Variational Data Assimilation: Latest SIC4DVar results and potential global applicability

Hind Oubanas, Igor Gejadze, Felix Billaud, **Pierre-Olivier Malaterre**

CONTENT

- SIC4DVAR ALGORITHM :
  - HYDRAULIC MODEL SIC².
  - VARIATIONAL DATA ASSIMILATION.

- EXPERIMENTS & RESULTS
  - PEPSI CHALLENGE 1 : GARONNE RIVER.
  - SWOT HYDROLOGY SIMULATOR : PO RIVER.

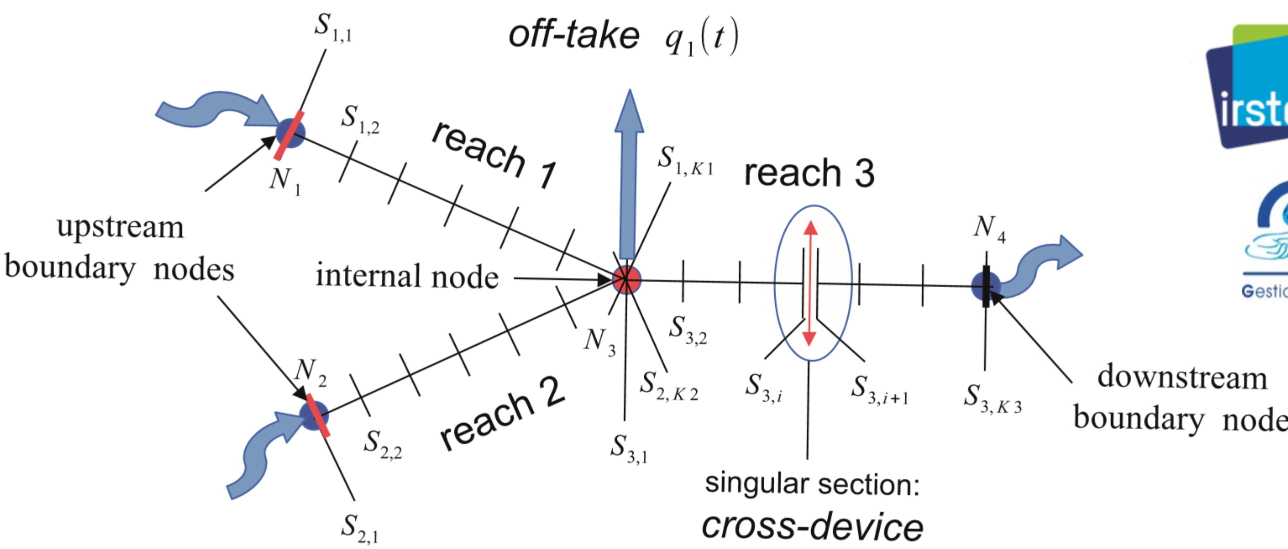
- RECENT DEVELOPMENTS & SET UPS
  - PEPSI CHALLENGE 2 : PRELIMINARY RESULTS.

- SIC4DVAR GLOBAL APPLICABILITY.

- PATH FORWARD.

SIC² HYDRAULIC MODEL

SIMULATION AND INTEGRATION OF CONTROLS FOR CHANNELS



- **1.5D Full Saint-Venant hydraulic model.**

$$\frac{\partial A}{\partial t} + \frac{\partial q}{\partial x} = Q_L$$

$$\frac{\partial q}{\partial t} + \frac{\partial q^2/A}{\partial x} + gA\frac{\partial z}{\partial x} = -gAS_f + C_L Q_L v$$

$$v = K_S R^{2/3} S_f^{1/2}$$

- Boundary conditions : e.g. $Q(t)$ upstream and $Q(Z)$ or $Z(t)$ downstream.

- $Q(x,t)$: Discharge
- $Z(x,t)$: Water level
- $A(x,t)$: Cross−sectional area

- $v(x,t)$: Mean velocity
- R : Hydraulic radius
- K_S : Strickler coefficient
- C_L : Lateral discharge coefficient

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

Iterative regularization method

Change of variables: $V = V_b + B_V^{1/2}\tilde{V}.$

For given $Y \in \mathcal{Y}$, find $\tilde{V} \in \tilde{\mathcal{U}}_V$ such that:

$$J_1(\tilde{V}, \tilde{\Lambda}) = \frac{1}{2}\left\| R^{-1/2}\left(\mathcal{G}\left(V_b + B_V^{1/2}\tilde{V}\right) - Y\right)\right\|_{\mathcal{Y}}^2 \rightarrow inf_{\tilde{V}} \qquad (1)$$

Residual (discrepency) principle : $J_1\left(\tilde{V}_i^{(1)}\right) \sim \mathcal{X}^2(M),$

$\left(\tilde{V}_i^{(1)}\right)$: the iterative solution of the minimization problem (1). M : the number of observations.

- $\mathcal{G} : \mathcal{U} \mapsto \mathcal{Y}$: Nonlinear mapping operator
- $\mathcal{H} : \mathcal{X} \mapsto \mathcal{Y}$: Observation operator
- $\mathcal{B}$: Background covariance matrix
- $\mathcal{R}$: Observation covariance matrix

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

Iterative regularization method

Change of variables: $V = V_b + B_V^{1/2}\tilde{V}$.

For given $Y \in \mathcal{Y}$, find $\tilde{V} \in \tilde{\mathcal{U}}_V$ such that:

$$J_1(\tilde{V}, \tilde{\Lambda}) = \frac{1}{2}\left\| R^{-1/2}\left(\mathcal{G}\left(V_b + B_V^{1/2}\tilde{V}\right) - Y\right)\right\|_{\mathcal{Y}}^2 \rightarrow inf_{\tilde{V}} \qquad (1)$$

Residual (discrepency) principle : $J_1\left(\tilde{V}_i^{(1)}\right) \sim \mathcal{X}^2(M)$,

$\left(\tilde{V}_i^{(1)}\right)$: the iterative solution of the minimization problem (1). M : the number of observations.

Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method :

$$\tilde{V}_{i+1} = \tilde{V}_i + \beta_i \tilde{H}_i^{-1}\frac{\partial J_1(\tilde{V}_i)}{\partial \tilde{V}}, \tilde{V}_0 = 0$$

- $\mathcal{G} : \mathcal{U} \mapsto \mathcal{Y}$: Nonlinear mapping operator
- $\mathcal{H} : \mathcal{X} \mapsto \mathcal{Y}$: Observation operator
- $\mathcal{B}$: Background covariance matrix
- $\mathcal{R}$: Observation covariance matrix

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

Iterative regularization method

Change of variables: $V = V_b + B_V^{1/2}\tilde{V}.$

For given $Y \in \mathcal{Y}$, find $\tilde{V} \in \tilde{\mathcal{U}}_V$ such that:

$$J_1(\tilde{V}, \tilde{\Lambda}) = \frac{1}{2}\left\|R^{-1/2}\left(\mathcal{G}\left(V_b + B_V^{1/2}\tilde{V}\right) - Y\right)\right\|_{\mathcal{Y}}^2 \rightarrow inf_{\tilde{V}} \qquad (1)$$

Residual (discrepency) principle : $J_1\left(\tilde{V}_i^{(1)}\right) \sim \mathcal{X}^2(M),$

$\left(\tilde{V}_i^{(1)}\right)$: the iterative solution of the minimization problem (1). M : the number of observations.

Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method :

$$\tilde{V}_{i+1} = \tilde{V}_i + \beta_i \tilde{H}_i^{-1}\frac{\partial J_1(\tilde{V}_i)}{\partial \tilde{V}}, \ \tilde{V}_0 = 0$$

Gradient of the cost function ⇒ Tangent Linear & Adjoint Models

$$\frac{\partial J_1(\tilde{V}_i)}{\partial \tilde{V}} = \left(\mathcal{B}^{\frac{1}{2}}\right)^* \left(\mathcal{G}'\left(V_b + B_V^{1/2}\tilde{V}\right)\right)^* R^{-1}\left(\mathcal{G}\left(V_b + B_V^{1/2}\tilde{V}\right) - Y\right)$$

- $\mathcal{G} : \mathcal{U} \mapsto \mathcal{Y}$: Nonlinear mapping operator
- $\mathcal{H} : \mathcal{X} \mapsto \mathcal{Y}$: Observation operator
- $\mathcal{B}$: Background covariance matrix
- $\mathcal{R}$: Observation covariance matrix

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

Iterative regularization method

Change of variables: $V = V_b + B_V^{1/2}\tilde{V}.$

For given $Y \in \mathcal{Y}$, find $\tilde{V} \in \tilde{\mathcal{U}}_V$ such that:

$$J_1(\tilde{V},\tilde{\Lambda}) = \frac{1}{2}\left\|R^{-1/2}\left(\mathcal{G}\left(V_b + B_V^{1/2}\tilde{V}\right) - Y\right)\right\|_{\mathcal{Y}}^2 \rightarrow inf_{\tilde{V}} \qquad (1)$$

Residual (discrepency) principle : $J_1\left(\tilde{V}_i^{(1)}\right) \sim \mathcal{X}^2(M),$

$\left(\tilde{V}_i^{(1)}\right)$: the iterative solution of the minimization problem (1). M : the number of observations.

Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method :

$$\tilde{V}_{i+1} = \tilde{V}_i + \beta_i \tilde{H}_i^{-1}\frac{\partial J_1(\tilde{V}_i)}{\partial \tilde{V}}, \tilde{V}_0 = 0$$

Gradient of the cost function ⇒ Tangent Linear & Adjoint Models

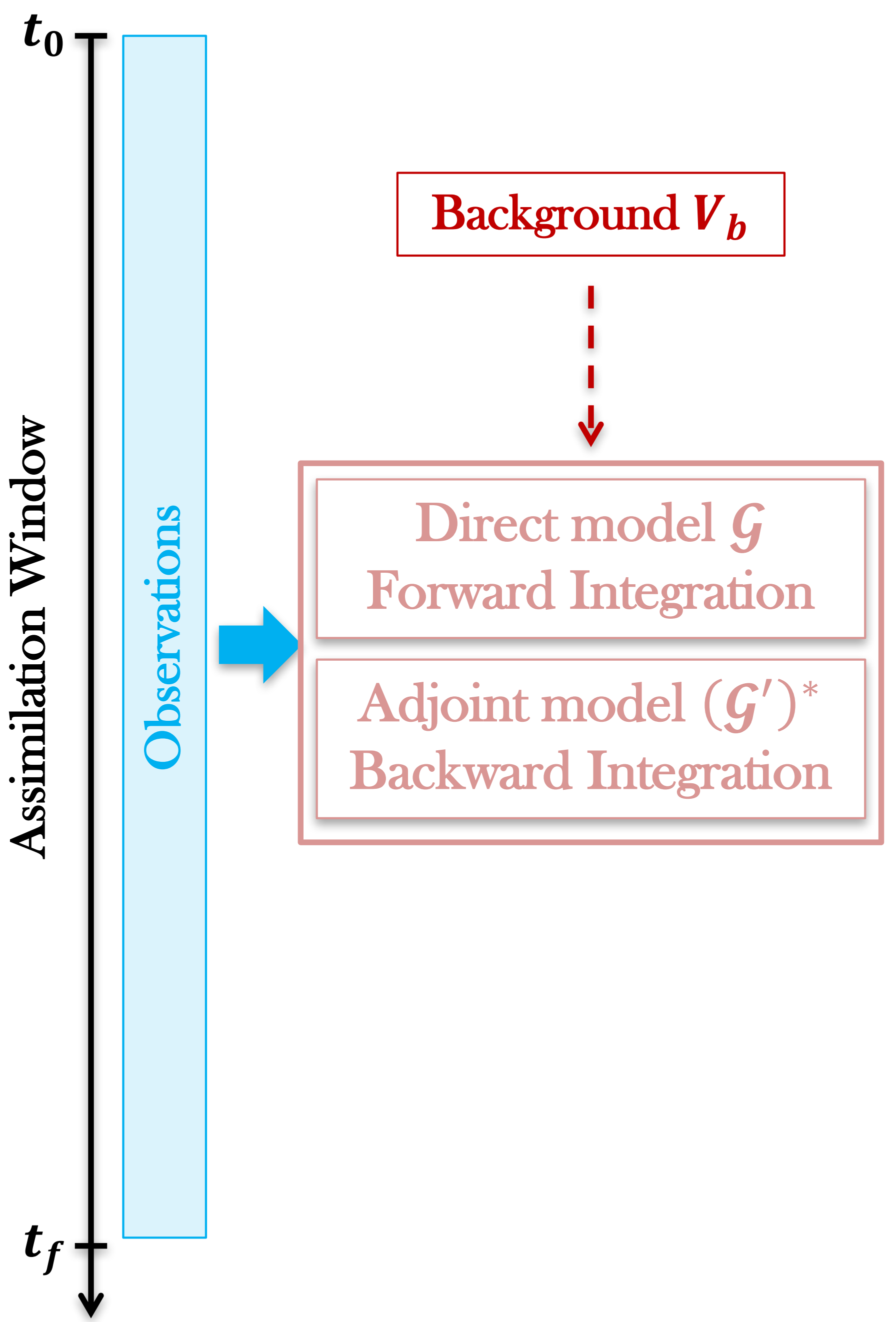
$$\frac{\partial J_1(\tilde{V}_i)}{\partial \tilde{V}} = \left(\mathcal{B}^{\frac{1}{2}}\right)^* \left(\mathcal{G}'\left(V_b + B_V^{1/2}\tilde{V}\right)\right)^* R^{-1}\left(\mathcal{G}\left(V_b + B_V^{1/2}\tilde{V}\right) - Y\right)$$

