Parametric reduced models for the nonlinear Schrödinger equation

John Harlim and Xiantao Li

Parametric Reduced Models for the Nonlinear Schrödinger Equation

John Harlim[1,2,*] and Xiantao Li[1,†]

[1]*Department of Mathematics, the Pennsylvania State University,*

University Park, PA 16802-6400, USA.

[2]*Department of Meteorology, the Pennsylvania State University,*

University Park, PA 16802-5013, USA.

(Dated: May 7, 2015)

Abstract

Reduced models for the (defocusing) nonlinear Schrödinger equation are developed. In particular, we develop reduced models that only involve the low-frequency modes given noisy observations of these modes. The ansatz of the reduced parametric models are obtained by employing a rational approximation and a colored noise approximation, respectively, on the memory terms and the random noise of a generalized Langevin equation that is derived from the standard Mori-Zwanzig formalism. The parameters in the resulting reduced models are inferred from noisy observations with a recently developed ensemble Kalman filter-based parameterization method. The forecasting skill across different temperature regimes are verified by comparing the moments up to order four, a two-time correlation function statistics, and marginal densities of the coarse-grained variables.

* jharlim@psu.edu

† xli@math.psu.edu

I.   INTRODUCTION

An important scientific problem in applied sciences is to forecast some quantity of interest of dynamical systems that exhibit multiscale behavior. Traditional approaches (see e.g., review paper [1]) often assume some knowledge about the underlying dynamics and proceed by deriving an effective equation for a set of preselected variables. Instead of working with the trajectories associated with the full solutions, one is interested in a reduced model in which only the quantities of interest are involved. These quantities of interest are generally referred to as the *coarse-grained* variables. In the case when the dynamics of the coarse-grained variables is of primary interest, the effective model provides an efficient means to simulate directly the coarse-grained variables, without having to keep track of the remaining degrees of freedom.

An elegant framework for deriving the effective model is the Mori-Zwanzig projection [2–4], which has recently become an extremely important tool to simplify complex dynamical systems [5–15]. In particular, this derivation led to a set of generalized Langevin equation (GLEs), a typical result of the Mori-Zwanzig procedure. A notable feature of the GLE is a memory term which represents the history-dependence of the effective dynamics, along with a random noise term, which incorporates the influence of the remaining degrees of freedom. Unfortunately, solving the resulting GLE still remains as a challenge. For example, the memory function has been expressed as an infinite series [16], and it may exhibit very slow decay. The implication is that a long history of the solution has to be kept in order to evaluate the integral in the GLE. The evaluation of the integral has to be done at every time step, which adds great complexity to the entire computation. Furthermore, incorporating the random noise term is not straightforward.

A simple approach proposed by [7, 17, 18] is to approximate the memory kernel in the GLE model with a delta function (but with a carefully chosen damping parameter). This certainly introduces additional modeling error that is difficult to quantify. In problems that arise from biological systems, the memory function in the GLE can be computed by matching the auto-correlation function of the coarse-grained variables. For instance, for the GLEs derived from Newton's equations of motion in classical mechanics, one can derive an integral equation for the memory function [10, 19, 20]. But this approach requires the computation of the velocity correlation function for the *full model*, which clearly is a challenge. Furthermore,

the solution procedure for the integral equation is often not reliable. Another approach to approximate the GLE is by using an extended Markovian system, which can be done using a projection to the Krylov subspace approximation [14, 21]. This approach, however, requires the knowledge of the full model, especially the interaction among all the degrees of freedom. But this approach suggests that the full GLE models with strong memory effects can be approximated by an extended system with a few auxiliary variables and this key result motivates the present work.

The main idea of the present work is to apply a rational approximation to the kernel function in the GLE and a colored noise approximation to the orthogonal dynamics in the GLE such that the resulting parametric model is Markovian. We subsequently use the stability conditions established in [22] as guidelines to ensure non blow-up solutions in the resulting models. Rather than deriving the explicit dependence of the parameters in the resulting Markovian models in terms of the true solutions (and/or the parameters in the original dynamics), we estimate these parameters by solving an inverse problem, filtering partially observed noisy measurement of the dynamics. This approach is often useful when (a) there is a large amount of training data, e.g., from experimental observation of part of the system; (b) the explicit form of the GLEs is difficult to obtain; (c) we don't have access to the exact solutions of the full dynamical systems. Computationally, since the resulting model is Markovian, we don't need to explicitly compute the memory terms and we don't need to store the solution history. More importantly, compared to direct numerical approximation of the integro-differential equations associated with the GLE model, solving the reduced parametric system requires much less computational cost.

We will demonstrate our modeling approach on the nonlinear Schrödinger equation (NLS), which finds many applications in various areas of applied physics. Of our particular interest is the statistical-mechanics aspects, which has been well studied theoretically [23–25]. Our goal is to predict the equilibrium statistical behavior of the low-order wave numbers. We should stress out that developing reduced models for the NLS equations is highly nontrivial in the following sense. Since the solutions of NLS equation exhibit strong correlation time with nontrivial autocorrelation function, the memory feedback from the unresolved scales is non-negligible and need to be appropriately accounted. Moreover, the equilibrium statistics of the solutions are highly non-Gaussian with bimodal distribution. We will proceed by applying a rational approximation and a colored noise approximation, subsequently, to the

kernel functions and orthogonal dynamics of a GLE, derived by Chorin and coworkers [16]. Subsequently, we estimate the parameters of the resulting model with an adaptive parameter estimation scheme that is recently developed in [26]. We will then validate the forecasting skill by comparing moments up to order four, a two-time correlation function statistics, and marginal densities of the coarse-grained variables.

The remaining part of the paper is organized as follows. In Section II, we state the problem and provide a short review of the GLE deduced by Chorin and coworkers [16]. In Section III, we construct the parametric models. The procedure for the parametric estimation method in [26] is formally described in Section IV. To be self-contained, we include a pseudo-algorithm in the appendix. In Sections V-VI, numerical results are then presented to demonstrate the effectiveness of the reduced models. We close this paper with a short summary and discussion in Section VII.

II. PROBLEM STATEMENT AND BACKGROUND

We consider the nonlinear Schrödinger equation (NLS),

$$iu_t = -u_{xx} + |u|^2 u, \tag{2.1}$$

in one space dimension. For simplicity, we apply a periodic boundary conditions on a non-dimensionalized domain $x \in [0, 2\pi]$. Here, the solutions of (2.1) can be described by the Fourier series,

$$u(x,t) = \sum_{k \in \mathbb{Z}} u_k(t) e^{ikx}. \tag{2.2}$$

This turns the PDE into a set of ODEs for the Fourier modes,

$$\frac{d}{dt} u_k = -i\omega_k u_k - i \sum_{k_1 \in \mathbb{Z}} \sum_{k_2 \in \mathbb{Z}} u_{k_1} u_{k_2} u_{k_1+k_2-k}^*, \tag{2.3}$$

with dispersion relation given by, $\omega_k = k^2$.

Of particular importance to the statistical mechanics interpretation of (2.1) is the Hamiltonian structure of the system, with the Hamiltonian given by,

$$E = E_0 + E_1,$$

where,

$$\begin{cases} E_0 = \sum_{k \in \mathbb{Z}} \omega_k |u_k|^2, \\ E_1 = \dfrac{1}{2} \sum_{k_1 \in \mathbb{Z}} \sum_{k_2 \in \mathbb{Z}} \sum_{k_3 \in \mathbb{Z}} u_{k_1} u_{k_2} u_{k_3}^* u_{k_1 + k_2 - k_3}^*. \end{cases}$$

With this Hamiltonian, we can rewrite (2.3) as follows,

$$i \frac{d}{dt} u_k = \frac{\partial E}{\partial u_k^*}. \tag{2.4}$$

Numerically, we can simulate the solutions of (2.4) with pseudo-spectral methods, e.g. [27], of (2.4) for finite wave numbers, $|k| \leq K$. The initial condition can be prepared using a Monte-Carlo algorithm. We assume that the resulting solutions are the underlying dynamics.

In this paper, we are interested to construct a low-dimensional parametric model to predict low-frequency modes of (2.4), given noisy observations of the corresponding modes at discrete-times. Namely,

$$v_{k,j} = u_k(t_j) + \varepsilon_j^o, \quad |k| \leq m, \tag{2.5}$$

where m denotes the upper bound of the observed/resolved modes that are much smaller than the dimensionality of the underlying dynamics K, $m \ll K$. In (2.5), the noises ε_j^o are i.i.d. Gaussian with mean zero and unknown error covariance, R. To achieve this goal, our strategy is to exhaust our physical knowledge of the model to deduce an appropriate ansatz for the parametric model and then apply a recently developed, adaptive ensemble Kalman filter based, parameter estimation method [26] to specify the parameters in the corresponding model as well as the observation noise covariance, R.

Before we discuss our main strategy, we briefly review a classical dimensional-reduction Mori-Zwanzig formalism [2–4], which underpins the choice of ansatz for our parametric models in the remaining of this section.

