



Introducción asimilación de datos Sesión práctica

Rodolfo Alvarado Montero

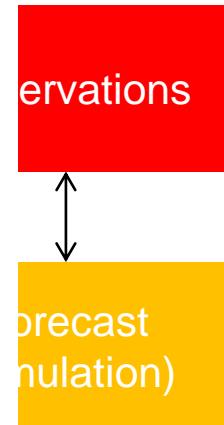
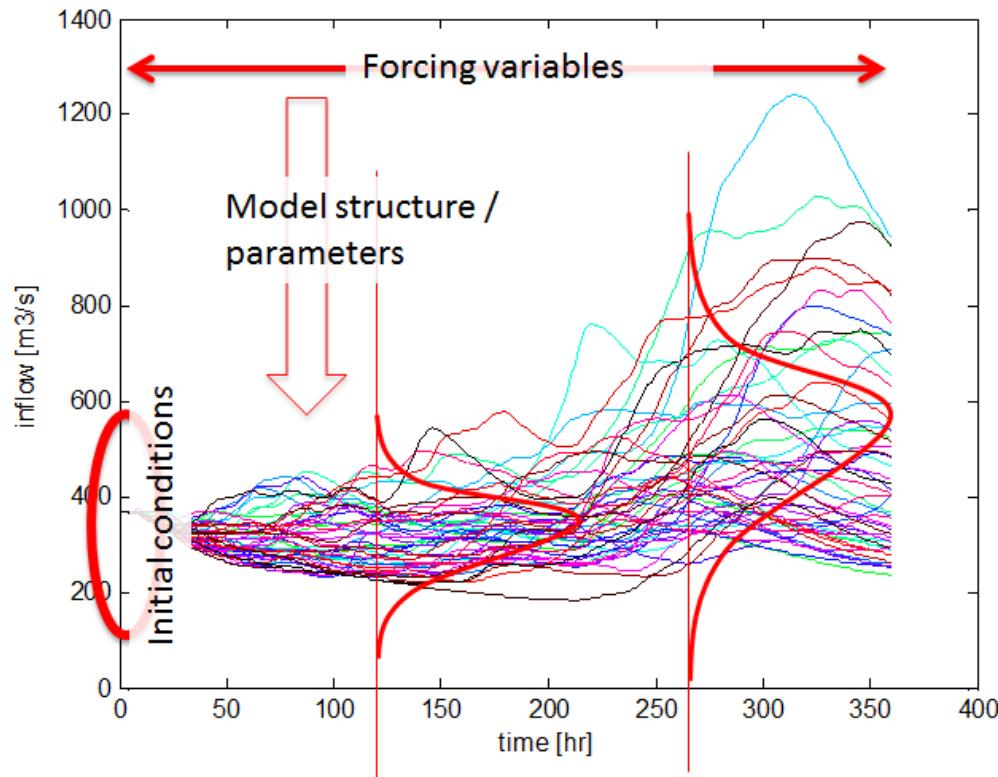
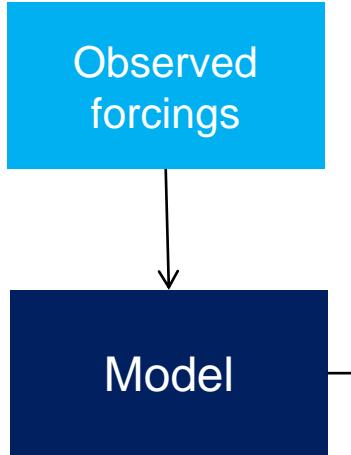
Días de usuarios Delft-FEWS Latinoamérica
16 december 2019



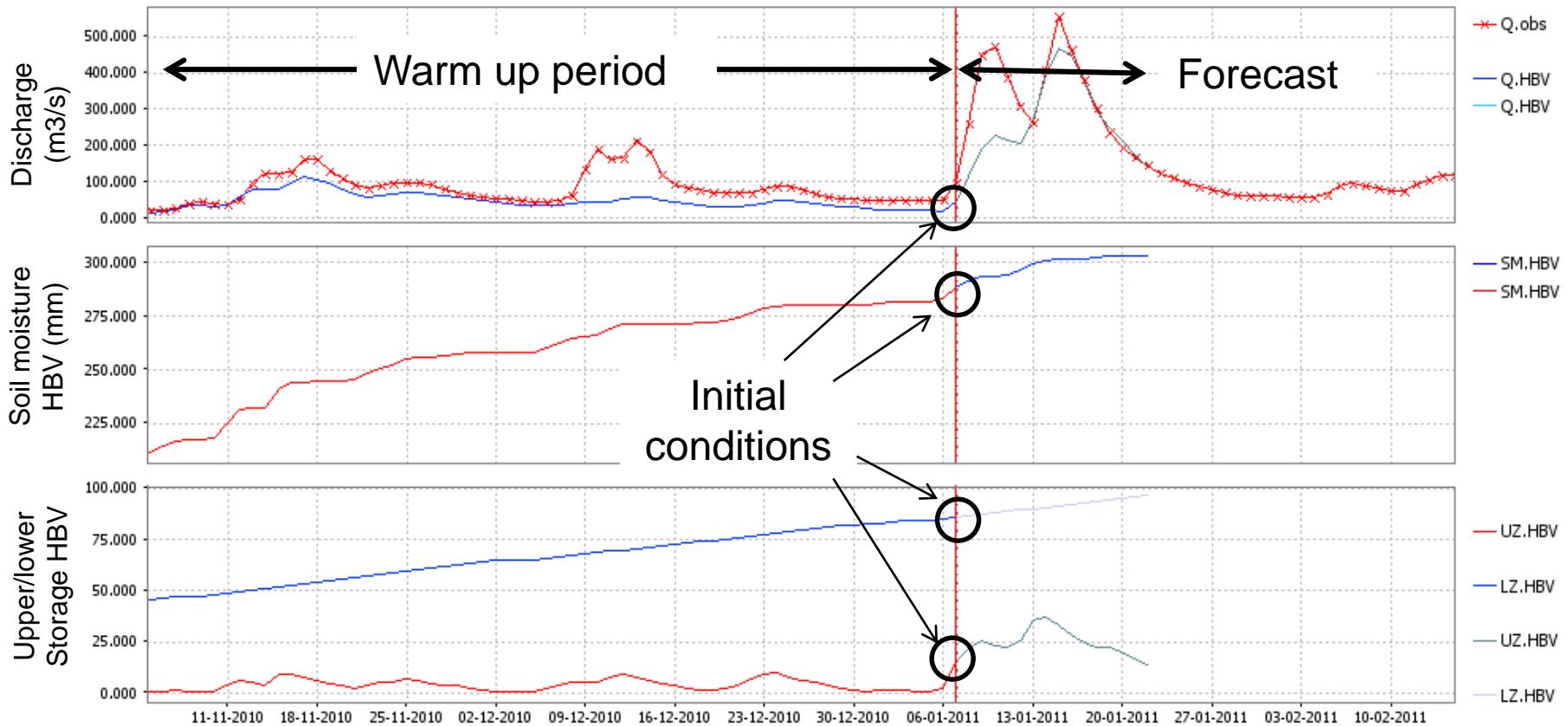
Asimilación de datos

„.... procedures that aim to produce physically consistent representations or estimates of the dynamical behavior of a system by merging the information present in imperfect models and uncertain data in an optimal way to achieve uncertainty quantification and reduction.“ (Liu and Gupta, 2007)

Cadena de pronóstico



Cadena de pronóstico



Fuentes de errores



En las observaciones:

- Representatividad de los datos
- Precisión de sensores
- Transformación de variable observada (ejemplo: curva clave / rating curve, reflectividad a humedad de suelo, etc)
- ...

