

# Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions

By D. B. STEPHENSON<sup>1\*</sup>, C. A. S. COELHO<sup>1</sup>, F. J. DOBLAS-REYES<sup>2</sup> and M. BALMASEDA<sup>2</sup>, <sup>1</sup>*Department of Meteorology, University of Reading, Earley Gate, PO Box 243, Reading RG6 6BB, UK;* <sup>2</sup>*ECMWF, Shinfield Park, Reading RG2 9AX, UK*

(Manuscript received 1 April 2004; in final form 22 September 2004)

## ABSTRACT

In this paper we present a unified conceptual framework for the creation of calibrated probability forecasts of observable variables based on information from ensembles of weather/climate model predictions. For the same reasons that data assimilation is required to feed observed information into numerical prediction models, an analogous process of forecast assimilation is required to convert model predictions into well-calibrated forecasts of observable variables. Forecast assimilation includes and generalizes previous calibration methods such as model output statistics and statistical downscaling. To illustrate the approach, we present a flexible variational form of forecast assimilation based on a Bayesian multivariate normal model capable of assimilating multi-model predictions of gridded fields. This method is then successfully applied to equatorial Pacific sea surface temperature grid point predictions produced by seven coupled models in the DEMETER project. The results show improved forecast skill compared to individual model forecasts and multi-model mean forecasts.

## 1. Introduction

The challenge of improving forecasts has generated many fascinating areas of research in atmospheric and oceanic science such as numerical weather prediction, data assimilation, statistical downscaling, single-model ensemble prediction and, more recently, multi-model ensemble prediction. Tremendous effort has gone into creating reliable and informative forecasts of future observable variables given knowledge of observable variables at earlier times. Such forecasts are inherently probabilistic because of the uncertainties in environmental data (e.g. measurement errors and sparse coverage) and structural and parametric uncertainty in numerical prediction models. Well-calibrated probability forecasts are required for forecast users to be able to make optimal decisions.

Recent studies have focused on how best to combine multi-model climate predictions (Krishnamurti et al., 1999, 2000a,b, 2001; Kharin and Zwiers, 2002; Rajagopalan et al., 2002; Coelho et al., 2003, 2004; etc.). There are many possible methods for combining forecasts (see Clemen, 1989, for a review) but no unique method can be prescribed that is ideal for all the types of weather/climate forecasting problems. However, there is a

need to develop a framework that can incorporate the different approaches for combining weather and climate predictions in order to provide the most informative forecasts of future observables. In addition to the issue of how best to combine multiple predictions, there is also an important issue as to how best to calibrate the predictions. There are many reasons why model predictions should not be taken on face value as forecasts of observable variables (e.g. drift in coupled models that leads to bias) and so calibration is an important step in obtaining useful information for forecast users and risk/impact assessments.

In Coelho et al. (2004), we used a simple Bayesian method to combine single model ensemble predictions with historical empirical data to produce calibrated probabilistic interval forecasts of a single variable. The method was also used by Coelho et al. (2003) to assess the skill of various seasonal forecasts of the Niño-3.4 index produced at the European Centre for Medium-Range Weather Forecasts (ECMWF). In this paper, we generalize the approach to deal with more than one model and more than one variable so that useful forecasts can then be produced from multi-model predictions of spatially gridded fields.

In Section 2 of this paper we introduce a unified framework for the forecasting process and present arguments for the need for forecast assimilation (FA). In Section 3 we describe the equatorial Pacific sea surface temperature (SST) predictions (hindcasts) produced by seven different coupled models

---

\*Corresponding author.  
e-mail: d.b.stephenson@reading.ac.uk

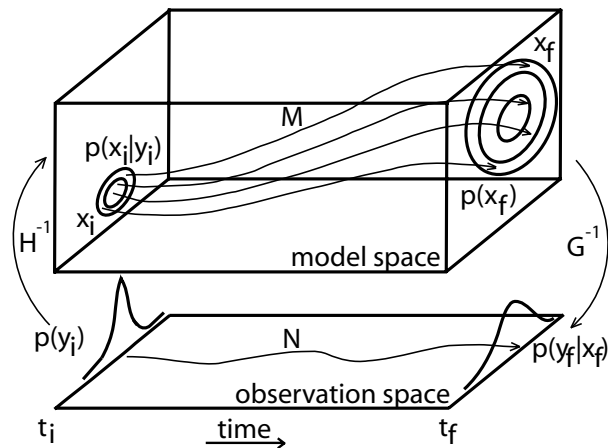


Fig 1. Schematic diagram showing the forecasting process. Time  $t_i$  is the initial time and time  $t_f$  is the final forecast target time. The evolution operator ( $N$ ) in observation space is not known and so numerical forecasting approximates it by mapping observations into model space, evolving model states in time in model space via the model operator ( $M$ ), and then mapping model predictions from  $p$ -dimensional model space back into  $q$ -dimensional observation space. Environmental forecasting is particularly challenging because of the complexity and high dimensionality of the model and observation spaces.

employed in the European Union (EU) Framework 5 project DEMETER<sup>1</sup> (<http://www.ecmwf.int/research/demeter/>). In Section 4 we show how forecast assimilation can be used to combine and calibrate the multi-model predictions and we compare the results to the simpler multi-model ensemble mean approach. In Section 5 we conclude with a summary and suggestions for possible future directions.

## 2. A unified framework for forecasting

### 2.1. Forecasting process

Figure 1 shows a highly simplified (low-dimensional) schematic diagram of the forecasting process. It is important to recognize that observable variables (e.g. temperature at a particular location) are not the same mathematical quantities as model grid point variables. The state vector of the real atmosphere moves dynamically around  $q$ -dimensional ‘observation state space’ whereas the model state vector moves around  $p$ -dimensional ‘model state space’. To initialize models with observations, information in observation state space has to be mapped into model state space using a procedure known as ‘data assimilation’ (Daley, 1991; Courtier, 1997; Boultier and Courtier, 1999). A set of numerical model predictions can then be made to produce an ensemble of possible future model states – a procedure known as ‘ensemble

prediction’ (Palmer, 2000; Stephenson and Doblas-Reyes, 2000, and references therein).

It is often falsely assumed that ensembles of model predictions are probability forecasts of the real world. Model variables are generally neither representative nor unbiased estimates of site-specific observable variables. Instead, model predictions should be considered as proxy information that can be used to infer the probability of future observables. The skill of forecasts depends on their ability to discriminate between observable outcomes (known as forecast resolution; Jolliffe and Stephenson, 2003) rather than their ability to closely match observations. For example, temperature forecasts that distinguish well between hot and cold days but that are always 20°C too warm are more skilful than less biased forecasts that distinguish less well between hot and cold days. To make inferences, one needs a probability model (e.g. a regression model) that can give the probability of observable quantities when provided with model forecast data. There needs to be a procedure for mapping the model predicted state back into observation space. To recognize its analogous role to data assimilation, we will refer to this important final step as ‘forecast assimilation’. In the past, forecast assimilation has been addressed by a wide variety of approaches such as bias-correction, statistical downscaling, model output statistics (MOS), perfect prognosis, etc. (Wilks, 1995). As apparent in Fig. 1, there is a strong duality between data assimilation and forecast assimilation, which will be elaborated mathematically in the following sections of this paper.

