

Improvement of operational flood forecasting through the assimilation of satellite observations and multiple river flow data

Fabio Castelli¹ and Giulia Ercolani¹

¹Department of Civil and Environmental Engineering, University of Florence, Florence, Italy

Correspondence to: Giulia Ercolani (giulia.ercolani@dicea.unifi.it)

Abstract. Data assimilation has the potential to improve flood forecasting. However, it is rarely employed in distributed hydrologic models for operational predictions. In this study, we present variational assimilation of river flow data at multiple locations and of land surface temperature (LST) from satellite in a distributed hydrologic model that is part of the operational forecasting chain for the Arno river, in central Italy. LST is used to estimate initial condition of soil moisture through a coupled surface energy/water balance scheme. We present here several hindcast experiments to assess the performances of the assimilation system. The results show that assimilation can significantly improve flood forecasting, although in the limit of data error and model structure.

1 Introduction

The potential of data assimilation in hydrology has been demonstrated by several studies (e.g., Clark et al., 2008; Seo et al., 2009; Brocca et al., 2010; Lee et al., 2012; Laiolo et al., 2015). However, the usage of data assimilation in distributed hydrologic models for operational flood forecasts is limited by many issues: non-linear and discontinuous model structure, non-Gaussian/multiplicative errors, large dimensionality of the inverse problem, model governed by different equations, complex topology of domains such as surface drainage and river network. Moreover, the majority of studies investigates capabilities of data assimilation through synthetic experiments, while applications conducted from an operational perspective are rare, although the need for an effective transition of research advances into operational forecasting systems has been increasingly claimed in recent years (Liu et al., 2012). This work presents variational assimilation of flow data at multiple locations and of Land Surface Temperature (LST) maps from satellite in a distributed hydrologic model that is part of the operational forecasting chain for the Arno river, in central Italy. We assess the actual gain that can be obtained in flood predictions through data assimilation in order to answer the unsolved doubts of whether the system has enough long memory, or hydrologic models are sufficiently adherent to reality, to significantly benefit from more accurate initial conditions. We show results from several hindcast experiments in the Arno basin.

2 The hydrologic model Mobidic

Mobidic (MOdello di Bilancio Idrologico DIstribuito e Continuo) is a physically-based hydrologic model Campo et al. (2006); Castelli et al. (2009); Yang et al. (2014a, b). It is continuous in time and distributed in space, with raster-based horizontal discretization. Mobidic solves a coupled mass and energy balance at the surface and employs a computationally efficient representation of soil moisture dynamics, that has been recently improved in Castillo et al. (2015). It has a single layer scheme for soil, whose peculiarity is the conceptual subdivision of the layer into two non-linear reservoirs, the capillary and gravitational one. They correspond to larger pores that drains under gravity and smaller pores that hold water through capillary forces. The two components control different sets of hydrologic fluxes. In particular, available water for evapotranspiration is capillary water, that is fed by gravitational water through an absorption flux. Interactions between surface and subsurface hydrology are explicitly taken into account. Groundwater dynamics may be modeled through 2-D Dupuit approximation or as a linear reservoir. The latter method is employed in the present work. Three options are available for flow routing through the network, i.e. the lag approach, the Muskingum-Cunge method and the cascade of liner reservoirs. Mobidic runs operatively at the hydrologic service of Tuscany region (Servizio Idrologico Regionale, Regione Toscana) for floods forecasting and water resources management purposes.

3 The assimilation scheme

For the assimilation system the variational approach has been selected, since it requires less restrictive hypothesis than Kalman and Montecarlo filters and smoothers. The payback is the need for an adjoint model, that is challenging to derive for distributed hydrologic models. Separate adjoint models have been derived for Mobidic's modules of flow routing through the river network and of water and energy balance at the soil surface. The first one is devoted to the assimilation of discharge data at multiple locations, while the second one deals with LST observations from satellite.

