

# An Ensemble Kalman Filter for Statistical Estimation of Physics Constrained Nonlinear Regression Models

John Harlim\*

*Department of Mathematics and Department of Meteorology, the Pennsylvania State University, University Park, PA 16802*

Adam Mahdi

*Department of Mathematics, North Carolina State University, Raleigh, NC 27695*

Andrew J. Majda

*Department of Mathematics and Center for Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012*

---

## Abstract

A central issue in contemporary science is the development of nonlinear data driven statistical-dynamical models for time series of noisy partial observations from nature or a complex model. It has been established recently that ad-hoc quadratic multi-level regression models can have finite time blow-up of statistical solutions and/or pathological behavior of their invariant measure. Recently, a new class of physics constrained nonlinear regression models were developed to ameliorate this pathological behavior. Here a new finite ensemble Kalman filtering algorithm is developed for estimating the state, the linear and nonlinear model coefficients, the model and the observation noise covariances from available partial noisy observations of the state.

---

\*Corresponding author.

*Email addresses:* jharlim@psu.edu (John Harlim), amahdi@ncsu.edu (Adam Mahdi), jonjon@cims.nyu.edu (Andrew J. Majda)

Several stringent tests and applications of the method are developed here. In the most complex application, the perfect model has 57 degrees of freedom involving a zonal (east–west) jet, two topographic Rossby waves, and 54 nonlinearly interacting Rossby waves; the perfect model has significant non-Gaussian statistics in the zonal jet with blocked and unblocked regimes and a non-Gaussian skewed distribution due to interaction with the other 56 modes. We only observe the zonal jet contaminated by noise and apply the ensemble filter algorithm for estimation. Numerically, we find that a three dimensional nonlinear stochastic model with one level of memory mimics the statistical effect of the other 56 modes on the zonal jet in an accurate fashion, including the skew non-Gaussian distribution and autocorrelation decay. On the other hand, a similar stochastic model with zero memory levels fails to capture the crucial non-Gaussian behavior of the zonal jet from the perfect 57-mode model.

*Keywords:* ensemble Kalman filter, nonlinear regression models, parameter estimation of stochastic differential equations, multi-level models.

---

## 1. Introduction

A central issue in contemporary science is the development of data driven statistical-dynamical models for the time series of a partial subset of observed variables,  $u_1(t) \in \mathbb{R}^{N_1}$ , which arise from observations of nature or from an extremely complex physical model [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11]. This is an important issue in systems ranging from bio-molecular dynamics to climate science to engineering turbulence. Examples of such data driven dynamical models are multi-level linear autoregressive models with external factors [2, 6]

as well as ad-hoc quadratic nonlinear regression models [12, 13, 14, 15]. Such purely data driven ad-hoc regression models are developed through various criteria to fit the data but by design, do not respect the underlying physical dynamics of the partially observed system or the causal processes in the dynamics; nevertheless, the goal of purely data driven statistical modeling is to provide simplified low order models with high predictive skill for central features of the underlying physical system and not just fit (or over-fit, see [2]) the given data. Indeed, the work in [16] provides rigorous mathematical theory and examples with straightforward numerical experiments where the ad-hoc quadratic multi-level regression models proposed in [12, 13, 14] necessarily have non-physical finite-time blow-up of statistical solutions and also pathological behavior of the related invariant measure even though these models match a long time series of the observed data produced from the physical model with high accuracy.

In [17] a new class of physics constrained multi-level regression models were proposed to systematically ameliorate the above difficulties. These models begin with an observed time series for a variable  $u_1 \in \mathbb{R}^{N_1}$  and augment  $u_1$  with a hidden variable  $u_2$  so that primary nonlinear energy conserving interactions occur in the  $u$  variables with  $u = (u_1, u_2) \in \mathbb{R}^N$ ,  $N = N_1 + N_2$ , while the stochastic memory effects of noise are linear, conditional on the state  $u$ . If  $\Pi_2(u) = (0, u_2)^\top$  denotes the projection on  $u_2$ , then the **Physics Constrained Multi-Level Regression Models** developed in [17] have the structural form,

$$\begin{aligned} \frac{du}{dt} &= Lu + B(u, u) + F + \Pi_2 r_1 \\ \frac{d\vec{r}}{dt} &= Qu + \tilde{A}\vec{r} + \sigma\dot{W}, \end{aligned} \tag{1}$$

where the nonlinear interactions  $B(u, u)$  conserve energy so that

$$\langle u, B(u, u) \rangle = 0, \quad (2)$$

for a suitable inner product  $\langle \cdot, \cdot \rangle$ ; and  $\dot{W}$  is a white noise vector with noise matrix  $\sigma$ . The matrix  $A$  defining the interactions of  $(r_1, \dots, r_p)$  with  $r_j \in \mathbb{R}^{N_2}$ , where  $p$  memory levels are triangular and are given by

$$\begin{aligned} \frac{dr_i}{dt} &= Q_i u + \sum_{j=1}^i \tilde{a}_{i,j} r_j + r_{i+1}, \quad 1 \leq i \leq p-1 \\ \frac{dr_p}{dt} &= Q_p u + \sum_{j=1}^p \tilde{a}_{p,j} r_j + \sigma \dot{W}. \end{aligned}$$

Here  $p$  denotes the number of memory levels and  $p = 0$  denotes the special **zero-memory level model**

$$\frac{du}{dt} = Lu + B(u, u) + F + \Pi_2 \sigma \dot{W}. \quad (3)$$

Rigorous theorems in [17] show that such models do not blow-up (under certain hypotheses) and have nice invariant measures. Guidelines involving stability of  $L$  as well as observability and controllability of (1) for the primary variables  $u = (u_1, u_2)$  when only  $u_1$  is observed are also given in [17] together with a counterexample with statistical blow-up with neutral stability.

The practical issue treated here involves the inference of an appropriate stochastic model in (1) with (2) from partial knowledge involving observations of  $u_1$  alone, contaminated by noise. This estimation problem involves determining the coefficients  $L, Q, \tilde{A}, B$  and estimating the noise amplitude  $\sigma$  in (1) and the observation noise covariance, besides estimating  $u_2$ .

The filtering or estimation algorithm implemented in [17] was based on using Extended Kalman Filter (EKF) combined with Belanger's method [18]

for noise and parameter estimation. This algorithm was used successfully in [17] to estimate reduced stochastic model for the first Fourier mode of the Truncated Burgers-Hopf (TBH) model [19] as well as related nonlinear oscillators with memory. In Section 2 of this paper we provide an instructive motivating example, showing the skill of this algorithm with complete observation of  $u$  and a spectacular failure of this algorithm due to linear instability with partial observation when only  $u_1$  is observed. A new finite Ensemble Kalman Filter (EnKF) based scheme for estimating all the parameters  $L, B, Q, \tilde{A}$  and the noise  $\sigma$  in the models from (1) is developed in Section 3.

An important area for application of the regression models is low frequency climate variability and a family of stringent paradigm models for this behavior [9, 20], which are studied through the new estimation algorithm in Section 4. The most complex problem considered in Section 4 involves a perfect model with 57 degrees of freedom with a large scale zonal (east-west) jet, two topographic Rossby waves, and 54 nonlinearly interacting Rossby waves; the perfect models have significant non-Gaussian statistics with blocked and unblocked regimes of the zonal jet due to stochastic backscatter from the Rossby waves even though the total perfect system exactly conserves energy. We only observe the zonal jet contaminated by noise and seek a physics constrained stochastic model with memory with form in (1) which mimics the statistics of the 57-mode model. We apply our EnKF based algorithm for parameter estimation introduced in Section 3 and find that a three dimensional nonlinear stochastic model with one level of memory,  $p = 1$ , is sufficient to mimic the statistical effects of the other 56 modes on the zonal jet  $u$  in an

accurate fashion, including both the skewed non-Gaussian distribution for  $u$  and its autocorrelation decay. On the other hand, a similar model with zero level memory fails to capture any of these crucial non-Gaussian behavior for the zonal jet from the perfect 57-mode model.

## 2. Motivating example for new parameter estimation algorithms

The existing EKF based algorithm [22, 23] was successfully used for estimating deterministic and stochastic parameters of various models (see e.g. [17]). In this section we will introduce a family of test models that will be used to show some of the limitations of the EKF based scheme, including a failure due to linear instability when only partial observations are available. This, in turn, motivates the development of a new, more robust, EnKF based algorithm (described in detail in Section 3), which we compare with the EKF filter.

Consider a system of stochastic differential equations (SDEs) in (1) with one memory level and  $F = 0$ , which in compact form can be written as

$$\frac{dx}{dt} = Ax + B(x, x) + \Sigma \dot{W}, \quad (4)$$

where  $x = (u_1, u_2, r) \in \mathbb{C}^3$  and  $\Sigma \dot{W} = (0, 0, \sigma \dot{W})^\top$ , with  $W$  denoting a complex valued Wiener process. In this example, the linear operator  $A \in \mathbb{C}^{3 \times 3}$  is given by

$$A = \left( \begin{array}{c|c} L & \begin{matrix} 0 \\ 1 \end{matrix} \\ \hline Q_1 & \tilde{a}_{1,1} \end{array} \right)$$

and is set to be symmetric negative definite; the nonlinear term is defined as follows,

$$B(x, x) = (-iu_2^*(c_1u_1^* + c_2u_2^*), iu_1^*(c_1u_1^* + c_2u_2^*), 0)^\top$$

to ensure globally stable non-Gaussian solutions with Lyapunov function  $V = \frac{1}{2}(|u_1|^2 + |u_2|^2 + |r|^2)$  when  $\sigma = 0$ . Here and throughout this article  $|x|^2 \equiv x^*x$ , where ‘\*’ denotes complex conjugate. In our simulations, we choose nonzero  $A, c_1, c_2, \sigma$  such that the system in (4) has all negative eigenvalues and comparable variance on each component (see Table 1 below). For a special choice with  $c_1 \neq 0$  and  $c_2 = 0$ , the model in (4) corresponds to the nonlinear oscillator test model in [17] constructed by adding a one memory level to Galerkin truncation of the TBH model up to the first two Fourier modes [21]. It is not too difficult to show that the model in (4) is controllable and thus geometrically ergodic [17], except on a set of measure zero parameters.

### 2.1. Numerical results with EKF based parameter estimation scheme

We consider here parameter estimation with Belanger’s method [18, 22] blended with EKF (see Appendix for the detail algorithm) for estimating all the model deterministic parameters  $A, B$ , the model stochastic noise amplitude,  $\Sigma$ , and the unknown observation noise covariance, given noisy partial observations of  $x$ . This is exactly the algorithm used in [17, 23].

First, we consider the case when we observe  $u_1$  and  $u_2$ , corrupted by independent white noises with variance 10% of the corresponding climatological (temporal average) variance. In this situation, we obtain very accurate estimates (see Figures 1, 2 for the marginal densities and autocorrelation function, respectively) for this non-Gaussian statistics. In Table 2 we also include

other statistical estimates such as the correlation times and variances. In Figures 3-5 we show the corresponding state and parameter estimates compared to the truth. Notice that the fitting for the observable state variables is quite skillful. The scheme estimates the deterministic parameters remarkably well (even the nonlinear coefficients  $c_1, c_2$  are estimated accurately), except for  $a_{22}, a_{33}, a_{23}$ .

Now consider noisy observations of only  $u_1$ . Unfortunately in this situation the parameterization strategy with EKF diverges in finite time. In Figure 6, we plot the largest absolute value of the eigenvalue of the linear tangent model used in EKF based scheme as well as the estimates for real part of the observed  $u_1$ , and unobserved  $u_2$ , variables, as functions of assimilation step, until the scheme diverges (in this simulation, the solutions blow-up to machine infinity after 728 iterations). Here, the stability of the linear tangent model is sensitive to initial conditions. When the linear tangent model largest magnitude of the eigenvalue is greater than one, the EKF is unstable and the state estimate for the unobserved variable,  $u_2$ , diverges. On the other hand, the estimates for the observed variable,  $u_1$ , are exactly the observations. Such a pathological behavior was also observed in a simpler context (see Chapter 3 of [24]).

## *2.2. Non blow-up parameter estimation scheme*

As a remedy to the divergence of solutions in the EKF based parameter estimation scheme with sparse observations, we propose an EnKF based parameter estimation algorithm, which we will explain, in more detail, in Section 3.

In Figures 7, 8 and Table 2, we show the statistical estimates obtained



from EnKF based scheme observing similar set of noisy observations  $u_1$  and  $u_2$  as those used in the EKF estimation experiment above. Notice that the density estimates are accurate but the autocorrelation function estimates are less accurate especially, beyond 20 time units. Compare to the EKF based scheme, the EnKF based algorithm is less accurate with noisy observations of  $u = (u_1, u_2)$ . Such degrading estimates in this particular regime can be attributed to various factors such as ensemble size, etc. However, our scheme produces reasonable statistical estimates when observing only  $u_1$  (see Figure 9 and Table 2) whereas the EKF based algorithm blow-up in finite time. Next, we will describe the proposed EnKF based algorithm in detail. Subsequently, we will apply this newly developed algorithm on a complex test model for topographic mean flow interaction in Section 4.

### 3. An ensemble based parameter estimation algorithm for stochastic differential equations

The basic parameter estimation scheme utilized in [17] blends the EKF and Belanger’s method [18] in adaptive fashion. When observation  $y_k$  becomes available at time step  $t_k$ , the parameter estimation scheme applies EKF to estimate simultaneously, the state variables and the parameters associated with the deterministic part of the SDE’s (for model in (1), these parameters are components of  $L, Q, \tilde{A}, B$ ). This step is called the primary filter in [23, 22]. Subsequently, it applies a secondary filter to estimate the stochastic parameter,  $\sigma$ , and the observation noise covariance, based on the estimated innovation vector,  $v_k \equiv y_k - Hx_k^-$ , defined as the difference between the observation  $y_k$  and prior mean estimate  $Hx_k^-$  in the observation

space. Here, the observation operator  $H$  projects the model state space to observation space. In the Appendix below, we review the EKF based parameter estimation algorithm step-by-step for convenience. In there, we write the full algorithm in eight steps: Steps 1 and 2 describe the primary filter and, the remaining Steps 3-8 describe the secondary filter. This estimation algorithm is equivalent to the recursive formulation introduced by Belanger for parameterizing linear SDEs [18].

To implement Belanger’s noise estimation scheme, one needs to specify a linear operator that maps the second order statistics forward in time. This forward operator is used to construct the observation operator and the noise covariance matrix used in the secondary filter (see Steps 3-7 in the Appendix). When EKF is used for the primary filter, it naturally provides linear tangent model as the forward operator (see Step 2 in the Appendix below).

While the EKF based estimation scheme is accurate in the presence of full observations,  $u = (u_1, u_2)$ , as we reported in Section 2.1, the estimation scheme diverges with only partial observation  $u_1$  due to linear instability of the linear tangent operator. As a remedy, we propose to use prior and posterior ensembles of solutions to approximate the linear tangent operator (this step will be discussed in detail shortly); this approach is also used in [25]. In general, one can choose any desirable EnKF filter; here we implement Ensemble Transform Kalman Filter (ETKF) [26].

In summary, the new parameter estimation scheme uses the same secondary filter (Steps 3-8 in the Appendix) and replaces the primary filter (Steps 1 and 2 in the Appendix) with the following two steps. Given a prior ensemble of solutions  $x_1^{(j),-}$ ,  $j = 1, \dots, K$ :

1. Compute the mean and perturbation about the mean

$$\begin{aligned}\bar{x}_k^- &= \frac{1}{K} \sum_{i=1}^K x_k^{(i),-} \\ U_k &= [x_k^{(1),-} - \bar{x}_k^-, \dots, x_k^{(K),-} - \bar{x}_k^-] \\ V_k &= HU_k.\end{aligned}$$

Apply the ETKF with the following step

$$\begin{aligned}v_k &= y_k - H\bar{x}_k^- \\ J &= (K-1)\mathcal{I}_K + V_k \left( \sum_{i=1}^p R_i a_{i,k} \right)^{-1} V_k^\top.\end{aligned}$$

Compute singular value decomposition (SVD) of  $J = X\Sigma X^\top$ . Construct

$$\begin{aligned}\tilde{K}_k &= U_k X \Sigma^{-1} X^\top V_k^\top \left( \sum_{i=1}^p R_i a_{i,k} \right)^{-1} \\ W_k &= \sqrt{K-1} U_k X \Sigma^{-1/2} X^\top, \\ \bar{x}_k^+ &= \bar{x}_k^- + \tilde{K}_k v_k, \\ x_k^{(j),+} &= \bar{x}_k^+ + W_k^{(j)}, \quad j = 1, \dots, K.\end{aligned}$$

Here  $W_k^{(j)}$  denotes the  $j$ th column of  $W_k$ .

