Remarks prepared by A. P. Dempster for:

Satellite Workshop on *Probabilistic Climate Prediction*

Programme on *Mathematical and Statistical Approaches to Climate Modelling and Prediction*, Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, 11 August - 22 December 2010

University of Exeter, 20-23 September 2010

Organisers: Dr R Chandler *(UCL)***, Dr M Collins** *(Exeter)***, Professor P Cox** *(Exeter)***, Dr K Horsburgh** *(National Oceanography Centre, Liverpool)***, Professor JM Huthnance** *(National Oceanography Centre, Liverpool)***, Dr JC Rougier** *(Bristol)***, Professor DB Stephenson** *(Exeter)* **and Professor J Thuburn** *(Exeter))*

Notes on Fundamental Approaches to Climate Prediction

A. P. Dempster
Department of Statistics
Harvard University

(1) Two Cultures

One concern is the basic problem of trying to get physicists and statisticians on the same page. Statisticians think of themselves as dealing with data as "information with context", as I saw quoted somewhere recently, and with statistical models as parts of a complex system of tools for extracting meaning from statistical data. Physicists on the other hand tend to think of models as approximations to scientific truth, with the ultimate goal of research being to arrive at representations and explanations of such truth. The two cultures are very different.

Almost in contradiction to their pure science backgrounds, it has become a basic function of physical climate modelers to inform policymakers and other real world stakeholders about possible alternative future climates. When used in this mode, climate models are treated as carriers of information, so move closer to statistical models. Specifically, physical models become interpretable as information when their equations are regarded as approximating relations among the values of actual real world variables at successive points in actual time. Statistical models, should similarly be regarded as describing probabilistic relations among unknown true values of such variables, including probabilistic time dependence. From this unifying perspective, probabilistic climate prediction is largely a matter of information fusion. There are multiple sources of information from physical theory that can be combined, and multiple sources of statistically summarized empirical data. Optimum predictive power requires pooling information of both types. Adopting this attitude may help physicists and statisticians bridge the two cultures, and in particular may open a path to a more central role for research statisticians.

Several factors get in the way of recognizing the fundamental synergy of the two types of information source. Both communities are deeply involved in technical developments expressed in terms that the other community can scarcely follow. Separate literatures deal with different formal mathematical structures having long histories expressed through non-overlapping technical jargons, mostly leaving real world connections implicit, or ignored altogether. An important way to help circumvent communication gaps could be to begin each attempt at a predictive climate analysis by setting up a common system of real world variables, using precise language that both communities can understand. For example, as an advocate of probabilistic reasoning, I might set a goal of reporting probabilities that the average annual surface temperature in the UK in 2035 will be at least 1 degree C higher than in 2010, assuming given world-wide emissions scenarios. For a reported numerical probability to be meaningful, it is necessary not only that the two average temperature variables must be precisely defined in operational terms, but an additional multitude of variables appearing in the analysis must be similarly specified and connected to the one-off real world in a meaningful way. My sense is that neither community is accustomed to meeting this standard at the level of system complexity posed by real world climate prediction.

(2) Approaching Probabilistic Models: What are the Issues?

Once a system description of adequate complexity for a planned predictive analysis is judged to be sufficiently well-defined, the tasks shift, first to specifying mathematical models that represent available information from both empirical and theoretical sources, and second to combining models to obtain a pooled model from which desired uncertain inferences can be obtained. Implementations of such a scheme to climate prediction is hard to discern in the norms of current practice, though maybe less hard for weather prediction on time scales of a few hours to a few days where Bayesian schemes are more manageable and verifiable. What are the problems and prospects for moving from weather to the longer time scales of climate change? What unknowns are predictable probabilistically on longer time scales, and which are not?

A fundamental issue concerns the nature of uncertainty. All can agree that predictions are uncertain. But what mathematics should be used when computing and reporting predictive uncertainties? Here the divergence of the two cultures is astonishing. Few physicists have training and hence knowledge of how the mathematics of probability, together with its relations to scientific uncertainty, has developed over 300 years into a formidable set of theoretical structures and tools. The identity of the academic discipline of statistics was transformed, especially over the middle decades of the 20th Century, by competing methodologies for addressing scientific uncertainty. This is not the place to delve into explaining what developed, and how there are differing viewpoints, with mine in particular lying outside the statistical mainstream. I believe it is fair to say, however, that how physicists approach scientific uncertainty has been scarcely touched by fundamental developments within statistics concerning mathematical representations of scientific uncertainty. An indication of the disconnect is provided by the guidelines used by the IPCC in its 2007 major report, where the terms "likelihood" and "confidence" were recommended for two types of uncertainty reports, apparently in complete ignorance of how these terms have been used for more than 60 years as basic textbook concepts in statistics, having nothing whatsoever in common with the recommended IPCC language (which I regard as operationally very confusing). Another indication is that experts from the statistical research community constituted according to one source only about 1% of the attendees at the recent Edinburgh conference on statistical climatology.

