Accepted Manuscript

Multi-model cross-pollination in time

Hailiang Du, Leonard A. Smith

PII: S0167-2789(16)30012-4
DOI: http://dx.doi.org/10.1016/j.physd.2017.06.001
Reference: PHYSD 31916

To appear in: *Physica D*

Received date : 8 January 2016
Revised date : 17 May 2017
Accepted date : 5 June 2017

Please cite this article as: H. Du, L.A. Smith, Multi-model cross-pollination in time, *Physica D* (2017), http://dx.doi.org/10.1016/j.physd.2017.06.001

Multi-model Cross-Pollination in Time

Hailiang Du[1,2] Leonard A. Smith[1,2,3]

[1]Center for Robust Decision Making on Climate and Energy Policy,
University of Chicago, Chicago, IL, US

[2]Centre for the Analysis of Time Series,
London School of Economics, London WC2A 2AE. UK

[3]Pembroke College, Oxford, UK

Email: h.l.du@lse.ac.uk; Tel: +44 07713582128

June 10, 2017

Abstract

The predictive skill of complex models is rarely uniform in model-state space; in weather forecasting models, for example, the skill of the model can be greater in the regions of most interest to a particular operational agency than it is in "remote" regions of the globe. Given a collection of models, a multi-model forecast system using the cross-pollination in time approach can be generalized to take advantage of instances where some models produce forecasts with more information regarding specific components of the model-state than other models, systematically. This generalization is stated and then successfully demonstrated in a moderate (~ 40) dimensional nonlinear dynamical system, suggested by Lorenz, using four imperfect models with similar global forecast skill. Applications to weather forecasting and in economic forecasting are discussed. Given that the relative importance of different phenomena in shaping the weather changes in latitude, changes in attitude among forecast centers in terms of the resources assigned to each phenomena are to be expected. The demonstration establishes that cross-pollinating elements of forecast trajectories enriches the collection of simulations upon which the forecast is built, and given the same collection of models can yield a new forecast system with significantly more skill than the original forecast system.

Keywords: multi-model ensemble; data assimilation; cross-pollination; structural model error.

1 Introduction

Nonlinear dynamical systems are frequently used to model physical processes including the evolution of the solar system, the motion of fluids and the weather. Uncertainty in the observations makes identification of the exact state of the system impossible for a chaotic nonlinear system. This suggests basing forecasts on an ensemble of initial conditions which reflects an inescapable uncertainty from the observations, thereby capturing the sensitivity of each particular forecast. When forecasting real systems like the Earth's atmosphere, there is no reason to believe that a perfect model exists. Generally, the model class from which the particular model equations are drawn does not contain a process that is able to generate trajectories consistent with arbitrarily long time series of observations. In order to take into account both the structural model error and uncertainties in initial conditions, the multi-model and ensemble techniques can be combined into multi-model ensemble forecast systems. A novel approach to this task is presented in this paper.

Multi-model ensembles [13, 17] have become popular tools to investigate, and to better account for shortcomings due to structural model error, in weather and climate simulation-based predictions on time scales from days to seasons and centuries ([15, 17, 30, 31]). While there have been some results suggesting that the multi-model ensemble forecasts outperform the single-model forecasts in an RMS sense (for example, [15, 31]) Smith, et al. [24] challenged the claim that the multi-model ensemble provides a "better" probabilistic forecast than the best single-model. The current multi-model ensemble forecasts are based on combining single-model ensemble forecasts only by means of statistically interpreting ensembles of model simulations to form forecasts of the target system variables. To the extent that each model is developed independently, every single-model is likely to contain different local dynamical information from that of other models; the use of such information has not previously been explored as it is below. In typical statistical processing, such information is only carried by the simulations under a single-model ensemble: no advantage is taken to influence simulations under the other models. This paper presents a novel methodology, named Multi-model Cross-Pollination in Time, a multi-model ensemble scheme with the aim of integrating the dynamical information from each individual model operationally in time. The proposed approach generates model-states in time via applying data assimilation scheme(s) to pseudo-observations drawn from the multi-model forecasts. Illustrated here using the moderate-order Lorenz model [16], the proposed approach is shown to allow significant improvement both upon the traditional statistical processing of multiple, single-model trajectories and upon larger ensembles from the best single-model. It is suggested this illustration could form the basis for more general results which in turn might be deployed in operational forecasting. In weather forecasting, there is a tendency to focus on model performance locally: North America for National Centers for Environmental Prediction (NCEP), Europe for European Centre for Medium-Range Weather Forecasts (ECMWF) and Eastern Asia for Japan Meteorological Agency (JMA). "Local" may correspond to such regional information, or to any neighborhood in a model-state space.

The multi-model ensemble forecast problem of interest is defined, and traditional statistical processing approaches are reviewed, in Section 2. A full review of simple Multi-model Cross-Pollination in Time (CPT I) approach is presented in Section 3. An advanced Multi-model Cross-Pollination in Time (CPT II) approach is presented in Section 4. The experiment, based on a Lorenz 96 system-models pair is designed and the results are presented in Section 5. Section 6 provides a discussion of wider applications and conclusions.

2 Problem description

Outside those problems defined within pure mathematics, there is arguably no perfect model for physical dynamical system [14, 25] evolving smoothly in time. Nevertheless, one may hypothesize a perfect model: a nonlinear system with state space $\mathbb{R}^{\tilde{m}}$, and the evolution operator of the system is $\tilde{G}$ (i.e. $\tilde{\mathbf{x}}(t+1) = \tilde{G}(\tilde{\mathbf{x}}(t))$ where $\tilde{\mathbf{x}}(t) \in \mathbb{R}^{\tilde{m}}$ is the state of the system). $\tilde{G}$, $\tilde{\mathbf{x}}$, and $\tilde{m}$ are unknown. It is often useful to speak as if a mathematically well-defined system existed, regardless of whether or not one actually does exist. An observation of the system state at time t is defined by $\tilde{\mathbf{s}}(t) = \tilde{h}(\tilde{\mathbf{x}}(t)) + \boldsymbol{\eta}(t)$ where $\tilde{\mathbf{s}}(t) \in \mathbb{O}$, $\tilde{h}$ is the observation operator that projects the system state into observation space and $\boldsymbol{\eta}(t)$ represents the observational noise. What is in hand are M models, each of which approximates the system. Each model has the form $\mathbf{x}(t+1;i) = G_i(\mathbf{x}(t;i)), i = 1, \ldots, M$, where $\mathbf{x}(t;i) \in \mathbb{R}^{m(i)}$. $\mathbb{R}^{m(i)}$ is the model-state space of the i^{th} model, G_i. In practice, model-state spaces usually differ from observation space, and it is likely that different models define different model-state spaces. The model-states can be projected into the observation space via an observation operator $h_i(\cdot)$; different models may, of course, also have different operators.

