School of Mathematical
and Physical Sciences

University of Reading

Department of Mathematics and Statistics

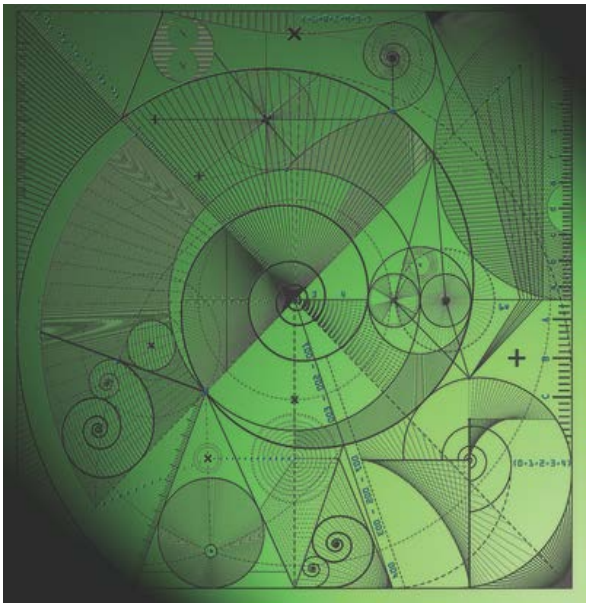
Preprint MPS-2014-18

18 December 2014

The Error of Representation: Basic Understanding

by

Daniel Hodyss and **Nancy Nichols**



The Error of Representation: Basic Understanding

Daniel Hodyss[1]

Naval Research Laboratory, Marine Meteorology Division,

Monterey, California, USA

Nancy Nichols

University of Reading, School of Mathematical and Physical Sciences,

Reading, UK

To appear in:

Tellus A

18 December 2014

[1] Dr. Daniel Hodyss, Naval Research Laboratory, Marine Meteorology Division, 7 Grace Hopper Ave., Stop 2, Monterey, CA 93943, email: daniel.hodyss@nrlmry.navy.mil

Abstract

Representation error arises from the inability of the forecast model to accurately simulate the climatology of the truth. We present a rigorous framework for understanding this kind of error of representation. This framework shows that the lack of an inverse in the relationship between the true climatology (true attractor) and the forecast climatology (forecast attractor) leads to the error of representation. A new gain matrix for the data assimilation problem is derived that illustrates the proper approaches one may take to perform Bayesian data assimilation when the observations are of states on one attractor but the forecast model resides on another. This new data assimilation algorithm is the optimal scheme for the situation where the distributions on the true attractor and the forecast attractors are separately Gaussian and there exists a linear map between them. The results of this theory are illustrated in a simple Gaussian multivariate model.

1. Introduction

Representation error in this manuscript will refer to the impact of the unavoidable misrepresentation of complex atmospheric flows by the inadequacies of the forecast model on the data assimilation. This misrepresentation of complex fluid flows arises for the most part from the inability of the forecast model, using the relatively coarse grids customarily employed in numerical weather prediction, to resolve small-scale properties of the boundary conditions as well as other small-scale properties of the turbulent flows in the interior of the fluid. This misrepresentation of the flow leads to an incompatibility between the observations of the true state, which see these small-scale processes, and the relatively coarser states achievable by the forecast model. The result of this incompatibility is that the forecast model can effectively consider the states implied by the observations to be inconsistent with its own attracting manifold, and therefore either ignore some or all of the information in the observations or react pathologically to them.

Recognition of this incompatibility between the observations of the true state and the states achievable by the forecast model goes back at least to Petersen and Middleton (1963) with more thorough and modern treatments in Daley (1983), Mitchell and Daley (1997a,b), Liu and Rabier (2002), Janjic and Cohn (2006) and Frehlich (2006). This body of work has correctly identified that because of the incompatibility mentioned above performance gains in the quality of the analysis can be made by inflating the observation error variances and accounting for the

implied correlations between observations owing to unresolved processes. In addition to these more theoretical works there have recently been attempts at estimating the structure of representation error from observational data and numerical model output (e.g. Richman et al. 2005, Oke and Sakov 2007, Frehlich 2008, Waller et al. 2013). This work has shown that there are a number of ways to see this error of representation in data. For example, the standard observations of temperature and humidity as well as aircraft measurements of turbulence have been compared with forecasts and shown to contain a component consistent with errors in representation.

This paper intends to conjoin this past work under a single unifying theme. The basic idea is to extend the Kalman (1960) filter to explicitly account for the fact that the climate (attracting manifold) of the Earth's atmosphere is distinct from that of the forecast model. To understand how we will accomplish this it will prove illustrative to review Kalman's original setup. In the work of Kalman the flaws in the model were accounted for using a white noise source, viz.

$$\mathbf{x}_f^{k+1} = \mathbf{M}\mathbf{x}_f^k + \mathbf{w}^k, \tag{1.1}$$

where $\mathbf{M}$ is a linear model, $\mathbf{x}_f^k$ is the forecast at time k and $\mathbf{w}^k$ is white noise drawn from $N(\mathbf{0},\mathbf{Q})$. The idea behind (1.1) is that the model $\mathbf{M}$ is not entirely adequate at producing the set of states that correctly describes all the potential true states (one of which is being observed by the observational instruments). In terms of the problem at hand, we may interpret (1.1) as implying that the forecast model's climatology (attracting manifold) is not sufficient to encompass the complete set of potential true states. The inclusion of the white noise source is then there to enhance the spread of states in state space (with $\mathbf{Q}$ chosen large enough to more than fill this gap between the true and forecast attractors) such that this noisy forecast model does encompass the complete set of potential true states. In other words, the action of this noise is such that the forecast model's attracting manifold, which is distinct from that of the true attracting manifold, is in essence "blurred" in phase-space until the states available to (1.1) encompasses the states on the true attracting manifold and many others for that matter.

There are at least two downsides to this choice to render the model stochastic. The first is that while probability forecasts using this noisy model now correctly assign non-zero probability to states on the true attracting manifold other forecasts from (1.1) will assign non-zero

probability to states off the true attracting manifold, which of course implies that implausible events are assigned a non-zero probability of occurrence. The second issue with this noisy model is that in fluids as complex as the general circulation of the atmosphere it is well-known that choosing the character of this noise is difficult and improper choices may lead to issues with the physical realism of the fluid evolution from the noisy forecast model (e.g., Hodyss et al. 2014).

While we believe that stochastic modeling can be useful the tack taken here is to explore what happens when one does not add noise to the model but addresses the climatology (attractor) differences between the true physical system and that of the forecast model through the use of maps between distributions on each attractor. We will derive the best, linear unbiased estimate (minimum error variance) of the state *on the forecast model attractor* given observations of states on the true attractor. This new data assimilation method will be optimal when the distributions on the true attractor and the forecast attractor are Gaussian and the map between them is linear. By deriving the data assimilation method that produces the best state estimates *on the forecast model attractor* we will see the error of representation arise as a natural consequence of a specific form of attracting manifold difference.

Finally, we would like to point out here that one common viewpoint for the connection between the data assimilation process and the forecast process is the expectation that the most accurate forecast arises from the most accurate estimate of the true state. This desire for the data assimilation algorithm to produce an estimate that is close to the true state is conceptually easy to rationalize when the model is perfect. However, when the model is flawed creating a data assimilation algorithm that produces an accurate estimate of the true state is likely to mean that this state is in some way incompatible (e.g. unbalanced) with the flawed forecast model. This notion that the true state might not be the best initial condition for the flawed forecast model has lead us to choose to develop a data assimilation system that attempts to produce a state estimate on the forecast attractor. The estimation of the true state is then one relegated to post-processing of the resulting forecasts. More discussion of the ramifications of this choice will be made throughout the development and in the conclusions.

In section 2 we derive the general theory for this new form of data assimilation algorithm. In section 3 we apply the theory of section 2 to representation error in the form of a smoothing operator that describes the differences between the true and forecast attractors. Section 4 applies the theory of section 3 to a multivariate Gaussian model to illustrate the basic ideas in their

simplest forms. Section 5 closes the manuscript with a summary of the results and a discussion of the conclusions.

2. The Two Attractor Problem

In this section we formulate a general theory for data assimilation in the situation where the observations are of states in one subset of state space but the forecast model resides in another. Our basic assumption throughout will be that our goal in such a situation is to develop a data assimilation method that will provide the best estimate of the state in the region of state space in which the forecast model resides. As we go we will illustrate the theory of this section using a very simple, example problem that we believe illustrates the basic ideas in their simplest form.

a) The True Posterior

We imagine the true state, $\mathbf{x}_t$, to be an N-vector and that it is drawn from a climatological distribution whose probability density function (pdf) we label $\rho(\mathbf{x}_t)$. By "climatological" we are referring to the pdf we would obtain if we ran the true model for a very long time, discretized state space, and counted the number of times the state entered each cell of our discretization. In the limit as this true model simulation becomes very long and the volume of each cell of our discretization of state space becomes very small we obtain this climatological pdf. We define from this climatological pdf the set $\mathbb{A}_t$ of states $\mathbf{x}_t$ with the property that $\rho(\mathbf{x}_t) > 0$ and will subsequently refer to the set $\mathbb{A}_t$ as the true "attracting manifold" of the physical system under consideration. Therefore, the set $\mathbb{A}_t$ consists of all the states in which the true state at any point in the future may be found.

We simply use the phrase "attracting manifold" throughout this manuscript as a convenient way to refer to the portion of state space in which the true physical system resides. We emphasize however that the following theory does not require the existence of a compactly supported region in state space with the properties normally associated with attracting manifolds as we will apply the theory to example problems which do not technically have this feature.

More discussion of this set and its relationship to the "attracting manifold" of the forecast model can be found in the next sub-section.

We obtain a sequence of p-vector observations, $\mathbf{y}_j = \mathbf{H}\mathbf{x}_t + \boldsymbol{\varepsilon}_o$ that were taken at various times $j = 1, 2, \ldots, J$. The object $\mathbf{H}$ is a vector-valued function and, for simplicity, will be assumed to be a linear matrix operator (p x N) with the instrument errors drawn from $\boldsymbol{\varepsilon}_o \sim N(\mathbf{0}, \mathbf{R}_i)$. We emphasize however that the results presented below will not depend on a linear observation operator. For simplicity we will assume that both the instrument error variance, $\mathbf{R}_i$, and the number of observations, p, per assimilation time, j, are fixed constants. Because we have observations at various times, j, this implies that the state must also be integrated through time. In the interest of simplicity of presentation we do not attach a label to $\mathbf{x}_t$ that denotes its relevant time j. However, because we will refer to the "filtering" data assimilation problem throughout, this should cause no confusion as $\mathbf{x}_t$ will always be considered to be at the time of the latest set of observations.

We begin by assimilating the $j = 1$ set of observations using Bayes' rule, viz.

$$\rho(\mathbf{x}_t|\mathbf{y}_1) = C_1 \rho(\mathbf{y}_1|\mathbf{x}_t)\rho(\mathbf{x}_t), \tag{2.1}$$

where $C_1 = 1/\rho(\mathbf{y}_1)$. The density $\rho(\mathbf{y}_1|\mathbf{x}_t)$ describes the conditional distribution of observations given a particular value of the state on the true attractor (often referred to as the observation likelihood). The interpretation of the act of employing Bayes' rule in (2.1) is simply as a "windowing" function through the observation likelihood that acts to reduce the view of the climatological distribution to a portion of $\mathbb{A}_t$ in the vicinity of the observation. This is important because, under the assumption that the observation likelihood is Gaussian, which implies that $\rho(\mathbf{y}_1|\mathbf{x}_t) > 0$ for all values of $(\mathbf{x}_t, \mathbf{y}_1)$, this means that the act of data assimilation does not change the states in $\mathbb{A}_t$ but simply re-calculates their relative probability of being the true state. We will use this fact later in the next sub-section.

