THE DEVELOPMENT OF SEASONAL AND INTER-ANNUAL CLIMATE FORECASTING

MIKE HARRISON

Hadley Centre, Met Office, London Road, Bracknell, Berks RG12 2SY
E-mail: mike.harrison@metoffice.com

Abstract. A review of the development and status of seasonal to inter-annual climate forecasting up until 2001 is presented covering not only the successes but also identifying some of the major challenges remaining. Included is discussion on the history of the enterprise; the scientific basis of modern seasonal to inter-annual prediction and its background of predictability theory; the current status of predictions and the measurement of their skill; the experiences and consequences of the 1997–1998 El Niño event; approaches to linking forecasts with applications; and a view to the future.

1. Introduction

Climate forecasting is probably one of the oldest professions in the world. A Babylyonic scroll dated to about 3000 BC may well be the oldest-known example of an attempt to predict the weather that would affect the following season's crops. Later attempts are listed in Holy Scriptures, such as in the story of Joseph in the Bible, while phenological approaches based on perceived behaviours of flora and fauna, or on the sun, moon and stars, have developed throughout the world across many succeeding centuries. Not surprisingly there are many similarities amongst these phenological approaches between countries and continents. Interest in phenology remains strong, and has prompted new investigation by the World Meteorological Organisation's Commission for Climatology, although rigorous examination does not always provide succour to the acolytes (Marriott, 1981).

Society has been exposed to the vagaries of climate variability throughout the millennia, but resilience to and protection from these variations as they affect food supplies, water availability and comfort levels has been gained only in some countries. It is hardly surprising therefore that people have long looked to phenology and to beliefs for ways of knowing the severity of the coming winter and the beneficence of the next summer well ahead in order to prepare. Phenological and belief approaches are still used in many countries, including developed ones, and it is likely that they will continue to be employed long after modern scientific approaches have matured well past their current stage.

It is not a simple task to identify the initial breakthrough that led to the current stage of development of modern scientific seasonal to inter-annual forecasts, although the work of Gilbert Walker[1] is certainly recognised as one of the prime inputs, even though its import was not fully appreciated at the time. Working in

Climatic Change (2005) 70: 201–220

India in the early 20th Century, and tasked as Director-General of the Indian Meteorological Department to find a method of predicting the monsoon rain total, Walker set out to examine statistical relationships in the global atmosphere. There were already clues that climate variations in different parts of the globe may be associated and Walker examined all data available to him in an attempt to identify predictive links for the monsoon rains. His work uncovered the North Pacific Oscillation and the North Atlantic Oscillation, apparent latitudinal "seesaws" in pressure across these basins. Neither aided the monsoon problem. However it was the third seesaw, the "Southern Oscillation", which imparted the stimulus to further study, and continues to provide one input to the prediction methods used by the Indian Meteorological Department today. It was many years, however, before the importance of the Southern Oscillation was comprehended in terms of its inherent global-scale atmosphere/ocean dynamics, and its links to El Niño recognised.

2. A Review of Modern Short-Range Climate Forecasting

The principles of modern short-range climate forecasting are straightforward. The atmosphere works on many spatial and temporal scales, the shortest of which only possess predictability out for a limited time period provided the current state of the system is adequately known. For synoptic scales, those associated with local day-to-day weather variations, this translates to predictability for no more than a few days, given rigorous observation of the state of the atmosphere around the globe. But underlying the atmosphere are the continents and oceans, surface conditions across which change only relatively slowly. Transfers of heat and momentum between these underlying surfaces and the atmosphere can, in appropriate circumstances, act as a slowly varying regulator for the atmosphere, providing a memory which limits the range of weather in affected areas, thus creating a predominant climate anomaly over a period of a season or more.

The best-known example of boundary layer changes forcing climate anomalies on scales up to the global is the El Niño/Southern Oscillation (ENSO),[2] of which Walker's work uncovered only part of the atmospheric component. Many texts outline the dynamics of ENSO,[3] and here it may suffice that when cold water upwelling off the equatorial west coast of South America is replaced by warmer waters from the west, then an El Niño occurs. At the same time the mass of tropical convection that is typically found in the South-East Asian/Indonesian region migrates into the central tropical Pacific Ocean, close to the date line, bringing drier conditions in the west and heavy rainfall to normally-dry islands. There is close linking between the change in sea surface temperatures, the location of the equatorial convective mass, and the changes referred to as the Southern Oscillation by Walker. Other changes in the global circulation accompany the movement of the equatorial convective mass (Glantz et al., 1991), which is the atmosphere's main heat source, and impacts on temperatures and rainfall occur over a season or more as a consequence in regions across, around and remote from the Pacific basin.

Sea surface temperature variations in other tropical oceans additionally affect countries around those basins (e.g. Hastenrath, 1991), although these influences appear to be less spatially extensive than those related to the Pacific and as well act predominantly on decadal time scales (Latif and Barnett, 1994). Mid and high latitude oceanic changes have lesser seasonal scale impacts than those in the tropics (e.g. Rodwell and Folland, 2002). Evidence is growing that changes in land surface conditions, particular the soil moisture content, can also affect atmospheric circulation on seasonal time scales.[4] Perhaps the best known example relates to the Indian monsoon, changes in which are linked to soil moisture deficits in the Himalayan region resulting from melt of winter snowfall (Ferranti and Molteni, 1999).

Seasonal to interannual prediction in general presently uses knowledge of sea surface temperature variations on which to base a forecast of temperature and rainfall conditions in teleconnected parts of the globe. Several prediction approaches are possible. Relationships between recent sea surface temperature anomalies and future climate conditions may be estimated empirically using historical data. Alternatively current sea surface temperature conditions may be persisted into the future as a proxy prediction, or future anomalies themselves predicted using empirical or dynamical methods, and in either case concurrent or additional dynamical methods then used to make the required climate prediction. Other predictors are used at times, such as the Himalayan snow cover mentioned above or various proxy indicators, but the principles remain the same and the focus in the remainder of this paper will stay with the oceanic anomalies.

