Ensemble of Kalman Filter

By
UWAMARIYA Denise

January 2012

Outline

Abstract

The ensemble Kalman Filter (EnKF) is a recursive filter suitable for problems with a large number of variables, such as discretizations of partial differential equations in geophysical models. The EnKF originated as a version of the Kalman filter for large problems (essentially, the covariance matrix is replaced by the sample covariance), and it is now an important data assimilation component of ensemble forecasting and the EnKF makes the assumption that all probability distributions involved are Gaussian. This report briefy described a derivation of EnKF.

INTRODUCTION

The Ensemble Kalman Filter (EnKF) is a Monte-Carlo implementation of the Bayesian update problem: Given a probability density function (pdf) of the state of the modeled system (the prior, called often the forecast in geosciences) and the data likelihood, the Bayes theorem is used to obtain pdf after the data likelihood has been taken into account (the posterior, often called the analysis).

cont'd

This is called a Bayesian update. The Bayesian update is combined with advancing the model in time, incorporating new data from time to time. The original Kalman Filter assumes that all pdf are Gaussian (the Gaussian assumption) and provides algebraic formulas for the change of the mean and covariance by the Bayesian update, as well as a formula for advancing the covariance matrix in time provided the system is linear. However, maintaining the covariance matrix is not feasible computationally for high-dimensional systems. For this reason, EnKFs were developed.

cont'd

EnKFs represent the distribution of the system state using a random sample, called an ensemble, and replace the covariance matrix by the sample covariance computed from the ensemble. One advantage of EnKFs is that advancing the pdf in time is achieved by simply advancing each member of the ensemble. For a survey of EnKF and related data assimilation techniques.

A derivation of Ensemble Kalman Filter

Let us review first the Kalman filter: Let x denote the n-dimensional state vector of a model, and assume that it has Gaussian probability distribution with mean μ and covariance Q i.e., its pdf is

$$p(x)\alpha \exp(-1/2(x-\mu)^T Q^{-1}(x-\mu)) \tag{1}$$

where α means proportional This probability distribution, called the prior distribution. was evolved in time by running the model and now is to be updated to account for new data. It is natural to assume that the error distribution of the data is known; data have to come with an error estimate, otherwise they are meaningless.

cont'd

Here, the data d is assumed to have Gaussian pdf with covariance R and mean H_x, where H is the so-called the observation matrix. The covariance matrix R describes the estimate of the error of the data; if the random errors in the entries of the data vector d are independent, R is diagonal and its diagonal entries are the squares of the standard deviation (error size) of the error of the corresponding entries of the data vector d. The value H_x is what the value of the data would be for the state x in the absence of data errors.

cont'd

Then the probability density $p(d|x)$ of the the data d conditional of the system state x, called the data likelihood, is

$$p(d|x)\alpha \exp(-1/2(d - H_x)^T R^{-1}(d - H_x)) \quad (2)$$

The pdf of the state and the data likelihood are combined to give the new probability density of the system state x conditional on the value of the data d the posterior) by the Bayes theorem

$$p(x|d)\alpha p(y|d)p(x) \quad (3)$$

cont'd

The data d is fixed once it is received, so denote the posterior state by $\hat{x}$ instead of $x|d$ and the posterior pdf by $p(\hat{x})$. It can be shown by algebraic manipulations that the posterior pdf is also Gaussian,

$$p(\hat{x})\alpha \exp(-1/2(\hat{x} - \hat{\mu})^T p^{-1}(\hat{x} - \hat{\mu})) \tag{4}$$

with the posterior mean $\hat{\mu}$ and covariance $\hat{Q}$ given by the Kalman update formulas

$$\hat{\mu} = \mu + K(d - H_\mu), \hat{Q} = (I - KH)Q, \tag{5}$$

where

$$K = QH^T(HQH^T + R)^{-1} \tag{6}$$

is the so-called kalman gain matrix.

cont'd

The Ensemble Kalman Filter: The EnKF is a Monte Carlo approximation of the Kalman filter, which avoids evolving the covariance matrix of the pdf of the state vector x. Instead, the distribution is represented by a sample,called an ensemble. So, let

$$x = [x_1, ..., x_n] = [x_i] \tag{7}$$

be an $n \times N$ matrix whose columns are a sample from the prior distribution.the matrix x is called the prior ensemble.Replicate the data d into an $m \times N$ matrix

$$D = [d_1, ..., d_N] = [d_i] \tag{8}$$

so that each column d_i consists of the data vector d plus a random vector from the n-dimensional normal distribution $N(0, R)$.

cont'd

Then the columns of

$$\hat{x} = x + K(D - Hx) \tag{9}$$

form a random sample from the posterior distribution. The EnKF is now obtained simply by replacing the state covariance Q in Kalman gain matrix by the sample covariance C computed from the ensemble members (called the ensemble covariance).

Thank you

MURAKOZE