En las simulaciones:

- Representación del proceso físico
- Calibración de modelos
- Condiciones iniciales del modelo
- Datos de entrada al modelo
- ...



Fuentes de errores



Cómo cuantificar la incertidumbre?

Incertidumbre
meteorológica



Variable forzantes
probabilísticas

Incertidumbre
modelación



Usar varios
modelos

Incertidumbre
cond. iniciales



Asimilación de
datos

Pregunta 1



Cuál es la humedad actual en la cuenca?

Valor observado: 200

Valor simulado: 300

Depende de la
incertidumbre
de cada uno



Sin más
información:
250

Pregunta 2



Cuál es la humedad actual en la cuenca si el valor observado es 3 veces más importante que el valor simulado?

Valor observado: 200

Valor simulado: 300

$$\text{Valor verdadero} = (200*3 + 300*1) / (3+1) = 225$$

Pregunta 3



Cuál es la humedad actual en la cuenca dada la incertidumbre (varianza) dada del valor observado y simulado?

Valor observado: 200, varianza = 20

Valor simulado: 300, varianza = 30

Nota: recordar que la varianza es el cuadrado de la desviación standard

$$\text{Valor verdadero} = (200 \cdot 1/20 + 300 \cdot 1/30) / (1/20 + 1/30) = 240$$

Qué hicieron?



Mínimos cuadrados

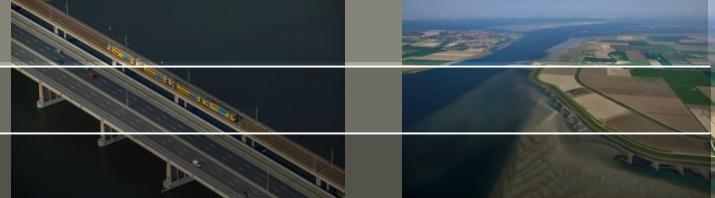
$$\hat{x} = \frac{\frac{x_1}{\sigma_1^2} + \frac{x_2}{\sigma_2^2}}{\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}}$$

$$\hat{x} = \frac{x_1 \sigma_2^2 + x_2 \sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

$$\hat{x} = x_1 + K(x_2 - x_1) \quad K = \sigma_1^2 / (\sigma_1^2 + \sigma_2^2)$$

Raul Rojas. *The Kalman Filter*; <http://www.inf.fu-berlin.de/lehre/SS05/Robotik/kalman.pdf>

Qué hicieron?



Bayesian methods: $P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)}$

Kalman filter:

$$J = \frac{(True - Obs)^2}{\sigma_{obs}^2} + \frac{(True - Model)^2}{\sigma_{Model}^2}$$

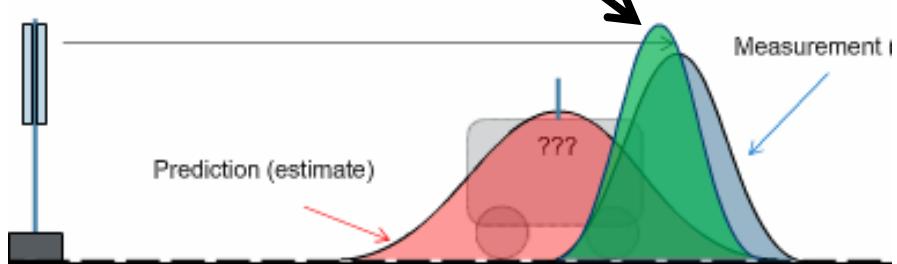
Objective function!

$$True = Model + K \cdot (Obs - Model)$$

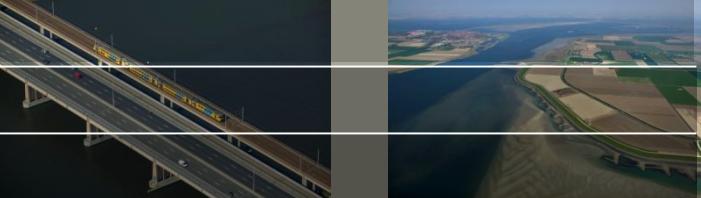
Improved estimate!

$$K = \frac{\sigma_{Model}^2}{(\sigma_{Model}^2 + \sigma_{obs}^2)}$$

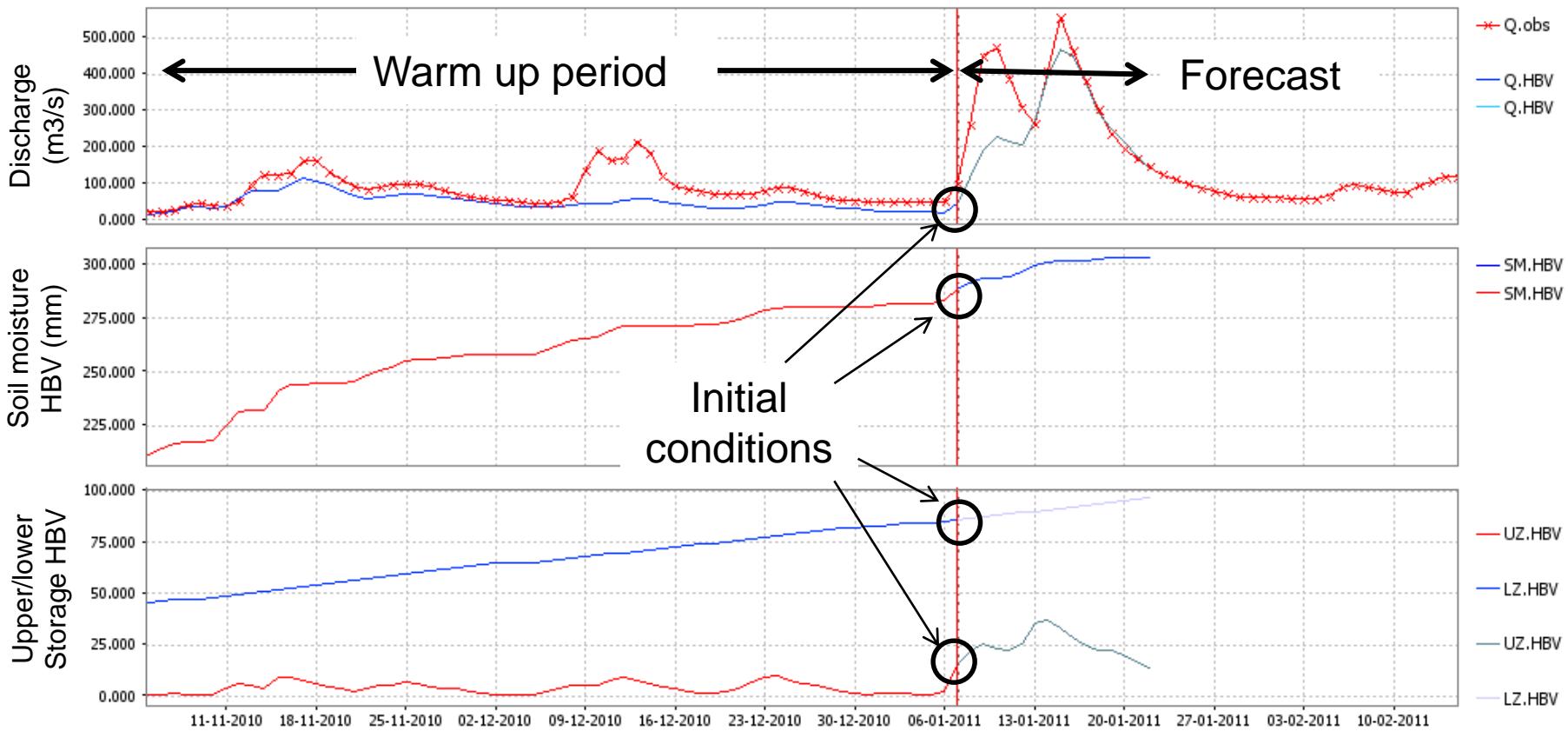
... estimates by merging models
and data in an optimal way to
achieve uncertainty quantification
and reduction