2. Compute the deterministic solutions for each ensemble member at the next time step with

$$\tilde{x}_{k+1}^{(j),-} = f(x_k^{(j),+}), \quad k = 1, \dots, K,$$

where  $f$  denotes the filter model deterministic operator. Notice that if we only have a deterministic linear model, the covariance update is given by

$$P_{k+1}^- = A_k P_k^+ A_k^\top, \quad (5)$$

where  $f(x_k) = A_k x_k$ . For nonlinear  $f$ , we propose to approximate  $A_k$  with ensemble of solutions, following the approach in [25]. The ensemble based prior and posterior error covariance matrices are defined by

$$\begin{aligned} P_{k+1}^- &= \frac{1}{K-1} U_{k+1}^d (U_{k+1}^d)^\top, \\ P_k^+ &= \frac{1}{K-1} W_k W_k^\top, \end{aligned} \tag{6}$$

where  $U_{k+1}^d = [\tilde{x}_{k+1}^{(1),-} - \bar{\tilde{x}}_{k+1}^-, \dots, \tilde{x}_{k+1}^{(K),-} - \bar{\tilde{x}}_{k+1}^-]$  and  $\bar{\tilde{x}}_{k+1}^-$  is the ensemble average of the deterministic solutions  $\{\tilde{x}_{k+1}^{(j),-}\}_{j=1}^K$ . Inserting (6) into (5) and multiplying both side by  $K-1$ , we obtain

$$U_{k+1}^d (U_{k+1}^d)^\top = A_k W_k W_k^\top A_k^\top,$$

or

$$U_{k+1}^d = A_k W_k.$$

Therefore, we can approximate  $A_k$  by

$$A_k = U_{k+1}^d W_k^\dagger,$$

where  $\dagger$  denotes pseudo-inverse. Now we account for the error covariance associated with the stochastic term through,

$$P_{k+1}^- = \frac{1}{K-1} U_{k+1}^d (U_{k+1}^d)^\top + \sum_{i=1}^p Q_i a_{i,k}.$$

Then generate  $U_{k+1}$  with mean zero and covariance  $P_{k+1}^-$ . One method is to first simulate a random matrix  $\tilde{U} = [u^{(1)}, \dots, u^{(K)}] \in \mathbb{R}^{n \times K}$  where

each component is sampled from  $\mathcal{N}(0, 1)$ . Since these are finite samples, take away the bias by

$$\hat{u}^{(j)} = u^{(j)} - \frac{1}{K} \sum_{j=1}^K u^{(j)}.$$

Define  $\hat{U} = [\hat{u}^{(1)}, \dots, \hat{u}^{(K)}]$ , take eigenvalue decomposition,  $\frac{1}{K-1} \hat{U} \hat{U}^\top = \Xi D \Xi^\top$ . Then we have,

$$U_{k+1} = P_{k+1}^- \Xi D^{-1/2} \Xi^\top \hat{U}.$$

Now we can apply the secondary filter (described in Step 3-8 in the Appendix) to update the noise parameter,  $\sigma$ , and the observation noise covariance. Then, return to Step 1 above for the next assimilation time.

#### 4. Physics Constrained Nonlinear Regression Models for Topographic Mean Flow Interaction

In this section, we apply the proposed parameter estimation scheme on a family of stringent paradigm models for topographic mean flow interaction [27, 28]. In particular, we consider fitting time series from topographic mean flow interaction that solves a barotropic quasi-geostrophic equation with a large scale zonal mean flow  $u$  on a two-dimensional periodic domain [29, 28]. The full equations are given as follows

$$\begin{aligned} \frac{\partial q}{\partial t} &= -\nabla^\perp \psi \cdot \nabla q - u(t) \frac{\partial q}{\partial x} - \beta \frac{\partial \psi}{\partial x}, \\ q &= \Delta \psi + h, \\ \frac{du}{dt} &= \frac{1}{4\pi^2} \int h \frac{\partial \psi}{\partial x}. \end{aligned} \tag{7}$$

Here  $q$  and  $\psi$  are the small-scale potential vorticity and stream function, respectively. The large-scale zonal velocity field is characterized by  $u(t)$  and the topography is defined by function  $h = h(x, y)$ . The parameter  $\beta$  is associated with the  $\beta$ -plane approximation to the Coriolis force. Notation  $\nabla^\perp \equiv (-\partial_y, \partial_x)$  and the integral above is a two-dimensional integral over a periodic box of length  $2\pi$ .

Now we shall construct a set of special solutions to (7), which inherit the nonlinear coupling of the small-scale flow with the large-scale mean flow via topographic stress. Consider the following Fourier decomposition,

$$\psi(\vec{x}, t) = \sum_{k \neq 0} \psi_k(t) e^{ik\vec{\ell} \cdot \vec{x}}, \quad h(\vec{x}) = \sum_{k \neq 0} h_k e^{ik\vec{\ell} \cdot \vec{x}} \quad (8)$$

where  $\vec{\ell} = (\ell_x, \ell_y)$  and  $\vec{x} = (x, y)$ . Note that we assume the topography has zero mean with respect to spatial average, that is,  $h_0 = 0$ . Substituting the ansatz in (8) into (7), we obtain the **layered topographic equations in Fourier form** [29, 28]:

$$\begin{aligned} \frac{d\psi_k}{dt} &= ik\ell_x \left( \frac{\beta}{k^2 |\vec{\ell}|^2} - u \right) \psi_k + i \frac{k\ell_x}{k^2 |\vec{\ell}|^2} h_k u, \\ \frac{du}{dt} &= -i\ell_x \sum_{k \neq 0} kh_k \psi_k^*, \end{aligned} \quad (9)$$

where  $\psi_k^* = \psi_{-k}$  and  $h_k^* = h_{-k}$ , since both quantities  $\psi$  and  $h$  are real. We note (cf. [29, 28]) that without loss of generality we can always rescale the system in (9) with  $\ell_x \neq 0$  to align to a special case with  $\vec{\ell} = (1, 0)$ . With this property, we will consider a single layer topography in this zonal direction,  $\vec{\ell} = (1, 0)$  in the remaining of this paper.

Now we consider truncation of (9) with  $0 < |k| \leq \Lambda = 1$ , i.e., restrict

ourselves with the solutions of (7) in the form,

$$\begin{aligned}\psi(x, t) &= \psi_1(t)e^{ix} + \psi_{-1}(t)e^{-ix}, \\ h(x) &= H(\cos(x) + \sin(x)) = h_1e^{ix} + h_{-1}e^{-ix},\end{aligned}\tag{10}$$

where  $h_1 = H/2 - H/2i$  and  $h_{-1} = h_1^*$ , and  $H$  denotes the topography amplitude.

#### 4.1. Parameter estimation of a test model for interactions of zonal jet with two topographic Rossby waves

Here we proceed with the following specific choice of the coefficients for function  $\psi(x, t)$ :

$$\psi_1(t) = \frac{1}{2}(b(t) - ia(t)) \quad \text{and} \quad \psi_{-1} = \psi_1^*.\tag{11}$$

Substituting (10) with (11) into (9) and rotating the variables  $(a, b)$  counterclockwise by  $45^\circ$  to coordinate  $(v_1, v_2)$ , we obtain the following system

$$\begin{aligned}\frac{du}{dt} &= \omega v_1, \\ \frac{dv_1}{dt} &= -2\omega u - \beta v_2 + uv_2 \\ \frac{dv_2}{dt} &= \beta v_1 - uv_1,\end{aligned}\tag{12}$$

where  $\omega \equiv H/\sqrt{2}$ . This model is similar to the Charney and DeVore [30] model for nonlinear regime behavior without dissipation and forcing. Now we consider (see [31, 32, 20, 28]) approximating the interaction with the truncated Rossby wave modes,  $|k| > 1$ , with an Ornstein-Uhlenbeck process, to obtain a test model that describes nonlinear interactions of zonal jet with two topographic Rossby waves:

$$\dot{x} = (L - D)x + B(x, x) + \Sigma\dot{W},\tag{13}$$

where  $x = (u, v_1, v_2)^\top$ ,  $B(x, x) = (0, A_1 u v_2, -A_1 u v_1)^\top$ ,  $D = \text{diag}(0, d_1, d_2)^\top$ ,  $\dot{W} = (\dot{W}_1, \dot{W}_2)^\top$ ,

$$L = \begin{pmatrix} 0 & \omega & 0 \\ -2\omega & 0 & -\beta \\ 0 & \beta & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 & 0 \\ \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}. \quad (14)$$

Before we move on discussing the numerical results, we establish some basic properties of system (13), (14). We define on  $\mathbb{R}^3$  the following inner product

$$\langle x, x \rangle_* = u^2 + \frac{1}{2}(v_1^2 + v_2^2).$$

Notice that  $L$  is skew-symmetric with respect to this inner product, that is  $\langle Lx, y \rangle_* = -\langle x, Ly \rangle_*$  for any  $x, y \in \mathbb{R}^3$  and  $\langle x, B(x, x) \rangle_* = 0$ . Note that the test model in (13), (14) is an example of the special zero-memory level model in (3). Furthermore, we have the following general properties.

**Proposition 1.** *Consider the  $3 \times 3$  stochastic test system in (13), (14), and assume  $d_1, d_2 > 0$ . We have the following properties:*

- (a) *Set the noise  $\Sigma$  to zero. The origin of the system is globally asymptotically stable.*
- (b) *The system is controllable provided  $\omega \neq 0$ .*
- (c) *For  $\sigma_1^2/d_1 = \sigma_2^2/d_2 = E$ , the system has a Gaussian invariant measure proportional to  $\exp(-E^{-1}\|x\|_*^2)$ .*
- (d) *Consider the two stochastic systems in (13), (14) with  $\omega$  replaced by  $-\omega$ , then these two systems have the same temporal statistics for  $u$  in statistical equilibrium.*



*Proof.* The proof of (a) is straightforward, noting that  $V(u, v_1, v_2) = \frac{1}{2}u^2 + \frac{1}{4}(v_1^2 + v_2^2)$  is a Lyapunov function and the origin is the only singular point.

Now we show (b). Let  $\mathcal{F}$  be the vector field associated with (13) and let  $g_1 = (0, 1, 0)^\top$ ,  $g_2 = (0, 0, 1)^\top$ . We construct a matrix

$$M = (g_1, g_2, [\mathcal{F}, g_1], [\mathcal{F}, g_2])^\top,$$

which can be given explicitly as

$$M = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\omega & -d_1 & -A_1u + \beta \\ 0 & A_1u - \beta & -d_2 \end{pmatrix}.$$

If we show that  $M$  has a full rank, the system is controllable. Let  $M_j$  be a matrix  $M$  with deleted  $j$ -th row. Then

$$\det(M_1) = -\omega(A_1u - \beta); \quad \det(M_2) = -d_2\omega; \quad \det(M_3) = 0; \quad \det(M_4) = -\omega.$$

Since  $\omega$  is a common factor of the above determinants the fact (b) of the proposition follows.

Statement (c) is a direct calculation following Section 10.2.2 of [24].

Statement (d) is a consequence of the fact that there is symmetry in the system; if  $(u, v_1, v_2)$  solves (13), (14) with a given  $\omega$ , then  $(u, -v_1, -v_2)$  solves this system with  $-\omega$  since  $-\sigma dW \cong \sigma dW$  in probability law.  $\square$

The conditions in (a) and (b) guarantee that the model under consideration is geometrically ergodic [33, 17] with a smooth invariant measure. The condition in (d) has the implication that partial observations of  $u$  alone cannot uniquely determine the coefficient of the model (see [16] for even simpler

linear systems with this property). One of the numerical experiments below illustrates this phenomena.

Following [20], we consider numerical experiments with the following parameters:  $\beta = 1$ ,  $A_1 = 1$ , and  $E = 1 = \frac{\sigma_1^2}{2d_1} + \frac{\sigma_2^2}{2d_2}$ . In particular, we choose  $\sigma_1^2 = \sigma_2^2 = 1/2$ ,  $d_1 = d_2 = 1/2$  such that  $E = \frac{\sigma_j^2}{d_j} = 1$ . According to property (c) of Proposition 1 the invariant measure is Gaussian for these parameters. In the rest of the paper we will be referring to the following regimes related to different values of the topography amplitude  $H$ , corresponding to maximum height of the topography  $\max |h|$ :

Regime 1:  $H = \sqrt{2}/4$ , corresponds to  $\max |h| = 0.5$ ,

Regime 2:  $H = 2\sqrt{2}/4$ , corresponds to  $\max |h| = 1$ ,

Regime 3:  $H = 3\sqrt{2}/4$ , corresponds to  $\max |h| = 1.5$ ,

Regime 4:  $H = 8\sqrt{2}/4$ , corresponds to  $\max |h| = 4$ .

In the numerical experiments below, we consider the following observation model,

$$v_m = u_m + \sigma_m^o, \quad \sigma_m^o \sim \mathcal{N}(0, r^o), \quad (15)$$

where  $u_m \equiv u(t_m)$  are solutions of (13) with  $L, \Sigma$  defined as in (14) and  $r^o = 10\% \text{Var}(u)$ . Unless stated we set the observation time interval,  $\Delta t = t_{m+1} - t_m = 0.01$ , number of lag,  $L = 8$ , in the secondary filter (see Step 8 in the Appendix), and the ensemble size  $K = 2n$ , where  $n = 8$  denotes the dimension of the augmented state space, concatenating the physical variables  $x \in \mathbb{R}^3$  and the parameters  $\beta, \omega, A_1, d_1, d_2$ . The equilibrium statistics,

including the mean, covariance, skewness, kurtosis, autocorrelation function, correlation time, and the marginal density for  $u$  are numerically compared with time series of length 300,000 with resolution  $\Delta t = 0.01$ . On the other hand, the parameter estimation is inferring from observations defined in (15) for only up to 20,000 time units.

#### *4.1.1. Statistical estimation of the three-dimensional test model*

In Figure 10 we show the estimated marginal density and autocorrelation function for various regimes (or topographic height). Notice that as the topographic height  $H$  increases, the estimates become more accurate. We suspect that the inaccuracy with smaller  $H$  is due to stiffness of matrix  $L - D$ ; in Regime 1, the eigenvalues of  $L - D$  are  $-0.04, -0.47 \pm 1.05i$ . While in Regime 2, the eigenvalues of  $L - D$  are  $-0.15, -0.42 \pm 1.20i$ . In Table 3 we report the other statistical estimates compared to the truth, confirming the improved statistical estimates as  $H$  increases. Notice that in Regime 3 the parameter and state estimates for  $\omega, v_1, v_2$  are reflected about zero (see Figures 11-12) while the fitting on the observed data,  $u$ , is nearly identical. This is a consequence of the non-uniqueness of the parameters as described by statement (d) of Proposition 1. In Figure 13, notice also the accuracy of the stochastic parameter estimates for  $\sigma_1, \sigma_2, r^o$ , even when the estimates for  $\omega, v_1, v_2$  differs by a negative sign from the true parameters.

#### *4.1.2. Sensitivity of statistical estimates to observation time interval*

In this section, we study the sensitivity of the statistical estimates when the observation time interval (or sampling time),  $\Delta t$ , varies. Here we compare three cases  $\Delta t = 0.01, 0.1, 0.25$ , while fixing the model integration time step

$\delta t = 0.01$ . In previous simulations, we only considered a dense observation time interval with  $\Delta t = \delta t = 0.01$ .

