

On Assessing the Relative Roles of Initial and Boundary Conditions in Interannual to Decadal Climate Predictability

Matthew Collins¹ and Myles R. Allen²

October 9, 2001

¹ Centre for Global Atmospheric Modelling, Department of Meteorology, University of Reading, Reading, UK. Also at Hadley Centre for Climate Prediction and Research, The Met. Office, Bracknell, UK.

² Space Science and Technology Department, Rutherford Appleton Laboratory, Didcot, Oxfordshire, UK.

Abstract

We seek to address the relative importance of initial conditions and boundary conditions in interannual to decadal climate predictability. We develop a simple framework in which: (i) ensembles of climate model simulations with changing external forcing can be measured against climatology to get an estimate of the time scale on which changing boundary conditions can provide predictive skill, and (ii) the rate of spread of ensembles of simulations with small perturbations to the initial conditions can be measured against climatology to assess the time scale at which the information in the initial conditions

is degraded by chaotic error growth. We present a preliminary test of the method on a limited number of climate model simulations.

1 Introduction

Predictability of climate on interannual to decadal time scales can arise in two distinct ways (Lorenz (1975)):

Predictability of the first kind focusses on the initial value problem: how uncertainties in the initial state of the climate system amplify and spoil the forecast at a given lead time. This is the dominant source of error in weather forecasting. Errors in the atmospheric analysis grow such that the current predictability horizon is often of the order of a week or so (e.g. Molteni *et al* (1996)). In climate forecasting, it is clearly not possible to forecast individual weather events beyond this limit. Slow processes, such as changes in ocean circulation, might, however, lead to predictability of the bulk statistical properties of weather and, in principle, give useful information on interannual time scales and beyond – e.g. forecasting of the El Niño Southern Oscillation (ENSO), (e.g. Stockdale *et al* (1998)). To first order, it is the specification of the initial ocean state which is crucial in the first kind predictability of climate since this provides the possible “memory” of the system (although other factors such as variations in land cover may be important).

Predictability of the second kind focusses on the boundary value problem: how predictable changes in the boundary conditions which affect climate can provide predictive power. A common class of second kind predictability studies use atmosphere models with prescribed sea surface temperatures (SSTs) in order to assess the upper limit of predictability associated with a perfect knowledge of the future ocean state (e.g. Zwiers (1996), Rowell (1998)). Here we focus on a different problem; that of future changes in external forcing agents of climate, principally anthropogenic increases in greenhouse gases and other radia-

tively active substances. Changes in radiative forcing can lead to changes in climate which are outside the “noise” of climate variability and, to the extent that these forcing changes are themselves predictable, thus provide a source of potentially useful skill (Stott *et al* (2000)). The usual focus of climate change detection studies has been on time scales of 50 years and greater, since their priority has been to attain high levels of confidence by maximising signal-to-noise – e.g. Tett *et al* (1999). Here we assess the prospects for useful skill on shorter time scales.

On a practical level, knowing the role of the initial conditions and boundary conditions will be crucial in the design of any “operational” climate forecasting system. Consider a climate forecast for a lead time of 10 years into the future. If predictability on this time scale turns out to be a first kind (initial value) problem, then there is a requirement for a huge multi-national investment in taking routine observations of the ocean in order to provide an accurate initial analysis with which to start the forecast. Conversely, if even very small errors in the initial conditions (i.e. errors which are smaller than the minimum errors achievable in any feasible ocean-atmosphere analysis system) rapidly saturate, leaving only the response to changing boundary conditions, then we need to focus on understanding the response to relatively predictable changes in radiative forcing, such as anthropogenic factors and (on sub-decadal to decadal timescales) solar irradiance changes. Some changes in boundary conditions, such as explosive volcanic eruptions, are unlikely ever to be predictable in a deterministic sense, but their role still needs to be understood since it places limits on attainable forecast skill.