Automatic differentiation TAPENADE (INRIA)

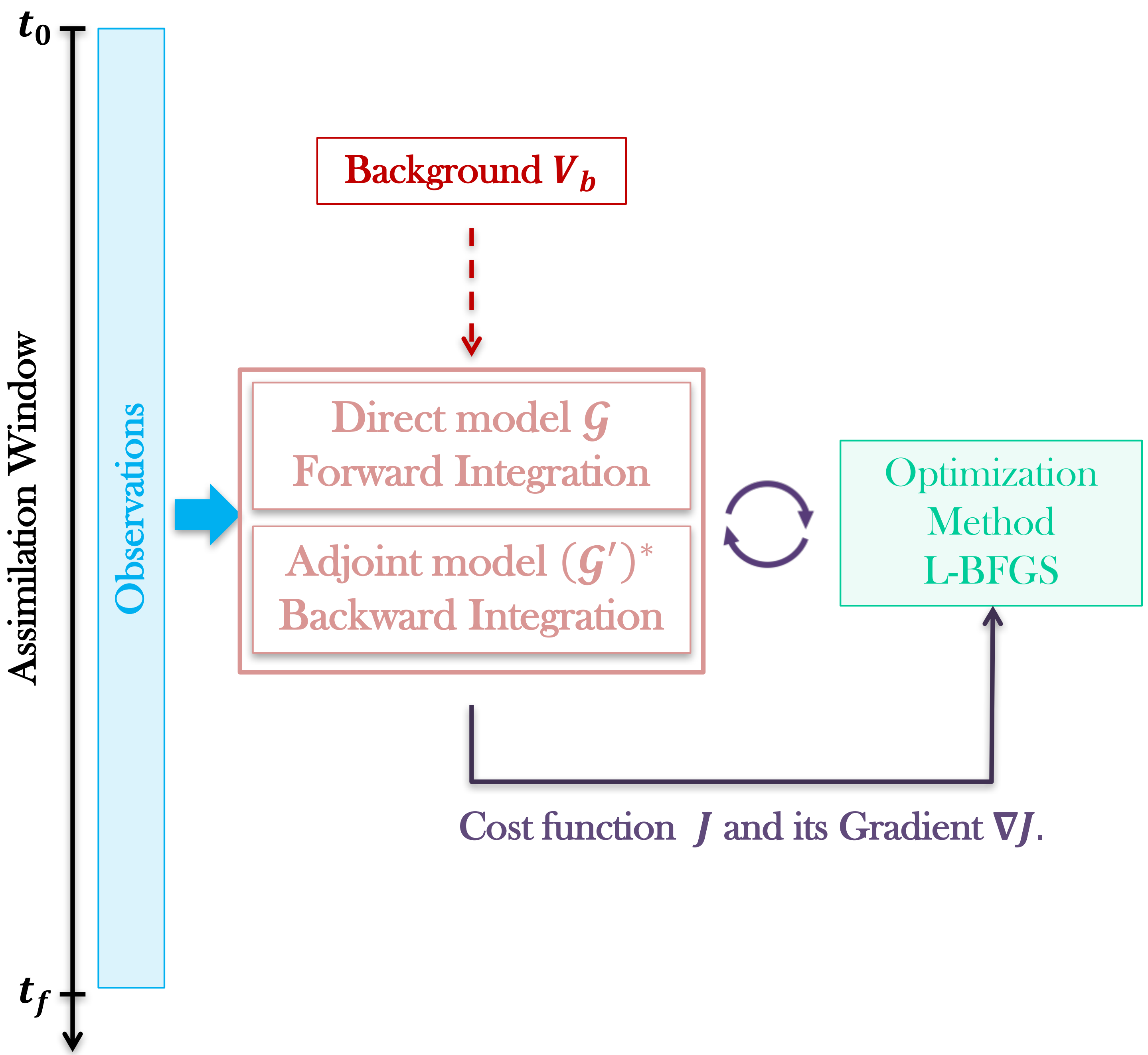
(Gejadze and Malaterre 2016)

- $\mathcal{G} : \mathcal{U} \mapsto \mathcal{Y}$: Nonlinear mapping operator
- $\mathcal{H} : \mathcal{X} \mapsto \mathcal{Y}$: Observation operator
- $\mathcal{B}$: Background covariance matrix
- $\mathcal{R}$: Observation covariance matrix

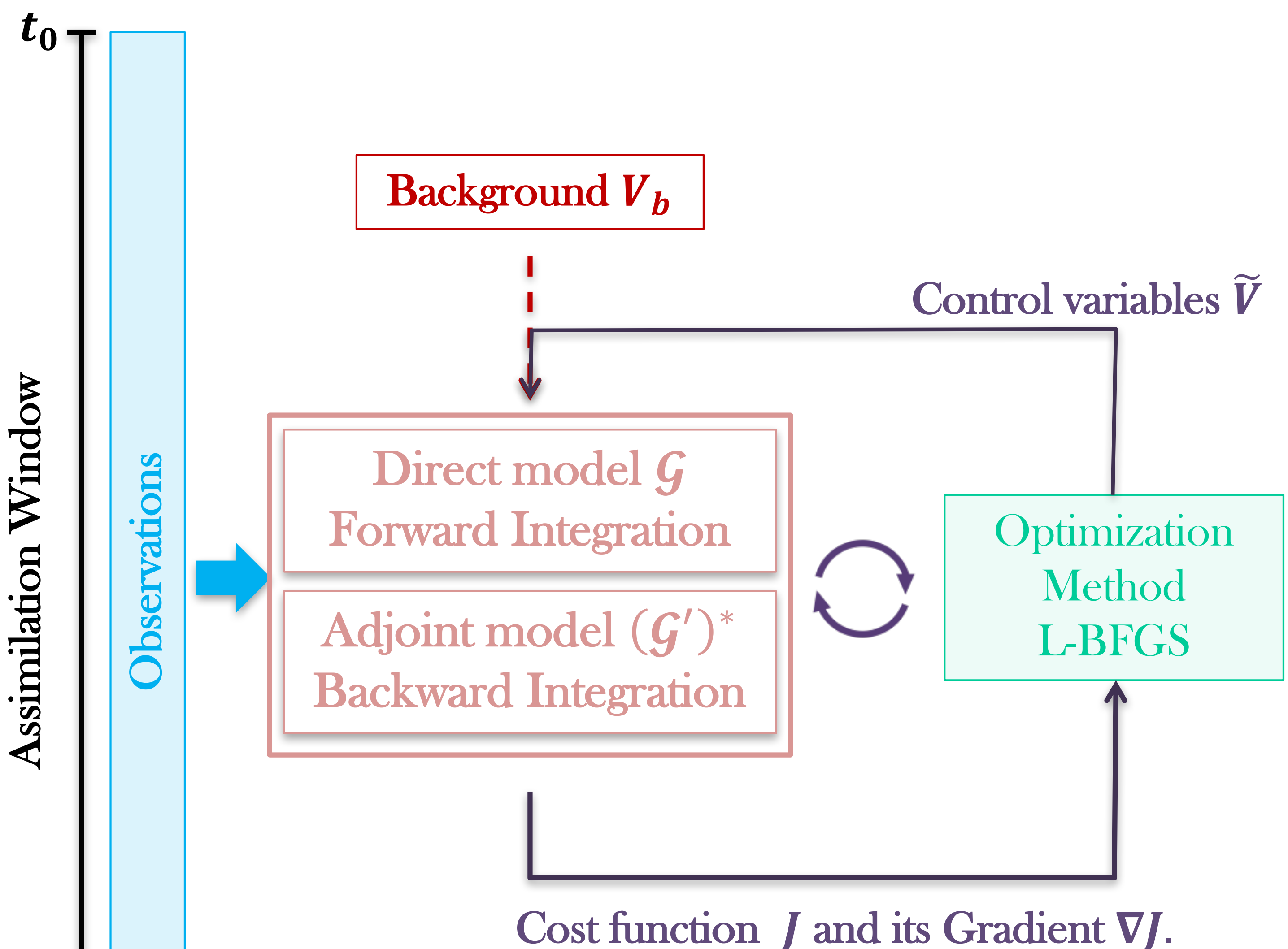
VARIATIONAL DATA ASSIMILATION – SIC4DVAR



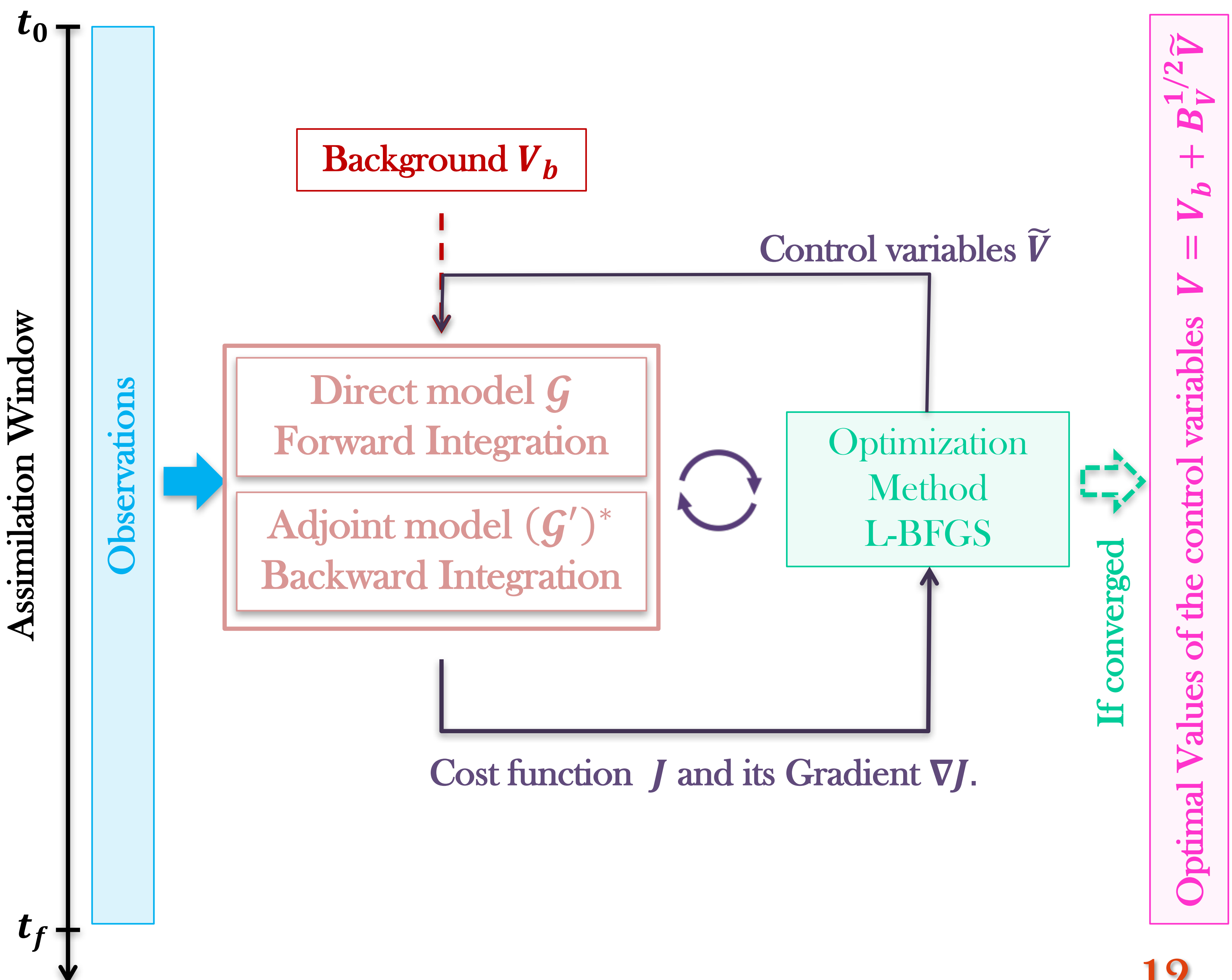
VARIATIONAL DATA ASSIMILATION – SIC4DVAR