A complete review of all types of models used in prediction is beyond the scope of this paper,[5] but these may be divided into two main groups. Empirical models are developed using historical data and attempt to represent statistical links between one or more predictors, usually sea surface temperature anomalies averaged across specific regions, and rainfall or temperature anomalies across the area of concern. Typically linear regression is used to derive the associations, but other methods have been employed. Computer-based dynamical models of the ocean and/or atmosphere, often similar to, or even the same as, the models used for weather or climate change prediction, form the other group. In between are a number of approaches that use simplified statistical or statistical/dynamical models for one of the components, either the atmosphere or the ocean. Fledgling coupled models of both the atmosphere and the ocean, recently introduced into operational short-period climate forecasting at a small number of leading centres, represent the most technically advanced approach.

In all cases the objective is to produce a prediction of the average climatic conditions throughout a season across a region measuring several hundred kilometres along each side (Goddard et al., 2001). In general the main predictands of concern to applications are rainfall and temperature, although others, such as numbers of tropical storms and start/end dates of rainfall seasons, have also been attempted. This is such an extensive area of activity that the focus in this paper will be on rainfall prediction, the central concern of agriculture and many other activities including energy conservation, water and fire management. Rainfall is a natural predictand for seasonal forecasts, as it is the changes in oceanic rainfall, forced by modified sea surface temperatures, that result in adjustments to the atmospheric circulation that lead to the variations in rainfall over the continents.[6] Nevertheless rainfall is variable in both space and time and many applications are more sensitive to the timing and amounts of rainfall through a season than they are to the total amount. For this reason users of seasonal forecasts have often requested both more temporal and spatial detail in predictions than is available directly from the global models. These requests, for example, have driven attempts over a number of years to predict local start dates of rainfall seasons using empirical techniques. However the modern approach to obtaining increased detail of information is to use regional climate models (RCMs) (Goddard et al., 2001). RCMs are the higher resolution brothers of the global atmospheric models. Covering areas perhaps the size of Europe, and using controlling information from runs of global models, RCMs provide more detail on both temporal and, particularly, spatial scales than is available from the global models themselves. In particular RCMs have the more detailed underlying topography and so in principle can provide information on rainfall differences associated with mountains and valleys, although it must be noted that development of these models is at an early stage and the practical utility of the information produced remains to be demonstrated.

3. Predictability

Were the links between oceanic sea surface temperatures and continental rainfall directly linear then seasonal prediction would be straightforward. In practice most empirical models make the underlying assumption of linearity and this provides a reasonable starting point. Nevertheless rainfall over land may be influenced by sea surface temperature variations across several parts of an ocean, by variations across more than one ocean, by adjustments in land surface moisture, and by a number of other factors which may have no clear predictability and which are often dismissed as noise. An excellent example of the complexity of the situation related to sea surface temperature changes alone was given during the 1997/98 El Niño event, as will be discussed later, but because of this complexity it is inevitable that even the best-calibrated empirical model will fail at times. Unfortunately, few, if any, empirical models can claim to be well calibrated, as insufficient historical data are normally available from which to develop the models and, additionally, evidence exists that the empirical relationships forming the foundations of the models may in any case change slowly in time (e.g. Kumar et al., 1999).

There is, however, a more key aspect of predictability than the basic complexity of the links, and that is ‘chaos’. A chaotic system is one that, if perturbed by a small amount (in meteorology these ‘small amounts’ can be below normal measurement capabilities), may end up in a different state than it would have had the perturbation not been present. In fact the perturbation may not exist as such but may, and usually is, a consequence of a limited global observing system leading to errors in information provided to the model. Such sensitivity to small changes is not present at all times at a given location in chaotic systems, but in the more complex chaotic systems, including that of the atmosphere, special techniques are required to determine what sensitivities are present and what their consequences are. In meteorology the ensemble approach is used to assess the level of sensitivity of a forecast to perturbations, the impacts on a prediction of which can be substantial even over only a few days. The standard way of producing an ensemble, of which there are a number of variants, is to run the model from a number of slightly different starting points, differences between which being consistent with the errors that can occur in measuring the starting point. Ensembles with over 50 members are now run operationally at some leading centres for periods out to ten days (Mullen and Buizza, 2002), with smaller ensembles generally used for seasonal predictions (Goddard et al., 2001). A further issue is that the models themselves are not perfect and errors developing in the models can provide additional sensitivities. One of the best approaches to encompassing model sensitivities is to use more than one model in an ensemble (Graham et al., 2000). On seasonal time scales the underlying sea surface temperatures tend to constrain the atmospheric circulation to a restricted set of options and so provide the sought-after predictability, but even in the ocean chaotic influences are important. Only a few systems take this oceanic chaos into account at present.

Because of computer processor restrictions operational seasonal forecast ensembles often have less than 10 members, but one system[7] has 40, a figure recently doubled to 80 with the incorporation of a second model from the Met Office in a multi-model configuration. Larger ensembles can be expected in the future. The range of rainfall predictions for a specific location across an ensemble of forecasts, whether from a single model or from multiple models, can be, but need not be, substantial, and from the applications perspective may be daunting. There are two fundamental approaches to presenting ensemble output. The first is to average across all ensemble members and to base the forecast on that average. In theory this is supposed to eliminate the less predictable, smaller-scale, details of the forecast, and certainly produces information easier to use from the applications side, but in practice it has all the limitations of any deterministic prediction, including the possibility of being entirely wrong. The second approach, and the one consistent with chaos theory, is to treat all members of an ensemble as equally likely (or weighted in some manner) and to produce a probability distribution. Although many users, and even some meteorologists, balk at the prospect of using probability forecasts, this approach has many advantages, expanded further later, not least of which is the possibility of providing information on less likely, but nonetheless possible, outcomes. Forecasts of the more extreme events also become more achievable through use of distributions.