To summarize, three important steps are needed in order to find the probability density function  $p(y_f|y_i)$  of a future observable variable  $y_f$ : data assimilation to find  $p(x_i|y_i)$ , model ensemble prediction to find  $p(x_f|x_i)$ , and forecast assimilation to find  $p(y_f|x_f)$ . The desired probability density  $p(y_f|y_i)$  is obtained by integrating over model states using a Monte Carlo approximation (a carefully chosen ensemble). For this to be a good approximation, the initial ensemble states should be sampled from the distribution  $p(x_i|y_i)$  – a condition not always satisfied in the design of current operational ensemble systems (Stephenson and Doblas-Reyes, 2000).

### 2.2. Conditional probabilities and Bayesian combination

Whereas data assimilation is concerned with how best to estimate the probability density function of model state  $x_i$  given observational data  $y_i$ , the dual problem of forecast assimilation is concerned with how best to estimate the probability density of a future observable  $y_f$  given model prediction data  $x_f$ . Both these activities involve the estimation of conditional probabilities:  $p(x_i|y_i)$  for data assimilation and  $p(y_f|x_f)$  for forecast assimilation. The resulting distributions are conditioned on the available data such as observations for data assimilation and model predictions for forecast assimilation. The identity known as Bayes theorem shows how to obtain these conditional

<sup>1</sup>Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction.

probabilities from the unconditional (uninformed) probability distributions (Gelman et al., 1995; Lee, 1997). For example, data assimilation uses

$$p(x_i|y_i) = \frac{p(y_i|x_i)p(x_i)}{p(y_i)} \quad (1)$$

to update the prior (background) distribution  $p(x_i)$  to obtain the posterior distribution  $p(x_i|y_i)$  (Lorenc, 1986; Bouttier and Courtier, 1999). Similarly forecast assimilation uses

$$p(y_f|x_f) = \frac{p(x_f|y_f)p(y_f)}{p(x_f)} \quad (2)$$

to update the prior (climatological) distribution  $p(y_f)$  to obtain the less uncertain posterior distribution  $p(y_f|x_f)$  (Coelho et al., 2003, 2004). Note that initial and final time subscripts are suppressed in subsequent equations. For perfect models with no prediction errors or for forecast with very short lead times, the data assimilation equations can be used to perform forecast assimilation (i.e. perfect prognosis). However, for all other model predictions containing errors, then forecast assimilation is required in addition to data assimilation. From this it can be seen that the true role of ensemble model predictions is to provide data that can be used to update the probability distribution of the observable variable from  $p(y)$  to  $p(y|x)$  rather than to provide an estimate of the distribution  $p(x)$ .

### 2.3. Multivariate normal model

It is useful to briefly review the key equations in data assimilation. The practical implementation of Bayes theorem requires the specification of a suitable probability model. For the sake of generality, consider a  $p$ -dimensional model space and a  $q$ -dimensional observation space where  $p, q \geq 1$ . The least-squares estimation used in variational assimilation is equivalent to maximum likelihood estimation when the  $p \times 1$  model state  $x$  and the  $q \times 1$  observations given a model state are (multivariate) normally distributed

$$x = x_b + \epsilon_B \quad (3)$$

$$y = Hx + \epsilon_R, \quad (4)$$

where the  $p \times 1$  vector  $x_b$  is the background model state (the first guess) and  $\epsilon_B$  and  $\epsilon_R$  are (multivariate) normally distributed errors with zero mean and covariances  $B$  and  $R$ , respectively. The  $q \times p$  matrix  $H$  is known as the observation or interpolation operator that predicts observables (e.g. satellite-measurable radiances) from model states (e.g. vertical temperature profiles). These equations can be written more informatively in probability notation as follows

$$x \sim N(x_b, B) \quad (5)$$

$$y|x \sim N(Hx, R), \quad (6)$$

where  $\sim N(\mu, \Sigma)$  means distributed as a multivariate normal distribution with mean  $\mu$  and covariance  $\Sigma$ . By use of Bayes

theorem (eq. 1), it can then be shown (e.g. Section 4; Bouttier and Courtier, 1999) that these equations lead to

$$x|y \sim N(x_a, A) \quad (7)$$

with the analysis model state  $x_a$  and the analysis error covariance  $A$  given by

$$x_a = x_b + K(y - Hx_b) \quad (8)$$

$$A = (I - KH)B. \quad (9)$$

The  $(p \times q)$  matrix  $K = BH^T(HBH^T + R)^{-1}$  is known as the gain/weight matrix. Therefore, observation data  $y$  can be used to update the background model state  $x_b$  to give an improved analysis estimate  $x_a$  provided one can estimate matrices  $B$ ,  $R$  and  $H$ . The analysis state  $x_a$  is the maximum a posteriori (MAP) estimate (Robert, 2001) that maximizes the probability density  $p(x|y)$ . In other words, the MAP estimate  $x_a$  can be found by minimizing  $-2 \log p(x|y)$  which is given up to a constant by

$$J_{x|y} = (x - x_b)^T B^{-1} (x - x_b) + (y - Hx)^T R^{-1} (y - Hx). \quad (10)$$

The quantity  $J_{x|y}$  is known in variational (e.g. 3-D VAR) data assimilation as the ‘cost function’ and various sparse matrix algorithms can be used to find the value of  $x$  that minimizes this function.

The equations for forecast assimilation of normally distributed predictions are the dual of those for data assimilation with  $x$  and  $y$  interchanged. One assumes that the observable state and the model predictions given an observable state are (multivariate) normally distributed

$$y = y_b + \epsilon_C \quad (11)$$

$$x = G(y - y_0) + \epsilon_S, \quad (12)$$

where  $y_b$  is the background observable state (e.g. the climatological mean value or a persistence forecast) and  $\epsilon_C$  and  $\epsilon_S$  are (multivariate) normally distributed errors with zero mean and background observable covariance  $C$  and forecast error covariance  $S$ , respectively. For generality, a bias term  $y_0$  has been included to take account of the mean bias often found in model predictions.<sup>2</sup> The  $(p \times q)$  matrix  $G$  is the forecast operator (or likelihood) that can be estimated by regression of the model predictions on the observed values. The equations can be rewritten more succinctly as the following probability models:

$$y \sim N(y_b, C) \quad (13)$$

$$x|y \sim N(G(y - y_0), S). \quad (14)$$

<sup>2</sup>A bias term is generally not required for variational data assimilation because the non-linear observation operator  $\mathcal{H}(x)$  is linearized about  $x=x_b$  to obtain the unbiased equation  $y' = Hx' + O(x'^2)$  where  $y' = y - \mathcal{H}(x_b)$  and  $x' = x - x_b$ .

Then, Bayes theorem (eq. 2) can be used to show that

$$y|x \sim N(y_a, D) \quad (15)$$

with the forecast observable state  $y_a$  and the forecast error covariance  $D$  given by

$$y_a = y_b + L[x - G(y_b - y_0)] \quad (16)$$

$$D = (I - LG)C = (G^T S^{-1} G + C^{-1})^{-1}. \quad (17)$$

The  $(q \times p)$  matrix  $L = CG^T(GCG^T + S)^{-1}$  is the forecast gain/weight matrix. Therefore, model prediction data  $x$  can be used to update the background observable state  $y_b$  (e.g. the climatological mean) to give an improved forecast of the observable  $y_a$  provided one can estimate matrices  $C$ ,  $S$  and  $G$  and bias vector  $y_0$ . The forecast observable state  $y_a$  is the MAP estimate that maximizes the probability  $p(y|x)$  or alternatively minimizes  $-2 \log p(y|x)$  given up to a constant by

$$J_{y|x} = (y - y_b)^T C^{-1} (y - y_b) + [x - G(y - y_0)]^T S^{-1} [x - G(y - y_0)]. \quad (18)$$

The quantity  $J_{y|x}$  is the cost function that needs to be minimized for forecast assimilation. It is the sum of two penalty terms: one that penalizes departures  $y - y_b$  from the background observable state and one that penalizes departures  $x - G(y - y_0)$  from calibrated model predictions.