Perhaps the simplest reaction to having M models, each of which provides N-member ensemble forecasts, is to identify the best model, and discard the others. If the models are of comparable quality[1], then it is likely that different models will tend to do better in different regions of state space (for weather models this could be either in different geographical locations or in different synoptic conditions or perhaps different variables). In part, this is due to variations both in tuning and in resource allocation during model construction, reflecting the particular processes that were considered most important.[2]

Let each model producing N-member ensemble forecasts by iterating an N-member initial condition ensemble forward. In practice, such a multi-model forecast system is evaluated using the observations not yet taken. The goal of this paper is to introduce a new multi-model ensemble forecast system (in time) to improve[3] forecast of the future states.

The Model Output Statistics (MOS) has a long and successful history of improving statistically single-model ensemble forecasts (see [32, 33] and references therein). For multi-model ensembles, statistical approaches have been proposed to combine ensembles of individual model simulations to produce a single, probabilistic, multi-model forecast distribution. Most of these approaches are based on weighting the model simulations according to some measure of past performance, see for example [5, 10, 19]. The output of these statistical processing approaches is a function of each individual forecast ensemble. Each single-model ensemble carries only the dynamic information as provided by that model forward in time. The multi-model ensemble framework is designed to reduce the impact of model inadequacy, as different models have different model structures; statistically processing the individual model outputs can not fully explore the local dynamical information available from each individual model. The extension of CPT presented in this paper integrates the dynamical information from each individual model in time; this results in new, truly multi-model trajectories which significantly increase the information in the ensemble of simulations beyond that available from the original multi-model ensemble forecast. This is achieved

[1] Or even in the case some models are inferior on average but more competitive on occasion.

[2] In practice, there is rarely enough data to identify which model will be the best in a particular (out of sample) instance; one reasonable alternative is to compute M N-member ensembles (one ensemble under each model) and treat each ensemble equally.

[3] The improvement is quantified by the information in probabilistic forecasts, as reflected in the Ignorance score $-\log_2(p(Y))$ (see [9] and Section 5).

by communicating information between different models regarding the likely future evolution of the system.

3 Multi-model Cross-Pollination in Time I

To the extent that the structural shortcomings of different models are independent, cross-pollinating trajectories between models to obtain truly multi-model trajectories can allow the ensemble of trajectories to explore important regions of state space that ensembles of individual models just can't reach.

Smith [23] introduced the Cross-Pollination in Time (hereafter CPT I) approach exploiting the assumption that all the models share the same model-state space[4]. Let Δt be the observation time where every Δt time step an observation is recorded. For simplicity, at every observation time all the models provide their model outputs[5]. Let τ be the cross-pollination time, that is, after each period of duration τ a cross-pollination event takes place. Given M N-member ensembles of trajectories, (one ensemble under each model), firstly consider the ensembles of states at $t = \tau$ as one large ensemble of $N \times M$ states in a model-state space. Secondly, use some pruning scheme to reduce this large ensemble to N-member states in order to maintain a manageable ensemble size. While the optimal pruning scheme is, at best, an object of research, the simple approach of identifying the nearest pair of states, and then deleting from the pair the one member with the smallest second nearest neighbor distance has been found [23] to be more effective than random selection in some simple examples.[6] In this paper, a pruning scheme based on the local forecast performance (see Section 5) is adopted to serve the purpose of demonstrating CPT II methodology. Thirdly, adopt this new set of N states as initial conditions at $t = \tau$ and propagate them forward under each of the M models to produce M N-member ensembles of trajectory segments from $t = \tau$ to $t = 2\tau$, the next cross-pollination time. These three steps are repeated until the forecast time of interest is reached; then the ensemble can be interpreted for decision support, for example, providing probabilistic forecasts [4].

[4]Or that there are known one-to-one maps which link their individual state space, given all the models are iterated discretely.

[5]Note it is often the case that the model iteration (simulation) step is much smaller than the observation time, and different models may have different iteration time steps and a different output frequency.

[6]Note that the aim of pruning is quite different than that of resampling from an estimated probability density function [2].

Inasmuch as the CPT I ensemble scheme contains all trajectories of each of its constituent models implicitly, the dynamical information of each model is explored and integrated individually. In practice, however, different models usually define different model-state spaces, and so one-to-one maps linking different model-state spaces may not exist. More relevant for the work below, however, is that CPT I traditionally considers the entire model-state, without careful regard for the fact[7] that some models might forecast some components with greater skill. Under CPT I, each trajectory segment is a trajectory of one of the M models, the cross-pollination is one of trajectory segments; CPT II aims to use the information in the dynamics of each model more effectively. Another challenge for CPT I is that for each model, the initial conditions produced by other models are neither likely to be consistent with that model's dynamics (not "on its attractor" if such a thing exists) nor to prove efficient in sampling initial conditions for the original model. Iterating initial conditions which are "out of balance" is expected to produce potentially odd, transient behaviour. The CPT II approach introduced in the next section frees the methodology from the assumptions above and overcomes some practical shortcomings as well.

4 Multi-model Cross-Pollination in Time II

The Multi-model Cross-Pollination in Time (CPT II) approach not only frees one from the assumption that all models share the same model-state space but also extracts and integrates the dynamical information from each model via exploring a sequence space.

The CPT II approach consists of three steps:

(i) Cast each trajectory in the multi-model ensemble into observation space, to create an ensemble of observation trajectories; these trajectories are then used to produce at least one sequence of states in the observation space.

