough data available to construct pure data-based forecast models (except for very short lead times for which only the local state of the atmosphere is important).

Towards the end of the 19th century a growing number of scientists took the view that the behaviour of the atmosphere could be modelled from first principles, that is, using the laws of physics. The leading proponent of this view was a Norwegian physicist called Vilhelm Bjerknes. Bjerknes believed that the problem of predicting the future evolution of the atmosphere could be formulated mathematically in terms of seven variables: three components of air velocity, pressure, temperature, density and humidity – each of these variables being a function of space and time. Furthermore, using the established laws of dynamics and thermodynamics, a differential equation could be formulated for each of these seven quantities (Richardson 1922).

The equations describe the flows of mass, momentum, energy and water vapour. These equations, however, form a set of non-linear partial differential equations (PDEs) and so an analytic solution was out of the question. Meanwhile, the First World War raged and Briton Lewis Fry Richardson was developing a scheme for solving Bjerknes' equations of atmospheric motion.<sup>1</sup>

In 1911, Richardson developed a method to obtain approximate solutions to PDE's. The method involved approximating infinitesimal differences as finite differences, that is dividing space into a finite number of grid boxes and assuming the variables are uniform within each grid box. The solution obtained is not exact, but becomes more accurate as the number of grid boxes increases. Richardson applied his approach to the atmospheric equations, producing a set of finite difference equations that could be solved by straightforward arithmetic calculations.

The difficulty of Richardson's achievement cannot be understated. In creating his scheme for numerical weather prediction (NWP), he had to ensure the problem was formulated in terms of quantities that could be measured, sometimes developing new methods of measurement when they were required. He also had to develop a way of dealing with turbulence. He related vertical transfer of heat and moisture to the vertical stability of the atmosphere, as measured by a dimensionless quantity now called the Richardson number. Richardson was truly years ahead of his time – and therein lay the problem. Although his recipe was straightforward, it was also incredibly tedious. It took him six weeks to produce a single six-hour forecast for just two European grid points! Richardson envisaged an army of clerks doing the calculations that would be necessary to generate an operational forecast, but this did not happen.

The Second World War saw the invention of the digital computer. After the War, the computer pioneer, John von Neumann, was trying to persuade the US government of the usefulness of this new device. Though not a meteorologist, von Neumann identified weather forecasting as an ideal application to demonstrate the power of the computer. Thanks to Richardson, the problem had been formulated in an algorithm that could be executed by a computer, but was impractical to do without one. Furthermore, the potential benefits of successful weather forecasting could be appreciated by laymen, generals and politicians. In 1950, von Neumann's team, led by meteorologist Jules Charney, ran the first numerical weather prediction program on the ENIAC computer. Thus began an intimate relationship between meteorology and the leading edge of computer science – a relationship that continues to this day. See Nebeker (1995) for a comprehensive account of the history of weather forecasting.

Since we never know the precise state of the atmosphere, it would be foolhardy to expect to produce a precise forecast of its future state. This has motivated operational forecast centres to develop probability forecasts, a set of possible



outcomes based on slightly different views of the current state of the atmosphere. These probability forecasts come closer to the type of information required for effective risk management and the pricing of weather derivatives.

### Modern Numerical Forecasting

Weather forecasting can be divided, somewhat arbitrarily, into three categories: short-, medium- and long-range. In the context of this chapter, short-range forecasting refers to forecasting the weather over the next day or two. Medium-range forecasting covers lead times of three days to about two weeks, while long-range, or seasonal, forecasting aims to predict the weather at lead times of a month or more. While short and medium-range forecasts are valuable to many users, including energy companies in planning their operations, it is seasonal forecasts that are of most interest to the weather derivatives markets.

Nowadays it is increasingly common for seasonal forecasts to be made using computer models which are essentially the same as those used to produce tomorrow's forecasts. In this section we shall outline the key features of these models so that the reader will have some understanding of the origin of forecast products. We also hope to familiarise the reader with some of the technical language that meteorologists use to describe their models.

