Data Assimilation Using Sequential Monte Carlo Methods
in Hydrological Applications

Data-assimilatie met behulp van sequentiële Monte Carlo-methoden
in hydrologische toepassingen

Douglas Antonio Plaza Guingla

Promotoren: prof. dr. ir. R. De Keyser, prof. dr. ir. V. Pauwels
Proefschrift ingediend tot het behalen van de graad van
Doctor in de Ingenieurswetenschappen: Werktuigkunde-Elektrotechniek

Vakgroep Elektrische Energie, Systemen en Automatisering
Voorzitter: prof. dr. ir. J. Melkebeek
Faculteit Ingenieurswetenschappen en Architectuur
Academiejaar 2013 - 2014

UNIVERSITEIT
GENT

ISBN 978-90-8578-631-3
NUR 959
Wettelijk depot: D/2013/10.500/64

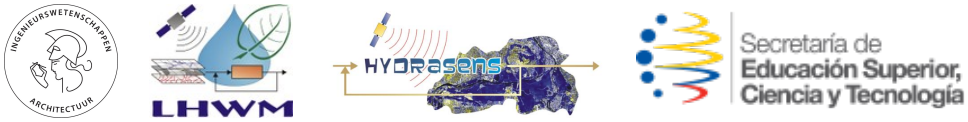
Members of the jury

prof. dr. ir. Robain De Keyser (Ghent University, supervisor)
prof. dr. ir. Valentijn Pauwels (Monash University, Australia, supervisor)
ereprof. Rene Boel (Ghent University)
prof. dr. ir. Niko Verhoest (Ghent Univerity)
prof. dr. Cesar de Prada (Universidad de Valladolid, Spain)
dr. ir. Clara Ionescu (Ghent University)
dr. ir. Hans Lievens (Ghent University)
dr. ir. Patrick Matgen (Public Research Centre-Gabriel Lippman, Luxembourg)
prof. dr. ir. Rik Van de Walle (Ghent University)

Affiliations

Department of Electrical energy, Systems and Automation (EeSA)
Faculty of Engineering and Architecture
Ghent University

Sint-Pietersnieuwstraat 41
B9000 Ghent
Belgium



Acknowledgments

This dissertation was possible through the collaboration of a number of people mainly from the Laboratory of Hydrology and Water Management and from the Department of Electrical energy, Systems and Automation. These lines are devoted to all of them.

My truthful thankfulness goes to my promoters. I would like to express my uppermost gratitude to my promoter prof. dr. ir. Valentijn Pauwels. I am very grateful for his timely and proficient reviews, discussions, ideas and suggestions during the course of my PhD studies. Thanks to Valentijn also for his patience and his distinctive mentor character which boosted my motivation. I would like to thank my promoter prof. dr. ir. Robain De Keyser for the important comments and ideas received during the development of my research. I truly received a good support, supervision and encouragement to finalize my studies successfully.

The research work of this thesis was performed mainly at the Laboratory of Hydrology and Water Management. In this sense, I have only words of gratefulness to the staff members and colleagues at the Laboratory of Hydrology. Thanks to prof. dr. ir. Niko Verhoest for the discussions and ideas about my research. Special thanks to Gabriëlle De Lannoy for allowing me to use part of the code which was developed during her PhD studies and for the valuable comments and contributions to my research topic. My gratitude to the technical and administrative staff: Rudi Hoeben, Marcella Van Hese and Davy Loete for their valuable contribution either by helping me with the set-up of my experiments or by helping with the administrative stuff. Thanks to my colleagues: Hans Lievens, Bruno Samain, Sander Vandenberghe and Lien Loosvelt for creating a nice and ideal environment to conduct my research.

The research conducted in this dissertation was financially supported by the Belgian Science Policy for the Integrating Radar Remote Sensing, Hydrologic and Hydraulic Modeling for Surface Water Management (HYDRASENS) project in the frame of the Support to the Explotation and Research in Earth Observation (STEREO II) programme. The last phase of this doctoral work,

which was related to the writing of the manuscript and the report of the final research results, was funded during the academic year 2012–2013 by Secretaría de Educación Superior en Ciencia y Tecnología (SENESCYT). SENESCYT is a research governmental organization from Ecuador.

Finally, this dissertation would not have come to a successful end without the emotional and spiritual support from my family. First, my gratitude to my wife Glenda Sanchez for encouraging me during the hard moments. My daughter Valentina helps me to keep in my mind that every day we have the opportunity to learn something new. Special thanks to my brother Vicente Plaza and his family for taking care of my parents during my stay abroad. The efforts and dedication involved in this dissertation were inspired by the amazing example of hard work of my parents Elena Rosa Guingla Pucha and Manuel Ignacio Plaza Quizhpi.

Ghent, September 2013

Douglas Plaza G.

Samenvatting

Het onderwerp van dit manuscript heeft betrekking tot natuurwaarnemingen. De focus van dit werk ligt op hydrologische systemen. Hydrologische waarnemingen zijn in deze context van fundamenteel belang voor toepassingen in de werkelijkheid zoals waterbeheer, waterbevoorrading, landbouwproductie, voorspelling van overstromingen, weers en klimaatvoorspelling en milieubehoud. Verbeterde voorspellingen van de hydrologische omstandigheden zijn nuttig voor landbouw, ecologie, burgerlijke ingenieurswetenschap, waterbeheer, voorspellingen voor neerslagafvoer en rampenmanagement.

Hydrologische omstandigheden zijn afhankelijk van tijd en ruimte. Waarnemingen verschaffen bijgevolg enkel informatie over het hydrologische systeem op het moment dat de waarnemingen gesampled worden. Hydrologische modellen kunnen gebruikt worden om de tijdelijke en ruimtelijke variaties te voorspellen; maar die voorspellingen zijn vaak zwak door fouten in de beginwaarden, foutieve parameters, foute input en simplificaties. Het is mogelijk om de schattingen bij te sturen door de sterktes van hydrologische modellering te combineren met waarnemingen. Die doelstelling wordt bereikt door data-assimilatie.

Data-assimilatie is een moderne methodologie die te maken heeft met de relatie tussen natuurlijke data en computermodellen. De methode bestaat eruit om algemene dynamieken van een model te combineren of te verbinden met een set van waarnemingen. Alle modellen hebben fouten en alle datasets hebben hun beperkingen. Het doel van data-assimilatie is het verschaffen van adequate schattingen.

De ensemble Kalman filter is de vaakst gebruikte assimilatiemethode in hydrologie. Recente studies rapporteren echter een probleem met de ensemble Kalman filter: de onderliggende veronderstelling van Gaussianiteit in de functies voor waarschijnlijkheidsverdeling. Aangezien het evident is dat deze veronderstelling niet realistisch is voor hydrologische systemen, zouden er niet Gaussiaanse assimilatiemethodes moeten worden ontwikkeld om deze beperking te overwinnen. Een methode die steeds meer aandacht krijgt in de hydrologie is de particle filter (ook bekend als de sequentiële Monte

Carlo-methoden). In de particle filter methodologie worden waarschijnlijkheidsverdelingen voorgesteld door de particle approximation. De particle filter laat toe om eender welke verdeling weer te geven. Met andere woorden, de veronderstelling van Gaussianiteit, waarvan men uitgaat in de Kalman filter, wordt uitgesloten als men de particle filters gebruikt.

In deze verhandeling wordt de werking van een aantal data assimilatiemethodes, gebaseerd op de particle filter, onderzocht. De hydrologische modellen die in deze studie gebruikt worden, zijn een eenvoudig conceptueel model en een ruimtelijk verdeeld fysisch gebaseerd model. De geassimileerde hydrologische variabelen zijn de afvoer aan het uistroompunt van het stroomgebied en het volumetrisch bodemvochtgehalte. Deze variabelen werden in verschillend georganiseerde experimenten geassimileerd. Omwille van deze reden hangen de doelstellingen van deze studie af van de organisatie van de experimenten.

De doelstelling van het eerste deel van deze dissertatie is het analyseren van de verbetering van de werking van de particle filter door het inbrengen van een resample-move step, of gebruik makend van de Gaussiaanse particle filter. Meer specifiek werd de structuur van de standaard particle filter aangepast door gebruik te maken van een Markovketen move step. De tweede keuze in deze studie gebruikt de momenten van een ensemble Kalman filter analyse om de dichtheidsfunctie te definieren binnen de structuur van de Gaussiaanse particle filter. Beide varianten van de standaard particle filter werden gebruikt om hoogfrequente afvoermetingen in een conceptueel neerslag-afvoermodel te assimileren. De resultaten geven aan dat het gebruik van een resample move step in de standaard particle filter en het gebruik van een optimale dichtheidsfunctie in de Gaussiaanse particle filter de efficientie van de particle filter verhogen. Verder werd een optimalisatie van het voorspellingsensemble gebruikt in deze studie, hetgeen leidde tot een betere werking van de aangepaste particle filter, vergeleken met de particle filter met resample move step.

Het doel van het tweede deel van dit manuscript is het verbeteren van de afvoervoorspellingen door de assimilatie van pseudo-afstandswaargenomen bodemvochtsmetingen. De Ensemble Kalman filter (EnKF) en de Sequential Importance Resampling (SIR) particle filter werden geeavalueerd, hetgeen leidde tot het besluit dat beide methodes tot gelijkaardige resultaten leiden. Verder werd bij beide filters een negatief effect op de basisafvoer waargenomen, veroorzaakt door een inconssistentie tussen de modelparameters en de toestandsvariabelen na de correctie door de filter. Om dit consistentieprobleem op te lossen werd een parameter resamplingmethode voorgesteld, om tot consistente parameterwaarden en toestandsvariabelen te leiden. Extreme particle replicatie, hetgeen tot een ensembleineenstorting kan leiden, werd vermeden door de verstoring van de parameterwaarden met

witte ruis. De resultaten van een synthetische studie tonen aan dat de SIR filter hierdoor tot veel betere resultaten leidt. Verder werd de correctie van de basisafvoer positief beinvloed. Een robuustheidstest op de voorgestelde methodologie leidt tot het besluit dat goede resultaten kunnen verwacht worden bij de assimilatie van echte afstandswaargenomen data.

In deze thesis worden enkele varianten van de originele standard particle filter voorgesteld. De verschillende versies worden afzonderlijk voorgesteld, naargelang het gaat om een conceptueel model of een fysisch gebaseerd model. Het is duidelijk dat de algemene bijdrage van dit werk betrekking heeft tot het lopende onderzoek naar verbeterde methodes voor sequentiële data-assimilatie onder niet-lineaire of niet-Gaussiaanse omstandigheden. De originele bijdragen van deze dissertatie worden hieronder beschreven:

- De assimilatiemethodes gebaseerd op de Kalman filter worden afgeleid en voorgesteld door middel van de recursieve aanpak van Bayes. Door de manier waarop de methodologieën worden afgeleid kunnen de Kalman filters en de particle filters met elkaar worden vergelijken.
- Er wordt een nieuwe assimilatiemethode geïntroduceerd. De methodologie bekend als ensemble Gaussiaanse particle filter beoogt een combinatie van de meest relevante kenmerken van de ensemble Kalman filter en de Gaussiaanse particle filter.
- Het potentiële gebruik van de ensemble Gaussiaanse particle filter in een niet-lineair/niet-Gaussiaans scenario wordt onderzocht. De ensemble Gaussianse particle filter wordt vergeleken met de particle filter met Markov Chain Monte Carlo move step als beiden toegepast worden op een hydrologisch schattingsprobleem. De particle filter met Markov Chain Monte Carlo move step is de state of the art filter voor niet-lineaire/niet-Gaussiaanse schattingsproblemen.
- De particle filter wordt toegepast op een schattingsprobleem voor variabelen en parameters samen. De belangrijkste bijdrage betreft de verbetering die verkregen werd door de impliciete identificatie van optimale parameterwaarden en de algemene bijsturing van de gemodelleerde hydrologische variabelen.

De methodologieën die in deze dissertatie worden voorgesteld kunnen aangewend worden in verder onderzoek naar hydrologische schattingsproblemen.

Summary

The topic of interest in this manuscript is related to observing nature. The focus of this work is on hydrologic systems. In this context, hydrologic observations are of critical importance to real-world applications such as water resource management, water supply, agricultural production, flood prediction, weather and climate forecasting, and environmental preservation. Improved hydrologic condition estimates are useful for agriculture, ecology, civil engineering, water resources management, rainfall-runoff prediction, atmospheric process studies, climate and weather prediction, and disaster management.

Hydrologic conditions are time and space dependent. Additionally, observations provide information of the hydrologic system only at the current time instant when the observations are sampled. Hydrologic system models can be used to predict the temporal and spatial hydrologic variations, but these predictions are often poor, due to model initialization, parameter and forcing, and model simplification errors. It is possible to correct the estimates by combining the strengths of hydrologic modeling with the observations. This goal is obtained through data assimilation.

Data assimilation is a modern methodology which concerns the relation between natural data and dynamical computer models. The procedure consists of the combination or merging of general dynamics of a model with a set of observations. All dynamical models have modeling errors and all data sets are finite. Thus, the aim of data assimilation is to provide proper estimates of nature.

The ensemble Kalman filter is the most frequently used assimilation method in hydrology. However, recent studies report that one problem with the ensemble Kalman filter is the underlying assumption of Gaussianity in the probability density functions. As it is evident that this assumption is not realistic for hydrologic systems, non-Gaussian assimilation methods should be developed in order to overcome this limitation. One method that is receiving increasing attention in hydrology is the particle filter (a.k.a., sequential Monte Carlo methods). In the particle filter methodology, all the

probability distribution functions are described by the point-mass approximation. This approximation allows to represent any kind of distribution function. In other words, the assumption of Gaussian distributions, which is held in the application of the Kalman filter, is relaxed when using particle filters.

In this dissertation, the performance of particle filter based assimilation methods in hydrology is assessed. The hydrologic models used in this study correspond to a lumped conceptual model and a distributed physically-based model. The hydrologic variables to be assimilated are the discharge of water at the outlet of the catchment and the volumetric soil moisture content. These variables are assimilated in experiments with different setups. Therefore, the objectives of this study depend on the experimental setup.

In the first part of this dissertation, the main objective is to analyze the improvement in the performance of the particle filter by including a resample-move step or by using a modified Gaussian particle filter. Specifically, the standard particle filter structure is altered by the inclusion of the Markov chain Monte Carlo move step. The second choice adopted in this study uses the moments of an ensemble Kalman filter analysis to define the importance density function within the Gaussian particle filter structure. Both variants of the standard particle filter are used in the assimilation of densely sampled discharge records into a conceptual rainfall-runoff model. The results indicate that the inclusion of the resample-move step in the standard particle filter and the use of an optimal importance density function in the Gaussian particle filter improves the effectiveness of particle filters. Moreover, an optimization of the forecast ensemble used in this study, allowed for a better performance of the modified Gaussian particle filter compared to the particle filter with resample-move step.

The aim of the second part of this manuscript is to improve discharge predictions by the assimilation of pseudo-remote sensed soil moisture observations. The Ensemble Kalman filter (EnKF) and the Sequential Importance Resampling (SIR) particle filter are evaluated for this purpose finding similar functioning and performances between each other. Furthermore, both filters experience a negative effect on the prediction of the discharge due to a consistency problem between the model parameter values and the states after the filter correction step. In order to overcome this consistency problem, the parameter resampling approach is used aiming to assign consequent parameter values to soil moisture conditions. Extreme parameter replication, which could generate a particle collapse, is avoided by the perturbation of the parameters with white Gaussian noise. Results of the synthetic study carried out in this study report a considerable improvement in the performance of the standard particle filter with the complementary parameter resampling stage. In addition to the improvement of the performance of the particle

filter, the baseflow correction influenced by the soil moisture assimilation is positively affected. A robustness test of the methodology is presented and the results obtained from this test foresee confident performance of the algorithm when using real remote sensed observations.

This dissertation presents some variants of the original particle filtering method. The variant versions are presented separately according to the application thus applied either to conceptual or to physically based hydrologic models. It is clear that the general contribution of this work is related to the ongoing research of improving sequential data assimilation methods under nonlinear and non-Gaussian conditions. The original contributions of this dissertation are described here:

- The Kalman filter based assimilation methods are derived and presented by using the recursive Bayesian approach. The way how the methodologies are derived allows for a comparison between the Kalman filters and the particle filters.
- A novel assimilation method is introduced. The methodology referred to as the ensemble Gaussian particle filter aims to combine the most relevant features of the ensemble Kalman filter and the Gaussian particle filter.
- The potential use of the ensemble Gaussian particle filter in a nonlinear and non-Gaussian scenario is explored. The ensemble Gaussian particle filter is compared to the particle filter with Markov Chain Monte Carlo move step when both are applied to a hydrologic estimation problem. The particle filter with Markov Chain Monte Carlo move step is the state of the art filter in nonlinear/non- Gaussian estimation problems.
- The particle filter is applied to a joint state-parameter estimation problem. The main contribution concerns the improvement obtained in the estimation by the implicit identification of optimal model parameter values and the general correction of the modeled hydrologic variables.

The methodologies presented in this dissertation can be used in further research related to hydrologic estimation problems.

Contents

List of Figures

List of Tables

List of Algorithms

Symbols and acronyms

Symbols

Symbol	Description
$\sim$	distributed according to
$\prod_{k=1}^{n}$	the product of a sequence of factors
$\simeq$	approximately equal
δ	Dirac delta function
$\propto$	proportional to
$\mathcal{M}(.)$	Multinomial distribution
$< . >$	temporal average
$\xrightarrow[N\to\infty]{}$	almost sure convergence
$\xRightarrow[N\to\infty]{}$	convergence in distribution
$E[.]$	Expected valued of a random field
$\overline{E[.]}$	empirical expectation
$\mathbf{F}_{t,t-1}$	linear/linearized(Jacobian) transition matrix at time step $t-1$ to t
$\mathbf{f}_{t,t-1}(\mathbf{x},\ldots)$	discrete non-linear function for the propagation of the states at time step $t-1$ to t
$\mathcal{F}''_{t,t-1}$	Hessian matrix of the system function $\mathbf{f}_{t,t-1}(.)$
$\mathbf{h}_t(\mathbf{x},\ldots)$	discrete non-linear function mapping the states to the observations at time step t
$\mathbf{H}_t$	linear/linearized(Jacobian) observation matrix which maps the system states to the observation system at time t
$\mathbf{M}^T$	Transpose of matrix $\mathbf{M}$
$\mathcal{N}(\mathbf{z}:\mu,\mathbf{\Sigma})$	Gaussian distribution of random field $\mathbf{z}$ with parameters: mean μ and covariance matrix $\mathbf{\Sigma}$
N	number of ensemble members/particles

Symbol	Description
$p(.)$	probability density function
$p(.\|.)$	conditional probability density function
$p(.,.)$	join probability density function
$p_N(.)$	empirical probability density function
$\mathbf{P}_t$	process error covariance matrix at time step t
$\mathbf{P}_t^a$	process error covariance matrix for the analysis step at time step t
$\mathbf{P}_t^f$	process error covariance matrix for the forecast step at time step t
$\mathbf{P}_t^{xy}$	cross covariance between the forecast ensemble and the modeled observation ensemble at time step t
$\mathbf{P}_t^{yy}$	sample covariance of the modeled observation ensemble at time step t
$\hat{\mathbf{P}}_t$	weighted covariance of the process error
$q(.)$	importance probability density function
$\mathbf{Q}$	positive definite covariance matrix of the process noise
$\mathbf{R}$	positive definite covariance matrix of the observation noise
$\mathbf{u}_t$	vector of the input signals at time step t
$\mathcal{U}(.)$	Uniform distribution
$\mathbf{v}_t$	observation noise vector at time step t
$var(.)$	variance of a random variable
$\mathbf{w}_t$	model noise vector at time step t
$w_{t,i}$	unnormalized importance weight of the ith particle at time step t
$\tilde{w}_{t,i}$	normalized importance weight of the ith particle at time step t
$\mathbf{x}_t$	state vector of a discrete stochastic system at time step t
$\mathbf{x}_t^a$	state vector for the analysis step at time step t
$\mathbf{X}_t^a$	analysis ensemble matrix at time step t
$\mathbf{x}_{t,i}^a$	ith member of the analysis state ensemble at time step t
$\mathbf{x}_t^f$	state vector for the forecast step at time step t
$\mathbf{X}_t^f$	forecast ensemble matrix at time step t
$\mathbf{x}_{t,i}^f$	ith member of the forecast state ensemble at time step t
$\mathbf{x}_{t,i}^*$	ith member of the particle set obtained after resampling at time step t
$\mathbf{x}_{1:t}$	sequence of state vectors at time step 1 up to time step t
$\hat{\mathbf{x}}_t$	weighted mean of the state vector
$\bar{\mathbf{x}}_t$	expectation of $\mathbf{x}$ at time step t

Symbol	Description
$\mathbf{y}_t$	state vector of observations at time step t
$\mathbf{y}_{1:t}$	sequence of observation vectors at time step 1 up to time step t
$\mathbf{Y}_t^s$	surrogate observations matrix at time step t
$\mathbf{y}_{t,i}^s$	ith member of the surrogate observations matrix at time step t

List of symbols for HBV model

Symbol	Description
A	area of the Zwalm catchment [m^2]
s_{soil}	water storage in the soil reservoir [m^3]
s_{slow}	water storage in the slow reacting reservoir [m^3]
s_{fast}	water storage in the fast reacting reservoir [m^3]
$s_{soil_{t=0}}$	initial condition of s_{soil} [m^3]
$s_{slow_{t=0}}$	initial condition of s_{slow} [m^3]
$s_{fast_{t=0}}$	initial condition of s_{sfast} [m^3]
s_{slow}	water storage in the slow reacting reservoir [m^3]
s_{fast}	water storage in the fast reacting reservoir [m^3]
E_{tr}	actual evapotranspiration [m^3/s]
E_{tp}	potential evapotranspiration [m^3/s]
R_{in}	infiltration [m^3/s]
R_{eff}	effective precipitation [m^3/s]
R_{fast}	input, fast reacting reservoir [m^3/s]
R_{slow}	input, slow reacting reservoir [m^3/s]
P_{er}	percolation [m^3/s]
Q_{fast}	output flow of the fast reacting reservoir [m^3/s]
Q_{slow}	output flow of the slow reacting reservoir [m^3/s]
Q_{dis}	total discharge of the catchment [m^3/s]
λ	model parameter [-]
S_{max}	model parameter [m^3]
b	model parameter [-]
α	model parameter [-]
P	model parameter [m^3s^{-1}]
β	model parameter [-]
γ	model parameter [-]
$S_{2,max}$	model parameter [m^3]
κ_2	model parameter [m^3s^{-1}]
κ_1	model parameter [s^{-1}]
Θ	vector containing the model parameters

List of symbols for CLM

Symbol	Description
$\boldsymbol{\rho_{liq}}$	density of liquid [kg.m^{-3}]
$\boldsymbol{\theta_{sat_j}}$	volumetric soil water at saturation in layer j [m^3.m^{-3}]
$\boldsymbol{\psi_{max}}$	maximum value of negative of leaves potential before dessication or wilting point potential [mm], model parameter ***smpmax***
$\boldsymbol{\psi_{sat}}$	saturated soil water potential [mm],
$\boldsymbol{\psi_{min}}$	restriction for minimal soil water potential [mm], model parameter ***smpmin***
$\boldsymbol{B}$	model parameter ***bws*** [-], defined by Clapp and Hornberger [1978] model parameter ***sucsat***
$\boldsymbol{f_z}$	water table depth, model parameter ***fz*** [m^{-1}]
$\boldsymbol{K_d}$	saturated soil hydraulic conductivity at the bottom, model parameter $\boldsymbol{k_d}$ [mm.s^{-1}]
$\boldsymbol{l_d}$	base flow parameter for saturated fraction, model parameter ***ld*** [mm.s^{-1}]
NwRs	last top layer contributing to the calculation of surface runoff [-], model parameter ***NwRs***
NwRb	first bottom layer contributing to the calculation of surface runoff [-], model parameter ***NwRb***
θ_j	volumetric soil moisture [m^3.m^{-3}]
Δz_j	thickness of soil layer j [m]
ψ_j	matrix potential limited by $\boldsymbol{\psi_{max}}$ [mm] parameter $\boldsymbol{watsat_j}$
$\psi_j - z_j$	hydraulic height [mm]
$E_{g,Surf}$	soil surface evaporation [kg.m^{-2}.s^{-1}]
$f_{root,j}E_{tr}$	water extracted by plant roots, transpiration [kg.m^{-2}.s^{-1}]
$f_{root,j}$	effective root fraction [-]
$f_{r,j}$	root fraction within soil layer j [-]
f_{sat}	partial contributing area, saturated fraction [-]
G_w	effective precipitation composed by throughfall, drip from canopy and snowmelt [kg.m^{-2}.s^{-1}]
K_j	unsaturated hydraulic conductivity in layer j [mm.s^{-1}]
$M_{ice,liq}$	rate of phase change (ice to liquid) [kg.m^{-3}.s^{-1}]
q_j	water flow between neighboring layers, positive downward [kg.m^{-2}.s^{-1}]
q_0	surface flow [kg.m^{-2}.s^{-1}]
R_s	surface runoff [kg.m^{-2}.s^{-1}]
R_b	baseflow drainage [kg.m^{-2}.s^{-1}]