Problema?

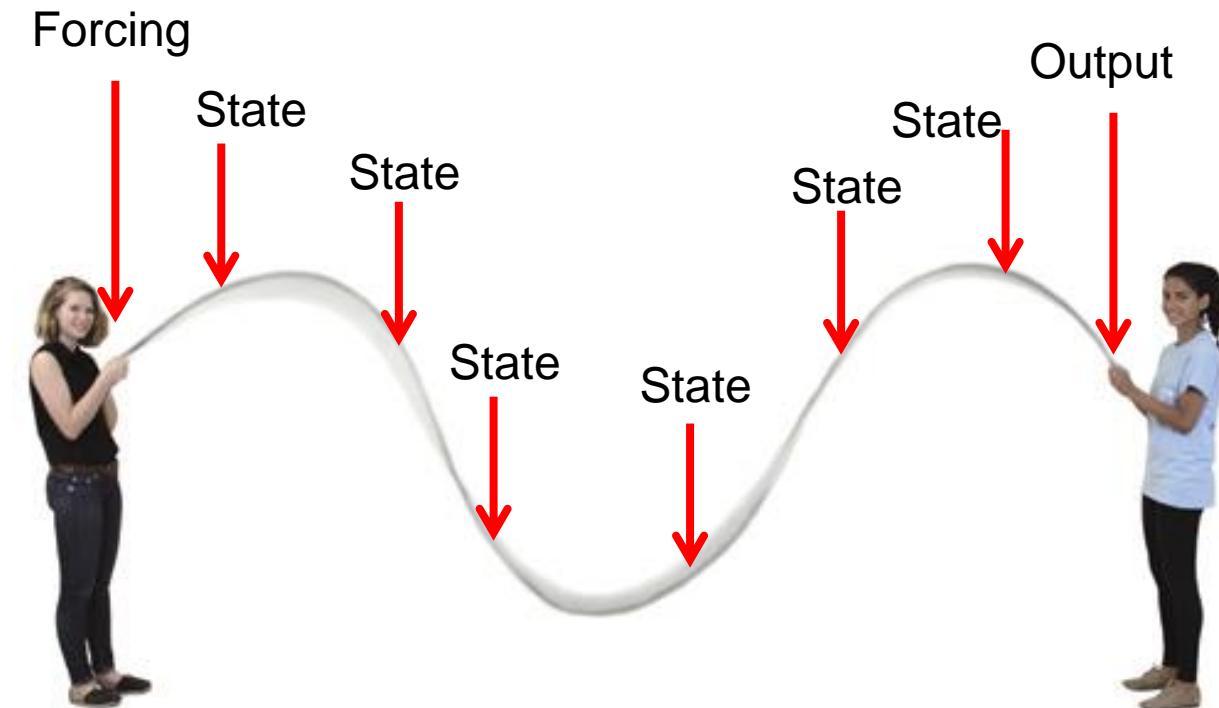


No observamos todas las variables de estado de un modelo hidrológico!

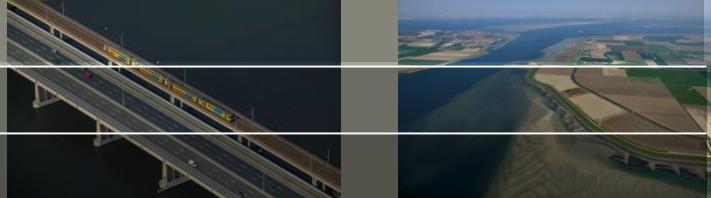




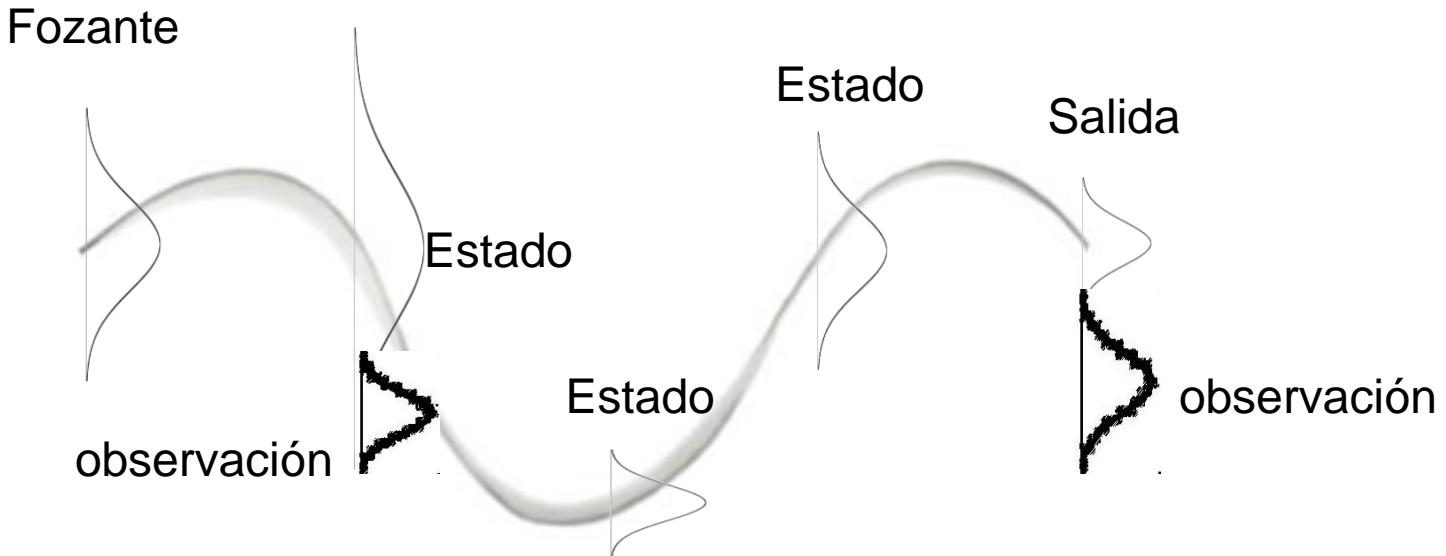
Modelo



Cómo calcular la covarianza?