By repeating the indicated operations in (2.1), and using the Fokker-Planck equation to propagate the resulting distributions forward in time between each set of observations, we may repeat the process in (2.1) up to the present time, $j = J$, for which we have observations $\mathbf{y}_J$ such that:

$$\rho\left(\mathbf{x}_t \middle| \mathbf{Y}_J\right) = C_J \rho\left(\mathbf{y}_J \middle| \mathbf{x}_t\right) \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right), \tag{2.2}$$

where the symbol $\mathbf{Y}_j$ denotes the set of all observations at all times up to and including the jth set and the $C_J = 1/\rho\left(\mathbf{y}_J \middle| \mathbf{Y}_{J-1}\right)$ is simply the normalization. The density $\rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right)$ will hereafter be referred to as the "prior." The density $\rho\left(\mathbf{x}_t \middle| \mathbf{Y}_J\right)$ describes the conditional distribution of all possible true states given all observations including the present set; this density will hereafter be referred to as the "posterior."

As alluded to in the beginning of this section we construct here a simple data assimilation example that we believe will illustrate the basic properties of the more complex concepts in the remainder of this section in the simplest way. To this end we assume the true states to be characterized by a two-vector whose climatological distribution is illustrated in figure 1a. This distribution is Gaussian with a variance of 3 on each variable and a covariance between the two variables of 1. Note that the set $\mathbb{A}_t$ in this case is the entire plane as a Gaussian vanishes nowhere. An observation of only one of the variables is made and defines an observation likelihood with instrument error variance equal to 1 (Figure 1b). Calculating equation (2.1) from the climatological distribution and observation likelihood for an observation of $y_1 = 1$ obtains the posterior in figure 1c. For simplicity, we further assume that the true model that is propagating the states forward in time is simply the identity. This implies that the posterior from (2.1) is simply the prior for the next assimilation step in (2.2) and the result of this calculation with $y_2 = 3$ is plotted in figure 1d. One can see in figures 1c and 1d that the assimilation of the observations has reduced the variance greatly for the observed variable but less so for the unobserved variable. In the next sub-section we will return to this example and illustrate its relationship to the forecast posterior.

b) The Forecast Posterior

Because of our fundamental inability to construct an exact model of the evolution of the general circulation of the atmosphere, the posterior distribution described by (2.2) can never be produced by a data assimilation algorithm. This is true for several reasons. First, even if we could give the forecast model the true state it would not produce the correct forecast evolution because of incorrectly specified parameters and altogether missing physics. Second, what exactly is the appropriate true state to give the forecast model as an initial condition is

ambiguous given the fact that this model can only represent states that are "smoothed", or said another way, truncated with respect to the truth (Frehlich, 2011). More on this notion will be presented in sections 3 and 4.

Because the forecast model cannot represent the true attractor, we begin by defining the state on the forecast attractor, $\mathbf{x}_f$, as an M-vector and hypothesizing that it too is drawn from a "climatological" distribution whose pdf we label $\rho\left(\mathbf{x}_f\right)$. We emphasize that the "climatology" here is different from that of the true distribution denoted above and results from running the forecast model for a very long time, discretizing state space, and counting the number of times the forecast state enters each cell of our discretization. We again define from this forecast climatological distribution a new and distinct set of states $\mathbb{A}_f$ with the property that $\rho\left(\mathbf{x}_f\right) > 0$ and will subsequently refer to the set $\mathbb{A}_f$ as the "attracting manifold" of the forecast model's *representation* of the physical system under consideration.

Next, we hypothesize the existence of a map (function) over $\mathbb{A}_t \to \mathbb{A}_f$. We do this by relating the states on the forecast attractor and the states on the true attractor through a vector-valued function:

$$\mathbf{x}_f = \mathbf{F}\left(\mathbf{x}_t\right). \tag{2.3}$$

The function, **F**, represents a mapping from the true attracting manifold to that of the forecast model. We emphasize that this is a mapping from one state-space ($\mathbb{A}_t$) to another ($\mathbb{A}_f$) and not a direct relationship between today's truth and today's forecast, which because of the noise in the observations could not possibly satisfy (2.3). We will assume that this function, **F**, has the property that for every $\mathbf{x}_t$ in $\mathbb{A}_t$ there is a corresponding element of $\mathbf{x}_f$ in $\mathbb{A}_f$. However, we will not in general assume that the converse is true and therefore we will discuss, **F**, as lacking an inverse, but we will also examine specific situations where this function does have an inverse. Please see figure 2.

From a numerical weather prediction perspective we view the equation (2.3) much like that of the algorithms in the ensemble post-processing and bias-correction literature (e.g. Glahn and Lowry 1972, Raftery et al. 2005) in which climatological information is used to build relationships between forecasts and observations. Additionally, we view equation (2.3) as having both a practical as well as a pedagogical application. From a practical perspective one may wish to deduce the relationship (2.3) from some climatological archive of forecast-truth

pairs in order to implement the data assimilation algorithm discussed in this manuscript. By contrast, and from a pedagogic perspective, one may wish to simply assert a particular relationship in (2.3) and subsequently use the framework presented below to understand its implications. This last tack will be the one illustrated in this manuscript as we will focus in sections 3 and 4 on a linear map in (2.3) that is to be interpreted as a smoothing operator as this was used in previous work in the representation error literature (e.g., Liu and Rabier 2002, Waller et al. 2013). In a sequel to this manuscript we will illustrate estimation techniques for **F** and the results of its application to different cycling data assimilation experiments.

The equation (2.3) allows for the definition of several new densities in this problem that will prove to be useful tools in the subsequent analysis. Equation (2.3) implies that the density that describes the distribution of states on the forecast attractor given a state on the true attractor, which we will refer to as a *conversion density*, is the Dirac delta function:

$$\rho\left(\mathbf{x}_f \middle| \mathbf{x}_t\right) = \delta\left(\mathbf{x}_f - \mathbf{F}\left(\mathbf{x}_t\right)\right). \tag{2.4}$$

This is because given a particular true state $\mathbf{x}_t$ there is one and only one forecast state $\mathbf{x}_f$ that it is related to and hence there is no uncertainty in the position on the forecast attractor given a particular realization of the state on the true attractor. It is important to realize that (2.3) implies that while $\rho\left(\mathbf{x}_f \middle| \mathbf{x}_t\right)$ has zero variance that in general the *converse* conversion density $\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)$ has a non-zero variance because **F** does not necessarily have an inverse. Last, the assumption that the relationship between the two attractors is of the form (2.3) allows for the subsequent simplification in (2.4) that the conversion density is simply the Dirac delta function, but this assumption is technically not required by the following theory and is largely used to allow greater focus on the relationship between the lack of an inverse in (2.3) and the representation error.

Bayes' rule tells us that the two conversion densities discussed above can be related to each other through their associated climatological distributions as

$$\rho\left(\mathbf{x}_f \middle| \mathbf{x}_t\right)\rho\left(\mathbf{x}_t\right) = \rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)\rho\left(\mathbf{x}_f\right). \tag{2.5}$$

We may understand a little bit about the structure of the converse conversion density $\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)$ by solving (2.5):

$$\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right) = w\left(\mathbf{x}_t, \mathbf{x}_f\right)\delta\left(\mathbf{x}_f - \mathbf{F}\left(\mathbf{x}_t\right)\right), \tag{2.6a}$$

where

$$w\left(\mathbf{x}_t, \mathbf{x}_f\right) = \frac{\rho\left(\mathbf{x}_t\right)}{\rho\left(\mathbf{x}_f\right)}. \tag{2.6b}$$

First, equation (2.6a) shows that when **F** does not have an inverse, the density $\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)$ is a weighted collection of Dirac delta functions on the surface defined by fixing $\mathbf{x}_f$ and determining the collection of states $\mathbf{x}_t$ for which $\mathbf{F}\left(\mathbf{x}_t\right) = \mathbf{x}_f$. Note that because **F** only produces states on $\mathbb{A}_f$ we never need evaluate $w\left(\mathbf{x}_t, \mathbf{x}_f\right)$ for values of $\mathbf{x}_f$ for which $\rho\left(\mathbf{x}_f\right) = 0$. Second, as discussed in sub-section 2.a, because the observations in the data assimilation cycles from times j = 1, 2, …, J do not change the set $\mathbb{A}_t$, and the map (2.3) is valid for all $\mathbb{A}_t$ and $\mathbb{A}_f$, we may apply the map (2.3), and its corresponding conversion densities, to all data assimilation cycles, j.

To illustrate the properties of these conversion densities we relate these densities to the two-vector example of sub-section 2.a. Here, we define a forecast (low-resolution) state to be a scalar that arises from a smoothing operator that is an arithmetic mean: $\mathbf{F} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$. We use this operator in (2.3) to define the forecast (scalar) states that are obtained from the high-resolution true (two-vector) states. Equation (2.4) is plotted in figure 3a for an example high resolution state of $\mathbf{x}_t = \begin{bmatrix} 1 & 3 \end{bmatrix}^T$, which implies a low-resolution state of $x_f = 2$ because $\mathbf{F}\mathbf{x}_t = 2$ in this case. Note that the delta function in figure 3a is placed at $x_f = 2$ and has amplitude 10. The amplitude of 10 arises because we have defined integration numerically here as the trapezoidal rule. For the trapezoidal rule the Dirac delta function is inversely proportional to the grid resolution that we are using to represent these densities. Here we have used a grid resolution of 0.1, which implies that the Dirac delta has amplitude 10. As described in the previous section, while $\rho\left(x_f \middle| \mathbf{x}_t = \begin{bmatrix} 1 & 3 \end{bmatrix}^T\right)$ is the Dirac delta function the converse conversion density $\rho\left(\mathbf{x}_t \middle| x_f\right)$ is not. As an example, the converse conversion density, $\rho\left(\mathbf{x}_t \middle| x_f = 1\right)$, has non-zero probability on a line defined by $\mathbf{F}\mathbf{x}_t = 1$ weighted by the climatological distributions through w and is shown in figure 3b.

These conversion densities are useful because they allow one to convert between the true and forecast densities. For example, the prior density on the attracting manifold of the forecast model is:

$$\rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) = \int_{-\infty}^{\infty} \rho\left(\mathbf{x}_f \middle| \mathbf{x}_t\right) \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_t \ . \tag{2.7}$$

The equation (2.7) describes the distribution of forecasts $\mathbf{x}_f$ that obtains from sampling $\rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right)$ for $\mathbf{x}_t$ and using these samples of $\mathbf{x}_t$ in (2.3) to obtain samples of $\mathbf{x}_f$. We may apply equation (2.7) to our simple example problem to find the prior distribution of forecasts for the j = 2 cycle (Figure 4b). Recall that the prior distribution for our simple example problem is the previous (j = 1) posterior, $\rho\left(\mathbf{x}_f \middle| y_1 = 1\right)$. Note that the mode of this low-resolution prior is not the mode of either variable in the high-resolution prior. In fact, the mode of the low-resolution prior is the arithmetic mean of the mode of each high-resolution variable in the high-resolution prior.

The conversion densities are central to our development because through them we are able to build one of the most important components of this work. These conversion densities allow us to show that the forecasts have their own posterior density defined as

$$\rho\left(\mathbf{x}_f \middle| \mathbf{Y}_J\right) = \int_{-\infty}^{\infty} \rho\left(\mathbf{x}_f \middle| \mathbf{x}_t\right) \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_J\right) d\mathbf{x}_t \ , \tag{2.8a}$$

which, upon using (2.2) and (2.4), implies

$$\rho\left(\mathbf{x}_f \middle| \mathbf{Y}_J\right) = C_J \rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right) \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right), \tag{2.8b}$$

where $\rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right)$ is the conditional density for the observation conditioned on the forecast state (which we shall refer to as the *forecast* observation likelihood) and $\rho\left(\mathbf{x}_f \middle| \mathbf{Y}_J\right)$ represents the density that describes the conditional distribution of forecast states given the entire observational record. This notion that the states on the forecast attractor have a posterior density that has been conditioned upon observations of the state on a different attractor is a unique aspect of this work and will allow us to explicitly show how the difference in the two attractors leads to an error of representation.