In Figures 14 and 15 we show the parameter estimates (circles) for simulation with  $\Delta t = 0.01, 0.1, 0.25$  for Regimes 1 and 3, respectively. The values of the parameters of the true model are also provided (dotted line). We note that in Regime 1 the estimates are within 20% of the errors (see the “error-bar” plot) only for  $\Delta t = 0.1$ . In Regime 3, on the other hand, the parameter estimates are all within 20% of the errors for  $\Delta t = 0.01, 0.1, 0.25$  and we expect insensitive statistical estimates in this regime. This suggests that, for Regime 1, the selection of an appropriate sampling time (not too dense nor too sparse) is essential in order to obtain accurate statistical estimates.

In Figure 16 we compare the marginal density and autocorrelation for Regimes 1 and 3 for  $\Delta t = 0.1, 0.25$ . In Tables 4 and 5, we also compare the marginal statistics for  $u$ . From this posterior verification we confirm, what we already observed before, that the accuracy of the statistical prediction in Regime 1 depends on the sampling time. This result is consistent with those reported in [8, 34] in the context of parameterizing multiscale dynamical systems; and in [16] in the context of parameterizing linear systems. One possible explanation for the sensitivity to sampling time in Regime 1 is due to the stiffness of the resulting dynamical system. For Regime 3, on the other hand, the statistical estimates are not sensitive at all to the sampling times studied here ( $\Delta t = 0.01, 0.1, 0.25$ ).

We also point out that the estimates in the numerically stiff Regime 1 are very sensitive to model errors. Numerically we find this sensitivity when model errors are introduced through sparse integration time step, equals

to the sampling time step,  $\delta t = \Delta t$ . In our numerical experiment with  $\delta t = \Delta t = 0.25$ , the parameter estimation scheme produces a completely inaccurate estimate for  $\sigma_2 = 12.4$ , while the true parameter is  $\sqrt{0.5}$  (detailed results are omitted).

#### *4.2. Parameter estimation of low order stochastic dynamics with partially observed 57-mode topographic stress model solutions*

In this section, we fit noisy solutions  $u$  of the topographic stress model in (9), obtained from Galerkin projection of  $\psi$  in (8) with  $0 < |k| \leq \Lambda = 17$  and a single-layer topography  $h$  in (10), resulting in a total of 57-mode model [20, 28], to physics constrained reduced stochastic models. Here, the true signals are solutions of nonlinear ordinary differential equations in (9), describing complex interactions of the zonal mean flow with two topographic Rossby modes and 54 Rossby modes. Notice that this regression problem is much more difficult than that in the test model in Section 4.1 because we introduce model errors whenever we use reduced stochastic models to approximate the marginal statistical estimates of the true 57-mode model. The difficulty here is because the 57-mode model produces significantly non-Gaussian statistics with blocked and unblocked regimes of the zonal jet due to stochastic backscatter from the Rossby waves even though the perfect system exactly conserves energy. Then it becomes unclear as how to choose an appropriate reduced stochastic model that will produce reasonable statistical estimates. Below, we will demonstrate that a model with zero memory level as in (3) is not sufficient for capturing any of these crucial non-Gaussian behavior. Subsequently, we find that models with one memory level produce reasonably good estimates with similar statistical effects.

#### 4.2.1. The zero memory level model

The test model (12) for interactions of zonal jet with two topographic Rossby waves considered in Section 4.1 was derived based on a specific choice of the coefficients of function  $\psi(x, t)$  given in (11). Motivated by extensive numerical experiments, we derive a slightly modified zero memory level model by considering (10) with the coefficients of  $\psi(x, t)$  given here by

$$\psi_1(t) = a(t) - ib(t) \quad \text{and} \quad \psi_{-1} = \psi_1^*. \quad (16)$$

Inserting the ansatz in (10) and (16) to (9) and approximating the interaction with the remaining 54 truncated Rossby modes with a two-dimensional Ornstein-Uhlenbeck process, we obtain a reduced SDE model (13) with  $x = (u, a, b)^\top$ ,  $B(x, x) = (0, A_1ub, -A_1ua)^\top$ ,  $D = \text{diag}(0, d_1, d_2)^\top$ ,  $\dot{W} = (\dot{W}_1, \dot{W}_2)^\top$ ,

$$L = \begin{pmatrix} 0 & -H & -H \\ \frac{H}{2} & 0 & -\beta \\ \frac{H}{2} & \beta & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 & 0 \\ \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}. \quad (17)$$

We refer to this three-dimensional system as the **zero memory level model** to be consistent with the definition in (3).

We fit noisy observations of  $u_m$ , defined in (15), where  $u_m \equiv u(t_m)$  are solutions of 57-mode Galerkin truncation of the barotropic model in (7), resolved at every  $\Delta t = t_{m+1} - t_m = 0.25$  and  $r^o = 10\% \text{Var}(u)$ . In our parameter estimation scheme, we set the integration time step to be  $\delta t = \Delta t = 0.25$ . In our numerical experiments, we found worse estimates with smaller  $\delta t$ . From the numerical simulations in Regime 3, we find a very accurate estimate on the autocorrelation function despite inaccurate estimate on the marginal distribution of  $u$  (see Figure 17). Here, the zero level memory model in (13),

(17) produces nearly Gaussian equilibrium statistics and fails to capture the skewness (and non-Gaussian feature of the true model) completely (see Figure 17 and Table 6). In Figures 18-20, we see that the corresponding observed variable,  $u$ , fits the true signal and the parameter estimates converge to some values. We find this result to be robust even with larger observation noise variance  $r^o = 25\%Var(u)$ .

The failure in estimating the skewness suggests that the zero memory level model in (13), (17) may have very different dynamics compared to the true model. We confirm this by comparing the joint pdf's of  $(u, a)$  and  $(u, b)$  from the truth and the zero memory level models (see Figure 21). From our numerical simulations we find that in Regimes 1 and 2, the model in (13), (17) does not produce reasonable statistical estimates either (results are not shown).

The main message here is that Ornstein-Uhlenbeck process is not a good model for the interactions with the remaining 54 Rossby modes. Second, accurate estimation of equilibrium marginal statistics is a necessary but not a sufficient condition for good statistical predictive skill. This is because there are non-unique models with very different dynamics that can produce identical equilibrium marginal statistics. This nonuniqueness even occurs in the setting of partially observed linear models [16]. Thus, we will consider one level memory models in the remaining of this paper.

#### *4.2.2. The one memory level model*

Next, we consider a slight modification of the model in (17) with one level of memory terms given by linear SDEs for residual terms  $r_1, r_2$ , conditional to  $u, a, b$ . Mathematically, this modification yields a five-dimensional SDEs

in (13) with  $x = (u, a, b, r_1, r_2)^\top$ ,  $B(x, x) = (0, A_1ub, -A_1ua, 0, 0)^\top$ ,  $D = \text{diag}(0, d_1, d_2, d_3, d_4)^\top$ ,  $\dot{W} = (\dot{W}_1, \dot{W}_2)^\top$ ,

$$L = \begin{pmatrix} 0 & -H & -H & 0 & 0 \\ \frac{H}{2} & 0 & -\beta & 1 & 0 \\ \frac{H}{2} & \beta & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}. \quad (18)$$

We refer to this five-dimensional system as the **one memory level model** to be consistent with (1). Notice that there are non-unique choices of parametric form for the class of models with one memory level, e.g., one can add nonzero terms on components (4,5) and (5,4) in  $L$  in (18) that preserves the skew symmetric condition so the model remains stable. We have tried many parametric forms and found that they do not produce reasonable statistical estimates as well as the parametric form in (18) and an alternative one memory level model that we will discuss in next section. For example, in a numerical experiment with a mildest violation of the skew symmetric condition through  $L_{4,5} \neq -L_{5,4} \neq 0$ , we find that the regression model diverges in finite time; this is consistent with the physics constrained condition advocated in [16, 17].

We find that the one memory level model in (13), (18) produces reasonable statistical estimates for the autocorrelation function and other statistics, except for the kurtosis (see Figure 22 and Table 7 for detail) in Regime 2. Notice also that this one memory level model produces comparable (not perfect) statistical behavior of the 57-mode; compare the joint pdf's of  $(u, a)$  and  $(u, b)$  in Figure 23. While the results in Regime 2 are very encouraging



in the presence of model errors, such accurate statistical estimates are not achieved in the other regimes with different topographic heights,  $H$ , using the regression model in (13) with parametric form (18).

*4.2.3. An alternative one memory level model with rotated topographic Rossby modes*

Here, we consider an alternative parametric form for one memory level model by applying a  $45^\circ$  rotation on  $(a, b)$  to new coordinate  $(v_1, v_2)$  in (18). This  $45^\circ$  rotation results in a parametric form that is similar to the test model in Section 4.1. In particular, we obtain a system of five-dimensional one memory level model in the form of (13) with  $x = (u, v_2, v_1, r_1, r_2)^\top$ ,  $B(x, x) = (0, -A_1uv_1, A_1uv_2)^\top$ ,  $D = \text{diag}(0, d_1, d_2, d_3, d_4)^\top$ ,  $\dot{W} = (\dot{W}_1, \dot{W}_2)^\top$ ,

$$L = \begin{pmatrix} 0 & -2\omega & 0 & 0 & 0 \\ \omega & 0 & \beta & 1 & 0 \\ 0 & -\beta & 0 & 0 & 1 \\ 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ \sigma_1 & 0 \\ 0 & \sigma_2 \end{pmatrix}, \quad (19)$$

where  $\omega \equiv H/\sqrt{2}$ .

This model produces reasonable statistical estimates (see Figure 24 and Table 8 for detail) in the numerically stiff Regime 1 with  $H = \sqrt{2}/4$ . Furthermore, we also see reasonable joint pdf's estimates, especially for  $(u, v_2)$  (see Figure 25). The statistical accuracy in Regime 1 here is not obtainable with the zero memory level model in (13), (17) nor the previous one memory level model in (13), (18).

In comparison with the one memory level model in (13), (18), the one memory level model with parametric form (19) improves the estimates for the

mean and kurtosis while degrades the estimates for the variance, skewness, and correlation times, as well as the joint pdf's especially for  $(u, v_1)$  (see Table 9, Figures 26 and 27).

We should also point out that the estimated stochastic noise amplitude,  $\sigma_2^2$ , can be negative when the parametric form in (19) is used in Regime 1. We suspect that such non-physical parameter estimates may be due to an inappropriate choice of parametric form within the class of physics constrained models. We should point out that our algorithm is designed to avoid such nonphysical estimates by setting  $\sigma_j^2 \geq 0$  (see Step 8 in the Appendix). In our numerical results in Regime 1 that was shown in Figure 24 and Table 8, the corresponding estimated stochastic parameters  $\sigma_2 \approx 0$  (see Figure 28). This suggests that with the parametric form in (19), the physics constrained model has a complete dissipative residual term,  $r_2$ . In this system of SDE's with stochastic forcing that appears only through  $r_1$  dynamics, the controllability condition is indeed satisfied except for a parameter set of measure zero.

## 5. Summary

The objective of this paper was to solve the following problem, which often appears in various scientific disciplines that involve predictability issue:

*Given partial observations of complex dynamical systems, develop low dimensional, data driven dynamical models for statistical prediction!*

Our approach is to break this broad issue into two problems:

- (i) How to choose appropriate low-dimensional stochastic models that respect certain physical causality with accurate statistical prediction?

(ii) Once we have chosen a class of regression models, how to estimate the parameters in these models?

In this paper, our main contribution was to address problem (ii) with a new algorithm for estimating the parameters in the regression models, given only partial observations of the underlying dynamical systems. In particular, a new estimation scheme was developed based on Belanger’s noise estimation algorithm [18], blended with finite ensemble Kalman filter. This scheme is computationally efficient when the dimension of the observation variables is small [22]. This new algorithm was designed to avoid the numerical blow-up that can occur with the existing EKF based estimation scheme when observations are sparse.

We also applied the newly developed parameter estimation scheme on a class of nonlinear models for topographic mean flow interaction. In the context with no model errors (Section 4.1), we found that the estimation scheme is indeed quite robust. We also found that when the underlying dynamical systems of interest are numerically stiff, we need to sample the observations at an appropriate time (see Section 4.1.2) for accurate statistical estimation. This result is consistent with those in [8, 34] in the context of parameterizing multiscale dynamical systems and in [16] in the context of parameterizing linear systems.

In the presence of model errors, choosing an appropriate low-dimensional models becomes a nontrivial task. In our work, we chose the regression models, relying mostly on the mathematical criteria for physics constrained multi-level regression models developed in [17] to avoid non-physical blow-up solutions. In our numerical simulations involving non-Gaussian statistical

behavior resulting from nonlinear interactions of a large scale zonal jet with 56 Rossby modes, we found that the zero level memory model is not sufficient to capture these non-Gaussian statistics (see Section 4.2.1). Our numerical results suggested that we need at least one memory level in the regression models to produce reasonable statistically estimates.

While the numerical results are encouraging, we need more systematic guidelines for choosing the most appropriate parametric form within the class of physics constrained multi-level models. Notice that the results in Section 4.2 were based on exhaustive numerical trials of non-unique parametric forms of regression models with zero and one memory levels. We plan to study this issue in the near future. Finally, while the algorithm that we introduced here is quite robust and useful, ideally one would want to estimate the distribution of the parameters. In that situation, it will be important to consider efficient Markov Chain Monte Carlo (MCMC) methods for parameter estimation.

### **Acknowledgment**

This project is primarily supported by the ONR MURI Grant N00014-12-1-0912. The research of J.H. is also partially supported by the ONR grant N00014-11-1-0310. The authors thank Christian Franzke for providing the time series of the 57-mode topographic stress model.

### **Appendix: Stochastic parameter estimation blending Belanger’s method with Extended Kalman Filter**

This Appendix provides a detailed review of application of Belanger’s method [18] with the classical EKF [23]. Consider the following filtering

problem,

$$\begin{aligned}x_k &= f(x_{k-1}) + \Gamma\epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, Q), \\y_k &= Hx_k + \epsilon_k^o, \quad \epsilon_k^o \sim \mathcal{N}(0, R),\end{aligned}$$

where  $x_k = (u_k, \theta_k)^\top$  consists of the augmented state variables,  $x_k$ , and parameters,  $\theta_k$ . In our case, we assume a persistent model for the parameters, that is  $\theta_{k+1} = \theta_k$ . We attempt to estimate  $x_k$  as well as  $Q$  and  $R$ , on-the-fly.

To solve this problem, define

$$\begin{aligned}Q &= \sum_{i=1}^p Q_i a_i, \\R &= \sum_{i=1}^p R_i a_i.\end{aligned}$$

and our aim is to estimate  $a_i$ ,  $i = 1, \dots, p$ .

Start with time index  $k = 1$ , we provide prior statistical estimates,  $x_k^-, P_k^- = \mathcal{I}$ , for the primary filter and  $\vec{a}_k = (a_{1,k}, \dots, a_{p,k})$ , and  $\Theta_k = \mathcal{I}_p$ , for the secondary filter. The primary filter for estimating  $x_k$  is described in Steps 1 and 2, while the secondary filter for estimating  $\vec{a}_k$  is described in Steps 3-8.

