Didier Auroux
University of Nice Sophia Antipolis
France
auroux@unice.fr

Nudging methods in geophysical data assimilation

Part 2 : Optimal nudging and comparisons

Mathematical and Statistical Foundations of Data Assimilation,
TIFR-CAM, Bangalore, India.
July 04-23, 2011

Talk overview

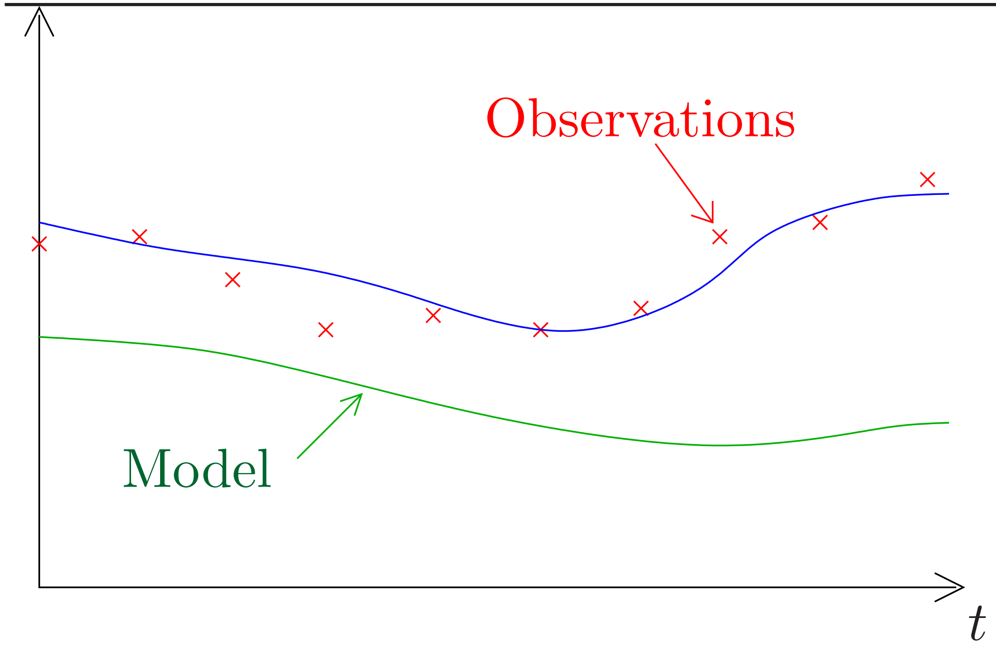
1. Optimal nudging (use of a variational method)

2. Kalman filter and optimal nudging

3. Comparison optimal nudging - 4D-VAR

4. Hybrid nudging - ensemble Kalman filter

Data assimilation



combination

model + observations

$\Downarrow$

identification of the initial condition

of a geophysical system

Fundamental for a chaotic system (atmosphere, ocean, ...)

Issue : These systems are generally irreversible

Goal : Combine models and data

Typical inverse problem : retrieve the system state from sparse and noisy observations

Talk overview

⇒ 1. Optimal nudging (use of a variational method)

2. Kalman filter and optimal nudging

3. Comparison optimal nudging - 4D-VAR

4. Hybrid nudging - ensemble Kalman filter

Nudging

Model :

$$\begin{cases} \dfrac{dX}{dt} = F(X), \quad 0 < t < T, \\ X(0) = X_0, \end{cases}$$

Standard nudging applied to this model :

$$\begin{cases} \dfrac{dX}{dt} = F(X) + K(Y_{obs} - H(X)), \quad 0 < t < T, \\ X(0) = X_0, \end{cases}$$

where K is the nudging (or gain) matrix.

Optimal determination of K

Goal : improve the standard nudging technique.

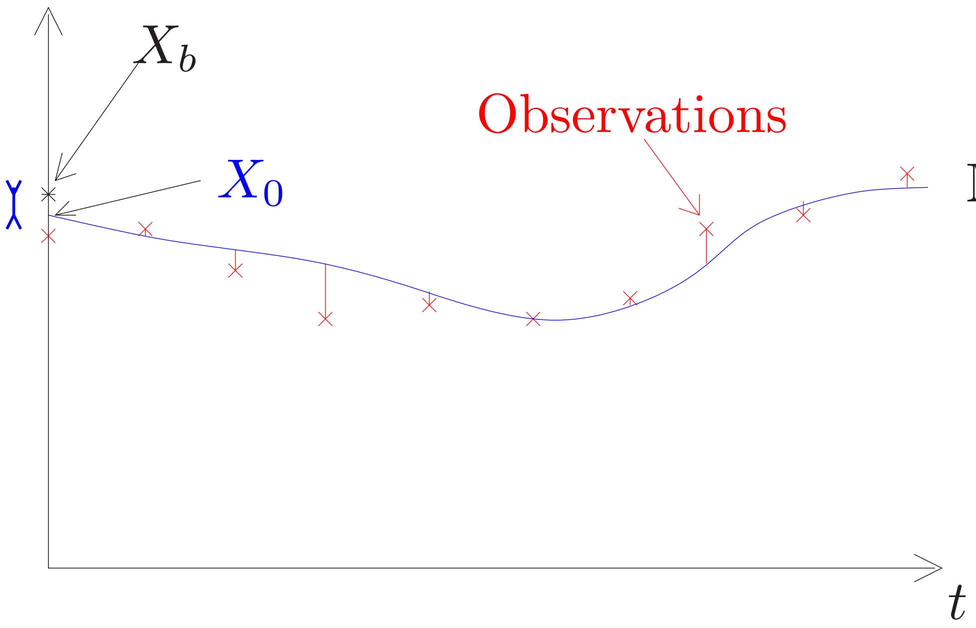
Use of a variational FDDA (Four Dimensional Data Assimilation) technique for parameter estimation in order to determine the best K.

$$\begin{cases} \dfrac{dX}{dt} = F(X) + K(Y_{obs} - H(X)), \quad 0 < t < T, \\ X(0) = X_0, \end{cases}$$

Use of a variational approach : we consider the 4D-VAR cost function, measuring the difference between the observations and the corresponding quantities computed by the model :

$$J(X_0) = \frac{1}{2} \int_0^T \langle R^{-1}(Y_{obs} - H(X)), Y_{obs} - H(X) \rangle \, dt + \frac{1}{2} \langle B^{-1}(X_0 - X_b), X_0 - X_b \rangle.$$

4D-VAR



Model governed by a system of ODE :

$$\begin{cases} \dfrac{dX}{dt} = F(X), \quad 0 < t < T \\ X(0) = X_0 \end{cases}$$

$Y_{obs}(t)$: observations of the system, H : observation operator,
X_b : background state (a priori estimation of the initial condition),
B and R : covariance matrices of background and observation errors.

$$\begin{aligned} J(X_0) &= \frac{1}{2}(X_0 - X_b)^T B^{-1}(X_0 - X_b) \\ &+ \frac{1}{2}\int_0^T [Y_{obs}(t) - H(X(t))]^T R^{-1} [Y_{obs}(t) - H(X(t))] \; dt \end{aligned}$$

Optimal determination of K

In the following, we do not take into account the background part of the cost function. We need to add a term that controls the nudging matrix K :

$$J(X_0, K) = \frac{1}{2} \int_0^T \Big(\langle R^{-1}(Y_{obs} - H(X)), Y_{obs} - H(X) \rangle + \langle G^{-1}(K - \hat{K}), K - \hat{K} \rangle \Big) \, dt$$

where R and G are weighting (covariance) matrices, and $\hat{K}$ denotes the estimated nudging coefficients.

This term can be seen as a regularization term, ensuring the strict (and even strong) convexity of the cost function, and then existence and uniqueness of the minimum. It also allows the minimization algorithm to be efficient.

Minimizing J with respect to K leads to the optimal nudging coefficients K^*, and then to an optimal nudging data assimilation scheme.

Minimization under constraints

The goal is to minimize $J(X_0, K)$ under the model constraint : $X(t)$ has to be a solution of the model (with nudging). We introduce then a Lagrange multiplier associated to the constraint :

$$\mathcal{L}(X_0, X, K, P) = J(X_0, K) + \int_0^T \langle P, \frac{dX}{dt} - F(X) - K(Y_{obs} - H(X))\rangle$$

$$= \frac{1}{2}\int_0^T \left(\langle R^{-1}(Y_{obs} - H(X)), Y_{obs} - H(X)\rangle + \langle G^{-1}(K - \hat{K}), K - \hat{K}\rangle\right) dt$$

$$+ \int_0^T \langle P, \frac{dX}{dt} - F(X) - K(Y_{obs} - H(X))\rangle$$

where P is a function of time, defined over $[0;T]$, and the Lagrange multiplier associated to the nudging model constraint. There is no more constraint, and X is assumed to be any function of time.

Minimization under constraints

It is straightforward to see that

$$\min_{X_0,K,\ X \text{ model solution}} J(X_0,K) = \min_{X_0,K} \max_{P} \mathcal{L}(X_0,X,K,P)$$

Indeed, for a given function X :

- either X is a solution of the nudging model, and then the second term of $\mathcal{L}$ is equal to 0, and $\mathcal{L} = J$;
- or X is not a solution of the nudging model, and then the maximum over all functions P is equal to $+\infty$.

Then the minimization can be restricted to X solution of the nudging model.

Optimality conditions

Optimality conditions : all partial derivatives of $\mathcal{L}$ should be equal to 0.

• **Maximum with respect to** P **:** the partial derivative $\dfrac{\partial \mathcal{L}}{\partial P}$ is equal to 0, which implies

$$\frac{dX}{dt} - F(X) - K(Y_{obs} - H(X)) = 0, \ \forall t,$$

and then X has to be a solution of the model with nudging.

Optimality conditions

• **Minimum with respect to X :** using an integration by parts,

$$\mathcal{L} = \frac{1}{2}\int_0^T \Big(\langle R^{-1}(Y_{obs} - H(X)), Y_{obs} - H(X)\rangle + \langle G^{-1}(K - \hat{K}), K - \hat{K}\rangle \Big)\ dt$$

$$+\langle P(T), X(T)\rangle - \langle P(0), X_0\rangle - \int_0^T \langle \frac{dP}{dt}, X\rangle - \int_0^T \langle P, F(X) + K(Y_{obs} - H(X))\rangle$$

and then, $\dfrac{\partial \mathcal{L}}{\partial X} = 0$ implies that P is solution of the following adjoint model :

$$H'(X)^T R^{-1}(H(X) - Y_{obs}) - \frac{dP}{dt} - [F'(X)]^T P + [H'(X)]^T K^T P = 0,$$

and that $P(T) = 0$.

Note that this model has to be solved backwards in time, from a final condition to the initial time.

Optimality conditions

• **Minimum with respect to X_0 :** using the previous expression of the Lagrangian, if X is a solution of the nudging model, then $\mathcal{L}(X_0, X, K, P) = J(X_0, K)$ for any P, and then $\dfrac{\partial J}{\partial X_0} = \dfrac{\partial \mathcal{L}}{\partial X_0}$, and

$$\frac{\partial \mathcal{L}}{\partial X_0} = -P(0).$$

Then the optimaly condition is

$$-P(0) = 0.$$

Optimality conditions

• **Minimum with respect to K :** in a similar way, $\dfrac{\partial J}{\partial K} = \dfrac{\partial \mathcal{L}}{\partial K}$, and

$$\frac{\partial \mathcal{L}}{\partial K}.\delta K = \int_0^T \langle G^{-1}(K - \hat{K}), \delta K \rangle \, dt - \int_0^T \langle P, \delta K (Y_{obs} - H(X)) \rangle \, dt = 0$$

for all δK.

Optimality system

Optimality system :

$$\frac{dX}{dt} - F(X) - K(Y_{obs} - H(X)) = 0, \quad X(0) = X_0$$

$$-\frac{dP}{dt} = [F'(X)]^T P + H'(X)^T R^{-1}(Y_{obs} - H(X)) - [H'(X)]^T K^T P, \quad P(T) = 0$$

$$-P(0) = 0$$

$$\int_0^T \langle G^{-1}(K - \hat{K}), \delta K \rangle \, dt - \int_0^T \langle P, \delta K(Y_{obs} - H(X)) \rangle \, dt = 0, \quad \forall \delta K$$

Optimal determination of K

Algorithm :

- For a given initial condition X_0 and a given nudging matrix K, solve the direct model with nudging $\Rightarrow X$.
- Linearize the model and observation operators around the trajectory X, and solve the adjoint model $\Rightarrow P$.
- Compute the gradient of the cost function (as a function of P), and update the controls X_0 and K.

$$\nabla_{X_0} J = -P(0),$$

$$\nabla_K J \,.\, \delta K = -\int_0^T \langle\, \delta K \,(Y_{obs} - H(X)), P\,\rangle \, dt + G^{-1}(K - \hat{K}) \,.\, \delta K.$$

Optimal determination of K

Drawback : computational cost of the optimization, need for a good a priori estimate.

[Zou-Navon-Le Dimet (92)]

Once J has been optimized, and the optimal nudging gain K^* identified, one may assume that this gain can be used for the next assimilation windows.

Talk overview

1. Optimal nudging (use of a variational method)

⇒ 2. Kalman filter and optimal nudging

3. Comparison optimal nudging - 4D-VAR

4. Hybrid nudging - ensemble Kalman filter

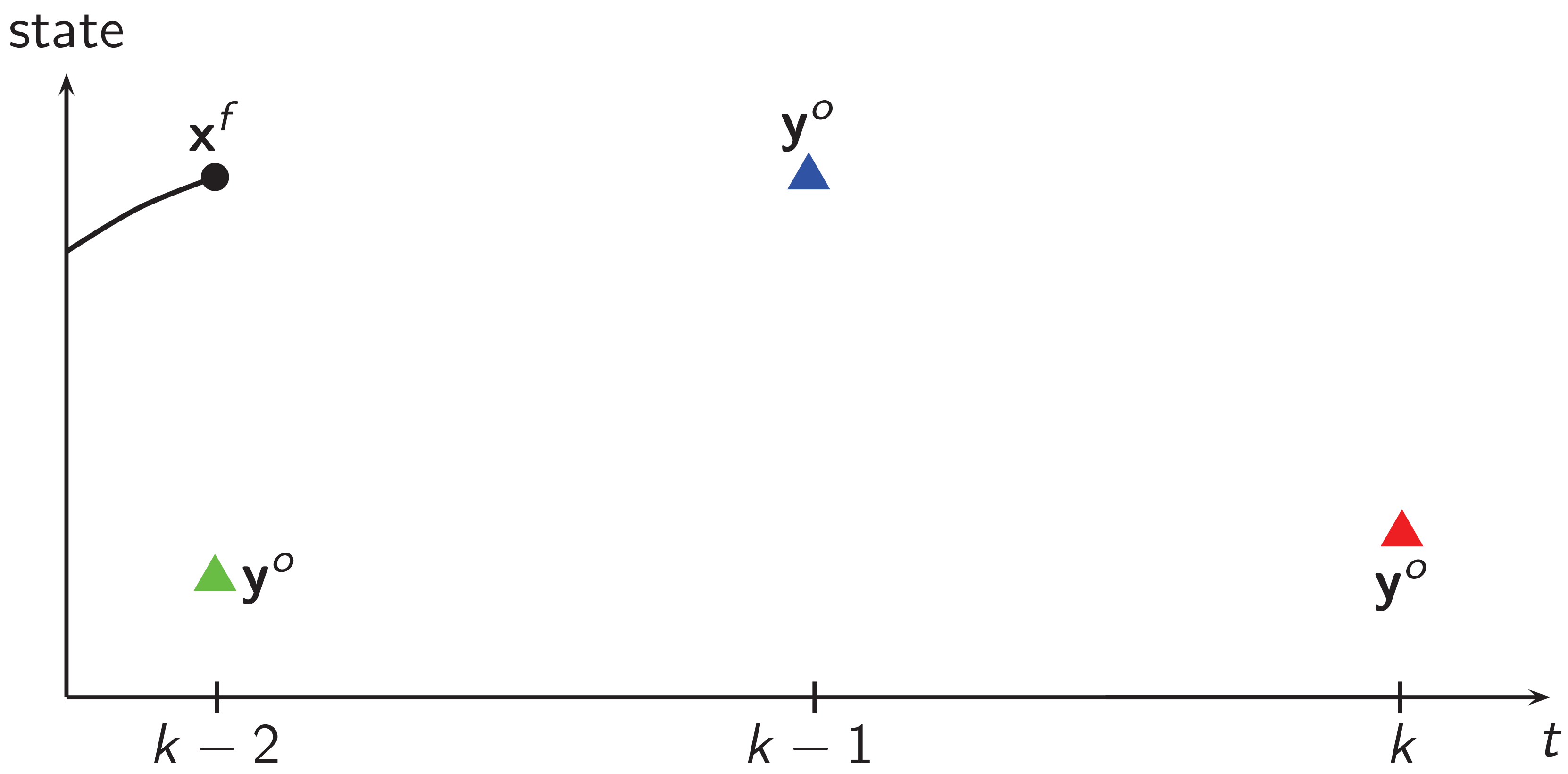
Kalman filter

Kalman filter : extension to dynamical systems of the optimal interpolation, or BLUE (Best Linear Unbiased Estimator) between the observation and the forecast state.