For each individual model, the forecast ensemble is obtained via iterating an initial condition ensemble forward from $t = 0$ to $t = \tau$, the first CPT time, thereby producing an ensemble of model trajectory segments, from time t_0 to $t_0 + \tau$. Although different models may define different model-state space, every model-state can be projected into observation space using the corresponding observation operator. A model trajectory segment of the i^{th} model,

[7] CPT I could, of course, employ weighting and model skill explicitly in the pruning algorithm. CPT II takes a much more considered approach.

projected into observation space, becomes a series of points,

$$\mathbf{S}(i) \equiv \{h_i(\mathbf{x}(t_0;i)), h_i(\mathbf{x}(t_0+\Delta t;i)), \ldots, h_i(\mathbf{x}(t_0+\tau;i))\},$$

where t_0 is the initial time and $\mathbf{x}(t+\Delta t;i) = G_i^{\Delta t}(\mathbf{x}(t;i))$. As each of the M models has an N-member ensemble, construct one large ensemble of $M \times N$ observation trajectories $\mathbf{S}$(i,j), i=1,...,M and $j = 1, \ldots, N$ in the observation space. There are various statistical processing approaches to cast sequence states into the observation space; traditional MOS approach is one example. In order to maintain a manageable ensemble size, one may prune this large ensemble back into N sequences of states using some pruning scheme (the pruning scheme used in this paper is described in the following section), that is:

$$\left.\begin{array}{l}\{\mathbf{S}(1,1),\ldots,\mathbf{S}(1,N)\}\\ \ldots\\ \{\mathbf{S}(M,1),\ldots,\mathbf{S}(M,N)\}\end{array}\right\} \rightarrow \{\mathbf{S}^\star(1),\ldots,\mathbf{S}^\star(N)\} \tag{1}$$

$\mathbf{S}^\star$ is the combined output of ensemble sequence states in the observation space,

$$\mathbf{S}^\star(j) \equiv \{\mathbf{s}^\star(t_0;j), \mathbf{s}^\star(t_0+\Delta t;j), \ldots, \mathbf{s}^\star(t_0+\tau;j)\}$$

where $\mathbf{s}^\star(t;j) \in \mathbb{O}$.

(ii) Data assimilation of (future) observation sets (defined with a timing resembling actual observations) is made with consideration of the local skill of the model which generated it. For each observation, the model(s) with more skill with regard to each particular observation are favored.

Given N sequences of states in the observation space, $\{\mathbf{S}^\star(1),\ldots,\mathbf{S}^\star(N)\}$, each individual model can apply a data assimilation scheme to each sequence of state to obtain a sequence of model-states in its model-state space; this corresponds to treating the sequence of states in the observation space as a sequence of observations (in the future):

$$\{\mathbf{S}^\star(1),\ldots,\mathbf{S}^\star(N)\} \rightarrow \left\{\begin{array}{l}\{\mathbf{Z}(1,1),\ldots,\mathbf{Z}(1,N)\}\\ \ldots\\ \{\mathbf{Z}(M,1),\ldots,\mathbf{Z}(M,N)\}\end{array}\right. \tag{2}$$

$\mathbf{Z}(i,j)$ is the j^{th} sequence of model-states in the i^{th} model-state space,

$$\mathbf{Z}(i,j) \equiv \{\mathbf{z}(t_0;i;j), \mathbf{z}(t_0+\Delta t;i;j), \ldots, \mathbf{z}(t_0+\tau;i;j)\}$$

where $\mathbf{z}(t;i;j) \in \mathbb{R}^{m(i)}$. $\{\mathbf{Z}(i,1),\ldots,\mathbf{Z}(i,N)\}$ is obtained by applying a data assimilation scheme (using the i^{th} model) to the observation trajectory $\{\mathbf{S}^\star(1),\ldots,\mathbf{S}^\star(N)\}$.

It is not necessary for each model to apply the same data assimilation scheme (in practice, it is likely that each operational forecast system has a unique data assimilation scheme); using existing data assimilation schemes would clearly avoid an extra cost of implementing the CPT II approach. It is, however, noted that applying data assimilation here is crucial in order to extract dynamical information from the model. It is desirable to use a nonlinear data assimilation scheme which is robust (or as robust as currently possible) to structural model error; Pseudo-orbit Data Assimilation (see Du and Smith [6, 7]) is one such scheme. A brief description is given in Appendix A. As no model is perfect, the data assimilation scheme need not aim to obtain model trajectories, but merely pseudo-orbits [7]. Projecting the end component of the model pseudo-orbits, obtained from the data assimilation, into the observation space would provide $N \times M$ states at $t = t_0 + \tau$.

(iii) Iterate new states (from ii) forward.

Take the end component of $\mathbf{Z}(i,j)$, specifically the point $\mathbf{z}(t_0+\tau;i;j)$ to be the initial condition for the j^{th} ensemble member under the i^{th} model, and iterate it forward using the i^{th} model until the next cross-pollination time $t_0+2\tau$. Repeating this over N members produces an ensemble of model trajectory segments, from time $t_0+\tau$ to $t_0+2\tau$.

Repeat steps (i),(ii) and (iii) above starting at $t = t_0+\tau$ to provide forecast states at $t = t_0+2\tau$ and so on, providing an ensemble of future model-states at $t = t_0 + k\tau$, $k = 1, 2, \ldots$.

Cross-Pollination in Time [23] differs fundamentally from other "forecast assimilation" techniques. Stephenson et al. [28] for example introduced a novel approach to forecast assimilation which generalizes earlier calibration methods including model output statistics (see [34]). This approach provides a map from the space of model simulations to the space of observations. In general, any map from the model-state space to the target observable space which uses (only) information available at the time the forecast simulations were launched is admissible. Van den Berge, et al. [29] introduced a multi-model ensemble approach that combines imperfect models into one super-model through the introduction of linear connection terms between the model equations (see Duane [8] and Shen et al. [22] for recent applications). Note that this "super-model" approach builds connections between imperfect models by modifying the model's equations while

CPT II creates the communications between models via data assimilation. Brocker and Smith [4] discuss other approaches to ensemble interpretations. None of these papers[8], however, enable the feedback of forecast-simulation information into the dynamics of the forecast itself. CPT II does precisely this.

5 Experiments based on Lorenz96

A system of nonlinear ordinary differential equations (hereafter, the Lorenz96 System) was introduced by Lorenz [16]. For the system containing n variables $x_1, ..., x_n$ with cyclic boundary conditions (where $x_{n+1} = x_1$), the equations are

$$\frac{dx_i}{dt} = -x_{i-2}x_{i-1} + x_{i-1}x_{i+1} - x_i + F_i, \tag{3}$$

Lorenz initially reported [16] the case where $F_i = F_{fix}$ for all i. The variation of F with i was a consideration at that time (Lorenz 1995, personal communication). The system represents a one-dimensional atmosphere; the n variables $x_1, ..., x_n$ are identified with the values of some unspecified scalar atmospheric quantity at n equally spaced points about a latitude circle called grid points. F_{fix} is a positive constant. It was also found [16] that for $n > 12$, Equation (3) are chaotic when $F_{fix} > 5$.