A.   Reduced Models from the Mori-Zwanzig formalism

A general framework for reducing the dimension associated with a complex dynamical system is the Mori-Zwanzig projection formalism [2–4], which was originally developed to deal with non-equilibrium processes in statistical mechanics. This approach relies on a projection operator, denoted by $\mathcal{P}$, which separates out the quantities of interest and identifies

terms of different nature. In particular, for a system of initial value problem in the form,

$$\dot{x} = f(x), \quad x(0) = z, \tag{2.6}$$

and an arbitrary reduced quantity, φ, which is a function of $x(t)$, the Mori-Zwanzig procedure yields an exact equation for φ [2, 3],

$$\frac{d}{dt}\varphi(t) = e^{t\mathcal{L}}\mathcal{P}\mathcal{L}\varphi(0) + \int_0^t e^{(t-s)\mathcal{L}}K(s)ds + \xi(t), \tag{2.7}$$

where the first term in (2.7) usually represents the reversible part of the dynamics and it represents the "Markovian" term. Here, the differential operator $\mathcal{L}$ corresponds to the generator of the dynamical system in (2.6) and it is defined with respect to initial condition z as follows,

$$\mathcal{L} = \sum_i f_i(z)\frac{\partial}{\partial z_i}, \tag{2.8}$$

and we use semigroup notation $e^{t\mathcal{L}}$ to denote the evolution operator that maps the solutions forward in time as follows, $\varphi(t) = e^{t\mathcal{L}}\varphi(0)$. The second term depends on φ at all times between 0 and t so it incorporates the *memory effect* as a result of coarse-graining, and it dictates a strong coupling with the remaining degrees of freedom through a memory kernel,

$$K(t) = \mathcal{P}\mathcal{L}\xi(t), \tag{2.9}$$

where

$$\xi(t) = e^{t\mathcal{Q}\mathcal{L}}\mathcal{Q}\mathcal{L}\varphi(0), \quad \mathcal{Q} = \mathcal{I} - \mathcal{P}. \tag{2.10}$$

The term in (2.10) is referred to as the *orthogonal dynamics* and if the initial condition z is random, then $\xi(t)$ is a stochastic forcing. Equation (2.7) is often called a *generalized Langevin equation* (GLE). The most appealing aspect of the GLE in (2.7) is that it is *exact*. However, solving the GLE in (2.7) directly is not much simpler than solving the full system in (2.6) since one has to estimate the orthogonal dynamics in (2.10) and the memory kernel function in (2.9).

We should point out that the GLE for the ODE in (2.6) is nonunique since there are different choices for the projection operator $\mathcal{P}$. For example, in the work of Mori [2], an orthogonal projection is employed, which is often appropriate when the problem can be formulated in a Hilbert space. For the NLS equation in (2.4), Chorin et al [16] used a projection operator that is the conditional expectation with respect to the canonical ensemble $\rho \propto e^{-\beta E}$ to deduce an effective equation for selected (low-frequency) Fourier modes,

$|k| \leq m$. This choice is motivated by the statistical mechanics aspect of the NLS [24]. Since the calculation is usually quite cumbersome, an expansion around the Gaussian distribution $\rho_0 \propto e^{-\beta E_0}$ was introduced, which is appropriate for systems at low temperature, $\beta \gg 1$. To see this, one can introduce a change of variables in the Gibbs distribution, $v = \sqrt{\beta}u$. As a result, the distribution can be written as $\rho \propto e^{-E_0(v)-E_1(v)/\beta}$. At low temperature when $\beta \gg 1$, the distribution is approximately Gaussian. Furthermore, higher order terms in the equation are much less important, since statistically, u is of the order $1/\sqrt{\beta}$.

In the simplest case when $m = 0$, only the zeroth Fourier mode is retained and the effective equation takes the form of [16],

$$\dot{u}_0 = -icu_0 - i|u_0|^2 u_0 + \int_0^t \kappa_0(t-\tau)u_0(\tau)d\tau + i\int_0^t \phi_0(t-\tau)|u_0(\tau)|^2 u_0(\tau)d\tau + \xi(t), \quad (2.11)$$

where c is a positive constant, and κ_0 and ϕ_0 are complex valued kernel functions with complicated expressions [16] (they are written as infinite series). Furthermore, in solving (2.11), the history of the solution has to be stored and the integral has to be approximated by appropriate quadrature formulas *at each step* of the time integration. All these operations add up to significant computational cost and it is also unclear how the approximation by ρ_0 affects the modeling error.

Rather than computing these kernels directly, we take a different approach here. In particular, we will model the memory terms in (2.11) and the stochastic process ξ with an appropriate ansatz of parametric equations. Subsequently, we estimate the corresponding parameters from noisy observations (2.5) such that the resulting Markovian model gives accurate equilibrium statistical estimates for the selected Fourier modes that are retained: $|k| \leq m$.

III. CONSTRUCTING PARAMETRIC MODELS

Here we discuss our approach in approximating the GLEs using parametric models that involve explicitly few parameters. These approximations are constructed in such a way that the approximate model can be re-written into a memory-less form, leading to a Markovian dynamics to facilitate the numerical implementation. To clarify the exposition, we first discuss the case where we only retain the zeroth mode. We assume the form of the GLE (2.11), but we approximate the memory terms using rational functions so that parameters

can be introduced. We then provide the resulting parametric form for the more general case which retain more Fourier modes, $0 < |k| \leq m$.

A. A reduced model for u_0 with scalar parametric approximation

Now we will construct an ansatz for approximating the first memory term in (2.11). To this end, we introduce a parameter $b \in \mathbb{C}$ and an auxiliary function f to denote the first memory term,

$$bf := \int_0^t \kappa_0(t-\tau) u_0(\tau) d\tau, \tag{3.12}$$

and our plan is to find a set of differential equations for solving f. First, taking the Laplace transform on (3.12), we arrive at,

$$b\widetilde{f}(s) = \tilde{\kappa}_0(s)\widetilde{u}_0(s), \tag{3.13}$$

where we denote $\widetilde{h}$ to be the Laplace transform (defined on frequency domain s) of any function h that is locally integrable on $\mathbb{R}^+$. The key idea is to approximate the kernel function, $\widetilde{\kappa}_0$, using a rational function,

$$\widetilde{\kappa}_0(s) \approx \frac{-|b|^2}{s-a}, \tag{3.14}$$

where $a \in \mathbb{C}$ is the second parameter to be determined. This particular form of the rational function is chosen to ensure the stability of the resulting parametric model, as we will explain below. In principle, these two coefficients, a and b, can be determined with Padé approximations (or more general rational approximations) of the exact kernel. In model reduction problems, this is known as the *moment matching* procedure [28, 29], where for linear dynamical systems, these parameters can be explicitly connected to properties of the original problem. The main departure of the current approach from those existing methods is mainly that we leave them as parameters and later infer them with a filtering procedure, learning from partially observed noisy time series.

Converting (3.13) and (3.14) back to the time domain, we find that f satisfies a differential equation,

$$\dot{f} = af - b^* u_0(t), \quad f(0) = 0. \tag{3.15}$$

For low temperature case, one can neglect the second memory term in (2.11) that involves ψ_0 since this higher-order term is negligible as explained before in Section II A. With this

perspective, we propose the following parametric model,

$$\begin{cases} \dot{u}_0 = -icu_0 - id|u_0|^2 u_0 + bf, \\ \dot{f} = af - b^* u_0 + \sigma_1 \dot{W}_f, \end{cases} \tag{3.16}$$

where we have introduced two additional non-negative parameters c and d. In principle, these parameters can be determined from the Mori-Zwanzig reduction procedure, which might involve lengthy calculations. However, since the derivation in [16] employed further approximations using the (conditional) Gaussian distribution, the resulting values for c and d may not be optimal. Therefore, we kept the form of the equations suggested by the Mori-Zwanzig formalism, but leave c and d as additional parameters, which we will determine using a filtering procedure.

We have also introduced a white noise $\dot{W}_f$. When the second equation is analytically solved and subsequently substituted into the first equation, this white noise will become a colored-noise approximation to the random process $\xi(t)$. When both a and b are real-valued parameters, the second equation represents an Ornstein-Uhlenbeck process [30]. But here f is a more general Gaussian process. We should also point out that the reduced system of parametric equations in (3.16) is a special case of the physics constrained nonlinear regression model described in [22] in the following sense. In compact form, we can write (3.16) as a system of four-dimensional real valued SDEs,

$$dx = [Ax + N(x)]dt + \Sigma dW, \tag{3.17}$$

where we define $x = (\mathrm{Re}\{u_0\}, \mathrm{Im}\{u_0\}, \mathrm{Re}\{f\}, \mathrm{Im}\{f\})^\top$, and W is a standard two-dimensional Wiener process. In addition, we define $a = a_1 + ia_2$ and $b = b_1 + ib_2$ such that,

$$A = \begin{pmatrix} 0 & c & b_1 & -b_2 \\ -c & 0 & b_2 & b_1 \\ -b_1 & -b_2 & a_1 & -a_2 \\ b_2 & -b_1 & a_2 & a_1 \end{pmatrix}, \quad N(x) = d \begin{pmatrix} |u_0|^2 \mathrm{Im} u_0 \\ -|u_0|^2 \mathrm{Re} u_0 \\ 0 \\ 0 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ \frac{\sigma_1}{\sqrt{2}} & 0 \\ 0 & \frac{\sigma_1}{\sqrt{2}} \end{pmatrix}. \tag{3.18}$$

One of the main results in [22] states that if the Fokker-Planck operator of the SDE in (3.17) is hypoelliptic, and suppose also that all eigenvalues of A have negative real part and there exists an appropriate norm under which the inner product $\langle N(x), x \rangle = 0$, then solutions of (3.17) is geometrically ergodic. For our parametric model above, the stability condition

is met when $a_1 \leq 0$ and the dissipation of the energy of the nonlinear terms is satisfied under an inner product with respect to $L = E + \frac{1}{2}|f|^2$.

We should note that one can repeat the same calculation for approximating the second memory term in (2.11) but the resulting nonlinear terms will not conserve energy and can be unstable (see Appendix A). Based on this consideration, we ignore approximating the second memory terms in this paper. Instead, we will only consider the parametric model in (3.16) which guarantees non-blowup solutions.

B. A reduced model for u_0 with multi-dimensional parametric approximations

A simple extension of the two-parameter scalar parametric approximation model in (3.16) is to allow b and f to be vectors, here denoted by $\boldsymbol{b}$ and $\boldsymbol{f}$, respectively, to emphasize the multi-dimensional representation. This leads to an extended model,

$$\begin{cases} \dot{u}_0 = -ic\frac{1}{2}u_0 - id|u_0|^2 u_0 + \boldsymbol{b} \cdot \boldsymbol{f}, \\ \dot{\boldsymbol{f}} = A\boldsymbol{f} - \boldsymbol{b}^* u_0 + \Sigma \dot{W}. \end{cases} \tag{3.19}$$

For instance, the matrices A and Σ can be chosen in the following form,

$$A = \begin{bmatrix} a_1 & a_2 \\ -a_2 & a_1 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix}. \tag{3.20}$$

The corresponding extended model has a parameter space of dimension 10. We will see that this model will give improved estimates compared to (3.16) in higher temperature case. Similar extensions can be found by increasing the dimension of A, $\boldsymbol{b}$ and $\boldsymbol{f}$.

