QUARTERLY JOURNAL
OF THE
ROYAL METEOROLOGICAL SOCIETY

Vol. 128 | APRIL 2002 Part A | No. 581

Q. J. R. Meteorol. Soc. (2002), **128**, pp. 747–774

The economic value of ensemble forecasts as a tool for risk assessment: From days to decades

By T. N. PALMER*
European Centre for Medium-Range Weather Forecasts, UK

(Received 18 June 2001; revised 30 November 2001)
(Symons Memorial Lecture: delivered on 16 May 2001)

SUMMARY

Despite the revolutionary development of numerical weather and climate prediction (NWCP) in the second half of the last century, quantitative interaction between model developers and forecast customers has been rather limited. This is apparent in the diverse ways in which weather forecasts are assessed by these two groups: root-mean-square error of 500 hPa height on the one hand; pounds, euros or dollars saved on the other.

These differences of approach are changing with the development of ensemble forecasting. Ensemble forecasts provide a qualitative tool for the assessment of weather and climate risk for a range of user applications, and on a range of time-scales, from days to decades. Examples of the commercial application of ensemble forecasting, from electricity generation, ship routeing, pollution modelling, weather-risk finance, disease prediction and crop yield modelling, are shown from all these time-scales.

A generic user decision model is described that allows one to assess the potential economic value of numerical weather and climate forecasts for a range of customers. Using this, it is possible to relate analytically, potential economic value to conventional meteorological skill scores. A generalized meteorological measure of forecast skill is proposed which takes the distribution of customers into account. It is suggested that when customers' exposure to weather or climate risk can be quantified, such more generalized measures of skill should be used in assessing the performance of an operational NWCP system.

KEYWORDS: Climate change Cost/loss ratio Probability forecasting Seasonal forecasts User application models

1. INTRODUCTION

Future generations may well view the second half of the 20th century as the most important period in the last millennium of meteorology; not least it marked the development of the comprehensive numerical weather and climate prediction (NWCP) model. It has long been recognized that numerical weather and climate predictions have enormous potential economic value. Mason (1966) estimated, for example, that the overall benefit/cost ratio of the Meteorological Office to the national economy (much of which derived from its weather and climate prediction capability) was about 20:1. It is likely that this figure is still relevant today.

On the other hand, the 20th century saw only limited quantitative interaction between the scientists who developed NWCP models (and associated data-assimilation schemes), and customers for these forecasts. This is apparent in the very different measures with which model developers and forecast end-users assess weather and

* Corresponding address: ECMWF, Shinfield Park, Reading, Berkshire RG2 9AX, UK. e-mail: nez@ecmwf.int

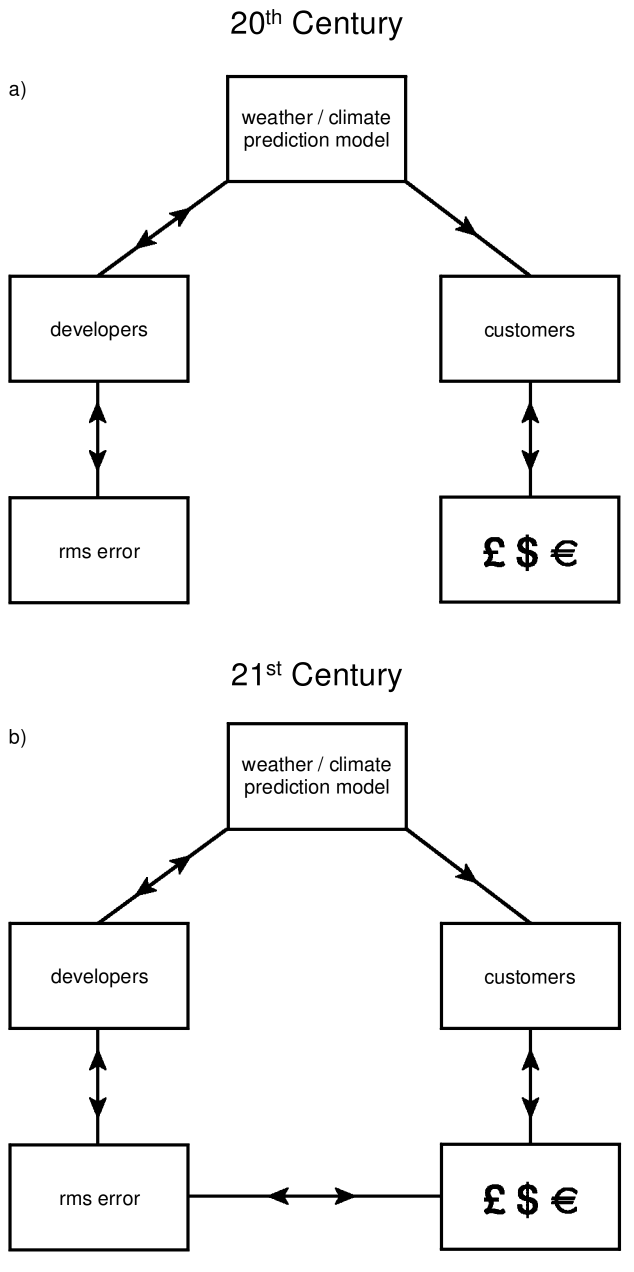

Figure 1. Schematic linkage between numerical weather and climate prediction and forecast customers: (a) situation that characterized much of the 20th century, with limited quantitative feedback between forecast model developers and forecast model users, and poorly-known correspondence between customers' and developers' measures of skill; and (b) the author's speculation for developer-user linkage in the 21st century based on the development of ensemble prediction systems for quantitative weather/climate risk assessment.

climate forecasts: developers typically seek to minimize the root-mean-square (r.m.s.) error of basic meteorological fields such as 500 hPa height, whilst commercial end-users require their investment in meteorological forecasts to be financially worthwhile (see Fig. 1(a)).

It is proposed in this paper that the development of ensemble weather and climate prediction systems (EPSs) will lead to the establishment of more quantitative links between forecast developers and forecast customers. This in turn, it is suggested, will enhance the economic value of weather and climate prediction. An ensemble forecast comprises multiple integrations of NWCP models; these integrations differ by uncertainties in the initial state and in the computational representation of the equations of motion. Because the equations of motion are nonlinear, the dispersion of an ensemble forecast will be dependent on the initial state: small dispersion indicating high predictability, and vice versa. In the past few years, ensemble prediction has become an established part of weather and climate prediction, from days to decades (see, for example, Palmer 2000).

The London Financial Times recently estimated that at least a quarter of the national economy is influenced by climate, either at risk from adverse weather and climate, or able to benefit from equable conditions. An ability to quantify weather and climate risk is therefore key to commercial decision making for much of the economy. Probabilistic forecasts provide the means to predict weather and climate-related risk; single deterministic forecasts, by contrast, are incapable of this. For example, if a conventional deterministic 5-day prediction is for settled weather, this does not imply that the risk of severe weather is zero. If all that was available in deciding whether or not to tow a multi-million dollar oil rig to site was a 5-day forecast of settled weather, then a rational decision might be to proceed. However, suppose an EPS indicated that there was nevertheless a 20% risk of severe weather. Then, based on the ratio of the cost of keeping the oil rig in port (until a more favourable forecast was available), to the loss of a capsized oil rig in mid-tow, a more informed decision might be to wait in port. This example illustrates a key principle to which I shall return many times: commercial decisions are often made, not on the basis of events which are likely to occur, but on the basis of events which are unlikely to occur, but which, if they did occur, would involve serious financial loss. Only reliable probabilistic forecasts can provide the means for this type of decision process in the weather and climate-sensitive commercial sector.

In this paper, the economic value of ensemble forecasts as a tool for risk assessment is discussed. Some specific examples of the application of ensemble forecasts are given in section 2, based on weather time-scales, seasonal time-scales, and climate change time-scales. An analysis of the potential economic value of ensemble forecasts is given in section 3. Relationships between economic value and meteorological skill are discussed in section 4. This will allow us to assert that the development of EPSs in the 21st century will lead to a more complete linkage between forecast developers and forecast customers, as suggested in Fig. 1(b). Concluding remarks are made in section 5.

2. Applications of ensemble predictions

(*a*) *The basics*

He believed in the primacy of doubt, not as a blemish on our ability to know, but as the essence of knowing. *Gleick (1992) on Richard Feynman.*

Feynman, one of the great 20th century physicists, embraced uncertainty as a key factor which distinguishes science from pseudo-science. Astrologers may well be vague, but they are never uncertain! As scientists, we should recognise uncertainty, and, where possible, quantify it.

As mentioned in the introduction, there are two principal reasons why weather and climate forecasts are uncertain—the initial data are both sparse and uncertain, and the process of representing computationally the equations of motion is not well defined (e.g. Palmer 2001). In fact, for climate prediction there is a third uncertainty; we don't know how rapidly the composition of the atmosphere will change over the coming decades—it depends, for example, on whether treaties to limit the burning of fossil fuels are ratified and acted on. If we can quantify uncertainty in the initial state and in the model equations (and in the rate of change of atmospheric composition) through some probability density function, then, formally at least, forecast uncertainty can be expressed using equations such as the Liouville and Fokker-Planck equations (e.g. Ehrendorfer 1994).

However, in practice, we don't know how to estimate these probability density functions sufficiently accurately to use directly these elegant equations. By contrast, ensemble prediction involves taking 'plausible' samples of initial conditions (each consistent with uncertainties in the initial data), and 'plausible' samples of possible equations of motion (each consistent with the known continuum equations of motion) and creating forecasts from perturbed initial states and with perturbed sets of equations. On time-scales of days there are a number of operational or quasi-operational EPSs worldwide (e.g. Houtekamer *et al.* 1996: Molteni *et al.* 1996: Toth and Kalnay 1997). The multi-model EPS (e.g. Palmer *et al.* 2000; Meehl *et al.* 2000) has proven a pragmatic approach to ensemble prediction on time-scales of seasons and decades. At present, typical ensemble sizes range between 10 and 100.

Critical questions are: what is the best way to determine the ensemble initial conditions and the different computational representations of the equations of motion, how many ensemble members are needed, and how should computer resources be divided between ensemble size and model resolution? In this paper it will be argued that these questions cannot be answered properly without considering the users of these forecasts—the customers.

Given an EPS, how should customers use ensemble forecasts? For commercial customers, provision of probability forecasts of weather elements in general will not be good enough. Weather is one of a number of factors that influences commercial decision making. So, for example, a commercial decision is not based on the probability that the weather itself will be abnormal, but on the risk that property damage, electricity demand, crop yield, transportation delay, and so on, will be abnormal. These user variables may not depend linearly on just one meteorological variable, but nonlinearly on a number of variables (e.g. temperature, precipitation, cloud cover, wind speed).