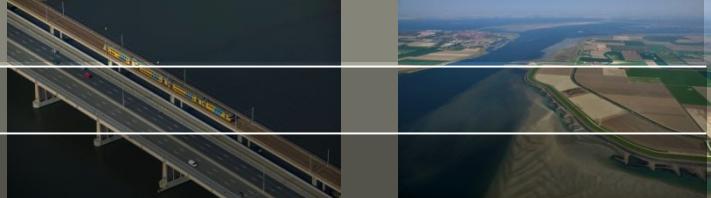
- EnKF estima la covarianza del modelo mediante la perturbación de variables forzantes del modelo y tomando muestras de las variables de estado



- Luego aplicar las ecuaciones de filtro de Kalman

$$True = Model + K \cdot (Obs - Model)$$

Pregunta 4



Calcular la covarianza del siguiente modelo:

Caudal	Humedad
1100	200
1050	220
900	210
1250	190

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$



La ganancia de Kalman (K), es ahora:

$$K = P^*H / (H^*P^*H + R)$$

Luego:

$$\text{Verdadero} = \text{Modelo} + K * (\text{Observado} - H^* \text{Modelo})$$

Pregunta 5



Cuál es la humedad actual (variable de estado) en la cuenca dado el valor observado y simulado de caudal?

Depende de la
correlación entre
caudal y humedad

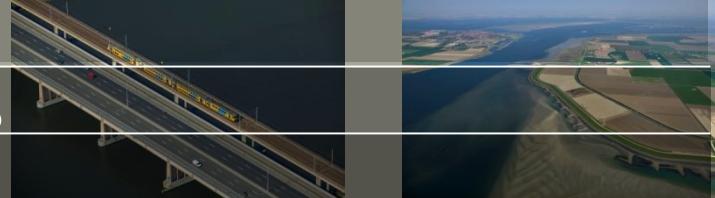


Covarianza del
modelo

Valor caudal observado: 1200, varianza observada: 50

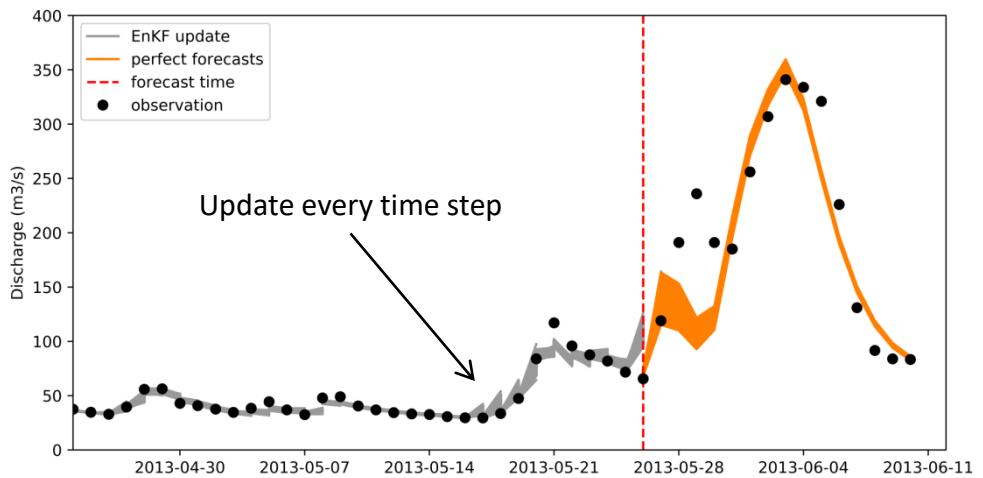
Valor humedad simulada: 200

Tipos de asimilación de datos



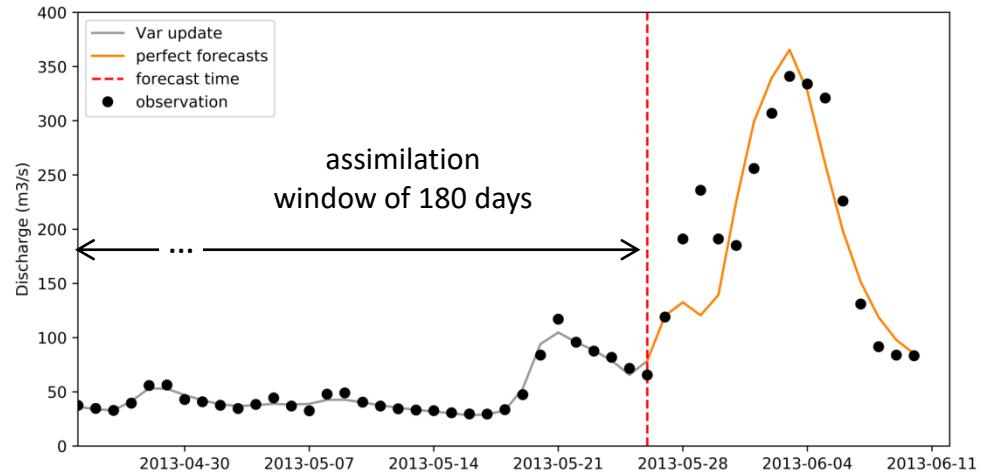
Sequential EnKF

Standard deviation Of 2.5%
Qobs, using 200 members



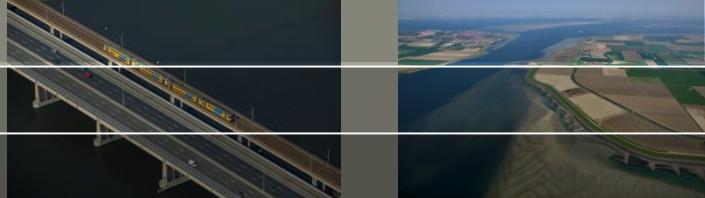
Variational MHE-DA

Deterministic forecast using
observation weight of 1.0



R. Alvarado-Montero, D. Schwanenberg, P. Krahe, D. Lisniak, A. Sensoy, A. Sorman, B. Akkol. Moving horizon estimation for assimilating H-SAF remote sensing data into the HBV hydrological model

Tipos de asimilación de datos



Ensemble KF:

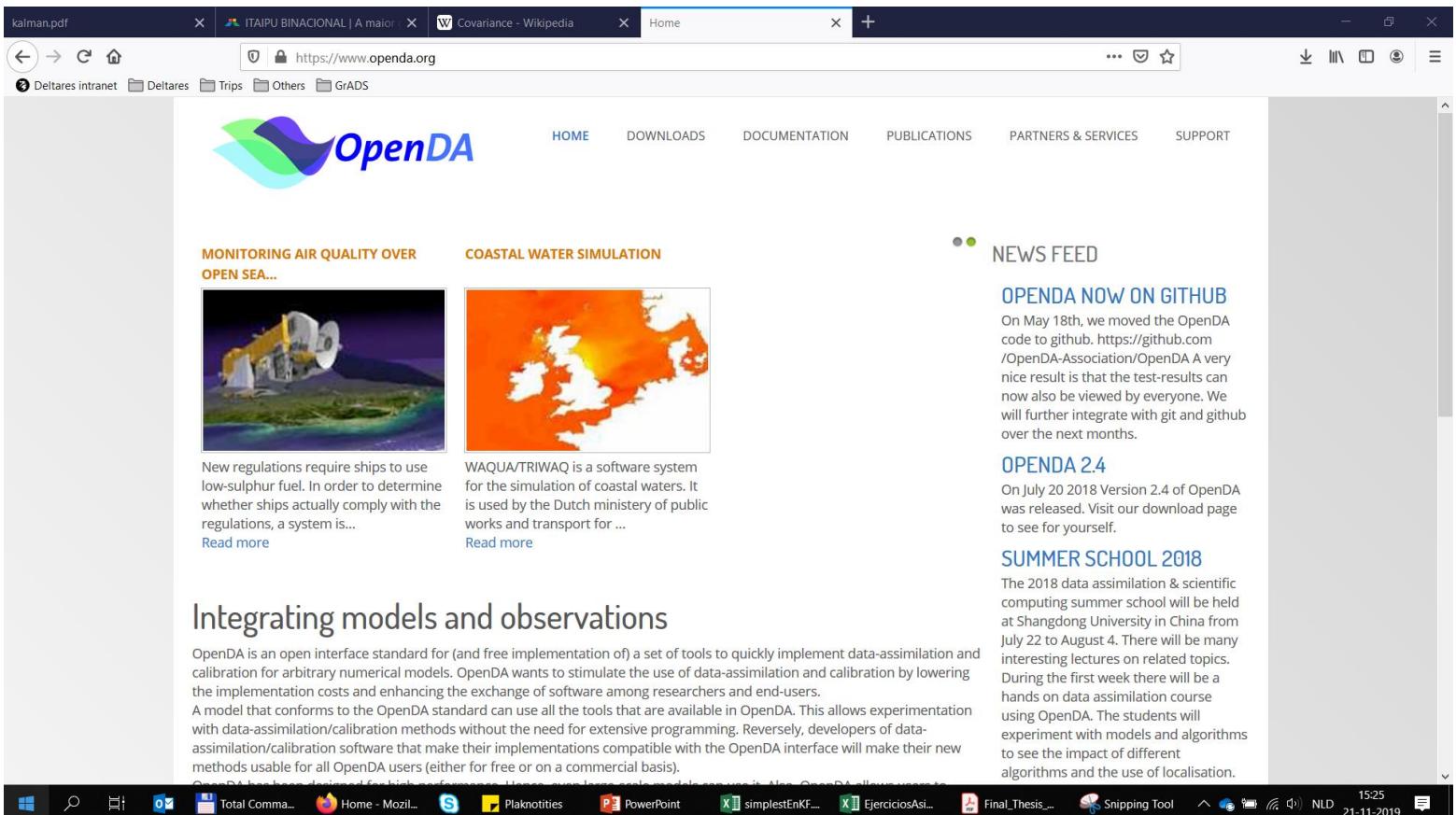
- + applicable on black-box models, simple to implement
- + probabilistic approach
- sequential technique, has issues with time lags

Var DA:

- + simultaneous technique over several time steps
- + suitable for reanalysis
- requires first-order sensitivities, i.e. adjoint code, and preferably a smooth model
- deterministic approach

R. Alvarado-Montero, D. Schwanenberg, P. Krahe, P. Helmke, , B. Klein. Multi Parametric Variational Data Assimilation for Hydrological Forecasting. Advances in Water Resources, 2017.

Software OpenDA



The screenshot shows a Microsoft Edge browser window displaying the OpenDA website. The address bar shows the URL <https://www.opendata.org>. The page features a header with the OpenDA logo and navigation links for HOME, DOWNLOADS, DOCUMENTATION, PUBLICATIONS, PARTNERS & SERVICES, and SUPPORT. Below the header, there are two main sections: 'MONITORING AIR QUALITY OVER OPEN SEA...' featuring an image of a satellite and text about ship fuel regulations, and 'COASTAL WATER SIMULATION' featuring an image of a map with a color-coded simulation result. To the right, there is a 'NEWS FEED' section with a heading 'OPENDA NOW ON GITHUB' and text about the move to GitHub. Below that is a section for 'OPENDA 2.4' and 'SUMMER SCHOOL 2018'. The bottom of the page includes a footer with various software icons and the date '16 december 2019'.

MONITORING AIR QUALITY OVER OPEN SEA...

New regulations require ships to use low-sulphur fuel. In order to determine whether ships actually comply with the regulations, a system is...

[Read more](#)

COASTAL WATER SIMULATION

WAQUA/TRIWAQ is a software system for the simulation of coastal waters. It is used by the Dutch ministry of public works and transport for ...

[Read more](#)

Integrating models and observations

OpenDA is an open interface standard for (and free implementation of) a set of tools to quickly implement data-assimilation and calibration for arbitrary numerical models. OpenDA wants to stimulate the use of data-assimilation and calibration by lowering the implementation costs and enhancing the exchange of software among researchers and end-users. A model that conforms to the OpenDA standard can use all the tools that are available in OpenDA. This allows experimentation with data-assimilation/calibration methods without the need for extensive programming. Reversely, developers of data-assimilation/calibration software that make their implementations compatible with the OpenDA interface will make their new methods usable for all OpenDA users (either for free or on a commercial basis).

OPENDA NOW ON GITHUB

On May 18th, we moved the OpenDA code to github. <https://github.com/OpenDA-Association/OpenDA> A very nice result is that the test-results can now also be viewed by everyone. We will further integrate with git and github over the next months.

OPENDA 2.4

On July 20 2018 Version 2.4 of OpenDA was released. Visit our download page to see for yourself.

SUMMER SCHOOL 2018

The 2018 data assimilation & scientific computing summer school will be held at Shangdong University in China from July 22 to August 4. There will be many interesting lectures on related topics. During the first week there will be a hands on data assimilation course using OpenDA. The students will experiment with models and algorithms to see the impact of different algorithms and the use of localisation.

Windows Taskbar icons: Total Comma..., Home - Mozilla Firefox, Plaknoties, PowerPoint, simplestEnKF..., EjerciciosAsi..., Final_Thesis..., Snipping Tool, NLD, 15:25, 21-11-2019