C. Models with more retained Fourier modes

In general, we can keep those modes k with $|k| \leq m$, and m indicates the range of the modes to be kept. In this case, we first define the coarse-grained energy,

$$E = \sum_{|k| \leq m} c_k |u_k|^2 + \frac{1}{2} \sum_{|k_1| \leq m} d_k \sum_{|k_2| \leq m |k_3|, |k_1+k_2-k_3| \leq m} u_{k_1} u_{k_2} u_{k_3}^* u_{k_1+k_2-k_3}^*. \tag{3.21}$$

The model *without* memory can be written as follows,

$$\dot{u}_k = -i\frac{\partial E}{\partial u_k^*}, \quad -m \leq k \leq m. \tag{3.22}$$

10

Motivated by (3.16) and the Mori-Zwanzig procedure [5], let us consider a parametric model as follows,

$$
\begin{cases}
\dot{u}_k = -i\dfrac{\partial E}{\partial u_k^*} + b_k f_k, \\[2mm]
\dot{f}_k = -b_k^*\dfrac{\partial E}{\partial u_k^*} + a_k f_k + \sigma_k \dot{W}_k, \quad -m \le k \le m.
\end{cases}
\tag{3.23}
$$

The auxiliary functions are assumed to be zero initially, i.e., $f_k(0) = 0$, since they are introduced to approximate the memory terms.

To see the energy dissipation mechanism, we can define the Lyapunov functional,

$$
V = E + \sum_k |f_k|^2.
\tag{3.24}
$$

Direct calculations yield,

$$
\frac{d}{dt}V = \sum_k \mathrm{Re}(a_k)|f_k|^2.
\tag{3.25}
$$

Consequently, we require that $\mathrm{Re}(a_k) \le 0$ to guarantee non-blow up solutions.

IV. THE PARAMETER ESTIMATION PROCEDURE

In this section, we describe formally how to estimate the parameters of the reduced models (e.g., (3.16), (3.19), or (3.23)), given noisy observations $v_j = (v_{-m,j}, \ldots, v_{m,j})^\top \in \mathbb{C}^{2m+1}$ of $u_j = (u_{-m,j}, \ldots, u_{m,j})^\top \in \mathbb{C}^{2m+1}$, where $u_{k,j} = u_k(t_j)$ are solutions of the full system in (2.3) for $|k| \le K$ and $K \gg m$ at discrete time step t_j:

$$
v_j = u_j + \epsilon_j, \quad \epsilon_j \sim \mathcal{N}(0, R), \quad |k| \le m,
\tag{4.26}
$$

with an unknown observation error covariance R, where R is an $(2m+1) \times (2m+1)$ diagonal matrix with k-th diagonal component, r_k. To simplify the discussion, let us classify the parameters in our reduced model to two types. We refer to the parameters in the deterministic terms in the reduced models as the "deterministic parameters", θ_d, and the amplitude of the stochastic forcings as the "stochastic parameters", θ_s. For example, in (3.16), the deterministic and stochastic parameters are, $\theta_d = \{a, b, c, d\}$, and $\theta_s = \{\sigma_1^2, R\}$, respectively. We split the parameters into two types because the algorithm that we use to estimate θ_d is simply a standard augmentation method while the algorithm to estimate θ_s is the adaptive method for estimating covariances which preserves the positivity of σ_1^2 and R. Let us also define vector $x_j = (u_{-m,j}, \ldots, u_{m,j}, f_{-m,j}, \ldots, f_{m,j})^\top \in \mathbb{C}^{4m+2}$ to simplify the notation below.

11

The main idea of the parameterization method is to apply Bayes' theorem to obtain a posterior distribution of the augmented state and parameters at each time step t_j when observations become available,

$$p(x_j, \theta_d, \theta_s | v_j) \propto p(x_j, \theta_d, \theta_s) p(v_j | x_j, \theta_d, \theta_s), \tag{4.27}$$

where $p(x_j, \theta_d, \theta_s)$ denotes the prior distribution of the augmented state and parameters at time t_j and $p(v_j | x_j, \theta_d, \theta_s)$ denotes the likelihood function of the augmented state and parameters, corresponding to the observation model in (4.26), that is, $p(v_j | x_j, \theta_d, \theta_s) = \mathcal{N}(u_j, R)$. The parameterization method can be formally described as follows: Since $p(x_j, \theta_d, \theta_s) = p(\theta_s) p(x_j, \theta_d | \theta_s)$ by definition of the conditional distribution, we can rewrite (4.27) as follows:

$$p(x_j, \theta_d, \theta_s | v_j) \propto p(\theta_s) p(x_j, \theta_d | \theta_s) p(v_j | x_j, \theta_d, \theta_s), \tag{4.28}$$

$$\propto p(\theta_s) p(x_j, \theta_d | v_j, \theta_s), \tag{4.29}$$

where we use another Bayes' theorem, $p(x_j, \theta_d | v_j, \theta_s) \propto p(x_j, \theta_d | \theta_s) p(v_j | x_j, \theta_d, \theta_s)$, to obtain (4.29). Here, the first step in the filtering algorithm is to estimate $p(x_j, \theta_d | \theta_s, y_j)$ by applying Bayes' theorem to the last two components of (4.28). Subsequently, we implement the Bayes' theorem one more time in (4.29) to obtain the posterior distribution of the augmented $(x_j, \theta_d, \theta_s)$.

To avoid unobservability of the stochastic parameters due to sparse observations with dimension less than the number of stochastic parameters, θ_s, in our implementation, we include information from past observations up to lag $L > 1$. At each time step t_j, instead of solving (4.28)-(4.29), we formally solve

$$p(x_j, \theta_d, \theta_s | v_j, \ldots, v_{j-L+1}) \propto p(x_j, \theta_d, \theta_s | v_{j-1}, \ldots, v_{j-L+1}) p(v_j | x_j, \theta_d, \theta_s)$$

$$\propto p(\theta_s) p(x_j, \theta_d | \theta_s, v_{j-1}, \ldots, v_{j-L+1}) p(v_j | x_j, \theta_d, \theta_s) \tag{4.30}$$

$$\propto p(\theta_s) p(x_j, \theta_d | \theta_s, v_j, \ldots, v_{j-L+1}), \tag{4.31}$$

where, similar as before, the first step is to estimate $p(x_j, \theta_d | \theta_s, v_j, \ldots, v_{j-L+1})$ by applying Bayes' theorem to the last two components of (4.30). Subsequently, we implement the Bayes' theorem one more time in (4.31) to obtain the posterior distribution of the augmented $(x_j, \theta_d, \theta_s)$.

In our numerical implementation, we use the method in [26] which uses Gaussian approximation to solve these inverse problems. At time $j \geq L$, we assume that we have prior ensemble estimates of $\{x_j, \theta_d, \theta_s\}$ at times $j, j-1, \ldots, j-L+1$ and the associated observations at these times. We assume that the deterministic parameter is persistence, that is, $\dot{\theta}_d = 0$. The first step is to apply ETKF method [31] to obtain posterior ensemble estimates of the augmented variable $\{x_j, \theta_d\}$, incorporating observations in (4.26), which is a Gaussian approximation of $p(x_j, \theta_d | \theta_s, v_j, \ldots, y_{j-L+1})$. To start the algorithm, one can just repeat this ETKF algorithm L-times to obtain the prior ensemble estimates at time $j = 1, \ldots, L$ with fixed parameters $\{\theta_s, \theta_d\}$. Now, at $j \geq L$, we start the secondary filter to update θ_s. The key idea of the secondary filter is to view the posterior density function $p(x_j, \theta_d | \theta_s, v_j, \ldots, v_{j-L+1})$ as a likelihood function of θ_s. Notice that while this posterior density is Gaussian with respect to variables (x_j, θ_d), its dependence on θ_s can be described non-uniquely (for example, see [26, 32–36]). Here, we will adopt the estimation method of [26] that is based on Belanger's formulation [34] with a likelihood function corresponding to the following pseudo-observation model,

$$\sigma_{j,\ell} = \mathcal{F}_{j,\ell} \theta_s + \eta_{j,\ell}, \quad \eta_{j,\ell} \sim \mathcal{N}(0, W_{j,\ell}), \quad \ell = j, \ldots, j-L+1 \tag{4.32}$$

Here, components of $\sigma_{j,\ell} = \{e_j e_{j-\ell}^\top\}$ are the product of the forecast error estimates in the observation space (which are also known as innovations),

$$e_j = v_j - \bar{u}_j^-, \tag{4.33}$$

where $\bar{u}_j^-$ denotes the mean prior estimate that is empirically estimated with an ensemble average. In (4.32), the observation operator $\mathcal{F}_{j,\ell}$ and the noise covariance matrix $W_{j,\ell}$ are functions of $\bar{x}_{j-\ell}^-$ and θ_d and they will be constructed recursively. We should also note that in our implementation, $W_{j,\ell}$ is approximated under a Gaussian assumption (see Appendix B below for detail). With the pseudo-observation model in (4.32), a secondary Kalman filter is implemented L-times to sequentially update the posterior mean and covariance estimate of θ_s, accounting for pseudo-observations $\{\sigma_{j,\ell}\}_{\ell=1\ldots,L}$ one at a time. To be self-contained, we provide a pseudo algorithm of this method in Appendix B below. We should note that there are other methods to approximate the secondary Bayes' update in (4.31) that use different observation model in (4.32) and do not use Kalman update (see e.g., [35, 36]).