Two steps, repeated for each observation time : prediction and correction.

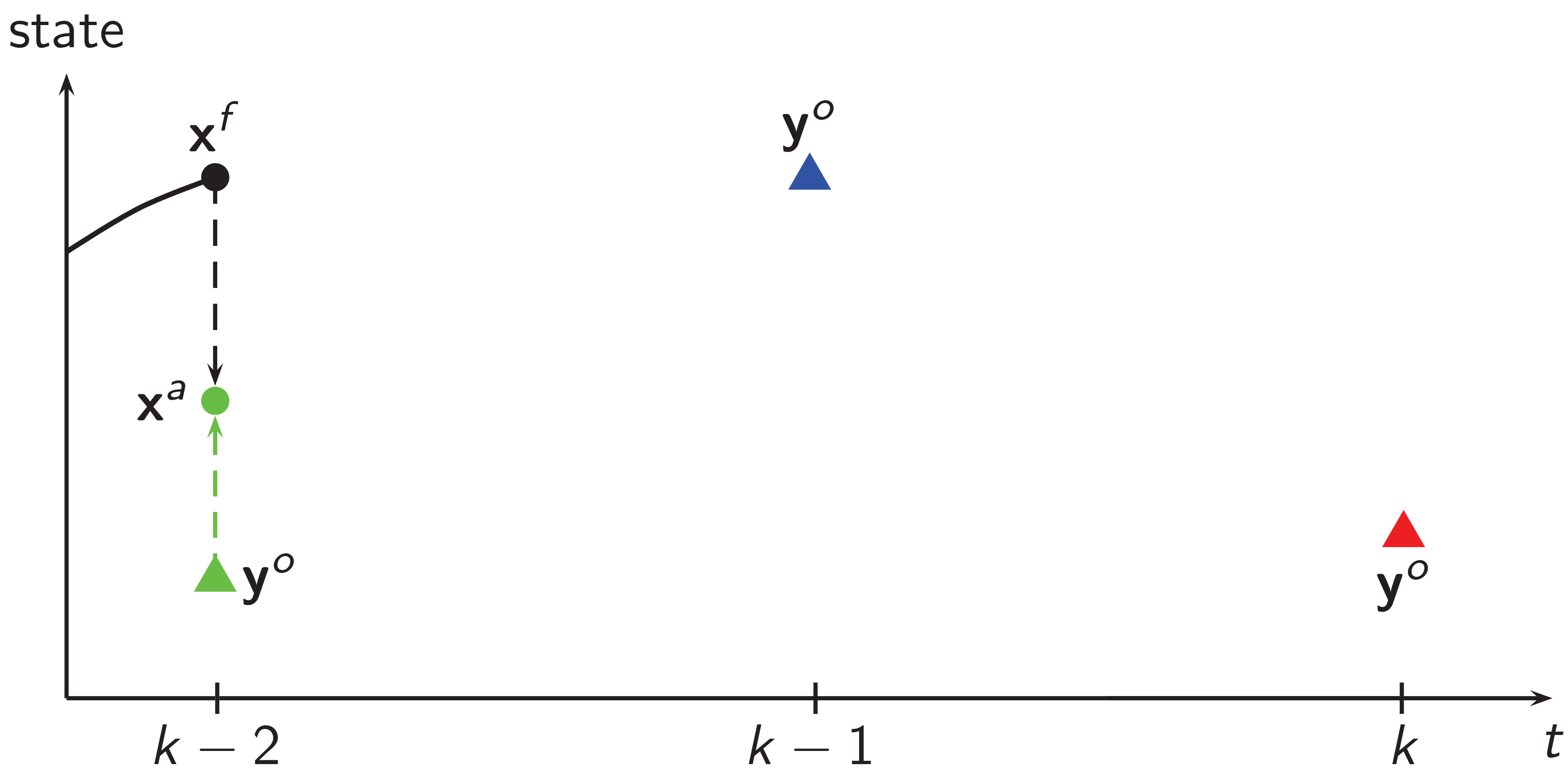
Kalman filter

Kalman filter :



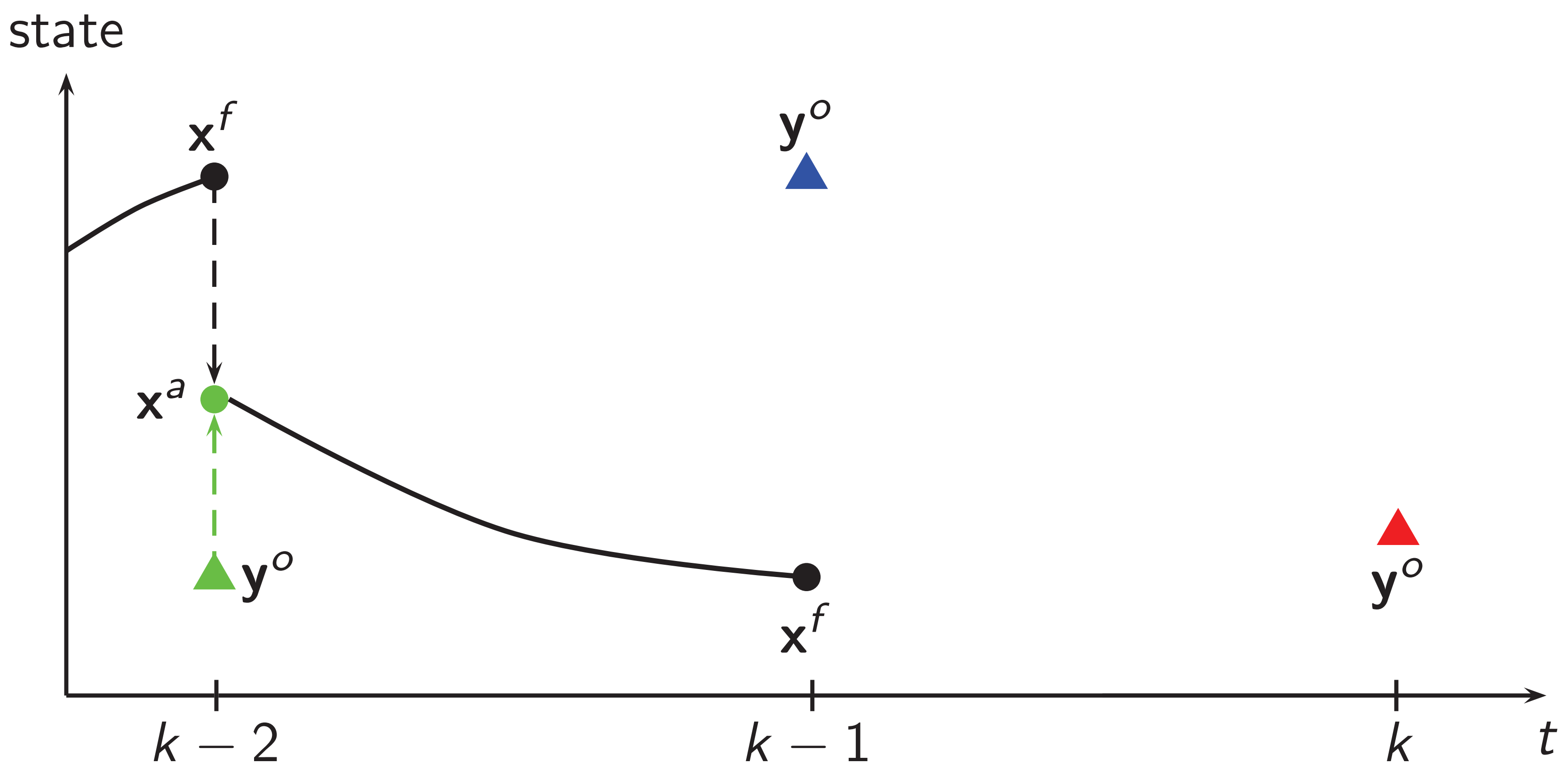
Kalman filter

Kalman filter :



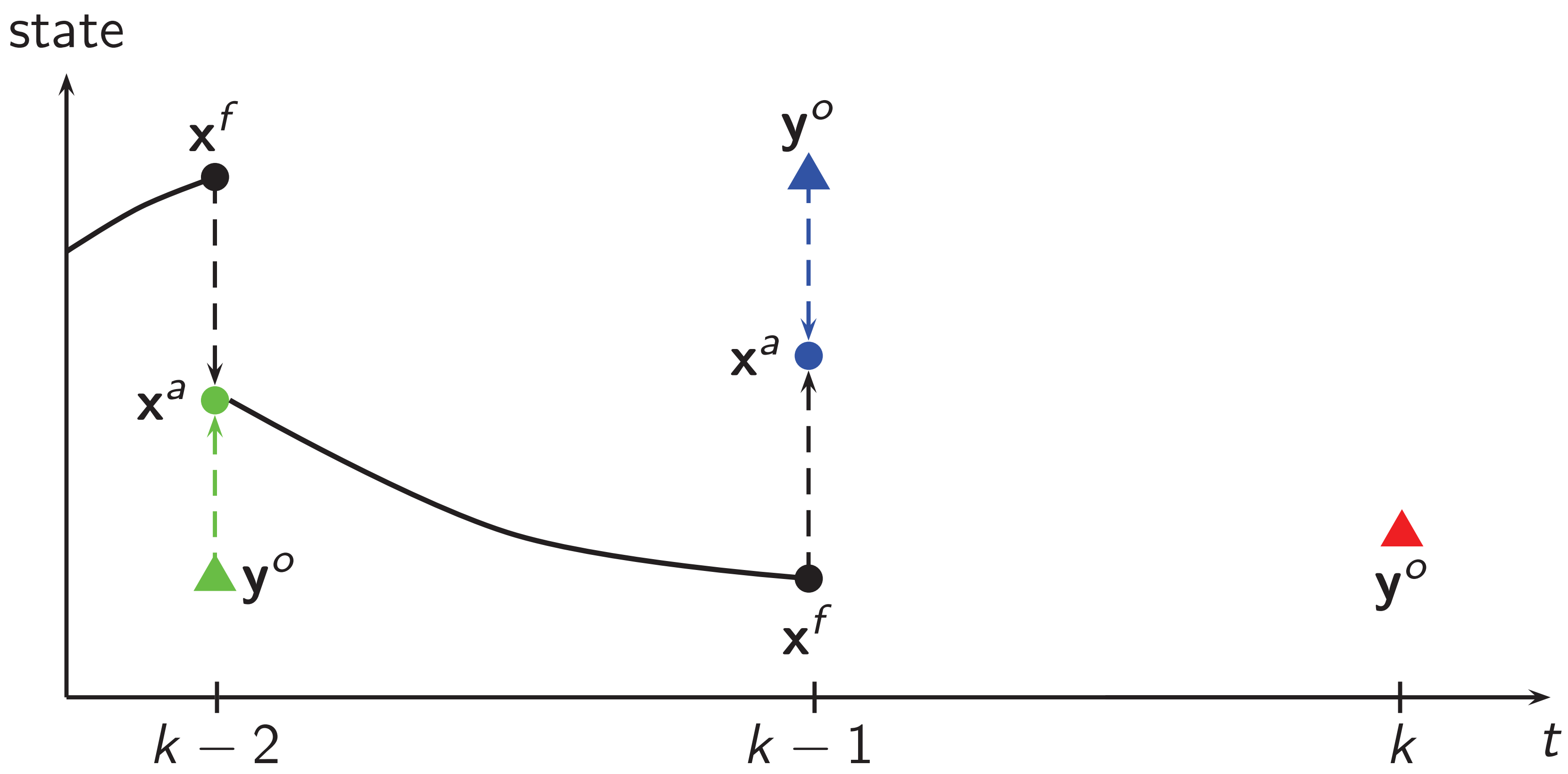
Kalman filter

Kalman filter :

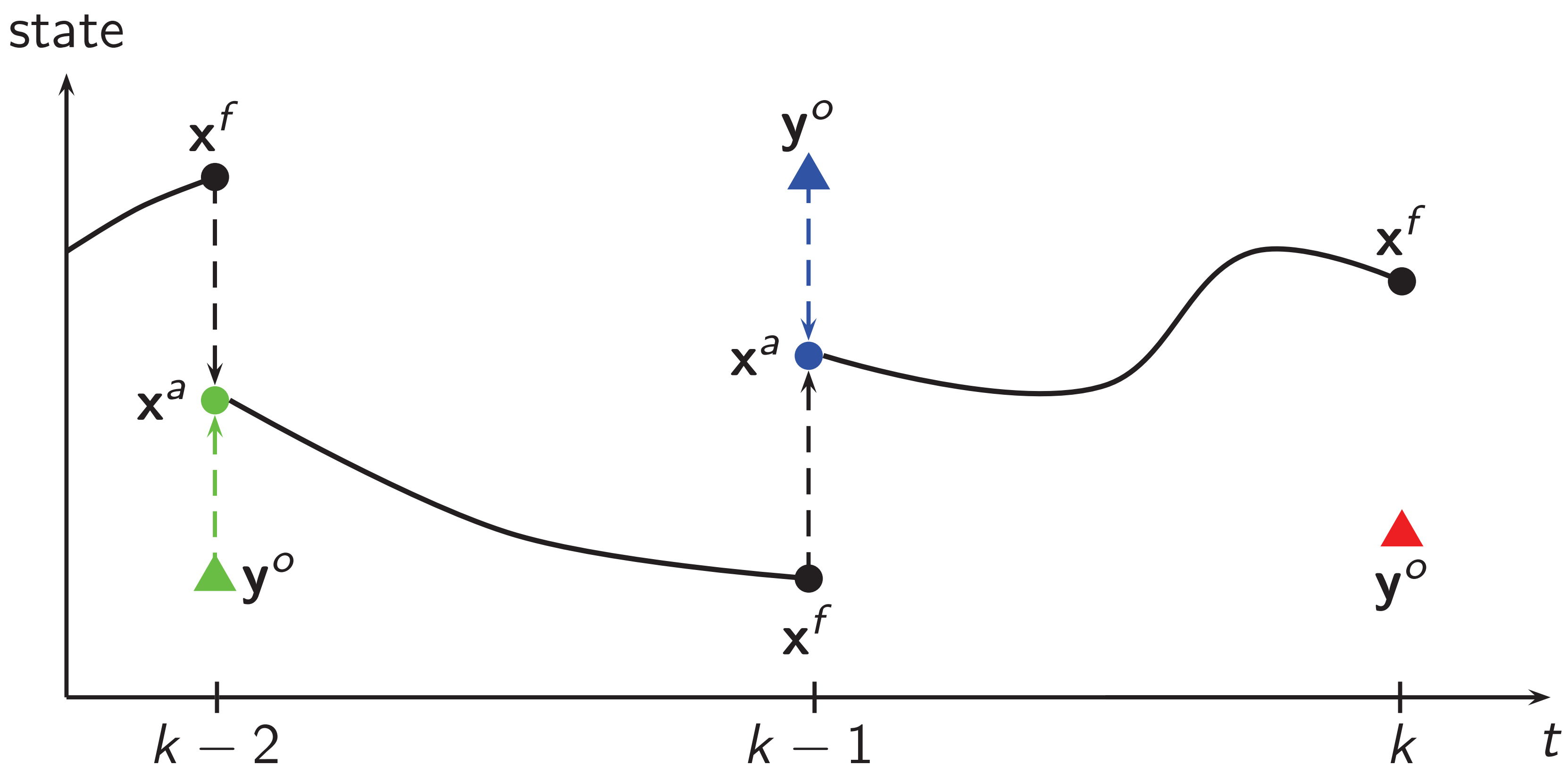


Kalman filter

Kalman filter :

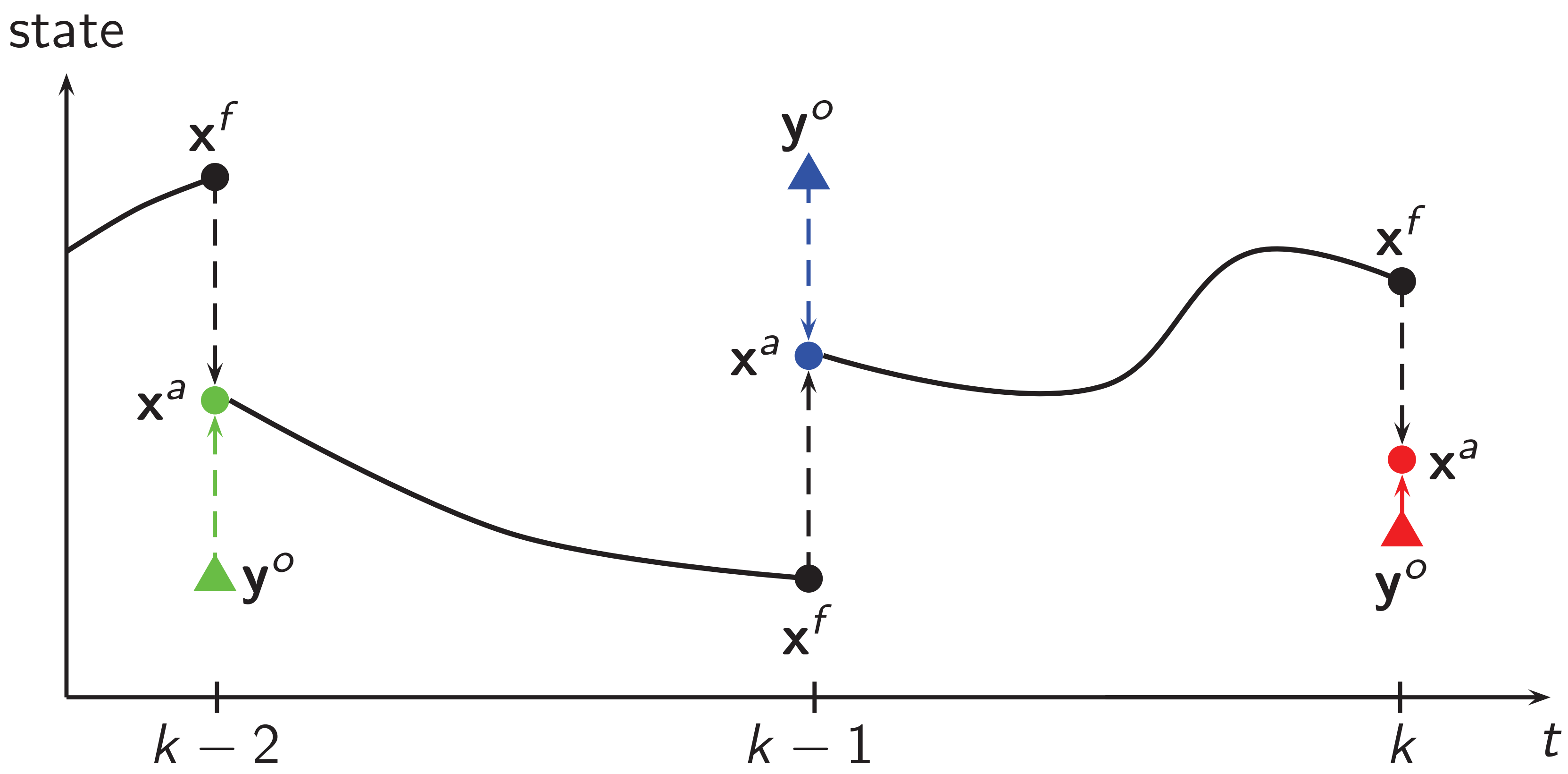


Kalman filter :



Kalman filter

Kalman filter :



Prediction step

We denote by X_n^a the **analyzed state** at time t_n, and M_n the system resolvent from time t_n to t_{n+1}.