2 Comparing two kinds of predictability

The question of predictability of the second kind can be addressed by measuring the mean of an ensemble of climate model experiments with changing radiative forcing and comparing

that with some estimate of climatological internal variability. The question of predictability of the first kind can be addressed by taking an ensemble of climate model simulations with small perturbations to the initial conditions and measuring the ensemble spread at increasing lead time against the spread expected from random sampling of climatological noise.

Assume, for example, that we have n experiments with changing radiative forcing (base experiments) from which we have produced m ensemble forecasts with slight changes in initial conditions at some date during the base experiments (so that we have n ensemble forecasts each with m members). Figure 1 displays this experimental set-up, with $n = 4$ and $m = 2$, for the simulations considered later in this paper. Let $x_{ij}(t)$ be a measure of the change in some climate variable over a fixed time interval, t , as observed in the i th member of the j th ensemble forecast. Absolute values could be considered as well as changes, but persistence then introduces additional complications. From here on we take the time dependence, t , as given so we drop all (t) 's from the equations.

We define r as the change in x attributable to the deterministic response to external forcing over this period (i.e. r is the mean change over this period in a hypothetical infinite ensemble). We also define s_j as the change in x attributable to the predictable component of internal variability given the initial conditions of the j th ensemble. The sum $r + s_j$ may be thought of as the mean of another hypothetical infinite ensemble with the same external forcing and initial conditions “close” to those of ensemble j . The term “close” is ambiguous, but because of the wide disparity of time scales between the atmosphere and ocean, the precise definition turns out to be irrelevant, and we take “close” to mean identical ocean initial states and arbitrary selected atmospheric initial states.

The residual, u_{ij} , is the change in x that cannot be predicted even in principle given both initial and boundary condition information. On the assumption of no interaction between

these different components, we have:

$$x_{ij} = r + s_j + u_{ij} \quad (1)$$

The components r , s and u are unknown, but we have an estimate of the likely level of variability in the sum $s_j + u_{ij}$ given by

$$\mathcal{E}((s_j + u_{ij})^2) = \sigma^2, \quad (2)$$

where σ^2 is the variance of changes in x measured over same-length segments, started at the same point in the seasonal cycle, extracted from a long control integration of the model with constant radiative forcing.

The assumption of no interaction between r , s_j and u_{ij} may be too restrictive as we expect that, as is the case in weather forecasting, climate predictability will be state dependent (e.g. predictability during the growth phase of a large El Niño event may be higher than that at the onset of a weak one). We proceed though with this assumption in order to illustrate the method, which can be applied to individual ensemble forecasts on a case-by-case basis to highlight this state-dependence.

Ultimately the decision of which climatological measures to use will be problem specific and may be determined by the end user. However, equations (1) and (2) could be applied to almost any linear diagnostic, and contain all the ingredients we need for our measures of first- and second-kind predictability. It is simplest to address the second kind first.

2.1 Measures of second-kind predictability

We can estimate r simply by averaging over all the base experiments which have a particular change in radiative forcing:

$$\hat{r} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n x_{ij}. \quad (3)$$

Predictability arising from changing boundary conditions might be considered useful when

$$r^2 > \sigma^2, \quad (4)$$

meaning that $>50\%$ of the variance in the forecast quantity can be predicted given accurate knowledge of the boundary conditions, a perfect model and a hypothetical infinite ensemble. This criterion is, of course, somewhat arbitrary: the actual skill required for a forecast to be useful will depend on the details of the application, feasible ensemble size, pay-off matrix and so forth. Nevertheless, it provides a reasonable benchmark.