Equation (2.8b) shows that the data assimilation procedure, starting from the climatological distribution and proceeding for J assimilation steps, applies to the forecast model in so far as one simply replaces the true states, $\mathbf{x}_t$, with the forecast states, $\mathbf{x}_f$, in equations (2.1) and (2.2) to define Bayesian data assimilation on the forecast attractor. It is however important

to recognize that while (2.8b) looks superficially similar to (2.2) it is in fact quite different. This is true for two reasons: 1) the data assimilation procedure for the forecast model begins with the forecast climatological distribution, which may be significantly different from the true climatological distribution, and 2) the forecast observation likelihood is actually very different from the true observation likelihood. This difference between the true observation likelihood and the forecast observation likelihood will be described next.

c) The Forecast Observation Likelihood

Central to understanding the forecast posterior distribution, and the manifestation of representation error within it, is the forecast observation likelihood. The observation likelihood in (2.8b) obtains from application of the chain rule of probability through the use of the conversion density as

$$\rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right) = \int_{-\infty}^{\infty} \rho\left(\mathbf{y}_J \middle| \mathbf{x}_t\right) \rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right) d\mathbf{x}_t \,. \tag{2.9}$$

An important difference between $\rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right)$ and $\rho\left(\mathbf{y}_J \middle| \mathbf{x}_t\right)$ is that while the mean of $\rho\left(\mathbf{y}_J \middle| \mathbf{x}_t\right)$ is the true state $\left(\mathbf{H}\mathbf{x}_t\right)$ and its variance is the instrument error, $\mathbf{R}_i$, this is not true of $\rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right)$. The mean of $\rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right)$ is, after use of (2.9):

$$\mathbf{y}_f\left(\mathbf{x}_f\right) = \int_{-\infty}^{\infty} \mathbf{y}_J \rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right) d\mathbf{y}_J = \mathbf{H}\overline{\mathbf{x}}_c \,, \tag{2.10a}$$

$$\overline{\mathbf{x}}_c\left(\mathbf{x}_f\right) = \int_{-\infty}^{\infty} \mathbf{x}_t \rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right) d\mathbf{x}_t \,, \tag{2.10b}$$

where we will explicitly write this as a state-dependent bias (from the perspective of the forecast model attractor and its associated prior) as

$$\mathbf{y}_f\left(\mathbf{x}_f\right) = \mathbf{H}_f \mathbf{x}_f + \mathbf{b} \,, \tag{2.11}$$

with this bias defined as

$$\mathbf{b} = \mathbf{H}\overline{\mathbf{x}}_c - \mathbf{H}_f \mathbf{x}_f \tag{2.12}$$

and $\mathbf{H}_f$ (p x M) is the observation operator in the space of the forecast model. At this point we will leave the definition and the distinction between $\mathbf{H}$ and $\mathbf{H}_f$ undefined. We emphasize here

however that the situation we will examine is one in which the difference between $\mathbf{H}$ and $\mathbf{H}_f$ is because $\mathbf{H}_f$ operates on a truncated state-vector and is not because $\mathbf{H}_f$ has been approximated or created in error. Nevertheless, we will see subsequently that this difference between these two observation operators will turn out to be a central feature of the analysis and therefore we shall return to discuss their differences at several points below.

Equation (2.10) shows that the observation conditioned on the forecast is biased with respect to the truth (because the observation is of an object on the true attracting manifold and not on the forecast attracting manifold) and that bias is described by the mean of the conversion density. In addition, the variance about the mean state, $\mathbf{y}_f\left(\mathbf{x}_f\right)$, is

$$\begin{aligned}\mathbf{R}_f\left(\mathbf{x}_f\right) &= \int_{-\infty}^{\infty}\left(\mathbf{y}_J-\mathbf{y}_f\right)\left(\mathbf{y}_J-\mathbf{y}_f\right)^T \rho\left(\mathbf{y}_J\middle|\mathbf{x}_f\right)d\mathbf{y}_J \\ &= \mathbf{R}_i+\mathbf{H}\mathbf{P}_c\mathbf{H}^T\end{aligned}, \tag{2.13}$$

where

$$\mathbf{P}_c\left(\mathbf{x}_f\right)=\int_{-\infty}^{\infty}\left(\mathbf{x}_t-\overline{\mathbf{x}}_c\right)\left(\mathbf{x}_t-\overline{\mathbf{x}}_c\right)^T \rho\left(\mathbf{x}_t\middle|\mathbf{x}_f\right)d\mathbf{x}_t\,, \tag{2.14}$$

is the covariance matrix of the conversion density. The equation (2.14) carries the information that relates the forecast model states to the true attracting manifold. The term $\mathbf{H}\mathbf{P}_c\mathbf{H}^T$ is the manifestation of the representation error in the observation covariance matrix and shows that representation error occurs when the function $\mathbf{F}$ does not have an inverse. To see this note that when the function $\mathbf{F}$ has an inverse then $w = 1$ and

$$\rho\left(\mathbf{x}_t\middle|\mathbf{x}_f\right)=\delta\left(\mathbf{x}_f-\mathbf{F}\left(\mathbf{x}_t\right)\right)=\delta\left(\mathbf{x}_t-\mathbf{F}^{-1}\left(\mathbf{x}_f\right)\right), \tag{2.15}$$

which clearly has a variance of zero and hence (2.14) would vanish. Furthermore, note that if we use (2.15) in (2.10) then we obtain $\mathbf{y}_f=\mathbf{H}\mathbf{F}^{-1}\left(\mathbf{x}_f\right)$, which is biased from the perspective of the forecast attractor, where that bias is $\mathbf{b}=\mathbf{H}\mathbf{F}^{-1}\left(\mathbf{x}_f\right)-\mathbf{H}_f\mathbf{x}_f$. We may however remove this bias by defining the observation operator in the space of the forecast model as, $\mathbf{H}_f=\mathbf{H}\mathbf{F}^{-1}$. This definition of the forecast observation operator is novel as it implies that we extend our view of what an observation operator does. The observation operator, $\mathbf{H}_f=\mathbf{H}\mathbf{F}^{-1}$, implies that the forecast observation operator should include a "bias" correction for the particular forecast model

that we are using. In section 4, this definition of the forecast observation operator will be generalized to linear equations (2.3) that do not contain an explicit inverse and subsequently shown to be the proper way to account for representation error. In the meantime, however, we will maintain the general form for $\mathbf{H}_f$ that we have been using thus far.

In order to understand the forecast observation likelihood better we apply our example problem to equation (2.9) to find the low-resolution observation likelihood. This low-resolution observation likelihood is plotted in figure 4a as a function of observation and conditioned on the low resolution state of $x_f = 1$. Because the observation is actually of the high-resolution state the low-resolution observation likelihood is biased (with respect to the low-resolution states), which can be seen by the mode of the distribution not being centered on the low-resolution state of $x_f = 1$. It is actually centered at $x_f = 1/2$, where this bias may be seen as coming from the bias in the forecast climatological distribution whose mean is at $\overline{x}_f = -1/2$. The forecast observation likelihood has a variance of 2, of which ½ of this variance (recall that the instrument error is 1) can be attributed to representation error and equation (2.14). Finally, the low-resolution posterior (2.8a,b) for the observation j = 1and 2 cycles are plotted in figure 4b. The posterior for the j = 1 cycle has its mode at 1/2 and the mode of the posterior for the j = 2 cycle is at 1.2. Again, the mode at each cycle is located at the arithmetic mean of the location of the modes in the true (high resolution) posteriors.

d) Data Assimilation

Because we are interested in doing data assimilation with these concepts we will now derive the minimum error variance estimate of a linear estimator on the forecast models attractor, which, in this context, is the Kalman filter (Kalman 1960) for the states on the forecast attractor. This formula is derived by minimizing the trace of the *expected* posterior forecast covariance matrix about a linear estimator. The expected posterior forecast covariance matrix may be written as:

$$\overline{\mathbf{P}}_a = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\mathbf{x}_f - \overline{\mathbf{x}}_a\right)\left(\mathbf{x}_f - \overline{\mathbf{x}}_a\right)^T \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_J\right) d\mathbf{x}_f \rho\left(\mathbf{y}_J \middle| \mathbf{Y}_{J-1}\right) d\mathbf{y}_J, \tag{2.16}$$

where the analysis update equation, whose "error" variance is being measured, takes the form of a linear estimation equation

$$\overline{\mathbf{x}}_a = \overline{\mathbf{x}}_f + \mathbf{G}\left[\mathbf{v} - \langle \mathbf{v} \rangle\right]. \tag{2.17}$$

where

$$\overline{\mathbf{x}}_f = \int_{-\infty}^{\infty} \mathbf{x}_f \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f = \int_{-\infty}^{\infty} \mathbf{F}\left(\mathbf{x}_t\right) \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_t \,, \tag{2.18}$$

The overbar on $\overline{\mathbf{P}}_a$ in (2.16) is there to make clear that this is the expression for the expected posterior covariance matrix as it has been averaged over all the observations (Please see Hodyss and Campbell (2013) for further discussion). The innovation in (2.17) is $\mathbf{v} = \mathbf{y}_J - \mathbf{H}_f \overline{\mathbf{x}}_f$, which we emphasize uses the observation operator in the space of the forecast model. The expected innovation, $\langle \mathbf{v} \rangle$, in (2.17) is there because the observations with respect to the forecasts are biased because the observations are taken on the true attracting manifold. This de-biasing of the innovation in (2.17) is critical to get the data assimilation to put the analysis at the center of $\rho\left(\mathbf{x}_f \middle| \mathbf{Y}_J\right)$ and therefore the posterior distribution on the desired forecast attractor.

It is important to realize that while the quantities derived in (2.10) through (2.13) are interesting descriptions of the corresponding densities, they are in fact not the ones that would be used in a data assimilation algorithm. This is because as shown in (2.16) those expressions need to be averaged over the forecasts in the derivation of the update equations. This averaging for (2.10) would take the form:

$$\overline{\mathbf{y}}_f = \int_{-\infty}^{\infty} \mathbf{y}_f\left(\mathbf{x}_f\right) \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f = \mathbf{H}_f \overline{\mathbf{x}}_f + \overline{\mathbf{b}} \,, \tag{2.19}$$

where $\overline{\mathbf{b}} = \langle \mathbf{v} \rangle = \mathbf{H}\overline{\mathbf{x}}_t - \mathbf{H}_f \overline{\mathbf{x}}_f$ and

$$\overline{\mathbf{x}}_t = \int_{-\infty}^{\infty} \mathbf{x}_t \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_t \,. \tag{2.20}$$

In addition, the observation error variance that would be used in a data assimilation algorithm is obtained from

$$\overline{\mathbf{R}}_f = \int_{-\infty}^{\infty} \mathbf{R}_f\left(\mathbf{x}_f\right) \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f \,, \tag{2.21}$$

Note that the result of the integration in (2.21) is an observation error variance that is not state-dependent. The derivation of the Kalman gain requires the use of an observation error covariance matrix that is the result of a weighted average over all possible observation error covariance matrices and therefore cannot depend on the state.

Equation (2.21) is interesting because we may use (2.13) in (2.21) to obtain:

$$\bar{\mathbf{R}}_f = \mathbf{R}_i + \int_{-\infty}^{\infty} \mathbf{H}\mathbf{P}_c\mathbf{H}^T \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f \ , \tag{2.22a}$$

where

$$\int_{-\infty}^{\infty} \mathbf{H}\mathbf{P}_c\mathbf{H}^T \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f = \mathbf{H}\mathbf{P}_t\mathbf{H}^T - \mathbf{H}_f\mathbf{P}_f\mathbf{H}_f^T - \mathbf{H}_f\mathbf{P}_{fb} - \mathbf{P}_{fb}^T\mathbf{H}_f^T - \mathbf{P}_{bb} \ , \tag{2.22b}$$

$$\mathbf{P}_t = \int_{-\infty}^{\infty} \left(\mathbf{x}_t - \bar{\mathbf{x}}_t\right)\left(\mathbf{x}_t - \bar{\mathbf{x}}_t\right)^T \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_t \ , \tag{2.22c}$$

$$\mathbf{P}_f = \int_{-\infty}^{\infty} \left(\mathbf{x}_f - \bar{\mathbf{x}}_f\right)\left(\mathbf{x}_f - \bar{\mathbf{x}}_f\right)^T \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f \ , \tag{2.22d}$$

$$\mathbf{P}_{fb} = \int_{-\infty}^{\infty} \left(\mathbf{x}_f - \bar{\mathbf{x}}_f\right)\left(\mathbf{b} - \bar{\mathbf{b}}\right)^T \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f \ , \tag{2.22e}$$

$$\mathbf{P}_{bb} = \int_{-\infty}^{\infty} \left(\mathbf{b} - \bar{\mathbf{b}}\right)\left(\mathbf{b} - \bar{\mathbf{b}}\right)^T \rho\left(\mathbf{x}_f \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_f \ . \tag{2.22f}$$

Equation (2.22b) shows that representation error is the difference between the true covariance, the covariance of the forecasts and the bias, and the covariance matrix of the state-dependent bias all mapped to observation space.