Prediction/forecast step : get a background state at time t_{n+1}

$$X_{n+1}^f = M_n X_n^a$$

We denote by P_n^f the covariance matrix of forecast error $X_n^f - X_n^t$ and P_n^a the covariance matrix of analysis error $X_n^a - X_n^t$.

$$P_{n+1}^f = E\left((X_{n+1}^f - X_{n+1}^t)(X_{n+1}^f - X_{n+1}^t)^T\right)$$
$$= E\left((M_n X_n^a - M_n X_n^t - \varepsilon_n)(M_n X_n^a - M_n X_n^t - \varepsilon_n)^T\right)$$

$$P_{n+1}^f = M_n P_n^a M_n^T + Q_n$$

where Q_n is the model error covariance matrix at time t_n.

Correction step

Analysis step : Correction of the background vector X^f_{n+1} with the innovation vector $Y_{obs\,n+1} - H_{n+1}X^f_{n+1}$:

$$X^a_{n+1} = X^f_{n+1} + K_{n+1}(Y_{obs\,n+1} - H_{n+1}X^f_{n+1})$$

The new analysis error $e^a_{n+1} = X^a_{n+1} - X^t_{n+1}$ is :

$$e^a_{n+1} = e^f_{n+1} + K_{n+1}(\epsilon_{n+1} - H_{n+1}e^f_{n+1})$$

where ϵ is the observation error (of covariance matrix R) and e^f is the forecast error (of covariance matrix P^f).

$$P^a_{n+1} = [I - K_{n+1}H_{n+1}]P^f_{n+1}[I - K_{n+1}H_{n+1}]^T + K_{n+1}R_{n+1}K^T_{n+1}$$

One chooses K_{n+1} such that the variance of analysis error is minimum :

$$K_{n+1} = P^f_{n+1}H^T_{n+1}[H_{n+1}P^f_{n+1}H^T_{n+1} + R_{n+1}]^{-1}$$

Then, $$P^a_{n+1} = [I - K_{n+1}H_{n+1}]P^f_{n+1}$$

KF algorithm

- Initialization :

$$X_0^f \text{ and } P_0^f \text{ given}$$

- Analysis :

$$\begin{aligned} K_n &= P_n^f H_n^T [H_n P_n^f H_n^T + R_n]^{-1} \\ X_n^a &= X_n^f + K_n (Y_{obs\,n} - H_n X_n^f) \\ P_n^a &= [I - K_n H_n] P_n^f \end{aligned} \tag{1}$$

- Forecast :

$$\begin{aligned} X_{n+1}^f &= M_{n;n+1} X_n^a \\ P_{n+1}^f &= M_{n;n+1} P_n^a M_{n;n+1}^T + Q_n \end{aligned}$$

Comparison nudging - Kalman

Two main problems arise in the Kalman filter. First, the computational complexity of advancing in time the error covariance matrices. Second, the noise covariance matrices are assumed to be perfectly known, but this is not the case in practice (particularly Q_n).

The nudging scheme is carried out by the following procedure :

$$X_n^a = M_{n-1;n} X_{n-1}^a + K_n (Y_{obs\,n} - H_n X_n^f).$$

The optimal nudging coefficients K_n are obtained by minimizing a cost function measuring the distance between the analysis and the observations :

$$J(K) = \sum_{n=0}^{N} (H_n(X_n^a) - Y_{obs\,n})^T G^{-1} (H_n(X_n^a) - Y_{obs\,n}),$$

where G is the same weighting matrix as before.

Comparison nudging - Kalman

The core of the KF is the optimal merging of observation and forecast information in the sense that the expected mean-square estimation error is minimized at every time step.

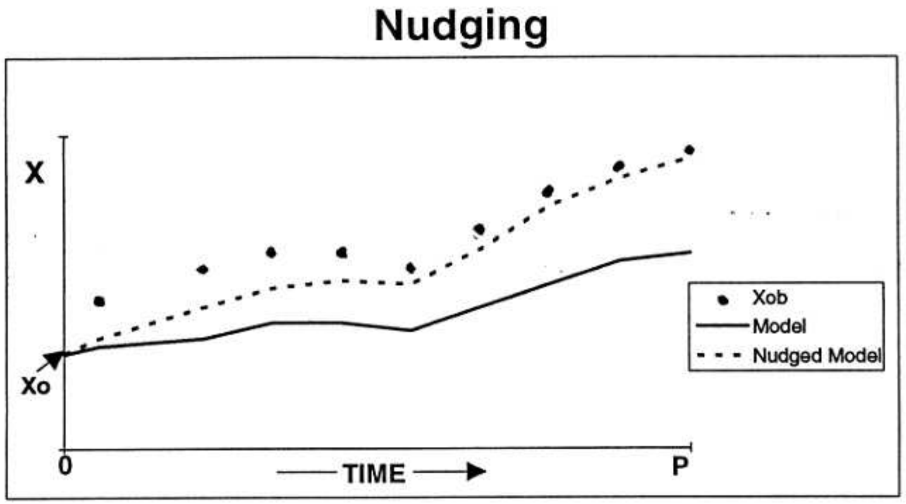
The optimal nudging is the optimal merging of observations and analysis in the sense that the total differences between them is minimized over the assimilation window.

Main difference : the gain matrix is determined sequentially in the KF, at each time step, while the nudging coefficients at every time step are obtained simultaneously. Moreover, the two problems of the KF described above disappear in the optimal nudging method : the computational cost is reduced, and the optimal nudging does not require any knowledge of the noise covariance matrices.

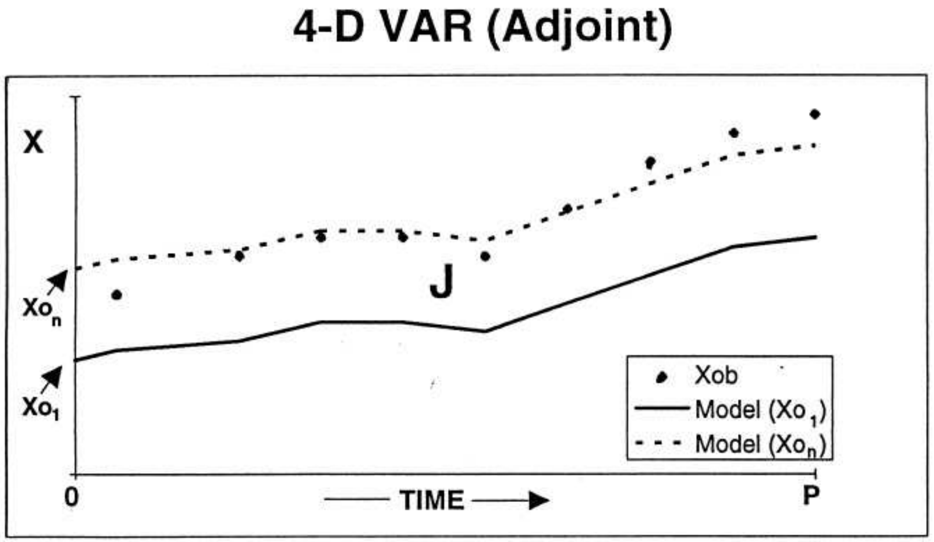
Talk overview

1. Optimal nudging (use of a variational method)
2. Kalman filter and optimal nudging
3. Comparison optimal nudging - 4D-VAR ⇒
4. Hybrid nudging - ensemble Kalman filter

Comparison nudging - 4DVAR



"... introduces an artificial tendency term to gradually correct the model solution where necessary."



"... fits a model to the data by modifying the initial state (Xo) to minimize model error (J) through the period (P)."

Back to 4DVAR formulation

Motivations : the nudging term provides a correction of the model with the observations, and the nudging does not assume the model to be perfect (weak-constrained model).

Adding a nudging term in the model equation, within the framework of 4D-VAR, relaxes the strong model constraint. This is an easy way to consider the model error η in VDA :

$$J(X_0,\eta) = \frac{1}{2}\int_0^T \langle R^{-1}(Y_{obs}-HX), Y_{obs}-HX\rangle\, dt + \frac{1}{2}\langle B^{-1}\delta X_0, \delta X_0\rangle + \frac{1}{2}\langle Q^{-1}\eta, \eta\rangle$$

by setting $\eta = K(Y_{obs} - HX)$.

[Vidard-Le Dimet-Piacentini (03)]

Comparison nudging - 4DVAR

The optimization is then performed with respect to X_0 and K. The computation of the gradient of the cost function is done like in the 4D-VAR (adjoint technique), but the optimization is harder.

Usually, it is not possible to consider a full rank matrix K. Simpler solutions : scalar coefficient (varying with time), or pseudo-diagonal matrix (corrections at the same location as observations).

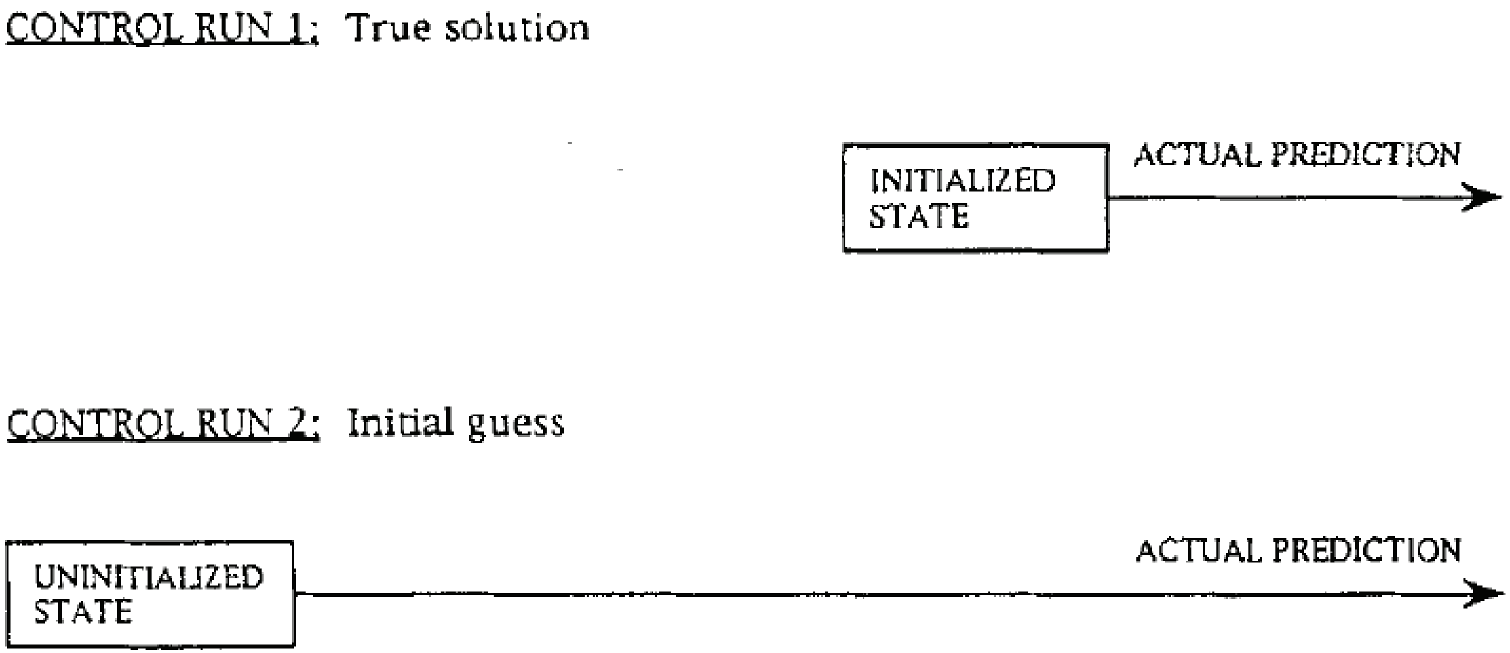
Much cheaper than the 4D-VAR with an imperfect model, but it does not take into account any model error term.

Comparison (optimal) nudging - 4DVAR

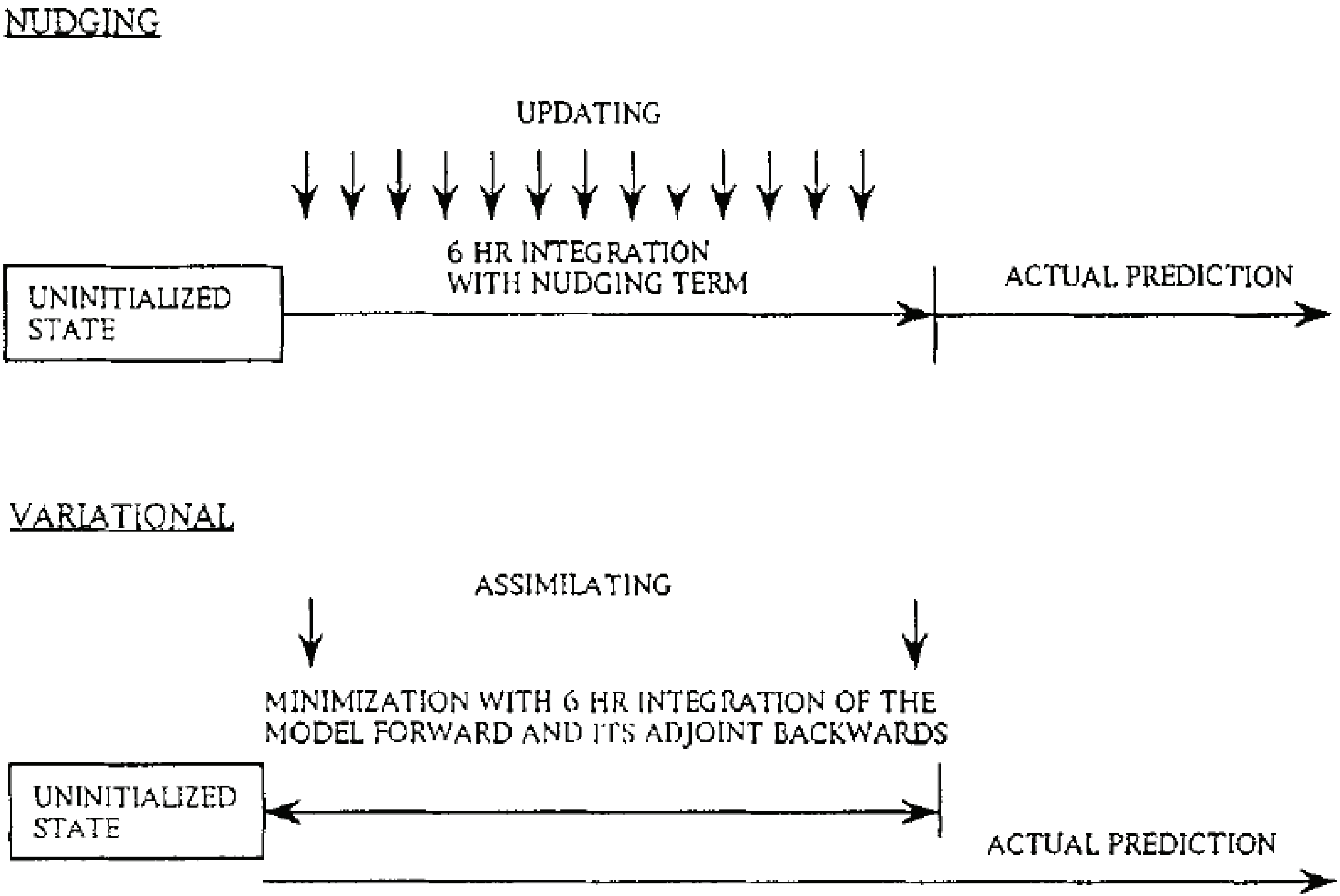
Comparisons between nudging, optimal nudging, and variational data assimilaiton :