V. NUMERICAL RESULTS ON PARAMETER ESTIMATION OF MODELS WITH A SINGLE FOURIER MODE u_0

In this section, we present the results from three numerical tests, where the parametric models for u_0 (equations (3.16) and (3.19)) are estimated and further assessed. We assume that the observation time interval $T_{obs} = 0.02$. The time series, consisting of 100,000 observations, is generated by using the Strang's splitting method in time, which has been implemented in [27] for the NLS equation. The data is generated from the solutions of the NLS in the Fourier domain (2.3), with $K = 32$.

We then perform a parameter estimation method with an ensemble Kalman filter based method [26], which was described in the previous section. The forecast is generated with the 4th order Runge-Kutta method with step size $\Delta t = 0.002$.

Following the estimation, we verify the forecasting skill of the reduced models with the estimated parameter set as follows. We take the estimated parameters and run the reduced models forward in time for a sufficiently long time. Then, based on the long trajectory, we compare the low-order statistics up to order-four and the time correlation function. The auto-correlation is computed based on the Wiener-Khinchin theorem. Namely, we take the Fourier transform of the data, $x(t)$, and compute the power spectrum, $|\hat{x}(\omega)|^2$. The correlation function is then given by the inverse Fourier transform of the power spectrum. This procedure is often more efficient than the direct approach, i.e.,

$$c(\tau) \approx \frac{1}{M} \sum_{m=1}^{M} x(t_m + \tau) x(t_m).$$

A. Low temperature $\beta = 10^4$.

We first consider a low temperature case with $\beta = 10^4$ ($k_B T = 0.0001$), and estimate the parameters in the model (3.16) for u_0. This model contains two (complex) deterministic parameters a and b, together with two real valued parameters c and d. We set the observation noise error with variance, $R = 0.01$. In Fig. 1, we show the estimated solutions along with the observed values during the estimation procedure. Very good agreement with the true values has been found. We also monitor the predicted values of the deterministic parameters, and the history is presented in Fig. 2. Meanwhile, the stochastic parameter σ_1 has settled to a constant value, and the variance of the observation noise has been correctly

14

predicted, as indicated in Fig. 3. Another observation is that the predicted values of the parameters c and d are $c = -0.0067$ and $d = 0.0024$, which are quite different from the values calculated from the Mori-Zwanzig's projection procedure (0.00063 and 1, respectively) corresponding to an approximate Gaussian measure, $e^{-\beta E_0}$. We should point out that if we fix these two parameters to be those from the Mori-Zwanzig projection and run the filtering procedure to estimate the remaining parameters, a, b, σ_1, R, the resulting estimates are completely innaccurate. This suggests that while the perturbation approach [16] suggests the explicit forms of the reduced model, it is more natural to adaptively estimate all the parameters which reconfirms the results in [37]. Moreover, in general, there can be non-unique parameters that provide the same equilibrium statistics (for e.g., see Proposition 1(d) in [26]).

Next we evaluate the forecasting skill and check the accuracy of the climatological statistics of the resulting solution u_0. We observe from Fig. 4 that the qualitative behavior of the path-wise solutions is well captured. The solution f, which was introduced to replace the memory term, exhibits much faster oscillations. Further, from Fig. 5, we observe that the distribution and the time correlation of the true solution are accurately predicted. We also report the accuracy of the first four moment estimates in Table I, where the errors are on the order of 10^{-2}.

B. Results for a higher temperature $\beta = 10$.

We now turn to observations obtained from a higher temperature simulation with $\beta = 10$ ($k_B T = 0.1$). In this case, the time series for u_0 exhibits faster oscillations and slightly larger amplitude. We also observe the amplitude and frequency of the oscillations are somewhat sensitive to the initial conditions since the system is not ergodic.

Based on the data, we estimate the first reduced model (3.16), and then check the statistics of the resulting model. Due to the higher temperature, the variance of u_0 is much bigger. Therefore, we set a higher value for observation noise variance, $R = 0.1$. We see from Fig. 6 that the accuracy is not as satisfactory as in the previous case. In particular, the peaks of the marginal density are not well captured and the variance is underestimated (see Table I) although the other statistics are accurately estimated. Also, the correlation function is inaccurate beyond the first oscillation.
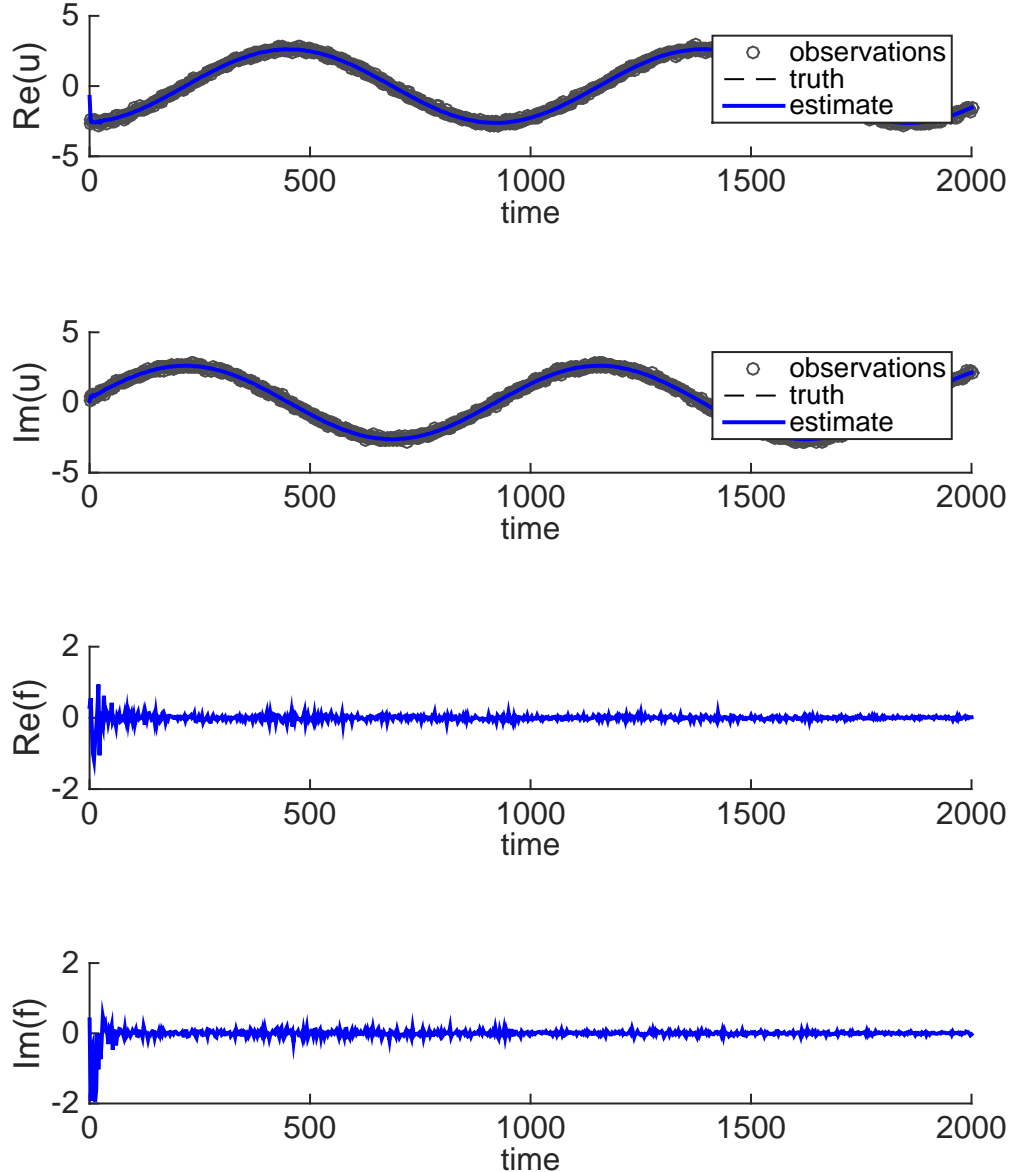
15

FIG. 1. (Color online) State estimates u and f for the parametric model (3.16).

