Filtering Nonlinear Spatio-Temporal Chaos with Autoregressive Linear Stochastic Model

Emily L. Kang\textsuperscript{a,*}, John Harlim\textsuperscript{b}

\textsuperscript{a}Department of Mathematical Sciences, University of Cincinnati, Cincinnati, OH, United States.
\textsuperscript{b}Department of Mathematics, North Carolina State University, Raleigh, NC, United States.

Abstract

Fundamental barriers in practical filtering of nonlinear spatio-temporal chaotic systems are model errors attributed to the stiffness in resolving multiscale features. Recently, reduced stochastic filters based on linear stochastic models have been introduced to overcome such stiffness; one of them is the Mean Stochastic Model (MSM) based on a diagonal Ornstein-Uhlenbeck process in Fourier space. Despite model errors, the MSM shows very encouraging filtering skill, especially when the hidden signal of interest is strongly chaotic. In this regime, the dynamical system statistical properties resemble to those of the energy-conserving equilibrium statistical mechanics with Gaussian invariant measure; therefore, the Ornstein-Uhlenbeck process with appropriate parameters is sufficient to produce reasonable statistical estimates for the filter model.

In this paper, we consider a generalization of the MSM with a diagonal autoregressive linear stochastic model in Fourier space as a filter model for chaotic signals with long memory depth. With this generalization, the filter prior model becomes slightly more expensive than the MSM, but it is still less expensive relative to integrating the perfect model which is typically unknown in real problems. Furthermore, the associated Kalman filter on each Fourier mode is computationally as cheap as inverting a matrix of size $D$, where $D$ is the number of observed variables on each Fourier mode.

*Corresponding Author.

Email addresses: kangel@ucmail.uc.edu (Emily L. Kang), jharlim@ncsu.edu (John Harlim)
(in our numerical example, $D = 1$). Using the Lorenz 96 (L-96) model as a testbed, we show that the non-Markovian nature of this autoregressive model is an important feature in capturing the highly oscillatory modes with long memory depth. Secondly, we show that the filtering skill with autoregressive models supersedes that with MSM in weakly chaotic regime where the memory depth is longer. In strongly chaotic regime, the performance of the AR(p) filter is still better or at least comparable to that of the MSM. Most importantly, we find that this reduced filtering strategy is not as sensitive as standard ensemble filtering strategies to additional intrinsic model errors that are often encountered when model parameters are incorrectly specified.

**Keywords:** Autoregressive models, Kalman filter, Lorenz 96 (L-96) model, non-Markovian linear stochastic model, Ornstein-Uhlenbeck process

1. **Introduction**

Given noisy observations from nature, filtering is the process of finding the best statistical estimate of the true signal. Filtering consists of a two-step predictor-corrector scheme that adjusts a prior estimate to be more consistent with the current observations. The revised estimate is then fed into the model as an initial condition for the future time prediction. This approach of generating initial conditions is also known as data assimilation. In practice, the demand of practical filtering methods for real-time prediction (or state estimation) problem escalates as the model resolution is significantly increased. For example, in the coupled atmosphere-ocean system, the current practical models for prediction of both weather and climate involve general circulation models where the physical equations for these extremely stiff complex flows are discretized in space and time and the effects of unresolved processes are parametrized according to various recipes; the result of these approximations is an extremely unstable chaotic prediction model with several billion degrees of freedom.

In the data assimilation community, many well developed algorithms based on the Bayesian hierarchical modeling [1] and reduced order Kalman filtering strategies [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12] have shown some successes in these extremely complex high dimensional nonlinear systems. For nonlinear and non-Gaussian problems, the current important scientific problem is to turn the particle filter, which was well developed theoretically [7, 13, 14, 15, 16] but suffers from ‘curse of dimensionality’ [17, 18, 19, 20], into a prac-
tically useful method for filtering complex systems described above. There are a few potentially useful particle filtering algorithms for high dimensional systems such as the one proposed in [21] which constraints the prior particles to be sufficiently close to the future observations whenever they are available and the rank histogram filter [22] that uses the order statistics to redistribute the posterior distribution. Having all the above successful ensemble based filtering strategies (despite that most of them are sensitive to variation of parameters such as the ensemble size, localization, variance inflation, etc, so one needs to tune them on an ad-hoc basis), there is still an inherently difficult practical issue for some high dimensional complex problems due to the large computational overhead in generating individual ensemble members through the forward dynamical operator [23]. Thus, it is worthwhile to develop an ensemble-less filtering method as a complementary strategy that is robust and not sensitive to variations of parameters, and this is the goal of this paper.

Recently, the second author with his collaborators proposed much simpler and computationally faster filtering strategies with linear stochastic models resulted from committing judicious model errors [24, 25]. In particular, the design of their reduced stochastic filters is attributed to the standard approach for modeling turbulent fluctuations [26, 27, 28, 29, 30, 31] with a linear damping and white noise (replacing nonlinear terms). The resulting filtering strategy, the Mean Stochastic Model (MSM), consists of diagonal Langevin equations in Fourier space where the parameters are obtained through an off-line regression fitting to a training data set (we will review this filtering method in Section 2 below). On the other direction, they also introduce the stochastic parameterized “extended” Kalman filter (SPEKF) which estimates the parameters in the MSM “on-the-fly” with an exactly solvable non-Gaussian prior statistics without linearized tangent models [32, 33, 34]. These cheap diagonal filters (MSM and SPEKF) have shown encouraging results in various applications involving more realistic higher dimensional problems with $\mathcal{O}(10^4)$ state variables. In [35], the MSM and SPEKF are superior to the ensemble Kalman filters in estimating midlatitude barotropic flows with baroclinic instability with sparse observations on a numerically stiff regime with shorter radius of deformation, mimicking the ocean turbulence. In [36], the MSM and SPEKF are used for estimating ocean eddy heat transport with satellite altimetry. In [37], the MSM model is also used to capture the initiation of moist convectively coupled waves in the tropics as well as the MJO-like traveling waves through partial observation network; there,
the MSM is designed for an eight-dimensional system with an appropriate eigenmode basis on each Fourier component (the eight variables include the first two baroclinic winds and potential temperatures, the equivalent boundary layer potential temperature, moisture, and heating rates of three cloud types above the boundary layer). Independently in [38], Law and Stuart assessed the performance of various data assimilation methods including the MSM (they called it the Fourier Diagonal Filter or FDF) on the 2D Navier-Stokes equations resolving $O(10^3)$ horizontal grid points; in particular, they found that in turbulent regime the FDF performs better than 3D-VAR and ensemble Kalman filter for longer observation time and they explain that the success of FDF in this regime is due to the stability of the forward model in MSM by design. For detailed discussions of Fourier domain filters for sparse observations, we recommend the review article [34] and the forthcoming book [39] for interested readers.

In [24, 25], the main result is that the MSM strategy produces high filtering skill beyond ensemble Kalman filters with perfect model for strongly chaotic and fully turbulent systems. The high filtering skill with MSM in this regime is because the statistical properties of the dynamical system in this regime resemble to those of the energy-conserving equilibrium statistical mechanics with Gaussian invariant measure; therefore, the Ornstein-Uhlenbeck process with appropriate parameters is sufficient to produce reasonable statistical estimates for the filter model. In weakly chaotic regime, some modes are oscillatory and behave like traveling waves with longer correlation times, and the MSM approximation produces unsatisfactory filtered solutions [24, 25]. Here, we propose to use autoregressive linear stochastic models to capture the non-Markovian long memories in these slowly decaying oscillatory modes. We subsequently apply the standard linear Kalman filter with this autoregressive model, “the AR(p) filter”; here, the computational cost is as expensive as inverting a matrix of size $D$ where $D$ is the number of observed variables on each Fourier mode ($D = 1$ in our numerical test example). Note that filtering with autoregressive models has been widely used in engineering community [40, 41, 42] but not in the context of stiff complex turbulent systems with multiple scales.