#### 2.4. Estimation and the need for dimension reduction

For multivariate normal forecast assimilation, one needs estimates of vectors  $y_b$  and  $y_0$  and matrices  $C$ ,  $S$  and  $G$ . Vector  $y_b$  and matrix  $C$  are parameters of the prior observable distribution  $y \sim N(y_b, C)$ . Reliable estimates of these parameters can be most simply obtained by calculating the climatological mean and sample covariance of past observations. More precise estimates of the prior can be obtained using empirical forecasts (if available) and so empirical forecasts can be elegantly merged with numerical model predictions (Coelho et al., 2004). The other parameters are obtained by performing a multivariate regression of the model predictions on the observations for a period when both predictions and observations are available (the calibration period). For model predictions we will use the seven ensemble mean forecasts obtained from the seven models rather than 63 ( $7 \times 9$ ) individual forecasts. The slope  $G$ ,  $x = 0$   $y$ -intercept  $y_0$ , and the prediction error covariance  $S$  can be estimated using ordinary least-squares estimation

$$G = S_{xy} S_{yy}^{-1} \quad (19)$$

$$y_0 = -(\bar{x} - \bar{y} G^T) G (G^T G)^{-1} \quad (20)$$

$$S = S_{xx} - S_{xy} S_{yy}^{-1} S_{xy}^T, \quad (21)$$

where  $S_{xx}$  is the  $(p \times p)$  covariance matrix of the model predictions,  $S_{yy}$  is the  $(q \times q)$  covariance matrix of the observ-

ables, and  $S_{xy}$  is the  $(p \times q)$  cross-covariance matrix. Use of individual forecasts rather than ensemble means for each model would give the same slope and intercept estimates but would lead to biased estimates of the prediction error covariance by failing to take into account the model grouping of the individual forecasts.

Calibration of gridded forecasts is particularly difficult because of the large dimensionality of gridded data sets compared to the number of independent forecasts and the strong dependency between values at neighbouring grid points. When the matrix  $S_{yy}$  is poorly conditioned (or even rank deficient, e.g. when  $q > n$ ) then the estimation of  $G = S_{xy} S_{yy}^{-1}$  becomes problematic (or impossible). Furthermore, for  $L$  to be well estimated then  $GCG^T + S$  also has to be a well-conditioned matrix. Put simply, one cannot simultaneously calibrate many predictions if one has only a small historical record of calibration data. This problem becomes even worse for multi-model predictions where the number of grid points is multiplied by the number of models (e.g. in our example,  $p = 7 \times 56 = 392$ ). To avoid this problem, one can use various multivariate dimension reduction techniques to reduce the dimensionality of the data sets. Instead of considering grid point variables, one can project the data on to a small set of spatial patterns to obtain a small number of indices. For example, one could perform principal component regression by using the leading principal components of the model predictions and the observations (Derome et al., 2001; Jolliffe, 2002, section 8.4). Alternatively, one can use either maximum covariance analysis (MCA, sometimes confusingly referred to as SVD) or canonical correlation analysis (CCA) to extract leading co-varying modes from the model prediction and observation data (von Storch and Zwiers, 1999). An MCA-based regression approach has been used in previous studies to improve single model seasonal forecasts (Fedderson et al., 1999; Fedderson, 2003). In the example here, we have tested both MCA and CCA dimension reduction approaches with up to eight retained modes. It was found that MCA with three modes gave the best cross-validated forecast scores (see discussion of Fig. 6 at the end of Section 4).

Forecast assimilation was performed as follows.

(i) In order to produce cross-validated forecasts on data not used in the estimation, the year to be forecast was removed from the data set.

(ii) The time mean was subtracted from the remaining observations and the model predictions to make anomalies stored in an  $(n \times q)$  data matrix  $Y$  of observations and an  $(n \times p)$  data matrix  $X$  of model predictions ( $n = 21$ ,  $q = 56$  and  $p = 392$ ).

(iii) An SVD analysis was performed of the matrix  $Y^T X = U \Sigma V^T$  to determine the leading MCA modes.

(iv) In order to estimate the prior distribution, the background observable covariance matrix  $C$  was calculated for the  $k = 3$  leading MCA modes of the observations. The mean of the prior distribution  $y_b = 0$  since we are treating anomalies about the long-term climatological mean.

(v) A multivariate regression of the  $k$ -leading MCA model prediction modes on the  $k$ -leading MCA observation modes was performed in order to estimate  $G$ ,  $y_0$  and  $S$ .

(vi) The estimated quantities  $C$ ,  $y_b$ ,  $G$ ,  $y_0$  and  $S$  were then used to forecast the observations on removed year using model predictions from that year.

### 2.5. Model output statistics as a special case of forecast assimilation

In the special case when the prior is estimated over the same calibration period used to estimate the prediction error covariance  $S$  (i.e. no extra information is used to estimate  $y_b$  or  $C$ ) then forecast assimilation gives the same forecasts as obtained by ordinary least-squares multivariate regression of the observations on the predictions

$$y = \bar{y} + S_{yx} S_{xx}^{-1} (x - \bar{x}). \quad (22)$$

The proof of this can easily be derived by noting that  $S = S_{xx} - GS_{yy}G^T$ , and so when  $C = S_{yy}$  one obtains  $L = S_{yy}G^T(GS_{yy}G^T + S_{xx} - GS_{yy}G^T)^{-1}$ , which simply becomes  $L = S_{yx}S_{xx}^{-1}$ . Multivariate regression of observations on predictions is the basis of the MOS approach to calibration that has been used in previous multi-model studies (Pavan and Doblas-Reyes, 2000; Kharin and Zwiers, 2002; Doblas-Reyes et al., 2005). Forecast assimilation incorporates this approach as a special case when the prior is estimated using only the calibration data. An important difference in forecast assimilation is that one models the likelihood by regression of the model predictions on the observables rather than directly modelling the observables as functions of model predictions. The likelihood regression in forecast assimilation minimizes uncertainty in the model predictions for given observations whereas the MOS approach minimizes uncertainty in observations for given model predictions. These two approaches will be compared in Section 4 by estimating the prior over only the calibration period 1980–2001 (a MOS approach) and over the extended period 1958–2001 (a Bayesian approach).

## 3. Model predictions of Pacific sea surface temperatures

In this section we present the six-month lead equatorial Pacific SST predictions produced by the seven coupled models participating in the DEMETER project.

The DEMETER project has produced an invaluable multi-model ensemble of global coupled model seasonal hindcasts (i.e. retrospective forecasts made after the events are observed). Coupled model hindcasts were produced four times a year using ERA-40 reanalysis initial conditions starting at 00 GMT on the first day of February, May, August and November (Palmer et al., 2004). Nine-member ensembles for each model were produced by perturbing SST and ocean subsurface initial conditions. We focus on the longest lead six-month ahead predictions for the four target months of July, October, January and April, and

we verify the predictions against the ERA-40 reanalysis SSTs (loosely referred to as observations). Results will be presented based on hindcasts produced by the nine-member ensembles from the seven different coupled models over the common period 1980–2001. The seven coupled models were developed by the following DEMETER project partners: Météo-France (MF), European Centre for Research and Advanced Training in Scientific Computation (CERFACS), Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC), ECMWF, Max-Planck-Institut (MPI) and the UK Met Office (UKMO).

