State and Parameter Estimation in Stochastic Dynamical Models

Timothy DelSole

George Mason University, Fairfax, VA, USA and
Center for Ocean-Land-Atmosphere Studies, Calverton, MD, USA
delsole@cola.iges.org

ABSTRACT

We review state and parameter estimation in the context of ensemble data assimilation methods. A common approach to estimating model parameters is to augment the state vector with the model parameters, and then to apply a data assimilation system to the resulting augmented vector. We show that this method proves very effective for estimating deterministic parameters that affect the mean forecast, but fails utterly for estimating stochastic parameters that affect the forecast variance. We propose a new approach based on Generalized Maximum Likelihood Estimation (GMLE) theory and show that this approach can estimate stochastic parameters in low dimensional, nonlinear, stochastic dynamical models.

1 Introduction

It has recently been shown that introducing parameterizations that are partly stochastic in coupled atmosphere-ocean models can enhance forecast skill [1, 2, 3]. However, such parameterizations raise the important question of how to estimate the parameters within the parameterizations. In general, the most rigorous approach is to estimate these parameters within a comprehensive data assimilation system. There are at least two approaches to doing this: augmentation methods [4] and adjoint methods [5]. It turns out that neither of these approaches are capable of estimating parameters that control the variance of a stochastic parameterization. The reasons for this shortcoming will be discussed in sec. 2. A new method for estimating stochastic parameters based on Generalized Maximum Likelihood Estimation is proposed and applied to simple low-order models in sec. 4. The relation between GMLE, the augmentation method, and adjoint parameter estimation is discussed in sec. 5. We conclude with a summary and discussion of our results.

The material reviewed here has been developed in collaboration with Dr. Xiaosong Yang. Most of the results of this review have appeared previously in [6] and [7], to which we direct the reader for more details. The only new result here is the demonstration in sec. 5.1 that Generalized Maximum Likelihood Estimation is effectively equivalent to the augmentation method when the model contains no stochastic parameters, thereby providing a formal rational for this ad hoc method.

2 State and Parameter Estimation

Let $\mathbf{x}_t$ be the estimated state vector at time t, and let the model parameters estimated at time t be collected into the vector β_t. A very common approach to estimating model parameters β_t is to augment the state vector with the unknown parameters and then apply the resulting augmented vector to a data assimilation system [8, 9, 10, 4, 11]. That is, we define the new "state vector"

$$\mathbf{z}_t = \begin{pmatrix} \mathbf{x}_t \\ \beta_t \end{pmatrix}, \tag{1}$$

and then estimate $\mathbf{z}_t$ using a standard data assimilation system. This approach requires defining an evolution model for β_t, which in most applications is assumed the "persistence model"

$$\beta_t = \beta_{t-1}. \tag{2}$$

Other parameter update models include random white noise and first order autoregressive models [12].

The augmentation method has been proven to be a convenient and effective way to estimate certain kinds of model parameters. A spectacular example is presented in [7], where the augmentation method is used to estimate the state and parameters of a modified form of the Lorenz-96 model [13]

$$\frac{dx_i}{dt} = (x_{i+1} - x_{i-2})\, x_{i-1} - \frac{x_i}{1+d_i} + 8 + f_i, \tag{3}$$

where $i = 1, 2, \ldots, 40$, with cyclic boundary conditions, and the "true" values of f_i and d_i are chosen randomly. Since the model is nonlinear, the covariances are not easily computable and hence we employ ensemble methods. However, the parameter d_i is multiplicative, and it turns out that applying the augmentation method to multiplicative parameters in the context of an ensemble filter is problematic. The reason for this difficulty is that multiplicative parameters change the dynamical properties of the model, and in particular can cause the model to become dynamically unstable for some ensemble members. Such model instability can be avoided if the usual persistence model for parameter update is replaced by a temporally smoothed version of the update model

$$\beta_t^f = \alpha \beta_{t-1}^f + (1-\alpha)\beta_{t-1}^a, \tag{4}$$

where α is a "smoothing" parameter, and superscripts f and a denote the forecast and analysis values, respectively. By smoothing the prediction of the β_t^f in time, "wild" excursions into unstable regions of parameter space are avoided.