As comparison, we consider the next parametric model, represented by the equation (3.19), which contains 4 complex deterministic parameters and two real ones. With the parameters obtained from the filtering procedure, we perform a similar statistical verification. The results, including the histogram and the time correlation functions, are illustrated in Fig. 7. It is clear that the extension has offered improved accuracy in the resulting histogram. Table I summarized the statistics (four moments) of u_0 obtained from the three tests, compared to the true values. We notice that for the higher temperature case, the model (3.19) yields much better estimates for the second moment. However, the estimated
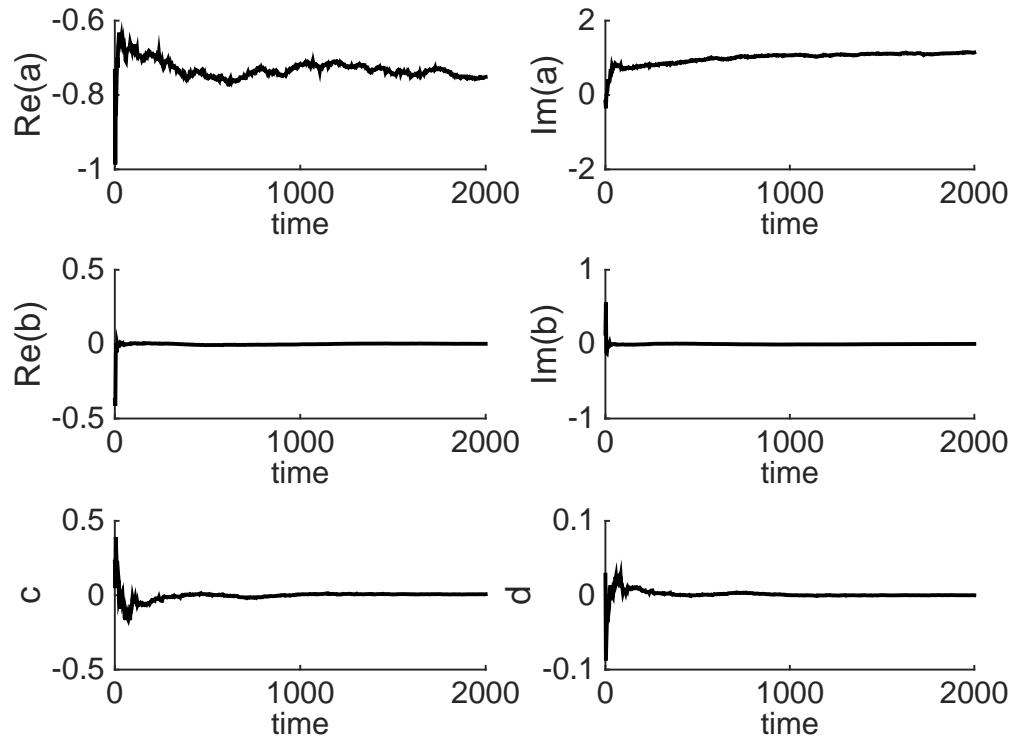
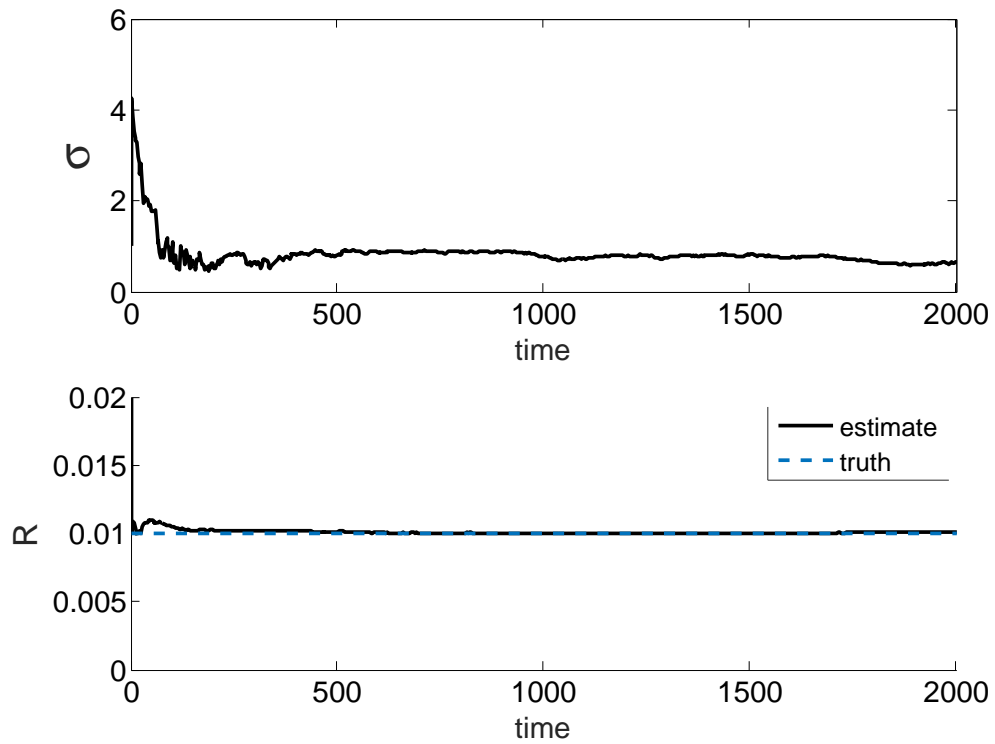FIG. 2. (Color online) Deterministic parameter estimates, a, b, c and d.

FIG. 3. (Color online) Stochastic parameter estimates, σ, R.
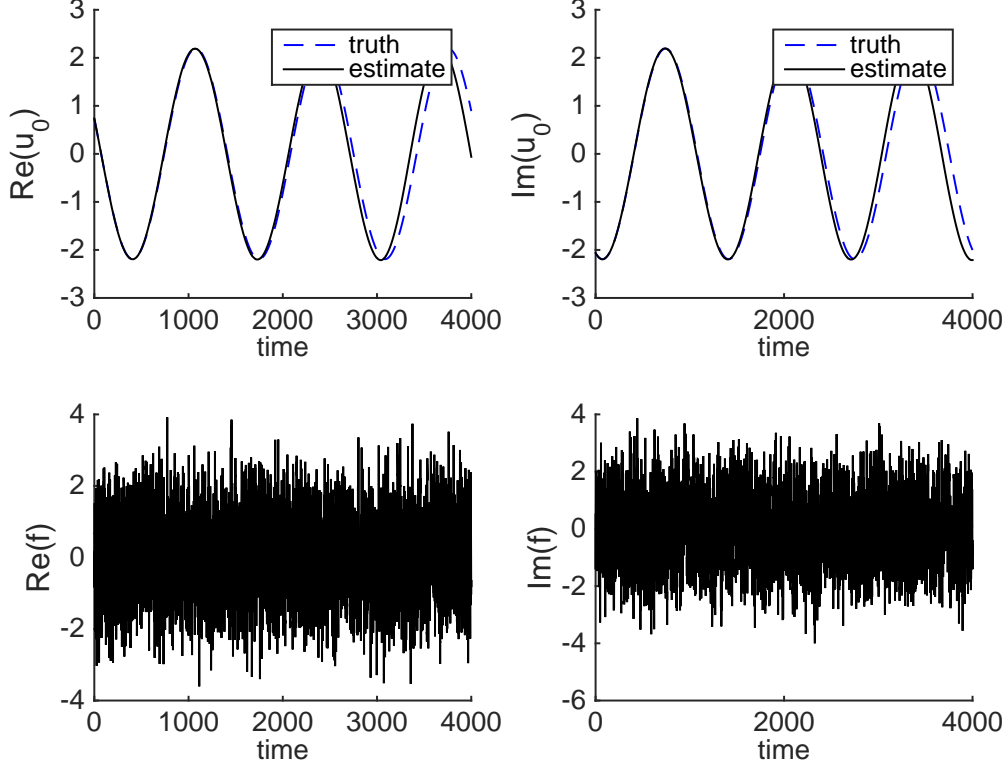
17

FIG. 4. (Color online) Solutions of the reduced model in (3.16), integrated with the estimated parameters.

correlation function is only slightly improved up to time 500 relative to the result from the model in (3.16).

TABLE I. Comparison of the equilibrium statistics of $\mathrm{Re}(u_0)$ for the three tests.

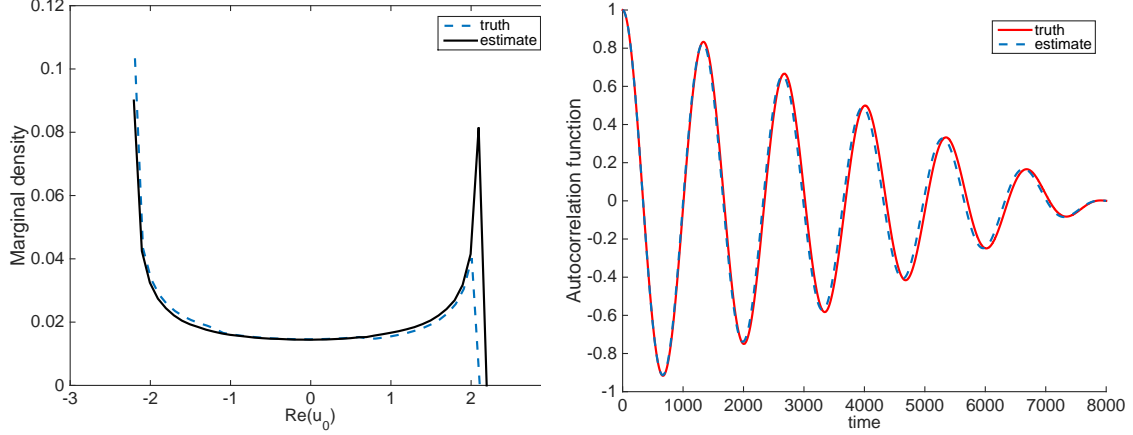|	Model (3.16), $\beta = 10^4$		Model (3.16), $\beta = 10$		Model (3.19), $\beta = 10$	
Statistics	Truth	Estimate	Truth	Estimate	Truth	Estimate
mean	-0.0037	-0.0717	0.0457	0.0368	0.0457	-0.0642
variance	2.4018	2.4570	8.0655	5.3181	8.0655	8.5148
skewness	0.0840	0.0658	-0.0251	-0.0204	-0.0251	0.0371
kurtosis	1.5071	1.5123	1.4998	1.5160	1.4998	1.5062

18

FIG. 5. (Color online) The marginal distribution (left) and the time correlation function (right) predicted by the reduced mode (3.16) for low temperature, $\beta = 10^4$. As comparison, the statistics of the true solution is also shown.
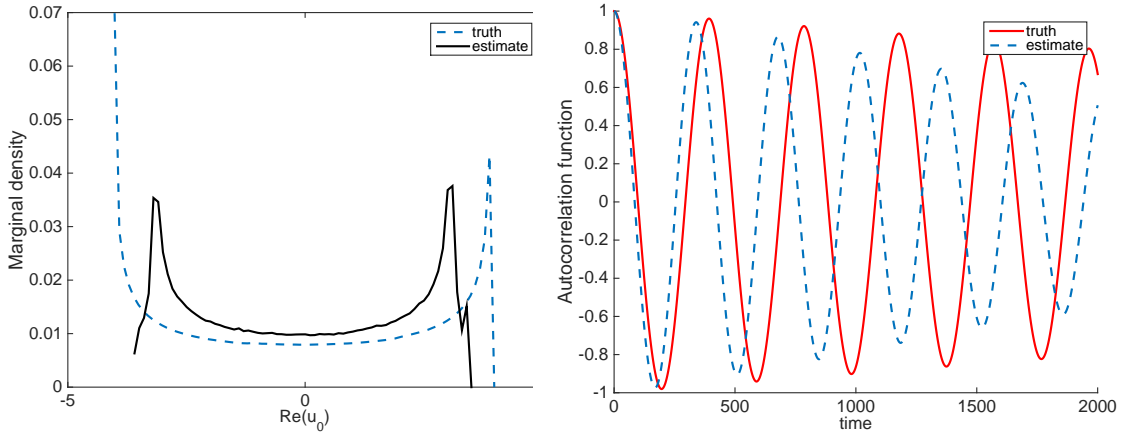

FIG. 6. (Color online) Predicted marginal distribution and correlation function for $\beta = 10$ using the first reduced model (3.16).