The *true* system (hereafter, system) used in the following experiments, contains 40 variables, $n = 40$, and the values of the parameter F_i varies with locations, i.e. $F_i = 8$ for $i = 1, \ldots, 10$; $F_i = 12$ for $i = 11, \ldots, 20$; $F_i = 14$ for $i = 21, \ldots, 30$; $F_i = 10$ for $i = 31, \ldots, 40$. Four models are each defined using the same dynamical equation as the system but with fixed value of parameter F_i, that is: model I, $F_i = 8$ for all i; model II, $F_i = 12$; model III, $F_i = 14$ and model IV, $F_i = 10$.

Both the system and the model are integrated using a standard fourth-order Runge-Kutta numerical simulation. The simulation time step is 0.01 time unit and the model time step Δt is 0.05, that is each model time step is conducted by 5 integration steps. Observations $\mathbf{s}(t) \in \mathbb{R}^{40}$ are generated by the system plus IID Gaussian noise, $N(0, \sigma^2_{Noise})$, $\sigma_{Noise} = 0.2$, at every system

[8]The super model approach modifies the dynamics through altering model structure to improve forecasts using information from a global linear fit; it does not, however, blend dynamic forecast-simulation information so as to alter the dynamics of the individual simulations.

time step (0.05 time unit).[9] The cross-pollination time $\tau = 0.4$ corresponds to 2 days[10].

For a forecast initial time t_0, a simple inverse noise method[11] is adopted to generate a 9-member initial condition ensemble for all the models, $\mathbf{IC}(t_0) \equiv \{\mathbf{x}(t_0, j) \in \mathbb{R}^{40}, j = 1, ..., 9\}$. The initial condition ensemble is evolved forward under each of the four models to time $t_0 + \tau$. This gives a 9-member ensemble of model trajectories for each model; for example, for model I: $\{\mathbf{S}(I, 1), \ldots, \mathbf{S}(I, 9)\}$ where $\mathbf{S}(I, j) \equiv \{\mathbf{x}(t_0; I; j), \mathbf{x}(t_0 + \Delta t; I; j), \ldots, \mathbf{x}(t_0 + \tau; I; j)\}$, $\mathbf{x}(t + \Delta t; I; j) = G_I^{\Delta t}(\mathbf{x}(t; I; j))$ and $\mathbf{x}(t; I; j) \equiv \{x_1(t; I, j), x_2(t; I, j), \ldots, x_{40}(t; I, j)\} \in \mathbb{R}^{40}$.

To demonstrate that the proposed CPT II approach is both effective and robust, a long time series of observations are generated using the true system plus IID Gaussian observational noise; the true states of the system are also recorded for the evaluation phase of the experiment. For each model, a large set of *pure* model ensemble forecasts is obtained by launching ensemble forecasts using the inverse noise initial condition ensemble [12] at different observational times. To assess each model's forecasts at various lead-times, the forecast ensemble is translated into a predictive distribution function by kernel dressing and blending with climatological distribution (for a full description see [4], and Appendix B). The interpretation-parameter values required for ensemble interpretation (including kernel width and blending weight, see Appendix B) are fitted using a training set of 2048 ensemble forecasts. The forecast skills of each model is assessed out-of-sample using an independent set of 2048 forecast-outcome pairs.

The forecast performance is evaluated with IJ Good's logarithmic score, Ignorance [9, 20]. Ignorance is the only proper local score for continuous variables [1, 3, 18]. Although there are other nonlocal proper scores, the authors prefer using Ignorance since it is both local and has a clear interpretation in terms of information theory (it can also be easily communicated in terms of an effective interest rate of return [11]).[12] Ignorance is defined by:

$$S(p(y), Y) = -\log_2(p(Y)), \tag{4}$$

[9] In this setting, the model-state space, system state space and the observation space are identical, although as proposed, CPT II is not constrained to operating in this setting.

[10] Assuming 1 time unit is equal to 5 days, the doubling time of the Lorenz96 system roughly matches the characteristic time-scale of dissipation in the atmosphere (see Lorenz [16]).

[11] Given a model of the observational noise, one can add random draws from the inverse of the observational noise model to the observation to define ensemble members [12]. As each ensemble member is an independent draw from the inverse observational noise distribution, each ensemble member is weighted equally.

[12] There are no compelling examples in favor of the general use of nonlocal scores and some nonlocal scores have been shown to produce counter-intuitive evaluations [27].

where Y is the outcome and $p(Y)$ is the probability of the outcome Y. The empirical average Ignorance of a forecast system given K forecast-outcome pairs $\{(p_i, Y_i) \mid i = 1, \ldots, K\}$ is then

$$S_E(p(y), Y) = \frac{1}{K}\sum_{i=1}^{K} -\log_2(p_i(Y_i)), \tag{5}$$

Relative Ignorance reflects the performance of (a set of) forecasts from one model relative to those of a reference forecast p_{ref}:

$$S_{rel}(p(y), Y) = \frac{1}{K}\sum_{i=1}^{K} -\log_2[(p_i(Y_i))/p_{ref_i}(Y_i)]. \tag{6}$$

The relative Ignorance of two forecast systems quantifies the information gain (in bits) the model forecast system provides over the reference system. In other words, Ignorance reflects the (average) increase in probability density that the model forecast placed on the outcome relative to that of the reference forecast. By convention, Ignorance is a negatively oriented score, which means the smaller (the more negative) the score the more skillful the forecast system.