To illustrate the method, we solve (3) numerically for a specific initial condition and call this "truth." The "observations" are generated by adding Gaussian white noise with zero mean and unit variance to this truth. Note that there are no observations of the model parameters– all information about the parameters are derived from observations of the state. In these experiments, we assume that every other grid point is observed, so that only 20 observations are assimilated. The "forecast model" is (3), but with f_i and d_i initially chosen to vanish. Thus, the forecast model differs from the model that generated the truth because some model parameters differ from those used in the true model. In addition, we apply covariance localization to both the state and the parameters f_i and d_i. Such localization is reasonable when the parameters have a spatial interpretation. We then assimilate the "observations" using ensemble forecasts generated by the forecast model. The goal is to see if the assimilation system can recover the true values of the model parameters. We emphasize the challenging nature of this experiment: there are only 20 observations, but there are 120 unknown quantities being estimated, namely 40 state variables $\mathbf{x}$, 40 values of f_i, and 40 values of d_i.

The performance of the resulting filter is shown as the solid line in fig. 1. For comparison, we also show the performance of the filter in which the model parameters are fixed to their initial values ("imperfect") and to their true values ("perfect"). We see that the filter in which model parameters are estimated produces much more accurate analyses than the filter in which model parameters are fixed at their imperfect values. Indeed, the analyses are as accurate as if the true forecast model had been used to produce a traditional analysis. Thus, these results illustrate how effective and powerful the augmentation approach to parameter estimation can be.

3 Estimation of Stochastic Parameters

Let us now apply the augmentation method to the simplest possible stochastic dynamical model

$$x_t = \phi x_{t-1} + \beta w_t, \tag{5}$$

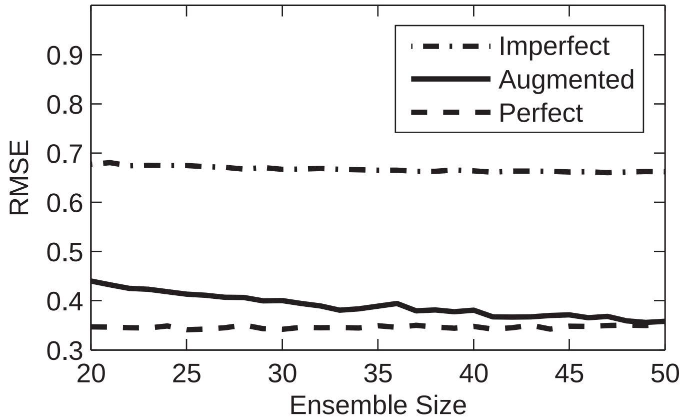

Figure 1: The root mean square difference between the analyzed and true state of the model (3) as a function of the ensemble size, where the analyzed state is derived from the augmentation method with smoothed parameter update equation (4) (solid curve denoted "Augmented"). Also shown is the root mean square difference between the analyzed and true state in which the model parameters are fixed to their initial values (dot-dashed curve denoted "Imperfect") and to their true values (dashed curve denoted "Perfect").

where $|\phi| < 1$, w is Gaussian white noise with variance σ_w^2, and β is a parameter to be estimated. We call β a "stochastic parameter" because it controls the variance of a stochastic process. The result of applying the augmentation method to estimate ϕ and β separately are shown in left panels of fig. 2. We see that the filter produces accurate estimates of the parameter ϕ, but fails utterly to produce accurate estimates of the stochastic parameter β. The reason for this failure is explained in mathematical detail in [6]. The essence of the argument is as follows. The distribution of x_t for fixed ϕ, β, and x_{t-1} is

$$P(x_t|\phi,\beta,x_{t-1}) \sim N(\phi x_{t-1}, \beta^2\sigma_w^2). \tag{6}$$

This equation reveals that variations of β affect the ensemble spread, but *not the ensemble mean*. As a result, it can be shown that the covariance between x_t and β_t decays rapidly with time step, and in fact vanishes if the initial covariance vanishes. This vanishing covariance implies that x_t and β_t are independent (under a normal distribution), and hence the filter cannot estimate one variable based only on knowledge of the other. This point does not appear to be widely recognized in the literature, as evidenced by the fact that some papers claim that the augmentation method can estimate stochastic parameters.