Empirical methods can be used to assess chaotic impacts on a prediction, but models currently in practical use do not do so. However a common method used to develop a probability distribution from a deterministic empirical model is to examine the historical performance of the model and to develop a distribution from that information using occasions when the model was providing similar forecasts to a current one (Folland et al., 2001). When this approach is adopted it becomes possible to build consensus forecasts from both numerical and empirical models, a technique often used in Regional Climate Outlook Forums (RCOFs—see below), and thought to produce the best prediction possible (Basher et al., 2001).

4. The Current Status of Seasonal and Inter-Annual Climate Forecasts

When confronted with any form of seasonal prediction the first question usually asked by potential users is “How good are your forecasts?”. Often this is followed by the deterministically-based query “How often is your forecast correct?”. These are two questions the climatological community has difficulty in responding to in a manner that satisfies most users.

Amongst the numerous issues confronting climate scientists when attempting to validate their models and verify their rainfall forecasts is the problem of data homogeneity when the model is providing spot information at scales of a few hundred kilometres. These spot values are typically averaged to produce a single figure for each model cell, averages which then need to be compared with observations from a set of points somewhat randomly scattered around the cell, data records with inevitable missed observations, and reflecting topography not resolved by the model. Undoubtedly the noise introduced by this process impacts on validation and verification results. The difficulties are compounded when there is a need to compare not a series of single deterministic predictions but a sequence of probabilistic ones. A further important consideration is that there are relatively few realisations of seasonal predictions with which to work. As far as is known to the author no empirical technique employs a data series that exceeds 100 years in length, and most are based on much shorter series, some perhaps no more than 30 years in length. Often with only one prediction per year in the series, these are statistically seriously limited data sets for assessing deterministic systems; for probabilistic systems they are undoubtedly inadequate. Numerical modelling studies, because of data problems and of computing costs, have fewer realisations still. One of the most extensive currently-available series is that from the PROVOST experiment, that reached across 15 years.[8] The succeeding DEMETER project is planned to cover 40 years,[9] bountiful by any current measure of numerical seasonal prediction experimentation, inadequate by any statistical measure.

Even accepting the caveats above, it still remains difficult to provide specific information on the quality of forecasts. Probably the major difficulty is that the diagnostics used are often based on those developed for short-range predictions and for informing modellers. Diagnostics that specifically inform users are thin on the ground, although some exist. Probably the most common diagnostic used in seasonal prediction is the correlation between predicted and observed, an approach particularly used for sea surface temperatures, including those measuring ENSO, and for predictions from empirical models. There is an argument, often discussed by meteorologists, that the minimum correlation required for 'skill' is (±)0.6, but many seasonal forecast systems fail to reach this level, although in some favoured areas far higher values have been obtained. Correlation, however, has many drawbacks as a diagnostic, and, while useful as a first indicator, it is not one that supplies value to the user question "What decisions can I base on these forecasts"?

Numerous other diagnostics are available, but most forecast centres tend to have in-house preferences that often differ from those of other centres. Thus it is difficult to cross-compare quality of forecasts between centres. Several standardisation activities are now underway in various areas of WMO, with one in particular being designed to assist intercomparison of operational predictions.[10] Three diagnostics are being used, the first of which, Mean Square Skill Score (MSSS), is an extension to the Mean Square Error (MSE) commonly used by meteorologists. MSE, the averaged sum of the squares of differences between forecasts and observations, provides an indication of improvements in forecasts as its value hopefully reduces in time, the ideal being zero. Beyond that its interpretation is difficult, particularly from the perspective of applications. In MSSS MSE values from the forecasts are compared with MSE values of a standard simple forecast in such a way that a zero value means both are identical by this measure. A positive MSSS value indicates that the forecast is an improvement compared to the standard; this improvement is measured in a linear sense so that perfect forecasts (those with an MSE of zero) give an MSSS of 100%. The standard forecast normally used is climatology or persistence, and MSSS at least gives the forecast user an indication of the extent to which the forecast improves on the chosen standard. It applies only to deterministic predictions.

The second diagnostic, Relative Operating Characteristics (ROC) (Mason and Graham, 2002), is of more use to applications, and has the advantage of being applicable to both deterministic and probabilistic predictions. ROC works through events that can be defined in any way consistent with the capabilities of the prediction system and preferably of interest to applications. For example, a specific event might be 'above-average rainfall', or 'temperatures in the lowest 20% of historic events', or any other definable binary value. Then for a deterministic prediction the hit rate (HR) is the proportion of observed events that were correctly predicted. Similarly the false alarm rate (FAR) is the proportion of occasions that the event did *not* happened but on which the event was predicted to occur. The extension to probability forecasts is achieved by calculating HR and FAR for a sequence of cases in which the event is taken to be predicted when the forecast probability equals or exceeds a specific value. Often this is done in 10% steps and when values of HR and FAR are then plotted against each other the ROC curve emerges. Provided the event is chosen to reflect the interest of applications, then the combination of HR and FAR is readily interpreted in terms of user actions, although it is immediately clear from the curve that attempts to maximise HR will also lead to increases in FAR.

The third diagnostic, reliability, is inherent in ROC and is applicable to probability forecasts only. Stated directly, a forecast is said to reliable if across all occasions on which an event is predicted to occur with, say, 40% probability then it does indeed occur on 40% of occasions. The reliability curve is normally plotted with pairs of values at 10% probability intervals in order to give an overall view of reliability.

Given the range of diagnostics used, and the inadequate time to achieve results as yet from the standardisation activities, it is difficult to create an overarching statement on the quality of the predictions. Nevertheless a number of generalisations may be made. Probably the first statement to be made from the meteorological perspective (the applications perspective is treated below) is that, whilst variable both spatially and between seasons, skill over and above climatology does exist across much of the globe. For example, early results indicating that skill over Europe was marginal and amongst the lowest on the planet have been overturned in the PROVOST project, in which some remarkably skilful predictions were achieved (Graham et al., 2000). Nevertheless high levels of skill are not available over Europe in all years. Projects such as PROVOST have permitted consistent assessment of skill on a global scale.

