Using a surrogate model in a hydrologic data assimilation system

Isabelle Mirouze[1], Sophie Ricci[1],
Matthias De Lozzo[1], Nicole Goutal[2], Siham El Garroussi[1],
Simon Saysset[1], Pamphile Roy[1]

1: UMR5318 CNRS / CERFACS, Toulouse, France
2: LNHE, LHSV, Chatou, France

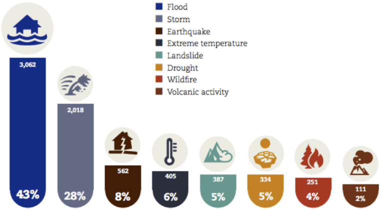
CliMathNet 19-21 September 2018, Reading, UK

Outline

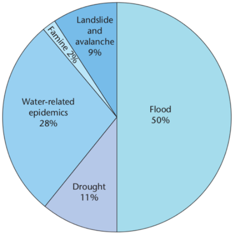
- Motivations
- Models and observations
- Data assimilation system
- Using a surrogate model
- Project status
- Summary and conclusions

Knowing the hydrologic system of rivers is essential for

- Activities: agriculture, hydroelectric production, fishing ...
- Safety: forecasting and preventing flood
  - environmental risk: safety and financial stakes
  - In France 16.8 million people leave in flood plain areas

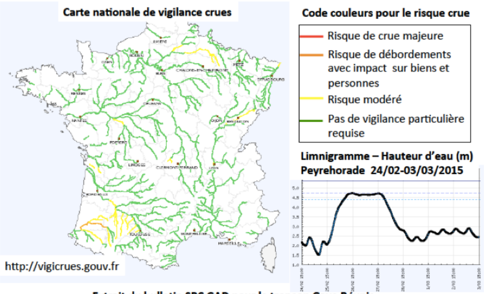


Source: WMO, 2011



Extreme events, such as rainfalls, floods, storms tend to intensify and be more frequent in the context of climate change.

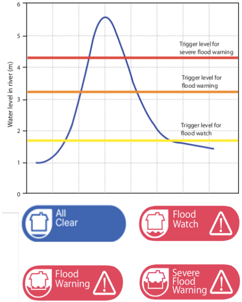
Objective: To alert civil and governmental authorities in charge of safety, and general public



Extrait du bulletin SPC GAD pour le tronçon Gave Réunis

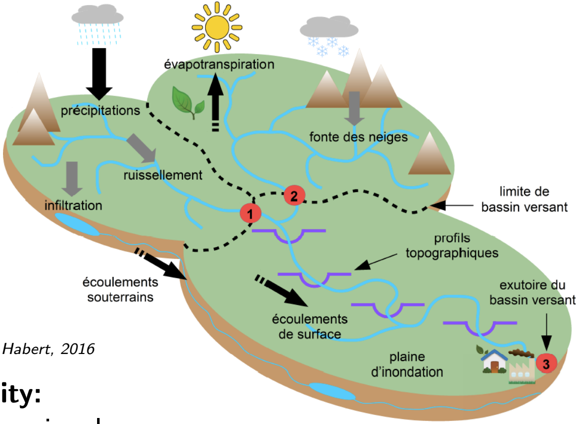
Conséquence des précipitations régulièrement observées et de la fonte nivale toujours active ce mardi matin sur les hauts bassins, les débits des Gaves de Pau et d'Oloron vont rester élevés ce mardi. Une hauteur comprise entre 2,80 m et 3,00 m est attendue à Peyrehorade pour la pleine mer de 18h00 ce mardi. Pour la pleine mer de mercredi à 6h00, la hauteur devrait être comprise entre 2,80 m et 3,30 m.

Source: PhD S. Barthélémy, 2015



Source: WMO, 2011

Decision Support System based on simulations (models) and observations

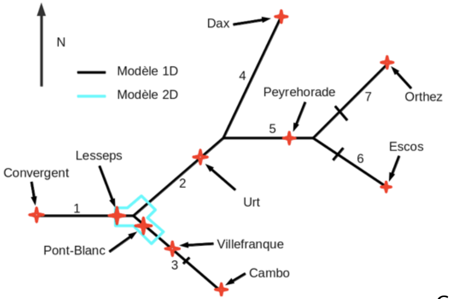


Source: PhD J. Habert, 2016

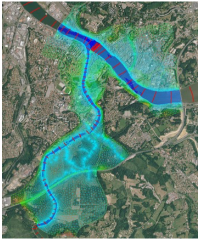
Complexity:
Multi dimensional,
Multi physics coupling (hydrodynamics, meteorology, sedimentology, . . .)