4 Generalized Maximum Likelihood Estimation

It is widely recognized that Bayes' theorem provides the most general and consistent basis for estimating unknown quantities. In the context of state and parameter estimation, we want to know the joint distribution of the state vector $\mathbf{x}$ and parameter vector β conditioned on all antecedent observations. For convenience, we separate observations at the current time step t, denoted $\mathbf{o}$, from the set of all observations prior to time t, denoted Θ; the complete set of observations would be denoted by $\mathbf{o}\Theta$. Then, Bayes' theorem states

$$\underbrace{p(\beta\mathbf{x}|\mathbf{o}\Theta)}_{Posterior} \propto \underbrace{p(\mathbf{o}|\mathbf{x}\beta\Theta)}_{likelihood}\ \underbrace{p(\mathbf{x}|\beta\Theta)}_{forecast}\ \underbrace{p(\beta|\Theta)}_{prior}, \tag{7}$$

where the traditional name for each term is indicated directly underneath the respective term. Most data assimilation schemes assume that the observation is some linear combination of the truth, plus noise:

$$\mathbf{o} = \mathbf{H}\mathbf{x} + \mathbf{r}, \tag{8}$$

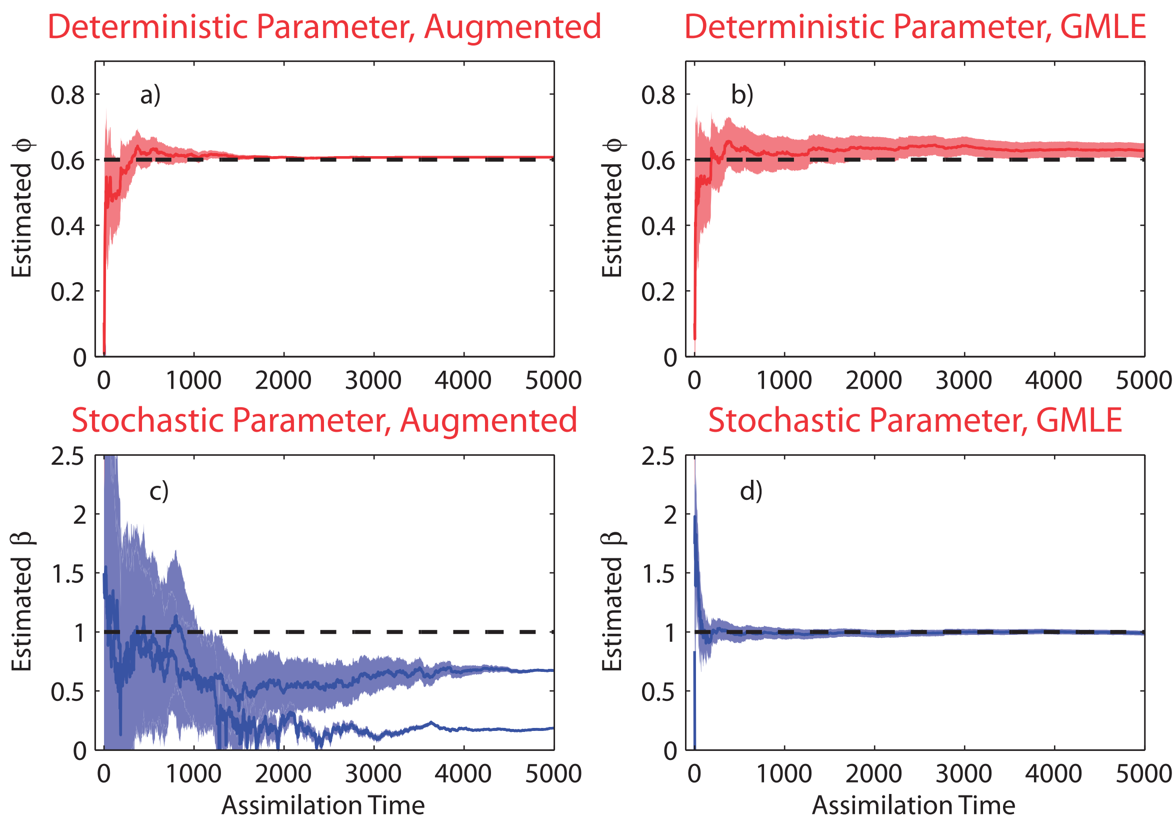

Figure 2: Results from four different data assimilation experiments with the AR(1) model (5). Specifically, the figure shows as a function of assimilation time: (a) the estimated value of the AR-parameter ϕ using the augmented EnSRF, (b) estimated value of ϕ using the maximum likelihood method, (c) estimated value of β using the augmented EnSRF, and (d) estimated value of β using the maximum likelihood method. In the case of estimating β (i.e., panels c and d), two independent assimilation experiments were performed using the same initial condition but different realizations of noise. The dashed lines indicate the true parameter values, and shading indicates two standard deviations of the predicted parameter uncertainty.