Symbol	Description
S_m	snowmelt [kg.m^{-2}.s^{-1}]
s_j	soil wetness [-]
$\bar{w}_b$	weighted soil wetness for bottom ***NwRb*** layers [m.mm.s^{-1}]
w_{fact}	parameter determined by the distribution of the topographic index, model parameter ***wtfact*** [-]
$\bar{w}_s$	soil layer thickness weighted soil wetness [-]
z_j	node depth of layer j, positive, increasing downward [mm]
z_w	mean water table depth [-]

Acronyms

Notation	Description
APF	Auxiliary Particle Filter
%BIAS	Percentage of bias
CLM	Community Land Model
CLM2.0	Community Land Model version 2.0
CDF	Cumulative Distribution Function
CCSM	Community Climate System Model
CAM	Community Atmosphere Model
CTD	Computation Time Demand
DA	Data Assimilation
DA event	Data Assimilation event
DA frequency	Data Assimilation frequency
EnKF	Ensemble Kalman Filter
EKF	Extended Kalman Filter
EnGPF	Ensemble Gaussian Particle Filter
GPF	Gaussian Particle Filter
HBV	Hydrologiska Byråns Vattenbalansavdelning model
HBV-96	Modified version of the original HBV model
IS	Importance Sampling
KF	Kalman Filter
LSM	Land Surface Model
MC	Monte Carlo
MCMC	Markov Chain Monte Carlo
MulR	Multinomial Resampling
MH	Metropolis Hastling
NSE	Nash-Sutcliffe Efficiency
NRR	Normalized Root Mean Square Error Radio
NCAR	US National Center for Atmospheric Research
PF	Particle Filter
pdf	probability density function

Notation	Description
RM	Resample Move
RBE	Recursive Bayesian Estimation
RMSE	Root Mean Square Error
RR	Rainfall-Runoff model
RPF	Regularized Particle Filter
ResR	Residual Resampling
SSM	State Space Model
SMC	Sequential Monte Carlo
SIS	Sequential Importance Sampling
SCE-UA	Shuffled Complex Evolution
SPF	Standard Particle Filter
SysR	Systematic Resampling
StrR	Stratified Resampling
SPF-RM	Standard Particle Filter-Resample Move
SVATS	Soil-Vegetation-Atmosphere Transfer Scheme
UPF	Unscented Particle Filter
UKF	Unscented Kalman Filter

Chapter 1

Introduction

1.1 Data assimilation in hydrology

It is widely recognized that hydrologic models are useful tools for a number of purposes, ranging from flood forecasting (Andersson, 1992) to numerical weather prediction and climate studies (Zhang et al., 2008). Due to uncertainties in the meteorological forcings and model parameters, and errors or oversimplifications in the model physics, these models are always prone to a certain level of uncertainty. One way to reduce the predictive uncertainty of hydrologic models is to regularly update these models using externally obtained data sets, which is commonly referred to as Data Assimilation (DA). The improvement of hydrologic model results through the assimilation of soil moisture data has been the subject of numerous studies (Entekhabi et al., 1994; Walker et al., 2002; Pauwels et al., 2002; De Lannoy et al., 2007a). The underlying idea of data assimilation is to calculate a weighted average between the observations and the model results. The simplest way to perform this is to simply replace the model results by the observations, which is defined as direct insertion (Heathman et al., 2003). More advanced assimilation methods include nudging of the model results to the observations (Houser et al., 1998; Pauwels et al., 2001; Paniconi et al., 2002) and optimal interpolation (Seuffert et al., 2004). These techniques are in fact simplifications of the Kalman filter (Kalman, 1960), in which the model error is calculated explicitly throughout the simulation.

Originally developed for linear systems, and later extended for nonlinear systems, a great deal of attention has been paid to the Kalman filter as a methodology for hydrologic data assimilation. The Kalman filter is a square error estimator for linear systems. In his seminal paper, Kalman (1960) used the state space representation in order to generalize the application to any kind of linear system. It was possible to extend the application of

the filter to different systems and to develop nonlinear versions from the original Kalman filter, such as the extended Kalman filter (Bellantoni and Dodge, 1967; Hoeben and Troch, 2000), unscented Kalman filter (Wan and Van Der Merwe, 2000) and the Ensemble Kalman Filter (EnKF) (Evensen, 1994). The extended Kalman filter, in which the forecast error covariance is calculated through a linearization of the model, and the ensemble Kalman filter, in which this model error covariance is calculated using the spread of an ensemble of model realizations, have been intercompared by Reichle et al. (2002).

At this point, it can be argued that the ensemble Kalman filter is the most frequently used assimilation method in hydrology. A variation to this method is the ensemble Kalman smoother (Dunne and Entekhabi, 2005), in which observations that are distributed in time are used to update the model state variables. This method is comparable to variational assimilation (Caparrini et al., 2004), in which observations within a predefined window are used to estimate the initial state variables. One problem with the frequently used ensemble Kalman filter is the underlying assumption of Gaussianity of both the forecast and observation error structure. As it is evident that this assumption is not realistic for hydrologic systems, assimilation methods have been developed that relax this assumption.

1.2 Data assimilation in flood forecasting systems

Every year, human and economic losses are reported all around the world due to the presence of floods. Therefore, the scientific community actively is investing in improving the current flood forecasting systems.

Conceptual rainfall-runoff models are an important component in operational flood forecasting systems. Generally, these models represent the study area by a number of water reservoirs through which different inflows and outflows (for example infiltration, evapotranspiration, discharge) interact dynamically. Examples of such models are the Hydrologiska Byrans Vattenbalansavdelning (HBV) (Lindström et al., 1997) model and the Probability Distributed Model (PDM) (Moore, 2007) or variations derived from these models. From a technical point of view, the simplicity of conceptual models is an advantage which offers flexibility in the implementation. However, the identification of the model parameters that lead to realistic model predictions is a complex task. Moreover, the uncertainties in the forcings, model parameters and the simplifications in the model physics affect the overall performance of the conceptual model (Kavetski et al., 2006). One way to reduce the predictive uncertainty of conceptual hydrologic models is the use of data assimilation to regularly update models using externally obtained

data sets (Vrugt et al., 2006; Moradkhani and Sorooshian, 2008a). Sequential data assimilation is nowadays also a key component in flood forecasting systems.

1.3 New trends in hydrologic data assimilation

One method that is receiving increasing attention in hydrology is the particle filter. In the particle filter methodology, the posterior of interest is described by the point mass approximation allowing for the representation of any kind of distribution. In other words, the assumption of Gaussian distributions, which is held in the application of the Kalman filter, is relaxed when using particle filters. This method has been used to assimilate discharge records into conceptual rainfall-runoff models (Moradkhani et al., 2005; Weerts and El Serafy, 2006) and to assimilate water stage records into hydraulic models (Matgen et al., 2010; Giustarini et al., 2011). This method has also been used for the assimilation of soil moisture data (Plaza et al., 2012), for the estimation of model parameters (Montzka et al., 2011), and the estimation of root-zone soil moisture conditions (Nagarajan et al., 2010). All these studies share a similar implementation of the particle filter which is known as the generic particle filter or the standard particle filter. The standard particle filter simplifies the computation of the importance weights allowing for a straightforward implementation. However, this simplification could affect the overall performance of the particle filter mainly when the observation error is small. In Weerts and El Serafy (2006), the EnKF and the standard particle filter are intercompared leading to the conclusion that the EnKF is more robust with respect to forecast and observation errors. Other studies using the particle filter are discussed in Leisenring and Moradkhani (2011); DeChant and Moradkhani (2012); Leisenring and Moradkhani (2012) and Liu et al. (2012).

Recently, the standard particle filter has been applied in combination with the Bayesian model averaging approach in order to update the model weight at each assimilation time step (Parrish et al., 2012). In the same context of model selection, particle Markov Chain Monte Carlo (MCMC) methods (Andrieu et al., 2010) have been used (Rings et al., 2012; Vrugt et al., 2012) in more sophisticated implementations of the particle filter. Moradkhani et al. (2012) reported the increase of the effectiveness of the standard particle filter by using MCMC moves in a joint state-parameter estimation study. According to Moradkhani et al. (2005), Nagarajan et al. (2010), and Montzka et al. (2011), it is clear that the trend towards the application of particle filters is not limited to only the state estimation problem, but it can also be used for the identification of model parameter values, by exploiting the advantage of the flexible structure of the particle filter algorithms.

1.4 Objectives

The work presented in this manuscript focuses on the two following main objectives.

- The first general goal of this work is the improvement of flood forecasting systems by adopting particle filter based assimilation methods. The hydrologic modeling framework involves the selection of the hydrologic model to be used and the implementation of the data assimilation method. In this context, a lumped conceptual model is used. With respect to the data assimilation method, an exploration of two possible options which can lead to an improvement in the operation of the particle filter when state estimation is performed in rainfall-runoff models is conducted. More specifically, a resample-move step based on MCMC methods is included in the standard particle filter in order to improve the spread of the particles. The second alternative consists of the enhancement of the importance sampling step in the Gaussian particle filter (Kotecha and Djuric, 2003a) by considering a posterior estimate from an EnKF to generate the importance density function. The characteristics of the proposed techniques are studied in a synthetic experiment where artificial discharge records are assimilated into a conceptual rainfall-runoff model. The methodologies are assessed by the assimilation of in-situ observed discharge data. A comparison is carried out between the proposed techniques, the EnKF and the standard particle filter.

- The second objective is to test the particle filter as an assimilation method when sparse soil moisture data are available at certain time instants and at different space-scales, aiming at a correction of the modeled output flows. For this, a proof of concept study is carried out with a distributed physically-based hydrologic model. Specifically, state and parameter estimation are performed within the framework of the particle filter, aiming at an improvement of the model performance in terms of both soil moisture and discharge, through the assimilation of soil moisture data. Moreover, instead of estimating all the model parameters, we propose a methodology where a limited model parameter set is used. Dual or joint estimation has been widely studied using either the Kalman filter (Moradkhani et al., 2005a; Hendricks Franssen and Kinzelbach, 2008; Wang et al., 2009) or recently, the particle filter (Moradkhani et al., 2005; Nagarajan et al., 2010; Montzka et al., 2011). The approach presented here differs from previous state-parameter estimation studies in the objective. More specifically, the particle filter with parameter resampling is applied aiming at an improvement of the modeled discharge as a result of soil moisture assimilation, and

the parameter values are not estimated explicitly.

In this thesis, some variants of the original particle filtering method are introduced. The variant versions are presented separetaly according to their application thus either applied to conceptual or physically-based hydrologic models. However, it is clear that this work contributes to the ongoing research of improving sequential data assimilation methods under nonlinear and possibly non-Gaussian conditions.

1.5 Thesis outline

Figure 1.1 shows a general overview of the manuscript. The methodologies used and developed in this dissertation are presented in chapter 4. Chapters 5 and 6 address the assessment of the DA methodologies in two hydrologic estimation problems. A detailed description of each chapter is given below.

Chapter 2: Basic concepts in state estimation and data assimilation

The aim of this chapter is to introduce the reader with some fundamental concepts of system theory along with a brief presentation of the concept of data assimilation. This chapter describes deterministic and stochastic systems, the representation of these systems and also describes a general framework for the derivation of the estimation methods. The framework is the recursive Bayesian approach.

Chapter 3: Kalman filtering

In this chapter, the linear Kalman filter and the variants for nonlinear systems are presented. The presentation of these techniques is useful in the understanding and presentation of the ensemble Kalman filter which is nowadays the most frequently used assimilation method in the Geosciences. Although this thesis deals with the application and development of particle filter based assimilation methods, the usefulness of the presentation of the ensemble Kalman filter is twofold: the particle filter based methods are intercompared to the ensemble Kalman filter and the development of the ensemble Gaussian particle filter (EnGPF). EnGPF is presented in Chapter 4 and it is based on the ensemble Kalman filter theory.

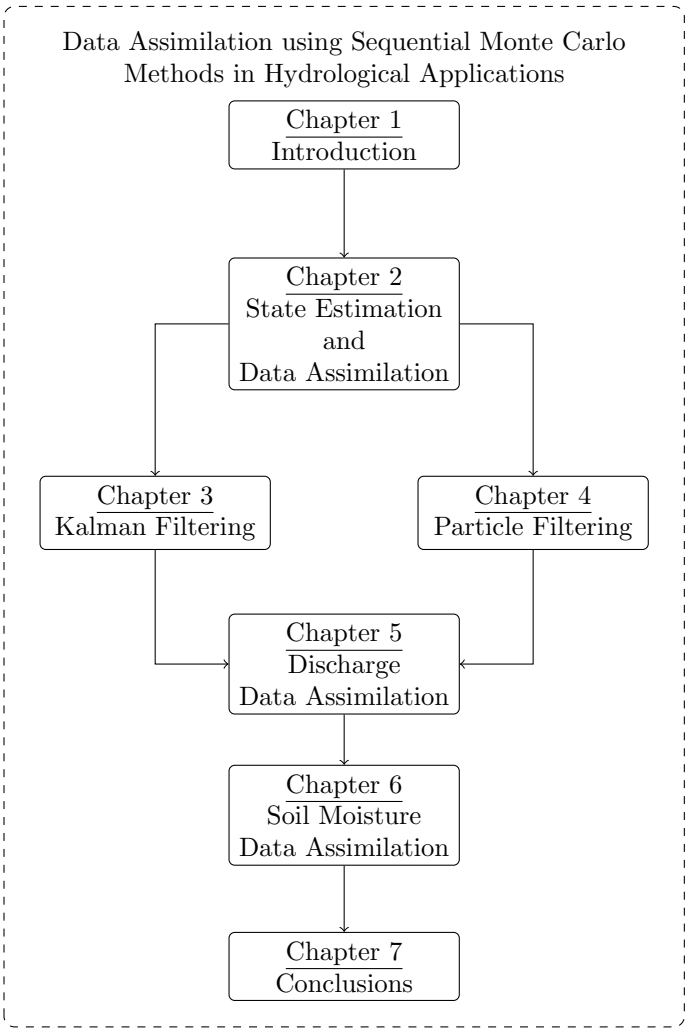


Figure 1.1: Outline of the manuscript

Chapter 4: Particle filtering

Chapter 4 covers most of the theory related to particle filters. The derivation of the formulation of the Kalman filter from the perspective of the recursive Bayesian estimation theory is useful for a better understanding of the particle filter which is naturally derived from this approach. In this chapter, the formulation of the ensemble Gaussian particle filter will be developed. The development in this chapter and the further assessment in chapter 5 of the ensemble Gaussian particle filter is one of the main contributions of this dissertation.

Chapter 5: Assimilation of discharge data

Chapter 5 is dedicated to exploring the performance of the data assimilation methods introduced in Chapters 3 and 4 when the methods are applied to a hydrologic assimilation problem. The assimilation problem consists in the assessment of the predictive uncertainty of modeled discharge (output flow) by the assimilation of discharge data records. The hydrologic modeling framework involves the selection, application and calibration of the hydrologic model and the selection and implementation of the assimilation methods. All these components of the hydrologic modeling cycle are presented in this chapter. A comparative study is performed aiming to assess the functioning of the novel ensemble Gaussian particle filter introduced in chapter 4.

Chapter 6: Assimilation of soil moisture data

This chapter is dedicated to the implementation of a particle filter based assimilation methodology with a distributed hydrologic model. In this chapter, state and parameter estimation with the standard implementation of the particle filter is presented and discussed. The methodology has not been previously presented in Chapter 4 since dual estimation is applied to the specific study case in this chapter. Although join or dual estimation with the particle filter in hydrologic problems is not a new topic, the novel contribution of this dissertation strives on the application and assessment of the methodology in distributed hydrologic systems.

Chapter 7: Conclusions and future directions

In this chapter the main conclusions of this thesis, which is related to the exploration of particle filter based assimilation methods in hydrologic systems, are elaborated and some further research perspectives are considered.

1.6 Main contributions

The original contributions of this dissertation are listed below.

- In chapter 3, the ensemble Kalman filter is derived by using the recursive Bayesian approach. The way how the methodologies are derived allows for a comparison to all kind of particle filters.
- In chapter 4, the particle filters are prensented. The main contribution in this section consist in the presentation of the Gaussian particle filter and the development of the Ensemble Gaussian particle filter (EnGPF). In chapter 4, the formulation of the EnGPF is derived while in chapter 5, the properties of the EnGPF are analyzed through a comparative study between the standard implementations of the ensemble Kalman filter and particle filter.
- In chapter 5, the potential use of the ensemble Gaussian particle filter in a nonlinear/non-Gaussian scenario is explored. In this chapter, the EnGPF is compared to the particle filter with MCMC move steps when both are applied to a hydrologic estimation problem. The particle filter with MCMC move step is the state of the art filter in nonlinear/non-Gaussian estimation problems.
- In chapter 6, the particle filter is applied to a joint state-parameter estimation problem. The main contribution concerns the improvement obtained by the implicit identification of optimal model parameter values and the general correction of the modeled hydrologic variables.

1.7 Published results

Some chapters of this manuscript are based on parts of the following publications:

- Plaza Guingla, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., and Pauwels, V. R. N. The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. *Hydrology and Earth System Sciences.*, 16(2):375–390, doi:10.5194/hess-16-375-2012, 2012.
- Plaza Guingla, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., and Pauwels, V. R. N. Improving particle filters in rainfall-runoff models: Application of the resample-move step and the ensemble Gaussian particle filter. *Water Resources Research*, 49, doi:10.1002/wrcr.20291, 2013.

- Plaza, D. A., De Keyser, R., and Pauwels, V. R. N. From the ensemble Kalman filter to the particle filter: a comparative study in rainfall-runoff models. In: *Proceedings of the 18th World Congress of the International Federation of Automatic Control (IFAC 2011)*, doi:10.3182/20110828-6-IT-1002.02326, Milan, Italy, August 28–September 2, 2011.

Additionally the author of this manuscript contributed to the following publications related to the application of the particle filter in hydraulic and hydrologic modeling:

- Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Plaza Guingla, D., Pauwels, V. R. N., De Lannoy, G. J. M., De Keyser, R., and Savenije H. H. G. Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: proof of concept. *Hydrology and Earth System Sciences*, 14(9), 17731785, doi:10.5194/hess-14-1773-2010, 2010.

- Giustarini, L., Matgen, P., Hostache, R., Montanari, M., Plaza, D., Pauwels, V. R. N., De Lannoy, G. J. M., De Keyser, R., Pfister, L., Hoffmann, L., and Savenije, H. H. G. Assimilating SAR-derived water level data into a hydraulic model: a case study. *Hydrology and Earth System Sciences*, 15, 2349-2365, doi:10.5194/hess-15-2349- 2011, 2011.

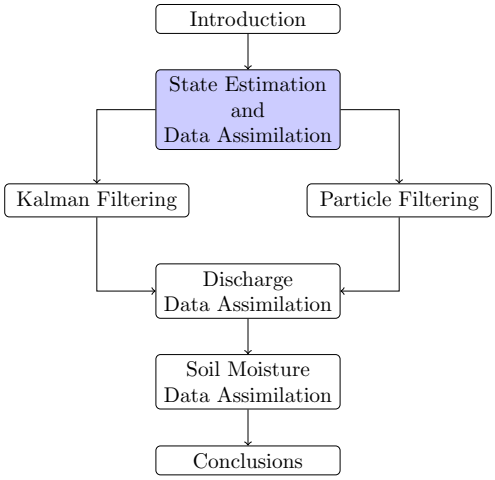
- Matgen, P., Fenicia, F., Heitz, S., Plaza, D., De Keyser, R., Pauwels, V. R.N., Wagner, W., and Savenije, H. H. G. Can ASCAT-derived soil wetness indices reduce predictive uncertainty in well-gauged areas? A comparison with in-situ observed soil moisture in an assimilation application. *Advances in Water Resources*, Vol 44, Pages 49-65, doi:10.1016/j.advwatres.2012.03.022, 2012.

- Matgen, P., Giustarini, L., Hostache, R., Plaza Guingla, D. A., Pauwels, V. R. N., and Verhoest, N. Joint assimilation of remote sensing-derived water stage and soil moisture data into coupled Hydrological-Hydraulic models. In *IEEE International Geosciences and Remote Sensing Symposium (IGARSS 2012)*, Munich, Germany, July 22–27, 2012.

- Pauwels, V. R. N., Lievens, H., Verhoest, N. E. C., De Lannoy, G., Plaza Guingla, D. A., van den Berg, M. J., Kerr, Y., Al Bitar, A., Merlin, O., Cabot, F., Gascoin, S., Wood, E., Pan, M., Sahoo, A., Walker, J., Dumedah, G. and Drusch, M. On the assimilation of multi-angular SMOS data into a coupled land surface and radiative transfer model for improving surface water management. In: *IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2012).* Munich, Germany, July 22–27, 2012.

Chapter 2

Basic concepts in state estimation and data assimilation



2.1 Introduction

The topic of interest in this manuscript is related to observing nature. There are different approaches for studying nature. In this manuscript, nature is studied based on the theory of systems. In systems theory, nature can be interpreted as a system and data collected from nature correspond to signals. In this context, the observed nature is studied within a well defined theoretical framework. The aim of this chapter is to present fundamental

concepts of systems theory as a preamble to the presentation of the methodologies. More specific, deterministic and stochastic systems are revisited with an emphasis in the state-space representation of the systems. Additionally, a mechanism to develop recursive state estimation methods which is based on the Bayes' theorem and is known as Recursive Bayesian Estimation is presented. Finally, the concept of data assimilation is presented. The presentation focuses on the magnification of the relation between state estimation and data assimilation.

2.2 Deterministic systems

Basically, a system studies the transformation of input signals into output signals. There are different approaches in the representation of a system. One of these approaches is mathematical modeling. The derivation of a mathematical model encompasses two methodologies which can be applied separately or combined in the study of a system.

First, a model can be obtained by the theoretical analysis of the system. In this analysis, the laws of conservation of mass, energy and momentum are applied and studied for the respective system and subsystems. As a result of the analysis, linear or non-linear, ordinary or partial differential equations form the mathematical model.

A second approach in the derivation of a mathematical model is through experimental analysis in which a model structure is proposed. The parameters of the model are estimated by surveying the behavior of the system under specific conditions when the model structure is known (parametric identification methods). On the other hand, non-parametric identification methods are used when the model structure is unknown.

In this chapter, deterministic and stochastic models are revisited as background and motivation for the discussion of state estimation techniques in chapters 3 and 4.

2.3 Deterministic discrete systems

A model is essentially a set of differential equations of different orders, which describes the relationship between the input $\mathbf{u}$ (dimension m_u) and the output $\mathbf{y}$ (dimension m_y) signals. The m_x-th order differential equations can be transformed to a system of m_x first-order differential equations. This way of representation is commonly known as the state-space representation of the system. In general, state-space models are classified into continuous time

models and discrete time models. For implementation in digital computers, continuous time models are transformed to discrete time models. In this research, discrete time models will be treated due to the implementation of the models in computers and the nature of the measurements which is discrete.

Linear state-space model

Although the focus of this study is on non-linear systems, linear systems are presented as a background for the presentation of the discrete linear Kalman filter in chapter 3.

The state-space representation for a discrete linear time-varying system is the following:

$$\mathbf{x}_t = \mathbf{F}_{t,t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} \tag{2.1}$$

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t \tag{2.2}$$

Equation 2.1 is a set of first order difference equations often referred to as the process model with $\mathbf{x}_t$ the m_x-dimensional state vector of the dynamic system. The state space of the system is given by the m_x-dimensional domain of the state vector. $\mathbf{F}_{t,t-1}$ is the $(m_x \times m_x)$-dimensional linear transition matrix which maps the state vector from discrete time step $t-1$ to t and $\mathbf{B}_{t-1}$ an $(m_x \times m_u)$-dimensional matrix relating the input to the state. The projection of the state space to the observation space is performed in Equation 2.2 with $\mathbf{H}_t$ the $(m_y \times m_x)$-dimensional observation matrix.

In control theory, the concept of controllability of linear systems is of major importance. Controllability concerns the manipulation of the inputs in order to move the system to a desired output. On the other hand, estimation theory concerns about the possibility of observing the system states based on input and output data. Moreover, according to the principle of separation, an optimal feedback controller for a linear system can be designed by separating the problem into a state estimation problem, where the optimal estimate of the state is encountered, followed by the development of an optimal deterministic controller for the system. In this sense, Kalman (Kalman, 1960) introduced the concept of observability and the principles of duality and separability of the estimation and control problems for linear dynamic systems.

The scope of this dissertation is limited to the estimation problem. The main reason for this limit in the research is that a hydrologic system is a representation of the nature and the input signals corresponds to meteorological forcings which can be considered as uncontrollable. Therefore, the aim of hydrologic studies is to model land processes as accurate as possible

in order to obtain a complete idea of the system rather than controlling nature directly.

