Doctoral Thesis

Data Assimilation of SAR derived flood extent maps into flood forecasting models via Particle Filters

submitted in satisfaction of the requirements for the degree of
Doctor of Science in Civil Engineering
of the TU Wien, Faculty of Civil Engineering

Dott.Mag. Ing. **Di Mauro Concetta**

[redacted]

Institute of Hydraulic Engineering and
Water Resources Management Vienna University of Technology

Supervisor: Prof. **Günter Blöschl**
Institut für Wasserbau und Ingenieurhydrologie
Technische Universität Wien
Karlsplatz 13/222, 1040 Wien, Österreich

Supervisor: Dr. **Renaud Hostache**
Institut de recherche pour le développement
Maison Teledetection IRD-MONTPELLIER-FRANCE
500 Rue J.F Breton, 34093 Montpellier, France

Supervisor: Dr. **Patrick Matgen**
Luxembourg Institute for Science and Technology
Environmental Research and Innovation
41 rue du Brill,L-4422 Belvaux, Luxembourg

Wien, 2022

EXAMINERS

Examiner: Prof. **Günter Blöschl**
Institut für Wasserbau und Ingenieurhydrologie
Technische Universität Wien
Karlsplatz 13/222, 1040 Wien, Österreich

Examiner: Associate Prof. **Jeffrey Neal**
School of Geographical Sciences, Cabot Institute for the Environment
University of Bristol
Bristol BS8 1TH, United Kingdom

Examiner: Associate Prof. **Alberto Viglione**
Dipartimento di Ingegneria dell'Ambiente, del Territorio e delle Infrastrutture (DIATI)
Politecnico di Torino
Corso Duca degli Abruzzi, 24, 10129 Torino, Italia

FUNDING INFORMATION

The research reported herein was funded by the National Research fund of Luxembourg through the HyDRO-CSI projects. Funding from the Austrian Science Funds as part of the Vienna Doctoral Programme on Water Resources System (DK W1219-N22) is acknowledged. Funding was also provided by the UK Engineering and Physical Sciences Research Council (EPSRC) DARE project (EP/P002331/1). Peter Jan van Leeuwen thanks the European Research Council (ERC) for funding of the CUNDA ERC 694509 project under the European Unions Horizon 2020 research and innovation programme. Nancy K. Nichols was funded in part by the UK Natural Environmental Research Council (NERC) National Centre for Earth Observation (NCEO). The work of Renaud Hostache was supported by the National Research Fund of Luxembourg through the CASCADE Project under Grant C17/SR/11682050.

The Lisflood-FP model can be freely downloaded at http://www.bristol.ac.uk/geography/research/hydrology/models/lisflood. The river cross-section data, the digital elevation model, and the gauging station water level, streamflow, and rating curve data are freely available upon request from the Environment Agency (enquiries@environmentagency.gov.uk). The ERA-5 data set is freely available at https://confluence.ecmwf.int/display/CKB/ERA5.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Professor Günter Blöschl and Patrick Matgen for the supervision of my PhD thesis. I am grateful for the constructive suggestions and ideas so that I could continuously improve my research. I would also thank my colleague and friend Ramona, for the precious help and the good chat. I wish to thank my family and friends for leading me to take this path. Above all, I want to thank two people, they believed and trusted me in the first place, without them all this could not have happened. I would like to thank Renaud Hostache for the relentless support and the infinite kindness. I want to say thank you to a great man, Michele, for the kilometers travelled, for the enormous patience, for the infinite love, for getting along such a long way together.

To my princess and my warrior Sophie,
Who taught me how to be brave,
the real meaning of resilience,
the art of pure love.

CONTENTS

ABSTRACT

Every year, floods cause global socio-economic losses of more than US$100 billion with an increasing trend due to growing economic activities in flood-prone zones and changes in the climate patterns. Flood warning can assist in reducing these losses based on flood forecasting models that simulate the physical rainfall-runoff-inundation processes in river basins. These models have to be as accurate and reliable as possible. However, the equations, the model parameters, the boundary conditions and the inputs are all affected by inherent uncertainty. Data assimilation (DA) techniques integrate ground based or satellite observations into the model in order to reduce the uncertainty associated with both model results and measurements. In this PhD study, flood extent maps derived from Synthetic Aperture Radar (SAR) observations are assimilated into flood forecasting models by using newly developed filtering techniques based on variants of the Particle Filter.

The thesis is organised into three parts. In the first, as a proof of concept of a DA framework previously introduced in the scientific literature, we evaluate a Sequential Importance Sampling (SIS) variant of a Particle Filter in a controlled environment where synthetically generated rainfall and SAR observations represent the only sources of uncertainty. We show that DA indeed improves the model performances in terms of water level, discharge and flood extent predictions. However, the state-of-the-art filter used in part 1 shows a tendency for degeneracy issues, as the number of particles with a significant weight reduces after the assimilation.

For that reason, in the second part of the thesis, we enhance the existing SIS to mitigate the degeneracy and sample impoverishment issues. We develop a DA framework based on a Tempered Particle Filter (TPF) and evaluate it in a synthetic twin experiment, where rainfall and SAR observations are known to be the only sources of uncertainties. The analysis finds that the model results are improved and the degeneracy issue is mitigated.

In the third part of the PhD study, the TPF is applied in a real-world experiment, where the uncertainties are no longer controlled. Three flood events of the River Severn in the UK are used as test cases. The filter performances differ depending on the gauging stations considered as a reference for the evaluation. For the gauging stations located downstream of the confluence of the three main rivers (i.e., Theme, Avon and Severn) close to the downstream boundary of the hydraulic model, the model results in terms of water level simulations are substantially improved. However, the improvements are less significant in terms of inflows at the upstream boundaries of the model, due to compensation effects between the contributing tributaries.

Overall, the study is considered an important step towards an enhanced Particle Filter by reducing degeneracy and sample impoverishment and improving predictions for the downstream gauging stations. The comparison of the real-world test case with the synthetic one shows that the assumption of the rainfall and SAR observations as the only

sources on uncertainty could be too simple, given the complexity of the hydrological processes involved, and it is suggested to take additional sources of uncertainty into account in future studies.

1

INTRODUCTION

Accurate, precise and timely forecasts are essential for reducing or mitigating the effects of floods. Rainfall-runoff and shallow water models are used to forecast the dynamic outputs (e.g., inundation area, water depth, and/or discharge) for flood prevention, warning, emergency response, and risk reduction. Spatially and temporally distributed observations of key hydrologic variables are often used to improve the accuracy and reliability of flood forecasts. In this study we use Remote Sensing observations, namely Synthetic Aperture Radar (SAR) data, for this purpose. The value of SAR data is linked to the capacity of radar sensors to penetrate clouds, retrieve information regardless of light conditions, and provide information over large areas. However, the SAR observations are inherently uncertain (Moradkhani et al., 2005). SAR observations are not exempt from errors in flood area detection, especially in densely vegetated areas and built-up environments and when the acquisitions take place during challenging weather conditions (strong wind or heavy rainfall) (Chen et al., 2018; Grimaldi et al., 2020; Zhao et al., 2021). Moreover, depending on the flashiness of the flood events, the use of SAR observations in near-real time can be still limited by the satellite revisit period, although this was reduced to 3 days over Europe with Sentinel-1. The latency (time between the acquisition and the availability of the data for the user) has been substantially reduced recently and reaches few hours, for example, for Sentinel-1. In addition, models can also present a source of significant uncertainty due to simplified model structures, insufficiently well determined or calibrated model parameters, erroneous input data, etc. Due to the uncertainties of predicted and observational datasets, much research has focused on statistically combining observation with model simulations in a framework that merges the knowledge of both datasets to yield a more accurate system state estimation, along with the uncertainty of the prediction. This process is called Data Assimilation (DA).

Data assimilation methods are often split into two families: variational methods and statistical methods. Variational methods (4D Var) are based on the definition of a cost function $J(x)$ which defines the discrepancy between the computational results and observations. 4D-Var can be considered as an unconstrained optimization problem that seeks an optimal control vector x to minimize the cost function $J(x)$. Statistical methods

are based on the Bayes' theorem and the uncertainty of the model, and the observations are represented by probability distributions functions (pdfs). Bayes' Theorem states that the conditional probability of the model based on the observations (posterior probability), is equal to the likelihood of the observations given the model multiplied by the probability of the model (prior probability) divided the marginal probability of the observations (i.e. normalising constant).

$$p(x \mid y) = \frac{p(y \mid x)}{p(y)} p(x) = \frac{p(y \mid x)}{\int p(y \mid x) p(x) \delta x} p(x) \tag{1.1}$$

Estimating these pdfs in large dimensional systems is impossible. Therefore, some methods use an ensemble of state vectors to represent a statistical sample of the forecast or analysis uncertainty. Each ensemble member represents one possible realisation of the true state of the system, given uncertainties in initial conditions and/or model parameters. The Ensemble Kalman Filter (EnKF), which is a variant of the Kalman Filter, is widely used in hydrological applications (Revilla-Romero et al., 2016; Cooper et al., 2018a; Annis et al., 2021; Wongchuig-Correa et al., 2020). It is based on an ensemble of N state vectors $x_i, (i = 1, 2, ..., N)$ that represents possible states of the system. 2 steps are performed, analysis and forecast. In the forecast step, the state vector is propagated from time t_k to time t_{k+1} using the model (M) such that: $x_i(t_{k+1}) = M(x_i(t_k))$. The analysis step is performed when an observation is available and the state vector is updated based on the observational information. The observation y_{obs} can be described by: $y_{obs} = H(x^t) + \epsilon$ where H maps the state vector into the observation space. The state is updated according to:

$$x^a = x^f + K(y^{obs} - y^f) \tag{1.2}$$

where x^a is the analysis state vector, x^f is the forecast state vector, y^f is the model equivalent of the observation, K is the Kalman gain and it contains the weights assigned to the $(y^{obs} - y^f)$ term to update the system. The EnKF solves the mean and covariance of the posterior. Even though for the EnKF the evolution in time of the system is not linear, when confronted with the prior and the observations, pdfs are assumed to be Gaussian and the analysis step is linear. Hydrological problems are typically non linear and the KF-like methods have problems in dealing with non-linear problems (Leeuwen et al., 2019).

Particle Filters (PFs) have gained recent attention from the research community because PFs hold promises for a fully non-linear DA and relax the assumption of Gaussian errors (Leeuwen et al., 2019). In PFs, the prior pdf is represented by an ensemble of size N of random model state x_n, called particles.

$$p(x) = \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n) \tag{1.3}$$

Particles are obtained by running the model forward in time until an observation is available. Then, inserting the prior into the Bayes' theorem we find that:

$$p(x \mid y) \approx \sum_{n=1}^{N} W_n \delta(x - x_n) \tag{1.4}$$

where δ is the Dirac delta function in which particle weights are given by Equation 1.5:

$$W_n = \frac{p(y \mid x_n)}{N \cdot p(y)} = \frac{p(y \mid x_n)}{N \cdot \int p(y \mid x) p(x) dx} \approx \frac{p(y \mid x_n)}{\sum_n p(y \mid x_n)} \tag{1.5}$$

Since all the terms of Equation 1.5 are known, it is possible to assign the weights to the particles. In the PFs, a potential issue in high-dimensional problems is that the weights carried by the different particles may vary significantly across the ensemble so that some particles gain higher weights than all the others (weight degeneracy). If the process of forecast and analysis is repeated a number of times, only one particle with the larger weight will remain and the statistical information of the ensemble is lost. In this case, the ensemble is no longer representative of the uncertainty of the system (Moradkhani et al., 2012). To avoid this issue, an enormous number of particles would be required resulting in a very high computational cost. Therefore, some techniques have been implemented to mitigate this issue without using a large amount of particles. The most recent techniques combine PFs with other techniques such as Monte-Carlo Markov Chains (MCMC) (Andrieu et al., 2010; Moradkhani et al., 2012), or combining PFs with Metaheuristic Algorithms (MA) from Computer Sciences (Kwok et al., 2005), Particle Swarm Optimization (Wang et al., 2006; Li et al., 2005), Immune Genetic Algorithms (IGA) (Han et al., 2011). Some recent developments, such as the Evolutionary PF-MCMC (EPFM) (Abbaszadeh et al., 2018) or the Hybrid Ensemble and Variational Data Assimilation framework for Environmental systems (HEAVEN) (Abbaszadeh et al., 2019), generate more accurate and reliable results in spite of the (still unrealistic) assumption of a Gaussian distribution of the pdfs.

In operational flood forecasting, a hydraulic model is used to compute water level and velocity in the river network, and when the storage capacity of the river is exceeded, in the floodplain. Some studies assimilate satellite data derived water levels to improve flood model forecasts [e.g. Andreadis et al., 2007, García-Pintado et al., 2015, Matgen et al., 2010, Revilla-Romero et al., 2016, Giustarini et al., 2011a, Hostache et al., 2010], but only few of them directly assimilate flood extents [e.g. Lai et al., 2014, Revilla-Romero et al., 2016, Cooper et al., 2018a, Cooper et al., 2018b, Hostache et al., 2018]. In this thesis, flood extent maps derived from SAR data are assimilated into a flood forecasting model via variants of the PFs following the study by Hostache et al., 2018. Hostache et al., 2018 use Sequential Importance Sampling (a variant of the PF) to assimilate probabilistic flood maps (a map where each pixel is assigned a probability to be flooded given the SAR image backscatter value) into a flood forecasting model. The flood forecasting model used in this study has 2 components. A hydrological (rainfal-runoff) model, fed with meteorological predictions, computes the run-off over the river reaches contributing area. The run-off is used as a boundary condition of the hydraulic (shallow-water) model. The main assumption in this case is that rainfall and SAR observations are the main sources of uncertainty which seems to be reasonable in operational cases because they are likely to be the major error sources (Koussis et al., 2003; Pappenberger et al., 2005). Improvements of the model simulation after the assimilation were not systematic, probably due to the underlying assumptions of the rainfall as unique model-related uncertainty (Hostache et al., 2018). As already mentioned, model uncertainty can derive from several sources and to better estimate the predictions and to reduce the predictive

uncertainty other sources of uncertainties should be considered to improve the quality of the forecasts (Moradkhani et al., 2005).

The main objective of the present study is to assess the strengths and the limitations of the DA framework proposed by Hostache et al., 2018 in a synthetic environment and to develop, apply, and evaluate a more robust DA framework for the assimilation of satellite Earth Observation data into flood forecasting systems. The main aim of this thesis is thus to advance the knowledge on PF techniques in hydrological sciences and to achieve higher accuracy and reliability of model forecasts.
In particular, this thesis addresses the following research questions:

- Is the Sequential Importance Sampling efficient in a synthetic experiment where the only sources of uncertainty are the rainfall and the SAR observations?
- Does the enhanced Particle Filter (Tempered Particle Filter) improve the model results while reducing the degeneracy and the sample impoverishment issues?
- Does the enhanced Tempered Particle Filter improve the prediction accuracy and increase the persistence of the assimilation benefits when applied to a real case study?

Each chapter of this thesis corresponds to a published or submitted scientific article (Chapters 2-4). In Chapter 2, we evaluate a previously introduced state of the art DA framework by the means of a synthetic experiment where sources of uncertainty are tightly controlled. In Chapter 3, we introduce a new algorithm based on the Tempering Particle Filter (TPF) technique in order to mitigate more effectively the still present issue of degeneracy and sample impoverishment. In Chapter 4, the TPF is applied in a real-world experiment (River Severn in the United Kingdom). Finally, Chapter 5 concludes the thesis with a summary of its key findings and a discussion of potential future extensions of the research work.

2

ASSIMILATION OF PROBABILISTIC FLOOD MAPS FROM SAR DATA INTO A COUPLED HYDROLOGIC-HYDRAULIC FORECASTING MODEL: A PROOF OF CONCEPT.

The present chapter corresponds to the following scientific publication in its original form: Concetta Di Mauro, Renaud Hostache, Patrick Matgen, Ramona Pelich, Marco Chini, Peter Jan van Leeuwen, Nancy K. Nichols, and Günter Blösch. "Assimilation of probabilistic flood maps from SAR data into a coupled hydrologic–hydraulic forecasting model: a proof of concept. Hydrol. Earth Syst. Sci., 25, 4081–4097, 2021. https://doi.org/10.5194/hess-25-4081-2021"

2.1. ABSTRACT

Coupled hydrologic and hydraulic models represent powerful tools for simulating streamflow and water levels along the riverbed and in the floodplain. However, input data, model parameters, initial conditions and model structure represent sources of uncertainty that affect the reliability and accuracy of flood forecasts. Assimilation of satellite-based Synthetic Aperture Radar (SAR) observations into a flood forecasting model is generally used to reduce such uncertainties. In this context, we have evaluated how sequential assimilation of flood extent derived from SAR data can help improve flood forecasts. In particular, we carried out twin experiments based on a synthetically gen-

erated data-set with controlled uncertainty. To this end, two assimilation methods are explored and compared: the Sequential Importance Sampling (standard method) and its enhanced method where a tempering coefficient is used to inflate the posterior probability (adapted method) and to reduce degeneracy. The experimental results show that the assimilation of SAR probabilistic flood maps significantly improves the predictions of streamflow and water elevation, thereby confirming the effectiveness of the data assimilation framework. In addition, the assimilation method significantly reduces the spatially averaged root mean square error of water levels with respect to the case without assimilation. The critical success index of predicted flood extent maps is significantly increased by the assimilation. While the standard method proves to be more accurate in estimating the water levels and streamflow at the assimilation time step, the adapted method enables a more persistent improvement of the forecasts. However, although the use of a tempering coefficient reduces the degeneracy problem, the accuracy of model simulation is lower than the one of the standard method at the assimilation time step.

2.2. INTRODUCTION

Floods represent one of the major natural disasters with a global annual average loss of US $104 billion (UNISDR, 2015). Extent of flood damages have risen during the last years due to climate-driven changes and an increase in the asset values of floodplains (Blöschl et al., 2019a). This emphasizes the need for reliable and cost-effective flood forecasting models to predict flood inundations in near real-time. Hydrologic and hydraulic models represent useful tools for simulating flood extent, discharge and water levels in the river bed and on the floodplain. However, both the models and the observations used as inputs for running, calibrating and evaluating the models are affected by uncertainty.

Data assimilation (DA) aims at improving model predictions by updating model states and/or parameters based on observations (Moradkhani et al., 2005). It optimally combines observations with the system state derived from a numerical model accounting for both model and observation errors.

Ideally, *in situ* data are systematically assimilated into flood forecasting models, but these observations are not always available (e.g. in *un-gauged* catchments) and only provide space-limited information (Grimaldi et al., 2016). Therefore, satellite Earth Observation (EO) data, and in particular Synthetic Aperture Radar (SAR) images, represent a valuable complementary dataset to *in situ* observations due to their capacities to provide frequent updates of flooded areas at a large scale. In addition, as the corresponding EO data archives are growing fast, historical observational data spanning an extended period of time can be assimilated into large scale hydrodynamic models.

SAR sensors are able to acquire images of flooded areas and permanent water bodies during day and night almost regardless of weather conditions. The backscattered signal depends on the dielectrical properties of the imaged objects. Smooth surfaces, such as open water bodies, interact with the transmitted pulse so that a very limited part of the signal is backscattered to the satellite resulting in dark areas in the acquired image.

Different information about water extent can be extracted from a SAR image and used to improve the forecasts using DA techniques. Directly assimilating flood extent maps is not straightforward because these do not correspond to a state variable of the model. Therefore, some studies suggested to transform the SAR backscatter information into

state variable prior to the assimilation. For instance, several studies have used EO-derived water levels to improve flood forecasts [e.g. Andreadis et al., 2007, García-Pintado et al., 2015, Matgen et al., 2010, Revilla-Romero et al., 2016, Giustarini et al., 2011a, Hostache et al., 2010]. The water levels are estimated by merging pre-selected flood extent limits extracted from the SAR satellite imagery with a digital elevation model (DEM). This step requires precise flood contour maps and high resolution DEMs which are not always available (Hostache et al., 2018).

In the existing literature only a few studies have used DA for directly assimilating flood extent maps into flood forecasting models [e.g. Lai et al., 2014, Revilla-Romero et al., 2016, Cooper et al., 2018a, Cooper et al., 2018b, Hostache et al., 2018]. Among the advantages of a direct use of the SAR backscatter values is that it reduces the data processing time that is a key-element in near-operational applications.

Cooper et al., 2018b have used an Ensemble Kalman Filter to update a 2D hydrodynamic model. In this case, the backscatter values are directly assimilated into the model without being transformed into state variables of the flood forecasting system. The dry and wet pixels of the simulated binary flood map are converted into equivalent SAR backscatter values corresponding to the spatial mean of the SAR backscatter observations.

Cooper et al., 2018b showed that the SAR backscatter-based assimilation method performs well compared to the assimilation method where the SAR backscatter is transformed into water levels.

Hostache et al., 2018 used a variant of the Particle Filter (PF) with Sequential Importance Sampling (SIS), to assimilate probabilistic flood maps (PFMs) derived from SAR data into a coupled hydrologic-hydraulic model with the assumption that the rainfall is the main source of uncertainty together with SAR observations.

Their study showed that the assimilation of PFMs is beneficial: the number of correctly predicted flooded pixels increases as compared to the case without any assimilation, hereafter called *Open Loop* (OL). Forecast errors are reduced by a factor of 2 at the assimilation time and improvements persist for subsequent time steps up to 2 days. However, the improvements are not systematic: for some cases the updated hydraulic output deviates from the observations. One of the reasons could be the assumption that rainfall represents the dominating source of uncertainty together with satellite observation errors, thereby excluding other possible sources of uncertainty in the model system such as input data, model parameters, initial conditions and model structure. Even though the assumption seems to be rather realistic and suitable in operational cases, given that rainfall uncertainty has been generally identified as one of the major causes of poorly performing models (Koussis et al., 2003; Pappenberger et al., 2005), coupled models may have additional sources of uncertainty affecting the results.

The present study is a follow up of the real world experiment by Hostache et al., 2018 and carries out a similar experiment in a controlled environment that considers the estimated rainfall together with SAR observations as the only sources of uncertainty.

Hostache et al., 2018 also highlighted that degeneracy may be a major issue of PFs: after the assimilation, the number of particles with high weights reduces to a few or only one particle so that the ensemble loses statistical significance. To overcome this issue, Hostache et al., 2018 used a site-dependent tempering coefficient which inflates the pos-

terior probability. In our study, we propose to adopt an enhanced tempering coefficient as a function of the desired effective ensemble size (EES) after the assimilation.

Moreover, in Hostache et al., 2018 speckle errors in the SAR observations, are taken into account through the Bayesian approach introduced by Giustarini et al., 2016. However, no conclusions are drawn concerning the effect of misclassified pixels. In fact, for some particular cases such as densely vegetated areas, the detection of flood water from SAR imagery is known to be prone to errors. Detecting and removing such errors represents one of the main scientific challenges of using SAR data for a systematic, fully automated, large-scale flood monitoring (and prediction).
The main objective of the present study is to assess the strengths and the limitations of the DA framework previously proposed by Hostache et al., 2018. To do that we evaluate the DA framework in a fully controlled environment via synthetic twin experiments as this shall allow us drawing unambiguous and comprehensive conclusions. In addition, we conduct a sensitivity analysis of the DA framework with respect to the critical tempering coefficient that was recently introduced for tackling degeneracy more efficiently. We also aim to evaluate the effect of misclassified SAR pixels on DA. Therefore, errors are artificially added within the SAR image with the aim of getting a better understanding on how robust the proposed method is with respect to this type of errors. Results are evaluated not only locally but also over the entire flood domain and for subsequent time steps to the assimilation. To carry out the experimental study we apply the DA framework to a forecasting system consisting of a loosely-coupled hydrologic model (SUPERFLEX) and hydraulic model (LISFLOOD-FP). The meteorological data that are used to run the experiments are derived from the ERA-5 archive with a spatial resolution of 25 km and a temporal resolution of 1 hour. The SAR data are synthetically generated with a pixel spacing of 75 m.

2.3. METHODS

The proposed methodology is based on numerical experiments conducted with synthetically generated data as illustrated in the flow chart given in Figure 2.1. In this framework, the following data inputs and models are employed:

1. *True* rainfall time series are used to generate the *true* hydrologic-hydraulic model simulation.

2. Synthetic SAR observations are generated from the *true* model run (i.e. from the simulated flood extent map).

3. The *true* rainfall time series are randomly perturbed and used as inputs of the hydrologic model. The simulated discharge data are then used as boundary conditions to realize an ensemble of hydraulic model runs.

4. The synthetic SAR observations are assimilated into the coupled hydrologic-hydraulic model via different variants of the Particle Filter (PF) .

The three conducted experiments are summarized as follows:

(a) An application of the standard PF where degeneracy occurs;

(b) An application of the adapted PF where a tempering coefficient is used to reduce degeneracy. We also investigated the sensitivity of the DA results to different values for the tempering coefficient, corresponding to EES of 5, 10, 20 and 50%;

(c) An application of both proposed methods with artificially introduced known errors into the SAR image classification in order to evaluate the impact of these errors on the DA performance metrics.

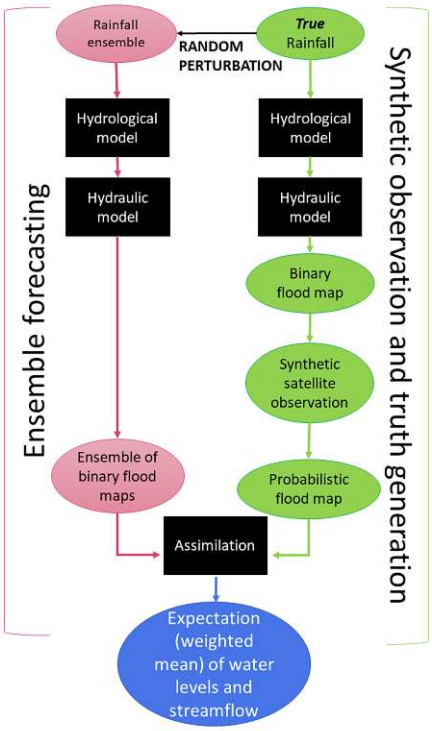


Figure 2.1: Flow chart of the synthetic experiment. The *true* rainfall is perturbed. The same flood forecasting model structure composed of a hydrologic model and a hydraulic model is used to obtain the probabilistic flood map and the ensemble of binary flood maps. The probabilistic flood map is assimilated into the ensemble of binary flood maps via the Particle Filter to obtain the weights with which the expectation of water levels, streamflow and flood extents are computed.

2.3.1. COUPLED HYDROLOGIC-HYDRAULIC MODEL: SYNTHETIC TRUTH AND ENSEMBLE

The coupled modelling system consists of a hydrologic model coupled with a hydraulic model. The hydrologic model is used to compute the run-off at the upstream boundaries

of the hydraulic model. The hydrologic model used in this study is SUPERFLEX which is a framework for conceptual hydrologic modelling introduced by Fenicia et al., 2011. The model structure is a combination of generic components: reservoirs, connection elements and lag functions. In this study, a lumped conceptual model and its structure as a combination of three reservoirs are used: an unsaturated soil reservoir with storage S_{UR}, a fast reacting reservoir with storage S_{FR}, and a slow reacting reservoir with storage S_{SR}. A lag function has been added at the outlet of the slow and fast reacting reservoirs.