VI.   NUMERICAL RESULTS FOR MULTIPLE RETAINED FOURIER MODES

In this section, we consider modeling three modes, u_{-1}, u_0, and u_1, in the Fourier series for much higher temperature case with $\beta = 1/20$ ($k_B T = 20$). In this numerical experiment, the parametric model in (3.23) has 6 dimensional complex valued variables and 18 real valued parameters. For this case, we found several numerical issues when including more Fourier models in the reduced models. First, Fourier modes u_{-1} and u_1 exhibit very different frequency compared with that of u_0, which can be seen in Fig. 8. As a consequence, the
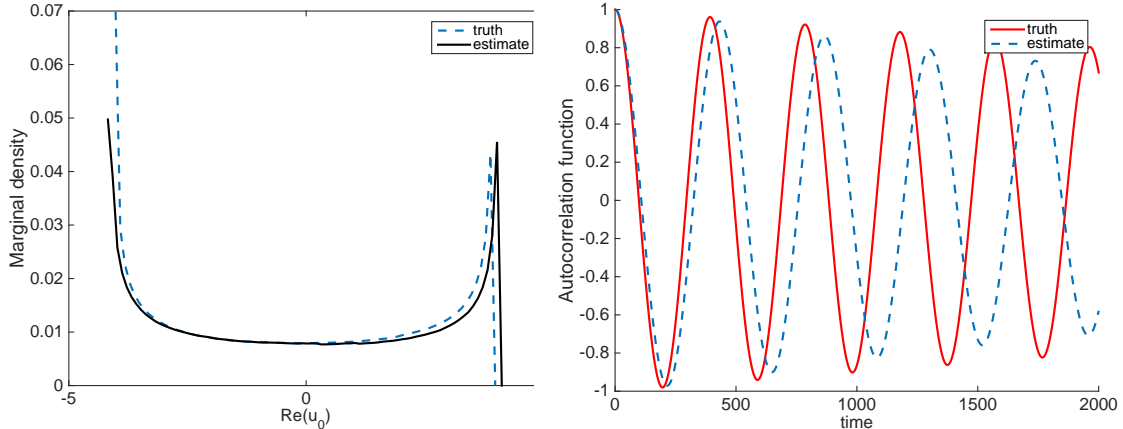
FIG. 7. (Color online) Predicted marginal distribution and correlation function for $\beta = 10$ using the reduced model (3.19) .

variance of each component has different scales (see Table II). The disparity in covariance scaling becomes exceedingly large when the temperature is low. As a result, much smaller time steps are needed in the estimation procedure to sample the observations for u_{-1} and u_1. On the other hand, the procedure has to be continued for a long time period to make sufficient observations of u_0. A more flexible estimation method would be more useful in this case.

A second related issue is that these three multiscale Fourier modes are correlated and this suggests that one may need a different ansatz for the parametric models. For example, one may need to consider fully correlated noises in the equations for f_k in (3.23), which means more parameters to fit. An alternative way to overcome this issue is to fit the model in (3.23) to the uncorrelated observations that can be obtained by rescaling the observations with the covariance matrix. In particular, we define our observations as follows,

$$v_j = \tilde{u}_j + \varepsilon_j, \quad \varepsilon \sim \mathcal{N}(0, R), \tag{6.34}$$

where $\tilde{u}_j \equiv C^{-1/2} u_j$ is the rescaled of u_j by the equilibrium covariance matrix C that can be computed empirically through time averaging of a long time series, assuming the stationarity and ergodicity of the underlying dynamics. Note that this rescaling improves the identifiability of u_0 that has much larger variance relative to u_1, u_{-1} (again, see Table II) since the equilibrium covariance of the rescaled variables $\tilde{u}_j$ is an identity covariance matrix, $\mathcal{I}$. In our numerical experiment below, we assume that the observation error covariance to be 10% of identity, $R = 0.1\mathcal{I}$ and we fit the rescaled observations in (6.34) to (3.23), where

the observation time interval is chosen to be $T_{obs} = 0.02$ and the training data set is 100000 data points.

To confirm the success of the filtering procedure, we see that the filter estimate for R converges to the true value, $R = 0.1\mathcal{I}$, and the filter estimates for $\tilde{u}_j$ have an equilibrium covariance that is indeed identity, exactly equals to the equilibrium covariance of $\tilde{u}_j$. The last point here, however, does not imply that the solutions of the reduced model in (3.23), integrated with the estimated parameters, will have an identity equilibrium covariance matrix. To verify the predictive skill of the resulting parametric model in (3.23), we rescale the solutions, $\breve{u}_j$, to the appropriate scaling of the underlying dynamics with the following covariance transformation, $\hat{u}_j = C^{1/2}\tilde{C}^{-1/2}\breve{u}_j$, where $\tilde{C}$ is the equilibrium covariance of $\breve{u}_j$. In our numerical experiment, we compute this statistics, $\tilde{C}$, by averaging the solutions, $\breve{u}_j$, of (3.23) at 6000-10000 model time units, at discrete time step of $T_{obs} = 0.02$. In particular, the solutions $\breve{u}_j$ are obtained by integrating the model in (3.23) with parameters determined by averaging over the last 1000 steps of the filter estimates.

In Fig. 8, an example of the solutions from the parametric model (3.23) is shown; here we compare the estimates $\breve{u}_j$ (black solid line) with the truth u_j (red dashes) at an arbitratry period of time interval. Notice that even if we don't expect a path-wise agreement, the qualitative behavior of the solutions are reasonably reproduced (in the sense that their magnitude and frequency are qualitatively comparable). In Fig. 9 the comparison of the histogram and the time correlation functions to those of the full model is demonstrated. Notice that despite the difference in amplitude and temporal scalings between modes u_0 and u_{-1}, u_1, the nontrivial marginal distributions are well captured. The correlation times for mode u_0 are well captured at least until 10 unit time; for the other modes, $\{u_{-1}, u_1\}$, the correlation times are in agreement for about one period of oscillation (approximately up to one unit time). We also report the first four moments estimates compared to those of the truth for each variables in Table II. The exact agreement in terms of variances are not surprising since we purposely scale the estimates to match the covariance of the true dynamics. However, the agreement in terms of the higher order moments such as skewness and kurtosis is nontrivial.
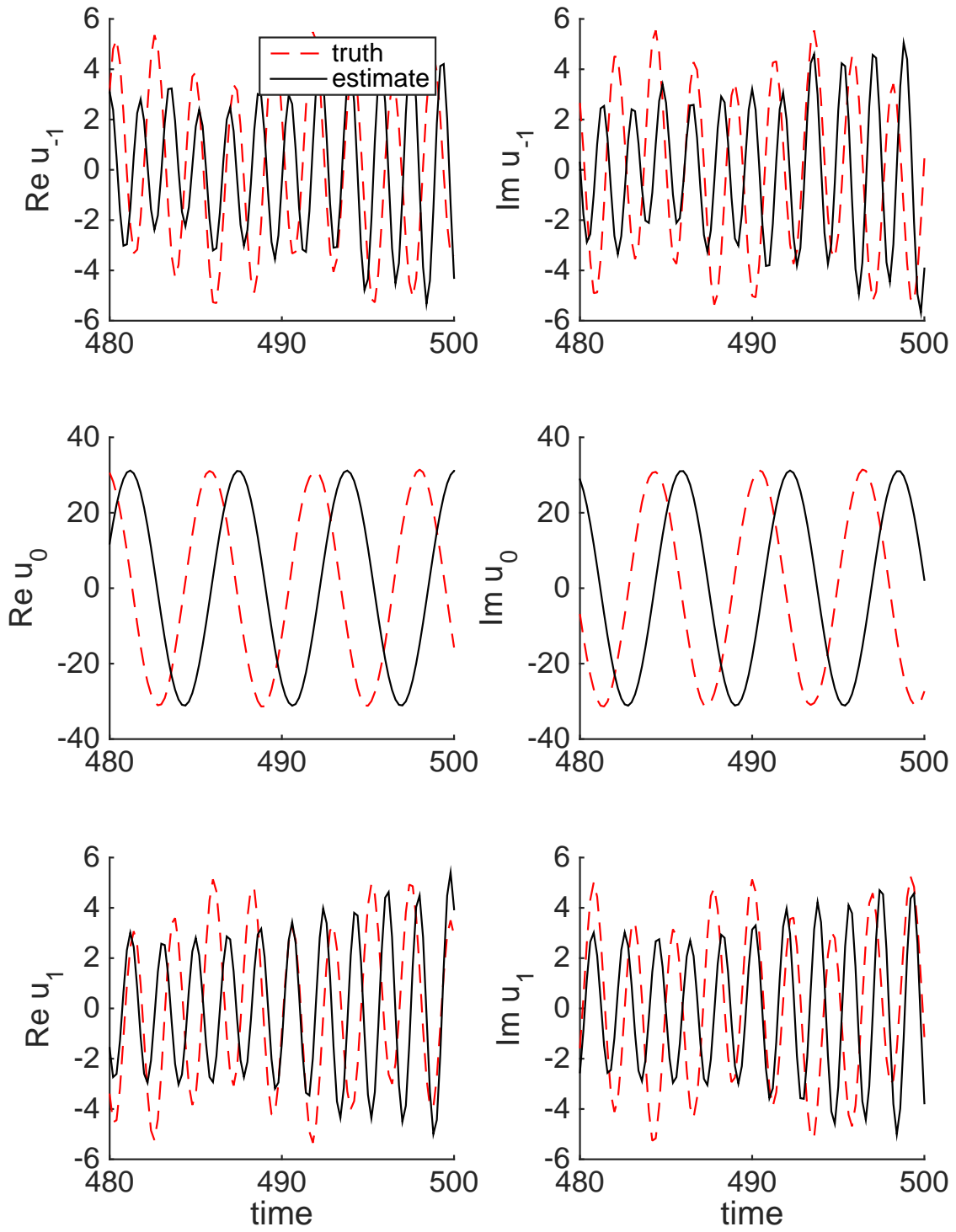
21

