Variational Kalman Filtering
in Lorenz'95 case

Harri Auvinen, Heikki Haario and Tuomo Kauranne
Department of mathematics
Lappeenranta University of Technology
`harri.auvinen@helsinki.fi`

December 2006

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Outline

🌙 Introduction to data assimilation

🌙 Kalman filter

🌙 Extended Kalman filter

🌙 Variational approach to the data assimilation

🌙 Variational Kalman filter

🌙 Simulated assimilation results

🌙 Conclusions and future work

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Introduction to data assimilation

The formulation of the general data assimilation (state estimation) problem for discrete time steps $t = 1, 2, 3, ..., n$ contains an evolution or prediction equation and an observation equation:

$$\mathbf{x}(t) = \mathcal{M}_t(\mathbf{x}(t-1)) + \mathbf{E}_t \tag{1}$$

$$\mathbf{y}(t) = \mathcal{K}_t(\mathbf{x}(t)) + \mathbf{e}_t, \tag{2}$$

where $\mathcal{M}_t$ is a model of the evolution operator, $\mathbf{x}_t$ is the state of the process at time t and $\mathbf{E}_t$ is a vector valued stochastic process.

The second equation connects the state estimate $\mathbf{x}(t)$ to the measurements $\mathbf{y}(t)$, with their associated observation errors $\mathbf{e}_t$, by way of an observation operator $\mathcal{K}_t$.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Basic linear Kalman filter

The basic Kalman filter operates on a linear version of the general assimilation problem, equations (1) and (2):

$$\mathbf{x}(t) = \mathbf{M}_t \mathbf{x}(t-1) + \mathbf{E}_t \tag{3}$$

$$\mathbf{y}(t) = \mathbf{K}_t \mathbf{x}(t) + \mathbf{e}_t, \tag{4}$$

where $\mathbf{M}_t$ is the linear evolution operator and similarly $\mathbf{K}_t$ is the linear observation operator. The Kalman filter operates sequentially in time.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Kalman filter algorithm (1/2)

Let $\mathbf{x}_{est}(t-1)$ be an estimate of state $\mathbf{x}(t-1)$ and $\mathbf{S}_{est}(t-1)$ be the corresponding error covariance matrix of the estimate. At time t the evolution operator is used to produce an *a priori* estimate $\mathbf{x}_a(t)$ and its covariance $\mathbf{S}_a(t)$:

$$
\begin{aligned}
\mathbf{x}_a(t) &= \mathbf{M}_t \mathbf{x}_{est}(t-1) & (5)\\
\mathbf{S}_a(t) &= \mathbf{M}_t \mathbf{S}_{est}(t-1)\mathbf{M}_t^{\mathrm{T}} + \mathbf{SE}_t, & (6)
\end{aligned}
$$

where $\mathbf{SE}_t$ is the covariance of the prediction error $\mathbf{E}_t$.

Kalman filter algorithm (2/2)

The next step is to combine $\mathbf{x}_a(t)$ with the observations $\mathbf{y}(t)$ made at time t to construct an updated estimate of the state and its covariance:

$$\mathbf{G}_t = \mathbf{S}_a(t)\mathbf{K}_t^{\mathrm{T}}(\mathbf{K}_t\mathbf{S}_a(t)\mathbf{K}_t^{\mathrm{T}} + \mathbf{Se}_t)^{-1} \tag{7}$$

$$\mathbf{x}_{est}(t) = \mathbf{x}_a(t) + \mathbf{G}_t(\mathbf{y}(t) - \mathbf{K}_t\mathbf{x}_a(t)) \tag{8}$$

$$\mathbf{S}_{est}(t) = \mathbf{S}_a(t) - \mathbf{G}_t\mathbf{K}_t\mathbf{S}_a(t), \tag{9}$$

where $\mathbf{G}_t$ is the Kalman gain matrix, which is functionally identical to the maximum *a posteriori* estimator.