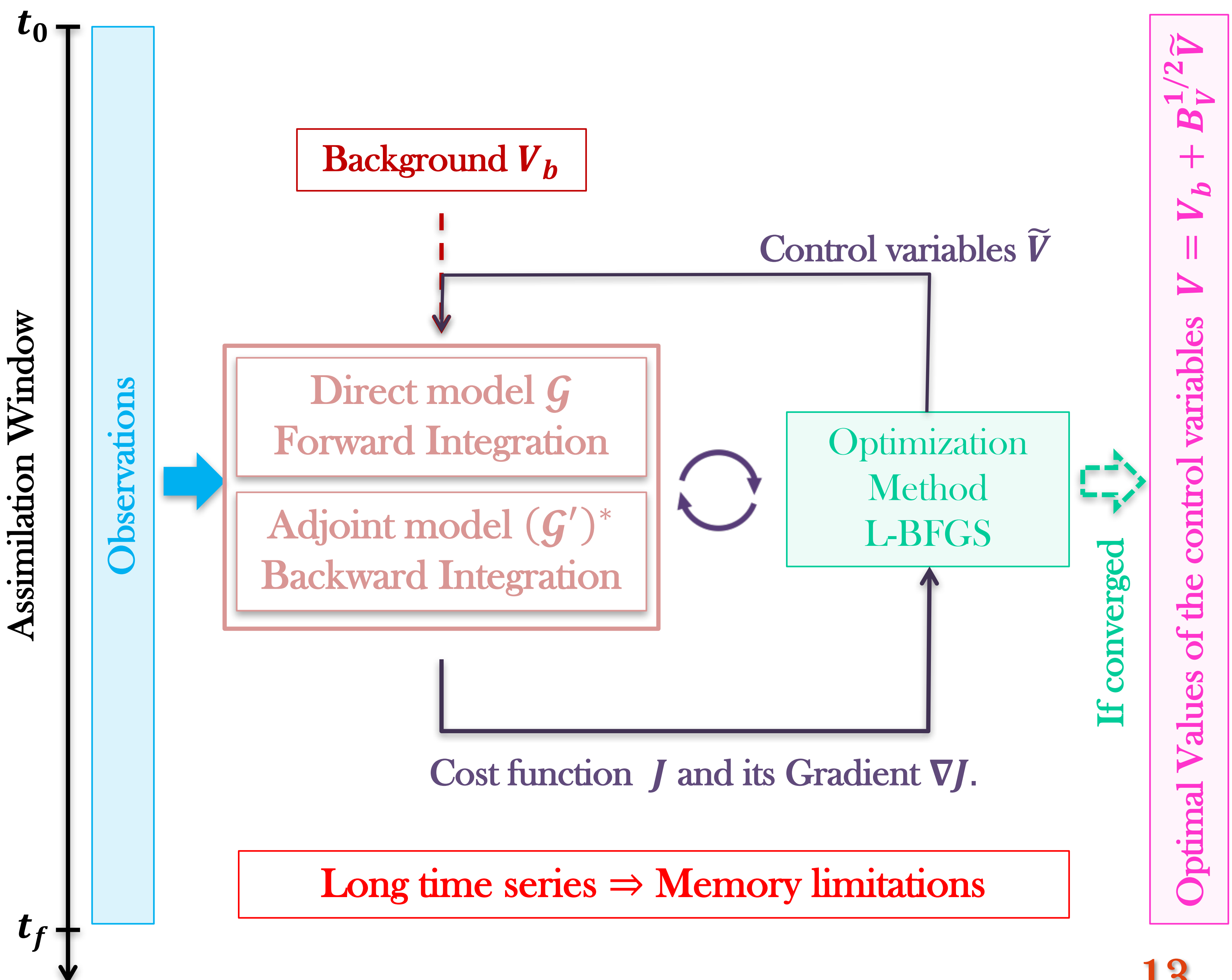
VARIATIONAL DATA ASSIMILATION – SIC4DVAR



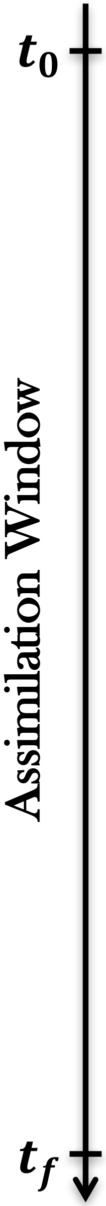
VARIATIONAL DATA ASSIMILATION – SIC4DVAR



VARIATIONAL DATA ASSIMILATION – SIC4DVAR

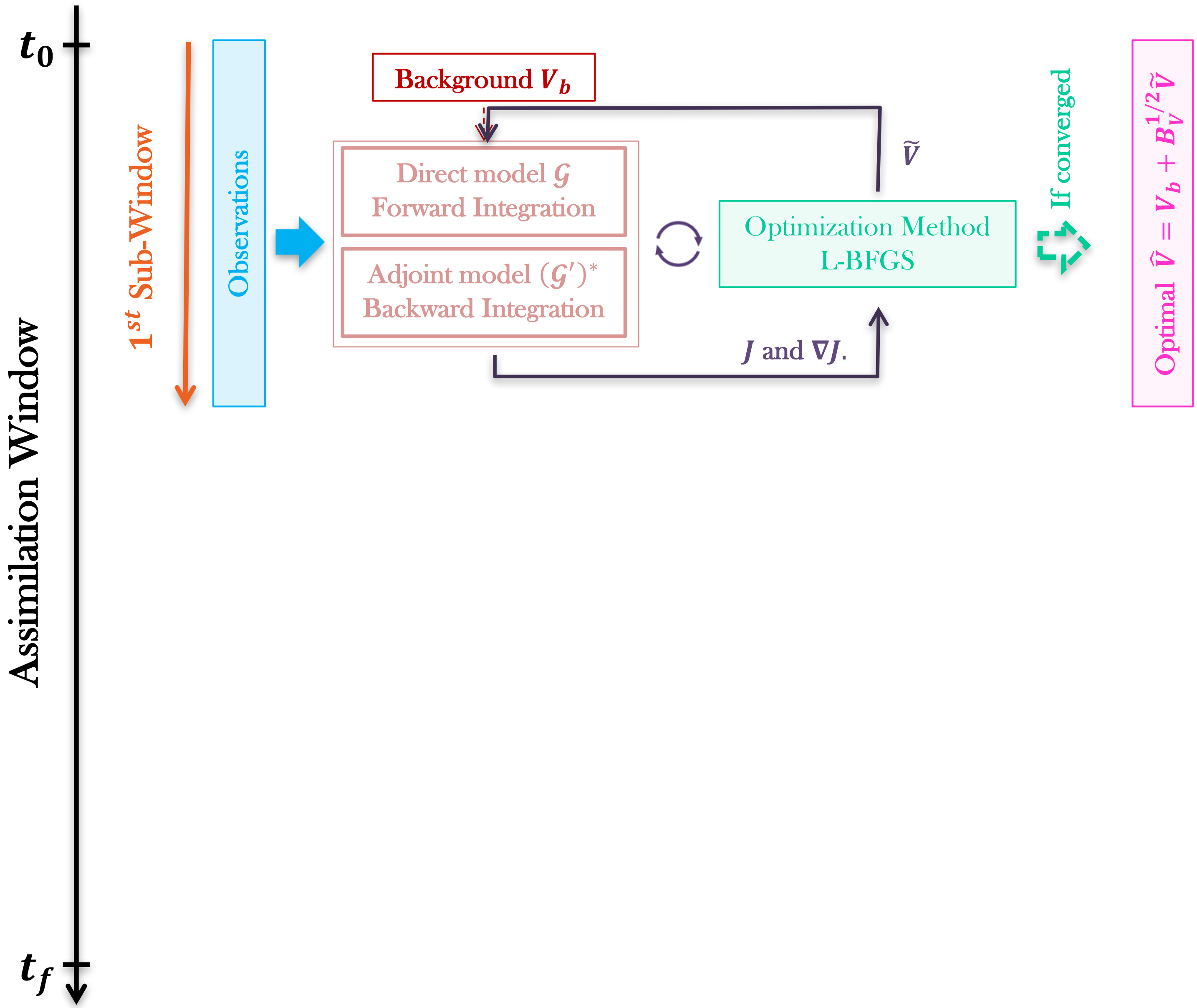


Sequential version of VDA



VARIATIONAL DATA ASSIMILATION – SIC4DVAR

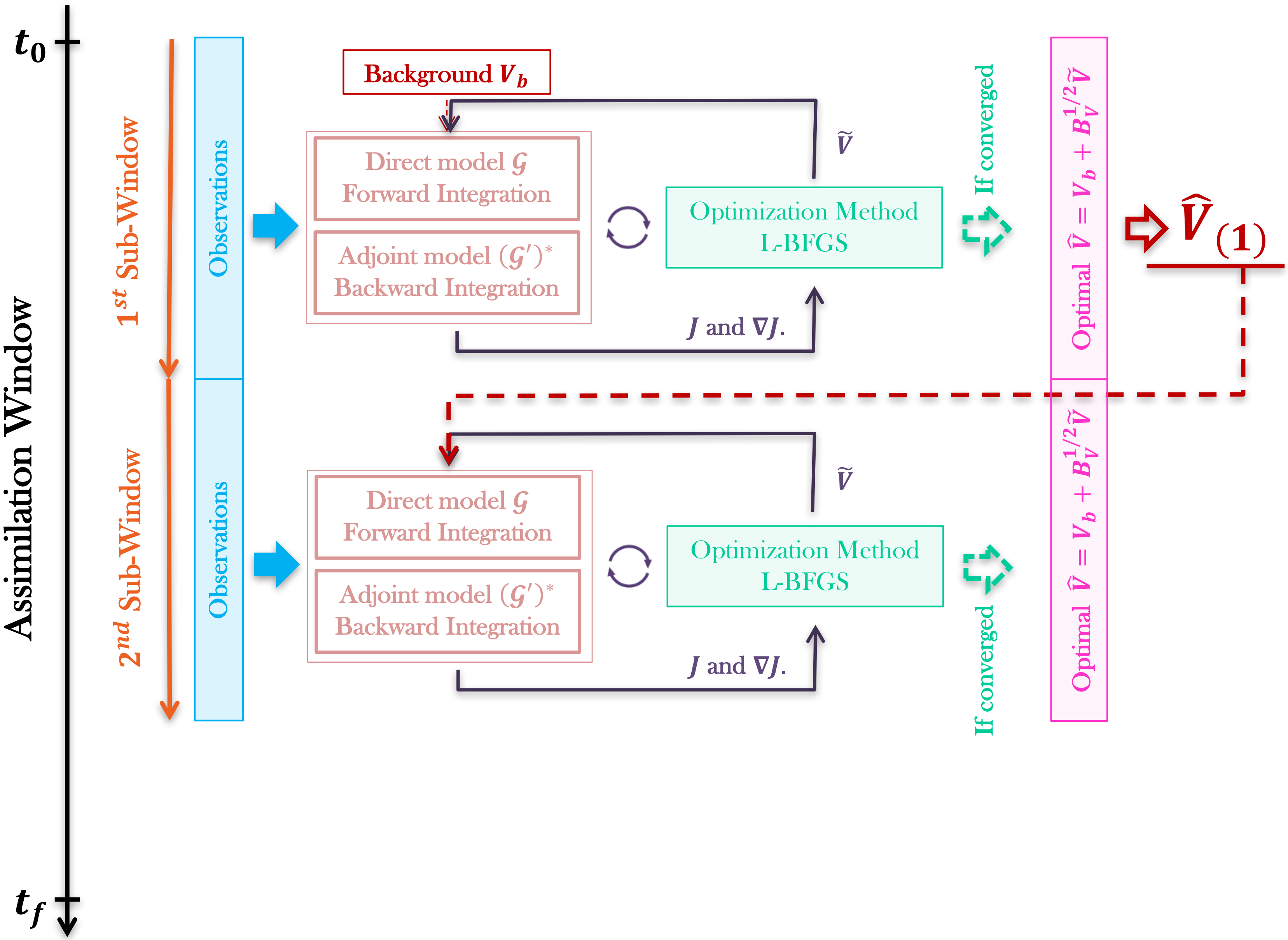
Sequential version of VDA



$\hat{V}_{(i)}$ is the optimal solution of the V at the sub-window i.