The variety of models targeted on predicting ENSO-related tropical sea surface temperature variations in the Pacific Ocean all have measurable skill out to perhaps nine months to a year, in some case perhaps a little longer. Attempts have been made to predict these variations for even longer. A number of estimates from numerical models for Pacific sea surface temperature predictions outside the tropical belt and for predictions across the other ocean basins have been made, and there are indications of some skill, although this is varies by region. Empirical prediction models designed for the Atlantic and Indian Ocean basins also give some skill, but overall highest skill levels are clearly achieved for the tropical Pacific. The creation of an array of moored data buoys across this basin, an outcome of the TOGA project, has underpinned this ability. Some moored buoys are present across the other two basins but these arrays are not yet suitable for use in operational prediction. A new global array of drifting buoys, ARGO, is expected to provide useful information for seasonal prediction in the near future.

Regarding the atmosphere, predictions of spatially relatively homogenous variables, such as temperature, are more skilful than those for the more heterogeneous variables such as rainfall. Nevertheless rainfall can be predicted with a high level of skill in some regions, predictions for the Nordeste of Brazil having a correlation approaching 0.9. It is likely that the high skill for this region results from strongly linear influences from both the Pacific and Atlantic oceans, although rainfall predictions for some tropical Pacific Rim areas approach similar levels of skill. Highest levels of skill in general are in those regions where variations are strongly linked to sea surface temperature variations, and these tend to be at tropical latitudes, particularly close to the Pacific Ocean. In general skill tends to fall off as distance from the equator increases, although regions such as parts of North America are favoured with relatively high levels of skill for their latitude because of the specific manner in which the global atmosphere responds to changes in the tropical Pacific. Results from the PROVOST experiment appear to indicate that predictability for Europe is mainly restricted to El Niño and La Niña periods.

Skill is also dependent on two further considerations. The highest skills are found in general for the shortest range predictions. As the forecast range and/or the lead time increases so skill tends to decline. The effect has been well demonstrated for sea surface predictions in the Pacific, but is less fully documented for temperature and rainfall predictions. Some empirical methods have been assessed for forecast lead periods exceeding a year, and apparently retain some skill over limited regions, but few statistics are available as yet for numerical models beyond four months. DEMETER is one project that will provide extensive statistics out to six months.

Seasonality is the second consideration, and again statistics are limited, but it is clear that skill varies in a seasonal manner that is difficult to generalise. In the middle northern latitudes skill in general is highest in winter and spring and lowest in autumn. Elsewhere skill for both temperature and rainfall predictions tends to maximise when linear links with sea surface temperatures are strongest. Predictions for Pacific Ocean sea surface temperatures tend to have lowest skill when projected through the April/May period, but skill recovers quickly in June. Known as the predictability barrier, the effect inhibits predictions of the build-up of El Niño events until the later part of the year.

Whilst projects such as DEMETER will provide substantial information concerning the quality of predictions from numerical models there is as yet no equivalent project for detailing the efficacy of empirical models. It is anticipated that the standardisation exercises will facilitate intercomparison of skill levels achievable through the various modelling approaches. At present it is probably fair to state that, in general, the empirical and numerical approaches provide forecasts of comparable skill, although there certainly are regions and/or seasons where one approach demonstrably outperforms the other. Numerical approaches have many potential advantages over empirical methods and are likely to become increasingly predominant. Empirical methods, on the other hand, will continue to play a fundamental role in exploration, in providing performance benchmarks for the numerical models, and in offering opportunities for developing prediction systems available to all.

5. The 1997/98 El Niño Event

Depending on the manner in which it is assessed, the El Niño event of 1997–1998 might be the strongest recorded. Certainly it was of comparable magnitude to the event of 1982–1983 which brought drought to many areas surrounding the Indian Ocean and heavy rainfall to continental areas adjoining the eastern Pacific Ocean. Without doubt the 1997–1998 event was the most heralded in history, as it offered the first opportunity for modellers to test their El Niño predictions using data from the then newly completed Pacific tropical buoy array. Discussion continues, though, over whether or not the event was predicted ahead of its presence being detected by the observing systems (Barnston et al., 1999).

There is a standard diagram (Ropelewski and Halpert, 1987) that details climate anomalies likely (but not certain) to occur during El Niño events, and countries threatened with drought or flood according to this diagram took some considered actions through the latter months of 1997, which certainly mitigated negative impacts in places (World Meteorological Organisation, 1999; Glantz, 2001). In hindsight, however, some actions were inappropriate to the climate conditions experienced and the event provides an object lesson in many of the issues surrounding prediction on seasonal to interannual time scales. Examples from southern Africa will be used to illustrate the point.

Coincidentally a major new initiative was planned for the period of the 1997–1998 El Niño event, the world's first Regional Climate Outlook Forum (RCOF). Entitled SARCOF, with SA standing for Southern Africa, this was a pilot for the 30+ RCOFs that have subsequently been held around the globe. RCOFs have proven to be a prime instrument through which climate services based on seasonal to interannual predictions have been introduced to the developing world. Forums bring together regional climatologists, intermediaries, end users, policy makers, the media, and various experts in an assembly at which presentations are made in a capacity building context, a consensus forecast is created using all available empirical and numerical inputs, and the forecast is discussed and interpreted in actions appropriate to the forthcoming rainfall season. The standard output product from a Forum is a map on which are given probabilities for rainfall terciles in individual parts of a region (Basher et al., 2001). A text prediction is also provided.

The first meeting in the 1997–1998 SARCOF series was held in Kadoma, Zimbabwe, during September 1997. The concept of probability forecasts was new to most attendees and there were consequent difficulties in defining the interpretation of the forecast, which placed highest probabilities in the driest tercile in most areas likely to experience drought during an El Niño but retained modest probabilities in the other two terciles, including the wettest. Against a media and an Internet barrage (a medium being widely exploited for the first time in delivering seasonal forecasts) calling for a severe drought in line with the canonical El Niño response in southern Africa, the carefully-placed Forum forecast apparently carried little weight. The text accompanying the Forum forecast accepted that drought, possibly severe, might occur but also pointed out that, historically, drought does not necessarily accompany an El Niño event. At a Forum update meeting in Windhoek, Namibia, in December, several of the numerical model predictions had reduced likelihoods of drought as compared to the September situation, a change reflected in the Forum forecast. Despite these activities it is probably fair to state that the general perception across southern Africa, and the frequent planning position, was that a drought was inevitable.