Quantities of interest:
Water height,
Discharge

Hydrology Modelling



Source: PhD S. Barthélémy, 2016

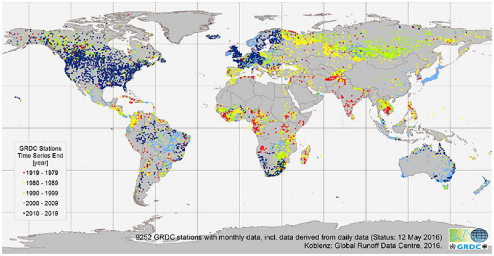


Confluence Nive / Adour, Bayonne

Random uncertainties:
Solvers, discretisation,
Hypotheses, simplifications,
Initial and boundary conditions,

Epistemic uncertainties:
Friction coefficients,
Bathymetry

Limnimetric in situ stations at bidos, Brazil (Paris, 2015)

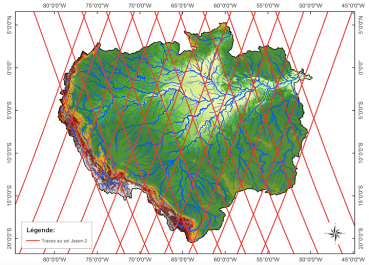


Global Runoff Data Center (GRDC) last data:
● 1919-1979 ● 1980-1989 ● 1990-1999 ● 2000-2009 ● 2010-2017

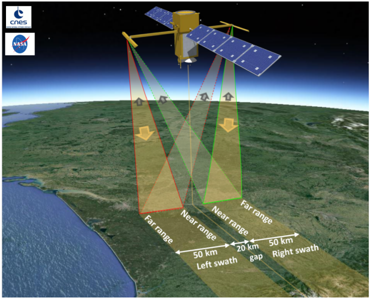
Source: PhD C. Emery, 2017

Mainly for water height. Few observations of discharge.
Network is sparse and number of available data is decreasing.
Accessing the observations can be difficult (charges, not shared).
Drones deployed punctually for a particular event.

Spatial altimetry for water height



JASON-2 altimeter tracks over the Amazon basin



Surface Water and Ocean Topography Mission

Current satellites: incomplete and low resolution spatial coverage, adapted to only major lakes and rivers.

SWOT: scheduled for a launch in 2021, will provide a global and high resolution coverage, adapted to river with width > 50 m

Data Assimilation system

Goal:
Improve water height and discharge prediction
from a few hours to 48 hours for operational forecasts

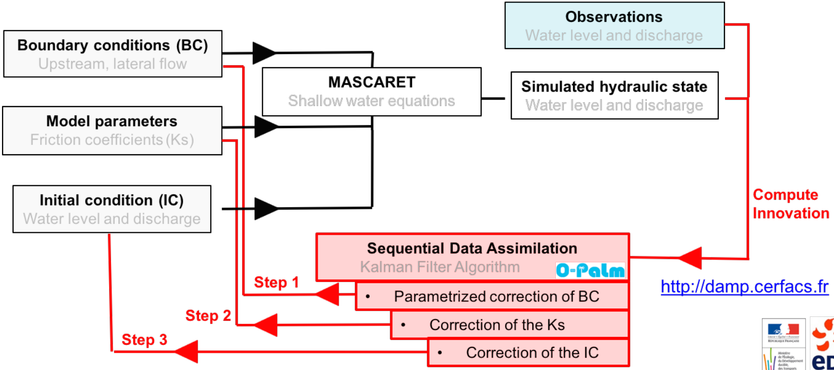
Models:
Shallow water Equations in 1D, 2D, 3D (opentelemac.org)

Observations:
Water height from station gauges
Simulated SWOT data

Control vector:	**Observation operator $\mathcal{H}$:**
Upstream and lateral flow (BC)	Complex and non-linear
Bathymetry (BC)	Complex and non-linear
Friction coefficients (KS)	Complex and non-linear
Water height and discharge (IC)	Interpolation

Scheme:
Ensemble Kalman Filter with surrogate models

Operational flood forecasting in French national services (SCHAPI, SPC):
Reduction of uncertainty in hydraulic model parameters, initial and boundary conditions to improve hydrodynamic simulation and forecast