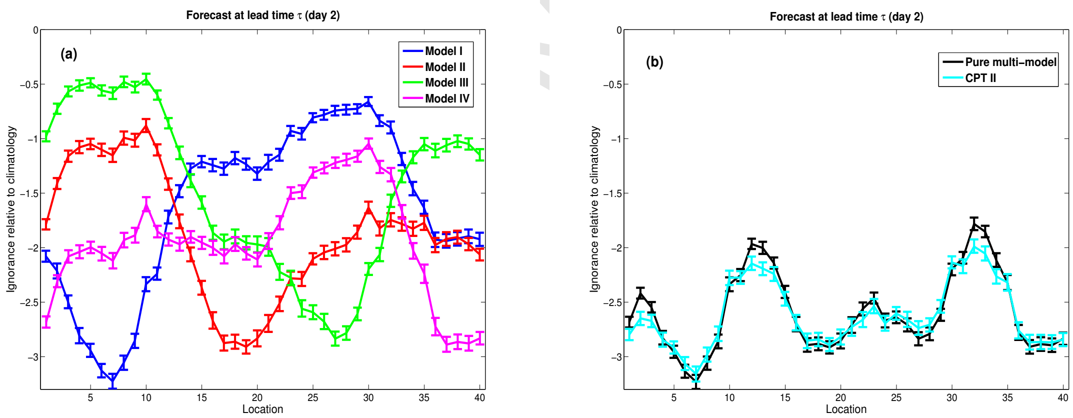

Figure 1: Ignorance score of forecasts as a function of location (model-state component) at lead-time τ = 0.4 time unit, a) forecasts from each individual model, b) pure multi-model forecast (Black) and CPT II forecast (Cyan).

Figure 1a, 2a and 3a shows the empirical Ignorance relative to climatology[13] for forecasts made by each model at different locations (model-state component) at lead-time τ, 2τ and 3τ. The empirical Ignorance is calculated based on 2048 forecast-outcome pairs and the climatological distribution is estimated using an independent 2048 historical observations.

[13] A climatological forecast (see [4]) based on the distribution of historical observations serves to define a zero-skill reference.

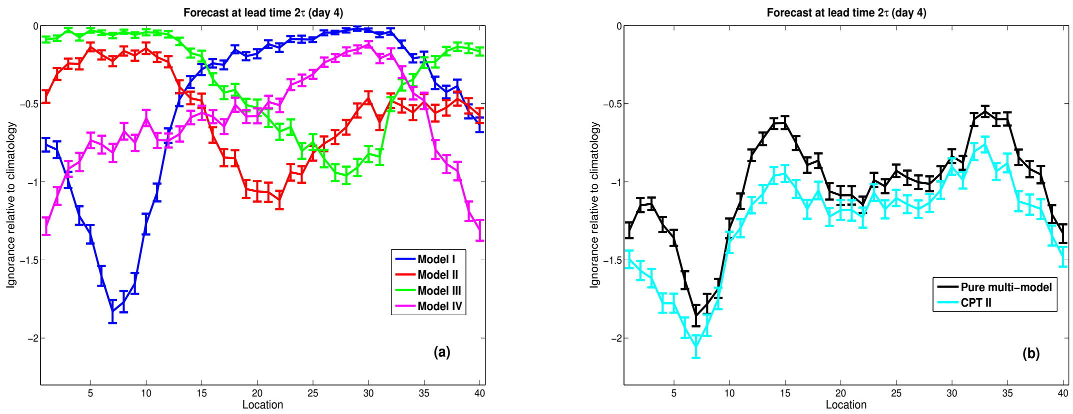

Figure 2: Ignorance score of forecasts as a function of location (model-state component) at lead-time $2\tau = 0.8$ time unit, a) forecasts from each individual model, b) pure multi-model forecast (Black) and CPT II forecast (Cyan). Note that, as expected, the improved skill under CPT II is greater at this longer lead-time.

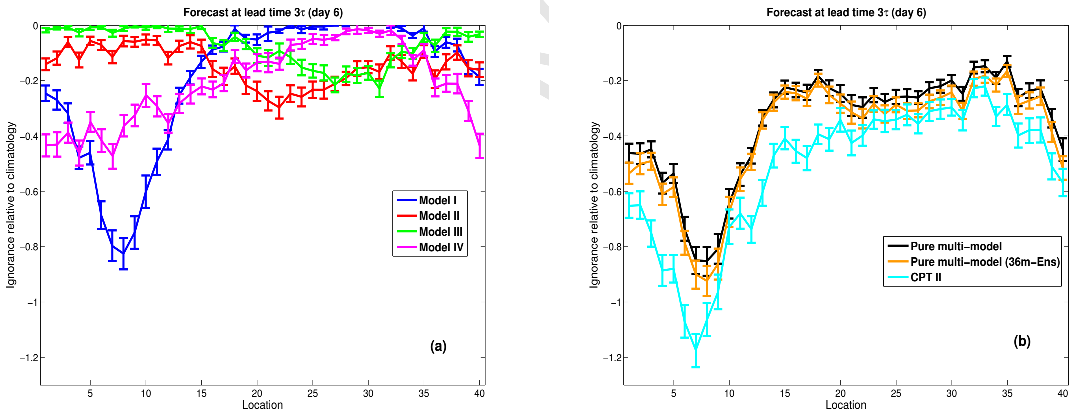

Figure 3: Ignorance score of forecasts as a function of location (model-state component) at lead-time $3\tau = 1.2$ time unit, a) forecasts from each individual model, b) pure multi-model forecast (Black), pure multi-model forecast with 36-member ensemble from each model (Brown) and CPT II forecast (Cyan).

The four imperfect models each produce a (9-member) ensemble of forecast trajectories, thus there are 36 forecast trajectories in total. To conduct cross-pollination, a simple pruning algorithm, based on local forecast performance, is adopted to maintain a manageable ensemble size.

Define the pruned ensemble observation trajectories to be $\{\mathbf{S}^\star(1), \ldots, \mathbf{S}^\star(9)\}$ where

$$\mathbf{S}^\star(j) \equiv \{\mathbf{s}^\star(t_0; j), \mathbf{s}^\star(t_0 + \Delta t; j), \ldots, \mathbf{s}^\star(t_0 + \tau; j)\}$$

and $\mathbf{s}^\star(t; j) \equiv \{y_1(t; j), y_2(t; j), \ldots, y_{40}(t; j)\} \in \mathbb{R}^{40}$. Assign the value of $x_i(t, B, j)$ to $y_i(t; j)$, where B is historically the local most informative model among (I, II, III, IV) (the one that produced lowest Ignorance forecasts at lead-time τ for model-state component location i in the training set). A probabilistic approach selecting the most informative model dynamically is easily implemented. The proposed pruning algorithm therefore prunes 36 forecast trajectories into a 9-member ensemble of observation trajectories.

To demonstrate the CPT II approach, the outputs from the pure model approach are compared with the results from the CPT II approach at lead-time τ, 2τ and 3τ. Note that both the outputs from pure model simulations and those from the CPT II approach at any lead-time t form a multi-model ensemble (the CPT II ensembles are obtained by pruning the model forecast trajectories and then conducting data assimilation). The multi-model ensemble is interpreted by combining the forecast distributions, generated from ensembles of individual model outputs to produce a single probabilistic multi-model forecast distribution for evaluation. A linearly weighted approach is adopted to combine the single-model forecast distribution (the model weights are determined using the training set, for a full description see [26], and Appendix C). All evaluation is out-of-sample.