1. Apply the Kalman filter,

$$\begin{aligned}v_k &= y_k - Hx_k^- \\ \tilde{K}_k &= P_k^- H^\top (HP_k^- H^\top + \sum_{i=1}^p R_i a_{i,k})^{-1} \\ P_k^+ &= (\mathcal{I} - \tilde{K}_k H) P_k^- \\ x_k^+ &= x_k^- + \tilde{K}_k v_k.\end{aligned}$$

2. Define linear tangent model,  $A_k \equiv \nabla f(x_k^+)$  and compute  $x_{k+1}^-$  and

$$P_{k+1}^- = A_k P_k^+ A_k^\top + \sum_{i=1}^p Q_i a_{i,k}$$

3. Define

$$\begin{aligned} K_k &= A_k \tilde{K}_k, \\ \phi_k &= A_k - K_k H. \end{aligned}$$

4. For each  $i$ , construct observation operator for  $v_k v_k^\top$ , starting with  $S_{i,1,0} = 0$ , let

$$\begin{aligned} M_{i,k,0} &= S_{i,k,0} H^\top, \\ F_{i,k,0} &= H M_{i,k,0} + R_i, \\ S_{i,k+1,0} &= \phi_k S_{i,k,0} \phi_k^\top + \Gamma Q_i \Gamma^\top + K_k R_i K_k^\top. \end{aligned}$$

5. For each  $i$ , construct observation operator for  $v_k v_{k-\ell}^\top$ , where  $k > 1$ . Set

$$\begin{aligned} M_{i,k,\ell} &= \phi_{k-1} M_{i,k-1,\ell-1} - K_{k-1} R_i \delta_{\ell,1} \\ F_{i,k,\ell} &= H M_{i,k,\ell} \end{aligned}$$

6. Approximate  $\mathbb{E}(v_k v_k^\top)$  with

$$\mathbb{E}(v_k v_k^\top) = \sum_{i=1}^p F_{i,k,0} a_{i,k}.$$

Suppose if  $v_k = (v_k^1, \dots, v_k^m)^\top$  is  $m$ -dimensional. Define

$$\begin{aligned} \sigma_{k,\ell} \equiv \text{vec}(v_k v_{k-\ell}^\top) &= (v_k^1 v_{k-\ell}^1, v_k^2 v_{k-\ell}^1, \dots, v_k^m v_{k-\ell}^1, \\ &\quad \dots, v_k^1 v_{k-\ell}^m, v_k^2 v_{k-\ell}^m, \dots, v_k^m v_{k-\ell}^m)^\top. \end{aligned}$$

7. Consider the observation model,

$$\sigma_{k,\ell} = \mathcal{F}_{k,\ell} \vec{a} + \eta_{k,\ell}, \quad \eta_{k,\ell} \sim \mathcal{N}(0, W_{k,\ell}),$$

where in our case,  $\sigma_{k,\ell} = \text{vec}(v_k v_{k-\ell}^\top) \in \mathbb{R}^+$ ,  $\mathcal{F}_{k,\ell} = (F_{1,k,\ell}, \dots, F_{p,k,\ell})$ ,  $\vec{a} = (a_1, a_2, \dots, a_p)^\top$  and for each pair of indices  $\{k, \ell\}$ , construct

$$W_{k,\ell} = \mathbb{E}(v_k v_k^\top) \mathbb{E}(v_{k-\ell} v_{k-\ell}^\top) + \mathbb{E}(v_k v_k^\top)^2 \delta_{\ell,0}.$$

Note that  $W$  is constructed, assuming Gaussian and independent noises,  $\eta_{k,\ell}$ . Components of matrix  $W$  in (20) can be rewritten as follows,

$$W_{k,l}^{\alpha,\beta,\gamma,\delta} = \mathbb{E}(v_k^\alpha v_k^\gamma) \mathbb{E}(v_{k-\ell}^\beta v_{k-\ell}^\delta) + \mathbb{E}(v_k^\alpha v_k^\delta) \mathbb{E}(v_k^\beta v_k^\gamma) \delta_{\ell,0}.$$

8. Define a counter

$$t \equiv t(k, \ell) = (k-1)(L+1) + \ell,$$

and perform correction for  $\vec{a}_k$  in sequential fashion for  $t = 1, \dots, L+1$ . Below, replace subscript- $(k, \ell)$  with  $t$ . Now, perform sequential update for  $a$  and  $\Theta$ , to obtain

$$\begin{aligned} \vec{a}_{k+1} &= \tilde{a}_{L+1}, \\ \Theta_{k+1} &= R_{L+1}, \end{aligned}$$

where

$$\begin{aligned} R_{t+1} &= R_t - R_t \mathcal{F}_t^\top (W_t + \mathcal{F}_t R_t \mathcal{F}_t^\top)^{-1} \mathcal{F}_t R_t, \\ \tilde{a}_{t+1} &= \tilde{a}_t + R_{t+1} \mathcal{F}_t^\top W_t^{-1} (\sigma_t - \mathcal{F}_t \tilde{a}_t), \end{aligned}$$

are computed with initial conditions,

$$\begin{aligned} \tilde{a}_1 &= \vec{a}_k, \\ R_1 &= \Theta_k. \end{aligned}$$

To ensure positive definite covariance estimates for  $Q$  and  $R$ , we set

$$\vec{a}_{k+1} = \max(\vec{a}_{k+1}, 0). \quad (20)$$

Now we can repeat Step 1 above for the new assimilation time.

## References

- [1] I. Horenko, Finite element approach to clustering of multidimensional time series, *SIAM J. Sci. Comp.* 32 (2010) 68–83.
- [2] I. Horenko, On the identification of nonstationary factor models and their application to atmospheric data analysis, *Journal of the Atmospheric Sciences* 67 (2010) 1559–1574.
- [3] D. Crommelin, E. Vanden-Eijnden, Subgrid-scale parameterization with conditional Markov chains, *J. Atmos. Sci.* 65 (2008) 2661–2675.
- [4] M. Priestly, *Non-linear and non-stationary time series analysis*, Academic Press, London, UK, 1988.
- [5] H. Tong, *Nonlinear time series: A dynamical systems approach*, Oxford University Press, Oxford, UK, 1990.
- [6] N. Cressie, C. Wikle, *Statistics for spatio-temporal data*, Wiley, New Jersey, 2011.
- [7] G. Pavliotis, A. Stuart, *Multiscale Methods: Averaging and Homogenization*, volume 53 of *Texts in Applied Mathematics*, Springer, 2000.



- [8] R. Azencott, A. Beri, I. Timofeyev, Adaptive sub-sampling for parametric estimation of gaussian diffusions, *Journal of Statistical Physics* 139 (2010) 1066–1089.
- [9] A. Majda, C. Franzke, A. Fischer, D. Crommelin, Distinct metastable atmospheric regimes despite nearly gaussian statistics: A paradigm model, *Proceedings of the National Academy of Sciences* 103 (2006) 8309–8314.
- [10] C. Franzke, I. Horenko, A. Majda, R. Klein, Systematic metastable atmospheric regime identification in an agcm, *Journal of the Atmospheric Sciences* 66 (2009) 1997–2012.
- [11] K. Janes, M. Yaffe, Data-driven modelling of signal-transduction networks., *Nat Rev Mol Cell Biol.* 11 (2006) 820 – 828.
- [12] S. Kravtsov, D. Kondrashov, M. Ghil, Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability, *Journal of Climate* 18 (2005) 4404–4424.
- [13] D. Kondrashov, S. Kravtsov, A. W. Robertson, M. Ghil, A hierarchy of data-based enso models, *Journal of Climate* 18 (2005) 4425–4444.
- [14] D. Kondrashov, S. Kravtsov, M. Ghil, Empirical mode reduction in a model of extratropical low-frequency variability, *Journal of the Atmospheric Sciences* 63 (2006) 1859–1877.
- [15] C. Wikle, M. Hooten, A general science-based framework for dynamical spatio-temporal models, *TEST* 19 (2010) 417–451.

- [16] A. Majda, Y. Yuan, Fundamental limitations of ad hoc linear and quadratic multi-level regression models for physical systems, *Discrete and Continuous Dynamical Systems B* 17 (2012) 1333–1363.
- [17] A. Majda, J. Harlim, Physics constrained nonlinear regression models for time series., *Nonlinearity* 26 (2013) 201–217.
- [18] P. Belanger, Estimation of noise covariance matrices for a linear time-varying stochastic process, *Automatica* 10 (1974) 267 – 275.
- [19] A. Majda, I. Timofeyev, Remarkable statistical behavior for truncated burgers-hopf dynamics, *Proceedings of the National Academy of Sciences* 97 (2000) 12413–12417.
- [20] A. Majda, I. Timofeyev, E. Vanden-Eijnden, Systematic strategies for stochastic mode reduction in climate, *Journal of the Atmospheric Sciences* 60 (2003) 1705–1722.
- [21] A. Majda, I. Timofeyev, Statistical mechanics for truncations of the Burgers-Hopf equation: A model for intrinsic behavior with scaling, *Milan Journal of Mathematics* 70 (2002) 39–96.
- [22] D. Dee, S. Cohn, A. Dalcher, M. Ghil, An efficient algorithm for estimating noise covariances in distributed systems, *Automatic Control, IEEE Transactions on* 30 (1985) 1057 – 1065.
- [23] K. Strounine, S. Kravtsov, D. Kondrashov, M. Ghil, Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance, *Physica D: Nonlinear Phenomena* 239 (2010) 145 – 166.

- [24] A. Majda, J. Harlim, *Filtering Complex Turbulent Systems*, Cambridge University Press, UK, 2012.
- [25] T. Berry, T. Sauer, Adaptive ensemble Kalman filtering of nonlinear systems, *Tellus A* (in press) (2013).
- [26] C. Bishop, B. Etherton, S. Majumdar, Adaptive sampling with the ensemble transform Kalman filter part I: the theoretical aspects, *Monthly Weather Review* 129 (2001) 420–436.
- [27] G. Carnevale, J. Frederiksen, Nonlinear stability and statistical mechanics of flow over topography, *Journal of Fluid Mechanics* 175 (1987) 157–181.
- [28] A. Majda, X. Wang, *Nonlinear dynamics and statistical theories for basic geophysical flows*, Cambridge University Press, UK, 2006.
- [29] M. J. Grote, A. J. Majda, C. Grotta Ragazzo, Dynamic mean flow and small-scale interaction through topographic stress, *Journal of Nonlinear Science* 9 (1999) 89–130.
- [30] J. Charney, J. DeVore, Multiple flow equilibria in the atmosphere and blocking, *Journal of the Atmospheric Sciences* 36 (1979) 1205–1216.
- [31] A. Majda, I. Timofeyev, E. Vanden-Eijnden, A mathematical framework for stochastic climate models, *Comm. Pure Appl. Math.* 54 (2001) 891–974.
- [32] A. Majda, I. Timofeyev, E. Vanden-Eijnden, A priori tests of a stochastic mode reduction strategy, *Physica D* 170 (2002) 206–252.

- [33] J. Mattingly, A. Stuart, D. Higham, Ergodicity for sdes and approximations: locally lipschitz vector fields and degenerate noise, *Stochastic Processes and their Applications* 101 (2002) 185 – 232.
- [34] G. Pavliotis, A. Stuart, Parameter estimation for multiscale diffusions, *J. Stat. Phys.* 127 (2007) 741–781.

variable	mean	variance	correlation time
$u_1$	-0.0004 - 0.0073i	0.0901	13.6284 (6.3701)
$u_2$	-0.0051 - 0.0136i	0.0994	3.6426 (3.4758)
$r$	-0.0025 - 0.0039i	0.1066	0.2793 (0.2734)

Table 1: Long time averaged statistics

statistics	truth	EKF ( $u_1, u_2$ )	EnKF ( $u_1, u_2$ )	EnKF ( $u_1$ )
$Re\{\mathcal{T}_{u_1}\}$	7.0940	7.0845	6.9049	7.9594
$Im\{\mathcal{T}_{u_1}\}$	7.0857	7.0752	7.0799	7.9374
$Re\{\mathcal{T}_{u_2}\}$	3.4624	3.4298	3.6103	3.6103
$Re\{\mathcal{T}_{u_2}\}$	3.4966	3.4629	3.3945	3.3945
$Var(u_1)$	0.0957	0.0955	0.0955	0.0888
$Var(u_2)$	0.0980	0.0918	0.0994	0.0994

Table 2: Correlation time,  $\mathcal{T}$ , and variance estimates for  $u_1$  and  $u_2$  from EKF and EnKF based estimation schemes. In brackets we indicate the observed variables in our estimation simulations.

Regime 1:  $H = \sqrt{2}/4$ .

statistics	truth	estimate
Mean	-0.0045	0.0085
Variance	0.2509	0.5176
Skewness	0.0027	-0.0460
Kurtosis	3.0277	2.9796
Corr time	17.5401	24.7415

Regime 2:  $H = 2\sqrt{2}/4$

statistics	truth	estimate
Mean	0.0041	0.0085
Variance	0.2515	0.2434
Skewness	0.0060	0.0617
Kurtosis	2.971	3.0214
Corr time	5.6839	5.0854

Regime 3:  $H = 3\sqrt{2}/4$ .

statistics	truth	estimate
Mean	0.0032	-0.010300
Variance	0.2540	0.2828
Skewness	0.0060	-0.0440
Kurtosis	2.9878	2.9904
Corr time	2.4503	2.5290

Regime 4:  $H = 8\sqrt{2}/4$

statistics	truth	estimate
Mean	0.0022	-0.0011
Variance	0.2947	0.3080
Skewness	0.0268	0.0047
Kurtosis	2.9913	2.9991
Corr time	0.2803	0.2605

Table 3: Marginal Statistics for  $u$  of the three-dimensional test model (13), (14) for observation (sampling) time interval  $\Delta t = 0.01$ .

statistics	truth	estimate ( $\Delta t = 0.1$ )	estimate ( $\Delta t = 0.25$ )
Mean	-0.0045	-0.0433	-0.0003
Variance	0.2509	0.2759	0.3627
Skewness	0.0027	-0.1448	0.0325
Kurtosis	3.0277	2.8622	2.9388
Corr time	17.5401	19.0139	20.6745

Table 4: Marginal statistics for  $u$  of the three-dimensional test model (13), (14) in Regime 1 for observation time step  $\Delta t = 0.1, 0.25$ .



statistics	truth	estimate ( $\Delta t = 0.1$ )	estimate ( $\Delta t = 0.25$ )
Mean	0.0032	0.0104	0.0116
Variance	0.2540	0.2809	0.2621
Skewness	0.0060	0.0876	0.0758
Kurtosis	2.9878	3.0077	3.0074
Corr time	2.4503	2.6881	2.2359

Table 5: Marginal statistics for  $u$  of the three-dimensional test model (13), (14) in Regime 3 for observation time step  $\Delta t = 0.1, 0.25$ .

statistics	truth	estimate
Mean	-0.3889	-0.2717
Variance	0.3586	0.3514
Skewness	0.5821	-0.0179
Kurtosis	2.8483	3.0375
Corr time	2.5068	3.5166

Table 6: Statistics for the zero memory level model (13), (17) in Regime 3.

statistics	truth	estimate
Mean	-0.3860	-0.3059
Variance	0.3515	0.3529
Skewness	0.6128	0.6731
Kurtosis	2.8817	4.5402
Corr time	5.5442	5.0872

Table 7: Statistics for the one memory level model (13), (18) in Regime 2.

statistics	truth	estimate
Mean	-0.3818	-0.2985
Variance	0.3407	0.2950
Skewness	0.6098	0.5877
Kurtosis	2.8968	3.4942
Corr time	16.9594	14.7060

Table 8: Statistics for the one memory level model (13), (19) in Regime 1.

statistics	truth	estimate
Mean	-0.3860	-0.3437
Variance	0.3515	0.3138
Skewness	0.6128	0.4853
Kurtosis	2.8817	3.3205
Corr time	5.5442	4.3553

Table 9: Statistics for the one memory level model (13), (19) in Regime 2.

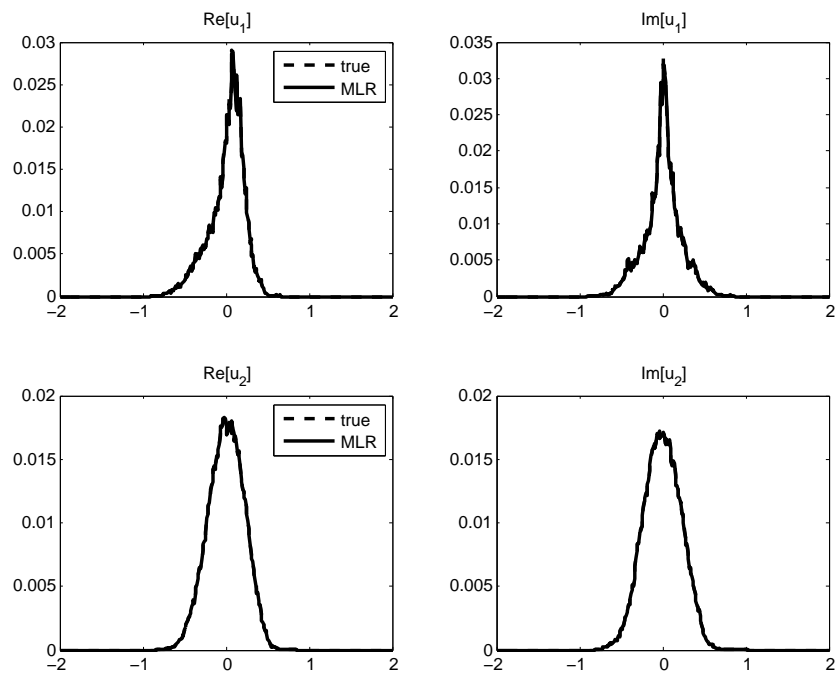


Figure 1: Marginal densities for the real and imaginary part of  $u_1$  and  $u_2$  from EKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (dashes) and estimates (solid); the two curves are on top of each other.