3.1 Adjoint model of flow routing in the river network

The adjoint model of flow routing has been developed for the method of linear reservoirs in cascade, which represents the optimal compromise between complexity and representativeness of the physical process. Hence, the routing is driven by:

$$\frac{dQ}{dt} = \mathbf{A}(q_L + \mathbf{U}Q - Q) = \mathbf{F}(\mathbf{A}, Q, q_L) \quad (1)$$

where, considering a network composed by n reaches, $Q \in \mathbb{R}^n$ are the discharges exiting each reach, $q_L \in \mathbb{R}^n$ are the lateral inflows (surface runoff plus groundwater flow), and $\mathbf{A} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with the inverse of the characteristic time of each river on the diagonal. Lastly, $\mathbf{U} \in \mathbb{R}^{n \times n}$ is a binary matrix accounting for network topology. Following a classical variational approach (Castelli et al., 1999; Caparrini et al., 2004; Sun, 2013), the assimilation scheme provides optimal estimates of discharge initial condition and input by minimizing a penalty functional J . J contains squared errors between states predictions and observations and between current and previous values of the quantities to optimize. The physical constraint is

imposed by adjoining Equation (1) to J through a vector of Lagrange multipliers $\lambda \in \mathbb{R}^n$.

$$J = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} (Q - Q^{obs})^T \frac{\mathbf{K}_Q}{2} (Q - Q^{obs}) dt + (Q - Q^{obs})^T \frac{\mathbf{K}_Q}{2} (Q - Q^{obs}) \Big|_{t_1} + (Q_0 - Q'_0)^T \frac{\mathbf{K}_{Q_0}}{2} (Q_0 - Q'_0) \\ + \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} (q_L - q_L')^T \frac{\mathbf{K}_{q_L}}{2} (q_L - q_L') dt + \int_{t_0}^{t_1} \lambda^T \left(\frac{dQ}{dt} - F(\mathbf{A}, Q, q_L) \right) dt \quad (2)$$

5 In eq. (2), $\mathbf{K}_Q, \mathbf{K}_{Q_0}, \mathbf{K}_{q_L} \in \mathbb{R}^{n \times n}$ are weighting factors applied to the various terms composing J , $Q^{obs} \in \mathbb{R}^n$ are the observed discharges available inside the assimilation window $[t_0, t_1]$, the sign $(\cdot)'$ indicates the previous value of the quantity. J is minimized if its first variation δJ vanishes, condition that, after some computations, leads to a system of ordinary differential equations that describes the time evolution of Lagrange multipliers. Furthermore, a terminal condition for the backward integration of the adjoint model and update equations for initial streamflow Q_0 and lateral input q_L that depends on λ are
10 obtained. The optimal estimate of Q_0 and q_L is obtained through an iterative procedure constituted by subsequent integrations of forward and adjoint model and corresponding updates. Iterations are interrupted when updates become negligible. To note that the corrections evaluated at discharge measurement stations spread upstream thanks to the coupling between equations of flow channel routing (see eq. (1)). Since our aim is to improve both flow routing and runoff formation processes, the equations driving these latter should be included in the derivation of the adjoint model. However, it would be an extremely challeng-
15 ing task, mainly because of the threshold processes that characterize soil moisture dynamics. Therefore, we employ a mixed variational-Montecarlo approach. First, through the flow routing adjoint model, we estimate the optimal temporal evolution of lateral inflow q_L . Then, on its basis, we infer key variables that determine runoff formation through a parsimonious Montecarlo approach. As key variables, we selected initial condition of water content in capillary soil and rainfall intermittence, that is represented by the parameter f_0 (frequency of no-rainfall during a certain time step, see Castelli (1996)). An ensemble cover-
20 ing the extent of the assimilation window is generated for q_L by reasonably varying initial soil moisture and f_0 . Both initial capillary water and f_0 are maintained spatially homogeneous, and hence the size of the ensemble remains small (typically around 100 realizations). At each iteration of the assimilation procedure, the realization with the minimum distance from the desired trajectory of q_L is selected for any single reach, and the corresponding initial capillary water and f_0 are adopted for the contributing cells. Hence, a spatially distributed estimate of both quantities is obtained.