16 december 2019



Rodolfo Alvarado Montero

Rodolfo.AlvaradoMontero@deltares.nl

Tel. +(31) 657 58 4342

Deltares,

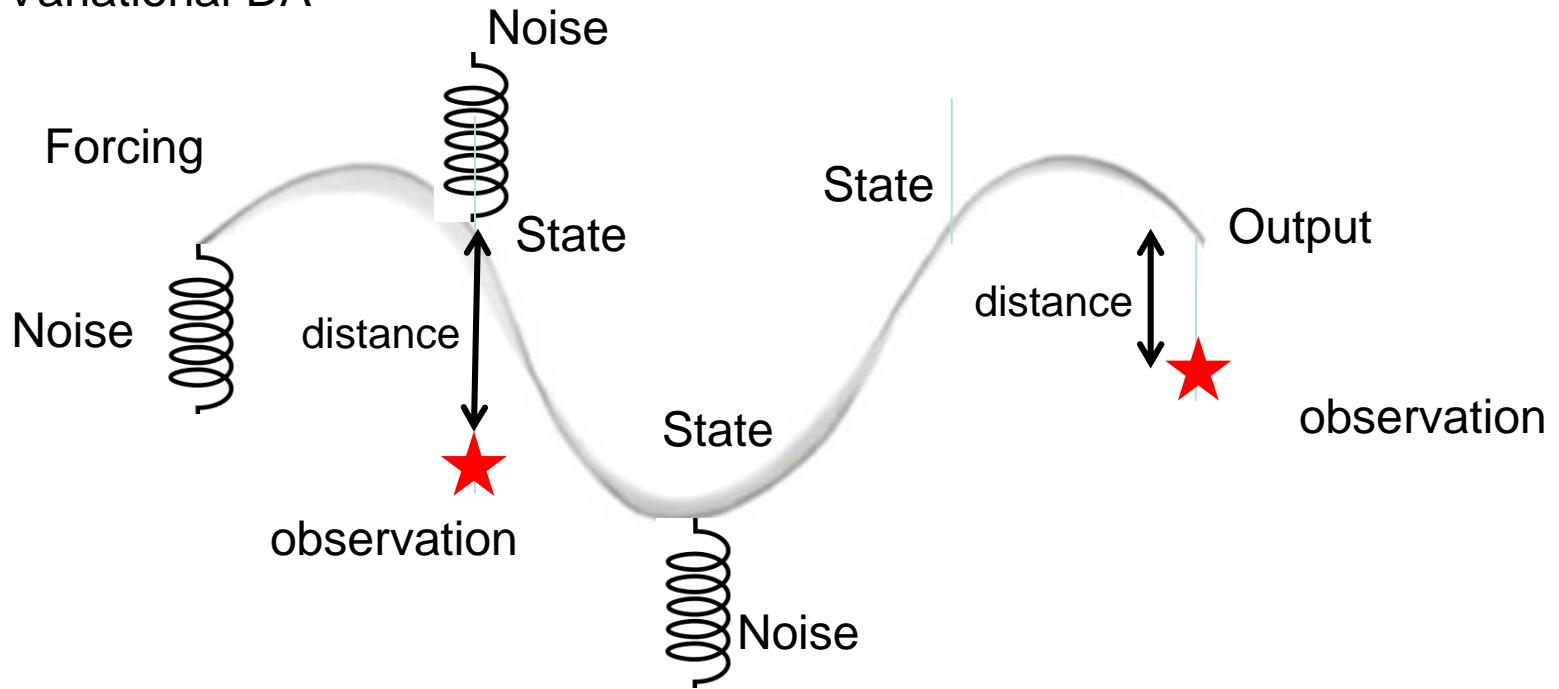
The Netherlands

GRACIAS / BRIGADO

Methodology (4)



- Variational DA



- Relies on the minimization of:

$$\min_{u,v} \sum_{k=-N+1}^0 [w_x \|\hat{x}^k - x^k(u)\| + w_y \|\hat{y}^k - y^k(u,v)\| + w_u \|u^k\| + w_v \|v^k\|]$$

Assimilation window

Distances

Noise

This equation represents the cost function used in variational DA to minimize the difference between observed data and model predictions, while also accounting for model error (noise) and control variables (u, v).

Methodology (8)



Deterministic approach:

$$\min_{u,v} \sum_{k=-N+1}^0 [w_x \|\hat{x}^k - x^k(u)\| + w_y \|\hat{y}^k - y^k(u,v)\| + w_u \|u^k\| + w_v \|v^k\|]$$

Observations Objective function

Probabilistic approach:

$$\min_{u,v} \sum_{m=1}^M p_m \cdot \sum_{k=-N+1}^0 [w_x \|\hat{x}^k - x^{k,m}(u)\| + w_y \|\hat{y}^k - y^{k,m}(u,v)\| + w_u \|u^k\| + w_v \|v^k\|]$$

Number of models Objective function

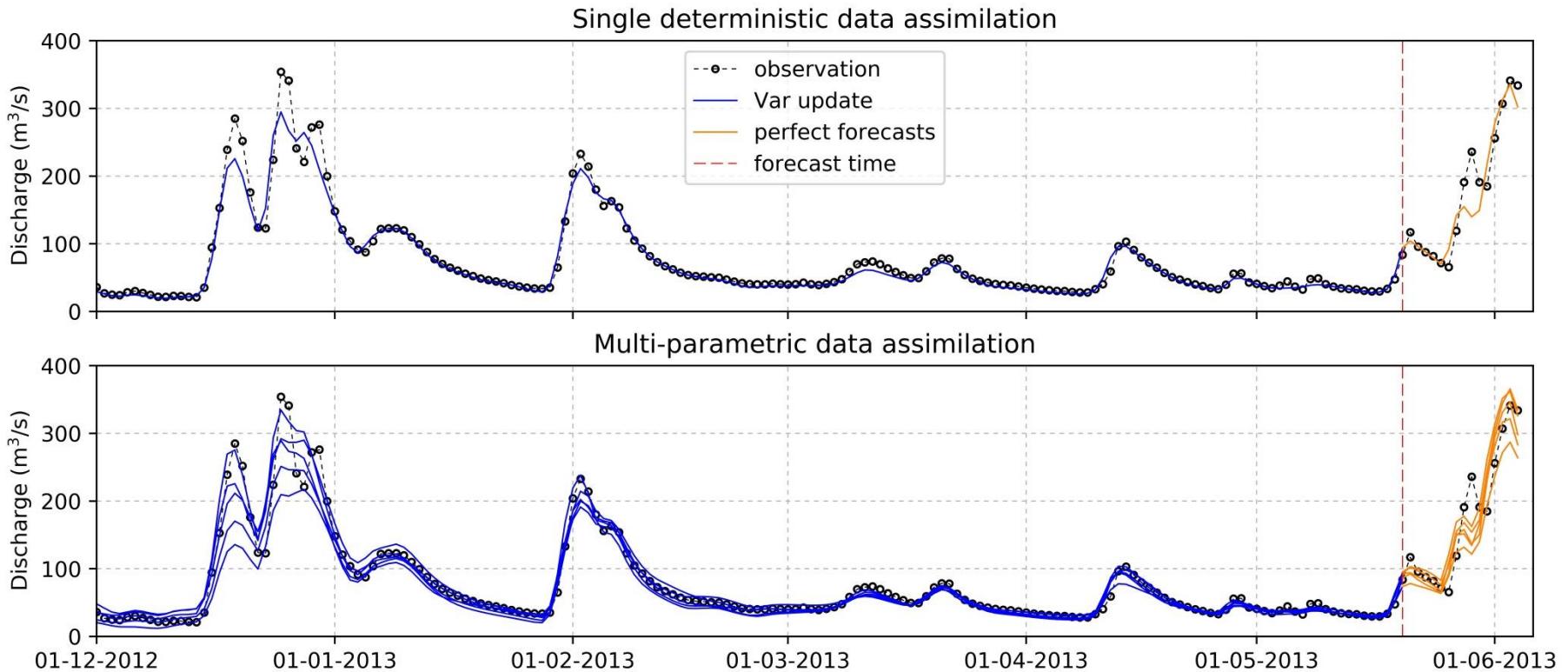
Probability of each model Subindex m for each model

Noise terms remain unchanged between models!