Non-linear state-space model

The state-space representation of a non-linear system is given by the following equations:

$$\mathbf{x}_t = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) \tag{2.3}$$
$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t) \tag{2.4}$$

with $\mathbf{f}_{t,t-1}()$ the non-linear function relating the state propagation through time and $\mathbf{h}_t()$ the non-linear observation function. Most estimation problems are nonlinear. Thus, the non-linear models are linearized for the application of linear control or estimation theory. A linearization approach is by using Taylor series expansion around a nominal state trajectory, i.e., $\mathbf{x}_t = \mathbf{x}_t^*$ and $\mathbf{u}_t = \mathbf{u}_t^*$. The expression for the linearized system is the following:

$$\mathbf{x}_t - \mathbf{x}_t^* = \mathbf{F}_{t,t-1}[\mathbf{x}_{t-1} - \mathbf{x}_{t-1}^*] + \mathbf{B}_{t-1}[\mathbf{u}_{t-1} - \mathbf{u}_{t-1}^*] \tag{2.5}$$
$$\mathbf{y}_t - \mathbf{y}_t^* = \mathbf{H}_t[\mathbf{x}_t - \mathbf{x}_t^*] \tag{2.6}$$

where $\mathbf{F}_{t,t-1}$, $\mathbf{B}_t$ and $\mathbf{H}_t$ are Jacobian matrices with dimensions $(m_x \times m_x)$, $(m_x \times m_u)$ and $(m_y \times m_x)$, respectively. The matrices are presented as follows.

$$\mathbf{F}_{t,t-1} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_{m_x}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{m_x}}{\partial x_1} & \cdots & \frac{\partial f_{m_x}}{\partial x_{m_x}} \end{bmatrix} |(\mathbf{x}_t^*, \mathbf{u}_t^*) \tag{2.7}$$

$$\mathbf{B}_t = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \cdots & \frac{\partial f_1}{\partial u_{m_x}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{m_x}}{\partial u_1} & \cdots & \frac{\partial f_{m_x}}{\partial u_{m_y}} \end{bmatrix} |(\mathbf{x}_t^*, \mathbf{u}_t^*) \tag{2.8}$$

$$\mathbf{H}_t = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_{m_x}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_{m_y}}{\partial x_1} & \cdots & \frac{\partial h_{m_y}}{\partial x_{m_x}} \end{bmatrix} |(\mathbf{x}_t^*) \tag{2.9}$$

where $\mathbf{F}_{t,t-1}$ and $\mathbf{B}_t$ are the derivatives of $\mathbf{f}_{t,t-1}$ with respect to $\mathbf{x}$ and $\mathbf{u}$, evaluated at $\mathbf{x}_t^*$ and $\mathbf{u}_t^*$, respectively, while $\mathbf{H}_t$ is the derivative of $\mathbf{h}_t$ with respect to $\mathbf{x}$, evaluated for $\mathbf{x}_t^*$.

It is important to remark the fact that in Equations 2.5 and 2.6 only the first order derivatives from the Taylor series are taken into account in the linearization. For a more precise approximation of highly non-linear functions,

the second order derivatives (Hessian matrices) should be considered. However, the computation of the Hessian matrices in high dimensional systems is impractical since the computational cost is very expensive.

2.4 Stochastic systems and state estimation

The terms stochastic systems and state estimation are mutually related, in this section the discussion about these concepts is extended. Based on the fact that most of the systems have certain random nature, the theory of random processes is used as a complement for deterministic systems in order to properly describe the uncertainties encountered in real systems. Random processes or stochastic systems represent the evolution over time of the uncertainty of our knowledge about physical systems. The representation of a physical system by a stochastic model includes the effects of measurements from the real system and the effects of uncertainties about the observation system and the dynamic system. For applications in state inference, the stochastic process has to satisfy the Markov property. A process satisfies the Markov property when it is possible to make future predictions of the system based solely on the information from the present state.

State-space modeling for stochastic discrete systems is performed through stochastic difference equations. These equations are mainly used in the study of Gaussian processes. In Gaussian processes, the properties of uncertain dynamic systems are characterized by statistical parameters such as mean, correlation and covariance, i.e., the first two moments of the probability density function. This allows for a finite representation of the problem which is important in the derivation of the optimal analytical solution to the estimation problem. A general representation of the state-space model for all non-Gaussian processes is through conditional probability density functions. However, the probability density functions have to be approximated in order to obtain an approximated solution to the estimation problem.

2.4.1 Linear state-space model

The linear system in discrete time, described by stochastic difference equations, is given by:

$$\mathbf{x}_t = \mathbf{F}_{t,t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{G}_{t-1}\mathbf{w}_{t-1} \tag{2.10}$$

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t \tag{2.11}$$

with $\mathbf{w}_t$ the process noise, and $\mathbf{G}_t$ a $(m_x \times m_x)$ dimensional matrix relating the process noise to the state and $\mathbf{v}_t$ the observation noise. $\mathbf{w}_t$ and $\mathbf{v}_t$ are

random processes and the stochastic difference equations become difference equations if the noises are removed. For a deterministic linear system, the solution of the difference equations is $\mathbf{x}_t$. On the other hand, for stochastic linear systems, the pdf of $\mathbf{x}_t$ gives a complete description of the state.

Gaussian processes

A process is referred to as Gaussian or normal when the probability distribution functions (pdfs) of the random state variables correspond to Gaussian distributions. Gaussian processes have some useful properties: any linear transformation of jointly Gaussian processes results in another Gaussian process, and the most important feature in the development of the Kalman filter is that all the statistics of a Gaussian process are completely determined by its first- and second-order statistics. These 2 moments can be directly obtained from Equation 2.10 with zero mean white Gaussian noise for $\mathbf{w}_t$, i.e. $E[\mathbf{w}_t] = 0$ and $E[\mathbf{w}_t\mathbf{w}_t^T] = \mathbf{Q}_t$. The 2 moments are presented as follows:

$$\hat{\mathbf{x}}_t = \mathbf{F}_{t,t-1}\hat{\mathbf{x}}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} \tag{2.12}$$

$$\mathbf{P}_t = \mathbf{F}_{t,t-1}\mathbf{P}_{t-1}\mathbf{F}_{t,t-1}^T + \mathbf{G}_{t-1}\mathbf{Q}_{t-1}\mathbf{G}_{t-1}^T \tag{2.13}$$

with $\hat{\mathbf{x}}_t$ an estimate of the true $\mathbf{x}_t$ and $\mathbf{P}_t$ is the covariance matrix which is an indicator of the uncertainty of $\mathbf{x}_t$.

2.4.2 Non-linear state-space model

A non-linear stochastic system in discrete time, can be represented as follows:

$$\mathbf{x}_t = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}, \mathbf{w}_{t-1}) \tag{2.14}$$

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{v}_t) \tag{2.15}$$

This representation corresponds to a functional representation since $\mathbf{f}()$ and $\mathbf{h}()$ are mathematical functions.

Additive Gaussian process and observation noise

By considering specific additive Gaussian process $\mathbf{w}_t$ and observation $\mathbf{v}$ noises, Equations 2.14 and 2.15 become:

$$\mathbf{x}_t = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}, \mathbf{u}_{t-1}) + \mathbf{G}_{t-1}(\mathbf{x}_{t-1})\mathbf{w}_{t-1} \tag{2.16}$$

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t \tag{2.17}$$

The evolution in time of the pdf of nonlinear stochastic processes is described by the Fokker-Planck or Kolgomorov's forward equation. Moreover, for nonlinear processes the pdfs are not sufficiently characterized by the first- and second-order moments of the pdf such as in the case of Gaussian processes. However, the first 2 moments give valuable information on the mean path and uncertainty for linearized systems.

In Jazwinski (1970), the expression of the first moment for a non-linear model is calculated by:

$$\hat{\mathbf{x}}_t = \mathbf{f}_{t,t-1}(\hat{\mathbf{x}}_{t-1}, \mathbf{u}_{t-1}) + \frac{1}{2}\mathcal{F}''_{t,t-1}(\hat{\mathbf{x}}_{t-1}, \mathbf{u}_{t-1})vec[\mathbf{P}_{t-1}] \tag{2.18}$$

with $\mathcal{F}''_{t,t-1}$ the Hessian matrix of the system function $\mathbf{f}_{t,t-1}$ and $vec[\ ]$ standing for a vector constructed by concatenation of the different columns in a one vector. It is clear from Equation 2.18 that the mean depends on the variance.

2.4.3 State-space model represented by pdfs

For the representation of the state-space model through pdfs, the evolution of the state over time is considered to be a stochastic process. Specifically, the dynamics are modeled according to a Markov process, which is a stochastic process such that given the present state of the process, the future evolution of the process is independent of its past. Based on the Markov property, the evolution of the state can be written as:

$$\mathbf{x}_{t+1}|\mathbf{x}_t \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t) \tag{2.19}$$

where the symbol $\sim$ denotes distributed according to, $p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{u}_t)$ is the conditional probability of moving from $\mathbf{x}_t$ to $\mathbf{x}_{t+1}$. This distribution is referred to as the state transition pdf. $\mathbf{u}_t$ corresponds to the inputs or forcings. In this probabilistic framework, $\mathbf{u}_t$ is considered as a random process.

Regularly, the state of the system cannot be observed completely, but a part or a function of the state is observed. This type of processes are referred to as Hidden Markov Models (HMM). Information about the underlying state process is obtained indirectly via the observation (measurement) process which describes the relationship between the state and the measurements. In terms of probability, the observation process is given by:

$$\mathbf{y}_t|\mathbf{x}_t \sim p(\mathbf{y}_t|\mathbf{x}_t) \tag{2.20}$$

where $p(\mathbf{y}_t|\mathbf{x}_t)$ is the conditional pdf of obtaining a certain measurement $\mathbf{y}_t$ given a specific state $\mathbf{x}_t$. This distribution is referred to as the observation pdf. Additional to the state transition and observation pdfs, a third pdf

should be considered which corresponds to the initial distribution of the state ($\mathbf{x}_0 \sim p(\mathbf{x}_0)$).

Equations (2.19) and (2.20) represents a nonlinear system in which the parameters θ of the model are explicitly known. Therefore the inference problem in this study corresponds to the state inference problem unless the contrary is specified.

2.5 Recursive Bayesian estimation

Recursive Bayesian Estimation (RBE) provides a probabilistic framework which recursively solves the state inference problem in terms of conditional probability density functions. State inference, which is also known as State Estimation or as Data Assimilation in system theory (Simon, 2006) and in atmospheric and earth sciences (Lahoz et al., 2010), respectively, is the estimation of the current state value or the sequence of values up to the present time. This estimation is based on the sequence of measurements observed up to the present time.

In the RBE context, inference about $\mathbf{x}_{1:t}$ given $\mathbf{y}_{1:t}$ relies upon the conditional posterior pdf:

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t})p(\mathbf{x}_{1:t})}{p(\mathbf{y}_{1:t})} \tag{2.21}$$

where

$$p(\mathbf{y}_{1:t}) = \int p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t})p(\mathbf{x}_{1:t})d\mathbf{x}_{1:t} \tag{2.22}$$

The recursive expression of equation (2.21) is the following:

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})\frac{p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{y}_t|\mathbf{x}_t)}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \tag{2.23}$$

where

$$p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{y}_t|\mathbf{x}_t)d\mathbf{x}_{t-1:t} \tag{2.24}$$

The problem of solving equation (2.23) is referred to as the optimal filtering problem. The posterior distribution $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ is also known as the smoothing distribution. Equation (2.23) can be decomposed into two parts or steps: The time update step (prediction or forecast) and the measurement update step(correction or analysis).

The time update step corresponds to the estimation of the distribution of the first t states given only $t-1$ observations.

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t-1}) = p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{2.25}$$

The measurement update step involves the correction of the predicted distribution based on the information from the current measurement.

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t-1})p(\mathbf{y}_t|\mathbf{x}_t)}{\int p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t-1})p(\mathbf{y}_t|\mathbf{x}_t)d\mathbf{x}_{1:t}} \tag{2.26}$$

Other distributions such as the smoothing distributions and the prediction distributions are also of interest. In smoothing, the aim is to estimate the distribution of some sequence of states conditionally on knowledge of the observation up to some stage in the future. Formally, smoothing is the estimation of $p(\mathbf{x}_{l:k}|\mathbf{y}_{1:t})$ when $l \leq k \leq t$ and such distribution is obtained from the principle smoothing distribution as follows:

$$p(\mathbf{x}_{l:k}|\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})d\mathbf{x}_{1:l-1}d\mathbf{x}_{k+1:t} \tag{2.27}$$

In prediction, the distribution of some group of future states is estimated conditionally on knowledge of the observations up to the present time. Prediction can be viewed as the estimation of $p(\mathbf{x}_{l:k}|\mathbf{y}_{1:t})$ when $t \leq k$ and $l \leq k$. For $l = 1$, the prediction distribution is computed as follows:

$$p(\mathbf{x}_{1:k}|\mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) \prod_{j=t+1}^{k} p_j(\mathbf{x}_j|\mathbf{x}_{j-1}) \tag{2.28}$$

The posterior of interest in this work is the marginal distribution which is known as the filtering distribution. This distribution is obtained by integrating out $\mathbf{x}_{1:t-1}$ from 2.25 and considering the Markov property. The result of this integration is the Bayes' filter which is presented in algorithm 1. A detailed derivation of the Bayes' filter is given in Appendix A.

Algorithm 1 presents a set of conditional pdfs which defines the conceptual Bayesian solution to the estimation problem. In the forecast step, the predictive distribution is obtained by a combination of our best knowledge of the system at present time through the model (transition pdf) and the knowledge of the system in the recent past which is given by the posterior pdf at time step $t-1$. Next, the filtering pdf is updated based on the information from a new measurement of the output at time t in the analysis step. The updated predictive pdf is referred to as the filtering pdf and it is the conditional posterior pdf of interest throughout this study.

Algorithm 1 Bayes' filter

For $t = 1$ to the number of time steps

1. Forecast step:

$$\overbrace{p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}^{\text{predictive}} = \int \overbrace{p(\mathbf{x}_t|\mathbf{x}_{t-1})}^{\text{transition}} p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}$$

2. Analysis step:

$$\overbrace{p(\mathbf{x}_t|\mathbf{y}_{1:t})}^{\text{filtering}} = \frac{\overbrace{p(\mathbf{y}_t|\mathbf{x}_t)}^{\text{likelihood}} p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) d\mathbf{x}_t}$$

Recursive Bayesian estimation methods, which are implementations of the theoretical Bayes' filter, can be classified into two groups: optimal and suboptimal estimation methods. Optimal filters are derived when the estimation problem is presented in a closed-form. Therefore, an analytical solution is tractable. From this group, the most representative method is the Kalman filter (Kalman, 1960) which is presented in chapter 3. The optimal solution does exist for linear Gaussian processes where the predictive and the filtering pdfs corresponds to Gaussian distributions at each time instant. In this case, the mean and covariance completely represent the two distributions.

On the other hand, for nonlinear stochastic processes, suboptimal algorithms have been developed in order to approximate the optimal Bayesian solution. In this case, the analytical solution of the estimation problem is intractable. One widely used suboptimal method in engineering is the extended Kalman filter (Simon, 2006, ch. 13) which approximates the Bayesian solution by a local linearization of the nonlinear system. For high dimensional systems, where the computational time demand is crucial, derivative-free local filters become more practical from the implementation point of view. Examples of this kind of filters are the Unscented Kalman filter (Julier et al., 2000) and the divided different filters (Ito and Xiong, 2000; Norgaard et al., 2000).

A set of suboptimal algorithms use Monte Carlo methods in order to approximate the integrals involved in the implementation of the Bayes' filter. Examples of this type of filters correspond to the Ensemble Kalman Filter (EnKF) (Evensen, 1994) and the Particle Filter (PF) (Gordon et al., 1993).

In the EnKF, the covariance matrix is approximated by the sample covariance of a set of random model realizations while the particle filter uses sequential Monte Carlo methods in order to approximate the distributions involved in the estimation by a point-mass representation.

2.6 Introduction to data assimilation

Data assimilation is a modern methodology which concerns the relation between natural data and dynamical computer models. The procedure consists of the combination or merging of general dynamics of a model with a set of observations. All dynamical models have modeling errors and all data sets are finite and to some extent limited by error bounds. Thus, the aim of data assimilation is to provide proper estimates of nature by combining observational data with dynamic models.

Data assimilation methods are classified as either sequential or retrospective methods. A sequential data assimilation method is an algorithm which make use of the information from the observations up to and including the current time when the system state is to be estimated. On the other hand, retrospective data assimilation algorithms incorporates not only observation from the past. In this sense, sequential DA methods are a combination of predictive and filtering algorithms, while a retrospective DA method is a smoothing algorithm.

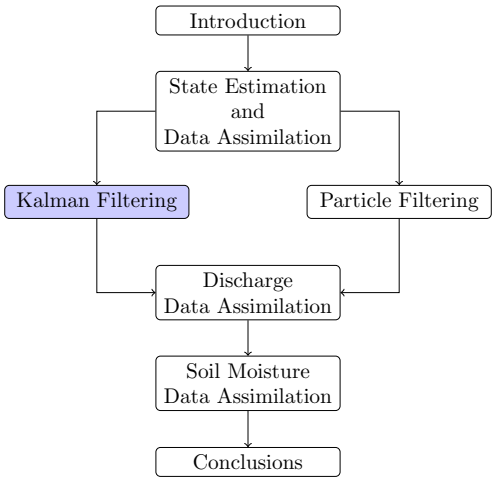
Data assimilation is applied in many fields of Geosciences, mainly in numerical weather forecasting (Anderson, 1996), oceanography (Oke et al., 2010) and hydrology (Lahoz et al., 2010). Also, in other fields, such as Bioinformatics (Nagasaki, 2006) and finance and econometrics (Wells, 1996). These dynamic systems have a large number of states and are highly nonlinear. Therefore, the RBE is a powerful tool in the development of DA methods. In DA theory, the optimal nonlinear filter can be obtained by following two steps: the forecast (background) step and the analysis step. Both steps are analogous as to the prediction (Equation 2.25) and correction (Equation 2.26) step, respectively.

2.7 Summary

Some concepts in Systems theory have been introduced in this chapter with an emphasis on mathematical modeling of systems. Additionally, the difference between deterministic and stochastic, continuous and discrete systems was explained. Hereafter, we deal with stochastic systems in discrete time. The state-space model is used to represent a system. In stochastic systems, the state of a system is a degree of knowledge about the system. Therefore, state estimation techniques should be applied in order to improve the degree of knowledge about the system. For this, observations plays an important role. Finally, a conceptual optimal filter was presented based on the recursive Bayesian estimation approach as well as some initial concepts in data assimilation.

Chapter 3

Kalman filtering



3.1 Introduction

The aim of this chapter is to present the formulation of the linear Kalman filter and the nonlinear versions of this filter such as the extended Kalman filter and the ensemble Kalman filter. The formulation is presented by following the recursive Bayesian approach. By following this approach, it is possible to demonstrate the different approximations of the predictive and filtering pdfs used for nonlinear systems.

3.2 The discrete Kalman Filter

3.2.1 Introduction

In this manuscript, the original Kalman filter for linear Gaussian processes is derived from a probabilistic framework which is the RBE. This approach allows for the derivation of the KF in terms of its pdfs rather than a direct analytical derivation, e.g. the least square estimation (Grewal and Andrews, 2001). More specifically, the type of system considered here is a linear stochastic dynamical system driven by Gaussian process and measurement noise. Moreover, these noises are white Gaussian, zero mean and uncorrelated for all t, with t the discrete time index. The state-space model derived from Equations (2.10) and (2.11) is presented as follows:

$$\mathbf{x}_t = \mathbf{F}_{t,t-1}\mathbf{x}_{t-1} + \mathbf{B}_{t-1}\mathbf{u}_{t-1} + \mathbf{w}_t \tag{3.1}$$
$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t \tag{3.2}$$

where the process and observation noises are distributed according to Gaussian distributions:

$$\mathbf{w}_t \sim \mathcal{N}(\mathbf{w}_t : \mathbf{0}, \mathbf{Q}_t) \tag{3.3}$$
$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{v}_t : \mathbf{0}, \mathbf{R}_t) \tag{3.4}$$

where the notation $\mathcal{N}(\mathbf{x}_t : \mu, \mathbf{\Sigma})$ is used for representing a Gaussian distribution with $\mathbf{x}$ the random vector with mean μ and covariance matrix $\mathbf{\Sigma}$. The statistical properties of the random vectors are the following:

$$E[\mathbf{w}_t] = \mathbf{0} \quad , \quad E[\mathbf{v}_t] = \mathbf{0}, \tag{3.5}$$
$$E[\mathbf{w}_t\mathbf{w}_{t+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{w}_t\mathbf{w}_t^T] = \mathbf{Q}_t, \tag{3.6}$$
$$E[\mathbf{v}_t\mathbf{v}_{t+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{v}_t\mathbf{v}_t^T] = \mathbf{R}_t, \tag{3.7}$$
$$E[\mathbf{w}_t\mathbf{v}_{t+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{w}_t\mathbf{v}_t^T] = \mathbf{0} \tag{3.8}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are positive definite covariances matrices for the process and observation noises, respectively, and k is a positive integer.

3.2.2 A Recursive Bayesian Estimation approach to the Kalman filter

The recursive algorithm is derived based on the available information at previous time steps. The following expression is true based on the Gaussian properties of the stochastic system under study.

$$p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_{t-1} : \mathbf{x}_{t-1}^a, \mathbf{P}_{t-1}^a) \tag{3.9}$$

where $\mathbf{x}_{t-1}$ is the true state vector, $\mathbf{x}^a_{t-1}$ and $\mathbf{P}^a_{t-1}$ are the state vector and the covariance matrix, obtained both from the analysis step at time step $t-1$. Although the analysis step has not been introduced yet, it is here considered that $\mathbf{x}^a_{t-1}$ and $\mathbf{P}^a_{t-1}$ are known. The analysis step is formally presented further.

Forecast step

According to the Bayes filter presented in Algorithm 1, the first step in the derivation of the optimal filter is the computation of the prior pdf at time step t:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \tag{3.10}$$

In Equation 3.10, the prior pdf is determined by the transition pdf and the filtering (posterior) pdf. In other words, the prior pdf is computed based on all the available information which is known beforehand, i.e. the information obtained from the model and the estimated information which is obtained one time step backward. The filtering pdf at time step $t-1$ was presented in Equation 3.9 and the transition pdf is calculated from 3.1 as follows:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_{t-1} : \mathbf{F}_{t,t-1}\bar{\mathbf{x}}_{t-1}, \mathbf{Q}_t) \tag{3.11}$$

where $\bar{\mathbf{x}}_{t-1}$ is the expectation of the state $E[\mathbf{x}_{t-1}]$. The term corresponding to the forcings in Equation 3.1 is not considered in the derivation for the sake of clarity. Moreover, the absence of this term does not affect the formulation thus this term is not considered in the remaining chapters. It can be shown that the predictive pdf (Equation 3.10) is Gaussian distributed when the Gaussian densities in Equations 3.9 and 3.11 are substituted by the exponential function of the pdfs and after some matrix algebra operations. The statistics of the predictive pdf are shown below.

$$\mathbf{x}^f_t = \mathbf{F}_{t,t-1}\mathbf{x}^a_{t-1}, \tag{3.12}$$

$$\mathbf{P}^f_t = \mathbf{F}_{t,t-1}\mathbf{P}^a_{t-1}\mathbf{F}^T_{t,t-1} + \mathbf{Q}_t \tag{3.13}$$

with $\mathbf{x}^f_t$ the forecast state vector and $\mathbf{P}^f_t$ the forecast error covariance matrix $\mathbf{P}^f_t = E[\mathbf{x}^f_t - \bar{\mathbf{x}}_t][\mathbf{x}^f_t - \bar{\mathbf{x}}_t]^T$.

Analysis step

In the analysis step, the prior pdf is updated with the last measurement which is obtained at time t. According to Algorithm 1, the update of the prior is given by Equation 3.14.

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \tag{3.14}$$

The prior is updated based on the information from the likelihood pdf. In this case the likelihood is Gaussian distributed, so that:

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t : \mathbf{H}_t\bar{\mathbf{x}}_t, \mathbf{R}_t) \tag{3.15}$$

Substituting Equation 3.15 into 3.14, the following expression is obtained:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{\mathcal{N}(\mathbf{y}_t : \mathbf{H}_t\bar{\mathbf{x}}_t, \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)}{\int \mathcal{N}(\mathbf{y}_t : \mathbf{H}_t\bar{\mathbf{x}}_t, \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)d\mathbf{x}_t} \tag{3.16}$$

It can be shown that the integration of the denominator in Equation 3.16 is Gaussian distributed, so that

$$\int \mathcal{N}(\mathbf{y}_t : \mathbf{H}_t\bar{\mathbf{x}}_t, \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)d\mathbf{x}_t = \mathcal{N}(\mathbf{y}_t; \mathbf{H}_t\mathbf{x}_t^f, \mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T + \mathbf{R}_t) \tag{3.17}$$

Here, with some algebraic operations, Equation 3.16 can be reduced to

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^a, \mathbf{P}_t^a) \tag{3.18}$$

where

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^f), \tag{3.19}$$

$$\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_t^f, \tag{3.20}$$

with

$$\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t^T[\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T + \mathbf{R}_t]^{-1} \tag{3.21}$$

Summary of the linear Kalman filter

The recursive steps of the Kalman filter for linear/Gaussian systems are summarized as follows.