In reality seasonal rainfall over southern Africa was generally around the long-term climatological normal, but with somewhat above-normal totals in favoured places. Certain areas experienced localised flooding. Departures from the canonical response also occurred elsewhere. Rainfall in East Africa was above average, as would be expected from the canonical response, but by an excessive amount. As in southern Africa, the droughts anticipated in India and Australia did not occur. However an unusually severe drought did happen in parts of South East Asia, and was responsible for creating suitable conditions for the haze that developed there.

Almost certainly one reason why these responses did not match the canonical expectation was due to the influence of sea surface temperature anomalies in the Indian Ocean, which were perhaps larger than any previously observed. While the El Niño was taking place in the Pacific Ocean, temperatures across the Indian Ocean resembled the equivalent of a La Niña event with warm anomalies in the west, a distribution that might have been linked with the El Niño itself (Yu and Rienecker, 1999). The warm waters in the western Indian Ocean probably assisted in increasing the rainfall across the whole of the eastern part of Africa. Most of the empirical forecast models used in Africa focused on Pacific Ocean sea surface temperatures as the main predictor, hence failing to incorporate any information from the Indian Ocean. Certain numerical models may have been picking up information from the Indian Ocean given their predicted reduced probabilities of dry conditions by December.

The credibility of seasonal forecasting suffered initially as a result of the 1997–1998 episode. In the view of the author the issue was one of understanding and of recognising the probabilistic nature of the forecasts. This message has been passed continually through ensuing Forums, although it is still probably fair to say that many are not yet comfortable with probabilities. Confidence returned to a certain extent in following years as an extended La Niña event occurred that was associated with generally canonical impacts. But in 2001 the possibility of a new El Niño event developing by the end of the year caused ripples in some parts of the Pacific region. No El Niño developed. At the time of drafting the original version of this paper a similar scenario to that of 2001 was being played out in early 2002, although against the background of El Niño signals certainly stronger than those during 2001. It is now clear, in late 2002, that an El Niño event has developed.

No final solution has yet been obtained in the creation of a general understanding of the limitations of seasonal to interannual forecasting. Many users meetings continue to call for improved forecast accuracy, which will certainly come as the models and understanding improves but which will not meet deterministic expectations. Education initiatives, together with clearer statements from climatologists on the limitations of the prediction systems, plus a willingness amongst users to examine how current predictions might be best used in particular applications, are amongst the requirements at the present time.

6. Interfaces to Applications

The work of Gilbert Walker was driven not by the need to understand the climate of India but by the need to plan for droughts in a largely self-dependent agricultural country. Much of the solution to the food problem in India was ultimately obtained through improved management of food stocks rather than through climate prediction, but India has the distinction of having the longest sequence of operational seasonal predictions using modern empirical techniques anywhere in the world. Elsewhere the long-term drought of the 1980's in the Sahel brought increased attention to the impact of sea surface temperature variations on rainfall and lead, at the UK Met Office, to the development of an empirical forecast model for rainfall over the region (Ward et al., 1989), and then to experiments with numerical models (Folland et al., 1991). A second area subsequently considered by the Met Office team was the Nordeste of Brazil (Ward and Folland, 1991), and the Met Office has provided forecasts regularly for this region since 1987. The Nordeste is subject to El Niño-related severe droughts that impact on both food and local industrial production. Such is the level of predictability that the Nordeste is favoured by having numerous prediction models available created by various research groups.

Applications of seasonal forecasts go back many years and a review of estimates of value obtained through use of the predictions has been published by Nicholls (1996). However as the science has reached a level of maturity over the past few years so the number of projects in which forecasts are fed into decision making in applications has increased and the question of value has become more critical. In the following discussion agricultural applications will provide the focus, although activities are underway in many other applications sectors including water management, energy planning and trading, manufacturing, insurance and health.

One of the longest consistent research projects studying the use of seasonal prediction to agriculture has been that focussed on corn production in the US Mid West (e.g. Mjelde et al., 1997). This series of studies has illustrated the potential value of predictions provided these are designed to fit directly into the needs of the farmer. The basic concept of “designed to fit” was taken as perfect deterministic predictions of temperature and rainfall anomalies delivered at a lead sufficient for decisions to be taken and acted upon. In later studies the concept of partially unreliable predictions was introduced. Substantial benefits were to be gained should prediction quality attain the levels assumed. However experience in the US and elsewhere has indicated that, despite the positive theoretical position established, multiple impediments to the use of seasonal to interannual predictions exist. These impediments lie both on the forecast producer side – including the provision of imperfect forecasts, in probabilistic terms, using non-intuitive presentation formats, at too short leads, in insufficient detail, and through limited-access delivery methods – and on the user side – including misinterpretation of probabilistic information, lack of full understanding of the limitations of forecasting, and want of optimal methods of incorporating real forecast information into decision processes.

Measurements of forecast quality, and its interpretation as value ultimately obtained in applications, remain amongst the key issues linking and offering at least partial solutions to many of the impediments. The issue of measurement of forecast quality has already been covered in brief. Most currently-used quality measures have been derived from the perspective of the forecaster, with consideration for the needs of the user often attached only as an add-on. A noteworthy sequence of studies in which theoretical attempts were made to link forecast quality with ensuing value was provided by Alan Murphy and collaborators (e.g. Ehrendorfer and Murphy, 1992). One central methodology used in these studies was the Cost/Loss model, in which value for a particular binary event (as defined above under ROC) was calculated for the four combinations of event predicted/not predicted with event observed/not observed. Attempts to relate quality and value were not entirely successful, however, until the introduction of ROC, as the approach used by ROC for examining quality permits transfer directly into the Cost/Loss model without transformation of data. This ROC-Cost/Loss approach was first used at a practical level within a seasonal forecasting context during the first SARCOF and produced results that illustrated the potential value in southern Africa of forecasts of then-existing quality (Harrison and Graham, 2001). Other studies of ROC and its use in estimating value have been made (e.g. Richardson, 2001).