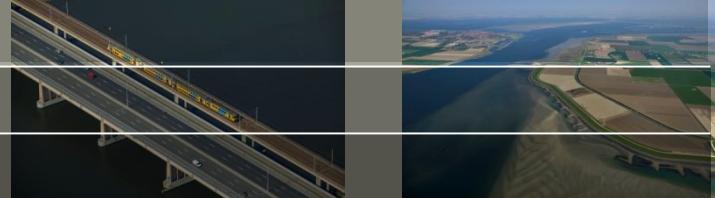
Results (2)



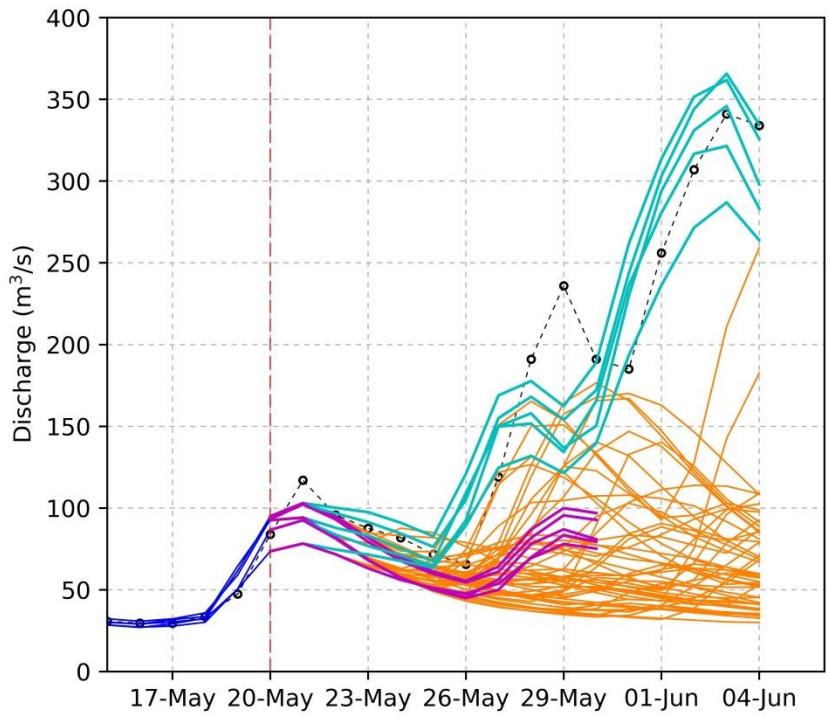
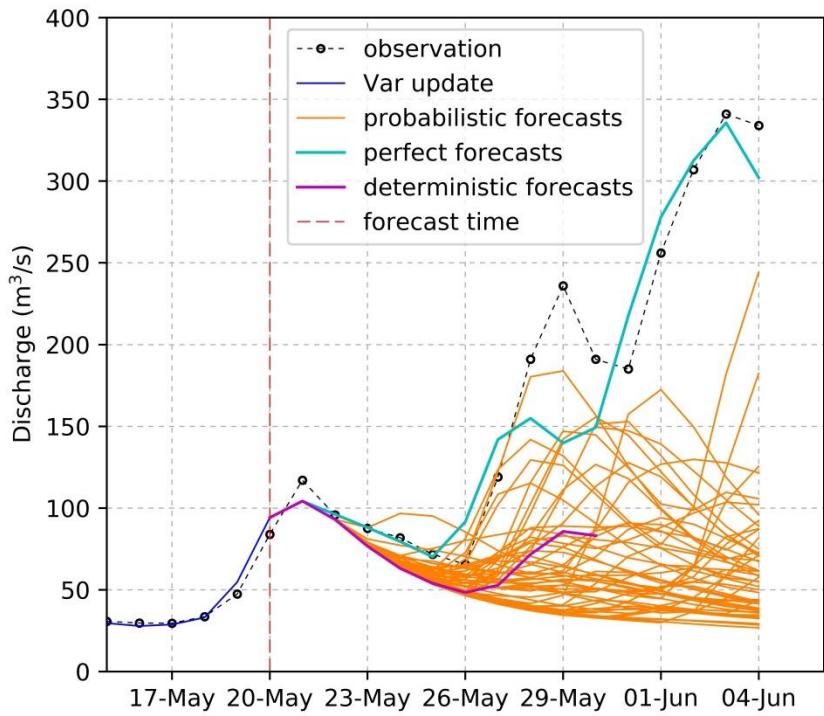
Example at a single forecast time using the multi-model variational DA approach (5 different parameter sets of HBV for this case)



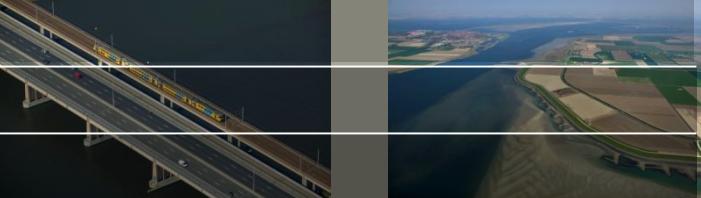
Results (3)



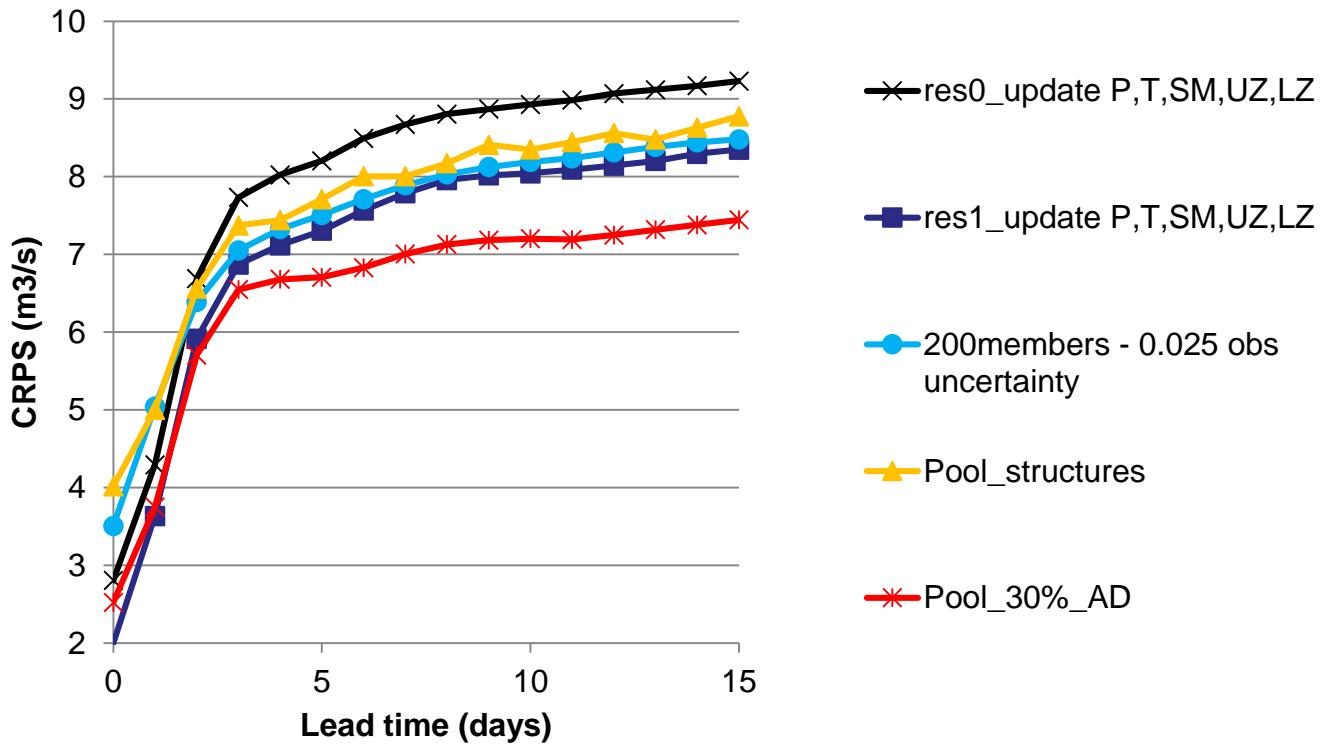
Example at a single forecast time using the multi-model variational DA approach (5 different parameter sets of HBV for this case)