In a more general case, when the evolution model and/or the observation model is non-linear, the Extended Kalman Filter (EKF) is required.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Extended Kalman filter algorithm (1/3)

The filter uses the full non-linear evolution model equation (1) to produce an *a priori* estimate: $\mathbf{x}_a(t) = \mathcal{M}_t(\mathbf{x}_{est}(t-1))$.

In order to obtain the corresponding covariance $\mathbf{S}_a(t)$ of the *a priori* information, the prediction model is linearized about $\mathbf{x}_{est}(t-1)$:

$$\mathbf{M}_t = \frac{\partial \mathcal{M}_t(\mathbf{x}_{est}(t-1))}{\partial \mathbf{x}} \tag{10}$$

$$\mathbf{S}_a(t) = \mathbf{M}_t \mathbf{S}_{est}(t-1)\mathbf{M}_t^{\mathrm{T}} + \mathbf{SE}_t. \tag{11}$$

The linearization in equation (10) is produced by taking finite differences between evolved small state vector perturbations, which is a computationally *very* expensive operation for large models.

Extended Kalman filter algorithm (2/3)

The observation operator is linearized at the time of the observation about the *a priori* estimate $\mathbf{x}_a(t)$ in order to obtain $\mathbf{K}_t$, which is then used for calculating the gain matrix:

$$\mathbf{K}_t = \frac{\partial \mathcal{K}_t(\mathbf{x}_a(t))}{\partial \mathbf{x}} \tag{12}$$

$$\mathbf{G}_t = \mathbf{S}_a(t)\mathbf{K}_t^{\mathrm{T}}(\mathbf{K}_t\mathbf{S}_a(t)\mathbf{K}_t^{\mathrm{T}} + \mathbf{Se}_t)^{-1}. \tag{13}$$

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Extended Kalman filter algorithm (3/3)

After this, the full non-linear observation operator is used to update $\mathbf{x}_a(t)$ and this is then used to produce a current state estimate and the corresponding error estimate:

$$\mathbf{x}_{est}(t) = \mathbf{x}_a(t) + \mathbf{G}_t(\mathbf{y}(t) - \mathcal{K}_t(\mathbf{x}_a(t))). \qquad (14)$$

$$\mathbf{S}_{est}(t) = \mathbf{S}_a(t) - \mathbf{G}_t\mathbf{K}_t\mathbf{S}_a(t). \qquad (15)$$

If the linearization of the observation operator at $\mathbf{x}_a(t)$ is not good enough to construct $\mathbf{x}_{est}(t)$, it is necessary to carry out some iterations of the last four equations. In which case the method is called Iterated Extended Kalman filter (IEKF).

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Variational approach to the data assimilation

The idea in variational approach is to avoid the computation of the gain completely by looking for the analysis as an approximate solution to the equivalent minimization problem defined by the cost function:

$$
\begin{aligned}
J(\mathbf{x}_t) &= (\mathbf{x}_a(t) - \mathbf{x}_t)^{\mathrm{T}}(\mathbf{S}_a(t))^{-1}(\mathbf{x}_a(t) - \mathbf{x}_t) \quad &(16)\\
&+ (\mathbf{y}_t - \mathcal{K}(\mathbf{x}_t))^{\mathrm{T}}\mathbf{Se}_t^{-1}(\mathbf{y}_t - \mathcal{K}(\mathbf{x}_t)), \quad &(17)
\end{aligned}
$$

where $\mathcal{K}$ is the nonlinear observation operator.

This approach is in literature called 3d-Var method. Opposite to the EKF method, the solution is retrieved iteratively by performing several evaluations of the cost function and of its gradient using a suitable descent algorithm. The approximation lies in the fact that only a small number of iterations are performed.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Variational Kalman filter method

We introduce Variational Kalman filter (VKF) method, which approximates the linear or Extended Kalman filter.