Short-range forecasts can be made using limited area models. These models use grid boxes to cover restricted parts of the globe. In a period of two weeks, however, weather systems can travel halfway round the globe. Therefore forecasting in the medium range or beyond requires a global model of the atmosphere. The most advanced global model used for operational forecasting belongs to the European Centre for Medium Range Weather Forecasting (ECMWF) funded by 19 countries and based in Reading, UK.<sup>2</sup> Today, ECMWF makes daily forecasts out to 10 days. These forecasts are distributed through the national meteorological offices of the member countries.

The current ECMWF global model is a T511 spectral model (see Panel 1); equivalent to a horizontal resolution of 40 km, with 60 vertical levels. The complete model state, at a given time, is described by approximately 10 million individual variables. The state is evolved forward in time by taking time steps of about 10

#### PANEL 1

### COMPUTER MODELS: GRID POINT VS SPECTRAL POINT

Computer models of the atmosphere come in two types: grid points and spectral points. Grid point models represent the atmosphere as finite boxes centred on a grid point and calculate the changes in mass, energy and momentum at each grid point as a function of time. The size of the boxes determines the *resolution* of the model.

Spectral models represent the distribution of atmospheric properties as sums of *spherical harmonics*. These harmonics are similar to sine and cosine functions, except they are two dimensional and exist on the surface of a sphere. The resolution of a spectral model is determined by the wavelength of the highest spherical harmonic used in the model. Spectral models are denoted by labels such as "T511"; this means that the highest harmonic used has 511 waves around each line of latitude. T511 model has a horizontal resolution of  $1/(2 \times 511)$  times the radius of the Earth (about 40 km). The state of the atmosphere, as represented in a spectral model, can be converted to a grid-point representation using a mathematical transformation. Both grid-point and spectral models divide the atmosphere vertically into layers.

minutes. The evolution of the atmosphere is thus represented as a trajectory in the ultra-high dimensional state space of the model.<sup>2</sup>

The approximations introduced by representing continuous fields on a finite grid are, in a sense, well-defined or mathematical approximations, sometimes referred to as “errors of representation”. There are, however, another set of physical approximations that all numerical weather models contain. These approximations are called “parameterisations”. Numerical models have a finite spatial resolution. As mentioned above, the ECMWF global model cannot represent the details of weather, or of topography, at scales less than 40 km. It would be nice to be able to know the distribution of rainfall on a much smaller scale, but even having information on weather averaged over tens of kilometres can be very useful.

Even if one is content to accept weather forecasts averaged over relatively large regions, weather on smaller “sub-grid” scales can have a profound impact on the weather at larger scales. For example, thunderstorms are too small for global models to resolve, but the convection and rainfall associated with them has a major impact on the energy balance of the atmosphere, and consequently on the weather over an area much larger than the storm itself. Therefore, the impact of sub-grid processes must be parameterised and included in the equations that describe the evolution of the atmosphere at the larger scales. Essentially, a parameterisation scheme for cumulus convection must predict the amount of convection in a grid box purely as a function of the meteorological variables averaged over the grid box (and possibly surrounding grid boxes), and then predict the affect that this amount of convection will have on the time evolution of those meteorological variables. Parameterisation schemes are usually designed based partly on a physical understanding of the processes involved and partly on the study of empirical observations. There are many processes that must be parameterised in numerical models of the atmosphere, eg, surface evaporation, drag due to topography and sub-grid turbulence. Designing better parameterisation schemes is one of the most active and important areas of modern meteorological research.

#### DATA ASSIMILATION

Before a forecast can be made with a numerical model of the atmosphere, the current state of the atmosphere, as represented within the model, must be estimated. The process of using observations to make this estimate of the model’s initial condition

### PANEL 2

## OBSERVING THE ATMOSPHERE

The World Meteorological Organization’s World Weather Watch (WMO WWW) manages the data from 10,000 land stations, 7,000 ship stations and 300 buoys fitted with automatic weather stations. All these stations are maintained by National Meteorological Centres.

The most important development in meteorological observation in the last 40 years has been the advent of weather satellites. The WMO WWW incorporates data from a constellation of nine weather satellites that provide global coverage. It is a testimony to the importance of these satellites that weather forecasts now tend to be more skilful in the southern hemisphere, where surface stations are quite sparse, than in the more densely observed northern hemisphere. The fact that southern hemisphere skill is actually slightly *higher* is probably because there is less land south of the equator. Topography and other continental effects make the job of forecasting northern hemisphere weather harder, even with the greater data coverage in the North.



is called "data assimilation". The estimate of the state of the atmosphere derived from data assimilation is called the analysis.