Hence, to be of commercial use, EPS forecasts should be expressed in terms of the relevant user variables. This can be achieved by feeding the output of each member of the ensemble of weather or climate forecasts into a quantitative application model as shown in Fig. 2. The resulting output can be cast into the form of a probability distribution for the required user variable. (If the application model is itself proprietary, then this last phase in the production of the probability forecast may have to be done by the user, being supplied with output from individual members of the ensemble.) For example, if the user variable represents demand for a certain type of soft drink (such demand is clearly temperature dependent in summer), then the first two probability distributions shown in Fig. 2 would probably lead to straightforward decisions by retailers on how much soft-drink stock to order. If the EPS predicted the third type of probability distribution, then the decision on whether or not to order extra stock would depend on the profit made on selling soft drinks, compared with the cost of stocking

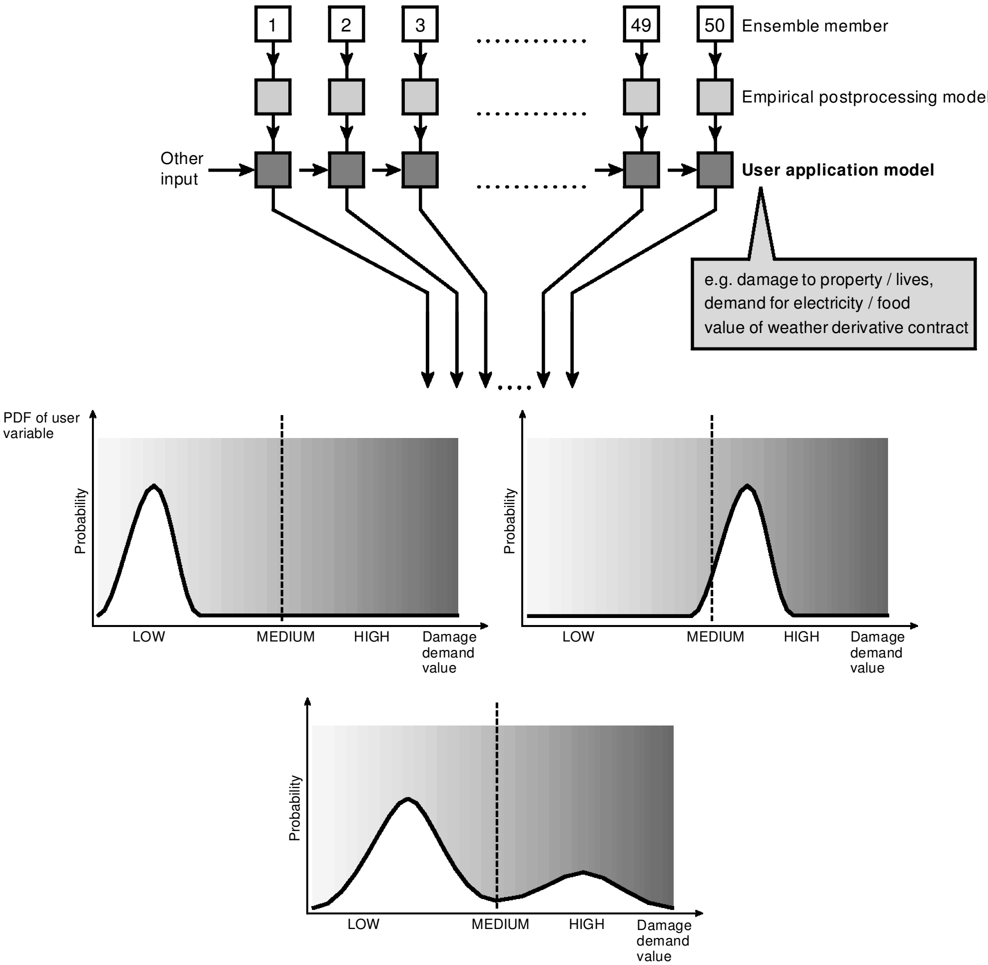

Figure 2. User applications depend, possibly nonlinearly, on a number of meteorological forecast variables. Hence many users will not find probability distribution function (PDF) forecasts of individual weather variables particularly useful. A general scheme for predicting weather/climate risk, tailored to the customers needs, is illustrated schematically.

unsold soft drinks. If, for example, the profit was 10 times the cost of stocking, then it would be worth ordering extra stocks, even if the secondary probability maximum was only, say, 20%.

As suggested in Fig. 2, it is necessary to postprocess the output of the ensemble, both to remove biases and to downscale to the scale of interest of the customer. This is not a trivial process, and may involve inflating the EPS-based probability distribution.

(*b*) *Ensemble forecasts on daily time-scales*

> Why have meteorologists such difficulty in predicting the weather with any certainty? Why is it that showers and even storms seem to come by chance . . . a tenth of a degree (C) more or less at any given point, and the cyclone will burst here and not there, and extend its ravages over districts that it would otherwise have spared. If (the meteorologists) had been aware of this tenth of a degree, they could have known (about the cyclone) beforehand, but the observations were neither sufficiently comprehensive nor sufficiently precise, and that is the reason why it all seems due to the intervention of chance. *Poincaré (1909).*

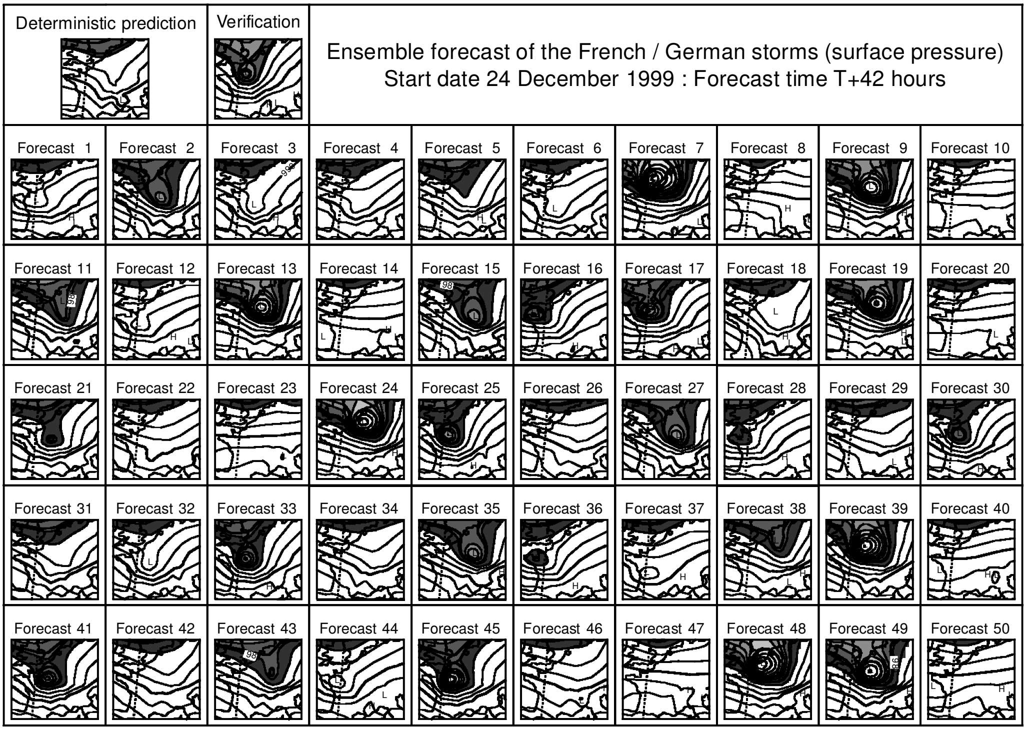

Figure 3. Individual forecasts from a 42-hour ensemble for one of the storms that devastated parts of mainland Europe on 26 December 1999. In this case, single deterministic forecasts using the ECMWF model did not forecast the risk of a severe storm, whilst the EPS indicated that there was a such a risk, vastly exceeding its climatological probability.

Poincaré was one of the greatest mathematicians of the late 19th/early 20th century. He understood the sensitive dependence of forecasts to initial conditions, what we would now call chaos (Lorenz 1993). In doing so he recognized that even if the best single deterministic forecast predicted that the cyclone would ‘burst here’, there is always a risk that the cyclone will in fact ‘burst there’. Figure 3 shows an example of his great insight applied to one of the most severe storms to pass over Poincaré’s home country, some 87 years after his death. By some accounts, this storm, together with one which passed over central Europe a couple of days later, led to the greatest insurance claim that Europe has ever seen.

The first row of Fig. 3 shows a 42-hour surface pressure forecast for 26 December 1999 from the (high-resolution) European Centre for Medium-Range Weather Forecasts (ECMWF) operational forecast and for the verifying analysis. This single deterministic forecast predicted a not-atypical winter’s day; this is the best single estimate of the outcome, given the available weather data on 24 December. However, as mentioned earlier, this does not imply that the risk of a substantially-atypical winter’s day is zero. The risk of severe weather can be assessed from the 50 members of the ECMWF EPS (here run at the current $\sim$80 km TL255 grid spacing) given in the other panels of Fig. 3. It can be seen that there are several members which predict some sort of storm, though its location is uncertain. Based on a simple gustiness parametrization, the probability of gusts exceeding 40 m s^{-1}, calculated from this ensemble of forecasts, is shown in Fig. 4. It can be seen that probabilities exceed 30% in places. It turns out that the

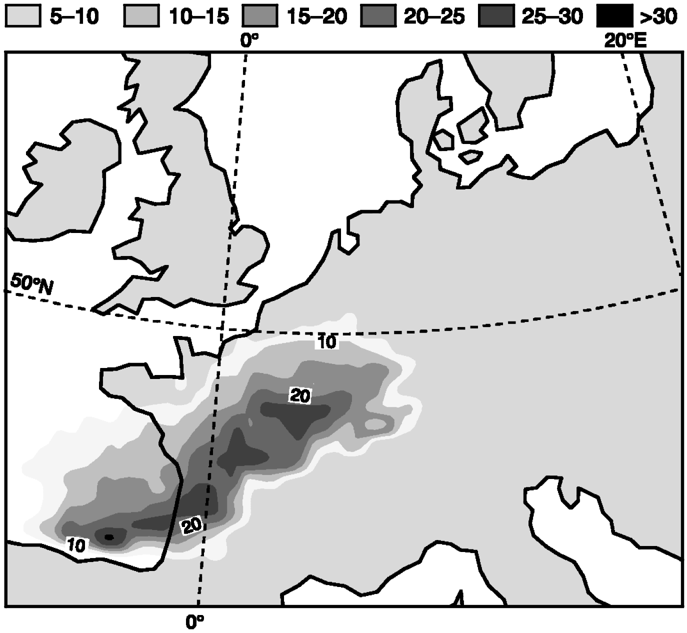

Figure 4. Probability (%) of gusts exceeding 40 m s^{-1} based on the ensemble forecasts shown in Fig. 3.

difference between the ensemble members which did have storms and those which did not, corresponded to small perturbations in temperature and wind over the west Atlantic (Poincaré's estimate was not too far wrong!). As mentioned in the introduction, it is easy to imagine commercial decisions that would be influenced by a predicted 30% chance of structural damage to buildings.

In discussing the economic value of ensemble forecasts, it is important to note their possible impact on personal decisions. It is often argued that the public will never be able to understand probability forecasts. However, 'understanding' may not be the key problem—we never hear bookies complaining that their customers do not understand the concept of 'odds'. Moreover, it is not hard to imagine personal decisions being made on a risk assessment of severe weather. Rather, the problem is that, most of the time, the ordinary person does not have the motivation to digest the extra information that is implicit in a probability weather forecast. No one wants to have to solve a nonlinear optimization problem to decide whether or not to take the umbrella to work! Hence the issue is whether this extra information can be put over succinctly and lucidly. On meteorological web sites, or on regional TV weather forecasts, a simplified version of a local 'EPS-gram' (see Fig. 5) is something that (I believe) the public could easily digest. At a glance, one can assess qualitatively the likelihood of good weather, or the risk of bad weather.