The steps required to perform the minimization of the trace of (2.16) are well known, can be found in Hodyss (2011), and will not be repeated here. The result of this minimization is that the *gain matrix*, **G**, can be written in two equivalent ways:

$$\mathbf{G}_1 = \mathbf{P}_{ft}\mathbf{H}^T\left[\mathbf{H}\mathbf{P}_t\mathbf{H}^T + \mathbf{R}_i\right]^{-1}, \tag{2.23a}$$

$$\mathbf{G}_2 = \left[\mathbf{P}_f\mathbf{H}_f^T + \mathbf{P}_{fb}\right]\left[\mathbf{H}_f\mathbf{P}_f\mathbf{H}_f^T + \mathbf{H}_f\mathbf{P}_{fb} + \mathbf{P}_{fb}^T\mathbf{H}_f^T + \mathbf{P}_{bb} + \bar{\mathbf{R}}_f\right]^{-1}, \tag{2.23b}$$

where

$$\mathbf{P}_{ft} = \int_{-\infty}^{\infty} \left(\mathbf{F}\left(\mathbf{x}_t\right) - \bar{\mathbf{x}}_f\right)\left(\mathbf{x}_t - \bar{\mathbf{x}}_t\right)^T \rho\left(\mathbf{x}_t \middle| \mathbf{Y}_{J-1}\right) d\mathbf{x}_t \ , \tag{2.24}$$

$$\mathbf{P}_{ft}\mathbf{H}^T = \mathbf{P}_f\mathbf{H}_f^T + \mathbf{P}_{fb}\,. \tag{2.25}$$

The first gain matrix, $\mathbf{G}_1$, uses information from the true attractor that is subsequently mapped back to the forecast attractor through the covariance between the states on the two attractors, $\mathbf{P}_{ft}$. This version of the gain is most similar to the traditional Kalman gain whereby the numerator relates the (true attractor's) observation space to the state-space to be updated (forecast attractor) using the covariance (2.24). This version of the gain matrix is of course impossible to implement in practice as it requires use of the true prior distribution.

The second gain matrix, $\mathbf{G}_2$, is novel and only uses the information on the forecast model attractor along with the function **F** to infer the relationship between the true and the forecast attractors. This second gain matrix has two terms in its numerator. The first term is the standard term that maps the observation space to the state space to be updated, but totally from within the forecast attractor. The second term is new and accounts for the possibility that on the forecast attractor there is no covariance between the observation space and state space to be updated but, because the observation is actually of a state on the true attractor and there may be a covariance between the true attractor and the forecast state-space to be updated, there should in fact be a correction at that location. This new term is shown in (2.25) to be precisely the difference between the forecast state-space/true observation-space covariance ($\mathbf{P}_{ft}\mathbf{H}^T$) and the forecast state-space/forecast observation-space covariance ($\mathbf{P}_f\mathbf{H}_f^T$). Equation (2.25) shows that the numerators of (2.23a) and (2.23b) would be identical if the forecast-bias covariance matrix, $\mathbf{P}_{fb}$, would vanish. In section 4 we show how to choose the forecast observation operator to eliminate this forecast-bias covariance matrix.

Last, it is important to realize that the gain matrices (2.23a,b) when used in (2.17) do not in fact provide an estimate of the true state on the true attractor. The gain matrices (2.23a,b) when used in (2.17) find the state estimate that is closest (in the sense of the function **F** and in mean-square) to the *forecast* that obtains from mapping the truth through equation (2.3). Hence, the gain matrices (2.23a,b) push the state estimate from the true attracting manifold onto the forecast attracting manifold and is actually likely to push the state estimate further from the observations than it would have been if we had not altered the numerator and denominator of the gain matrix to account for this error in representation. The benefit from using these new gain

matrices is therefore not their distance from the true attractor but actually the fact that they produce a state estimate near to the forecast attractor, which is likely to produce a better, and more balanced, forecast.

3. Smoothing Operators

Previous work in the representation error literature has examined the situation where the relationship between the states being observed by the observational instruments and that produced by the forecast model are related by a smoothing operator (e.g. Mitchell and Daley 1997a,b, Liu and Rabier 2002, Waller et al. 2013). Following this work we will assume that the forecast model resolution is coarse as compared to the true model resolution, i.e. $M < N$. In addition, we will assume the function **F** is linear but we make no assumptions on the shape of the prior distributions on either attractor.

The function **F** will be assumed to be a linear matrix operator that acts to "smooth" the true state to the resolution of the forecast model, viz.

$$\mathbf{x}_f = \mathbf{S}\mathbf{x}_t , \tag{3.1}$$

where **S** is an M x N matrix whose singular value decomposition results in

$$\mathbf{S} = \mathbf{U}\left[\mathbf{\Lambda}^{1/2} \quad \mathbf{0}\right]\mathbf{V}^T , \tag{3.2}$$

with the left singular vectors contained in **U** (M x M), right singular vectors in **V** (N x N), $\mathbf{\Lambda}$ (M x M) the diagonal matrix of singular values, and **0** [M x (N-M)] the zero matrix. The bracket in (3.2) is a truncation operator and along with $\mathbf{\Lambda}$, which we assume to be either constant (white) or steeply sloped (red), we view as a "smoothing" operator. We attach to this notion of "smoothing" the philosophy prevalent in numerical modelling (e.g. Lilly 1962) in which forecast model output is assumed to represent grid-cell averages of the true state around each nodal point of the model's grid representation of the spatial domain. Note that the statement (3.1) states that the only difference between the forecast and the truth is through the smoothing of that state, which ignores the fact that a truncated model will differ in its cascades of energy and enstrophy (i.e., the nonlinear interaction between scales and their subsequent evolution).

Because **S** is M x N and $M < N$, the matrix **S** will not generally have an inverse. This has important implications for the structure of $\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)$ as it implies that there exists a hyperplane defined by all of the states, $\mathbf{x}_t$, that satisfy (3.1) for a given forecast state $\mathbf{x}_f$. Along this plane, $\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)$ has non-zero probability and this results in a non-zero variance of the converse conversion density, which as shown in equation (2.14) leads to the error of representation. An explicit example of $\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)$ that had non-zero variance was presented in figure 3 and in this case one could see in that figure that the hyperplane alluded to above was reduced to a line in the two-dimensional plane of $\mathbf{x}_t$.

The smoothing operator in (3.2) will lead to a bias in the observation mean (2.19). This bias leads to the expected innovation being

$$\langle \mathbf{v} \rangle = \bar{\mathbf{b}} = \left(\mathbf{H} - \mathbf{H}_f \mathbf{S}\right) \bar{\mathbf{x}}_t . \tag{3.3}$$

The bias results from the difference between the true prior mean and the "smoothed" one obtained after use of (3.1).

Similarly, the use of (3.1) in the *gain matrix*, **G**, (2.23a,b) obtains:

$$\mathbf{G}_1 = \mathbf{S}\mathbf{P}_t \mathbf{H}^T \left[\mathbf{H}\mathbf{P}_t \mathbf{H}^T + \mathbf{R}_i\right]^{-1} = \mathbf{S}\mathbf{G}_t , \tag{3.4a}$$

$$\mathbf{G}_2 = \left[\mathbf{P}_f \mathbf{H}_f^T + \mathbf{P}_{fb}\right]\left[\mathbf{H}_f \mathbf{P}_f \mathbf{H}_f^T + \bar{\mathbf{R}}_f^*\right]^{-1} , \tag{3.4b}$$

where $\mathbf{P}_t$ is the covariance matrix of the true prior,

$$\mathbf{P}_f = \mathbf{S}\mathbf{P}_t \mathbf{S}^T , \tag{3.5a}$$

$$\mathbf{P}_{fb} = \mathbf{S}\mathbf{P}_t \left(\mathbf{H} - \mathbf{H}_f \mathbf{S}\right)^T , \tag{3.5b}$$

$$\bar{\mathbf{R}}_f = \mathbf{R}_i + \left(\mathbf{H} - \mathbf{H}_f \mathbf{S}\right) \mathbf{P}_t \left(\mathbf{H} - \mathbf{H}_f \mathbf{S}\right)^T - \mathbf{P}_{bb} . \tag{3.5c}$$

$$\bar{\mathbf{R}}_f^* = \mathbf{H}_f \mathbf{P}_{fb} + \mathbf{P}_{fb}^T \mathbf{H}_f^T + \mathbf{P}_{bb} + \bar{\mathbf{R}}_f = \mathbf{R}_i + \mathbf{H}\mathbf{P}_t \mathbf{H}^T - \mathbf{H}_f \mathbf{P}_f \mathbf{H}_f^T . \tag{3.5d}$$

In (3.4a) the gain matrix $\mathbf{G}_t$, which is the true gain matrix for the true attractor, is mapped into the forecast space through application of the smoothing matrix **S**. In (3.4b) the gain matrix is calculated using only quantities known from the forecast distributions and the matrix **S**. The quantity $\bar{\mathbf{R}}_f^*$ is an abstraction of the observation error covariance matrix to include all the terms except the forecast covariance matrix in observation space in the denominator of (3.4b). We will

refer to this quantity as the *effective* observation error covariance matrix. The effective observation error covariance matrix reduces to the sum of the instrument error and the difference between the true and forecast covariance matrices in observation space. Note that the difference between the effective observation covariance matrix, $\bar{\mathbf{R}}_f^*$, and the actual observation covariance matrix, $\bar{\mathbf{R}}_f$, is entirely a result of the matrices $\mathbf{P}_{fb}$ and $\mathbf{P}_{bb}$. Given (3.1) and (3.2) it may be shown that the $trace\left(\mathbf{P}_t/N\right) \geq trace\left(\mathbf{P}_f/M\right)$ and therefore the diagonal of $\bar{\mathbf{R}}_f^*$ is equal to or greater than the instrument error, $\mathbf{R}_i$. This fact about smoothing matrices of the form (3.2) implies that high-resolution models should as a general rule have greater variance than low-resolution models.

This matrix $\bar{\mathbf{R}}_f^*$ is important because it makes the connection between the Kalman gains $\mathbf{G}_1$ and $\mathbf{G}_2$. The gain matrix (3.4a) operates on the same innovation as the gain matrix (3.4b). This implies that the innovation variance in the denominator, as calculated both ways, must be equal, viz.

$$\mathbf{H}_f\mathbf{P}_f\mathbf{H}_f^T + \bar{\mathbf{R}}_f^* = \mathbf{H}\mathbf{P}_t\mathbf{H}^T + \mathbf{R}_i \,. \tag{3.6}$$

This relationship may be proven by using (3.5d) in (3.6). This shows that if we define the error of representation as a property of the co-variance matrix of the forecast observation likelihood then one cannot actually deduce it straightforwardly from the innovation variance (e.g., Hollingsworth and Lönnberg 1986, Desroziers et al. 2006). The object that can be deduced directly from the innovation variance is $\bar{\mathbf{R}}_f^*$ and not $\bar{\mathbf{R}}_f$. More discussion of the differences between $\bar{\mathbf{R}}_f^*$ and $\bar{\mathbf{R}}_f$ will be presented in section 4.