Given the encouraging results in various applications on high-dimensional problems with the MSM method as we discussed above [35, 36, 38, 37], our aim here is to understand the merit of using autoregressive model over the MSM as a filter model given only time series of turbulent signals at statistical steady state. In particular, our goal is to understand the strengths
and weaknesses of the proposed AR(p) filter in assimilating different characteristic of chaos. For this purpose, we choose Lorenz 96 (L-96) model [43] as a testbed since this model is simple yet produces different characteristic of chaos ranging from weakly chaotic to strongly chaotic to fully turbulent regime as the external forcing is varied. The solutions of this model are rich enough in the sense that the most energetic mode does not coincide with the mode with the largest correlation time (longest memory) as typically observed in nature. Despite its simplicity, the design of L-96 model is to mimic the large scale midlatitude atmospheric dynamics with an inhomogeneous (or anisotropic) spectrum with no power law cascade. From dynamical system point of view, this model has an “absorbing property” (that is, a dissipative dynamical system with bounded solutions), and its solutions distribution is unimodal, slightly skewed from Gaussian distribution [44]. Note that this model has also been used as a testbed for various problems ranging from climate modeling [45] to estimating the mean linear response operator to the change in external forcing which is a problem of contemporary interest in climate change sciences [46, 47] to, obviously, filtering problems [48, 25, 49].

To emphasize the effect of errors in the forward model, we only consider full observation network here. There are various versions of Fourier domain diagonal filters with MSM that address sparse regularly spaced [48, 35] and irregularly spaced observations [50] and their application with AR(p) model is subject to our on-going research. Here, we will compare the filtered solutions of the AR(p) filter for appropriate choice of $p$ as well as for $p = 1$ with those of the MSM [24, 35] and the perfect model simulation with the ensemble adjustment Kalman filter [5] for various chaotic regimes.

The remainder of this paper is organized as follow: In Section 2, we discuss the design of autoregressive linear stochastic model in the context of turbulent modeling with L-96 model as a prototype example. There, we also review the MSM model. Subsequently, we describe the corresponding AR(p) filter in Section 3 and provide numerical results in Section 4. Finally, we conclude the paper with a short summary and discussion in Section 5.
2. Autoregressive Linear Stochastic Model for Complex Turbulent Systems

The simplest approach in climate modeling is to fit the statistics of the given turbulent signals to the following linear regression model [51, 30],

\[
\frac{d u_R}{dt} = L_R u_R + \sigma_R \dot{W} + F,
\]

where \( \dot{W} \) is a vector white noise, \( \sigma_R \) is the noise strength, \( F \) denotes an external forcing, and \( L_R \) has eigenvalues with negative real part. Since \( L_R \) has a decaying spectrum, one strategy for filtering [24] is to use solution from this model as a surrogate for the actual prior estimate, which has a vastly reduced computational overhead compared to the true model that is often unknown in real applications. However, there are inherent realizability issues so that there often is not a linear operator, \( L_R \), with negative real eigenvalues which reproduces the lagged covariances [51, 52] so that there are severe model errors for filtering [35, 39]. A different linear regression strategy involving Ornstein-Uhlenbeck processes was introduced recently [46, 35, 47, 39], the Mean Stochastic Model (MSM), which is guaranteed to always be realizable by matching the variances and the integral of the lagged covariances with data from the turbulent dynamical system. Below, we will review the MSM model to motivate autoregressive modeling. For simplicity in exposition, we use Lorenz 96 (L-96) model [43] as a canonical example; we will review the statistical behavior of this model in detail to motivate the used of this toy model in this paper. The same derivations for more realistic higher dimensional problems can be found in [35, 39, 37].

2.1. The Lorenz-96 Model

The Lorenz-96 (L-96) model [43] was introduced as a toy model for an "atmospheric variable" \( u \) at \( J \) equally spaced grid points on a circle of constant latitude. Specifically, the \( j \)th component evolves in time as follows:

\[
\frac{d u_j}{dt} = (u_{j+1} - u_{j-2})u_{j-1} - u_j + F,
\]

for \( j = 0, \ldots, J - 1 \), with periodic boundary condition, \( u_{j+J} = u_j \). As seen in (2), L-96 model has a quadratic nonlinear interaction advective-like term, a linear damping term, and an external forcing term. As in [43, 53], we set \( J = 40 \) so that the distance between two grid points is roughly the midlatitude
Rossby radius ($\approx 800$ km), where the circumference of the midlatitude belt is assumed to be around 30,000 km.

In this paper, we investigate the filtering skills on L-96 model in regimes ranging from weakly chaotic ($F = 5$) to fully turbulent ($F = 32$), given full observations. Note that this 40-mode model has 9 positive Lyapunov exponents for $F = 5$ and 16 positive Lyapunov exponents for $F = 16$ (see Table 1 of [54] for complete list of chaotic measures including the largest Lyapunov exponent and the Kolmogorov-Sinai entropy). Snapshots of the time series for (2) with $F = 6, 8, 16$, as depicted in Fig. 1, qualitatively confirm the above quantitative intuition with weakly chaotic patterns for $F = 6$, strongly chaotic wave turbulence for $F = 8$, and fully developed wave turbulence for $F = 16$.

Figure 1: Space-time plot of numerical solutions of L-96 model for weakly chaotic ($F = 6$), strongly chaotic ($F = 8$), and fully turbulent ($F = 16$) regime.

To compare different regimes, we consider the following rescaling in [24, 25, 53],

$$u_j = \bar{u} + E_p^{1/2} \tilde{u}_j, \ t = E_p^{-1/2} \tilde{t},$$

where $\bar{u}$ represents the (temporal) mean state and $E_p$ represents the average variance in the energy fluctuation. In practice, these quantities are computed by integrating (2) and using formulas for empirical long time averages as follows:

$$\bar{u} = \frac{1}{T} \int_{T_0}^{T_0+T} u_j dt,$$

$$E_p = \frac{1}{2T} \sum_{j=0}^{J-1} \int_{T_0}^{T_0+T} (u_j - \bar{u})^2 dt.$$
In our numerical implementation, we integrate (2) using the fourth-order Runge-Kutta method (RK4) with time step $\delta t = 1/64$ and initial startup time $T_0 = 1000$ up to $T = 10,000$. The values of $\bar{u}$ and $E_p$ for various $F$ are summarized in Table 1. After rescaling, $\{\tilde{u}_j\}$ has mean state zero and non-dimensionalized energy fluctuation of one unit, independent of $F$.