However, whilst this type of product is potentially valuable for the general customer, it may not, for reasons mentioned earlier, be useful for commercial customers. Based on the schematic shown in Fig. 2, it is easy to envisage an EPS-gram where the variable is not temperature or wind, but some specific user variable. For example, Fig. 6 shows a box-and-whiskers plot for mean England and Wales electricity demand (Taylor and Buizza 2002). These forecasts make use of surface temperature, wind speed and cloud cover, at noon and midnight. Consistent with the schematic in Fig. 2, the demand

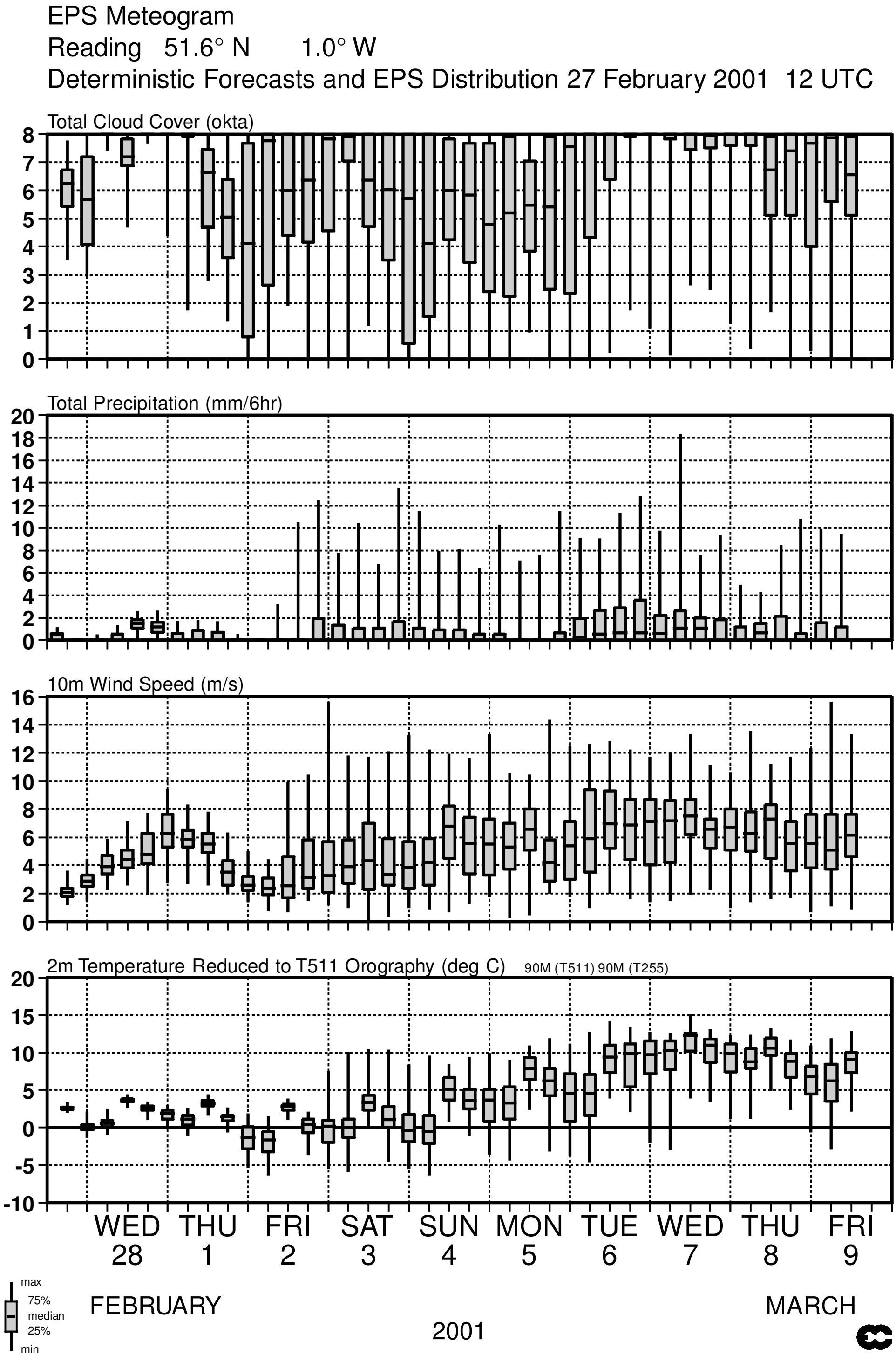

Figure 5. An example of an ECMWF 'EPS-gram' for Reading giving probability forecast distributions of cloud cover, total precipitation, 10 m wind speed and 2 m temperature, based on the ECMWF ensemble predication system (EPS). 50% of the forecast ensemble members lie within the box, 100% lie within the extremes of the whiskers. The median value is shown as a bar within the box.

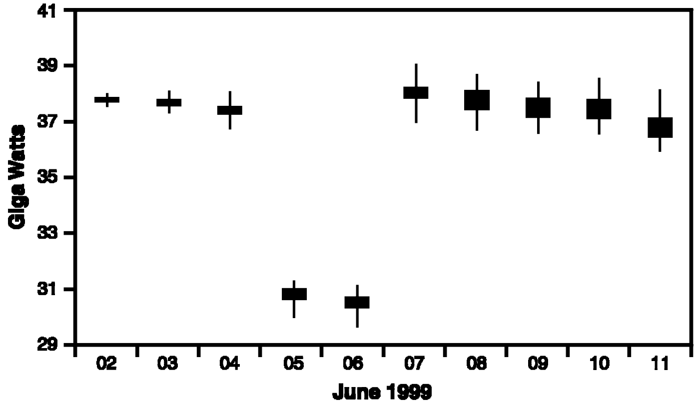

Figure 6. An example of an 'electricity-gram', giving forecast distributions of electricity demand for England and Wales based on a state-of-the art demand forecast model run using output from the ECMWF ensemble prediction system. 50% of the ensemble of demand forecasts lie within the box, 100% lie within the extremes of the whiskers. (From Taylor and Buizza 2002.)

is calculated using output from each of the 51 members of the ensemble. Electricity demand is dependent on non-meteorological factors, as is clearly demonstrated in Fig. 6—demand is lower at weekends. Taylor and Buizza (2002) show that significantly more skilful demands can be predicted from the ensemble than from a single deterministic NWCP forecast, or from an empirical forecast model of the sort that is used in the electricity industry. In addition, Taylor and Buizza show that the spread in the demand prediction itself carries useful information—this is relevant to electricity generators and electricity traders, both of whom need to assess the likelihood of high or low demand.

Smith *et al.* (2001) studied the economic value of ensemble forecasts for electricity generators who produce electricity using wind power. The example is motivated by the UK's New Electricity Trading Agreements (NETA). Consider the power P produced by a wind turbine. Suppose the generator must decide in advance the power PP, sold at a price SP, that they will provide on a given day. If actual production AP exceeds the promised amount, they cannot sell the excess. However, if AP falls short of the promised amount, the generator must replace the missing power at cost RC, which can be substantial. Assume that the wind generator's income IN is given by

$$IN = PP.SP \tag{1}$$

if $AP \geqslant PP$, and

$$IN = PP.SP - (PP - AP).RC \tag{2}$$

if $AP < PP$. Note that if the shortfall is significant, the replacement cost can completely wipe out any profit arising from generation. With AP proportional to the daily-mean cubed wind speed at a site in southern England, Fig. 7 shows the relative profits of wind generators, using different forecasts based on 4-day forecasts (too far in advance to be directly relevant to the NETA arrangements, but arguably pertinent for forward trading of electricity). In this application, the ECMWF EPS has a clear advantage over the deterministic ECMWF forecast, even when the latter is converted to a 'poor man's' probability forecast by adding a fixed empirical Gaussian (based on past forecast errors) to the single deterministic forecast. In this example, the relative benefit of the ensemble (over that of a single deterministic forecast) is larger, the higher is the net replacement cost ($RC - SP$) of electric power. This is essentially because the ECMWF EPS estimates

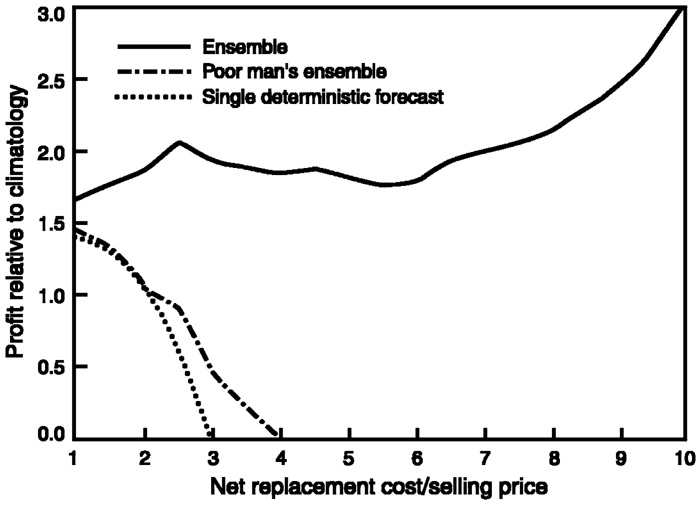

Figure 7. The relative income of a wind-power producer using different forecasts to predict average mean cubed wind speed at a forecast time of 4 days, based on the ECMWF ensemble forecasts, single deterministic forecasts, and a 'poor-man's' ensemble where the single deterministic forecast is cloaked with a fixed probability distribution deduced from historical forecast errors. Relative income is shown as a function of the ratio of the net replacement cost of electricity promised but not provided, to the selling price of the electricity. The results are averages for the period January 1999 to July 2000. All the ECMWF ensemble forecasts were for Heathrow, while the actual production was taken to be daily mean cubed wind speed at the Rutherford Appleton Laboratory, Oxfordshire. (From Smith *et al.* 2001.)

reasonably reliably the risk of calm weather. The decision on how much power to contract to sell will depend on the size of this risk. When the net replacement cost is sufficiently large, the decision on how much power to sell will be influenced by even very small forecast probabilities for low wind-speed conditions.

An actual situation where decisions were based on a low-probability event predicted by the ECMWF EPS is shown in Fig. 8 (see ECMWF 2000). On 12 December 1999 the oil tanker Erika was shipwrecked off the French Cape Pen'March. On 16 December Météo-France had to assess whether there was any possibility that the pollution would reach the coast before Christmas. The most likely forecast (Fig. 8(a)) indicated that the oil slick would still be out to sea by Christmas. However, a specific clustering of EPS forecasts was made which gave pessimistic scenarios in terms of pollution drift. This suggested (Fig. 8(b)) that there was a quantifiable risk of the oil reaching the coast by Christmas, and indicated that the authorities could only safely rely on a week to put into place emergency plans for the protection of the threatened coastline. The oil slick reached the coast on Christmas Day.

These examples illustrate an important principle in the development of EPSs. For users who are especially sensitive to weather and climate, the ensembles should not be under-dispersive. For the case of the Erika oil spill, an under-dispersive ensemble could have had disastrous consequences. For example, based on ensemble forecasts close to the control forecast, there was no risk of the oil reaching the coast by Christmas Day. For the electricity generator, an under-dispersive ensemble (underestimating the risk of calm weather) could mean buying in power at penalty rates so often that it would lead to insolvency. This sort of practical consideration was central to the philosophy that led to the choice of so-called singular vectors (Buizza and Palmer 1995) to generate the initial perturbations for the ECMWF EPS. As mentioned, for a number of reasons the probability distribution of initial error is not well known, so it is not possible to

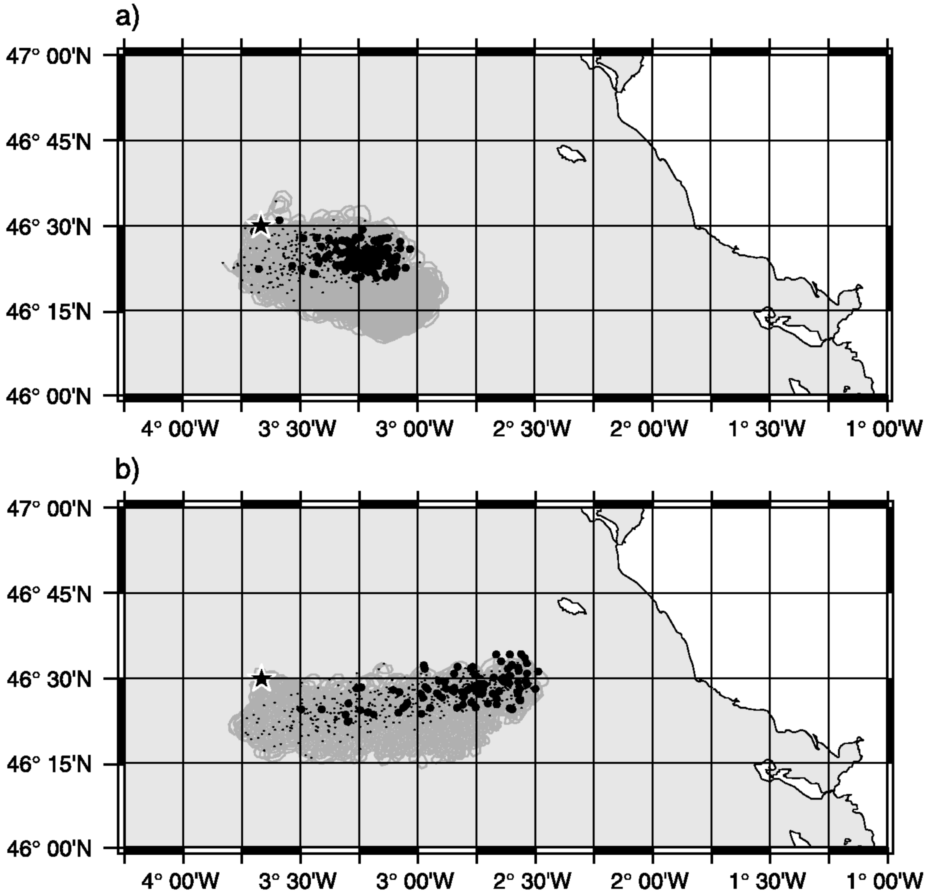

Figure 8. (a) 10-day prediction of the spread of the oil slick from the shipwrecked tanker 'Erika' based on the 'most likely' forecast scenario. (b) Risk of pollution spreading from a 'pessimistic' cluster of ensemble forecasts. Forecasts initialized on 17 December 1999, valid on 23 December 1999. (From ECMWF 2000.)

define an operational procedure which would give a reliable random sampling of such a distribution. As such, the actual sampling procedure is deliberately biased towards perturbations whose structures are likely to grow significantly during the forecast period.