4. Gaussian Statistics

In this section we add to the development of section 3 the assumption of Gaussianity to the climatological distributions and linearity in the true and forecast model evolution equations. This implies that the prior distributions of both the forecast and true systems must also be Gaussian. It is important to point out that the assumption of Gaussian error statistics in the

climatological distributions and linear error evolution implies that the sets $\mathbb{A}_t$ and $\mathbb{A}_f$ are no longer bounded in state space as is implied by figure 2 but now extend to include all possible states in their domain.

a) Theory

A simple way to construct a Gaussian problem that is amenable to analysis is through the use of a discrete Fourier series representation. To this end we assert a Gaussian covariance model for the true (high-resolution) states of the form

$$\mathbf{x}_t = \overline{\mathbf{x}}_t + \mathbf{Z}\boldsymbol{\eta}, \tag{4.1}$$

where $\overline{\mathbf{x}}_t$ is an N-vector, $\mathbf{Z}$ is the square-root of the true covariance matrix,

$$\mathbf{P}_t = \mathbf{Z}\mathbf{Z}^T, \tag{4.2}$$

and $\boldsymbol{\eta}$ is an N-vector of random numbers drawn from $N(\mathbf{0},\mathbf{I})$. We construct (4.2) using a sinusoidal basis in which the columns of $\mathbf{E}$ (N x N) contain the sinusoids such that

$$\mathbf{P}_t = \mathbf{E}\boldsymbol{\Gamma}\mathbf{E}^T, \tag{4.3}$$

$\boldsymbol{\Gamma}$ is a diagonal matrix whose i^{th} element of the diagonal is defined from $\Gamma_i = a\,\mathrm{e}^{-\alpha^2 k_i^2}$, k_i is the wavenumber associated with the ith basis function of $\mathbf{E}$, and $a = N \Big/ \sum_{i=1}^{N} \mathrm{e}^{-\alpha^2 k_i^2}$. The parameter α determines the slope of the true spectrum, where large α is associated with a red spectrum and small α is associated with a white spectrum.

We connect the true (high-resolution) states to the forecast (low-resolution) states through a smoother that operates as:

$$\mathbf{S} = \mathbf{E}_L \left[\mathbf{D}^{1/2}\mathbf{T} \quad \mathbf{0} \right] \mathbf{E}^T, \tag{4.4}$$

where $\mathbf{D}$ (M x M) is a diagonal matrix whose diagonal elements are $d_i = \mathrm{e}^{-\beta^2 k_i^2}$, $\mathbf{E}_L$ (M x M) is the low-resolution basis whose columns are also the sinusoids, $\mathbf{T}$ (M x M) is a diagonal matrix with the value $\sqrt{M/N}$ along the diagonal. The matrix $\mathbf{D}$ represents the climatological "model error" on the resolved scales and would be equal to the identity matrix if the forecast model's climate at the resolved scales was identical to the true model's climate at those same scales. The matrix implied by the bracket in (4.4) performs a truncation of the high-resolution basis to the M-

dimensional subspace while the matrix **T** assures that the Fourier coefficients calculated from the high-resolution basis are reweighted consistently with respect to the low-resolution basis.

Equation (4.4) allows for the creation of the low-resolution forecast states from the true states in (4.1) using (3.1). This implies that the forecast error covariance matrix may be written as

$$\mathbf{P}_f = \mathbf{S}\mathbf{P}_t\mathbf{S}^T = \mathbf{E}_L\left[\mathbf{D}^{1/2}\mathbf{T} \quad \mathbf{0}\right]\mathbf{\Gamma}\left[\mathbf{D}^{1/2}\mathbf{T} \quad \mathbf{0}\right]^T \mathbf{E}_L^T. \tag{4.5}$$

Because $\mathbf{\Gamma}$ are the true (high-resolution) eigenvalues, equation (4.5) shows that the forecast covariance matrix would be correct up to its M eigenvalues if the climatological model error **D** could be removed. We show next how to remove this climatological model error by accounting for the error of representation.

Because the true states and the forecast states are Gaussian and their relationship is described by a linear operator, we know that the mean of the converse conversion density $\left[\rho\left(\mathbf{x}_t \middle| \mathbf{x}_f\right)\right]$ is a linear function of the vector it is conditioned upon. This presents us with a direct method to calculate the prediction of the observation by the forecast state:

$$\mathbf{y}_f\left(\mathbf{x}_f\right) = \mathbf{H}\overline{\mathbf{x}}_c = \mathbf{H}\overline{\mathbf{x}}_t + \mathbf{H}\mathbf{G}_p\left[\mathbf{x}_f - \overline{\mathbf{x}}_f\right], \tag{4.6}$$

where

$$\mathbf{G}_p = \mathbf{Z}\left[\mathbf{S}\mathbf{Z}\right]^\dagger = \mathbf{S}^\dagger. \tag{4.7}$$

The superscript † in (4.7) denotes the Moore-Penrose pseudo-inverse (Golub and Van Loan 1996). The pseudo-inverse is constructed by finding the singular value decomposition, viz.

$$\mathbf{S}\mathbf{Z} = \mathbf{E}_L\left[\mathbf{D}^{1/2}\mathbf{T} \quad \mathbf{0}\right]\mathbf{\Gamma}^{1/2}\mathbf{E}^T, \tag{4.8}$$

and then noting that its pseudo-inverse is therefore

$$\left[\mathbf{S}\mathbf{Z}\right]^\dagger = \mathbf{E}\mathbf{\Gamma}^{-1/2}\left[\mathbf{D}^{-1/2}\mathbf{T}^{-1} \quad \mathbf{0}\right]^T \mathbf{E}_L^T. \tag{4.9}$$

The equation (4.9) is the *pseudo*-inverse of (4.8) in the sense that $\mathbf{S}\mathbf{Z}\left[\mathbf{S}\mathbf{Z}\right]^\dagger \mathbf{r} = \mathbf{r}$ but $\left[\mathbf{S}\mathbf{Z}\right]^\dagger \mathbf{S}\mathbf{Z}\mathbf{q} \neq \mathbf{q}$ for arbitrary vectors **r** and **q**. Please see Appendix A for a brief derivation of equation (4.6). It is important to note in (4.7) that the matrix **Z** is cancelled by the pseudo-inverse operation such that the only information required to determine $\mathbf{G}_p$ is the smoothing matrix (4.4). In practice, one would build (4.4) by estimating its components using an archive of truth-forecast pairs as is done in bias-correction algorithms (Glahn and Lowry 1972). Moreover,

this implies that we may view equation (4.6) as simply a bias-correction algorithm that we build into the data assimilation system to account for the fact that the forecast model's estimate of the observation is, in effect, biased.

Equation (4.6) is remarkable in so far as this construction allows us to calculate the important quantities from sections 2 and 3 without the need to explicitly develop the conversion densities. For example, the representation error is therefore:

$$\bar{\mathbf{R}}_f - \mathbf{R}_i = \mathbf{H}\mathbf{E}\boldsymbol{\Theta}\mathbf{E}^T\mathbf{H}^T, \tag{4.10}$$

where $\boldsymbol{\Theta}$ is a diagonal matrix with the value of the diagonal of $\boldsymbol{\Theta}$ satisfying:

$$\Theta_i = \begin{cases} 0, & i = 1,\ldots,M \\ \Gamma_i, & i = M+1,\ldots,N \end{cases}. \tag{4.11}$$

The term on the right-hand side of (4.10) arises from (3.5c). This term clearly shows that the representation error (4.10) is simply the portion of the high resolution spectrum that is missing from the forecast states. Note that for $M = N$, and therefore no truncation, the elements of $\boldsymbol{\Theta}$ vanish and there is no representation error. This is again a result of there being, in that case, an inverse available when $M = N$ and, as discussed in section 2, this implies that the representation error must vanish. This point is important because it shows that the climatological model error on the resolved scales (**D**) is irrelevant to both the existence of representation error and to the structure of the representation error. By contrast, equation (3.5d) for which we are interested in $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ does depend on the climatological model error on the resolved scales (**D**) through equation (4.5).

An important quantity in our development of section 2 was the state-dependent bias of the forecast models estimate of the observation (2.12), **b**, which given (4.6), implies:

$$\mathbf{b} = \left[\mathbf{H}\mathbf{S}^\dagger - \mathbf{H}_f\right]\mathbf{x}_f + \mathbf{H}\left[\bar{\mathbf{x}}_t - \mathbf{S}^\dagger\bar{\mathbf{x}}_f\right]. \tag{4.12}$$

The first term in (4.12) is interesting because it corresponds to the mismatch between the forecast model's estimate of the true observation operator $\mathbf{H}\mathbf{S}^\dagger$ in (4.6) and the observation operator we are actually using $\mathbf{H}_f$. We emphasize here that this mismatch is *not* one in which we are implying that $\mathbf{H}_f$ is incorrect in the sense that if, for example, we had a point measurement that there would be some form of inaccuracy in the interpolation to the observation location. Indeed, even in this case of a point measurement, in which we are assuming we have a perfect

interpolation, the mismatch inferred by (4.12) is between the statistically derived observation operator $\mathbf{HS}^{\dagger}$, which now corresponds to more than just an interpolation, and the operator, $\mathbf{H}_f$.

The equation (4.12) suggests that if we define $\mathbf{H}_f$ such that,

$$\mathbf{H}_f \equiv \mathbf{HS}^{\dagger}, \tag{4.13}$$

we may remove this state-dependent bias term, which subsequently renders $\mathbf{P}_{fb} = \mathbf{0}$ and $\mathbf{P}_{bb} = \mathbf{0}$. This implies that in this data assimilation system the observation operator does *not* simply map the truth to the observation, but rather it maps the forecast to the observation and, because the forecast is in a different portion of state space than the truth, this requires the matrix operator, $\mathbf{HS}^{\dagger}$ rather than just **H**.

One of the most important results of choosing (4.13), and subsequently rendering $\mathbf{P}_{fb} = \mathbf{0}$ and $\mathbf{P}_{bb} = \mathbf{0}$, is that this results in $\bar{\mathbf{R}}_f^* = \bar{\mathbf{R}}_f$ and therefore the choice (4.13) has removed the impact of the climatological model error on the resolved scales, **D**, in the forecast covariance matrix and therefore also in $\bar{\mathbf{R}}_f^*$. This has two important consequences: 1) innovation based techniques (e.g. Hollingsworth and Lönnberg 1986, Desroziers et al. 2006) for the estimation of $\bar{\mathbf{R}}_f^*$ are therefore self-consistent estimators of the representation error *only when the choice (4.13) has been implemented* and 2) as we show next this provides a direct way to remove the deleterious impact of the climatological model error on the state estimate from the data assimilation algorithm.

Subsequently, by employing (4.13) it can be shown that the gain matrix written in terms of the forecast quantities (3.4b) is now:

$$\mathbf{G}_2 = \mathbf{P}_f \mathbf{S}^{\dagger T} \mathbf{H}^T \left[\mathbf{HS}^{\dagger} \mathbf{P}_f \mathbf{S}^{\dagger T} \mathbf{H}^T + \bar{\mathbf{R}}_f^* \right]^{-1}. \tag{4.14}$$

This new gain matrix is the most important result of this manuscript. The calculation of the operator $\mathbf{G}_p$ has allowed for the creation of a new observation operator $\mathbf{HS}^{\dagger}$ that represents the forecast model's estimate of the observation of states on the true attractor. Subsequently, this has allowed the calculation of the correct *numerator* in (4.14) that represents the covariance between the states on the forecast attractor and the observations of the true attractor. The equation (4.13) results in

$$\mathbf{H}_f\mathbf{P}_f\mathbf{H}_f^T = \mathbf{H}\mathbf{S}^\dagger\mathbf{P}_f\mathbf{S}^{\dagger T}\mathbf{H}^T = \mathbf{H}\mathbf{E}\begin{bmatrix}\mathbf{\Gamma}_M & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{E}^T\mathbf{H}^T, \tag{4.15a}$$

$$\langle\mathbf{v}\rangle = \mathbf{H}\mathbf{E}\begin{bmatrix}\mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\mathbf{E}^T\overline{\mathbf{x}}_t, \tag{4.15b}$$

where $\mathbf{\Gamma}_M$ denotes the first M eigenvalues of $\mathbf{\Gamma}$ and $\mathbf{0}$ is the zero matrix. Equation (4.15a) shows that the choice (4.13) results in $\mathbf{H}_f\mathbf{P}_f\mathbf{H}_f^T$ being correct up to the resolution of the forecast model in the sense that the climatological model error denoted by $\mathbf{D}$ has been removed. The equation (4.15b), which describes the bias in the innovation owing to the unresolved scales of motion, is a result of the scales in the true prior mean that are missing in the forecast prior mean.

b) Spatially Extended Example

We now explore the theory discussed above for the spatially extended problem described in this section by employing some simple example problems. We will not re-define the observation operator in order to clearly show the differences between $\overline{\mathbf{R}}_f - \mathbf{R}_i$ and $\overline{\mathbf{R}}_f^* - \mathbf{R}_i$, which will underscore the importance of applying (4.13), as it was already proven above to remove this difference. We will define $\mathbf{H}$ to be the operator that observes each point of the low-resolution (forecast) state. Hence, in these experiments $\mathbf{H}_f$ will simply be the identity operator for the point measurements we have available. We emphasize again that employing (4.13) would lead to an observation operator for these point measurements that is not the identity operator even though these are point measurements, which as we have proven above corrects the problems to be illustrated next.