To initialise the model, one must effectively know the value of all the relevant meteorological variables as represented on the grid points of the model. Even with the vast amounts of weather data that are collected every day (see Panel 2) there are still massive gaps in the observational data set. The simplest approach to overcoming these gaps is interpolation, as Richardson did in his early in numerical forecasting experiments (Richardson 1922). More sophisticated approaches actually combine the observations with the knowledge of atmospheric dynamics that is contained in the numerical model itself. Any state of the model can be converted into an estimate of the observations that would result if the atmosphere were in that particular state, by using the "observation function".

At ECMWF a data assimilation technique called "variational assimilation" is used. This method involves trying to find a model trajectory that leads to the closest match of the model to the actual observations that were made. The trajectory of the model is the path the model state traces out in time within the state space of the model. This space is a high dimensional space, defined by the millions of variables that describe the model state. Optimisation by variational assimilation is performed by trying to find the state of the model that leads to the best match with observations over the subsequent assimilation period. The state of the model somewhere in the middle of the assimilation period is then used as the analysis with which to initialise the forecast. A lower resolution, T159 model, is used for the data assimilation.

Data assimilation is also used by forecasting centres to produce reanalysis products. These products are historical reconstructions of the state of the atmosphere, projected into the model grid-point representation. They are constructed using similar techniques that are used to generate the analysis used to initialise forecasts. When producing a reanalysis, however, it is possible to use observations made *after* the time for which the state estimate is required, in addition to those from before, to estimate the model state. Because reanalysis products are reconstructions that are complete in time and space, they can be used to estimate weather at locations for which direct historical observations are not available. Reanalysis projects have been undertaken by ECMWF and NCEP.<sup>4</sup>

#### CHAOS AND WEATHER FORECASTING

In the late 1950s the meteorologist, Edward Lorenz, was experimenting with a numerical model of the atmosphere at the Massachusetts Institute of Technology (MIT). During a set of experiments, he started a new run by resetting the state of the model to the state obtained halfway through a previous run. To his surprise, the behaviour of the atmosphere in second half of the new run was substantially different from the second half of the initial run. Eventually Lorenz realised that, while the computer was evaluating the model to six decimal-place accuracy, it was printing out the model state to only three decimal-place accuracy, thus resetting the model with the printed output had introduced a tiny discrepancy (less than one part in 1,000) between the two runs. This error was enough to cause a big difference in their evolutions (Lorenz 1993). While such behaviour had been known for centuries, the existence of digital computing made such behaviour more amenable to study. In 1975 the word "chaos" was coined to describe such sensitivity of these models to initial conditions, a property popularly known as the "butterfly effect".

To study the model sensitivity to initial conditions, experiments in which artificial errors, consistent with known observational uncertainty, are introduced into computer models of the atmosphere. The model results using these alternative initial conditions can then be compared with the original runs; the comparisons suggest that it is unlikely that we will ever predict the *precise evolution* of the atmosphere for longer than a few weeks (Lorenz 1982).

The impact of chaos on meteorology has not been entirely negative. Instead, it

has led to a shift in emphasis – a shift likely to be useful to weather risk management professionals. The atmosphere is not uniformly sensitive to initial conditions; its sensitivity depends on its current state. On some days the atmosphere can be approximated by a relatively predictable linear system over a short enough period of time. On other days, the inevitable uncertainty that exists in the analysis can lead to rapid error growth in the forecast. The key point is that *the predictability of the atmosphere depends upon the state of the atmosphere*. Predicting predictability has now become a major component of operational weather forecasting.

#### ENSEMBLE FORECASTING

Both the ECMWF and the US National Centre for Environmental Prediction (NCEP) have been running daily ensemble forecasts since 1992. The idea behind ensemble forecasting is simple; run several forecasts using slightly different initial conditions generated around the analysis. The relative divergence of the forecasts indicates the predictability of the atmospheric model in its current state – the greater the divergence the lower the confidence in the forecast. Thus, ensemble forecasts should provide *a priori* information on the reliability of that day's forecast. The difficulty arises in deciding what perturbations (errors introduced into the initial condition) to make to the analysis. Limited computing power means that only a few dozen forecasts can be produced (even when the ensemble forecasts are produced using a lower resolution model than the main forecasts).