Ensembles of medium-range forecasts provide forecast information for a range of variables, at a range of forecast times and over the whole globe. One application in which the space–time multivariate nature of the ensemble is utilized is ship routeing. The wind, ocean wind waves and swell are crucial parameters in finding an optimal ship route between two given points. Hoffschildt *et al.* (1999) studied the potential benefit of ensemble forecasts for ship routeing. This application makes use of the fact that the ECMWF atmospheric forecast model is coupled to an ocean wave model. Figure 9 shows ensembles of optimal ship routes for the Atlantic crossing from Brest to New York on two different dates. In the first example, deviations from the shortest navigable route are large. However, the ensemble dispersion in ship routes is small, and the ship's captain can deviate from the shortest navigable route, confident that this route will save on fuel. In the second example, there is considerable spread in the ensemble of ship routes, and, moreover, the ensemble median route deviates substantially from the route suggested by the operational deterministic forecast. The ship's captain may decide to stay in port,

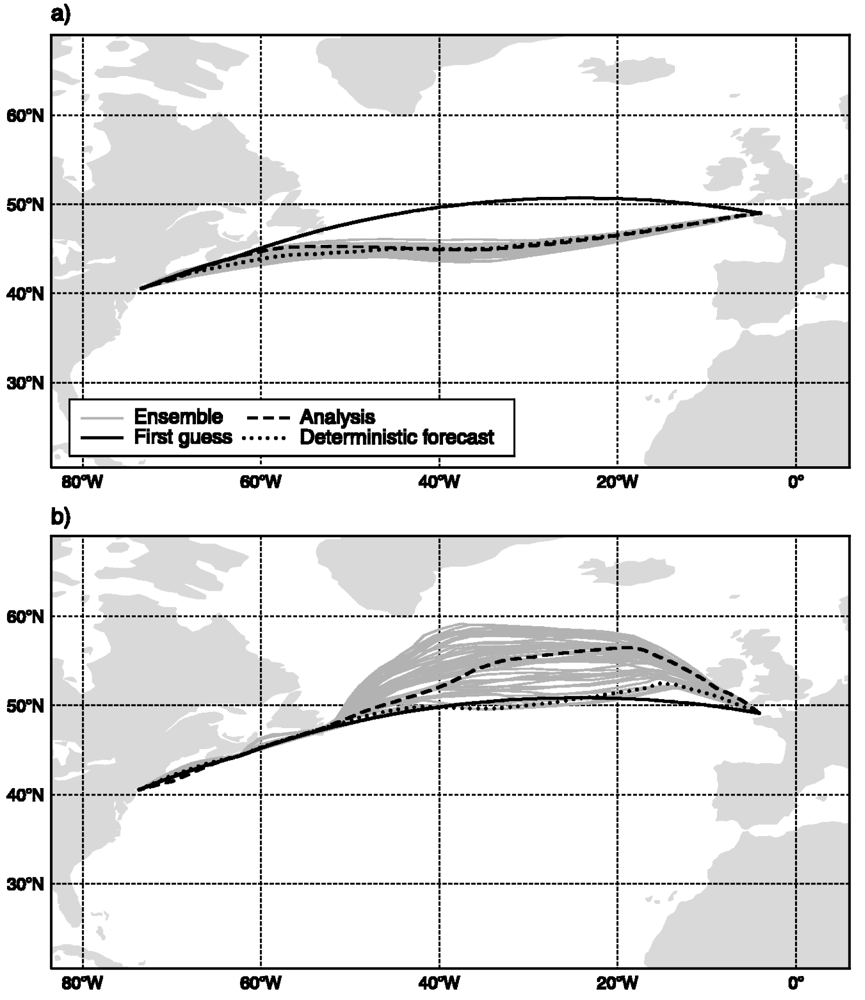

Figure 9. (a) Ship routes for the crossing leaving Brest at 12 UTC 22 February arriving in New York at 00 UTC 1 March 1999. The grey lines mark the optimal routes of the 50 ensemble members, the climatological (approximately great-circle) route is heavy solid, the verifying analysis dashed, and the single deterministic operational forecast dotted. (b) As (a) but for a crossing departing 12 UTC 30 December 1998 and arriving 00 UTC 7 January 1999. (From Hoffschildt *et al.* 1999.)

or possibly proceed along the most likely route, awaiting later forecasts. Based on a large number of cases, Hoffschildt *et al.* (1999) show that on average there would be significant fuel savings if the optimal ship route is computed using the ensemble of ship routes.

(*c*) *Ensemble forecasts on seasonal time-scales*

You asked, 'What is this transient pattern?' If we tell the truth of it, it will be a long story; it is a pattern that came up out of an ocean, and in a moment returned to that ocean's depth. *The Ruba'iyat of Omar Khayyam*.

The scientific basis for seasonal forecasts has been discussed many times (see, for example, Palmer and Anderson 1994). However, perhaps the 12th century quote from the Ruba'iyat sums it up, even though Omar Khayyam may himself not have been familiar with the El Niño phenomenon!

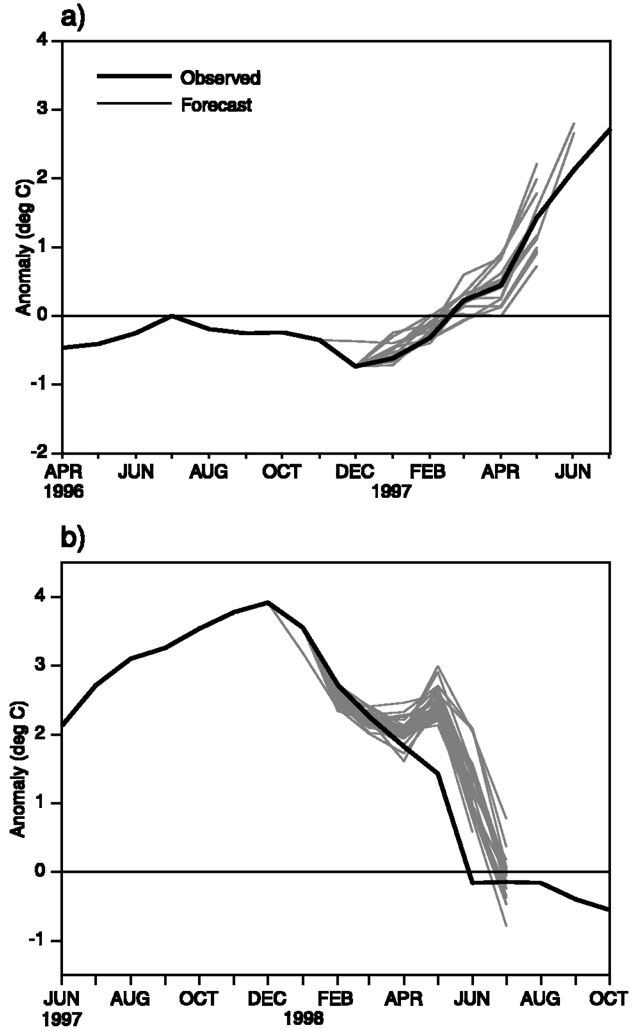

Figure 10. An ensemble of 6-month forecasts of El Niño sea surface temperature anomalies (NINO3 region in the eastern Pacific) using the ECMWF coupled ocean–atmosphere model from (a) December 1996, and (b) January 1998. The thin lines show the forecast ensemble, the thick line gives the observed anomalies. (From Stockdale *et al.* 1998.)

Ensemble forecasting has become an established and necessary part of seasonal prediction. Figure 10 shows a set of forecast plumes of sea surface temperature for the so-called NINO3 region of the equatorial Pacific, based on the ECMWF coupled ocean–atmosphere model (Stockdale *et al.* 1998). Figure 10(a) shows the probability of warm sea temperatures in the eastern Pacific 'arising from the deep' based on 6-month forecasts made at the end of 1996. Figure 10(b) shows the probability of these warm temperatures 'returning to the deep'. In other words, these forecasts clearly show, at the end of 1996, a significant probability of an El Niño event occurring in 1997, and the transition to La Niña occurring in 1998.

However, the ensembles are not perfectly reliable. None of the ensemble members correctly predicted the intensity of the ocean warming in the middle of 1997, and there is clearly a timing problem with the onset of La Niña. There is a growing body of opinion that such forecast failures cannot be attributed solely to inadequacies in our knowledge of the initial state of the coupled ocean–atmosphere system, that a significant component of error arises from inadequacies in the model equations. For example, the current ECMWF model has a poor representation of tropical intraseasonal atmospheric variability. There is substantial observational evidence that such variability played a role in determining the intensity of the 1997/98 El Niño.

Considerations of this sort lead one to the idea that different models have different forecast characteristics, due primarily to different representations of subgrid variability. Some models have strengths in one area, other models have strengths in other areas. As such, a more reliable ensemble system, especially for seasonal prediction, could arise by combining output from different models, the so-called multi-model ensemble. This notion was suggested by results from the PROVOST project (PRediction Of climate Variations On Seasonal to interannual Time-scales; funded by the European Union IVth framework Environment and Climate Programme (Doblas-Reyes *et al.* 2000; Graham *et al.* 2000; Palmer *et al.* 2000)), and from other studies (Krishnamurti *et al.* 2000). In fact, there is evidence that multi-model ensembles may also be beneficial in the short and medium range (Harrison *et al.* 1999; Evans *et al.* 2000; Wandishin *et al.* 2001).

In section 3, it will be demonstrated, based on PROVOST data, that multi-model seasonal ensembles have enhanced potential economic value over individual-model ensembles. As a result of this, work has begun on a successor project named DEMETER (Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction (see http://www.ecmwf.int/research/demeter/)) to install a number of different European coupled ocean–atmosphere models on a single computer, and to produce multi-model ensemble seasonal 'hindcasts' for much of the latter part of the 20th century. The DEMETER project has three application partners who will assimilate the output of the multi-model ensemble data into their application model, as indicated in the Fig. 2 schematic. One application is the prediction of a malaria outbreak in tropical Africa (Thomson *et al.* 2000; Thomson and Connor 2001). As shown in Fig. 11, the seasonal forecasts play a major role in the malaria prediction models. However, there are certainly other input factors which need to be taken into account—the immune and nutrition status of the local population for example.

Crop yield prediction is another application activity in the DEMETER project. For example, the European Commission's Joint Research Centre is charged with predicting agricultural crop yields across Europe. This is an important component of the workings of the Common Agricultural Policy which set crop quotas and intervention pricing. Currently, crop yield estimates are based on multiple regression analysis between crop conditions in April and crop yield, based on historical data. In DEMETER, crop yield will be estimated using, in addition, each member of the multi-model forecast ensemble.