The typical statistical measures for turbulent systems are the energy spectrum and correlation time:

$$E_k \equiv Var(\hat{u}_k) = \overline{(\hat{u}_k(t) - \bar{u}_k)(\hat{u}_k(t) - \bar{u}_k)^*}$$  \hspace{1cm} (5)

$$R_k \equiv \int_0^\infty Corr_k(\tau)d\tau = T_k - i\theta_k,$$  \hspace{1cm} (6)

where $\hat{u}_k$ is the corresponding Fourier component of $\tilde{u}_j$ with the discrete Fourier transform:

$$\hat{u}_k = \frac{1}{J} \sum_{j=0}^{J-1} \tilde{u}_j e^{-2\pi i kj/J}, \quad |k| \leq J/2,$$  \hspace{1cm} (7)

and $Corr_k(\tau)$ is the autocorrelation function defined as follows,

$$Corr_k(\tau) \equiv \frac{(\hat{u}_k(t) - \bar{u}_k)(\hat{u}_k(t + \tau) - \bar{u}_k)^*}{E_k}.$$  \hspace{1cm} (8)

In (5) and (8), $\langle \cdot \rangle$ denotes an empirical average and ‘*’ denotes conjugate transpose. In Fig. 3, we show the energy spectra for the L-96 solutions in normalized coordinate for various $F$ including the inviscid-unforced L-96 model (IL96). For the IL96 case, we ignore the linear damping term, $-u_j$, in (2) and set $F = 0$ and therefore the system corresponds to an equipartition energy equilibrium statistical mechanics with Gaussian invariant measure [53, 47]. Notice that the equilibrium variance, $E_k$, which represents the climatological energy spectrum, is more uniformly distributed approaching white noise spectrum of the IL96 model as the forcing $F$ increases. Notice also that the most energetic mode in each chaotic regime does not coincide with mode with the longest correlation time. Furthermore, if one checks the autocorrelation function of the most energetic mode, it decays slowly (with longer memory depth) with oscillatory behavior (e.g., see mode $k = 8$ for the case of $F = 6$ in Fig. 4) and thus the correlation time $T_k$ is small due to cancellations in integrating this oscillatory correlation function. In physical space, this behavior is apparent with nearly regular wave train of eight westward moving waves in
Table 1: The (temporal) mean $\bar{u}$, energy fluctuation $E_p$ for different dynamical regimes, and the largest correlation time, $T \equiv \max_{0 \leq k \leq J/2} T_k / \sqrt{E_p}$, where $T_k$ is real component of the correlation time of each Fourier mode of the rescaled variables (see Eq. (6)).

<table>
<thead>
<tr>
<th>$F$</th>
<th>5</th>
<th>6</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{u}$</td>
<td>1.63</td>
<td>2.01</td>
<td>2.34</td>
<td>3.09</td>
<td>3.89</td>
</tr>
<tr>
<td>$E_p$</td>
<td>109.80</td>
<td>160.53</td>
<td>264.85</td>
<td>797.39</td>
<td>2186.54</td>
</tr>
<tr>
<td>$T$</td>
<td>7.60</td>
<td>1.92</td>
<td>0.70</td>
<td>0.28</td>
<td>0.16</td>
</tr>
</tbody>
</table>

weakly chaotic regime with $F = 6$ in Fig. 1. In Table 1, we also include $T$ to indicate the slowest correlation time $T_k$ as a reference for choosing relevant observation times for the numerical experiments in Section 4. Notice that $T$ decreases as $F$ increases. This confirms our intuition that L-96 model has shorter memory depth as the system becomes more chaotic. In addition, notice the different scales of correlation times for different $F$’s in Fig. 3.

2.2. The Mean Stochastic Models

In this section, we review the design of the Mean Stochastic Models (MSM) in [24, 46]. After rescaling with (3), we obtain the following rescaled L-96 model:

$$\frac{d\tilde{u}_j}{dt} = (\tilde{u}_{j+1} - \tilde{u}_{j-2})\tilde{u}_{j-1} + E_p^{-1/2}((\tilde{u}_{j+1} - \tilde{u}_{j-2})\bar{u} - \tilde{u}_j) + E_p^{-1}(F - \bar{u}).$$

(9)

The MSM is designed by replacing the nonlinear term, $(\tilde{u}_{j+1} - \tilde{u}_{j-2})\tilde{u}_{j-1}$, with an Ornstein-Uhlenbeck (OU) process on each Fourier mode [24, 25, 35, 46, 34],

$$(\bar{u}_{j+1} - \bar{u}_{j-2})\bar{u}_{j-1} \rightarrow \sum_{k=0}^{J-1} \left((-d_k + i\omega_k)\hat{u}'_k + \tilde{\sigma}_k \hat{W}_k\right) e^{2\pi i k j / J},$$

(10)

where $\hat{W}_k$ denotes a complex white noise in time for each Fourier mode $k$. Here, we choose Fourier coordinate since this model has a periodic geometry; such an approach can be applied to any generalized spectral coordinate. We avoid to use OU process in physical space to minimize the number of parameters to be fitted. Note that there are, of course, no guarantee that such an approximation can be justifiable mathematically but there are strong
evident that for the purpose of filtering, such a “poor-man’s” approach is advantageous in fully turbulent regime \([24, 25]\) where the system’s invariant measure is close to Gaussian as we pointed out in Section 1 above.

In Fourier space, the MSM model is expressed through the following diagonal Langevin equation for \(\hat{u}'_k\) that approximates \(\hat{u}_k\),

\[
\frac{d\hat{u}'_k}{dt} = (-d_k + \omega_1(k) + i(\omega_k + \omega_2(k))) \hat{u}'_k + \frac{1}{E_p} (F - \bar{u}) \delta_k + \bar{\sigma}_k W_k,
\]

where \(\omega_1(k) = \text{Re}(A_k)\), \(\omega_2(k) = \text{Im}(A_k)\),

\[
A_k = E_p^{-1}[(e^{2\pi ik/J} - e^{-4\pi ik/J})\bar{u} - 1],
\]

and \(\delta_k = 0\) for \(k \neq 0\), and \(\delta_0 = 1\).

Ignoring the forcing term, the stochastic differential equation in (11) has a Gaussian equilibrium probability measure with explicit covariance (or energy spectrum) and correlation times \([55]\):

\[
E_k \equiv \lim_{\tilde{t} \to \infty} \left\langle \left( \hat{u}'_k(\tilde{t}) - \langle \hat{u}'_k \rangle \right) \left( \hat{u}'_k(\tilde{t}) - \langle \hat{u}'_k \rangle \right)^\ast \right\rangle = \frac{\hat{\sigma}_k^2}{2(d_k - \omega_1(k))} \quad \text{(13)}
\]

\[
R_k \equiv \lim_{\tilde{t} \to \infty} \int_0^\infty \left\langle \left( \hat{u}'_k(\tilde{t} + \tau) - \langle \hat{u}'_k \rangle \right) E_k^{-1} \left( \hat{u}'_k(\tilde{t}) - \langle \hat{u}'_k \rangle \right)^\ast \right\rangle d\tau = \frac{1}{d_k - \omega_1(k) + i(\omega_k + \omega_2(k))}. \quad \text{(14)}
\]

Matching these statistics to the empirically estimated statistics in (5) and (6), we obtain

\[
d_k = \omega_1(k) + \frac{T_k}{T_k^2 + \theta_k^2},
\]

\[
\omega_k = -\omega_2(k) + \frac{\theta_k}{T_k^2 + \theta_k^2},
\]

\[
\hat{\sigma}_k^2 = 2E_k \frac{\theta_k}{T_k^2 + \theta_k^2},
\]

for each wave number. Notice that the MSM regression fitting in (15) always guarantees realizable solutions with \(d_k - \omega_1(k) > 0\) as pointed out in the beginning of Section 2 and in \([46, 47]\); this realizability is not guaranteed for standard linear regression strategy that typically parameterizes the frequency with the linearized frequency by setting \(\omega_k = 0\) in (11) \([30, 51]\).
regression fitting through (15) produces a perfect fit on the energy spectrum and correlation times for any regime, by design (results are not shown in Fig. 3). However, if one looks at the autocorrelation functions (see Figs. 4 and 5), there is no reason to believe that the oscillatory and “exotic-looking” true autocorrelation functions (such as those of wavenumbers 13 and 18) can be reproduced with the analytical correlation functions of the MSM in the following form,

$$\text{Corr}_k(\tau) = e^{-(d_k - \omega_1(k) + i(\omega_2(k)))\tau}. \quad (16)$$

As we already pointed out earlier, modes with longer memory depth (that decay slowly) may have small correlation times due to cancellations in integrating the autocorrelation functions whenever they are oscillatory; in that case, the MSM model with regression fitting in (15) introduces severe model errors. This is the central issue that motivates the next section.