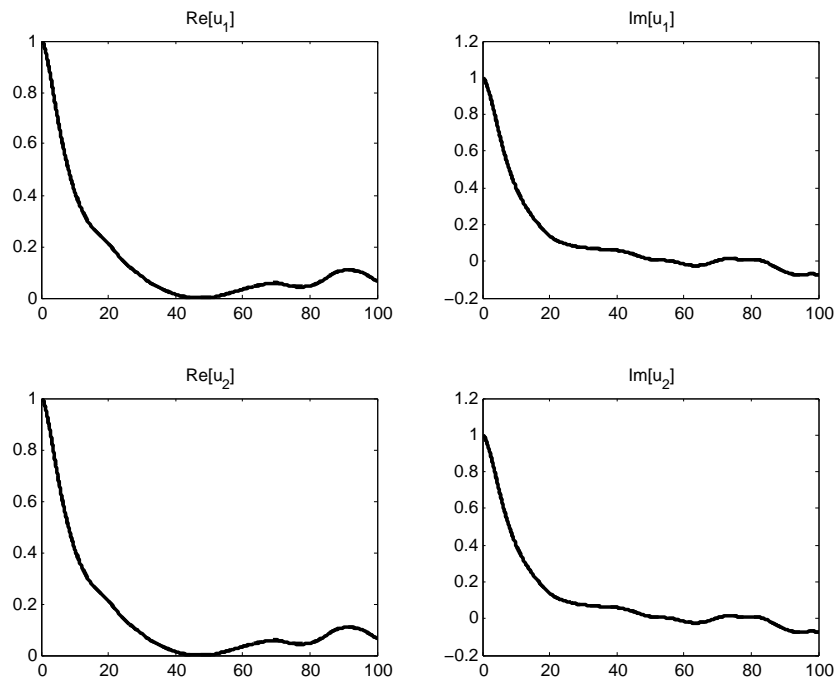


Figure 2: Autocorrelation functions for the real and imaginary part of  $u_1$  and  $u_2$  from EKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (dashes) and estimates (solid); the two curves are on top of each other.

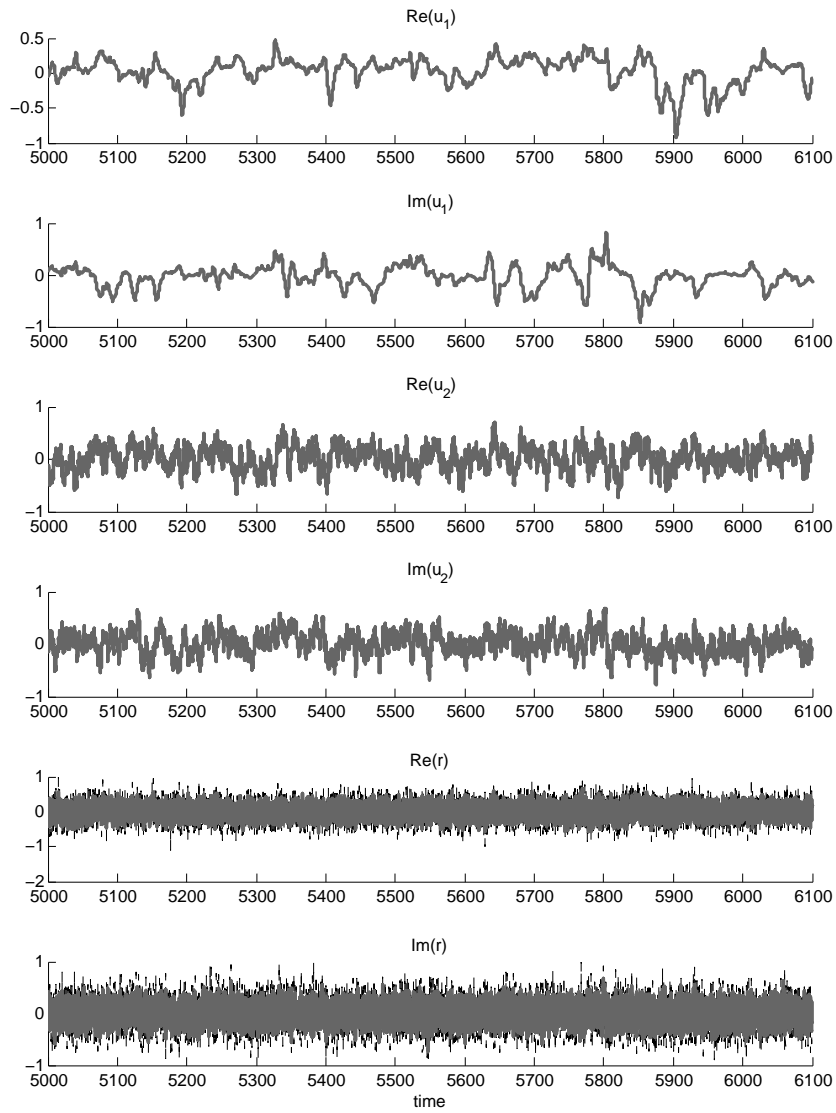


Figure 3: State estimates from EKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (black dashes) and estimates (grey).



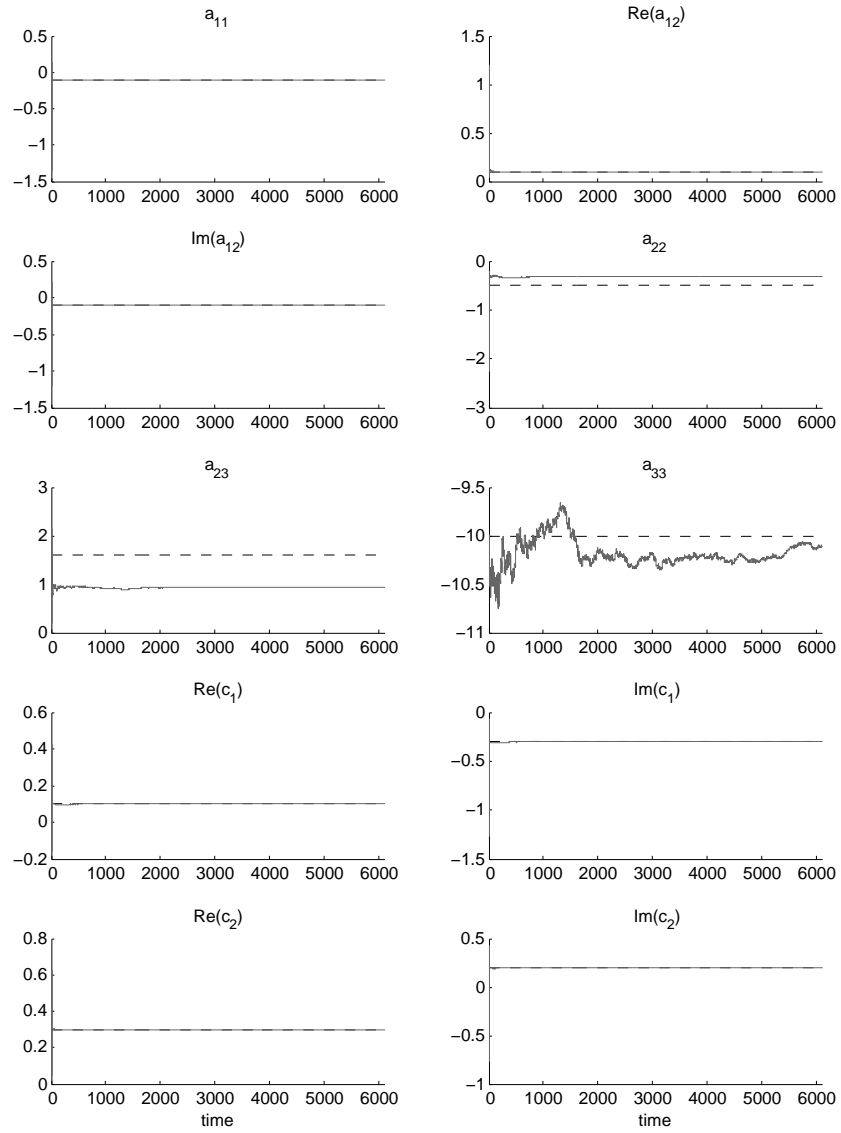


Figure 4: Deterministic parameter estimates (components of  $A$  and  $c_1, c_2$ ) from EKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (black dashes) and estimates (grey).

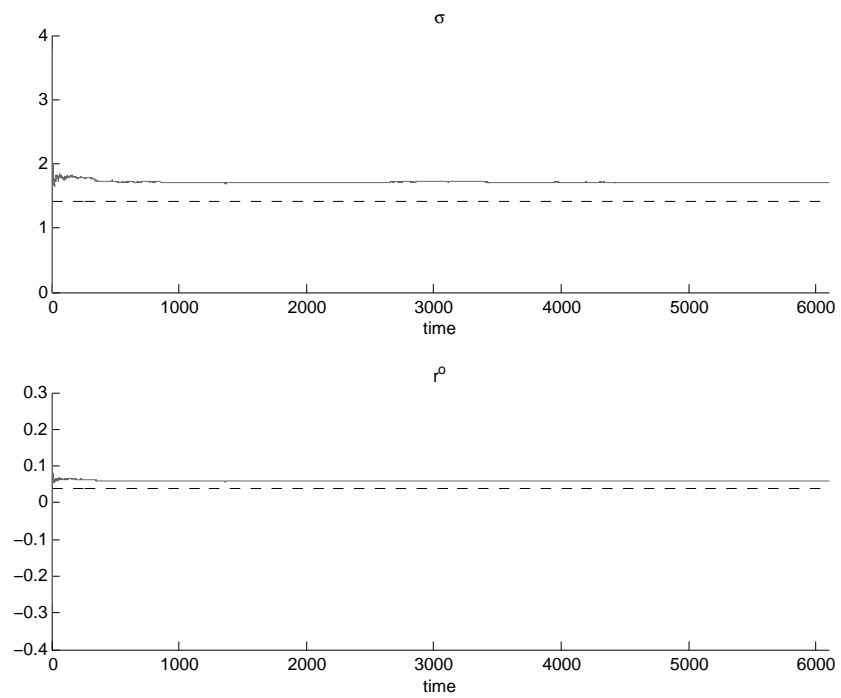


Figure 5: Stochastic parameter estimates from EKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (black dashes) and estimates (grey).

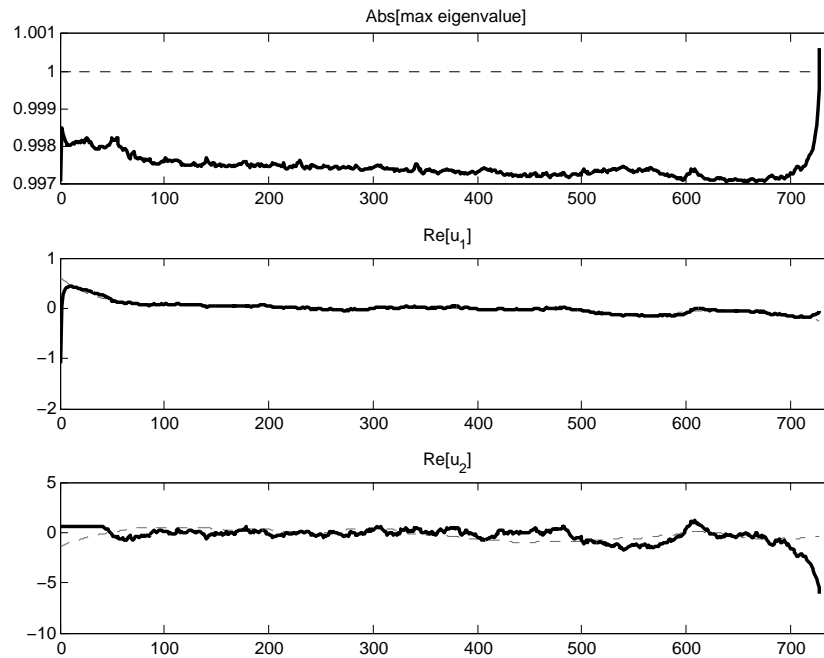


Figure 6: The evolution of the largest magnitude of the eigenvalue of the linear tangent model used in the EKF based parameter scheme from observing only  $u_1$  (top panel). In the last two panels, we plot the state estimates of the real part of  $u_1$  and  $u_2$  as functions of the assimilation step. The plot shows the first 728 iterations before the scheme diverges.

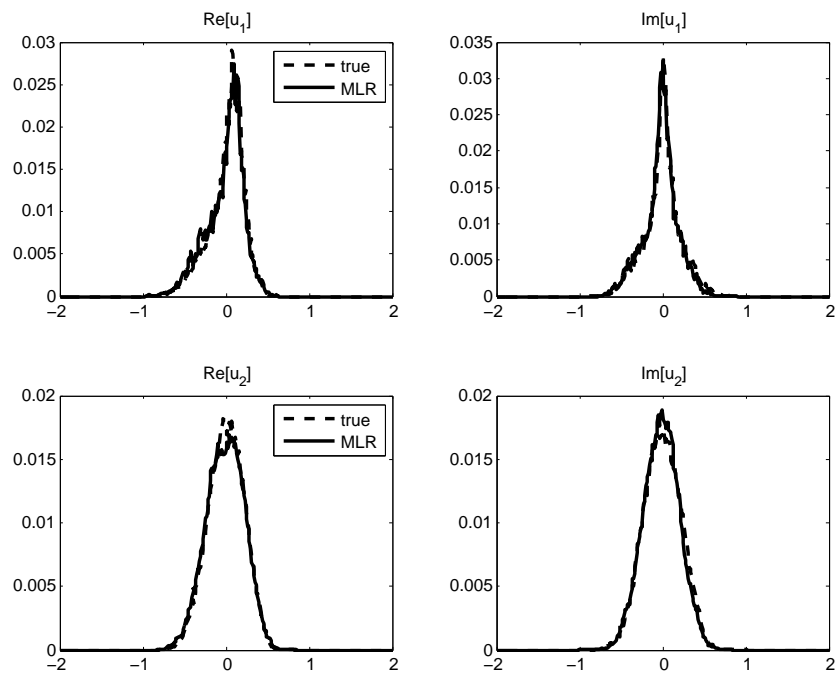


Figure 7: Marginal densities for the real and imaginary part of  $u_1$  and  $u_2$  from EnKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (dashes) and estimates (solid).

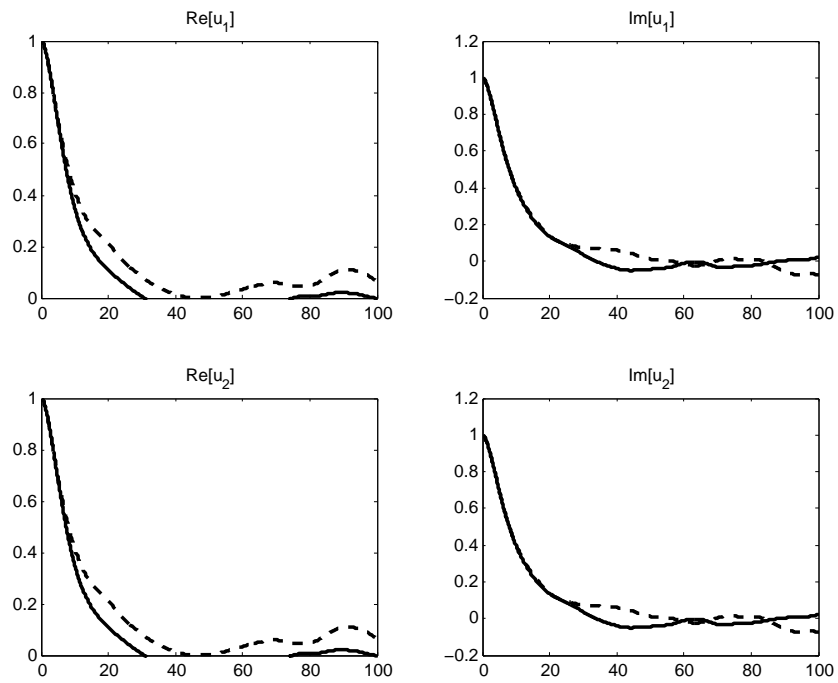


Figure 8: Autocorrelation functions for the real and imaginary part of  $u_1$  and  $u_2$  from EKF based parameter estimation scheme with noisy observations of  $u_1$  and  $u_2$ . Truth (dashes) and estimates (solid).