Our estimate of r will contain noise due to the small sample sizes, so

$$r^2 + \frac{\sigma^2}{mn} < \mathcal{E}(r^2) < r^2 + \frac{\sigma^2}{n} \quad (5)$$

with the lower and upper limits corresponding to long and short forecast lead times respectively. In the cases considered here, boundary condition predictability does not become significant until initial condition predictability is almost negligible, so we will use the long-forecast approximation (lower limit). Provided $n \geq 4$, the difference in the resulting criterion on \hat{r} is at most $O(0.1)$. Substituting for r in the left-hand inequality of equation (5) thus provides a criterion of useful skill arising from time-evolving boundary conditions:

$$\mathcal{B} \equiv \frac{mn}{mn + 1} \frac{\hat{r}^2}{\sigma^2} > 1. \quad (6)$$

2.2 Measures of first-kind predictability

An estimate of the change in x in the j th ensemble that is predictable given only knowledge of the forecast initial conditions is provided by

$$\hat{s}_j = \frac{1}{m} \sum_{i=1}^m x_{ij} - \hat{r}, \quad (7)$$

giving an estimate of the completely unpredictable component

$$\hat{u}_{ij} = x_{ij} - \hat{s}_j - \hat{r} = x_{ij} - \frac{1}{m} \sum_{i=1}^m x_{ij}. \quad (8)$$

Focussing on the role of initial condition information, thus assuming r is negligible (as will generally be the case early in the forecast), a similar criterion to equation (6) is provided by

$$\mathcal{E}(u_{ij}^2) < \frac{\sigma^2}{2}. \quad (9)$$

Again, if this criterion is satisfied, more than half the climatological variance is, in principle, predictable given perfect knowledge of initial conditions, a perfect model and an arbitrarily large ensemble.

Taking into account small ensemble sizes, we have

$$\mathcal{E}(\hat{u}_{ij}^2) = \frac{m-1}{m} \mathcal{E}(u_{ij}^2). \quad (10)$$

Using summation to approximate the role of the expectation operator and substituting for u_{ij} in equation, (9) the criterion becomes

$$\mathcal{I} \equiv 1 - \frac{2}{n(m-1)} \sum_{i=1}^m \sum_{j=1}^n \frac{\hat{u}_{ij}^2}{\sigma^2} > 0. \quad (11)$$

In the specific case described later on where $m = 2$, this reduces to

$$\mathcal{I} \equiv 1 - \frac{1}{n} \sum_{j=1}^n \frac{(x_{1j} - x_{2j})^2}{\sigma^2} > 0. \quad (12)$$

This is intuitively plausible, making the forecast useless when the average spread between members of individual ensembles exceeds the climatological variance in the quantity in question (see also Boer (2000)).

3 A Preliminary Application of the Method

In this section we illustrate the methods developed above by applying them to a limited set of coupled climate model simulations. These simulations were not performed specifically to address the problem of initial and boundary condition predictability, but nevertheless they present a suitable test-bed for our ideas.

The climate model used is version two of the Hadley Centre Coupled Ocean-Atmosphere model. The model has a horizontal resolution of 2.5° latitude and 3.75° longitude in both the atmosphere and ocean components with 19 and 20 levels in the vertical respectively and the two components are coupled using flux adjustments – see Johns *et al* (1997) for more details.

We use nine simulations of the coupled model. Four simulations from 1860 to 1996 with increases in greenhouse gases and sulphate aerosols (Mitchell and Johns (1997)) which we denote GS1-4. From each of these ensemble members, the initial conditions from September 1974 were used to start a further set of simulations with identical greenhouse and sulphate forcing to GS1-4 plus an estimate of the downward trend in stratospheric ozone since 1974: these we denote GSO1-4. Hence in September 1974, GS1 and GSO1 have exactly the same initial conditions in both the atmosphere and ocean, but have slight changes in radiative forcing which can be thought of as a small perturbation to the atmospheric initial state (as do GS2 and GSO2 etc.). Global mean temperatures from these experiments are shown in fig. 1. We also use a control experiment which was run for 200 years with fixed concentrations of greenhouse gases etc. This represents the background internal variability of the climate system.

Fortunately, for our purposes, the impact of stratospheric ozone depletion on surface temperatures is small, at least for the first 10 years, so we can sensibly treat GS1 and GSO1 as two members of an initial-condition ensemble. Moreover, Allen and Tett (1999) have shown that the most accurate reproduction of observed tropospheric temperature trends is obtained by reducing the amplitude of the ozone contribution by approximately a factor of 2, so we can examine potential predictability arising from the total anthropogenic response using the mean of the GS and GSO ensemble members. The precise details of the externally-forced signal used are unimportant, since the magnitude of this signal is changing over time: our aim here is simply to provide an indication of its importance for predictability relative to

initial condition information.