- In this scheme a 3D-Var method is used to find both the state estimate and the error covariance matrix.
- During 3D-Var step we collect search directions $\mathbf{s}_i$, $i = [0, ..., k-1]$ and corresponding gradients $\mathbf{z}_i$ of a BFGS optimization process.
- The inverse Hessian $(Hess(t))^{-1}$ of the 3D-Var cost function can be approximated by using limited memory BFGS (L-BFGS) formula, collected search directions and corresponding gradients.
- This approximated (limited memory) inverse Hessian can be then used to evolve the covariance matrix $\mathbf{S}_{est}(t)$ in order to compute $\mathbf{S}_a(t+1)$.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Limited memory BFGS formula

Let $\mathbf{H}_0$ be symmetric and positive definite and assume that the k pairs of search directions and gradients $\{\mathbf{s}_i, \mathbf{z}_i\}$ satisfy $\mathbf{s}_i^{\mathrm{T}} \mathbf{z}_i > 0$. Let $\mathbf{H}_k$ be obtained by updating $\mathbf{H}_0$ k times using the inverse BFGS formula and pairs $\{\mathbf{s}_i, \mathbf{z}_i\}_{i=0}^{k-1}$. Then

$$\mathbf{H}_k = \mathbf{H}_0 + [\mathbf{S}_k \ \mathbf{H}_0 \mathbf{Z}_k] \begin{bmatrix} \mathbf{R}_k^{-\mathrm{T}}(\mathbf{D}_k + \mathbf{Z}_k^{\mathrm{T}}\mathbf{H}_0\mathbf{Z}_k)\mathbf{R}_k^{-1} & -\mathbf{R}_k^{-\mathrm{T}} \\ -\mathbf{R}_k^{-1} & 0 \end{bmatrix} \begin{bmatrix} \mathbf{S}_k \\ \mathbf{H}_0 \mathbf{Z}_k \end{bmatrix},$$

(18)

where $\mathbf{S}_k = [\mathbf{s}_0, ..., \mathbf{s}_{k-1}]$ and $\mathbf{Z}_k = [\mathbf{z}_0, ..., \mathbf{z}_{k-1}]$ are $ns \times k$ matrices, $\mathbf{D}_k = diag[\mathbf{s}_0^{\mathrm{T}}\mathbf{z}_0, ..., \mathbf{s}_{k-1}^{\mathrm{T}}\mathbf{z}_{k-1}]$ is $k \times k$ matrix and $\mathbf{R}_k$ is $k \times k$ matrix:

$$(\mathbf{R}_k)_{i,j} = \begin{cases} \mathbf{s}_{i-1}^{\mathrm{T}}\mathbf{z}_{i-1} & \text{if } i > j \\ 0 & \text{if } i \le j \end{cases}.$$

(19)

Variational Kalman filter algorithm (1/4)

The VKF method uses in general case prediction equation (1) to construct an *a priori* estimate from the previous state estimate:

$$\mathbf{x}_a(t) = \mathcal{M}_t(\mathbf{x}_{est}(t-1)). \tag{20}$$

The corresponding approximated covariance $\mathbf{S}_a(t)$ of the *a priori* information is available from previous time step of VKF method.

In order to avoid the computation of the Kalman gain we perform a 3D-Var optimization with equivalent cost function. As a result of the optimization we get the state estimate $\mathbf{x}_{est}(t)$ for present time t.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Variational Kalman filter algorithm (2/4)

The error estimate $\mathbf{S}_{est}(t)$ of the state estimate follows from the formula:

$$\mathbf{S}_{est}(t) \approx 2(Hess(t))^{-1}, \tag{21}$$

where the matrix $(Hess(t))^{-1}$ can be approximated by using following inverse L-BFGS formula:

$$(Hess(t))^{-1} \approx \mathbf{H}_0 + \mathbf{X} * \mathbf{U} * \mathbf{X}^{\mathrm{T}}, \tag{22}$$

where $\mathbf{X} = [\mathbf{S}_k \ \ \mathbf{H}_0\mathbf{Z}_k]$ and $\mathbf{U}$ is the middle matrix of equation (18). In practice the initial inverse Hessian $\mathbf{H}_0 = \alpha\mathbf{I}$.