25 3.2 Adjoint model of soil water and energy balance

As described in Section 2, Mobidic solves a coupled water and energy balance at the surface. Hence, it reproduces the partitioning of available energy between latent and sensible heat flux that is determined by soil moisture. This characteristic allows to estimate soil moisture initial condition through the assimilation of LST observations from satellite. The advantage of assimilating LST maps instead of soil moisture products lies in their higher spatio-temporal resolution and accuracy (Campo et al.,
30 2012).

Among Mobidic's driving equations, a set of three ODEs can be identified as significant for the assimilation of LST. The system is here reported in synthetic form to highlight the relevant dependencies.

$$\frac{dT_s}{dt} = F_1(T_s, T_d, H(T_s, \dots), LE(T_s, W_c, \dots)) \quad (3)$$

$$\frac{dT_d}{dt} = F_2(T_s, T_d, \dots) \quad (4)$$

$$5 \quad \frac{dW_c}{dt} = F_3(W_c, W_g, LE(T_s, W_c, \dots)) \quad (5)$$

where H and LE are surface turbulent heat fluxes of sensible and latent heat respectively, T_s is the land surface temperature, T_d is the temperature of a deeper layer of soil, W_c and W_g are capillary and gravitational soil water content. The system can be solved pixel by pixel, being the states of a specific cell independent from those of the others. In the developed assimilation framework T_s , W_c and W_g are the analyzed states, whose optimal estimation is obtained by minimization of the penalty

10 functional J :

$$\begin{aligned} J = & \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} (\mathbf{T}_s - \mathbf{T}_s^{obs})^T \frac{\mathbf{K}_{T_s}}{2} (\mathbf{T}_s - \mathbf{T}_s^{obs}) dt + (\mathbf{T}_s - \mathbf{T}_s')^T \frac{\mathbf{K}_{T_s}}{2} (\mathbf{T}_s - \mathbf{T}_s') \Big|_{t_0} + (\mathbf{T}_d - \mathbf{T}_d')^T \frac{\mathbf{K}_{T_d}}{2} (\mathbf{T}_d - \mathbf{T}_d') \Big|_{t_0} \\ & + (\mathbf{W}_c - \mathbf{W}_c')^T \frac{\mathbf{K}_{W_c}}{2} (\mathbf{W}_c - \mathbf{W}_c') \Big|_{t_0} + \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} (\mathbf{W}_g - \mathbf{W}_g')^T \frac{\mathbf{K}_{W_g}}{2} (\mathbf{W}_g - \mathbf{W}_g') dt \\ & + \int_{t_0}^{t_1} \lambda_1^T \left(\frac{d\mathbf{T}_s}{dt} - \mathbf{F}_1 \right) + \lambda_2^T \left(\frac{d\mathbf{T}_d}{dt} - \mathbf{F}_2 \right) + \lambda_3^T \left(\frac{d\mathbf{W}_c}{dt} - \mathbf{F}_3 \right) dt \quad (6) \end{aligned}$$

15 where all the state variables are vector in $\in \mathbb{R}^n$, being n the number of cells composing the basin. \mathbf{K}_{T_s} , \mathbf{K}_{T_d} and $\mathbf{K}_{W_c} \in \mathbb{R}^{n \times n}$ are weighting factors applied to the various terms of the penalty functional, $\mathbf{T}_s^{obs} \in \mathbb{R}^n$ are measurements of LST available inside the assimilation window $[t_0, t_1]$ and the sign $(\cdot)'$ indicates the previous value of the quantity. The penalty functional is formed by quadratic errors in respect to both measurements and previous values of the quantity to estimate. The last term is the physical constrain adjoined through Lagrange multipliers λ_1 , λ_2 and $\lambda_3 \in \mathbb{R}^n$, each one corresponding to a specific

20 state variable. J is minimized when its first variation vanishes, condition that, after some computations, leads to ODEs driving the time evolution of λ_1 , λ_2 and λ_3 . Terminal conditions for their backward integration are also obtained, as well as update equations for initial condition of T_s , T_d , W_c and W_g that depend on Lagrange multipliers. Forward and adjoint model are solved iteratively and corresponding updates are computed. Iterations are interrupted when updates become negligible.