### 2.3. The AR(p) Models for Complex Turbulent Systems

Our goal is to come up with new reduced stochastic models that can capture the “exotic-looking” autocorrelation feature in weakly chaotic regime and still maintain the skill of the MSM in strongly chaotic and fully turbulent regimes. To meet this requirement, we consider fitting the training data set to the following constantly forced complex-valued autoregressive model of order $p$ (AR(p) model) on each Fourier coefficient:

$$\hat{u}_{k,n} = \phi_1 \hat{u}_{k,n-1} + \phi_2 \hat{u}_{k,n-2} + \cdots + \phi_p \hat{u}_{k,n-p} + f_k + \eta_{k,n}, \quad |k| \leq J/2. \quad (17)$$

In (17), index $k$ denotes wavenumber, index $n$ denotes discrete time $t_{n+1} = t_n + \Delta t$, and parameter $p \in \mathbb{Z}$ is called the order of this autoregressive model. Larger $p$ corresponds to longer memory depth of the timeseries that we are fitting. We shall see below that $p$ can be significantly different for various wavenumbers depending on the correlation functions and, obviously, the forcing strength.

In (17), $\eta_{k,n}$ are complex-valued i.i.d. Gaussian noises with mean zero and variance $\sigma_k^2$. For each wavenumber, we have $p + 3$ parameters to be specified, including the constant forcing term $f_k$ that can be estimated by fitting to the empirical mean state, $\tilde{u}_k$, directly, coefficients $\{\phi_j, j = 1, \ldots, p\}$, the noise covariance $\sigma_k^2$, and the autoregressive model order $p$. Given time series of $\{\hat{u}_{k,n}, n = 1, \ldots, N\}$, we first take away the empirical mean and define an $(N - p)$-dimensional vector $\mathbf{Y} \equiv (X_{k,p+1}, \ldots, X_{k,N})^T$, with

$$X_{k,n} = \hat{u}_{k,n} - \tilde{u}_k, \quad (18)$$
and an \((N - p) \times p\) matrix \(X\) as follows:

\[
X \equiv \begin{pmatrix}
X_{k,p} & X_{k,p-1} & \cdots & X_{k,1} \\
X_{k,p+1} & X_{k,p} & \cdots & X_{k,2} \\
\vdots & \vdots & \ddots & \vdots \\
X_{k,N-1} & X_{k,N-2} & \cdots & X_{k,N-p}
\end{pmatrix}.
\]

Then the AR(p) model in (17) can be written in a matrix form as:

\[
Y = X\phi + \eta,
\]

(19)

where \(\phi \equiv (\phi_1, \ldots, \phi_p)^T\) and \(\eta \equiv (\eta_{k,p+1}, \ldots, \eta_{k,N})^T \sim \mathcal{N}(0, \sigma^2_k I)\). The maximum likelihood estimators of \(\phi\) and \(\sigma^2_k\) are in the following form:

\[
\hat{\phi} = (X^*X)^{-1}X^*Y,
\]

(20)

\[
\hat{\sigma}^2_k = \frac{|Y - X\hat{\phi}|^2}{N - p}.
\]

(21)

These estimators are also called the Yule-Walker estimators for the parameters of an AR(p) process (see [56]). Although the likelihood function for an AR(p) model without conditioning on the first \(p\) observations can be derived, explicit formulas of the corresponding (exact) maximum likelihood estimators are not feasible, and numerical procedures are needed for the maximization. Additionally, it has been pointed out that these Yule-Walker estimators have approximately the same distribution as the corresponding (exact) maximum likelihood estimators for a large data set (see [56]). Given the large sample size we have in our off-line training phase, we choose to use the Yule-Walker estimators in (20) and (21) instead of the (exact) maximum likelihood estimators.

To select the appropriate order \(p\) of an autoregressive model, various model selection criteria can be applied. For our numerical analysis in the off-line training phase, we use the following FPE criterion proposed in [57]:

\[
\text{FPE}_k(p) = \hat{\sigma}^2_{k,p} \frac{N + p}{N - p},
\]

(22)

where \(\hat{\sigma}^2_{k,p}\) is the corresponding estimator in (21) for the AR(p) model on mode \(k\) and given \(p\). Then the order \(p\), which gives the minimum of \(\text{FPE}_k(p)\) within \(p = 0, 1, 2, \ldots, P\) (\(P\) is a pre-set upper limit of the order) will be
adopted as the order of the model for prediction. The step-by-step algorithms for choosing $p$ and parameterizing the AR(p) model are provided in Appendix A. In our numerical experiments below, we also show results with AR(1) where we purposely fix $p = 1$ for diagnostic purpose.

In the off-line training phase, we fit the AR(p) model with training data sets with discrete time steps $\Delta t = 4/64, 8/64,$ and $16/64$. Fig. 2 shows the chosen values of $p$ for Fourier modes in different regimes, with $F$ ranging from 5 to 32. Notice that $p$ decreases as the system becomes more chaotic ($F$ increases); this is not so surprising because the correlation time (and absolute correlation time) decreases as $F$ increases (see [53, 47]). Higher $p$ implies longer memory depth as we expected in weakly chaotic regime. Secondly, $p$ also decreases as $\Delta t$ increases; this is because the values of FPE$_k(p)$ which indirectly depend on $\Delta t$, increase as functions of $p$. For $\Delta t = 16/64$ and $F = 32$, $p$ is chosen to be 1 for more than half of the Fourier modes since these modes decay quickly (see the corresponding slowest correlation time $T$ in Table 1).

In Fig. 3, we compare the energy spectrum and correlation time from the true with those reproduced by both the AR(p) and AR(1) models with $F = 6, 8, 32$. It can be seen that both the AR(p) and AR(1) models recover the true energy spectrum accurately. On the other hand, the correlation times are only accurately recovered by the AR(p) model and not by the
Figure 3: Upper-left: Energy spectrum for $F = 6$, 8, and 32 from the true L-96 model and the IL96 model, the AR(1) model, and the AR(p) model with $\Delta t = 4/64$, 8/64, and 16/64, respectively. Correlation times for $F = 6$ (upper-right), $F = 8$ (lower-left), and $F = 32$ (lower-right) from the L-96 model, the AR(1) model, and the AR(p) model.

AR(1) model. Additionally, the reproduced energy spectrum and correlation time from the AR(p) model show low sensitivity to the time step $\Delta t$. Even with $\Delta t$ large ($\Delta t = 16/64$), the correlation times reproduced by the AR(p) model are reasonably accurate (though a little worse than those from the AR(p) models with smaller $\Delta t$). Similar results are found for $F = 8, 32$ in Fig. 3 and for $F = 5, 16$ (results are not shown).

In Fig. 4, we show the autocorrelation functions with $F = 6$ for wavenumbers $k = 0, 8, 13,$ and 18. Here, we present only the AR(p) model from temporally coarse training data sets, $\Delta t = 16/64$ (similar results are found with the AR(p) model from temporally finer training data sets, $\Delta t = 4/64$ and 8/64). Notice that the autocorrelation functions reproduced by the AR(p) model are much better compared to those from AR(1) and MSM. Similar results are found for $F = 8$ (see Fig. 5) and for $F = 5, 16, 32$ (results are not shown).
Figure 4: Autocorrelation functions as a function of rescaled time lag $\tau$ with $F = 6$ for modes $k = 0, 8, 13, \text{ and } 18$, computed from the true L-96 model, the AR(1) model, the MSM, and the AR(p) model with $\Delta t = 16/64$.