Courtesy of X. Zou, I. M. Navon, and F.-X. Le Dimet (Q. J. R. Meteorol. Soc. 1992)

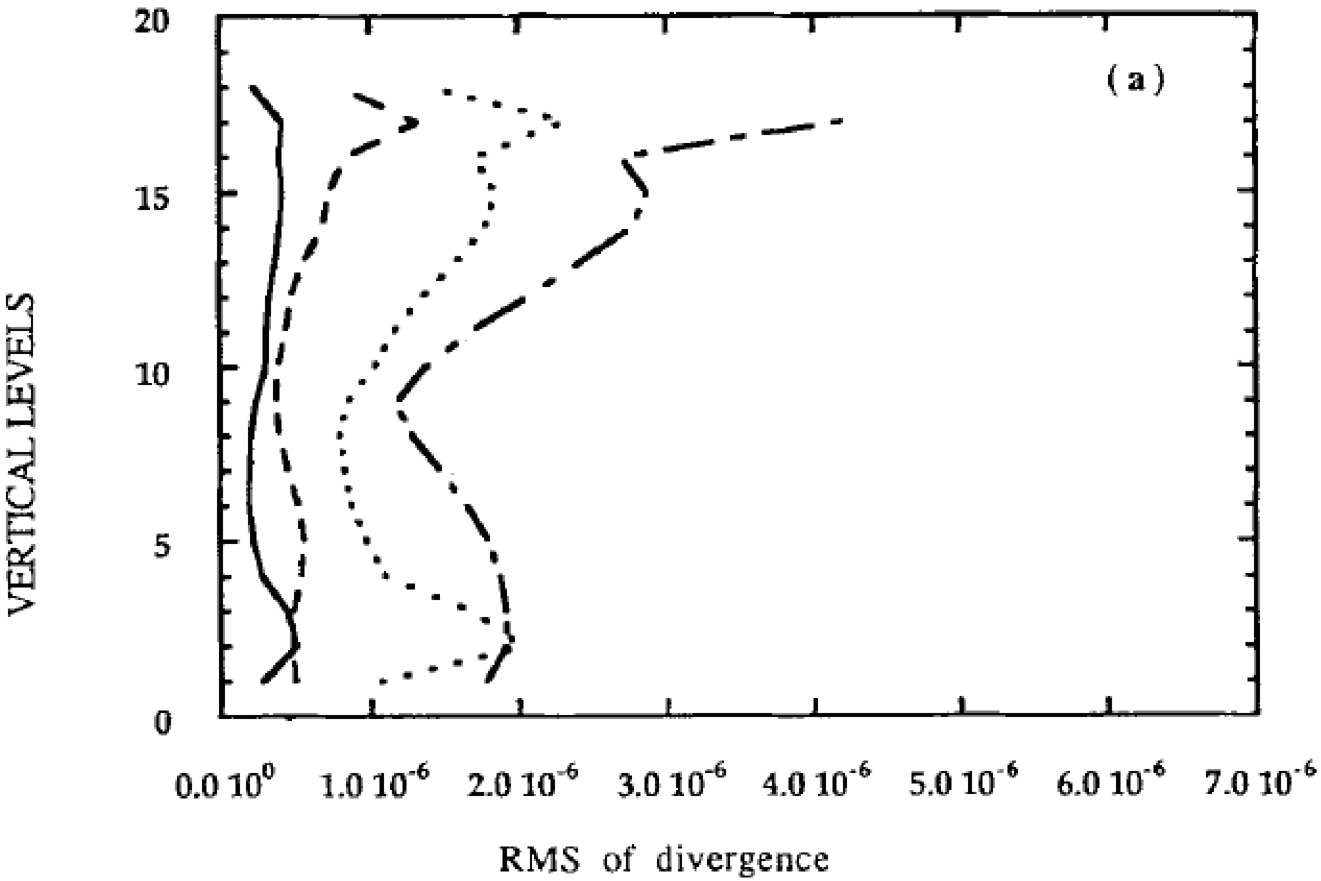
Tests on the national meteorological center model with 18 vertical layers.



Comparison (optimal) nudging - 4DVAR

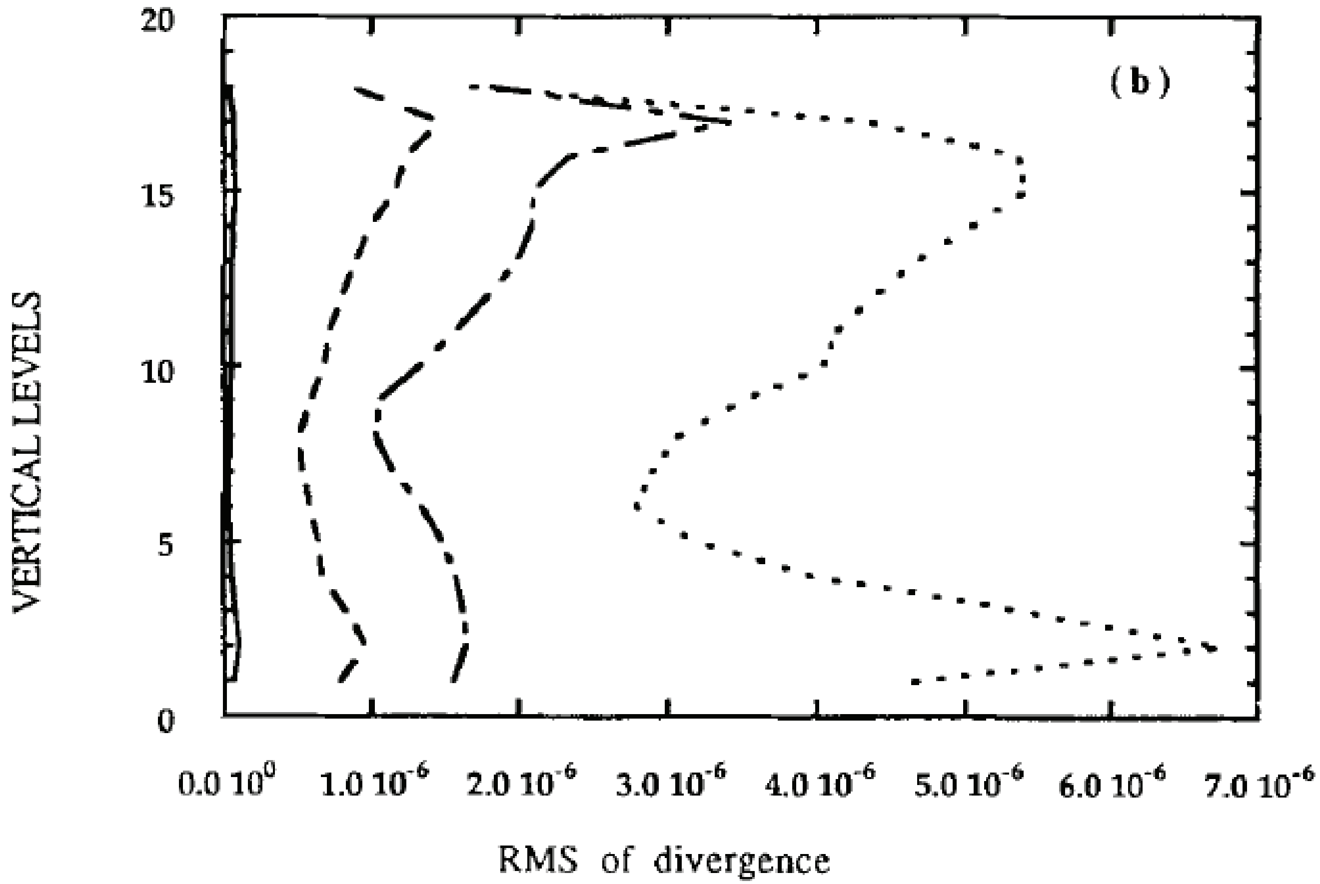


Comparison (optimal) nudging - 4DVAR



RMS difference for the divergence field between the true solution and the assimilated fields at the end of 6-hour assimilation intervals : control run (-.-), nudging (...), variational data assimilation (—), and optimal nudging (plain).

Comparison (optimal) nudging - 4DVAR



RMS difference for the divergence field between the true solution and the assimilated fields at the end of 12-hour assimilation intervals : control run (-.-), nudging (...), variational data assimilation (—), and optimal nudging (plain).

(Sub-)optimal nudging

Other experiments : in Vidard, Le Dimet, Piacentini, Tellus 2003

1D Burgers equation :

$$\begin{cases} \dfrac{\partial u}{\partial t} + u\dfrac{\partial u}{\partial x} - \nu\dfrac{\partial^2 u}{\partial x^2} = f, \\ u(0,t) = u(1,t) = 0, \\ u(x,0) = \sin \pi x \end{cases}$$

With discrete (in time and space) observations, the additional nudging term is

$$+\sum_{n=1}^{N} K_n [u_n^{obs} - u_n]\, \delta(x - x_n)\delta(t - t_n)$$

where the nudging gains K_n have to be optimized.

(Sub-)optimal nudging

Several gain matrices are considered :

- K_n is a full rank matrix, $\forall n$: at each point, the correcting term is a linear combination of all forecast/observation misfits at the same time ;
- K_n is a pseudo-diagonal matrix : the model equations are only corrected at observation locations, and the correction coefficient is different for each observation ;
- K_n is a scalar coefficient, just varying with time.

It is also possible to add a linear time interpolation of correcting terms, as a temporal smoothing. And a convolution product with a gaussian function is used as spatial smoothing.

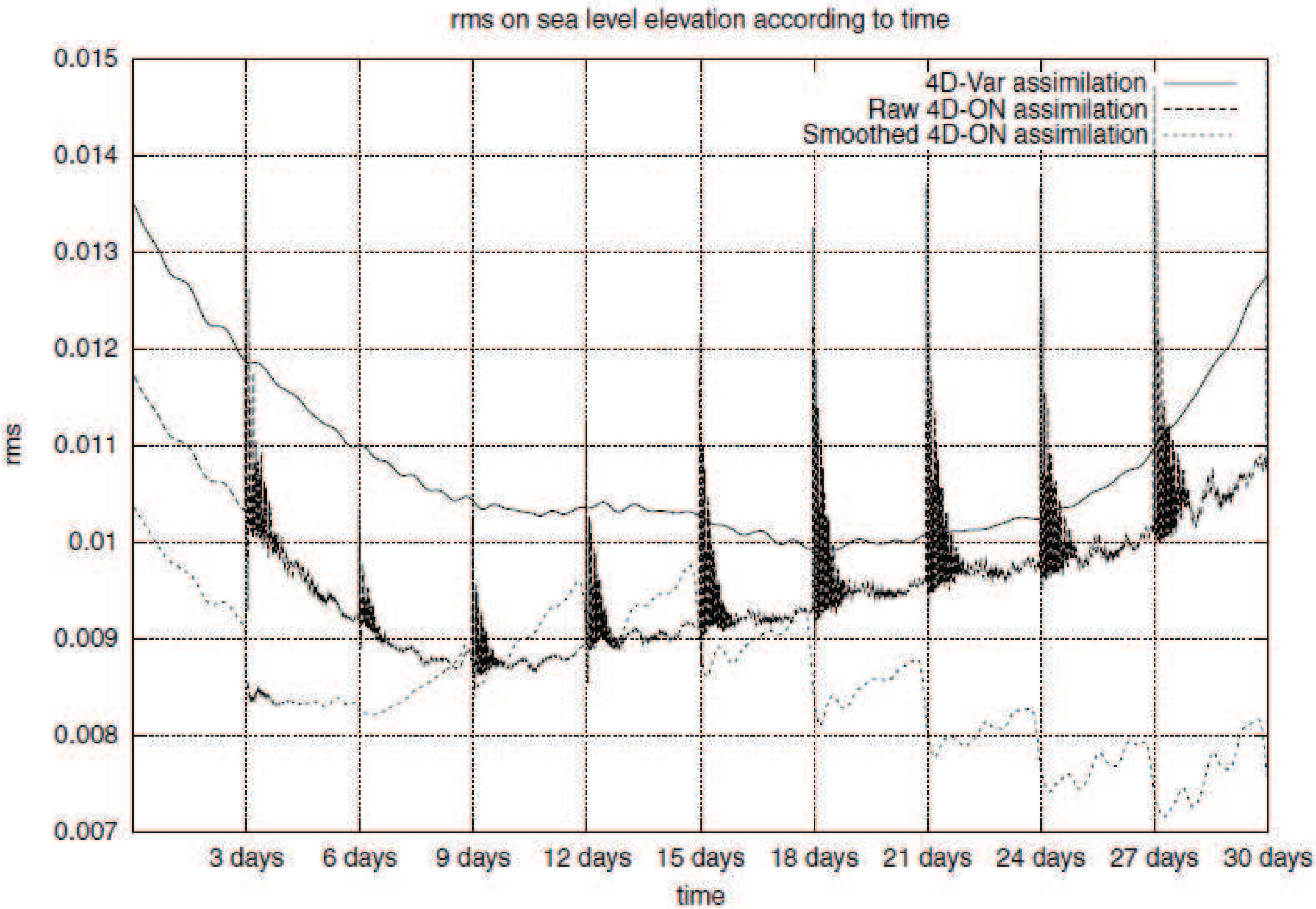
Indeed, the correction of the model only at observation locations is not consistent with the other coordinates, and then gravity waves can be created to return to a more physical state ($\rightsquigarrow$ extreme fluctuations).

(Sub-)(optimal) nudging vs 4DVAR

Form of **G**	Size of control	Raw RMS	Raw CPU	Interpolated RMS	Interpolated CPU
G is a full matrix	$E \times M \times N$	45.06	57.80s	38.49	57.69s
G is a diagonal matrix	$M \times N$	65.30	6.90s	17.57	7.43s
G is a scalar coefficient	N	64.23	9.74s	17.33	3.63s
4D-Var	E	73.73	5.82s		
Without assimilation		248.21			

Norm of the error and CPU time for 4D-VAR and optimal nudging on 1D Burgers experiment. E is the size of the control vector, M is the number of observations at a given time, and N is the number of observation times.

(Sub-)(optimal) nudging vs 4DVAR



RMS of the SSH error for smoothed optimal nudging compared to 4D-VAR and raw optimal nudging ($\rightsquigarrow$ gravity waves)

Talk overview

1. Optimal nudging (use of a variational method)

2. Kalman filter and optimal nudging

3. Comparison optimal nudging - 4D-VAR

⇒ 4. Hybrid nudging - ensemble Kalman filter

Hybridization

Hybrid Nudging - EnKF :

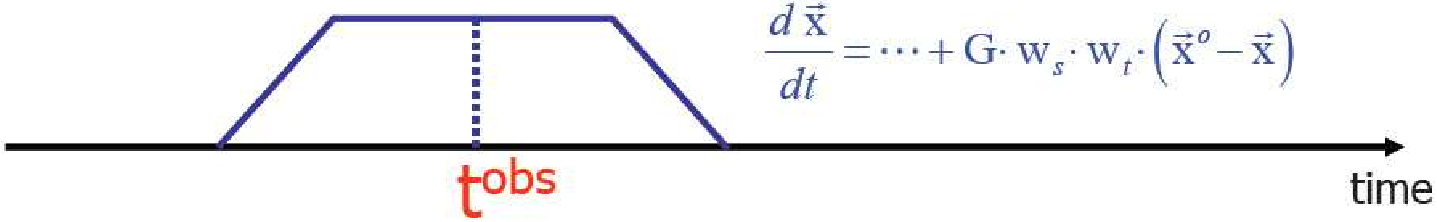
In the Ensemble Kalman filter (EnKF), the background error covariances are computed from an ensemble forecast. Like the standard KF, it performs an analysis at each observation time and switches back to standard model integration between analysis times. This cycle often causes discontinuities or error spikes between the observation times.

The EnKF analyses can be improved by applying more continuously in time the corrections. The idea is to use nudging-type terms to gradually apply the correction terms in time, in order to minimize the insertion of shocks.