The hydraulic model is based on LISFLOOD-FP (Bates et al., 2000) and simulates flood extent, water level and streamflow along the river and on the floodplain. The models solves the local inertial form of the shallow water equations using a finite difference method on a 2D grid. A sub-grid channel routing is used for simulating the channel flow (Neal et al., 2012).

The *true* meteorological data (i.e., temperature and rainfall) are used as input of the hydrologic-hydraulic model to simulate streamflow and water level time series and to provide binary flood maps, where each pixel is classified as flooded (with value 1) or not flooded (with value 0), at each assimilation time step. These computational results represent the synthetic *truth* that will be used to evaluate the performance of the proposed assimilation framework. The *true* binary flood maps are also used to generate the synthetic SAR observations as described in the next section.

2.3.2. SYNTHETIC OBSERVATIONS

In the proposed synthetic experiment, we generate synthetic SAR images at each assimilation time step. The SAR images are generated with the same spatial resolution of the LISFLOOD-FP maps. Similarly to the Van Wesemael, 2019 study, we make use of a real SAR image, acquired during a flood event in the past, and of the LISFLOOD-FP model to generate true binary flood maps. The histogram of the SAR image backscatter values can be approximated with two Gaussian curves relative to the flooded and not flooded pixel classes. Generally, the class of flooded pixels is often represented just by a fraction of the SAR image scenes. Therefore, to identify and characterize areas where the flooded and not flooded classes are more balanced, the hierarchical split based approach (HSBA, Chini et al., 2017) is applied to the selected SAR image. The parameters of the Gaussian pdfs are determined by fitting the histogram values of the HSBA selected areas.

Then random backscatter values, derived from the calibrated Gaussian pdfs, are associated to the pixels of the *true* binary flood map indicating the presence of water and no-water areas. Once the synthetic SAR images are generated, the Giustarini et al., 2016 procedure is applied and synthetic PFMs are derived. The probability to be flooded given the recorded backscatter values for each pixel of a SAR image $p(F|\sigma^0)$ is obtained via the Bayes' theorem:

$$p(F|\sigma^0) = \frac{p(\sigma^0|F)\,p(F)}{p(\sigma^0)} = \frac{p(\sigma^0|F)\,p(F)}{p(\sigma^0|F)\,p(F) + p(\sigma^0|\overline{F})\,p(\overline{F})} \tag{2.1}$$

In the equation 2.1, $p(\sigma^0|F)$ and $p(\sigma^0|\overline{F})$ represent, respectively, the probability of the backscatter σ^0 values of the flooded and non-flooded pixels, $p(F)$ is the prior probability of a pixel being flooded and $p(\overline{F})$ is the prior probability of a pixel being non-flooded before any backscatter information is taken into account. The conditional prob-

abilities are derived from the histogram of the backscatter values estimated from the synthetically generated SAR image. The prior probabilities can be estimated from the flood extent model output or through visual interpretation of an aerial photography in real cases. However, in general such information is not always available and the prior probabilities are unknown. Consequently, Giustarini et al., 2016 set the prior probability of equation 2.1 to 0.5 so that both flooded and non flooded pixels are equally likely. While this study is based on a synthetic experiment, *true* binary flood extent maps are available. Therefore, the assimilation is realized using both the estimated prior probability (as the ratio between the flooded area and the total area) and the prior probability equal to 0.5. Given the similarity of the results for both cases, in the following sections we only discuss the experiment using the estimated prior probability.

The method proposed by Giustarini et al., 2016 aims at characterizing the speckle-induced uncertainty. However, it does not consider any other phenomena leading to a wrong classification in SAR-based flood maps. Particular atmospheric conditions (e.g. wind, snow, precipitation), water-look-a-like areas (e.g. asphalt, sand, shadow) or obstructing objects (e.g. dense canopy, buildings), as mentioned in Giustarini et al., 2015, can lead to a wrong classification in the flood maps. Therefore, the areas where such errors could occur should be masked out from the SAR-based flood maps in order to provide a reliable flood detection.
In the first part of this study, SAR observations are considered without errors. In the second part, these kinds of errors are integrated in the synthetic SAR observations to evaluate their effect on the DA. Specifically, the pixels along the flood edge of each particle are selected. From this set, a given number of those pixels effectively flooded in the *true* binary flood map are artificially corrupted so that they belong to dry pixels. The number of corrupted pixels depends on the magnitude of the error that we want to introduce in the SAR observations.

2.3.3. ENSEMBLE GENERATION

In a PF the prior and posterior pdfs are approximated by a set of particles. Here, we hypothesize that the rainfall is the only source of uncertainty affecting the model-based flood extent simulations. Due to this reason, an ensemble of rainfall time series is used as input of the coupled hydrologic-hydraulic model. Each rainfall time series is obtained by perturbing, with a multiplicative random noise from a log-normal error distribution, the *true* rainfall time series following the approach proposed in Hostache et al., 2018. 128 rainfall time series are obtained and forwarded in time via the hydrologic model.

It is important to note that the same hydrologic-hydraulic model in terms of structure, initial conditions and parameters is used for all model runs. The reliability of the rainfall ensemble is verified with the statistical metrics proposed by De Lannoy et al., 2006. According to the verification measurement VM_1 in equation 2.2:

$$VM_1 = \frac{< ensk >}{< ensp >} \approx 1 \tag{2.2}$$

with <> meant as the average over available observations (in time). The ensemble spread in the equation 2.5 (where $x_{k,n}$ represents the value of the variable x at time k for each

pixel n)

$$ensp = \frac{1}{N}\sum_{n=1}^{N}(x_{k,n} - \overline{x}_k)^2 \tag{2.3}$$

has to be close to the ensemble skill (equation 2.4)

$$ensk = (\overline{x}_k - y_k)^2 \tag{2.4}$$

which is the difference between the mean $\overline{x}_k$ over the N particles of the ensemble and the observation y_k at time k. VM_1 given by the following Equation:

$$VM_1 = \frac{< ensp >}{< ensk >} \tag{2.5}$$

should be equal to 1, larger values would indicate a too small spread of the ensemble. Another method to define the similarity of the observation to any of the ensemble members is based on the calculation of expected value of the ratio of the time-averaged root mean square (rms) error of the ensemble mean to the time-averaged mean rms error of the ensemble members:

$$E(Ra) = \frac{\sqrt{< ensk >}}{\sqrt{< mse >}} \approx \sqrt{\frac{(N+1)}{2N}} \tag{2.6}$$

with *mse* estimated as:

$$mse = \frac{1}{N}\sum_{n=1}^{N}(x_{k,n} - y_k)^2 \tag{2.7}$$

If the observation is statistically indistinguishable from a particle of the ensemble, the expected value E[Ra] is equal to $\sqrt{\frac{(N+1)}{2N}}$ as shown by Murphy, 1988. VM_2, also known as the normalized error, which is a measure of the ensemble dispersion and it is given by the ratio of Ra to the expected value E[Ra], should be equal to 1 (Anderson, 2001; Moradkhani et al., 2005).

2.3.4. DATA ASSIMILATION FRAMEWORK

The DA framework consists of two main steps: *prediction*, i.e model simulations, and *analysis*, i.e update of particle probabilities when an observation is available. The prior probability of the model state x at a given time k is represented by a set of N independent random particles x_n sampled from the prior probability distribution $p(x)$ (Equation 1.3). where δ is the Dirac delta function. In this study, the prior probability distribution is assumed to be uniform. The observations of flooded/not flooded pixels y are related to the true state x^t according to the Equation $y = H[x^t] + \epsilon$ where H is the observation operator that maps the state vector into the observation space and ϵ represents the observation errors. According to the Bayes' theorem, the observations are assimilated by multiplying the prior pdf $p(x)$ and the likelihood $p(y \mid x)$, which is the probability density of the observation given the model state, and dividing by the total probability $p(y)$, resulting in Equation 1.1 where $p(x \mid y)$ is the posterior probability, i.e. the probability density function of the model state given the observations. By inserting the equation 1.3 into the

equation 1.1 we obtain the Equation 1.4. where W_n represents the relative importance in the probability density function (i.e. global weight) given by Equation 1.5.

In this study, the likelihood (global weight, W_n) is represented by the product of the pixel-based likelihood (local weight, w_i), assuming the L pixel observation errors to be independent from each other. At time k of the observation, local weights $w_{i,n}$ are defined for each particle *n* and for each pixel i according to Hostache et al., 2018:

$$w_{i,n} = p_i(F \mid \sigma_0)M_{i,n} + [1 - p_i(F \mid \sigma_0)](1 - M_{i,n}) \tag{2.8}$$

$w_{i,n}$ is equal to the probability of a pixel being flooded as derived from the synthetically generated SAR image. $M_{i,n}$ is equal to "1" if the model predicts the pixel as flooded, whereas $M_{i,n}$ is equal to "0" if the model predicts the pixels as non-flooded. We convert the model-based water depth maps into binary flood extent maps by considering that a pixel is flooded if its water level is above 10 cm. $p_i(F \mid \sigma_0)$ equals the probability of a pixel being flooded according to the observations, viceversa $1 - p_i(F \mid \sigma_0)$ equals the probability of not being flooded. By applying the equation 2.8 we assign higher probabilities to those pixels where model predictions and observations agree. Next, W_n is estimated for each particle by the normalization of the product of the local weights ensuring that the sum of the global weights is equal to 1 (equation 2.9, *standard method*).

$$W_n = \frac{\prod_{i=1}^{L} w_{i,n}}{\sum_{n=1}^{N}\prod_{i=1}^{L} w_{i,n}} \tag{2.9}$$

The expectation of the OL is equivalent to the mean of the ensemble because the relative importance of each particle is the same. The global weights are used to compute the expectation of the streamflows (Q) and water levels (h) at time (k) and per pixel (i) after the assimilation (see equations 2.10, 2.11).

$$\overline{h}_i = \sum_{n=1}^{N} W_n \cdot h_{i,n} \tag{2.10}$$

$$\overline{Q}_i = \sum_{n=1}^{N} W_n \cdot Q_{i,n} \tag{2.11}$$

The particles keep these global weights until the next assimilation time. The particles are then set to the same equal weight before a new *analysis* step is performed.
Unless the number of particles increases exponentially with the dimension of the system-state, the particle-filter is likely to degenerate because high probability is assigned to a single particle while all other members will result in small weights (Leeuwen et al., 2019). PFs are often subject to degeneracy issues when, due to computational reasons, the number of particles is not sufficiently high (Zhu et al., 2016). After the application of the standard PF, the variance of the weights tend to increase and only a few particles of the ensemble have a non-negligible weight. To mitigate this problem, in Hostache et al., 2018, the global weight defined in the equation 2.9 has been adapted using a *tempering* coefficient (α, as described by the following equation 2.12).

$$W_n(\alpha) = \frac{\prod_{i=1}^{L} w_{i,n}{}^{\gamma}}{\sum_{n=1}^{N}\prod_{i=1}^{L} w_{i,n}{}^{\gamma}} \tag{2.12}$$

Since α and weights values are lower than one, adding the power of α in the weights formula allows for shifting all weight values closer to one. This therefore decreases the variance of the weights and inflates the posterior probability. After the assimilation, the number of particles with significant weight depends on the α value. The smaller α, the higher the variance of the posterior pdf. Consequently, as argued in Hostache et al., 2018, when the α coefficient is small enough, this adaptation of the PF helps reduce the degeneracy of the ensemble. While in the previous study by Hostache et al., 2018, the α value was defined so that the worst model solution would have had a non-zero global weight, in this study we propose to define α based on the desired effective ensemble size (EES). The coefficient α in Hostache et al., 2018 is site-dependent as it relies on the number of flooded pixels, whereas in this study α is a function of the EES which is a measure of degeneracy based on the global weights (Arulampalam et al., 2002):

$$EES(\alpha) = \frac{1}{\sum_{n=1}^{N}(W_n(\gamma))^2} \cdot \frac{1}{N} \cdot 100 \tag{2.13}$$

The EES is lower than N and its value indicates the level of degeneracy. α is equal to one when the standard method is used. Decreasing the α coefficient leads to an increase of the EES.

2.3.5. PERFORMANCE METRICS

To carry out the evaluation of the PFM statistics we have used the reliability plots. The results of the different assimilation scenarios are evaluated on a spatio-temporal scale with the following performance metrics:

- Contingency maps and the confusion matrix;
- Critical success index (CSI);
- Root mean square error (RMSE);

To evaluate the simulations of water level and discharge after the DA application the time series of these variables will be also plotted.

RELIABILITY PLOTS

Reliability diagrams are employed to statistically evaluate the synthetically generated PFMs. In such diagrams, the probability range [0;1] is subdivided into intervals of average probability P_i and width ΔP_i. We identify the pixels Ω_i having a probability value of $P_i \pm \Delta P_i$ in the PFM. The fraction of Ω_i pixels effectively flooded in the binary *truth* map are identified with F_i. The reliability diagram plots P_i on the x-axis and F_i on the y-axis. A reliability diagram indicating an alignment of data points close to the 1:1 line means that the PFM is statistically reliable.

CONTINGENCY MAPS AND CONFUSION MATRIX

First, we use contingency maps to graphically compare the expected flood map with the synthetic *truth* map at each assimilation time step. Pixel classification errors can be of two types: overprediction (type error I) when the pixels in the *truth* map are not flooded but are predicted as flooded, and underprediction (type error II) in the opposite case.

Then, the confusion matrix numerically summarizes the results of the contingency map. It is a 2 rows by 2 columns matrix that reports the number of false positives (type I error), false negatives (type II error), true positives and true negatives.

CSI

The CSI evaluates the goodness of fit between the *truth* map and the predicted flood extent map (Bates et al., 2000):

$$CSI = \frac{A}{A+B+C} \tag{2.14}$$

It represents the ratio between the number of pixels correctly predicted as flooded (A) over the sum of all the flooded pixels including the false positives (B, overdetection) and false negatives (C, underdetection). CSI ranges between 0 and 1 (best score). We also used it to evaluate the results at the assimilation time step and the effect of the assimilation at subsequent time steps. It has been also used to evaluate performances when errors are added in the SAR observations.

RMSE

The RMSE is considered an excellent error metric for numerical predictions. The RMSE measures the square root of the average square error of the predicted water levels (h_k^p) against the truth (h_k^t) per pixel k over the total number of pixels L of the flood domain.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{L}(h_i^p - h_i^t)^2}{L}} \tag{2.15}$$

In this study, the RMSE is a measure of the *global* accuracy of the flood forecasting model predictions of water levels allowing to compare prediction errors of the different assimilation scenarios over the flood domain. The RMSE is evaluated at the assimilation time and also at subsequent time steps. It has been also used to evaluate performances when errors are added in the SAR observations.

2.4. STUDY AREA AND DATA

Our synthetic experiment is grounded on a real test site and an actual storm event: the river Severn in the middle-west of UK (figure 2.2) and the July 2007 flood event, rispectively. This area has experienced several floods along the river valleys (Environment Agency, 2009) generally due to intense precipitation.

While seven upstream catchments contribute to the flow along the river Severn, in our study only one upstream catchment is considered: the Severn at the Bewdley gauging station. Our first objective is to evaluate whether the model correctly predicts the output in the simplest case, i.e. when a unique run-off input to the hydraulic model determines the flood extent and no additional contributing tributaries interfere.

The ERA5 dataset (Hersbach et al., 2019) referring to the period of July 2007 has been used in this experiment. ERA5 is a global atmospheric re-analysis dataset provided by the European Centre for Medium-Range Weather Forecasts (ECMWF). Rainfall and 2 m air temperature at a spatial resolution of approximately 25 km and a temporal resolution of 1 hour are used as input to the hydrologic model. The *true* rainfall time series is used to

generate the *true* run-off before being perturbed in order to obtain 128 different particles as inputs to the hydrologic model. The boundary condition of the hydraulic model is imposed in correspondence to the red dots in Figure 2.2. Channel width, channel depth, slope of terrain, friction of the flood domain and channel bathymetry are defined in each cell of the model domain as described in Wood et al., 2016. A uniform flow condition is imposed downstream. No lateral inflow in the hydraulic model is assumed. Finally, at each time step a stack of 128 wet/dry maps is obtained. Discharges and water levels

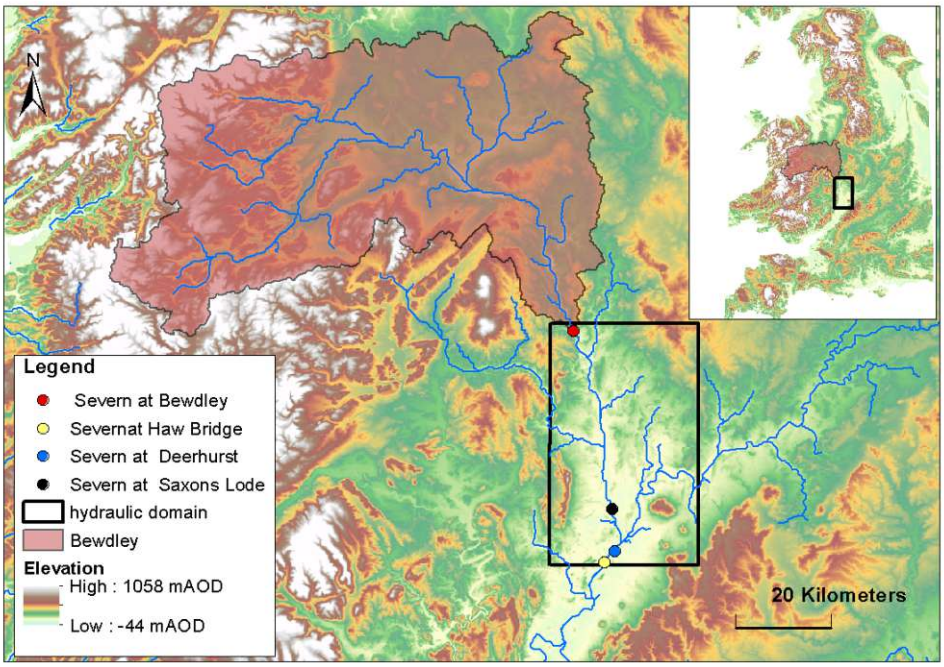


Figure 2.2: Study area: River Severn (UK). Only the boundary condition in Bewdley is taken into account. Within the sub-catchment upstream of Bewdley (area of 4325 km^2) a lumped hydrologic model is used to determine the input of the hydraulic model along the river Severn downstream. The dots represent the existing gauging stations where the performances of the DA framework are evaluated. The black square is the hydraulic domain where LISFLOOD-FP runs.

recorded at different gauging stations (corresponding to the existing ones, dots in figure 2.2) along the river are used to evaluate the performance of the DA.

2.5. RESULTS

2.5.1. SYNTHETIC SAR AND ENSEMBLE GENERATION AND EVALUATION

The virtual satellite acquisition dates are aligned with the actual Sentinel-1 acquisition frequency. The revisit time over Europe, considering both ascending and descending orbits, is around 3-4 days meaning that on average 2 satellite images are available per week. In order to adopt a realistic Sentinel-1-like observation scenario we chose to as-

similate four synthetic observations over a period of 10 days.

In the Figures 2.3 and 2.4 , the area corresponds to the hydraulic model domain. The hydrologic model, covering the upstream catchment, is used to compute the input boundary conditions of the hydraulic model. Results are computed and compared within the hydraulic model domain. The synthetic SAR observations are shown in Figure 2.3. The corresponding PFMs are shown in Figure 2.4 and reliability plots are provided in Figure 2.5. In the reliability plots, the points aligned along the 1:1 line indicate a statistically reliable PFM.

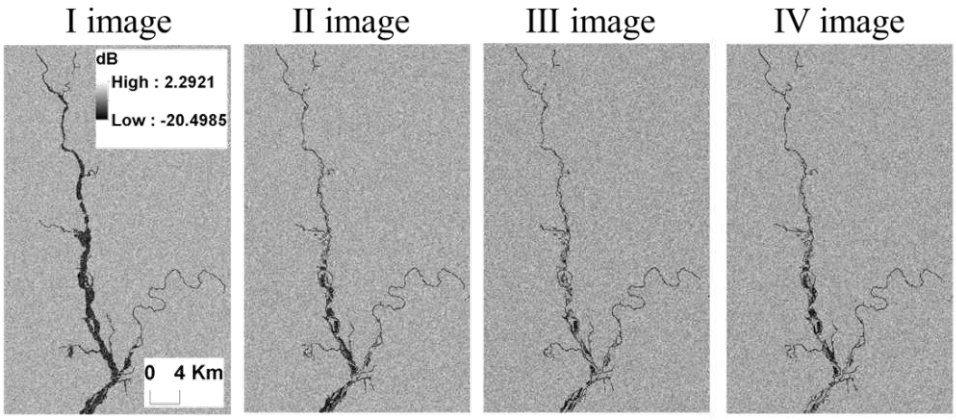


Figure 2.3: A detail of the synthetic SAR images corresponding to the 4 assimilation time steps. Darker pixels correspond to lower backscatter.

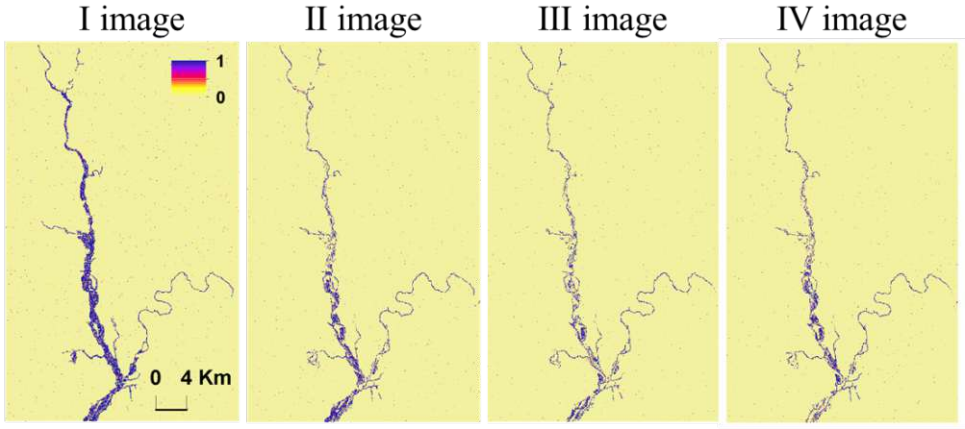


Figure 2.4: A detail of the synthetic probabilistic flood maps derived from synthetic SAR images. Probabilities to be flooded knowing the backscatter go from low value (yellow) to high values (blue).

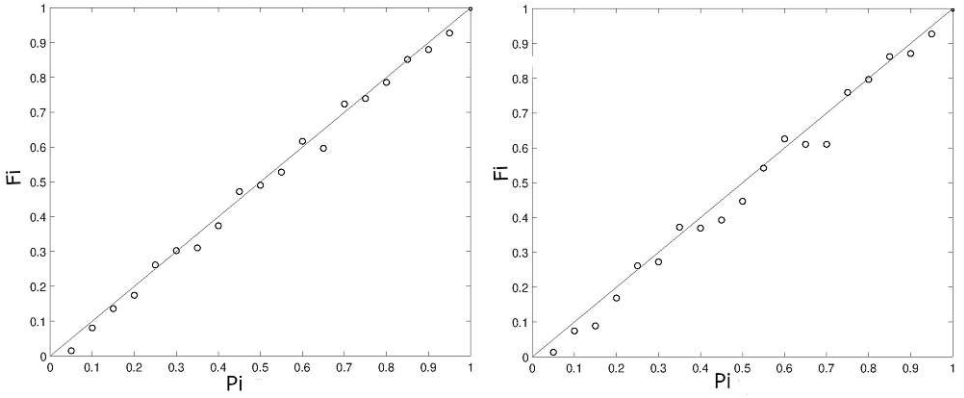


Figure 2.5: Example of the reliability plots for the verification of the synthetic probabilistic flood maps of the first two synthetic SAR images. On the x-axis probability range (P_i), on the y-axis the fraction of pixels within the probability range of the probabilistic flood map observed as flooded in the *true* binary flood map (F_i). The probabilistic flood maps are statistically reliable because the points align along the 1:1 line.

The verification measurements VM_1 and VM_2 (equations 2.2 and 2.6) of the ensemble discharge in Bewdley (figure 2.6) are equal to 1.047 and 0.527, respectively. These values are close to the ideal ones of 1 and 0.5.

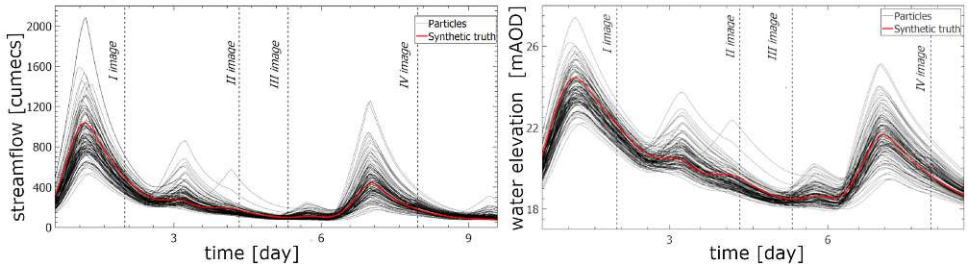


Figure 2.6: Streamflow time series (left) and water elevation time series (right) at the gauge station in Bewdley. Black lines represent the 128 particles while the red line corresponds to the synthetic truth.

2.5.2. EVALUATION OF THE FLOOD EXTENT MAP ESTIMATED AT THE ASSIMILATION TIME

The CSI is computed over the entire hydraulic model domain at each assimilation time step.

The general trend of the assimilation effect is positive, as errors tend to decrease at all the assimilation steps with different assimilation methods. Even though the CSI is already high with the OL, the assimilation further improves the results and this becomes particularly clear at the last assimilation time step. From Table 2.1 it can be noticed that the CSI, approximately equal to 0.80 with the OL in the worst case (assimilation of the IV image), exceeds 0.96 for the different assimilation types and reaches the maximum value

Table 2.1: Critical success index values at each assimilation time step. The Open Loop where no assimilation is computed is compared with the standard method and the adapted method with an increasing effective ensemble size (EES).

Assimilation times	Open Loop	Assimilation				
		standard	*5% EES*	*10% EES*	*20% EES*	*50% EES*
I image	0.9573	0.9887	0.9914	0.9866	0.9805	0.9779
II image	0.9202	0.9873	0.9800	0.9758	0.9658	0.9645
III image	0.9437	0.9921	0.9753	0.9690	0.9622	0.9636
IV image	0.7976	0.9881	0.9754	0.9638	0.9577	0.9610

of 0.99 with the standard method.

In Figure 2.7, we provide the contingency maps of the OL and of the 5% EES approach (results of the standard method are similar to those of the 5% EES approach and therefore not shown). For each pair of images, we show on the left the results of the OL and on the right the results obtained after the assimilation.

In this study, it can be observed that the OL has a tendency to over-detection; the number of red pixels is higher than the black ones and after the assimilation the number of over-detected pixels decreases confirming the results obtained with the CSI.