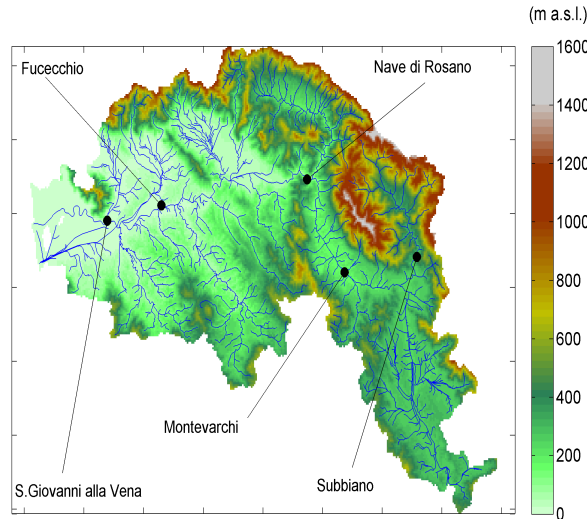


Figure 1. DEM and river network of Arno basin used in the simulations. Black circles indicate measurement stations available for experiments of flow data assimilation.

4 Application: flood forecasts in the Arno river basin

The assimilation scheme is tested in the Arno river basin, central Italy. The basin extends over about 8300 km². Figure 1 shows Digital Elevation Model, river network and flow measurement stations employed in this work. Flood forecasting is a relevant issue in Arno basin, since Arno passes through major Tuscan cities, as Florence and Pisa.

- 5 The performances of the assimilation system are assessed through several hindcast experiments. Simulations are run with the spatial and temporal resolutions that are employed operationally, i.e. 500 m and 15 minutes. The analysis is performed mainly in terms of flow peak prediction accuracy, since it is one of the most important skills for floods early warning.

4.1 Results from assimilation of flow data

Experiments devoted to test assimilation of multiple flow data include both high flow and false alarm (high rainfall but low
10 flows) events that occurred in the period 2009-2014. Flow observations are available from the 5 measurement stations shown in Figure 1, with a temporal resolution of 15 minutes. Sequential assimilations are realized on windows of 6 hours, employing the data from all the stations simultaneously. Analysis of initial discharge in the river network and capillary water in soil, as well as the optimal estimate of the parameter f_0 , are obtained for each assimilation window, and then used to run the corresponding prediction simulation. To evaluate the assimilation system we focus on results at S. Giovanni alla Vena, that is the closest
15 station to the outlet and can be considered as representative of the overall functioning of the scheme. Figure 2 summarizes the results obtained for all the 16 examined events. Peak flows at S. Giovanni alla Vena are plotted against the corresponding total volume of rainfall. Observations and open loop are marked with black stars and white circles respectively. Forecasts from data assimilation are white triangles and gray circles. These latter correspond to the last assimilation window, that includes observed peaks at the upstream locations (Subbiano and Montevarchi, and in some cases also Nave di Rosano), while gray circles are

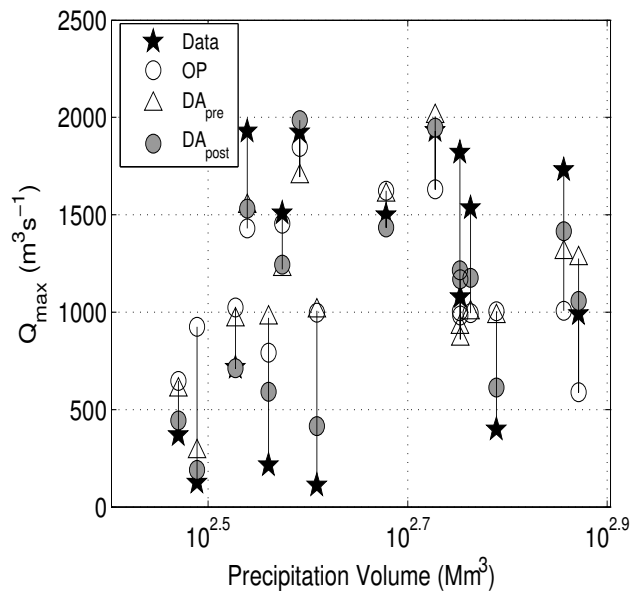


Figure 2. Flow peak versus precipitation volume at S. Giovanni alla Vena for the experiments of streamflow assimilation. Observations and open loop are black stars and white circles respectively. Values from data assimilation are white triangles/gray circles in case peaks at upstream locations are included/not included in the assimilation window.