In order to maintain a connection with previous work (e.g. Mitchell and Daley 1997a,b, Liu and Rabier 2002, Waller et al. 2013) we begin by defining the low-resolution (forecast) states as obtained from a smoother that operates as a pure spectral truncation such that we set $\beta = 0$ in (4.4), which obtains $\mathbf{D} = \mathbf{I}$ and

$$\mathbf{S} = \mathbf{E}_L\begin{bmatrix}\mathbf{T} & \mathbf{0}\end{bmatrix}\mathbf{E}^T. \tag{4.17}$$

We set $\alpha = 1/12$ and plot the central column of $\mathbf{P}_t$ in figure 5a, which represents the one-point covariance function for the point $x = \pi$. The length of the state-vector for the high-resolution (true) states will be $N = 256$. In figure 5a we also plot the central column of $\mathbf{P}_f$ for two different

truncation matrices **S**: one is for $M = 16$ and the other is for $M = 8$. In figure 5b is shown the corresponding central column of **S**, which maps the high resolution (true) states to the low-resolution (forecast) states. The smoothing functions in figure 5b show that truncating the true spectrum results in smoothing kernels that average the true state globally to produce the local low-resolution (forecast) state. In addition, note that the smoothing kernel weights the true state both positively and negatively, which we will see shortly results in long distance negative correlations. Panels 5c and 5d show the one-point covariance function (again for the central column) for $\bar{\mathbf{R}}_f - \mathbf{R}_i$ and $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ for the $M = 16$ and M = 8 cases. In both cases, the representation error $\bar{\mathbf{R}}_f - \mathbf{R}_i$ and the effective representation error $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ are very nearly equal in the center of the domain but differ in the far-field. We will show by contrast below that this equality between these two matrices is due to the pure truncation in this case. When we invoke a Gaussian smoother (i.e. $\mathbf{D} \neq \mathbf{I}$), which as noted above corresponds to a climatological model error on the resolved scales, $\bar{\mathbf{R}}_f - \mathbf{R}_i$ and $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ will become substantially different. Also, note that there exists a strong far-field positive-negative wave pattern in $\bar{\mathbf{R}}_f - \mathbf{R}_i$ and less so for $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$. This is due to the aforementioned positive-negative weightings in the smoothing kernels of figure 5b. An example of a randomly constructed high-resolution (true) state (Equ. 4.1) is shown in figure 6. Using equation (4.17) we may produce the corresponding low-resolution (forecast) state from the $M = 8$ forecast state shown in figure 6. Also, shown in figure 6 is the estimate of the true state, $\bar{\mathbf{x}}_c$, given the low-resolution forecast state from equation (4.6) using the $M = 8$ forecast state shown in that figure. This "best estimate" from (4.6) makes use of the linear regression equation in (4.6) but does not make use of the observation operator, **H**. In this sense, this "best estimate" is simply a state-dependent bias correction of the forecast.

The next set of experiments will make use of the matrix **D**. Here, we study (4.4) for $\beta = 1/6$, which implies a Gaussian smoother on the degrees of freedom that are retained *after* truncation. The same true covariance model is used in this experiment, $\mathbf{P}_t$, for which the one-point covariance function is repeated in figure 7a. Again, in figure 7a we also plot the central column of $\mathbf{P}_f$ for two different truncation matrices **T**: one is for $M = 256$ and the other is for $M = 16$. Note that the $M = 256$ case corresponds to no truncation at all as the forecast state vector and the true state vector are equal ($M = N = 256$). The $M = 256$ case provides an example of a

forecast model error that does not arise from a reduction (truncation) of the number of degrees of freedom in the forecast model. The smoothing kernels for these two cases are shown in figure 7b. The truncated smoothing kernel ($M = 16$) shows the positive-negative oscillations in the far-field that we saw previously. Contrast this with the non-truncated smoothing kernel ($M = 256$) that is completely local.

The application of these smoothing kernels produces distinctly different responses in $\bar{\mathbf{R}}_f - \mathbf{R}_i$ and $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$. For the $M = 256$ case, the $\mathbf{S}$ is full-rank, has an inverse, and results in $\bar{\mathbf{R}}_f - \mathbf{R}_i = \mathbf{0}$ as shown in figure 7c. However, because the smoothing matrix $\mathbf{S}$ still produces a difference between $\mathbf{P}_t$ and $\mathbf{P}_f$, which implies that $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ is non-zero as shown in figure 7c. Hence, when the forecast model is full-rank, but fails to reproduce the climatology of the true attractor, the representation error nonetheless vanishes. This type of climatological error in the model must still be accounted for using $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ and $\mathbf{G}_p$. In the case where $M = 16$, which implies both a Gaussian smoothing and a truncation, one can see in figure 7d that both representation error and $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ are non-zero and quite different. This difference between $\bar{\mathbf{R}}_f - \mathbf{R}_i$ and $\bar{\mathbf{R}}_f^* - \mathbf{R}_i$ illustrates that one can only estimate the representation error from innovation statistics when (4.13) is implemented because this is the only way to render $\mathbf{P}_{fb} = \mathbf{0}$ and $\mathbf{P}_{bb} = \mathbf{0}$.

5. Summary and Conclusions

A framework has been presented to understand the origin of the representation error as well as to properly frame attempts at estimating and accounting for its effects. We have extended the work of Kalman (1960), whose data assimilation algorithm is optimal for Gaussian systems for which the flaws in the model were accounted for using a white noise forcing, to the case where the observations are of states on a true "attractor" and the model evolution produces states on a forecast "attractor" with both states Gaussian distributed and a linear map existing between them, but with no applied white noise forcing. In practical terms, when the distributions

are not Gaussian and the mapping between the two attractors is not linear, we have derived the best linear, unbiased estimation technique. For this case we have shown that in this data assimilation algorithm the observation operator does *not* simply map the truth to the observation, but rather it maps the forecast to the observation and because the forecast is in a different portion of state space than the truth this requires a modified observation operator. We emphasize that the operation of this new data assimilation framework only requires the prior distribution on the forecast model attractor and the function **F,** and does not need the prior distribution on the true attractor, to correctly assimilate observations of states on the true attractor. We view this map in equation (2.3) much like that of the algorithms in the ensemble post-processing and bias-correction literature (e.g. Glahn and Lowry 1972, Raftery et al. 2005) in which climatological information is used to build relationships between forecasts and observations. As such, this new framework has shown how to properly include the information normally obtained from "bias-correction" algorithms within the data assimilation system.

As discussed in the introduction, the notion that the true state might not be the best initial condition for the flawed forecast model has lead us to choose to develop a data assimilation system that attempts to produce a state estimate on the forecast attractor. This obviously produces a state estimate that may not be as close to the truth as is required in some applications. Note however that we may use the state estimation procedure described in Appendix A to map our forecast back to the true attractor in order to produce a state estimate on the true attractor. This of course is nothing more than the well-known post-processing of a forecast but using the infrastructure that we have already developed for the data assimilation algorithm.

In any event, this new framework shows that the representation error arises from the lack of an inverse in the relationship between the true attracting manifold and that of the forecast models. This lack of an inverse in their relationship results when the forecast model is a truncated representation of the true states. In other words, representation error does not occur when the forecast model is simply in error in its representation of climatology. The error that the model must make for representation error to exist is one in which the forecast model has been truncated to represent fewer degrees of freedom than the number of degrees of freedom describing the true states. In this specific case, the forecast observation likelihood [$\rho\left(\mathbf{y}_J \middle| \mathbf{x}_f\right)$] will have a variance greater than that of the error of the instrument and is also likely to have a correlation between observations. In the Gaussian case, it was shown explicitly that this inflation

and correlation structure arises as the covariance matrix of the portion of the true spectrum that goes missing from the truncated forecast model. Lastly, it was shown that innovation based techniques (e.g. Hollingsworth and Lönnberg 1986, Desroziers et al. 2006) for the estimation of representation error covariance matrices are self-consistent estimators of the representation error *only when the observation operator has been modified to account for the attractor differences*.

Applying this new framework to specific problems in the geosciences will require estimation of the map between the true attractor and that of the forecast model (Equ. 2.3). We imagine a proxy for such a model could be developed from the difference between high-resolution and low-resolution simulations. After development of the map (2.3) one must develop the observation gain ($\mathbf{G}_p$) for each observation in which one is interested in accounting for errors in representation. We suggest performing this step using an observation-by-observation approach as this will likely lead to a reduction in the size of the matrices in the calculation (4.7). Research into performing this estimation of **S** is underway and will be reported in a sequel.

Acknowledgements
We gratefully acknowledge support from the Chief of Naval Research PE-0601153N and from the NERC National Centre for Earth Observation in the UK and the European Space Agency. We thank Jeff Anderson for a thorough reading and insightful comments on this work.

Appendix A

The Moore-Penrose Inverse and the Minimum Variance Estimate

We begin with (3.1) and decompose it into prior mean and perturbation:

$$\overline{\mathbf{x}}_f = \mathbf{S}\overline{\mathbf{x}}_t\,,\ \boldsymbol{\varepsilon}_f = \mathbf{S}\boldsymbol{\varepsilon}_t\,, \tag{A.1a,b}$$

where $\boldsymbol{\varepsilon}_f = \mathbf{x}_f - \overline{\mathbf{x}}_f$ and $\boldsymbol{\varepsilon}_t = \mathbf{x}_t - \overline{\mathbf{x}}_t$. Equation (4.1) tells us that

$$\boldsymbol{\varepsilon}_t = \mathbf{Z}\boldsymbol{\eta}\,, \tag{A.2}$$

where $\boldsymbol{\eta}$ is a N-vector whose elements are drawn from N(0,1). Therefore, we attempt to minimize the cost function

$$J(\boldsymbol{\eta}) = \frac{1}{2}\left[\boldsymbol{\varepsilon}_f - \mathbf{SZ}\boldsymbol{\eta}\right]^T\left[\boldsymbol{\varepsilon}_f - \mathbf{SZ}\boldsymbol{\eta}\right], \tag{A.3}$$

to find that the minimum of (A.3) is defined by an infinite number of solutions of the form:

$$\hat{\boldsymbol{\eta}} = \left[\mathbf{SZ}\right]^{\dagger} \boldsymbol{\varepsilon}_f + \mathbf{E}_2 \boldsymbol{\xi}, \tag{A.4}$$

where

$$\left[\mathbf{SZ}\right]^{\dagger} = \mathbf{E}\boldsymbol{\Gamma}^{-1/2}\left[\mathbf{D}^{-1/2}\mathbf{T}^{-1} \quad \mathbf{0}\right]\mathbf{E}_L^T, \tag{A.5}$$

$$\mathbf{E} = \left[\mathbf{E}_1 \quad \mathbf{E}_2\right], \tag{A.6}$$

and $\boldsymbol{\xi}$ is a vector of length (*N-M*) that is composed of random noise with the property that it is a random draw from $N\left(\mathbf{0}, \boldsymbol{\Gamma}_{N-M}\right)$ with $\boldsymbol{\Gamma}_{N-M}$ denoting the last (*N-M*) elements of the diagonal of $\boldsymbol{\Gamma}$. In (A.6) $\mathbf{E}$ has been sub-divided such that we define $\mathbf{E}_1$ as the first *M* columns of $\mathbf{E}$ and then place the remaining columns into $\mathbf{E}_2$. We choose the best solution from the set (A.4) by requiring the solution at the minimum of (A.3) to also minimize $\boldsymbol{\varepsilon}_t^T \mathbf{P}_t^{-1} \boldsymbol{\varepsilon}_t$, which may be shown to imply that $\boldsymbol{\eta}^T \boldsymbol{\eta}$ is also a minimum. The addition of this constraint chooses $\boldsymbol{\xi} = \mathbf{0}$, which then defines the solution as

$$\hat{\boldsymbol{\varepsilon}}_t = \mathbf{Z}\left[\mathbf{SZ}\right]^{\dagger} \boldsymbol{\varepsilon}_f. \tag{A.7}$$