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

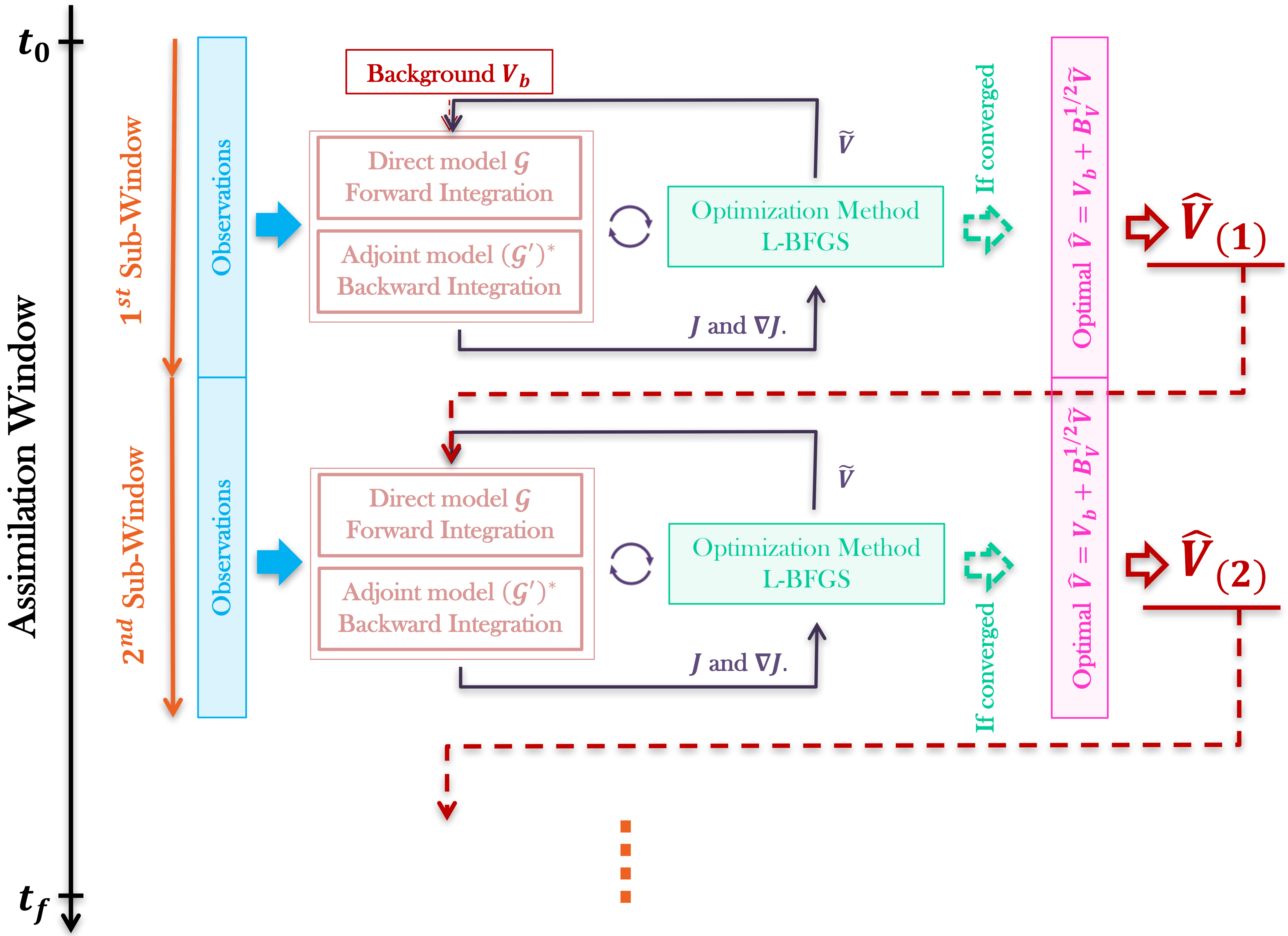
Sequential version of VDA



$\hat{V}_{(i)}$ is the optimal solution of the V at the sub-window i.

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

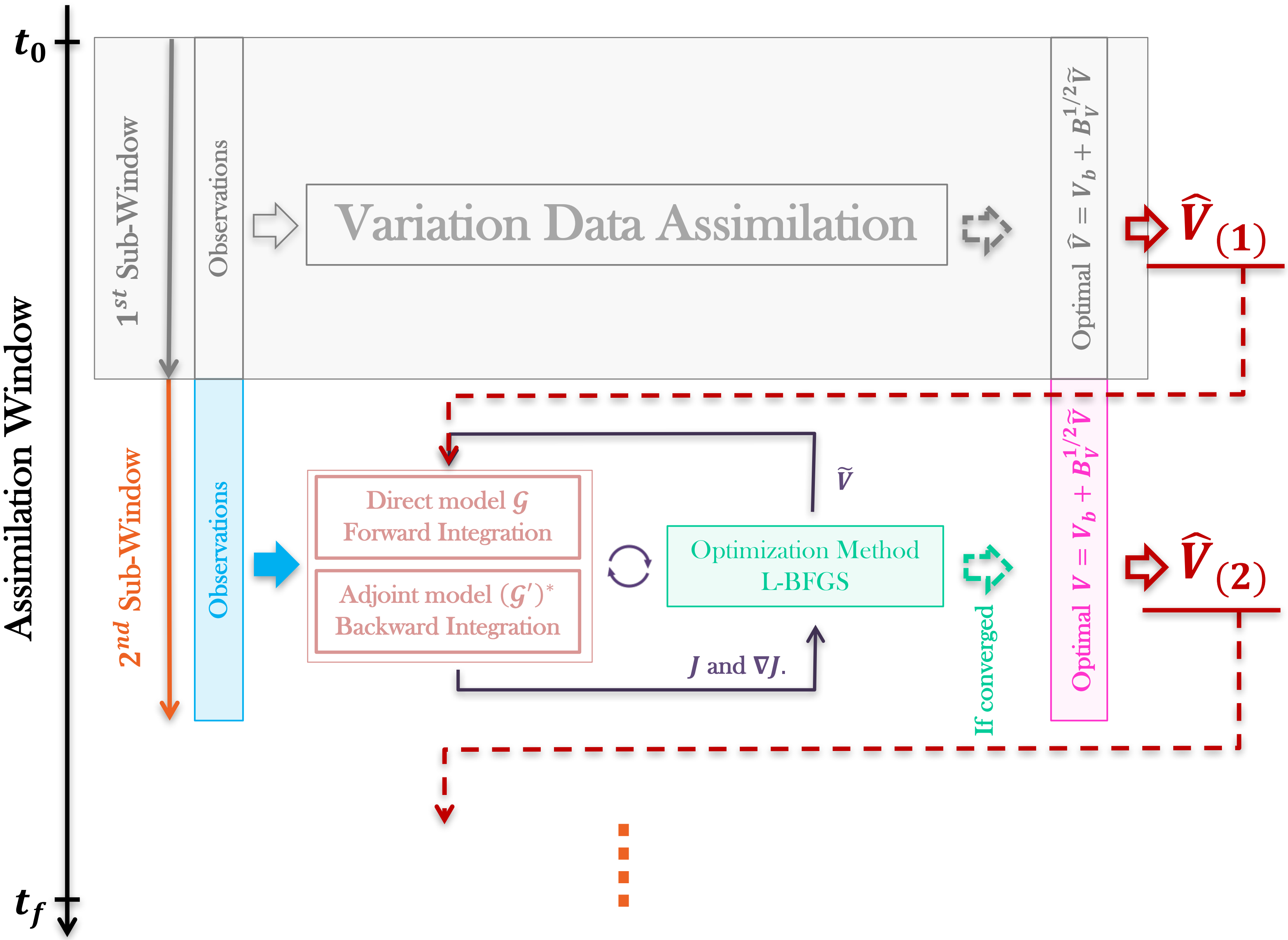
Sequential version of VDA



$\hat{V}_{(i)}$ is the optimal solution of the V at the sub-window i.

VARIATIONAL DATA ASSIMILATION – SIC4DVAR

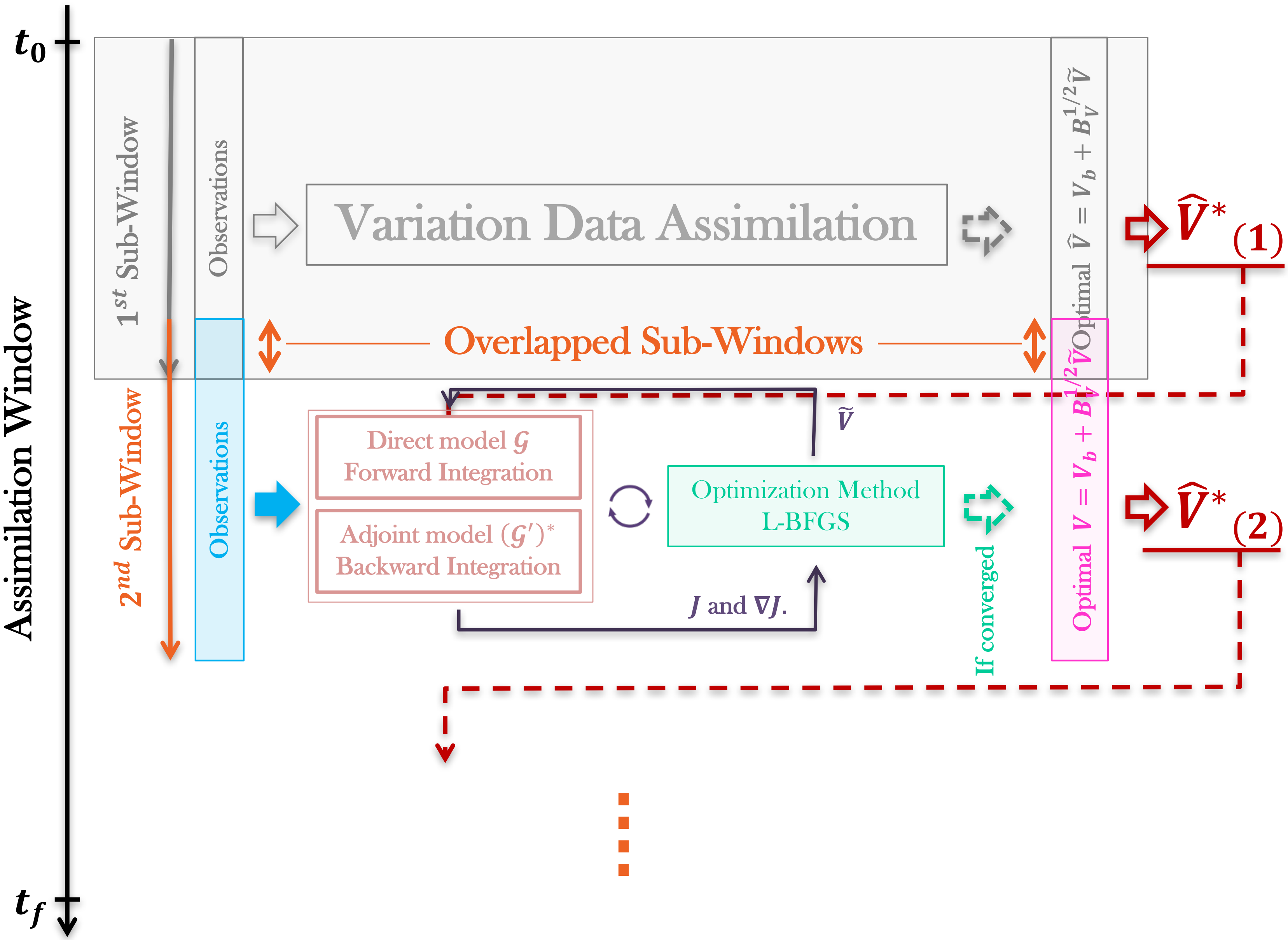
Sequential version of VDA



$\hat{V}_{(i)}$ is the optimal solution of the V at the sub-window i.

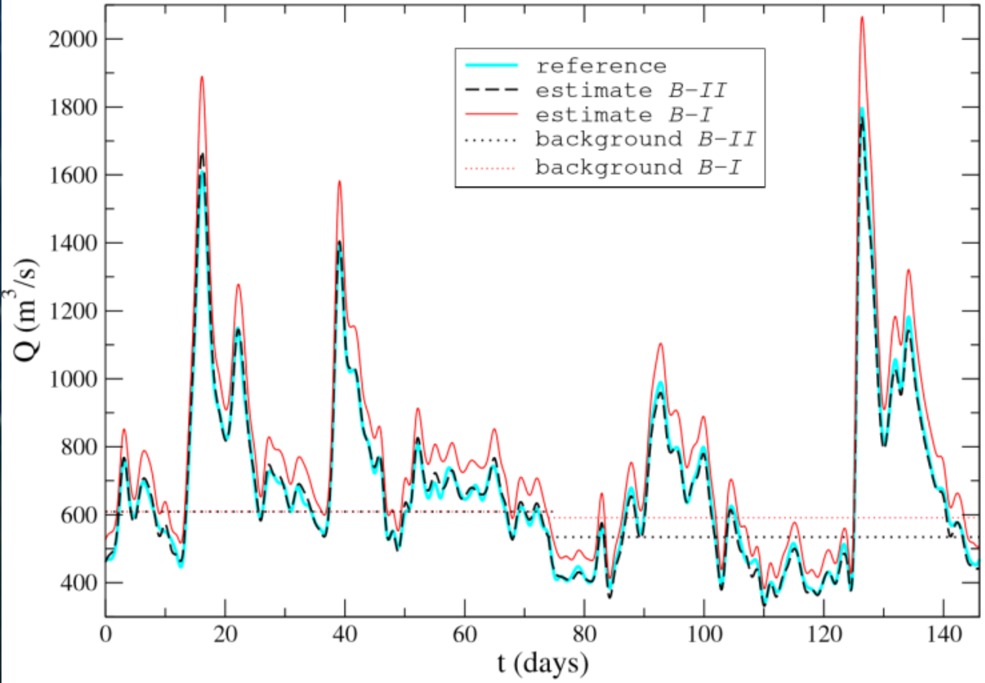
VARIATIONAL DATA ASSIMILATION – SIC4DVAR