Figure 2 shows latitude–time Hovmöller plots of SST anomalies along the equator in the Pacific sector from Indonesia to the west coast of South America (sampled four times a year: Jan, Apr, Jul, Oct). The observations and model predictions have all been bi-linearly interpolated on to the same  $2.5^\circ \times 2.5^\circ$  grid. The equatorial Pacific section contains 56 grid points along the equator running from  $140^\circ\text{E}$  to  $82.5^\circ\text{W}$ . The observed SST anomalies (Fig. 2a) clearly reveal four major positive anomaly El Niño events (1982/83, 1986/87, 1991/92, 1997/98) separated by negative anomaly La Niña episodes. The individual coupled model DEMETER ensemble-mean hindcasts are shown in Figs. 2b–h— anomalies were produced by subtracting the long-term mean for each calendar month for each of the models. All of the six-month lead model forecasts capture the gross features of the observed El Niño Southern Oscillation (ENSO) events in the Pacific SSTs. However, more careful inspection reveals that all the models except MPI tend to underestimate the peak magnitude of the ENSO events. Furthermore, the models appear to overestimate the spatial extent of certain El Niño events such as the event in 1997/98. With the exception of the MPI model, which shows more variance than the other models and observations, most of the model forecasts are rather similar.

Table 1 summarizes the forecasts using statistics calculated by pooling over all the anomaly data shown in the Hovmöller plots of Fig. 2. The verification scores were constructed by averaging over all 56 grid points and 88 time points shown in the Hovmöller plots. With the exception of the MPI model, the model predictions have similar root mean square error (RMSE) scores (0.82–0.90). The MPI predictions have more variance and lower correlations with the observations and this leads to a much larger RMSE score. Apart from the MPI and UKMO predictions, the models tend to underestimate the maximum values and overestimate the minimum values compared to observations. The scores are similar to those obtained for ENSO forecasts using empirical regression methods (Coelho et al., 2003, 2004).

## 4. Combined/calibrated forecasts

In this section we present combined forecasts obtained using the forecast assimilation equations described in Section 2.3.

Figures 3a–d show latitude–time Hovmöller plots of equatorial Pacific SST anomalies for observations and three different combined forecasts.

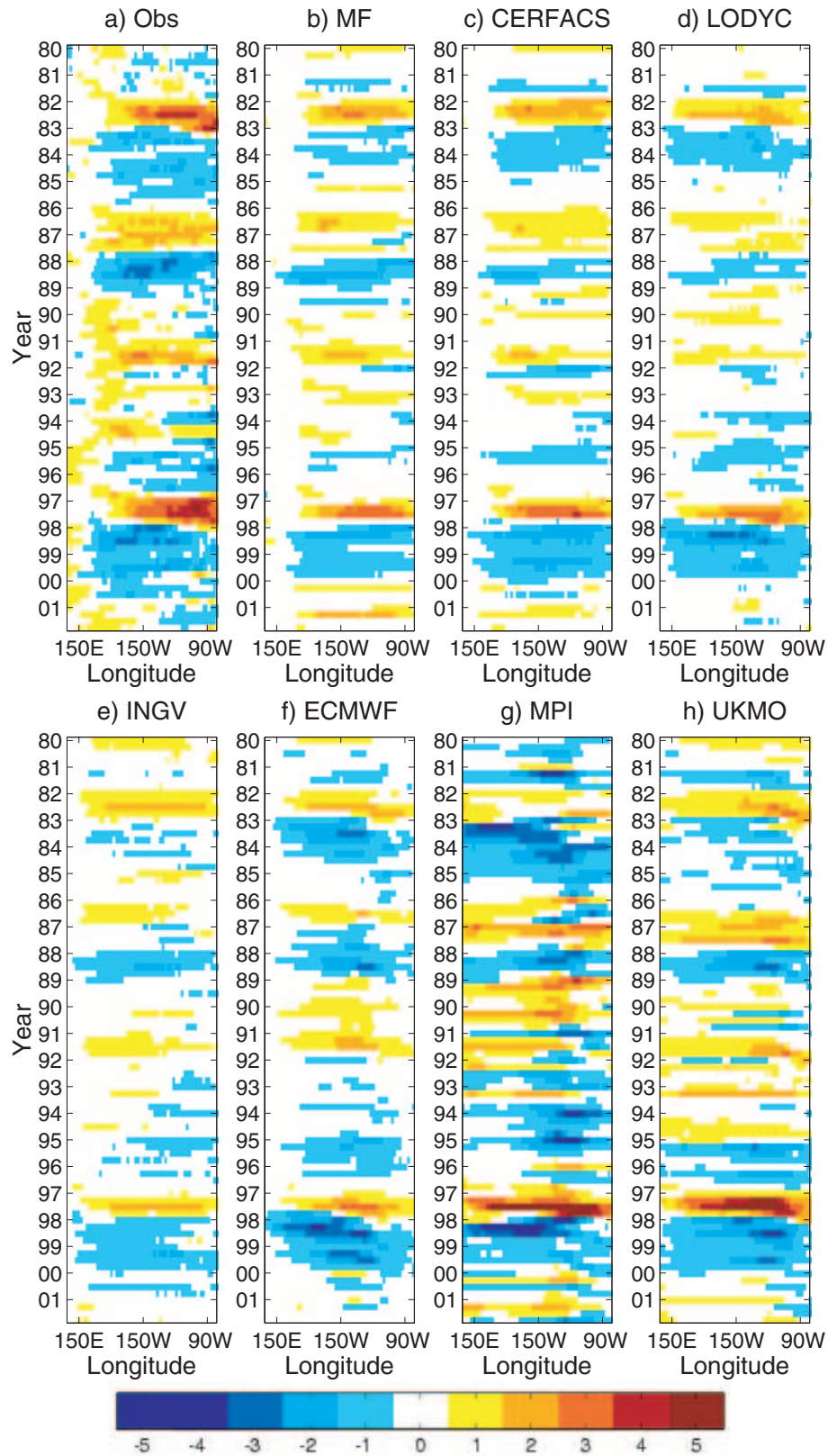


Fig 2. Hovmöller plots of Pacific SST anomalies (in °C) along the equator from July 1980 to July 2001: (a) observed anomalies; (b) MF forecasts; (c) CERFACS forecasts; (d) LODYC forecasts; (e) INGV forecasts; (f) ECMWF forecasts; (g) MPI forecasts; (h) UKMO forecasts.