3. Kalman Filtering with AR(p) Models

Consider full observation network with discrete observation time step $T_{obs} = t_{n+1} - t_n$, for $j = 1, \ldots, J$ and $n = 1, 2, \ldots$,

$$v_{j,n} = u_{j,n} + \epsilon_{j,n}, \quad j = 1, \ldots, J, \quad n = 1, 2, \ldots,$$

where $\epsilon_{j,n}$ are spatially and temporally i.i.d. Gaussian white noises with mean zero and variance $\sigma^2$. In this paper, we consider this idealistic observation network to emphasize the effect of model errors on the filtering skill. Approaches for sparse observations will be reported in the future utilizing different versions of Fourier diagonal filtering schemes as proposed in [48, 35, 50].

After the rescaling in (3), the Fourier coefficients of the rescaled observation model are given as follows,

$$\hat{v}_{k,n} = \hat{u}_{k,n} + \hat{\epsilon}_{k,n}, \quad |k| \leq J/2, \quad n = 1, 2, \ldots,$$
Figure 5: Autocorrelation functions as a function of rescaled time lag $\tau$ with $F = 8$ for modes $k = 0, 8, 13, \text{ and } 18$, computed from the true L-96 model, the AR(1) model, the MSM, and the AR($p$) model with $\Delta t = 16/64$.

where the rescaled observation time step is $\tilde{T}_{\text{obs}} = \sqrt{E_p} T_{\text{obs}}$; $\hat{v}_{k,n}$ are the Fourier coefficients of the rescaled observations, $\tilde{v}_{j,n} = (v_{j,n} - \bar{u})/\sqrt{E_p}$; $\{\hat{\epsilon}_{k,n}\}$ are complex-valued white noises with mean zero and variance $\hat{r}^2 = r^2/(J \cdot E_p)$.

We now describe the approximate filtering scheme with the AR($p$) model in (17) as the prior model on each Fourier coefficient. Assume for a moment that the observation time step $T_{\text{obs}}$ is the same as the time step in the off-line training phase; that is, $T_{\text{obs}} = \Delta t$. We will discuss the case of $T_{\text{obs}} = M \Delta t$, where $M > 1$ in the end of this section. First, let’s specify the forcing term $f_k$ in the AR($p$) model in (17) with an AR($p$) model for the mean-corrected term defined through (18),

$$X_{k,n} = F_k X_{k,n-1} + \eta_{k,n},$$

(25)

where $X_{k,n} \equiv (X_{k,n}, X_{k,n-1}, \ldots, X_{k,n-p+1})^T$ is a $p$-dimensional complex val-
ued vector for wavenumber $k$,

$$
F_k \equiv \begin{pmatrix}
\phi_1 & \phi_2 & \phi_3 & \cdots & \phi_p \\
0 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1
\end{pmatrix},
$$

and $\eta_{k,n} \equiv (\eta_{k,n}, \eta_{k,n-1}, \ldots, \eta_{k,n-p+1})^T \sim \mathcal{N}(\mathbf{0}, Q_k)$ with a $p \times p$ diagonal covariance matrix $Q_k$ with diagonal elements $(\sigma_k^2, 0, \ldots, 0)$. Substituting (18) to (25) for the unbiased terms, we have

$$
\hat{u}_{k,n} = F_k \hat{X}_{k,n-1} + \hat{u}_k + \eta_{k,n} = F_k (\hat{u}_{k,n-1} - \hat{\tilde{u}}_k) + \hat{u}_k + \eta_{k,n},
$$

(26)

where $\hat{u}_{k,n} \equiv (\hat{u}_{k,n}, \hat{u}_{k,n-1}, \ldots, \hat{u}_{k,n-p+1})^T$, $\tilde{u}_k = \tilde{u}_k \mathbf{1}$ with $\mathbf{1}$ denoting the $p$-dimensional vector of 1’s. Then we have an approximate filter with AR(p) as the forward model,

$$
\hat{u}_{k,n} = F_k \hat{X}_{k,n-1} + (I - F_k) \tilde{u}_k + \eta_{k,n},
\hat{v}_{k,n} = G \hat{u}_{k,n} + \tilde{\varepsilon}_{k,n},
$$

(27)

where $G \equiv (1, 0, \ldots, 0)$.

The classical Kalman filter consists of two steps: First, we propagate the best estimate of the true state at time $\hat{t}_n$, $\hat{u}_{k,n|n}$, and the associated error covariance $R_{k,n|n}$, with

$$
\hat{u}_{k,n+1|n} = F_k \hat{u}_{k,n|n} + (I - F_k) \tilde{u}_k,
R_{k,n+1|n} = F_k R_{k,n|n} F_k^* + Q_k,
$$

(28) \hspace{1cm} (29)

where $\hat{u}_{k,n+1|n}$ denotes the prior (forecast) mean state which is the best estimate before including observations at time $\hat{t}_{n+1} = \hat{t}_n + \hat{T}_{obs}$; the $p \times p$ matrix $R_{k,n+1|n}$ denotes the associated prior error covariance matrix. The posterior update includes observations $\hat{v}_{k,n+1}$ at time $\hat{t}_{n+1}$ through the following formulas,

$$
\hat{u}_{k,n+1|n+1} = G \hat{u}_{k,n+1|n+1}
\hat{u}_{k,n+1|n+1} = \hat{G}(\hat{u}_{k,n+1|n} + K_{k,n+1}(\hat{v}_{k,n+1} - \hat{G} \hat{u}_{k,n+1|n}))
R_{k,n+1|n+1} = (I - K_{k,n+1} G) R_{k,n+1|n},
$$

(30) \hspace{1cm} (31)
where
\[ K_{k,n+1} = R_{k,n+1|n} G^T (GR_{k,n+1|n} G^T + \hat{\sigma}_n)^{-1}, \] (32)
is the \( p \)-dimensional Kalman gain vector; \( R_{k,n+1|n+1} \) is the \( p \times p \) posterior error covariance matrix. The first equality in the posterior mean update in (30) is to ensure that we perform an “honest” filter, that is, we use observation \( v_{k,n+1} \) only to update the state \( \hat{u}_{k,n+1|n+1} \) and not the entire vector \( \hat{u}_{k,n+1|n+1} \) which components include posterior states at different lags; the second equality in (30) is the standard Kalman formula for mean posterior update. This completes one data-assimilation cycle in estimating the Fourier component \( \hat{u}_k \) at time \( \tilde{t}_{n+1} \).

We carry out the above filtering procedure for all \( J/2 + 1 \) independent Fourier modes. This is called as the Fourier diagonal Kalman filter (FDKF) based on the AR(p) model, and we call it “the AR(p) filter” in this paper. Due to the independence between the dynamics (i.e., the AR(p) model) of the Fourier modes, the computation of filtering all \( J/2 + 1 \) Fourier modes can be carried out in parallel. Notice that we only need to invert a \( D \)-dimensional matrix, where \( D \) is the number of observed variables, on each Fourier component to obtain the Kalman gain in (32). In contrast, the physical space ensemble filters, such as the ensemble adjusted Kalman filter (EAKF) [5] implemented in Section 4, require inverting a \( K \times K \) matrix (\( K \) denotes the size of the ensemble) to compute the Kalman gain in each data-assimilation cycle. In our numerical example here with L-96 model, the AR(p) filter inverts only scalars since \( D = 1 \). Hence, the AR(p) filter is computationally much cheaper than the standard physical space ensemble filters. For sparse observation networks, some Fourier components will be coupled with appropriate observation models in the posterior updating step [48, 35] and thus we have slightly higher dimensional filtering problems depending on the number of coupled modes.

