Ensemble Techniques for Seasonal Prediction


This paper was prepared for submittal to the Proceedings of the 1995 Tropical and Ocean Global Atmosphere (TOGA) Conference Melbourne, Australia April 2-7, 1995

April 1995

Lawrence Livermore National Laboratory
DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial products, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.
DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.
ENSEMBLE TECHNIQUES FOR SEASONAL PREDICTION

by

T.N.Palmer\textsuperscript{1}, D.Battisti\textsuperscript{2}, C.Brankovic\textsuperscript{1}, R.Buizza\textsuperscript{1}, Y.-Q.Chen\textsuperscript{2}, K.Sperber\textsuperscript{3}

\textsuperscript{1}ECMWF, UK
\textsuperscript{2}Department of Atmospheric Sciences, University of Washington, USA.
\textsuperscript{3}PCMDI, Lawrence Livermore Laboratories, USA.

Why use ensembles?

As a legacy of the TOGA programme, it is likely that prediction of interannual climate fluctuations using dynamically-based models will increasingly become an operational activity in many countries. And yet we should remember Tennekes et al.'s (1987) Popperian assertion that "no forecast is complete without a forecast of forecast skill". The role that ensemble techniques can play in estimating the predictability of seasonal forecasts is addressed here through a series of questions. Within this discussion we will consider the specific role that adjoint techniques can play in seasonal predictability studies.

How are ensembles used in practice?

There are two distinct uses of ensemble techniques in seasonal prediction research. A seasonal forecast is an initial value problem (prediction of the first kind) in which predictability is believed to derive primarily from the ocean and land surface. Predictability decays with forecast time as uncertainties in initial conditions grow linearly and saturate nonlinearly. In the first use of the ensemble technique, the initial ensemble is a finite sample of the probability distribution of the initial state (associated with errors and uncertainties in measuring it). The initial ensemble is integrated with the full nonlinear prediction equation to give an estimate of the forecast probability distribution. This technique is widely used in medium-range weather forecasting (Palmer et al, 1993; Toth and Kalnay, 1993), and can readily be extended to seasonal prediction. For the latter, it will be important for the ensemble to represent errors both in the initial state of the atmosphere and in the initial state of the ocean. An important problem is that the dimension of the phase-space associated with the equations of motion vastly exceeds the ensemble sample size. One therefore needs a strategy for choosing the initial sample of perturbations for these ensembles (of the first kind).

In research mode, ensembles are also used to study the impact of prescribed anomalies in the atmosphere's lower boundary forcing. Commonly, such anomalies are made in sea surface temperature field (SST). These are predictions of the second kind. Hence, in dynamical system's language, an ensemble of the second kind is made to estimate changes to the geometry in the atmospheric chaotic attractor as a result of the imposed forcing anomaly. As such, the principal requirement is that the trajectories from the ensemble of integrations adequately cover the attractor. Hence, the choice of initial perturbations for ensembles of the second kind is not critical.

Why not use stochastic-dynamic equations?

If we are interested in the evolution of some probability distribution function, why not frame
the prognostic variables in terms of the moments of this distribution function? This question is often asked when the cost of ensemble integration is considered. However, one of the principal problems of stochastic-dynamic prediction is moment closure. Assumptions that higher order moments are small or slowly varying are often hard to justify. Consider for example, the Lorenz (1963) model. It can be shown that truncation to second order moments can lead to a prediction of negative ensemble variances, obviously unphysical (C. Nicholis, personal communication, 1994).

How large does an ensemble have to be?

This question can only be answered by actually making ensemble integrations. At ECMWF we have integrated the T63L19 model over 120-day periods using prescribed observed SSTs during the 1980s. Early results are discussed in Brankovic et al (1994). For each season we have constructed a 9-member ensemble using as initial conditions, consecutive 24 hour analyses (this is an ensemble of the second kind, hence results should not be sensitive to the method for producing initial perturbations). By comparing differences between ensembles using SSTs from different years, we can determine how many members are needed before simulated atmospheric fields are statistically significantly different.

Fig 1 shows the level of statistical significance for different ensemble sizes and for different regions, and for different years. For brevity we concentrate only on the northern summer season (JJA). The field under study is rainfall, a variable of primary practical importance. The solid circles are for rainfall differences between seasons in which the ENSO was strong and opposite (one year was chosen from the 1987 warm event, the other from the 1988 cold event). The open circles are for rainfall differences between years in which ENSO was approximately neutral (1990 and 1989).