*Table 1.* Forecast statistics for the various models based on all the gridded anomalies shown in the Hovmöller plots: MIN is the minimum forecast value, MAX is the maximum forecast value (in °C), RMSE is the root mean square forecast error (in °C) and BS is the Brier score. The Brier score is given for probability forecasts of cold events defined by anomalies not exceeding zero. The statistics were calculated by pooling over all the space–time points in the Hovmöller plots. The Brier score for the observations is that obtained by forecasting the climatological value of  $p = 0.5$  for each event at each grid point. The RMSE for the observations is that obtained by forecasting an anomaly of zero

Model	MIN	MAX	RMSE	BS
MF	−2.5	3.4	0.84	0.20
CERFACS	−2.5	3.5	0.82	0.20
LODYC	−3.0	3.0	0.82	0.22
INGV	−1.7	2.3	0.88	0.22
ECMWF	−3.8	3.2	0.89	0.21
MPI	−5.2	9.8	1.46	0.29
UKMO	−4.1	5.7	0.90	0.22
Ens. mean	−3.4	4.4	0.77	0.19
FA 1980–2001	−3.5	4.8	0.75	0.17
FA 1958–2001	−3.4	4.8	0.75	0.17
Obs. 1980–2001	−2.7	4.5	1.13	0.25

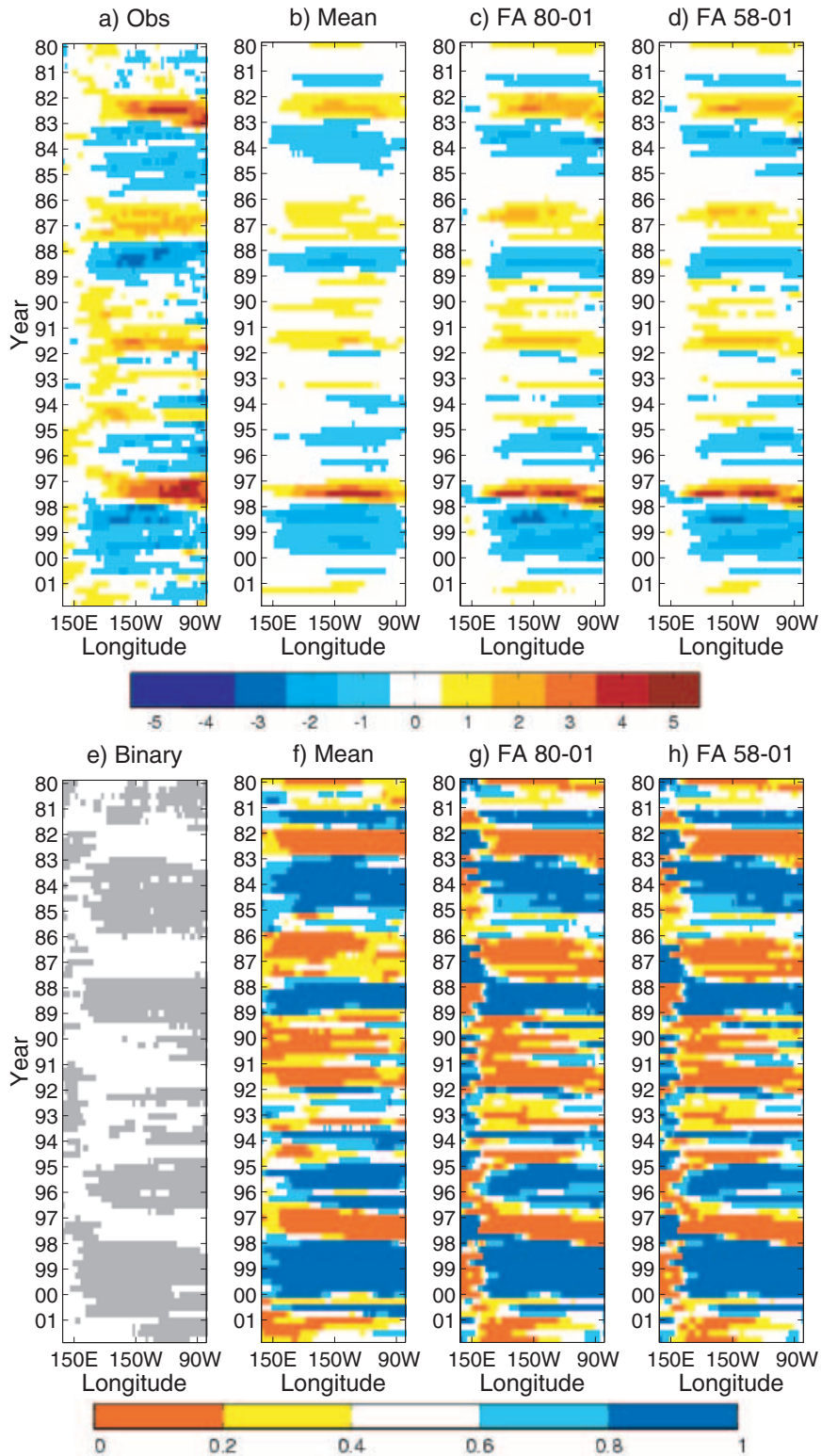
One of the simplest and most naïve ways to combine multi-model predictions is to calculate the ensemble mean  $\bar{X}$  of all the model predictions (Doblas-Reyes et al., 2003; Kharin and Zwiers, 2002; Krishnamurti et al., 2001). By comparing the ensemble mean in Fig. 3b with the observations in Fig. 3a, it can be seen that the ensemble mean captures the main ENSO events. However, more careful examination reveals that the peak magnitudes of the El Niño events in the ensemble mean forecast are slightly smaller than those in the observations and the peak magnitudes of the La Niña events are slightly larger. With the exception of MPI and UKMO, the models generally underestimate the magnitude of the maximum anomalies whereas only one model – Istituto Nazionale di Geofisica e Vulcanologia (INGV) – underestimates the magnitude of the minimum anomalies when compared to observations (Table 1). This leads to the ensemble mean forecast underestimating interannual variance and the positive skewness seen in the observations. Underestimation of skewness in Pacific SST is a known deficiency of many coupled models (Burgers and Stephenson, 1999; Hannachi et al., 2003). The reduction in interannual variability is not a necessary consequence of averaging over the model predictions. In the additive signal plus noise model of Kharin and Zwiers (2002), each of the model forecasts  $X_i$  is considered to be the sum of the observed signal plus some noise,  $X_i = Y + \epsilon_i$ , and so the ensemble mean forecast is then given by  $\bar{X} = Y + \epsilon$  where  $\epsilon$  is the noise. In this situation, averaging over the models reduces the variance of

the noise term but does not reduce the amplitude of the signal component which varies in time.

Figure 3c shows the combined forecast obtained using the forecast assimilation equations described in Section 2.3 with the likelihood and prior estimated over the common period 1980–2001 (FA 1980–2001). As explained in Section 2.5, this special case of forecast assimilation is identical to the more traditional MOS multivariate regression of observations on forecasts. The combined forecast qualitatively resembles the ensemble mean forecast, although there are some important differences in sign in the western Pacific. For comparison, Fig. 3d shows the forecast assimilation forecast where the prior has been estimated over the extended period 1958–2001 (FA 1958–2001). Figure 3d closely resembles Fig. 3c, implying that prior information about observed SSTs does not make a large difference to the results in this particular example. There is sufficient information in the calibration period 1980–2001 to provide a good estimate of the probability density of observed SSTs without the need for more prior information about the observations. However, prior observational information will improve the final probability forecasts in applications where the prior is more informative relative to the model predictions. For example, in applications where the model predictions are less skilful or where the prior is more skilful (e.g. an empirical forecast instead of climatology). From Table 1 it can be seen that all the combined forecasts have smaller RMSE than the individual model predictions and that the forecast assimilation forecasts give slightly smaller RMSE than the ensemble mean forecasts. Furthermore, the combined forecasts give minimum and maximum values that are in closer agreement with the observations than those obtained for individual models.