One important result that appears consistent across all studies to date relates to the fact that ROC can be used to measure quality in both deterministic and probabilistic systems in a common manner. As far as is known highest value always results, at least in theory, from use of probability forecasts according to the ROC approach.[8] This theoretical result is a consequence of the balance between value gained from correct decisions against value lost in incorrect ones. No forecast system can consistently guide perfect decision making, and the evidence suggests that the management and consequences of those occasions on which incorrect decisions are taken determine the final overall value obtained across many years. In the view of the author no quality measure approach that fails to provide detail on prediction limitations sufficient to satisfy value estimates through the range of potential outcomes is adequate for applications needs. One rider to this is that deterministic forecasts, despite offering lower value, can be used provided adequate recognition of the possibilities and consequences of an erroneous prediction is taken. However, as the consequences as far as an application is concerned may depend on the magnitude of the error, a distribution of outcomes is still desirable even in this case for planning purposes. Despite their potential importance few studies of approaches to contingencies in decision making within the context of seasonal forecasts have yet been undertaken. These, plus additional theoretical and practical value studies, are needed in order to design improved decision processes in applications. While the ROC approach to value is a useful starting point, it handles only limited aspects of issues related to individual decisions, it implicitly assumes certain vital aspects of the decision process, and it ignores socio-economic interactions on both the micro and macro scales. ROC's merit in providing guidance in the use of seasonal forecasts in agriculture has, nonetheless, been established.

One additional benefit of using ROC illustrated in the SARCOF project is that it can be employed to indicate optimal strategies for forecast use by helping identify the selection of model parameters that maximises benefit (Harrison and Graham, 2001). However this carries the vital rider that it is quite possible to use a given set of forecasts in a way that would provide negative value rather than extracting the available positive value, a point that is frequently not considered in planning applications. The benefit gained from the use of seasonal to interannual predictions is therefore not dependent upon the quality of the predictions alone but also relies on the manner in which decisions are made based upon the predictions. Limited research attention has so far been focussed on this latter aspect by comparison with the forecast quality issue.

Relative Operating Characteristics provided much of the basis for a 2000–2001 project in the United Kingdom in which the value of forecasts on a wide range of time scales for the complete food chain, from production to retail, was examined.[11] Despite the relatively low level of seasonal predictability over the United Kingdom, certainly by comparison with many tropical regions, substantial benefit was shown to be achievable, and confirmed through a calibration process using ROC. One key ingredient in the project's success was the close collaboration undertaken between meteorologists and all those involved on the agricultural/food side in each of the four sub-projects (dealing with field vegetables, sugar beet, apples and tomatoes), and this stands as a prime example of the type of integrated approach needed. One important project conclusion was that optimal benefits were obtained by coordination of responses to predictions throughout the food chain rather than as a series of separate decisions.

A number of other approaches to the decision process have been examined in both theory and practice, although it is not currently clear what level of benefit is achievable through most. The simplest, and probably the most common, is the straightforward treatment of a deterministic prediction in terms of experience. The benefits and hazards of this approach were illustrated during the 1997–1998 El Niño event. A second approach, the use of analogues, is basically a development of the deterministic approach in that it identifies historical sequences similar to those recently observed and/or predicted that provide appropriate analogues to guide decision making. The technique can be extended to probabilistic predictions through the development of analogues across the full range of possibilities that can then be interpreted as a group of weighted decision options. Analogues require substantial historical data and resource to develop, and so are not appropriate in all situations. A third alternative is the approach of 'least regret'. Following careful examination of the needs, perceptions and restrictions upon the end user, a decision based on forecast information is created that will provide least future regret should it turn out to be incorrect. By this approach a buffer against the more risky decisions, often those with greater rewards if correct but at greater costs if incorrect, is created. This approach also acknowledges that some agricultural forecast users, particularly in developing countries, do not have the scope to accept incorrect decisions or the flexibility to develop contingencies in the manners that probability forecasts inevitably impose.

In recent years research projects have been developed to provide more direct approaches to decision making through use of climate model outputs to drive agricultural models in a variety of manners. It is expected in these projects that these methods will provide the types of direct information required by agriculture. Further it is hoped that these methods will handle the issue that agricultural responses depend as much, if not more, on the detailed sequence of weather events during a season than on the averages across the season. Statistical weather generators have been used as one method of providing realistic daily weather sequences to feed crop models, as have daily outputs from the climate models. There is insufficient knowledge as yet about the statistics of climate models on a daily basis to be certain that this option is an improvement over the statistical generators, and results will be awaited with interest. Major activities in this area include CLIMAG,[12] PROMISE[13] and DEMETER.[14] As discussed earlier, regional climate models (RCMs) are also being examined as a means of providing climate information on spatial and temporal scales below those provided by the climate models and to provide information for agricultural decision making.

Most of the above discussion has focused on the use of forecast information alone, with limited reference to the pertinence of historical information in guiding the decision process. The substantial benefit that can be obtained, even by use of information alone without forecast input, has been well demonstrated for agriculture by the work in Australia at BoM[15] and APSRU,[16] both organisations having developed extensive information systems based on historical data to assist in decision making. These systems cover not only climate information but also pertinent agricultural, economic and financial information. A basic approach developed at APSRU, and based in part on observed phases of ENSO in recent months, has been successfully adopted in countries outside Australia (Meinke and Stone, 2004).