Figure 1b, 2b and 3b compares the probabilistic forecasts from the pure multi-model outputs (Black) with those from CPT II (Cyan) at lead-time τ, 2τ and 3τ. At lead-time τ, the dynamical information from each of the individual trajectory are combined and embedded into the forecasts, which are also the initial states of the forecasts for the next cross-pollination period. Such additional information in the initial conditions reveals its value at the next forecast period, where significant[14] improvement in skill is shown at lead-time 2τ and 3τ. Note in Figure 3b that the CPT II forecast also significantly outperforms the pure multi-model forecast (Brown) based on an ensemble four times larger (this allows a comparison when each forecast is based on a 36-member ensemble).

Is CPT II also an improvement on CPT I? Figure 4 shows the forecast skill of CPT II (Cyan) and CPT I (Purple) relative to the probabilistic forecasts from the pure multi-model outputs at

[14]The resampling bars in each figure represent $10\% - 90\%$ bootstrap resampling intervals.

lead-time 2τ and 3τ. Both CPT II and CPT I forecasts improve significantly the forecast skill over that of the pure multi-model forecast, while CPT II places an additional 5% more probability mass (a difference of ~ 0.065 bits in expected Ignorance) on the outcomes than CPT I.

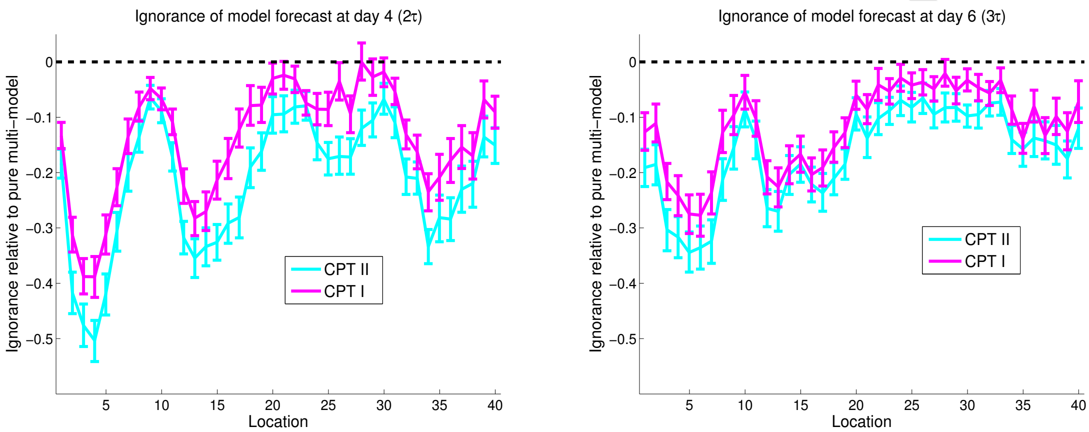

Figure 4: Ignorance score of forecasts from CPT II (Cyan) and CPT I (Purple) relative to pure multi-model forecast (Black dashed zero line) as a function of location (model-state component) at lead-time a) $2\tau = 0.8$ time unit and b) $3\tau = 1.2$ time unit.

CPT II successfully exploits the sophisticated aspects of the PDA data assimilation scheme [6, 7] to allow selective inclusion of locations (state space components) in the forecast simulation of each model. Doing so allows the CPT forecast system to generate forecasts with support beyond that of any single-model forecast systems and also beyond any traditional multi-model forecast systems. As demonstrated above, CPT II can increase forecast skill.

6 Conclusion

Suppose, for a moment, one has two models which simulate supply and demand, given the current supply and demand. Model A produces significantly better forecasts of supply, while Model B yields significantly better forecasts of demand. The traditional multi-model approach is to consider an ensemble of simulation under Model A and a second, independent, ensemble of simulations under model B. In this case, the specific model inadequacy of each model will result in a decay in the relevance of the probabilistic forecasts with lead-time. Cross-Pollination in Time II aims to forestall this decay, extending the lead-time on which forecasts are informative, by assimilating forecasts of the near-term future to generate an enhanced ensemble of forecasts in the medium range, and iterating the process into the long range. In this simple case, taking the "supply" forecast from Model A and the "demand" forecast from Model B would produce a new initial condition to be folded into the forecast ensembles of each model at lead-time one, and propagated into the future, improving forecast skill while using only the models already in hand.

By contrast, the state of a modern weather model consists of (more than) ten million components, and no two operational models need actually share the same model-state space. As implemented above, CPT II overcomes this challenge by using a data assimilation algorithm designed to start with a pseudo-orbit, and an initial pseudo-orbit extracted from the initial simulation trajectory of each model. Here, CPT II yields probabilistic forecasts more skillful than traditional approaches, even when the ensemble size of the traditional approach is increased by a factor of four (the number of models used above). As noted above, challenges remain in deploying and interpreting CPT II forecast systems; this concrete example of success is intended to motivate exploration of more realistic cases.

APPENDIX

A Pseudo-orbit Data Assimilation

A brief description of the PDA approach is given in the following paragraphs (more details can be found in Du and Smith [6, 7]). Let the dimension of the model-state space be m and the number of observation times in the assimilation window be n (for experiments presented in this paper, $n = \frac{\tau}{\Delta t}$). A *pseudo-orbit*, $\mathbf{U} \equiv \{\mathbf{u}_{-n+1}, ..., \mathbf{u}_{-1}, \mathbf{u}_0\}$, is a point in the $m \times n$ dimensional sequence space for which $\mathbf{u}_{t+1} \neq G(\mathbf{u}_t)$ for any component of $\mathbf{U}$. Define the component of the *mismatch* of a pseudo-orbit $\mathbf{U}$ at time t to be $\mathbf{e}_t = \mid G(\mathbf{u}_t) - \mathbf{u}_{t+1} \mid$, $t = -n+1, ..., -1$ and take the mismatch cost function to be

$$C(\mathbf{U}) = \sum \mathbf{e}_t^2 \tag{7}$$

The *Pseudo-orbit Data Assimilation* minimizes the mismatch cost function for $\mathbf{U}$ in the $m \times n$ dimensional sequence space. In this paper a gradient descent (GD) minimization algorithm is used. For CPT II, such minimization is initialized with the combined model output of sequence states in the observation space, $\mathbf{S}^\star(j)$ in Equation (1) and (2). An important advantage of PDA is that the minimization is done in the sequence space: information from across the entire assimilation window is used simultaneously.