Seasonal forecasts (particularly of El Niño) are of great interest in the finance sector, in providing information for pricing so-called weather derivatives. At present many of the weather derivative contracts are for so-called heating or cooling degree days (either positive or negative differences between the daily mean temperature and a reference temperature of 18 °C, summed over the days in the contract period). In principle, an ensemble of seasonal forecasts can give a probability distribution of either heating or cooling degree days for a particular location. In practice, as with all the examples discussed so far, one needs to take care to remove model biases, and to downscale to the specific site for where the contract is made. It should be noted that 10-day ensemble forecasts can also play a role in the weather derivative market, since, based on the ensemble, the probability of the contract exceeding or falling short of the strike value, 10 days in advance, can be estimated.

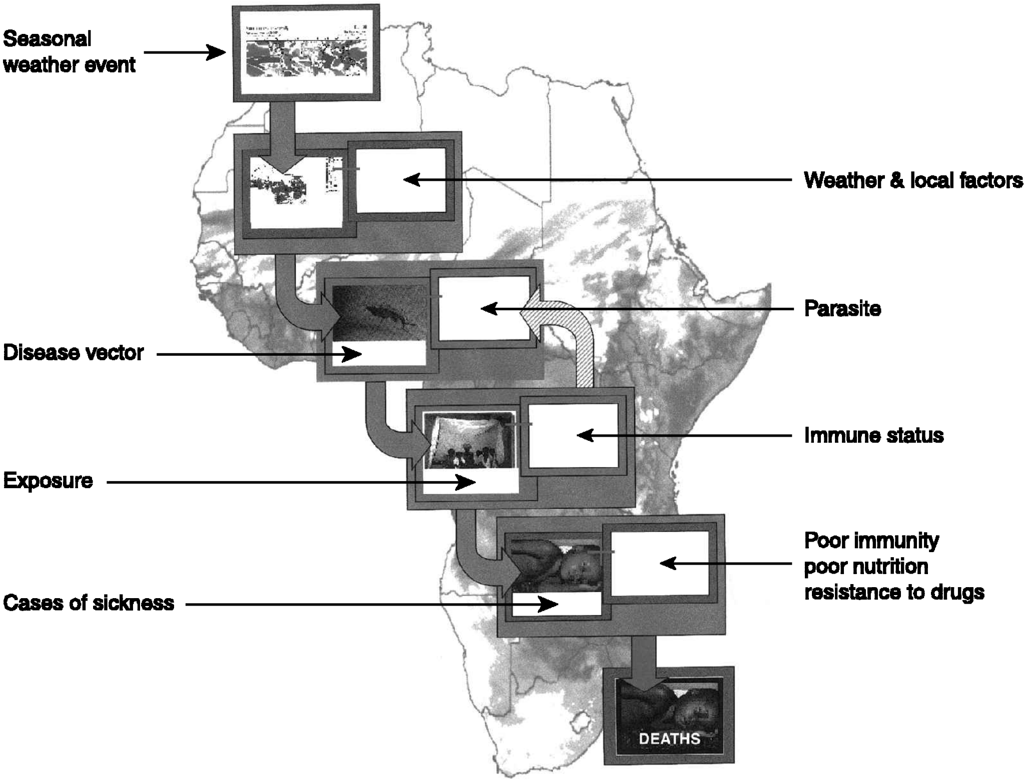

Figure 11. A schematic showing the DEMETER application model for the prediction of the incidence of malaria in tropical Africa. One of the inputs is the seasonal forecast, though there are many other factors which determine mortality. If the risk of an epidemic is forecast to be sufficiently high, then spraying of stagnant waters, and targeting of anti-malarial drugs, can be undertaken in advance. (From Thomson and Connor 2001.)

(*d*) *Ensemble forecasts on decadal time-scales*

Ministers 'too simplistic' on climate change. *Headline: The Times, 16 April 2001.*

In autumn 2000 and winter 2000/2001, parts of the British Isles suffered what was reported to be the wettest seasons on record. Many parts of the country were under water, causing misery to many thousands whose houses lay in the path of flood water. The question on everyone's mind was: has global warming contributed to these floods? Some politicians tended to claim that these floods were somehow proof of global warming, whilst scientists argued that it is impossible to attribute specific climate anomalies to global warming. To support this argument further, it was noted that the flooding was associated with some pattern of natural variability. The scientists claimed that the fact that the previous 12 months had been the wettest on record does not constitute proof of global warming.

However, whilst some of the politicians' sound bites may well have been simplistic, the scientists' response somehow seems evasive, and fails to address an issue of legitimate concern. However, there is a 'third way' in which the layman's concerns can be addressed in a scientifically rigorous manner: by analysing climate change ensemble integrations from the perspective of risk assessment.

An example of a probability forecast of climate change is given in Fig. 12 (based on a methodology described by Räisänen and Palmer (2001)). It has been calculated using

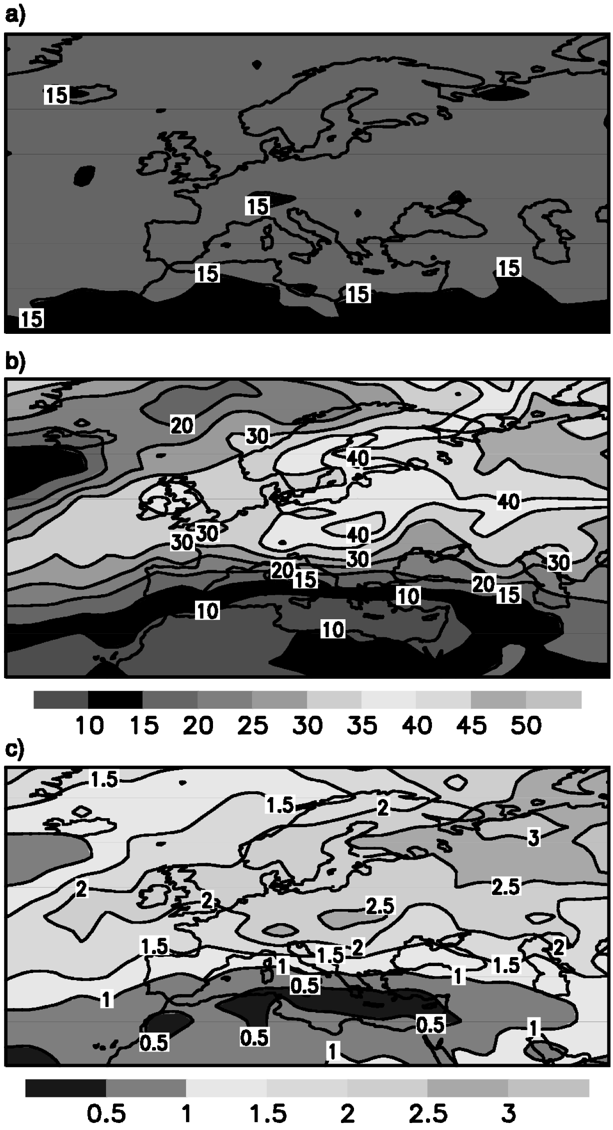

Figure 12. A probabilistic analysis of climate change based on CMIP2 multi-model ensemble integrations with current and with increasing levels of CO_2. (a) The probability of a wet winter defined from the control ensemble with 20th century levels of CO_2 (based on the event E_1 (see text): December to February rainfall greater than the mean plus one standard deviation). (b) The probability of a wet winter (also based on the event E_1), but using data from the ensemble with transient increase in CO_2, and calculated around the time of CO_2 doubling (years 61–80 from present). (c) The ratio of the (b) to (a) values, giving the change in the risk of a wet winter, arising from man's impact on climate.

data from CMIP2 (the second phase of the Coupled Model Intercomparison Project (Meehl *et al.* 2000)), comprising a 19-member ensemble of model integrations with both present day and increasing (1% year^{-1}) levels of CO_2 (the increase in CO_2 is the same in all the experiments, so that forcing-related uncertainty is not represented). Figure 12(a) shows the (CMIP2 ensemble-based) probability that December–February (DJF) mean precipitation is at least one standard deviation greater than normal, based on the integrations with 20th century CO_2. (Similar results for two standard deviations above normal are described by Palmer and Räisänen (2002).) We will call this a 'wet winter'; the probability of occurrence is about 15%. Figure 12(b) shows the probability for the same event (i.e. using a mean and standard deviation based on 20th century seasonal rainfall values) but for the climate-change ensemble between 60 and 80 years into the integrations, around the time that CO_2 has doubled. It can be seen that over the British Isles and much of central and northern Europe the probability of occurrence of this event has increased significantly.

Figure 12(c) shows the ratio of the probability in Fig. 12(b) to the probability in Fig. 12(a). It can be interpreted as giving the increase in the risk of wet winters in the second half of the 21st century. On the basis of these results, a statement such as: 'The probability of the coming winter being wet will double over the next 50 years', is headline catching, scientifically sound and addresses the concerns of the man in the street.

Of course, a wet winter according to this definition does not imply a winter with widespread flooding. However, on the basis of the schematic in Fig. 2, a hydrological model for a particular catchment area could be driven by the NWCP model ensemble output, and the risk of flooding in the 20th and 21st centuries calculated. However, for these more exceptional events, reliable probabilities cannot be estimated directly from 19-member ensembles. Moreover, it is unlikely that with current resolution, climate-model precipitation fields are sufficiently well simulated that realistic risk estimates could be calculated. Both of these factors, ensemble size and resolution, argue for the need for larger computers.

The use of ensembles as a tool for risk analysis of climate change will be shown later to be more economically valuable than the more conventional ensemble-mean forecast (as shown, for example, in the Intergovernmental Panel for Climate Change assessments (e.g. IPCC 2001)).

3. THE POTENTIAL ECONOMIC VALUE OF ENSEMBLE FORECASTS

> The most appropriate system seems therefore to be to leave to the clients concerned by the warning to form an idea of the value of loss/cost and to issue the warnings in such a form that the larger or smaller probability of the event gets clear from the formulation. The client may then himself consider if it is worth while to make arrangements of protection, or to disregard a given warning.
> *Ångström, 1922*.

In the previous section, we discussed specific applications of ensemble forecasts on time-scales of days to decades. Overall, the ensemble method is becoming well established, and, rather than query whether such systems will be part of the meteorological and climatological prediction in the future, we can look forward to improvements in EPSs, as computer power increases. There are numerous ways in which improvements can be envisaged: higher-resolution forecast models, improved estimates of initial and model uncertainty, and larger ensemble size. Which improvement should be chosen for operational implementation? In principle the application models could be run on one set of ensembles with increased resolution, and another set with increased ensemble size, and some synthesis of increased economic value made. However, this proposal might not be practicable; not least the application models may be proprietary.

Rather, one could ask whether there is a more idealized and generic application model, that in some sense captures the essence of commercial decision making, but where different users are represented by different values of some model parameter. With such a model, a synthesis of increased value can be made by averaging over the values of the user parameter. With a sufficiently simple model, it may then be possible to relate potential economic value to more conventional meteorological skill scores.

There is such a model. It was first proposed by Anders Ångström (Ångström 1922), grandson of the eponymous physicist after whom the atomic unit of distance is named. Ångström's work was rediscovered by Liljas and Murphy (1994). Over the years, Ångström's model has been refined and developed by Thompson (1952), Murphy (1977), Katz and Murphy (1997) and Richardson (2000a, 2001).

Consider a decision process for which there are just two alternatives, to take some form of precautionary action or to do nothing, the choice depending on an assessment of the likelihood that a given weather or climate event E will occur or not. Taking action incurs a cost C irrespective of the outcome. If E does occur and no action has been taken, then a loss L is incurred.

Let us consider a practical example. The weather event could be the occurrence of ice on roads, and the decision for a local authority is whether or not to grit the roads. In this case, C would be the cost of the gritting procedure while L would be the economic loss due to traffic delays and increased accidents on icy roads. According to Thornes and Stephenson (2001) the cost/loss ratio C/L for this application is about 0.1. Other applications, on the seasonal and climate-change time-scales, are discussed below.

In general, the decision maker wishes to pursue the strategy that will minimize expense over a large number of situations where decisions have to be made. Suppose the event E has a climatological frequency $\overline{o}$. This climatological information can be used to decide (every time a decision needs to be made), either always to take action or never to take action. Always taking action incurs a cost C on every occasion; never taking action incurs, on average, a loss L on a fraction $\overline{o}$ of occasions. If we define M_{c} as the mean expense associated with a knowledge of climatology only, then $M_{\mathrm{c}} = \min(C, \overline{o}L)$. Notice that climatology is of no help at all in the decision process if $C/L = \overline{o}$.