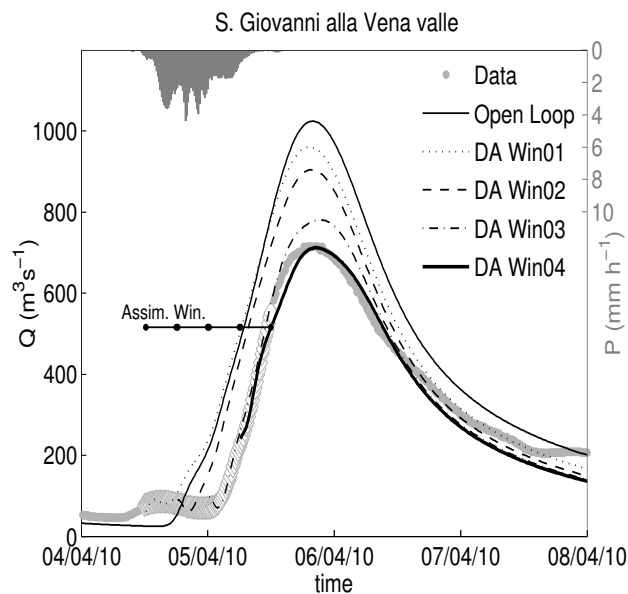


Figure 3. Hydrographs at S. Giovanni alla Vena measurement station for the false alarm event of 06/04/2010.

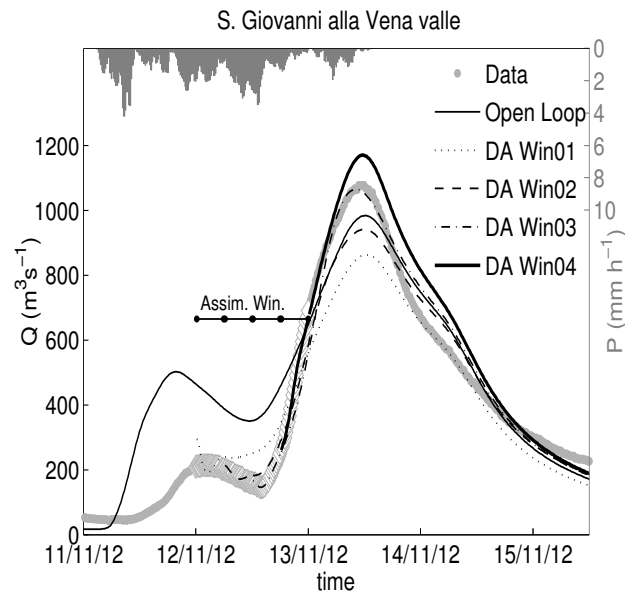


Figure 4. Hydrographs at S. Giovanni alla Vena measurement station for the flood event of 13/11/2012.

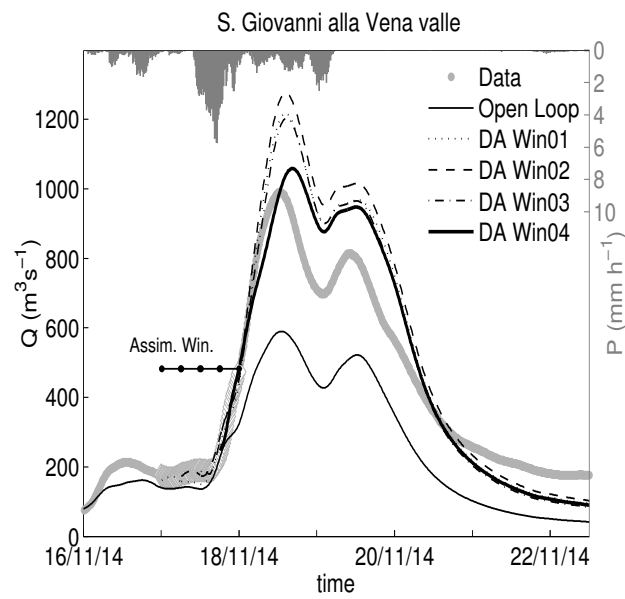


Figure 5. Hydrographs at S. Giovanni alla Vena measurement station for the flood event of 18/11/2014.

from an antecedent window. The general behavior is, as desired, of enhanced flow peak predictions from data assimilation in respect to open loop. Gray circles show a remarkable better adherence to observations than white triangles, suggesting that the gain increases significantly when upstream peaks are included in the assimilation window. However, in some experiments data assimilation has a slightly negative impact on predictions. This occurs when open loop is already very close to the observations.