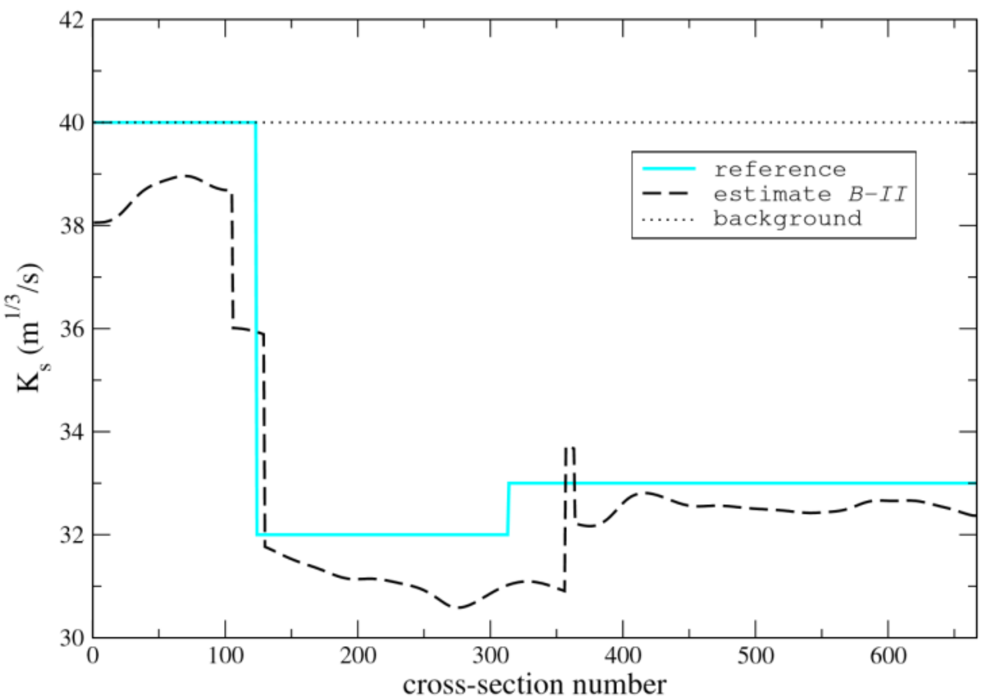
Sequential version of VDA



$\hat{V}_{(i)}$ is the optimal solution of the V at the sub-window i.

PEPSI CHALLENGE 1 : GARONNE RIVER



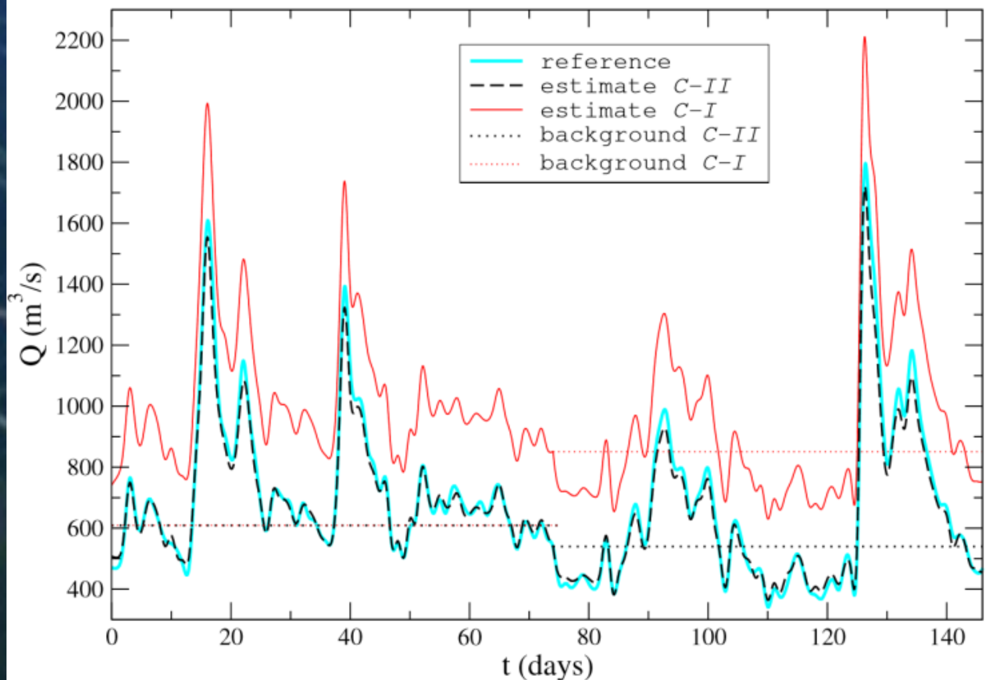


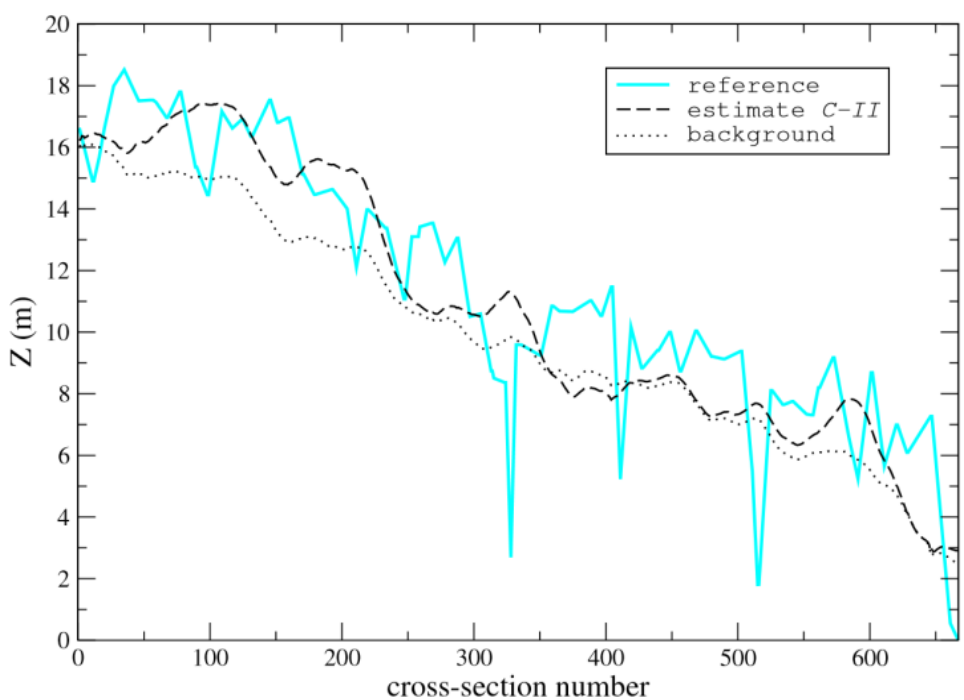
Oubanas et al. JoH 2018.

- Estimation of discharge and roughness coefficient $\{Q, K_S\}$.
- Bathymetry assumed **known.**
- **1-day** obs frequency.
- SWOT observation errors ($\sigma_z = 10\ cm$).

	B-I	**B-II**
Q_{rRMSE}	12.9%	2.6%
$K_{S\,rRMSE}$	20.4%	3.4%

PEPSI CHALLENGE 1 : GARONNE RIVER





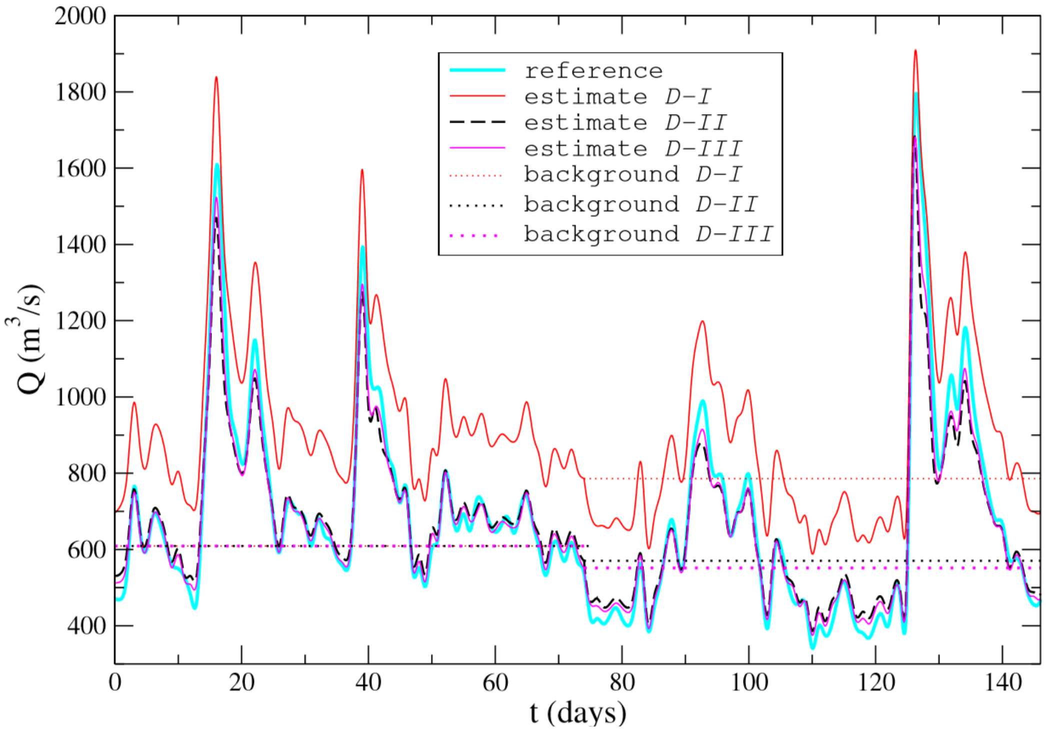
Oubanas et al. JoH 2018.

- Estimation of discharge and bed elevation $\{Q, Z_b\}$.
- Roughness assumed **known.**
- **1-day** obs frequency.
- SWOT observation errors ($\sigma_z = 10\ cm$).

	C-I	C-II
Q_{rRMSE}	50%	3,8%
$Z_{B_{rRMSE}}$	5,7%	4,9%

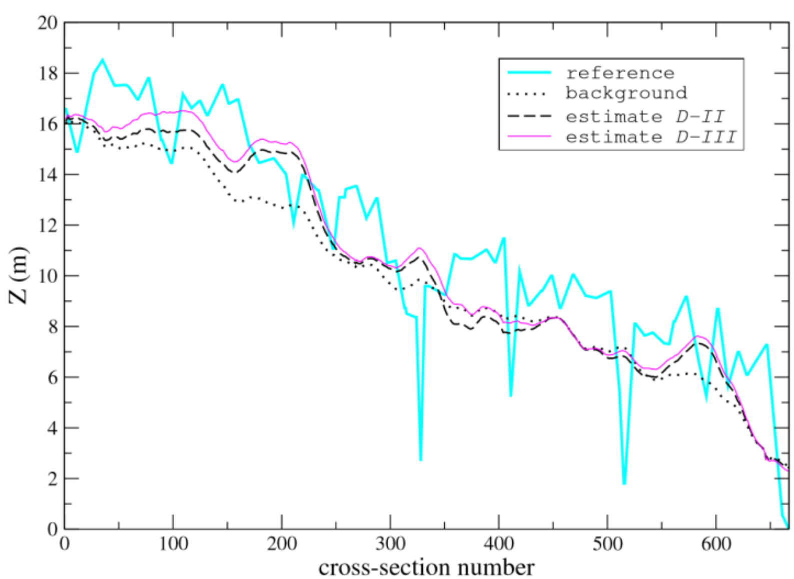
RIVER DISCHARGE ESTIMATION

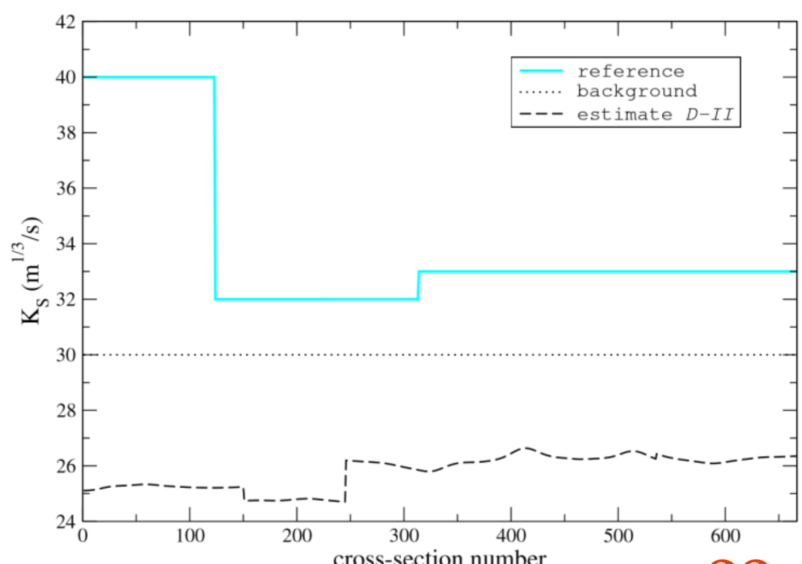
PEPSI CHALLENGE 1 : GARONNE RIVER



Oubanas et al. JoH 2018.