where $\mathbf{H}$ is a (linear) operator that maps the state space to observations space and $\mathbf{r}$ is Gaussian white noise with zero mean and covariance matrix $\mathbf{R}$. This model immediately implies

$$p(\mathbf{o}|\mathbf{x}\beta\Theta) = (2\pi)^{-M_o/2}|\mathbf{R}|^{-1/2}\exp\left(-(\mathbf{o}-\mathbf{Hx})^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{Hx})/2\right), \quad (9)$$

where M_o is the dimension of the observation vector $\mathbf{o}$. Note that neither β nor Θ appear in this expression, indicating that $\mathbf{o}$ is conditionally independent of these parameters given $\mathbf{x}$. The forecast distribution also is assumed to be normally distributed with mean μ_x^f and covariance matrix $\mathbf{P}^f$:

$$p(\mathbf{x}|\beta\theta) = (2\pi)^{-M_x/2}|\mathbf{P}^f|^{-1/2}\exp\left(-\left(\mathbf{x}-\mu_x^f\right)^T\mathbf{P}^{f^{-1}}\left(\mathbf{x}-\mu_x^f\right)/2\right), \quad (10)$$

where M_x is the dimension of the state vector $\mathbf{x}$. Finally, the prior distribution is assumed to be Gaussian with mean μ_β^f and covariance matrix Σ_β:

$$p(\beta) = (2\pi)^{-M_\beta/2}|\Sigma_\beta|^{-1/2}\exp\left(-\left(\beta-\mu_\beta^f\right)^T\Sigma_\beta^{-1}\left(\beta-\mu_\beta^f\right)/2\right), \quad (11)$$

where M_β is the dimension of the parameter vector β.

A standard result in statistical theory is that as the number of observations increases, the posterior distribution for the parameters tends toward a normal distribution ([14], p224). Since the mean of a Gaussian also maximizes the density, the mean can be estimated by setting the derivative of the posterior to zero and solving. This resulting mean is called the *Generalized Maximum Likelihood Estimate (GMLE)*. In differentiating the posterior distribution, the question arises as to which terms vary with $\mathbf{x}$ and β. As we found in the case of the simple stochastic dynamical model (5), model parameters can influence both the mean μ_x^f and the covariance matrix $\mathbf{P}^f$. Therefore, the derivative of these terms will appear in the GMLE.

In practice, it is simpler to deal with -2 times the log of the posterior, which is

$$-2\log p(\beta\mathbf{x}|\mathbf{o}\Theta) =$$

$$\begin{array}{lr} (\mathbf{o}-\mathbf{Hx})^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{Hx})+\log|\mathbf{R}|+M_o\log 2\pi+ & \textit{Likelihood} \\ \left(\mathbf{x}-\mu_x^f\right)^T\mathbf{P}^{f^{-1}}\left(\mathbf{x}-\mu_x^f\right)+\log|\mathbf{P}^f|+M_x\log 2\pi+ & \textit{Forecast} \\ \left(\beta-\mu_\beta^f\right)^T\Sigma_\beta^{-1}\left(\beta-\mu_\beta^f\right)+\log|\Sigma_\beta|+M_\beta\log 2\pi & \textit{Prior} \end{array}. \tag{12}$$