5 To give further insight, hydrographs at S. Giovanni alla Vena for 3 events are reported in Figures 3, 4 and 5. They are characterized by different levels of performance. In the false alarm event of 06/04/2010 (Figure 3), forecasts progressively improve as the assimilation window advances in time. Predictions corresponding to the last window match observations almost perfectly. Conversely, in the flood event of 13/11/2012 (Figure 4), data assimilation initially reduces the adherence of the forecasted hydrograph to observations, especially in terms of peak flow. The reason is that assimilation attempts to lower streamflow, that

10 in open loop is significantly overestimated during the first window. Improvements are obtained in the 3rd step, whose window corresponds to the actual beginning of the event rising limb. It results into an excellent reproduction of data, that is slightly deteriorated by the subsequent assimilation. The final predicted peak is affected by an error almost equivalent to that of the open loop (about 8.4%), although overestimation instead of underestimation. The behavior for the flood event of 18/11/2014 (Figure 5) is intermediate. Predictions from data assimilation are always better than open loop. However, a progressive im-

15 provement as in the false alarm event of 06/04/2010 is not observed. For instance, the 2nd assimilation slightly worsen the forecasts corresponding to the 1st one. The subsequent assimilation recovers the gap and predictions from the last window are definitely a significant enhancement of the open loop. In summary, the assimilation scheme can considerably improve flood forecasts and reduce false alarms. Nevertheless, in presence of already accurate prediction, the assimilation system can not provide additional enhancement. This fact indicates that the gain obtainable through the assimilation scheme is limited by model

20 structure, namely, model errors can not be overcome by the assimilation, as well as by possible errors in data. Conversely, an exacerbate forcing of modeled outputs toward observations can lead to a negative impact on the forecasts.

4.2 Results from assimilation of LST observations

Assimilation of LST observations is tested on a late summer event (September 16-18, 2006) characterized by locally heavy rainfall (up to 250 mm) but quite low streamflow values. The employed LST observations are from MeteosatSG-SEVIRI

25 (LandSAF product at 15 minutes with a spatial resolution of about 3 km). Assimilation is performed for September, 13th in order to evaluate a proper initial soil moisture before rainfall starts. Figure 6 shows soil moisture at the end of September, 13th for the open loop and its analysis obtained through data assimilation. Figure 7 compares streamflow observations at Nave di Rosano with hydrographs from simulations initialized with a complete dry soil, soil saturation level equal to 50% of field capacity and soil moisture estimated from the assimilation of LST. The peak flow is well reproduced by the latter, while half-

30 field capacity run significantly overestimates discharges and the dry run suffers underestimation. Figure 8 summarizes results for all the available flow measurement stations. Peak flows are plotted against total volume of rainfall in the drainage basin. The significant overestimation of the half-field capacity simulation that has been observed for Nave di Rosano affects also the other stations. Dry and assimilation run are closer to observations (black stars), with the latter better predicting the peak at some locations. Nevertheless, model results do not manage to reproduce the variability in the dependence of flow peak from

upstream precipitation that characterizes observations. This fact reveals that the assimilation system does not add sufficiently flexibility in the model, although it improves the estimate of initial soil moisture. However, tests on several events would be necessary to properly evaluate LST assimilation.