- Estimation of discharge roughness coefficient and bed elevation $\{Q, K_S, Z_b\}$.
- **1-day** obs frequency.
- SWOT observation errors ($\sigma_z = 10\ cm$).

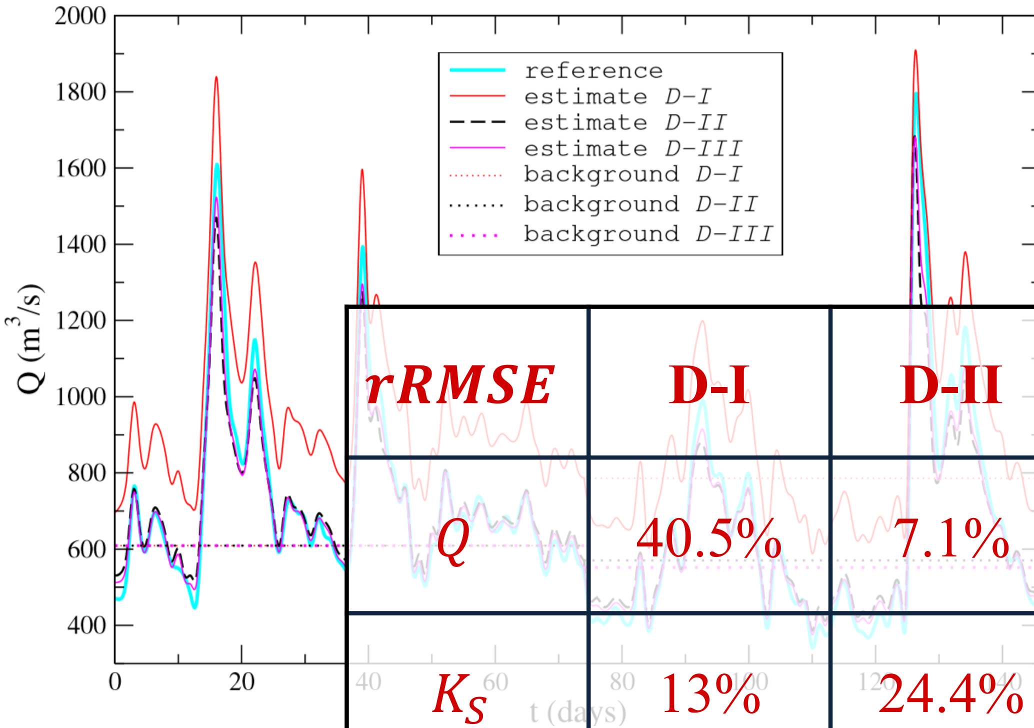




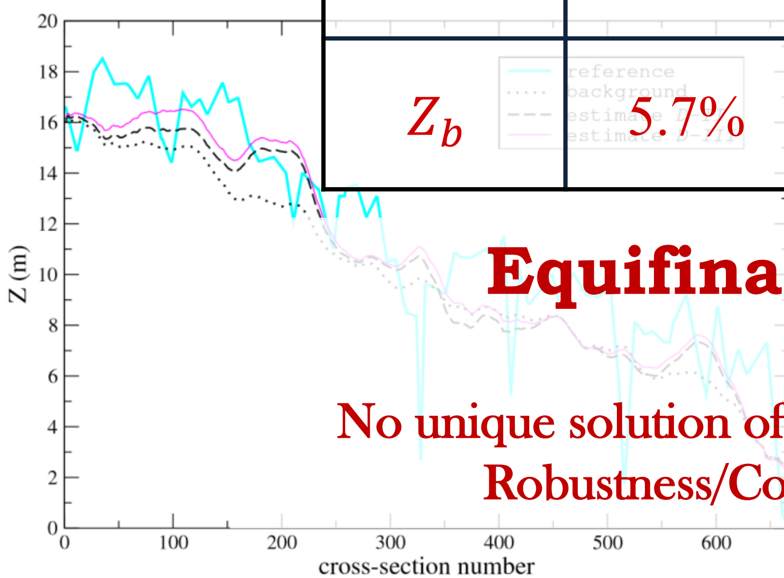
PEPSI CHALLENGE 1 : GARONNE RIVER

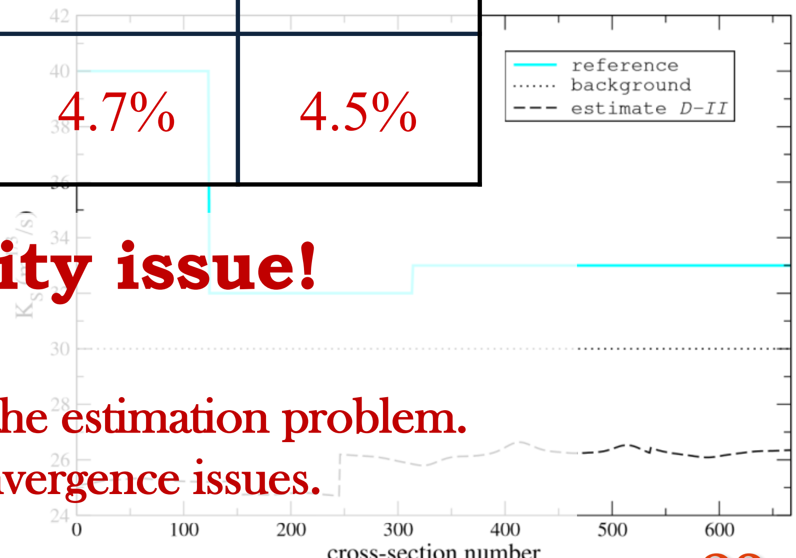
Oubanas et al. JoH 2018.

- Estimation of discharge roughness coefficient and bed elevation $\{Q, K_S, Z_b\}$.
- 1-day obs frequency.
- SWOT observation errors ($\sigma_z = 10\ cm$).



rRMSE	D-I	D-II	D-III
Q	40.5%	7.1%	5.1
K_S	13%	24.4%	13%
Z_b	5.7%	4.7%	4.5%

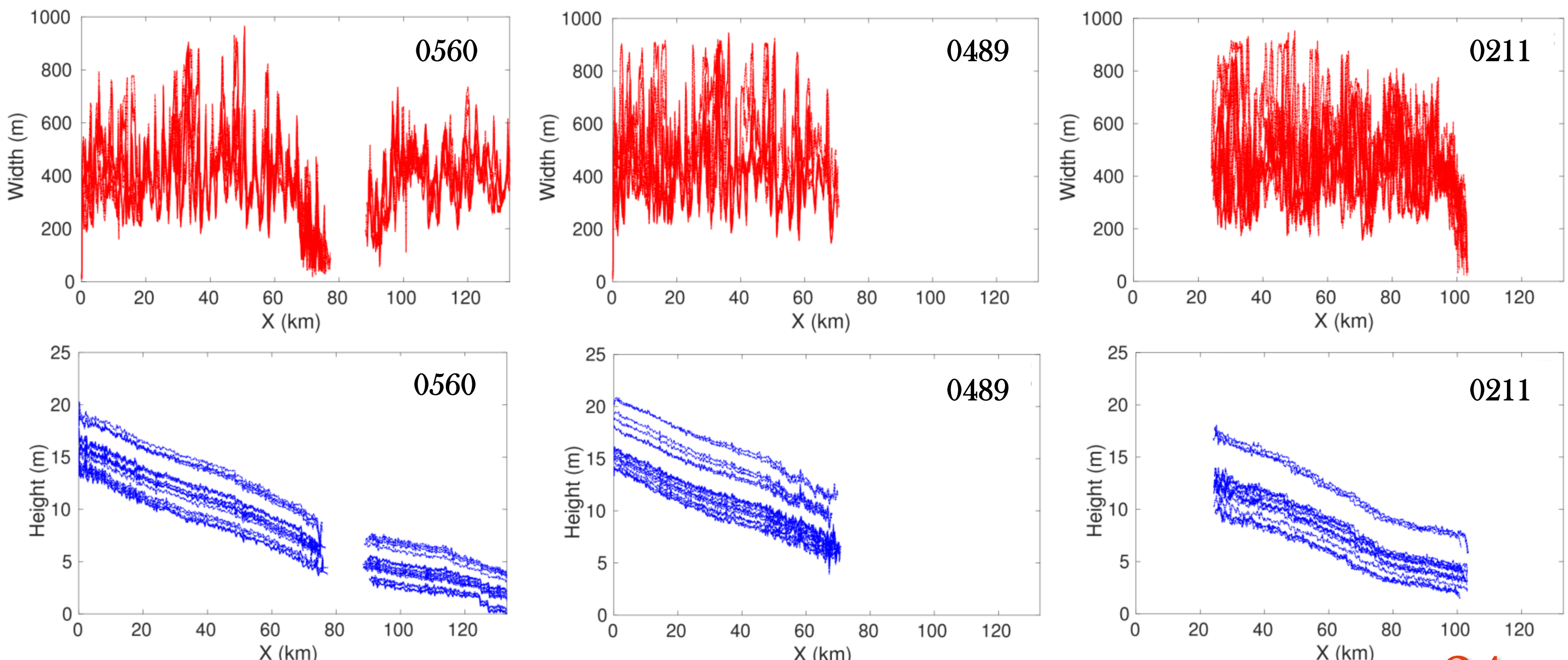




Equifinality issue!

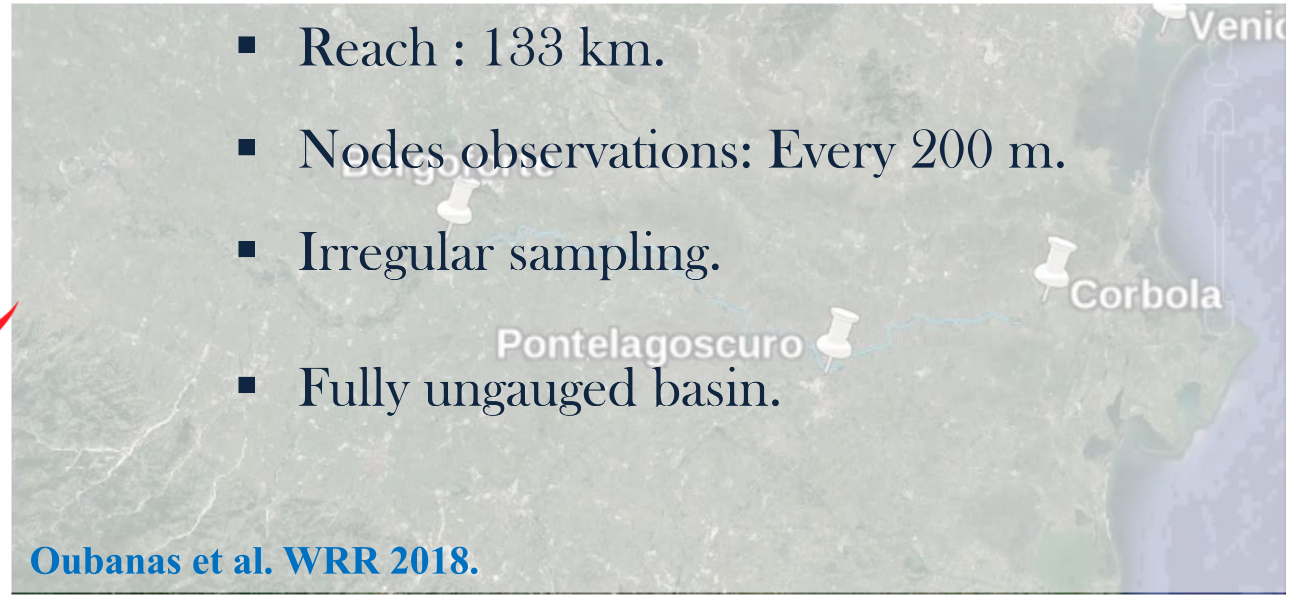
No unique solution of the estimation problem.
Robustness/Convergence issues.