The generalized maximum likelihood estimate is found by differentiating (12) with respect to $\mathbf{x}$ and β and setting the result to zero. This procedure yields

$$2\frac{\partial\Pi}{\partial\mathbf{x}} = -2\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{o}-\mathbf{Hx})+2\mathbf{P}^{f^{-1}}(\mathbf{x}-\mu_x^f)=0 \tag{13}$$

$$2\frac{\partial\Pi}{\partial\beta_j} = -(\mathbf{x}-\mu_x^f)^T\mathbf{P}^{f^{-1}}\frac{\partial\mathbf{P}^f}{\partial\beta_j}\mathbf{P}^{f^{-1}}(\mathbf{x}-\mu_x^f)+\mathrm{tr}\left[\mathbf{P}^{f^{-1}}\frac{\partial\mathbf{P}^f}{\partial\beta_j}\right] \\ +2\left(\Sigma_\beta^{-1}(\beta-\mu_\beta^f)\right)_j-2\left(\frac{\partial\mu_x^f}{\partial\beta_j}\right)^T\mathbf{P}^{f^{-1}}(\mathbf{x}-\mu_x^f)=0, \tag{14}$$

where the following identities have been used:

$$\frac{\partial\mathbf{P}^{f^{-1}}}{\partial\beta_j} = -\mathbf{P}^{f^{-1}}\frac{\partial\mathbf{P}^f}{\partial\beta_j}\mathbf{P}^{f^{-1}} \quad \text{and} \quad \frac{\partial\log|\mathbf{P}^f|}{\partial\beta_j}=\mathrm{tr}[\mathbf{P}^{f^{-1}}\frac{\partial\mathbf{P}^f}{\partial\beta_j}]. \tag{15}$$

The vector $\mathbf{x}$ that satisfies (13) is called the analysis μ_a and is given by the standard Kalman Filter estimate

$$\mu_x^a = \mu_x^f+\mathbf{P}^f\mathbf{H}^T\left(\mathbf{R}+\mathbf{HP}^f\mathbf{H}^T\right)^{-1}\left(\mathbf{o}-\mathbf{H}\mu_x^f\right). \tag{16}$$

The fact that the Kalman Filter estimate emerges from maximum likelihood estimation is well known [15]. Using (16) to eliminate $\mathbf{x}=\mu_a$ from (14) gives

$$-(\mathbf{o}-\mathbf{H}\mu_x^f)^T\left(\mathbf{R}+\mathbf{HP}^f\mathbf{H}^T\right)^{-1}\mathbf{H}\frac{\partial\mathbf{P}^f}{\partial\beta_j}\mathbf{H}^T\left(\mathbf{R}+\mathbf{HP}^f\mathbf{H}^T\right)^{-1}(\mathbf{o}-\mathbf{H}\mu_x^f) \\ -2\left(\frac{\partial\mu_x^f}{\partial\beta_j}\right)^T\mathbf{H}^T\left(\mathbf{R}+\mathbf{HP}^f\mathbf{H}^T\right)^{-1}(\mathbf{o}-\mathbf{H}\mu_x^f) \\ +\mathrm{tr}\left[\mathbf{P}^{f^{-1}}\frac{\partial\mathbf{P}^f}{\partial\beta_j}\right]+2\left(\Sigma_\beta^{-1}(\beta-\mu_\beta^f)\right)_j=0 \tag{17}$$

This last equation depends only on β, and thus can be solved (in principle) for the maximum likelihood value. Once the maximum likelihood value of β is determined, the forecast covariance matrix $\mathbf{P}^f$ is specified and the Kalman filter equations can be solved to obtain μ_a.

The above equations effectively specify the mean analysis. To specify the analysis covariance matrix, another set of equations must be derived by taking the second derivative of the posterior distribution. The result of this calculation and the synthesis of these equations to formulate an ensemble data assimilation system is given in [6]. A key step in the new data assimilation scheme is estimation of the derivative of the background mean and covariance matrix with respect to the parameters. We propose estimating these derivatives with ensemble methods. Specifically, we generate an ensemble with fixed parameter $\beta + \Delta\beta$, where $\Delta\beta$ is a small perturbation to β, and another ensemble with fixed parameter $\beta - \Delta\beta$. Then, we estimate the first derivative of $\mathbf{P}^f$ as

$$\frac{\partial \mathbf{P}^f}{\partial \beta} = \frac{\mathbf{P}^f(\beta + \Delta\beta) - \mathbf{P}^f(\beta - \Delta\beta)}{2\Delta\beta}, \tag{18}$$

and similarly for μ_x^f. The result of applying the new method to estimating ϕ and β in (5) are shown in the right panels of fig. 2. Comparison with the left panels shows that the new method outperforms augmentation methods for estimating stochastic parameters.