For the African Sahel, rainfall differences between strong and opposite ENSO years could be simulated reliably with just 3-member ensembles (with 3 or more members the level of confidence that the ensemble populations are different exceeds 99% using a t-test). On the other hand, the full 9-member ensemble was required, in order to obtain 90% significance during non-ENSO years. By extrapolation, more than 20 members might be required, in order to obtain 99% significance. For the Indian summer monsoon, rainfall differences for strong and opposite ENSO years are only significant at the 99% level with (an estimated) 10 members, and the level of significance is very poor for the non-ENSO years.

These results confirm the importance of El Nino on seasonal predictability of the Asian and African monsoons. The relatively poorer level of predictability for the Indian monsoon may be associated with the fact that the Asian monsoon has more internal chaotic low-frequency dynamical variability on the intraseasonal timescale, associated with the irregular meridional movement of the ITCZ associated with active and break monsoon spells (Palmer 1994).

In Fig 1c,d we show rainfall predictability significance estimates for two extratropical regions in northern summer. For southern Europe, the 9-member ensemble differences are significant at 99% for strong and opposite ENSO years. However, for the neutral ENSO years, extrapolation of the t-values suggests that significance could be achieved with less than 20 members. For the eastern USA it is interesting to note that a higher level of significance was achieved for the neutral ENSO years than for the strong ENSO years. This might be indicative of the importance either of in situ land surface anomalies or of extratropical SST
anomalies. A full documentation of these results is underway.

**Is there any advantage in generating ensembles by perturbing model formulation?**

In principle, if we had a perfect model, then, in general, an ensemble (of the second kind) generated by sufficiently small and random perturbations to model formulation should be equivalent to an ensemble generated by perturbations to initial conditions. In practice, however, all current GCMs have significant non-random model error. Since climate is nonlinear, it is a forlorn hope to expect that estimates of intrinsic predictability, as shown in Fig 1, are independent of model error.

A second means of constructing an ensemble of the second kind is to use different available GCMs, each forced with the same specified SSTs. Such a dataset is available through the Atmospheric Model Intercomparison Project (AMIP). Fig 2-4 shows timeseries of rainfall simulations for the Brazilian Nordeste (MAM), the African monsoon (JAS) and the Indian summer monsoon (JJAS), using observed specified SST from 1979 to 1988. In each case, the top figure shows timeseries of model rainfall estimates from each member of the ensemble. Each rainfall estimate is a seasonal average, standardised with respect to the individual model climatology. The observed rainfall is shown as the solid dark line.

Also shown for each region, are time-series of members of a sub-ensemble created from those models which passed a quality control procedure. This procedure essentially eliminated all models which had a poor climatology for the region under investigation, or were unable to simulate climatological teleconnections between rainfall in the region of interest, and large-scale patterns of SST (observed teleconnections are shown in the second panels of Fig 2-4). Finally, timeseries of 6 ECMWF AMIP integrations are shown, made by varying initial conditions.

First of all we assess the potential predictability associated with these ensembles by comparing the ratio of the variance of the ensemble mean fields, within the 10 years (the part that is assumed to be associated with interannual variations in SST), with the contemporaneous variance within the ensemble, averaged over the 10 years (the part associated with internal dynamics independent of SST). The interpretation is that the larger the ratio the more potentially predictable the rainfall fluctuations are. This number is shown in the top left hand corner of each diagram.

It can be seen that the quality control has the effect of increasing the potential predictability of seasonally-averaged rainfall for all regions. However, compared with the quality-controlled set, the ECMWF ensemble has significantly more potential predictability. A conclusion of such a study is that assessments of potential predictability from multi-model ensembles without quality control are unduly pessimistic. However, whilst estimates from the single-model ECMWF ensembles could be unreliably optimistic, present results do not indicate that this is a serious problem.