FIG. 8. (Color online) Solutions of the reduced model in (3.23), compared to those of the full model.
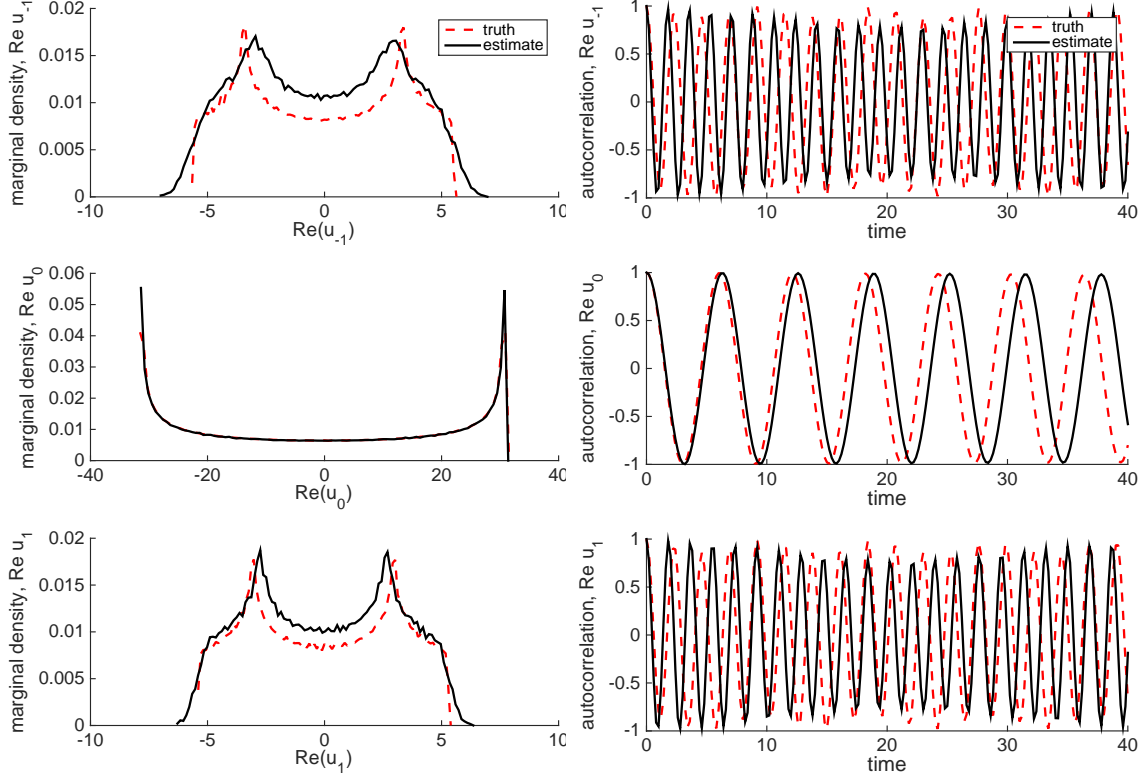
FIG. 9. (Color online) Predicted marginal distributions and correlation functions for $\beta = 1/20$ from the reduced model in (3.23)

TABLE II. (Color online) Equilibrium statistics predicted by the model in (3.23) compared to those of the full model.

statistics	Reu_{-1}		Reu_0		Reu_1	
	truth	estimate	truth	estimate	truth	estimate
mean	0.0008	0.0005	-0.0036	0.0164	0.0019	-0.0004
variance	10.5646	10.5646	487.4128	487.4128	9.3998	9.3998
skewness	-0.0002	-0.0003	0.0004	-0.0009	-0.0007	-0.0002
kurtosis	1.6594	1.8106	1.5000	1.5000	1.6999	1.8017

VII. SUMMARY AND DISCUSSION

This paper presented a modeling approach that blends some physical knowledge about the underlying dynamics and the availability of training data to predict low-frequency modes

of the NLS equation. In particular, we use the Mori-Zwanzig formalism as guidelines to construct effective parametric models and apply an adaptive ensemble Kalman filter to estimate the parameters. The novelty here is that we approximate the memory term and the orthogonalized dynamics of a generalized Langevin equation obtained from the Mori-Zwanzig expansion with a rational function and a colored noise, respectively. It turns out that the resulting parametric model here is an example of the physics constrained nonlinear regression modeling approaches proposed in [22, 26]. This serendipity allows one to use the stability conditions established in [22] to ensure non-blow up solutions of the resulting parametric model. Compared to the full GLE, these models have advantages in practical implementations because they do not involve memory.

The climatological forecasting skill of the proposed parametric model was verified in terms of the first four moments, marginal densities, and correlation functions for various temperatures. For low temperature case, high predictive skill of Fourier mode u_0 is obtained with a reduced model with a scalar parameterization for the memory term (3.16). For higher temperature case where the scale-gap is smaller than the low temperature case, the problem becomes more challenging. In this situation, we showed that one can improve the estimates either with a two-dimensional parameterization for the memory term in (3.19) or with fitting more modes into a model with more retained modes in (3.23).

With the encouraging results in this paper, we plan to apply this modeling strategy on other applications such as on coarse-grained biomolecular models [9–11, 14] in our future research. In general problems, however, the success of this modeling approach will depend mostly on the choice of the ansatz for modeling the memory terms. As it has been theoretically established in [37], if the ansatz is adequate, then it is possible to obtain, both, accurate climatological statistical forecasting and optimal filtering. Our NLS example in this paper empirically suggested that our ansatz is optimal in this case. Other potential issue is in the parameter estimation strategy which can be expensive when more observations are included. While many cheaper parameterization methods are available (such as regression-based or maximum likelihood-based algorithms), these methods are often inferior to the adaptive method applied in the present work even when adequate ansatz is used as shown in [37]. Therefore, improving the numerical efficiency of the adaptive parameter estimation scheme that we used here [26] or its variant (see e.g., [35, 36]) will be the key for successful applications in more complex problems.

---

[1] D. Givon, R. Kupferman, and A. Stuart, Nonlinearity **17**, 55 (2004).

[2] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

[3] R. Zwanzig, J. Stat. Phys. **9**, 215 (1973).

[4] R. Zwanzig, Lectures in Theoretical Physics **3**, 106 (1961).

[5] A. J. Chorin, A. Kast, and R. Kupferman, Proc. Nat. Acad. Sci. USA **96**, 4094 (1998).

[6] P. Espanol, in *Novel Methods in Soft Matter Simulations* (Springer, 2004) pp. 69–115.

[7] C. Hijón, P. Español, E. Vanden-Eijnden, and R. Delgado-Buscalioni, Faraday discuss. **144**, 301 (2010).

[8] A. Chorin and P. Stinis, Communications in Applied Mathematics and Computational Science **1**, 1 (2007).

[9] S. Izvekov and G. A. Voth, J. Chem. Phys. **125**, 151101 (2006).

[10] O. F. Lange and H. Grubmüller, J. Chem. Phys. **124**, 214903 (2006).

[11] B. Oliva, X. Daura, E. Querol, F. X. Avilés, and O. Tapia, Theor. Chem. Acc. **105**, 101 (2000).

[12] E. Darve, J. Solomon, and A. Kia, Proceedings of the National Academy of Sciences **106**, 10884 (2009).

[13] D. Venturi and G. E. Karniadakis, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science **470**, 20130754 (2014).

[14] M. Chen, X. Li, and C. Liu, J. Chem. Phys. **141**, 064112 (2014), arXiv:1403.6543.

[15] P. Stinis, arXiv preprint arXiv:1211.4285 (2012).

[16] A. J. Chorin, O. H. Hald, and R. Kupferman, Proceedings of the National Academy of Sciences **97**, 2968 (2000).

[17] C. Hijón, M. Serrano, and P. Español, J. Chem. Phys. **125**, 204101 (2006).

[18] D. Kauzlarić, J. T. Meier, P. Español, S. Succi, A. Greiner, and J. G. Korvink, J. Chem. Phys. **134**, 064106 (2011).

[19] M. Berkowitz, J. Morgan, and J. A. McCammon, J. Chem. Phys. **78**, 3256 (1983).

[20] M. Berkowitz, J. D. Morgan, D. J. Kouri, and J. A. McCammon, J. Chem. Phys. **75**, 2462 (1981).

[21] X. Li, Int. J. Numer. Meth. Engng. **99**, 157 (2014).

[22] A. Majda and J. Harlim, Nonlinearity **26**, 201 (2013).

[23] J. Bourgain, Communications in mathematical physics **176**, 421 (1996).

[24] J. L. Lebowitz, H. A. Rose, and E. R. Speer, Journal of statistical physics **50**, 657 (1988).

[25] N. Tzvetkov, Ann. Inst. Fourier (Grenoble) **58**, 2543 (2008).

[26] J. Harlim, A. Mahdi, and A. Majda, Journal of Computational Physics **257, Part A**, 782 (2014).

[27] W. Bao, S. Jin, and P. A. Markowich, SIAM Journal on Scientific Computing **25**, 27 (2003).

[28] Z. Bai, Applied Numerical Mathematics **43**, 9 (2002).

[29] R. W. Freund, Acta Numerica **12**, 267 (2003).

[30] G. E. Uhlenbeck and L. S. Ornstein, Physical review **36**, 823 (1930).

[31] B. Hunt, E. Kostelich, and I. Szunyogh, Physica D **230**, 112 (2007).