Algorithm 2 is the optimal solution to the state estimation problem for Gaussian processes of the type described in Equations (3.1 - 3.8). The analytical solution of the predictive and filtering pdfs was tractable due to the full descriptions of these pdfs through their means and covariances. The updated state $\mathbf{x}_t^a$ is given by the addition of the forecast step $\mathbf{x}_t^f$ and the weighted residuals $(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^f)$. The residuals are weighted by the Kalman gain $\mathbf{K}_t$. For a proper computation of the Kalman gain, the error covariance

Algorithm 2 Discrete Kalman Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_t^f = \mathbf{F}_{t,t-1}\mathbf{x}_{t-1}^a$
   $\mathbf{P}_t^f = \mathbf{F}_{t,t-1}\mathbf{P}_{t-1}^a\mathbf{F}_{t,t-1}^T + \mathbf{Q}_t$
2. Analysis step:
   $\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t^T[\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T + \mathbf{R}_t]^{-1}$
   $\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t^f)$
   $\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_t^f$

matrices ($\mathbf{P}_t^a$, $\mathbf{P}_t^f$) should be symmetric and positive definite. However, in systems with a large number of states, these properties may not be preserved due to the finite computational precision. Thus, the performance of the filter can be improved by the adoption of the square root Kalman filter (Simon, 2006, ch. 6) which use the square root matrices of the error covariance matrices and the Kalman filter with fading memory (Simon, 2006, ch. 7) which use the principle of covariance inflation.

3.3 Extended Kalman Filter

3.3.1 Introduction

The observation and posterior estimation of states in nature is a complex task since most of the estimation problems are nonlinear. However, by considering smooth nonlinearities, it is possible to apply methods of linear estimation theory, e.g. Kalman filter, after the linearization of the system around a nominal trajectory. For some problems, the nominal trajectory is known beforehand. For these kind of systems, the estimation problem can often be effectively linearized about the nominal trajectory and the Kalman gains can be precomputed to relieve the real-time computational burden. On the other hand, when the nominal trajectory is not known beforehand, it can be defined at each time step as the current best estimate of the actual trajectory. The major disadvantage is the added computational cost of linearization about an unpredictable trajectory. This approach is called the Extended Kalman Filter.

The following nonlinear system is considered in the derivation of the EKF by using the RBE approach:

$$\mathbf{x}_t = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}) + \mathbf{w}_t \tag{3.22}$$

$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t \tag{3.23}$$

The distributions and statistics of the process and observation noises are presented as follows:

$$\begin{aligned}
\mathbf{w}_t &\sim \mathcal{N}(\mathbf{w}_t : \mathbf{0}, \mathbf{Q}_t) && (3.24)\\
\mathbf{v}_t &\sim \mathcal{N}(\mathbf{v}_t : \mathbf{0}, \mathbf{R}_t) && (3.25)\\
E[\mathbf{w}_t] &= \mathbf{0} \quad , \quad E[\mathbf{v}_t] = \mathbf{0}, && (3.26)\\
E[\mathbf{w}_t \mathbf{w}_{t+k}^T] &= \mathbf{0} \quad , \quad E[\mathbf{w}_t \mathbf{w}_t^T] = \mathbf{Q}_t, && (3.27)\\
E[\mathbf{v}_t \mathbf{v}_{t+k}^T] &= \mathbf{0} \quad , \quad E[\mathbf{v}_t \mathbf{v}_t^T] = \mathbf{R}_t, && (3.28)\\
E[\mathbf{w}_t \mathbf{v}_{t+k}^T] &= \mathbf{0} \quad , \quad E[\mathbf{w}_t \mathbf{v}_t^T] = \mathbf{0} && (3.29)
\end{aligned}$$

the noises are white, Gaussian distributed with zero mean and uncorrelated $\forall t$ with $k \geq 0$.

3.3.2 A RBE approach to the extended Kalman filter

Based on the fact that the nonlinear system is linearized before the analysis step, the posterior pdf at time step $t-1$ can be approximated by a Gaussian distribution, i.e.

$$p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_{t-1} : \mathbf{x}_{t-1}^a, \mathbf{P}_{t-1}^a) \tag{3.30}$$

The main difference between Equation 3.9 and Equation 3.30 is the state itself. In Equation 3.9, $\mathbf{x}_t$ represents all the possible state trajectories, while in Equation 3.30, the state represents only the particular state trajectory at which the system is linearized. Moreover, the posterior distribution is approximated as a Gaussian. The linearization is elaborated in the next section.

Forecast step

The forecast step is given by the computation of the predictive pdf:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} \tag{3.31}$$

The posterior at time $t-1$ is given by Equation 3.30. The transition pdf at the right hand side of Equation 3.31 can be obtained from Equation 3.22 and by considering that this pdf follows a Gaussian distribution:

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_{t-1} : E[\mathbf{f}_{t,t-1}(\mathbf{x}_{t-1})], \mathbf{Q}_t), \tag{3.32}$$

thus, the predictive pdf is defined as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int \mathcal{N}(\mathbf{x}_{t-1} : \mathbf{x}^a_{t-1}, \mathbf{P}^a_{t-1})\mathcal{N}(\mathbf{x}_{t-1} : E[\mathbf{f}_{t,t-1}(\mathbf{x}_{t-1})], \mathbf{Q}_t)d\mathbf{x}_{t-1} \tag{3.33}$$

The integral in Equation 3.33 is complicated due to the presence of the non-linear function $\mathbf{f}_{t,t-1}(.)$. The problem formulation should be changed in order to obtain an analytical solution such as in the case of the discrete KF (Equations 3.12 and 3.13). The approach adopted in the extended KF is the linearization by using Taylor series. Additionally, a state trajectory should be defined around which the function is linearized. At this point in the development of the extended KF, the best state estimate is given by $\mathbf{x}^a_{t-1}$.

Taylor series expansion of $\mathbf{f}_{t,t-1}(\mathbf{x}_{t-1})$ in a neighborhood of $\mathbf{x}^a_{t-1} = E[\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}]$ is considered in the linearization of the system. For the linear approximation, only the first two terms are considered:

$$\mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}) \simeq \mathbf{f}_{t,t-1}(\mathbf{x}^a_{t-1}) + \mathbf{F}_{t,t-1}[\mathbf{x}_{t-1} - \mathbf{x}^a_{t-1}] \tag{3.34}$$

where

$$\mathbf{F}_{t,t-1} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_{m_x}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_{m_x}}{\partial x_1} & \cdots & \frac{\partial f_{m_x}}{\partial x_{m_x}} \end{bmatrix} |(\mathbf{x}^a_{t-1}) \tag{3.35}$$

is a Jacobian matrix with dimension $(m_x \times m_x)$. The right-hand side of Equation 3.34 is elaborated as follows:

$$\mathbf{F}_{t,t-1}\mathbf{x}_{t-1} + \mathbf{f}_{t,t-1}(\mathbf{x}^a_{t-1}) + \mathbf{F}_{t,t-1}\mathbf{x}^a_{t-1} \tag{3.36}$$

where the first term is linear with respect to $\mathbf{x}_{t-1}$ and the following two terms are known values. The substitution of Equation 3.34 into 3.33,

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int & \mathcal{N}(\mathbf{x}_{t-1} : \mathbf{x}^a_{t-1}, \mathbf{P}^a_{t-1}) \dots \\ & \mathcal{N}(\mathbf{x}_{t-1} : \mathbf{f}_{t,t-1}(\mathbf{x}^a_{t-1}) + \mathbf{F}_{t,t-1}[\mathbf{x}_{t-1} - \mathbf{x}^a_{t-1}], \mathbf{Q}_t)d\mathbf{x}_{t-1} \end{aligned} \tag{3.37}$$

allows for an analytical solution of the integral. The resulting prior pdf can be obtained in the Gaussian form as

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_t : \mathbf{x}^f_t, \mathbf{P}^f_t) \tag{3.38}$$

Although the prior pdf is represented by a Gaussian distribution, it is important to remark the fact that this representation is only a local approximation around a nominal trajectory, thus it is an approximation of the true pdf which is not explicitly known. The forecast state vector $\mathbf{x}^f_t$ and forecast

error covariance matrix $\mathbf{P}_t^f$ are mostly similar as in the KF and are given by:

$$\mathbf{x}_t^f = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}^a) \tag{3.39}$$

$$\mathbf{P}_t^f = \mathbf{F}_{t,t-1}\mathbf{P}_{t-1}^a\mathbf{F}_{t,t-1}^T + \mathbf{Q}_t \tag{3.40}$$

The forecast state vector is obtained by the propagation of the state through the nonlinear model and for the computation of the covariance matrix, the Jacobian matrix $\mathbf{F}_{t,t-1}$ is used.

Analysis step

The analysis step consists in the update of the prior according to:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \tag{3.41}$$

For the extended Kalman filter, the prior is approximated by a Gaussian distribution which is characterized by the mean 3.39 and the covariance 3.40. The likelihood pdf is obtained from Equation 3.23 by considering that this pdf is Gaussian.

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t : E[\mathbf{h}_t(\mathbf{x}_t)], \mathbf{R}_t) \tag{3.42}$$

By substituting 3.38 and 3.42 into 3.41, the filtering pdf is given by:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{\mathcal{N}(\mathbf{y}_t : E[\mathbf{h}_t(\mathbf{x}_t)], \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)}{\int \mathcal{N}(\mathbf{y}_t : E[\mathbf{h}_t(\mathbf{x}_t)], \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)d\mathbf{x}_t} \tag{3.43}$$

where the first issue aiming to calculate the filtering pdf is to solve the integral in the denominator. In order to find an analytical solution, the observation system is linearized around a state trajectory. At this point $\mathbf{x}_t^f$ is the best state estimation thus the nonlinear observation system is linearized through Taylor series expansion in the neighborhood of $\mathbf{x}_t^f$. The nonlinear function is expanded as follows:

$$\mathbf{h}_t(\mathbf{x}_t) \simeq \mathbf{h}_t(\mathbf{x}_t^f) + \mathbf{H}_t[\mathbf{x}_t - \mathbf{x}_t^f] \tag{3.44}$$

where $\mathbf{H}_t$ is a $(m_y \times m_x)$ dimensional Jacobian matrix given by:

$$\mathbf{H}_t = \begin{bmatrix} \frac{\partial h_1}{\partial x_1} & \cdots & \frac{\partial h_1}{\partial x_{m_x}} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_{m_y}}{\partial x_1} & \cdots & \frac{\partial h_{m_y}}{\partial x_{m_x}} \end{bmatrix} |(\mathbf{x}_t^f) \tag{3.45}$$

By substituting 3.44 into 3.43, the expression for the posterior is the following:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{\mathcal{N}(\mathbf{y}_t : \mathbf{h}_t(\mathbf{x}_t^f) + \mathbf{H}_t[\mathbf{x}_t - \mathbf{x}_t^f], \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)}{\int \mathcal{N}(\mathbf{y}_t : \mathbf{h}_t(\mathbf{x}_t^f) + \mathbf{H}_t[\mathbf{x}_t - \mathbf{x}_t^f], \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)d\mathbf{x}_t} \quad (3.46)$$

Now, the solution of the integral is tractable and the resulting expression for the posterior pdf in terms of a distribution is:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t^a, \mathbf{P}_t^a); \quad (3.47)$$

with

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t^f)), \quad (3.48)$$

$$\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_t^f, \quad (3.49)$$

and

$$\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t^T[\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T + \mathbf{R}_t]^{-1} \quad (3.50)$$

Summary of the extended Kalman filter

The extended Kalman filter is described in algorithm 3.

Algorithm 3 Extended Kalman Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_t^f = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}^a)$
   $\mathbf{P}_t^f = \mathbf{F}_{t,t-1}\mathbf{P}_{t-1}^a\mathbf{F}_{t,t-1}^{\mathrm{T}} + \mathbf{Q}_t$
2. Analysis step:
   $\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}}[\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^{\mathrm{T}} + \mathbf{R}_t]^{-1}$
   $\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t^f))$
   $\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_t^f$

In algorithm 3, the probability density functions involved in the estimation are represented by the first and second moments of the density functions. Although this characterization might not be sufficient for specific systems, the EKF has shown proper performance in a widespread range of applications.

An alternative to the extended Kalman filter is the second order extended Kalman filter (Simon, 2006, ch. 13). Again, the second order filter is based on the Taylor series expansion but in this case the second order terms are preserved (Hessian matrices). Although this type of filter improves the local

approximation of the filtering pdfs, it is computationally expensive mainly when the system is high dimensional.

In the last decade, novel local filtering techniques have been proposed (Julier et al., 2000; Norgaard et al., 2000). These filtering methods are based either on another type of polynomial approximation of the nonlinear functions in the system description (Norgaard et al., 2000; Schei, 1997; Ito and Xiong, 2000; Crassidis, 2006) or on the so called unscented transformation (Julier et al., 2000; Ito and Xiong, 2000; Julier, 2002; Julier and Uhlmann, 2004; Šimandl et al., 2006). The polynomial approximation can be seen as the Taylor series expansion where the first and second derivatives are approximated by divided differences. On the other hand, the basic idea of the unscented transformation is to perform an approximation of the pdfs involved in the estimation by a set of deterministically chosen points.

3.4 Ensemble Kalman Filter

The ensemble Kalman filter originally introduced by Evensen (1994) is a data assimilation method, initially developed for estimation problems in weather forecast applications where the numerical models are highly nonlinear, high dimensional and having a chaotic nature. The name of the filter is adopted from the theoretical foundations which are used: ensemble forecasting and Kalman filtering.

3.4.1 Background information

Basically, the extended Kalman filter consists of the linearization of the nonlinear system about a nominal state trajectory and the posterior approximation of the pdfs as Gaussian. In the derivation of the EnKF, a different approach in the approximation of the predictive and filtering pdfs is used. Before presenting the EnFK and for the sake of clarity in the derivation, the concepts of Monte Carlo simulation and Ensemble forecasting are presented in this section as a background for the presentation of the EnKF.

Monte Carlo simulation

Monte Carlo methods are a set of computational algorithms that rely on random sampling to obtain numerical results of complex problems. Since truly random numbers are not possible to obtain through algorithms implemented in computers, random sampling in MC methods implies the generation of pseudo-random numbers. Monte Carlo techniques are particularly useful

in scenarios where it is of interest to perform calculations that involve implicitly or explicitly a probability density function p for which closed-form calculations cannot be performed due to the algebraic complexity of the problem. The algebraic representation of p is approximated by a sample representation.

Consider the pdf $p(\mathbf{x})$ from which it is possible to draw N independent and identically distributed random samples $\{\mathbf{x}^{(i)}; i = 1, \dots N\}$ (perfect MC simulation). Then, the pdf $p(\mathbf{x})$ can be approximated by the empirical pdf $p_N(\mathbf{x})$ as follows:

$$p_N(\mathbf{x}) = \frac{1}{N} \sum_{i=1}^{N} \delta(\mathbf{x} - \mathbf{x}^{(i)}) \tag{3.51}$$

where δ is the Dirac delta function, defined by:

$$\delta(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad \mathbf{x} = \mathbf{0}, \\ \mathbf{0} & \text{if} \quad \mathbf{x} \neq \mathbf{0} \end{cases} \tag{3.52}$$

The Dirac delta function has the following fundamental property:

$$\int \mathbf{f}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{c}) d\mathbf{x} = \mathbf{f}(\mathbf{c}) \tag{3.53}$$

with $\mathbf{f}(\mathbf{x})$ a vector-valued function and $\mathbf{c}$ a constant vector. This property allows for the mapping of integrals to discrete sums. For example, consider the problem of estimating $E[\mathbf{f}(\mathbf{x})]$ with the pdf of $\mathbf{f}(\mathbf{x})$ given by $p(\mathbf{x})$. The problem is posed as follows:

$$E[\mathbf{f}(\mathbf{x})] = \int \mathbf{f}(\mathbf{x}) p(\mathbf{x}) d\mathbf{x} \tag{3.54}$$

where the approximated solution consists of the replacement of $p(\mathbf{x})$ with its sample representation presented in equation 3.51, so that

$$E[\mathbf{f}(\mathbf{x})] \simeq \frac{1}{N} \sum_{i=1}^{N} \int \mathbf{f}(\mathbf{x}) \delta(\mathbf{x} - \mathbf{x}^{(i)}) = \frac{1}{N} \sum_{i=1}^{N} \mathbf{f}(\mathbf{x}^{(i)}) \tag{3.55}$$

The power of MC techniques lies in the fact that the rate at which the approximation converges towards the true value of the expectation is immune to the dimension of the state-space. In numerical weather forecast models, it is possible to represent the modeling and data uncertainties through MC simulation.

Ensemble forecasting

Ensemble forecasting is a numerical method which is based on MC simulation. The methodology attempts to generate a representative sample of the pdf of the forecast state vector. The representative sample is referred to as ensemble. The ensemble representation of the forecast pdf is conducted to account for the two common sources of uncertainty in forecast models:

- the errors introduced by the use of imperfect initial conditions, amplified in the case of weather models by the chaotic nature of the evolution equations of the dynamical system, which is often referred to as the sensitive dependence on the initial conditions.
- the errors introduced because of the imperfections in the model formulation, including the possible noise in the model input data and the errors in the model parameters.

3.4.2 RBE approach to the EnKF

In this section the equations of the EnKF are presented. Consider the following scenario:

$$\mathbf{x}_t = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1}) + \mathbf{w}_t \tag{3.56}$$
$$\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t) + \mathbf{v}_t \tag{3.57}$$

where the distributions and statistics of the process and observation noises are:

$$\mathbf{w}_t \sim \mathcal{N}(\mathbf{w}_t : \mathbf{0}, \mathbf{Q}_t) \tag{3.58}$$
$$\mathbf{v}_t \sim \mathcal{N}(\mathbf{v}_t : \mathbf{0}, \mathbf{R}_t) \tag{3.59}$$
$$E[\mathbf{w}_t] = \mathbf{0} \quad , \quad E[\mathbf{v}_t] = \mathbf{0}, \tag{3.60}$$
$$E[\mathbf{w}_t \mathbf{w}_{t+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{w}_t \mathbf{w}_t^T] = \mathbf{Q}_t, \tag{3.61}$$
$$E[\mathbf{v}_t \mathbf{v}_{t+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{v}_t \mathbf{v}_t^T] = \mathbf{R}_t, \tag{3.62}$$
$$E[\mathbf{w}_t \mathbf{v}_{t+k}^T] = \mathbf{0} \quad , \quad E[\mathbf{w}_t \mathbf{v}_t^T] = \mathbf{0} \tag{3.63}$$

with white noises, Gaussianly distributed with zero mean and uncorrelated $\forall t$ with $k \geq 0$. The noise assumptions may not always be realistic in practice. However, these assumptions reduce the complexity in the derivation of the EnKF. Realistic errors are considered in the section concerning the application of the EnKF.

Forecast step

The EnKF is a recursive filter where the forecast step at time instant t is carried out based on the information obtained from the analysis step at instant $t-1$. The EnKF methodology relies on the ensemble representation of the predictive and filtering pdfs which are known as the forecast ensemble and the analysis ensemble.

Consider the sample representation of the posterior at time $t-1$:

$$p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) \simeq \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x}_{t-1}-\mathbf{x}^a_{t-1,i}) \tag{3.64}$$

with $\mathbf{X}^a_{t-1} = \{\mathbf{x}^a_{t-1,i} : i = 1,\dots N\}$ the analysis ensemble and taking into account that the transition pdf is Gaussianly distributed,

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_{t-1} : E[\mathbf{f}_{t,t-1}(\mathbf{x}_{t-1})], \mathbf{Q}_t), \tag{3.65}$$

then, the predictive pdf is approximated as follows:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \simeq \frac{1}{N}\sum_{i=1}^{N}\mathcal{N}(\mathbf{x}_t : E[\mathbf{f}_{t,t-1}(\mathbf{x}^a_{t-1,i})], \mathbf{Q}_t) \tag{3.66}$$

which is a sum of a set of Gaussian pdfs. The forecast ensemble is obtained by propagating the analysis ensemble through the nonlinear model:

$$\mathbf{x}^f_{t,i} = \mathbf{f}_{t,t-1}(\mathbf{x}^a_{t-1,i}), \quad i = 1, 2, \dots, N. \tag{3.67}$$

thus, it is possible to calculate the mean and the covariance of the approximated predictive pdf in Equation 3.66 by the sample mean and sample covariance:

$$\mathbf{x}^f_t = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}^f_{t,i} \tag{3.68}$$

$$\mathbf{P}^f_t = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}^f_{t,i}-\mathbf{x}^f_t)(\mathbf{x}^f_{t,i}-\mathbf{x}^f_t)^T \tag{3.69}$$

An important fact regarding the sample representation of the predictive pdf is the possibility of exploring a wide range of the state-space rather than a local exploration such as in the case of the extended Kalman filter. Another advantage is the increase in the computational efficiency by avoiding the evaluation of the Jacobian matrices.

Analysis step

In the analysis step, the prior pdf is updated according to the posterior pdf:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \tag{3.70}$$

Using Equations 3.57 and 3.59, the likelihood pdf can be represented by a Gaussian distribution as follows:

$$p(\mathbf{y}_t|\mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t : E[\mathbf{h}_t(\mathbf{x}_t)], \mathbf{R}_t) \tag{3.71}$$

From Equations 3.68 and 3.69, the prior pdf is a Gaussian pdf which is the result of a sum of N Gaussian pdfs (Equation 3.66):

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f) \tag{3.72}$$

Therefore, the posterior pdf is given by:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{\mathcal{N}(\mathbf{y}_t : E[\mathbf{h}_t(\mathbf{x}_t)], \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)}{\int \mathcal{N}(\mathbf{y}_t : E[\mathbf{h}_t(\mathbf{x}_t)], \mathbf{R}_t)\mathcal{N}(\mathbf{x}_t : \mathbf{x}_t^f, \mathbf{P}_t^f)d\mathbf{x}_t} \tag{3.73}$$

In Equation 3.73, the presence of the non-linear function $\mathbf{h}(.)$ may not allow for a closed form solution of the integral in the denominator. In the derivation of the EKF, the linearization of the observation system was adopted and the posterior could be approximated by a Gaussian. Basically, the linearization consisted of the computation of the Jacobian Matrix $\mathbf{H}_t$ (Equation 3.45), evaluated over the best estimate of the state trajectory, in this case the best estimate corresponds to $\mathbf{x}_t^f$. By following the same approach, the posterior pdf is presented as follows:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_t^a, \mathbf{P}_t^a); \tag{3.74}$$

with

$$\mathbf{K}_t = \mathbf{P}_t^f\mathbf{H}_t^T[\mathbf{H}_t\mathbf{P}_t^f\mathbf{H}_t^T + \mathbf{R}_t]^{-1} \tag{3.75}$$

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t^f)), \tag{3.76}$$

$$\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{K}_t\mathbf{H}_t\mathbf{P}_t^f, \tag{3.77}$$

The derived methodology is completely valid for smooth non-linear observation systems. However, unlike the derivation of the EKF, an analysis ensemble $\mathbf{X}_t^a$ should be generated for the computation of the forecast ensemble (Equation 3.67). A way to carry out the generation of the analysis ensemble can be by sampling from the posterior pdf (Equation 3.74):

$$\mathbf{x}_{t,i}^a \sim \mathcal{N}(\mathbf{x}_t|\mathbf{y}_{1:t} : \mathbf{x}_t^a, \mathbf{P}_t^a); \quad i = 1, 2, \ldots N \tag{3.78}$$

A disadvantage of this approach could be the possible non proper representation of the analysis ensemble due to the insertion of sampling errors in the generation of random numbers. Another approach, which is widely used, is the Monte Carlo approximation of the observation system. In order to introduce this approach, the following expressions are presented. The sample mean of the modeled observation ensemble:

$$\mathbf{y}_t = \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_t(\mathbf{x}_{t,i}^f), \tag{3.79}$$

the cross covariance between the forecast ensemble and the modeled observation ensemble:

$$\mathbf{P}_t^{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_{t,i}^f - \mathbf{x}_t^f)(\mathbf{h}_t(\mathbf{x}_{t,i}^f) - \mathbf{y}_t)^T, \tag{3.80}$$

and the sample covariance of the modeled observation ensemble.

$$\mathbf{P}_t^{yy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{h}_t(\mathbf{x}_{t,i}^f) - \mathbf{y}_t)(\mathbf{h}_t(\mathbf{x}_{t,i}^f) - \mathbf{y}_t)^T, \tag{3.81}$$

Expressions 3.80 and 3.81 allows for a complete derivative-free derivation of the Kalman gain and the sample covariance of the analysis ensemble:

$$\mathbf{K}_t = \mathbf{P}_t^{xy}\left(\mathbf{P}_t^{yy} + \mathbf{R}_t\right)^{-1} \tag{3.82}$$

$$\mathbf{P}_t^a = \mathbf{P}_t^f - \mathbf{K}_t(\mathbf{P}_t^{xy})^T \tag{3.83}$$

A detailed formulation of this approach is presented in Houtekamer and Mitchell (2001). As can be seen in equation 3.82, the computation of the Kalman gain does not depend on the forecast covariance matrix and neither from the analysis covariance at one time step backwards. Therefore, this is an important property of the EnKF when the algorithm is implemented in practice since it is not necessary to evaluate the covariances $\mathbf{P}_t^f$ and $\mathbf{P}_t^a$ in Equations 3.69 and 3.83. Moreover, with the computation of the Kalman gain based on sample covariances, different types of process noise can be considered in the application of the EnKF. For instance, possible errors in the driving forces and model parameters can be taken into account with the EnKF.