SWOT HYDROLOGY SIMULATOR : PO RIVER

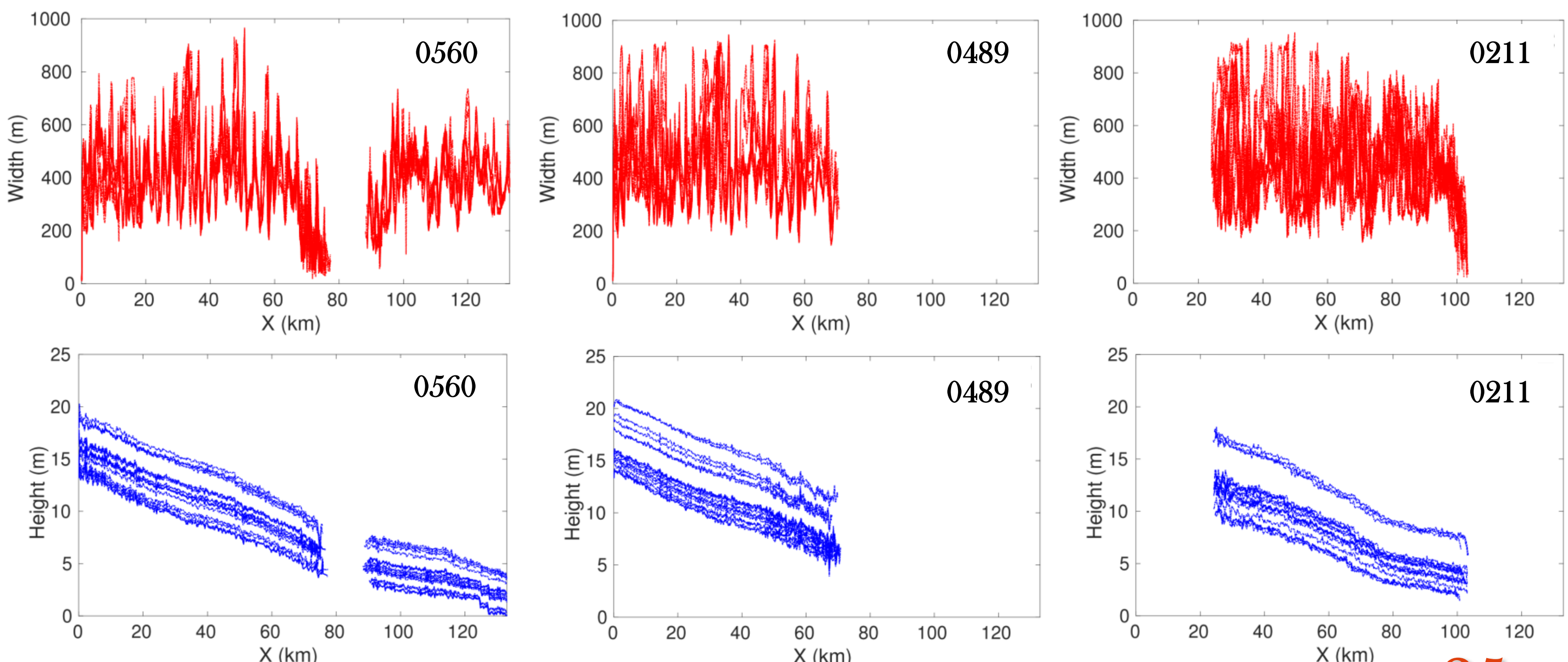


SWOT HYDROLOGY SIMULATOR : PO RIVER

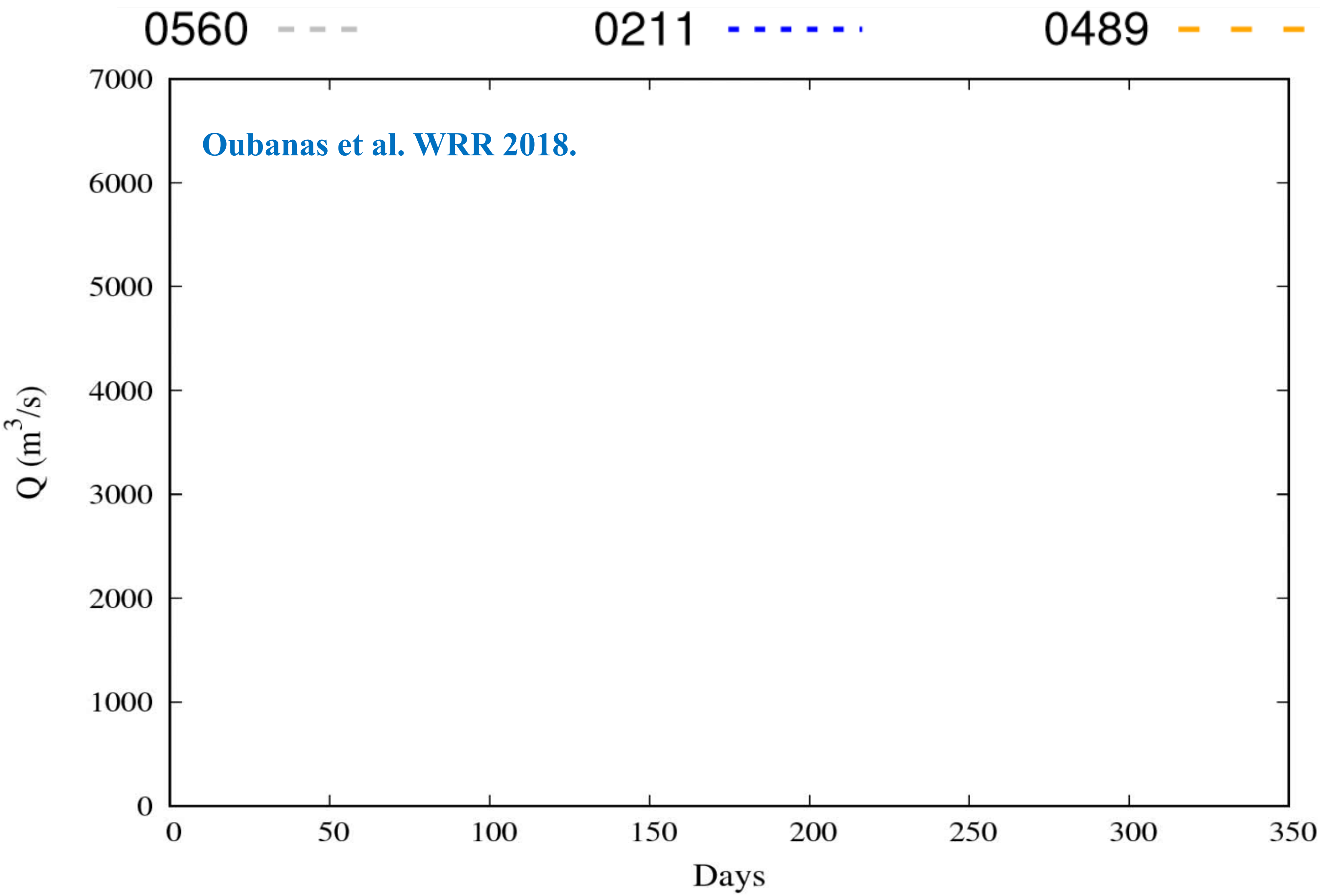
- Reach : 133 km.
- Nodes observations: Every 200 m.
- Irregular sampling.
- Fully ungauged basin.



Oubanas et al. WRR 2018.



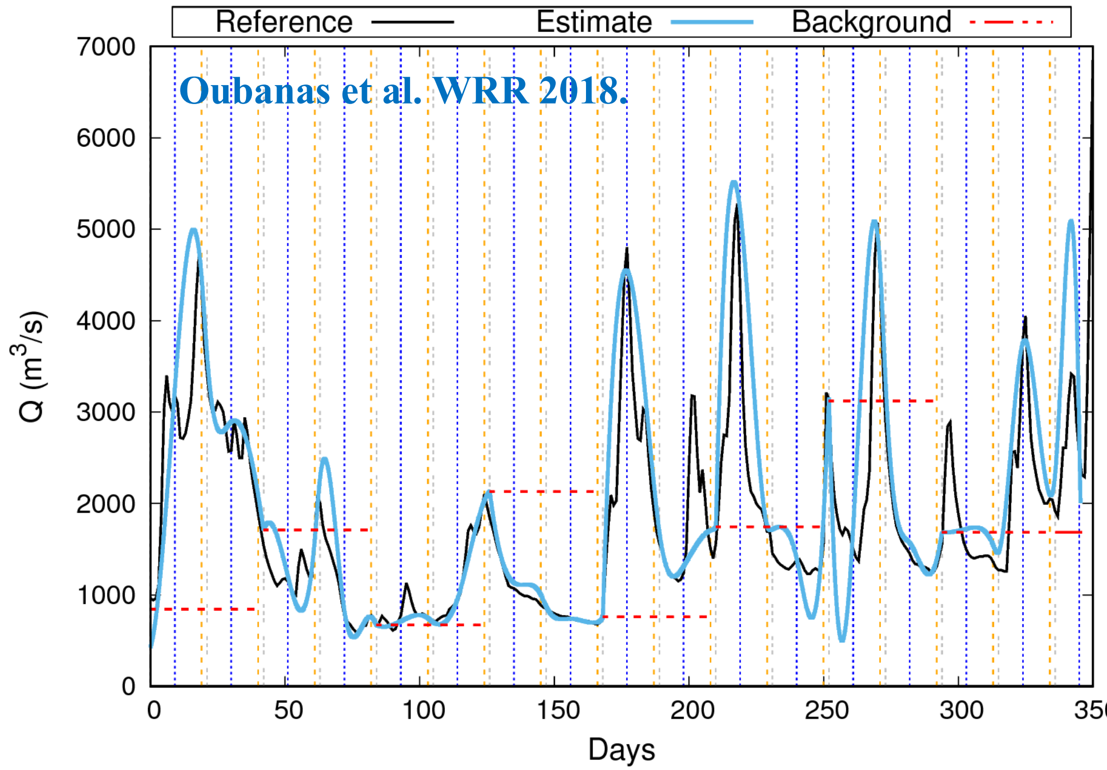
SWOT HYDROLOGY SIMULATOR : PO RIVER

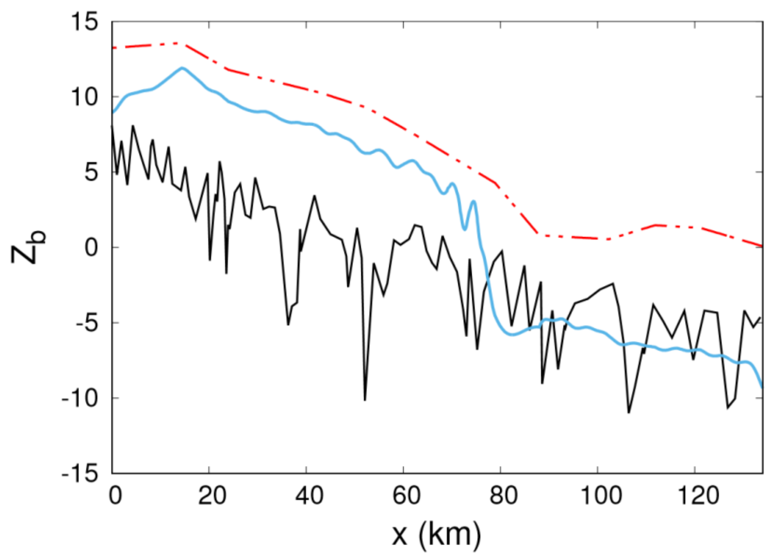


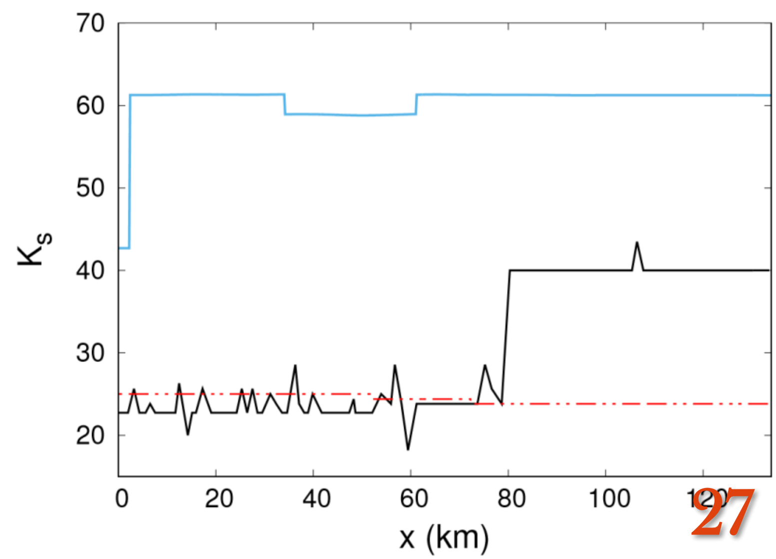
RIVER DISCHARGE ESTIMATION

SWOT HYDROLOGY SIMULATOR : PO RIVER

	1-day	ΔT_{obs}
RMSE ($m^3 s^{-1}$)	657,9	221,4
rRMSE (%)	29,8	12,6
NRMSE (%)	36,5	12,1
NSE	0,54	0,94
VE	0,77	0,91