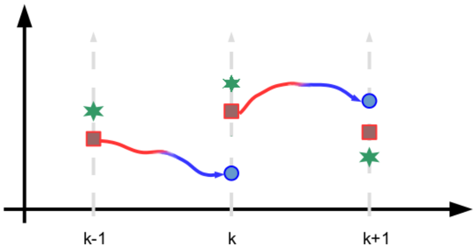


Source: Ricci et al., HESS 2011

(Extended) Kalman Filter

Propagation: $\boldsymbol{x}^{f,(k)} = \mathcal{M}\boldsymbol{x}^{a,(k-1)}$

Analysis: $\boldsymbol{x}^{a,(k)} = \boldsymbol{x}^{f,(k)} + \mathbf{K}\left(\boldsymbol{y}^{o,(k)} - \mathcal{H}\boldsymbol{x}^{f,(k)}\right)$



$\boldsymbol{x}^{a}$: analysis vector

$\boldsymbol{x}^{f}$: forecast vector

$\mathcal{H}\boldsymbol{x}^{f}$: in obs. space

$\boldsymbol{y}^{o}$: obs. vector

$\mathbf{K} = \mathbf{C_{xy}}\left(\mathbf{C_{yy}} + \mathbf{R}\right)^{-1}$: Kalman gain

$\mathbf{R}$: Observation error covariance matrix

$\mathbf{C_{yy}}$: Forecast error covariance matrix in observation space

$\mathbf{C_{xy}}$: Covariance matrix for forecast error in model/obs. space

From Kalman Filter to Ensemble Kalman Filter

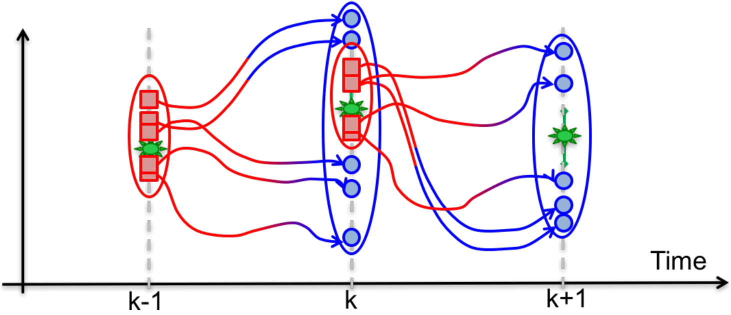
Issues for the Kalman Filter:

Compute $\mathbf{y}^f = \mathcal{H}\boldsymbol{x}^f$ with $\mathcal{H}$ complex and non-linear

Compute $\mathbf{C}_{\mathbf{yy}}$, $\mathbf{C}_{\mathbf{xy}}$

To use an Ensemble of Kalman Filters

Monte Carlo sampling with N_e members



Source: S. Ricci

Generate the ensemble with perturbations $\boldsymbol{\delta}_i$ for each member i

$$\boldsymbol{x}_i^{a,(k-1)} = \overline{\boldsymbol{x}^{a,(k-1)}} + \boldsymbol{\delta}_i$$

Integrate the model to obtain

$$\boldsymbol{x}_i^{f,(k)} = \mathcal{M}\boldsymbol{x}_i^{a,(k-1)}$$

$$\mathbf{y}_i^{f,(k)} = \mathcal{H}\mathcal{M}\boldsymbol{x}_i^{a,(k-1)}$$

Calculate stochastically (exponent (k) has been dropped)

$$\mathbf{C}_{\mathbf{yy}} = \frac{1}{N_e - 1}\sum_{i=1}^{N_e}\left(\mathbf{y}_i^f - \overline{\mathbf{y}^f}\right)\left(\mathbf{y}_i^f - \overline{\mathbf{y}^f}\right)^{\mathrm{T}}$$

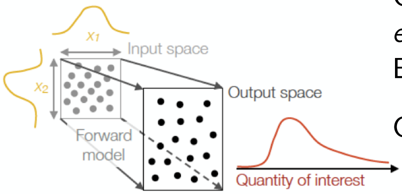
Calculate stochastically (exponent (k) has been dropped)

$$\mathbf{C}_{\mathbf{xy}} = \frac{1}{N_e - 1}\sum_{i=1}^{N_e}\left(\boldsymbol{x}_i^f - \overline{\boldsymbol{x}^f}\right)\left(\mathbf{y}_i^f - \overline{\mathbf{y}^f}\right)^{\mathrm{T}}$$

Issues for the Ensemble Kalman Filter:

N_e model integrations can be costly to be accurate enough

To use a surrogate model to determine $\tilde{\mathbf{y}}_i^f \approx \mathbf{y}_i^f$