It should be noted that the above method is general and does not require distinguishing between deterministic or stochastic parameter estimation. In general, deterministic parameter estimation is characterized by nonzero $\partial \mu_x^f / \partial \beta$, while stochastic parameter estimation is characterized by nonzero $\partial \mathbf{P}^f / \partial \beta$.

5 Relation Between GMLE and Other Parameter Estimation Methods

Since the above method can estimate either stochastic and deterministic parameters, or both, it is of interest to understand its relation to other parameter estimation methods.

5.1 Relation Between GMLE and Augmentation Methods

As discussed earlier, the augmentation method is based on defining the augmented state vector (1) and then applying a standard data assimilation system to the augmented vector. The forecast covariance matrix for the augmented vector is

$$\mathbf{P}_z^f = \begin{pmatrix} \mathbf{P}_x^f & \mathbf{P}_{x\beta}^f \\ \mathbf{P}_{\beta x}^f & \mathbf{P}_\beta^f \end{pmatrix}, \tag{19}$$

where $\mathbf{P}_{x\beta}^f$ is the cross-covariance matrix between $\mathbf{x}$ and β. In the context of parameter estimation, the parameter itself is never observed, implying that the mapping operator between model and observation space is

$$\mathbf{H}_z = \begin{pmatrix} \mathbf{H}_x & \mathbf{0} \end{pmatrix}. \tag{20}$$

The standard update equation for the mean is

$$\mu_z^a = \mu_z^f + \mathbf{P}_z^f \mathbf{H}_z^T \left(\mathbf{H}_z \mathbf{P}_z^f \mathbf{H}_z^T + \mathbf{R} \right)^{-1} \left(\mathbf{o} - \mathbf{H}_z \mu_z^f \right). \tag{21}$$

It is straightforward to substitute (19) and (20) in the mean update equation (21) to derive the mean update for β:

$$\mu_\beta^a = \mu_\beta^f + \mathbf{P}_{\beta x}^f \mathbf{H}_x^T \left(\mathbf{H}_x \mathbf{P}_x^f \mathbf{H}_x^T + \mathbf{R} \right)^{-1} \left(\mathbf{o} - \mathbf{H}_x \mu_x^f \right). \tag{22}$$

This equation therefore defines the mean update for the model parameters β in the augmentation method.

The GMLE for β is the solution to (17). However, we have seen that the augmentation method does not work for stochastic parameters, so we consider only the case in which no stochastic parameters occur. This assumption is tantamount to assuming $\partial \mathbf{P}^f / \partial \beta = \mathbf{0}$, in which case (17) can be manipulated to the form

$$\mu_\beta^a = \mu_\beta^f + \Sigma_\beta \frac{\partial \mu_x^f}{\partial \beta}^T \mathbf{H}^T \left(\mathbf{R} + \mathbf{H}\mathbf{P}^f\mathbf{H}^T \right)^{-1} \left(\mathbf{o} - \mathbf{H}\mu_x^f \right). \tag{23}$$

Comparison between (22) and (23) shows that the two expressions would be equal if

$$\Sigma_\beta \frac{\partial \mu_x^{f\,T}}{\partial \beta} = \mathbf{P}^f_{\beta x}. \tag{24}$$

Recall that the forecast is merely the conditional distribution of $\mathbf{x}_t$ given β and $\mathbf{x}_{t-1}$; that is,

$$\mu_x^f = E[\mathbf{x}_t|\beta \mathbf{x}_{t-1}]. \tag{25}$$

It is a standard result that if $\mathbf{x}_t, \beta, \mathbf{x}_{t-1}$ are joint normally distributed, then

$$E[\mathbf{x}_t|\beta \mathbf{x}_{t-1}] - \mu_x = \mathbf{P}^f_{x\beta}\Sigma_\beta^{-1}\left(\beta - \mu_\beta\right) + \text{linear terms in } \mathbf{x}_{t-1}. \tag{26}$$

Therefore,

$$\frac{\partial \mu_x^f}{\partial \beta} = \Sigma_\beta^{-1}\mathbf{P}^f_{\beta x}, \tag{27}$$

which is equivalent to (24). Thus, these considerations show that the GMLE of deterministic models is effectively equivalent to the augmentation method, provided that the covariance of the parameter ensemble in the augmentation method is identified with Σ_β.

