Physica D 298–299 (2015) 21–41

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems

Di Qi *, Andrew J. Majda

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

HIGHLIGHTS

- We develop blended filtering methods that exploit the structure of dynamical systems.
- Non-Gaussian features are captured adaptively in a subspace through particle methods.
- The remaining parts of the phase space are amended by conditional Gaussian mixtures.
- The performance of the blended algorithms is compared in various dynamical regimes.

ARTICLE INFO

Article history:
Received 1 November 2014
Received in revised form
1 February 2015
Accepted 4 February 2015
Available online 11 February 2015
Communicated by T. Wanner

Keywords:
Data assimilation
Hybrid filters
Curse of dimensionality
Turbulent dynamical systems

ABSTRACT

It is a major challenge throughout science and engineering to improve uncertain model predictions by utilizing noisy data sets from nature. Hybrid methods combining the advantages of traditional particle filters and the Kalman filter offer a promising direction for filtering or data assimilation in high dimensional turbulent dynamical systems. In this paper, blended particle filtering methods that exploit the physical structure of turbulent dynamical systems are developed. Non-Gaussian features of the dynamical system are captured adaptively in an evolving-in-time low dimensional subspace through particle methods, while at the same time statistics in the remaining portion of the phase space are amended by conditional Gaussian mixtures interacting with the particles. The importance of both using the adaptively evolving subspace and introducing conditional Gaussian statistics in the orthogonal part is illustrated here by simple examples. For practical implementation of the algorithms, finding the most probable distributions that characterize the statistics in the phase space as well as effective resampling strategies is discussed to handle realizability and stability issues. To test the performance of the blended algorithms, the forty dimensional Lorenz 96 system is utilized with a five dimensional subspace to run particles. The filters are tested extensively in various turbulent regimes with distinct statistics and with changing observation time frequency and both dense and sparse spatial observations. In real applications perfect dynamical models are always inaccessible considering the complexities in both modeling and computation of high dimensional turbulent system. The effects of model errors from imperfect modeling of the systems are also checked for these methods. The blended methods show uniformly high skill in both capturing non-Gaussian statistics and achieving accurate filtering results in various dynamical regimes with and without model errors.

1. Introduction

Accurate predictions of the future states of high-dimensional turbulent dynamical systems are a formidable problem with significant practical impact in a wide range of areas throughout science and engineering. Filtering or data assimilation refers to the process of obtaining the best estimation of a natural system by combining uncertain model predictions with noisy observations of the true signal from nature. Examples for important contemporary applications involve the real time filtering and prediction of weather and climate systems as well as the spread of hazardous plumes and pollutants or the prediction of storm surges in environmental science and engineering. These turbulent dynamical systems always include a large number of active degrees of freedom under various kinds of nonlinear non-Gaussian

* Corresponding author.
E-mail address: qidi@cims.nyu.edu (D. Qi).

http://dx.doi.org/10.1016/j.physd.2015.02.002
0167-2789/

scenarios even with an irreducible high dimensional attractor in phase space. Considering the complexity of these systems, model errors in the approximation models are always unavoidable due to both the imperfect understanding of the real nature systems and the limited computational ability at hand. Thus an important emerging scientific issue is the development of statistically accurate filtering methods through models with high skill in capturing both Gaussian and non-Gaussian features as well as being robust to external perturbations and imperfect prior estimates from model errors for filtering turbulent dynamical systems.

The complexity of the turbulent dynamical systems makes it impossible to evaluate the distributions of state variables analytically in explicit form for most situations. To approximate the filter distributions computationally instead, particle filters based on Monte-Carlo simulations are a popular class of numerical methods in characterizing the nonlinear non-Gaussian structures of low-dimensional systems [1,2]. Compared with the standard Kalman filtering methods [3], the principal advantage of the particle methods is that they do not rely on any local linearization of the models and in principle can get the entire statistical information about any higher order moments with large enough ensemble size. However the price that must be paid for these advantages of higher accuracy is the high computational expense required to evolve and update the entire ensemble of particles. Especially, as the dimensionality of the system increases, the required ensemble size that is sufficient for describing the full statistics in the phase space increases exponentially and becomes computationally impossible even with moderate dimensionality about order of 100. Insufficient ensemble size will end up with frequent particle collapse even with proper resampling strategies [4,5].