Source: M. Rochoux

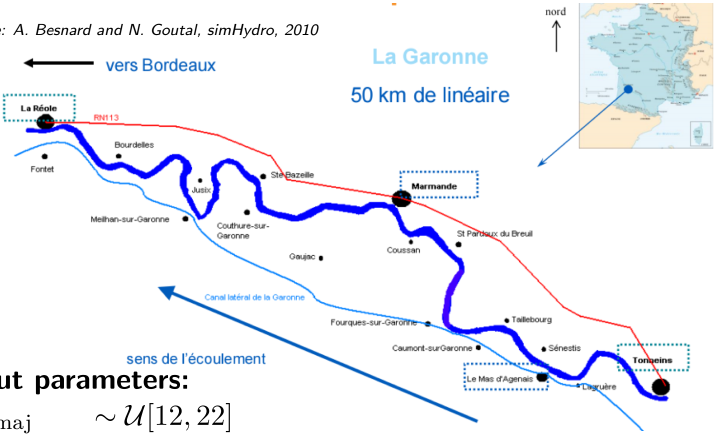
Polynomial Chaos order P:
Choose $\phi\left(\boldsymbol{x}^{a,(k-1)}\right)$
e.g. Legendre, Hermite, ...
Estimate w^j from N_{ls} samples

Calculate

$$\tilde{\mathbf{y}}_i^{f,(k)} = \sum_{j=1}^{P} w_j \phi_j \left(\boldsymbol{x}_i^{a,(k-1)}\right)$$

Requires N_{ls} model integrations with $N_{\mathrm{ls}} << N_e$

Source: A. Besnard and N. Goutal, simHydro, 2010



Input parameters:

$Ks_{\mathrm{maj}} \sim \mathcal{U}[12, 22]$

$Ks_{\mathrm{min,up}} \sim \mathcal{U}[40, 50]$

$Ks_{\mathrm{min,mid}} \sim \mathcal{U}[33, 43]$

$Ks_{\mathrm{min,down}} \sim \mathcal{U}[35, 45]$

$Q_{\mathrm{up}} \sim \mathcal{N}(8490, 700)$

High Flow

Quantity of interest:

Water height at Marmande

Water height on 41416 mesh

Surrogate model: test case for Telemac2D

Learning samples: $\mathbf{y}_j^{\mathrm{ls}} = \mathcal{M}\boldsymbol{x}_j^{\mathrm{ls}}$ for $j = 1, \ldots, N_{\mathrm{ls}} << N_e$

Gaussian process:

- Choose correlation kernel $r(\boldsymbol{x}, \boldsymbol{x}'; \zeta)$ with length scale ζ
- Estimate $\mu(\boldsymbol{x}; a) = a_0 + a_1\boldsymbol{x}$
- Estimate (w_j) for $1 \leq j \leq P$ from a_0, a_1, ζ, max. likelihood
- Surrogate model $\mathcal{S}$:

$$\tilde{\mathbf{y}}^f = \mu(\boldsymbol{x}^a) + \sum_{j=1}^{P} w_j \; r\left(\boldsymbol{x}^a, \boldsymbol{x}_j^{ls}\right)$$

Polynomial Chaos:

- Choose orthogonal function $\phi\left(\boldsymbol{x}\right)$
- Estimate (w_j) for $1 \leq j \leq P$, least square
- Surrogate model $\mathcal{S}$:

$$\tilde{\mathbf{y}}^f = \sum_{j=1}^{P} w_j \; \phi\left(\boldsymbol{x}^a\right)$$

Validation samples: $\begin{aligned}\mathbf{y}_j^{\mathrm{vs}} &= \mathcal{M}\boldsymbol{x}_j^{\mathrm{vs}} \\ \tilde{\mathbf{y}}_j^{\mathrm{vs}} &= \mathcal{S}\boldsymbol{x}_j^{\mathrm{vs}}\end{aligned}$ for $j = 1, \ldots, N_{\mathrm{vs}}$

$$\mathrm{RMSE} = ||\mathbf{y}^{\mathrm{vs}} - \tilde{\mathbf{y}}^{\mathrm{vs}}||_2 \qquad Q_2 = 1 - \frac{||\mathbf{y}^{\mathrm{vs}} - \tilde{\mathbf{y}}^{\mathrm{vs}}||_2^2}{||\mathbf{y}^{\mathrm{vs}} - \overline{\mathbf{y}^{\mathrm{vs}}}||_2^2}$$