The approach presented in Burgers et al. (1998) is adopted for the generation of the analysis ensemble. This approach consists in the generation of surrogate observations $\mathbf{Y}_t^s = \mathbf{y}_{t,i}^s; i = 1, \ldots N$ where $\mathbf{y}_{t,i}$ are random samples drawn from the Gaussian distribution with mean $\mathbf{y}_t$ and covariance $\mathbf{R}_t$, as follows:

$$\mathbf{y}_{t,i}^s \sim \mathcal{N}(\mathbf{y}_t : \mathbf{y}_t, \mathbf{R}_t); \tag{3.84}$$

The analysis ensemble is obtained by replacing the observation by the set of surrogates observations into Equation 3.76:

$$\mathbf{x}_{t,i}^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_{t,i}^s - \mathbf{h}_t(\mathbf{x}_t^f)), \quad i = 1\ldots, N. \tag{3.85}$$

Equation 3.85 indicates that all the ensemble members of a particular state are updated by the same Kalman gain but a different innovation term. In order to take into account possible sampling errors in the generation of the surrogate observations, the observation covariance matrix can be replaced by the sample covariance,

$$\mathbf{R}_t^s = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}_{t,i}^s - \mathbf{y}_t)(\mathbf{y}_{t,i}^s - \mathbf{y}_t)^T, \tag{3.86}$$

and used in the computation of the Kalman gain.

Summary of the EnKF

The EnKF algorithm is presented as follows:

Algorithm 4 Ensemble Kalman Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_{t,i}^f = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1,i}^a), \quad i = 1, 2, \ldots, N,$
   $\mathbf{x}_t^f = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{t,i}^f$
   $\mathbf{y}_t = \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_t(\mathbf{x}_{t,i}^f),$
   $\mathbf{P}_t^{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_{t,i}^f - \mathbf{x}_t^f)(\mathbf{h}_t(\mathbf{x}_{t,i}^f) - \mathbf{y}_t)^T,$
   $\mathbf{P}_t^{yy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{h}_t(\mathbf{x}_{t,i}^f) - \mathbf{y}_t)(\mathbf{h}_t(\mathbf{x}_{t,i}^f) - \mathbf{y}_t)^T,$
2. Analysis step:
   $\mathbf{y}_{t,i}^s \sim \mathcal{N}(\mathbf{y}_t : \mathbf{y}_t, \mathbf{R}_t);$
   $\mathbf{R}_t^s = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}_{t,i}^s - \mathbf{y}_t)(\mathbf{y}_{t,i}^s - \mathbf{y}_t)^T,$
   $\mathbf{K}_t = \mathbf{P}_t^{xy}\,(\mathbf{P}_t^{yy} + \mathbf{R}_t)^{-1}$
   $\mathbf{x}_{t,i}^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_{t,i}^s - \mathbf{h}_t(\mathbf{x}_t^f)), \quad i = 1\ldots, N.$

The ensemble Kalman filter in Algorithm 4 is referred to as the stochastic EnKF since in the implementation of this filter surrogate observations are

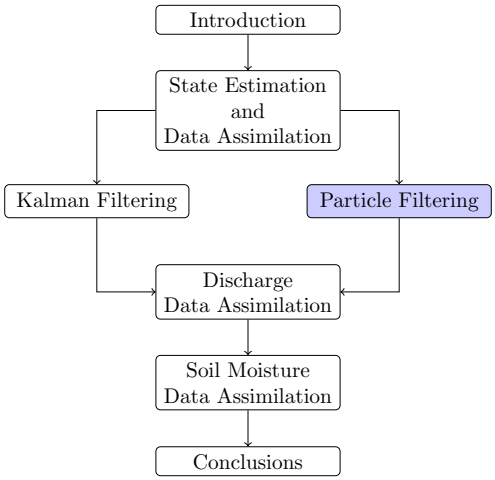
drawn for the generation of the analysis ensemble. An alternative to this implementation is the ensemble square root Kalman filter (Anderson, 2001; Bishop et al., 2001; Whitaker and Hamill, 2002) in which the analysis ensemble is generated based on the analysis sample mean and a square root matrix of the analysis error covariance.

3.5 Summary

For linear systems, the discrete Kalman filter is the optimal implementation of the Bayes filter. The problem is solved analytically due to the Gaussian conditions in the process and observations models. For non-linear systems, the predictive and the filtering pdfs are approximated as Gaussian either by linearizing the system around a local trajectory or by using the Monte Carlo integration approach. The first corresponds to the extended Kalman filter in which the system is allowed to be non-linear affected by additive zero mean Gaussian noise and the second is the Ensemble Kalman filter in which the non-linear system may be affected not only by zero mean white Gaussian noise.

Chapter 4

Particle Filtering



4.1 Introduction

The aim of this chapter is to present a set of filtering methods for non-linear/non-Gaussian systems which are based on Sequential Monte Carlo methods (SMC). SMC methods have been introduced in the fifties. These implementations were based on the sequential importance sampling principle which is presented in this chapter. However, advances in SMC methods were possible due to inclusion of a resampling stage in the nineties (Gordon et al., 1993). As indicated in Doucet and Johansen (2009), new improvements have been proposed lately.

4.2 Perfect Monte Carlo Simulation

Monte Carlo simulation was briefly introduced in section 3.4.1. In this section, Monte Carlo methods are used in the approximation of the posterior density function. Assume at this point that it is possible to obtain N random samples from the posterior pdf. Hereafter, the random samples used in the approximation of density functions are referred to as particles.

$$\mathbf{x}_{0:t,i} \sim p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}); \quad i = 1, 2, \ldots, N \tag{4.1}$$

It is clear that a perfect Monte Carlo sampling is only hypothetical case since the posterior is unknown but the assumption is used in order to explain the MC approach. According to equation 3.51, the posterior pdf can be approximated by the following empirical estimate:

$$p_N(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t,i}) \tag{4.2}$$

Moreover, any expectation of interest, for example

$$E[g(\mathbf{x}_{0:t})] = \int g(\mathbf{x}_{0:t})p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})d\mathbf{x}_{0:t} \tag{4.3}$$

can be represented as follows:

$$\overline{E[g(\mathbf{x}_{0:t})]} = \frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x}_{0:t,i}) \tag{4.4}$$

Here, the particles $\mathbf{x}_{t,i}$ should be independent and identically distributed (i.i.d) for the approximation to be valid. The law of large numbers allows for the demonstration of convergence:

$$\overline{E[g(\mathbf{x}_{0:t})]} \xrightarrow[N\to\infty]{} E[g(\mathbf{x}_{0:t})] \tag{4.5}$$

where the symbol $\xrightarrow[N\to\infty]{}$ denotes almost sure convergence. Equation 4.5 indicates that the particle approximation of the expectation converges to the true value as the number of particles is increased. Additionally, according to the central limit theorem and if the variance of $g(\mathbf{x}_t)$ is bounded, i.e. $(var(g(\mathbf{x}_t)) < \infty)$, then the approximation error converges to a Gaussian distribution with zero mean:

$$\sqrt{N}\left(\overline{E[g(\mathbf{x}_{0:t})]} - E[g(\mathbf{x}_{0:t})]\right) \xRightarrow[N\to\infty]{} \mathcal{N}(0, var(g(\mathbf{x}_{0:t})) \tag{4.6}$$

with the symbol $\xRightarrow[N\to\infty]{}$ indicating convergence in distribution.

4.3 Importance sampling

Perfect Monte Carlo sampling is often impossible to be carried out. However, importance sampling is a methodology which circumvents the difficulty of sampling from unknown distributions. Basically, the methodology consists of sampling from a known proposal distribution $q(\mathbf{x}_t|\mathbf{y}_{1:t})$. The proposal distribution is also referred to as the importance distribution. A general recommendation is that the sampling from the proposal should be easy. The importance sampling approach for the expectation in Equation 4.3 is derived and presented in the following expressions. First, the proposal density function is inserted into Equation 4.3:

$$E[g(\mathbf{x}_{0:t})] = \int g(\mathbf{x}_{0:t}) \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t} \tag{4.7}$$

In Equation 4.7, $\mathbf{x}_{0:t}$ represents the full path of the state vector. The unknown posterior pdf is then replaced by considering the Bayes' theorem:

$$E[g(\mathbf{x}_{0:t})] = \int g(\mathbf{x}_{0:t}) \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t)p(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t})q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t} \tag{4.8}$$

Equation 4.8 is simplified by grouping pdfs as follows:

$$w(\mathbf{x}_{0:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \tag{4.9}$$

with $w(\mathbf{x}_{0:t})$ the unnormalized importance weights. The new expression for the expectation with the importance weights is:

$$E[g(\mathbf{x}_{0:t})] = \int g(\mathbf{x}_{0:t}) \frac{w(\mathbf{x}_{0:t})}{p(\mathbf{y}_{1:t})} q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t} \tag{4.10}$$

The unknown normalizing pdf $p(\mathbf{y}_{1:t})$ is elaborated in the following expressions:

$$\begin{aligned} E[g(\mathbf{x}_{0:t})] &= \frac{1}{p(\mathbf{y}_{1:t})} \int g(\mathbf{x}_{0:t}) w(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t} \\ &= \frac{\int g(\mathbf{x}_{0:t}) w(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t}}{\int p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t}) p(\mathbf{x}_{0:t}) \frac{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} d\mathbf{x}_{0:t}} \\ &= \frac{\int g(\mathbf{x}_{0:t}) w(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t}}{\int w(\mathbf{x}_{0:t}) q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) d\mathbf{x}_{0:t}} \end{aligned} \tag{4.11}$$

Equation 4.11 can be represented as the ratio of two expectations:

$$E[g(\mathbf{x}_{0:t})] = \frac{E_{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}[w(\mathbf{x}_{0:t})g(\mathbf{x}_{0:t})]}{E_{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}[w(\mathbf{x}_{0:t})]} \tag{4.12}$$

with the notation $E_{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}$ indicating that the expectations are taken over the proposal distribution $q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})$. Finally, the expectation of interest is obtained by drawing samples from the proposal pdf and the corresponding sample representation of the expectations in Equation 4.12:

$$\begin{aligned}\overline{E[g(\mathbf{x}_{0:t})]} &= \frac{\frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}_{0:t,i})\mathbf{g}(\mathbf{x}_{0:t,i})}{\frac{1}{N}\sum_{i=1}^{N} w(\mathbf{x}_{0:t,i})} \\ &= \sum_{i=1}^{N} g(\mathbf{x}_{0:t,i})\tilde{w}(\mathbf{x}_{0:t,i}) \end{aligned} \tag{4.13}$$

where $\tilde{w}$ are the normalized importance weights which are given by

$$\tilde{w}(\mathbf{x}_{0:t,i}) = \frac{w_t(\mathbf{x}_{0:t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{0:t,i})} \tag{4.14}$$

The asymptotic convergence for $\overline{E[g(\mathbf{x}_{0:t})]}$ have been demonstrated (Doucet, 1998; Geweke, 1989; Crisan, 2000) under the assumptions that the particles drawn from the proposal distribution are i.i.d and the expectations $E[g(\mathbf{x}_{0:t})]$, $E[w(\mathbf{x}_{0:t})]$, $E[w(\mathbf{x}_{0:t})g(\mathbf{x}_{0:t})]$ do exist with a finite dimensional space.

From equation 4.13, it can be inferred that the posterior pdf can be approximated well by the weighted empirical estimate:

$$p_N(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = \sum_{i=1}^{N} \tilde{w}(\mathbf{x}_{0:t,i})\delta(\mathbf{x}_{0:t} - \mathbf{x}_{0:t,i}) \tag{4.15}$$

4.4 Sequential Importance Sampling (SIS)

It is possible to derive a sequential formulation for the importance weights. In this case, the information captured at time step $t-1$ is important. Specifically, the state $\mathbf{x}_{t-1}$ and the observations $\mathbf{y}_{1:t-1}$ are used in the formulation. The derivation of the sequential version of the importance sampling approach implies the modification of the unnormalized importance weights (Equation 4.9), as follows:

$$\begin{aligned} w(\mathbf{x}_{0:t}) &= \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_{0:t}|\mathbf{y}_{1:t-1})p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\ &= \frac{p(\mathbf{y}_t|\mathbf{x}_{0:t}|\mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \end{aligned} \tag{4.16}$$

From Equation 4.16, by considering that it is possible to factorize the importance pdf as follows:

$$q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t}) = q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1}), \tag{4.17}$$

and by considering $\{\mathbf{x}_{0:t}\}$ as a Markov process, the recursive expression for the importance weights is given by:

$$\begin{aligned} w(\mathbf{x}_t) &= \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})q(\mathbf{x}_{0:t-1}|\mathbf{y}_{1:t-1})} \\ &= w(\mathbf{x}_{t-1})\frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{x}_{t-1})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})} \end{aligned} \tag{4.18}$$

with $w(\mathbf{x}_{t-1})$ indicating the importance weights computed at time step $t-1$. Hereafter, the following notation is used: $w_t = w(\mathbf{x}_t)$ and $w_{t-1} = w(\mathbf{x}_{t-1})$. Equation 4.18 provides a procedure to sequentially update the importance weights, given an appropriate choice of the proposal distribution. In this context, the methodology for approximating the posterior pdf is to draw particles from the proposal distribution, i.e. $\mathbf{x}_{t,i} \sim q(\mathbf{x}_t|\mathbf{x}_{0:t-1}, \mathbf{y}_{1:t})$ for $i = 1, \ldots, N$, and evaluate the likelihood and transition pdfs iteratively. Therefore, Equation 4.18 is used to assign a weight to each particle as follows:

$$w_{t,i} = w_{t-1,i}\frac{p(\mathbf{y}_t|\mathbf{x}_{t,i})p(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i})}{q(\mathbf{x}_{t,i}|\mathbf{x}_{0:t-1,i}, \mathbf{y}_{1:t})} \tag{4.19}$$

This procedure is referred to as sequential importance sampling. The computation of the importance weights allows for the approximation of any expectation of interest according to equation 4.13. In this dissertation, the expectation of interest is the conditional mean of $\mathbf{x}_t$, i.e $E[\mathbf{x}_t|\mathbf{y}_{1:t}]$, with the approximation given by:

$$\overline{E[\mathbf{x}_t]} = \sum_{i=1}^{N} \mathbf{x}_{t,i}\tilde{w}_{t,i} \tag{4.20}$$

with $\{\mathbf{x}_{t,i};\, i = 1, \ldots, N\}$ the set of particles which are drawn according to the proposal pdf. The SIS methodology is summarized in the next section.

4.4.1 Summary of the SIS filter

In this section, the SIS filter is presented in Algorithm 5. In the forecast step, a set of i.i.d particles $\mathbf{x}_{t,i}$ is drawn from the proposal pdf $q(\mathbf{x}_t|\mathbf{x}_{t-1,i}\,, \mathbf{y}_{1:t})$. At this point, the implementation of the SIS filter in practice faces the challenge of selecting a proper proposal pdf. The choice of an appropriate proposal pdf is treated in Section 4.8.

In the analysis step, a weight is assigned to each particle. The computation of the weight is possible after the arrival of the observation $\mathbf{y}_t$. The pdfs involved in the computation of the weights are the likelihood pdf, the transition pdf and the proposal pdf. The likelihood pdf is given by the observation model while the transition pdf is given by the process model. Both densities form the state-space model revisited in section 2.4.3.

Algorithm 5 General SIS Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_{t,i} \sim q(\mathbf{x}_t|\mathbf{x}_{t-1,i}\,,\,\mathbf{y}_{1:t}), \quad i = 1, 2, \ldots, N$
2. Analysis step:
   $$w_t(\mathbf{x}_{t,i}) = w_{t-1}(\mathbf{x}_{t-1,i})\frac{p(\mathbf{y}_t|\mathbf{x}_{t,i})p(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i})}{q(\mathbf{x}_{t,i}|\mathbf{x}_{t-1,i}\,,\,\mathbf{y}_{1:t})} \quad i = 1, 2, \ldots, N$$
   $$\tilde{w}_t(\mathbf{x}_{t,i}) = \frac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$$
   $$\overline{E[\mathbf{x}_t]} = \sum_{i=1}^{N} \mathbf{x}_{t,i}\tilde{w}_t(\mathbf{x}_{t,i})$$

The implementation in practice of Algorithm 5 suffers from two major limitations. One limitation is the complexity in the form of the proposal pdf and the corresponding difficulty in the selection of a proper proposal pdf. Another limitation, from the implementation point of view, is the evaluation of the particles by the transition pdf for which we should know precisely the mathematical function of the distribution. For example, for a Gaussian density function $\mathcal{N}(x; \mu, \sigma)$, it is defined by the exponential function $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{s\sigma^2}}$.

An initial approach to tackle the difficulties in the implementation of the SIS filter is the selection of the transition pdf as the proposal pdf, so that:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t}) = p(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{4.21}$$

The main disadvantage in this approach to the selected proposal (equation 4.21) is the lack of incorporating the information from the sequence of observations $\mathbf{y}_{1:t}$ in the proposal. On the other hand, this approach facilitates the implementation of the SIS filter due to the resulting simplified expression for the computation of the importance weights. Additionally, the particles should only be drawn from the transition pdf and not evaluated. For the generation of the particles, the nonlinear stochastic model can be used. The SIS filter with the corresponding modifications is presented in Algorithm 6.

The application of Algorithm 6 is straightforward, however, a limitation in the performance of the filter is expected due to the almost likely mismatch

Algorithm 6 Natural SIS Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_{t,i} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1}), \quad i = 1, 2, \ldots, N$
2. Analysis step:
   $w_t(\mathbf{x}_{t,i}) = w_{t-1}(\mathbf{x}_{t-1,i})p(\mathbf{y}_t|\mathbf{x}_{t,i}) \quad i = 1, 2, \ldots, N$
   $\tilde{w}_t(\mathbf{x}_{t,i}) = \dfrac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$
   $\overline{E[\mathbf{x}_t]} = \sum_{i=1}^{N} \mathbf{x}_{t,i}\tilde{w}_t(\mathbf{x}_{t,i})$

between the true posterior pdf, which is unknown, and the selected proposal pdf. More specific, a degeneration of the weights over time is experienced when the implementation of the natural SIS filter is carried out.

4.4.2 Degeneracy of the particle weights

The SIS filter discussed in section 4.4.1 has a main limitation in terms of performance. The limitation is related to the evolution of the variance of the importance weights over time. The evolution indicates a large increase of the variance which degenerates the importance weight values after a few iterations of the filter. An elaborated demonstration of the degeneracy phenomenon is reported in Doucet (1998).

To understand the harmful effect of the weights variance increase, assume the best possible choice of the proposal pdf, i.e. a proposal pdf very close, in probabilistic terms, to the posterior density. Here, we modify equation 4.9 as follows:

$$\begin{aligned} w_{0:t} &= \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{0:t})p(\mathbf{x}_{0:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\ &= \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})p(\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \\ &\propto \frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})} \end{aligned} \tag{4.22}$$

where the last line in equation 4.22 indicates that the unnormalized importance weight is proportional to a ratio of pdfs which is called the importance ratio. The importance ratio is useful in the analysis of the mean and the variance of the importance weights. The mean is given by:

$$E_{q(.|\mathbf{y}_{1:t})}\left[\frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}\right] = 1 \tag{4.23}$$

and the variance:

$$E_{q(.|\mathbf{y}_{1:t})}\left[\left(\frac{p(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}{q(\mathbf{x}_{0:t}|\mathbf{y}_{1:t})}-1\right)^2\right]=0 \tag{4.24}$$

Equation 4.24 indicates that for the variance to be close to zero, we need a proposal very close to the posterior. However, the best possible proposal, which is:

$$q(\mathbf{x}_t|\mathbf{y}_{1:t})=p(\mathbf{x}_t|\mathbf{y}_{1:t}), \tag{4.25}$$

is difficult to be obtained in practice since the posterior is unknown. Additionally, any mismatch in the importance ratio leads to a different variance than 0, thus, it is impossible to avoid the degeneracy problem.

In practice, the degeneracy is observed by monitoring the importance weights at every model time step. Normally, what we observe is that, from the entire set of particles, the normalized importance weight of only one particle tends to the value of 1 (maximum value), while the remaining weights tend to marginal values close to 0. Therefore, a large number of particles with numerical insignificant weights leads to a wrong sample approximation of the posterior.

The plots in the upper part of figure 4.1 show the importance weight transition from a uniform distribution at $t = 0$ to a normal distribution according to the Gaussian likelihood pdf at $t = 1$. While in the plots located in the lower part, it is clearly noticeable that after a few model time steps, only one of the normalized importance weights reaches the value of 1, and the remaining set of weights are reduced to negligible values. Consequently, a large number of samples are removed from the sample space, because their weights become numerically insignificant, generating a wrong approximation of the posterior pdf.

There exists some approaches for the mitigation of the degeneracy problem. An heuristic approach which is the increase of the numbers of particles N, is impractical since the particle set size should be increased considerably with a corresponding increase in the computational cost. Two approaches which are feasible to be implemented in practice and treated in this manuscript, concern a proper choice of the proposal distribution and the resampling of the weighted particle set.

4.4.3 Choice of the proposal distribution

The selection of the proposal $q(\mathbf{x}_t|\mathbf{x}^i_{1:t-1}, \mathbf{y}_{1:t})$ is important in the design stage of the SIS filter. The filter performance mainly depends on how well

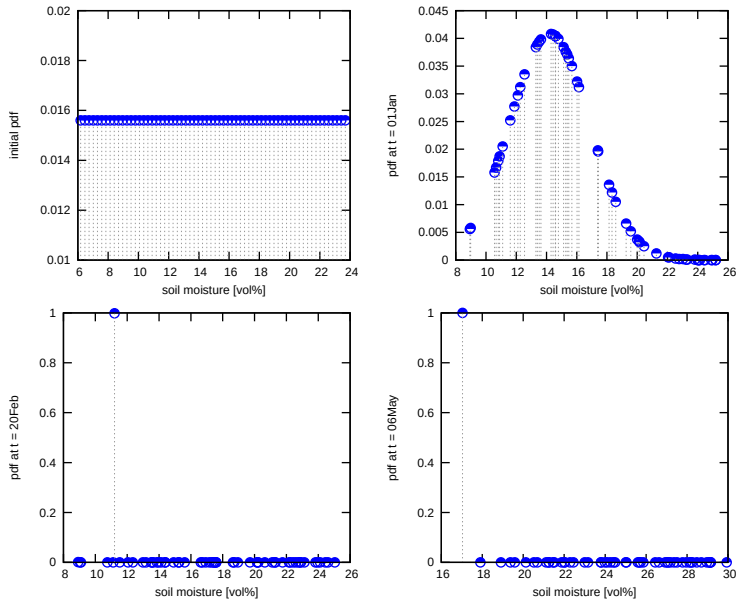


Figure 4.1: Monitoring the importance weights: axis x represents the particles location (volumetric soil moisture [vol%] with $N = 64$) and axis y the importance weights values at four different daily model time steps 0, 1 (01/Jan), 51 (20/Feb), 126 (06/May).

the proposal approximates the posterior. In Doucet et al. (2000), an optimal choice for the proposal density function is proposed

$$q(\mathbf{x}_t|\mathbf{x}_{1:t-1,i}, \mathbf{y}_{1:t})_{opt} = p(\mathbf{x}_t|\mathbf{x}_{t-1,i}, \mathbf{y}_t) \tag{4.26}$$

This optimal importance distribution has also been recommended by other researchers (Kong et al., 1994; Liu and Chen, 1995). The proposal distribution $p(\mathbf{x}_t|\mathbf{x}_{t-1,i}, \mathbf{y}_t)$ is optimal in the sense that it minimizes the variance of the importance weights conditionally upon $\mathbf{x}_{1:t-1,i}$ and $\mathbf{y}_{1:t}$. However, the application of this type of pdf (equation (4.26)) is complex from the implementation point of view. It is possible to sample from the optimal proposal when the system is defined by a Jump-Markov linear system. An application of such system is reported in Doucet et al. (2001) where the particle filter is applied along with the Kalman filter. Analytical evaluation is possible when the proposal pdf is Gaussian (Doucet et al., 2001; Moral, 1996). This is the case when the process model is non-linear but the observation model is linear. For complex models, the analytical evaluation of the optimal proposal pdf is not possible. However, it is possible to apply suboptimal approximations to the proposal pdf. Moreover, the optimal pdf is assumed to be Gaussian with statistic parameters obtained through a local linearization either by a Taylor series expansion (Doucet et al., 2000) or

by the Unscented transform (Van Der Merwe et al., 2001). The discussion about approximations to the optimal proposal pdf is extended in the next chapter.

