University of Reading

# *Linking the anomaly initialization approach to the mapping paradigm: a proof-of-concept study*

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work.  See Guidance on citing.

the End User Agreement.

# www.reading.ac.uk/centaur

**CentAUR**

Central Archive at the University of Reading

Reading's research outputs online

SPECIAL

6th WMO Data Assimilation Symposium

COLLECTION

# Linking the Anomaly Initialization Approach to the Mapping Paradigm: A Proof-of-Concept Study

ROBIN J. T. WEBER

*Institut Catalá de Ciéncies del Clima (IC3), Barcelona, Spain*

ALBERTO CARRASSI

*Nansen Environmental and Remote Sensing Center, Bergen, Norway, and Institut Catalá de Ciéncies del Clima (IC3), Barcelona, Spain*

FRANCISCO J. DOBLAS-REYES

*Institut Catalá de Ciéncies del Clima (IC3), and Institució Catalana de Recerca i Estudis Avançats, and Barcelona Supercomputing Center-Centro Nacional de Supercomputación, Barcelona, Spain*

ABSTRACT

Seasonal-to-decadal predictions are initialized using observations of the present climatic state in full field initialization (FFI). Such model integrations undergo a drift toward the model attractor due to model deficiencies that incur a bias in the model. The anomaly initialization (AI) approach reduces the drift by adding an estimate of the bias onto the observations at the expense of a larger initial error.

In this study FFI is associated with the fidelity paradigm, and AI is associated with an instance of the mapping paradigm, in which the initial conditions are mapped onto the imperfect model attractor by adding a fixed error term; the mapped state on the model attractor should correspond to the nature state. Two diagnosis tools assess how well AI conforms to its own paradigm under various circumstances of model error: the degree of approximation of the model attractor is measured by calculating the overlap of the AI initial conditions PDF with the model PDF; and the sensitivity to random error in the initial conditions reveals how well the selected initial conditions on the model attractor correspond to the nature states. As a useful reference, the initial conditions of FFI are subjected to the same analysis.

Conducting hindcast experiments using a hierarchy of low-order coupled climate models, it is shown that the initial conditions generated using AI approximate the model attractor only under certain conditions: differences in higher-than-first-order moments between the model and nature PDFs must be negligible. Where such conditions fail, FFI is likely to perform better.

## 1. Introduction

A geophysical prediction is made by integrating the model in time from its initial condition. The quality of the forecast will rely on the quality of the initial condition, and the quality of the model, given by the implementation of the set of equations describing nature. Predictions evaluated beyond the deterministic limit of two weeks typical of numerical weather prediction are useful mainly when considering the statistical properties of the natural system over forecast ranges of several weeks, months, or perhaps years. Where the accuracy of numerical weather predictions are determined by error in the initial conditions, centennial climate projection evaluations are determined by boundary conditions such as atmospheric greenhouse gas concentrations, while the signature of the initial condition is lost (Hawkins and Sutton 2009). Seasonal-to-decadal prediction spans time horizons of up to approximately 10 years, falling between numerical weather prediction and centurial projections. Correct initialization of the model is known to improve forecast

*Corresponding author address*: Robin J. T. Weber, Climate Forecasting Unit, Institut Catalá de Ciéncies del Clima (IC3), Doctor Trueta 203 3a Planta, Barcelona 08005, Spain.
E-mail: robin.weber@ic3.cat

quality on horizons of several years (Doblas-Reyes et al. 2013). Such predictions require Earth system models that include coupled feedbacks between the atmosphere and the bio-, hydro-, and cryospheres across all time scales, and that incorporate greenhouse gases. The largest uncertainty on seasonal-to-decadal time scales is attributed to model error (Hawkins and Sutton 2009).

Models are affected by errors due to incorrect representation of dynamics, its projection onto a finite grid, and unresolved scales insufficiently represented through parameterizations. These errors affect the long-term statistical properties of the system (i.e., its climate). The result is a "model climate" different from the natural climate (e.g., in its variability and time mean). Initial conditions obtained directly from observations of the climate system often do not represent a feasible model state, so that forecasts drift toward the model's own equilibrium. Drift is defined as the time evolution of the prediction error averaged over the initial conditions. Initialization shocks can be seen as specific cases of drift in which the imbalance of initial conditions leads to a rapid short-term readjustment of the system.

Full field initialization (FFI) makes use of the best possible available estimate of the real state. In this way FFI reduces the initial error, but the unavoidable presence of model deficiencies causes the model trajectory to drift away from the observations regardless of initial error size (e.g., Stockdale 1997). Anomaly initialization (AI) assimilates, in place of the observations, the observed climate anomalies on top of an estimate of the model mean climate. This initial state, at the expense of an initial error of the size of the model bias, is expected to be closer to the model's own attractor (e.g., Smith et al. 2007), so that drift is reduced.

Comprehensive comparisons between FFI and AI using Earth system models for seasonal-to-multiyear time horizons have recently appeared (Magnusson et al. 2013; Smith et al. 2013; Hazeleger et al. 2013). These studies represent a first attempt to assess their respective performance using exactly the same observational and model setup and are therefore of central importance in guiding future development of initialized climate prediction systems. Results have been far from conclusive, varying regionally and among models. Carrassi et al. (2014) introduced a formalism based on data assimilation notation that helps identify universal error characteristics of either scheme, and illustrated the error features of FFI and AI using a low-order climate model.

Anomaly initialization has been devised by the climate prediction community in an ad hoc fashion to deal with problems related to drift (e.g., by allowing for more statistically robust bias correction techniques). Many scientists working in the community also dislike the idea of an unbalanced initial state, observing how the model readjusts toward equilibrium over the course of the prediction. The initial condition after AI is both close to the mean state of the model, and includes observational information. However, conclusive evidence associating AI with improved forecast quality is lacking.

In a pioneering study, Toth and Peña (2007) have introduced a conceptual framework that differentiates between initialization schemes according to two competing paradigms. The "fidelity" paradigm reduces the initial error of the system by starting the forecast as close to nature as possible. Its name derives itself from the goal of adhering to the truth. It represents the traditional approach of trusting the observational information as much as possible, and assumes that the model is perfect. The "mapping" paradigm recognizes that models are imperfect, and links corresponding states of nature on the model attractor by means of a mapping vector. A corresponding state on the model attractor imitates the natural state to the model's best ability, without leaving the attractor of the model. A corresponding state can be interpreted as an "image" of nature on the imperfect model attractor. A mapping vector will have identified the images corresponding to their natural states, as far as they exist. Such images can represent "better" initial conditions for an imperfect model in terms of improved forecast skill (Toth and Peña 2007). Using this framework, we associate FFI and AI with practical realizations of the fidelity and mapping paradigms, respectively. The bias term added onto the observations in AI can be seen as a state-independent mapping vector. This interpretation allows for important inferences and becomes useful in understanding the merits and limitations of AI.

This study assesses the performance of AI as a function of its ability to conform to the mapping paradigm objective. We derive two simple indicators of the quality of a mapping scheme: the degree to which the initial conditions approximate the model attractor, and the sensitivity of the forecast skill toward random errors in the initial conditions. Using a hierarchy of low-order dynamical systems of increasing realism, the forecast quality of AI with respect to both indicators is assessed. We simulate model error through parameter misspecification, and evaluate, with respect to a target trajectory, initialized predictions using imperfect configurations. The degree of approximation of the model attractor is assessed by calculating the overlap of the AI initial conditions and model probability distribution function (PDF) using the Bhattacharyya coefficient. As a useful reference, we carry out the same analysis for FFI and compare the performance.

The paper is organized as follows. Section 2 formalizes AI and FFI. Section 3 connects AI to the mapping paradigm as introduced by Toth and Peña (2007). Section 4 describes the experimental setup, along with the hierarchy of low-order models. Results and the conclusions are given in sections 5 and 6, respectively.

## 2. Initialization methods

We assume that nature can be formally expressed as an autonomous dynamical system $\mathbf{F}$ so that

$$\frac{d\mathbf{x}^{\text{nat}}}{dt} = \mathbf{F}(\mathbf{x}^{\text{nat}}), \tag{1}$$

where $\mathbf{x}^{\text{nat}} \in \mathbb{X}_N$ is the unknown nature state of dimension $I$ on the nature attractor represented by the set $\mathbb{X}_N$; index $N$ stands for nature. We assume that predictions targeting nature are made using an imperfect model $\mathbf{G} \neq \mathbf{F}$:

$$\frac{d\mathbf{x}^m}{dt} = \mathbf{G}(\mathbf{x}^m), \tag{2}$$

where $\mathbf{x}^m \in \mathbb{X}_M$ is the model state, supposed to have the same dimension $I$ as the unknown nature state, on the model attractor represented by the set $\mathbb{X}_M$; index $M$ stands for model. Observations $\mathbf{y}_i = \mathbf{y}(t_i)$ are assumed to be available at equally spaced times $t_i = i$, $i = 0, 1, \ldots$ In real applications the vector of observations has a much smaller dimension $O$ than the model state vector ($O \ll I$). The observations have an observational error $\boldsymbol{\epsilon}^o$ that is assumed to be Gaussian and white in time, so that we can write

$$\mathbf{y} = H(\mathbf{x}^{\text{nat}}) + \boldsymbol{\epsilon}^o. \tag{3}$$

where $H$ is the observation operator mapping from nature to the phase space of the observations; $\boldsymbol{\epsilon}^o$ accounts for both the instrumental and the representativity error connected to the specification of the operator $H$ (Janjić and Cohn 2006).

Where observations are not available, the initialization procedures fall back on information of the model state obtained after a long transient spinup run also known as the control. As in Carrassi et al. (2014) we interpret the control as a background field that contains information about the real system (nature) prior to the assimilation of the observation; the control will be indicated as $\mathbf{x}^b$. We emphasize that $\mathbf{x}^b$ is not obtained from a short forecast as is the practice in numerical weather prediction, because initialization of climate predictions cannot fall back on such information. The initial condition, obtained after the initialization step that incorporates the observational

information, is interpreted as the analysis field, indicated as $\mathbf{x}^a$. The initialization step in FFI can be written as

$$\mathbf{x}^a_{\text{FFI}} = \mathbf{x}^b + \mathbf{H}^{\text{T}}[\mathbf{y} - \mathbf{H}\mathbf{x}^b], \tag{4}$$

with $\mathbf{H}^{\text{T}}$ being the transpose of the linear observation operator. We have made the additional assumption of a linear observation operator. This implies that the observation operator is given as an $O \times I$ matrix, $\mathbf{H}$.

Note that (4) does not include a Kalman gainlike operator as is typical for data assimilation update equations (Kalnay 2002). This is because in standard FFI the observations are assimilated as if they were perfect (i.e., the model state is substituted by the observational values). FFI does not use a criteria of optimality, such as the least squares method at the basis of Kalman filtering (Kalman 1960) appearing in the form of a gain matrix. More sophisticated initialization procedures based on the data assimilation approach might account for the relative accuracy of model and observations and give weight to either accordingly (see Smith and Murphy 2007; Carrassi et al. 2014).

In AI (Smith et al. 2007) the model state at initial time is replaced by the observed anomaly assimilated on top of the model climatology (i.e., the bias is added onto the observations). To obtain the AI equation, replace $\mathbf{y}$ in (4) with

$$\mathbf{y}^{\text{AI}} = \mathbf{y} - (\overline{\mathbf{y}} - \mathbf{H}\overline{\mathbf{x}^b}), \tag{5}$$

where the overbars indicate a long time average over past observational/model statistics. Then,

$$\mathbf{x}^a_{\text{AI}} = \mathbf{x}^b + \mathbf{H}^{\text{T}}[\mathbf{y}^{\text{AI}} - \mathbf{H}\mathbf{x}^b]. \tag{6}$$

In the perfect model case $\boldsymbol{\epsilon}^{\text{bias}} = \mathbf{H}\overline{\mathbf{x}^b} - \overline{\mathbf{y}} = 0$ and FFI and AI [(4) and (6)] coincide. The fact that the model bias $\boldsymbol{\epsilon}^{\text{bias}}$ can only be evaluated on the basis of a finite set of observations introduces an additional source of error in AI, proportional to the mismatch between the estimated and the actual bias, $\boldsymbol{\epsilon}^{\text{bias}} = \mathbf{H}\overline{\mathbf{x}^b} - \overline{\mathbf{y}} \neq \mathbf{H}\overline{\mathbf{x}^b} - \overline{\mathbf{x}^{\text{nat}}}$.