To avoid this curse of dimensionality for particle methods, a number of ensemble based Kalman filters [6–9] are developed, which use an ensemble of particles to estimate the mean and covariance only and then apply Kalman filter instead in the analysis step. These ensemble Kalman filtering methods show promising results for some high dimensional nonlinear systems, for example in synoptic scale midlatitude weather dynamics, with careful choices of inflation and localization parameters for the particles. But still with only Gaussian statistics of the particles exploited in the analysis step, these methods are implicitly Gaussian and are sensitive to model resolution and different kinds of observations, and model parameters need to be changed according to different dynamical regimes. For another direction, reduced order filtering strategies [10–13] and Bayesian hierarchical modeling methods [14] have been developed trying to apply particles in a sufficiently low dimensional subspace. One idea might be that only considering the subspace which contains most of the energy in the system is sufficient to obtain desirable filtering performance. One strategy related to this idea is to filter the solution in an evolving low-dimensional subspace which captures the leading variances adaptively while ignoring the remaining degrees of freedom [15]. Whereas it turned out that the ignored degrees of freedom could be crucial in the model prediction skills. Simple examples with non-normal linear systems [16] demonstrate the limited skill of such an approach for reduced filtering methods in general as shown here in Section 2.2.1. Despite the difficulties in both the Kalman filter based Gaussian methods and the reduced order particle strategies, these two approaches with distinct features both have some advantages in achieving promising filtering performance and are often complementary with each other. Ideas about hybrid methods [17,18] have been developed and show promising results.

This paper builds on the mathematical framework described in [19] to develop new blended particle filtering methods for chaotic dynamical systems. The idea is a hybrid method combining the merits of the adaptive reduced order method as well as conditional Gaussian mixtures in which Kalman filter update formulas can be applied in the analysis step. The large dimensional phase space is decomposed into two subspaces. The particle filter is applied in one subspace low dimensional enough for accuracy and efficiency and evolving adaptively in time for capturing the directions with principal variances. On the other hand, the information in the high dimensional orthogonal subspace is corrected by a conditional Gaussian mixture representation where Kalman filter updates can be used with efficiency. The adaptive space decomposition required in this framework for the forecast model is achieved here with the help of recently developed statistically accurate models [16,20,21] in this paper, while the general framework is more flexible and need not be limited to these forecast models only. See [22,23] for another application of the blended filtering ideas using multi-scale forecast models with superparameterization. Two essential aspects concerning different statistics in the two separate subspaces are highlighted here for the effectiveness of this method. The low dimensional subspace with particle filters used to resolve full statistics needs to evolve adaptively in time in order to keep tracking the most energetic directions of the system with largest variances; while the statistics in the orthogonal subspace with conditional Gaussian mixtures applied are nevertheless important and should never be neglected. The importance of these two aspects can be illustrated by simple test models. The choice for the Gaussian distributions in the orthogonal subspace is guided by information theory and can be approximated by solving a linear system conditional on particle values in the other subspace. The efficiency of the method is guaranteed by a universal form of the prior conditional covariance matrix to which the Kalman filter update is applied, requiring only one time calculation of the large scale Kalman gain matrix. In addition, to avoid particle degeneration while maintaining the minimum amount of noise added in the resampling step, an adaptive inflation is introduced to the resampled particles which has potential to further stabilize the scheme. In [19], we have checked the blended filters in several difficult regimes using the Lorenz-96 system with no model error included. Here, a more extensive discussion about the performances of these methods is carried out in both Gaussian and non-Gaussian regimes of the same system with various kinds of observation networks. Furthermore, the effects of model error are checked by introducing imperfect external forcing term in this system. It could make the problem much more challenging due to the distinct statistical dynamics between the perfect model and the imperfect model with error. The ability of these methods to capture significant non-Gaussian features with imperfect model is demonstrated below in Section 4.

In the following part of this paper, in Section 2, the general ideas and basic framework for the blended particle filtering algorithms will be described. We illustrate the importance of the two indispensable parts of this method with several simple examples in Section 2. Section 3 describes the detailed strategies for finding the conditional Gaussian distribution and the resampling tricks, which are essential to the realizability and stability of these hybrid filtering schemes. Various tests of these methods using the Lorenz-96 system are reported in Section 4 for models with or without model errors including the capability of the methods to capture non-Gaussian features. We finish in Section 5 with a brief summary and a discussion about the future directions.

2. Algorithms for blended particle filtering

The blended particle filtering algorithms exploit the physical structure of turbulent dynamical systems and capture non-Gaussian features in an adaptively evolving low dimensional subspace through particles interacting with evolving Gaussian statistics on the remaining portion of phase space. This framework is set to be flexible so that any proper forecast models with