Hybrid nudging - EnKF

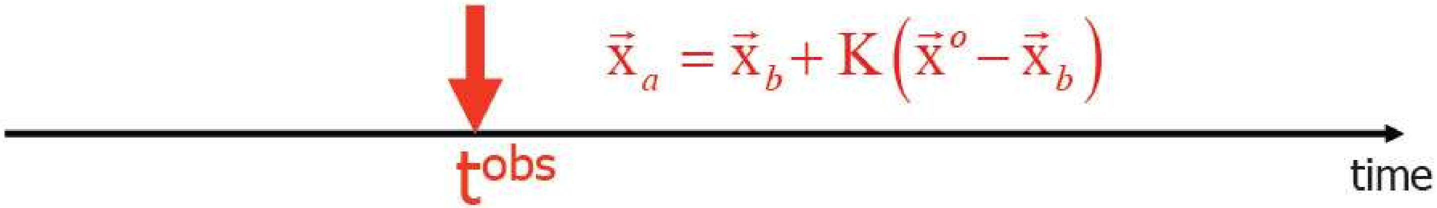
Nudging:

$$\frac{d\vec{x}}{dt} = \cdots + G \cdot w_s \cdot w_t \cdot \left(\vec{x}^o - \vec{x}\right)$$



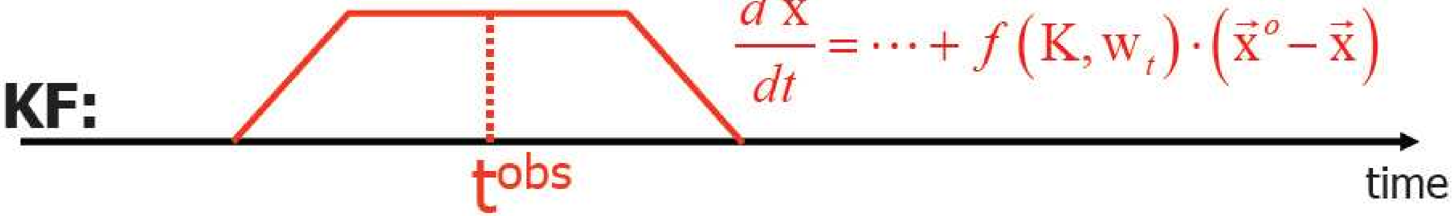
EnKF:

$$\vec{x}_a = \vec{x}_b + K\left(\vec{x}^o - \vec{x}_b\right)$$

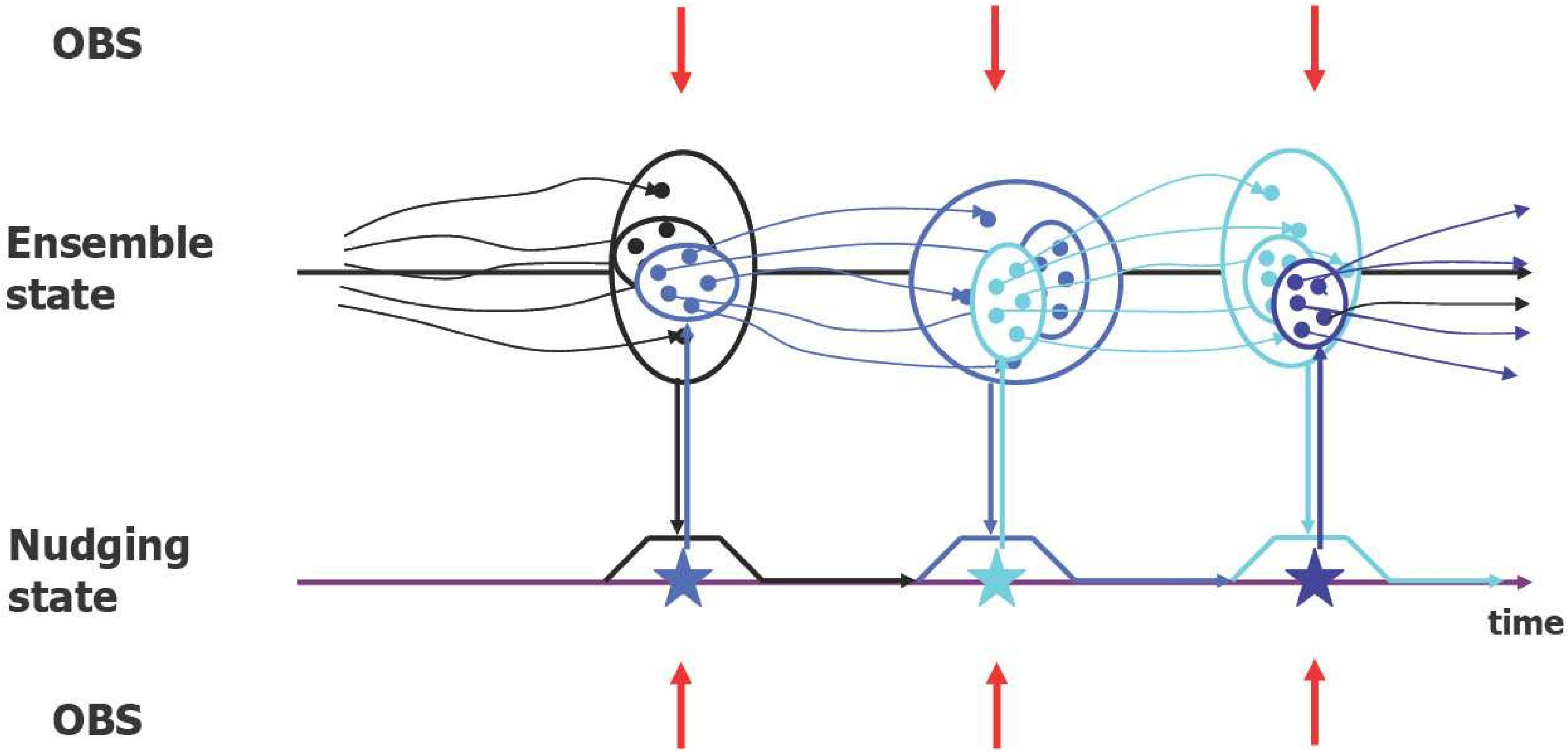


Hybrid nudging-EnKF:

$$\frac{d\vec{x}}{dt} = \cdots + f\left(K, w_t\right) \cdot \left(\vec{x}^o - \vec{x}\right)$$

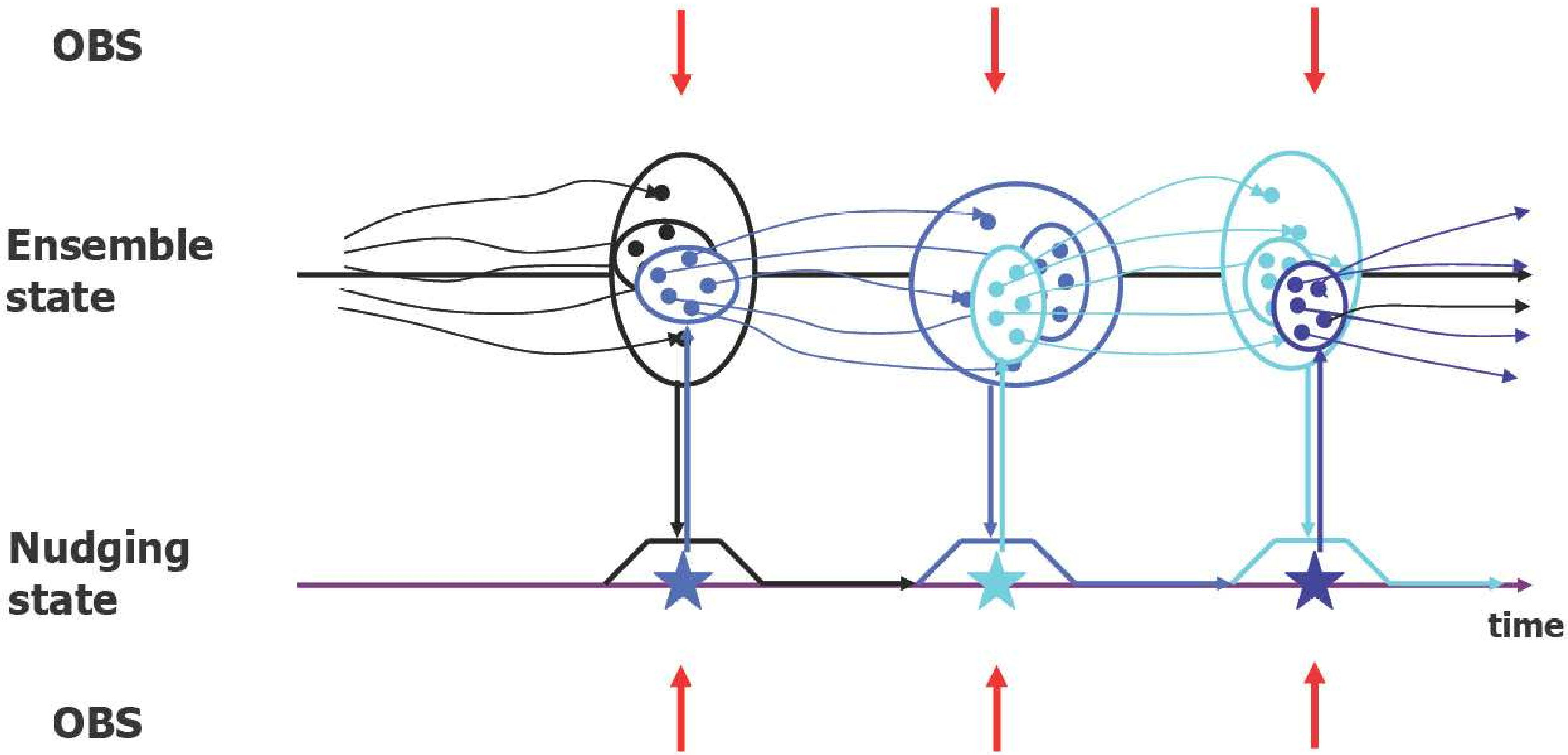


Hybrid nudging - EnKF



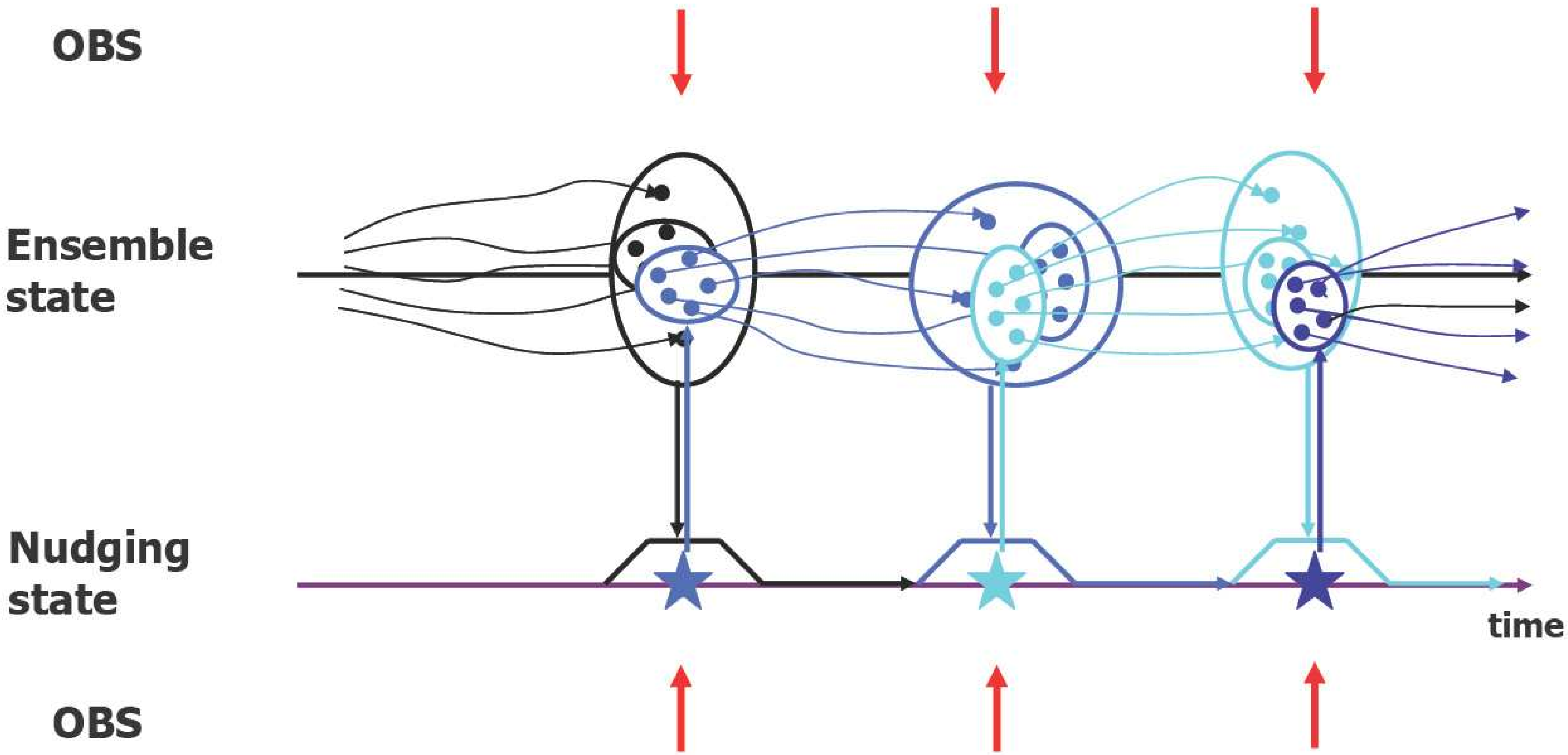
1. Compute the nudging coefficients using the EnKF gain matrix (use the EnKF computed gain matrix, and add the regularization and spreading in time).

Hybrid nudging - EnKF



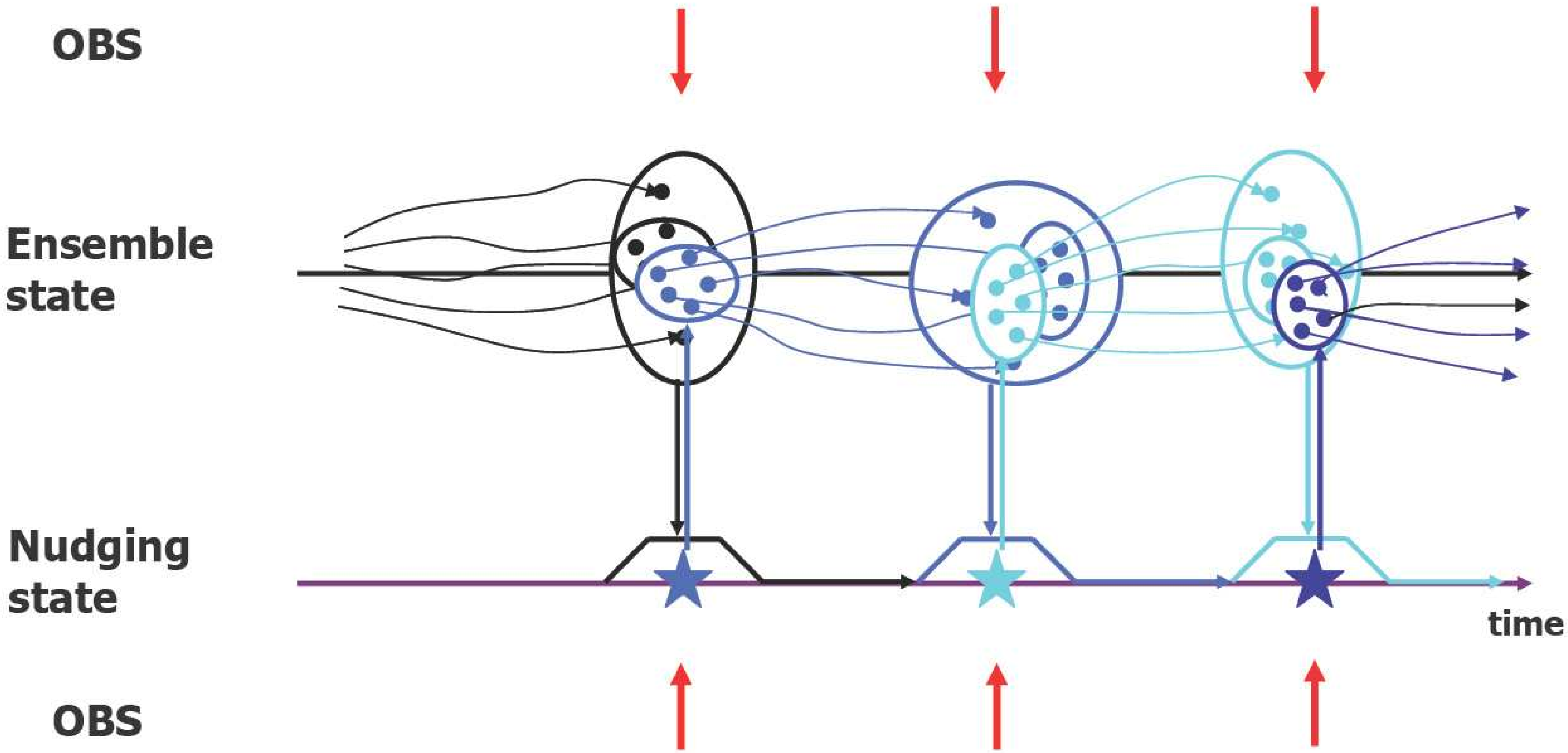
2. Compute the nudging state by continuously applying nudging with these coefficients.