The ensemble formation methods discussed above attempt to account for errors in the initial condition. Other ensemble forecasts use models with different combinations of parameterisation schemes for sub-grid processes, in an attempt to estimate how the uncertainty in the choice of scheme affects the final forecast (Stensrud 2000). Another way in which ensemble members may differ is by introducing random terms into the dynamical equations of the model. This approach attempts to model the uncertainty in the future evolution of the atmosphere at each time step of the model. In such a stochastic parameterisation, the impact of sub-grid processes on the resolved flow is not assumed to be a deterministic function of the model state. Instead, it is a random variable. The parameters of the distribution from which this variable is selected (such as its mean and variance) can be determined by the model state.

Due to the non-linearity of NWP models, introduction of these stochastic terms can actually improve the mean state of the model in addition to helping to assess the uncertainty in its evolution (Palmer 2001). Stochastic parameterisation is a new feature of numerical weather prediction, and is partly a manifestation of meteorology's willingness to accept uncertainty and its attempt to quantify it.

For the purpose of pricing weather derivatives, ensemble forecasts are much more useful than traditional single forecasts. For example, each member of an ensemble forecast can be used to calculate the number of heating degree-days (HDDs) accumulated in a period: this provides a rudimentary distribution of future HDDs. At present, however, the relatively small size of the ensembles and the fact that they represent quantities averaged over tens of kilometres rather than at individual weather stations, means that using ensembles in this manner would be ill-advised. We shall discuss how the predictability information contained in current ensemble forecast might be extracted in the next section.

Once uncertainty in the atmospheric state has been accepted as a fact of life that will not go away, the extension of medium-range forecasting techniques to longer-range seasonal forecasting is not a major leap. As noted above, the sensitivity of the evolution of the atmosphere's state to its initial conditions prohibits the precise prediction of the trajectory of this state for longer than, at best, a few weeks into the future. This means that there is little hope of forecasting whether it will rain on a specific day in a few months time. This does not mean, however, that useful forecasts at lead times of several months are not possible. It is possible to predict whether a



season will be wetter or colder than average at lead times of over three months and to estimate the probability of magnitudes of change – the probability it will be at least 1°C warmer than average, for example.

The crucial extra ingredient required for seasonal forecasting is a computer model of the ocean. The time scales on which the state of the ocean changes are quite long compared to the lead time of a medium-range forecast. Because of this, when making such a forecast, the state of the ocean can be held fixed. Beyond a couple of weeks, however, the changing state of the ocean is an important influence on the behaviour of the atmosphere. To make progress in seasonal forecasting, a model of the ocean must be coupled to the atmosphere model. The atmosphere forces the ocean through wind stress at its surface, while the ocean forces the atmosphere by exchanging heat with it, especially through the evaporation of water – which forms clouds – and radiation.

The sea surface temperature (SST) is an important influence on the behaviour of the atmosphere, particularly in the tropics. It is the coupling of the ocean and atmosphere that lies behind much inter-annual climate variation such as El Niño-Southern Oscillation (ENSO).<sup>5</sup> ENSO is characterised by a large-scale cycle of warming and cooling in the Eastern tropical Pacific that repeats on a time scale of 2–7 years. The warm SST of the El Niño phase of the cycle influences the atmospheric circulation over large parts of the globe. In particular, El Niño events are associated with heavy rainfall in Peru and Southern California, mild winters in the Eastern US and drought in Indonesia and Northern Australia (Glantz 1996). These are all probabilistic associations – ENSO is just one influence of the atmosphere's behaviour, although an important one at lower latitudes (see Philander, 1990 for further reading).