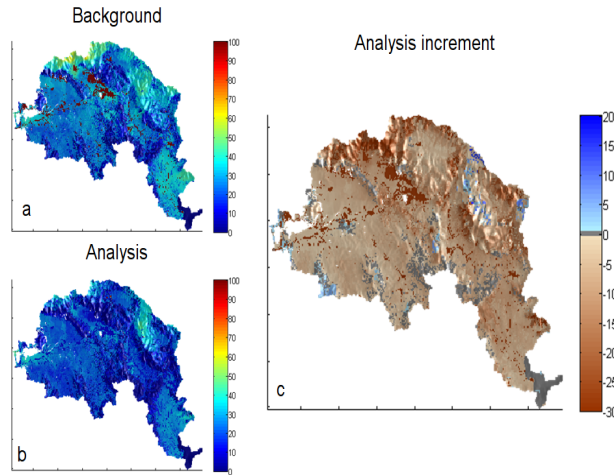


Figure 6. Soil moisture maps of background (a), analysis (b) and increment between them (c) at the end of the assimilation window. Values are expressed as saturation level (%).

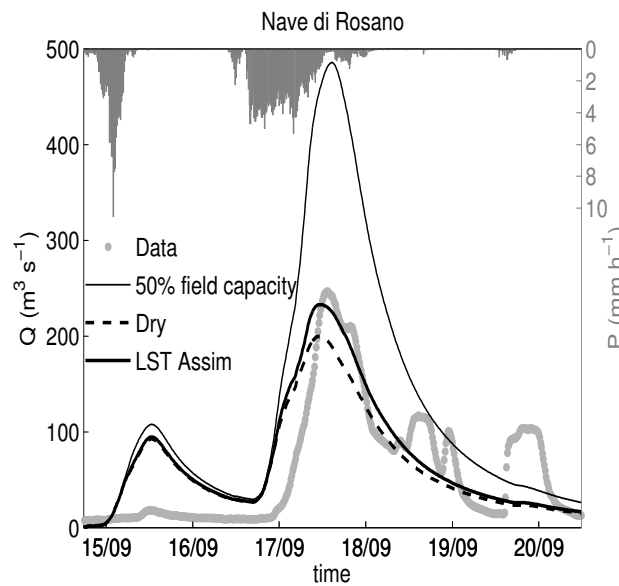


Figure 7. Hydrographs at Nave di Rosano measurement station for the test event of LST assimilation.

5 Conclusions

- 5 This work presents variational assimilation of flow data at multiple locations and of LST in the distributed hydrologic model Mobidic, that is part of the operational forecasting chain of the Arno river in central Italy. Flow data are employed to optimally estimate initial discharge in the river network, initial capillary water in soil and rainfall intermittence parameter f_0 . LST is

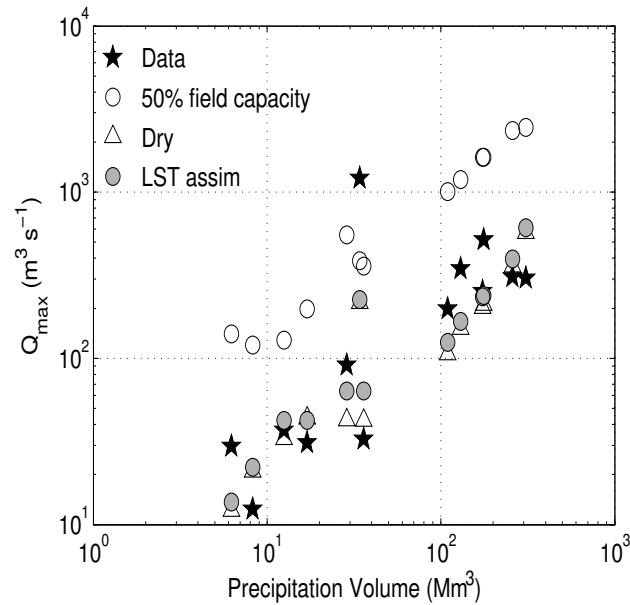


Figure 8. Flow peak versus precipitation volume at the available measurement stations for the test event of LST assimilation.

exploited to infer initial soil moisture (capillary and gravitational). The performances of the developed assimilation system are assessed in several hindcast experiments. In particular, the scheme for assimilation of flow data is tested through 16 experiments. The scheme produces relevant improvements in flood forecasting, especially when the assimilation window includes peak flows at upstream locations. Some negative impacts of the assimilation are observed in case open loop results are already