The pseudo-orbit $\mathbf{U}$ is updated on every iteration of the GD minimization. Call these steps along GD minimization ${}^{\alpha}\mathbf{U}$, where α indicates algorithmic time in GD. Due to model imperfection, the minimization is applied with a stopping criteria in order to obtain more consistent pseudo-orbits [7]; this is because the mismatches will reflect the point-wise model error when the model is imperfect. In the experiments presented in this paper the stopping criteria targeted forecast performance at lead-time τ, 2τ and 3τ.

B Ensemble Interpretation

An ensemble of simulations is transformed into a probabilistic distribution function by a combination of kernel dressing and blending with climatological distribution (see [4]). Denote an N-member ensemble at time t as $X_t = [x_t^1, ..., x_t^N]$, where x_t^i is the ith ensemble member. For simplicity, all ensemble members under a model are treated as exchangeable. Kernel dressing defines the model-based component of the density as:

$$p(y : X, \sigma) = \frac{1}{N\sigma} \sum_{i}^{N} K\left(\frac{y - (x^i - \mu)}{\sigma}\right), \qquad (8)$$

where y is a random variable corresponding to the density function p and K is the kernel, taken here to be

$$K(\zeta) = \frac{1}{\sqrt{2\pi}} exp(-\frac{1}{2}\zeta^2). \qquad (9)$$

Thus each ensemble member contributes a Gaussian kernel centred at $x^i - \mu$. Here μ is an offset, which accounts for any systematic "bias". For a Gaussian kernel, the kernel width σ is simply the standard deviation.

For any finite ensemble, there remains the chance of $\sim \frac{2}{N}$ that the outcome lies outside the range of the ensemble even when the outcome is selected from the same distribution as the ensemble itself. Given the nonlinearity of the model, such outcomes can be very far outside the range of the ensemble members. Not only is N finite in practice, of course, but also the simulations are not drawn from the same distribution as the outcome; the ensemble simulation system is not perfect. To improve the out-of-sample skill of the probabilistic forecasts, the kernel dressed ensemble is blended with an estimate of the climatological distribution of the system (see [4] for more details, see [21] for alternative kernels and see [18] for a Bayesian approach). The blended forecast distribution is then written as

$$p(\cdot) = \alpha p_m(\cdot) + (1 - \alpha)p_c(\cdot), \tag{10}$$

where p_m is the density function generated by dressing the model ensemble as in Equation (8) above, and p_c is the estimate of climatological density. The blending parameter α determines how much weight is placed in the model. Specifying the three values (the offset μ, the kernel width σ, and the blended parameter α) defines the forecast distribution. These parameters are fitted simultaneously for each forecast system by optimizing the empirical Ignorance score in the training set.

C Weighting Multi-model Ensemble

There are many ways to combine forecast distributions generated from ensembles of individual model simulations in order to produce a single, probabilistic, multi-model forecast distribution. One approach is to assign equal weight to each model and simply sum the equally weighted distributions generated from each model to obtain a single probabilistic distribution (see [10]). In general, different forecast models do not provide equal amounts of information; more skillful forecasts are obtained by weighting the models according to some measure of past performance, see, for example, [5, 19]. The combined multi-model forecast is the weighted linear sum of the constituent distributions,

$$p_{mm} = \sum_i \omega_i p_i, \tag{11}$$

where the p_i is the forecast distribution from model i and ω_i its weight, with $\sum_i \omega_i = 1$. The weighting parameters may be chosen by minimizing the Ignorance score, for example; although fitting ω_i in this way can be costly, other approaches are typically complicated by different models sharing information. The weights of individual models are, of course, expected to vary as a function of lead-time.

To avoid poorly determined weights, a simple iterative method to combine models is adopted avoiding any attempt to determine all the weights simultaneously. For each lead-time, the best (in terms of Ignorance) forecast system is first combined with the second-best forecast system to form a combined forecast distribution (by assigning weights to both models). The combined forecast distribution is then combined with the third-best forecast system to update the combined forecast distribution. Repeat this process until the least skillful model has been considered.

Acknowledgment

This research was supported by the LSE's Grantham Research Institute on Climate Change and the Environment and the ESRC Centre for Climate Change Economics and Policy, funded by the Economic and Social Research Council and Munich Re; it was supported by the EPSRC-funded Blue Green Cities (EP/K013661/1). Additional support for H.D. was also provided by the National Science Foundation Award No. 1463644 "DMUU: Center for Robust Decision Making on Climate and Energy Policy (RDCEP)". L.A.S. gratefully acknowledges the continuing support of Pembroke College, Oxford.

References

[1] J. M. Bernardo. Expected information as expected utility. *Ann. Stat.*, 7:686690, 1979.

[2] C. H. Bishop, B. J. Etherton, and S. J. Majumdar. Adaptive Sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical Aspects. *Mon. Wea. Rev.*, 129(3):420–436, 2001.

[3] J. Brocker and L.A. Smith. Scoring probabilistic forecasts: On the importance of being proper. *Wea. Forecasting*, 22:382–388, 2007.

[4] J. Brocker and L.A. Smith. From ensemble forecasts to predictive distribution functions. *Tellus*, 60:663–678, 2008.

[5] F. J. Doblas-Reyes, R. Hagedorn, and T. N. Palmer. The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: Calibration and combination. *Tellus A*, 57:234, 2005.

[6] H. Du and L.A. Smith. Pseudo-orbit data assimilation part I: the perfect model scenario. *Journal of the Atmospheric Sciences*, 71(2):469–482, 2014.

[7] H. Du and L.A. Smith. Pseudo-orbit data assimilation part II: assimilation with imperfect models. *Journal of the Atmospheric Sciences*, 71(2):483–495, 2014.

[8] Gregory S. Duane. Synchronicity from synchronized chaos. *Entropy*, 17(4):1701, 2015.

[9] I. J. Good. Rational decisions. *Journal of the Royal Statistical Society*, XIV(1), 1952.

[10] R. Hagedorn, F. J. Doblas-Reyes, and T. N. Palmer. The rationale behind the success of multi-model ensembles in seasonal forecasting. Part I: Basic concept. *Tellus A*, 57:219–233, 2005.