Hybrid nudging - EnKF



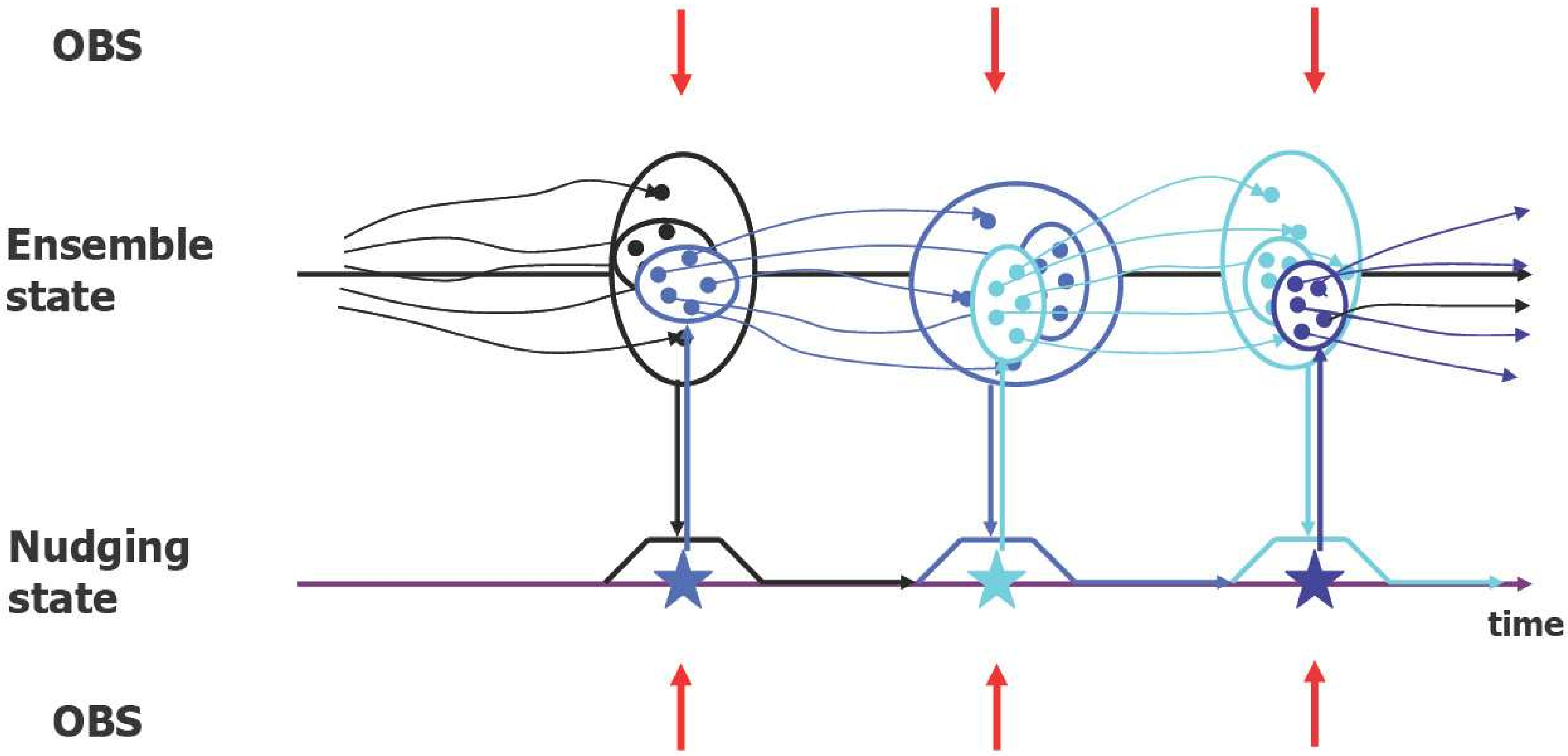
3. Update each ensemble member using the EnKF.

Hybrid nudging - EnKF



4. Shift the ensemble from its ensemble mean to the analysis of the nudging state while retaining the ensemble spread.

Hybrid nudging - EnKF



5. Integrate the ensemble and nudging state forward to the next analysis time.

Hybrid nudging - EnKF

In [Lei-Stauffer (AMS Conf Proc. 2009)], experiments on Lorenz and shallow-water models : similar (or even better) results as the EnKS, for a reduced computational cost (Nudging-EnKF cost similar to that of the EnKF).

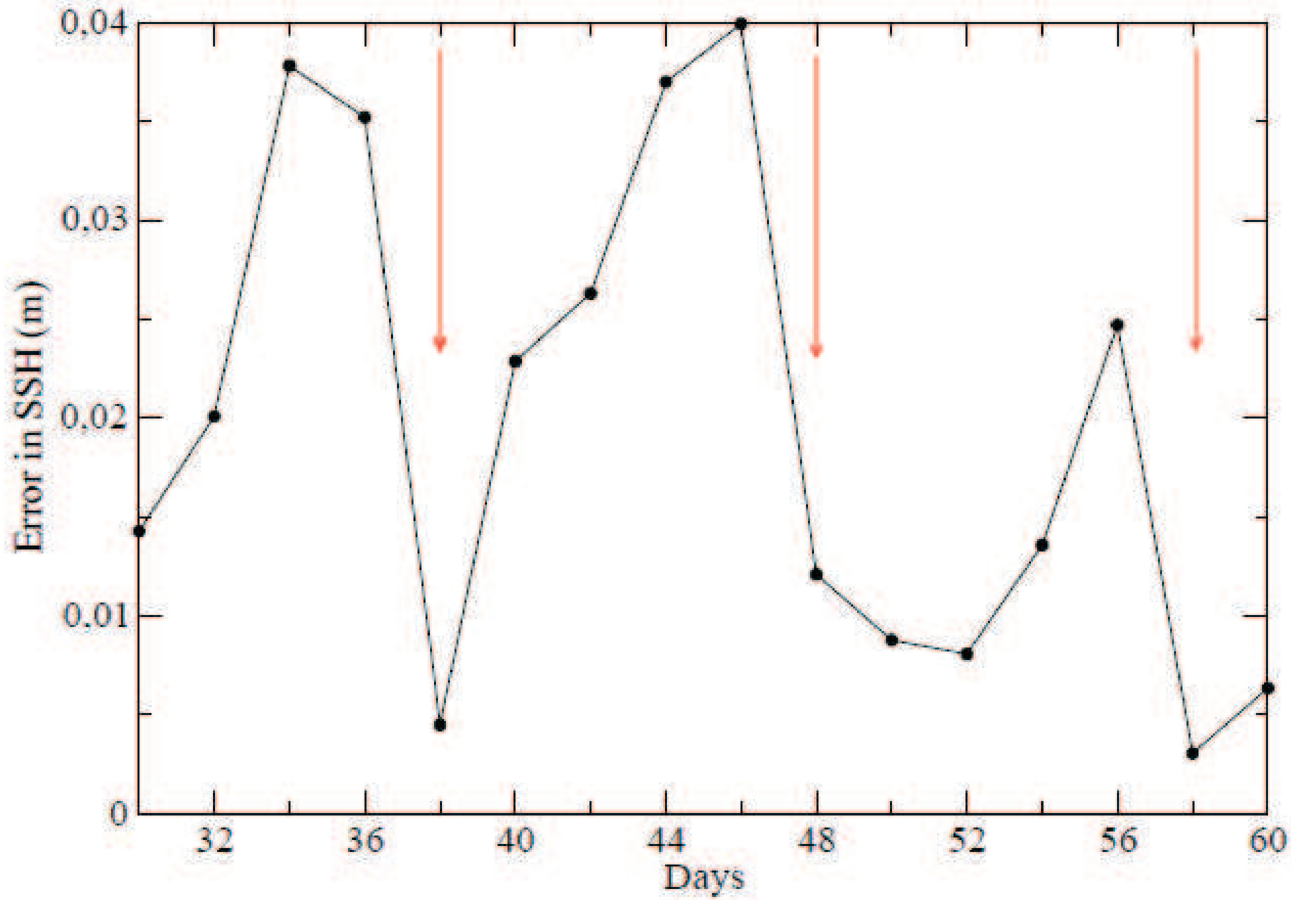
The hybrid nudging-EnKF retains the spatial (flow-dependent) error correlation weighting function from the EnKF and the gradual corrections of the continuous nudging approach (digital filter unnecessary) to avoid the strong corrections and discontinuities (error spikes) at the analysis steps.

In the hybrid nudging-EnKF, the model equations assist in the data assimilation process.

Comparison with Kalman smoother

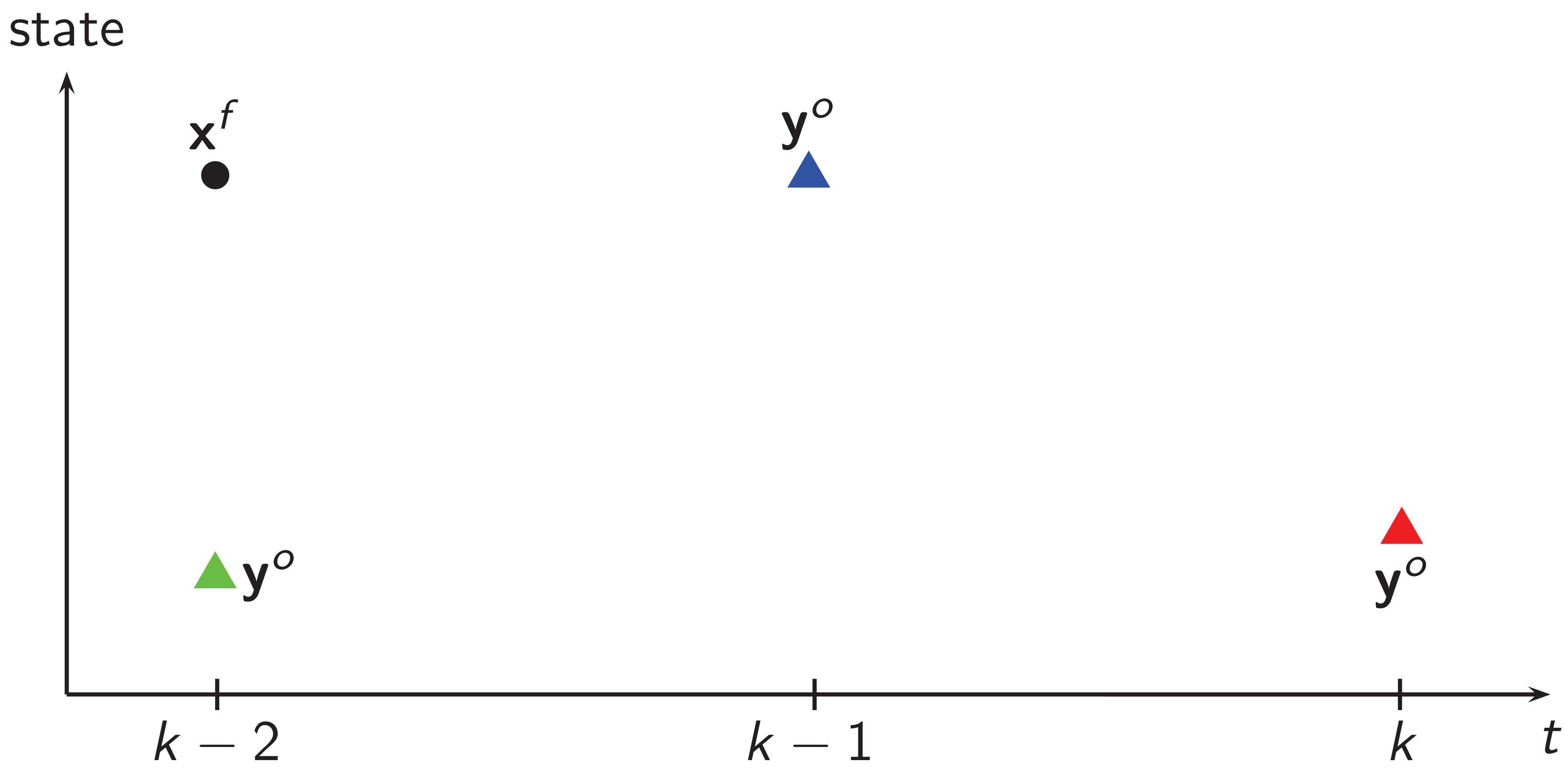
Use of future observation in Kalman filter analyses :

The idea is to incorporate the future observations in the filtered estimate, and not only the past information.

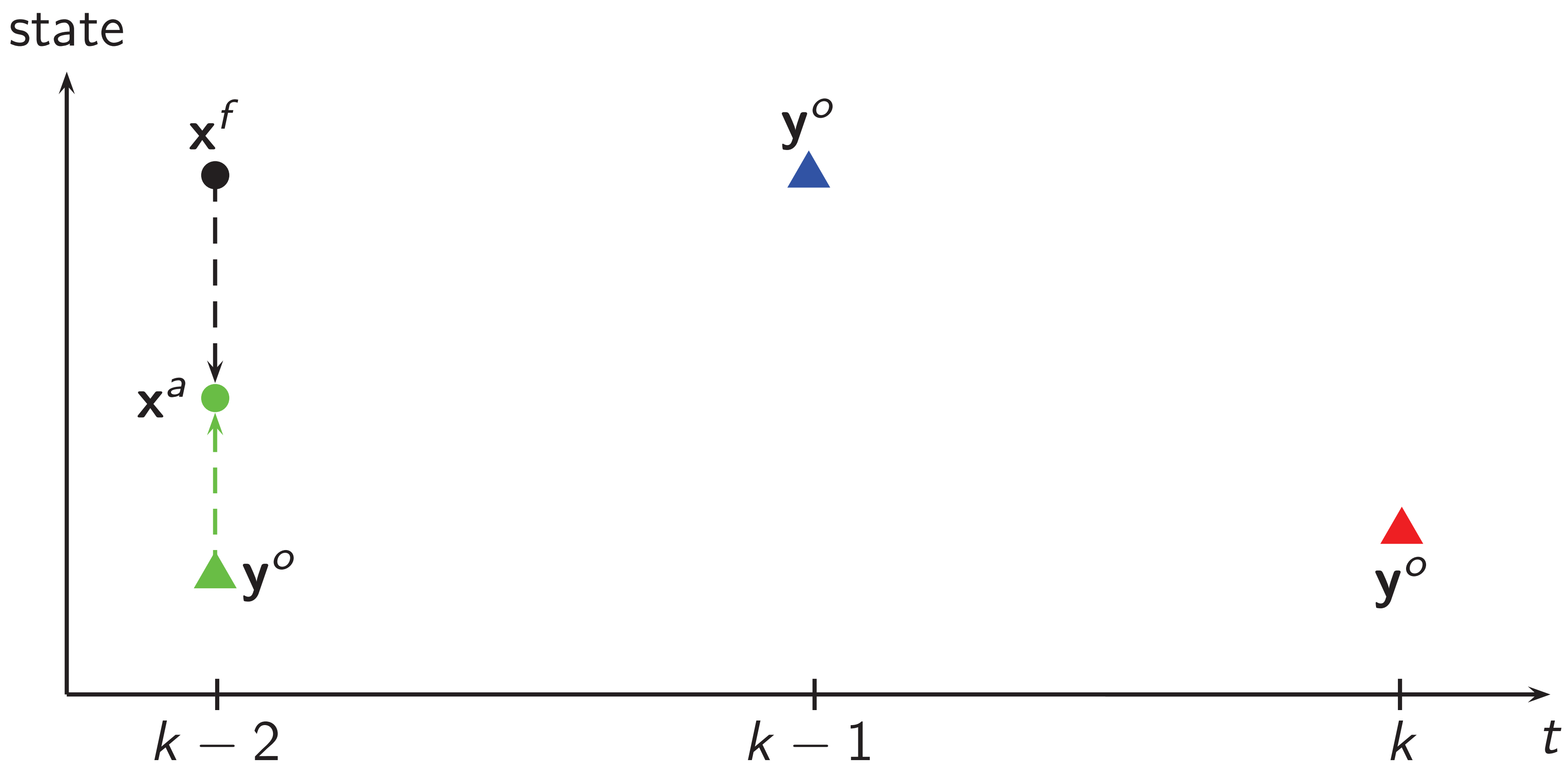


Might also reduce the shocks in the trajectory.