Results (4)



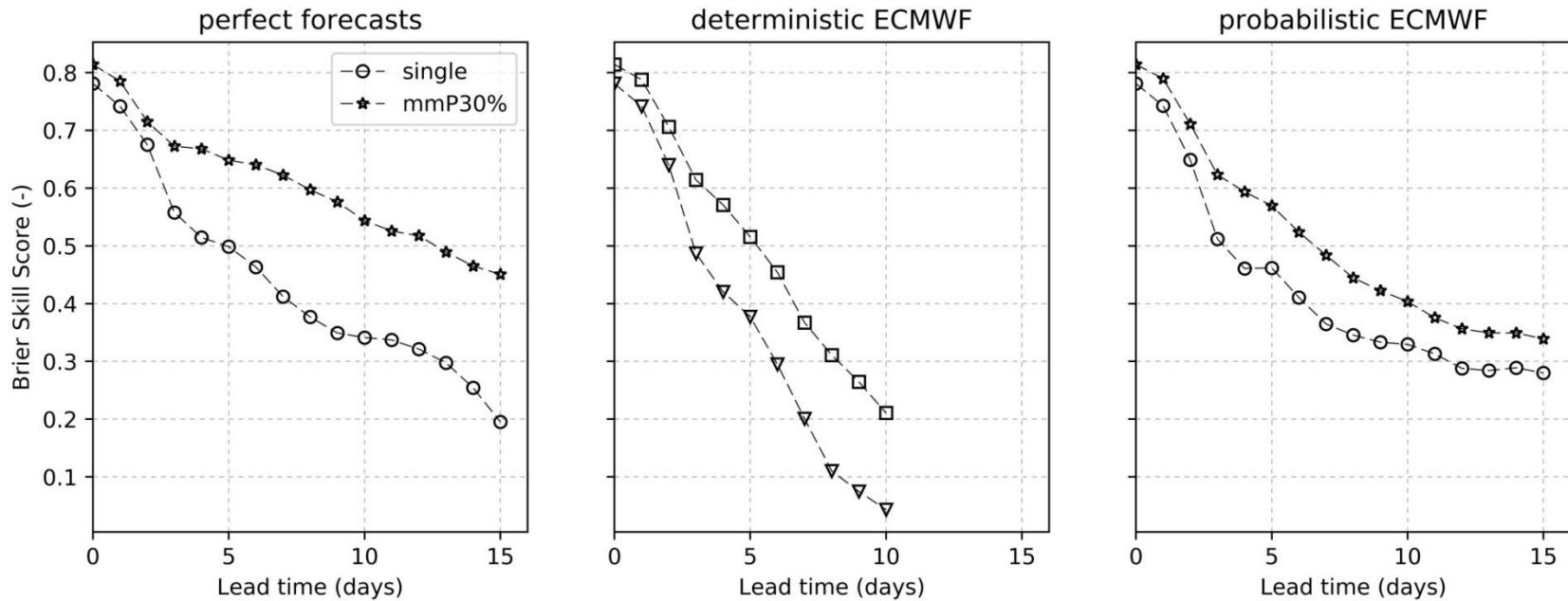
- Results at Main watershed
- Assimilation of streamflow
- 10 years of hindcast experiment (daily forecasts)



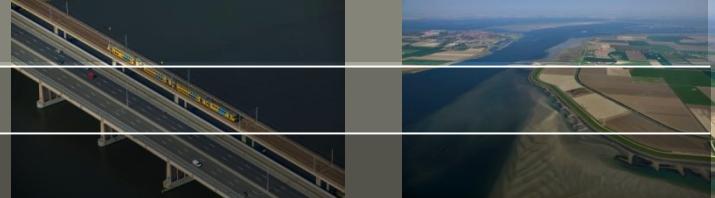
Results (5)



- Assimilating multiple variables (Q and SM, SCA, SWE of H-SAF)
- Hindcast of 2 years approximately 3.5 years using real forecasts
- BSS performance using perfect, deterministic and probabilistic forecasts



Results (6)

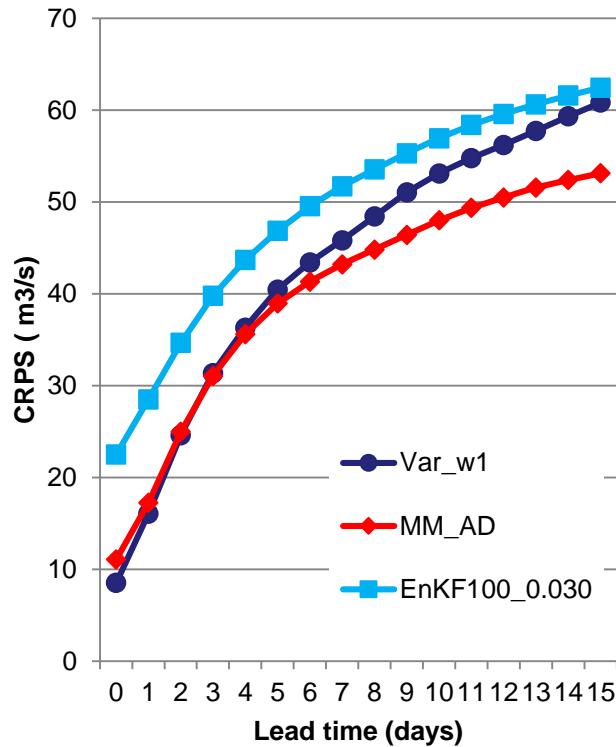


Mistassibi basin, Canada

Assimilation of Q, hindcast 5 years

Comparison of best performances:

- EnKF at 3% obs uncertainty
- Deterministic variational DA
- Probabilistic multi-parametric approach using aggregated distance reduction



Improvement with respect to EnKF

Lead time (days)	0	5	10	15
Deterministic Var DA	62%	14%	7%	3%
Probabilistic Var DA	51%	17%	16%	15%