How the accuracy of the initial conditions will impact the prediction skill at seasonal-to-decadal time scales is unclear. Weather forecast practice with a horizon of two weeks often neglects model error and prioritizes the control of chaotic error growth (e.g., Palatella et al. 2013). In contrast, in seasonal-to-decadal prediction the bias caused by model deficiencies is more important. Forecasts initialized close to the observed state as in FFI will drift toward the model climate following a nonlinear state-dependent evolution, which can appear as an initial dynamical shock caused by the displacement of the model state onto the observed values lying outside the

model attractor. At the expense of larger initial errors, the objective of AI is to keep the initial state close to the model attractor and reduce the drift. The mean forecast error is less dependent on lead time and, as argued by Magnusson et al. (2013), the use of standard a posteriori bias correction techniques is more robust. Anomaly initialization can reduce initialization shocks, but is unable to avoid shocks caused by inconsistencies (e.g., geographical mismatches) between the model climate and observed anomalies (Magnusson et al. 2013).

## 3. Anomaly initialization as a specific case of the mapping paradigm

### a. Linking full field/anomaly initialization to the fidelity/mapping paradigms

We interpret FFI and AI using the general framework introduced by Toth and Peña (2007). They have proposed categorizing initialization schemes as belonging to the traditional fidelity paradigm or the mapping paradigm. The fidelity paradigm focuses on initializing the model as close to nature as possible. A premise of traditional thinking is that forecast errors are reduced by bringing the initial conditions closer to nature. The quality of the initial conditions is particularly important in chaotic systems like the atmosphere, because initial errors contribute to a loss of predictability. Given the degree of instability of the dynamics, the actual forecast time at which this loss of predictability occurs is determined by the size and structure of initial errors. The conventional paradigm thus consists of estimating the state of nature as precisely as possible (i.e., the analysis) and running numerical model forecasts from the analysis field. The systematic error in the numerical forecasts—which is state dependent—is assessed and removed in postprocessing.

We associate FFI with the fidelity paradigm because of its strict adherence to the observations. The analysis field is obtained according to (4). Systematic error is assessed by initializing many so-called hindcast experiments from past start dates and comparing such hindcasts with available reanalysis fields. The systematic error is removed a posteriori according to both the forecast start date within the year and the forecast horizon.

Toth and Peña (2007) recognize that the fidelity paradigm is justified only in the special case when perfect models are used for prediction, becoming progressively worse as models deviate from nature in practical applications. States obtained via the fidelity paradigm lead to a transient drift as they are not on the attractor of an imperfect model, which would not occur had an appropriate initial state on (or near) the model attractor been chosen.

The mapping paradigm explicitly recognizes the systematic difference between corresponding states of nature and a numerical model of it. It "maps" an initial state from nature close to the attractor of a similar system, which is a numerical model of nature. Mapping is thus defined as an operation linking each natural state to a state near the model attractor; such a corresponding state on the model attractor can be interpreted as an image of nature. Initialization using the mapped state (or image) leads to a model forecast that best reproduces the time evolution of nature. Before evaluation the forecast is remapped back to the nature attractor. The emphasis is laid on a "shadowing" (Toth and Peña 2007) of the nature trajectory within the numerical model. For example, initializing the model close to the observations and outside the model attractor can lead to a quasi-systematic excitation of model climate dynamics (i.e., the drift is projected onto the main climate internal modes, which is dissociated from the actual nature evolution intended to be captured). This is observed in experiments with Earth system models in which initialization of the Pacific Ocean nudged to reanalysis leads to an artificial sequence of El Niño/La Niña events in the model during the first 4 years of integration (Sanchez-Gomez et al. 2015). In such a case, a mapped state would avoid an excitation of the model dynamics and remapped forecasts would reproduce (i.e., "shadow") the timing and structure of natural anomalies better. The proximity of the mapped initial conditions to the model attractor reduces the drift and associated forecast errors.

The mapping paradigm starts with an estimation of the mapping vector $\mathbf{d}_j$ defined as the actual difference between nature and its numerical model at time $t_j$, so that $\mathbf{d}_j = \mathbf{x}_k^m - \mathbf{x}_j^{\mathrm{nat}}$, where $\mathbf{x}_k^m$ is the corresponding model state to (i.e., the image of) $\mathbf{x}_j^{\mathrm{nat}}$; $\mathbf{d}_j$ is, of course, unknown, and predictions with the mapping paradigm will need the best possible estimate of $\mathbf{d}_j$ at the initial time and the best possible estimate of the remapping vector at the end of the prediction. After an assimilation step in which the mapped observations are merged with the model, the model is run from the mapped initial condition.

We attribute AI, as defined in (6), to the mapping paradigm, where the model bias in (5) plays the role of the mapping vector. Thus, AI becomes a specific case of a mapping scheme in which the mapping vector is a constant (given by the estimated bias) applied to all initial conditions indiscriminately. The bias correction at forecast evaluation can be seen as a remapping. Anomaly initialization as defined in (6) is qualitatively equivalent to the climate mean mapping described by Toth and Peña (2007) if the model bias is calculated based on the entire set of observations. In an Earth System model, the bias term is calculated for each forecast month of the year and

for each start date of the year, taking into account the seasonal cycle of the bias. Thus, in AI the mapping vector is dependent on the month in which the forecast is issued, retaining some dependence of state.

Attributing AI to the mapping paradigm should not suggest that predictions initialized with AI do not drift. Anomaly initialization is far from an ideal mapping scheme able to identify the image of nature on the model attractor, and is subject to transient dynamics for reasons described in the following section (section 3b).

Finally, a common practice in hindcast experiments causes the implemented AI scheme to defer from the strict mapping procedure: the fields are first conventionally analyzed and then mapped onto the model, reversing the order of analysis and mapping.

### b. Anomaly initialization: A state-independent mapping scheme

In this study we focus on a core limitation of AI in pursuit of the mapping paradigm objective: the actual difference $\mathbf{d}_j$ between the model and nature is assumed to be state independent (i.e., $\mathbf{d}_j = \mathbf{d} = \boldsymbol{\epsilon}^{\text{bias}}$). Anomaly initialization resembles a *first-order* (i.e., linear) *mapping* of the initial conditions. It implicitly assumes that higher-order differences in the model and nature PDFs are negligible. We highlight here the circumstances under which such a mapping is justified, as well as its limitations, with the help of an ideal experiment.

A perfect forecast is achievable when both the model and the initial conditions are perfect. A perfect forecast using FFI would thus necessitate a perfect model, full observational coverage, and no observational error. A perfect forecast using AI would impose a slightly weaker requirement on the model: *under the central condition that all else remains identical*, the model can differ from nature in its mean climate.

Figure 1 is a schematic of the linear mapping principle behind AI using the attractor of the three-variable Lorenz model (Lorenz 1963). The pink line depicts the phase space evolution of the (defined) "nature" state, the green line depicts the evolution of an imperfect model of the nature, the black crisscross represents a perfect "observation" $\mathbf{y} = \mathbf{x}_{t=0}^{\text{nat}}$, and the red crisscross represents the AI initial condition $\mathbf{x}_{\text{ic}}^{\text{AI}}$ on the model attractor after application of the mapping vector $\boldsymbol{\epsilon}^{\text{bias}}$. The short black trajectory represents the evolution of the nature following the observation; the short red trajectory represents the model integration after initialization. Here, the nature differs from the model by the addition of a constant vector $\mathbf{d} = \boldsymbol{\epsilon}^{\text{bias}}$ at each time step; Fig. 1 represents entirely an ideal experiment in that the imperfect model differs from the target system only in its mean state.
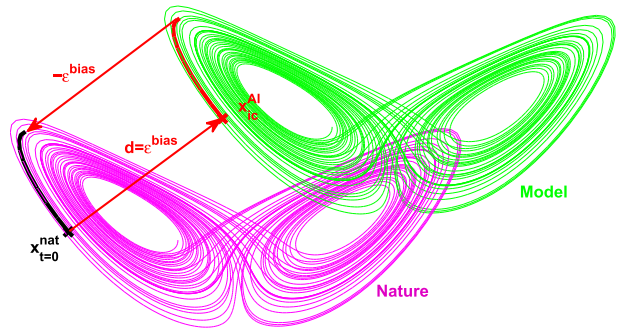


FIG. 1. Schematic of the linear mapping principle behind AI using the attractor of the Lorenz (1963) model. Depicted are the phase space evolution of the nature state (pink line), the evolution of an imperfect model of the nature (green line), a perfect observation (black crisscross), and initial condition after AI (red crisscross). The nature trajectory has been obtained through the addition of a constant vector $\boldsymbol{\epsilon}^{\text{bias}}$ onto the model state vector at each time step. The red arrows depict the mapping ($\boldsymbol{\epsilon}^{\text{bias}}$) and remapping vectors ($-\boldsymbol{\epsilon}^{\text{bias}}$).

In Fig. 1 the model integration shadows the nature evolution in a different region of phase space. At the desired forecast horizon, the bias $\boldsymbol{\epsilon}^{\text{bias}}$ is removed from the model state to obtain the perfect forecast. If the direct observation of the nature were used, as in the fidelity/FFI approach, the prediction skill would be lower. Note also that, in this ideal setup where the mapping is perfect, the accuracy of AI will depend on the observational error, as the latter will affect the shadowing of the nature. The dependency on the observational error is a key factor in the assessment of AI.

Figure 1 is useful in that it resolves an apparent contradiction of both paradigms: although the initial condition after AI has a large initial error, AI still profits from a reduction of the observational error. Here, a perfect forecast using AI is achievable only for zero observational error. Thus, the seemingly opposing challenges of an initialization scheme between reduction of initial error and a mapping onto the model attractor through addition of an error term are, at least in this hypothetical case, not contradictory. The mapping paradigm seeks a good representation of the true state on the model attractor, so far as such a representation exists, which itself will profit from better knowledge of the true state. In the case of AI in reality, this effect is very much "blurred" by the fact that imperfect models differ in many more ways from nature than only their mean climate, so that AI has been shown to profit less from observational improvements (Carrassi et al. 2014). The justification of the application of AI will therefore depend on how well the bias explains the difference between the model and nature attractors.

## c. Diagnostic tools of a mapping scheme

Let $\mathbf{x}_0^{\mathrm{nat}}$ be the nature state at $t = 0$ and $\mathbf{x}_0^m$ be the corresponding state (i.e., the image of $\mathbf{x}_0^{\mathrm{nat}}$) on the model attractor, so that the pair $\{\mathbf{x}_0^m, \mathbf{x}_0^{\mathrm{nat}}\}$ can be described as "twin" states. Let us furthermore define the map $\mathbf{M}: \mathbb{X}_N \rightarrow \mathbb{X}_M$ so that $\mathbf{M}(\mathbf{x}_0^{\mathrm{nat}}) = \mathbf{x}_0^m$. An intentional map $\mathbf{m}$, such as AI, should attempt to be as close as possible to the map $\mathbf{M}$. How well a predefined intentional map $\mathbf{m}$ will approach $\mathbf{M}$ will depend on the mismatch between model and nature (i.e., the model error). Let us identify some desirable properties connected to an intentional map $\mathbf{m}$.

1) The map $\mathbf{m}$ is a map from nature to the model attractor $\mathbf{m}: \mathbb{X}_N \rightarrow \mathbb{X}_M$, so that $\{\forall \mathbf{x}^{\mathrm{nat}} \in \mathbb{X}_N \mid \mathbf{m}(\mathbf{x}^{\mathrm{nat}})\} \subseteq \mathbb{X}_M$.

2) If $\mathbf{m}$ is differentiable in $\mathbf{x}_0^{\mathrm{nat}}$, then $\mathbf{m}(\mathbf{x}_0^{\mathrm{nat}} + \boldsymbol{\epsilon}^o) \approx \mathbf{m}(\mathbf{x}_0^{\mathrm{nat}}) + \mathbf{m}'(\mathbf{x}_0^{\mathrm{nat}})\, \boldsymbol{\epsilon}^o = \mathbf{x}_0^m + \mathbf{m}'(\mathbf{x}_0^{\mathrm{nat}})\, \boldsymbol{\epsilon}^o$. In particular, for a linear map, $\mathbf{m}(\mathbf{x}_0^{\mathrm{nat}} + \boldsymbol{\epsilon}^o) = \mathbf{m}(\mathbf{x}_0^{\mathrm{nat}}) + \mathbf{m}(\boldsymbol{\epsilon}^o) = \mathbf{x}_0^m + \mathbf{m}(\boldsymbol{\epsilon}^o)$.

3) The variable $\mathbf{m}^*$ is the reverse map from the model to the nature attractor $\mathbf{m}^*: \mathbb{X}_M \rightarrow \mathbb{X}_N$, equally subject to properties 1 and 2 after exchanging $\mathbf{x}^{\mathrm{nat}}$ and $\mathbf{x}^m$, $\mathbb{X}_M$ and $\mathbb{X}_N$, $\mathbf{m}$ with $\mathbf{m}^*$, and the observational error $\boldsymbol{\epsilon}^o$ with the forecast error $\boldsymbol{\epsilon}^f$.