Optimal decision-making requires an estimate of prediction uncertainty in addition to a forecast of the mean. Prediction uncertainty  $\hat{\sigma}$  is the before-the-event prediction of the RMSE of the forecast. When accurately estimated, the long-term mean prediction uncertainty should equal the standard deviation of the forecast errors. The prediction uncertainty for the ensemble mean forecasts is the standard deviation of the seven model ensemble mean forecasts. This naïve approach will be adopted here, although it is likely to give an underestimate of the true prediction uncertainty due to dependency between different model forecasts. The prediction uncertainty for the forecast assimilation forecasts takes account of some of the model dependency by using the square root of the diagonal elements of matrix  $D$  (eq. 17). One of the big advantages of the forecast assimilation approach over the ensemble mean approach is that it is capable of giving more realistic prediction intervals and hence (as will be shown) more reliable probability forecasts.

To test the skill of probabilistic forecasts, we have taken the simplest case of the binary event defined by when observed SST anomalies at each grid point are less than or equal to zero (Fig. 3e). For each type of combined forecast, we have used the predicted forecast mean and uncertainty to calculate a



*Fig 3.* The upper four panels show Hovmöller plots of Pacific SST anomalies (in  $^{\circ}\text{C}$ ) along the equator from July 1980 to July 2001: (a) observations; (b) multi-model ensemble mean forecast; (c) the forecast assimilation forecast with prior estimated over 1980–2001; (d) the forecast assimilation forecast with prior estimated over 1958–2001. The lower four panels show the binary event defined by when the observed SST anomaly is less than or equal to zero (grey shading in panel e), and the corresponding probability forecasts for the binary event based on the multi-model ensemble mean and standard deviation (panel f), forecast assimilation with the prior from 1980–2001 (panel g), and forecast assimilation with the prior from 1958–2001.



probability for the anomaly being less than or equal to zero

$$p = \text{Prob}(y_a \leq 0) = \Phi\left(\frac{0 - y_a}{\hat{\sigma}}\right), \quad (23)$$

where  $\Phi(z)$  is the area under the standard normal curve to the left of  $z$ . Note that the probability forecast depends on both the forecast of the mean  $y_a$  and the forecast of the prediction uncertainty  $\hat{\sigma}$ . These probability forecasts are shown for the combined forecasts in Figs. 3f–h. Note that the forecast assimilation but not the ensemble mean probabilities are attempting to forecast correctly the reversed events visible to the west of the date line at  $180^\circ$  in Fig. 3e.

The skill of the probability forecasts has been assessed using the well-known Brier score (Wilks, 1995; Jolliffe and Stephenson, 2003). Brier scores for the three combined forecasts have been calculated by pooling over all the space–time points in the Hovmöller plots and are given in Table 1. With the exception of the MPI model, all forecasts have smaller Brier scores than 0.25 and so are more skilful than climatological forecasts that always issue  $p = 0.5$ . The relative improvement of the Brier scores compared to the no-skill Brier score for climatology appears small, but this is a common feature of the Brier score known to occur for even quite skilful forecasting systems. The three combined forecasts have smaller Brier scores than any of the individual models, and so combination has improved the skill of the probability forecasts. The forecast assimilation forecasts have the smallest Brier scores and so provide the most skilful probability forecasts. The use of prior information from 1958–2001 rather than 1980–2001 has little effect on the Brier score of the forecast assimilation forecasts.

To investigate the improvement in the Brier score, Fig. 4 shows the Brier score and its resolution and reliability components as a function of longitude (see Wilks, 1995, p. 262, for a definition and interpretation of Brier score components). Figure 4a shows that the forecasts have the smallest Brier scores of around 0.1 and therefore have the most skill in the central Pacific. The Brier scores are smaller than the score of 0.25 obtained for climatological forecasts with  $p = 0.5$ . To the west of the date line, the forecast assimilation score is markedly smaller than the ensemble mean score, which is worse here than the score for climatology. The forecast assimilation score is also slightly less than that of the ensemble mean forecast eastwards of  $130^\circ\text{W}$ . Figure 4b shows that the improvement in the western Pacific is due to an improvement in the reliability of the forecasts. As can be noted from Figs. 3e–h, the ensemble mean fails to capture the reversed sign of the events in the western Pacific, whereas the forecast assimilation forecasts are able to capture this behaviour. A possible cause for the westward displacement of the forecast anomalies is that cold upwelling biases in four of the models (not shown) suggest that these models may have excessively strong easterly winds. Figure 4c shows that the resolution of the combined forecasts is similar, although there is evidence of slightly

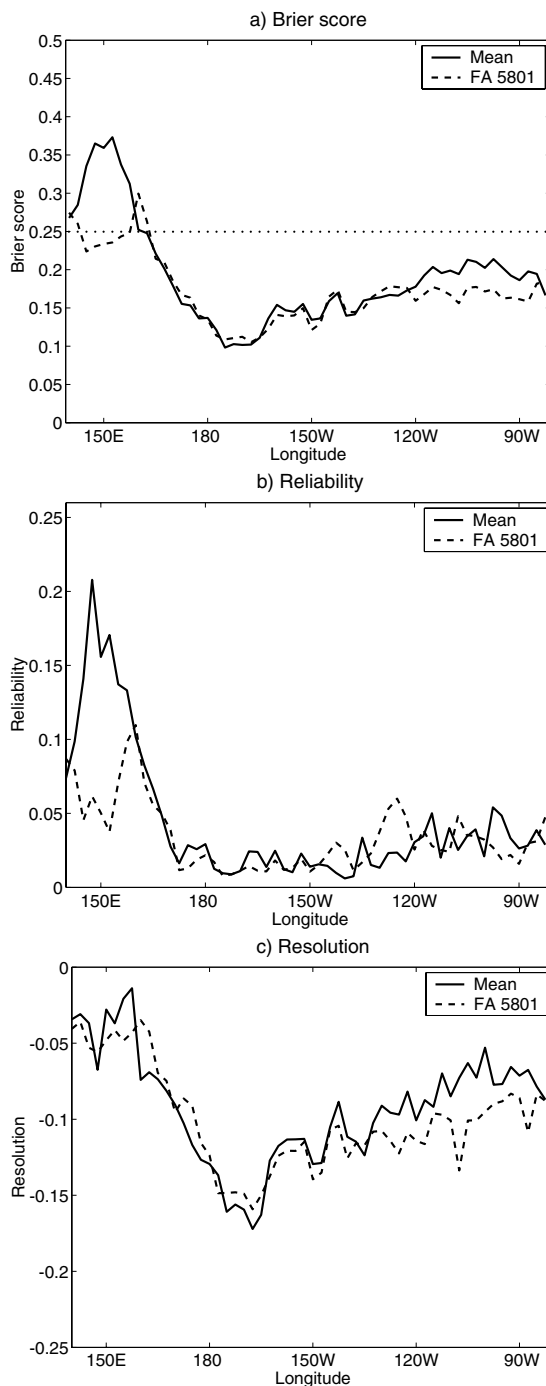


Fig. 4. The Brier score and its components as a function of longitude for the multi-model ensemble (solid line) and forecast assimilation (dashed line) probability forecasts of negative SST anomaly events: (a) the Brier score; (b) the reliability component; (c) the negated resolution component. The reliability and negated resolution components were estimated at each longitude using 10 equally spaced probability bins from 0 to 1. The Brier score is the sum of the reliability, negated resolution, and uncertainty (close to 0.25 at all longitudes) terms. Smaller values indicate more forecast skill.