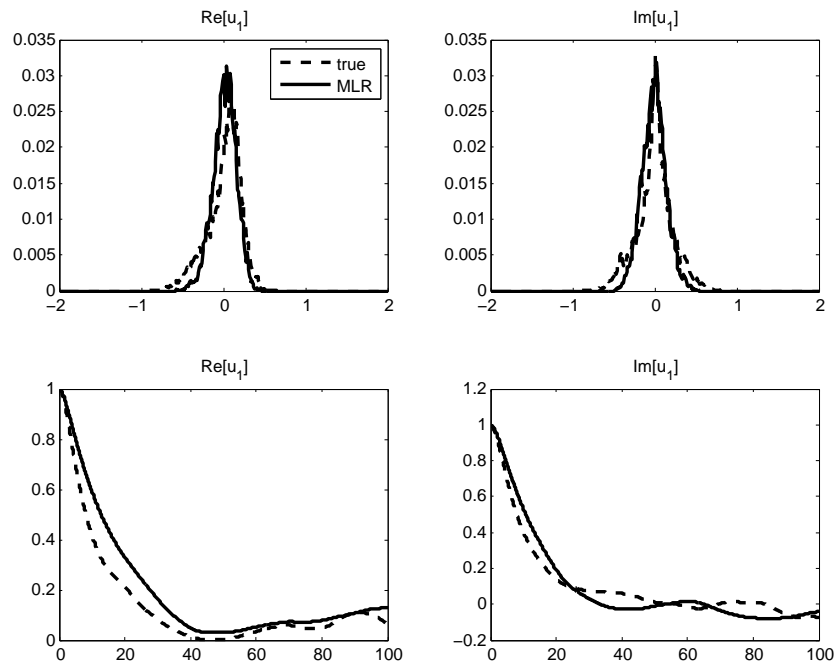


Figure 9: Marginal densities for the real and imaginary part of  $u_1$  from EnKF based parameter estimation scheme with noisy observations of  $u_1$ . Truth (dashes) and estimates (solid).

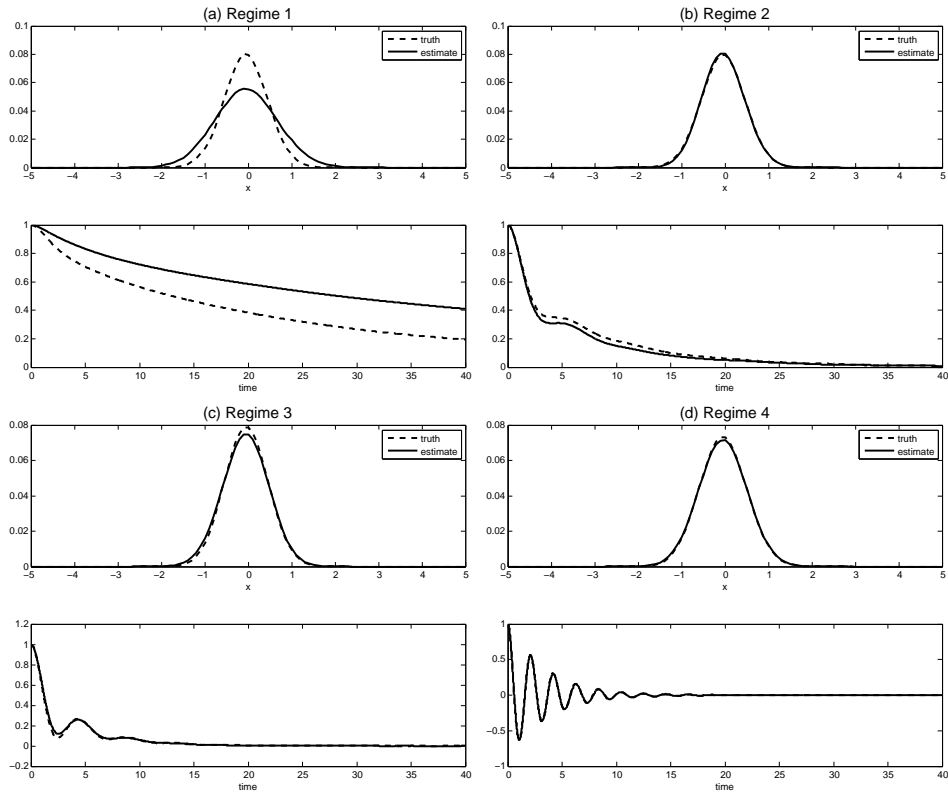


Figure 10: Marginal density (or histogram) and autocorrelation function of the three-dimensional test model (13), (14) in all four regimes for observation (sampling) time interval  $\Delta t = 0.01$ .

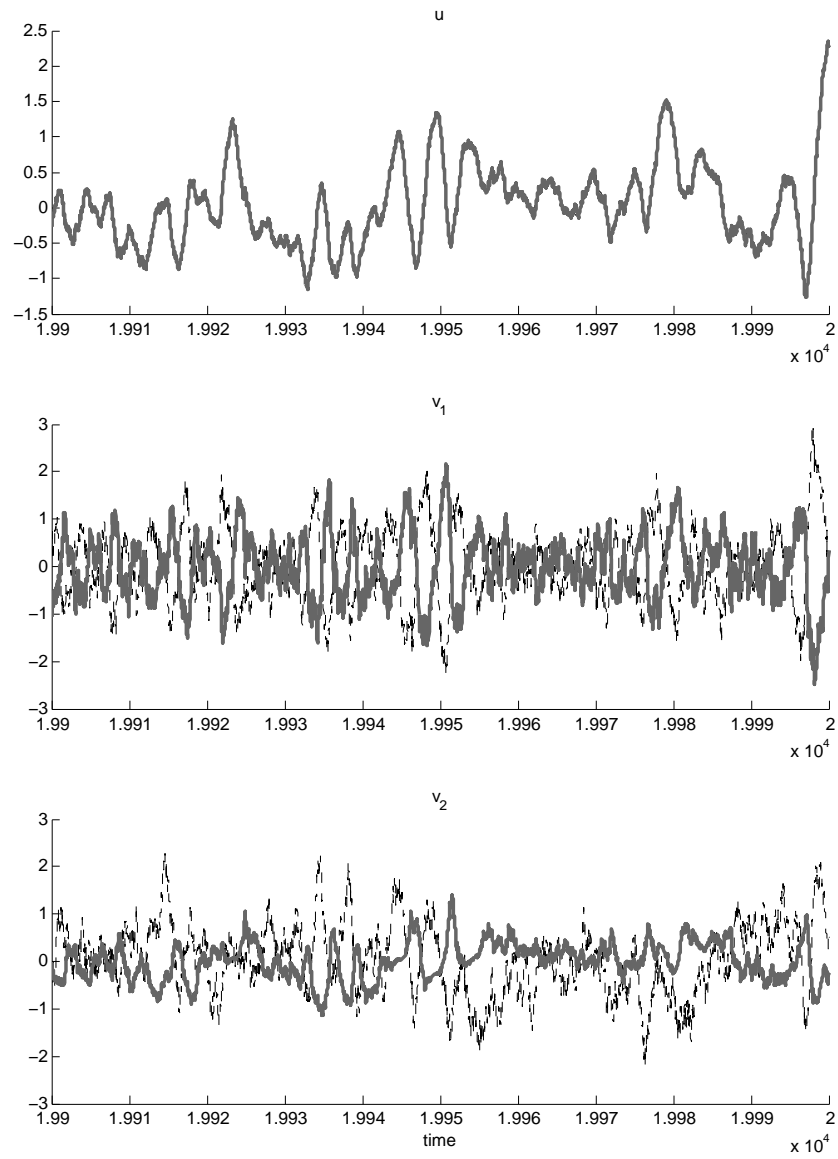


Figure 11: State estimates of the three-dimensional test model (13), (14) in Regime 3.



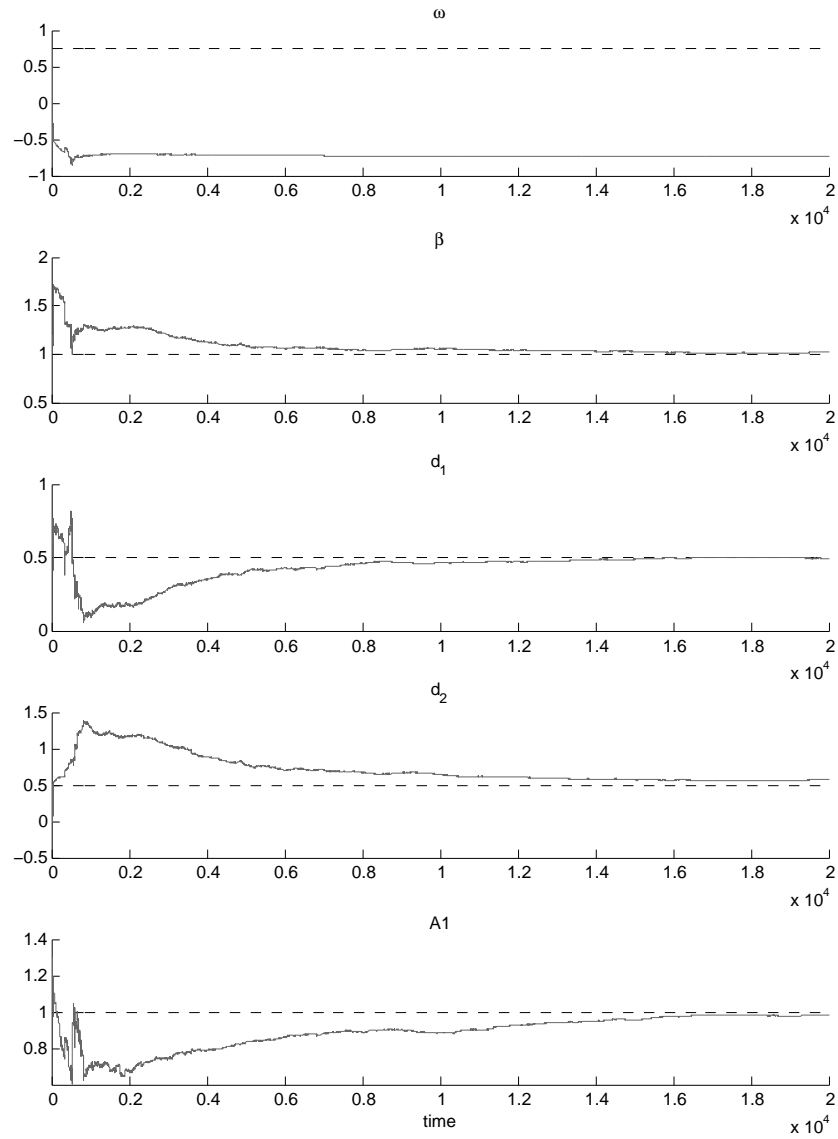


Figure 12: Deterministic parameter estimates of the three-dimensional test model (13), (14) in Regime 3.

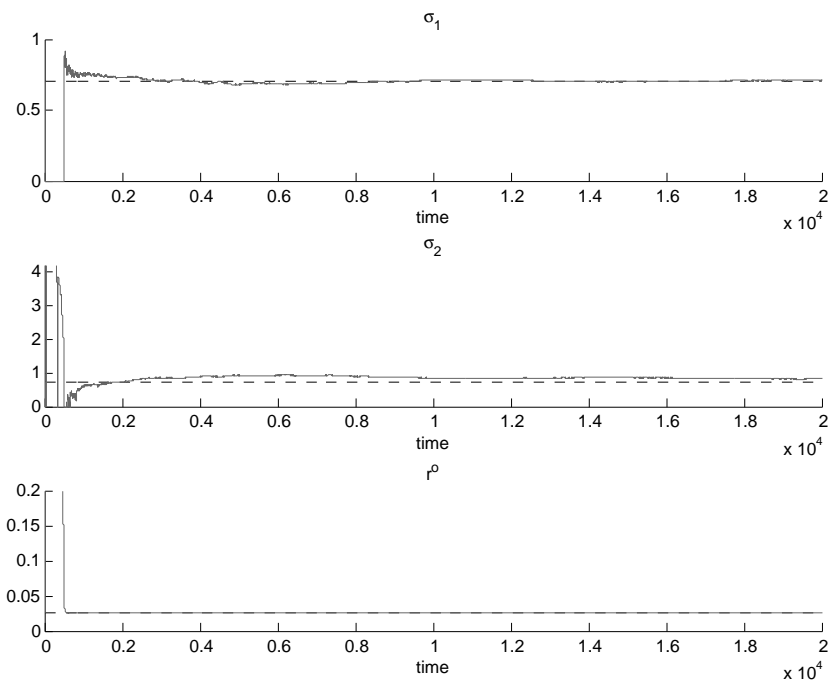


Figure 13: Stochastic parameter estimates of the three-dimensional test model (13), (14) in Regime 3.

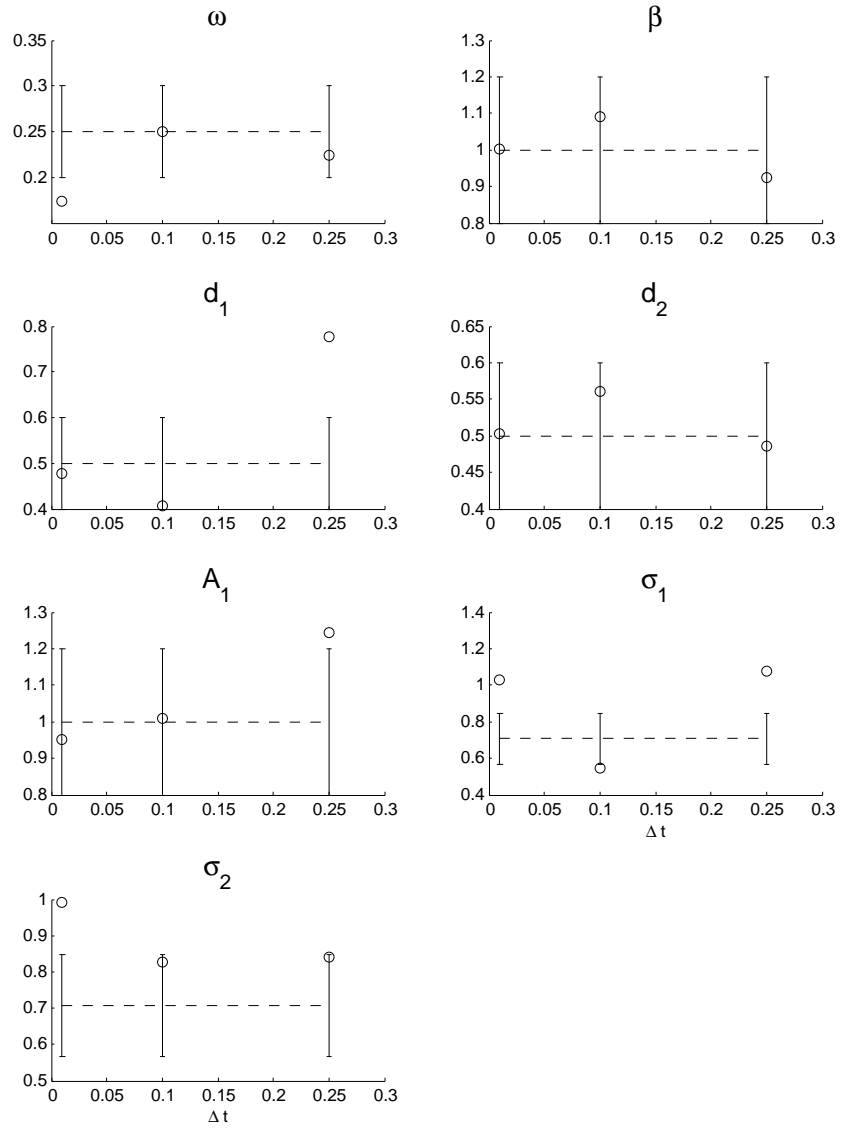


Figure 14: Parameter estimates of the three-dimensional test model (13), (14) as functions of observation time step,  $\Delta t$  in Regime 1. The true parameters are plotted in dashes with 20% of “error-bar” plot. The estimates are denoted in circles.