In many ways, the existence of ENSO is a blessing, it imposes some degree of regularity on tropical climate that helps seasonal prediction. The numerical ocean models used in seasonal forecasting are not fundamentally different from their atmospheric cousins; they rely on the division of the oceans into finite elements, horizontally and vertically; the equations of mass, momentum and energy conservation are integrated numerically, and sub-grid processes are parameterised. Seasonal forecasts at mid-latitudes are not as skilful as in the tropics. Although there are thought to be mid-latitude climate cycles, such as the North Atlantic Oscillation (NAO), they are not as regular and well defined as ENSO. The existence of these cycles has allowed the development of statistical seasonal forecasting models which have skill at lead times of up to six months (eg, Penland and Margorian, 1993). Like all statistical models, however, the availability of historical data is a major constraint on their refinement. Improvements in seasonal forecasting will require better information about the state of the ocean. Observations of the ocean are not as dense as atmospheric observations. Better ocean data, such as that obtained from the TOPEX-POSEIDON satellite, and its successor JASON-1, which measure sea surface height, should enable better estimation of the state of the ocean, and thus improved seasonal forecasts.<sup>6</sup>

Beyond seasonal time scales, forecasts that have more skill than climatology are elusive. It is possible, however, that coupled numerical models of the ocean-atmosphere system can help to improve the climatological distributions that are used to assess weather risk. The instrumental records for most locations are quite short, often only extending back a few decades. Extended runs of oceanic-atmospheric general circulation models (GCMs) could enable better estimates of the risk of extreme events that may only have occurred on a handful of occasions in recent history. This is particularly true if secular changes to forcings of the ocean-atmosphere system – such as enhanced radiative forcing due to increased carbon dioxide and other greenhouse gases (Harries et al. 2001, IPCC, 2001) – reduce the relevance of the historical record. Before GCMs can be used for this type of risk

assessment, it must be demonstrated that they can reproduce observed historical climate variability on regional scales, not just average global temperatures.

### Interpretation and Post-processing Model Output

The output generated by numerical models, should not be considered as "weather forecasts". The output merely represents the state of the model, which contains information about the weather but is not immediately relevant to quantities that are actually observed. There are many ways in which extra processing of the raw forecast products produced by forecast centres can substantially increase the value of these forecasts. For example, even the highest resolution global models cannot resolve the details of mountain topography or small islands, yet these physical features can have a substantial impact on the local weather conditions. One of the ways in which human forecasters can add value to a forecast is by using their experience of the weather in a particular locale to predict the likely conditions there, given the larger scale weather picture that the numerical forecast provides.

The finite resolution of numerical models, that is, the grid size, limits its application to areas no smaller than a grid. The user of the model forecast, however, is likely to need to know the values of forecast variables on a scale smaller than a model grid: pricing a weather derivative may require the temperature at the London Heathrow weather station, for example, but the ECMWF model predicts a temperature that is averaged over a grid box of 40 km by 40 km. "Downscaling" is the term used to describe a variety of quantitative methods that use the values of forecasted model variables for estimating the values of specific variables on scales smaller than the model grid.

One common method is the use of model output statistics (MOS) (see Glahn and Lowry 1972). A small number of model variables is chosen as the set of predictors of the desired variable. These predictors are extracted from past numerical forecasts, and correlated with the corresponding observational record of the desired variable. A statistical model can then be constructed that predicts the desired variable, using the forecast variables as predictors. This statistical model should remove any systematic biases in the numerical model. How MOS should be extended to ensemble forecasts is not obvious.

One could use each member of the ensemble as the predictor in a traditional statistical MOS model.

However, the forecast uncertainty represented by the ensemble must also be downscaled. This is because the forecasts are averaged over grid boxes that are tens of kilometres in size. The uncertainty of a forecast averaged over a grid box should be lower than the uncertainty of a forecast at a single weather station (that averaging reduces variability is a much exploited fact in both science and finance). So, while traditional MOS can be used to downscale the mean of the ensemble, obtaining an appropriate estimate of uncertainty is trickier. Methods for extracting the predictability that exists in the ensemble forecasts are now being developed (eg, Roulston *et al.*, 2002).

Another method for downscaling a forecast is "nested modelling". This uses a local area numerical model, covering a much smaller region than the global model, but with a much higher resolution. The local model is "nested" inside the global model. This means that it is integrated forward like the global model, but that on the edges of the region it covers the values of its variables are obtained by using the corresponding values from the global model (interpolated onto the local model's higher resolution grid). An example of using a nested model in a forecast application can be found in Kuligowski and Barros (1999).