A common choice of the proposal is the transition prior function. This proposal has been used and analyzed in different studies (Berzuini et al., 1997; Gordon et al., 1993; Kitagawa, 1996; Liu and Chen, 1998)

$$q(\mathbf{x}_t|\mathbf{x}_{1:t-1,i}, \mathbf{y}_{1:t}) = p(\mathbf{x}_t|\mathbf{x}_{t-1,i}) \tag{4.27}$$

The choice of the transition prior as the proposal simplifies equation (4.18) resulting in an expression where the importance weights depend on their past values and on the likelihood $p(\mathbf{y}_t|\mathbf{x}_{t,i})$. A common choice of the likelihood density function is the Gaussian distribution that describes the misfit between the observation predictions and the observations, scaled by the (usually a priori defined) observation error.

The consequence of simplifying the optimal proposal pdf according to Equation 4.27 is the magnification of the degeneracy of the importance weight values, which in turn degenerates the overall performance of the SIS filter. The high Monte Carlo (MC) variation in the weights leads to a depletion of the particle set which can be mitigated by the suppression of the particles with small importance weights and the replication of those with large importance weights. The latter is obtained by applying resampling with replacement to the particle set. Note that in SIS, the state variables are not updated, i.e. only the weights are updated.

4.5 Resampling

The aim of the resampling strategy is to apply methodologies (Monte Carlo methods) which should map the weighted Dirac random measure $\{\mathbf{x}_{t,i}, \mathbf{w}_{t,i}\}$ into an equally weighted random measure $\{\mathbf{x}_{t,i}, N^{-1}\}$. A first attempt to resampling is the Sequential Importance Resampling (SIR) method (Gordon et al., 1993) which has the following rationale. The SIS filter generates a random sample approximation of the posterior pdf. For any function of interest $g(\mathbf{x}_t)$, the following expression is valid according to the law of large numbers:

$$\lim_{N\to\infty} \int g(\mathbf{x}_t) p_N(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \to \int g(\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \tag{4.28}$$

If the aim is to obtain an unweighted collection of particles which approximate the distribution $p_N(\mathbf{x}_t|\mathbf{y}_{1:t})$, then crude Monte Carlo approximation can be applied. This is carried out by sampling (Bootstrap resampling) N new particles $\tilde{\mathbf{x}}_{t,i}$, from the empirical estimate of the posterior, i.e

$\tilde{\mathbf{x}}_{t,i} \sim p_N(\mathbf{x}_t|\mathbf{y}_{1:t})$, and then using the simple MC estimate of the integral. It is then straightforward to apply the law of large numbers to verify that:

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} g(\tilde{\mathbf{x}}_{t,i}) \to \int g(\mathbf{x}_t) p_N(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \tag{4.29}$$

The main disadvantage of this approach is that it is generally recommended to use as many particles as computational resources permit in order to compensate the additional Monte Carlo errors introduced by the additional resampling step. Moreover, the SIR methodology may be affected by the introduction of bias and the large MC variance of the resampled particle set.

Since resampling involves the replication of particles according to their importance weights, we could alternatively consider the number of replicates of each of the original particles which are present in the resampled set. Let $M_i = \{j : \mathbf{x}^*_{t,j} = \mathbf{x}_{t,i}\}$ be the number of replicates of $\mathbf{x}_{t,i}$ which are present in the resampled set. Then, a resampling strategy is not affected by bias when the expected number of replicates of a particle after resampling is precisely the weight associated with the sample, so that

$$E\left[\frac{M_i}{N}|\mathbf{x}_{t,1:N}\right] = w_t(\mathbf{x}_{t,i}) \tag{4.30}$$

There are efficient selection techniques in terms of a reduction of the resampled particles variance with unbiased sampling which have been developed such as: Multinomial Resampling (MR), the Stratified resampling (StrR) (Carpenter et al., 1999), Systematic resampling (SysR) (Kitagawa, 1996) and Residual resampling (ResR) (Higuchi, 1997; Liu and Chen, 1998). A theoretical description of these resampling strategies and their characteristics is presented in Douc et al. (2005); Robert and Casella (1999). In this section, we present a brief review of these techniques.

4.5.1 Multinomial resampling (MulR)

The resampling approach which has been described previously, in which each $\mathbf{x}^*_{t,i}$ is sampled independently from the empirical distribution with the collection $\{w_t(\mathbf{x}_{t,i}), \mathbf{x}_{t,i}; i = 1, \ldots, N\}$ is equivalent to drawing the vector of replicate counts M_i from a multinomial distribution with N trials and parameter vector $w_t(\mathbf{x}_{t,1:N})$, so that:

$$M \sim \mathcal{M}(M_i|N, w_t(\mathbf{x}_{t,1:N})) \tag{4.31}$$

with the multinomial distribution having a probability mass function given by:

$$\mathcal{M}(M_i|N, w_t(\mathbf{x}_{t,1:N})) = \begin{cases} \dfrac{N!}{\prod_{i=1}^{N} M_i!} \prod_{i=1}^{N} [w_t(\mathbf{x}_{t,i})]^{M_i} & \text{if } \sum_{i=1}^{N} M_i = N, \\ 0 & \text{otherwise.} \end{cases} \tag{4.32}$$

Due to the similarity between the SIR and the multinomial sampling, which is described above, both strategies are referred to as multinomial resampling. The variance of the resampled set obtained from multinomial resampling is $var(M_i) = Nw_t(\mathbf{x}_{t,i})[1 - w_t(\mathbf{x}_{t,i})]$. According to Carpenter et al. (1999) and Liu and Chen (1998), it is possible to design resampling schemes with lower variance.

4.5.2 Residual resampling (ResR)

The main reason why resampling is not carried out deterministically is that in general $Nw_t(\mathbf{x}_{t,i})$ is not an integer. However, it is possible to remove the integer component of $Nw_t(\mathbf{x}_{t,i})$ for each weight and then to assign the remainder (residual) of the mass by multinomial resampling. This is a brief description of the basis of the strategy known as residual resampling.

The rationale behind RR is a combination of deterministic replication of those particles that we expect at least one replicate in the resample set and introducing randomness only to allow as to deal with non-integer part of $Nw_t(\mathbf{x}_{t,i})$, the result is the retention of unbiased behavior from multinomial resampling while substantially reducing the degree of randomness introduced by the overall procedure. Moreover, with this approach, the variance is smaller than the one obtained from SIR and the procedure is computationally cheaper.

4.5.3 Stratified (StrR) and Systematic Resampling (SysR)

Another approach is motivated by the stratified sampling technique. The approach consists of the partition of the empirical distribution by taking the Cumulative Distribution Function (CDF), dividing it into a number of equally weighted segments and then, drawing one sample from each segment. The drawing of the sample can be carried out by drawing a uniform random variable from the range of each of the CDF partitions and applying inversion sampling. Furthermore, it is possible to reduce the randomness in this strategy, by drawing a single uniform random number and using that

to select the value sample from all of the segments of the CDF. This technique is widely used in practice and is known as systematic resampling. In order to implement the systematic resampling, one simply draws a random number from a uniform distribution, U, in the interval $[0, 1/N]$ and then the resampled set is obtained by applying inversion sampling as follows:

$$\mathbf{x}^*_{t,i} := F_N^{-1}(i/N - U), \quad \text{for each } i \text{ in } 1, \ldots, N. \tag{4.33}$$

with F_N denoting the CDF of the empirical distribution. This strategy introduces a variance on M_i even smaller than the residual resampling scheme with low computational cost.

4.5.4 Standard particle filter

A generic particle filter can be constructed. This filter is based on the SIS filter and the additional resampling step. In this manuscript, the generic particle filter is referred to as Standard Particle Filter (SPF). The term standard comes from the implementation point of view. The SPF is easy to implement since it follows the assumptions considered in the natural SIS filter (Algorithm 6) and the systematic resampling step is added in order to tackle the degeneracy problem with minimum MC resampling variation and low computational cost. The SPF is described in Algorithm 7.

Algorithm 7 Standard Particle Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_{t,i} \sim p(\mathbf{x}_t|\mathbf{x}^*_{t-1,i}), \quad i = 1, 2, \ldots, N$
2. Analysis step:
   $w_t(\mathbf{x}_{t,i}) = p(\mathbf{y}_t|\mathbf{x}_{t,i}) \quad i = 1, 2, \ldots, N$
   $\tilde{w}_t(\mathbf{x}_{t,i}) = \dfrac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$
   Obtain the resample set $\mathbf{x}^*_{t,i}$ according to SysR.
   $\overline{E[\mathbf{x}_t]} = \dfrac{1}{N}\sum_{i=1}^{N} \mathbf{x}^*_{t,i}$

The additional systematic resampling step maps the weighted sample into an equally weighted sample. Consequently, the algorithm does not take into account the past values of the importance weights and the conditional mean of the particles is given by the sample mean. The performance of the SPF filter can be affected by a possible extreme particle resampling. This problem, which is known as sample impoverishment, is severe in the case of small process noise where all the particles will collapse to a single

particle value within few iterations. For the estimation of smoothing distributions, the problem is magnified as reported in Godsill et al. (2004) while in the estimation of filtering distributions, the application of Markov chain Monte Carlo (MCMC) mitigates the problem by introducing variability to the particle set (Carlin et al., 1992).

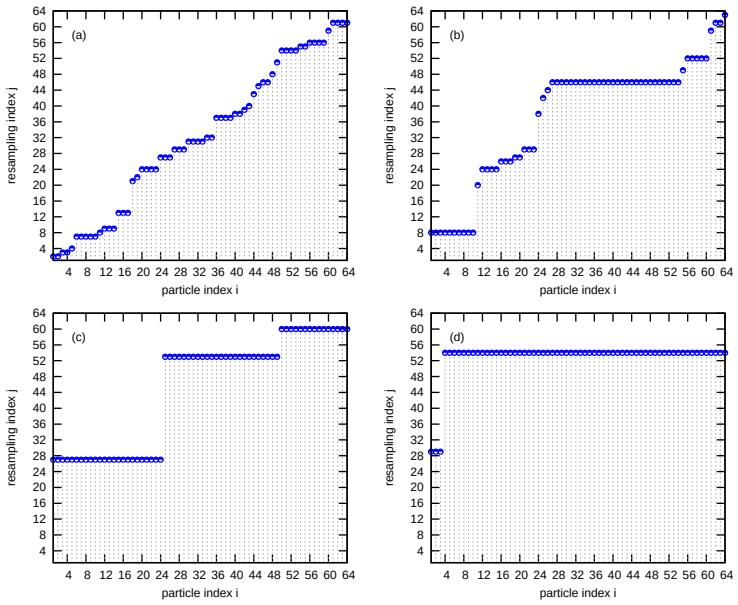


Figure 4.2: Extreme particle replication example with $N = 64$: Figures (a) and (b) present a proper resampling performance whereas figures (c) and (d) show the the extreme replication problem. The sample set collapses to particles located at 27th, 53th and 60th positions in figure (c) and almost all the set collapses to the value of the particle at the 54th position in figure (d).

Figure 4.2 shows the resampling index i, which indicates the location of the particles to be resampled, at 4 data assimilation events. Subfigures (a) and (b) indicate a proper performance of the resampling algorithm where the particle replication is not extreme. On the other hand, the resampling index i, as a result of the application of a hypothetical and too narrow likelihood pdf, is presented in subfigures (c) and (d), subfigure (c) indicates that the sample collapses to the particle values located at positions 27th, 53th, and 60th. The extreme replication problem is noticeable mostly in subfigure (d) where almost all the particles collapse to the value of the particle located at the 54th position.

4.6 Resample-move

An approach to mitigate the impoverishment of the particles is by applying a resample-move step to the resampled set (Gilks and Berzuini, 2001; Doucet et al., 2001; Fearnhead, 2002). Resample-move consists of the application of MCMC along with SMC algorithms. MCMC methods are traditionally used when random samples from complex or multidimensional probability distributions are needed. The methodology consists of the construction of Markov chains through the generation of collections of correlated samples which approximates a target distribution.

In the context of particle filters, the MCMC step is applied as a way to introduce particle variability and thus reducing the depletion of the resampled particles. The main idea is to construct a Markov transition kernel $\kappa(\mathbf{x}^*_{1:t}|\mathbf{x}_{1:t})$ of invariant distribution $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ with the following property:

$$\int p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})\kappa(\mathbf{x}^*_{1:t}|\mathbf{x}_{1:t})dx_{1:t} = p(\mathbf{x}^*_{1:t}|\mathbf{y}_{1:t}) \tag{4.34}$$

For this Markov kernel, if the resampled particles $\mathbf{x}_{1:t}$ are distributed according to the posterior then the new particle set $\mathbf{x}^*_{1:t}$ is still distributed according to $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ with the additional fact that the obtained particle set might have more diversity. Even in the case when the set $\mathbf{x}_{1:t}$ is not distributed according to the posterior, the application of the MCMC step assures that the new set can only have a distribution closer to the posterior.

In order to construct a Markov kernel, the Gibbs sampler or the Metropolis Hasting (MH) algorithms can be used. It is well known that the MH approach has an extra degree of freedom since this method allows for the sampling of the candidates according to some proposal and accept the candidate with the MH acceptance probability α. For the particular case of the SPF where the prior is identical to the proposal, the idea is to sample candidates from the transition prior and accept according to the following MH probability:

$$\alpha = min\left(1, \frac{p(\mathbf{y}_t|\mathbf{x}^*_t)}{p(\mathbf{y}_t|\mathbf{x}_t)}\right) \tag{4.35}$$

According to Doucet and Johansen (2009), the condition of ergodicity regarding the resample-move kernel is not longer required in order to be able to implement efficient recursive particle MCMC algorithms.

In this paper, the SPF with MCMC is applied to a rainfall-runoff model in order to analyze the performance and compare it to other approaches. The standard particle filter with MCMC move step is presented in Algorithm 8. This algorithm is the result of the implementation of independent MCMC steps on each resampled particle along with the standard particle filter. A

Algorithm 8 Standard Particle Filter - Resample Move

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_{t,i} \sim p(\mathbf{x}_t|\mathbf{x}^*_{t-1,i}), \quad i = 1, 2, \ldots, N$
2. Analysis step:
   $w_t(\mathbf{x}_{t,i}) = p(\mathbf{y}_t|\mathbf{x}_{t,i}) \quad i = 1, 2, \ldots, N$
   $\tilde{w}_t(\mathbf{x}_{t,i}) = \dfrac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$
   Obtain the resample set $\mathbf{x}^*_{t,i}$ according to SR.
   Apply MCMC move step: sample $\mathbf{x}_{t,i} \sim \kappa(\mathbf{x}^*_{t,i}|\mathbf{x}_{t,i})$.
   – Sample $v \sim \mathcal{U}[0, 1]$
   – Sample the proposal candidates $\mathbf{x}^*_{t,i} \sim p(\mathbf{x}_t|\mathbf{x}_{t-1,i})$
   – Compute the acceptance probability $\alpha = min\left(1, \frac{p(\mathbf{y}_t|\mathbf{x}^*_{t,i})}{p(\mathbf{y}_t|\mathbf{x}_{t,i})}\right)$
   – Set the state according to $\mathbf{x}_{t,i} = \begin{cases} \mathbf{x}^*_{t,i} & v \leq \alpha \\ \mathbf{x}_{t,i} & \text{otherwise} \end{cases}$
   $\overline{E[\mathbf{x}_t]} = \dfrac{1}{N}\sum_{i=1}^{N} \mathbf{x}^*_{t,i}$

possible drawback of the methodology is that a limited MCMC proposal comes with limited MCMC candidates to explore areas in the state-space that could possibly lead to more accurate estimates.

4.7 Gaussian Particle Filter (GPF)

Kotecha and Djuric (2003a) introduced the Gaussian Particle Filter (GPF). Basically, GPF approximates the mean and covariance of the state vector involved in the estimation by using importance sampling. The strengths of this approach are: non-Gaussian and nonadditive noise applications and unlike the SIS filter, resampling is not required.

The approach adopted in the GPF is the approximation of the filtering and predictive pdf by Gaussian pdfs. Basically, a different strategy is used by the GPF in the representation of the importance weights. Based on the expression of the expectation given by Equation 4.7 and considering the common case when only the filtered density is required, the modified equation is presented as follows:

$$E[g(\mathbf{x}_t)] = \int g(\mathbf{x}_t)\frac{p(\mathbf{x}_t|\mathbf{y}_{1:t})}{q(\mathbf{x}_t|\mathbf{y}_{1:t})}q(\mathbf{x}_t|\mathbf{y}_{1:t})d\mathbf{x}_t \tag{4.36}$$

From Equation 4.36, the filtering pdf is elaborated by using the Bayes'

theorem:

$$\begin{aligned} E[g(\mathbf{x}_t)] &= \int g(\mathbf{x}_t) \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_t)p(\mathbf{x}_t)}{p(\mathbf{y}_{1:t})q(\mathbf{x}_t|\mathbf{y}_{1:t})} q(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \\ &= \int g(\mathbf{x}_t) \frac{p(\mathbf{y}_t|\mathbf{x}_t,\mathbf{y}_{1:t-1})p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})q(\mathbf{x}_t|\mathbf{y}_{1:t})} q(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \\ &= \int g(\mathbf{x}_t) \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})q(\mathbf{x}_t|\mathbf{y}_{1:t})} q(\mathbf{x}_t|\mathbf{y}_{1:t}) d\mathbf{x}_t \end{aligned} \tag{4.37}$$

where similar as in section 4.3 and from the last line of Equation 4.37, the expression for the unnormalized importance weights in the GPF formulation is the following:

$$w_t = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_t|\mathbf{y}_{1:t})} \tag{4.38}$$

According to Equation 4.38, the importance weights are determined by the likelihood, predictive and importance pdfs. Here, assuming that it is possible to draw N particles $\mathbf{x}_{t,i}$ from the proposal pdf, then the importance weights of the particles are given by:

$$w_{t,i} = \frac{p(\mathbf{y}_t|\mathbf{x}_{t,i})p(\mathbf{x}_{t,i}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_{t,i}|\mathbf{y}_{1:t})} \tag{4.39}$$

The normalization of the importance weights is carried out similarly as in Equation 4.14 The calculation of the importance weights faces the issue that the shape of the predictive pdf is unknown. However, at this point, it is assumed that the evaluation of the particles according to the predictive pdf is possible. Then, an empirical measure of the posterior pdf can be obtained by means of the set of particles along with the importance weights $\{\tilde{w}_{t,i}, \mathbf{x}_{t,i};\ i = 1, \ldots, N.\}$. Although the point-mass representation of the posterior is valid for any kind of distribution when the number of samples is large enough, in the GPF, the posterior is approximated by a Gaussian distribution. This approximation restricts the applicability of the GPF to the Gaussian scenario. Nevertheless, Kotecha and Djuric (2003a) demonstrated the application of the GPF in possible non-Gaussian scenarios. In this dissertation, the GPF is used in the estimation of hydrological states from a conceptual hydrologic model.

For the approximation of the posterior pdf, the sample mean and covariance are computed from the weighted set as follows:

$$\hat{\mathbf{x}}_t = \sum_{i=1}^{N} \tilde{w}_{t,i} \mathbf{x}_{t,i} \tag{4.40}$$

$$\hat{\mathbf{P}}_t = \sum_{i=1}^{N} \tilde{w}_{t,i} (\mathbf{x}_{t,i} - \hat{\mathbf{x}}_t)(\mathbf{x}_{t,i} - \hat{\mathbf{x}}_t)^T \tag{4.41}$$

The final step in the representation of the posterior pdf is the generation of a new set of particles which should follow the Gaussian distribution with parameters $\hat{\mathbf{x}}_t$ and $\hat{\mathbf{P}}_t$:

$$\mathbf{x}^*_{t,i} \sim \mathcal{N}(\mathbf{x}^*_{t,i}; \hat{\mathbf{x}}_t, \hat{\mathbf{P}}_t) \tag{4.42}$$

with $\mathbf{x}^*_{t,i}$, the notation used to distinguish the new particle set. Note that the same notation used for the resampled set in Algorithm 7 is applied in Equation 4.42 since the sampling from a Gaussian pdf can be interpreted as a resampling step from a continuous density function.

The Gaussian representation of the posterior allows for an approximation of the predictive pdf which was assumed to be known when evaluating the importance weights 4.39. The predictive pdf is given by:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1} \tag{4.43}$$

with the posterior at time step $t-1$ approximated by the Gaussian distributed set: $\mathbf{x}^*_{t-1,i}$:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) \mathcal{N}(\mathbf{x}^*_{t-1,i}; \hat{\mathbf{x}}_{t-1}, \hat{\mathbf{P}}_{t-1}) d\mathbf{x}_{t-1} \tag{4.44}$$

The set of particles in Equation 4.44 allows for the approximation of the integral by applying Monte Carlo as follows:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \frac{1}{N} \sum_{i=1}^{N} p(\mathbf{x}_{t,i}|\mathbf{x}^*_{t-1,i}) \tag{4.45}$$

The predictive pdf is approximated by a Gaussian pdf as follows:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{N}(\mathbf{x}_t; \bar{\mathbf{x}}_t, \bar{\mathbf{P}}_t) \tag{4.46}$$

with the first two moments of the Gaussian pdf given by:

$$\bar{\mathbf{x}}_t = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{t,i} \tag{4.47}$$

$$\bar{\mathbf{P}}_t = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x}_{t,i} - \bar{\mathbf{x}}_t)(\mathbf{x}_{t,i} - \bar{\mathbf{x}}_t)^T \tag{4.48}$$

Note that the approximation of the predictive pdf can be used in the evaluation of the importance weights and no additional sampling is involved unlike the approximating of the posterior pdf in which a new particle set is sampled from the Gaussian posterior pdf. Finally, the formulation of the GPF is summarized in Algorithm 9 where the implementation of the GPF is referred to as General Gaussian Particle filter.

Algorithm 9 General Gaussian Particle Filter

For $t = 1$ to the number of time steps

1. Forecast step:
   $\mathbf{x}_{t,i} \sim q(\mathbf{x}_{t,i}|\mathbf{y}_{1:t}), \quad i = 1, 2, \ldots, N$
2. Analysis step:
   $w_t(\mathbf{x}_{t,i}) = \dfrac{p(\mathbf{y}_t|\mathbf{x}_{t,i})p(\mathbf{x}_{t,i}|\mathbf{y}_{1:t-1})}{q(\mathbf{x}_{t,i}|\mathbf{y}_{1:t})} \quad i = 1, 2, \ldots, N$
   $\tilde{w}_t(\mathbf{x}_{t,i}) = \dfrac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$
   $\overline{E[\mathbf{x}_t]} = \sum_{i=1}^{N} \mathbf{x}_{t,i}\tilde{w}_t(\mathbf{x}_{t,i})$

Finally, a simple choice of the importance density is the predictive pdf:

$$q(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) \approx \mathcal{N}(\mathbf{x}_t; \bar{\mathbf{x}}_t, \bar{\mathbf{P}}_t) \tag{4.49}$$

As a result of this approximation, the computation of the importance weights is simplified with the likelihood pdf as the unique pdf to be evaluated in Equation 4.39. In this context, the particles are weighted similarly as in the case of the standard particle filter. Moreover, the performance of the GPF can be affected by the simple choice of the proposal pdf since the proposal pdf is not optimal with the presence of a high Monte Carlo variation in the point-mass representation of the posterior pdf. On the other hand, the performance of the GPF is not limited neither by the degeneracy of the importance weights nor by the impoverishment of the samples due to extreme resampling. The latter two characteristics make the GPF an interesting methodology which is analyzed in the next section. A natural implementation of the GPF methodology is summarized in Algorithm 10.

Algorithm 10 Natural Gaussian Particle Filter

For $t = 1$ to the number of time steps

1. Prediction step:
   $\mathbf{x}_{t,i} \sim p(\mathbf{x}_t|\mathbf{x}^*_{t-1,i}), \quad i = 1, 2, \ldots, N$
   Compute $\bar{\mathbf{x}}_t$ and $\bar{\mathbf{P}}_t$ in case $q(\mathbf{x}_t|\mathbf{y}_{1:t}) \neq p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$.
2. Correction step:
   $w_{t,i} = p(\mathbf{y}_t|\mathbf{x}_{t,i}) \quad i = 1, 2, \ldots, N$
   $\tilde{w}_{t,i} = \dfrac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$
   Compute $\hat{\mathbf{x}}_t$, $\hat{\mathbf{P}}_t$ and obtain the resample set $\mathbf{x}^*_{t,i}$ according to Eq (4.42).
   $\overline{E[\mathbf{x}_t]} = \dfrac{1}{N}\sum_{i=1}^{N} \mathbf{x}^*_{t,i}$

4.8 The ensemble Gaussian particle filter

4.8.1 Improving the importance density function

The choice of the proposal distribution is one of the critical design issues in particle filters. A proper performance of the PF is expected when the following key assumptions are valid: the point-mass approximation should represent the posterior distribution adequately and the proposal distribution should approximate the posterior distribution as accurately as possible (Arulampalam et al., 2002).