5 very close to measurements, suggesting that the obtainable gain is limited by the structure of the hydrologic model and flow data errors. The scheme for LST assimilation is verified on one single event with locally heavy rainfall but low streamflows. LST is assimilated 2 days in advance in respect to the event. Comparison of simulation results with observations of flow at various locations shows that the initial condition of soil moisture estimated through LST assimilation improves model performances in respect to dry soil and half-field capacity runs. However, the assimilation does not add sufficiently flexibility in

10 the model to fully reproduce the observed variability in runoff formation process. Future work will analyze in more detail the potential of LST assimilation by performing additional hindcast experiments.

References

- Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., and Hasenauer, S.: Improving runoff prediction through the assimilation of the ASCAT soil moisture product, *Hydrology and Earth System Sciences*, 14, 1881–1893, 2010.
- Campo, L., Caparrini, F., and Castelli, F.: Use of multi-platform, multi-temporal remote-sensing data for calibration of a distributed hydro-
5 logical model: an application in the Arno basin, Italy, *Hydrological processes*, 20, 2693–2712, 2006.
- Campo, L., Castelli, F., Caparrini, F., and Entekhabi, D.: An assimilation algorithm of satellite-derived LST observations for the operational production of soil moisture maps, in: *Geoscience and Remote Sensing Symposium (IGARSS), 2012 IEEE International*, pp. 1314–1317, IEEE, 2012.
- Caparrini, F., Castelli, F., and Entekhabi, D.: Estimation of surface turbulent fluxes through assimilation of radiometric surface temperature
10 sequences, *Journal of Hydrometeorology*, 5, 145–159, 2004.
- Castelli, F.: A simplified stochastic model for infiltration into a heterogeneous soil forced by random precipitation, *Advances in water resources*, 19, 133–144, 1996.
- Castelli, F., Entekhabi, D., and Caporali, E.: Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance, *Water Resources Research*, 35, 3115–3125, 1999.
- 15 Castelli, F., Menduni, G., and Mazzanti, B.: A distributed package for sustainable water management: a case study in the Arno Basin, *International Assoc. of Hydrological Sci, IAHS) Publ.*, 327, 52–61, 2009.
- Castillo, A., Castelli, F., and Entekhabi, D.: Gravitational and capillary soil moisture dynamics for distributed hydrologic models, *Hydrology and Earth System Sciences*, 19, 1857–1869, 2015.
- Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J.: Hydrological data assim-
20 lation with the ensemble Kalman filter: Use of streamflow observations to update states in a distributed hydrological model, *Advances in water resources*, 31, 1309–1324, 2008.
- Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Rudari, R., Pulvirenti, L., Boni, G., Fascetti, F., Pierdicca, N., et al.: Impact of different satellite soil moisture products on the predictions of a continuous distributed hydrological model, *International Journal of Applied Earth Observation and Geoinformation*, 2015.
- 25 Lee, H., Seo, D.-J., Liu, Y., Koren, V., McKee, P., and Corby, R.: Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal scale of adjustment, *Hydrology and Earth System Sciences*, 16, 2233–2251, 2012.
- Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H., Kumar, S., Moradkhani, H., Seo, D., Schwanenberg, D., Smith, P., Van Dijk, A., et al.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrology and Earth System Sciences*, 16 (10), 2012, 2012.
- 30 Seo, D.-J., Cajina, L., Corby, R., and Howieson, T.: Automatic state updating for operational streamflow forecasting via variational data assimilation, *Journal of Hydrology*, 367, 255–275, 2009.
- Sun, N.-Z.: *Inverse problems in groundwater modeling*, vol. 6, Springer Science & Business Media, 2013.
- Yang, J., Castelli, F., and Chen, Y.: Multiobjective sensitivity analysis and optimization of distributed hydrologic model MOBIDIC, *Hydrology and Earth System Sciences*, 18, 4101–4112, 2014a.
- 35 Yang, J., Entekhabi, D., Castelli, F., and Chua, L.: Hydrologic response of a tropical watershed to urbanization, *Journal of Hydrology*, 517, 538–546, 2014b.