Property 1 holds that all mapped states ideally lie on the model attractor. Property 2 implies that the accuracy of the mapped state is bound to depend linearly on the observational error at the first order of approximation. The dependence on the observational error is a necessary condition for the intentional map $\mathbf{m}$ to be successful in identifying corresponding states of nature on the model attractor. Note that both properties hold for the map $\mathbf{M}$. Property 3 adds that forecasted states can be remapped from the model back to the nature attractor after applying $\mathbf{m}^*$.

From properties 1 and 2 we can derive corresponding diagnostic tools with regards to the applicability of a given mapping scheme. The overlap of the mapped states and the model PDF reproduces the degree of approximation of the model attractor (property 1). This is done here using the Bhattacharyya coefficient as detailed in section 4. Then, the sensitivity to observational error is a necessary condition for the mapped states to be good corresponding states of nature (property 2).

Finally, note also that the fidelity paradigm corresponds to $\mathbf{m} = \mathbb{I}$. In the limit of the perfect model ($\mathbf{M} = \mathbb{I}$), the fidelity paradigm and the mapping paradigm become equal, and both reduce the initial error. This is verified in (5); for $\boldsymbol{\epsilon}^{\mathrm{bias}} \rightarrow 0$, $\mathbf{y}^{\mathrm{AI}} \rightarrow \mathbf{y}$. Seeing FFI as a specific case of the mapping paradigm justifies using the same diagnostic tools for FFI as well.
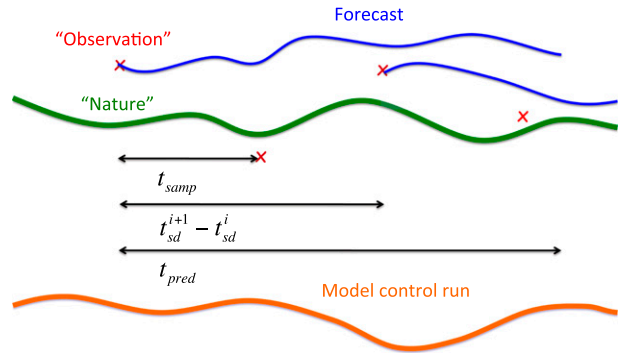


FIG. 2. Schematic of the hindcast setup. The observations (red crisscrosses) are sampled each month from the nature (green line) and used to initialize the forecasts (blue lines). The model control run is shown (orange line) systematically below the nature, indicating a model bias.

## 4. Experimental setup

Experiments are conducted under the framework of an observing system simulation experiment (OSSE; Bengtsson et al. 1981) test bed in which a nature is represented by a trajectory, a solution of the model (1), and targeted by predictions using imperfect configurations (2). A model of the nature is simulated by introducing error in the (model) equations describing the nature.

The objective is to assess the forecast quality after initialization of the imperfect configuration (using FFI/AI) compared to noninitialized trajectories (i.e., the control run of the imperfect configuration). To maximize the effect of initialization the full system is observed, so that the linear observational operator from (4) is equal to the identity, $\mathbf{H} = \mathbb{I}$. We then relate forecast quality to measures of properties 1 and 2 from section 3c.

Figure 2 is a schematic of a "hindcast" setup. The nature is obtained after a spinup run, and comprises an integration time of $t_{\mathrm{hind}}$. Observations of each variable are sampled from the nature trajectory at the sampling interval $t_{\mathrm{samp}}$; "observational error" is sampled from a univariate zero-mean Gaussian with standard deviation equal to $\sigma^o(\%)$ percent of the corresponding variable's standard deviation. Predictions—using the imperfect configuration (2)—with a horizon of $t_{\mathrm{pred}}$ are initialized at each start date $t_{\mathrm{sd}}^0, t_{\mathrm{sd}}^1, \ldots$, using either AI or FFI, over the period of $t_{\mathrm{init}} = t_{\mathrm{hind}} - t_{\mathrm{pred}}$. Similarly to the nature trajectory, the control run is obtained by integrating the imperfect configuration for $t_{\mathrm{hind}}$ after a spinup period. The control run represents a reference with regards to the assessment of forecast skill, because it contains no information on the initial condition.

As a skill measure we use the unbiased root-mean-square skill score (RMSSS) defined as

$$\mathrm{RMSSS}_i(t) = 100 \times \left[ 1 - \frac{\mathrm{RMSE}_i^f(t)}{\mathrm{RMSE}_i^b(t)} \right] \quad (\%), \quad (7)$$

where

$$\mathrm{RMSE}_i^f(t) = \sqrt{\overline{[\varepsilon_i^f(t) - \overline{\varepsilon_i^f(t)}]^2}} \quad (8)$$

is the unbiased RMSE of the forecast at lead time $t$, and $\mathrm{RMSE}_i^b(t)$ is the unbiased RMSE of the control run (i.e., background) of the imperfect configuration at the same time $t$ for the variable $i$. Unbiased means, as in (8), that the forecast bias $\overline{\varepsilon_i^f(t)}$ at time $t$ is subtracted from the forecast error $\varepsilon_i^f(t)$. The overbars indicate the average over all start dates of the hindcast period. The unbiased RMSE in (8) implicitly contains a remapping of the forecast state to the nature state according to property 3 of section 3c. This is similar to the bias correction done in practice in the postprocessing of real climate predictions. An RMSSS$(t) > 0\%$ means that the sample of initialized predictions has better skill (lower RMSE) than the equivalent sample of noninitialized predictions.

The approximation of the initial conditions of AI/FFI toward the model attractor, property 1 of section 3c, is measured as follows. The Bhattacharyya coefficient (BC; Bhattacharyya 1943) is an approximate measure of the amount of overlap between two statistical samples:

$$\mathrm{BC}(p, q) = \sum_{x \in \mathbb{X}} \sqrt{p(x)q(x)}, \quad (9)$$

where $p$ and $q$ are normalized, discrete probability distributions over the same domain $\mathbb{X}$. Therefore, $0 \leq \mathrm{BC} \leq 1$, where 1 corresponds to maximum overlap or similarity. Note that if the distributions do not overlap, the BC does not account for the level of separation and therefore does not distinguish between distributions "far apart" or "very far apart" from each other; both cases result in BC $= 0$. We measure the similarity between the full model PDF $[p(x)]$, obtained from the control run of the imperfect configuration, and the PDF of the initial conditions of AI or FFI $[q(x)]$ (i.e., $\mathbb{X}_N$, $\mathbb{X}_M \in \mathbb{X}$). This is done by generating 100 equally spaced intervals $X_k$ ($k = 1, 2, \ldots, 100$) over the common domain $\mathbb{X}$ of both PDFs (i.e., $X_1 \cap X_2 \cap \cdots \cap X_{100} \subseteq \mathbb{X}$); for each interval $k$ the probabilities $p(x \in X_k)$ and $q(x \in X_k)$ are given, and the BC is a value derived over the sum of all intervals according to (9).

The three models used in this study are briefly described in the following. Table 1 gives an overview of the models and the experimental setup.

TABLE 1. The three OSSEs as well as the associated erroneous parameters, the number of configurations, the hindcast period, the number of start dates/forecasts (i.e., sample size), and the observational error. The prediction horizon is $t_{\mathrm{pred}} = 10$ yr for all OSSEs.

| OSSE | OSSE No. | Model error | No. of configurations | $t_{\mathrm{hind}}$ (yr) | No. of start dates (sample size) | $\sigma^o$ (%) |
|---|---|---|---|---|---|---|
| L63 | 1 | $\Delta z$ | 20 | 40 | 360 | 2.5 |
| PK04 | 2a | $r^m$ | 40 | 40 | 360 | 2.5 |
| PK04 | 2b | $c^m/c_z^m$ | 109 | 40 | 360 | 2.5 |
| VD14 | 3 | $\delta^m/\theta^m$ | 16 | 110 | 200 | 5.0 |

### a. Lorenz (1963) model

The first OSSE is based on the Lorenz (1963) model (the model is hereafter referred to as L63):

$$\frac{dx^{\mathrm{nat}}}{dt} = \sigma(y^{\mathrm{nat}} - x^{\mathrm{nat}}),$$

$$\frac{dy^{\mathrm{nat}}}{dt} = rx^{\mathrm{nat}} - y^{\mathrm{nat}} - x^{\mathrm{nat}}z^{\mathrm{nat}},$$

$$\frac{dz^{\mathrm{nat}}}{dt} = x^{\mathrm{nat}}y^{\mathrm{nat}} - bz^{\mathrm{nat}}, \quad (10)$$

with parameters $\sigma = 10$, $r = 28$, and $b = 8/3$. We introduce error in the imperfect configurations by adding an offset $\Delta z$ to the $z$ variable:

$$\frac{dx^m}{dt} = \sigma(y^m - x^m),$$

$$\frac{dy^m}{dt} = rx^m - y^m - x^m(z^m + \Delta z),$$

$$\frac{dz^m}{dt} = x^m y^m - b(z^m + \Delta z), \quad (11)$$

where $\Delta z$ is a constant offset that varies as $\Delta z = 1, \ldots, 20$. It is a tuning parameter creating a bias in the $z$ variable. This enables us to investigate scenarios similar to that of Fig. 1 in which the model differs from nature mainly in its mean climate, with negligible differences elsewhere. Note that Fig. 1 has not been obtained using (11). The system is integrated using a second-order Runge–Kutta scheme with time step $dt = 0.01$. One "month" corresponds to 20 time steps as with OSSE 2; this choice is related to the similarity of models in OSSE 1 and OSSE 2, and will be justified in section 4b.

As listed in Table 1: $t_{\mathrm{hind}} = 40$ yr, $t_{\mathrm{samp}} = 1$ month, $t_{\mathrm{pred}} = 10$ yr; therefore, $t_{\mathrm{init}} = 30$ yr; $t_{\mathrm{sd}}^{i+1} - t_{\mathrm{sd}}^i = t_{\mathrm{samp}}$, so that $t_{\mathrm{sd}}^0, t_{\mathrm{sd}}^1, \ldots, t_{\mathrm{sd}}^{360}$ [i.e., the number of forecasts (sample size) is 360]; $\sigma^o = 2.5\%$.

### b. Peña and Kalnay (2004) model

The second OSSE, with equal hindcast format, is based on the Peña and Kalnay (2004) model (the model is hereafter referred to as PK04):

$$\frac{dx_e}{dt} = \sigma(y_e - x_e) - c_e(Sx_t + k_1),$$

$$\frac{dy_e}{dt} = rx_e - y_e - x_e z_e + c_e(Sy_t + k_1),$$

$$\frac{dz_e}{dt} = x_e y_e - b z_e,$$

$$\frac{dx_t}{dt} = \sigma(y_t - x_t) - c(SX + k_2) - c_e(Sx_e + k_1),$$

$$\frac{dy_t}{dt} = rx_t - y_t - x_t z_t + c(SY + k_2) + c_e(Sy_e + k_1),$$

$$\frac{dz_t}{dt} = x_t y_t - b z_t + c_z Z,$$

$$\frac{dX}{dt} = \tau\sigma(Y - X) - c(x_t + k_2),$$

$$\frac{dY}{dt} = \tau(rX - Y - SXZ) + c(y_t + k_2),$$

$$\frac{dZ}{dt} = \tau(SXY - bZ) - c_z z_t. \tag{12}$$

The model couples three copies of the three-variable L63 model (Lorenz 1963) with different spatial and temporal scales to mimic the extratropical/tropical atmospheres and ocean. The "atmospheric" variables are denoted with the lowercase variables, with the subscripts $e/t$ referring to the "extratropical"/"tropical atmosphere;" the "ocean" variables are denoted with capital letters. The two "atmospheres" are coupled through the variables $x$ and $y$ at a strength given by the parameter $c_e$; the tropical atmosphere and ocean are coupled through all variables with a strength given by the parameters $c$, for the $x$ and $y$, and $c_z$ for the $z$ component. The parameters $\sigma$, $r$, and $b$ are the same as in (10) and (11). The "uncentering" $k_1 = 10$ and $k_2 = -11$ introduce a phase lag between model compartments, while $S$ and $\tau$ modulate the amplitude and time scale of the ocean. Following Peña and Kalnay (2004) we set $S = 1$ and $\tau = 0.1$ implying that the ocean variables will have the same amplitude as the atmospheres but a slower rate by one order of magnitude (although the relative amplitude is in fact determined by the strength of the coupling). 