In case the first assumption is not completely valid, the MCMC move step has been proposed as a methodology to increase the spread of particles improving the resolution of the particle set and the corresponding point-mass representation of the posterior.

For the second assumption, some approaches have been reported in the literature, e.g. the auxiliary particle filter (APF) (Pitt and Shephard, 1999), regularized particle filter (RPF) (Musso et al., 2001) and the unscented particle filter (UPF) (Van Der Merwe et al., 2001) among others which are derived from these techniques.

In the APF, approximated samples from the optimal importance density are obtained by using an auxiliary variable, whilst in the RPF, samples are obtained from a continuous approximation of the posterior rather than from a discrete density improving the performance of the resampling step.

The UPF belongs to a set of techniques which approximate the optimal importance density by incorporating the current observation with the optimal Gaussian approximation of the state. In this context, the analysis statistics from extended Kalman filter and the unscented Kalman filter are valid approximations to the optimal proposal. In the UPF, the optimal proposal is approximated as follows.

$$q(\mathbf{x}_t|\mathbf{x}_{1:t-1,i}, \mathbf{y}_{1:t})_{opt} = \mathcal{N}(\mathbf{x}_{t,i}; \bar{\mathbf{x}}_t, \bar{\mathbf{P}}_t) \tag{4.50}$$

The samples $\{\mathbf{x}_{t,i} : i = 1, \ldots, N\}$ are drawn from a Gaussian distribution with mean $\bar{\mathbf{x}}_t$ and covariance $\bar{\mathbf{P}}_t$ given by the unscented Kalman filter and computed for every $i - th$ particle

In the same line of optimal proposals, the EnKF (Evensen, 1994) has shown high efficiency in terms of accuracy and computational time demand as a nonlinear filter outperforming the extended and unscented Kalman filters in most cases. Therefore, a proper combination of the EnKF and the particle filter assures a higher performance over the standard particle filter and ensemble Kalman filter. In the geophysical sciences, examples of this combination correspond to: the adaptive Gaussian mixture filter (Hoteit et al.,

2008; Andreas et al., 2011), the weighted ensemble Kalman filter (Papadakis et al., 2010) and the particle Kalman filter (Hoteit et al., 2012) among others.

In this dissertation, we modify the structure of the Gaussian particle filter (GPF) (Kotecha and Djuric, 2003a) by the inclusion of the EnKF to provide the importance density function. The GPF is selected as the particle filter structure based on some interesting features which are discussed below. The combination of the GPF and the EnKF is referred to as the Ensemble Gaussian Particle Filter (EnGPF).

4.8.2 Development of the ensemble Gaussian particle filter (EnGPF)

As presented in section 4.7, the GPF structure has interesting features. In this dissertation, the selection of the EnKF to provide the proposal distribution is the major contribution to the original GPF algorithm.

In GPF, the prior $p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ and posterior $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ density functions involved in the analysis step are considered as Gaussian distributions. The considerations make it possible to simplify the computation of the importance weights. Moreover, the importance weights in the GPF methodology are directly obtained from the importance sampling approach unlike the SIS method where the recursive expression of the weights is used.

The EnGPF is the result of the combination between the EnKF and the GPF. Therefore, the first main step in the structure of the EnGPF is the application of the EnKF. In the EnKF, the forecast ensemble ($\{\mathbf{x}_{t,i}^f; i = 1, 2, \ldots, N\}$) represents the a priori knowledge about the state of the system while the analysis ensemble ($\{\mathbf{x}_{t,i}^a; i = 1, 2, \ldots, N\}$) corresponds to the a posteriori knowledge (after correction). The formulation of the EnKF is derived in section 3.4 and the equations are presented in Algorithm 4.

The second main step in the EnGPF structure is the application of the GPF. This step can be interpreted as an extension of the analysis step from the EnKF where the particles are weighted according to the importance sampling approach and the predictive and filtering pdfs are approximated as Gaussian distributions.

In the EnGPF, the unnormalized importance weights are given by:

$$w_{t,i} = \frac{p(\mathbf{y}_t|\mathbf{x}_{t,i}) \left[p(\mathbf{x}_{t,i}|\mathbf{y}_{1:t-1}) = \mathcal{N}\left(\mathbf{x}_{t,i}; \mathbf{x}_t^f, \mathbf{P}_t^f\right)\right]}{q(\mathbf{x}_{t,i}|\mathbf{y}_{1:t})} \tag{4.51}$$

where $\{\mathbf{x}_{t,i}; i = 1, \ldots, N\}$ are particles drawn from the importance density function $q(\mathbf{x}_{t,i}|\mathbf{y}_{1:t})$ and the parameters $\mathbf{x}_t^f$, $\mathbf{P}_t^f$,which are used in the

approximation of the prior, are obtained from the transition prior density function $p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$ (same as the forecast step in the EnKF) as follows:

$$\mathbf{x}_t^f = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{t,i}^f \tag{4.52}$$

$$\mathbf{P}_t^f = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_{t,i}^f - \mathbf{x}_t^f)(\mathbf{x}_{t,i}^f - \mathbf{x}_t^f)^T \tag{4.53}$$

Here, the analysis ensemble from the EnKF is used in order to represent the particles $\mathbf{x}_{t,i}$ along with the sample mean $\mathbf{x}_t^a$ and sample covariance $\mathbf{P}_t^a$ of the particle set. Therefore, the proposal distribution can be approximated as a Gaussian distribution as follows:

$$q(\mathbf{x}_{t,i}|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_{t,i}; \mathbf{x}_t^a, \mathbf{P}_t^a) \tag{4.54}$$

The particles are weighted according to Equation (4.51). This expression involves the evaluation of the likelihood pdf with the predictive and importance pdfs represented by Gaussian pdfs. After the computation of the importance weights, the posterior can be approximated as a Gaussian distribution as follows:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \mathcal{N}(\mathbf{x}_{t,i}^*; \widehat{\mathbf{x}}_t, \widehat{\mathbf{P}_t}) \tag{4.55}$$

where $\{\mathbf{x}_{t,i}^*; i = 1, 2, \ldots, N\}$ is a new set of particles sampled from the Gaussian distribution with parameters: $\widehat{\mathbf{x}}_t$, $\widehat{\mathbf{P}_t}$. $\widehat{\mathbf{x}}_t$ is the weighted mean and $\widehat{\mathbf{P}_t}$ is the weighted covariance which are computed from the particle set as follows.

$$\widehat{\mathbf{x}}_t = \sum_{i=1}^{N}\mathbf{w}_{t,i}\mathbf{x}_{t,i} \tag{4.56}$$

$$\widehat{\mathbf{P}_t} = \sum_{i=1}^{N}\mathbf{w}_{t,i}(\mathbf{x}_{t,i} - \widehat{\mathbf{x}}_t)(\mathbf{x}_{t,i} - \widehat{\mathbf{x}}_t)^T \tag{4.57}$$

From the implementation point of view, the approximation of the posterior involves the replacement of the particle set, which is obtained from the application of the EnKF, by a new particle set which is generated according to a Gaussian distribution with parameters $\widehat{\mathbf{x}}_t$, $\widehat{\mathbf{P}_t}$. The generation of the new particle set can be seen as a particle-move step with the particles moved to more interesting areas of the state space. The move step might introduce variability to the particles avoiding the problem of particle impoverishment thus eliminating the need of a resampling stage. Moreover, since the importance weights do not depend on their past values, the filter does not suffer from particle degeneracy.

The EnGPF algorithm is presented in Algorithm 11. A limitation of the filter performance could arise when the propagation of the mean and covariance is insufficient for the approximation of the posterior. However, the representation of the posterior by finite Gaussian mixtures overcomes this limitation by the propagation of higher moments of the distribution (Kotecha and Djuric, 2003b).

Algorithm 11 Ensemble Gaussian Particle Filter

For $t = 1$ to the number of time steps

1. Forecast step:

$$\mathbf{x}_{t,i}^{f} = \mathbf{f}_{t,t-1}(\mathbf{x}_{t-1,i}^{a}), \quad i = 1, 2, \ldots, N,$$

$$\mathbf{x}_{t}^{f} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_{t,i}^{f}$$

$$\mathbf{y}_{t} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{h}_{t}(\mathbf{x}_{t,i}^{f}),$$

$$\mathbf{P}_{t}^{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{x}_{t,i}^{f} - \mathbf{x}_{t}^{f})(\mathbf{h}_{t}(\mathbf{x}_{t,i}^{f}) - \mathbf{y}_{t})^{T},$$

$$\mathbf{P}_{t}^{yy} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{h}_{t}(\mathbf{x}_{t,i}^{f}) - \mathbf{y}_{t})(\mathbf{h}_{t}(\mathbf{x}_{t,i}^{f}) - \mathbf{y}_{t})^{T},$$

2. Analysis step:

$$\mathbf{y}_{t,i}^{s} \sim \mathcal{N}(\mathbf{y}_{t} : \mathbf{y}_{t}, \mathbf{R}_{t});$$

$$\mathbf{R}_{t}^{s} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{y}_{t,i}^{s} - \mathbf{y}_{t})(\mathbf{y}_{t,i}^{s} - \mathbf{y}_{t})^{T},$$

$$\mathbf{K}_{t} = \mathbf{P}_{t}^{xy}\left(\mathbf{P}_{t}^{yy} + \mathbf{R}_{t}\right)^{-1}$$

$$\mathbf{x}_{t,i} = \mathbf{x}_{t}^{f} + \mathbf{K}_{t}(\mathbf{y}_{t,i}^{s} - \mathbf{h}_{t}(\mathbf{x}_{t}^{f})), \quad i = 1. \ldots, N.$$

$$w_{t,i} = \frac{p(\mathbf{y_t}|\mathbf{x}_{t,i})\,\mathcal{N}\left(\mathbf{x}_{t,i}; \mathbf{x}_{t}^{f}, \mathbf{P}_{t}^{f}\right)}{\mathcal{N}\left(\mathbf{x}_{t,i}; \mathbf{x}_{t}^{a}, \mathbf{P}_{t}^{a}\right)} \qquad \mathbf{w}_{t,i} = \frac{w_{t,i}}{\sum_{i=1}^{N} w_{t,i}}$$

$$\widehat{\mathbf{x}}_{t} = \sum_{i=1}^{N}\mathbf{w}_{t,i}\mathbf{x}_{t,i}$$

$$\widehat{\mathbf{P}_{t}} = \sum_{i=1}^{N}\mathbf{w}_{t,i}(\mathbf{x}_{t,i} - \widehat{\mathbf{x}}_{t})(\mathbf{x}_{t,i} - \widehat{\mathbf{x}}_{t})^{T}$$

$$\mathbf{x}_{t,i}^{*} \sim \mathcal{N}\left(\mathbf{x}_{t,i}^{*}; \widehat{\mathbf{x}}_{t}, \widehat{\mathbf{P}_{t}}\right)$$

$$\mathbf{x}_{t,i}^{a} := \mathbf{x}_{t,i}^{*}$$

In this dissertation, we assume that the introduction of an approximated optimal proposal can improve the overall performance of the GPF, thus outperforming the standard ensemble Kalman filter and possibly the SPF. The assumption is validated by a comparative study developed in chapter

5 where the performance of the EnGPF is analyzed in a hydrologic data assimilation problem.

4.9 Summary

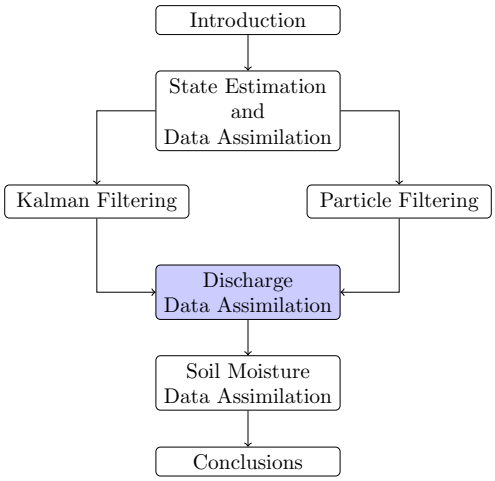
Different particle filters have been presented in this chapter, starting from the plain sequential importance sampling filter to the most sophisticated particle filter with resample-move step. It is clear that the derivation of these filters are based on the sequential importance sampling approach which presents the disadvantage of the particle degeneracy. This deficiency is overcome by the resampling of the particles. However, the particle diversity could be affected by extreme replication of the most likely particles. The particle filter with resample-move step aims to improve the diversity of the set. A general insight about the functioning of these filters is that the resampling of the particles might involve side-effects issues.

In this section, the Gaussian particle filter is presented. GPF relies on the assumption that the predictive and filtering pdfs are Gaussian distributed. This assumption limits the application of the filter to Gaussian or close to Gaussian scenarios. However, the approximation of the pdfs to Gaussian pdfs simplifies the structure of the GPF avoiding the need of the resampling of the particle set.

Finally, the formulation of the EnGPF is derived aiming to improve the functioning of the GPF in possible non-Gaussian scenarios by the adoption of an importance density function closer to the posterior pdf. Most of the techniques presented in this chapter are evaluated over hydrologic systems in the following chapters.

Chapter 5

Assimilation of discharge data



5.1 Introduction

The methodologies introduced in Chapter 3 and 4 are assessed in this chapter in a hydrologic assimilation framework. This chapter focuses on the analysis of the performance of the ensemble Gaussian particle filter. The formulation of EnGPF was introduced at the end of chapter 4 while in this chapter, the potential use of EnGPF in a non-linear/non-Gaussian estimation problem is studied. The assimilation is performed with a Rainfall-Runoff model. Every component in the hydrologic modeling process (model selection, model calibration and data assimilation) is described in this chapter.

5.2 Hydrologic Modeling

Hydrologic modeling concerns the representation of part or all of the hydrologic cycle through models which are implemented in computers. Hydrologic cycle is the name given to the movement of water in the atmosphere, hydrosphere and lithosphere. It is an intricate cycle which involves complex phenomena.

Figure 5.1 shows the hydrologic cycle. The hydrologic cycle is described as follows: water evaporates from the oceans and land and becomes a part of the atmosphere. The water vapor is either carried in the atmosphere or it returns to the earth in the form of precipitation. A portion of precipitation falling on land may be intercepted by vegetation and returned back directly to atmosphere as evaporation. Precipitation that reaches the earth may evaporate or be transpired by plants; or it may flow over the ground surface and reach streams as surface water; or it may infiltrate the soil. The infiltrated water may flow over the upper soil regions and reach surface water or it may percolate into deeper zones and become groundwater. Groundwater may reach the streams naturally or it may be pumped, used, and discarded to become a part of the surface water system. This description of the water cycle is indicated in Dooge et al. (1973)

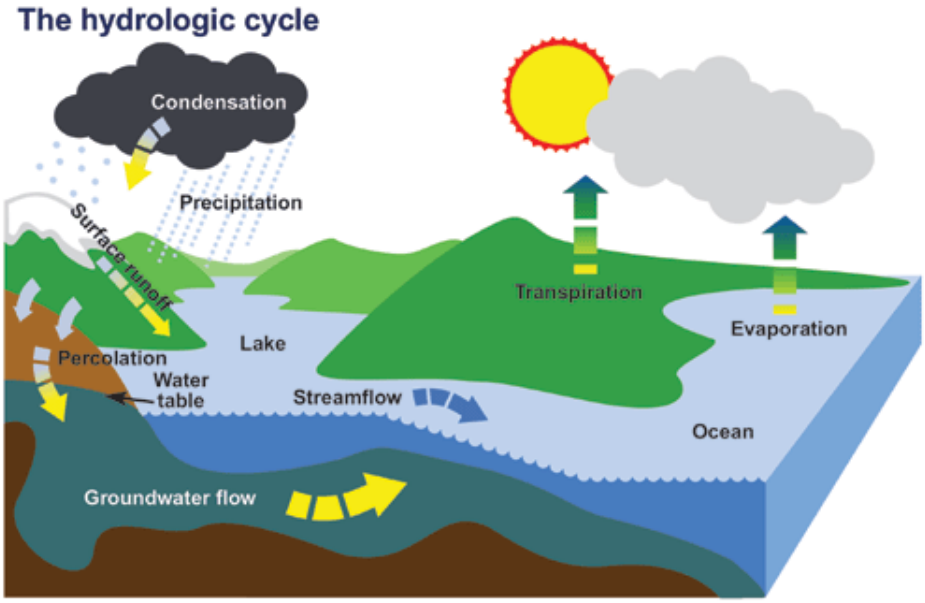


Figure 5.1: The figure shows the hydrologic cycle

Hydrologic models can be developed by following two main approaches: the scientific and the engineering approaches (Pauwels, 2010). The scientific

approach cares about a better understanding of the physical processes that influence the relationship between precipitation and discharge. While in the engineering approach, the focus is on a better prediction of discharge peaks. In the same context, hydrologic modeling can be classified into predictive modeling (engineering approach) and descriptive modeling (scientific approach). Based on experimental or theoretical aspects, hydrologic models can be subclassified from physically-based (white-box) to black-box or empirical and to conceptual models, with the most representative classification, from lumped models to distributed models. Lumped models try to relate the forcing data, mainly precipitation inputs, to system outputs (runoff) without considering the spatial variability of processes, patterns and characteristics. On the other hand, distributed models account for spatial variations of variables and parameters. A detailed review is presented in (Moradkhani and Sorooshian, 2008b; Pechlivanidis et al., 2011).

In this dissertation, a conceptual lumped model and a distributed physically-based hydrologic model are applied to two different catchments. More specific, in chapter 5, discharge data are assimilated into the HBV conceptual model. While, in chapter 6, soil moisture data are assimilated into the CLM2.0 model which is a semi-distributed physically-based model.

5.3 The Hydrologiska Byråns Vattenbalansavdelning model

5.3.1 Introduction

The Hydrologiska Byråns Vattenbalansavdelning (HBV) model was originally developed by Bergström (1976) as a semi-distributed conceptual rainfall-runoff model. HBV model covers the most important runoff generation processes using a simple and robust structure with a small number of model parameters. Lindström et al. (1997) improved the structure of the HBV model for making use of spatially distributed data. This version is referred to as HBV-96 model. A lumped and simplified version of the HBV-96 model is presented in Matgen et al. (2006). In this study, the lumped version of the HBV-96 model is adopted in order to be able to assess the performance of some data assimilation techniques presented in chapters 3 and 4.

The model inputs (driving forcings) correspond to the observed precipitation R_{tot}, the potential evapotranspiration E_{tp} and the model output is runoff Q_{dis}. The units of these flows are m^3/s.

HBV model is developed based on the general water balance equation of the

catchment which is described as:

$$\frac{ds_{tot}}{dt} = R_{tot} - E_{tp} - Q_{dis} \tag{5.1}$$

where s_{tot} is the total amount of water in the catchment in m^3. Based on the concepts of linear/non-linear reservoir modeling, the catchment is conceptually represented by three reservoirs. Figure 5.2 shows a schematic of the three reservoirs: a soil reservoir, a fast reacting reservoir, and a slow reacting reservoir. The slow flow unit characterizes the water that flows through the ground and eventually ends up in the discharge point. The fast flow unit represents the water that flows directly into the discharge point. In Figure 5.2, the arrows represent the different modeled flows and the rectangular boxes corresponds to the water storages. The equations in continuous time governing the water mass balance in each of the reservoirs are presented as follows:

$$\frac{ds_{soil}(t)}{dt} = R_{in}(t) - E_{tr}(t) - P_{er}(t) \tag{5.2}$$

$$\frac{ds_{slow}(t)}{dt} = R_{slow}(t) - Q_{slow}(t) + P_{er}(t) \tag{5.3}$$

$$\frac{ds_{fast}(t)}{dt} = R_{fast}(t) - Q_{fast}(t) \tag{5.4}$$

where t is the time in seconds and s_{soil}, s_{slow}, s_{fast} in m^3 are the states of the system. R_{tot} is the total precipitation in m^3/s and E_{tr} the actual evapotranspiration in m^3/s which is computed based on the potential evapotranspiration E_{tp} (m^3/s).

The simulated flows such as the actual evapotranspiration E_{tr}, the infiltration R_{in}, the effective precipitation R_{eff}, the percolation P_{er}, the fast reacting reservoir input R_{fast}, the output flow of the fast reacting reservoir Q_{fast}, the slow reacting reservoir input R_{slow} and the output flow of the slow reacting reservoir Q_{slow} depend on the model states and model parameters. All these flows are given in m^3/s. The detailed formulation of the HBV model is presented in the next section.

5.3.2 Implementation of the HBV model

The formulation of the HBV model starts by computing the actual evapotranspiration:

$$E_{tr}(t) = \frac{s_{soil}(t)}{\lambda S_{max}} E_{tp}(t) \tag{5.5}$$

where λ is a dimensionless parameter, and S_{max} is the storage capacity of the soil reservoir in m^3. The infiltration is calculated as:

$$R_{in}(t) = \left(1 - \frac{s_{soil}(t)}{S_{max}}\right)^b R_{tot}(t) \tag{5.6}$$

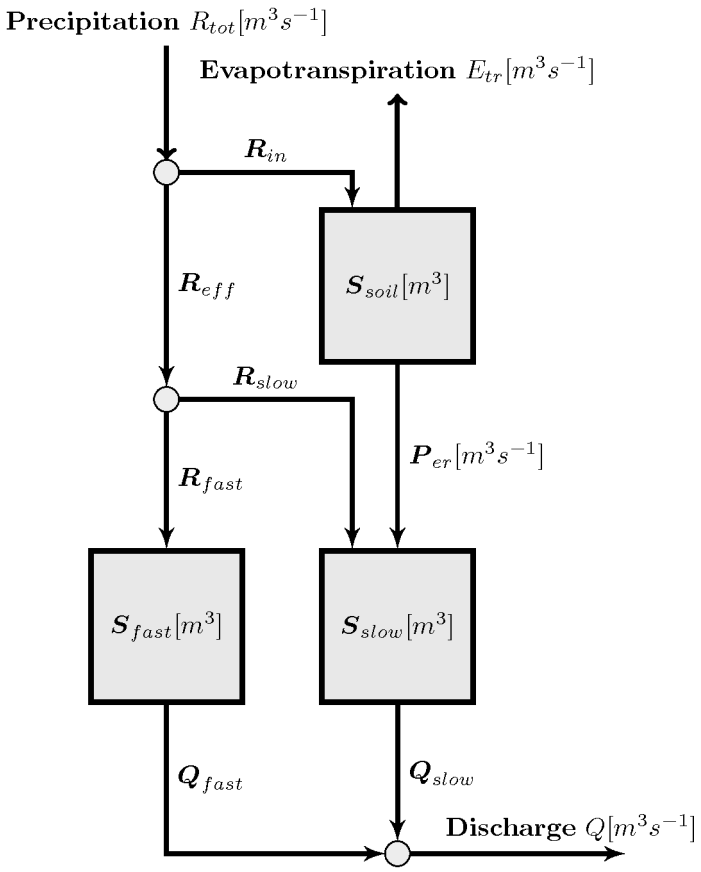


Figure 5.2: A schematic overview of the HVB rainfall-runoff model

with b a dimensionless parameter. The effective precipitation is determined as follows:

$$R_{eff}(t) = R_{tot}(t) - R_{in}(t) \tag{5.7}$$

The calculation of the percolation is presented in the following equation:

$$P_{er}(t) = P\left(1 - e^{-\beta\frac{s_{soil}(t)}{S_{max}}}\right) \tag{5.8}$$

with P a percolation parameter in m^3/s and β a dimensionless parameter. At this point, all the flows involved in the differential equation (5.2) have been computed. Thus, Equation 5.2 can solved by using the Euler integration method:

$$\frac{s_{soil}(t+\triangle t) - s_{soil}(t)}{\triangle t} = R_{in}(t) - E_{tr}(t) - P_{er}(t) \tag{5.9}$$

with $\triangle t$ the model time step in seconds, and $s_{soil}(t+\triangle t)$ the storage in the soil reservoir at the end of the time step. Equation 5.9 can be arranged as follows:

$$s_{soil}(t+\triangle t) = s_{soil}(t) + (R_{in}(t) - E_{tr}(t) - P_{er}(t))\triangle t \tag{5.10}$$

with the state in the left hand side of the equation 5.10.