Equation (A.7) defines the minimum variance estimate $\boldsymbol{\varepsilon}_t$ under the constraint that $\boldsymbol{\varepsilon}_t^T \mathbf{P}_t^{-1} \boldsymbol{\varepsilon}_t$ is also a minimum. Hence, (A.7) finds the minimum variance estimate of (A.1a,b) for the state $\boldsymbol{\varepsilon}_t$ given $\boldsymbol{\varepsilon}_f$. The minimum variance estimate of a linear estimator will be equal to the mean of the conversion density when that density is Gaussian. When that density is not Gaussian it then reduces to the best *linear,* unbiased estimate of the mean of the conversion density.

References

Daley, R. 1993: Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.* **11**, 634-647.

Desroziers, G., L. Berre, B. Chapnik, P. Poli, 2006: Diagnosis of observation, background and analysis-error statistics in observation space. *Q. J. R. Meteorol. Soc.*, **126**, 3385-3396.

Frehlich, R. and R. Sharman, 2004: Estimates of turbulence from numerical weather prediction output with applications to turbulence diagnosis and data assimilation. *Mon. Wea. Rev.*, **132**, 2308-2324.

Frehlich, R., 2006: Adaptive data assimilation including the effect of spatial variations in observation error. *Q. J. R. Meteorol. Soc.*, **132**,1225-1257.

Frehlich, R., 2008: Atmospheric turbulence component of the innovation covariance. *Q. J. R. Meteorol. Soc.*, **134**, 931-940.
Frehlich, R., 2011: The definition of 'truth' for numerical weather prediction error statistics. *Q. J. R. Meteorol. Soc.*, **137**,84-98.
Glahn, H.R. and D.A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203-1211.
Golub, G. and C. F. Van Loan, 1996: Matrix Computations. The Johns Hopkins University Press, 687 pgs.
Hodyss, D., 2011: Ensemble state estimation for nonlinear systems using polynomial expansions in the innovation. *Mon. Wea. Rev.*, **139**, 3571-3588.
Hodyss, D. and W.F. Campbell, 2013: Square root and perturbed observation ensemble generation techniques in Kalman and quadratic ensemble filtering algorithms. *Mon. Wea. Rev.*, **141**, 2561-2573.
Hodyss, D., J. G. McLay, J. Moskaitis, E. A. Serra, 2014: Inducing tropical cyclones to undergo Brownian motion: A comparison between Itô and Stratonovich in a numerical weather prediction model. *Mon. Wea. Rev.*, **142**, 1982-1996.
Hollingsworth, A. and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field., *Tellus*, **38A**, 111-136.
Janjic, T. and S.E. Cohn, 2006: Treatment of observation error due to unresolved scales in atmospheric data assimilation. *Mon. Wea. Rev.*, **134**, 2900-2915.
Kalman, R.E., 1960: A new approach to linear filtering and prediction problems. *J. Basic. Eng.*, **82**, 35-45.
Liu, Z.-Q. and F. Rabier, 2002: The interaction between model resolution, observation resolution and observation density in data assimilation" A one-dimensional study. *Q. J. R. Meteorol. Soc.*, **128**, 1367-1386.
Mitchell, H.L. and R. Daley, 1997a: Discretization error and signal/error correlation in atmospheric data assimilation (I). All scales resolved. *Tellus*, **49A**, 32-53.
Mitchell, H.L. and R. Daley, 1997b: Discretization error and signal/error correlation in atmospheric data assimilation (II). The effect of unresolved scales. *Tellus*, **49A**, 54-73.
Oke, P.R. and P. Sakov, 2008: Representation error of oceanic observations for data assimilation. *J. Atmos. Oceanic Technol.*, **25**, 1004-1017.
Petersen, D.P. and D. Middleton, 1963: On representative observations. *Tellus*. **15**, 387-405.
Richman, J.G., R.N. Miller, Y.H. Spitz, 2005: Error estimates for assimilation of satellite sea surface temperature data I ocean climate models. *J. Geophys. Res.*, **32**, L18608, doi:10.1029/2005GL023591.
Waller, J.A., S.L. Dance, A.S. Lawless, N.K. Nichols, and J.R. Eyre, 2013: Representivity error for temperature and humidity using the Met Office high-resolution model. *Q. J. R. Meteorol. Soc.*, DOI: 10.1002/qj.2207.

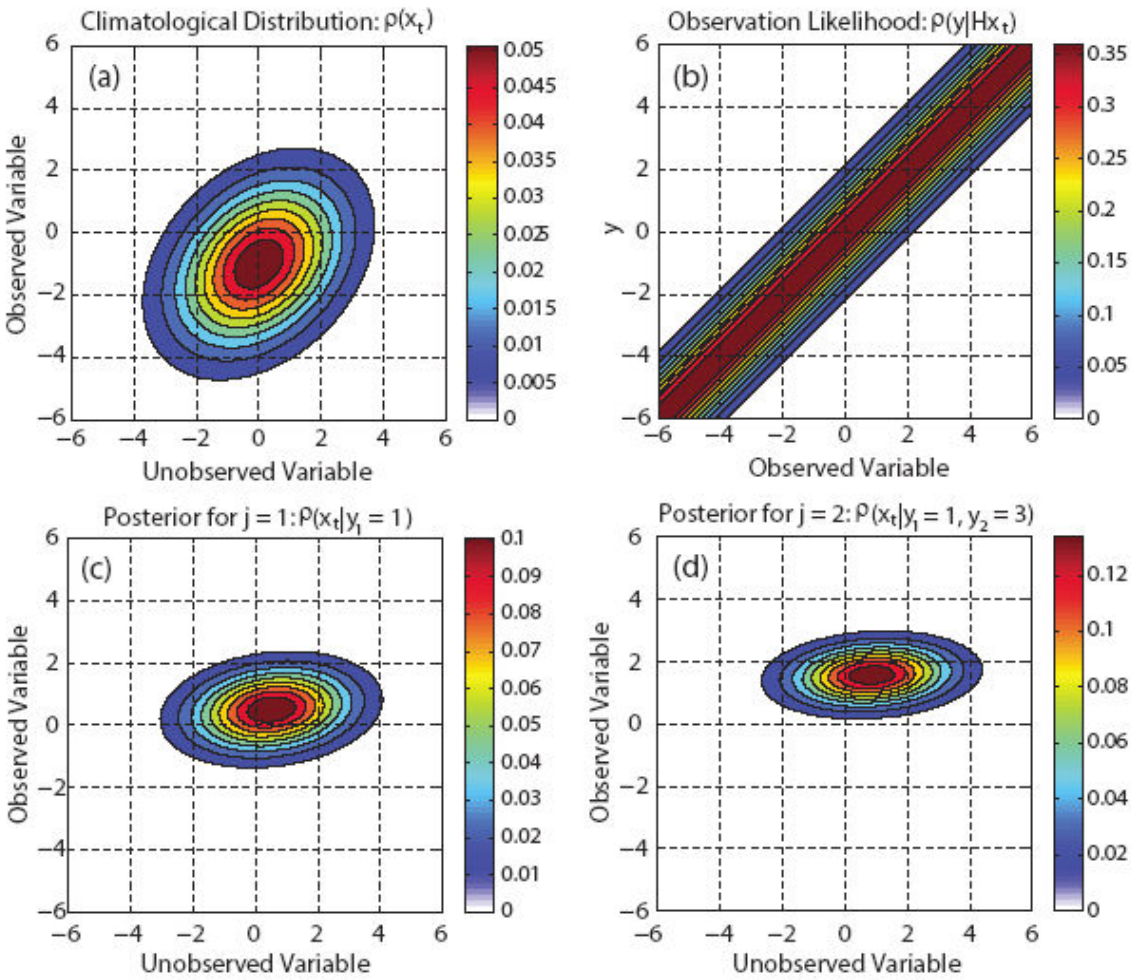


Figure 1. Data assimilation on the high-resolution (true) attractor. (a) High-resolution climatological distribution, (b) high-resolution observation likelihood, (c) high-resolution posterior for $j = 1$, and (d) high-resolution posterior for $j = 2$.

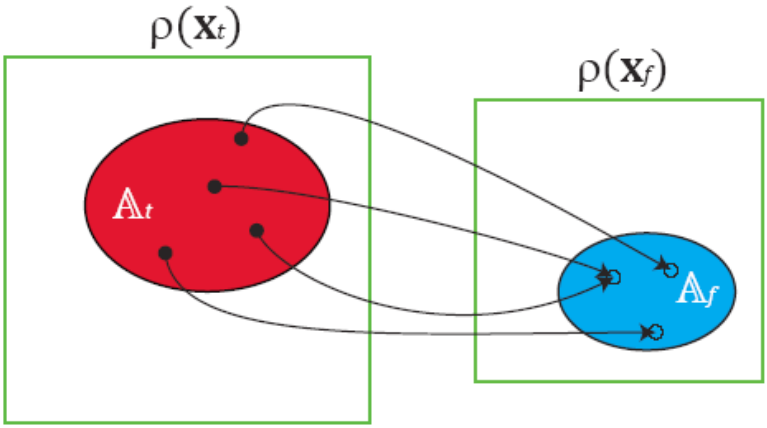


Figure 2. The attractors and their map. The two green squares represent the domain of the true states (large square) and the forecast states (small square). The region for which $\rho(\mathbf{x}_t) > 0$ $\left(\rho(\mathbf{x}_f) > 0\right)$ is denoted as the red (blue) shaded region. An example of the function in equation (2.3) is denoted by the arrows, which travel from states in $\mathbb{A}_t$ denoted by filled circles to states in $\mathbb{A}_f$ denoted by open circles. Note that multiple states in $\mathbb{A}_t$ may map to the same point in $\mathbb{A}_f$. We will show that this results in an error of representation.

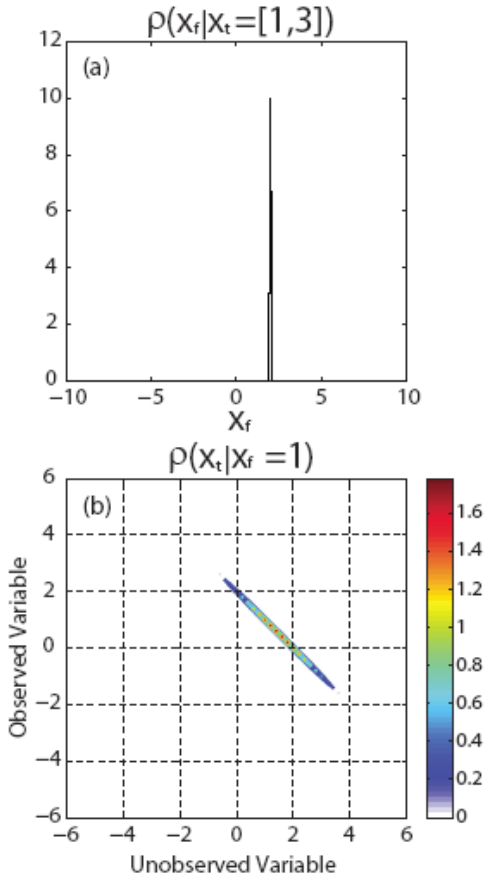


Figure 3. The conversion densities. In (a) is the conversion density describing the distribution of forecast states given a state on the true attractor ($\mathbf{x}_t = [1 \quad 3]^T$) and in (b) is the conversion density describing the distribution of true states given a state on the forecast attractor ($x_f = 1$).