Variational Kalman filter algorithm (3/4)

For these matrices holds:

$$\mathbf{M}_t * \mathbf{S}_{est}(t) * \mathbf{M}_t^{\mathrm{T}} \approx 2(\mathbf{M}_t * \mathbf{H}_0 * \mathbf{M}_t^{\mathrm{T}} + (\mathbf{M}_t * \mathbf{X}) * \mathbf{U} * (\mathbf{M}_t * \mathbf{X})^{\mathrm{T}}). \quad (23)$$

Therefore we can approximate the covariance of the *a priori* information at time $t + 1$ by following formula:

$$\mathbf{S}_a(t + 1) \approx 2(\mathbf{M}_t * \mathbf{H}_0 * \mathbf{M}_t^{\mathrm{T}} + (\mathbf{M}_t * \mathbf{X}) * \mathbf{U} * (\mathbf{M}_t * \mathbf{X})^{\mathrm{T}}) + \mathbf{SE}(t + 1)$$
$$(24)$$

where $\mathbf{H}_0 = \alpha \mathbf{I}$.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

The inverse of covariance $S_a(t+1)$ **(1/2)**

In order to use the Sherman-Morrison-Woodbury (S-M-W) formula we set:

$$k_1 = 2k, \tag{25}$$

$$\mathbf{A} = 2(\mathbf{M}_t * \mathbf{H}_0 * \mathbf{M}_t^{\mathrm{T}}) + \mathbf{SE}(t+1), \tag{26}$$

$$\mathbf{W} = 2\mathbf{U}, \tag{27}$$

$$\mathbf{X}_1 = \mathbf{M}_t * [\mathbf{S}_k \ \ \mathbf{H}_0 \mathbf{Z}_k] = \mathbf{M}_t * \mathbf{X}, \tag{28}$$

$$\mathbf{X}_2 = \mathbf{X}_1. \tag{29}$$

To calculate the inverse of $\mathbf{A}$ efficiently we take advantage of the fact that the matrix $\mathbf{A}$ is sparse. In Lorenz'95 case it is enough to take a count the main diagonal and tree diagonals above and below the main diagonal. There is also other ways to approximate $\mathbf{A}^{-1}$.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

The inverse of covariance $\mathrm{S}_a(t+1)$ **(2/2)**

The tangent linear model $\mathbf{M}_t$ is assumed to be available in implicit form, as a code, which gives us a possibility to program a subroutine to compute only required diagonal and off-diagonal elements of $\mathbf{M}_t\mathbf{M}_t^{\mathrm{T}}$.

The inverse of $\mathbf{A}$ can then be computed efficiently, since $\mathbf{A} = 2\alpha\mathbf{M}_t\mathbf{M}_t^{\mathrm{T}} + \mathbf{SE}(t+1)$ is hepta-diagonal (in Lorenz'95 case).

Then the inverse of covariance $\mathbf{S}_a(t+1)$ follows directly from S-M-W formula:

$$(\mathbf{S}_a(t+1))^{-1} \approx \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{X}_1(\mathbf{W}^{-1} + \mathbf{X}_2^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{X}_1)^{-1}\mathbf{X}_2^{\mathrm{T}}\mathbf{A}^{-1}. \quad (30)$$

Variational Kalman filter algorithm (4/4)

- ☽ Now we have computed all necessary elements in order to proceed to the next observation time $t + 1$.

- ☽ In Variational Kalman filter algorithm we make two approximations: the L-BFGS approximation of the Hessian and tri/penta/heptadiagonal approximation of the matrix $\mathbf{A}$.

- ☽ We also assume that tangent linear model $\mathbf{M}_t$ is available as a code, which is necessary for example 4d-Var applications.