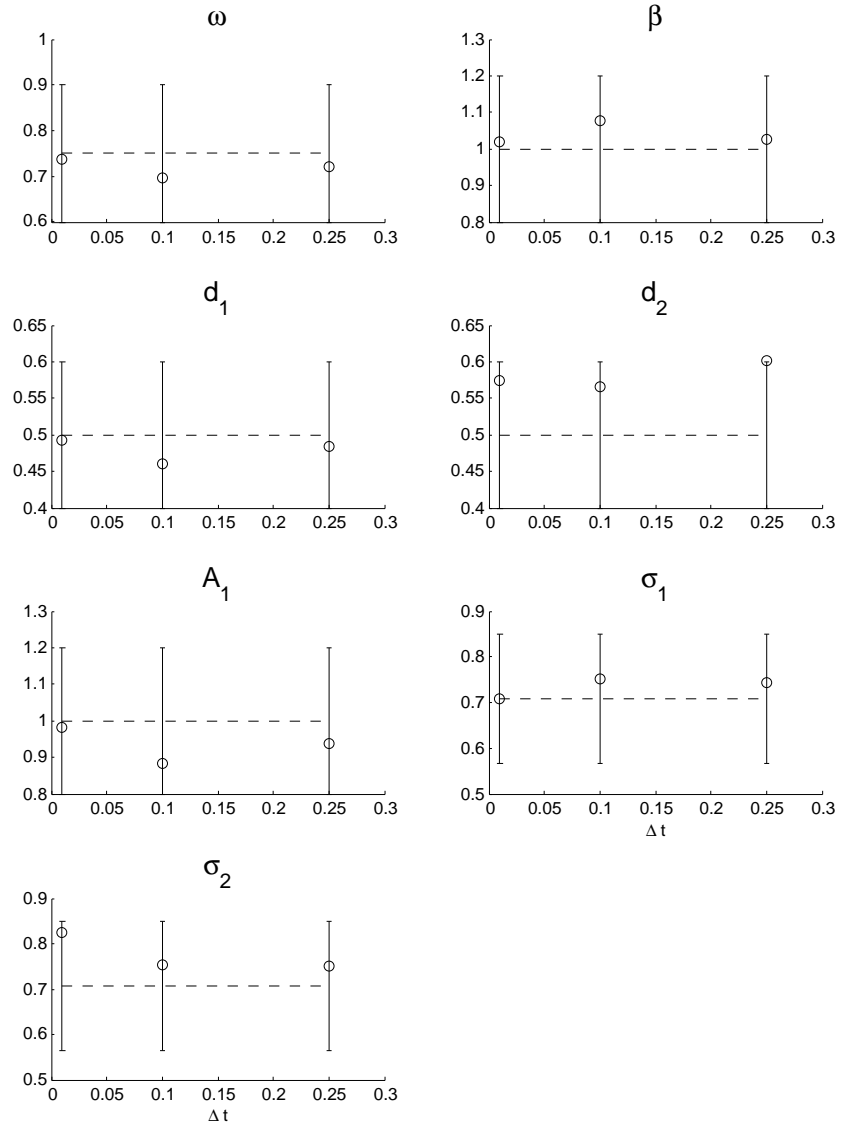


Figure 15: Parameter estimates of the three-dimensional test model (13), (14) as functions of observation time step,  $\Delta t$  in Regime 3. The true parameters are plotted in dashes with 20% of “error-bar” plot. The estimates are denoted in circles.

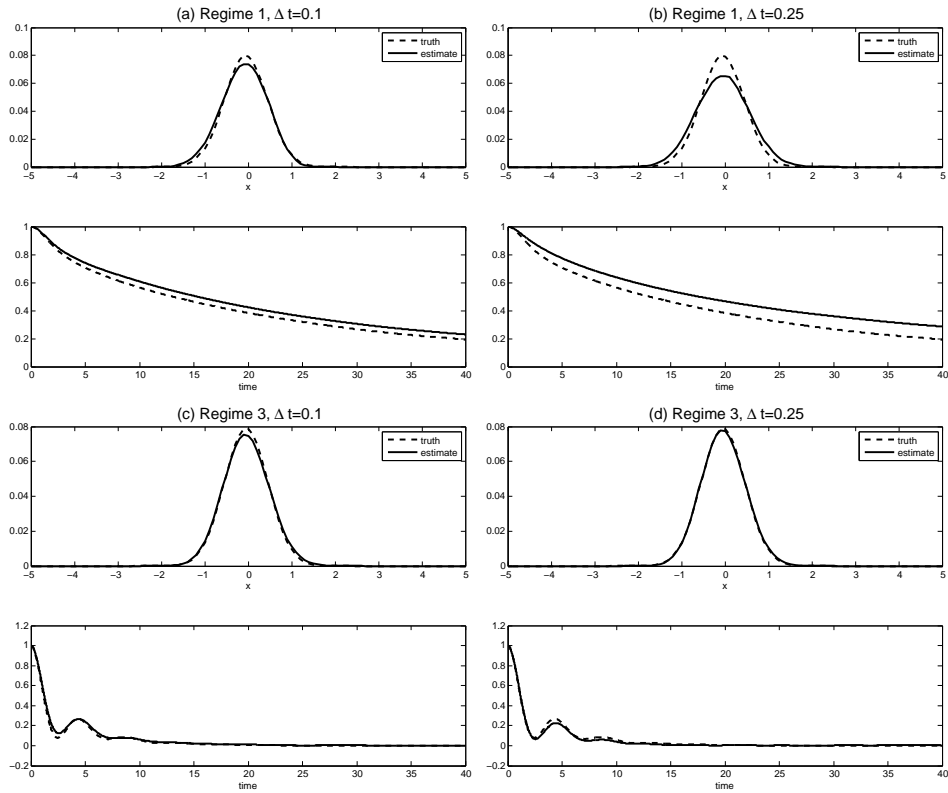


Figure 16: Density (or histogram) and autocorrelation function of the three-dimensional test model (13), (14) in Regimes 1 and 3 for observation (sampling) time intervals  $\Delta t = 0.1, 0.25$ .

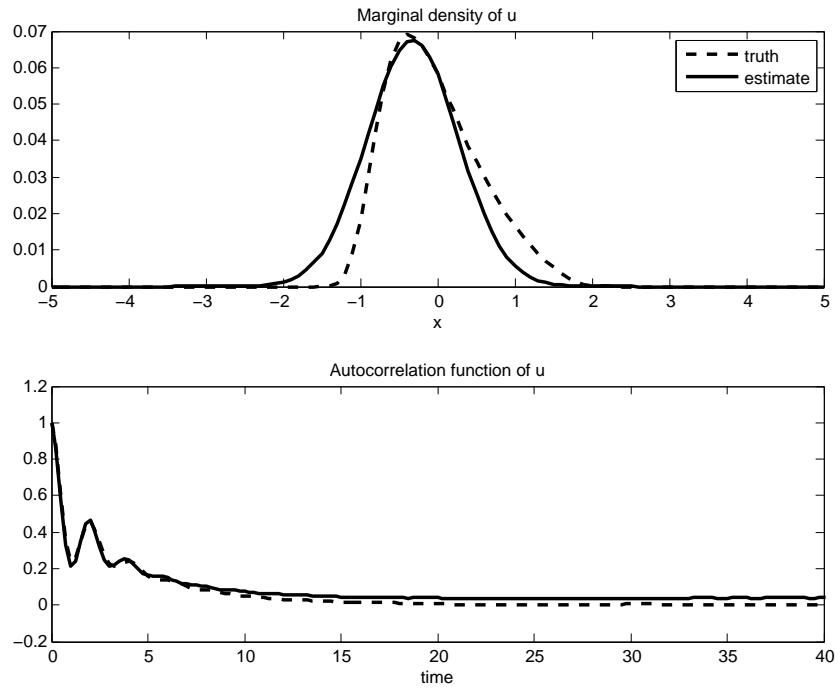


Figure 17: Marginal density (or histogram) and autocorrelation function of the solutions for the zero memory level model in (13), (17) in Regime 3.

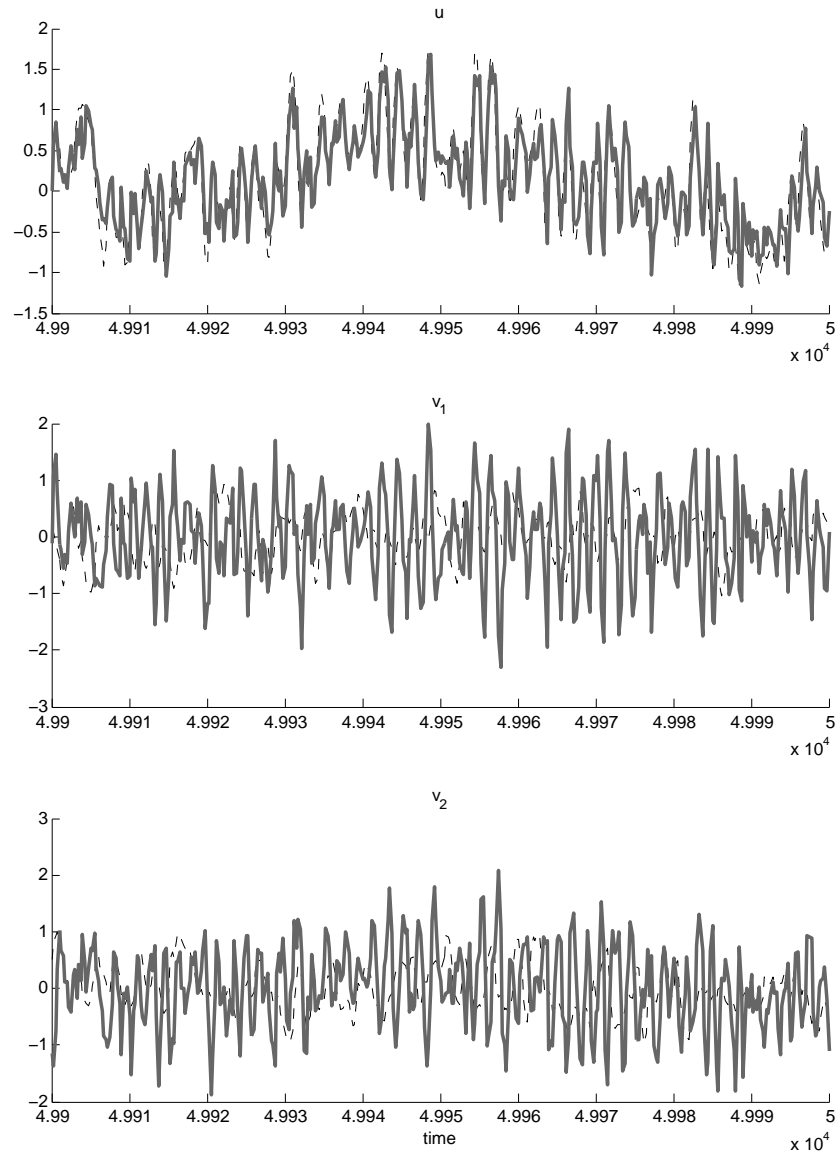


Figure 18: State estimates for the zero memory level model in (13), (17) in Regime 3.

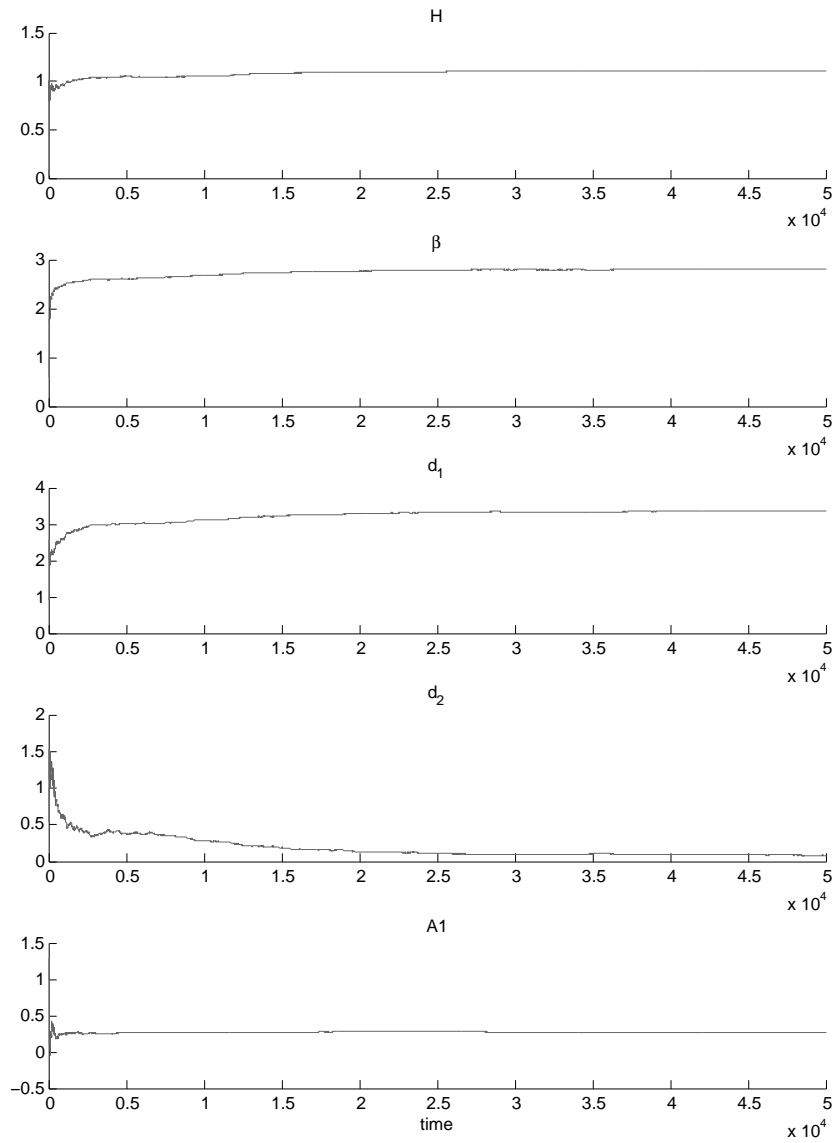


Figure 19: Deterministic parameter estimates for the zero memory level model in (13), (17) in Regime 3.



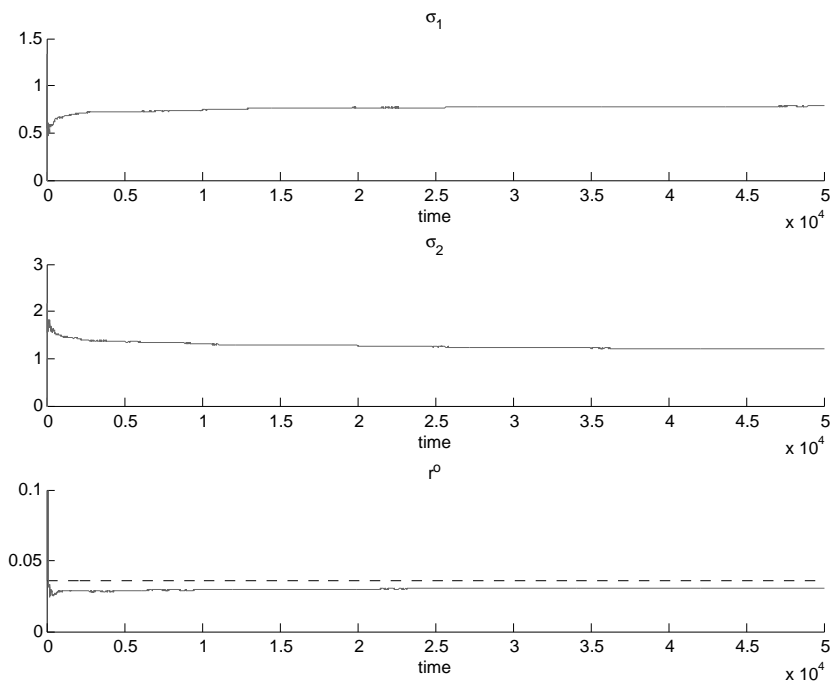


Figure 20: Stochastic parameter estimates for the zero memory level model in (13), (17) in Regime 3.

