A General Ensemble Filtering Framework Using Quantiles

Jeff Anderson, NCAR Data Assimilation Research Section
PRE-RELEASE, Please do NOT Share without permission
Outline

• Review of Ensemble Kalman Filters

• General Ensemble Filtering Framework using Quantiles
  Can fit arbitrary continuous prior to ensemble
  Product with a likelihood gives continuous posterior
  Posterior ensemble with same quantiles as prior

• Useful families for continuous priors and likelihoods

• Idealized examples for priors useful for earth science
  Normal: identical to existing ensemble Kalman filters
  Rank histogram
  Sum of normal kernels
  Gamma / inverse gamma
  Beta

• Extension to multivariate application
Building a Forecast System with Data Assimilation

Prediction Model

Observing System

Forecasts

Initial Conditions

Data Assimilation

Analysis

Observations
1. Use model to advance ensemble (3 members here) to time at which next observation becomes available.
2. Get prior ensemble sample of observation, $y = h(x)$, by applying forward operator $h$ to each ensemble member.

Theory: observations from instruments with uncorrelated errors can be done sequentially.

Can think about single observation without (too much) loss of generality.
3. Get **observed value** and **observational error distribution** from observing system.
4. Find the increments for the prior observation ensemble (this is a scalar problem for uncorrelated observation errors).
5. Use ensemble samples of $y$ and each state variable to linearly regress observation increments onto state variable increments.
6. When all ensemble members for each state variable are updated, there is a new analysis. Integrate to time of next observation ...
First part focuses on the scalar problem for an observed variable, $y$. All other model state variables updated by (linear or rank) regression.
Given a prior ensemble estimate of an observed quantity, $y$
Generalized Ensemble Filter Framework using Quantiles

Fit a continuous PDF from an appropriate distribution family and find the corresponding CDF

This example uses a normal PDF
Generalized Ensemble Filter Framework using Quantiles

Compute the quantile of ensemble members; just the value of CDF evaluated for each member.

This example uses a normal PDF
Continuous likelihood for this observation.

This example uses a normal PDF.
Bayes tells us that the continuous posterior PDF is the product of the continuous likelihood and prior.

Normal times normal is normal.
Posterior ensemble members have same quantiles as prior. This is quantile function, inverse of posterior CDF.

This example uses a normal PDF.
For normal prior and likelihood, this is identical to existing deterministic Ensemble Adjustment Kalman Filter (EAKF)
Useful families for continuous priors and likelihoods

Different families of distributions for continuous priors and likelihoods can lead to analytic continuous posterior.

This is similar to the notion of conjugate priors for estimating parameters of distributions.

A list of prior / likelihood pairs that may be useful for scientific application follows.
## Useful families for continuous priors and likelihoods

<table>
<thead>
<tr>
<th>Prior</th>
<th>Likelihood</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>Normal</td>
<td>Normal</td>
</tr>
<tr>
<td>Lognormal</td>
<td>Lognormal</td>
<td>Lognormal</td>
</tr>
<tr>
<td>Gamma</td>
<td>Gamma</td>
<td>Gamma</td>
</tr>
<tr>
<td>Inverse Gamma</td>
<td>Inverse Gamma</td>
<td>Inverse Gamma</td>
</tr>
<tr>
<td>Beta</td>
<td>Beta</td>
<td>Beta</td>
</tr>
<tr>
<td>Beta prime</td>
<td>Beta prime</td>
<td>Beta prime</td>
</tr>
<tr>
<td>Exponential</td>
<td>Exponential</td>
<td>Exponential</td>
</tr>
<tr>
<td>Pareto</td>
<td>Pareto</td>
<td>Pareto</td>
</tr>
<tr>
<td>Genl. Gamma given p</td>
<td>Genl. Gamma given p</td>
<td>Genl. Gamma given p</td>
</tr>
<tr>
<td>Any</td>
<td>Uniform</td>
<td>Any</td>
</tr>
<tr>
<td>Gamma</td>
<td>Poisson</td>
<td>Gamma</td>
</tr>
</tbody>
</table>
## Useful families for continuous priors and likelihoods (2)

<table>
<thead>
<tr>
<th>Prior</th>
<th>Likelihood</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delta function</td>
<td>Any</td>
<td>Delta function</td>
</tr>
<tr>
<td>Skew normal</td>
<td>Normal</td>
<td>Skew normal</td>
</tr>
<tr>
<td>Truncated normal</td>
<td>Normal</td>
<td>Truncated normal</td>
</tr>
<tr>
<td>Any</td>
<td>Piecewise constant</td>
<td>Piecewise weighted</td>
</tr>
<tr>
<td>Rank histogram</td>
<td>Any</td>
<td>Rank histogram (except tails)</td>
</tr>
<tr>
<td>Huber</td>
<td>Huber</td>
<td>Piecewise normal and exponential</td>
</tr>
<tr>
<td>Weighted sum of two normals</td>
<td>Normal</td>
<td>Weighted sum of two normals</td>
</tr>
<tr>
<td>Sum of N normals same variance</td>
<td>Normal</td>
<td>Weighted sum of N normals same variance</td>
</tr>
<tr>
<td>Jeffreys</td>
<td>Various</td>
<td>Various</td>
</tr>
</tbody>
</table>
Examples: EAKF, Rank Histogram, Sum of Normal Kernels

Three different continuous priors are shown for the same prior ensemble.

Continuous likelihood is normal. Piecewise constant approximation is used for RHF.

Posterior ensembles differ qualitatively.
Bayes' Rule Example: Gamma prior, Gamma Likelihood

Physical quantities may be bounded. For instance, amount of water vapor is non-negative.

Gamma prior enforces non-negativity.

Gamma likelihood leads to gamma posterior.
Example: Beta Prior and Likelihood

Sea ice concentration is bounded between 0 and 1.

A beta distribution can enforce these bounds.

A beta likelihood leads to a beta posterior.
There is statistical support for this method.

The Kolmogorov-Smirnov statistic is the same for the prior ensemble/continuous PDF as for the posterior.

The method is also related to the Q-Q plot for comparing distributions.
Extension to multivariate application

Using regression to update state variables means we lose control over the analysis distribution.

Can’t enforce boundedness for example.

Could apply GEFFQ directly to a state variable $x$,
Need a continuous likelihood as function of $x$,
Only know likelihood for ensemble members,
Piecewise constant approx. gives continuous likelihood.

This is same method used in the rank histogram filter (RHF).

Can directly update marginal ensemble for all state variables.
Useful for reanalysis purposes, too.
Proof of Concept Results: L63

Section 1:

N=80 Prior Period 3600 sec

N=80 Prior Period 10800 sec

N=80 Prior Period 21600 sec

N=80 Prior Period 1 day

N=80 Prior Period 2 day

N=80 Prior Period 4 day

RMSE vs Observation Error Variance

10 Expts. Mean

Data Assimilation Research Testbed

www.image.ucar.edu/DAReS/DART