### FORECAST UNCERTAINTY

Many applications of weather forecasts require an estimate of the uncertainty associated with the forecast. A single, best-guess forecast can be converted into a



## PANEL 3

## EVALUATION OF FORECASTS

When choosing forecast products to help manage their weather risk, users should obviously be concerned with the skill level of the forecasts. Unfortunately, it is difficult to assess skill when forecasting multiple variables with numerical models.

A common measure of skill used by meteorologists is the mean square error (MS error). This is just the mean of the square of the difference between the forecasted value and the observed value.

The mean can be an average over time for a univariate forecast (eg, temperature at Heathrow airport), or a time-space average for a multivariate forecast (eg, the 500hPa height field over the northern hemisphere). The MS error, however, is not a good way to evaluate an ensemble, or probabilistic forecast. The ensemble average can be calculated and assessed this way but such approach completely fails to take into account the information about uncertainty inherent in the ensemble forecast, information of great interest to risk managers.

Verifying probabilistic forecasting systems requires different measures of skill. The class of measures most relevant to risk managers is probabilistic "scoring rules". To use these rules the number of possible outcomes is usually a finite number of classes. Examples of such classes might be rain/no rain or a set of temperature ranges. A probabilistic forecast will assign a probability to each of the possible outcomes. Scoring rules are functions of the probability that was assigned to the outcome that actually occurred. If this probability is  $p$  then the quadratic score of the forecast is  $p_-$ . The logarithmic score is  $\log p$ .<sup>7</sup> The quadratic scoring rule forms the basis for the "brier score" and the "ranked probability score" (Brier 1950). The logarithmic scoring rule is connected to the information content of the forecast, and also to the returns a gambler would expect if they bet on the forecast (Roulston and Smith, 2002).

Another method for evaluating probabilistic forecasts is the cost-loss score (Richardson, 2000).

The cost-loss score is based on the losses that would be incurred by someone using a probabilistic forecast to make a simple binary decision. Consider the situation where the user must decide whether to grit the roads tonight. The cost of gritting is  $C$ . If it does not freeze, no loss is suffered if the roads are not gritted. If it does freeze, ungritted roads will lead to a loss,  $L$ . If  $p$  is the forecast probability of it freezing tonight, then the expected loss if the user does not grit is  $pL$ . If  $p > C/L$  then this expected loss exceeds the cost of gritting and a rational user would send out the gritting trucks. The cost-loss score illustrates how important it is to have a probabilistic forecast. A user with a  $C/L$  ratio far from 0.5 would be ill-advised to take the course of action suggested by a best-guess forecast. Best-guess forecasts can be easily converted into probabilistic forecasts by estimating the historical forecast error distribution (ie, the errors of previous forecasts).

Such a conversion should always be performed before evaluating best-guess forecasts using any skill score designed for probabilistic forecasts. Not doing so will artificially inflate the advantage of using an ensemble system. Users that can formalise their decision-making can directly estimate the value of forecasts by determining their impact on decisions. Decision making processes for real users will be more complex than the cost-loss scenario described above but, the principle of utility maximisation should still apply. Katz and Murphy (1997) contains detailed articles on common evaluation methods for weather forecasts.

probabilistic forecast by adding a distribution of historical forecast errors. The historical forecast errors are the differences between previous forecasts and the corresponding value of the variable that was later observed. Historical error distributions can be refined by allowing them to vary seasonally. For example, in the winter, only forecast errors from previous winters should be used to estimate the distribution. In principle, one can go further and choose only the historical errors that occurred when the atmospheric state was similar to the current state. One must then confront, however, similar problems faced when trying to build purely statistical forecasting models, namely too little data and too few similar days. The data problem for forecast error prediction is even more acute than for statistical weather prediction. Frequent upgrades of operational numerical models mean that the time series of past forecast errors for current models are seldom longer than a couple of years.

Ensemble forecasts contain information about the state-dependent forecast uncertainty. Ideally, this information should be propagated through any downscaling procedure in an "end-to-end" forecast.