The confusion matrix given in Table 2.8 provides more details on the 4^{th} assimilation time step. On the one hand, the number of pixels wrongly predicted as flooded in the OL is 1196 and more than 90% of these are correctly classified as non flooded after the assimilation for both standard and 5% EES methods. On the other hand, a few pixels correctly predicted as flooded in the OL are classified not flooded after the assimilation. However, it can be argued that the number of 201 wrongly classified pixels after the assimilation is rather low compared to the 1253 pixels of the OL.

2.5.3. EVALUATION OF THE FLOOD MAP ESTIMATED IN TIME

The flood is simulated using an hourly time step. Consequently, it is possible to evaluate the evolution of the performance metrics CSI (Figure 2.9).

This figure shows that the OL's performance is consistently poor and the standard assimilation performs best compared to the other assimilation runs at all the assimilation time steps.

The assimilation runs with different EES values lie within these two extremes. It can be noted that the more particles are neglected, which is equivalent to say the lower is the EES, the higher is the performance at the assimilation time step.
Moreover, markedly different CSI time series for the different assimilation experiments are shown in Figure 2.9.

27 hours after the first assimilation, the performances of the standard and 5% EES methods, which perform better than the other methods, start decreasing. The lowest values are reached 54 hours after the assimilation. One explanation is that the weights assigned to the particles at the 1^{st} assimilation time are no longer valid when hydraulic conditions change and need to be recomputed.

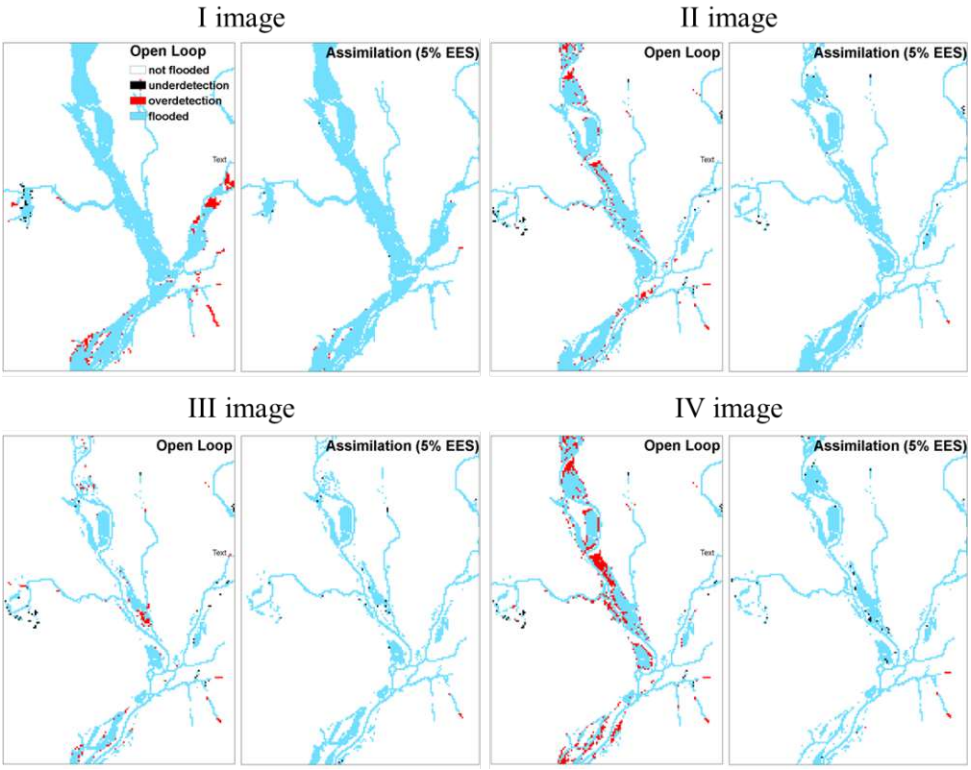


Figure 2.7: Contingency maps before (open loop) and after assimilation at 5 % EES at each time step. Two types of errors can be distinguished: overdetection (red pixels) when the model predicts the pixel as flooded but the pixel is observed as not-flooded and underdetection (black pixels) when the contrary occurs. In case model and observation agree pixels are correctly classified as not-flooded (white pixels) and flooded (blue pixels).

However, things change after the 2^{nd} assimilation, when the performances of the standard and the 5% EES assimilation methods remain stable until the end of the simulation time.

The decrease of performances attributed to the standard and 5% assimilation methods after the 1^{st} time step is due to a drastic change in the flood extent. The total number of flooded pixels reduces from 8539 to 5494 because the flood started receding.

The spread of the posterior pdf with the standard and 5% EES methods is small, meaning that only a few particles retain significant *importance* weight. Consequently, when the flood extent changes and particles evolve in time, it may happen that the uncertainty bounds of the posterior pdf do not encompass the *true* model state after several time steps. On the contrary, when more particles are considered (higher EES), more particles are used to draw the posterior pdf. This gives more chances to the ensemble to encompass the synthetic *truth* and increases the overall robustness of the method. This becomes particularly relevant when the hydraulic boundary conditions change and no new observation is available.

	Open Loop		Standard		Assimilation (5% EES)	
	PF	**PN**	**PF**	**PN**	**PF**	**PN**
OF	4826	57	4748	135	4815	68
ON	1196	264833	66	265963	41	265988

Figure 2.8: Confusion matrix of the OL and of the 5% EES assimilation at the 4^{th} assimilation time step: OF=observed flooded pixels in the truth map, ON=observed non-flooded pixels in the truth map, PF= predicted flooded pixels, PN=predicted non-flooded pixels.

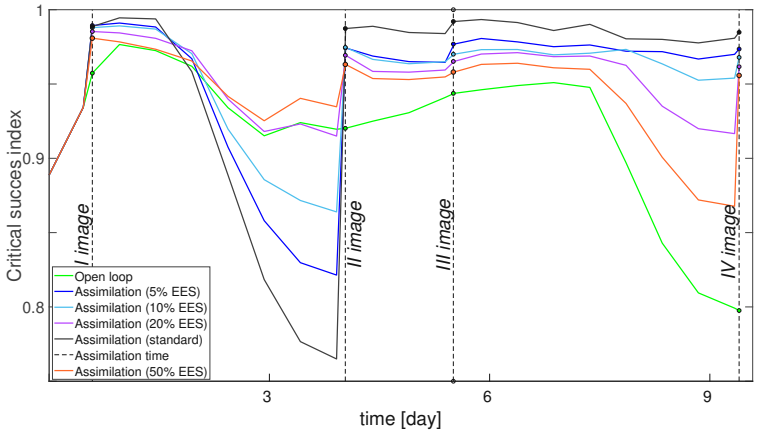


Figure 2.9: Time series of CSI of flood extent values for the different assimilation methods: open loop (green), standard assimilation (black), assimilations with 5% EES (blue), 10% EES (cyan), 20% EES (purple), 50% EES (orange).

2.5.4. EVALUATION OF THE WATER LEVELS IN TIME OVER A GLOBAL SCALE

The RMSE, reported in Table 2.2. decrease by factors larger than 2 and 3 with the standard assimilation and the 5% EES assimilation, respectively.

After the 1^{st} assimilation, carried out close to the flood peak in Saxons Lode, the accuracy of the water level is improved by approximately 20 cm over the entire flood domain. The assimilation of the 2^{nd} and 4^{th} images has a negative effect in case the adapted method 50% EES of the assimilation particle filter is applied: the RMSE increases compared to the OL. As already shown in the Table 2.2 the standard assimilation and 5% EES predictions of water levels provide more accurate results (figure 2.10). When moving away from the first assimilation, the RMSE of the best performing assimilation methods increases. For instance, after 54 hours the RMSE of the standard method is increased by 65% compared to the RMSE of the OL. In case different EES are considered, the RMSE values fluctuates significantly in between two assimilations and it becomes difficult to draw any general conclusions. As the number of *important* particles increases, water levels vary significantly, especially in the area close to the flood edge even though the flood extent does not change too much from a particle to another.

Table 2.2: Root mean square error (RMSE [m]) of water levels at each assimilation time step. The Open Loop (figure 2.6) where no assimilation is computed is compared with the standard method and the adapted method with an increasing effective ensemble size (EES).

Assimilation times	Open Loop	Assimilation				
		standard	*5% EES*	*10% EES*	*20% EES*	*50% EES*
I image	0.2608	0.0742	0.0608	0.0785	0.1501	0.1762
II image	0.1246	0.0526	0.1046	0.1278	0.1553	0.1704
III image	0.1604	0.0645	0.1103	0.1665	0.2154	0.2270
IV image	0.1702	0.0541	0.0619	0.1084	0.1899	0.2205

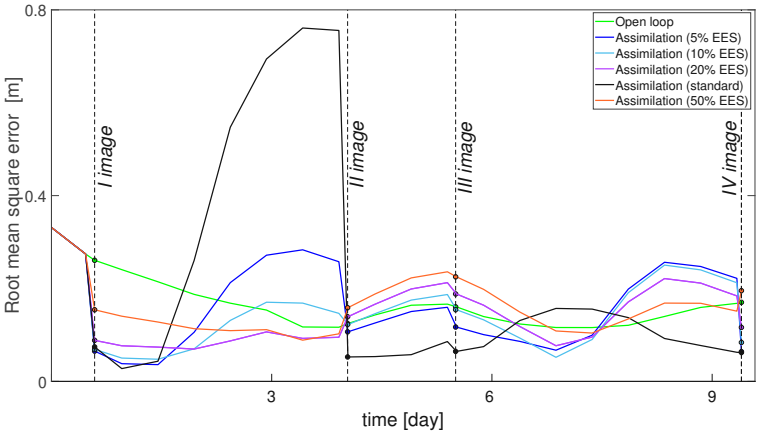


Figure 2.10: Time series of root mean square error (RMSE [m]) values for the different assimilation experiments: open loop (green), standard assimilation (black), assimilations with 5% EES (blue), 10% EES (cyan), 20% EES (purple), 50% EES (orange).

2.5.5. EVALUATION OF DISCHARGE AND WATER LEVEL TIME SERIES

The different assimilation runs are also compared considering the discharges and water levels at different gauge stations along the river Severn. In the right panels of Figures 2.11 and 2.12 the different assimilation experiments are compared against the synthetic truth (red line). In the left panels of Figures 2.11 and 2.12 the standard method and the 5% EES assimilation with the *important* particles and the synthetic truth are shown. The plotted *important* particles represent the 5% of the ensemble with the largest weight. All the 128 particles are equally weighted until the first observation is assimilated. After the first assimilation the number of important particles decreases. At the second assimilation time step, weights are recomputed and the new *important* particles are selected again and so on. The assimilation of the PFMs improves the predictions of water levels and streamflow at specific points of the river Severn, as in Bewdley and in Saxons Lode (Figure 2.11, 2.12), for the majority of the assimilation time steps in both underprediction and overprediction cases. The standard method and similarly the 5 % EES assimilation

method are the most accurate in forecasting the values of water levels and streamflows. The improvements due to the assimilation persist for a long time: up to 27 hours after the first assimilation predictions are still close to the synthetic *truth*. The local results of water levels suggest that the inaccuracy of the global RMSE values in time is likely due to the evaluation over the entire flood domain.

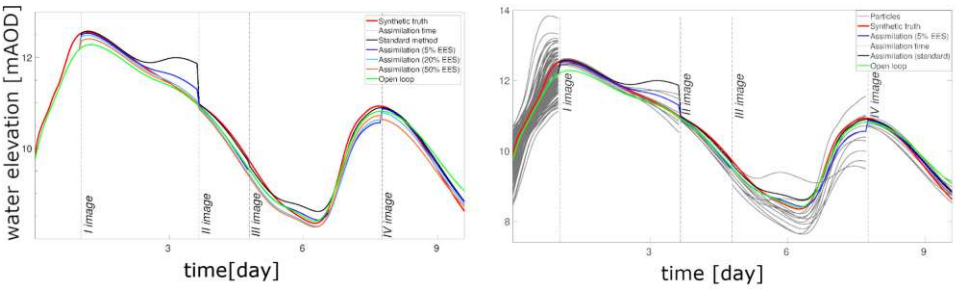


Figure 2.11: Water level time series at Saxons Lode. Left: assimilation runs with an EES of 5% (blue), 20% (cyan) and 50% (orange), OL (green), standard assimilation (black). Right: particles carrying significant weight after the assimilation at 5% EES (grey). Dashed lines correspond to the assimilation times.

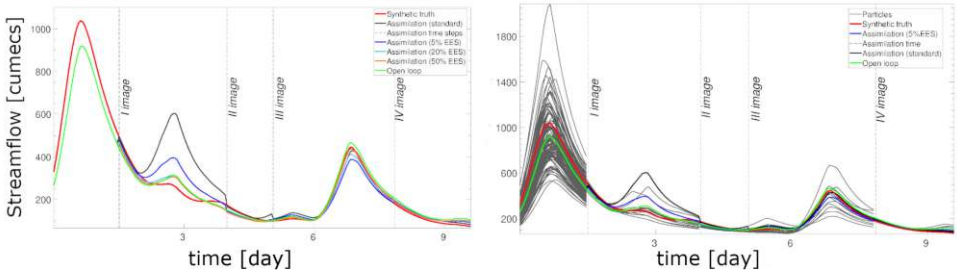


Figure 2.12: Streamflow time series at Bewdley. Left: assimilation runs with an EES of 5% (blue), 20% (cyan) and 50% (orange), OL (green), standard assimilation (black). Right: particles carrying significant weight after the assimilation at 5% EES (grey). Dashed lines correspond to the assimilation times.

2.5.6. IMPACT ASSESSMENT OF ERRORS IN SAR OBSERVATIONS

In the previous section, speckle uncertainty in SAR observations is considered. However, in reality, SAR observations are also susceptible to errors due to the misclassification of wet/dry pixels caused by features on the ground as already mentioned. Therefore, errors are added to the synthetic SAR observations as described in the methodology to investigate the impact on the DA assimilation framework. Figure 2.13 shows the RMSE and the CSI obtained at different assimilation time steps. The best performing assimilation methods (i.e. standard and 5% EES) with no error in the observations are compared with the ones where error is introduced. With the misclassification of 20% of the pixels, the assimilation still has beneficial effects: the CSI increases at each assimilation time step with respect to the OL. The RMSE values also tend to be satisfactory after each as-

simulation. With an increase in the error of 40% the performances of the DA framework start decreasing. The assimilation of the first image still has a positive effect on the predictions. In fact, CSI and RMSE are improved with respect to the OL even through the improvements are not as significant as in the previous cases. The explanation is arguably to be found in the high number of flooded pixels. It is large enough to counterbalance the misclassified pixels in the SAR image. Performances decrease with the assimilation of the remaining SAR observations when the number of flooded pixels is reduced by half.

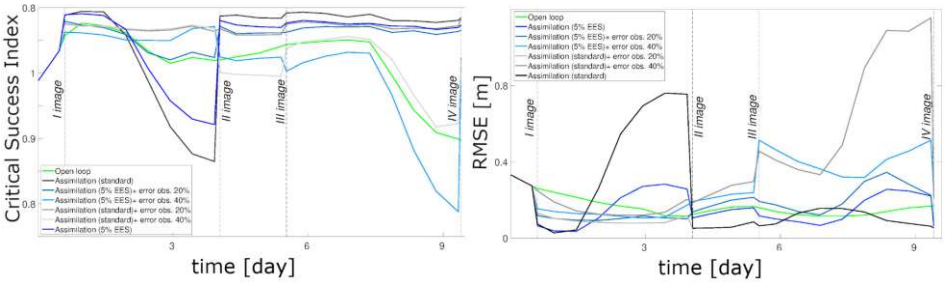


Figure 2.13: CSI values (on the left) and RMSE (on the rigth) after the standard assimilation of SAR observations free of errors (black), with 20% of errors (grey), with 40% of errors (ligth grey) and after the 5% EES assimilation free of errors (blue), with 20% of errors (ligth blue), with 40% of errors (cyan).

2.6. DISCUSSIONS

The results of our study confirm the effectiveness of the proposed DA framework when the hypothesis of the rainfall as the main source of uncertainty is verified. Consequently, for those cases where rainfall represents the main source of uncertainty, more obviously but not only in poorly and un-gauged catchments and when using medium-range forecasting models, our study results indicate that the application of the approach described in the manuscript may lead to improved results of the model simulations. For those cases where the uncertainty of other sources becomes more relevant and may be even dominant, it is clear that such sources need to be taken into account explicitly. However, the required adaptations of the proposed DA framework still need to be developed. In this context it is also worth mentioning that the limitations identified in the previously published real case study by Hostache et al., 2018 were explained by additional sources of uncertainties not taken into account.

Using probabilistic flood maps or backscatter values increases the number of observations to be assimilated when compared to a method that only derives the flood edge from satellite observations as reported in Cooper et al., 2018b. Moreover, the nearly-direct use of the SAR information enables a faster end-to-end processing from the acquisition of the image to the assimilation of the SAR data into the model which is beneficial for an operational usage.

In our experiments, the improvements of model forecasts of water level and streamflow are significant at the assimilation time step and the improvements persist over subse-

quent time steps (for example up to 27 hours after the first assimilation the model results outperform the open loop simulation). The persistence of these improvements depends on the flashiness of the flood event (i.e., the rapidity with which hydrologic conditions change). More frequent image acquisitions could help keep model predictions on track, especially when the system is highly dynamic. The update of a state variable of the forecasting model could as well increase the persistence of the improvements. In our study none of the model state variables is updated as only the particle weights are computed, based on the SAR observations and on the simulated flood extent maps and used to calculate the expectation of water levels and streamflow. In previous studies (Andreadis et al., 2007; Matgen et al., 2010; Cooper et al., 2018a), inflow updating was identified as a condition leading to more persistent improvements. For instance, one of the conclusions from the study by Matgen et al., 2010 was that updating the fluxes at the upstream boundary conditions, rather than the water levels, is more effective because of the high uncertainty of the inflow due to the poorly known rainfall distribution over the catchment. Therefore, as a future perspective, we aim to update hydrologic model states because it might have a positive impact on the long-term runoff simulations and consequently on the persistence of DA benefits.

Some modifications of the DA framework are still required to fully overcome the issue of degeneracy. Although the use of a smaller tempering coefficient leads to a larger effective ensemble size (e.g. 50 %) and helps avoid degeneracy, the results are less accurate compared to the standard method or the adapted method with 5% EES. As described in Neal, 1996 and in Leeuwen et al., 2019, the tempering procedure consists of several steps, but in this study the tempering coefficient is applied only to flatten the likelihood, therefore down weighting the observations. This most likely explains why the data assimilation performs better when the effective ensemble size (the number of particles not negligible after the assimilation) is smaller. As already mentioned, the present study has the aim of assessing and validating the method proposed by Hostache et al., 2018 in a synthetic environment. Our DA framework can be applied to a variety of flood inundation forecasting chains. In fact, the forecast updating is carried out via a sequential importance sampling only (i.e. importance weights). Only the particle weights are updated based on the observations and used to compute the expectation (i.e. weighted mean) of the augmented state vector including hydraulic state variables of water depth, plus flood extent and boundary conditions. In this study the hydrologic and hydraulic models are loosely coupled with a one-way transfer of information as in many other studies [e.g., Peckham et al., 2013, Hoch et al., 2017, Rajib et al., 2020]. The weights define the relative importance of the particles and thus of the inherent streamflow and stage along the entire river. We acknowledge that the observed flood extent is more closely linked to the past boundary conditions rather than the boundary conditions corresponding to the assimilation time steps. In spite of this limitation we argue that in this synthetic experiment, the particles that performed best in the past are also those that reach the highest performance level at the time of the assimilation. This is illustrated in the Figures 2.11 and 2.12 where the use of updated weights is shown to enable the correction of the state variables of the hydraulic model both upstream and downstream. However, we recognize that further improvements could be developed to address issues such as spurious relations that may occur between SAR observations and model variables due to a rather

small ensemble size. Enlarging the ensemble size could be necessary if this occurs.
We also argue that the method used in the manuscript has the potential to support EO-based modelling at large scale. This potential is particularly high in large, natural floodplains where flood inundation remains present over long time periods. In spite of the increased frequency of satellite observations, the persistence of a flood over many days increases the chance of its detection and mapping by satellite sensors. Another condition that needs to be satisfied is that there should be an unambiguous relationship between the flood extent observed by the spaceborne sensors and river discharge. This also means that areas where backscatter variations are not impacted by the appearance of floodwater (e.g. densely vegetated floodplains) should be rather small. Indeed, these constraints must be satisfied to enable a successful application of the proposed framework and to take advantage of the analysis carried out in this manuscript. As a conclusion, based on the above elements, we argue that our approach is valid regardless of the type of model coupling that is performed and is thus applicable to many different forecasting systems. However, more research is needed to fully understand the role of floodplain and water basin characteristics and SAR data properties on the DA performance. In a future study it is envisaged that to avoid degeneracy and keep a larger effective ensemble size, the full tempering scheme will be applied. Possible ways to adapt and advance the proposed DA framework are currently under development (e.g. updating a state variable of the model, using an enhanced version of the adapted filter).

2.7. CONCLUSIONS

Satellite images provide valuable information about flood extent that can complement or substitute in situ measurements. The fact that several space agencies provide free access to high resolution satellite Earth Observation data paves the way for improving Earth Observation-based flood forecasting and reanalyses worldwide. This study represents a follow-up of the previous real case study from Hostache et al., 2018 with the objective to further proceed in the evaluation of the proposed DA framework once the assumptions are effectively satisfied. This study has been set up in a controlled environment using a synthetically generated data-set in order to make sure that the rainfall and SAR observations are the only source of uncertainty. A common issue in Particle filters is degeneracy: the ensemble could collapse after the assimilation because higher probabilities are assigned to a limited number of particles. The tempering coefficient can be used to reduce degeneracy because it inflates the posterior probability and reduces the peak of the likelihood. In this study, we have evaluated the effect of variations of the α tempering coefficient on the DA performance. Different PFs are compared with the OL and the synthetic truth: the SIS (with only a few particles from the ensemble potentially carrying non- negligible weights) and the adapted method with 5-10-20-50% EES (with the number of particles with non-negligible weights increasing with the EES). This methodology leads to slightly biased estimates because the observation is down-weighted. In addition, we investigated the impact of errors in the observations (i.e. errors in the SAR derived pfms due to dry water-look alike pixels or emerging objects) on the assimilation. Indeed, the main issue of using SAR observations in flood forecasting models is the difficulty of detecting flooded area for specific cases (e.g. urban or vegetated areas). At first, following the study from Hostache et al., 2018 only speckle uncertainty of the SAR image

is taken into account in the pfms. In a second step, a error to reproduce misclassified pixels is introduced in the synthetic SAR observations.
The following key conclusions can be drawn from our experiments:

1. The best performing method is the standard method (i.e. SIS). Importance weights are assigned to a limited number of particles that better agree with the observations. At the time of the assimilation, results tend to be very accurate: the forecasts move close to the synthetic truth. The main weakness of the standard filter is to significantly suffer from degeneracy.

2. The 5% effective ensemble size assimilation (meaning that only the 5% of the ensemble will have a not-negligible weight after the assimilation) is slightly less accurate at the time of the assimilation but it has the advantage of reducing the degeneracy problem. Even though larger effective ensemble size prevents degeneracy, water levels and discharge are less accurate and performances of the predictions are degraded.

3. Our study further shows that it is important to characterize and mask out errors in the SAR observations. A large number of misclassified pixels substantially degrades the DA performance. In our case study, results suggest that an improvement of model simulations (i.e. water level and streamflow) in terms of CSI and RMSE performance metrics is achieved as long as errors in the observations are rather limited, i.e. when no more than 20% of the pixels are affected. However, if the misclassification goes beyond 40% of affected pixels, the assimilation has no effect and may even lead to a degradation of the model predictions.

3

A TEMPERED PARTICLE FILTER TO ENHANCE THE ASSIMILATION OF SAR DERIVED FLOOD EXTENT MAPS INTO FLOOD FORECASTING MODELS.

The present chapter corresponds to the following scientific publication in its original form: Concetta Di Mauro, Renaud Hostache, Patrick Matgen, Ramona Pelich, Marco Chini, Peter Jan van Leeuwen, Nancy K. Nichols, and Günter Blösch. "A tempered particle filter to enhance the assimilation of SAR derived flood extent maps into flood forecasting models. Water Resources Research, 2022, https://doi-org.proxy.bnl.lu/10.1029/2022WR031940"

3.1. ABSTRACT

Data Assimilation (DA) is a powerful tool to optimally combine uncertain model simulations and observations. Among DA techniques, the Particle Filter (PF) has gained attention for its capacity to deal with non-linear systems and for its relaxation of the Gaussian assumption. However, the PF may suffer from degeneracy and sample impoverishment. In this study, we propose an innovative approach, based on a Tempered Particle Filter (TPF), aiming at mitigating PFs issues, thus extending over time the assimilation benefits. Probabilistic flood maps (PFMs) derived from Synthetic Aperture Radar data are assimilated into a flood forecasting model through an iterative process including a particle mutation in order to keep diversity within the ensemble. Results show an improvement of the model forecasts accuracy, with respect to the Open Loop (OL): on average the RMSE of water levels decrease by 80% at the assimilation time and by 60%

two days after the assimilation. A comparison with the Sequential Importance Sampling (SIS), is carried out showing that although SIS performances are generally comparable to the TPF ones at the assimilation time, they tend to decrease more quickly. For instance, on average TPF-based RMSE are by 20% lower compared to the SIS-based ones two days after the assimilation. The application of the TPF determines higher CSI values compared to the SIS. On average the increase in performances lasts for almost 3 days after the assimilation. Our study provides evidence that the application of the variant of the TPF enables more persistent benefits compared to the SIS.

3.2. INTRODUCTION

Every year, floods cause important social and economic losses and the trend is increasing. Tellman et al., 2021 show that worldwide the population exposed to floods has increased by 20%–24% from 2000 to 2015, thereby highlighting the need for accurate and timely forecasts of water depth, discharge, flood wave propagation, and flood extent to help reducing or preventing the adverse effects of floods. Flood forecasting models are commonly used to generate short- to mid-term predictions. However the accuracy of such predictions can be affected by multiple factors contributing to the overall model uncertainty. This challenge represents one of the major unsolved scientific problems (Blöschl et al., 2019b). The assimilation of independent observations, such as field gauging data or satellite observations, can help reducing these uncertainties (Liu et al., 2007). The last decade has seen a substantial increase in the number of Earth Observation (EO) satellites providing a synoptic overview of the flooding situation at increasingly high frequency. Despite possible errors in the interpretation of the Synthetic Aperture Radar (SAR) data (Chen et al., 2018; Grimaldi et al., 2020; Zhao et al., 2021) that should be masked out before any use of these data, frequent observations of flood extent and water depth represent substantial added value, especially over poorly gauged or ungauged catchments. For example, SAR data are relevant for observing inundation extent because of their day-night and quasi all-weather capability. As a consequence, several methods enabling an effective assimilation of such observations [e.g., Revilla-Romero et al., 2016; Hostache et al., 2018; Andreadis et al., 2014; Garcia-Pintado et al., 2015] for improving the predictive capability of flood models have been introduced and investigated in recent years. The most widely used methods are based on the Kalman Filter and its variants [e.g. Revilla-Romero et al., 2016; Annis et al., 2021; Wongchuig-Correa et al., 2020] and they assume that the distributions of observation and model errors are Gaussian, which is not often the case when dealing with real word data (Leeuwen et al., 2019).