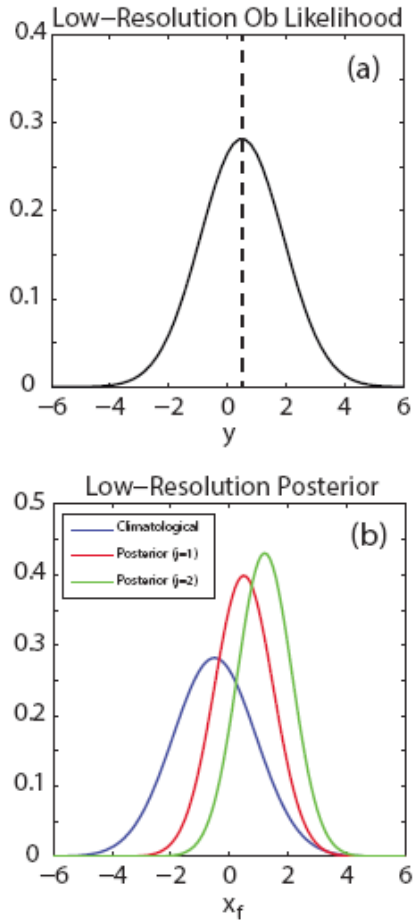


Figure 4. Data assimilation on the low-resolution (forecast) attractor. The low-resolution a) observation likelihood and b) climatological and posterior distributions for j = 1 and 2.

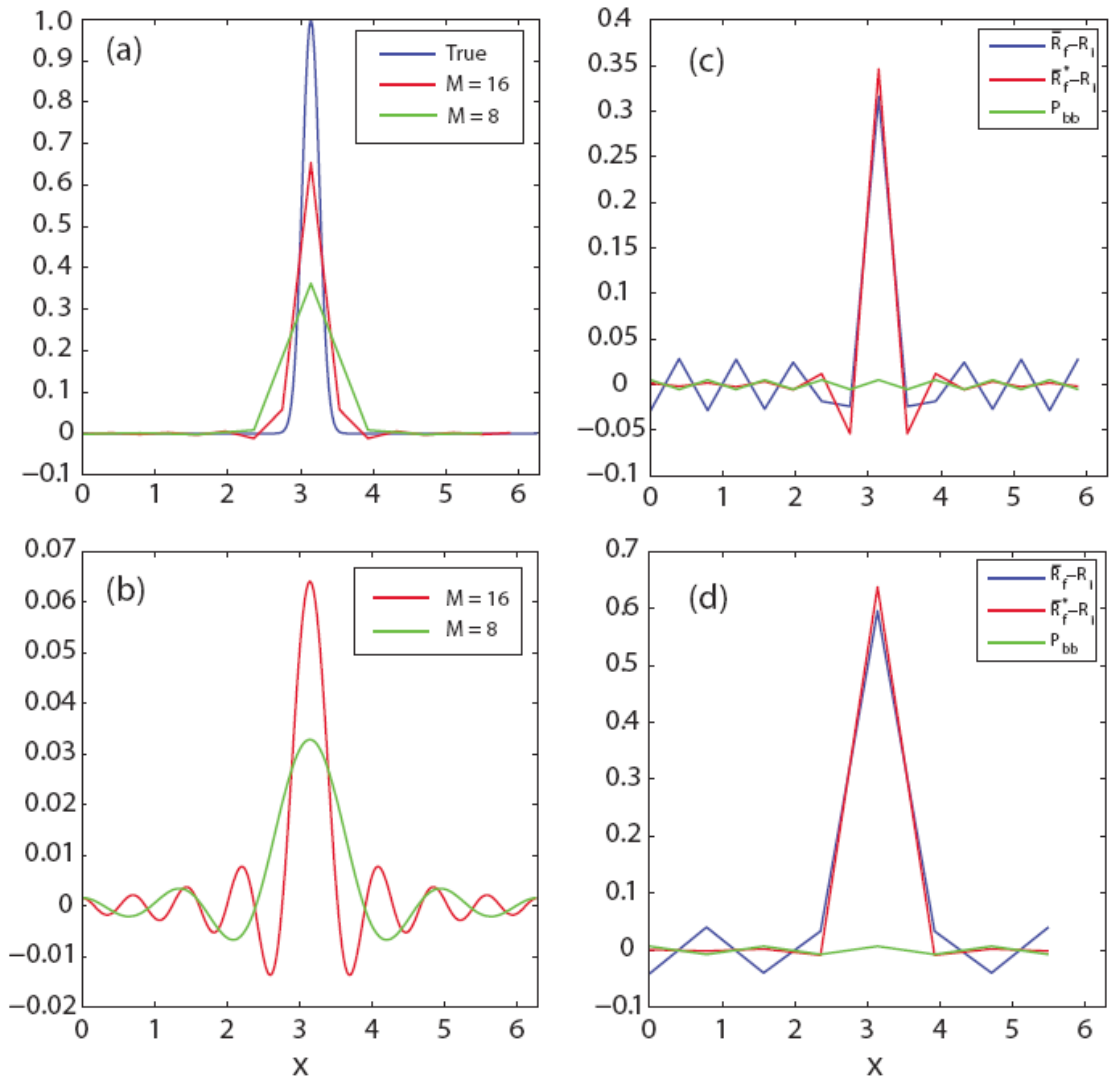


Figure 5. Pure spectral truncation. In (a) are three one-point prior covariance functions: blue is the high-resolution (true) covariance model, red is the $M = 16$ low-resolution (forecast) covariance model, green is the $M = 8$ low-resolution (forecast) covariance model. In (b) are one column of the smoother matrix (4.17) for $M = 16$ (red) and $M = 8$ (green). In (c) and (d) is shown the main components of the theory for $M = 16$ and $M = 8$, respectively: blue is the representation error covariance ($\bar{\mathbf{R}}_f - \mathbf{R}_i$), red is the *effective* representation error covariance ($\bar{\mathbf{R}}^*_f - \mathbf{R}_i$), and green is the bias covariance matrix ($\mathbf{P}_{bb}$).

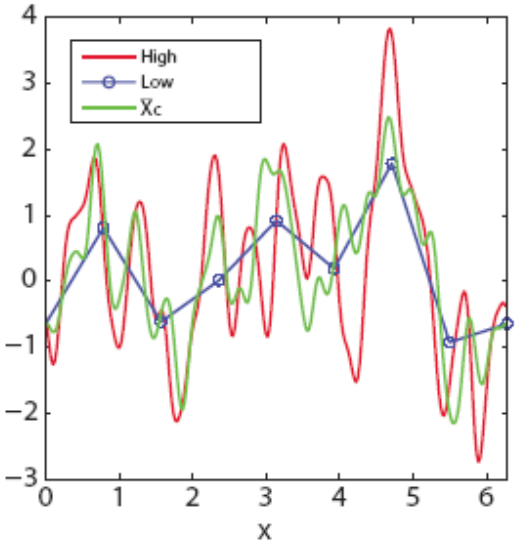


Figure 6. Example high-resolution state, low-resolution state (M = 8), and mean of the conversion density, which is labeled above as the "best estimate". The low-resolution state is defined only at the grid-points denoted by the open circles.

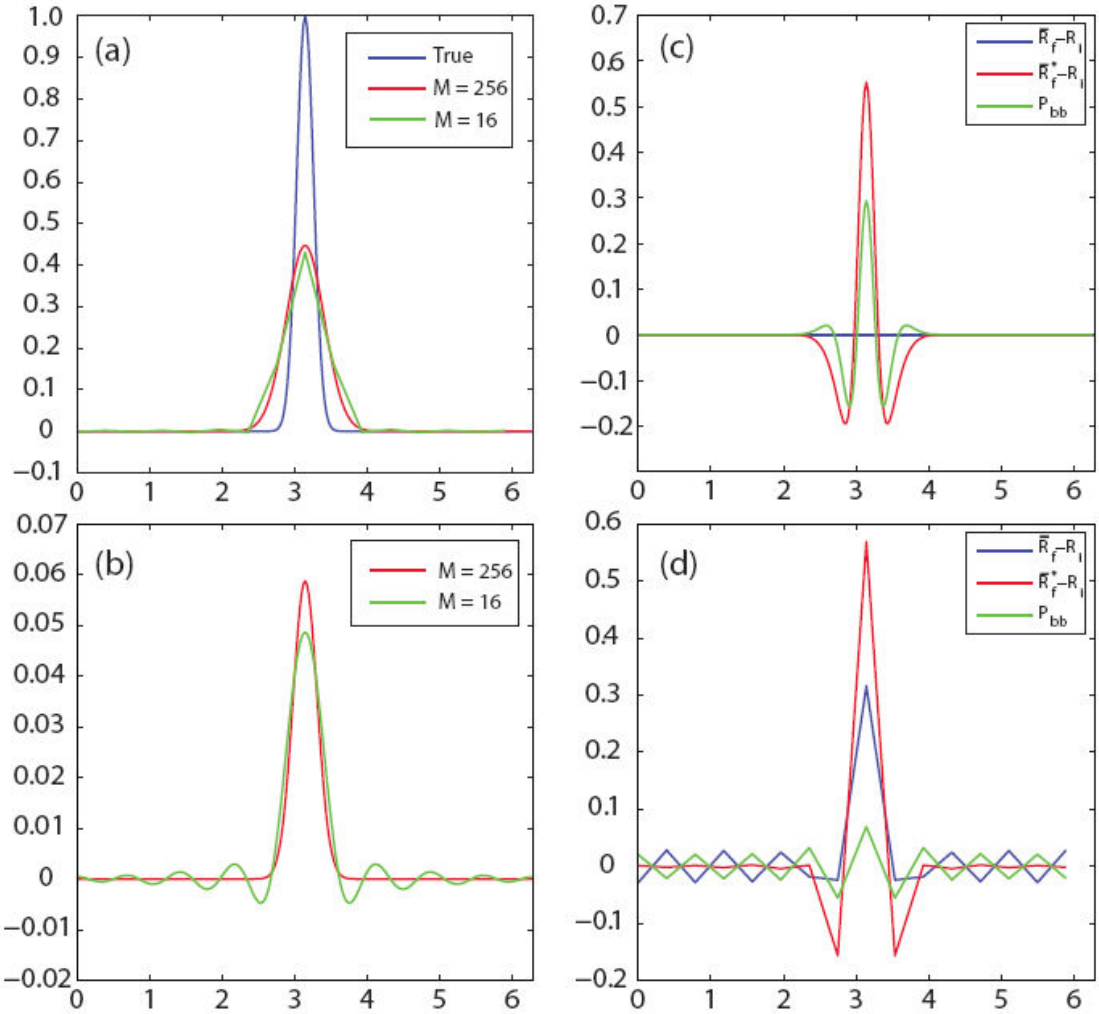


Figure 7. Gaussian Smoothing. In (a) are three one-point prior covariance functions: blue is the high-resolution (true) covariance model, red is the M = 256 low-resolution (forecast) covariance model, green is the M = 16 low-resolution (forecast) covariance model. In (b) are one column of the smoother matrix for M = 256 (red) and M = 16 (green). In (c) and (d) is shown the main components of the theory for M = 256 and M = 16, respectively: : blue is the representation error covariance ($\bar{\mathbf{R}}_f - \mathbf{R}_i$), red is the *effective* representation error covariance ($\bar{\mathbf{R}}^*_f - \mathbf{R}_i$), and green is the bias covariance matrix ($\mathbf{P}_{bb}$).