Consider a deterministic forecast system producing categorical (i.e. yes/no) forecasts of E (it could even be a persistence forecast, for example). Using this, the decision maker could decide to take action when E was forecast, otherwise to take no action. The 'potential economic value' of this forecast system can be measured by the relative measure

$$V = \frac{M_{\mathrm{c}} - M_{\mathrm{f}}}{M_{\mathrm{c}} - M_{\mathrm{p}}} \tag{3}$$

where M_{f} is the mean expense associated with the forecast system and M_{p} is the mean expense associated with a perfect deterministic forecast (i.e. $M_{\mathrm{p}} = \overline{o}C$). Hence $V = 1$ for a perfect forecast, whilst climatological information has $V = 0$. As shown, for example by Richardson (2000a), it is straightforward to express V in terms of C/L, $\overline{o}$, and the so-called hit rate and false alarm rate of the single deterministic forecasts.

On the other hand, suppose the forecast system is an EPS, producing probability forecasts of the event E. For a user with $C/L \sim 1$ (the cost of action is expensive, and/or the losses are not significant), it would be appropriate to take action only when it is relatively certain that E will occur, i.e. when the EPS predicts E with high probability. By contrast, users with $C/L \ll 0$ (the cost of action is cheap, and/or the loss is substantial) would want to take action even if E is forecast with only a small probability. In general, given a probability forecast of E, users will take action when the probability exceeds some threshold p_*, determined by C/L. For a well-calibrated EPS, $p_* = C/L$.

Figure 13 shows an example of potential economic value for probabilistic medium-range predictions of rainfall based on the operational ECMWF EPS system. Two different rainfall events are shown: precipitation greater than 1 mm day^{-1} and greater than 20 mm day^{-1}. There are three lines on each plot, showing the value of the

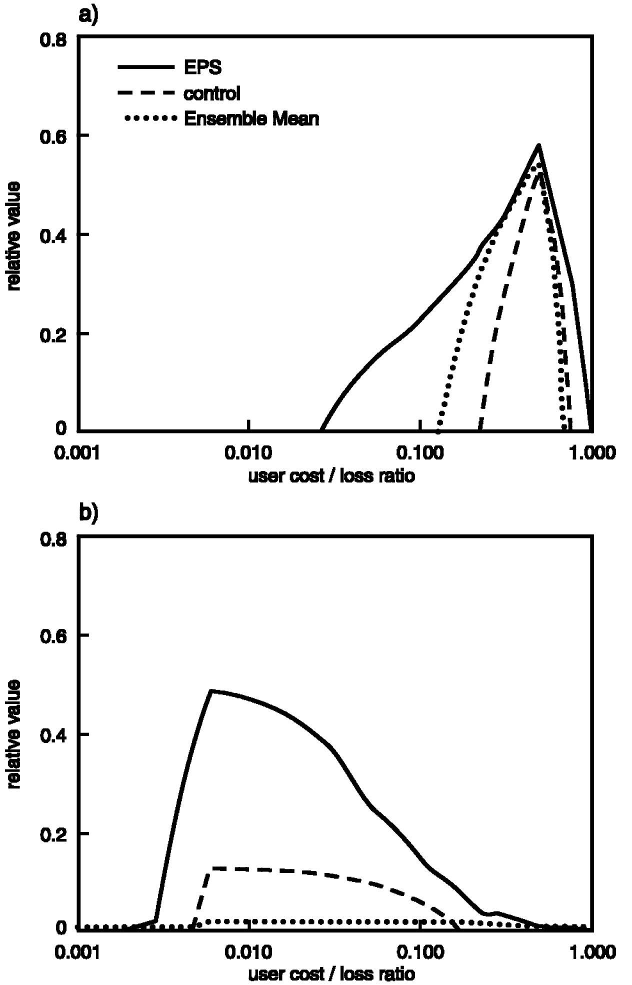

Figure 13. Potential economic value (normalized) of ECMWF forecasts as a function of user cost/loss ratio, for day 5 forecasts (December to February 1999/2000) for two events: (a) rainfall greater than 1 mm day^{-1}, and (b) rainfall greater than 20 mm day^{-1}. Solid: ensemble forecasts, dashed: single deterministic forecasts, and dotted: ensemble mean forecast.

EPS itself, the value of the single deterministic control forecast, and the value of the ensemble-mean consensus forecast. Note that the C/L associated with maximum value varies with the event. This reflects the fact that in general the forecasts are most valuable to users whose C/L is close to the climatological frequency of the event.

For the 1 mm event, it can be seen that the three forecast systems have similar maximum value, but the ensemble forecast benefits a broader range of users. For the more extreme precipitation event, the EPS has clear value over that of either of the two single deterministic forecasts; indeed, for this event, the consensus forecast is quite useless, even compared with the single deterministic control forecast. The reason is straightforward; the consensus forecast hardly ever predicts an extreme event, and, therefore, based on decisions made using the ensemble-mean forecasts, the decision maker will hardly ever take protective action. For users with small C/L, decisions based on the ensemble-mean forecast could be economically disastrous.

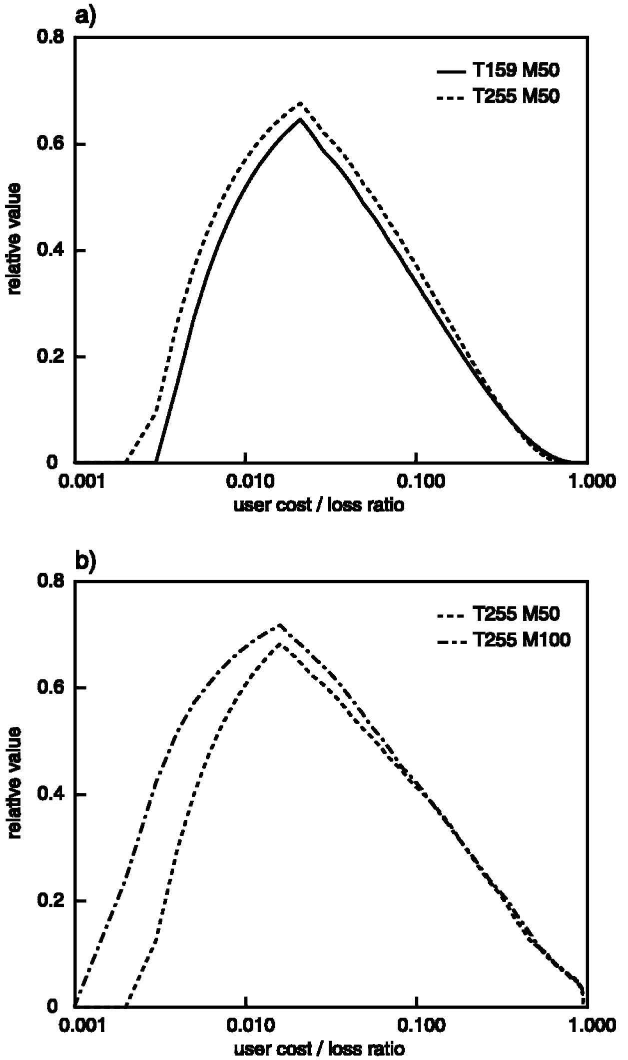

Figure 14. The impact of (a) increasing model resolution (50-member ensembles at TL159 and TL255) and (b) increasing ensemble size (50-member to 100-member ensembles at TL255 resolution) on the potential economic value of ensemble forecasts for the extreme event: 500 hPa geopotential height anomaly more than two standard deviations below normal. Based on approximately 30 different initial dates in November/December 1999.

We can use potential economic value to quantify the impact of increasing both model resolution and ensemble size. For example, Fig. 14(a) shows the impact of decreasing model grid spacing from TL159 (spectral truncation), equivalent to about 120 km, to TL255, equivalent to about 80 km. The event we consider here is the 'extreme' event that the 500 hPa height anomaly is more than two standard deviations below normal. It can be seen that the value is enhanced (equivalent to about a 12-hour gain in value) for a range of users. The principal users for whom value is not enhanced are those whose $C/L \sim 1$. These users will only take precautionary action when almost all ensemble members agree that E will occur. On this basis, it can be said that increasing model resolution has no real impact for highly predictable forecast situations.

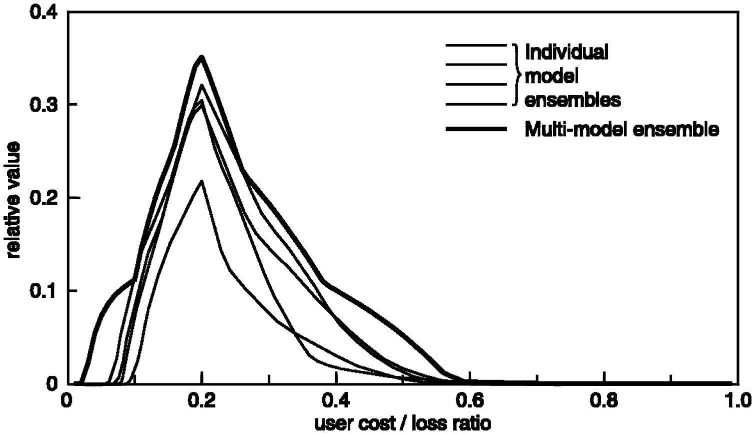

Figure 15. Potential economic value of seasonal integrations (with prescribed sea surface temperature anomalies) from four individual models, and for the multi-model ensemble for the event: 850 hPa temperature anomaly more than 1 K below normal. (From Palmer *et al.* 2000.)

By comparison, Fig. 14(b) shows the impact on value of increasing ensemble size from 50 to 100 members (at TL255 truncation). Compared with increasing resolution, increasing ensemble size only has an impact on users with very small cost/loss ratio. However, for such users, the impact on value is greater than could be obtained through an increase in model resolution. (Note that the computational cost of increasing resolution from TL159 to TL255 is about twice the cost of increasing ensemble size from 50 to 100; hence a fairer comparison would be with an increase in ensemble size to 150.) Which EPS enhancement should be implemented? Obviously it rather depends on whether EPS customers have C/L ratios which are relatively uniformly distributed between 0 and 1, or whether they are all weighted towards values $\ll 1$. We will return to this point in the next section.

As another example of the use of potential economic value, Fig. 15 shows value based on data from the seasonal forecast PROVOST* (PROVOST 2000) project. In PROVOST, ensembles of 120-day integrations with prescribed observed sea surface temperature were made with four different general-circulation models. Shown is the value for the event that 850 hPa temperature is more than 1 degC below normal, based on 15 years of data. The results show that the multi-model ensemble has greater value, for a range of users, than the individual single-model ensembles (Palmer *et al.* 2000). Data from the PROVOST project have been used to show the potential value of seasonal forecasts for the UK food sector (Foresight 2001).

Ångström's model can also be applied to the climate-change problem (Räisänen and Palmer 2001; Palmer and Räisänen 2002). Here we make use of the CMIP2 integrations shown in Fig. 12. Imagine a set of individuals faced with a long-term investment decision: buy either domestic property A (which is sited in a visually-attractive location, but one which might be prone to flooding), or domestic property B (which is insensitive to weather but lies in an area which is likely to be earmarked for high-density development). We assume all values are measured relative to some inflation-adjusted index, and base the calculations on anticipated climate change over the next 80 years. With E_n denoting seasonal-mean precipitation at least n standard deviations above normal (with statistics based on 20th century climate), then if no events

* PRediction Of climate Variations On Seasonal to interannual Time-scales.