Particle Filters (PFs) have gained attention within the research community because of their ability to handle non-linear and non-Gaussian systems (Leeuwen et al., 2019). PFs approximate the prior and the posterior probability distribution functions (pdfs) with an ensemble of model states also called particles. An equal weight is assigned to each particle a priori. Next, as a result of the assimilation, weights are updated to represent the posterior probability given the observations. The principal limitation of PFs is the difficulty to deal with high-dimensional systems. The weights may vary significantly across particles and in the ultimate case only one particle will have a weight close to unity while the other particles will have negligible weight. As a result the ensemble may collapse. This well-known issue in PFs is often referred to as degeneracy. Degeneracy could

lead to an erroneous approximation of the posterior distribution (García-Pintado et al., 2013) and a sub-optimal use of the assimilation filter. Resampling methods [e.g Gordon et al., 1993] have been used to prevent the collapse of the ensemble: particles with significant weights are replicated and non-significant particles are discarded. Even though resampling is powerful in reducing degeneracy, it often comes with a sample impoverishment and a poor representation of the actual uncertainty of the system (Moradkhani et al., 2012). After few iterations, replicated particles will hardly diversify and particles will again collapse into a single or few particles. According to Snyder et al., 2008, the number of particles should grow exponentially with the dimension of the system, otherwise the PF may suffer from degeneracy. Of course, a higher number of particles implies an increased computational cost which may hamper the use of DA in near real-time application. As a consequence, it is important to minimize the weight variance so that each particle keeps a significant weight.

In Chapter 2 and in the study by Hostache et al., 2018, a data assimilation framework based on Sequential Importance Sampling (a variant of PFs) has been developed following a similar previous work by Giustarini et al., 2011b, that enables an efficient assimilation of SAR data into a hydrodynamic model. In their experiment, the rainfall forcing and the SAR data are assumed to represent the only sources of uncertainty. While in Chapter 2 is shown that the SIS method provides good results when the assumptions are indeed satisfied, they also highlight the need for a method to mitigate degeneracy and sample impoverishment. The assimilation via a SIS tends to degenerate with only a few particles getting significant weights as a result of the assimilation. A preliminary attempt to mitigate the degeneracy consisted in using a tempering coefficient for the inflation of the posterior probability. The likelihood was raised to the power of a coefficient whose value enables a substantial increase of the likelihood variance. However, using this coefficient to inflate the likelihood only partially solved the degeneracy issue, and sometimes at the cost of a decrease in prediction accuracy.

To mitigate the mentioned PF-related issues, the following approaches have been introduced in the literature:

1. Using a one-step proposal density to steer particles in such a way that they obtain similar weights (Doucet et al., 2001; Van Leeuwen, 2009);

2. Moving the particles from the prior to the posterior by applying a smooth iterative transition process using model transitional densities (Beskos et al., 2014).

3. Using particles filters within Monte-Carlo Markov Chains (PMCMC) (Andrieu et al., 2010)

These methodologies are exact in the limit of an infinite ensemble size. Many approximate algorithms exist that have been used in hydrological sciences such as: localizing PFs, in which observations are only allowed to influence nearby elements of the state vector (Van Leeuwen, 2009; Reich, 2013), bringing in approximate elements of Ensemble Kalman Filters into the PF (Potthast et al., 2019; Frei et al., 2013), using approximate MCMC steps within the PF proposal step (PF-MCMC) (Moradkhani et al., 2012), combining the PF with Metaheuristic-Algorithms (MA) from Computer Science, such as Genetic Algorithm (GA) (Park et al., 2010; Kwok et al., 2005), Particle Swarm Optimization (PSO)

(Wang et al., 2006; Li et al., 2005), and the Immune Genetic Algorithm (IGA) (Han et al., 2011), combining the MCMC with GA algorithms and use it within the importance sampling step of the PF-MCMC, known as Evolutionary Particle Filter with Markov Chain Monte Carlo (EPFM) (Abbaszadeh et al., 2018), using 4DVar as an extra proposal density in an EPFM, known as Hybrid Ensemble and Variational Data Assimilation framework for Environmental systems method (HEAVEN) (Abbaszadeh et al., 2019). The evolutional swarm-like PFs contain several steps and assumptions for mutation and crossover without guaranteeing convergence to the full posterior pdf in the limit of an infinite ensemble size. Less significant approximations are needed in the Evolutionary PF-MCMC (EPFM) method described in Abbaszadeh et al., 2018 where GA-MCMC is used to define the importance sampling step. EPFM outperforms the PF-MCMC providing more accurate and reliable results and overcomes the limitations of the recent standard PF-GA algorithm where parameters of crossover and mutation steps need to be tuned. The EPFM method uses crossover and mutation step to generate new proposal model states. The crossover step consists in a linear combination of parent particles. The mutation process is carried out to increase the diversity among the particles. Afterwards, the proposal particles are further refined with the MCMC approach. A Gaussian distribution of the proposal state is assumed to calculate metropolis acceptance ratio in the MCMC step. The HEAVEN (Abbaszadeh et al., 2019) integrates the EPFM algorithm and the 4D-VAR to also account for model structure uncertainty other than model parameters and input uncertainties. Abbaszadeh et al., 2019 show that HEAVEN outperforms EPFM and better simulates streamflow in high flow regimes. In this study, we adopt and evaluate an enhanced PF following the results of the previous studies in Chapter 2 and Hostache et al., 2018. The DA approach, hereafter called Tempered Particle Filter (TPF), applies tempering coefficients to inflate the likelihood within an iterative process so that the Bayes' formula is respected (Beskos et al., 2014). The method is based on the method first proposed by Neal, 1996, combined with ideas from Herbst et al., 2019. The iterative assimilation approach is based on successive Sequential Importance Resamplings (SIRs) and particle mutations (Han et al., 2011; Li et al., 2005; Abbaszadeh et al., 2018). The mutations enable the ensemble to regain diversity after each resampling step in each iteration and are based on a Metropolis Hasting (MH) algorithm. We hypothesize that the proposed DA methodology enables the mitigation of some PF limitations, sample degeneracy and sample impoverishment, while preserving the assimilation performances in terms of flood extent, discharge and water level simulations.

In this study, we also further investigate additional benefits that come from this new approach. According to Dasgupta et al., 2021, degeneracy plays a crucial role in the persistence of the assimilation benefits over several time steps. Therefore the TPF approach could also help with improving the persistence of the assimilation benefits. Moreover, DA algorithms often assume that the observations as well as the model predictions are unbiased. Many authors pointed out the importance of bias removal before the DA, but it is not a straightforward procedure, especially in model forecasts (De Lannoy et al., 2007). Bias can depend on the model structure or parameters, on the initial conditions, or on forcing errors (especially when the forcings are derived from a forecast model, as in this study). In this context, we hypothesize that the new approach based on a TPF enables the reduction of bias in the model predictions and we test this hypothesis. To

enable a meaningful evaluation and to verify whether the new approach outperforms the previous one, the TPF performance is compared to that of the SIS.

We carry out twin experiments based on a synthetically generated data-set with controlled uncertainty. The SAR observations are synthetically generated from the simulated flood extent maps and assimilated into a coupled hydrologic-hydraulic model. Two different background ensembles, i.e., Open Loops, are drawn and used: in the first case the ensemble encompasses the synthetic truth most of the time, in the second case the ensemble is most of the time outside the ensemble range.

The objectives of this study are therefore i) to evaluate whether a principled method, in which the only approximation is the finite ensemble size, can mitigate degeneracy, ii) to evaluate whether the proposed framework improves the prediction accuracy and increases the persistence of the assimilation benefits, iii) to evaluate the efficiency of the method in reducing forecast bias. The paper is structured as follows: section 3.3 describes the materials and methods, section 3.4 showcases and discusses the results and 3.5 draws the conclusions of the study.

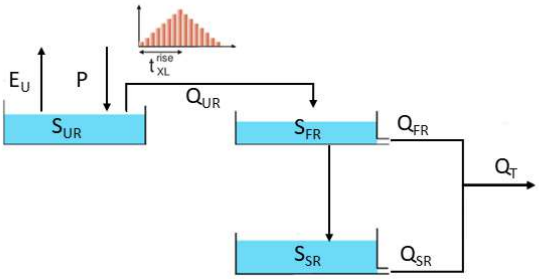
3.3. MATERIALS AND METHODS

The first part of this section presents the structure of the flood forecasting system. The second part describes the proposed assimilation framework based on a TPF. The experimental design, case study, and the performance metrics used within this experiment are introduced in the last part.

3.3.1. THE FLOOD FORECASTING MODEL

We use the ERA5 data-set (Hersbach et al., 2019) to derive the forcing of the flood forecasting system. Rainfall and 2 m air temperature at a spatial resolution of approximately 25 km and a temporal resolution of 1 hour are used as inputs to the flood forecasting system. A conceptual hydrological modelling framework (SUPERFLEX) coupled with a hydraulic model (LISFLOOD-FP) approach has been adopted: the run-off estimated with the hydrological model is used as input to the shallow water hydraulic model. In this study, the rainfall-runoff model SUPERFLEX (Fenicia et al., 2011) is a lumped conceptual model. The state variables and the parameters used are listed in Figure 3.1. The conceptualization model is composed of three reservoirs: an unsaturated soil reservoir with a storage S_{UR} representing the root zone, a fast reservoir with storage S_{FR} representing the fast responding components (e.g., the riparian zone and preferential flow paths), and a slow reservoir with storage S_{SR} representing slow responding components (e.g., deep groundwater). A lag function is used at the outlet of the unsaturated soil reservoir to enable a delayed hydrological response of the basin under intense rainfall conditions. The hydraulic model is based on LISFLOOD-FP (Bates et al., 2000; Neal et al., 2012) and simulates flood extent, water level, and discharge within the hydraulic model domain. The roughness coefficient and the bathymetry of the hydraulic model have been previously calibrated (Wood et al., 2016)

ERA5 rainfall time series are used to generate the synthetic truth and are also perturbed to generate an Open Loop (OL) simulations consisting in 32 particles. These 32 particles are then used as input to the flood forecasting model to obtain the ensemble



Parameters	Units	Definition
t_{rise}	h^{-1}	Hydrograph lag-time
K_{FR}	h^{-1}	FR time scale
K_{SR}	h^{-1}	SR time scale
S_{max}	mm	Maximum UR storage
β	-	limit for potential transpiration
K_{UR}	h^{-1}	UR outflow rate scale
α_{FR}	-	FR power coefficient
α_{SR}	-	SR power coefficient
State variables	**Units**	**Definition**
S_{UR}	mm	UR storage reservoir
S_{FR}	mm	FR storage reservoir
S_{SR}	mm	SR storage reservoir
Q_{UR}	$mm \cdot h^{-1}$	UR discharge
Q_{FR}	$mm \cdot h^{-1}$	Fast discharge
Q_{SR}	$mm \cdot h^{-1}$	Slow discharge

Figure 3.1: Scheme of the SUPERFLEX model used in this study. The hydrological model is based on three reservoirs: an unsaturated soil reservoir (S_{UR}), a fast run-off reservoir (S_{FR}), and a slow run-off reservoir (S_{SR}). The discharge deriving from the 3 reservoirs are: Q_{UR}, Q_{FR}, Q_{SR}. A triangular lag function with a base length equal to $2 \cdot t^{rise}$ is applied at the outflow of the unsaturated soil reservoir. E_U and P represents the potential evaporation and rainfall respectively.

of flood extent maps. We adopt the method proposed and detailed in Chapter 2 to generate synthetic observations from model results. The flood extent map of the synthetic truth together with a real SAR observation are used to compute Probabilistic Flood Maps (PFMs) where each pixel represents the probability to be flooded given the recorded backscatter values (Giustarini et al., 2016). During the analysis (i.e., assimilation) step, the generated PFMs are assimilated into the ensemble of wet-dry maps via the TPF to obtain the updated particles. The following section describes the data assimilation framework.

3.3.2. DATA ASSIMILATION FRAMEWORK

PFs are based on Bayes' theorem (Equation 1.1)

The observation y at a given time k, which is the probability to be flooded given the SAR backscatter value, is combined with the forecasts of the numerical model x at a given time k. The posterior probability $p(x \mid y)$ is computed by multiplying the prior probability density function $p(x)$, which is the probability of the model before any observation is taken into account, with the likelihood $p(y \mid x)$ that is the probability density that the model state x_n produces the observation. In PFs the prior Pdf is drawn from an ensemble of model states of size N called particles. Equation 1.3 represents the computation of the prior probability. Inserting Eq. 1.3 into Eq. 1.1 leads to the posterior probability formula (Equation 1.4).

The weights W_n, hereafter called global weights, were computed by the multiplication of the pixel-based local weights w_i^n, according to the formula 2.8, assuming that

observation errors are independent across space. The set of particles tends to degenerate: after the assimilation, the number of particles with significant weight is reduced to a few and the posterior distribution is poorly approximated. In Chapter 2, a first attempt is made to reduce degeneracy, within this DA framework, using a tempering coefficient γ according to the formula:

$$p(x \mid y) = \left(\frac{p(y \mid x)}{p(y)} \right)^{\gamma} p(x) \quad with\ \gamma \in [0,1] \tag{3.1}$$

This technical solution enables inflating the posterior variance so that several particles keep significant weight. However, it is an approximate solution as not all information from the observations is taken into account.

In the current study we aim to further improve the application of the likelihood tempering. The proposed method relies on the factorisation of the likelihood through an iterative approach according to the following formula:

$$\frac{p(y \mid x)}{p(y)} = \prod_{s=1}^{S} \left(\frac{p(y \mid x)}{p(y)} \right)^{\gamma_s} \tag{3.2}$$

where $0 < \gamma_s < 1$ for each iteration, s and $\sum_{s=1}^{S} \gamma_s = 1$.

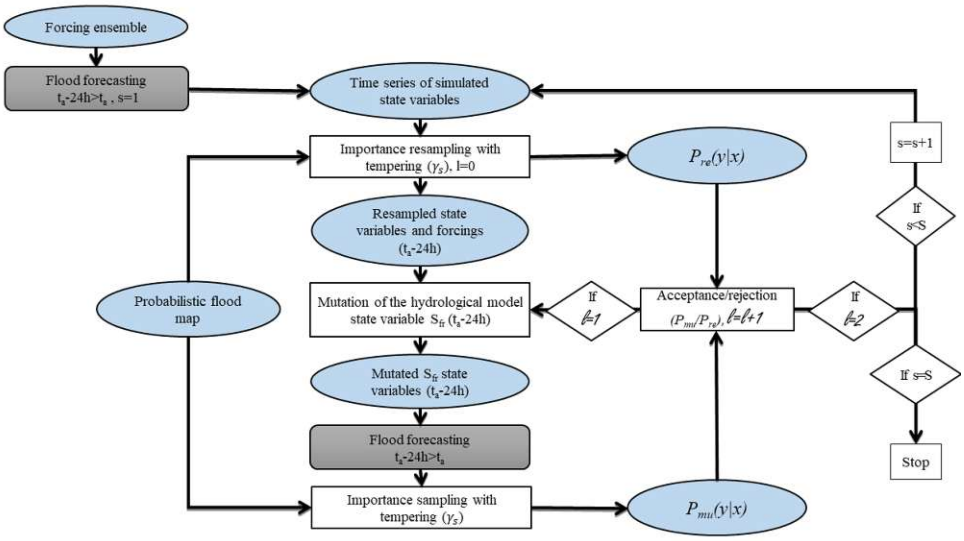


Figure 3.2: Flow chart of the DA framework where synthetic probabilistic flood maps are generated from flood extents, derived from a truth run, and assimilated within the same flood forecasting model. The flood forecasting model is represented with a gray rectangle, mathematical operations with a white rectangle, state variables, input, and observations with a blue ellipse.

This factorization enables application of the Bayes' theorem iteratively so that the transition from the prior to the posterior probability is smoothly processed. The iterative methodology leads to the following equation after one iteration:

$$p(y \mid x) = \prod_{s=2}^{S} \left(\frac{p(y \mid x)}{p(y)} \right)^{\gamma_s} \left(\frac{p(y \mid x)}{p(y)} \right)^{\gamma_1} p(x) \tag{3.3}$$

leading to

$$p(y \mid x) = \prod_{s=2}^{S} \left(\frac{p(y \mid x)}{p(y)} \right)^{\gamma_s} p_1(x \mid y) p(x) \tag{3.4}$$

and:

$$p_1(x \mid y) \approx \sum_{n=1}^{N} W_n^{(1)} \delta(x - x_n) \quad \text{with } \mathrm{W_n^{(1)}} = \left(\frac{\mathrm{p(y \mid x)}}{\mathrm{p(y)}} \right)^{\gamma_1} \tag{3.5}$$

At each iteration s, the tempering coefficient γ_s enables inflation of the likelihood variance and reduction of the weight variance, therefore reducing degeneracy. The exponent γ_s allows to keep a substantial number of particles with significant weights. At each iteration s, the γ_s value is increased and represents the solution to Equation 3.6:

$$InEff(\gamma_s) = r^* \tag{3.6}$$

where the ensemble inefficiency ratio ($InEff$) is given by Equation 3.7:

$$InEff(\gamma_s) = \frac{1}{N} \sum_{n=1}^{N} (W_n^s(\gamma_s))^2 \tag{3.7}$$

and a target value r^* of the $InEff$ is previously defined. Iterations are stopped when $InEff(1) < r^*$ and $\gamma_S = 1 - \sum_{s=1}^{S-1} \gamma_s$, where S is the total number of iterations.

After each iteration s, the particles with high weights are resampled using the SIR algorithm proposed by Gordon et al. (1993). Particles are replicated proportionally with their weights: those with an associated low importance weight are replaced with replicas of those having higher weight. After resampling, particles are equally weighted.

Next, a mutation is applied to the fast run-off reservoir level (S_{FR}), a variable of the hydrological model, 24 hours prior to the assimilation to regain diversity within the particle ensemble and the mutated value is used as initial condition for a subsequent model simulation over the 24 hours preceding the assimilation time. Mutating the hydrological state variable 24 hours prior to the assimilation time and carrying out the related model simulations is done in order to update the hydrological and hydraulic models in a more consistent way since the water depths simulated by the hydraulic model at a certain time are the result not only of the current but also of the past upstream streamflow conditions.

This mutation is carried out using a MH algorithm, based on a random perturbation via the steps of Markov chain Monte Carlo (MCMC) methods. Since the model is deterministic a mutation of the state 24-hours back in time leads to a corresponding unique mutation at present time. This allows us to write $p(y^k \mid x_n^k) = p(y^k \mid x_n^{k-1})$ for each particle j. Hence, the MH is based on two steps: first, draw a new particle from a proposal density as $x^* \sim q(x|x_n^{k-1})$, and then calculate the MH acceptance ratio:

$$\alpha = \min \left\{ 1, \left(\frac{p(y^k \mid x^*)\, p(x^*)}{p(y^k \mid x_n^{k-1})\, p(x_n^{k-1})} \frac{q(x_n^{k-1} \mid x^*)}{q(x^* \mid x_n^{k-1})} \right) \right\} \tag{3.8}$$

Many possibilities are available for choosing the proposal density $q()$. As detailed below, a symmetric proposal density cancel the proposal density ratio leading to the so-called MH. Furthermore, since the prior 24 hours back is much wider than the likelihood, we can safely ignore the ratio $p(x^*)/p(x_n^{k-1})$, and in this case the acceptance ratio becomes:

$$\alpha = \min\left\{1, \left(\frac{p(y^k \mid x^*)}{p(y^k \mid x_n^k)}\right)\right\} \tag{3.9}$$

where x_n^k represents the particles with high weight that have been resampled. A random variable $u \sim U[0,1]$ is drawn and the mutated particle is accepted if $\alpha > u$, otherwise we keep the particle as before its mutation.

As proposed by Herbst et al., 2019, the mutation is carried out based on a proposed innovation $p(x^* \mid x^{k-1}) = N(x^{k-1}, c_s^2 \cdot \sigma^2)$, with c_s being a scaling factor given by the following equation:

$$c_s = c_{s-1}\left(0.95 + 0.10 \cdot \frac{e^{20\cdot(\alpha-0.4)}}{1+e^{20\cdot(\alpha-0.4)}}\right) \tag{3.10}$$

c_s at the first iteration is set to 0.2. The mutation step is repeated for $l = 1, .., N^{MH}$. In our study $N^{MH} = 2$.
In detail, the method is structured according to the following steps (Figure 3.2):

- Ensemble forcing are used as input to the flood forecasting model;
- The hydrodynamic simulations are carried out over the 24 hours prior to the assimilation.
- Calculate $p(y|x_i)$ for each particle i and find γ_1 such that $InEff(\gamma_1) = r^*$.
- Particles are resampled using the tempered weights. The particles after resampling that are duplicates of particles with high weights are perturbed at time t_a-24 hours.
- New hydrodynamic simulations with the mutated levels of the S_{FR} are carried out during the 24 hours prior to the assimilation.
- The likelihood of the mutated particles $p_{mu}(y \mid x)$ is compared to the likelihood of the resampled particles $p_{re}(y \mid x)$.
- The resampled particles are replaced by the mutated particles if the ratio of the two is larger than a value randomly taken from the interval [0, 1].
- The mutation step is repeated twice.
- The iteration with a new tempering coefficient is realized.
- The entire process is repeated until the sum of the tempering coefficients is equal to unity.

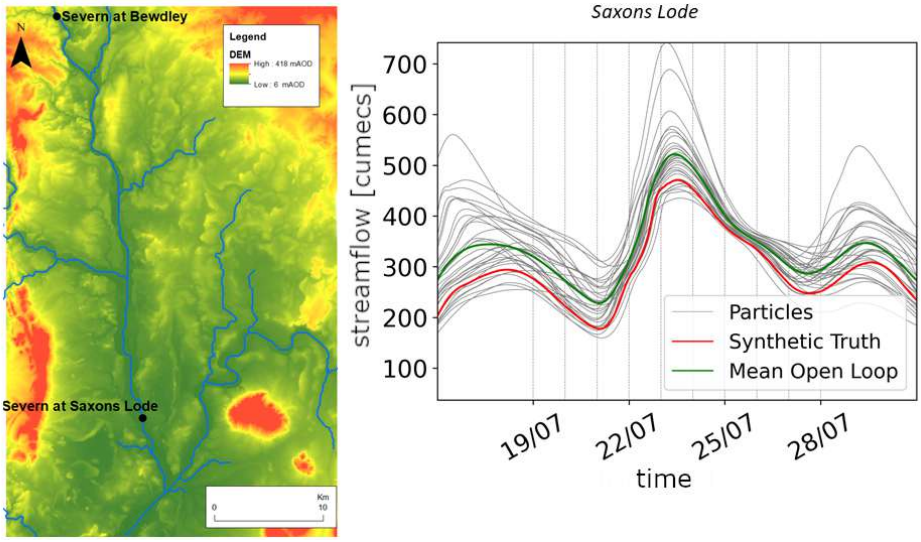


Figure 3.3: Study area of the synthetic experiment (left). Black dots correspond to the points where evaluation of the DA performances is carried out ("Severn at Bewdley" and "Severn at Saxons Lode"). Ensemble time series of discharge in Saxons Lode and assimilation times (right). Gray lines correspond to the Open Loop (OL), the red line corresponds to the synthetic truth, the green line corresponds to the mean of the OL. The dashed lines correspond to the different assimilation time steps performed independently every 24 hours from 19/07 00:00 to 28/07 00:00.

3.3.3. EXPERIMENTAL DESIGN, CASE STUDY, AND PERFORMANCE METRICS

The study area is the lower river Severn located in the United Kingdom (Figure 3.3, on the left). To analyze the filter performances at different assimilation times, SAR images have been synthetically generated (see Chapter 2) every 24 hours from 07/19 00:00 to 07/28 00:00 (Figure 3.3, on the right) and the 10 corresponding independent assimilations are carried out and evaluated.

The flood event has been simulated using the rainfall and temperature (ERA-5 dataset) time series corresponding to the July 2007 event as input data to the flood forecasting system.

Further details concerning the hydrological and hydraulic model set-up as well as the study area of the synthetic experiment, are provided in our previous study (Chapter 2). In this study, the ensemble contains 32 particles. The proposed TPF is characterised by a particle mutation at each iteration. The mutation step could have a key-role, especially when the ensemble is biased with respect to the observations. On the one hand, in the SIS case the weighted mean (also called expectation) is based on the initial particles of the ensemble meaning that if the truth falls outside the ensemble range the expectation cannot reach the synthetic truth. On the other hand, in the TPF case the particles can mutate and move outside the initial ensemble range. This way the expectation can potentially reach the synthetic truth. For evaluating the capability of the TPF to compensate for bias within the ensemble, two different cases are investigated. The difference between the OL and the synthetic truth (O) rainfall time series averaged over the flood

event period (K) represents the mean bias error (MBE, equation 3.11) and it is used to estimate the bias. For a "markedly" biased case MBE is 0.92 $\frac{mm}{h}$ while for a "limited" bias case the MBE is 0.14 $\frac{mm}{h}$, meaning that the error of the markedly biased case is 6.56 times larger than for the other case.

$$MBE = \frac{1}{K}\sum_{k=1}^{K}(OL_k - O_k) \tag{3.11}$$

In the limited case the synthetic truth is most of the time within the ensemble range; in the other case the ensemble is conspicuously biased and the synthetic truth falls outside the ensemble range most of the time. The assimilation steps are performed at the same time for both cases and the same observations are used.
Results are analyzed according to different spatial (global and local) and temporal scales (at the assimilation time and for the subsequent time steps). The filter performances are evaluated in terms of predicted flood extent and water depth maps, as well as local discharge and water levels time series. The performance metrics are assessed by comparing the results of the TPF with those of the OL. Moreover, the TPF is compared with the SIS method applied in our previous study (Chapter 2). The local evaluation of the prediction accuracy of water levels and discharge is performed by comparing the simulated discharge and water level time series with respect to the synthetic truth.
The following performance metrics are used:

- Confusion matrices: a matrix providing the number of false negatives (under-prediction) and false positives (over-prediction), together with correct positives and negatives;
- Contingency maps: maps comparing the simulated flood map with the synthetic truth map;
- Critical success index (CSI): a metric that evaluates the accuracy of the flood map predictions and is defined as the ratio between the number of pixels correctly predicted as flooded over the sum of predicted flooded pixels (correct positives, false positives and false negatives). It ranges from 0, complete disagreement, to 1, perfect match;
- Root mean square error (RMSE): it is given by the square root of the mean of the squares of the deviations of the predicted water levels against the synthetic truth over the hydraulic model domain. It evaluates the prediction errors of a state variable, in our case the water levels.
- 95% Exceedance Ratio (ER_{95}): it measures the reliability of the ensemble prediction quantiles and it is given by the formula: $(N_{exceedence}/T)\cdot 100$, where $N_{exceedence}$ is the number of times during the total simulation T where observations fall outside the 95% predictive bounds. The ideal ensemble should fall outside the 95% predictive bounds only the 5% of the time (Moradkhani et al., 2006).
- Normalized RMSE Ratio (NRR): it is a normalized measure of the ensemble dispersion. It is defined as the ratio of the time-averaged RMSE of the ensemble mean to the time averaged RMSE of the single members of the ensemble over the value

$\sqrt{(N+1)/2N}$ and it should be equal to one. $NRR > 1$ indicates an insufficient spread, while $NRR < 1$ indicates the opposite. (Anderson, 2001; Moradkhani et al., 2005).