[32] R. Mehra, Automatic Control, IEEE Transactions on **15**, 175 (1970).

[33] R. Mehra, Automatic Control, IEEE Transactions on **17**, 693 (1972).

[34] P. Belanger, Automatica **10**, 267 (1974).

[35] T. Berry and T. Sauer, Tellus A **65**, 20331 (2013).

[36] Y. Zhen and J. Harlim, J. Comput Phys. **294**, 619 (2015).

[37] T. Berry and J. Harlim, Proc. Roy. Soc. A 20140168 (2014).

Appendix A: Remarks on the second memory terms in (2.11)

Mathematically, one can also include the second memory term using the similar rational approximation for the high temperature case when this term is not negligible. Denoting the other kernel function as,

$$\beta g(t) = i \int_0^t \phi_0(t-\tau) u_0(\tau) |u_0(\tau)|^2 d\tau, \tag{1.1}$$

where β is an additional parameter and approximate the Laplace transform of the kernel function,

$$\widetilde{\phi}_0(s) \approx r_\phi(s) = \frac{-\beta^2}{s-\alpha}, \tag{1.2}$$

where we assume that α and β are real valued parameters. The function $g(t)$ follows the differential equation,

$$\dot{g} = \alpha g - i\beta |u_0|^2 u_0. \tag{1.3}$$

Adding white noises into (1.3), we obtain a parametric model given by,

$$\begin{cases} \dot{u}_0 = -ic\dfrac{1}{2}u_0 - id|u_0|^2 u_0 + bf + \beta g \\ \dot{f} = af - b^* u_0 + \sigma_1 \dot{W}_f \\ \dot{g} = \alpha g - i\beta |u_0|^2 u_0 + \sigma_2 \dot{W}_g, \end{cases} \tag{1.4}$$

where we have added an equation of g to represent the second memory term in (2.11). The problem here is that the nonlinear terms do not conserve energy since we can not control the nonlinear terms in the equation for g unless for $\beta \neq 0$. We suspect that there probably exists different approximations (other than the rational functions) for these kernel functions that give stable parametric models and these are beyond the scope of this paper. Based on this consideration, we do not implement the parametric model in (1.4) in this paper.

Appendix B: Pseudo-algorithm for parameter estimation

This Appendix provides a pseudo-algorithm of the estimation method proposed in [26]. Consider the following filtering problem,

$$\tilde{x}_j = f(\tilde{x}_{j-1}) + \Gamma\epsilon_k, \quad \epsilon_k \sim \mathcal{N}(0, Q), \tag{2.1}$$

$$v_j = H\tilde{x}_j + \epsilon_j^o, \quad \epsilon_j^o \sim \mathcal{N}(0, R),$$

where $\tilde{x}_j = (x_j, \theta_{d,j})$ denote the augmented state and deterministic parameters. Here, we assume a persistence model for the deterministic parameters, $\theta_{d,j} = \theta_{d,j-1}$. We attempt to estimate $\tilde{x}_j$ as well as Q and R, on-the-fly. Essentially, Q and R are the stochastic parameters through the following relation,

$$Q = \sum_{i=1}^{p} Q_i \theta_{s,i}, \quad R = \sum_{i=1}^{p} R_i \theta_{s,i}.$$

and our aim is to estimate $\theta_{s,i}$, $i = 1, \ldots, p$. For the model in (3.17), the augmented state-parameters are $\tilde{x} = (\mathrm{Re}\{u_0\}, \mathrm{Im}\{u_0\}, \mathrm{Re}\{f\}, \mathrm{Im}\{f\}, a_1, a_2, b_1, b_2, c, d)^\top$, the number of stochas-

tic parameters are $p = 2$, where $\theta_{s,1} = \sigma_1^2$, $\theta_{s,2} = R$, and

$$\Gamma = \begin{pmatrix} 0\,0\,\frac{1}{\sqrt{2}}\;0\;0\ldots0 \\ 0\,0\;0\;\frac{1}{\sqrt{2}}0\ldots0 \end{pmatrix}^{\top}, \quad Q_1 = \mathcal{I}_2, \quad Q_2 = R_1 = 0, \quad R_2 = 1. \tag{2.2}$$

Starting with time index $j = 1$, we provide an ensemble of prior statistical estimates, $\{\tilde{x}_j^{k,-}\}_{k=1}^{K}$, of size K for the primary filter and prior mean $\{\theta_{s,i,j}\}_{i=1}^{p}$ and covariance $\Theta_j = \mathcal{I}_p$, for the secondary filter. The primary filter for estimating $\tilde{x}_j$ is described in Steps 1-3, while the secondary filter for estimating θ_s is described in Steps 4-9.

1. Apply the ETKF to obtain the analysis ensemble estimate, $\{\tilde{x}_j^{k,+}\}_{k=1}^{K}$. Let's denote the corresponding Kalman gain and innovation as follows,

$$\tilde{K}_j = P_j^- H^\top (H P_j^- H^\top + \sum_{i=1}^{p} R_i \theta_{s,i,j})^{-1}$$

$$\epsilon_j = v_j - H \bar{x}_j^-,$$

where $\bar{x}_j^- = K^{-1} \sum_{k=1}^{K} \tilde{x}_j^{k,-}$ denotes the prior ensemble average. See [31] for the detail ETKF algorithm.

2. Propagate each ensemble member with the deterministic part of the model in (2.1) to obtain,

$$\tilde{x}_{j+1}^{k,d} = f(\tilde{x}_j^{k,+}), \quad k = 1,\ldots,K,$$

and form the posterior ensemble by adding a Gaussian noise,

$$\tilde{x}_{j+1}^{k,+} = \tilde{x}_{j+1}^{k,d} + \psi^k, \quad \psi^k \sim \mathcal{N}\big(0, \Gamma(\sum_{i=1}^{p} Q_i \theta_{s,i,j})\Gamma^\top\big), \quad k = 1,\ldots,K.$$

3. Define an ensemble approximation for the linear tangent model,

$$A_j \equiv \nabla f(\bar{x}_j^+) \approx U_{j+1}^d W_j^\dagger, \tag{2.3}$$

where each column vectors of U_{j+1}^d and W_j are the deterministic forecast ensemble perturbations and the analysis ensemble perturbations, consecutively. In (2.3), we denote pseudo-inverse by $\dagger$.

4. Define $K_j = A_j \tilde{K}_j$ and $\phi_j = A_j - K_j H$.

5. For each $i=1,\ldots,p$, construct an observation operator for $\epsilon_j \epsilon_j^\top$, starting with $S_{i,1,0}=0$, let

$$M_{i,j,0}=S_{i,j,0}H^\top,$$

$$F_{i,j,0}=HM_{i,j,0}+R_i,$$

$$S_{i,j+1,0}=\phi_k S_{i,j,0}\phi_j^\top +\Gamma Q_i \Gamma^\top +K_j R_i K_j^\top.$$

6. For each $i=1,\ldots,p$, construct an observation operator for $\epsilon_j \epsilon_{j-\ell}^\top$, where $k>1$. Set

$$M_{i,j,\ell}=\phi_{j-1}M_{i,j-1,\ell-1}-K_{j-1}R_i \delta_{\ell,1}$$

$$F_{i,j,\ell}=HM_{i,j,\ell}$$

7. Approximate $\mathbb{E}(v_j v_j^\top)=\sum_{i=1}^p F_{i,j,0}\theta_{s,i,j}$. Suppose if $\epsilon_j=(\epsilon_j^1,\ldots,\epsilon_j^m)^\top$ is m-dimensional. Define

$$\sigma_{j,\ell}\equiv vec(\epsilon_j \epsilon_{j-\ell}^\top)=(\epsilon_j^1 \epsilon_{j-\ell}^1,\epsilon_j^2 \epsilon_{j-\ell}^1,\ldots,\epsilon_j^m \epsilon_{j-\ell}^1,\ldots,\epsilon_j^1 \epsilon_{j-\ell}^m,\epsilon_j^2 \epsilon_{j-\ell}^m,\ldots,\epsilon_j^m \epsilon_{j-\ell}^m)^\top.$$

8. Consider the pseudo observation model for the secondary filter,

$$\sigma_{j,\ell}=\mathcal{F}_{j,\ell}\theta_s+\eta_{j,\ell},\quad \eta_{j,\ell}\sim\mathcal{N}(0,W_{j,\ell}),\quad \ell=1,\ldots,L, \tag{2.4}$$

where in our case, $\sigma_{j,\ell}=vec(\epsilon_j \epsilon_{j-\ell}^\top)\in\mathbb{R}^+$, $\mathcal{F}_{j,\ell}=(F_{1,j,\ell},\ldots,F_{p,j,\ell})$, $\theta_s=(\theta_{s,1},\ldots,\theta_{s,p})^\top$, and for each pair of indices $\{k,\ell\}$, construct

$$W_{j,\ell}=\mathbb{E}(\epsilon_j \epsilon_j^\top)\mathbb{E}(\epsilon_{j-\ell}\epsilon_{j-\ell}^\top)+\mathbb{E}(\epsilon_j \epsilon_j^\top)^2 \delta_{\ell,0}.$$

Note that W is constructed, assuming Gaussian and independent noises, $\eta_{j,\ell}$. Components of matrix W in (2.5) can be rewritten as follows,

$$W_{j,l}^{\alpha,\beta,\gamma,\delta}=\mathbb{E}(\epsilon_j^\alpha \epsilon_j^\gamma)\mathbb{E}(\epsilon_{j-\ell}^\beta \epsilon_{j-\ell}^\delta)+\mathbb{E}(\epsilon_j^\alpha \epsilon_j^\delta)\mathbb{E}(\epsilon_j^\beta \epsilon_j^\gamma)\delta_{\ell,0}.$$

9. Perform a secondary Kalman filter L-times to sequentially update $\theta_{s,i,j+1}$ with observation models in (2.4) one at the time, assuming that the dynamics of these parameters are persistence, $\dot{\theta}_{s,i}=0$. Now we can repeat Step 1 above for the new assimilation time.