- ☽ It is also possible to derive a Variational Kalman smoother, since the covariance of the state estimate of VKF method is available in computationally efficient form.
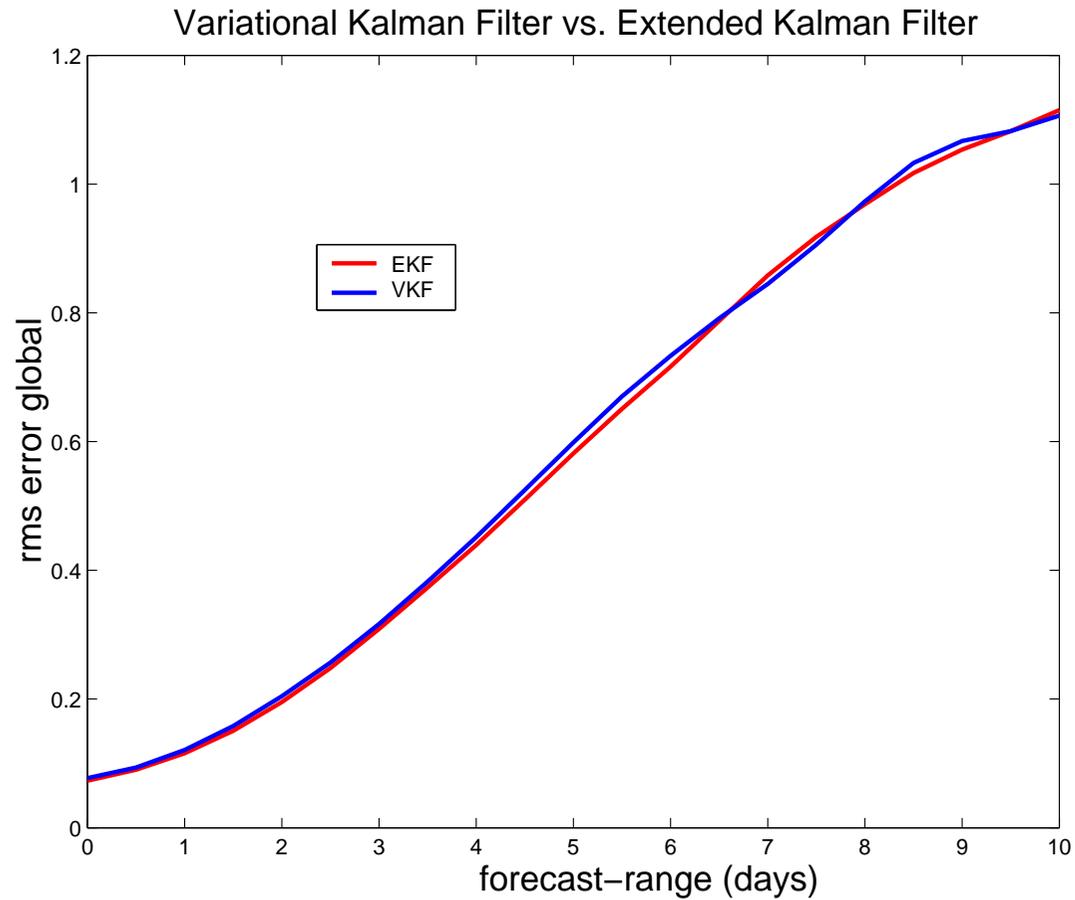
LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