For the AR(1) model, FDKF can be implemented similarly using the formulas in (28)-(32) with \( p = 1, F = \phi_1, \) and \( G = 1 \). We call it “the AR(1) filter” in this paper. For the MSM in (11), the scalar filter model (the Ornstein-Uhlenbeck process) is defined with analytical solution [24, 25, 35, 55]:
\[ \hat{u}_k(t) = e^{\tilde{A}_k t} \hat{u}_k(0) + \mu_k + \int_0^t \sigma_k e^{\tilde{A}_k (t-s)} dW_s, \] (33)
where \( \tilde{A}_k \equiv A_k - d_k + iw_k, \) and \( \mu_k \equiv E_p^{-1} (F - \bar{u}) \delta_k (e^{\tilde{A}_k 1} - 1)/\tilde{A}_k. \) In particular, the MSM filter will use the analytical mean and variance given
by

\[
\tilde{u}_k(\tilde{t}) = e^{\tilde{A}_k \tilde{t}} \tilde{u}_k(0) + \mu_k, \tag{34}
\]

\[
r_k(\tilde{t}) = \frac{\sigma_k^2}{2Re\{\tilde{A}_k\}} (e^{2Re\{\tilde{A}_k\} \tilde{t}} - 1). \tag{35}
\]

In discrete time formulation with observation time step \(T_{\text{obs}}\), the MSM approximate filter is

\[
\hat{u}_{k,n} = F_k \hat{u}_{k,n-1} + \mu_k + \eta_{k,n},
\]

\[
\hat{v}_{k,n} = G \hat{u}_{k,n} + \tilde{\epsilon}_{k,n}, \tag{36}
\]

where \(F_k = e^{\tilde{A}_k T_{\text{obs}}}\) and \(G = 1\); \(\eta_{k,n} \sim \mathcal{N}(0, r_k(T_{\text{obs}}))\). Here, the standard Kalman filter formula can be implemented for scalar problem in (36), and we call it “the MSM filter”.

One advantage of the MSM filter is that the filter model for MSM is independent of the time-step in the off-line training phase, since the filter model has explicit solution (33). For the linear stochastic models constructed in the discrete time formulation, such as the AR(1) and AR(p) models, the observation time interval \(T_{\text{obs}}\) and model timestep \(\Delta t\) are related, and their relationship will determine the form of the filter models. The approximate filter defined in (27) is based on the assumption of \(T_{\text{obs}} = \Delta t\). We will also consider the case of nonmatching \(T_{\text{obs}}\) and \(\Delta t\). We assume for simplicity that \(T_{\text{obs}}\) is an integer multiple of \(\Delta t\), say, \(T_{\text{obs}} = M \Delta t\) with integer \(M > 1\). In this simple setup, we simply propagate the posterior statistics \(M\) times to obtain the corresponding prior statistics at the future observation time. More complicated algorithms such as the Expectation-Maximization (EM) algorithms [58] can be applied to estimate the missing data within observation time interval, although the computation will be more complicated. In Section 4, we will show that this AR(p) filter with multiple times propagation performs reasonably well and provides skillful and robust filtered solutions.

Notice that with all three linear stochastic models (AR(p), AR(1), and MSM), we commit modeling errors in the filtering strategies since the true signal \(\hat{u}_{k,n}\) comes from (2), while the filter prior statistics are based on linear stochastic models. The work in [24, 25, 35] has shown that the MSM filter provides skillful filtering results in fully turbulent regime. In Section 4 we will show that the AR(p) filter proposed in this section performs even better than both the AR(1) filter and the MSM filter, especially in weakly chaotic regime.
4. Numerical Results

We compare the filter performance of the three linear stochastic models, the AR(p) model, the AR(1) model, and the MSM, on solutions of the L-96 model in various forcing regimes $F = 5, 6, 8, 16, 32$. The true trajectory of $\{u_j\}$ is generated by integrating L-96 model in (2) with the fourth order Runge-Kutta method (RK4) with time step $\delta t = 1/64$. Noisy observations are generated by randomly perturbing the true trajectory of L-96 model with Gaussian white noise. We choose observation noise variance $r_o$ based on the ratio between $r_o$ and the variance of $u_j$, that is, the inverse of the signal-to-noise ratio $\text{SNR} = 2E_p/(r_o J)$; specifically, we choose $\text{SNR}^{-1} = 0.1, 0.25, 0.5, 1.0$ to reflect small to large observation noises. We choose the observation time steps $T_{\text{obs}} = 4/64, 8/64, 16/64$, which are smaller compared to the slowest correlation time $T$ except when $F = 32$ (see Table 1). We consider mainly $T_{\text{obs}} = \Delta t$ in our experiment, but in the last part of this section, we also investigate the case of $T_{\text{obs}} = M\Delta t$ ($M > 1$), as described in the end of Section 3.

In the off-line training phase, we generate the discrete time series data by integrating L-96 model for total time $10^5$ with time step $\delta t$ and storing $\{u_j\}$ with time step $\Delta t$. In real application, this time series is a given data set and we have no knowledge about any truth model. The discrete time series of the Fourier modes are then obtained by the rescaling in (3) followed by the discrete Fourier transformation in (7). The detail parameterization algorithms for the AR(p), AR(1), and MSM are explained in Appendix A. Besides the filtering strategies based on the three linear stochastic models, for benchmark purposes, we also include the ensemble adjusted Kalman filter (EAKF) [5] with the original nonlinear L-96 model in (2) in our numerical simulations. Specifically, the EAKF is implemented with an ensemble size $K = 40$ and variance inflation coefficient 0.05.

To measure the filter performance, we compute the time series of the root-mean-squared (RMS) errors, $\text{RMSE}(t_n)$, defined as:

$$\text{RMSE}(t_n)^A \equiv \sqrt{\frac{1}{J} \sum_{j=1}^{J} (u_n^A - u_n)^2},$$

(37)

where $u_n = (u_{1,n}, \ldots, u_{J,n})^T$ denotes the truth and $u_n^A$ denotes the filtered (posterior mean) solution at time $t_n$ from method $A$. In Fig. 6, we plot $\text{RMSE}(t)$ for simulations with $F = 5, 6$ and $16$, $T_{\text{obs}} = 4/64$, $\text{SNR}^{-1} = 0.25$. 

20
In Figs. 7 and 8, we show spatial patterns of the filtered solutions at a fixed time for \( F = 6 \) and \( F = 16 \), respectively. For \( F = 5 \), the filtered solutions from the MSM filter diverge with errors larger than observation noise error (the RMS errors are beyond the scale displayed in the upper panel of Fig. 6). The AR(p) filter always produces more accurate filtered solutions than the AR(1) filter, and the MSM filter, especially when the system is weakly chaotic (see the upper and middle panels in Fig. 6). These three filters are all better than simply trusting the observations. The performance of the AR(p), AR(1), and MSM filters become similar when \( F \) is large \( (F \geq 16) \), and thus in the lower panel of Fig. 6 we only plotted the time series of the RMS errors of the AR(p) filter. Notice that EAKF with perfect model is not robust; in the regimes with small \( F \), EAKF produces the best filtered solutions, as can be seen from the upper and middle panels of Figs. 6 and 7. However, when the system is fully turbulent with \( F = 16 \), the filtered solutions from EAKF diverge with errors larger than simply trusting the observations (see the lower panel of Figs. 6 and 8).

We also quantify the filter performance with the time average of the rescaled RMS errors, RRMSE, and the time average of the spatial correlations, PC, defined as follows:

\[
RRMSE^A \equiv \frac{\langle \text{RMSE}^A(t_n) \rangle}{\sqrt{2E_p/J}}, \tag{38}
\]

\[
PC^A \equiv \frac{\langle (u_n^A - \bar{u}_n^A)^T (u_m - \bar{u}_m) \rangle}{\|u_m - \bar{u}_m\|_2 \|u_m - \bar{u}_m\|_2}, \tag{39}
\]

where the angle bracket \( \langle \cdot \rangle \) denotes the temporal average over all the assimilation cycles \( \{t_m : m = 1, \ldots, 40000\} \); \( \bar{u} \) denotes the spatial average of the \( J \)-dimensional vector \( \mathbf{u} \); and \( \| \mathbf{a} \|_2 = \mathbf{a}^T \mathbf{a} \) denotes the standard Euclidean 2-norm. Note that RRMSE for simply trusting the observations is equal to \( \sqrt{r_o J/(2E_p)} \), that is, \( \text{SNR}^{-1/2} \). We show the PC score in (39) to measure the relevance of the filtered solutions in physical space which is not necessarily obvious especially when the filters are performed in Fourier domain with diagonal approximation.