An objective measure of actual skill for these multi-model ensembles has been calculated using a two-category Brier score (years where the observed rainfall anomaly was close to zero were excluded). The higher the Brier score, the less skilful is the ensemble in estimating the probability of either above or below average rainfall, compared with observations. A climatological forecast would give a Brier score of 0.5.
The results show that, taking all models, interannual fluctuations in Nordeste rainfall simulations were most skilful, Sahel rainfall simulations were least skilful (and less skilful than climatology). On the other hand, if the quality-controlled integrations only are assessed, then for all three areas the Brier score is more skilful than climatology, though the impact of quality control is weakest for the Indian monsoon.

The relatively large impact that quality control had on Sahel rainfall, together with the results in Fig 1, suggest that the Sahel is not a region of intrinsically low predictability. However, it is a region where model error may tend to be large. By contrast, the Indian monsoon region may be an area where both model error can be large, and predictability low. Finally, predictability is high and model error low in the Nordeste region.

Are singular vectors useful in studying seasonal predictability?

Singular vector analysis is the most general technique for studying the linear growth of perturbations on a unrestricted time evolving basic state. Specifically, let $X(t)$ be the state vector associated with a nonlinear solution of the equations of motion, and let $L(t_0,t)$ be the linear propagator which maps small perturbations of the state vector at $t_0$ to small perturbations at $t$, according to the equations of motions. Then the singular vectors and the corresponding singular values are given by the eigenvectors and (square root of the) eigenvalues of the operator $L* L$ where * denotes the matrix transpose or operator adjoint. This product operator is normal so that its eigenvectors are orthogonal. Hence, in particular, the singular vector with largest singular value bounds all possible perturbation growth between $t_0$ and $t$.

Fig 5 shows the spectrum of the first 18 singular values for a particular (but typical) 2-day period based on the T42L19 ECMWF atmospheric model (eg Buizza and Palmer, 1995). Fig 5a shows growth optimised for the extratropics; Fig 5b shows growth optimised for the tropics. The solid line shows the energy of the singular vectors at optimisation time, the dashed lines shows the energy at initial time (multiplied by a factor of 40 to make them more visible on the diagram).

The spectra of the extratropical singular vectors show a profoundly upscale energy cascade, indicating that an inability to correctly forecast small sub-synoptic scale perturbations will influence the prediction of synoptic (and ultimately planetary-scale) components of the atmospheric circulation. This behaviour characterises the "butterfly effect" in a quantitative but completely linear context. By contrast, the tropical singular vectors have energy spectra which are much more similar at initial and final time. This illustrates from a theoretical linear calculation the practical experience that the inability to forecast the detailed synoptic or sub-synoptic flow in the tropics does not profoundly influence the predictability of the tropical planetary scales. These singular vector calculations complement the earlier remarks of Charney and Shukla (1981) on the reasons why the large-scale tropical circulation is more predictable than the extratropics: not only is the tropical circulation less generically unstable, (unpredictable) energy does not cascade significantly from sub-synoptic to planetary scale.

Another application of singular vector analysis to seasonal prediction studies was given in Molteni and Palmer(1993)'s analysis of the barotropic finite-time instability of large-scale weather regimes of the northern winter circulation, specifically positive and negative PNA
patterns. Results accord with experience that low-frequency intraseasonal variability of the PNA region is larger during La Nina events than during El Nino episodes. The dominant 8-day optimised singular values for the positive and negative PNA basic states are 5.9 and 11.2 respectively.

As mentioned above, singular vectors are used as a strategy for selecting potentially important initial perturbations for medium-range ensemble prediction where the dimension of the initial probability distribution of analysis error is too large to be sampled systematically. They can also be used for studying the predictability of coupled ocean-atmosphere dynamics (Xue et al, 1994). Singular vectors from the Battisti (1988) intermediate coupled ocean atmosphere model have been estimated both for smooth climatological and unsmoothed time evolving nonlinear solutions of the model. Some preliminary results have been given in Palmer et al (1994).

An example of the SST associated with the dominant singular vector from this coupled model, linearised about a climatological basic state, is given in Fig 6. The calculations are done for 6 month intervals starting in April (a-b) or in October (c-d). The dominant singular value is 46 for the April start, 10 for the October start. This behaviour accords well with perturbation growth in 'identical-twin' coupled GCM experiments, with the skill of intermediate models as a function of time of year (Latif et al 1994), and with the general notion of the spring predictability barrier. Note that at optimisation time the structure of the singular vector resembles a canonical El Nino SST anomaly. At initial time the singular vector structure shows a dipole pattern; any initial error in measuring SST which projects onto this initial structure could have a serious impact on the skill of the SST forecast.