3.4. RESULTS AND DISCUSSIONS

3.4.1. TPF-BASED ASSIMILATION PERFORMANCES

FLOOD EXTENT MAP PREDICTIONS

The flood extent maps are evaluated via different performance metrics: the contingency maps, the CSI and the confusion matrix. The contingency map is derived from the comparison between the simulated flood extent map (i.e. expectation) and the validation map which is derived from the synthetic truth simulation in our case. The contingency maps, corresponding to 3 different assimilation time steps (rising limb, peak, falling limb), are shown in Figure 3.4.

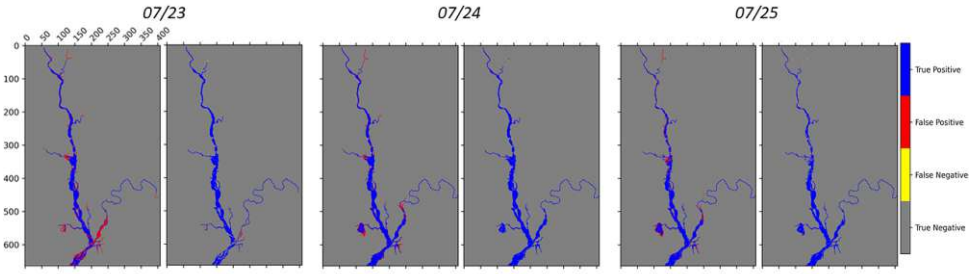


Figure 3.4: Contingency maps of the Open Loop (left) and after the assimilation (right) for three different assimilations at time 07/23 00:00, 07/24 00:00, 07/25 00:00. Red pixels correspond to over-prediction (false positives) errors, yellow pixels to under-prediction (false negatives) errors, pixels correctly classified as not-flooded are in gray and when the contrary occurs pixels are in blue.

Yellow and red pixels correspond to errors of under-prediction (when the model wrongly predicts the pixels as not-flooded) and over-prediction (the opposite case), respectively. In Figure 3.4, the reported images for each assimilation time correspond to the OL (on the left) and the TPF analysis (on the right). Over-prediction represents the most frequent type of error and it is significantly reduced as a result of the TPF-based assimilation.

The decrease of wrongly predicted pixels is quantified in the confusion matrix reported in Table 3.1. In line with Figure 3.4, after any of the three assimilation time steps, the number of over-prediction errors is reduced by 90% or more, while the number of under-predicted pixels increases in the upstream part of the river. However, they represent only 0.3% or less of the total number of flooded pixels.

Time series of CSI are also used to evaluate the TPF performances (Figure 3.5). They allow to evaluate the predicted flood extent maps not only at the assimilation time step (as for the contingency maps and the confusion matrices) but also for subsequent time steps. Time series of CSI provide an assessment of the persistence of the improvements over longer lead times after the assimilation. Figure 3.5 shows the time series of CSI

Method		07/23 00:00		07/24 00:00		07/25 00:00	
		PF	**PN**	**PF**	**PN**	**PF**	**PN**
Open Loop	**TF**	7497	0	9374	0	8390	1
	TN	2441	260974	1356	260182	1219	261302
TPF	**TF**	7475	22	9374	22	8378	13
	TN	204	263211	78	261460	30	262491

Table 3.1: Confusion matrix of the Open Loop and Tempered Particle Filter analysis for three different time steps (07/23 00:00, 07/24 00:00, 07/25 00:00): TF= flooded pixels in the truth map, TN= not-flooded pixels in the truth map, PF= predicted flooded pixels, PN=predicted non-flooded pixels.

before (black line) and after (blue line) the assimilation of SAR images taken during the rising limb (07/23 00:00), at the peak (07/24 00:00) and during the falling limb (07/25 00:00) of the flood event. This figure shows an improvement of the analysis compared to

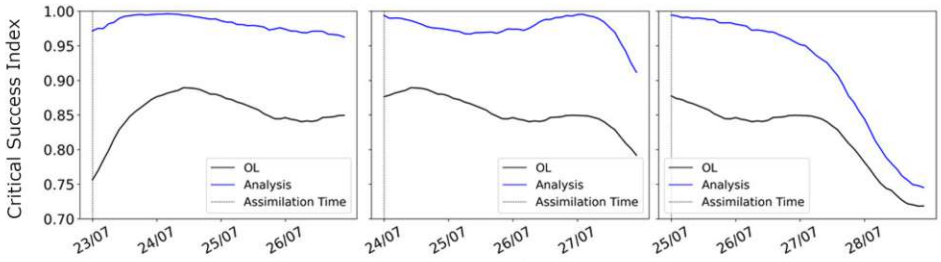


Figure 3.5: Hourly time series of the Critical Success Index of the Open Loop (black line) and Tempered Particle Filter analysis (blue line) due to the assimilation of 3 different images: during the rising limb (07/23 00:00), at the peak (07/24 00:00) and during the falling limb (07/25 00:00).

the OL not only at the assimilation time but also over subsequent time steps: on average, CSI improvements persist for more than 3 days after the TPF application.

WATER LEVEL AND DISCHARGE PREDICTIONS

To further investigate the TPF assimilation performance we evaluate water level and discharge predictions. This evaluation is carried out first at specific points along the river Severn: in Bewdley (the gauge station located at the upstream boundary of the hydraulic model domain), and in Saxons Lode (within the hydraulic domain). In Figures 3.6, the discharge at Bewdley (on the left) and at Saxons Lode (on the right) are plotted. The analysis expectation of discharge (blue line) moves closer to the synthetic truth (red line) at the two stations as a result of the assimilation showing a substantial improvement of the predictions. Here we show the results from the assimilation on July 23^{th} 00:00 as an illustrative example since the other assimilations produce similar effects. In Figure 3.6, it can be observed that the degeneracy is mitigated. At the assimilation time, the analysis

particles are very similar and close to the synthetic truth, but rapidly regain diversity, thereby avoiding degeneracy. After more than 3 days, the particles returns to their initial trajectories (i.e. the OL) mainly because precipitation uncertainty seems to prevail in the forecasts from that moment on.

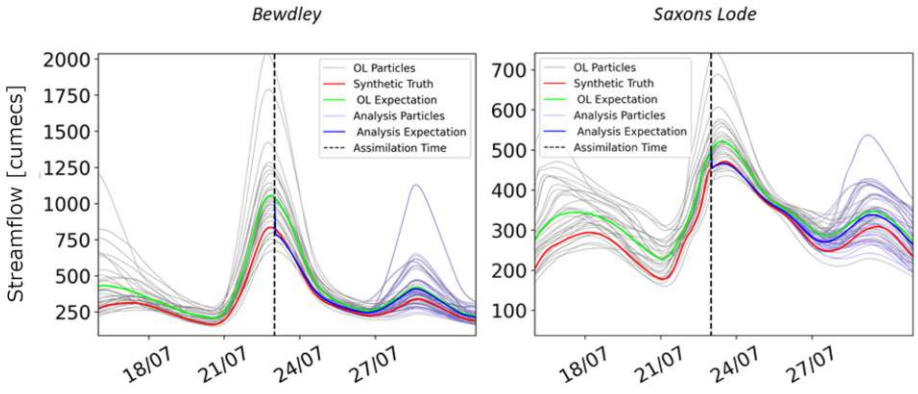


Figure 3.6: Time series of discharge at the peak at Bewdley and at Saxons Lode with the assimilation of an image at 07/23 00:00. The vertical dashed lines indicate the time of the assimilation. The gray lines correspond to the OL particles, the green line to the OL mean, the light blue lines to the analysis particles and the blue line to the analysis expectation. The synthetic truth is represented by a red line.

To generalize the evaluation made for the gauging stations, we evaluate the accuracy of water level predictions globally, using time series of RMSE computed over the entire hydraulic model domain. This index has been calculated at the assimilation time and for subsequent time steps, in order to assess if the assimilation benefits persist in time. In Figure 3.7, the RMSE of the analysis is lower than the OL and this improvement lasts for more than 3 days following the assimilation. The accuracy of the results is higher when assimilation is performed after the flood peak, when rainfall has stopped, and inflow errors are dominating. Flood extents during the falling limb become more sensitive to changes in water depth due to the connectivity between the river channel and its floodplain (Dasgupta et al., 2021). Because of this high sensitivity, during the falling limb, flood extents change faster and weights should be updated more frequently to be consistent with the new hydraulic conditions. This could explain the reason why, as for the CSI plots (Figure 3.5), DA performances start dropping more quickly for the assimilation at the falling limb. The performances of the TPF experiment have been compared to those of the OL for lead time up to 7 days. After one week, we observe that the TPF-CSI is 10% greater than the OL-CSI whereas the TPF-RMSE is 20% lower than the OL-RMSE. These results show that the TPF still outperforms the OL after one week. The standard deviation of the errors has also been computed in order to evaluate the dispersion of the errors (given as the difference between the expectation and the true water levels) and it is shown in Figure 3.8. Results show that the TPF application determines less dispersed and more clustered results around the synthetic truth.

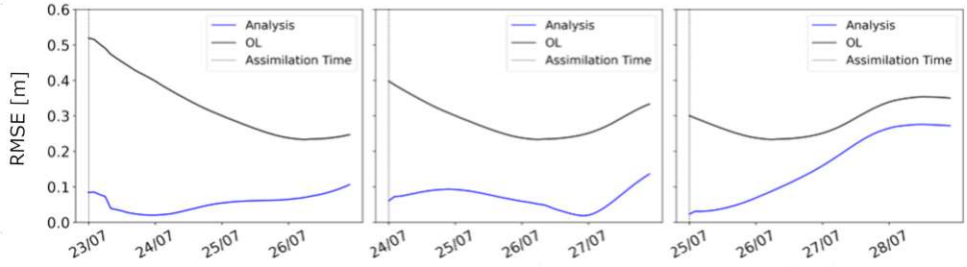


Figure 3.7: Hourly time series of the RMSE of water levels. Black line refers to the OL and blue line to the analysis results after the assimilations of 3 different images (07/23 00:00, 07/24 00:00, and 07/25 00:00).

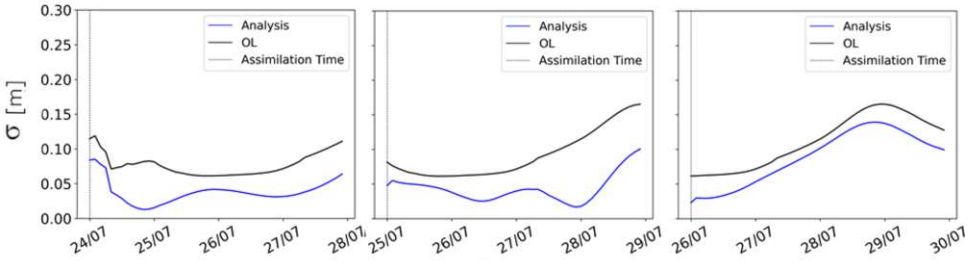


Figure 3.8: Hourly time series of the standard deviation of the errors due to the assimilation of 3 different images: 07/24 00:00, 07/25 00:00, and 07/26 00:00. The standard deviation of the errors as difference between the OL and the true water levels (black line) and as difference between the analysis expectation and the true water levels (blue line).

3.4.2. COMPARISON BETWEEN TPF- AND SIS-BASED ASSIMILATION EXPERIMENTS WITH UNBIASED BACKGROUND

We showed in section 3.4.1 that the TPF improves the predictions of water levels and discharge, as well as flood extent. In this section, the new TPF-based DA framework is compared with the SIS approach previously proposed in Chapter 2 . To do so, we apply the SIS method as proposed Chapter 2 on the same 32 background particles (i.e., OL) and the same synthetically generated flood extent observations. The choice of comparing the TPF with this SIS is related to the fact that other methods reported in Chapter 2 were providing comparable performances, and therefore, SIS has been chosen as a benchmark. In terms of flood extent, the comparison is realized using the hourly time series of the CSI index (Figure 3.9).

In Figure 3.9, the blue line corresponds to the CSI of the forecast obtained from the TPF-based case, the orange line to the one obtained from the SIS-based case and the black line to the one of the OL. The 3 plots correspond respectively to the assimilation on July 23 00:00, July 24 00:00 and July 25 00:00. The CSI values obtained when assimilating an image during the rising limb are systematically higher for the TPF. When the image is assimilated close to the peak and during the falling limb, CSI values of the TPF and

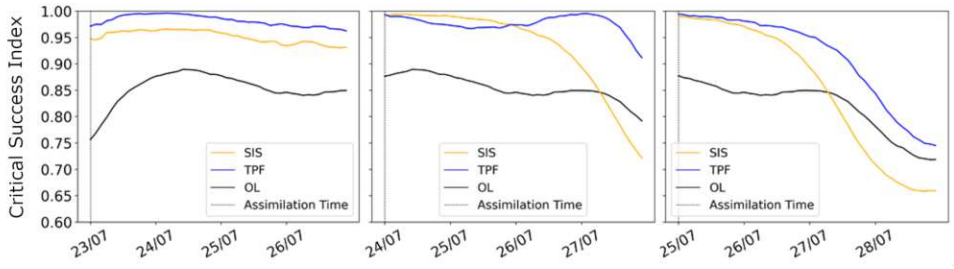


Figure 3.9: Comparison of the hourly time series of the Critical Success Index of the OL (black line), TPF analysis (blue line) and SIS analysis (orange line) due to the assimilation of 3 different images: 07/23 00:00, 07/24 00:00, and 07/25 00:00.

SIS-based assimilation are very similar at the assimilation time and for subsequent time steps. After 2 days, the performance of the SIS becomes substantially worse than that of the TPF. SIS suffers from degeneracy, the number of particles with a significant weight as a result of the assimilation is very limited. These particles produce accurate results at the assimilation time, but are not necessarily efficient after a few hours or days, especially when hydraulic conditions have changed in the meantime.

We have also compared the performances of the SIS and the TPF using time series of RMSE (Figure 3.10). As expected, the RMSE time series exhibit very similar trend to the CSI: the RMSE is lower with the TPF experiment when assimilating an image during the rising limb. For the other two assimilation steps RMSE values are comparable, but performances of the SIS decrease more rapidly, especially after 2 days. Overall, Figures 3.9 and 3.10 clearly show the beneficial effects of the TPF assimilation on the long-term.

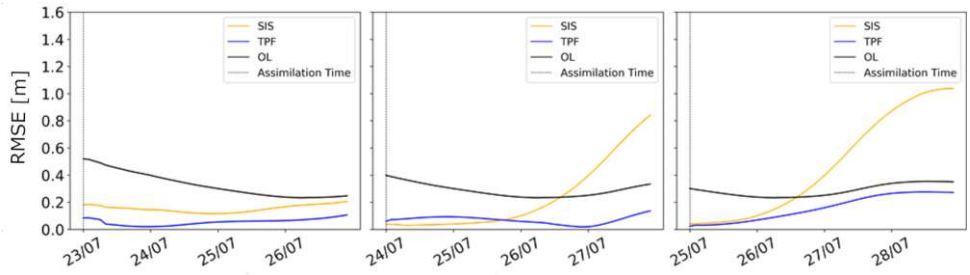


Figure 3.10: Hourly Root Mean Square Error (RMSE) time series. The black line represents the RMSE of the OL, the blue line the TPF-based RMSE and the orange line the SIS-based RMSE. 3 different assimilation cases are plotted: 07/23 00:00, 07/24 00:00, and 07/25 00:00.

Table 3.2 reports the ratios between the analysis-RMSE and the OL-RMSE for each assimilated SAR image and for different lead times. These ratios were calculated at each hour and for all the different assimilation dates. In the table the values at the assimilation time and for lead times of 6 hours, 1 day, 2 days, 3 days, and 4 days are reported. The ratios obtained with the TPF method are shown in the gray cells. The cyan cells contain

the ratios obtained with the SIS experiment. The last row of the table shows the mean of the RMSE ratios over the different assimilation times at given prediction lead times. The lower the RMSE ratio values, the better the performance. Ratios of RMSEs lower than unity indicate that the assimilation improves forecasts. Table 3.2 shows that the TPF-based ratios are most of the time substantially lower than those of the SIS-based ones. For instance, the SIS-based mean ratios for 3 and 4 days of lead times are almost twice that of the TPF-based one. The benefit of the TPF-based assimilation persists for more than 4 days after the assimilation time. Moreover, the TPF-based ratios are always lower than unity, whereas the SIS-based ratios get also values higher than unity.

Table 3.2: Ratios between the analysis and Open Loop RMSE for each assimilation date and for various lead times. Gray cells refer to the TPF-based method, cyan cells to the SIS-based method.

Image date	**Lead time**											
	0		6 hours		1 day		2 days		3 days		4 days	
	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS
07/19	0.25	0.24	0.25	0.24	0.23	0.26	0.20	0.22	0.59	0.57	0.80	0.83
07/20	0.23	0.26	0.22	0.26	0.19	0.22	0.60	0.57	0.83	0.85	0.90	1.08
07/21	0.19	0.22	0.28	0.24	0.62	0.57	0.77	0.85	0.79	1.10	0.76	1.26
07/22	0.27	0.25	0.30	0.29	0.35	0.35	0.31	0.36	0.23	0.39	0.27	0.67
07/23	0.16	0.35	0.15	0.36	0.05	0.36	0.18	0.39	0.27	0.70	0.43	0.84
07/24	0.15	0.09	0.19	0.09	0.31	0.13	0.25	0.42	0.08	1.58	0.41	2.52
07/25	0.08	0.13	0.11	0.16	0.29	0.42	0.63	1.58	0.78	2.57	0.78	2.96
07/26	0.17	0.23	0.17	0.25	0.25	0.20	0.54	0.24	0.63	0.38	0.64	0.72
07/27	0.11	0.18	0.12	0.16	0.26	0.24	0.38	0.41	0.49	0.69	0.56	1.20
07/28	0.15	0.24	0.23	0.29	0.36	0.41	0.54	0.69	0.63	1.26	-	-
Mean	0.17	0.21	0.19	0.22	0.25	0.29	0.39	0.48	0.44	0.85	0.58	1.1

Model performances have also been statistically evaluated using the 95% Exceedance Ratio (ER_{95}) and the Normalized Root Mean Square Error Ratio (NRR). Both metrics have been used to evaluate the water level ensemble at two different gauge stations (Bewdley and Saxons Lode). ER_{95} evaluates the ensemble spread by quantifying the percentage of time the observation falls outside the 95% confidence interval derived from the ensemble. ER_{95} values should be ideally around 5%, meaning that the observation falls outside of the 95% predictive bounds only 5% of the time. NRR also evaluates the spread of the ensemble, ideal values should be around the unity and lower or higher values indicate a too narrow or too wide ensemble, respectively. Table 3.3 reports these statistical performances for the SIS and TPF experiments. While TPF- and SIS-NRR are both close to the unity for the different assimilation time steps, ER_{95} varies with the different assimilation time steps. In particular, we found that on average, over the different assimilations, the value of ER_{95} for the TPF is around 7% in Bewdley and 9% in Saxons Lode, which are values close to the target values (5%). Moreover, if we compare these values with those of the SIS that are around 25%, it is clear that TPF substantially outperforms SIS. This highlights a marked degeneracy in the SIS, that is substantially reduced by TPF.

Table 3.3: NRR and ER_{95} of water levels at the different assimilation times and at two different gauge stations (Bewdley and Saxons Lode). SIS statistical performance measures are shown in the cyan column and TPF performance measures in the gray column. The average of the measures over the different assimilation time is also reported in the last row of the table.

Image date	Bewdley				Saxons Lode			
	ER_{95} [%]		NRR		ER_{95} [%]		NRR	
	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS
07/19	4.97	26.98	0.88	0.93	5.77	26.58	0.94	0.95
07/20	2.42	36.64	0.88	0.95	2.82	36.24	0.88	0.95
07/21	12.21	33.42	0.9	0.95	15.44	33.02	0.91	0.94
07/22	7.5	30.20	0.85	0.95	10.1	29.80	0.86	0.95
07/23	6.85	26.98	0.88	0.94	8.19	26.58	0.89	0.95
07/24	7.92	23.76	0.87	0.95	8.99	23.36	0.88	0.94
07/25	7.11	20.54	0.88	0.95	9.53	20.13	0.89	0.94
07/26	3.62	17.32	0.88	0.93	5.37	16.91	0.89	0.93
07/27	9.8	14.09	0.72	0.93	13.15	13.69	0.77	0.93
07/28	7.11	10.87	0.92	0.98	10.34	10.47	0.93	0.98
Mean	6.94	24.08	0.87	0.95	8.97	23.68	0.88	0.95

3.4.3. COMPARISON BETWEEN TPF- AND SIS-BASED ASSIMILATION EXPERIMENTS WITH BIASED BACKGROUND

In this last experiment, we use the same set-up as in the previous experiment but with the exception of a modified OL. We have introduced a perturbation error to the ERA-5 rainfall time series so that the bias in the ensemble is 6.56 times larger than in the previous case. The ensemble has significant bias and the synthetic truth is most of the time located outside of the ensemble range as can be see in Figure 3.11. For the evaluation of the results, the same performance indices and the same plots are used. The ratios between the analysis-RMSE and the OL-RMSE for each assimilated SAR image and for different lead times are reported in the Table 3.4. At the assimilation time and for more than one day after that, the TPF-based assimilation is capable of substantially reducing the forecast bias. The SIS is less efficient in that respect, as RMSE ratios are larger for the SIS-based assimilation. For longer lead times, the error in water levels increases due to the bias in the rainfall ensemble and the RMSE ratios of the TPF-based and the SIS-based assimilation become similar. This is clearly visible in Figure 3.12 that shows the RMSE time series on July 23^{th}, 24^{th}, and 25^{th} at 00:00. When the bias is limited and the synthetic truth falls inside the ensemble range most of the time, as in the previous case (Figure 3.7), the forecast improvement lasts for longer lead times. However, when the ensemble is markedly biased (Figure 3.12), the TPF improves the results at the assimilation time but the level of improvement degrades more quickly compared to the limited biased case.

At the assimilation time, the TPF always improves the accuracy of the results of the flood forecasts (in terms of flood extent, water levels, discharge) with respect to the OL

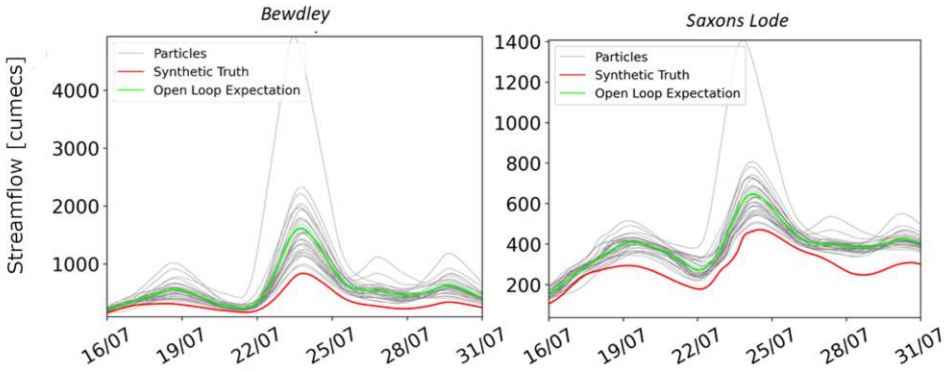


Figure 3.11: Discharge time series ensemble at Bewdley (on the left) and at Saxons Lode (on the right). The OL particles are represented with gray lines, the synthetic truth is represented by the red line. The OL expectation is in green. In this case, the ensemble is markedly biased; the synthetic truth falls outside the ensemble range most of the time.

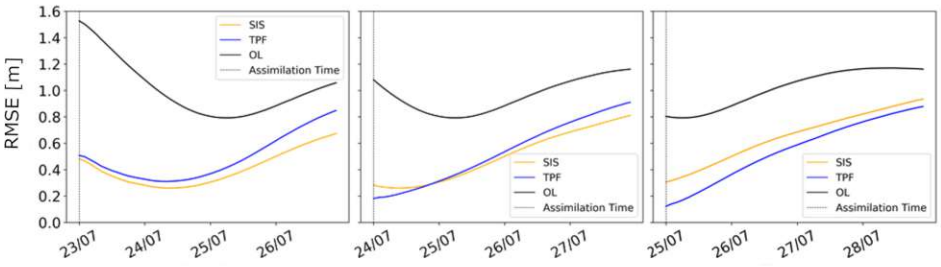


Figure 3.12: Hourly RMSE time series for a markedly biased ensemble case. The black line represents the RMSE of the OL, the blue line the RMSE after the TPF application and the orange line the RMSE after the SIS application. Assimilation at 07/23 00:00, 07/24 00:00, and 07/25 00:00 are plotted.

and it is comparable to the SIS performances. An important aspect that emerges from the results is the persistence of the assimilation benefits. They remain significant even 3 days after the TPF assimilation when compared to the SIS performances; nonetheless, performances start degrading with the onset of rainfall over the headwater catchment and rainfall uncertainty prevails in the forecast uncertainty. We argue that the marked improvement in the forecast skill of the TPF, compared to the SIS, is due to the update of the initial conditions of the hydrological model including S_{FR} 24 h prior to the assimilation time. In the TPF, better initial conditions of the model forecast are defined at each assimilation time via the different iteration and mutation steps, whereas the SIS only defines the relative importance of each particle, without carrying out any better definition of the initial conditions of the model. The runoff that is used as upstream boundaries of the hydraulic model is a function of the storage S_{FR} of the hydrological model. Updating the S_{FR}, and consequently the fast run-off, represents an effective way to increase the

Table 3.4: Ratio between the analysis and Open Loop of the the RMSE for each assimilation date and for various lead times for a markedly biased case. Gray cells refer to the TPF-based method, cyan cells to the SIS-based method.

Image date	Lead time											
	0		6 hours		1 day		2 days		3 days		4 days	
	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS	TPF	SIS
07/19	0.19	0.42	0.13	0.42	0.10	0.44	0.26	0.52	0.82	0.53	0.94	0.58
07/20	0.29	0.44	0.25	0.46	0.21	0.52	0.72	0.53	0.88	0.59	0.91	0.67
07/21	0.47	0.52	0.49	0.53	0.54	0.53	0.70	0.59	0.82	0.68	0.71	0.82
07/22	0.47	0.53	0.49	0.52	0.53	0.59	0.70	0.68	0.82	0.83	0.88	0.95
07/23	0.32	0.31	0.31	0.29	0.30	0.26	0.47	0.38	0.71	0.57	0.81	0.64
07/24	0.17	0.26	0.20	0.27	0.39	0.38	0.61	0.57	0.71	0.64	0.78	0.70
07/25	0.15	0.38	0.21	0.43	0.41	0.57	0.55	0.64	0.65	0.71	0.76	0.80
07/26	0.16	0.57	0.18	0.59	0.28	0.64	0.44	0.71	0.61	0.81	0.68	0.87
07/27	0.24	0.52	0.16	0.55	0.34	0.70	0.68	0.96	0.83	1.05	0.78	1.04
07/28	0.34	0.70	0.36	0.77	0.51	0.96	0.65	1.05	0.58	1.04	-	-
Mean	0.26	0.46	0.24	0.48	0.34	0.56	0.55	0.66	0.72	0.74	0.81	0.79

long-lasting effects of DA since runoff has the highest uncertainty deriving from poorly known rainfall as already pointed out by Matgen et al., 2010. This aspect, together with the mitigation of degeneracy, as hypothesized by Dasgupta et al., 2021, could explain the longer-term persistence of DA benefits via the TPF.