Figure 2 shows \mathcal{B} , the measure of predictability arising from changing boundary conditions, computed for a “forecast” starting in September 1974. We consider two diagnostics: the linear trend versus the net change over the forecast period. For global mean temperatures, the boundary conditions are unimportant at a lead times of less than a decade, but thereafter the signal emerges from the noise and we see the influence of a significant warming trend. For Northern Hemisphere land temperatures, the signal emerges at a similar lead time, although the net-change diagnostic highlights a seasonal dependence on predictability with winter being predictable and summer not. This shows the influence of feedbacks in the climate system (here a snow-albedo feedback) which can significantly modulate the climate change signal.

There is no indication of any second kind predictability at lead times of less than 20 years for North Atlantic SSTs nor for NINO3 SSTs. The climate change signal is more likely to emerge on global scales because of the “noise” being reduced relative to the “signal” by spatial averaging (Stott and Tett (1998)). Also land areas are likely to warm at a greater rate than the ocean because of the difference in thermal inertia and because, over water, much of the excess energy from the increased concentrations of greenhouse gases is taken up in greater evaporation rather than heating.

There appears to be some discrepancy between figure 2 and the typical conclusions of climate change detection and attribution studies, which have often argued that 30 years or more is required for the anthropogenic signal to emerge robustly from climatological noise (Hegerl *et al* (1997)). In fact, there is none: because of the economic implications of a “false positive”, these studies typically demand a high level of confidence before concluding that a signal is clearly detectable. A much lower level of confidence may none-the-less be useful for practical predictability: here we only require that a trend be larger than the climatological

spread of trends of equal length (equivalent to “detection” at the 50% level).

Figure 3 shows \mathcal{I} , the measure of predictability arising from perfect knowledge of the initial conditions, computed for lead times up to 5 years. Plots of \mathcal{I} based in simple differences in temperature show large variations from month-to-month whereas the measure based on trends smoothes this indicator. Clearly larger ensembles are required to make firm conclusions, but it seems that, in general, global temperatures may be predictable up to lead times of a couple of years, North Atlantic SSTs may be predictable on decadal time scales but there is little predictability past a month or so for temperatures over land. These results are comparable to those of Griffies and Bryan (1997), Grötzner *et al* (1999) and Boer (2000). We suspect that skillful lead times may, in fact, be highly variable and will be “state dependent” in the sense of different initial states may be more or less amenable to prediction. That then leaves the problem of either relating ensemble spread to skill, or of determining physical precursors of predictable states.

The \mathcal{I} measure for the NINO3 index shows a rapid drop-off of skill on the seasonal time scale and a return of skill at a lead time of 1-2 years. This seasonal modulation is consistent with the experience of ENSO forecasters, but the suggestion of potential predictability on interannual timescales may indicate a lack of realism in the model ENSO. Sampling uncertainty with these small ensembles is also an issue.

4 Discussion

This study has served to illustrate the importance of two kinds of ensembles: an ensemble of perturbations on a given oceanic initial condition to give an accurate assessment of forecast divergence; and an ensemble of initial conditions to provide information on the state-dependence of this divergence. For standard statistical measures to be easily applicable, these individual ensembles should contain at least four members, unlike the two-member

ensembles shown here. Also, other diagnostic measures of predictability should be applied to variables other than surface air temperature.

Ultimately, first and second kind climate predictability will be limited by our ability to estimate the precise state of the ocean at any given time and to predict the future evolution of external factors affecting climate (e.g. explosive volcanic eruptions). Hence studies of this type must be interpreted carefully as they represent only the potential predictability. The message, however, is clear: both initial and boundary condition information will need to be considered when designing operational climate forecasting systems in the future.