7. The Future

Seasonal to interannual prediction is no longer the marginal activity that it was only a few years ago. With the ability to observe the tropical Pacific Ocean both at the surface and at depth in real time, to observe the other oceans in increasing detail through satellites and moored buoys and through the ARGO system coming on stream, and with the rapidly increasing computing power available, global seasonal to interannual prediction has become an operational activity at a number of major meteorological centres. Workstations and powerful Personal Computers now allow the smallest research groups the opportunities to run RCMs, and even the global models. Empirical models can be developed and run by all. In the future further developments will continue along similar lines. Assimilation of ocean data into the coupled models, lead by the GODAE project, will improve with the implementation of ARGO and further moored buoys systems. Increasingly complex coupled models will be run in larger ensembles. The benefits of RCMs will be examined in detail. And improved empirical models will be developed within a framework that permits contrast of forecast skill with the numerical models and optimal consensus forecast building.

Nevertheless, substantial leaps in skill are unlikely to be gained. Rather a steady improvement in forecast quality will be achieved. Gains will be obtained through the use of improved quality models, through improved oceanic observations and related improved assimilation of ocean data, and perhaps through the use of land surface initialisation of models following results from the GEWEX project. RCMs may prove to provide additional valuable information for some parts of the world.

Predictability gains will be limited by the chaotic nature of the atmosphere-ocean system. The gains listed above will contribute to improvements in predicted probability distributions, with the larger ensembles providing more detail on the outlying parts of these distributions. These outlying parts contain information on the more extreme events and so are key to agricultural decision making. RCMs may be incorporated to help provide probability distributions on reduced spatial and temporal scales.

Alongside the developments in prediction science will be a marked intensification of activities related to decision making processes in agriculture, and it is in this area that the major advances are likely to come in the short term. The use of crop models will be developed and should become integrated into the ensemble outputs of the dynamical models. If necessary, weather generators will be used to obtain crop model output from empirical predictions. A focus in research will be on the integration of all prediction and historical information and its optimal use in agriculture. In support of these activities will be increased and imaginative use of existing information, both climatological and agricultural. This in turn will lead to development of the improved, coordinated climatological/agricultural data bases necessary to provide the information required. Further experiments of the type run in Algeria whereby crop patches were treated with or without use of forecast information will be used widely to assess the value of the forecasts.[17]

All evidence to date suggests that the potential benefits of the science are substantial, to both agriculture and to many other economic sectors, but that the effort involved in extracting optimum benefit is equally substantial and requires coordinated multidisciplinary research. Further activities towards development of such co-ordination will continue under the umbrella of the various multinational organisations in the United Nations system. These will cover not only climate science, but also all involved areas through to the social sciences. Activities such as the current Regional Climate Outlook Forums will provide a model for part of this coordination, but this will be extended through initiatives such as the Regional Climate Centres (Nicholls, 2002) and Res Agricola (Meinke and Stone, 2004). Over and above all of this will be an integrated capacity building activity designed to introduce all concerned to the opportunities and potential pitfalls of the developing science.

8. Conclusions

Seasonal to interannual prediction has a long history in a variety of guises, but it is in the five years during and since the 1997/98 El Niño event that it has become a major research and application issue. It has been demonstrated unequivocally that short-range climate prediction is achievable with skill in many parts of the world. The level of skill available varies geographically and with season, highest levels tending to be found in the tropics, and in some places may only be useable during windows of opportunity. In all cases predictability is available only over large temporal and spatial scales as well as in probabilistic terms, and it is these facts that create difficulties in applications. Attempts through the use of downscaling via RCMs are amongst those being made to overcome the scale problem. Addressing the probability problem is an issue that requires improved methods of interpreting forecast information into applications actions.

An increasing number of projects are being run to assess the opportunities of the predictions in agriculture in numerous crops and disparate regions of the world. Across the range of projects have been initiatives to understand how climate information is used in agriculture, to provide detailed historical information to guide decision making, and to interpret forecasts through a variety of approaches involving crop models and other analysis tools. There is little doubt that benefit has been gained by agriculture and that the potential for increasing this benefit in the future exists. However the optimal approach for using the predictions in a consistent manner remains to be assessed.

Climate variability sits at the lower end of the temporal spectrum occupied at longer scales by climate change. Insofar as climate change deals with trends, so within those trends lie the climate variations that have formed the subject matter of this paper. Adaptation to climate change incorporates adaptation not only to impacts of trends but also to impacts of changes in variability. A new multidisciplinary project, GECaFS,[18] is examining the adaptation needs of agricultural under a changing climate. As such many of GECaFS' considerations are directly equivalent to those required for handling climate variations. The synergies and opportunities for integration of climate and agricultural sciences will continue to grow, and it is likely that we will see a cascading of similar benefits to other climate sensitive industries and activities.

Acknowledgements

There have been numerous colleagues whose work has contributed to the contents of this paper. I must acknowledge broadly the small climate variability community whose enterprise over the past few years has made this paper possible. I also extend my thanks to an anonymous reviewer for providing a fine and relevant critique of the first draft of this paper. And finally my thanks to the officers of the Commission for Agricultural Meteorology for inviting me to prepare this paper.

Notes

1. There are numerous references, amongst which is The Halley Lecture delivered on 31 May 1929 and entitled 'Some Problems of Indian Meteorology'.
2. Numerous texts cover ENSO, including Glantz et al. (1991).
3. An authoritative overview is given in: Anderson et al. (1998).
4. See details on the Global Energy and Water Cycle Experiment web site-http://www.gewex.org/.
5. However see: Goddard L., et al.(2001).
6. For a more detailed explanation see Glantz et al. (1991).
7. See: http://www.ecmwf.int/products/forecasts/seasonal/documentation/index.html.
8. See Graham R.J. et al. (2000), and other articles in the same Special Edition on PROVOST of the *Quart. J. R. Meteor. Soc.*
9. See: http://www.ecmwf.int/research/demeter/.
10. For details see the Global Data-processing System pages, in the World Weather Watch area of the World Meteorological Organisation web pages at www.wmo.ch – follow 'Verification systems for Long-range forecasts'.
11. On http://www.foresight.gov.uk/default1024ns.htm, follow → Foresight 1999-2002 → Food Chain & Crops for Industry, for details. Copies of the report are also available from World Meteorological Organisation WCP CLIPS Project Office.
12. See: http://www.start.org/Projects/Climag/climag.html.
13. See: http://ugamp.nerc.ac.uk/promise/.
14. See: http://www.ecmwf.int/research/demeter/.
15. See: http://www.bom.gov.au/silo/.
16. See: http://www.apsru.gov.au/.
17. See: Report by Galem,O., to the CLIPS Working Group, March 2001. Available from the World Meteorological Office WCP CLIPS Project Office.
18. See: http://www.gecafs.org/.