After the TPF application, particles move towards the synthetic truth also in the case the truth falls outside the predictive bounds of the OL ensemble. Despite the improvements due to the TPF, performances are not as good as in the previous case. As a consequence, results obtained using the TPF are sometimes similar to those obtained using the SIS, or even slightly less satisfying when rainfall uncertainty dominates the system. The improvements resulting from the update of the initial conditions are vanished after a few days because of the bias in the ensemble and the model moves back to the OL state. The update of the state level of the reservoir has a time–limited benefit. It is a state variable highly influenced by the inputs, and thus by the rainfall. In our experiment, the rainfall ensemble is obtained by perturbing the deterministic ERA-5 product using a multiplicative noise. Therefore, when there is low intensity rainfall simulated in ERA5 the uncertainty is very limited. Moreover, as the rainfall ensemble is not updated, the ensemble analysis goes back to the OL trajectory after a while. This return of the analysis back to the OL is even more rapid when higher rainfall intensity is imposed to the model: the influence of the initial conditions is rapidly overruled by the forcing uncertainty. To increase the time window of the assimilation benefits, the update of hydrological model state variable could be completed by a forcing update or by a parameter update, as in Cooper et al., 2018b where channel friction is updated together with a state variable, but with the consequent risk of multiple acceptable solutions of the system according to the equifinality concept (Beven et al., 2001).

3.5. CONCLUSIONS

In this paper, we have proposed a new approach based on a Tempered Particle Filter (TPF) to assimilate flood extent maps into a flood forecasting system. The objective of this new data assimilation framework is to mitigate degeneracy and sample impoverishment, well known issues in particle filtering. In the proposed TPF method, the number of tuning parameters is small with respect to methods such as PF-MCMC (Andrieu et al., 2010; Moradkhani et al., 2012) thus rendering the TPF easily transferable to other situations. We also argue that this makes the approach potentially robust. Moreover, the TPF does not need cross-over steps or assumptions on the prior pdf used in the MH acceptance ratio as in the Evolutionary PF-MCMC (Abbaszadeh et al., 2018) or in the HEAVEN (Abbaszadeh et al., 2019).

We have evaluated the performances of the filter in two different cases: with a limited forecast bias and with a more important forecast bias. The TPF has been compared against the standard Particle Filter, namely the Sequential Importance Sampling (SIS) as used in previous studies such as Hostache et al., 2018 and Chapter 2 of this thesis. The following key conclusions are drawn from our experiments:

1. At the time of the assimilation, forecasts are very accurate locally: the forecast overlaps the synthetic truth for all the different assimilation cases and for both analysed locations. Results are very satisfying evaluating the performances over the entire hydraulic domain: RMSE and CSI improve systematically as a result of the assimilation. On average, RMSE values decrease by 80% whereas CSI values increase by 30% as a result of the assimilation;

2. Results are also satisfying across time: the CSI and RMSE are improved up to 3 days after the assimilation;

3. Performances are improved compared to the OL and the SIS filter. The benefits of the newly introduced TPF-based assimilation are longer persisting when compared to the effects obtained with assimilation techniques used in the previous studies;

4. The new assimilation framework significantly outperforms the SIS. SIS performance indices are generally comparable to the TPF ones at the assimilation time, but they tend to drop more rapidly, in general 2 days after the assimilation. For example, TPF-based RMSE are 20% lower compared to the SIS-based ones, 2 days after the assimilation;

5. When the ensemble is markedly biased results are significantly improved by the TPF at the assimilation times and for few days after. Afterwards, TPF and SIS based results are similar because the model state update cannot compensate for a too large bias in the precipitation ensemble.

The proposed data assimilation framework based on a TPF holds promise for improving prediction accuracy for longer lead times. In this study, we have shown a synthetic experiment where rainfall and SAR observations are the only sources of uncertainty. In a future study, it will be interesting to apply and evaluate this enhanced approach on a real test case in a weakly controlled environment.

4

APPLICATION OF THE TEMPERED PARTICLE FILTER TO A REAL CASE STUDY

The present chapter corresponds to the following scientific publication in its original form: Concetta Di Mauro, Renaud Hostache, Patrick Matgen, Ramona Pelich, Marco Chini, Peter Jan van Leeuwen, Nancy K. Nichols, and Günter Blösch. "Application of the tempered particle filter to a real case study." Submitted to "Advances in Water Resources", 2022.

4.1. ABSTRACT

Predicting and mitigating floods is fundamental to prevent or reduce the socio-economic consequences of them. Data assimilation (DA) can help in improving the flood forecast accuracy by reducing the uncertainties stemming from the observations and the model. In a previous study a DA technique, namely Tempered Particle filter (TPF), has been tested in a synthetic experiment with controlled uncertainty. This study aims to evaluate the TPF in a real world experiment. 3 different flood events of the River Severn (UK) are used as a test case. Results show that the errors are decreased by 60%-70% at the assimilation time at some gauging stations. Improvements in water levels accuracy persist also for several hours to days after the assimilation. In addition, the flood extent map is improved, unless the flood extent map of the Open Loop is already good. The main actual limitation of the TPF is linked to the poor predictions of the boundary inflows. This could be due to the fact that other possible sources of uncertainty, other than the rainfall and the satellite data, need to be taken into account.

4.2. INTRODUCTION

Early warning and prediction of flood events are crucial in river basin management strategies. In this context hydrodynamic models are fundamental to estimate water level and flood extent that represent necessary information to prevent and/or mitigate devastating consequences of floods. However flood models are subject to several concurring sources of uncertainties (e.g. boundary conditions, model parameters, observations, model structure) that often lead to inaccurate predictions. DA can help in reducing the predictive uncertainty via regular and frequent updates of model states and/or parameters based on observations (Moradkhani et al., 2005).

Ideally, DA in flood models should exploit observed data from distributed gauge networks measuring for example water level and/or discharge. However this type of information is not always available either because of a lack of gauging station data or because of measurements made during high flood events being unreliable (Grimaldi et al., 2016). Satellite data represent valuable sources of information, especially in data-scarce regions. In particular, Synthetic Aperture Radar (SAR) sensors have the capacity to capture flooded area, to penetrate clouds, and to produce observations during day and night due to their active microwave sensing technique.

The Ensemble Kalman Filter (EnKF) and its variants are currently the most widely used DA techniques in flood prediction (e.g., Revilla-Romero et al., 2016; Cooper et al., 2018b; Annis et al., 2021). However, EnKFs make strong hypotheses regarding model and observation error distributions (i.e., Gaussian errors), linear analysis and do not necessarily conserve the water balance (Matgen et al., 2010; Noh et al., 2011; Plaza et al., 2012). DeChant et al., 2012 show that EnKF limitations can potentially lead to filter divergence and can produce overconfident results. Given these limitations, Particle Filters (PFs) have gained interest in the research community especially because they relax the assumption of normally distributed errors. PFs represent the prior and the posterior Probability Density Functions (pdfs) by a set of model states called particles. While representing a valuable alternative in non-linear and non-Gaussian systems, PFs have some important limitations such as degeneracy. The degeneracy occurs when the number of particles with significant weight as a result of the assimilation is very limited, the ensemble could collapse and may lead to wrong approximation of the posterior distribution. Gordon et al., 1993 proposed a resampling of the particles to mitigate the degeneracy but this was generating a loss of diversity among particles, a problem known as sample impoverishment. Various additional techniques have been proposed in the literature to mitigate these PF-related issues. Van Leeuwen, 2009 provided an extensive review of recent developments in PFs. For instance, some recent techniques combine a PF with a Markov Chain Monte Carlo method (MCMC) (Andrieu et al., 2010; Moradkhani et al., 2012), with Metaheuristic techniques derived from Computer Sciences such as Genetic Algorithms (GA) (Kwok et al., 2005), Particle Swarm Optimization (Wang et al., 2006; Li et al., 2005), Immune Genetic Algorithms (IGA) (Han et al., 2011). From the combined use of GA and PFs, Evolutionary Particle Filters (EPFs) have been developed. EPFs are inspired by Darwin's evolution theory: particles with high weights are more likely to be selected. Degeneracy and sample impoverishment are tackled with crossover and mutation operators. This approach requires a tuning of parameters for cross-over and mutations. In a recent study, Abbaszadeh et al., 2018 evolved the evolutionary PF based

on the combination of GA with MCMC (EPFM). In Abbaszadeh et al., 2018 study, EPFM is applied to hydrological problems without any tuning of the parameters for steps like cross over or mutations. In Chapter 3, it has been introduced a DA framework based on a Tempered Particle Filter (TPF) where a smooth transition trough consecutive tempering steps is performed. The proposed methodology appears to be more general and transferable given that no parameters need to be tuned and physically realistic states are obtained by design. Differently from other studies on DA applied to hydrological problems, in the studies introduced in Chapters 2 and 3, the SAR-derived probabilistic flood maps (PFMs) are assimilated into a hydraulic model. The maps attribute to each pixel its probability to be flooded given the recorded SAR backscatter value (Giustarini et al., 2015). The proposed approach showed promising results in a synthetic twin experiment: estimates of water levels and discharge were improved when compared to the Open Loop (i.e., when no observation is assimilated) and to a standard state-of-the-art Sequential Importance Sampling (SIS) particle filter. Moreover the predictions were improved for prediction lead times up to 3 days following the time step of the actual assimilation. However, this method has been exclusively applied in a synthetic experiment where the multiple sources of uncertainty affecting the prediction system were fully controlled by the users (i.e. rainfall and SAR observations were assumed to be the only sources of known uncertainty).

In this study, we would like to investigate the limitations and the advantages of TPF-based approach in a real world experiment and we would like to compare it with the synthetic experiment previously applied in Chapter 3 . We would like to evaluate if, in a real world experiment where sources of uncertainty are not tightly controlled, the hypothesis of rainfall and SAR observations as the only sources of uncertainty is still valid when the TPF is applied. For the sake of consistency with the previously published synthetic experiment, in this real case experiment, rainfall and SAR observations are still considered as the dominating sources of uncertainty. We argue that in operational flood forecasting the amount of precipitation tends to be a dominant source of uncertainty (Koussis et al., 2003; Pappenberger et al., 2005) but we acknowledge that, in a real-world experiment, uncertainties arise from different sources and we want to investigate what are the implications of not considering these uncertainties. The PFMs derived from SAR data are assimilated into a flood forecasting chain composed of a cascade of coupled hydrological and hydraulic models. Three flood events of markedly different magnitude which were observed by different satellites are considered as test cases in order to get more insights on the potential and limitations of the newly proposed TPF.

4.3. MATERIAL AND METHODS

4.3.1. THE TEMPERED PARTICLE FILTER

The TPF is a variant of Particle Filters which use a set of particles to approximate a probability distribution function. Likewise many other Data Assimilation (DA) methods, TPF is based on the Bayes' formula which states that the probability of an event given an observation y at time k is given by Equation 1.1. In this study, the information derived from the SAR observations y and the forecasts of the hydraulic model x at time k can be combined to reduce the uncertainties of both. PFs do not make any assumptions on the

prior probability distribution (pdf) that is drawn from an ensemble of model states of size N (called particles). In this study, the number of particles is $N = 32$. The particles are generated from the perturbation of the rainfall time series, assuming that uncertainty only stems from rainfall estimates and SAR-based observations. The prior pdf,at time k, is defined according to the Equation 1.3. where δ is the Dirac delta function. Replacing prior probability in Equation 1.1, we obtain the Equation 1.4. where W_n is given by Equation 2.9. W_n represents the particle weight and $p(y^k \mid x^k)$ the likelihood. In our experiment, we sequentially assimilate probabilistic flood maps (PFMs) derived from SAR observations into a flood forecasting model. SAR-based observations of flood extent are resampled in order to have a resolution that is equal to the one of the mesh of the hydraulic model (i.e. 75 m). Global weights W_n, at time k are calculated according to the Equation 2.9 where L is the total number of pixels of the flood domain assuming that observational errors are spatially independent. The global weight W_n assigned to each particle is calculated with the multiplication of the likelihoods w defined for each pixel i of the hydraulic domain according to the Equation 2.8 by Hostache et al., 2018. The water depth maps derived from the hydraulic model simulations of the 32 particles are converted into binary maps. If the water depth is above the threshold of 10 cm the pixel is considered as flooded and a value of *1* is assigned to $M_{i,n}$. In the opposite case, the pixel is not considered as flooded and $M_{i,n}$ is equal to *0*. $p_i(F \mid \sigma_0)$ represents the probability of a pixel being flooded (F) according to the SAR observations (backscatter value σ_0), while $1 - p_i(F \mid \sigma_0)$ represents the probability of it not being flooded.
Degeneracy consists in only a few particles gaining significant weight with respect to others and it is a common issue in PFs. To mitigate degeneracy issues, a tempering coefficient is used in the TPF to inflate the likelihood, thereby increasing the number of particles with significant weight. The TPF relies on the factorisation of the likelihood of Equation 1.1 using small tempering factors γ trough consecutive iterations s enabling the inflation of the likelihood in order to reduce degeneracy. Therefore, the ratio $\frac{p(y|x)}{p(y)}$ of the equation 1.1, at time k, becomes the Equation 3.2

where $0 < \gamma_s < 1$ for each iteration s. This process is consecutively repeated so that the $\sum_{s=1}^{S} \gamma_s = 1$ in order to respect the Bayes' formula. Each γ_s is chosen such that the so-called inefficiency (*InEff*, Equation 3.7) is smaller than a threshold value r*, where r* is chosen of the order 2/N.

After each iteration s, particles are resampled (and duplicated) according to their weights meaning that particles with higher weights are replicated several times. After resampling, particles are equally weighted. At this stage, the ensemble of particles contains duplicates and a mutation is performed, in order to regain diversity in the ensemble. In this DA framework, we decided to perturb the fast run-off reservoir level (S_{FR}) of each particle, which is a state variable of the hydrological model, 24 hours prior to the assimilation and re-run the model cascade for the 24 h preceding the assimilation time. Each resampled particles is replaced by the related mutated particle if the ratio of the mutated and resampled particle likelihoods is larger than a value randomly taken from the interval $[0, 1]$. A more detailed description of the the methodology is proposed in Chapter 3
.

4.3.2. FLOOD FORECASTING SYSTEM

The model cascade is composed of a hydrological model coupled to a hydraulic model. The hydrological model (SUPERFLEX) is used to estimate the run-off within the 3 main catchments with the outlets in Bewdley, Evesham and Knightsford, respectively (Figure 4.1). The runoff is then used as boundary conditions of to the hydraulic model (LISFLOOD-FP) to estimate the flood extent, the streamflow and the water level in the floodplain located downstream.

We use a lumped version of SUPERFLEX (Fenicia et al., 2007). The structure is composed of three reservoirs (Figure 4.2): an unsaturated soil reservoir with a storage S_{UR} representing the root zone, a fast reservoir with storage S_{FR} representing the fast responding components (e.g., the riparian zone and preferential flow paths), and a slow reservoir with storage S_{SR} representing slow responding components (e.g., deep groundwater). A triangular lag function is used ahead of the S_{FR} and S_{SR} reservoirs to enable a delayed hydrological response of the basin. Precipitation P infiltrates in the soil R_U whereas the amount in excess R_F directly contributes to the S_{FR} reservoir (Equations 4.1 and 4.2).

$$R_F = C_r \cdot R \quad \text{and} \quad R_U = (1 - C_r) \cdot R \tag{4.1}$$

$$C_r = \frac{1}{1 + exp(\frac{\frac{S_U R}{S_c} + 0.5}{\beta})} \tag{4.2}$$

The UR reservoir linearly percolates into the SR reservoir according to Eq. 4.3:

$$P_S = P_{max} \cdot \frac{S_{UR}}{S_{max}} \tag{4.3}$$

The potential evapotranspiration is converted into actual evapotranspiration using Eq. 4.4:

$$T_P = Ta \cdot min(1, \frac{S_{UR}}{S_{max}} \frac{1}{L_P}) \tag{4.4}$$

where L_P is the fraction of S_{max} below which T_P is constrained by S_{UR}.
Q_{SR} is given by Equation 4.5:

$$Q_{SR} = S_{SR} / K_{SR} \tag{4.5}$$

Q_{FR} is given by Equation 4.6:

$$Q_{FR} = S_{FR} / K_{FR} \tag{4.6}$$

The 8 parameters of the SUPERFLEX model are listed in Table 4.1.

The inputs to the hydrological model are downloaded from the ERA5 dataset (Hersbach et al., 2019). In this study, rainfall and 2 m air temperature at a spatial resolution of 25 km and a temporal resolution of 1 hour have been used. The temperature is transformed into potential evapotranspiration using the Hamon formula (Hamon, 1963). The parameters of the hydrological model were calibrated for each catchment using in situ measured river discharge over the period of 2013-2019 years. The results were validated over the period 2019-2020. 5000 different parameter sets have been randomly generated

Table 4.1: SUPERFLEX parameters.

Parameter	Unit	Definition
S_{max}	mm	maximum S_{UR} storage
L_P	-	limit for potential transpiration
β	-	runoff shape parameter
P_{max}	-	limit for percolation
lag_{FR}	-	maximum time of triangular lag function FR
lag_{SR}	-	maximum time of triangular lag function SR
K_{FR}	-	FR time scale
K_{SR}	-	SR time scale

Table 4.2: Nash-Suthcliffe and Kling-Gupta efficiencies for the different catchments within the calibration and validation period.

Catchment outlet	Calibration		Validation	
	KGE	NSE	KGE	NSE
Knightsford	0.73	0.65	0.55	0.60
Evesham	0.79	0.76	0.86	0.78
Bewdley	0.83	0.77	0.7	0.78

from uniform distributions with predefined interval of plausible values for each parameter to carry out the calibration. These resulted in 5000 simulations of the hydrological model. The selected parameter set is the one corresponding to the highest values of the KGE with respect to hourly discharge measurements.
For the calibration and the validation, the simulations were compared to the gauge data at the outlet of each of the 3 catchments. Gauge data have been provided by the Environment Agency of England and Wales (`https://www.gov.uk/government/organisations/environment-agency`). Performances are reported in Table 4.2.

The hydraulic model used in this study is the LISFLOOD-FP model (Bates et al., 2000; Neal et al., 2012). The model domain footprint is shown in Figure 4.1. LISFLOOD-FP simulates flood extent, water level and streamflow along the river and in the floodplain. A sub-grid is used for the channel flow. When the storage capacity of the river is exceeded, the water spills into the floodplain and a 2D shallow water model neglecting the convective acceleration (Almeida et al., 2013; Bates et al., 2010) is used for the floodplain flow simulation. The upstream boundary conditions in Bewdley, Evesham and Knightsford are defined as the runoff (Q_t) simulated by the hydrological models. Downstream at Haw Bridge, a "free" downstream boundary condition is applied with an average slope of 0.00007. The riverbed geometry and the roughness coefficient were calibrated in a previous study (Wood et al., 2016). The model computational grid has a spatial resolution of 75 m.

4.3.3. EVALUATION OF THE DA FRAMEWORK

In our experiment, the rainfall ensemble is obtained by perturbing the deterministic ERA-5 product with a multiplicative random noise from a log-normal error distribution.

The rainfall was considered spatially uniform over the catchment. The reliability of the rainfall ensemble is verified with the statistical metrics proposed by De Lannoy et al., 2006: VM_1 and VM_2 and described in section 2.3.3.

Several metrics are used to evaluate the performance of the DA during the three flood events.

Observations of water level and discharge were available at the different gauging stations shown with black dots in Figure 4.1. To estimate the effect of the assimilation on the forecast skill, the model expectation of water levels and discharge, before (background or OL, when particles are equally weighted) and after (analysis, when the particles are weighted based on the information contained in the observations) the assimilation, is compared against the observations. Discharge simulation accuracy was evaluated at the boundaries of the hydraulic model, i.e. Bewdley, Knightsford, and Evesham. The background and the analysis of water levels are also plotted to evaluate the quality of water levels forecasts at the following locations: Saxons Lode, Mythe Bridge, and Haw Bridge.

In addition, errors were estimated, before and after the assimilation, as the deviation of the predicted water levels and discharge against the observations and plotted over time for each gauging station located further downstream (Saxons Lode, Mythe Bridge, and Haw Bridge).

In order to evaluate the accuracy of the flood extent map before and after the assimilation, simulated flood extent maps have been compared against the flood extent map derived from optical images. This evaluation has been performed using the CSI (Equation 2.14).

In Equation 2.14, A represents the Correct water (hits), B the Over-prediction (false alarms) and C the Under-prediction (misses). The Correct no-water (correct rejections, D) are not taken into account because the number of non-flooded pixels in a SAR image is comparatively large and considering them would lead to extremely optimistic results. Contingency maps are plotted so that the location of the different predicted pixels is explored (A, B, C, and D).

4.3.4. STUDY AREA AND AVAILABLE DATA

The study focuses on the area of the lower Severn in the South-West of the United Kingdom covering a 30.6 x 49.8 km (1524 km^2) domain.

The PFMs are derived from SAR observations attributing to each pixel a probability of being flooded knowing the backscatter value. The histogram of the backscatter values is approximated with two overlapping Gaussian probability distribution functions associated with flooded and non-flooded pixels.

In this study, 3 flooding events are considered. During the first event in July 2007, two Envisat images with 150 m spatial resolution were acquired on 23/07/2007 at 10:30 and at 22:00. One TerraSAR-X image with 3 m spatial resolution, was obtained on 25/07/2007 at 06:30. For the second event that occurred in 2012, 6 COSMO-SkyMed images with a spatial resolution of 5 m are available. The satellite images have been acquired at the following times: 27/11/2012 at 19:20, 28/11/2012 at 18:00, 29/11/2012 at 18:20, 30/11/2012 at 19:30, 1/12/2012 at 5:40, 2/12/2012 at 6:00 and 4/12/2012 at 18:15. The third event considered in this analysis occurred in February 2020. In this case 5 Sentinel-1 images

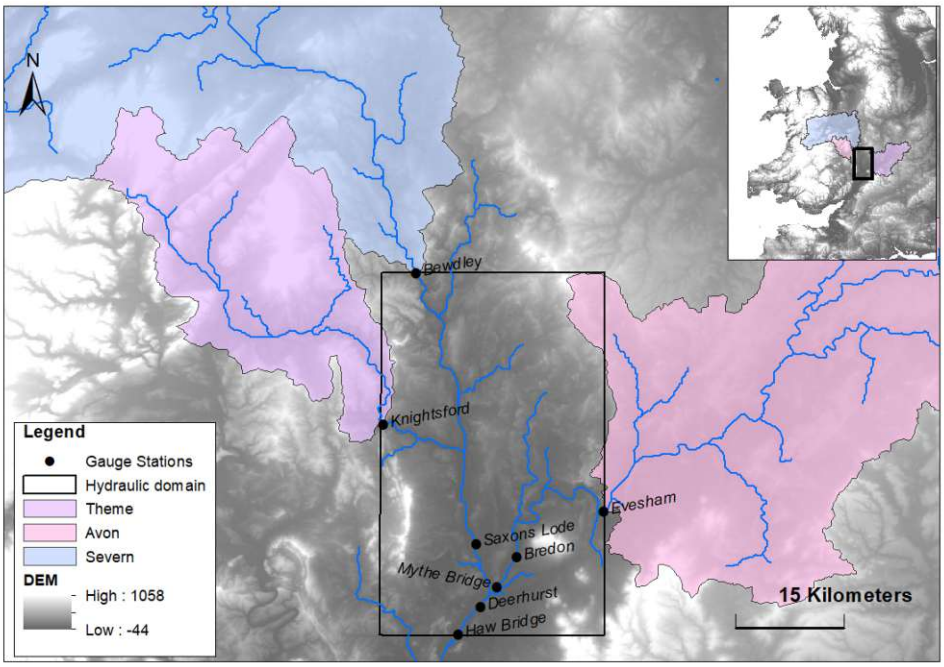


Figure 4.1: Study area: the downstream part of the river Severn (hydraulic domain borders are represented with a black frame). The 3 catchments are represented as colored polygons; their outlets are located at the gauging stations in Bewdley, Evesham and Knightsford. The gauging stations are represented as black dots.

with a spatial resolution of 20 m were acquired at the following times: 17/2/2022 at 6:20, 20/2/2022 at 18:00, 26/2/2022 at 18:00, 29/2/2022 at 6:20, 3/3/2020 at 18:00, and 6/3/2020 at 18:00.

For an independent evaluation of the flood model maps, an aerial photograph of 3 m spatial resolution on 24^{th} July 2007 was used to manually extract binary flood maps. Additionally, a Sentinel-2 image on 29^{th} of February at 11:20 was used to generatea flood binary map busing the "AUTOWADE On Demand S2" algorithm (Pulvirenti et al., 2020).

4.4. RESULTS

Table 4.3 reports the values of the verification measures of the ensemble precipitation for the 3 catchments and for the 3 flood events. The values of VM_1 and VM_2 are around the unity meaning that the ensemble spread is correct and that the observed precipitation is statistically indistinguishable from a member of the ensemble). The ensemble of discharge simulations, generated using the ensemble of precipitation, are mostly satisfactory. In Figure 4.3 the in situ measured data are well encompassed by the background ensemble, for instance the observations in Bewdley (red line falls within the ensemble for each one of the 3 flood events. For the event 2007, the ensemble has a peak that does not correspond to the observations whose discharge are constant during the entire

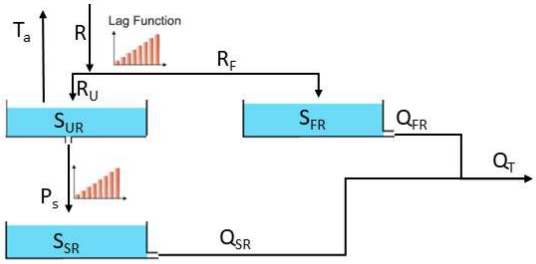


Figure 4.2: SUPERFLEX scheme. P=precipitation, Ta= actual transpiration, S_{UR}= storage unsaturated reservoir, P_S= percolation, S_{SR}=Storage Slow Run-Off, S_{FR}= Storage Fast Run-Off, R_u= rainfall infiltrating into the soil, R_F= rainfall excess, Q_{FR}=Fast Run-Off, Q_{SR}= Slow Run-Off, Q_T=Total run-off

period. This is probably due to the fact that the 2007 event was a summer flood due to local rainfall falling mostly over the floodplain. However, for some experiments there are observations close to the ensemble bounds and even sometimes outside the ensemble for the gauging stations located further downstream. Indeed, the observed water levels at Haw Bridge, Mythe Bridge, Saxons Lode or Deerhurst are outside or on the edge of the ensemble (Figure 4.4). This behaviour seems to be systematic and therefore could be due to a possible bias caused, for example, by an incorrect river bed geometry or lateral inflows along the river that the model does not account for.