We define the parametric values of the nature to be $r^{nat} = 28$ and, as in Peña and Kalnay (2004), $c^{nat} = c_z^{nat} = 1$ and $c_e^{nat} = 0.08$. This configuration implies that the tropical atmosphere and the ocean are strongly coupled, while the two atmospheres are only weakly coupled. The system is integrated using a second-order Runge–Kutta scheme with a time step of $dt = 0.01$. According to Peña and Kalnay (2004), the model (12) represents an El Niño–Southern Oscillation (ENSO)-like configuration with an almost slave, small amplitude atmosphere whose regime changes are modulated by the slow ocean component. Following their convention, a simulated year is made to correspond to an ocean regime, and the system oscillates between the normal regime, lasting between 3 and 12 yr, and an El Niño regime, lasting only 1 yr, equivalent to 240 time steps in the present experimental setup (Peña and Kalnay 2004).

Model error is simulated by altering the model parameters with respect to the nature. We have generated two distinct sets of parametric error: in the forcing parameter $r$ and in the coupling parameters $c$ and $c_z$. To simulate parametric errors originating at the level of the forcing, we have modified the parameter $r^m$ so that $r^m = r^{nat} + i$, $i = 1, \ldots, 40$ with the superscript $m$ standing for model. All three compartments are affected by this error. In a second set of experiments, to simulate parametric errors originating at the level of the coupling between the different model compartments, we have modified simultaneously the tropical atmosphere/ocean coupling parameters $c^m$ and $c_z^m$, so that $c^m/c_z^m = [(c^{nat}/c_z^{nat})/10] \times i$, $i = 1, \ldots, 15$. We have assumed that the coupling between the two atmospheres is known, so that $c_e^m = c_e^{nat}$. To place ourselves in the situation in which the model is able to reproduce the qualitative behavior of nature, we have restricted our analysis to the 109 parameter combinations for which the model stability properties, as measured by the first three Lyapunov exponents, are not too different from those of the nature.

### c. Vannitsem and De Cruz (2014) model

Our third OSSE is based on the 24-variable coupled model by Vannitsem and De Cruz (2014, the model is hereafter referred to as VD14). [See also e.g., Vannitsem (2014).] In spite of the reduced number of variables, the model possesses a high degree of realism and represents the most challenging framework of our analysis. The atmospheric component is a two-layer quasigeostrophic flow defined on a beta plane, developed by Charney and Straus (1980) and extended by Reinhold and Pierrehumbert (1982). It consists of 20 variables. The ocean component is based on the reduced-gravity quasigeostrophic shallow-water model (Vallis 2006), and consists of four variables. The atmosphere and ocean are coupled through momentum transfer at the interface only. For full details see Vannitsem and De Cruz (2014).

Here again model error is simulated through mismatch in the coupling $\delta$ and in the thermal forcing

parameter $\theta$. Based on Vannitsem and De Cruz (2014), $\theta$ has the range of validity, [0, 0.2], while $\delta$ (normalized by the Coriolis parameter) [0.0001, 0.01]. For the nature trajectory, the parameters are chosen to be $\theta^{nat} = 0.1$ and $\delta^{nat} = 0.001\,938$. A total of 16 configurations with erroneous values within the above given ranges in either or both parameters are sampled, see Table 2.

A second-order numerical scheme known as the Heun scheme [see Kalnay (2002)] with a time step of 0.01 time unit is used. The dimensional time unit is equal to 0.112 15 days. An initial spinup of $2 \times 10^6$ time steps is discarded. As listed in Table 1, the hindcast format differs from OSSEs 1 and 2: $t_{hind} = 110\,yr$, $t_{samp} = 1$ month, and $t_{pred} = 10\,yr$; therefore, $t_{init} = 100\,yr$; $t_{sd}^{i+1} - t_{sd}^{i} = 6$ months, so that $t_{sd}^0, t_{sd}^1, \ldots, t_{sd}^{200}$ [i.e., the number of forecasts (sample size) is 200]; $\sigma^o = 5\%$.

## 5. Results

In most subsequent figures the forecast horizon has been chosen to be the first forecast month, over which the RMSSS in (7) is averaged. The RMSSS is obtained for each OSSE from a sample size given in Table 1. Using the models of this study, the differences in forecast quality between AI and FFI are most distinguished in the initial forecast stage; a larger initial skill corresponds to a later (e.g., "seasonal") forecast time at which all skill is lost. This is not true in general; the expectation behind AI is to improve skill at the seasonal-to-decadal time horizon by reducing drift, perhaps at the expense of larger error in the initial forecast stage. This behavior is not expected in OSSE 1 because the L63-variable model is uncoupled; it is neither observed in OSSE 2, although the slow ocean compartment modulates the fast tropical atmosphere (Peña and Kalnay 2004). In OSSE 3 we consider longer forecast horizons, because anomaly initialized forecasts show decadal skill in the ocean compartment.

### a. OSSE 1: Lorenz (1963) model

We begin by showing results from the Lorenz (1963) system given by (10) (nature) and (11) (model configurations). The sample size is 360, and the number of configurations is 20, see OSSE 1 in Table 1. Recall that this first system is the closest to the idealized schematic in Fig. 1 in that the model attractor is different from the nature attractor mainly in its first moment (i.e., its mean), which is controlled by the tuning parameter $\Delta z$. Here $\Delta z$ can be seen as a "measure" of the model error; the bias is observed mainly in the $z$ variable.

The top panels in Fig. 3 show the first month average RMSSS as a function of the bias for the $x$, $y$, and $z$ variables, respectively. The biases have been normalized by

TABLE 2. The 16 configurations with erroneous forcing/coupling parameters $\theta^m/\delta^m$ of OSSE 3; $\theta^{nat} = 0.1$ and $\delta^{nat} = 2 \times [1/(f_0 \times 10^7)]$. Here $f_0 = 1.032 \times 10^{-4}$ is the Coriolis parameter at midlatitudes ($\phi = 45°$).

| $\theta^m$ | $\delta^m\{\times[1/(f_0 \times 10^7)]\}$ |
| --- | --- |
| 0.100 | 1 |
| 0.100 | 3 |
| 0.100 | 4 |
| 0.100 | 8 |
| 0.100 | 10 |
| 0.100 | 16 |
| 0.077 | 2 |
| 0.120 | 2 |
| 0.140 | 2 |
| 0.180 | 2 |
| 0.140 | 4 |
| 0.140 | 6 |
| 0.140 | 8 |
| 0.140 | 10 |
| 0.120 | 6 |
| 0.180 | 6 |

the respective nature variance for better comparison among variables. The black (red) lines refer to FFI (AI). FFI skill in the $z$ variable is proportional to the bias (which itself is proportional to $\Delta z$ with a negative proportionality constant). On the other hand, the biases of the $x$ and $y$ variables are characterized by a nonlinear response to the model error $\Delta z$ (not shown), and are an order of magnitude smaller. FFI skill as a function of the (negligible) biases in the $x$ and $y$ variables is therefore noisy. Anomaly initialization skill outperforms FFI systematically for biases of large magnitude in the $z$ variable, remaining insensitive to the amplitude of the model error $\Delta z$. In the limit of bias $\to 0$, FFI and AI have similar skill as expected.

The bottom panels in Fig. 3 display the BC as a function of the bias, and are otherwise analogous. The BC [(9)] measures the overlap of the FFI/AI initial conditions distributions and the model PDF for each variable separately. Each BC value corresponds to a single configuration (making a total of 20 points). The constant BC about the value of one in the $x$ and $y$ variables (which have negligible bias) for FFI indicates the high coincidence of the observations with the model PDF in these variables (the bias is mainly in the $z$ variable). For the $z$ variable the BC of FFI decreases monotonously with the magnitude of the bias corresponding with decreasing overlap of the initial conditions and model PDF. On the other hand, the BC for AI (red line) remains constant as a function of the bias magnitude in all three variables. Thus, the AI initial conditions lie well within the space confined by the model attractor according to property 1 of section 3c.
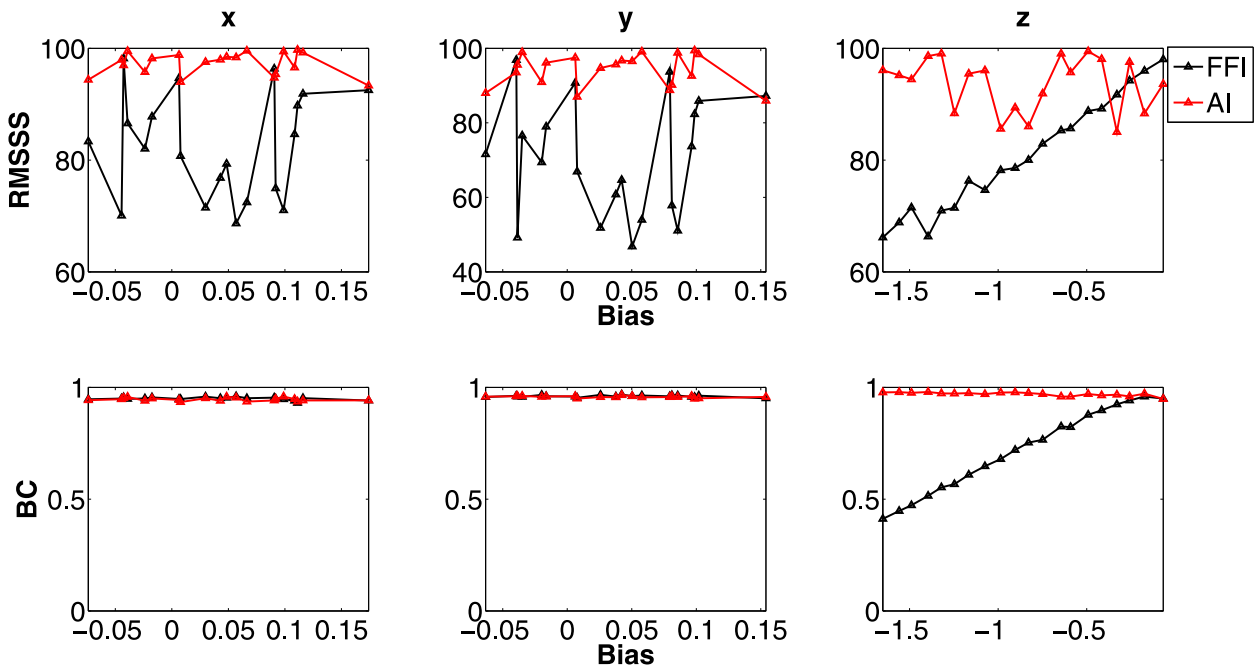
FIG. 3. RMSSS averaged over the first forecast (top) month and (bottom) BC of FFI (black) and AI (red) as a function of the bias for the $x$, $y$, and $z$ variables of (11) (OSSE 1). The bias is normalized by the respective variance (of the nature) for better comparison among variables. Displayed are 20 points corresponding with configurations characterized by a constant offset $\Delta z$.

Figure 3 illustrates two main points: the forecast skill decreases monotonously the further the initial conditions lie outside the model attractor, as is the case with FFI in the presence of this type of model error. Anomaly initialization initial conditions approximate the model attractor irrespective of $\Delta z$ (the mapping and remapping are successful), resulting in a somewhat constant forecast quality.

Figure 4 displays the RMSSS averaged over the first forecast month as a function of the observational error for all three variables of (11). The results of FFI (black) and AI (red) of three configurations with increasing model error $\Delta z = 0, 21,$ and $41$, are shown. For $\Delta z = 0$ the model is perfect. In principle FFI and AI should be identical in this case; the small discrepancy in RMSSS is due to the limited sampling of the observations resulting in an erroneous (nonzero) bias estimate in AI (see section 2). We observe that AI maintains a large sensitivity to the observational error even for very large model error, a necessary condition for $\boldsymbol{\epsilon}^{\text{bias}}$ to be a good mapping vector according to property 2 described in section 3c. We further observe that in the presence of large model error, FFI loses sensitivity to the observational error. Under the framework of the mapping paradigm in which FFI is seen as a specific case ($\mathbf{m} = \mathbb{I}$), this result is expected: the perfect model assumption behind FFI is too strong here. The presence of a large model error masks the otherwise natural sensitivity of FFI to the observational accuracy. Finally, Fig. 4 shows that the sensitivity to the observational error can reveal information about whether FFI or AI is more suitable for the model used, being in favor of AI in this case.

## b. OSSE 2: Peña and Kalnay (2004) model

In the following we show results from the second system (12) characterized by error in the forcing parameter $r^m$. The sample size is 360, and the number of configurations is 40, see OSSE 2a in Table 1. Let us begin by summarizing the changes in the model attractor incurred by this parametric error. We focus on the tropical atmosphere because it is coupled to the slow ocean, and the changes in its attractor incurred by the parametric errors are interesting case studies.