Kalman Smoother :

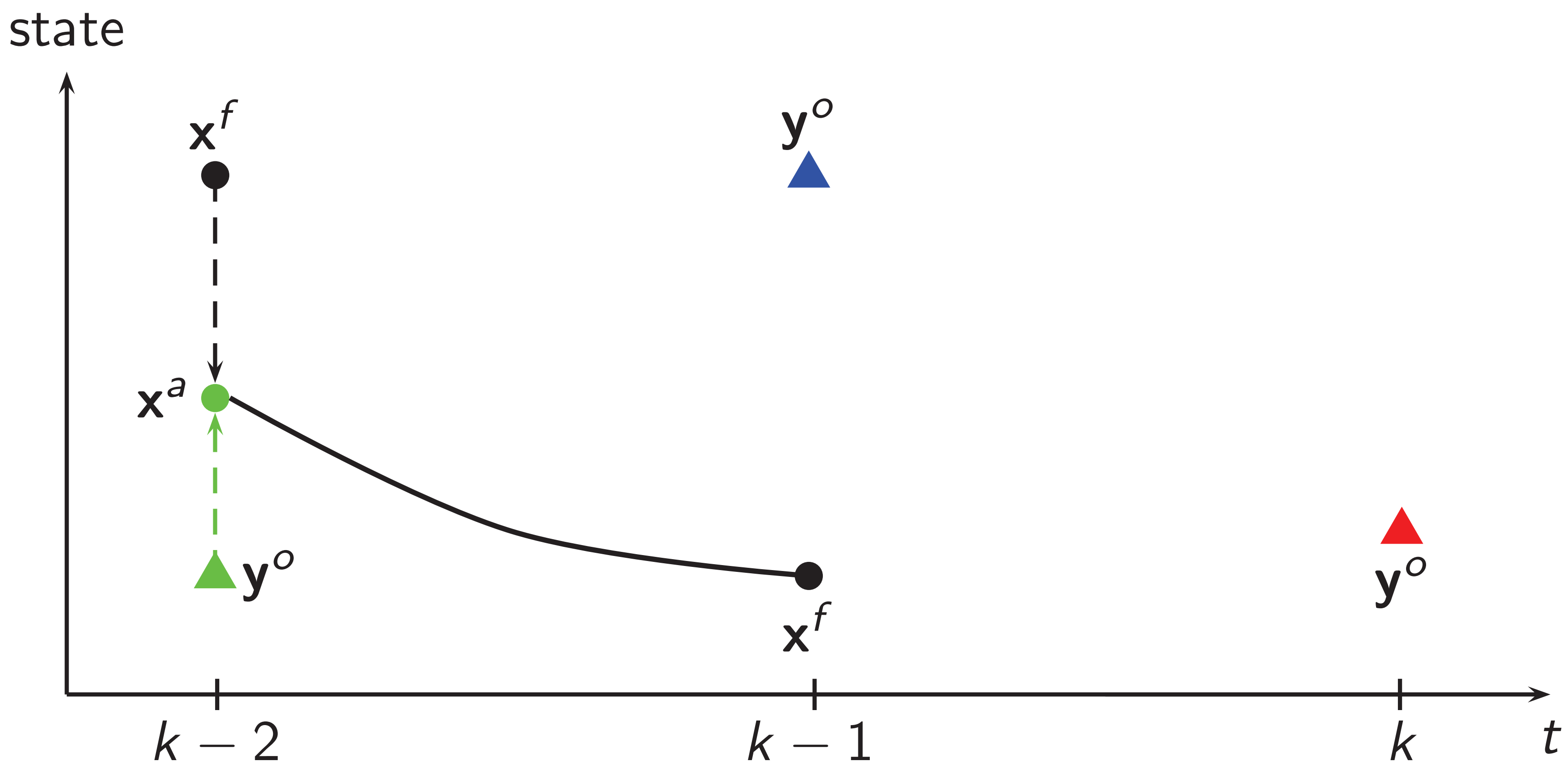


Kalman Smoother :

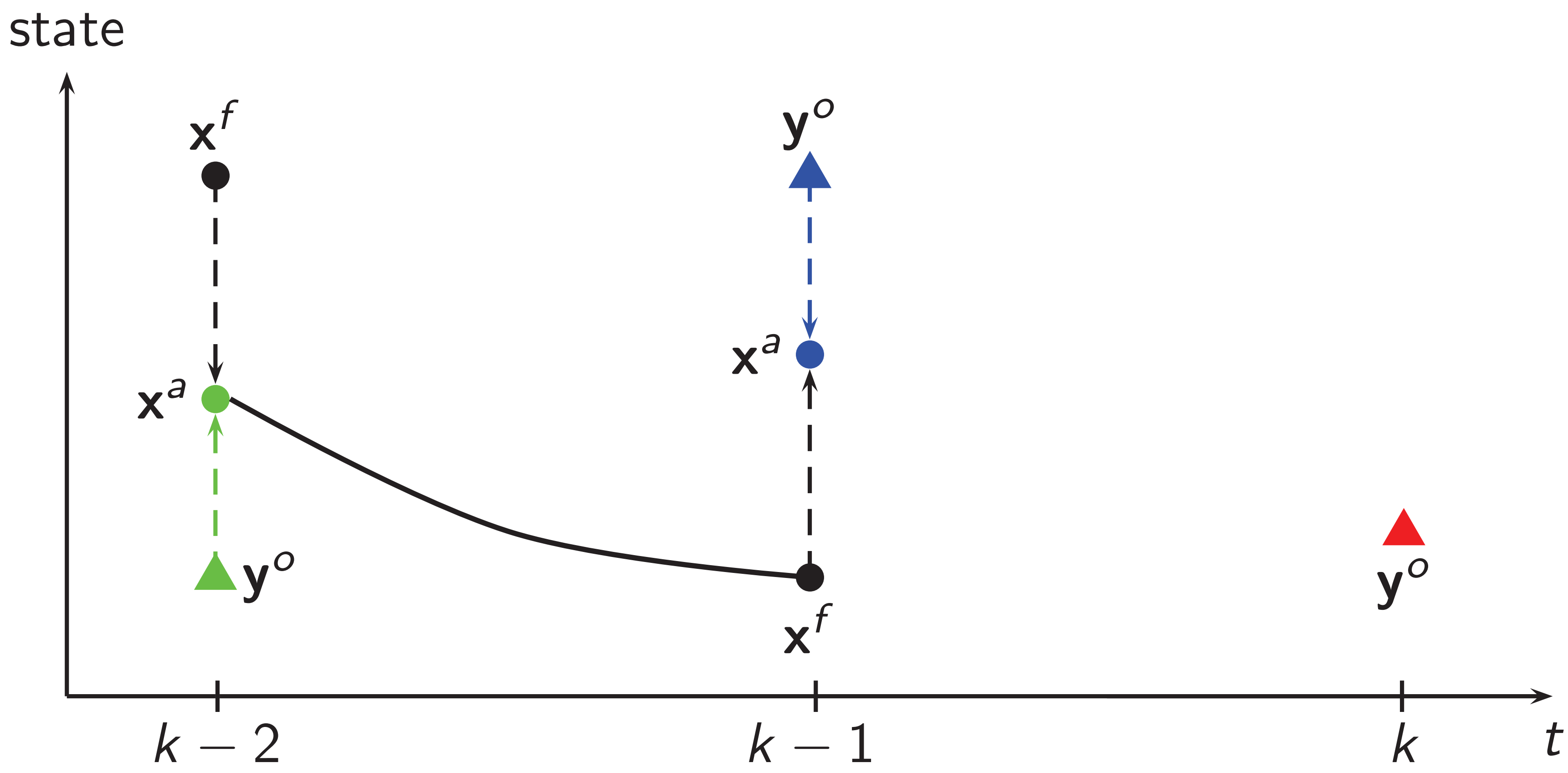


Kalman smoother

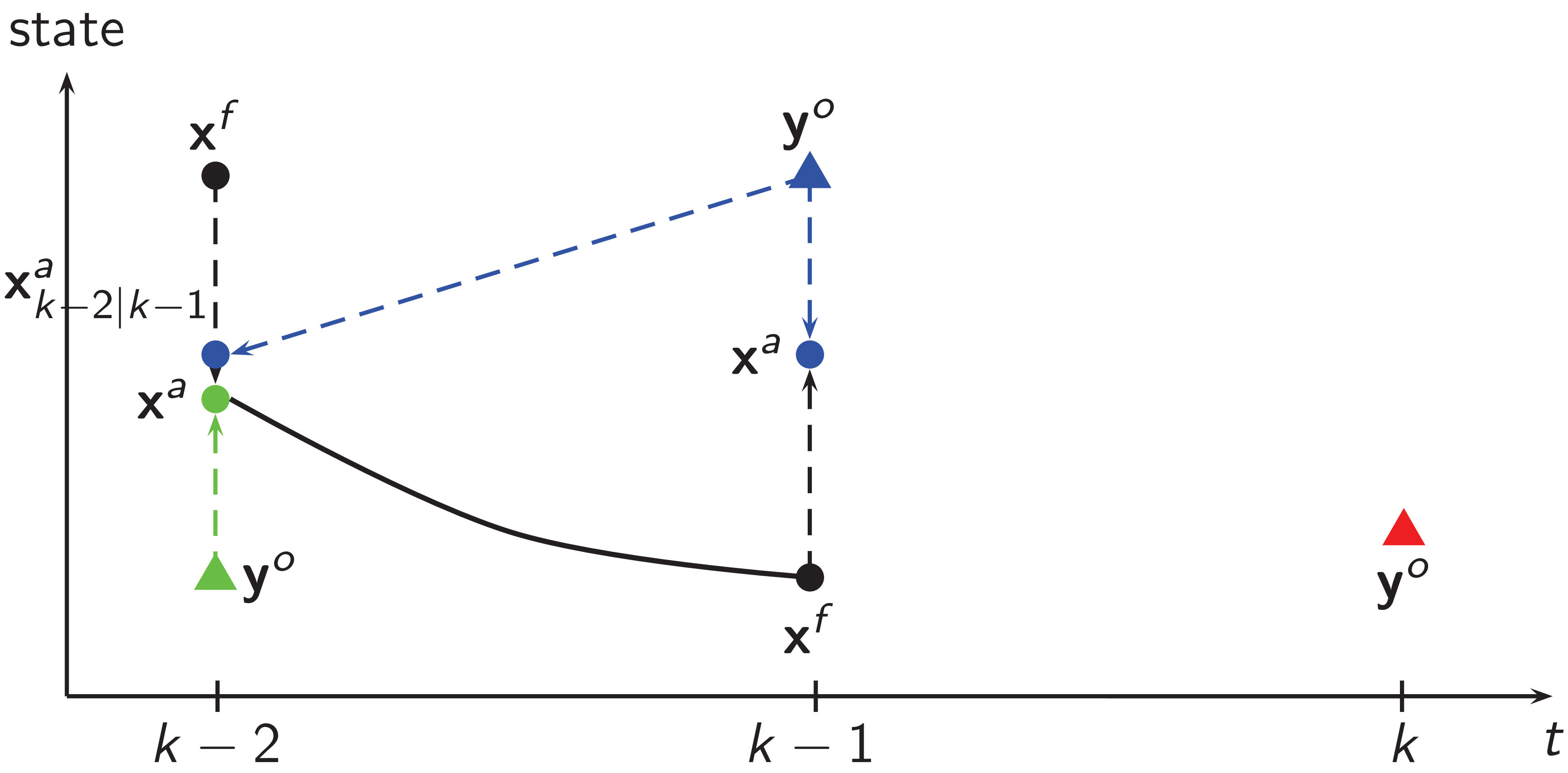
Kalman Smoother :



Kalman Smoother :

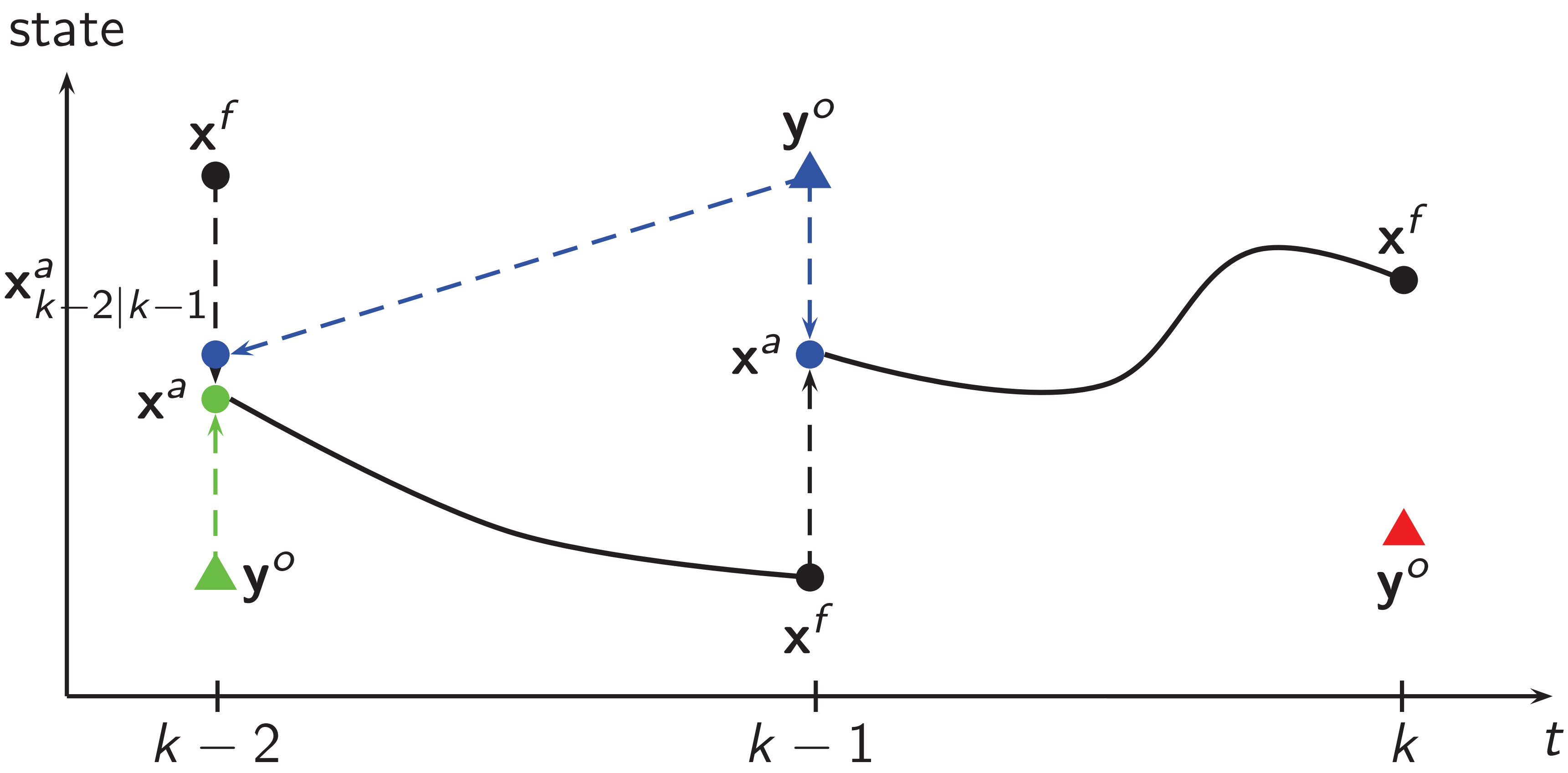


Kalman Smoother :

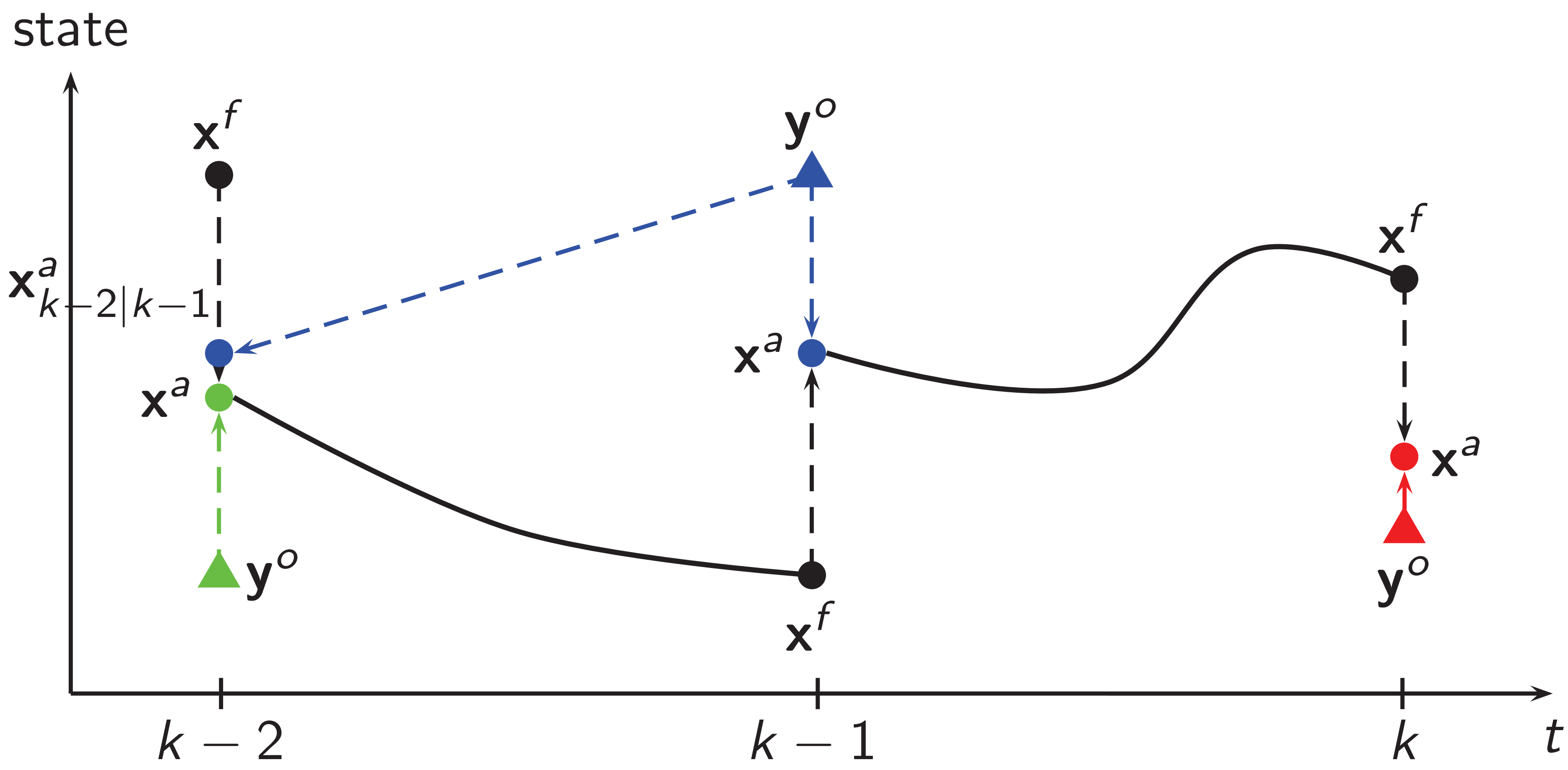


Kalman smoother

Kalman Smoother :