In Figs. 9 and 10, we show the RRMSE\(^A\) and PC\(^A\) for all the filtering strategies with various values of \( T_{obs} \) and \( \text{SNR}^{-1} \). For \( F = 5 \), the MSM filter performs extremely worse (as we discussed before) compared to the other filtering strategies or even simply trusting observations; the filtered solutions have nearly zero spatial correlations (see the upper panels in Fig. 10) and
much larger rescaled RMS errors (not shown in the upper panels in Fig. 9 since they are beyond the plotting scale). The AR(p) filter performs the best among the three filtering strategies based on linear stochastic models; the AR(p) filter has the smallest rescaled RMS errors with spatial correlation closer to one. More substantial improvement with the AR(p) filter over the AR(1) and MSM can be seen in weaker turbulent regime (or smaller $F$). When $F$ is large, the AR(p) filter performs slightly better than the AR(1) and the MSM filters when $T_{\text{obs}}$ is small, and they are all comparable when $T_{\text{obs}}$ is large (see Figs. 9 and 10). Between the AR(1) and the MSM filters, the former produces more skillful filtered solutions when $F$ is small and $T_{\text{obs}}$ is large.

Although EAKF generally produces better filtered solutions in weaker turbulent regime ($F = 5, 6$), it is not robust. For $F = 16$ and $F = 32$, the filtering skill of EAKF is worse than those of the other three filtering strategies (see the last two rows in Figs. 9 and 10). Additionally, EAKF produces large errors when observation noise variance is small; e.g., see the regimes $F = 6, 8$ with $T_{\text{obs}} = 16/64$ and $\text{SNR}^{-1} = 0.1$. These large errors are attributed to the failure in inverting the ill-conditioned ensemble based covariance matrix since the ensemble members “collapse”; that is, each ensemble member looks almost like the observations. The poor performance with EAKF in these tough regimes can be fixed with appropriate choices of ensemble size, variance inflation, and also by performing sequential analysis [6]; but our main point here is that the ensemble filter is sensitive in various regimes and so we do not tune EAKF. Moreover, EAKF requires more expensive computations compared to the filtering strategies based on the linear stochastic models where Fourier modes are decoupled; in each assimilation cycle, EAKF involves an inversion of a $K \times K$ matrix to obtain the Kalman gain matrix and integrates each ensemble solution with the perfect L-96 model. In our numerical implementation, the EAKF with the perfect L-96 model costs about 50 times more computational time (CPU wall-clock) compared to the AR(p), AR(1) and MSM filter for $T_{\text{obs}} = 8/64$, and about 80 and 150 times more computational times for $T_{\text{obs}} = 8/64$ and $T_{\text{obs}} = 16/64$, respectively.

Recall that since the AR(p) model is built with the discrete time formulation, the time step $\Delta t$ in its off-line training phase is related to its filter model. In the next numerical experiment we investigate the performance of the AR(p) filter for $T_{\text{obs}} = M\Delta t$ with integer $M > 1$. Specifically, we choose $\Delta t = 4/64$ in the off-line training phase and consider $T_{\text{obs}} = 8/64$ and $16/64$ in data assimilation phase, that is, $M$ takes the values 2 and 4. As described
in Section 3, the prior statistics are computed by propagating the mean and covariance in (30), (31) \( M \) times; in this paper, we call this scheme “the AR(p)-M filter”. In Fig. 11, we compare the rescaled RMS errors from the AR(p)-M filter with the errors from all the other filters discussed earlier for \( F = 6, 8 \). Although the filtered solutions from the AR(p)-M filter deteriorate compared to the original AR(p) filter with \( M = 1 \), the AR(p)-M filter performs reasonably well and produces better filtered solutions compared to the AR(1) and MSM filters and is not sensitive to small observation noise variance. Notice also that the rescaled RMS errors of the AR(p)-M increase as \( M \) increases (since the observation time interval is larger). This result suggests that the AR(p) model obtained from a temporally finer training data set may not produce the best filtered solutions when observations are not frequently available, \( T_{\text{obs}} > \Delta t \). This slight deterioration is perhaps due to additional errors in the prior statistics when the AR(p) model is run iteratively \( M \) times and we will study this issue more carefully in future report.

Finally, we investigate the impact of additional modeling error due to slightly incorrect specification of the parameter \( F \) in the off-line training phase; in practice, we may be given a training data set without knowing the true forcing parameter. In our numerical experiments, for data-assimilation in forcing regimes \( F = 6, 16 \), we carry out the AR(p) filter with parameters obtained from the off-line training data in regime with larger forcing, \( F = 8 \). For benchmark purposes, we also perform EAKF with incorrectly specified L-96 model in (2) with \( F = 8 \) for data assimilation in forcing regimes \( F = 6 \) and \( F = 16 \), respectively. In Fig. 12, we can see that this additional model error degrades the AR(p) filter by slightly. On the other hand, the EAKF errors increase significantly in the presence of additional model errors. This result suggests that reduced stochastic filter with the AR(p) model is relatively robust which is very important since model errors are unavoidable in many real problems.

5. Summary and discussion

In this paper, we introduce the AR(p) filter for assimilating weakly chaotic dynamical systems with longer memory depth. The AR(p) filter generalizes the MSM filter [24, 35], which was found to be advantageous in fully turbulent regime. The AR(p) model significantly improves the autocorrelation statistical estimates over the MSM through its non-Markovian nature. Using AR(p) models of order \( p \) as surrogate filter prior models, we reduce the
problem of filtering a nonlinear spatial extended system to filtering decoupled autoregressive linear stochastic models in Fourier space. The filtering on each Fourier mode is then carried with the linear Kalman based filter (FDKF) as in [59, 24, 34]. Computationally, the AR(p) filter is much cheaper than the ensemble filtering strategies such as EAKF in the physical space, since the Fourier modes are decoupled, and no ensemble is needed with linear stochastic filters. As we pointed out earlier, such approach can be implemented to any “spectral-friendly” problems; in contrast, difficulties can be arise for e.g. in ocean modeling with irregular coastal boundary in which Fourier approach is not feasible.

We compare the climatological statistical and real-time filtered solutions of the proposed AR(p) model to the AR(1) model and the MSM. In terms of their climatological statistical solutions, we show that although the MSM can reproduce the energy spectrum and the correlation time perfectly, it does not reproduce the autocorrelation functions as accurate as the AR(p) model. In terms of the filtered solutions, the AR(p) filter produces more accurate estimates compared to the AR(1) and MSM filters, especially in weakly chaotic regime where the system has long memory depth and invariant measure that is not very different (slightly skewed) from Gaussian distribution (see Fig 11.2 in [39]). In fully turbulent regime where the system has shorter memory depth with also slightly skewed from Gaussian invariant measure, the AR(p) filter solutions are comparable to MSM.