References

Anderson, D. L. T., Sarachik, E. S., Webster, P. B. and Rothstein, L. M. (eds.): 1998, 'The TOGA decade: Reviewing the progress of El Nino research and prediction', *J. Geophys. Res.* **103C**, 14167.

Barnston, A. G., Glantz, M. H. and He, Y.: 1999, 'Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset', *Bull. Am. Meteorol. Soc.* **80**, 217.

Basher, R., Clark, C., Dilley, M. and Harrison, M. S. J. (eds.): 2001, *Coping with Climate: A Way Forward – Summary and Proposals for Action; Preparatory Report and Full Workshop Report*, both published by the International Research Institute for Climate Prediction on behalf of the World Meteorological Organisation, NOAA, the South African Weather Bureau, USAID and the World Bank.

Ehrendorfer, M. and Murphy, A. M.: 1992, 'On the relationship between the quality and value of weather and climate forecasting systems', *Idojaras* **96**, 187.

Ferranti, L. and Molteni, F.: 1999, 'Ensemble simulations of Eurasian snow depth anomalies and their influence on the summer Asian monsoon', *Quart. J. R. Meteorol. Soc.* **125**, 2597.

Folland, C. K., Owen, J. A., Ward, M. N. and Colman, A. W.: 1991, 'Prediction of seasonal rainfall in the Sahel region using empirical and dynamical methods', *J. Forecast.* **10**, 21.

Folland, C. K., Colman, A. W., Rowell, D. P. and Davey, M. K.: 2001, 'Predictability of northeast Brazil rainfall and real-time forecast skill, 1987–98', *J. Clim.* **14**, 1937.

Glantz, M. H. (ed.): 2001, *Once Burned, Twice Shy?: Lessons Learned from the 1997–98 El Niño*, United Nations University, 294pp.

Glantz, M. H., Katz, R. W. and Nicholls, N.: 1991, *Teleconnections Linking Worldwide Climate Anomalies,* Cambridge University Press.

Goddard, L., Mason, S. J., Zebiak, S. E., Ropelewski, C. F., Basher, R. and Cane, M. A.: 2001, 'Current approaches to seasonal to inter-annual climate predictions', *Int. J. Climatol.* **21**, 1111.

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', *Quart. J. R. Meteorol. Soc.* **126**, 2211.

Harrison, M. S. J. and Graham, N. E.: 2001, 'Forecast quality, forecast applications and forecast value: cases from southern African seasonal forecasts', *WMO Bull.* **50**, 228.

Hastenrath, S.: 1991, *Climate Dynamics of the Tropics*, Kluwer Academic Publishers.

Kumar, K. K., Rajagopalan, B. and Cane, M. A.: 1999, 'On the weakening relationship between the Indian Monsoon and ENSO', *Science* **284**, 2156.

Latif, M. and Barnett, T. P.: 1994, 'Causes of decadal climate variability over the North Pacific and North America', *Science* **266**, 634.

Marriott, P. J.: 1981, *Red Sky at Night, Shepherd's Delight? Weather Lore of the English Countryside,* Sheba Books.

Mason, S. J. and Graham, N. E.: 2002, 'Areas beneath the relative operating characteristics (ROC) and levels (ROL) curves: Statistical significance and interpretation', *Quart. J. R. Meteorol. Soc.* **128**, 2145.

Meinke, H. and Stone, R.: 2004, this volume.

Mjelde, J. W., Thompson, T. N., Nixon, C. J. and Lamb, P. J.: 1997, 'Utilising a farm-level decision model to help prioritise future climate prediction research needs', *Meteorol. Appl.* **4**, 161.

Mullen, S.L. and Buizza, R.: 2002, 'The impact of horizontal resolution and ensemble size on probabilistic forecasts of precipitation by the ECMWF ensemble prediction system', *Weather Forecast.* **17**, 173.

Nicholls, J. M.: 1996, *Economic and Social Benefits of Climatological Information and Services: A Review of Existing Assessments*, WMO/TD No. 780.

Nicholls, J. M.: 2002, 'WMO approach to seasonal-to-interannual forecasting and organised Regional Climate Centre services', *WMO Bull.* **51**, 158.

Richardson, D. S., 2001: 'Measurements of skill and value of ensemble prediction systems, their interrelationship and the effect of ensemble size', *Quart. J. R. Meteorol. Soc.* **127**, 2473.

Rodwell, M. J. and Folland, C. K.: 2002, 'Atlantic air–sea interaction and seasonal predictability', *Quart. J. R. Meteorol. Soc.* **128**, 1413.

Ropelewski, C. F. and Halpert, M. S.: 1987, 'Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation', *Mon. Weather Rev.*, **115**, 1606.

Ward, M. N., Folland, C. K. and Parker, D. E.: 1989, *Statistical Prediction of Seasonal Rainfall in the Sahel and Other Tropical Regions*, WMO/TD No. 261, 82.

Ward, M. N. and Folland, C. K.: 1991, 'Prediction of seasonal rainfall in the north Nordeste of Brazil using eigenvectors of sea-surface temperature', *Int. J. Climatol*. **11**, 711.

World Meteorological Organisation: 1999, *The 1997–1998 El Niño Event: A Scientific and Technical Retrospective*, WMO No. 905.

Yu, L. and Rienecker, M. M.: 1999, 'Mechanisms for the Indian Ocean warming during the 1997–98 El Niño', *Geophys. Res. Lett.* **26**, 735.

(Received 15 December 2003; in revised form 14 June 2004)