In an end-to-end forecast the output of a weather forecasting model is processed to obtain a final forecast for the (weather-dependent) variable in which the user is actually interested. This could be electricity demand or wind energy supply, for example (Smith *et al.*, 2001). A statistical downscaling model based on MOS can be applied to each member of an ensemble forecast. This, however, will probably lead to an underestimate of forecast uncertainty. The uncertainty associated

with a weather variable measured at a particular location will be higher than the uncertainty in that variable averaged over the resolution of the numerical model that generated the ensemble. Therefore, the forecast uncertainty contained in the ensemble must also be downscaled. One possible approach is to include statistics describing the entire ensemble (eg, ensemble spread) in the MOS predictors that are used to construct the statistical downscaling model.

### Conclusion

Over the last 2,000 years, the prediction horizon of state-of-the-art weather forecasts has advanced significantly from seeing one day ahead based on the colour of the sky to almost a two-week outlook. The risk management community moved more quickly, as in only a few years, the very concept of a weather forecast has changed from a single best-guess of the future to a distribution of likely future weather scenarios.

In this chapter the methods used to produce modern numerical weather forecasts have been outlined. It was also claimed that the modelling techniques used for making short-range and medium-range forecasts are not fundamentally different from the approaches that must be adopted to forecast climate on seasonal scales and beyond. The main difference is that, for longer-range forecasts, a model of ocean dynamics must be coupled to the atmospheric model. This chapter has also stressed the importance of ensemble forecasting, for quantifying forecast uncertainty. Ensemble forecasting is on its way to becoming the standard approach to forecasting the evolution of the atmosphere and the ocean. This is a welcome development for the risk management community because ensemble forecasts contain information about the uncertainty of forecasts, uncertainty that is synonymous with risk.

The ultimate aim of accurate probability forecasts of commercially relevant variables is being pursued on time scales from a few hours to several months (Palmer 2002). For those who understand how to interpret it, this information will give a competitive edge at a range of lead times, from pricing next week's electricity futures to pricing next year's weather derivatives. Skilled and timely forecasts can be financially rewarding.



- 1 Richardson's work led to his book *Weather Prediction by Numerical Processes*, which is considered to be the foundation of modern meteorology (Richardson, 1922).
- 2 All operational forecast centres frequently upgrade their forecast models. Details of these upgrades, and of the present configuration of the ECMWF model can be found on their website <http://www.ecmwf.int>.
- 3 A description of the ECMWF and its output products can be found in Persson (2000).
- 4 Details of reanalysis projects and data can be found at <http://ecmwf.int/research/era/> and [http://www.cdc.noaa.gov/ncep\\_reanalysis](http://www.cdc.noaa.gov/ncep_reanalysis).
- 5 As noted above Gilbert Walker discovered the atmospheric Southern Oscillation in the 1920s. The relationship between this variation and the oceanic El Nino phenomenon was discovered in the 1960s by Jacob Bjerknes, son of Vilhelm.
- 6 Information on these satellites is available at <http://topex-www.jpl.nasa.gov/mission/jason-1.html>.
- 7 Readers may be curious as to why the simplest linear skill score  $=p$  has been omitted. It turns out that this score is "improper" in the sense that a forecaster increases their expected score by reporting probabilities that differ from what they believe to be true.