Figure 5 displays the (normalized) distributions of the first-, second-, third-, and fourth-order moments (from top to bottom) of the model PDFs of 40 configurations characterized by an erroneous forcing parameter $r^m$ (pink). The distributions are given for all three variables of the tropical atmosphere. The black dotted lines indicate the respective values associated with the nature. Figure 5 shows that the biases are largest in the $z_t$ variable (and negligible in the $x_t$ and $y_t$ variables in comparison); the variances of the configurations are almost always larger compared to the nature; changes in the third- and fourth-order moments are fairly negligible compared to the changes in the first and second moments. Thus, the
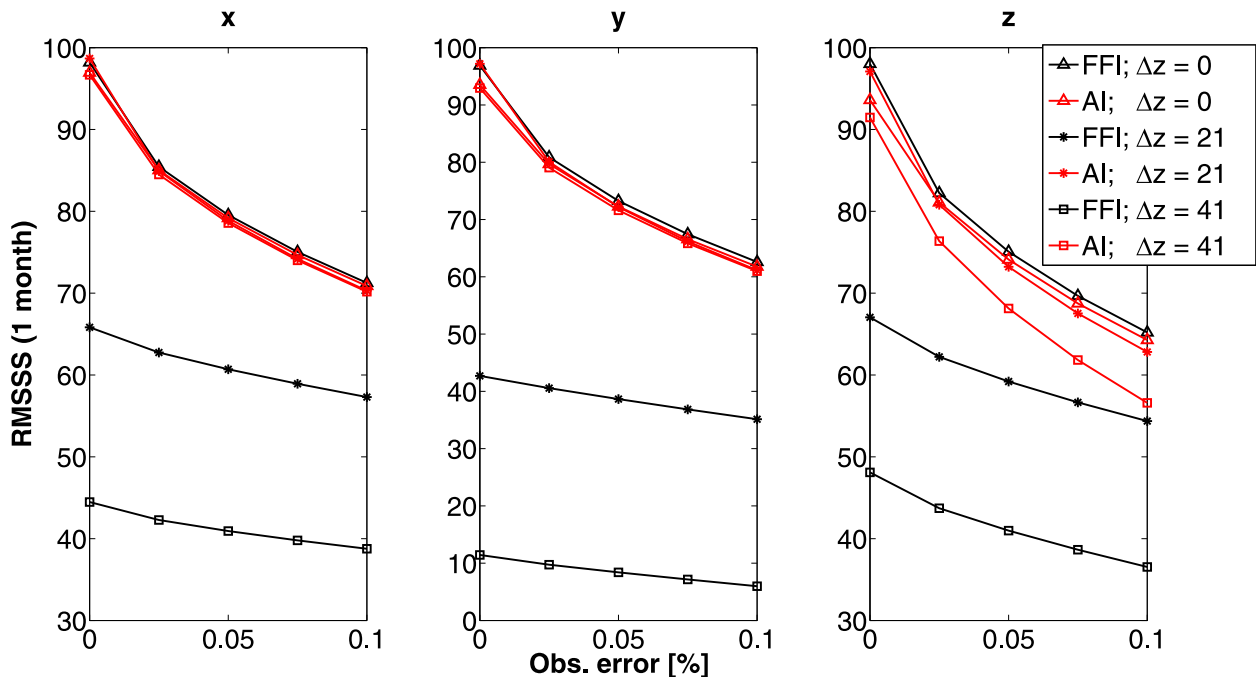
FIG. 4. RMSSS averaged over the first forecast month as a function of the observational error for all three variables of (11). Displayed are FFI (black) and AI (red) of three configurations with increasing model error $\Delta z = 0, 21,$ and 41. For $\Delta z = 0$ the model is perfect.

incurred changes in the model statistics on account of the error are similar to those of OSSE 1, with the exception that the variance increases too.

In analogy to Fig. 3, Fig. 6 shows results obtained from the tropical atmosphere of OSSE 2a in (12). The RMSSS (top) and BC (bottom) are given as a function of the bias of each variable. The biases have been normalized by the respective nature variance for better comparison among variables. Figure 6 shows 40 points corresponding to configurations characterized by erroneous forcing parameters $r^m$. In the limit of bias $\rightarrow 0$, FFI and AI have similar skill as expected. The skill of FFI varies abruptly as a function of the bias for the $x$ and $y$ variables as in Fig. 3, but note that the biases remain almost negligible in comparison to the $z$ bias. Here too, the biases of the $x$ and $y$ variables are characterized by a nonlinear response to the parametric model error in $r^m$, and are an order of magnitude smaller. In the $z$ variable we observe a monotonous decrease of skill similar to Fig. 3, because the relationship between the bias and the parametric model error in $r^m$ is linear. In contrast, the forecast quality associated with AI remains constant at a high level irrespective of the magnitude of the bias as in Fig. 3.

The BC reveals a relationship similar to Fig. 3 with respect to all three variables. Here again, the BC related to AI is about the value of one for all three variables; the approximation of AI initial conditions toward the model

attractor is successful, and results in better forecast quality. We can thus observe comparable forecast skill results (cf. Fig. 3) in the face of comparable changes in the imperfect model statistics, regardless of the origin of such a difference (i.e., parametric error $r^m$ in Fig. 6 or erroneous equation with $\Delta z$ in Fig. 3). The model statistics (PDF) reflected in the BC are a valuable source of information, and can be computed using data from the model control run.

OSSE 2b in Table 1 introduces error in the coupling parameters $c^m$ and $c_z^m$ between the model's tropical atmosphere and ocean. As before, Fig. 5 displays the (normalized) distributions of the first-, second-, third-, and fourth-order moments (from top to bottom) of the model PDFs of 109 configurations characterized by erroneous coupling parameters $c^m/c_z^m$ (blue). Note that biases are larger in the $x_t$ and $y_t$ variables, but do not reach the magnitude of the $z_t$ bias in the erroneously forced configurations (pink); the variances of the configurations are generally smaller compared to the nature; the third- and fourth-order moments are broadly distributed, and no longer negligible compared to the mean.

In analogy to Fig. 6, Fig. 7 shows results obtained from the tropical atmosphere of OSSE 2b in (12). Displayed are 109 points corresponding to configurations characterized by erroneous coupling parameters $c^m$ and $c_z^m$. In the limit of bias $\rightarrow 0$, FFI and AI have similar skill as in
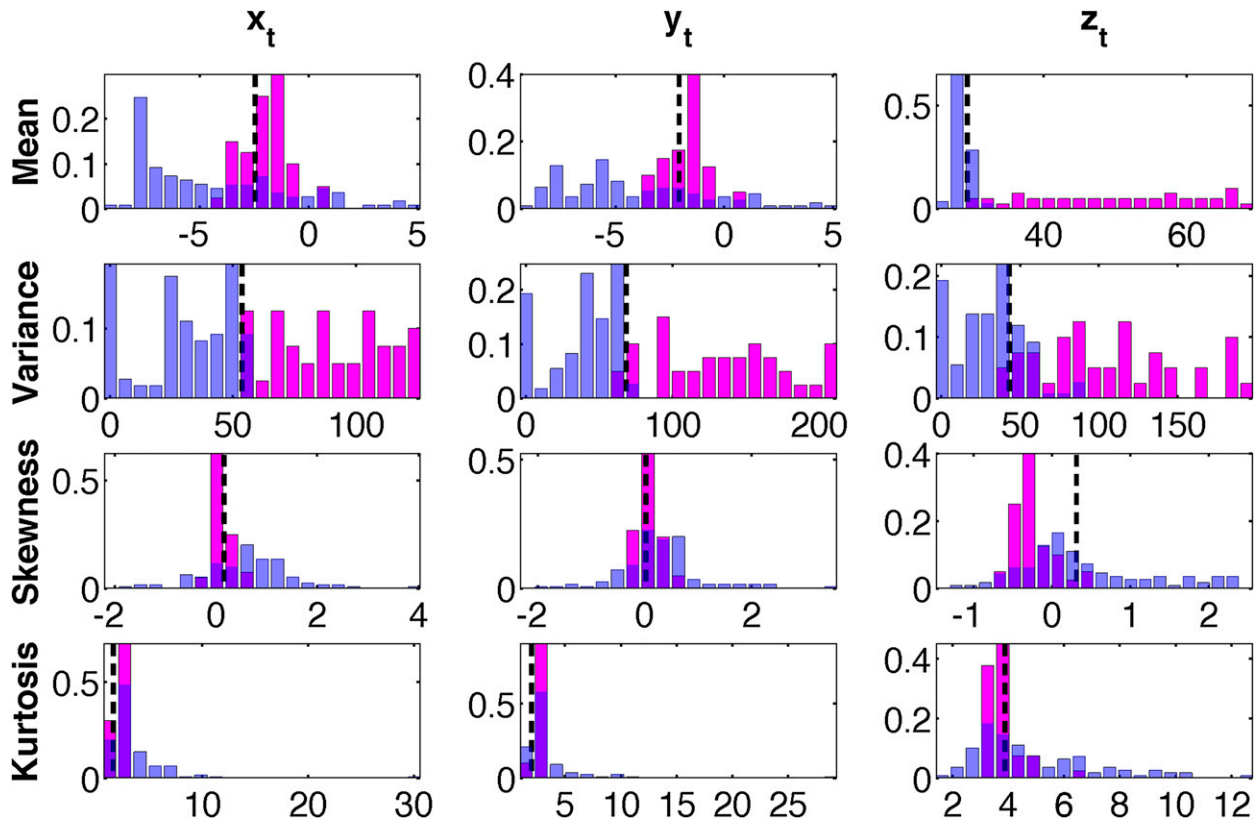
FIG. 5. (from top to bottom) Distributions of the first-, second-, third-, and fourth-order moments of the model PDF given for each variable of the tropical atmosphere (OSSE 2) for configurations characterized by an erroneous forcing parameter $r^m$ (pink) and erroneous coupling parameters $c^m/c_z^m$ (blue). The black dotted lines indicate the respective values associated with the nature.

Fig. 6. We also see that small changes cause major deterioration in the skill of AI. This stands in contrast to the results in Figs. 3 and 6, which is because the bias no longer explains the difference in the model and nature statistics. This is reflected in the bottom panels, which reveal that for the same variables the BC of either scheme deteriorates as a function of the bias magnitude. Notice that the distributions of BC points for both AI and FFI are similar, in spite of the fact that AI is designed to approximate the model attractor.

Comparing Figs. 6 and 7, we observe that AI performs badly when property 1 does not hold (i.e., when AI does not approximate the model attractor successfully). The statistical differences between the nature and model PDFs can apply to the moments of any order, a description based only on the bias (as implicitly assumed by AI) is satisfactory for OSSE 2a and unsatisfactory for OSSE 2b. In the latter case the fixed error term added onto the observational values ($\epsilon^{\mathrm{bias}}$) deteriorates the quality of the initial condition because it is an unsuitable mapping vector.

Figure 8 illustrates why for some models an approximation of the attractor using AI is successful and for

others it is not. The figure shows the PDFs of the FFI (black) and AI (red) initial conditions, as well as the PDFs of the model (blue) and the nature (green). The panels display the $z_t$ variable from an erroneously forced ($r^m = 42$) configuration (top), and the $x_t$ variable from an erroneously coupled ($c^m/c_z^m = 0.8/0.9$) configuration (bottom). In the top panel we observe that the PDF of the AI initial conditions is a good approximation of the model attractor, whereas the FFI initial conditions drawn from the observations of the nature, and by construction overlap with the nature PDF, partially lie outside the model attractor. In the bottom panel we observe that the AI initial conditions are a worse approximation of the model attractor compared to the FFI initial conditions.