(3) Roles for Statistical Modeling

Physical modelers often refer to two basic sources of uncertainty when interpreting the output of a climate simulator, namely, uncertainty about initial conditions, and uncertainty from discretization, or transform truncation, of space/time variables. From my outsider's perspective, I would prefer an emphasis on attempting to model and analyze only the unique actual climate system, instead of the current practice of running and analyzing a series of mathematical and therefore artificial climate systems. Of course, the same pair of uncertainty sources arise in the combined physical/ statistical modeling approach that I am advocating.

My suggested model type is captured by the term "hidden Markov model". The thing that is hidden is the actual past, present, and future of the real climate system, which is the domain of physical thinking and modeling that proceeds forward in time, such as may be represented, for example, by the equations of AOGCMs. Since the real processes are hidden, they cannot be directly simulated. Alongside the hidden system there is empirical data also linked to real space/time, and partially obscured from the actual system by observational error. The goal of hidden Markov analysis is to update posterior probability assessments of the true system, including limited ranges of past, present, and future, given stepwise accrual of empirical data. It is these probability assessments that should be updated sequentially. Once models are specified, this becomes a defined computational task for Bayesian or DS analysis. Fast algorithms are known. They may look like simulations from a physical model, but are conceptually different because they sample the posterior probability assessment from fused empirical and theoretical information sources, typically using MCMC methodologies.

One task for the statistical research community is to formulate and implement probabilistic space/time models, from which principles of statistical inference determine posterior probabilistic assessments of the true climate, to whatever level of detail the assumed state space permits. For the past up to the present, hidden Markov analyses, such as the familiar Kalman filter, or more complex versions thereof, fuse the information from the past concerning the actual process with information from the current empirical record. For predicting the future climate, there is no data, so statistical error models are no longer operational. The climate proceeds on its own with probabilistic uncertainty entering only through probabilistic uncertainty about the present state of the actual system. Predictive analysis then proceeds by forward propagation of probabilistic uncertainty.

The necessary discreteness of physical models suggests that they might best be regarded as tracking local averages across neighboring space/time regions. Because they are approximations to relations believed to hold for infinitesimal changes across time and space, they reflect model errors arising from the inability to represent natural processes in space/time domains smaller than discretization can capture. Physical modelers typically introduce "parameters" that attempt to adjust difference equations for the missing processes. I sense that statisticians could become more deeply involved in the development of probabilistic representations of such discretization errors, which in effect turn even the non-empirical component of the model into a parametrized stochastic process whose parameters need to be assessed through formal statistical inference tools (e.g., Bayes or DS).

(4) The Problem of "Chaos"

Routine weather forecasts for seven days ahead can be downloaded from various agencies. These pretty much fall back to long term seasonally adjusted averages over recent decades and nearby locations. Also, if I were to ask for a forecast probability that my local daily maximum temperature will exceed such an average by at least 2 degrees C seven days hence, I believe that I would find myself pretty much looking at historical statistical variation around an average daily maximum. Although I am an outsider in the world of weather physics, my understanding is that sophisticated physical models are able to contribute little additional forecast skill seven days ahead due to instabilities associated with nonlinearities in the equations, even when fed with a large amount of reasonably accurate prior information about the state of today's weather system and its recent past.

Presumably most detailed physical climate models of the atmosphere predict chaotic instabilities in the real world climate system, analogous to those that make longer range weather forecasting a low skill enterprise.   How might these instabilities impact the task of devising credible probabilistic predictions of long term trends in future climates? My response is to make a radical proposal, linking the recognized difficulty of predicting chaotic systems as the future time horizon grows with a fundamental change in how probabilities should be similarly degraded on a similar time scale. The proposal is based on a weakening of Bayesian theory that I originally developed in a series of papers in the 1960s. Further developments were spearheaded by Glenn Shafer in the 1970s and 1980s ,who gave the theory an AI spin, and named it the theory of belief functions. I now prefer to call the DS (for Dempster-Shafer) calculus. It appears to be gradually gaining increased recognition and respect.

A detailed exposition of DS is not possible in this note, but I wish to draw attention to two basic features of the DS system. The first is that probabilities are no longer additive. By this I mean that if p denotes probability "for" the truth of a particular assertion, here some statement about a specific aspect of the Earth's climate in the future under assumed forcing, while q denotes probability "against" the truth of the assertion, there is no longer a requirement that $p + q = 1$. Instead these probabilities are allowed to be subadditive, meaning that in general $p + q < 1$. The difference $1 - p - q$ is labeled r, so that now $p + q + r = 1$, with r referred to as the probability of "don't know". (Note: each of p, q, and r is limited to the closed interval $[0,1]$.)