## BIBLIOGRAPHY

- Brier, G. W., 1950, "Verification of Forecasts Expressed in Terms of Probabilities", *Monthly Weather Review*, 78, pp. 1-3.
- van den Dool, H. M., 1994, "Searching for Analogues, How Long must we Wait?" *Tellus A*, 46, pp. 314-24.
- Glahn, H. R., and Lowry, D. A., 1972, "The Use of Model Output Statistics (MOS) in Objective Weather Forecasting", *Journal of Applied Meteorology*, 11, 1203-11.
- Glantz, M. H., 1996, *Currents of Change: El Nino's Impact on Climate and Society*, (Cambridge University Press).
- Harries, J. E., Brindley, H. E., Sagoo, P. J. and Bantges, R. J., 2001, "Increases in Greenhouse Forcing Inferred from the Outgoing Longwave Radiation Spectra of the Earth in 1970 and 1997", *Nature*, 410, pp. 355-7.
- IPCC, Intergovernmental Panel on Climate Change, 2001, *Climate Change 2001: Synthesis Report*, (ed) R. T. Watson, (Cambridge University Press).
- Katz, R. W., and Murphy, A. H., 1997, *Economic Value of Weather and Climate Forecasts*, (Cambridge University Press).
- Kuligowski, R. J. and A. P. Barros, 1999, "High-Resolution Short-Term Quantitative Precipitation Forecasting in Mountainous Regions Using a Nested Model", *Journal of Geophysical Research (atmospheres)*, 104, pp. 31, 553-64.
- Lempfert, R. G. K., 1932, "The Presentation of Meteorological Data", *Quarterly Journal of the Royal Meteorological Society*, 58, pp. 91-102.
- Lorenz, E. N., 1982, "Atmospheric Predictability Experiments with a Large Numerical Model", *Tellus*, 34, pp. 505-13.
- Lorenz, E. N., 1993, *The Essence of Chaos*, (London: UCL Press).
- Murphy, A. H., 1997, "Forecast Verification", in: *Economic Value of Weather and Climate Forecasts*, (eds) R. W. Katz and A. H. Murphy, (Cambridge University Press), pp. 19-70.

Nebeker, F., 1995, *Calculating the Weather*, (San Diego: Academic Press).

Palmer, T. N., 2000, "Predicting Uncertainty in Forecasts of Weather and Climate", *Reports on Progress in Physics*, 63, pp. 71–116.

Palmer, T. N., 2001, "A Nonlinear Dynamical Perspective on Model Error: A Proposal for Non-Local Stochastic-Dynamic Parameterization in Weather and Climate Prediction Models", *Quarterly Journal of the Royal Meteorological Society*, 127, pp. 279–304.

Palmer, T. N., 2002, "The Economic Value of Ensemble Forecasts as a Tool for Risk Assessment: From Days to Decades", *Quarterly Journal of the Royal Meteorological Society*, 128, pp. 747–74.

Penland, C., and T. Magorian, 1993, "Prediction of Nino 3 Sea Surface Temperature Using Linear Inverse Modeling", *Journal of Climate*, 6, pp. 1067–76.

Persson, A., 2000, "User Guide to ECMWF Forecast Products", European Centre for Medium Range Weather Forecasting, Reading.

Philander, S. G., 1990, *El Nino, La Nina, and the Southern Oscillation*, (San Diego: Academic Press).

Richardson, L. F., 1922, *Weather Prediction by Numerical Processes*, (Cambridge University Press).

Richardson, D. S., 2000, "Skill and Relative Economic Value of the ECMWF Ensemble Prediction System", *Quarterly Journal of the Royal Meteorological Society*, 126, pp. 649–67.

Roulston, M. S. and L. A. Smith, 2002, "Evaluating Probabilistic Forecasts using Information Theory", *Monthly Weather Review*, 130, pp. 1653–60, 2002.

Roulston, M. S., D. T. Kaplan, J. Hardenberg and L. A. Smith, 2002, "Using Medium Range Weather Forecasts to Improve the Value of Wind Energy Production", Forthcoming, Renewable Energy.

Smith, L. A., Roulston, M. S. and Hardenberg, J., 2001, "End to End Ensemble Forecasting: Towards Evaluating the Economic Value of the Ensemble Prediction System", ECMWF Technical Memorandum 336, ECMWF, Reading.

Stensrud, D. J., J. W. Bao, and T. T. Warner, 2000, "Using Initial Condition and Model Physics Perturbations in Short-Range Ensemble Simulations of Mesoscale Convective Systems", *Monthly Weather Review*, 128, pp. 2077–107.

Walker, G. T., 1923, "Correlation in Seasonal Variations of Weather VIII", Memorandum of the Indian Meteorological Department, 24, pp. 75–131.

Walker, G. T., 1924, "World Weather IX", Memorandum of the Indian Meteorological Department, 24, pp. 275–332.

Walker, G. T., 1928, "World Weather III", Memorandum of the Royal Meteorological Department, 17, pp. 97–106.

Walker, G. T., 1937, "World Weather VI", Memorandum of the Royal Meteorological Society, 4, pp. 119–39.