5.2 Relation Between GMLE and Adjoint Parameter Estimation

As reviewed in [5], adjoint parameter estimation attempts to estimate the state and parameters that minimize the cost function

$$J = (\mathbf{o} - \mathbf{Hx})^T \mathbf{R}^{-1} (\mathbf{o} - \mathbf{Hx}) + \left(\mathbf{x} - \mu_x^f\right)^T \mathbf{P}^{f\,-1} \left(\mathbf{x} - \mu_x^f\right) + \left(\beta - \mu_\beta^f\right)^T \Sigma_\beta^{-1} \left(\beta - \mu_\beta^f\right). \tag{28}$$

This expression is precisely the leading terms of (12). If all covariances are independent of $\mathbf{x}$ and β, then J is the only part of the log-posterior that depends on $\mathbf{x}$ and β. Thus, the maximum likelihood method leads to the same cost function that is used in traditional parameter estimation when the covariances are fixed. However, when the parameter being estimated is stochastic, the covariance matrix $\mathbf{P}^f$ varies with the parameter and the full posterior distribution needs to be used. These considerations show that adjoint methods based on (28) are fundamentally incapable of estimating stochastic parameters, because these methods ignore variations in forecast covariances due to variations in the model parameters.

6 Limitations

The most challenging aspect of the proposed data assimilation method is that it requires estimating the derivative of the forecast covariance matrix with respect to the model parameters. We proposed estimating these derivatives by generating extra ensembles with different (but fixed) parameters, and then taking suitable finite differences of the resulting sample covariance matrices. For small sample sizes, the differences between estimated covariance matrices are likely to be dominated by sampling errors and hence very inaccurate.

Acknowledgements

The material presented here is based on the papers [6] and [7]. This research was supported by the National Science Foundation (ATM0332910, ATM0830062, ATM0830068), National Aeronautics and Space Administration (NNG04GG46G, NNX09AN50G), and the National Oceanic and Atmospheric Administration (NA04OAR4310034, NA06OAR4310001, NA09OAR4310058). The views expressed herein are those of the authors and do not necessarily reflect the views of these agencies.

References

[1] J. Berner, F. J. Doblas-Reyes, T. N. Palmer, G. Shutts, and A. Weisheimer. Impact of a quasi-stochastic cellular automaton backscatter scheme on the systematic error and seasonal prediction skill of a global climate model. *Phil. Trans. R. Soc. A*, 366:2561–2579, 2008.

[2] J. Berner, G. J. Shutts, M. Leutbecher, and T. N. Palmer. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, 66:603–626, 2009.

[3] G. J. Shutts. A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, 131:3079–3102, 2005.

[4] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970. 376 pp.

[5] I.M. Navon. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans*, 27:55–77, 1997.

[6] T. DelSole and Xiaosong Yang. State and parameter estimation in stochastic dynamical models. *Physica D*, 239:1781–1788, 2010.

[7] X. Yang and T. DelSole. Using the ensemble Kalman Filter to estimate multiplicative model parameters. *Tellus*, 61:601–609, 2009.

[8] J. L. Anderson. An ensemble adjustment filter for data assimilation. *Mon. Wea. Rev.*, 129:2884–2903, 2001.

[9] S.-J. Baek, B. R. Hunt, E. Kalnay, E. Ott, and I. Szunyogh. Local ensemble Kalman filtering in the presence of model bias. *Tellus*, 58A:293–306, 2006.

[10] A. Gelb, editor. *Applied Optimal Estimation*. MIT Press, Cambridge, MA, 1974.

[11] D. Zupanski and M. Zupanski. Model error estimation employing an ensemble data assimilation approach. *Mon. Wea. Rev.*, 134:1337–1354, 2006.

[12] B. Gershgorin, J. Harlim, and A. J. Majda. Improving filtering and prediction of spatially extended turbulent systems with model errors through stochastic parameter estimation. *J. Comput. Phys.*, 229:32–57, 2010.

[13] E. N. Lorenz and K. A. Emmanuel. Optimal sites for supplementary weather observations: Simulation with a small model. *J. Atmos. Sci.*, 55:399–414, 1998.

[14] J. O. Berger. *Statistical Decision Theory and Bayesian Analysis*. Springer, 2nd edition, 1985.

[15] P. S. Maybeck. *Stochastic models, estimation, and control*, volume 1. Academic Press, 1982.