The reason for the varying approximation of AI initial conditions toward the model attractor lies in a comparison of the nature and model PDFs: we observe that they differ from each other in their higher (than first) order moments, so that the first-order moment correction constituted by AI delivers a bad approximation. It is often overlooked in the implementation of AI that, being a linear correction of the observations with the
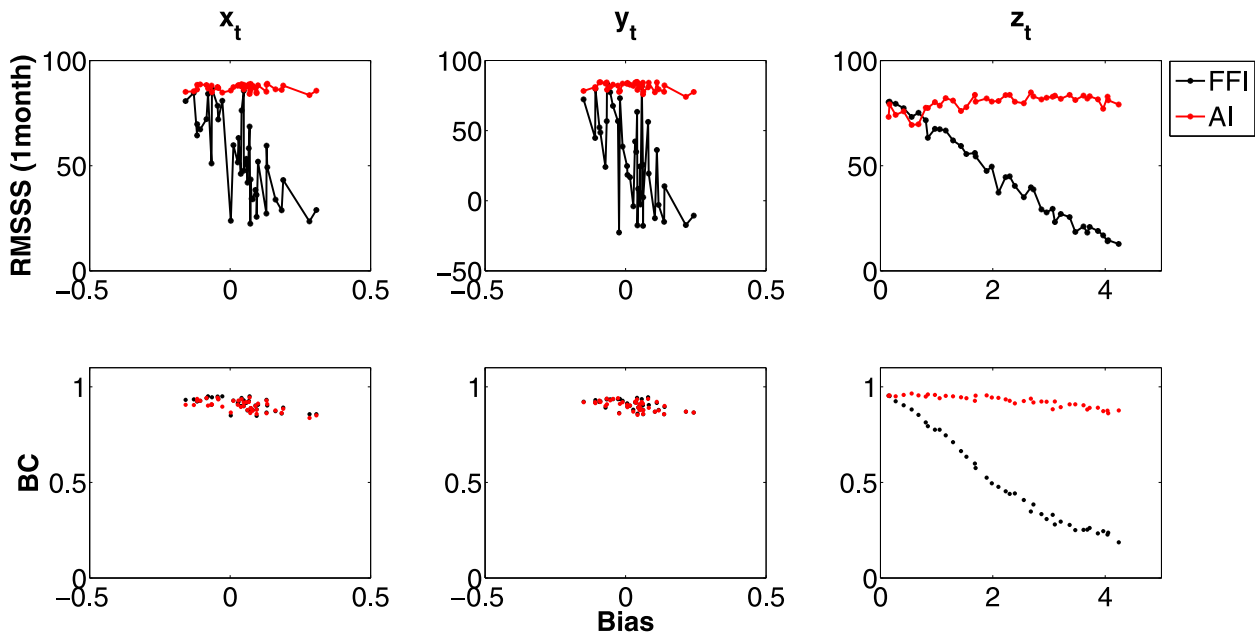
FIG. 6. RMSSS averaged over the first forecast (top) month and (bottom) BC of FFI (black) and AI (red) as a function of the bias of each variable (tropical atmosphere, OSSE 2a). The bias is normalized by the respective variance (of the nature) for better comparison among variables. Displayed are 40 points corresponding with configurations characterized by an erroneous forcing parameter $r^m$.

goal of approximating the model attractor, the assumption of negligible higher-order differences between model and nature attractors must apply. This can be further understood in the light of the ideal experiment described in the introduction, in which we stated that AI imposes a weaker requirement on the model in that it can differ from nature in its mean climate *under the condition that all else remains identical.*

Figure 9 shows the RMSSS averaged over the first month of FFI (black) and AI (red) as a function of the BC. The panels correspond to the nine variables of (12), and the 109 points correspond to configurations defined by erroneous coupling parameters $c^m/c_z^m$. Anomaly initialization skill shows a clear dependence on the degree of approximation of the model attractor measured by the BC: better skill corresponds to larger BC. FFI skill shows only a slight dependence, due to the fact that it focuses on reducing the initial error. However, in the case of FFI initial conditions that are given by the observations, the BC effectively measures the similarity of the model and nature PDFs (i.e., how well the model is capable of reproducing the nature). Therefore, a dependence is still noticeable, because a better model (corresponding to a larger BC) will improve skill. Notice also that, for configurations whose model PDF are close to the nature PDF (seen in the large FFI BC values) as is always the case in the extratropical atmosphere with known coupling $c_e^m = c_e^{nat}$, the distribution of BC values associated with AI cluster. This is because in cases in

which the model PDF is already close to the nature PDF, a mapping scheme will become less effective (with variable effects on skill).

Figure 10 displays, in analogy to Fig. 4, the RMSSS averaged over the first month as a function of the observational error for the $x_t$, $y_t$, and $z_t$ variables of the tropical atmosphere. The results of FFI (black) and AI (red) of two configurations corresponding to $r^m = 42$ (full lines) and $c^m/c_z^m = 0.8/0.9$ (dashed lines) are shown. Comparing FFI and AI for either configuration, we observe that the prevailing scheme is characterized only by a slightly higher sensitivity to observational error, so that agreement with property 2 is not robust. This can be explained by the large changes in the second moment of the model PDFs, which compromise the effectiveness of the mapping vector $\epsilon^{bias}$ in conforming with property 2 (but not necessarily property 1). For the $z_t$ variable and $r_m = 42$ the prevailing scheme (i.e., AI) is less sensitive to observational error, which might be caused by a smaller skewness in the $z_t$ variable (see Fig. 5).

Figure 11 shows the drift after FFI of the $z_t$ variable in the $r^m = 42$ configuration (black full line). An initialization shock occurs that can be associated with initial conditions lying outside the model attractor due to the large bias (see Fig. 8). The first peak is associated with the majority of trajectories following the boundaries of the attractor "wings" (cf. L63 attractor in Fig. 1), and the following minimum is associated with the majority of trajectories moving from one "wing" to
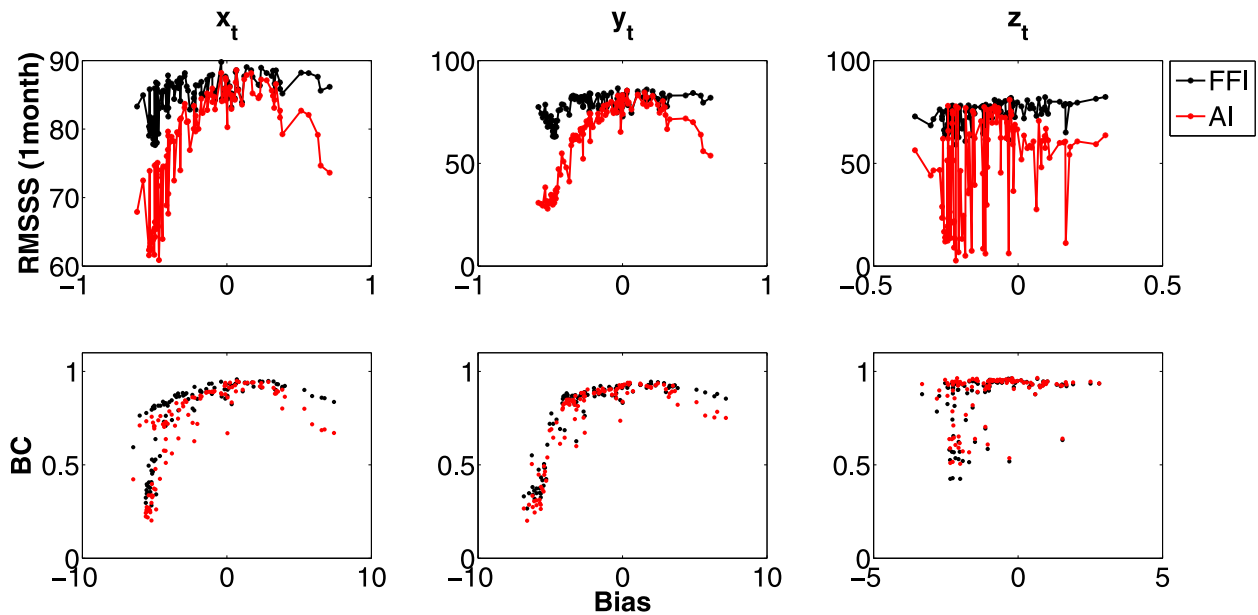
FIG. 7. RMSSS averaged over the first forecast (top) month and (bottom) BC of FFI (black) and AI (red) given as a function of the bias of each variable (tropical atmosphere, OSSE 2b). The bias is normalized by the respective variance (of the nature) for better comparison among variables. Displayed are 109 points corresponding with configurations characterized by erroneous coupling parameters $c^m/c_z^m$.

the other (not shown); thus, the drift follows the internal frequency associated with changing modes (wings), and later oscillations are progressively damped (depending on the strength of the coupling). Anomaly initialization (red full line) reduces the initial shock, because the large bias explains the difference between the model and nature statistics quite well. The initial shock is not entirely reduced, because the differences in the second-order moment between model and nature statistics are large, too.

Sanchez-Gomez et al. (2015) observe—in the Pacific region of an Earth system model after initialization of the ocean nudged toward reanalysis—a quasi-systematic excitation of ENSO warm events for all starting dates, followed by weak cold ENSO events in the second forecast year with spurious oscillatory behavior being progressively damped. The features of the drift in OSSE 2a seen in Fig. 11 appear to be in qualitative agreement with the behavior observed in Earth system models.

For the $x_t$ variable of the $c^m/c_z^m = 0.8/0.9$ configuration (black dashed line), the drift is gradual as a function of forecast lead time. Here, the bias is smaller, and the majority of the initial conditions do not lie outside the model attractor. Differences in higher (than first) order moments in the model and nature statistics cause the integrations to converge toward a different mean value, explaining the gradual drift; AI (red dashed line) does not overcome this problem, falling short of property 2. As a result, the drift after AI is larger in the first forecast

year; for longer lead times its mean error stabilizes while the drift after FFI continues to grow.

## c. OSSE 3: Vannitsem and De Cruz (2014) model

Figure 12 is analogous to Fig. 9, but shows results from OSSE 3 in Table 1, which has 16 configurations and a sample size of 200. The first 20 panels (from top left) correspond to atmospheric variables, the last 4 correspond to ocean variables. The 16 points in each panel correspond to erroneous configurations as given in Table 2. A polynomial is fitted to the data using least squares, and the error bars are estimates of the standard deviation of the error in predicting the skill at a given BC
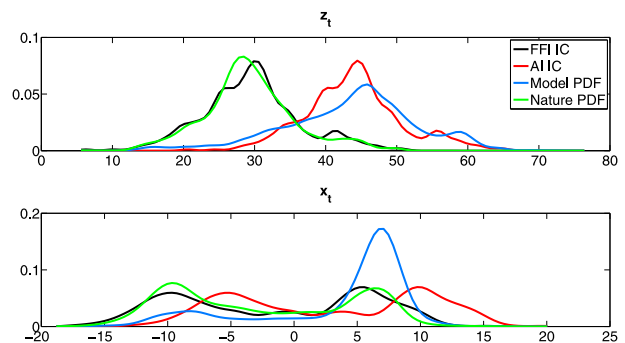


FIG. 8. PDFs of FFI (black) and AI (red) initial conditions, and PDFs of the model (blue) and nature (green) attractors. (top) The $z_t$ variable from an erroneously forced ($r^m = 42$) configuration and (bottom) the $x_t$ variable from an erroneously coupled ($c^m/c_z^m = 0.8/0.9$) configuration.
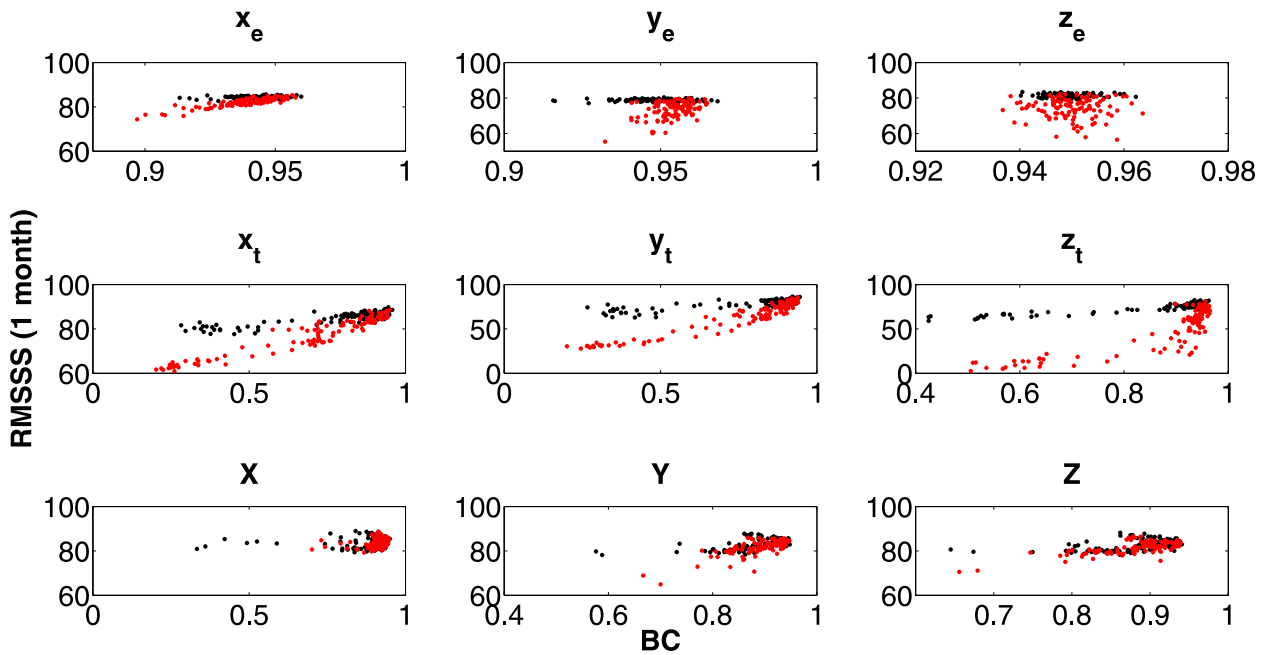
FIG. 9. RMSSS averaged over the first forecast month of FFI (black) and AI (red) as a function of the BC. The panels correspond to the nine variables of (12) (OSSE 2b), and the 109 points correspond to configurations defined by erroneous coupling parameters $c^m/c_z^m$.