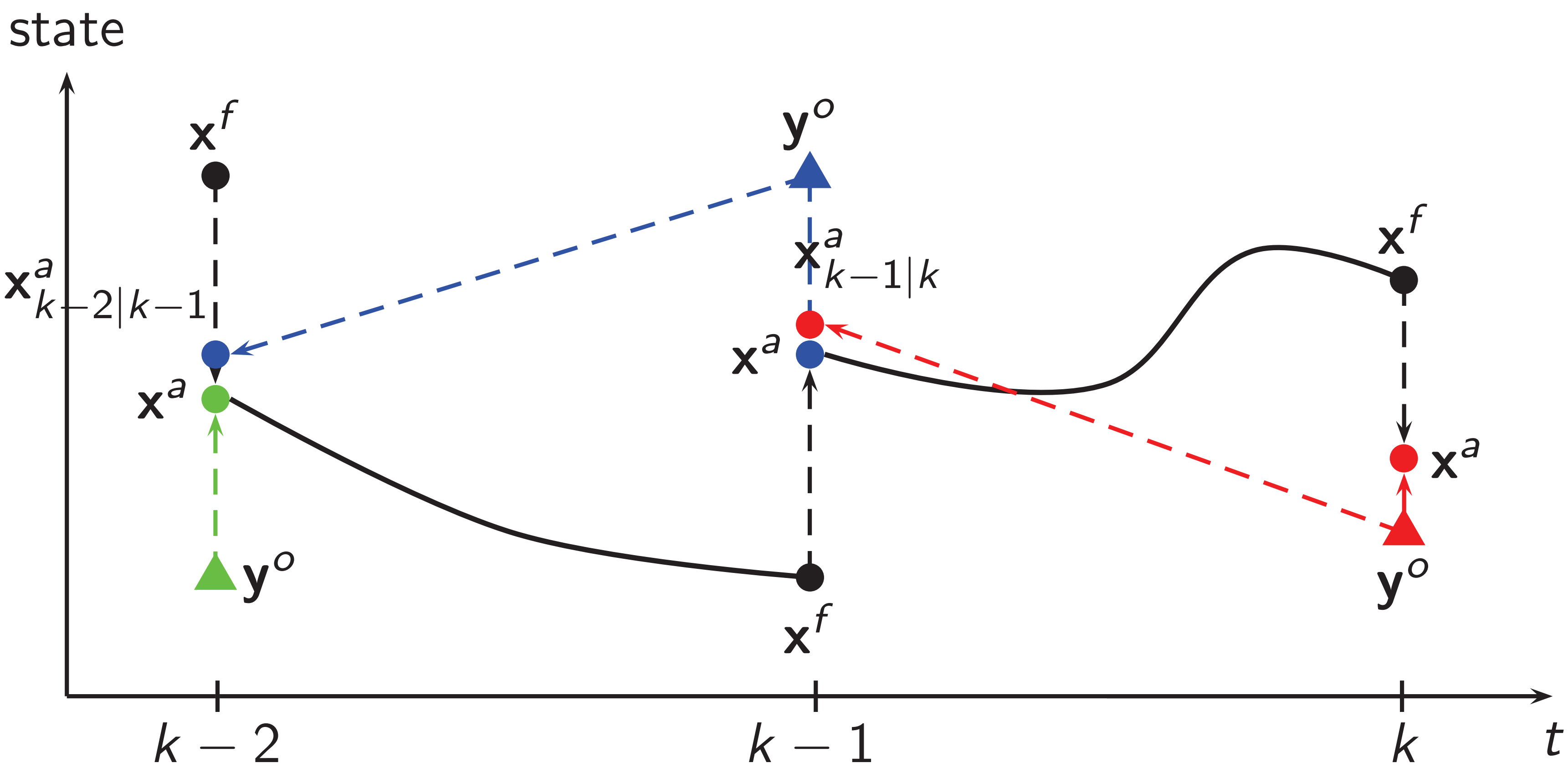


Kalman Smoother :



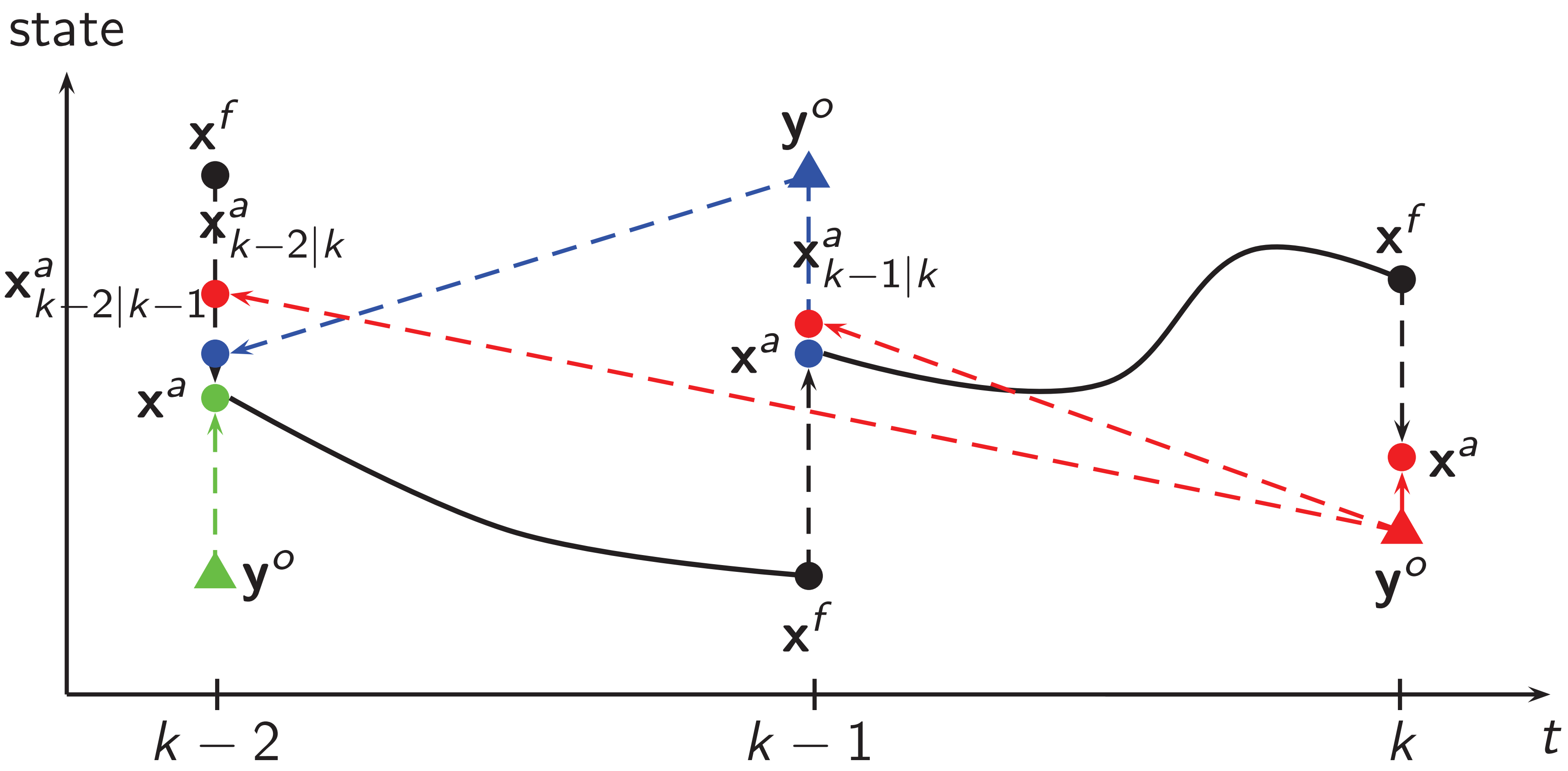
Kalman smoother

Kalman Smoother :

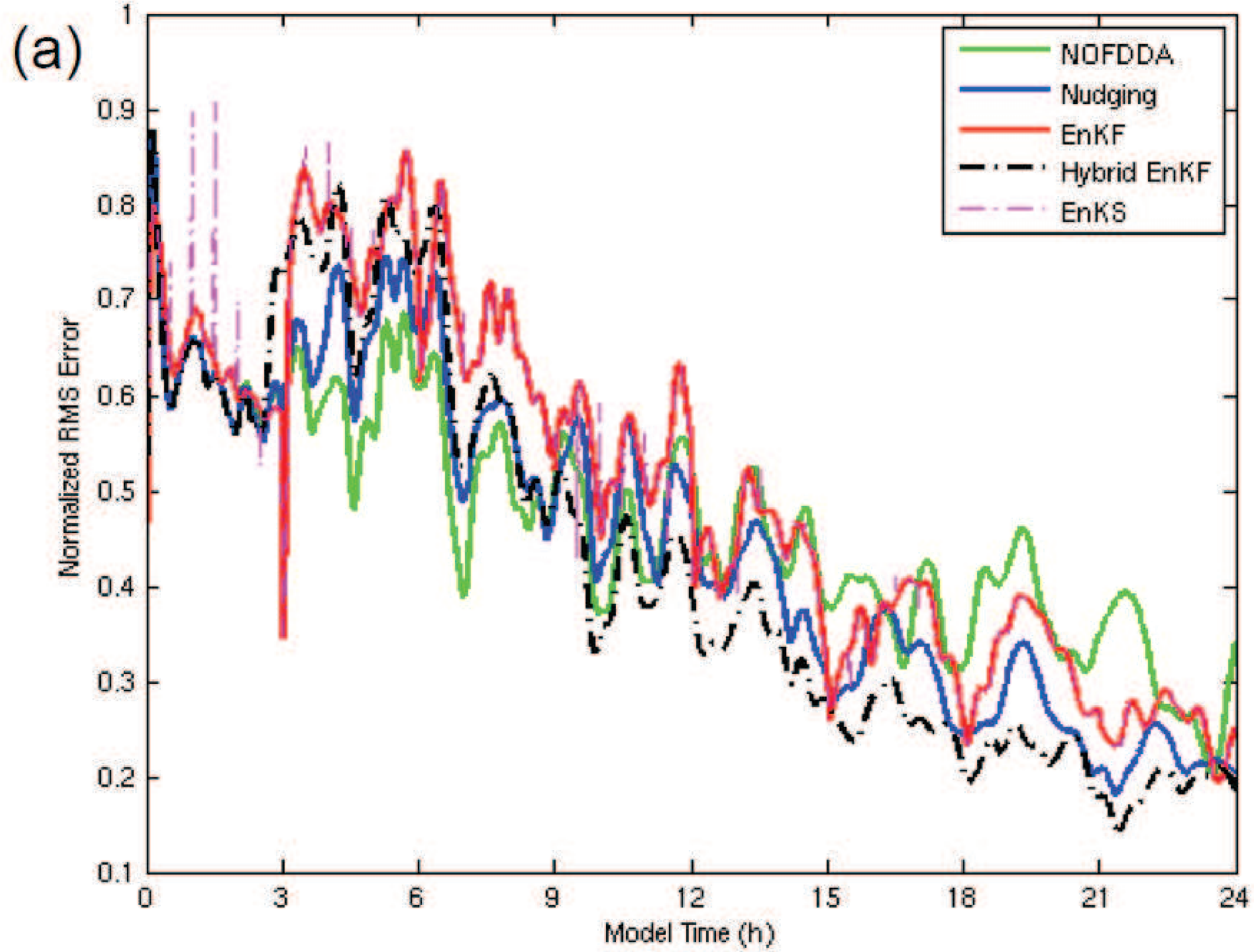


Kalman smoother

Kalman Smoother :

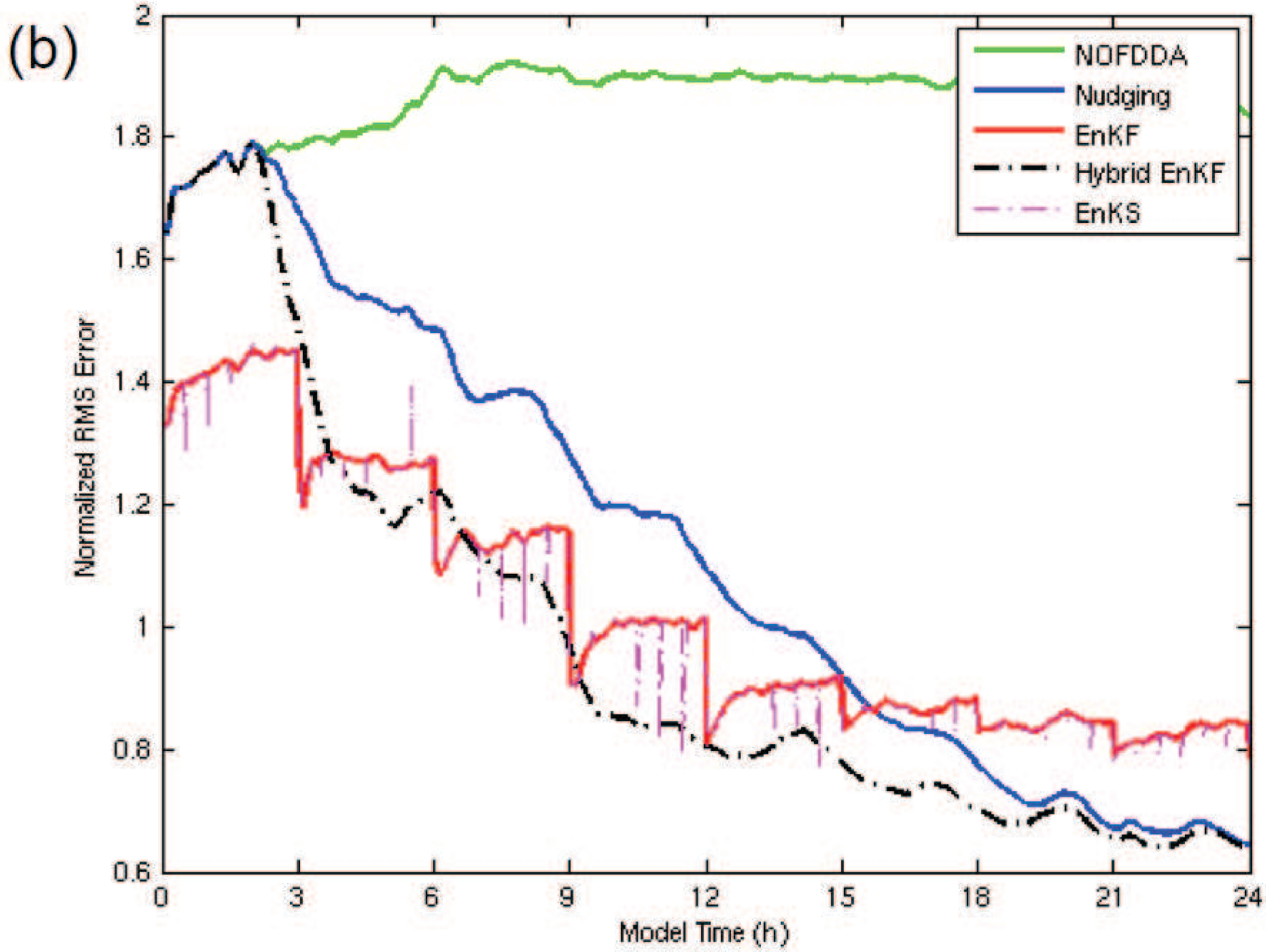


Comparison hybrid nudging-EnKF



RMS error for the height field (2D shallow water) : no data assimilation (green), nudging (blue), EnKF (red), hybrid nudging-EnKF (black), EnKS (pink).

Comparison hybrid nudging-EnKF



RMS error for the wind field (2D shallow water) : no data assimilation (green), nudging (blue), EnKF (red), hybrid nudging-EnKF (black), EnKS (pink).

Conclusions

Conclusions :

- Optimal nudging is a very efficient four-dimensional data assimilation scheme
- Similar or better results than 4D-VAR or Kalman filter
- However : the optimization of the coefficients is expensive
- Hybridization with filters or smoothers in order to prevent shocks