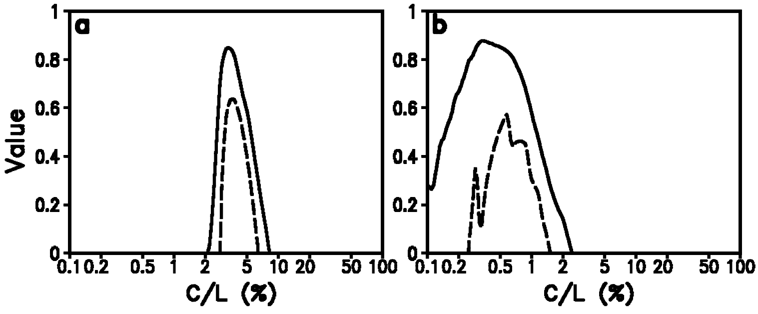

Figure 16. The potential economic value of probabilistic projections of climate change based on a generic binary decision model (see text for details). C and L are possible financial losses, relative to some reference index, associated with two alternative investments. Solid: value based on the CMIP2 perturbed ensemble probability of E_n (seasonal mean precipitation at least n standard deviations above normal) over the next 80 years. Dashed: value based on an estimate of E_n from a single member of the CMIP2 perturbed ensemble. When the value is zero or less, the corresponding estimate of the probability of E_n is no better than an estimate based on the control ensemble. A value of one would correspond to a perfect decision strategy. (a) $n = 2$, and (b) $n = 3$. (From Palmer and Räisänen 2002.)

E_n occur, the value of A is expected to exceed that of B by an amount C. On the other hand, the value of A will linearly decrease by an amount ΔL every time E_n occurs.

We assume, for simplicity and to minimize sampling effects, that the devaluation ΔL applies if E_n occurs, irrespective of season. For each grid point, let $p(t_{ij})$ denote an estimate of the probability of occurrence of E_n for season i $(i = 1, \ldots, 4)$ and year j $(j = 1, \ldots, 80)$. Then the expected devaluation of A due to E_n is $\sum_{j=1}^{80} \sum_{i=1}^{4} p(t_{ij})\Delta L = \overline{p}L$, where $\overline{p}$ is the average probability of occurrence of E_n over the 80-year period, and $L = \Delta L \times 320$. Hence, a rational risk-neutral decision strategy is: buy A if $C > \overline{p}L$, buy B if $C < \overline{p}L$, and toss a coin in the rare case that $C = \overline{p}L$.

Imagine three groups of individuals, each uniformly distributed around the (land points of the) globe. The individuals in group 1 estimate $\overline{p}$ from the probability of occurrence of E_n in the full perturbed ensemble. By contrast, the individuals in group 2 estimate $\overline{p}$ from the number of occurrences of E_n in a single randomly-chosen member of the perturbed ensemble. (For example, they might rely on a specific national model.) The individuals in group 3 estimate $\overline{p}$ from the number of occurrences of E_n in the perturbed-ensemble consensus projection.

We estimate the value V_i of the investments made by these three groups, assuming that the true number of occurrences of E_n during the 80-year period is given by a randomly chosen member of the perturbation ensemble (which is excluded from the climate prediction dataset). If $V_i = 0$, then subset i's estimate of $\overline{p}$ is no better than an estimate which assumes 21st century climate will be the same as 20th century climate. On the other hand, perfect investment decisions have a value $V_i = 1$. In calculating V_i we average over all choices of verifying model. For V_2 we also average over all choices of forecast model.

Figure 16 shows V_1 and V_2 for $n = 2$ and $n = 3$, as functions of C/L. We do not show V_3 since it is never greater than zero. The reason for this is straightforward: the number of occurrences of extreme climate events is severely underestimated in the consensus projection. This arises because ensemble averaging tends to produce smooth fields which are unrealistically biased towards the climate-mean state.

As before, if C/L is either sufficiently small, or sufficiently large, then accurate ensemble forecasts are clearly not needed to make a good investment decision (buy B or A respectively). Hence, the perturbed ensemble only has positive economic value for a subset of possible C/L. It tends to be most valuable when C/L is comparable with the control ensemble estimate of $\overline{p}$. Since $\overline{p}$ is smaller for E_3 than for E_2, then the perturbed ensemble provides value for smaller C/L, when decisions are based on the occurrences of E_3 rather than E_2. It can be seen that the value of the perturbed ensemble is never less than the value of a single deterministic projection, and that, overall, the difference in value between the ensemble and the single deterministic projection is larger for E_3 than for E_2. This is because the relative unreliability of the single deterministic projection is larger for E_3 than for E_2.

In fact, much larger ensembles are needed to provide reliable estimates of the probability of E_3 on a regional basis (e.g. over the British Isles). In addition, the resolution of present-day climate models (with grid sizes of hundreds of kilometres) is also inadequate for extreme-precipitation risk analysis. For example, it would clearly be of more direct relevance for the decision problem discussed above, if E was defined by some specific river bursting its banks during a particular season. However, estimating the future risk of this type of event would require feeding climate-model output into a basin-specific hydrological model. In general, climate-model grid sizes on the order of ten kilometres would be necessary to simulate precipitation statistics adequately over a typical catchment area (though for rivers with large catchment basins, such as the Brahmaputra, coarser climate-model grids may be adequate). An ability to estimate reliably the probability of extreme climate events is an important factor in defining future computational requirements for the climate-change problem.

4. On the relationship between value and skill

Let us now return to Fig. 1. We have discussed a simple measure of economic value which characterizes users by their cost/loss ratios. Since value is user dependent, there can be no unique relationship between value and meteorological skill that is relevant for all individual users.

Despite this, there is a relation between potential economic value, summed over all users, and the Brier score. The Brier score (e.g. Wilks 1995) is one of the standard measures used to evaluate probabilistic forecasts. In some sense it is an adaptation of the concept of a 'mean-square' measure of skill applied to a probability forecast. As before, we consider an event E which, for a particular ensemble forecast, occurs in a fraction p of ensemble members. If, when the verifying analysis is available, E actually occurred, then let $v = 1$. Repeat this over a sample of N different ensemble forecasts, so that p_i is the probability of E in the ith ensemble forecast $1 \leqslant i \leqslant N$, and $v_i = 1$ or $v_i = 0$, depending on whether E occurred or not in the ith verification. In practice the N ensemble forecasts are taken not only from different starting dates, but also over different individual grid points. The Brier score is

$$b = \frac{1}{N} \sum_{i=1}^{N} (p_i - v_i)^2, \quad 0 \leqslant p_i \leqslant 1,\ v_i \in \{0, 1\}. \tag{4}$$

Hence $b = 0$ for a perfect deterministic forecast. Usually, the Brier score is converted to a so-called Brier skill score

$$B = 1 - \frac{b}{b_{\mathrm{c}}} \tag{5}$$

where b_c is the Brier score of a forecast where the climatological probability $\overline{o}$ of the event is always forecast. Hence $B = 1$ for a perfect deterministic forecast, and $B \leqslant 0$ for a forecast which is no better than climatology.

On the other hand, the mean potential economic value integrated over all users, assuming their C/L ratios are uniformly distributed between 0 and 1, is directly related to the Brier skill score of the probability forecast (Murphy 1977; Murphy and Ehrendorfer 1987; Ehrendorfer and Murphy 1988; Wilks and Hamill 1995). Hence, a probabilistic generalization of the familiar meteorological r.m.s. error can be shown to give the integrated potential economic value, providing a typical range of customers would have C/L ratios uniformly distributed in the range between 0 and 1. Is this a reasonable assumption? There are indications that it is not: for example, Murphy (1977), Roebber and Bosart (1996) and Thornes and Stephenson (2001) all describe applications where $C/L \ll 1$.

Suppose that the C/L ratios of the customers for an EPS are well-defined and known. Can we define a meteorological skill score which reflects the overall value of the forecast system to these customers? Yes! If the Brier skill score is equal to mean potential economic value integrated with respect to a uniform distribution of C/L ratios, then we can define a generalized (Brier) skill score which is equal to the mean potential economic value integrated with respect to a specified non-uniform distribution of C/L ratios, consistent with the range of actual customers.

As discussed by Richardson (2001), the impact of changes in EPS design on skill scores can be quite sensitive to which generalized skill score is used. For example, Fig. 17(a) shows estimates of the Brier skill score as a function of ensemble size, for a range of events with different predictabilities. It can be seen that beyond ensemble sizes of 50, there is little impact of ensemble size on Brier skill score, even for events with small intrinsic predictability. (It should be noted that this calculation is based on a given ensemble probability density function which is sampled with different finite samplings. In practice, ensemble sizes can also be increased, e.g. as in Fig. 14(b), by increasing the number of phase space directions which are sampled at initial time. The calculation shown here does not include this aspect and should therefore be thought of as a lower bound on the impact of ensemble size on skill.)

On the other hand, Fig. 17(b) shows the same calculation but for a user C/L distribution based on a β-distribution (Wilks 1995) for which most users have small C/L (mean $= 0.2$, standard deviation $= 0.36$). In can now be seen that increasing ensemble size will certainly have an impact on skill, for ensemble sizes in excess of 50.

This example makes the point that whilst we need skill scores in order to be able to decide whether or not to implement possible changes to the operational forecast system, these skill scores should reflect the range of customers for whom the forecast system is serving. However, this puts a premium on being able to quantify the characteristics of these customers not only in terms of their C/L (recognizing that in practice this may well be difficult to quantify), but also in terms of the weather events to which their operations are sensitive.

As mentioned, the Brier skill score corresponds to mean value for a range of customers with uniform distribution of cost/loss ratio. Suppose we were able to identify such a range of customers and started selling these ensemble forecasts. Assume, not unreasonably, that the price of the forecasts is the same for all customers. After a while, the customers who do not perceive sufficient value, will start to drop out. As mentioned, the value of a forecast system is likely to be greatest (compared with climatology) for users whose $C/L \sim \overline{o}$. Hence the customers that drop out will tend to be those for whom C/L is very different from $\overline{o}$. Hence, after a certain period of time, the distribution of C/L

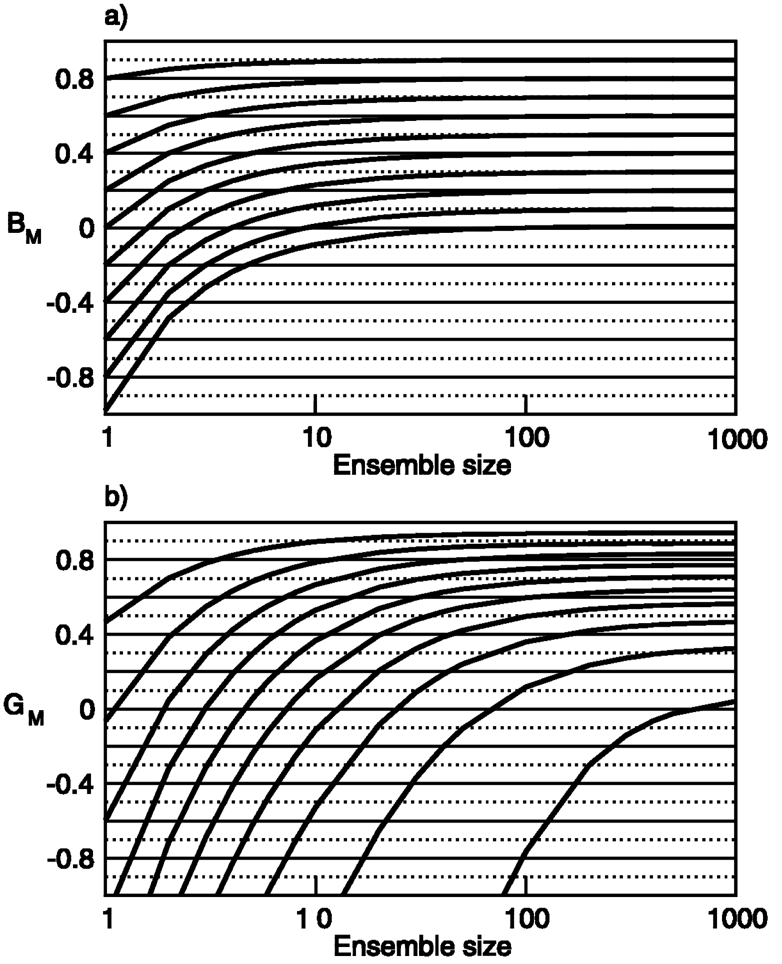

Figure 17. (a) Theoretical calculation of the impact of ensemble size on Brier skill score for a range of events. (b) Impact on ensemble size on a generalized Brier skill score, defined by integrating potential economic value with respect to a set of users with small cost/loss ratio. (From Richardson 2001.)

for customers who keep buying the forecasts will be peaked close to $\overline{o}$. In some limiting sense, this distribution can be approximated by a delta function at $\overline{o}$. The potential economic value, integrated with respect to such a delta function, gives the maximum possible value, and, as shown by Richardson (2000b), there is a close (almost linear) relationship between maximum value and the meteorological skill score known as 'area under the relative operating characteristic (ROC) curve' (Stanski *et al.* 1989). ROC area is a common probabilistic skill score in addition to the Brier skill score, and measures the relative fraction of so-called hit and false-alarm rates associated with probability forecasts of the meteorological event in question. Insofar as a distribution of users with $C/L \sim \overline{o}$ is more realistic than a distribution of users uniform in C/L, then the ROC area may reflect value better than does the Brier skill score.