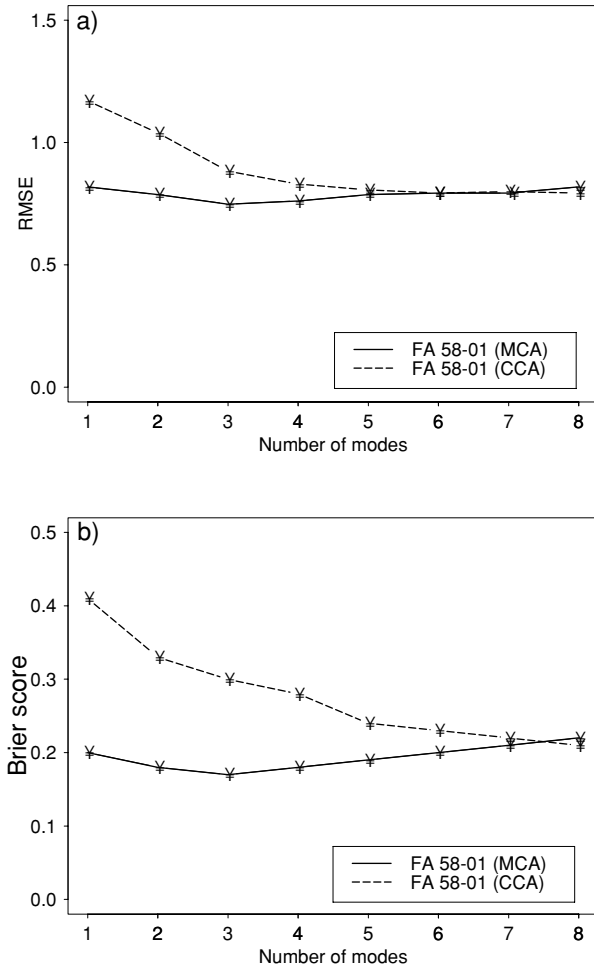


Fig 5. Scores of FA 1958–2001 cross-validated forecasts versus the number retained modes for MCA (solid line) and CCA (dashed line) data reduction: (a) mean square error (in °C); (b) Brier score. Scores are calculated using all longitude and time values in the Hovmöller plots shown in Figs. 3d and 3h.

improved resolution in the forecast assimilation forecasts in the eastern Pacific (east of 135°W).

In order to show how dimension reduction affects the dependency of the skill of the combined forecast, Fig. 5 shows the RMSE and Brier scores for cross-validated combined forecasts as a function of the number of retained modes. For less than five modes, MCA outperforms CCA by producing forecasts with smaller RMSE and Brier scores. The CCA modes are generally noisier both spatially and temporally than the MCA modes. For more modes, MCA and CCA lead to similar forecast scores. The smallest RMSE and Brier score are obtained using MCA with three modes and the scores are not highly sensitive to the addition of more modes. The three leading MCA modes consist of a basin-wide pattern, an east–west dipole over the central-east equatorial Pacific basin, and an east–west basin-wide dipole (not shown). A large fraction of the squared covariance between ob-

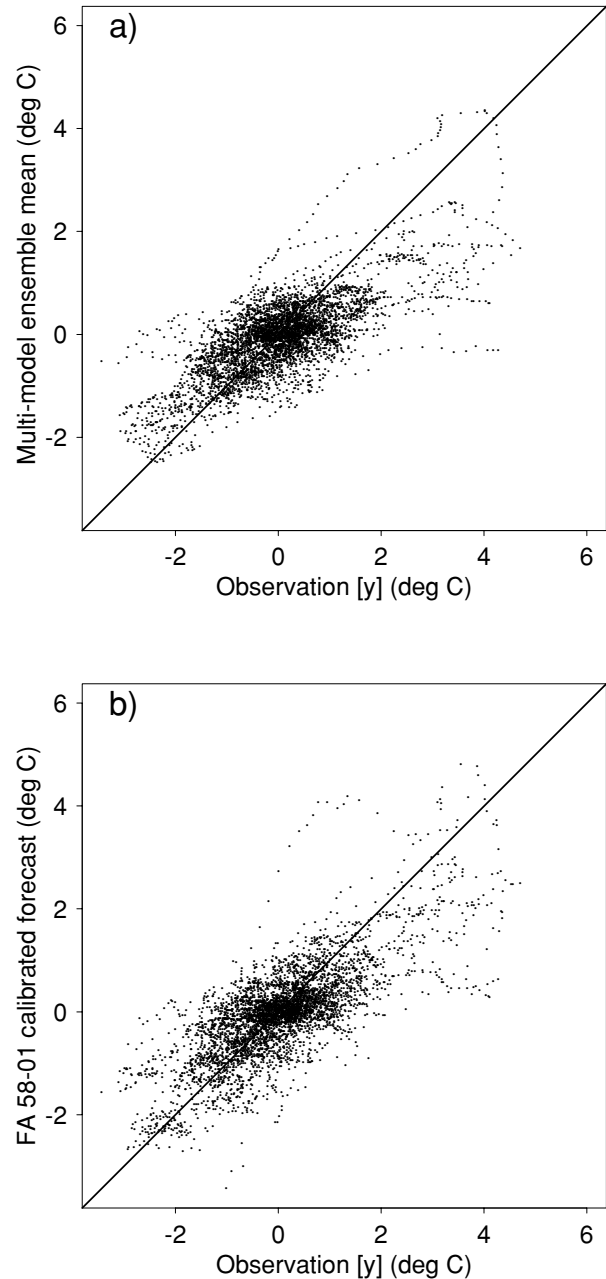


Fig 6. Scatter plots of combined forecasts versus observed SST anomalies (in °C): (a) ensemble mean forecast; (b) forecast assimilation combined forecast using data from 1958–2001.

servations and model predictions (99.7%) is explained by the three leading MCA modes.

Figure 6 shows scatter plots of the combined forecast SST anomalies versus the observed SST anomalies for (a) the ensemble mean forecasts and (b) the 1958–2001 forecast assimilation combined forecasts. The cloud of forecast assimilation forecasts lies much closer to the line  $y = x$  than does the cloud of ensemble mean forecasts. Some positive skewness can also be discerned

in the scatter plots (i.e. more positive anomaly points) caused by the inherent positive skewness in eastern equatorial Pacific SSTs (Burgers and Stephenson, 1999; Hannachi et al., 2003). A more sophisticated forecast assimilation model could be developed to take account of this deviation from normality. The string of outlier points at warm temperatures are caused by warm ENSO events such as the 1997/98 event extending across the equatorial Pacific.

## 5. Conclusion

In this paper we have developed a unifying framework for the production of calibrated probability forecasts of observable variables from multi-model ensemble predictions. In analogy with data assimilation, the concept of forecast assimilation has been introduced. Forecast assimilation is an inherently Bayesian procedure for making improved forecasts of observable variables based on information provided by model predictions. It incorporates many previous techniques such as MOS and statistical downscaling as special cases.

Forecast assimilation has been demonstrated by successful application to DEMETER multi-model equatorial Pacific SST predictions. The resulting combined forecasts reproduce well the temporal and longitudinal variations observed in equatorial Pacific SSTs. In this example, the combined forecast of the mean resembles the multi-model ensemble forecast of the mean. The ensemble mean works well in this case because the majority of the model predictions are for the most part rather similar to one another and closely resemble the observations. Applications with more disparate model predictions are likely to show enhanced skills for the combined forecast compared to the simple approach based on equal-weight averaging the ensemble mean predictions. The prediction uncertainty estimated by forecast assimilation is larger and more realistic than the prediction uncertainty estimated by the standard deviation of the ensemble-mean model predictions. It is important for risk assessment purposes that climate forecasts are able to provide good estimates of forecast uncertainty in addition to providing forecasts of the mean. The approach is easily applied to two-dimensional gridded data, and recent results demonstrate that forecast assimilation works well at reproducing spatial variations in South American rainfall anomalies (work in progress).