value by the polynomial. FFI skill (black) clearly correlates with the BC, showing that the skill improves as the model improves. In the atmospheric variables AI skill (red) is consistently worse in comparison, which can be explained as in Fig. 7 by the fact that the BCs are

similarly distributed to those of FFI, indicating that the approximation is unsuccessful. In many variables AI skill increases with increasing BC; in others, the linear regressions show opposite signs. However, the reader can compare with Fig. 9 to notice that for BCs close to
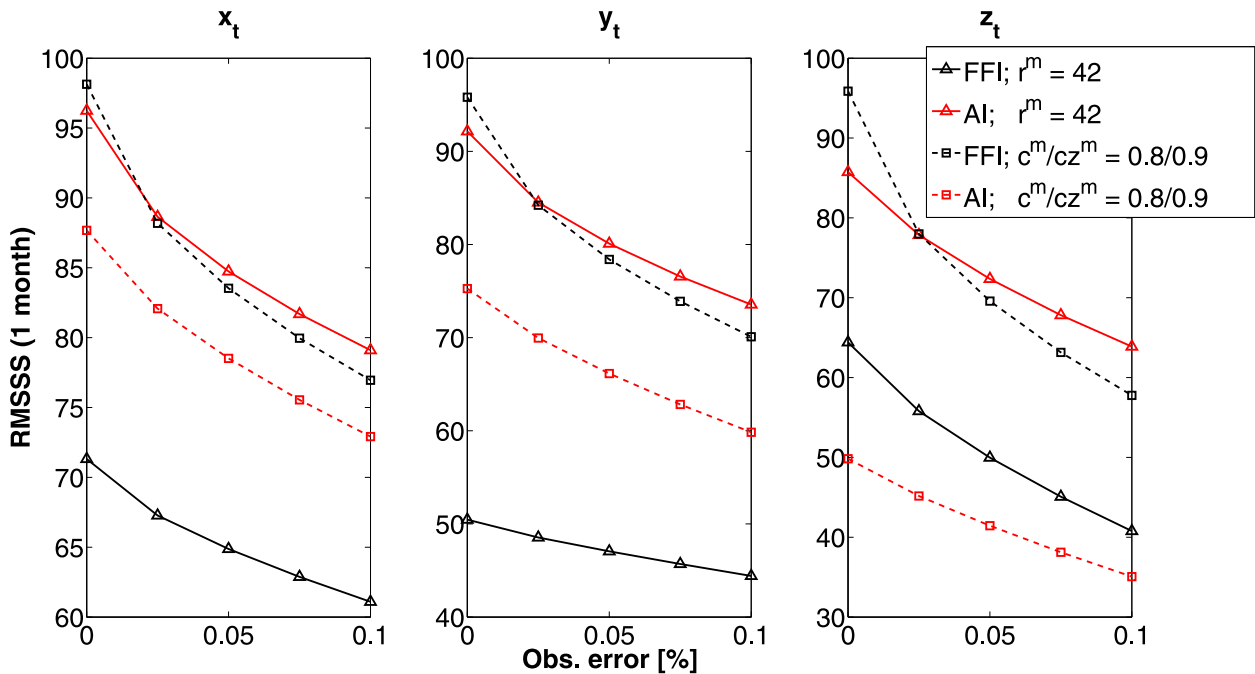


FIG. 10. RMSSS averaged over the first forecast month as a function of the observational error for the variables of the tropical atmosphere. Displayed are FFI (black) and AI (red) of two configurations corresponding to $r_m = 42$ (full lines) and $c^m/c_z^m = 0.8/0.9$ (dashed lines).
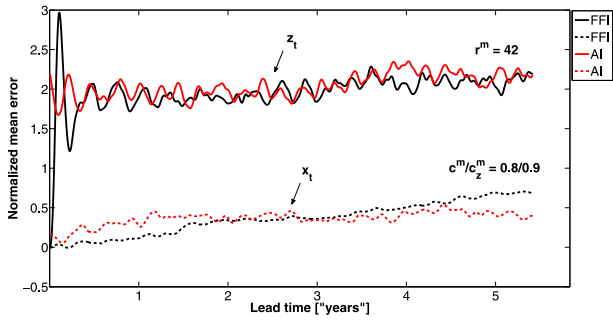
FIG. 11. Mean error over the initial conditions (normalized by the variable's nature variance for comparison) of FFI (black) and AI (red) as a function of lead time in years for the $z_t$ variable of the $r^m = 42$ configuration (full lines) and $x_t$ variable of the $c^m/c_z^m = 0.8/0.9$ configuration (dashed lines).

one, the linear dependence is less clear and the distributions of points are wider. This is true for Fig. 12 as well; many configurations are close to the nature (seen in large BC values associated with FFI), so that a mapping scheme is less effective (with variable effects on skill). In the ocean a preference among schemes cannot be inferred, which will be looked at in the following.

The top panels of Fig. 13 display the RMSSS of FFI (black) and AI (red) as a function of the lead time

in months for the four ocean variables of a particular configuration corresponding to $\theta^m = 0.1$ and $\delta^m = 8 \times [1/(f_0 \times 10^7)]$. Notice that FFI/AI skill is comparable at early forecast times, and at later horizons either AI prevails (first and third ocean variables) or FFI (second and fourth ocean variables). The difference in skill is significant for the first and third ocean variables after about the 40th month; the difference in skill in the second and fourth ocean variables at late forecast stages is not significant (see appendix for significance test). Neither properties 1 or 2 coherently explain this difference in skill among the variables themselves. Nevertheless, pinpointing skill of a specific variable to the properties of the same variable is anyhow unlikely in a multivariate nonlinear system.

In analogy to Figs. 4 and 10, in the bottom panels the RMSSS of FFI (black) and AI (red) as a function of the observational error for the same variables and configuration is shown. We can see that AI shows similar sensitivity to observational error as FFI for all four variables, explaining AI's competitive performance. In the atmospheric variables, the sensitivity to observational error of AI is consistently worse than FFI (not shown), agreeing with the result that AI performs consistently worse than FFI in the atmosphere.
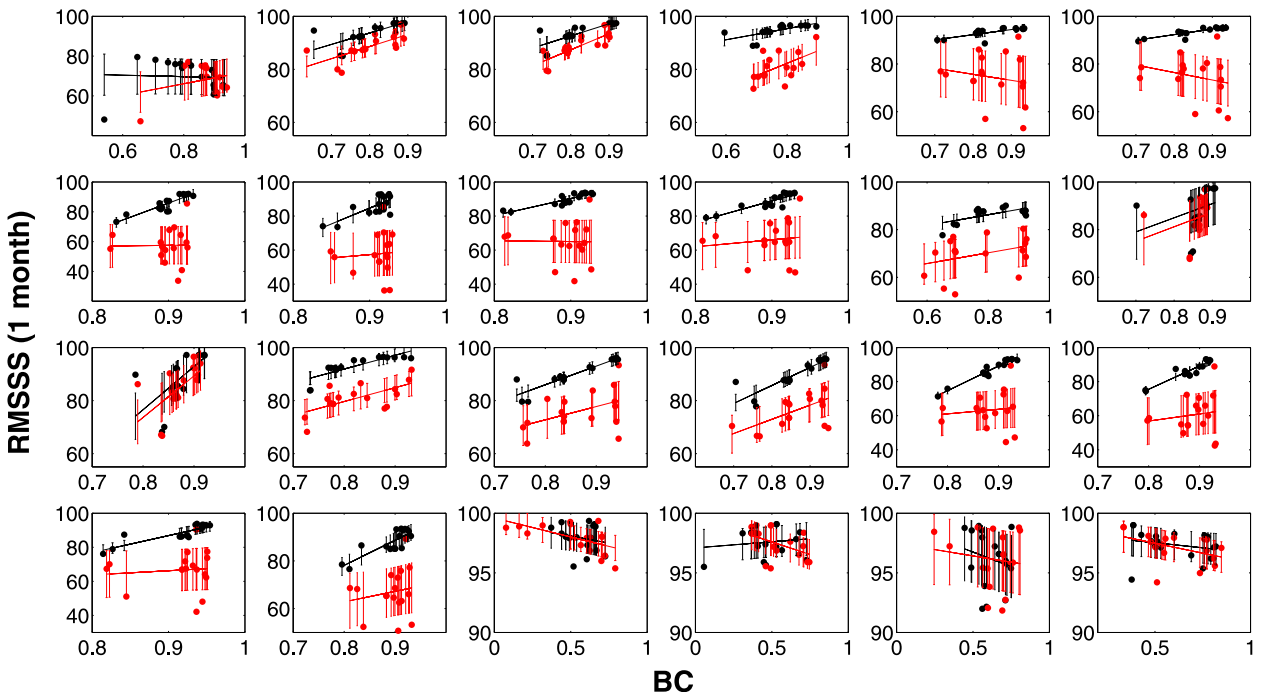


FIG. 12. RMSSS averaged over the first forecast month as a function of the BC. (from top left) The first 20 panels correspond to atmospheric variables and the last 4 correspond to ocean variables. Displayed are FFI (black) and AI (red), and the 16 points correspond to erroneous configurations of OSSE 3. A polynomial is fitted to the data using least squares, and the error bars are estimates of the standard deviation of the error in predicting the skill at a given BC value by the polynomial.
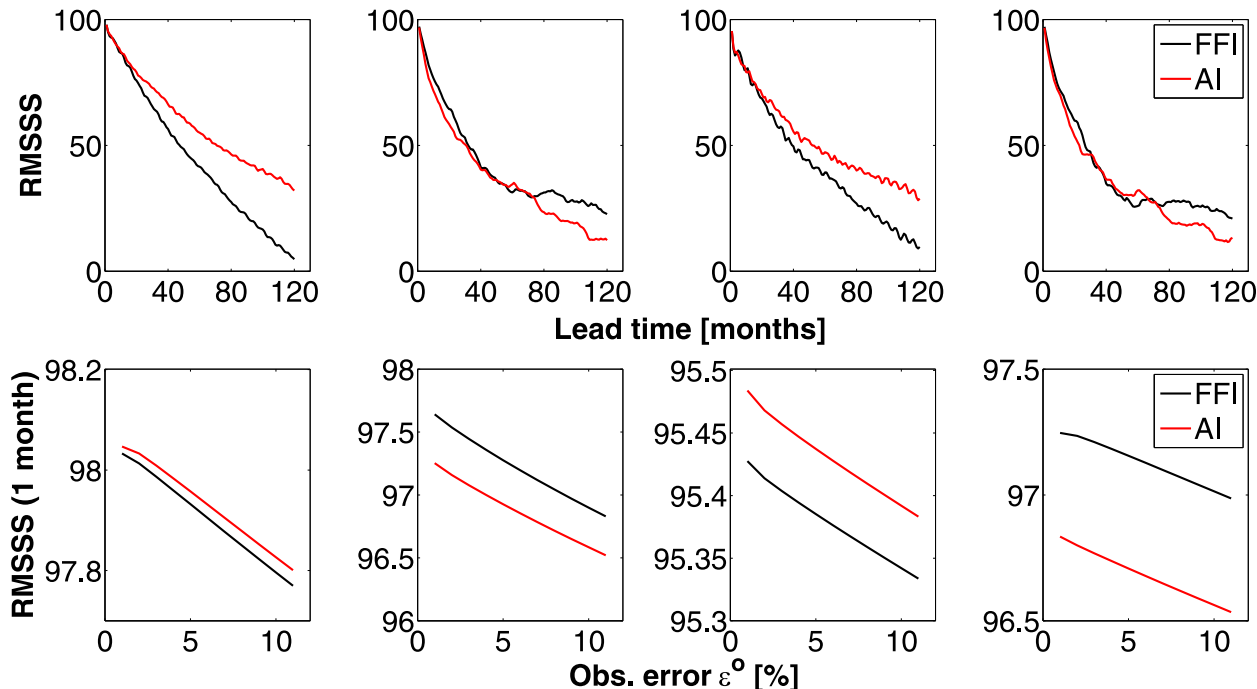
FIG. 13. (top) RMSSS of FFI (black) and AI (red) as a function of the lead time in months for the four ocean variables of a configuration corresponding to $\theta^{*,m} = 0.1$; $\delta^m = 8 \times [1/(f_0 \times 10^7)]$. (bottom) RMSSS of FFI (black) and AI (red) as a function of the observational error for the same variables and configuration.