Acknowledgments

We thank Simon Tett who ran the experiments. This work was supported by UK Department of the Environment, Transport and the Regions (PECD/7/12/37) and the UK Natural Environment Research Council through the COAPEC thematic programme (MC) and an Advanced Research Fellowship (MRA), with additional support from the European Commission QUARCC project.

REFERENCES

- Allen, M. R., and S. F. B. Tett, 1999: Checking for model consistency in optimal fingerprinting. *Climate Dynamics*, **15**, 419–434.
- Boer, G., 2000: A study of atmosphere-ocean predictability on long time scales. *Climate Dynamics*, **16**, 469–472.
- Griffes, S. M., and K. Bryan, 1997: A predictability study of simulated North Atlantic multidecadal variability. *Climate Dynamics*, **13**, 459–488.
- Grötzner, A., M. Latif, A. Timmermann, and R. Voss, 1999: Interannual to decadal predictability in a coupled ocean-atmosphere general circulation model. *J. Climate*, **12**, 2607–2624.

- Hegerl, G., K. Hasselmann, U. Cubasch, J. Mitchell, E. Roeckner, R. Voss, and J. Waszkewitz, 1997: Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. *Clim. Dyn.*, **13**(9), 613–634.
- Johns, T. C., R. E. Carnell, J. F. Crossley, J. M. Gregory, J. F. B. Mitchell, C. A. Senior, S. F. B. Tett, and R. A. Wood, 1997: The second Hadley Centre coupled ocean-atmosphere GCM: Model description, spinup and validation. *Climate Dynamics*, **13**, 103–134.
- Lorenz, E. N., 1975: The physical bases of climate and climate modelling. *Climate predictability*, number 16 in GARP Publication Series, WMO, 132–136.
- Mitchell, J. F. B., and T. C. Johns, 1997: On the modification of global warming by sulphate aerosols. *J. Climate*, **10**, 245–267.
- Molteni, F., R. Buzzia, and T. N. Palmer, 1996: The ECMWF ensemble prediction system: Methodology and validation. *Q. J. Roy. Met. Soc.*, **122**, 73–119.
- Rowell, D. P., 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *J. Clim.*, **11**, 109–120.
- Stockdale, T. N., D. L. T. Anderson, J. O. S. Alves, and M. A. Balmaseda, 1998: Global seasonal rainfall forecasts using a coupled ocean-atmosphere model. *Nature*, **392**.
- Stott, P. A., and S. F. B. Tett, 1998: Scale-dependent detection of climate change. *J. Climate*, **11**, 3282–3294.
- , ——, G. S. Jones, M. R. Allen, J. F. B. Mitchell, and G. J. Jenkins, 2000: External control of 20th century temperature by natural and anthropogenic factors. *Science*, **290**, 2133–2137.
- Tett, S. F. B., P. A. Stott, M. R. Allen, W. J. Ingram, and J. F. B. Mitchell, 1999: Causes of twentieth century temperature change near the earth's surface. *Nature*, **399**, 569–572.
- Zwiers, F. W., 1996: Interannual variability and predictability in an ensemble of AMIP climate simulations conducted with the CCC GCM2. *Climate Dynamics*, **12**, 825–847.

List of Figures

- 1 Global mean temperature anomalies for the GS and GSO experiments (see text for more details). The GSO simulations have initial conditions taken from the GS ensemble on 1st September 1974. 14
- 2 \mathcal{B} (see equ. 6), a measure of predictability due to changing boundary conditions, which is 0 when there is no significant response and greater than 1 when the response is significantly greater than climatological noise (as indicated by the shaded region). Curves are shown for four different temperature indices as indicated on the figure. The black curve is \mathcal{B} computed using linear trends over the forecast period and the grey curve is that computed by taking a simple temperature difference between initial and final months. . . 15
- 3 \mathcal{I} (see equ. 9), a measure of the initial condition predictability, which is 1 for a perfect forecast and 0 when the ensemble spread is of the same magnitude as the climatological spread (as indicated by the shaded region). Curves are shown for four different temperature indices as indicated on the figure. The black curve is \mathcal{B} computed using linear trends and the grey curve is that computed by taking simple differences in monthly mean temperature. . . . 16