RECENT DEVELOPMENTS

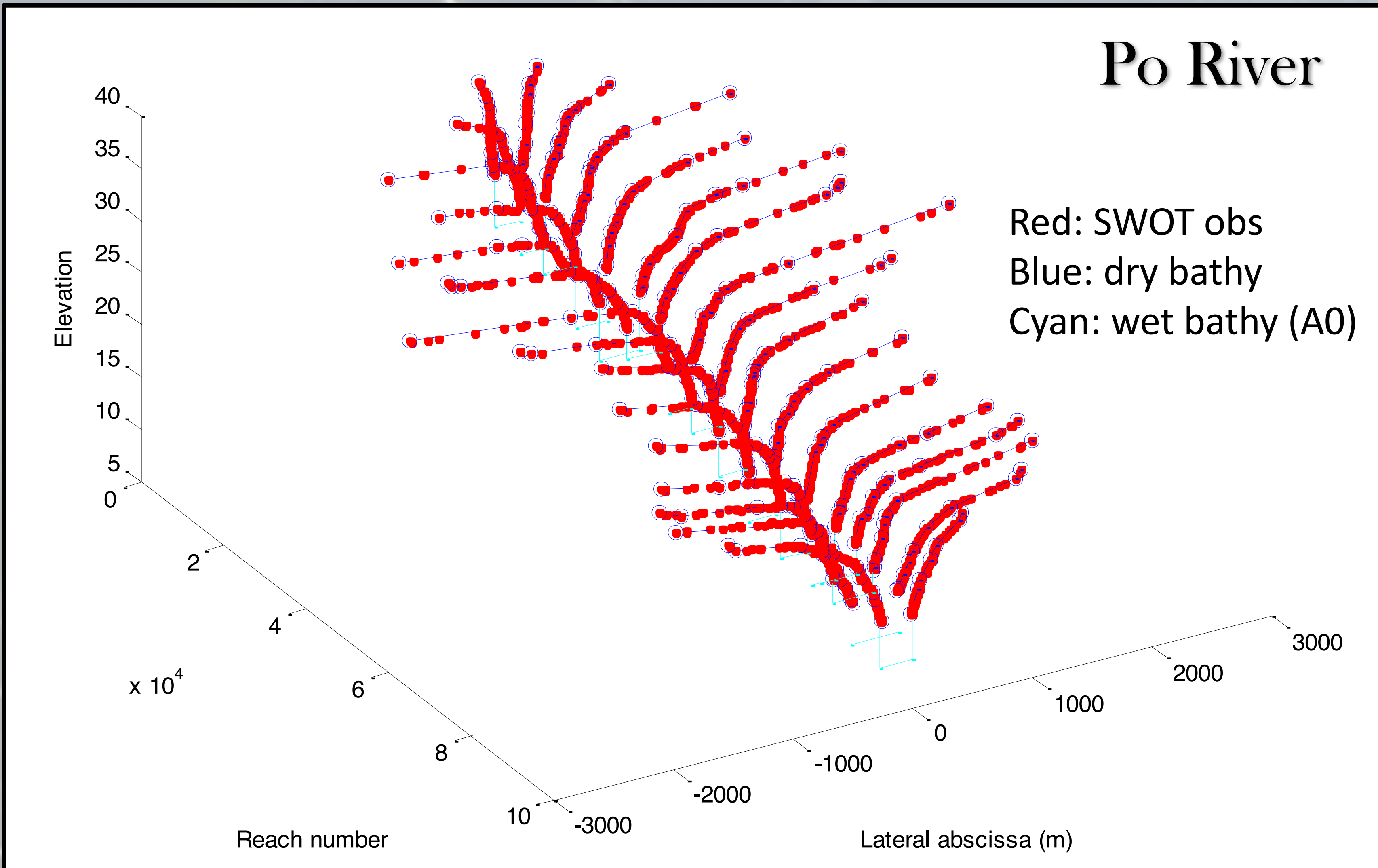
- SIC² hydrodynamic model :
  - Automatic building of the river model.
    - From SWOT type data : Pepsi 1 & 2
    - From GWD-LR (Yamazaki) & Hydroweb : Congo, Ubangi, Kasai, Sangha, etc.

- 4D-Var Algorithm :
  - New approach based on nuisance parameters (Gejadze et al 2017).
  - Robustness and stability.

- Prior/Background knowledge :
  - More realistic bathymetry prior (dry and wet bathy).
  - Method based on Manning's equation (S_f=S_w) & Q_{WBM} (A0 & n priors) (Durand et al 2016, Oubanas et al 2018) and new alternative one.
  - Daily discharge prior (S_f=S_w & Q_{WBM}).

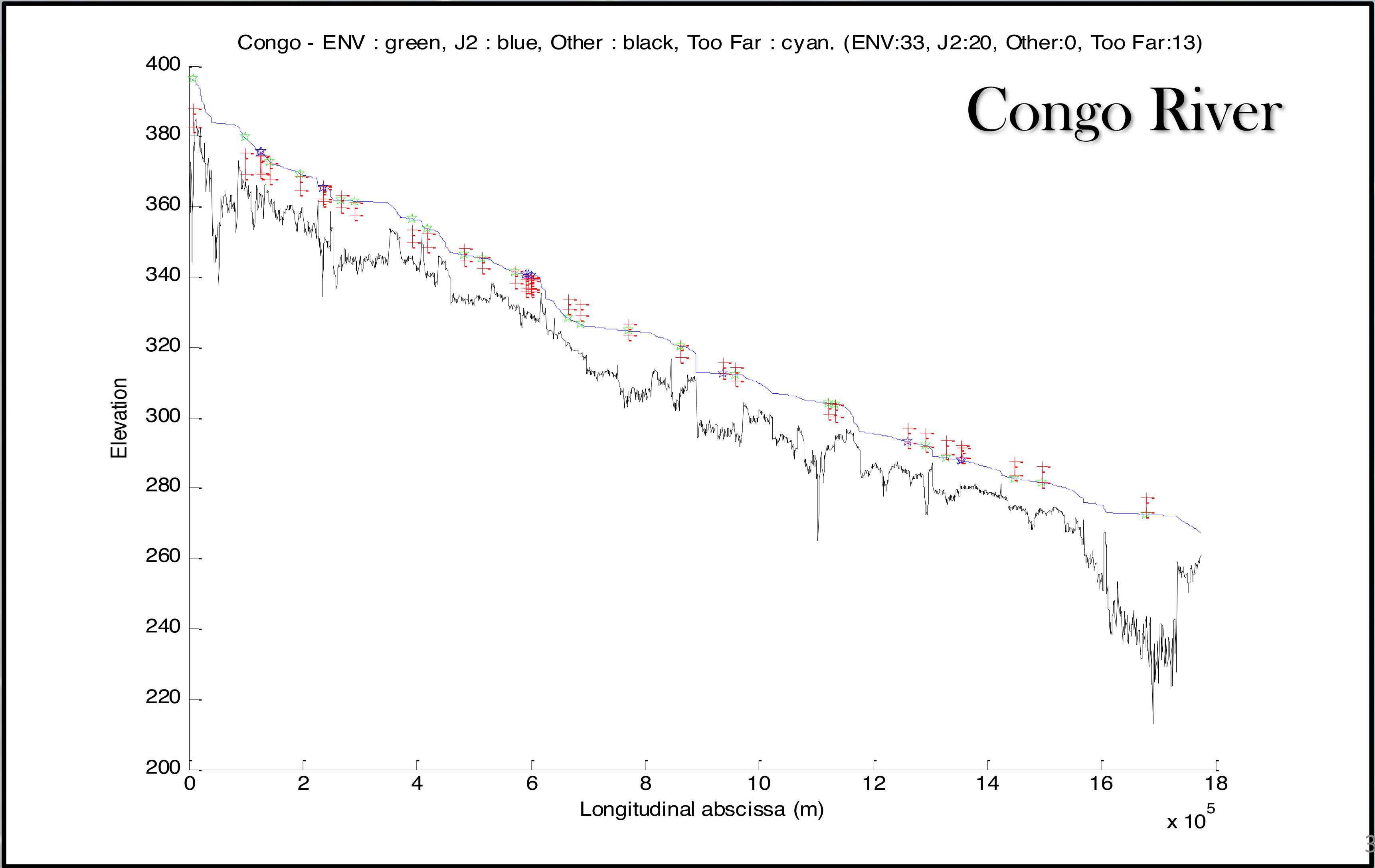
RECENT DEVELOPMENTS

Wet & dry bathymetry from SWOT type data



RECENT DEVELOPMENTS

from GWD-LR Yamazaki & Hydroweb



INVESTIGATED EXPERIMENTAL SET-UPS

- River **discharge** estimation under uncertainty in **bathymetry** and **roughness** with **exact or noisy** observations ($\sigma = [.05\ 5\ 0.1e-5]$, cf. Durand et al 2016).
- Choice of the control vector (Q), (Q, bathy), (Q, friction), (Q, bathy, friction) (cf Oubanas et al 2018, JoH). Choice of sliding windows size, etc. …
- Generation of river **prior/background variables** (wet bathy & friction) using different approaches based on the Manning's equation (S_f=S_w & Q_{WBM}).
- **River discharge prior/background** (S_f=S_w & Q_{WBM}, constant, etc.).
- Management of "good" and "**bad**" reaches
- **Bathymetry updates** using different approaches (fixed/changing cross-section shape of the wet/dry bathymetry).

GLOBAL APPLICABILITY

- SIC^2 hydrodynamic model
  - Ok for Multiple beds (minor, medium, major).
  - Ok for Artificial structures / River channel tributaries.
  - Ok for intermediate pools.
  - Need simplified configurations for braided rivers.
  - No strong 2D effects.
  - No dry bed / No supercritical flows (or not too much, cf Fr).

GLOBAL APPLICABILITY

- SIC² hydrodynamic model
  - Ok for Multiple beds (minor, medium, major).
  - Ok for Artificial structures / River channel tributaries.
  - Ok for intermediate pools.
  - Need simplified configurations for braided rivers.
  - No strong 2D effects.
  - No dry bed / No supercritical flows (or not too much, cf Fr).

- 4DVar Algorithm :
  - Sequential version : No limit (memory, CPU)

GLOBAL APPLICABILITY

- SIC² hydrodynamic model
  - Ok for Multiple beds (minor, medium, major).
  - Ok for Artificial structures / River channel tributaries.
  - Ok for intermediate pools.
  - Need simplified configurations for braided rivers.
  - No strong 2D effects.
  - No dry bed / No supercritical flows (or not too much, cf Fr).

- 4DVar Algorithm
  - Sequential version : No limit (memory, CPU)

- System selection / Prior
  - Selection of a series of reaches based on hydraulic conditions.
  - No artificial managed structures (or rule to be known).
  - No significant tributaries (or needed info on them).

CPU
Time Estimation

- Less than 1 mn per day on a 100-200 km long river on a average laptop (32 GB RAM Intel Core i7):
  - SWOT reach data base = 200000 reaches ?
  - River portions = 200000/20 = 10000 river portions (about 200 km each, 20 SWOT reaches each).
  - CPU=10000*21=210000 minutes=3500 hours per 21 day-cycle.
  - If CPU is spread over the 21 days : 7 laptops
  - If we run a sliding windows of 21 days every day (RT) : 7*21 laptops ≈ 150 cores

CPU
Time Estimation

- Less than 1 mn per day on a 100-200 km long river on a average laptop (32 GB RAM Intel Core i7).
  - SWOT reach data base = 200000 reaches ?
  - River portions = 200000/20 = 10000 river portions (about 200 km each, 20 SWOT reaches each).
  - CPU=10000*21=210000 minutes=3500 hours per 21 day-cycle.
  - If CPU is spread over the 21 days : 7 laptops
  - If we run a sliding windows of 21 days every day (RT) : 7*21 laptops ≈ 150 cores

- So far the algorithm is not optimized on the CPU nor memory aspects.

GLOBAL APPLICABILITY CONCLUSION?

- Not a CPU nor memory problem
- More:
  - Automatic generation of models (connectivity database)
  - Performance of the SIC4DVar algo (under SWOT conditions, space and time frequency, noise)
  - Robustness of SIC^2 (OK)
  - Robustness of 4DVar (not so bad, still under improvement)

  - All this is under investigation (Pepsi 1 & 2, Congo …)

GENERAL CONCLUSION

- With some in-situ data: “easy” problem
  - Ex: Bathymetry
  - Ex: in-situ rating curve somewhere along the river

- With **no** in-situ data: **difficult** problem
  - That is what makes it very interesting
  - Push the limits of SIC4DVar …
  - We still have difficulties …
  - … but we still have ideas to solve them …
  - … and are interested by yours …

PAPERS

Oubanas, H., Gejadze, I., Malaterre, P.O., Durand, M., Wei, R., Frasson, R.P.M. and Domeneghetti, A., **2018b**. Discharge Estimation in Ungauged Basins Through Variational Data Assimilation: The Potential of the SWOT Mission. Water Resources Research. https://doi.org/10.1002/2017WR021735

Oubanas. H., Gejadze, I., Malaterre, P.O. and Mercier, F., **2018a**. River discharge estimation from synthetic SWOT-type observations using variational data assimilation and the full Saint-Venant hydraulic model. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2018.02.004

Gejadze, I., Oubanas, H. and Shutyaev, V., **2017**. Implicit treatment of model error using inflated observation error covariance. Quarterly Journal of the Royal Meteorological Society, 143(707), pp.2496-2508. https://doi.org/10.1002/qj.3102

Gejadze, I.Y. and Malaterre, P.O., **2016a**. Design of the control set in the framework of variational data assimilation. Journal of Computational Physics, 325, pp.358-379.https://doi.org/10.1016/j.jcp.2016.08.029

Gejadze, I. and Malaterre, P.O., **2016b**. Discharge estimation under uncertainty using variational methods with application to the full Saint-Venant hydraulic network model. International Journal for Numerical Methods in Fluids, 83(5), pp.405-430. https://doi.org/10.1002/fld.4273

MERCI !

THANK YOU !

Many thanks to TOSCA CNES, CLS, Irstea, AFD funds