Surrogate model: test case for Telemac2D

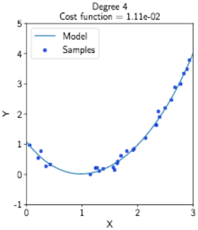
Water height at Marmande:

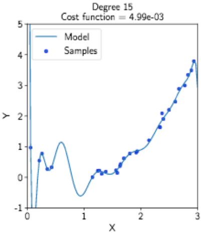
$N_{\text{ls}} = 55$	Gaussian Process		Polynomial Chaos	
$N_{\text{vs}} = 23$	Gauss	M(5/2)	$P = 2$	$P = 3$
Q_2	0.998	0.999	0.999	−105.453
RMSE (m)	0.008	0.004	0.005	2.869
Construction	20 mn	20 mn	0.69 s	0.87 s

Water height at Marmande:

$N_{\text{ls}} = 55$	Gaussian Process		Polynomial Chaos	
$N_{\text{vs}} = 23$	Gauss	M(5/2)	$P = 2$	$P = 3$
Q_2	0.998	0.999	0.999	−105.453
RMSE (m)	0.008	0.004	0.005	2.869
Construction	20 mn	20 mn	0.69 s	0.87 s

Overfitting risk:



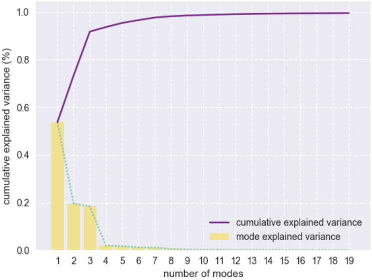


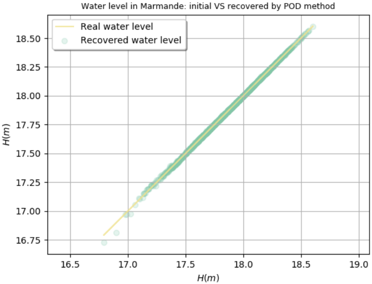
Source:

Output space reduction: Proper Orthogonal Decomposition

Construction of a surrogate model based on PC with $P = 2$:

1 point $\rightarrow 0.69$ s $\quad 41416$ points $\rightarrow\sim 8$ h ?



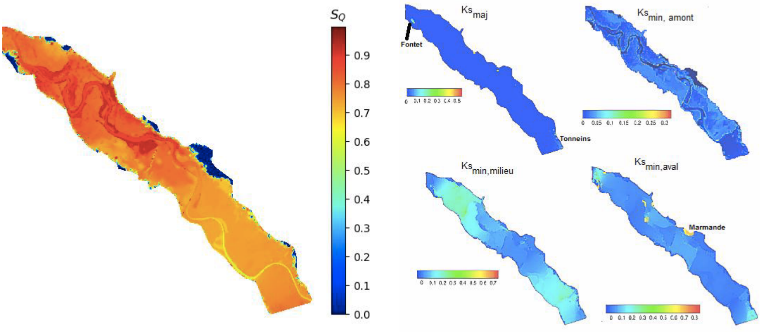


Source: S. El Garoussi, 2018

The first 4 modes explain 95% of the water height variance

Surrogate model: test case for Telemac2D

Sobol indices (first order)



Source: S. El Garoussi, 2018

The sensitivity depends on the range of variations

Project status

- **Models:**
  - Mascaret, Telemac: Fortran
  - TelAPI: Application Programming Interface, Python
- **Data Assimilation:**
  - Development of a modular code in python
    - stochastic Ensemble Kalman Filter for Mascaret
- **Surrogate models:**
  - Batman: Bayesian Analysis Tool for Modelling and uncertAinty quaNtification, in python, based on OpenTurns
    - Roy *et al.* (2018). BATMAN: Statistical analysis for expensive computer codes made easy. *Journal of Open Source Software*
    - https://batman.readthedocs.io/en/latest
    - openturns.org

Summary

- Knowing the hydrologic system of rivers is essential
- Prediction of water height and discharge:
  - Models (more or less complex) have uncertainties
  - Observations mainly for water height, sparse and imperfect
  - Data assimilation can help reducing the uncertainty
- On-going development of an Ensemble Kalman Filter
  - Monte Carlo: requires important amount of model integration
  - Surrogate models can help reducing the cost
- On-going studies on surrogate models for Mascaret (1.5D) and Telemac (2D)
  - Surrogate models for different range of discharge
  - Aggregating the different models
  - Surrogate model for unsteady flow (flood)