It is only for the last two years that I have focused on trying to explain what is meant by the DS concept of "don't know". I was helped when I ran across a reference to the following remarks by economist John Maynard Keynes, remarks that I believe have not been taken sufficiently seriously:

By "uncertain" knowledge, let me explain, I do not mean merely to distinguish what is known for certain from what is only probable. The game of roulette is not subject, in this sense, to uncertainty; nor is the prospect of a Victory bond being drawn. Or, again, the expectation of life is only slightly uncertain. Even the weather is only moderately uncertain. The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth-owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know. Nevertheless, the necessity for action and for decision compels us as practical men to do our best to overlook this awkward fact and to behave exactly as we should if we had behind us a good Benthamite calculation of a series of prospective advantages and disadvantages, each multiplied by its appropriate probability, waiting to he summed. (Excerpted from "The General Theory of Employment", *Quarterly Journal of Economics*, February, 1937, pages 209-223.)

By "good Benthamite calculation", Keynes meant what would now be called Bayesian decision analysis, which was evidently pretty well understood by the likes of Gauss and Laplace 200 years ago. In the phrase, "no scientific basis on which to form any calculable probability", Keynes was referring to the difficulty of finding a credible empirical footing for the numerical probabilities required by Bayesian models. A major concern of statisticians who argue in favor of Bayesian methods is the specification of "priors" whose often *ad hoc* nature began to be questioned in the mid-19th Century. Part of my motivation for DS in the 1960s was to follow up a proposal made by R. A. Fisher in 1930 to permit probabilistic inferences without Bayesian priors in carefully prescribed situations. The DS proposal amounts to paying for absence of the Bayesian prior by admitting the existence of probabilities of "don't know", with my concept of "don't know" complementing Keynes's proposal for a new species of uncertainty not covered by the traditional theory of additive probabilities.

A second fundamental feature of the DS calculus is a particular "rule of combination", or principle, for combining information from different sources, such as physical models and empirical data concerning past and present. The DS rule is linked to a DS concept of independence, which might therefore be regarded as a severe restriction on its applicability, except that many special cases of the rule are routinely used with little overt concern about independence, including Bayesian combination of likelihood and prior, and Boolean logical combination. Independence in the mathematics of ordinary additive probabilities is another special case. Most models in the burgeoning field of applied probability can be viewed as constructed from many independent components.

The DS rule of combination is a powerfully inclusive tool of probabilistic analysis, with potentially important applications to probabilistic climate prediction. In particular, I hypothesize that DS-style probabilities of "don't know" could come to be a basic way to separate unpredictable from predictable aspects of climate change. From the DS perspective, Bayesian inference cannot do this in a satisfactory way. An simple illustration may help to support my argument. It is easy to find on the web beautiful discussions and simulations of simple chaotic systems, beginning from the simple 3D model popularized by the late Ed Lorenz. The basic problem is that small perturbations in initial conditions often grow into large perturbations over "time". A Bayesian approach puts a prior on the initial position, which may be tiny, yet soon is projected forward to a much more spread out marginal distribution. The result is typically a limiting predictive distribution over the whole system. In a weaker DS framework, the prior may simply be a small region about initial point, where you "don't know" where the true initial condition is in the small region, i.e., you have r = 1 for the (p, q, r) of that region. As time progresses, your predicted "don't know" region with r = 1 grows, possibly taking over the whole system. It looks to me pretty obvious that the DS option of a logical (i.e., nonprobabilistic) analysis is needed to represent the fade-out of predictability for chaotic systems. Additive Bayesian predictive posterior distributions are unable to function in this way.

(5) The Problem of "Complex Systems"

Climate prediction not only has the problem of nonlinearity in dynamical systems, but also shares with the analysis of typical real complex systems the equally great trouble associated with the presence of a huge panoply of variables, subsystems, and possible feedbacks in play. I remember Rol Madden commenting, back in the 90s when I used to visit NCAR, that it would be a sheer accident if numerical experiments with GCMs were to give credible quantitative representations of the real climate, presumably including future effects of increasing GHG concentrations. Suppose that the system was defined at the outset to include the full carbon and hydrological systems. Then add atmospheric and ocean chemistry, and then living and breathing systems everywhere. Much "don't know" abounds simply about the present and recent past, let alone the fundamental problem of getting quantitative about the future.

What does this say about research priorities? Given the real climate system, characterized not only by fundamental "don't know" coming from dynamical nonlinearities, but also fundamental "don't know" coming simply from an inability to supply meaningful evidence-based priors in the presence of complexity, I believe that communities faced with needs for real predictions of complex systems should be investing in models of probabilistic prediction that provide measures of the type of "don't know" described by Keynes, including the DS approach as a leading candidate for numerical implementations.