Data assimilation and the example of forecast assimilation here assume that the data are normally distributed and treat gridded fields as multivariate vectors instead of taking account of the spatial nature of the data sets. It would be interesting in future studies to extend the methods to non-normally distributed data and to develop flexible functional methods that can exploit the smoothness of the spatial fields as an additional constraint. There is also scope for improving forecast assimilation by including state-dependent ensemble spread information in the estimation of the prediction error covariance matrix  $S$ . Another interesting avenue for future research would be to develop forecast assim-

ilation approaches that use knowledge of the data assimilation operator  $H$  to help estimate  $G$ .

Forecast assimilation is an inherent yet often poorly treated aspect of the forecasting process and it is our hope that this paper will stimulate more coordinated activity in this area for weather and climate predictions on all lead times.

## 6. Acknowledgments

We wish to thank colleagues in the Climate Analysis Group at Reading for useful discussions concerning this work, in particular, Professor Ian Jolliffe for his useful suggestion to use MCA instead of principal component analysis for data reduction. CASC was supported by funding from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) process 200826/00-0 and FJDR was supported by the EU-funded DEMETER project (EVK2-1999-00024).

## References

- Bouttier, F. and Courtier, P. 1999. Data assimilation concepts and methods. *ECMWF Meteorological Training Course Lecture Series*, 59 pp.
- Burgers, G. and Stephenson, D. B. 1999. The 'normality' of El Niño. *Geophys. Res. Lett.* **26**, 1027–1030.
- Clemen, R. T. 1989. Combining forecasts: a review and annotated bibliography. *Int. J. Forecast.* **5**, 559–583.
- Coelho, C. A. S., Pezzulli, S., Balmaseda, M., Doblas-Reyes, F. J. and Stephenson, D. B. 2003. Skill of coupled model seasonal forecasts: A Bayesian assessment of ECMWF ENSO forecasts. *ECMWF Technical Memorandum* **426**, 16 pp.
- Coelho, C. A. S., Pezzulli, S., Balmaseda, M., Doblas-Reyes, F. J. and Stephenson, D. B. 2004. Forecast calibration and combination: a simple Bayesian approach for ENSO. *J. Climate* **17**, 1504–1516.
- Courtier, P. 1997. Dual formulation of four-dimensional variational assimilation. *Q. J. R. Meteorol. Soc.* **123**, 2449–2461.
- Daley, R. 1991. *Atmospheric Data Analysis*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, Cambridge, 457 pp.
- Derome, J., Brunet, G., Plante, A., Gagnon, N., Boer, G. J. et al. 2001. Seasonal predictions based on two dynamical models. *Atmosphere-Ocean* **39**, 485–501.
- Doblas-Reyes, F. J., Deque, M. and Piedelievre, J. P. 2000. Multi-model spread and probabilistic seasonal forecasts in PROVOST. *Q. J. R. Meteorol. Soc.* **126**, 2069–2087.
- Doblas-Reyes, F. J., Pavan, V. and Stephenson, D. B. 2003. The skill of multi-model seasonal forecasts of the wintertime North Atlantic Oscillation. *Climate Dyn.* **21**, 501–514.
- Doblas-Reyes, F.J., Hagedorn, R. and Palmer, T. N. 2005. The rationale behind the success of multi-model ensembles in seasonal forecasting – II. Calibration and combination. *Tellus* **57A**, 234–252.
- Fedderson, H. 2003. Predictability of seasonal precipitation in the Nordic region. *Tellus* **55A**, 385–400.
- Fedderson, H., Navarra, A. and Ward, M. N. 1999. Reduction of model systematic error by statistical correction for dynamical seasonal predictions. *J. Climate* **12**, 1974–1989.

- Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. 1995. *Bayesian Data Analysis*, 2nd edition, Chapman and Hall, London, 526 pp.
- Hannachi, A., Stephenson, D. B. and Sperber, K. R. 2003. Probability-based methods for quantifying nonlinearity in the ENSO. *Climate Dyn.* **20**, 241–256.
- Jolliffe, I. N. 2002. *Principal Component Analysis*, 2nd edition, Springer Verlag, Berlin, 487 pp.
- Jolliffe, I. N. and Stephenson, D. B. 2003. *Forecast Verification: A Practitioner's Guide in Atmospheric Science*, Wiley, New York, 240 pp.
- Kharin, V. V. and Zwiers, F. W. 2002. Climate predictions with multi-model ensembles. *J. Climate* **15**, 793–799.
- Krishnamurti, T. N., Kishtawal, C. M., LaRow, T., Bachiochi, D., Zhang, Z. et al. 1999. Improved weather and seasonal climate forecasts from multi-model superensemble. *Science* **285**, 1548–1550.
- Krishnamurti, T. N., Shin, D. W. and Williford, C. E. 2000a. Improving tropical precipitation forecasts from a multianalysis superensemble. *J. Climate* **13**, 4217–4227.
- Krishnamurti, T. N., Kishtawal, C. M., Zhang, Z., LaRow, T., Bachiochi, D. et al. 2000b. Multi-model ensemble forecasts for weather and seasonal climate. *J. Climate* **13**, 4196–4216.
- Krishnamurti, T. N., Surendran, S., Shin, D. W., Correa-Torres, R. J., Kumar, T. S. V. V. et al. 2001. Real-time multianalysis-multi-model superensemble forecasts of precipitation using TRMM and SSM/I products. *Mon. Wea. Rev.* **129**, 2861–2883.
- Lee, P. M. 1997. *Bayesian Statistics: An Introduction*, 2nd edition, Arnold, London, 344 pp.
- Lorenc, A. 1986. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194.
- Palmer, T. N. 2000. Predicting uncertainty in forecasts of weather and climate. *Rep. Prog. Phys.* **63**, 71–116.
- Palmer, T. N., Alessandri, A., Andersen, U., Cantelaube, P., Davey, M. et al. 2004. Development of a European multi-model ensemble system for seasonal to interannual prediction (DEMETER). *Bull. Am. Meteorol. Soc.* **85**, 853–872.
- Pavan, V. and Doblas-Reyes, F. J. 2000. Multi-model seasonal hindcasts over the Euro-Atlantic: skill scores and dynamic features. *Climate Dyn.* **16**, 611–625.
- Rajagopalan, B., Lall, U. and Zebiak, S. E. 2002. Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. *Mon. Wea. Rev.* **130**, 1792–1811.
- Robert, C. P. 2001. *The Bayesian Choice*, Springer-Verlag, Berlin, 604 pp.
- Stephenson, D. B. and Doblas-Reyes, F. J. 2000. Statistical methods for interpreting Monte Carlo ensemble forecasts. *Tellus* **52A**, 300–322.
- von Storch, H. and Zwiers, F. W. 1999. *Statistical Analysis in Climate Research*, Cambridge University Press, Cambridge, 484 pp.
- Wilks, D. S. 1995. *Statistical Methods in the Atmospheric Sciences: An Introduction*, 1st edition, Academic, New York, 467 pp.