[11] R. Hagedorn and L. A. Smith. Communicating the value of probabilistic forecasts with weather roulette. *Meteor. Appl.*, 16:143155, 2009.

[12] J.A. Hansen and L.A. Smith. Probabilistic noise reduction. *Tellus A*, 53:585598, 2001.

[13] M. S. J. Harrison, T. N. Palmer, D. S. Richardson, R. Buizza, and T. Petroliagis. Joint ensembles from the UKMO and ECMWF models. In *ECMWF Seminar Procedings: Predictability*, volume 2, page 61120, ECMWF, Reading, UK, 1995.

[14] K. Judd and L.A. Smith. Indistinguishable states ii: The imperfect model scenario. *Physica D*, 196:224242, 2004.

[15] B. P. Kirtman, D. Min, J. M. Infanti, J. L. Kinter, D. A. Paolino, Q. Zhang, H. van den Dool, S. Saha, M. P. Mendez, E. Becker, P. Peng, P. Tripp, J. Huang, D. G. DeWitt, M. K. Tippett, A. G. Barnston, S. Li, A. Rosati, S. D. Schubert, M. Rienecker, M. Suarez, Z. E. Li, J. Marshak, Y. Lim, J. Tribbia, K. Pegion, W. J. Merryfield, B. Denis, and E. F. Wood. The North American Multimodel Ensemble: Phase-1 Seasonal-to-Interannual Prediction; Phase-2 toward Developing Intraseasonal Prediction. *Bull. Amer. Meteor. Soc.*, 95(4):585–601, August 2013.

[16] E.N. Lorenz. Predictability: a problem partly solved. In *Seminar on Predictability, 4-8 September 1995*, volume 1, pages 1–18, Shinfield Park, Reading, 1995. ECMWF, ECMWF.

[17] T. N. Palmer, F. J. Doblas-Reyes, R. Hagedorn, A. Alessandri, S. Gualdi, U. Andersen, H. Feddersen, P. Cantelaube, J. M. Terres, M. Davey, R. Graham, P. Délécluse, A. Lazar, M. Déqué, J. F. Guérémy, E. Díez, B. Orfila, M. Hoshen, A. P. Morse, N. Keenlyside, M. Latif, E. Maisonnave, P. Rogel, V. Marletto, and M. C. Thomson. Development of a european multimodel ensemble system for seasonal-to-interannual prediction (demeter). *Bull. Amer. Meteor. Soc.*, 85(6):853–872, 2004.

[18] A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, 133:11551174, 2005.

[19] B. Rajagopalan, U. Lall, and S. E. Zebiak. Categorical climate forecasts through regularization and optimal combination of multiple gcm ensembles. *Mon. Wea. Rev.*, 130:17921811, 2002.

[20] M. S. Roulston and L. A. Smith. Evaluating probabilistic forecasts using information theory. *Mon. Wea. Rev.*, 130:16531660, 2002.

[21] M. S. Roulston and L. A. Smith. Combining dynamical and statistical ensembles. *Tellus*, 55:16–30, 2003.

[22] Mao-Lin Shen, Noel Keenlyside, Frank Selten, Wim Wiegerinck, and Gregory S. Duane. Dynamically combining climate models to supermodel the tropical pacific. *Geophysical Research Letters*, 43(1):359–366, 2016. 2015GL066562.

[23] L. A. Smith. Disentangling uncertainty and error: On the predictability of nonlinear systems. In Alistair I. Mees, editor, *Nonlinear Dynamics and Statistics*, chapter 2, pages 31–64. Birkhäuser Boston, 2000.

[24] L. A. Smith, Hailiang Du, Emma B. Suckling, and Falk Niehrster. Probabilistic skill in ensemble seasonal forecasts. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1085–1100, 2015.

[25] L.A. Smith. What might we learn from climate forecasts? In *Proc. National Acad. Sci.*, volume 4, pages 2487–2492, USA, 2002.

[26] L.A. Smith, S. Higgins, and H. Du. On the design and use of ensembles of multi-model simulations for forecasting. *In preparation for Nonlinear Processes in Geophysics*, 2016.

[27] L.A. Smith, E.B. Suckling, E.L. Thompson, T. Maynard, and H. Du. Towards improving the framework for probabilistic forecast evaluation. *Climatic Change*, 2015.

[28] D.B. Stephenson, C. Coelho, F. Doblas-Reyes, and M. Alonso Balmaseda. Forecast assimilation: a unified framework for the combination of multimodel weather and climate predictions. *Tellus A*, 57:253–264, 2005.

[29] L. A. van den Berge, F. M. Selten, W. Wiegerinck, and G. S. Duane. A multi-model ensemble method that combines imperfect models through learning. *Earth System Dynamics*, 2(1):161–177, 2011.

[30] B. Wang, J. Lee, I. Kang, J. Shukla, C. K. Park, A. Kumar, J. Schemm, S. Cocke, J. S. Kug, J. J. Luo, T. Zhou, B. Wang, X. Fu, W. T. Yun, O. Alves, E. Jin, J. Kinter, B. Kirtman, T. Krishnamurti, N. Lau, W. Lau, P. Liu, P. Pegion, T. Rosati, S. Schubert, W. Stern, M. Suarez, and T. Yamagata. Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dynamics*, 33(1):93–117, 2009.

[31] A. Weisheimer, F. J. Doblas-Reyes, T. N. Palmer, A. Alessandri, A. Arribas, M. Déqué, N. Keenlyside, M. MacVean, A. Navarra, and P. Rogel. ENSEMBLES - a new multi-model ensemble for seasonal-to-annual predictions: Skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophysical Research Letters*, 36(21), 2009.

[32] D. S. Wilks. Comparison of ensemble–MOS methods in the Lorenz'96 setting. *Meteorological Applications*, 13:243–256, 2006.

[33] D. S. Wilks and T. M. Hamill. Comparison of Ensemble-MOS Methods Using GFS Reforecasts. *Mon. Wea. Rev.*, 135(6):2379, 2007.

[34] D.S. Wilks. *Statistical Methods in the Atmospheric Sciences.* Academic Press, second edition, 2005.

1. A multi-model ensemble scheme for integrating dynamical information is proposed.

2. Truly multi-model trajectories at future time are obtained via data assimilation.

3. The proposed approach yields more skillful probabilistic forecasts.