We also found that the AR(p) filter built from the temporally finer training data set does not produce the best solution when observation time interval is larger than the data set time step. This scenario is important in practice because we often collect observations at much coarser time steps beyond the model integration time step. Currently, we are in the progress of finding necessary conditions for optimal autoregressive filtering for this case, $T_{\text{obs}} \gg \Delta t$. In the future, we will report an improved approximate AR(p) filter that respects the relevant necessary conditions. Fundamentally, the MSM parameters in (15) have physical direct interpretation whereas those in the AR(p) model do not, and therefore the autoregressive filter can be interpreted as a “fully data-driven” (or “purely statistical”) approach since it only requires knowledge of training data set and some observations. However, we can also view the deterministic terms of the autoregressive models in (17) as a multistep discrete approximation for solving the deterministic part of the Langevin equations in (11) except now that the coefficients are not the outcomes of any particular discretization scheme. In this sense, the ad-
vantageous of the physically more meaningful Ornstein-Uhlenbeck processes retain with the “purely statistical” autoregressive models in the fully turbulent regime.

The major drawback of the AR(p) modeling here is that the parameterization strategy via Yule-Walker estimators may require longer timeseries that may not be available in various problems. Different parameterization strategies than the Yule-Walker estimators may be useful in such cases. Additionally, from numerical experiments in Section 4, we find that the AR(p) filter with wrong training data set is not as sensitive as the EAKF in the presence of physical model errors due to incorrectly specified forcing parameter, $F$. Such robustness with the AR(p) filter is very important since in many real problems the true model may not be available or as we mentioned above the available time series can be too short. Secondly, as in [24, 25], we also find that small observation error variances and large forcing can produce divergent filtered solutions with EAKF if the filter parameters are not appropriately tuned. This is in contrast to our reduced filtering strategy which by-design does not need to tune more parameters once the AR(p) parameters are learned in the offline phase. This result is consistent with the stability of the MSM filter as explained in [38].

With such encouraging results, we will consider applications on more realistic problems with sparse observations in future research. Particularly, we will consider the sparse regularly spaced observations as in [48, 25] and irregularly spaced observations as in [50] for ocean turbulent dynamics [35].

Appendix A. Algorithm of the Off-line Training Phase

In this appendix, we describe the step-by-step algorithm for the off-line training phase in the context of L-96 model.

1. Training data set
   In real problems, a training data set may be available; for e.g., the reanalysis data in atmospheric and ocean sciences are produced by the National Centers for Environmental Prediction in the US and the European Centre for Medium-Range Weather Forecasts. In our experimental setup, we generate data for $\{u_j\}$ by integrating L-96 model for total time $T = 10^5$ with time step $\delta t$, and the resulting data are stored with time step $\Delta t$. After the rescaling in (3), we obtain the discrete time series of the Fourier modes $\{\hat{u}_{k,n} : n = 1, 2, \ldots, N\}$ where $N = T/\Delta t$. 
2. Parameterization of the AR(p) model

For a given mode \( k \), we subtract the temporal mean from the training data set \( \{ \hat{u}_{k,n} \} \) to obtain the “mean-corrected” time series \( \{ X_{k,n} : n = 1, 2, \ldots, N \} \) as defined in (18). Then we fit an autoregressive model of order \( p \) for \( \{ X_{k,n} \} \), which consists of the following two steps: Choosing the order \( p \) and estimating the parameters \( \{ \phi_1, \ldots, \phi_p, \sigma_k^2 \} \).

(a) Choosing the order \( p \)

i. We choose an upper limit of the order \( P = 30 \) and a fixed length of time \( T^* = 200 \ll T = 10^5 \). Then we divide the time series \( \{ X_{k,n} \} \) of total length \( N \) into segments of length \( N^* = T^*/\Delta t \), which results in a total number of \( T/T^* = 50 \) subsamples.

ii. For the \( i \)-th subsample, we fit the the AR(p) model of order \( p \) for \( p = 0, \ldots, P \), respectively. Specifically, for given \( p \), we compute the Yule-Walker estimates in (20) and (21) for each subsample timeseries of length \( T^* \) and then calculate the FPE criterion in (22), denoted as \( \text{FPE}_k^i(p) \). Then we define \( p_o^i = \min \{ \text{FPE}_k^i(p) : p = 0, \ldots, P \} \); that is, \( p_o^i \) is the chosen order for the \( i \)-th sample based on the FPE criterion.

iii. We repeat the above step for all the 50 subsamples and obtain \( \{ p_o^i : i = 1, \ldots, 50 \} \), and \( p_o \equiv \text{mode} \{ p_o^i : i = 1, \ldots, 50 \} \) is chosen as the order of the autoregressive model.

(b) Parameters estimation:

With the chosen \( p = p_o \), we fit an autoregressive model of order \( p \) for the entire time series \( \{ X_{k,n} : n = 1, \ldots, N \} \); that is, we apply the Yule-Walker estimators to parameterize the AR(p) model.

The above steps are carried out on each the Fourier mode independently and this completes the off-line training phase for the AR(p) filter.

In the training phase of the AR(1) filter, we skip the steps for choosing \( p \) and simply use the entire time series to estimate the parameters in the AR(1) model. For the MSM, the parameterization is carried through the explicit formulas in (15).

Acknowledgments

We thank Ilya Timofeyev for pointing out the autoregressive model in the early stage of this research. The research of J. Harlim is partially supported
by the NC State startup fund and by the Office of Naval Research Grant N00014-11-1-0310. During the course of this work, E. L. Kang is supported as a Postdoctoral Fellow at Statistical and Applied Mathematical Sciences Institute (SAMSI) and through Harlim’s NC State startup fund.

References


Figure 6: RMS errors as functions of time for $F = 5$, $6$, and $16$. Dashes for observation errors; green solid lines with square for the AR(1) filter; thick solid lines for the AR(p) filter; grey solid lines for EAKF; solid lines with circles for the MSM filter. For $F = 5$, the errors for MSM are way beyond the presented scale. For $F = 16$, all three AR(1), AR(p), and MSM produce similar RMS errors, so we only plot the errors of AR(p) filter.
Figure 7: Spatial patterns of the filtered solutions (solid line) from the AR(p) filter, the AR(1) filter, EAKF, and the MSM filter for $F = 6$, $\text{SNR}^{-1} = 0.25$, and $\Delta t = 4/64$, compared to those from the true L-96 model (dashes) and observations (circles).
Figure 8: Spatial patterns of the filtered solutions (solid line) from the AR(p) filter, the AR(1) filter, EAKF, and the MSM filter for $F = 16$, $\text{SNR}^{-1} = 0.25$, and $\Delta t = 4/64$, compared to those from the true L-96 model (dashes) and observations (circles).
Figure 9: Rescaled average RMS errors as functions of SNR$^{-1}$, for $F = 5, 6, 8, 16, \text{and } 32$. In each panel, we plot the observation errors (dashes), MSM filter (dashes with squares), AR(1) filter (dotted dashes with ‘*’), AR(p) filter (solid lines with ‘o’), and EAKF (solid).
Figure 10: Spatial correlations as functions of SNR$^{-1}$, for $F = 5, 6, 8, 16,$ and 32. In each panel, we plot the observation errors (dashes), MSM filter (dashes with squares), AR(1) filter (dotted dashes with ‘*’), AR(p) filter (solid lines with ‘o’), and EAKF (solid).
Figure 11: Rescaled RMS errors as functions of SNR$^{-1}$ for $F = 6$ (first row) and $F = 8$ (second row). Dashes for observation errors; dashes with square for the the MSM filter; dotted dashes with asterisk for the AR(1) filter; solid lines with circle for the AR(p) filter with $\Delta t = T_{obs}$; blue solid lines with dots for AR(p)-M with $\Delta t = 4/64$; solid lines for EAKF.
Figure 12: Rescaled RMS errors as functions of SNR$^{-1}$ for $F = 6$ (first row) and $F = 16$ (second row). Dashes for observation errors; solid lines with circle for the AR(p); red dashes with circle for AR(p) with model errors (parameters are estimated from a data set with $F = 8$); solid lines for EAKF with perfect model; red dotted dashes for EAKF with model errors ($F = 8$ in the filter model).