Fig 6e shows the variation of 3-month dominant singular values with basic state. Results suggest that the basic state is more unstable leading up to a warm event than leading up to a cold event. As such, the failure of models to forecast well the development of the 1988/89 cold event could be an indication of the predominance of model error, rather than the event having intrinsically small predictability. This situation could be compared with that of simulating interannual fluctuations in Sahel rain in the AMIP study (see above). It can be noted that some coupled GCMs (Dr T.Stockdale, personal communication, 1995) do show increased ENSO variability as the mean temperature in the Pacific basin increases.

The ability to calculate singular vectors in large-dimension models is intimately linked with the use of adjoint models. Such adjoint models are also an essential feature of 4-D variational data assimilation schemes (Thépaut et al, 1994). In practice, the use of singular vector techniques for coupled ensemble prediction may depend on whether variational schemes are used to determine the ocean-atmosphere initial state.

Another relevant application of the adjoint technique is in the diagnostic calculation of the gradient of forecast error with respect to the initial state (computed by integrating forecast error back to the initial time using the adjoint model). This has proved a valuable technique in numerical weather prediction in tracing back forecast errors to the initial time, and identifying critical regions where the initial analysis was in error (Rabier et al, 1994). It can be shown analytically that the gradient function so obtained will project onto the dominant unstable subspace determined by the associated singular vectors. In principal, this technique can be applied to coupled model forecasts. By backtracking actual SST forecast errors, one can assess the extent to which initial errors projected onto dominant singular vectors such as
shown in Fig 6.

What are the prospects for operational ensemble seasonal prediction in the next millennium?

We have shown that whilst seasonal forecasting is an initial value problem, estimates of seasonal predictability are likely to be model dependent, due to model-dependent error. It would be naive to suppose that all model error will have been eliminated by the turn of the century. On the other hand, many institutes will be producing real-time seasonal forecasts. Hopefully all will be based on ensemble techniques. From a scientific point of view the results in this paper indicate that users of seasonal forecasts would be best served through probability forecasts made from a 'hyper-ensemble' based on a (quality-controlled) multi-model multi-initial-condition ensemble. As a scientist, I hope that political constraints do not prevent such a strategy from being realised.

References


The t statistic associated with the null hypothesis that the seasonal mean JJA rainfall samples from two $9 \leq n \leq 3$ member ensembles from two different years are indistinguishable. The solid symbols are associated with differences between two years in which El Nino was strong and opposite (i.e., a warm event and a cold event). The open symbols are associated with differences over land points between two years in which El Nino was essentially neutral. Shown here are:

a) The African Sahel

b) India

c) Southern Europe

d) Eastern USA.

The 90% and 99% confidence lines (that the hypothesis can be rejected) are shown.
Figure 2

Timeseries of standardised rainfall for GCMs taking part in AMIP for the Brazilian Nordeste (MAM).

a) all models

c) a quality controlled subset of AMIP GCMs based on whether a model could reproduce the observed teleconnection between observed rainfall and global SST (the teleconnection pattern is shown in b)).

d) 6 ECMWF AMIP integrations made by varying initial conditions.

Associated with each timeseries is the "potential predictability", giving the ratio of ensemble-mean variance explained by SST, to intra-ensemble variance considered to be unpredictable. The number in the top right hand corner gives a two-category Brier score against the observed rainfall (heavy solid line).
Figure 3
As Figure 2 but for the African Sahel (JAS).
Figure 5
Energy spectra of the first 18 atmospheric singular vectors for a typical 2-day period, optimised for:

a) extratropical growth

b) tropical growth.

The dashed line gives the energy spectra at initial time (multiplied by 40), the solid line gives the energy at optimisation time (2 days later). Calculated from T42L19 integrations of the ECMWF forward and adjoint tangent models.
Figure 6

a)-d) SST of a dominant singular vector of the Battisti intermediate coupled model, for a 6 month optimisation time, linearised about a climatological basic state.

a) at initial time (April),    b) at final time (October)
c) at initial time (October)  d) at final time (April).

The relevant singular value is given at the top of a) and c).

e) dominant singular value for 3 month optimisation for a time-varying