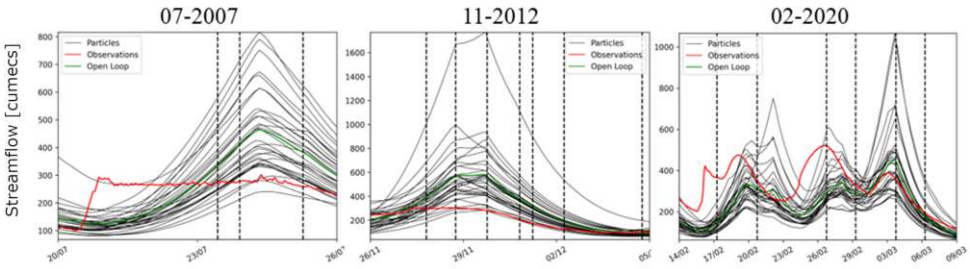


Figure 4.3: Background ensemble of discharge at Bewdley gauge station for the 3 flood events. Black lines represent the particles, red line the observed discharge and green line the Open Loop. Vertical dashed lines correspond to the assimilation time steps. The dates reported on the x-axis correspond to the time 00:00.

4.4.1. JULY 2007

Figure 4.4 shows the water levels at 3 gauging stations after the assimilation of the SAR images on 25^{th} July at 10:27 and 21:53, and on 25^{th} July at 06:34. In Figure 4.4, the analysis expectation (blue line) moves closer to the the observations (red line) with respect to the background (green line) especially in Mythe Bridge and Haw Bridge after the assimilation of the two SAR images on 23^{rd} of July at 10:27 and 21:53. These plots show that

Table 4.3: Verification Measures of the ensemble precipitation.

Catchment outlet	07-2007		11-2012		02-2020	
	VM_1	VM_2	VM_1	VM_2	VM_1	VM_2
Knightsford	1.03	1.02	1.03	1.02	0.92	0.98
Evesham	1	0.89	1	0.89	1	1
Bewdley	1.07	0.96	1.07	0.96	0.92	1

there is an improvement of the water level estimation following the assimilation.

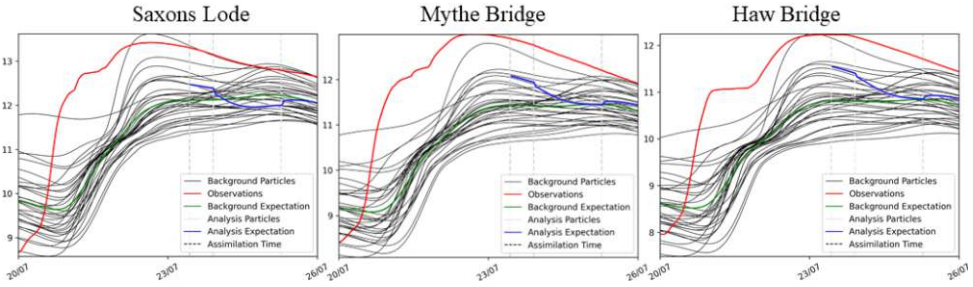


Figure 4.4: Flood event, July 2007. Water level time series plots for 3 gauging stations downstream. Black lines correspond to the Open Loop particles, green line to the Open Loop, the grey lines to the particles after the assimilation, the blue line to the expectation after the assimilation. Red line corresponds to the gauge data. Vertical dashed line correspond to the assimilation time steps. The dates reported on the x-axis correspond to midnight. On the y-axis, the water levels in mAOD is reported.

The DA improves the water level results but there is no spread left after the assimilation. The analysis particle collapse so that they will be all equal to the analysis expectation.

In Figure 4.5 for all of the 3 gauges there is a significant improvement of the predicted water levels as a result of the assimilation. For instance, for the first assimilation in Haw Bridge the analysis errors are decreased by 60% when compared to the errors of the OL. Moreover, the analysis errors are consistently lower than the background errors at the gauging stations of Haw Bridge, Mythe Bridge, and Deerhurst meaning that the benefits of the assimilation persist for several hours after the assimilation.

In the upper part of the river network (i.e. gauging stations at Bewdley, Evesham and Knightsford), there is no improvement after the assimilation. For instance, in Bewdley the discharge is underestimated whereas discharge is overestimated in Knightsford Bridge and Evesham.

A comparison of the simulated flood extent map before and after the assimilation with an aerial photograph acquired on 24^{th} of July shows that the assimilation has limited effect on the predicted flood extent since the OL flood extent is already quite accurate. There is no significant variation of the CSI between the background and the analysis. The contingency maps (Figure 4.6) show that the number of false negative pixels (yellow pixels) increases after the assimilation in the upstream part of the river Severn slightly reducing the CSI from 0.68 to 0.63.

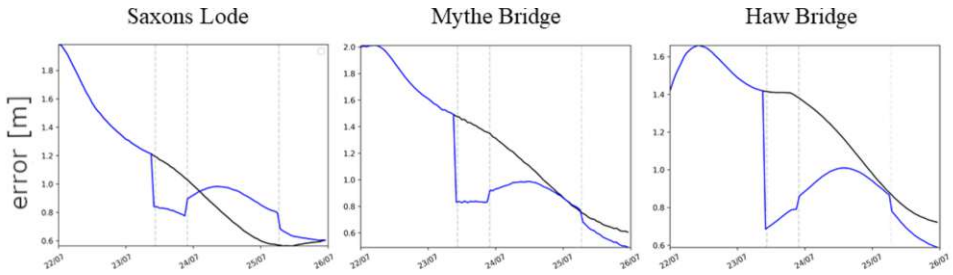


Figure 4.5: Background (black line) error of water level against analysis error (blue line) for 3 gauging stations for the 2007 flood event. Vertical dashed lines indicate the assimilation time steps. The dates reported on the x-axis correspond to the time 00:00.

4.4.2. NOVEMBER 2012

Figure 4.7 shows that the expectation of water levels after the assimilation (blue line) is closer to the observation (red line) and for some cases observations and simulations are overlapping. The water levels at Haw Bridge, Mythe Bridge, Saxons Lode, and Deerhurst are improved after the assimilation. The improvement is more significant than in the 2007 flood event possibly due to the fact that in this case the observed gauge data fall within the background ensemble (Figure 4.3).

However, the discharge in Evesham is consistently overestimated and, to compensate for this error, we observe that the assimilation leads to an underestimation of discharge in Bewdley. The discharge boundary conditions are not improved as in the 2007 flood event. The error of water levels of the background and the analysis is plotted in the Figure 4.8. After each assimilation, the results consistently improve for the gauging stations located downstream of the confluence of the rivers Theme, Avon and Severn. For some assimilations, there is a decrease of the errors of up to 68%. In addition, the analysis errors are lower than the OL most of the times, meaning that the improvements of forecast accuracy of water levels due to the assimilation persist for several hours after the assimilation. The results in Figure 4.8 show that assimilation improves the accuracy of the predicted water levels.

4.4.3. FEBRUARY 2020

Figure 4.9 shows that, for all the 5 assimilations carried out, there is an improvement of the analysis compared to the background and the improvements persist over time as confirmed also in the error times series in Figure 4.10. We observe that the water levels simulated at Haw Bridge, Mythe Bridge, Saxons Lode, and Deerhurst are improved for a majority of time steps when compared to the Open Loop simulations.

The error time series at the gauging stations of Saxons Lode, Mythe Bridge, Haw Bridge and Deerhurst can provide a numerical idea of the improvement achieved via the assimilation (Figure 4.10). There is a substantial decrease of errors at each assimilation time step. For some assimilation time steps and for some gauging station the errors decrease by around 70%. As in the previous cases, the analysis errors are improved not only at the assimilation times but also for the following hours. The degradation of the forecast performance in between the 02/20 and the 02/26 is likely due to the rapid change in the

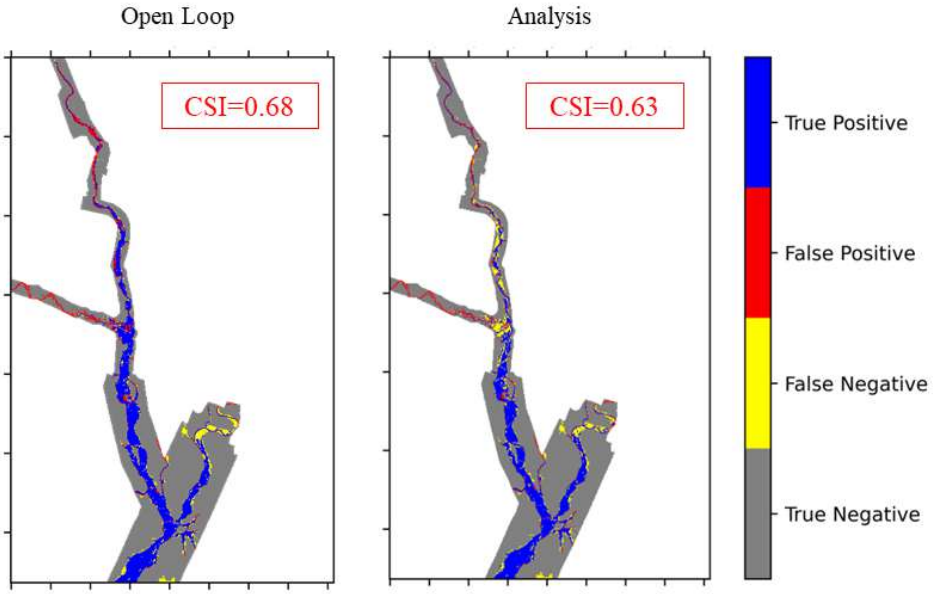


Figure 4.6: 24^{th} July 2007 at 12:00 flood event. Contingency map and Critical Success Index (Equation 2.14) before and after the assimilation. True positive pixels (blue) occur when both the flood simulation and the aerial photograph observe the pixel as flooded, true negative when both the flood simulation and the aerial photograph observe the pixel as not-flooded (grey). When the simulation predicts the pixel as flooded but the pixel is observed as not-flooded a false positive occurs (red), when the contrary takes place a false negative occurs (yellow). The aerial photograph is cut therefore white areas correspond to elements where no observed value is available.

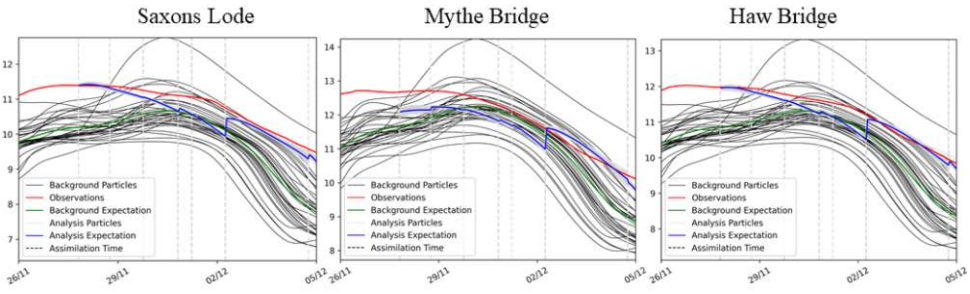


Figure 4.7: Flood event, November 2012. Water levels time series plots for 3 different gauge stations downstream. Black lines correspond to the OL particles, green line to the OL, the grey lines to the particles after the assimilation, the blue line to the expectation after the assimilation. Red line corresponds to the gauge data. Black dashed line correspond to the assimilation time steps.The dates reported on the x-axis correspond to the time 00:00. On the y-axis, the water levels in mAOD is reported.

hydrological conditions. The upstream discharge values are not correctly predicted and

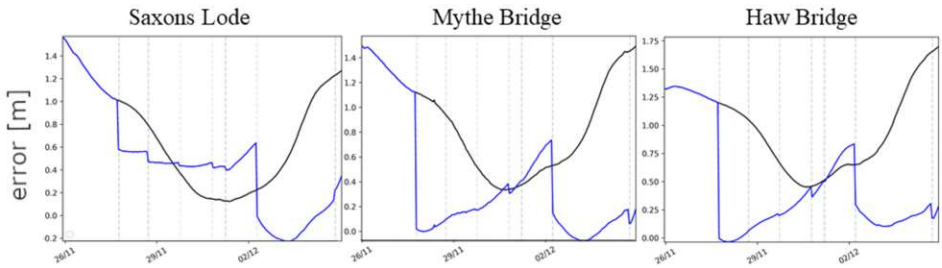


Figure 4.8: Background (black line) error of water level against analysis error (blue line) for 4 gauge stations for the 2012 flood event. Vertical dashed lines indicate the assimilation time steps. The dates reported on the x-axis correspond to the time 00:00.

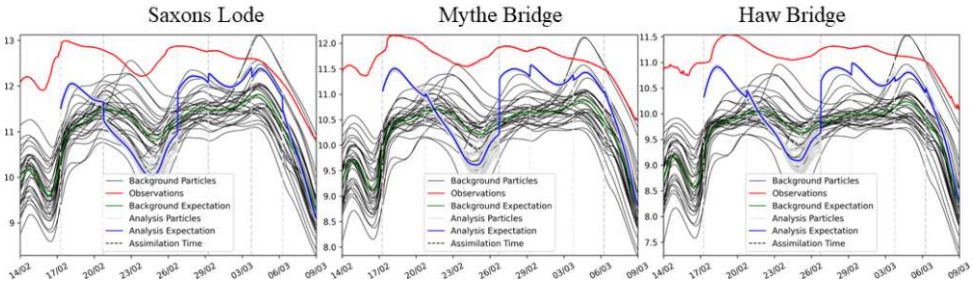


Figure 4.9: Flood event, February 2020. Water levels time series plots for 3 different gauge stations downstream. Black lines correspond to the Open Loop particles, green line to the Open Loop, the grey lines to the particles after the assimilation, the blue line to the expectation after the assimilation. Red line corresponds to the observed gauge data. Vertical dashed lines correspond to the assimilation time steps. The dates reported on the x-axis correspond to the time 00:00. On the y-axis, the water levels in mAOD is reported.

a compensation effect seems to occur: discharge in Evesham is over-estimated while discharge in Bewdley is under-estimated, similarly to the other flood events.

In Figure 4.11, the comparison of the background and analysis flood extent map with an aerial Sentinel-2 image acquired on February 29^{th} at 11:20 shows that the number of false negatives (yellow pixels) decreases after the assimilation while the number of the correctly predicted flood pixels (blue pixels) increases not only in the floodplain but also in the upstream part of the rivers Avon and Severn. This result is also confirmed by the CSI value that goes from 0.37 before the assimilation to 0.52 after the assimilation. The contingency maps show that there is an improvement of the flood extent map accuracy after the assimilation (Figure 4.11).

4.5. DISCUSSION

Our results show that for all 3 flood events the assimilation consistently improves the performances of the forecast water levels, in particular in the area located around the confluence of the 3 main rivers. However, the simulation of discharge at the boundary of the model domain is not always correctly updated. In the literature, different DA methods have been applied and evaluated on the same test site (i.e. River Severn floodplain).

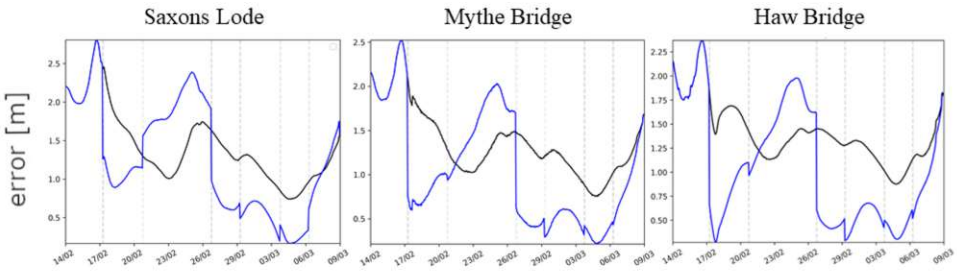


Figure 4.10: Time series of background (black line) error against analysis error (blue line) for 3 gauge stations downstream for the 2020 flood event. Vertical dashed line correspond to the assimilation time steps. The dates reported on the x-axis correspond to the time 00:00.

Hostache et al., 2018 evaluates the performances of a SIS-PF assimilating SAR-derived flood extent into a flood forecasting model including the July 2007 flood event. Similarities can be found between their and this study: the assimilation in 2007 has a limited effect on the improvement of the contingency maps, and the forecast quality of the upstream discharge in Bewdley and Evesham is not always satisfactory. For instance, the discharge is overestimated in Evesham and Knightsford Bridge and this is counterbalanced by discharge underestimation in Bewdley. We believe that errors in discharge will not largely affect the flooding, being this watershed a big valley filling event. The slightly limited effect of the DA in the current study and in the one by Hostache et al., 2018, could be explained with the simplified riverbed geometry of the Avon used in the hydraulic model or with errors in the rating curves (Bates et al., 2006, Bermúdez et al., 2017, Neal et al., 2011). Another explanation provided by Hostache et al., 2018 is that there could be a substantial amount of rainfall falling on the floodplain and causing some flooding. This effect is not taken into account in the hydraulic models. Indeed, the study by Bermúdez et al., 2017 reported that the storm in July 2007 was mostly occurring over the hydraulic domain. To correctly predict the flood peak, such direct rainfall should have been taken into account as it may be a the origin of at least part of the flooding observed by the satellites.

García-Pintado et al., 2015 examined the performances of an EnKF where water level observations obtained during the flood event in November 2012 are assimilated into a hydrodynamic model. As in our study, the assimilation is not improving the estimation of the gauged inflows on the upstream boundary of the hydraulic model. According to the authors, this might be due to unaccounted inflows/outflows that can be related to the neglecting of lateral inflows from small tributaries inflows and due to the resurgence of groundwater. García-Pintado et al., 2015 claim that the results at the boundaries cannot be interpreted as a failure of the assimilation filter but rather as a way of the DA of improving the estimation of the total inflows using satellite observations. In our experiment, the flood extents provide information mainly in the downstream area and therefore on the amount of water that should have entered the riverine system over a given period before the time of the assimilation. García-Pintado et al., 2015 shows that spatial localization is recommended to avoid spurious correlations leading to filter divergence.

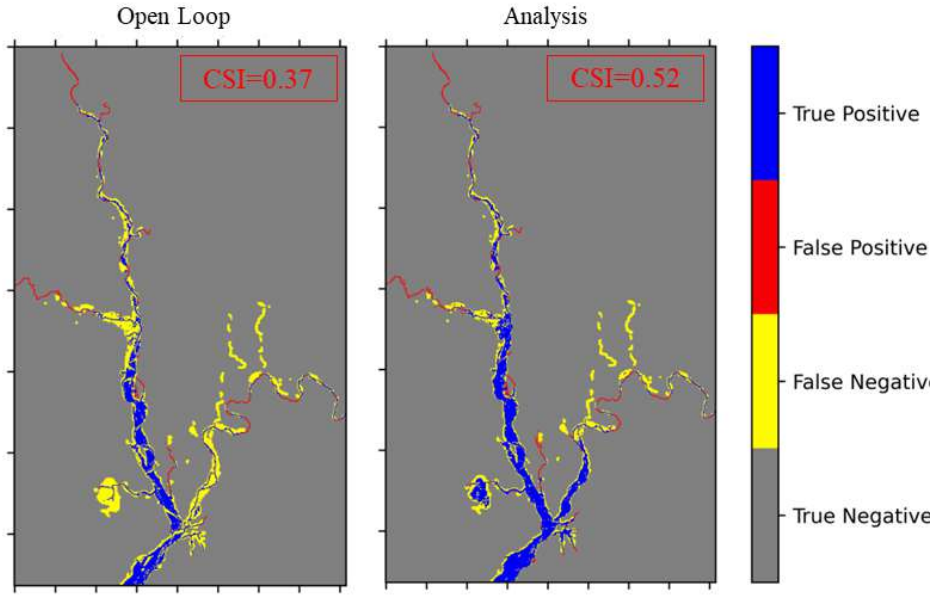


Figure 4.11: 29^{th} of February at 11:20 flood event. Contingency map and Critical Success Index (Equation 2.14) before and after the assimilation. True positive pixels (blue) occur when both the flood simulation and the aerial photograph observe the pixel as flooded, true negative when both the flood simulation and the aerial photograph observe the pixel as not-flooded (grey). When the simulation predicts the pixel as flooded but the pixel is observed as not-flooded a false positive occurs (red), when the contrary takes place a false negative occurs (yellow).

In the previous synthetic study described in Chapter 3, the flood forecasting system was simplified, given that only Bewdley was considered as the contributing upstream catchment, and forecasts were moved towards the observations for both gauges upstream and downstream. In this real case, a new level of complexity is added: input discharge enters the hydraulic model domain from 3 different catchments. Therefore, since no filter localization was applied and weights are globally calculated, the reason for unsatisfactory performances on the boundaries of the hydraulic model after the application of the TPF filter could be due to a compensation effects between the various inflow boundaries. This real world experiment follows a synthetic case experiment defined in Chapter 3 . The main hypothesis in the synthetic experiment was that rainfall and SAR observations were the only sources of uncertainty. The hypothesis seems to be realistic for operational flood forecasting but applying the TPF in a real world experiment shows performances levels lower than that of a the synthetic experiment. Reasons could be found in additional sources of uncertainty not taken into account, for instance such as the geometry of the river bed.

4.6. CONCLUSION

This study is the follow-up of the study in Chapter 3, where a TPF has been validated in a synthetic environment with rainfall as the only source of uncertainty besides SAR observations of flood extent. In this study, the TPF has been applied to 3 flood events occurring along the River Severn (UK) in 2007, 2012 and 2020. The key conclusions that can be drawn from our experiment can be formalized as follows:

- TPF represents a valuable methodology for improving the forecast quality. Indeed the DA experiment shows a substantial and systematic improvement of the predicted water levels when compared to the OL. The flood extent map is correctly updated for the flood event in 2020, the number of flooded pixels corrected after the assimilation increases with respect to the OL, especially along the floodplain.

- Improvements of the water level forecast accuracy persist for several hours and in some cases for several days after the assimilation unless hydrological conditions do not change too rapidly.

- Poor performances of the TPF are shown for the estimation of the inflows at the upstream gauges. The inflow estimation is not correct because of several factors possibly affecting model simulations and consequently DA performances such as erroneous river bed geometries, neglecting lateral inflows, and erroneous rating curves. Moreover, satellite observations can represent another possible cause of this issue.

- The TPF applied in this study computes global weights which are then used to independently update the hydrological model variables. Global weights are calculated based on the flood extent which is the product of the combination of the hydrological conditions of the 3 independent upstream catchment. Given the complexity of the fluvial system the update of the variables based on global weights may lead to errors and inaccurate estimations of inflows at the boundaries of the river because different set of particles can generate very similar flood extent maps.

- The TPF has been previously validated in a synthetic experiment with the underlying hypothesis that rainfall and SAR observations are the only sources of uncertainty. This real-world experiment proves that the TPF improves the performances compared to the Open Loop but with some limitations probably due to several sources of uncertainty not taken into account such as the river bed geometry.

We argue that it is realistic to consider the precipitation as the principal source of uncertainty in most flood forecasting systems. However it seems that other sources of uncertainties can effectively play a decisive role and may have to be taken into account through a more advanced joint parameter-state estimation procedure. In our study, the DA does not specifically take into account these potential sources of uncertainty generating unrealistic inflows. In this study, the application of the TPF to some actual case studies leads to substantial improvements in terms of accuracy of the generated flood extent maps and simulations of water levels at the gauging stations located downstream of the confluence of the 3 rivers Avon, Theme and Severn. However, the method needs to

be tested and refined before being used operationally in near-real time applications. The main limitation remains the unsatisfactory update of the inflows on the boundaries. Further investigations should be conducted, for example developing a joint parameter/state variable estimation for this DA framework, and including new sources of uncertainty such as the river bed geometry. Moreover, an interesting perspective could be to test our DA framework while integrating localization.

5

SUMMARY AND CONCLUSIONS

Accurate, reliable and timely forecasts of flood extent, water level, and discharge are fundamental for mitigating and/or preventing the damaging effects of floods. Data Assimilation (DA) helps in reducing the predictive uncertainty of flood forecasting systems. DA thus represents a useful tool not only for the research community but also for operational water management and emergency response.

The aim of this thesis is to investigate advantages and limitations of integrating flood extent maps derived from SAR data into a hydrologic-hydraulic model using a state-of-the-art Particle Filter. A second objective is to develop, test and evaluate a DA framework that allows mitigating the degeneracy and sample impoverishment issues. Degeneracy and sample impoverishment issues were identified as major limiting factors in state-of-the-art filters that are applied for similar purposes because of poor approximation of the posterior distribution and the loss of diversity among particles, respectively.

In Chapter 2 , we have validated a DA framework previously proposed in Hostache et al., 2018 by applying it in a synthetic experiment where uncertainty is tightly controlled. Hostache et al., 2018 demonstrates that the DA of SAR-derived flood extents improves the prediction accuracy of flood forecast outputs. However, these improvements were not obtained systematically, possibly because of a simplified modelling system with the assumption that observations and rainfall are the only sources of uncertainty. Therefore, a similar experiment has been conducted here in a synthetic environment. A new framework has been developed to convert a real SAR backscatter image into a synthetic probabilistic flood map and to assimilate it into a flood forecasting model in order to carry out the synthetic experiment. In addition, we implement a first procedure to reduce the degeneracy with the use of an inflation coefficient that depends on the size of the ensemble to be retained after the assimilation. Results were extremely satisfactory for the time of the assimilation, both for the enhanced method and for the Sequential Importance Sampling (standard PF). Improvements persist for several days after the assimilation when applying an inflation coefficient with the enhanced method. However degeneracy still represents an issue that undermines the correct use of the Particle Fil-

ter because a limited number of particles cannot be representative of the uncertainty of the model. Degeneracy is the main limitation of the PFs and it occurs when most of the particles have negligible weight, such that the variance of importance weights increases over time and the particles lose their ability to correctly approximate the posterior distribution. In the research community, there exist several techniques to overcome the degeneracy and the loss of the sample diversity (Andrieu et al., 2010; Moradkhani et al., 2012; Kwok et al., 2005; Abbaszadeh et al., 2018; Abbaszadeh et al., 2019).

Chapter 3 therefore explores a newly developed DA technique based on the Tempering Particle Filter (TPF). To overcome the degeneracy the TPF relies on successive applications of tempering coefficients to inflate the likelihood and to reduce the variance among the particle weights. Moreover, at each step, mutation and resampling are applied. The main objective of the chapter is to evaluate whether a method with finite ensemble size can mitigate degeneracy without losing prediction accuracy. The TPF is again applied in a synthetic environment with rainfall and observations as the only sources of uncertainty. The results show that the forecasts are very accurate for the time of the assimilation (e.g., RMSE values decrease by 80% whereas CSI values increase by 30%) and also for subsequent time steps up to 3 days after the assimilation. The TPF outperforms the standard PF (SIS) by Hostache et al., 2018, especially when the performances are compared over time. The improvements achieved with the TPF tend to be more persistent than those obtained with the original SIS and the results show that degeneracy is mitigated. It is rather pragmatic to assume that rainfall is the only source of uncertainty in operational flood forecasting models (Pappenberger et al., 2005; Koussis et al., 2003). In reality, in a flood forecasting model there are several sources of uncertainty such as model parameters, model structure and input (Moradkhani et al., 2005).