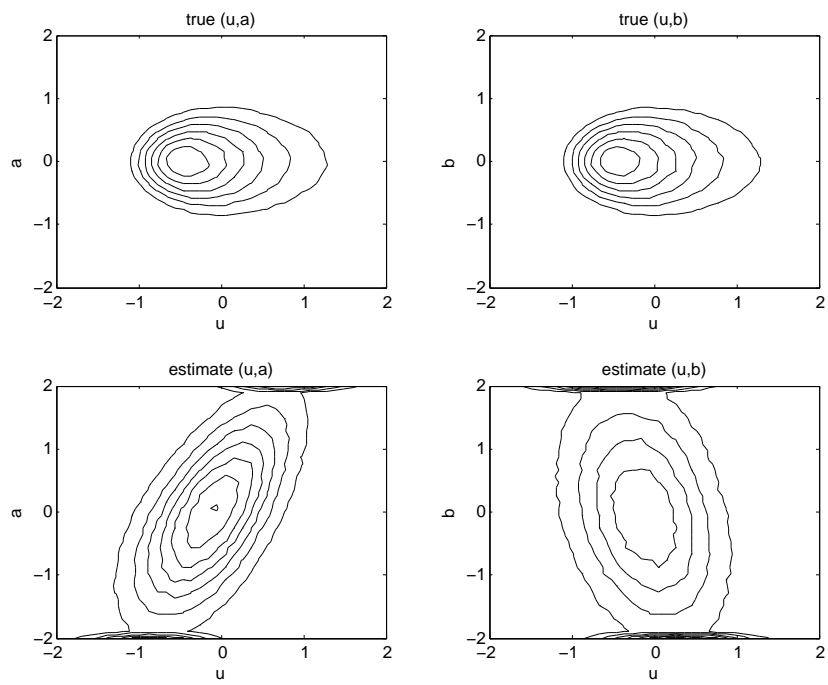


Figure 21: Contour plot of the joint pdf of the zero memory level model in (13), (17) in Regime 3.

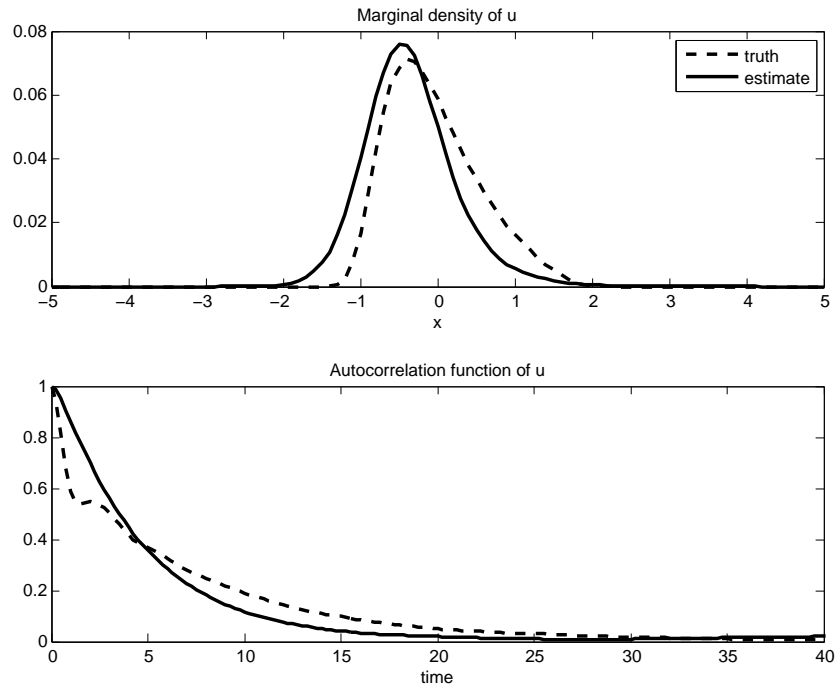


Figure 22: Marginal density (or histogram) and autocorrelation function of the solutions for the one memory level model (13), (18) in Regime 2.

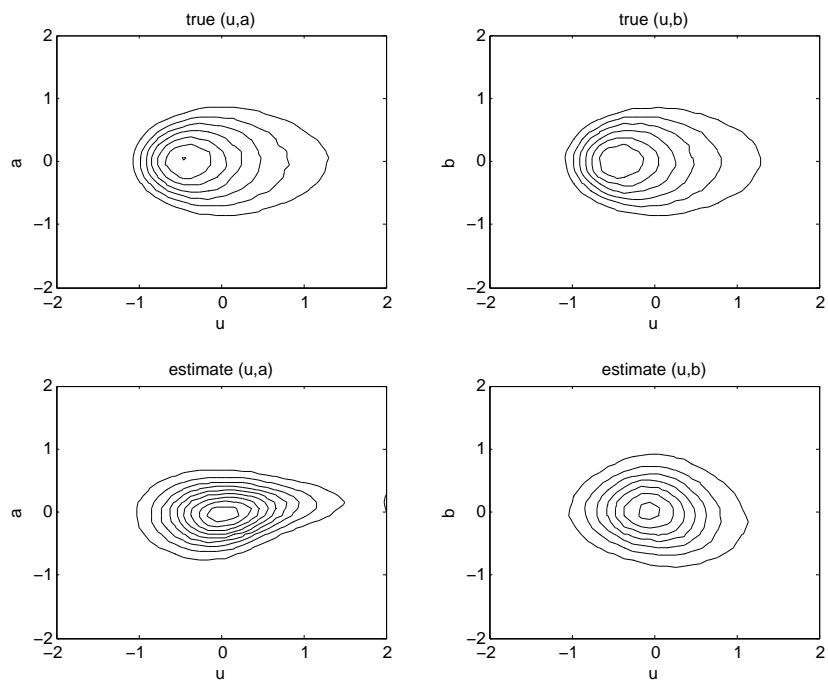


Figure 23: Contour plot of the joint pdf of the one memory level model in (13), (18) in Regime 2.

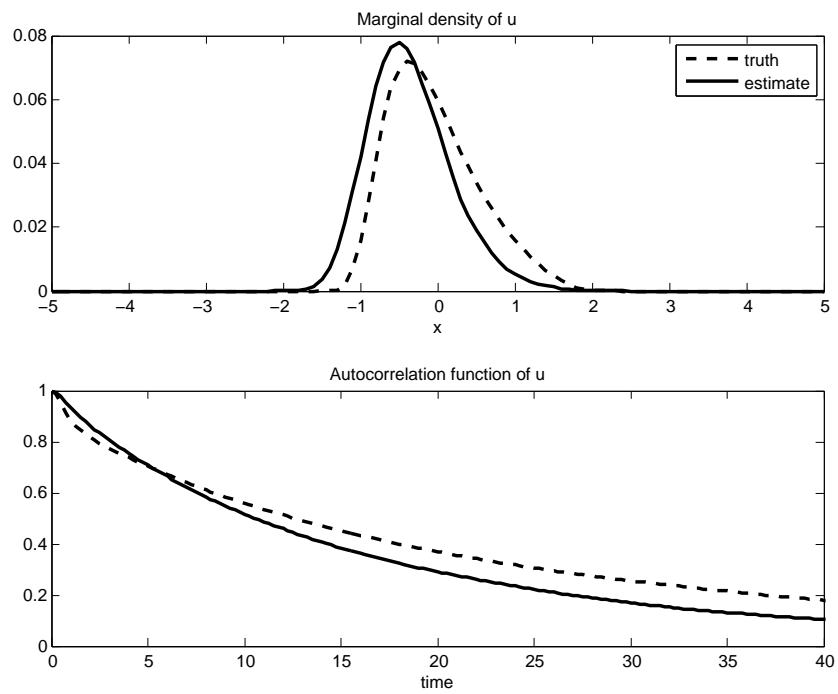


Figure 24: Marginal density (or histogram) and autocorrelation function of the solutions for the one memory level model (13), (19) in Regime 1.

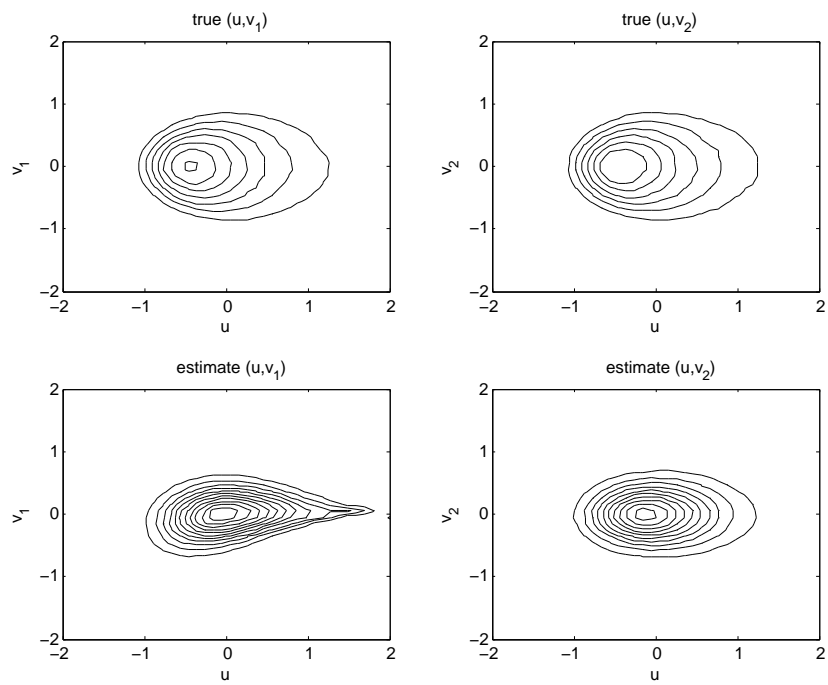


Figure 25: Contour plot of the joint pdf of the one memory level model in (13), (19) in Regime 1.

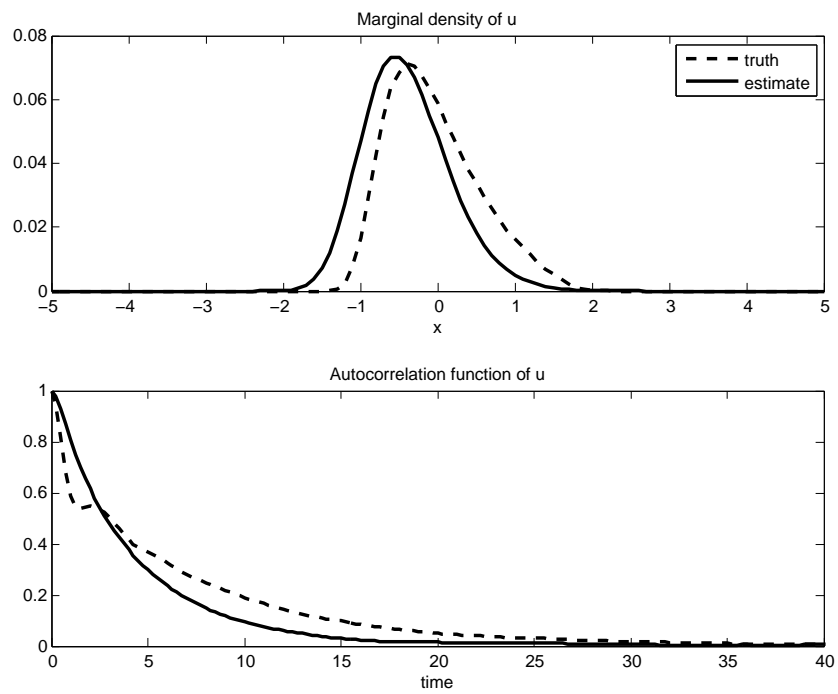


Figure 26: Marginal density (or histogram) and autocorrelation function of the solutions for the one memory level model (13), (19) in Regime 2.

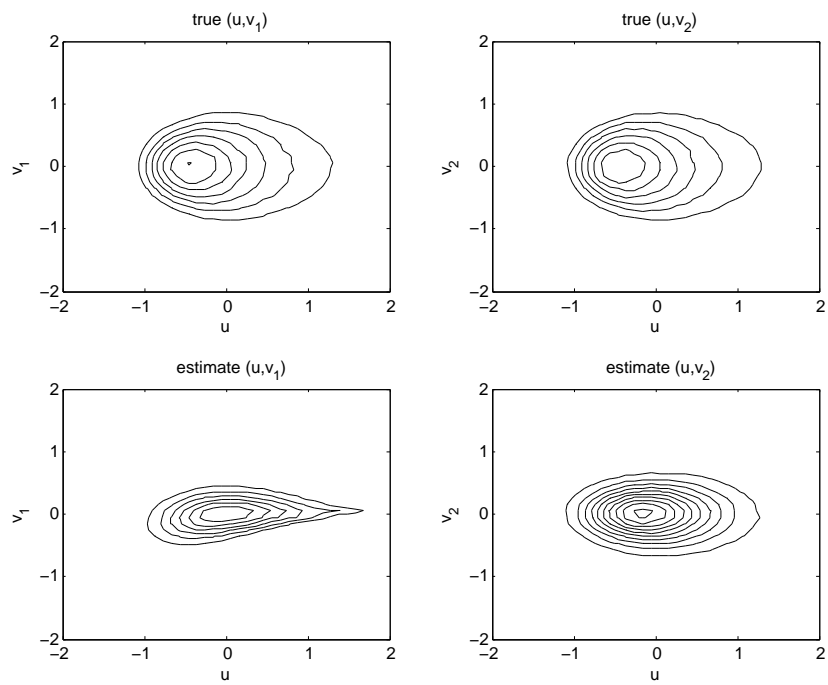


Figure 27: Contour plot of the joint pdf of the one memory level model in (13), (19) in Regime 2.



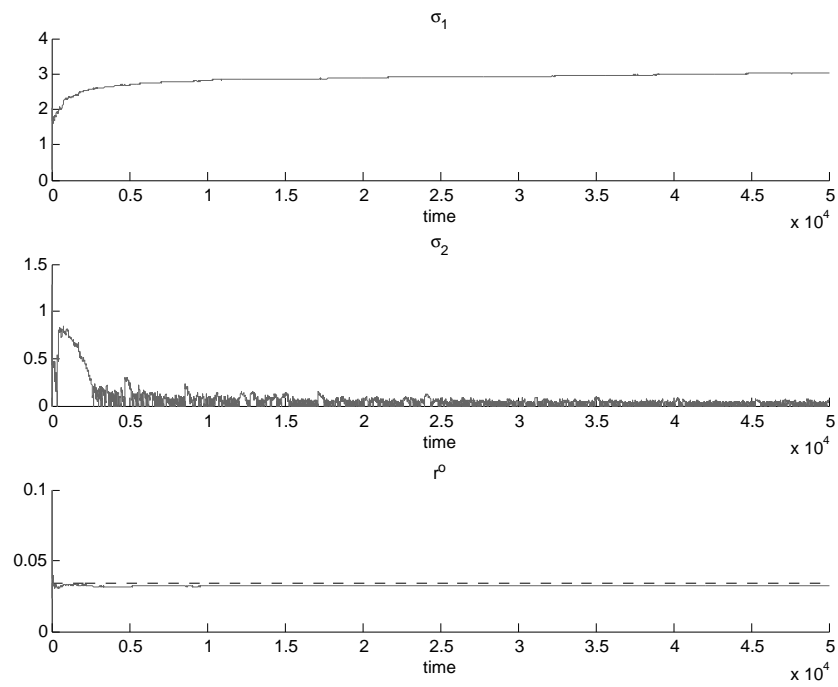


Figure 28: Stochastic parameter estimates for the one memory level model (13), (19) in Regime 1.