## 6. Conclusions

This study proposes to associate FFI to the fidelity paradigm and AI to an instance of the mapping paradigm upon which we focus. Anomaly initialization is interpreted as a state-independent mapping scheme, which intends to initialize the model on an image of nature on the model attractor. The observational information is used in order to seek out a model state—the image—corresponding to the nature state. Ideally, forecasts will no longer drift because they are initialized on feasible model states. The motivation behind this work is to move away from an ad hoc implementation of AI toward an approach based on considerations of the model at hand. To the best of our knowledge, this study is the first of its kind to research the circumstances under which AI initial conditions approximate the model attractor (which is often implicitly assumed), as well as the extent to which this characteristic might improve its forecast skill.

We have diagnosed AI according to two properties: how well the initial conditions approximate the model attractor, measured by calculating the overlap of the initial conditions PDF and the model PDF using the BC in (9) for each model variable; and how well the selected model state is a reflection of the nature state, indicated by the sensitivity of forecast skill to random

(observational) error. The mapping hypothesis has been tested using a hierarchy of low-order models (see Table 1) of increasing complexity: L63, PK04, and VD14. Our results are based on the remapped fields given after a bias correction.

In OSSE 1, the tuning parameter $\Delta z$ causes the model attractors to vary from the "nature" attractor mainly in the mean, and differences in higher-than-first-order moments are negligible. Therefore, the bias $\epsilon^{\mathrm{bias}}$ explains the differences between model and nature PDFs well, and represents an adequate mapping vector. The BC values associated with the AI initial conditions show that the latter approximate the model attractor; the sensitivity of forecasts initialized with AI to observational error indicate that good images of the nature on the model attractor are selected. In agreement with these findings, the skill of AI is insensitive to the size of the bias and remains on a high level. On the other hand, the BC of FFI decreases as a function of the bias as the initial conditions lie farther outside the model attractor, and the sensitivity of forecasts initialized with FFI to observational error decreases when the bias is very large. Properties 1 and 2 are unmet, explaining why FFI skill decreases as a function of the bias.

In OSSE 2a, characterized by error in the forcing parameter $r^m$, the tropical atmosphere model attractors vary from the nature mostly in the first- and second-order

moments. The BC values associated with AI remain large, so that AI skill is high regardless of the size of the bias. FFI's initial conditions lie farther outside the model attractor as a function of the bias, indicated in the BC values. This explains why FFI skill suffers with increasing bias. We also note that AI's sensitivity to observational error (although marginally better than that of FFI) is somewhat compromised because of the larger model variance which affects property 2 (but not property 1).

In OSSE 2b, characterized by errors in the coupling parameters $c^m/c_z^m$, the differences in higher (than first) order moments between the model and nature PDFs are no longer small in comparison with the magnitude of the bias. The BC values associated with AI decrease as a function of the bias, and are no larger than the BC values associated with FFI. This result clearly indicates that AI initial conditions fail to approximate the model attractor, and explains why AI's skill collapses for even small biases—$\boldsymbol{\epsilon}^{\text{bias}}$ is not a sufficient mapping vector for such configurations. Consequently, the application of the faulty mapping vector in AI results in worse initial conditions because of a larger initial error, so that FFI performs better in comparison. The fact that AI results in worse initial states is supported by its somewhat reduced sensitivity to random error.

Finally, OSSE 3 shows a similar distribution of BC values of both AI and FFI in the atmospheric variables. Anomaly initialization performs worse than FFI, and has a larger spread in skill. Again, $\boldsymbol{\epsilon}^{\text{bias}}$ is an insufficient mapping vector with variably negative effects on skill. In the atmosphere, this is supported by AI's reduced sensitivity to observational error. In the ocean, AI shows similar sensitivity to observational error as FFI; here, forecast skill shows that AI is competitive.

We have thus shown that the mapping paradigm is a useful concept that unifies complementary schemes such as AI and FFI into a single framework. The apparent contradiction between minimizing the initial error in FFI and adding an initial fixed error term onto the initial conditions in AI is solved when considering whether the model is close to perfect (FFI) or if the model PDF differs from the PDF of nature mostly in its first-order moment but is otherwise similar to nature (AI). The model PDF is a valuable source of information about the model error that can be estimated. The assumption that AI initial conditions approximate the model attractor must be applied cautiously under evaluation of the model and reanalysis statistics. Table 3 summarizes the prevailing schemes associated with each OSSE.

It is clear that a successful estimation of the mapping vector is presently feasible only in highly idealized systems. Yet, only by being clear about the desired goal of initialization can we seek to improve it. Diagnosis tools such as

TABLE 3. The prevailing initialization scheme corresponding to each OSSE.

| OSSE | OSSE No. | Prevailing scheme |
|------|----------|-------------------|
| L63  | 1        | AI                |
| PK04 | 2a       | AI                |
| PK04 | 2b       | FFI               |
| VD14 | 3        | FFI               |

those derived from the mapping framework in this study can a priori help assess situations in which implementing either FFI or AI is preferable. Using information from the control run of an Earth system model, we can in principle always analyze the differences in the statistics between the model and the reanalysis—avoiding many experiments—opting for either AI or FFI. In practice, measuring the approximation of the model attractor based on BC values and/or the skill sensitivity to random error of each model variable is likely to be a daunting task. The choice of either AI or FFI is a dichotomous question that would require analyzing a number of BC values as large as the number of model variables. Mixed initialization with AI/FFI simplifies such an analysis, although is likely to cause inconsistencies in the initial conditions. A more promising approach could rely on the evaluation of key (independent) variables or regions of the model.

We hope that this work can stimulate research toward initialization schemes of the mapping paradigm, which take into account the state dependence of the system. One logical continuation would be to use higher (than first) order moments in the design of the mapping vector. Another will require devising practical implementation strategies for evaluating model and reanalysis PDFs that can deal with the complexity of an Earth system model and/or the development of further properties of a successful mapping scheme along the line of properties 1 and 2. Investigating the emergence of skill at seasonal-to-decadal time horizons will be key, which we have only analyzed briefly (in anomaly initialized forecasts of the ocean compartment of VD14). The use of data assimilation algorithms that are now under consideration in the initialization of seasonal-to-decadal prediction is also expected to substantially improve initialization. This is an active field of research reflected by some recent studies (Counillon et al. 2014; Tardif et al. 2014) and is seen as a main area of development with several research initiatives worldwide.

# APPENDIX

## Significance Test

We have tested the significance of the difference in skill between AI and FFI in Fig. 13 by performing a one-way ANOVA test, which tests the hypothesis that the samples are drawn from populations with the same mean against the alternative hypothesis that the population means are not all the same. The populations were given by the distributions of the squared errors of either scheme at a defined forecast horizon. The hypothesis was rejected when the $p$ value was above 0.05, meaning in such cases that the mean squared error of either scheme is not significantly different.

## REFERENCES

Bengtsson, L., M. Ghil, and E. Källen, Eds., 1981: *Dynamic Meteorology: Data Assimilation Methods.* Springer Verlag, 330 pp.

Bhattacharyya, A., 1943: On a measure of divergence between two statistical populations defined by their probability distribution. *Bull. Calcutta Math. Soc.*, **35**, 99–110.

Carrassi, A., R. Weber, V. Guemas, F. Doblas-Reyes, M. Asif, and D. Volpi, 2014: Full-field and anomaly initialization using a low-order climate model: A comparison and proposals for advanced formulations. *Nonlinear Processes Geophys.*, **21**, 521–537, doi:10.5194/npg-21-521-2014.

Charney, J. G., and D. M. Straus, 1980: Form-drag instability, multiple equilibria, and propagating planetary waves in baroclinic, orographically forced, planetary wave systems. *J. Atmos. Sci.,* **37**, 1157–1176, doi:10.1175/1520-0469(1980)037<1157: FDIMEA>2.0.CO;2.

Counillon, F., I. Bethke, N. Keenlyside, M. Bentsen, L. Bertino, and F. Zheng, 2014: Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: A twin experiment. *Tellus*, **66A**, 21074, doi:10.3402/tellusa.v66.21074.

Doblas-Reyes, F., J. García-Serrano, F. Lienert, A. P. Biescas, and L. Rodrigues, 2013: Seasonal climate predictability and forecasting: Status and prospects. *Wiley Interdiscip. Rev.: Climate Change*, **4**, 245–268, doi:10.1002/wcc.217.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, doi:10.1175/2009BAMS2607.1.

Hazeleger, W., V. Guemas, B. Wouters, S. Corti, I. Andreu-Burillo, F. Doblas-Reyes, K. Wyser, and M. Caian, 2013: Multiyear climate predictions using two initialization strategies. *Geophys. Res. Lett.*, **40**, 1794–1798, doi:10.1002/grl.50355.

Janjić, T., and S. Cohn, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Wea. Rev.*, **134**, 2900–2915, doi:10.1175/MWR3229.1.

Kalman, R. E., 1960: A new approach to linear filtering and prediction problems. *J. Fluids Eng.*, **82**, 35–45.

Kalnay, E., 2002: *Atmospheric Modeling, Data Assimilation and Predictability.* Cambridge University Press, 364 pp.

Lorenz, E. N., 1963: Deterministic nonperiodic flow. *J. Atmos. Sci.*, **20**, 130–141, doi:10.1175/1520-0469(1963)020<0130: DNF>2.0.CO;2.

Magnusson, L., M. Alonso-Balmaseda, S. Corti, F. Molteni, and T. Stockdale, 2013: Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dyn.*, **41**, 2393–2409, doi:10.1007/s00382-012-1599-2.

Palatella, L., A. Carrassi, and A. Trevisan, 2013: Lyapunov vectors and assimilation in the unstable subspace: Theory and applications. *J. Phys. A: Math. Theor.*, **46**, 254020, doi:10.1088/1751-8113/46/25/254020.

Peña, M., and E. Kalnay, 2004: Separating fast and slow modes in coupled chaotic systems. *Nonlinear Processes Geophys.*, **11**, 319–327, doi:10.5194/npg-11-319-2004.

Reinhold, B. B., and R. T. Pierrehumbert, 1982: Dynamics of weather regimes: Quasi-stationary waves and blocking. *Mon. Wea. Rev.*, **110**, 1105–1145, doi:10.1175/1520-0493(1982)110<1105: DOWRQS>2.0.CO;2.

Sanchez-Gomez, E., C. Cassou, Y. Ruprich-Robert, E. Fernandez, and L. Terray, 2015: Drift dynamics in a coupled model initialized for decadal forecasts. *Climate Dyn.*, doi:10.1007/s00382-015-2678-y, in press.

Smith, D., and J. Murphy, 2007: An objective ocean temperature and salinity analysis using covariances from a global model. *J. Geophys. Res.*, **112**, C02022, doi:10.1029/2005JC003172.

——, A. Cusack, A. Colman, C. Folland, G. Harris, and J. Murphy, 2007: Improved surface temperature prediction for the coming decade from a global climate model. *Science*, **317**, 796–799, doi:10.1126/science.1139540.

——, R. Eade, and H. Pohlmann, 2013: A comparison of fullfield and anomaly initialization for seasonal to decadal climate prediction. *Climate Dyn.*, **41**, 3325–3338, doi:10.1007/s00382-013-1683-2.

Stockdale, T., 1997: Coupled ocean–atmosphere forecasts in the presence of climate drift. *Mon. Wea. Rev.*, **125**, 809–818, doi:10.1175/1520-0493(1997)125<0809:COAFIT>2.0.CO;2.

Tardif, R., G. J. Hakim, and C. Snyder, 2014: Coupled atmosphere–ocean data assimilation experiments with a low-order climate model. *Climate Dyn.*, **43**, 1631–1643, doi:10.1007/s00382-013-1989-0.

Toth, Z., and M. Peña, 2007: Data assimilation and numerical forecasting with imperfect models: The mapping paradigm. *Physica D*, **230**, 146–158, doi:10.1016/j.physd.2006.08.016.

Vallis, G., 2006: *Atmospheric and Oceanic Fluid Dynamics.* Cambridge University Press, 745 pp.

Vannitsem, S., 2014: Stochastic modelling and predictability: Analysis of a low-order coupled ocean-atmosphere model. *Philos. Trans. Roy. Soc. London*, **A372**, 2018, doi:10.1098/rsta.2013.0282.

——, and L. De Cruz, 2014: A 24-variable low-order coupled ocean-atmosphere model: OA-QG-WS v2. *Geosci. Model Dev.*, **7**, 649–662, doi:10.5194/gmd-7-649-2014.