5. Conclusions

It has been argued that ensemble forecasts provide a quantitative tool for weather and climate risk assessment, and have value which exceeds that of more conventional single deterministic forecasts. It has been stressed that commercial decisions are not necessarily based on events which are likely to occur, but on events that may be intrinsically unlikely, but which if they did occur, would imply great financial loss.

The full value of an ensemble prediction cannot be utilized without adequate statistical post-processing of the NWCP output. Whilst such post-processing is needed to remove possible biases that exist in the numerical output, it is also important to downscale from the gridbox to some site-specific location. Part of this downscaling will imply an inflation of the forecast probability distribution. For example, if the gridbox mean rainfall is predicted to be 20 mm day^{-1} by one ensemble member, and this rainfall is mainly convective, then at a point within the gridbox, there may be some probability that the actual rainfall rate will exceed say 100 mm day^{-1}.

An automated approach to ensemble forecasting has been advocated, one where the individual members of the ensemble are linked with the user application models. However, this does not imply that the human forecaster has been taken out of the chain. For the foreseeable future, there will be a need for a human forecaster to interpret the results, and most importantly to bolster customers' confidence in making use of the numerical forecasts. However, the sort of questions that the forecaster should be able to answer quantitatively, will not only include, 'What is the weather going to be like?', but also, 'What is the risk of a certain type of weather occurring?'. In this respect, the forecaster will have to know the weather conditions to which a given customer's operations are particularly sensitive.

One time-scale that has not been explicitly discussed in this paper is the very short range—say up to a day ahead. It might be thought that weather forecasting is essentially predictable at this range. Whilst this may be true for cyclone scales, it is certainly not true on the mesoscale. For such scales there is evidence that uncertainty in the model equations is a major source of forecast uncertainty, and that this needs to be taken into account in an ensemble forecast, possibly though a multi-model ensemble approach.

In concluding this paper, one could ask whether there is a case for including the estimation of uncertainty as a core part of the school science curriculum. There is certainly a good case to be made for this, and weather and climate prediction provides an excellent example of the importance of estimating uncertainty. But there is a possible drawback. A simplistic analysis of the problem might suggest that science itself is uncertain. This is, of course, far from the truth. The sort of uncertainty that has been discussed in this paper arises from the *application* of basic science, not from basic science itself. In this so-called post-modern world, where even mathematical theorems are sometimes judged to have a sociological bias (e.g. Graham 2001), this distinction between science and its application needs to be firmly stressed.

Acknowledgements

My thanks to J. Barkmeijer, R. Buizza, F. Grazzini, E. Klinker A. Persson and D. Richardson (ECMWF), J. Räisänen (The Rossby Centre, Sweden), L. Smith (London School of Economics), J. Taylor (University of Oxford, Said Business School), J.-M.Terres (Joint Research Centre, European Commission, Italy), M. Thomson (University of Liverpool School of Tropical Medicine) for many helpful discussions and for supplying me with unpublished material. My thanks also to R. Hine and A. Bowen for the most excellent graphical support, and to A. Mernandez for the quote at the beginning of section 2(c).

REFERENCES

Ångström, A. 1922 On the effectivity of weather warnings. *Nordisk Statistisk Tidskrift*, **1,** 394–408

Buizza, R. and Palmer, T. N. 1995 The singular vector structure of the atmospheric global circulation. *J. Atmos. Sci.*, **52,** 1434–1456

Doblas-Reyes, F. J., Déqué, M. and Piedlievre, J.-P. 2000 Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Q. J. R. Meteorol.Soc.*, **126,** 2069–2088

ECMWF 2000 'The first twenty-five years (1975–2000)'. European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, RG2 9AX, UK

Ehrendorfer, M. 1994 The Liouville equation and its potential usefulness for the prediction of forecast skill. Part I. Theory. *Mon. Weather Rev.*, **122,** 703–713

Ehrendorfer, M. and Murphy, A. H. 1988 Comparative evaluation of weather forecasting systems: sufficiency, quality and accuracy. *Mon. Weather Rev.*, **116,** 1757–1770

Evans, R. E., Harrison, M. S. J., Graham, R. J. and Mylne, K. R. 2000 Joint medium-range ensembles from the Met Office and ECMWF systems. *Mon.Weather Rev.*, **128,** 3104–3127

Foresight 2001 'Seasonal weather forecasting for the food chain'. Final report of the Foresight Seasonal Weather Forecasting for the Food Chain Steering Group. Ed. J. Griffin. (Available at www.foresight.gov.uk/fcci)

Gleick, J. 1992 *Genius. Richard Feynman and modern physics.* Little, Brown and Company, London

Graham, L. R. 2001 Do mathematical equations display social attributes? *Math. Intelligencer*, **22,** 31–36

Graham, R. J., Evans, A. D. L., Mylne, K. R., Harrison, M. S. J. and Robertson, K. B. 2000 An assessment of seasonal predictability using atmospheric general circulation models. *Q. J. R. Meteorol. Soc.*, **126,** 2211–2240

Harrison, M., Palmer, T. N., Richardson, D. S. and Buizza, R. 1999 Analysis and model dependencies in medium-range ensembles: Two transplant case-studies. *Q. J. R. Meteorol. Soc.*, **125,** 2487–2515

Hoffschildt, M., Bidlot, J.-R., Hansen, B. and Janssen, P. A. E. M. 1999 'Potential benefit of ensemble forecasts for ship routeing'. ECMWF Technical Memorandum 287. ECMWF, Reading, UK

Houtekamer, P. L., Lefaivre, L., Derome, J., Richie, H. and Mitchell, H. L. 1996 A system approach to ensemble prediction. *Mon. Weather Rev.*, **124,** 1225–1242

IPCC 2001 'Climate change 2001: The scientific basis'. Contribution of Working Group I to the third Assessment Report of the Intergovernmental Panel on Climate Change. J. T. Houghton, Y. Ding, D. J. Griggs, M. Noguer, P. J. Van der Linden, X. Dai, K. Maskell and C. A. Johnson, Eds. Cambridge University Press, Cambridge and New York

Katz, R. W. and Murphy, A. H. (Eds.) 1997 *Economic value of weather and climate forecasts*. Cambridge University Press

Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D. and Williford, E. 2000 Multimodel ensemble forecasts for weather and sesaonal climate. *J. Climate*, **13,** 4196–4216

Liljas, E. and Murphy, A. H. 1994 Anders Ångström and his early papers on probability forecasting and the use/value of weather forecasts. *Bull. Am. Meteorol. Soc.*, **75,** 1227–1236

Lorenz, E. N. 1993 *The essence of chaos*. University of Washington Press

Mason, B. J. 1966 The role of meteorology in the national economy. *Weather*, **21,** 382–393

Meehl, G. A., Boer, G. J., Covey, C., Latif, M. and Stouffer, R. J. 2000 The coupled model intercomparison project. *Bull. Am. Meteorol. Soc.*, **81,** 313–318

Molteni, F., Buizza, R., Palmer, T. N. and Petroliagis, T. 1996 The ECMWF ensemble prediction system: Methodology and validation. *Q. J. R. Meteorol. Soc.*, **122,** 73–119

Murphy, A. H. 1977 The value of climatological, categorical and probabilistic forecasts in the cost–loss ratio situation. *Mon. Weather Rev.*, **105,** 803–816

Murphy, A. H. and Ehrendorfer, M. 1987 On the relationship between the accuracy and value of forecasts in the cost–loss ratio situation. *Weather and Forecasting*, **2,** 243–251

Palmer, T. N. 2000 Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.*, **63,** 71–116

2001 A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Q. J. R. Meteorol. Soc.*, **127,** 279–304

Palmer, T. N. and Anderson, D. L. T. 1994 The prospects for seasonal forecasting. *Q. J. R. Meteorol. Soc.*, **120,** 755–793

Palmer, T. N. and Räisänen, J. 2002 Quantifying the risk of extreme seasonal precipitation events in a changing climate. *Nature*, **415,** 512–514

Palmer, T. N., Brankovic, C. and Richardson, D. S. 2000 A probability and decision-model analysis of PROVOST seasonal multi-model ensemble integrations. *Q. J. R. Meteorol. Soc.*, **126,** 2013–2033

Poincaré, H. 1909 *Science and method*. Reprinted by Dover Publications, 1952

PROVOST 2000 Special issue, No. 567. *Q. J. R. Meteorol. Soc.*, **126,** 1989–2350

Räisänen, J. and Palmer, T. N. 2001 A probability and decision-model analysis of a multi-model ensemble of climate change simulations. *J. Climate*, **14,** 3212–3226

Richardson, D. S. 2000a Skill and relative economic value of the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.*, **126,** 649–668

2000b 'Applications of cost–loss models'. Proceedings of seventh workshop on meteorological operational systems. ECMWF, Reading, UK

2001 Measures of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size. *Q. J. R. Meteorol. Soc.*, **127,** 2473–2489

Roebber, P. J. and Bosart, L. F. 1996 The complex relationship between forecast skill and forecast value: a real-world analysis. *Weather and Forecasting*, **11,** 544–559

Smith, L. A., Roulston, M. S. and von Hardenberg, J. 2001 'End-to-end ensemble forecasting: towards evaluating the economic value of the ensemble prediction system'. ECMWF Technical Memo No. 336. ECMWF, Reading, UK

Stanski, H. R., Wilson, L. J. and Burrows, W. R. 1989 'Survey of common verification methods in meteorology'. WMO WWW Tech. Report No. 8, WMO, TD No. 358

Stockdale, T. N., Anderson, D. L. T., Alves, J. O. S. and Balmaseda, M. A. 1998 Global seasonal rainfall forecasts using a coupled ocean–atmosphere model. *Nature*, **392,** 370–373

Taylor, J. W. and Buizza, R. 2002 Using weather ensemble predictions in electricity demand forecasting. *Int. J. Forecasting.* In press

Thompson, J. C. 1952 On the operational deficiencies in categorical weather forecasts. *Bull. Am. Meteorol. Soc.*, **33,** 223–226

Thomson, M. C. and Connor, S. J. 2001 The development of malaria early warning systems for Africa. *Trends in parasitology*, **17,** 9438–9445

Thomson, M. C., Palmer, T. N., Morse, A. P., Cresswell, M. and Connor, S. J. 2000 Forecasting disease risk with seasonal climate predictions. *Lancet*, **355,** 1559–1560

Thornes, J. E. and Stephenson, D. B. 2001 How to judge the quality and value of weather forecast products. *Meteorol. Apps.*, **8,** 307–314

Toth, Z. and Kalnay, E. 1997 Ensemble forecasting at NCEP and the breeding method. *Mon. Weather Rev.*, **125,** 3297–3319

Wandishin, M. S., Mullen, S. L., Stensrud, D. J. and Brooks, H. E. 2001 Evalulation of a short-range multi-model ensemble system. *Mon. Weather Rev.*, **129,** 729–747

Wilks, D. S. 1995 *Statistical methods in the atmospheric sciences*. Academic Press, London

Wilks, D. S. and Hamill, T. M. 1995 Potential economic value of ensemble forecasts. *Mon. Weather Rev.*, **123,** 3565–3575