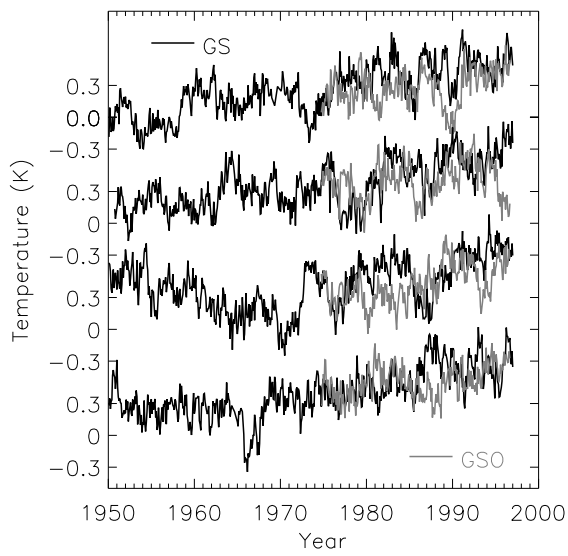


Figure 1: Global mean temperature anomalies for the GS and GSO experiments (see text for more details). The GSO simulations have initial conditions taken from the GS ensemble on 1st September 1974.

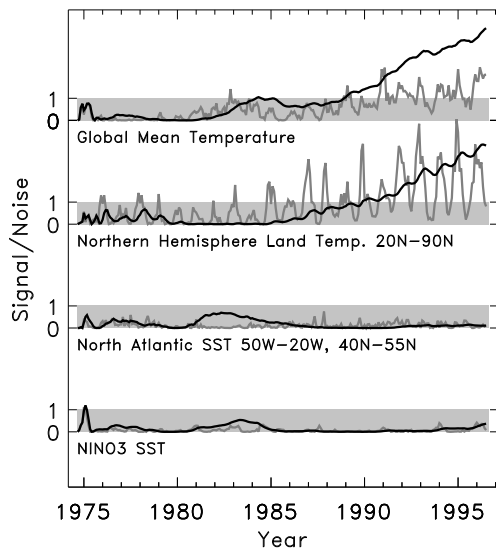


Figure 2: \mathcal{B} (see equ. 6), a measure of predictability due to changing boundary conditions, which is 0 when there is no significant response and greater than 1 when the response is significantly greater than climatological noise (as indicated by the shaded region). Curves are shown for four different temperature indices as indicated on the figure. The black curve is \mathcal{B} computed using linear trends over the forecast period and the grey curve is that computed by taking a simple temperature difference between initial and final months.

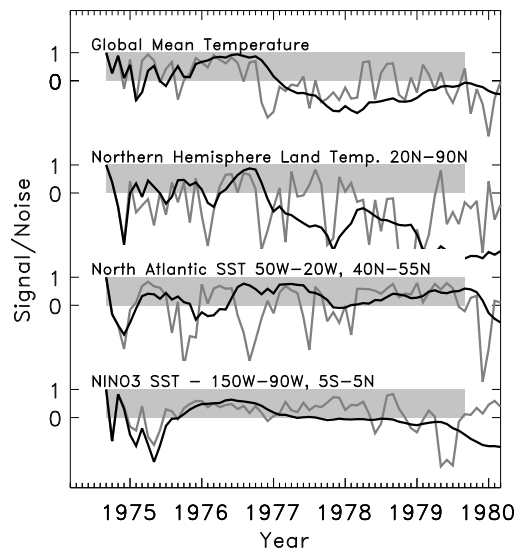


Figure 3: \mathcal{I} (see equ. 9), a measure of the initial condition predictability, which is 1 for a perfect forecast and 0 when the ensemble spread is of the same magnitude as the climatological spread (as indicated by the shaded region). Curves are shown for four different temperature indices as indicated on the figure. The black curve is \mathcal{B} computed using linear trends and the grey curve is that computed by taking simple differences in monthly mean temperature.