In Chapter 4, the TPF is applied to a real-world case where uncertainties are no longer controlled by the user. The aim is to to evaluate its validity when assumptions are not respected and to understand how the different sources of uncertainty could affect the model simulations. Three flood events are considered for the River Severn in the UK to evaluate the performance of the DA framework. The probabilistic flood maps are derived from the satellite images acquired during the flood events in 2007, 2012 and 2020. The results show that there is a consistent improvement of the water level estimation at the gauging stations located in the downstream part of the floodplain. Moreover, improvements are persistent for several hours after the assimilation when compared to the Open Loop (OL, i.e., no assimilation). The Critical Success Index (CSI) tends to be improved by the assimilation, unless the CSI of the OL was already very high before the assimilation was carried out. We noted that the discharge estimation at the gauging station in the upstream part of the river is not satisfactory and the estimation of inflows at the boundaries of the hydraulic model is inaccurate. This could be due to the fact that the flood extent is the result of inflow from three independent catchment areas. The observation of flood extent does not allow an unambiguous determination of the run-off contributed by each tributary and this may explain why the updates are not correct in some parts of the modelled river system. In the synthetic experiment described in Chapter 3, a single contributing catchment upstream of the modelled river reach is taken into account and

this explains that there is a tight connection in the hydraulic conditions between the upstream and downstream parts of the river. In the real world experiment, on the other hand, compensation effects may arise due to the topography of the area. An additional explanation can be that performances tend to decrease in the real-world experiment probably because the assumption of the rainfall as the only source of uncertainty is too simplistic and other sources of uncertainty might be taken into account in order to improve the model predictions. Indeed, other studies show that through the inclusion of parameter uncertainty in the data assimilation process the total uncertainty in the prediction can be more accurately characterized (Moradkhani et al., 2005; Peter et al., 2009; Abbaszadeh et al., 2019; Karthik et al., 2011; Smith et al., n.d.). For instance, the joint estimation of the uncertainty of the parameter of the forecasting model together with the state variable can provide more accurate predictions (Moradkhani et al., 2005; DeChant et al., 2011; Peter et al., 2009; Cooper et al., 2018a). Moreover, the results of the thesis show that, in the real world experiment, degeneracy is mitigated but not completely avoided. According to Peter et al., 2009, considering parameter uncertainty besides precipitation uncertainty could increase the spread of the posterior distribution.

We suggest that the introduction of some filter localization can help preventing the DA framework from having this compensation effect. Moreover, further investigations should be conducted to understand if a joint parameter-state estimation can effectively improve the accuracy and the persistence of the model simulations. The study of this thesis contributes to a further understanding of the PF applications. It has been realized with the aim of improving the existing DA framework, of validating the DA framework when the assumptions are met (synthetic case) and when they are not fully met (real-world case), and of better understanding the advantages and limitations of the DA framework application. This DA framework is generalizable and therefore applicable to other regions, particularly in data-scarce areas, to more accurately estimate floods and to contribute to better define management strategies in a probabilistic context.

BIBLIOGRAPHY

Abbaszadeh, Peyman, Hamid Moradkhani, and Dacian N. Daescu (2019). "The Quest for Model Uncertainty Quantification: A Hybrid Ensemble and Variational Data Assimilation Framework". In: *Water Resources Research* 55.3, pp. 2407–2431. DOI: https://doi.org/10.1029/2018WR023629. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018WR023629. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR023629.

Abbaszadeh, Peyman, Hamid Moradkhani, and Hongxiang Yan (2018). "Enhancing hydrologic data assimilation by evolutionary Particle Filter and Markov Chain Monte Carlo". In: *Advances in Water Resources* 111, pp. 192–204. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2017.11.011. URL: https://www.sciencedirect.com/science/article/pii/S0309170817303421.

Almeida, Gustavo A. M. de and Paul Bates (2013). "Applicability of the local inertial approximation of the shallow water equations to flood modeling". In: *Water Resources Research* 49.8, pp. 4833–4844. DOI: https://doi.org/10.1002/wrcr.20366. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/wrcr.20366. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/wrcr.20366.

Anderson, Jeffrey L. (2001). "An Ensemble Adjustment Kalman Filter for Data Assimilation". In: *Monthly Weather Review* 129.12, pp. 2884 –2903. DOI: 10.1175/1520-0493(2001)129<2884:AEAKFF>2.0.CO;2. URL: https://journals.ametsoc.org/view/journals/mwre/129/12/1520-0493_2001_129_2884_aeakff_2.0.co_2.xml.

Andreadis, Konstantinos M. and Guy J-P. Schumann (2014). "Estimating the impact of satellite observations on the predictability of large-scale hydraulic models". English. In: *Advances in Water Resources* 73.C, pp. 44–54. DOI: 10.1016/j.advwatres.2014.06.006.

Andreadis, Konstantinos M. et al. (2007). "Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model". In: *Geophysical Research Letters* 34.10. DOI: https://doi.org/10.1029/2007GL029721. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2007GL029721. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2007GL029721.

Andrieu, Christophe, Arnaud Doucet, and Roman Holenstein (2010). "Particle Markov chain Monte Carlo methods". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 72.3, pp. 269–342. DOI: https://doi.org/10.1111/j.1467-9868.2009.00736.x. eprint: https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2009.00736.x. URL: https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2009.00736.x.

Annis, A., F. Nardi, and F. Castelli (2021). "Simultaneous assimilation of water levels from river gauges and satellite flood maps for near-real time flood mapping". In: *Hydrology and Earth System Sciences Discussions* 2021, pp. 1–37. DOI: 10.5194/hess-2021-125. URL: https://hess.copernicus.org/preprints/hess-2021-125/.

Arulampalam, M. S. et al. (2002). "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on Signal Processing* 50.2, pp. 174–188. DOI: 10.1109/78.978374.

Bates, Paul D., Matthew S. Horritt, and Timothy J. Fewtrell (2010). "A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling". In: *Journal of Hydrology* 387.1, pp. 33–45. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2010.03.027. URL: https://www.sciencedirect.com/science/article/pii/S0022169410001538.

Bates, Paul D. et al. (2006). "Reach scale floodplain inundation dynamics observed using airborne synthetic aperture radar imagery: Data analysis and modelling". In: *Journal of Hydrology* 328.1. Measurement and Parameterization of Rainfall Microstructure, pp. 306–318. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2005.12.028. URL: https://www.sciencedirect.com/science/article/pii/S0022169406000047.

Bates, P.D and A.P.J De Roo (2000). "A simple raster-based model for flood inundation simulation". In: *Journal of Hydrology* 236.1, pp. 54 –77. ISSN: 0022-1694. DOI: https://doi.org/10.1016/S0022-1694(00)00278-X. URL: http://www.sciencedirect.com/science/article/pii/S002216940000278X.

Bermúdez, María et al. (2017). "Quantifying local rainfall dynamics and uncertain boundary conditions into a nested regional-local flood modeling system". In: *Water Resources Research* 53.4, pp. 2770–2785. DOI: https://doi.org/10.1002/2016WR019903. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2016WR019903. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016WR019903.

Beskos, Alexandros, Dan Crisan, and Ajay Jasra (2014). "On the stability of sequential Monte Carlo methods in high dimensions". In: *The Annals of Applied Probability* 24.4, pp. 1396 –1445. DOI: 10.1214/13-AAP951. URL: https://doi.org/10.1214/13-AAP951.

Beven, Keith and Jim Freer (2001). "Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology". In: *Journal of Hydrology* 249.1, pp. 11–29. ISSN: 0022-1694. DOI: https://doi.org/10.1016/S0022-1694(01)00421-8. URL: https://www.sciencedirect.com/science/article/pii/S0022169401004218.

Blöschl, G. et al. (2019a). "Changing climate both increases and decreases European river floods". In: *Nature* 573.7772, pp. 108–111. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1495-6.

Blöschl, Günter et al. (2019b). "Twenty-three unsolved problems in hydrology (UPH) – a community perspective". In: *Hydrological Sciences Journal* 64.10, pp. 1141–1158. DOI: 10.1080/02626667.2019.1620507. eprint: https://doi.org/10.1080/02626667.2019.1620507. URL: https://doi.org/10.1080/02626667.2019.1620507.

Chen, Xiaohong, Qian Sun, and Jun Hu (2018). "Generation of Complete SAR Geometric Distortion Maps Based on DEM and Neighbor Gradient Algorithm". In: *Applied Sciences* 8.11. ISSN: 2076-3417. DOI: 10.3390/app8112206. URL: https://www.mdpi.com/2076-3417/8/11/2206.

Chini, M. et al. (2017). "A Hierarchical Split-Based Approach for Parametric Thresholding of SAR Images: Flood Inundation as a Test Case". In: *IEEE Transactions on Geoscience and Remote Sensing* 55.12, pp. 6975–6988. ISSN: 0196-2892. DOI: 10.1109/TGRS.2017.2737664.

Cooper, E. S. et al. (2018a). "Observation operators for assimilation of satellite observations in fluvial inundation forecasting". In: *Hydrology and Earth System Sciences Discussions* 2018, pp. 1–32. DOI: 10.5194/hess-2018-589. URL: https://www.hydrol-earth-syst-sci-discuss.net/hess-2018-589/.

Cooper, E.S. et al. (2018b). "Observation impact, domain length and parameter estimation in data assimilation for flood forecasting". In: *Environmental Modelling and Software* 104, pp. 199–214. ISSN: 1364-8152. DOI: https://doi.org/10.1016/j.envsoft.2018.03.013. URL: https://www.sciencedirect.com/science/article/pii/S1364815217303602.

Dasgupta, Antara et al. (2021). "A Mutual Information-Based Likelihood Function for Particle Filter Flood Extent Assimilation". In: *Water Resources Research* 57.2. e2020WR027859 2020WR027859, e2020WR027859. DOI: https://doi.org/10.1029/2020WR027859. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020WR027859. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020WR027859.

De Lannoy, Gabriëlle J. M. et al. (2007). "Correcting for forecast bias in soil moisture assimilation with the ensemble Kalman filter". In: *Water Resources Research* 43.9. DOI: https://doi.org/10.1029/2006WR005449. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006WR005449. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006WR005449.

De Lannoy, GabriÃ«lle J. M. et al. (2006). "Assessment of model uncertainty for soil moisture through ensemble verification". In: *Journal of Geophysical Research: Atmospheres* 111.D10. DOI: 10.1029/2005JD006367. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2005JD006367. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2005JD006367.

DeChant, C. M. and H. Moradkhani (2011). "Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation". In: *Hydrology and Earth System Sciences* 15.11, pp. 3399–3410. DOI: 10.5194/hess-15-3399-2011. URL: https://hess.copernicus.org/articles/15/3399/2011/.

DeChant, Caleb M. and Hamid Moradkhani (2012). "Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting". In: *Water Resources Research* 48.4. DOI: https://doi.org/10.1029/2011WR011011. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011WR011011. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011WR011011.

Doucet, Arnaud., Nando. De Freitas, and Neil Gordon (2001). "Sequential Monte Carlo methods in practice". In: xxvii, 581 p. : DOI: https://doi.org/10.1007/978-1-4757-3437-9.

Environment Agency, Environment Agency (2009). "River Severn Catchment Flood Management Plan." In: URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/289103/River_Severn_Catchment_Management_Plan.pdf.

Fenicia, Fabrizio, Dmitri Kavetski, and Hubert H. G. Savenije (2011). "Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development". In: *Water Resources Research* 47.11. DOI: 10.1029/2010WR010174. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010WR010174. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010WR010174.

Fenicia, Fabrizio et al. (2007). "A comparison of alternative multiobjective calibration strategies for hydrological modeling". In: *Water Resources Research* 43.3. DOI: https://doi.org/10.1029/2006WR005098. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006WR005098. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006WR005098.

Frei, M. and H. R. Kunsch (2013). "Bridging the ensemble Kalman and particle filters". In: *Biometrika* 100.4, 781–800. ISSN: 1464-3510. DOI: 10.1093/biomet/ast020. URL: http://dx.doi.org/10.1093/biomet/ast020.

Garcia-Pintado, Javier et al. (2015). "Satellite-supported flood forecasting in river networks: a real case study". In: *Journal of Hydrology* 523, pp. 706–724. DOI: 10.1016/j.jhydrol.2015.01.084. URL: http://centaur.reading.ac.uk/39388/.

García-Pintado, Javier et al. (2013). "Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling". In: *Journal of Hydrology* 495, pp. 252–266. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2013.03.050. URL: https://www.sciencedirect.com/science/article/pii/S0022169413002783.

García-Pintado, Javier et al. (2015). "Satellite-supported flood forecasting in river networks: A real case study". In: *Journal of Hydrology* 523, pp. 706 –724. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2015.01.084. URL: http://www.sciencedirect.com/science/article/pii/S0022169415001031.

Giustarini, L. et al. (2011a). "Assimilating SAR-derived water level data into a hydraulic model: a case study". In: *Hydrology and Earth System Sciences* 15.7, pp. 2349–2365. DOI: 10.5194/hess-15-2349-2011. URL: https://www.hydrol-earth-syst-sci.net/15/2349/2011/.

— (2011b). "Assimilating SAR-derived water level data into a hydraulic model: a case study". In: *Hydrology and Earth System Sciences* 15.7, pp. 2349–2365. DOI: 10.5194/hess-15-2349-2011. URL: https://hess.copernicus.org/articles/15/2349/2011/.

Giustarini, L. et al. (2015). "Accounting for image uncertainty in SAR-based flood mapping". In: *International Journal of Applied Earth Observation and Geoinformation* 34, pp. 70 –77. ISSN: 0303-2434. DOI: https://doi.org/10.1016/j.jag.2014.06.01

7. URL: http://www.sciencedirect.com/science/article/pii/S0303243414001512.

Giustarini, L. et al. (2016). "Probabilistic Flood Mapping Using Synthetic Aperture Radar Data". In: *IEEE Transactions on Geoscience and Remote Sensing* 54.12, pp. 6958–6969. ISSN: 0196-2892. DOI: 10.1109/TGRS.2016.2592951.

Gordon, N.J., D.J. Salmond, and A.F.M. Smith (1993). "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE Proc. F Radar Signal Process. UK* 140.2, p. 107. DOI: 10.1049/ip-f-2.1993.0015. URL: http://dx.doi.org/10.1049/ip-f-2.1993.0015.

Grimaldi, S. et al. (2020). "Flood mapping under vegetation using single SAR acquisitions". In: *Remote Sensing of Environment* 237, p. 111582. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2019.111582. URL: https://www.sciencedirect.com/science/article/pii/S0034425719306029.

Grimaldi, Stefania et al. (2016). "Remote Sensing-Derived Water Extent and Level to Constrain Hydraulic Flood Forecasting Models: Opportunities and Challenges". In: *Surveys in Geophysics* 37.5, pp. 977–1034. ISSN: 1573-0956. DOI: 10.1007/s10712-016-9378-y. URL: https://doi.org/10.1007/s10712-016-9378-y.

Hamon, W. Russell (1963). "Estimating Potential Evapotranspiration". eng. In: *Transactions of the American Society of Civil Engineers* 128.1, pp. 324–338. ISSN: 0066-0604.

Han, Hua et al. (2011). "An evolutionary particle filter with the immune genetic algorithm for intelligent video target tracking". In: *Computers and Mathematics with Applications* 62.7. Computers and Mathematics in Natural Computation and Knowledge Discovery, pp. 2685–2695. ISSN: 0898-1221. DOI: https://doi.org/10.1016/j.camwa.2011.06.050. URL: https://www.sciencedirect.com/science/article/pii/S089812211100530X.

Herbst, Edward and Frank Schorfheide (2019). "Tempered particle filtering". In: *Journal of Econometrics* 210.1. Annals Issue in Honor of John Geweke "Complexity and Big Data in Economics and Finance: Recent Developments from a Bayesian Perspective", pp. 26–44. ISSN: 0304-4076. DOI: https://doi.org/10.1016/j.jeconom.2018.11.003. URL: https://www.sciencedirect.com/science/article/pii/S0304407618302033.

Hersbach, Hans et al. (Apr. 2019). "Global reanalysis: goodbye ERA-Interim, hello ERA5". In: (159), pp. 17–24. DOI: 10.21957/vf291hehd7. URL: https://www.ecmwf.int/node/19027.

Hoch, J. M. et al. (2017). "GLOFRIM v1.0 – A globally applicable computational framework for integrated hydrological–hydrodynamic modelling". In: *Geoscientific Model Development* 10.10, pp. 3913–3929. DOI: 10.5194/gmd-10-3913-2017. URL: https://gmd.copernicus.org/articles/10/3913/2017/.

Hostache, Renaud et al. (2010). "Assimilation of spatially distributed water levels into a shallow-water flood model. Part II: Use of a remote sensing image of Mosel River". In: *Journal of Hydrology* 390.3, pp. 257 –268. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2010.07.003. URL: http://www.sciencedirect.com/science/article/pii/S0022169410004166.

Hostache, Renaud et al. (2018). "Near-Real-Time Assimilation of SAR-Derived Flood Maps for Improving Flood Forecasts". In: *Water Resources Research* 54.8, pp. 5516–5535.

DOI: 10.1029/2017WR022205. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2017WR022205. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2017WR022205.

Karthik, Nagarajan et al. (2011). "Particle Filter-based assimilation algorithms for improved estimation of root-zone soil moisture under dynamic vegetation conditions". In: *Advances in Water Resources* 34.4, pp. 433–447. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2010.09.019.

Koussis, Antonis D. et al. (2003). "Flood Forecasts for Urban Basin with Integrated Hydro-Meteorological Model." In: *Journal of Hydrologic Engineering* 8.1, p. 1. ISSN: 10840699. URL: http://proxy.bnl.lu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=8687464&site=ehost-live&scope=site.

Kwok, N.M., Gu Fang, and W. Zhou (2005). "Evolutionary particle filter: re-sampling from the genetic algorithm perspective". In: pp. 2935–2940. DOI: 10.1109/IROS.2005.1545119.

Lai, X. et al. (2014). "Variational assimilation of remotely sensed flood extents using a 2-D flood model". In: *Hydrology and Earth System Sciences* 18.11, pp. 4325–4339. DOI: 10.5194/hess-18-4325-2014. URL: https://www.hydrol-earth-syst-sci.net/18/4325/2014/.

Leeuwen, Peter Jan van et al. (2019). "Particle filters for high-dimensional geoscience applications: A review". In: *Quarterly Journal of the Royal Meteorological Society* 145.723, pp. 2335–2365. DOI: 10.1002/qj.3551. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.3551. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.3551.

Li, Ming et al. (2005). "Particle Filter Improved by Genetic Algorithm and Particle Swarm Optimization Algorithm". In: *Journal of Software* 8.3, pp. 666–672.

Liu, Yuqiong and Hoshin V. Gupta (2007). "Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework". In: *Water Resources Research* 43.7. DOI: https://doi.org/10.1029/2006WR005756. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006WR005756. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006WR005756.

Matgen, P. et al. (2010). "Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: proof of concept". In: *Hydrology and Earth System Sciences* 14.9, pp. 1773–1785. DOI: 10.5194/hess-14-1773-2010. URL: https://www.hydrol-earth-syst-sci.net/14/1773/2010/.

Moradkhani, Hamid, Caleb M. DeChant, and Soroosh Sorooshian (2012). "Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method". In: *Water Resources Research* 48.12. DOI: https://doi.org/10.1029/2012WR012144. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012WR012144. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012WR012144.

Moradkhani, Hamid et al. (2005). "Dual state–parameter estimation of hydrological models using ensemble Kalman filter". In: *Advances in Water Resources* 28.2, pp. 135–147. ISSN: 0309-1708. DOI: https://doi.org/10.1016/j.advwatres.2004.09.002. URL: https://www.sciencedirect.com/science/article/pii/S0309170804001605.

Moradkhani, Hamid et al. (2006). "Investigating the impact of remotely sensed precipitation and hydrologic model uncertainties on the ensemble streamflow forecasting". In: *Geophysical Research Letters* 33.12. DOI: https://doi.org/10.1029/2006GL026855. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2006GL026855. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006GL026855.

Murphy, J. M. (1988). "The impact of ensemble forecasts on predictability". In: *Quarterly Journal of the Royal Meteorological Society* 114.480, pp. 463–493. DOI: https://doi.org/10.1002/qj.49711448010. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.49711448010. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.49711448010.

Neal, J. et al. (2011). "Evaluating a new LISFLOOD-FP formulation with data from the summer 2007 floods in Tewkesbury, UK". In: *Journal of Flood Risk Management* 4.2, pp. 88–95. DOI: https://doi.org/10.1111/j.1753-318X.2011.01093.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1753-318X.2011.01093.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1753-318X.2011.01093.x.

Neal, Jeffrey, Guy Schumann, and Paul Bates (2012). "A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas". In: *Water Resources Research* 48.11. DOI: 10.1029/2012WR012514. eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2012WR012514. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2012WR012514.

Neal, R. M. (1996). "Sampling from multimodal distributions using tempered transitions". In: *Statistics and Computing* 6, pp. 353–366. DOI: 10.1175/2009MWR2835.1.

Noh, S. J. et al. (2011). "Applying sequential Monte Carlo methods into a distributed hydrologic model: lagged particle filtering approach with regularization". In: *Hydrology and Earth System Sciences* 15.10, pp. 3237–3251. DOI: 10.5194/hess-15-3237-2011. URL: https://hess.copernicus.org/articles/15/3237/2011/.

Pappenberger, F. et al. (2005). "Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS)". In: *Hydrology and Earth System Sciences* 9.4, pp. 381–393. DOI: 10.5194/hess-9-381-2005. URL: https://www.hydrol-earth-syst-sci.net/9/381/2005/.

Park, SH., YJ Kim, and MT Lim (2010). "Novel adaptive particle filter using adjusted variance and its application". In: *International Journal of Control, Automation and Systems* 8, pp. 801–807. DOI: 10.1007/s12555-010-0412-4.

Peckham, Scott D., Eric W.H. Hutton, and Boyana Norris (2013). "A component-based approach to integrated modeling in the geosciences: The design of CSDMS". In: *Computers and Geosciences* 53. Modeling for Environmental Change, pp. 3–12. ISSN: 0098-3004. DOI: https://doi.org/10.1016/j.cageo.2012.04.002. URL: https://www.sciencedirect.com/science/article/pii/S0098300412001252.

Peter, Salamon and Feyen Luc (2009). "Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation

with the particle filter". In: *Journal of Hydrology* 376.3, pp. 428–442. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2009.07.051.

Plaza, D. A. et al. (2012). "The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter". In: *Hydrology and Earth System Sciences* 16.2, pp. 375–390. DOI: 10.5194/hess-16-375-2012. URL: https://www.hydrol-earth-syst-sci.net/16/375/2012/.

Potthast, Roland, Anne Walter, and Andreas Rhodin (2019). "A Localized Adaptive Particle Filter within an Operational NWP Framework." In: *Monthly Weather Review* 147.1, pp. 345 –362. ISSN: 00270644. URL: http://proxy.bnl.lu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=134406412&site=ehost-live&scope=site.

Pulvirenti, Luca, Giuseppe Squicciarino, and Elisabetta Fiori (2020). "A Method to Automatically Detect Changes in Multitemporal Spectral Indices: Application to Natural Disaster Damage Assessment". In: *Remote Sensing* 12.17. ISSN: 2072-4292. DOI: 10.3390/rs12172681. URL: https://www.mdpi.com/2072-4292/12/17/2681.

Rajib, Adnan et al. (2020). "Towards a large-scale locally relevant flood inundation modeling framework using SWAT and LISFLOOD-FP". In: *Journal of Hydrology* 581, p. 124406. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2019.124406. URL: https://www.sciencedirect.com/science/article/pii/S0022169419311412.

Reich, S. (2013). "A nonparametric ensemble transform method for bayesian inference". In: *SIAM Journal on Scientific Computing* 35.4. cited By 64, A2013–A2024. DOI: 10.1137/130907367.

Revilla-Romero, Beatriz et al. (2016). "Integrating remotely sensed surface water extent into continental scale hydrology". In: *Journal of Hydrology* 543, pp. 659–670. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2016.10.041. URL: https://www.sciencedirect.com/science/article/pii/S0022169416306862.

Smith, Tyler Jon and Amanda Marshall Lucy (n.d.). "Bayesian methods in hydrologic modeling: A study of recent advancements in Markov chain Monte Carlo techniques". In: *Water Resources Research* 44.12 (). DOI: https://doi.org/10.1029/2007WR006705.

Snyder, Chris et al. (2008). "Obstacles to High-Dimensional Particle Filtering." In: *Monthly Weather Review* 136.12, pp. 4629 –4640. ISSN: 00270644. URL: http://proxy.bnl.lu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=36092236&site=ehost-live&scope=site.

Tellman, B. et al. (2021). "Satellite imaging reveals increased proportion of population exposed to floods." In: *Nature* 596, 80–86. DOI: 10.1093/biomet/ast020. URL: https://doi-org.proxy.bnl.lu/10.1038/s41586-021-03695-w.

UNISDR (2015). "United Nations Office for Disaster Risk Reduction. Making Development Sustainable: The Future of Disaster Risk Management. Global Assessment Report on Disaster Risk Reduction". In: URL: www.unisdr.org/we/inform/publications/42809.

Van Leeuwen, P. J. (2009). "Particle filtering in Geophysical Systems". In: *Mon. Wea. Rev.* 137, pp. 4089–4114. DOI: 10.1175/2009MWR2835.1.

Van Wesemael, Alexandra (2019). "Assessing the value of remote sensing and in situ data for flood inundation forecasts". eng. PhD thesis. Ghent University, pp. XXVI, 209. ISBN: 9789463572552.

Wang, Qicong et al. (2006). "Enhancing Particle Swarm Optimization Based Particle Filter Tracker". In: ICIC'06, 1216–1221.

Wongchuig-Correa, Sly et al. (2020). "Assimilation of future SWOT-based river elevations, surface extent observations and discharge estimations into uncertain global hydrological models". In: *Journal of Hydrology* 590, p. 125473. ISSN: 0022-1694. DOI: https://doi.org/10.1016/j.jhydrol.2020.125473. URL: https://www.sciencedirect.com/science/article/pii/S0022169420309331.

Wood, M. et al. (2016). "Calibration of channel depth and friction parameters in the LISFLOOD-FP hydraulic model using medium-resolution SAR data and identifiability techniques". In: *Hydrology and Earth System Sciences* 20.12, pp. 4983–4997. DOI: 10.5194/hess-20-4983-2016. URL: https://hess.copernicus.org/articles/20/4983/2016/.

Zhao, Jie et al. (2021). "Deriving exclusion maps from C-band SAR time-series in support of floodwater mapping". In: *Remote Sensing of Environment* 265, p. 112668. ISSN: 0034-4257. DOI: https://doi.org/10.1016/j.rse.2021.112668. URL: https://www.sciencedirect.com/science/article/pii/S0034425721003886.

Zhu, Mengbin, Peter Jan van Leeuwen, and Javier Amezcua (2016). "Implicit equal-weights particle filter". In: *Quarterly Journal of the Royal Meteorological Society* 142.698, pp. 1904–1919. DOI: 10.1002/qj.2784. eprint: https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/qj.2784. URL: https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2784.