The Lorenz'95 model

The simulated assimilation results were generated using an simple non-linear model introduced by Lorenz (1995) and Lorenz and Emmanuel (1998). As the model is computationally light and has a realistic analogue of mid-latitude atmospheric dynamics, it is commonly used for testing different assimilation schemes. The variables of the model could represent some quantity of the atmosphere on a latitude circle. The model consist of the coupled ordinary differential equations
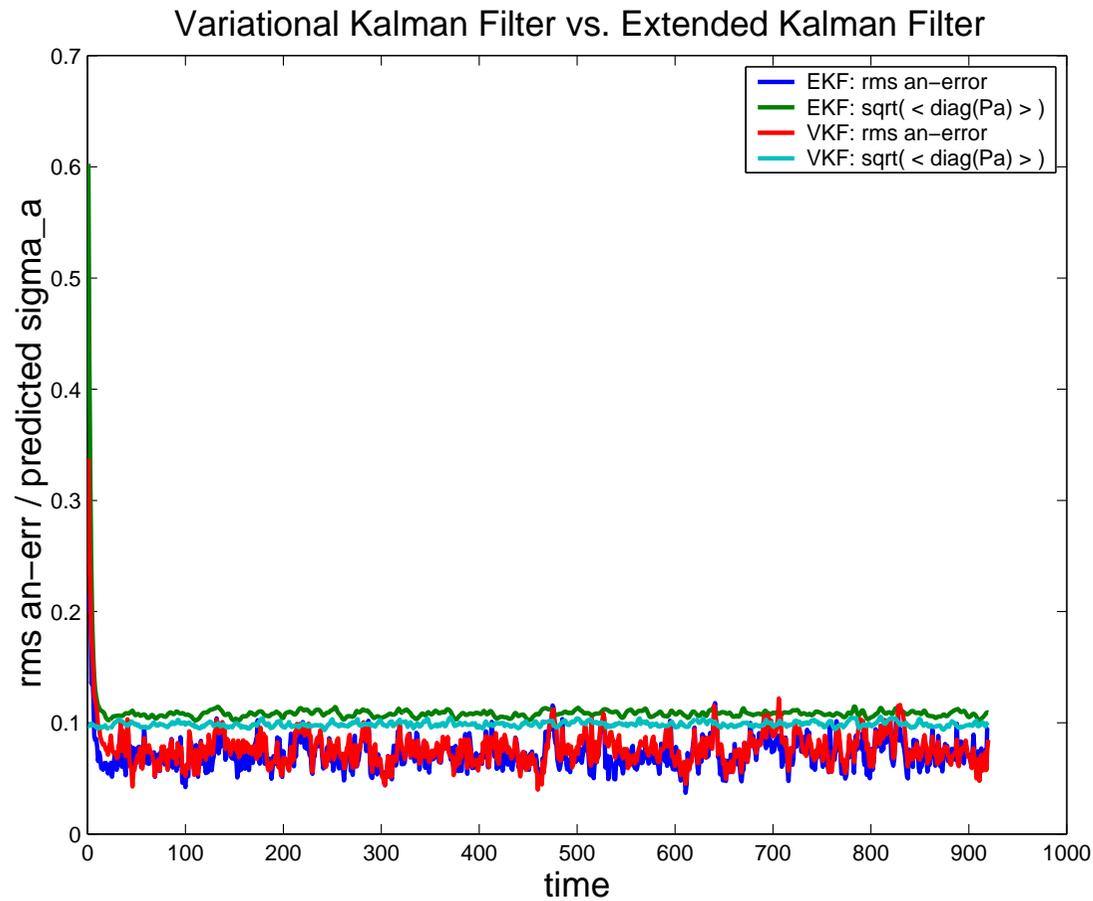
$$\frac{\partial c_i}{\partial t} = c_{i-2}c_{i-1} + c_{i-1}c_{i+1} - c_i + F, \tag{31}$$

where $i = 1, 2, ..., n$ and F is a constant. The number of grid points is controlled by the number $n = 40$. The domain is set to be cyclic, so that $c_{-1} = c_{n-1}, c_0 = c_n$ and $c_{n+1} = c_1$.

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

Simulated assimilation results in Lorenz'95 case

Variational Kalman Filter vs. Extended Kalman Filter

Simulated assimilation results in Lorenz'95 case

Conclusions and future work

🕭 We have derived the Variational Kalman filter (VKF) method to approximate a linear or extended Kalman filter.

🕭 It is possibly to use VKF method in Lorenz'95 case and archive equivalent results compared to EKF without computing the Kalman gain, if the rank of the approximated Hessian matrix is high enough.

🕭 In near future we will apply VKF method for high dimensional non-linear assimilation problem (Lorenz'95 model II).

LAPPEENRANTA
UNIVERSITY OF TECHNOLOGY

21