The input in the fast reservoir is given by:

$$R_{fast}(t) = \alpha\frac{s_{soil}(t)}{S_{max}}R_{eff}(t) \tag{5.11}$$

with α a dimensionless parameter. The outflow from the fast reservoir is calculated as follows:

$$Q_{fast}(t) = \kappa_2\left(\frac{s_{fast}(t)}{S_{2,max}}\right)^{\gamma} \tag{5.12}$$

Equation 5.12 has 3 model parameters: a recession coefficient κ_2 in m^3/s, the storage capacity of the fast reservoir $S_{2,max}$ in m^3, and a dimensionless parameter γ. By following the same approach as in Equation 5.9, the storage in the fast reservoir at the end of the time step is given by:

$$s_{fast}(t+\triangle t) = s_{fast}(t) + (R_{fast}(t) - Q_{fast}(t))\triangle t \tag{5.13}$$

The input in the slow reservoir is determined:

$$R_{slow}(t) = R_{eff}(t) - R_{fast}(t) \tag{5.14}$$

The outflow from the slow reacting reservoir is calculated as:

$$Q_{slow}(t) = \kappa_1 S_{slow}(t) \tag{5.15}$$

with κ_1 a model parameter in s^{-1}. Finally, the storage in the slow reservoir at the end of the time step is given by:

$$s_{slow}(t + \triangle t) = s_{slow}(t) + (R_{slow}(t) - Q_{slow}(t) + P_{er}(t))\triangle t \tag{5.16}$$

Normally, the unit hydrograph of the catchment, which is the impulse response of the entire catchment at the discharge outlet, is required in order to obtain the modeled discharge Q_{dis}. For small catchments where the time needed for water to flow from the most remote point of the catchment to the catchment outlet (time of concentration) is less than the default model time step of 1 day in the HBV, the total discharge Q_{dis} is simply the sum of the slow Q_{slow} and fast Q_{fast} components of the runoff.

State-space representation of the HBV model

The state-space representation of the HBV model is obtained by taking into account the discrete nature of the measurements (inputs and observations) and by the definition of a fixed model time step which for the HBV model is 1 day. Therefore, the process model is given by the non-linear equation:

$$\mathbf{x}_{t+1} = \mathbf{f}_{t+1,t}(\mathbf{x}_t, \mathbf{u}_t, \Theta) \tag{5.17}$$

where the state vector is the following:

$$\mathbf{x}_t = [s_{soil_t}\ s_{fast_t}\ s_{slow_t}]^T, \tag{5.18}$$

the input vector is defined by:

$$\mathbf{u}_t = [R_{tot_t}\ E_{tp_t}]^T, \tag{5.19}$$

and the vector of the model parameters is:

$$\Theta = [\lambda\ S_{max}\ b\ \alpha\ P\ \beta\ \gamma\ S_{2,max}\ \kappa_1\ \kappa_2]^T, \tag{5.20}$$

finally, the vector valued function $\mathbf{f}_{t+1,t}$ in Equation 5.17 is given implicitly through the non-linear model (Equations 5.10, 5.13 and 5.16).

The observation model is presented as follows:

$$\mathbf{y_t} = \mathbf{h}_t(\mathbf{x}_t) \tag{5.21}$$

the observations correspond to daily discharge measurements, i.e $\mathbf{y}_t = Q_{dis}$, with the observation model given by: $\mathbf{h}(\mathbf{x}_t) = Q_{slow_t}(s_{slow_t}) + Q_{fast_t}(s_{fast_t})$.

5.4 Hydrologic model calibration

A conceptual rainfall-runoff model has been presented in the former section. It can be seen that the dynamics of the hydrologic model are determined by the states of the system (water storages) but also some model parameters, which should be identified for a proper performance of the model, plays an important role in the characterization of the system. The identification of the parameter values are referred to as model calibration or parameter estimation. According to Moradkhani and Sorooshian (2008b), parameters may be classified into physical and process parameters. Physical parameters are those parameters with a physical meaning and which can be measured directly independent of the observed model outputs. On the other hand, process parameters are those which can not be measured directly and need to be estimated.

In general terms, in order for a model to properly simulate the observed response of a catchment over some historical period for which model input and output observations are available, the model parameters need to be tuned or calibrated. In this context, model calibration methods have been proposed to ensure conformity between the mode simulations of systems behavior and observations. A revision of different calibration techniques is given in Moradkhani and Sorooshian (2008b); Pechlivanidis et al. (2011). For long time, manual calibration methods have been used. A manual calibration method consists in the trial and error procedure which is a very labor-intensive procure requiring extensive training and expertise. This expertise is not only difficult to be obtained, but also hard to be transfered from one modeler to another. The difficulties involved in the manual calibration justified the need for development of automatic calibration techniques. In automatic calibration, the estimation of the parameters is formulated as an optimization problem through the objective function which is sometimes called loss or cost function. Three global optimization methods commonly used in watershed model calibration are: Simulated Annealing, Genetic Algorithm and Shuffled Complex Evolution. A revision of the three methodologies is presented in Duan (2003). In this manuscript, the Shuffled Complex Evolution (SCE-UA) method (Duan et al., 1993a) is used in the identification of the model parameters.

A detailed description of the SCE-UA method is given in Duan et al. (1994). Here, a brief description of the method is given. SCE-UA combines the best features of different global search optimization procedures: the Downhill Simplex, Controlled Random Search, Competitive evolution and Complex Shuffling. The method is based on the evolution of a predefined number of parameter combinations. These parameter combinations are referred to as points. The points are ranked in increasing order according to the objective

function to be minimized and subdivided into a number of complexes. All the points are arranged and assigned a complex. After this, each complex is evolved according to the Competitive Complex Evolution algorithm. Finally the points are re-assigned a complex and the procedure is repeated until either convergence has been achieved or a predefined number of iterations have been reached.

5.5 Application of the HBV model

In this section the HBV will be used to represent a catchment. The parameters in the HBV model are process parameter which do not have a physical meaning thus the parameters values are identified by using historical observations.

5.5.1 Site and Data description

The study site corresponds to the Zwalm catchment. This catchment is situated in East-Flanders, Belgium, and it is a tributary of the Scheld river.The total drainage area of the catchment at the gaging station is 114 km^2 and the total length of the perennial channels is 177.5 km. Figure 5.3 shows the location of the catchment. The topography of the basin is characterized by rolling hills and mild slopes. The maximal elevation difference in the catchment is equal to 150 m. The catchment has humid temperature based on the climatic conditions observed, with the yearly mean air temperature 10 °C, the average of the coldest month (January) being 3 °C and the average of the warmest month (July) being 18 °C. The mean yearly rainfall is 775 mm and can be considered to be distributed almost uniformly over the year. The annual evaporation is approximately 450 mm.

Meteorological forcing data with a daily resolution (same as the model time step) from 1994 and 2002 is used in this study. The climatological station in Kruishoutem, which is located approximately 5 Km outside the catchment, provided the precipitation needed by the HBV model. Potential evapotranspiration is obtained through the Penman-method and by using the station observations of air temperature, humidity, radiation and wind speed. Daily discharge values at the outlet of the catchment were available for the entire study period.

Figure 5.4 shows the daily measurements corresponding to the precipitation, the potential evapotranspiration and the discharge. The data are presented for year 1998.

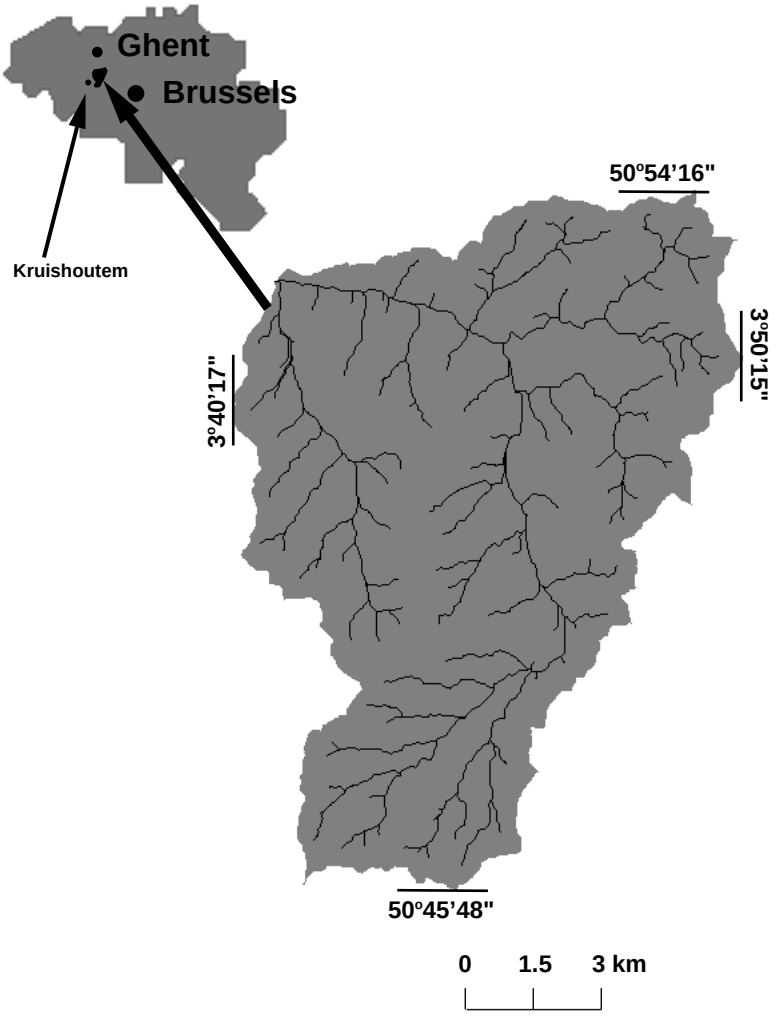


Figure 5.3: The location of the study area at the outlet of the Zwalm catchment

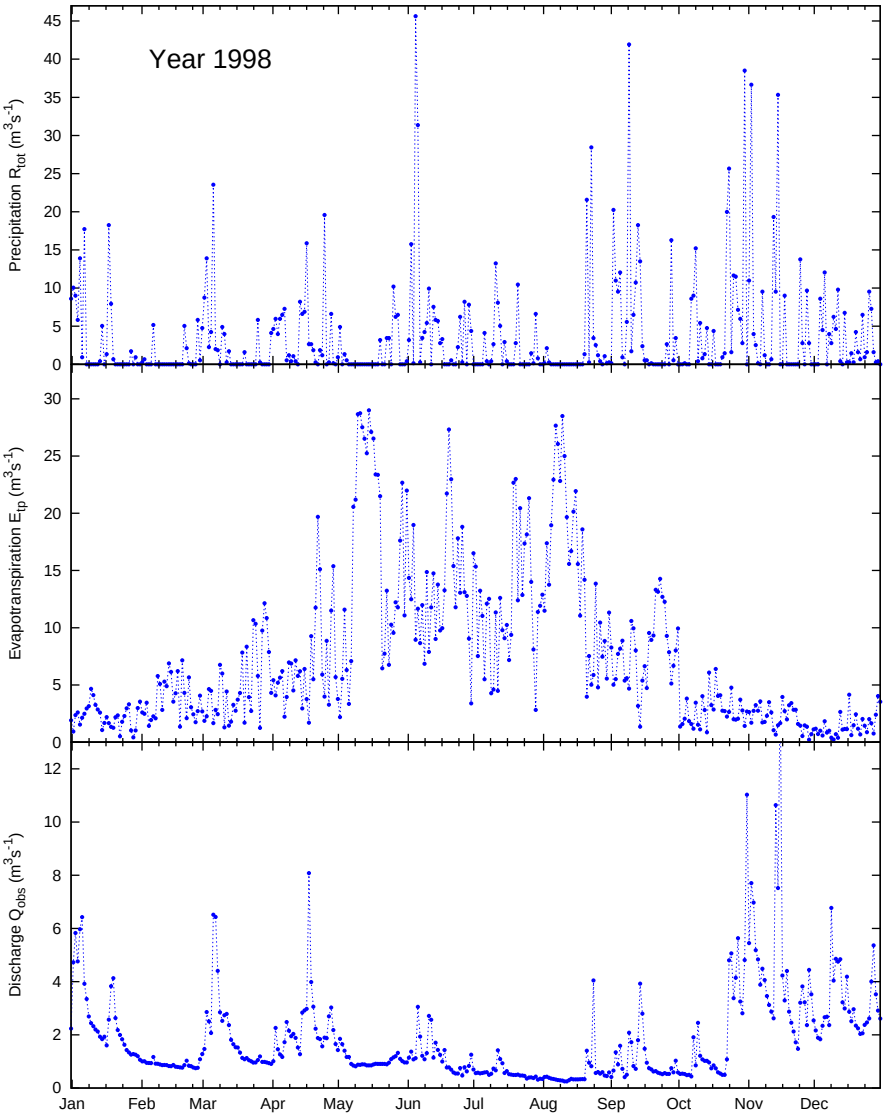


Figure 5.4: Observed forcings and discharge at the Zwalm

5.5.2 Model application

The HVB model is used in the representation of the Zwalm catchment. For this, the input data correspond to the period: 01/01/1994 - 31/12/2002 (9 years of data). The model time step is set to 1 day which is also the sampling time of the measurements. Table 5.1 shows the model parameters, their units, their initial values and the initial conditions of the water storages which are defined based on the capacity of the soil and fast reacting reservoirs.

Table 5.1: Model parameters, units, initial parameter values and initial state conditions.[a]

Parameter	**Units**	**Value**
λ	-	13
S_{max}	m^3	0.1*A
b	-	5
α	-	0.5
P	m^3s^{-1}	0.03*A/(24*3600)
β	-	0.1
γ	-	5
$S_{2,max}$	m^3	2.5*S_{max}
κ_2	m^3s^{-1}	0.2*A/(24*3600)
κ_1	s^{-1}	0.1/(24*3600)
$s_{soil_{t=0}}$	m^3	0.5*S_{max}
$s_{fast_{t=0}}$	m^3	0.5*$S_{2,max}$
$s_{slow_{t=0}}$	m^3	0.5*$S_{2,max}$

[a] - Indicates a dimensionless parameter and A is the area (114.3 Km^2).

Figure 5.5, shows the discharge data and the model simulations for the first three years (1994, 1995 and 1996) of the complete period of simulation. As can be seen from these figures, better parameter values should be identified since the misfit, which can be observed by visual inspection, between the observations and the simulations is considerably large with an overestimation of the low flows and underestimation of most of the discharge peaks.

For the purpose of further comparison, the scatter plot and the corresponding regression line between the observations and simulations are presented in Figure 5.6. The parameters of the regression line indicate the misfit between the observations and simulations with a correlation coefficient (R) of 0.51 indicating a moderate linear relationship between observations and simulations.

The Root Mean Squared Error (RMSE) index is computed as a measure of the goodness of fit between the observations and simulations. RMSE is a

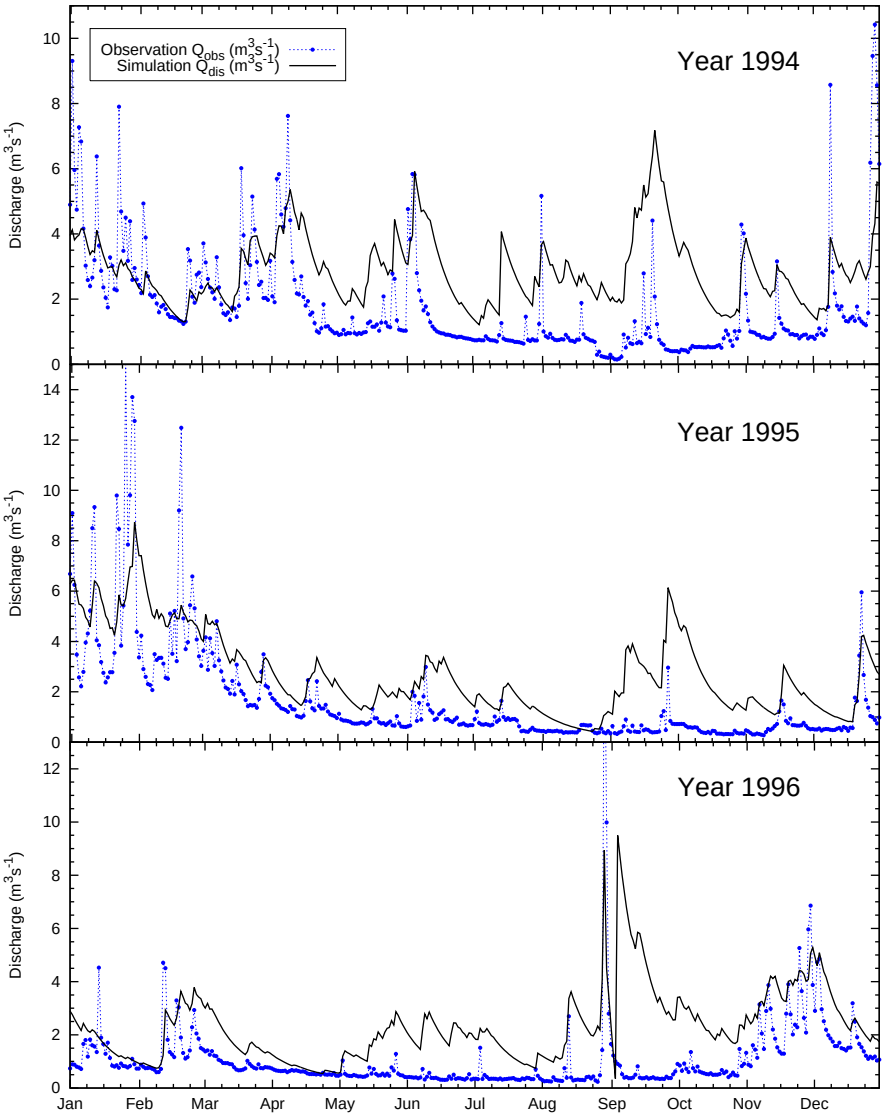


Figure 5.5: Observed and modeled discharge for years 1994, 1995 and 1996

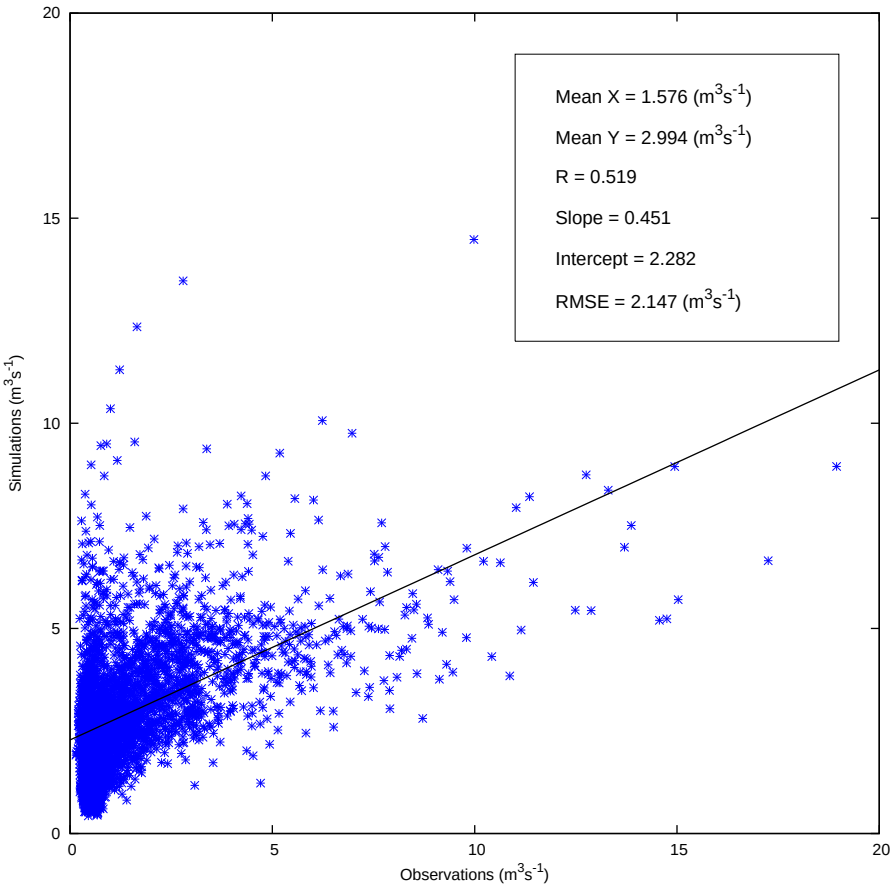


Figure 5.6: Scatter plot and regression line. Simulation period: 1994 - 2002

measure of the residuals and it is given by:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(Q_{obs_t} - Q_{dis_t})^2}{n}} \tag{5.22}$$

where n is the number of model time steps. RMSE can be computed for different time periods. For the period corresponding to 9 years with a model time step of 1 day, the RMSE is 2.147 m^3s^{-1}. This value indicates a large difference in the residuals. Therefore, the parameter values are identified by using an automatic calibration method.

5.5.3 Model Calibration

In section 5.4, the identification of the parameters in a model is interpreted as an optimization problem where an objective function is maximized or minimized. The SCE-UA is used in this work with the RMSE as the objective function to be minimized. The calibration period corresponds to the first 5 years of the full time series, i.e. from year 1994 to year 1998. The initial values of the parameters are those given in Table 5.1. The identified parameter values are validated with data corresponding to the last four years of the time data series. Table 5.2 presents the identified parameter values by the SCE-UA automatic algorithm.

Table 5.2: Model parameters, units and initial parameter values.[a]

Parameter	**Units**	**Value**
λ	-	2.602
S_{max}	m^3	1.279×10^{07}
b	-	0.667
α	-	0.704
P	m^3s^{-1}	68.89
β	-	0.001
γ	-	4.591
$S_{2,max}$	m^3	3.623×10^{07}
κ_2	m^3s^{-1}	247.490
κ_1	s^{-1}	1.984×10^{-06}

[a] - Indicates a dimensionless parameter.

Figure 5.7 shows the observed and modeled discharge when the identified parameter values are applied.

By looking at Figure 5.7, the first visual impression is a better fit of the simulations with respect to the observations. However, it can be seen that most of the discharge peaks in the observations are underestimated. The underestimation reflects the complexity in the problem of identifying optimal

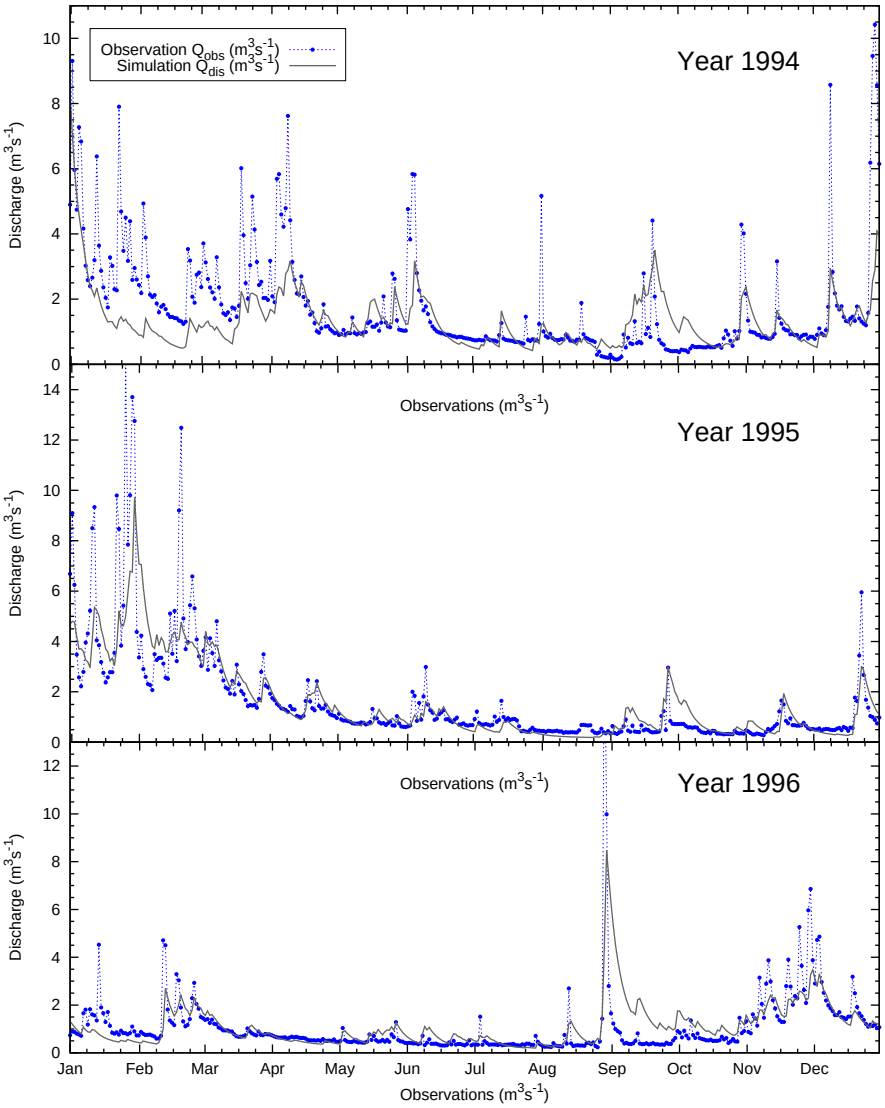


Figure 5.7: Observed and modeled discharge for years 1994, 1995 and 1996

parameter values when the model is highly nonlinear and the amount of model parameters is considerable.

Figure 5.8 presents the scatter plot along with the regression line of the observations and simulations when HBV model is applied with the identified parameter values. The parameters of the regression line (scope and intercept) and the RMSE value are presented in the right top panel of figure 5.8.

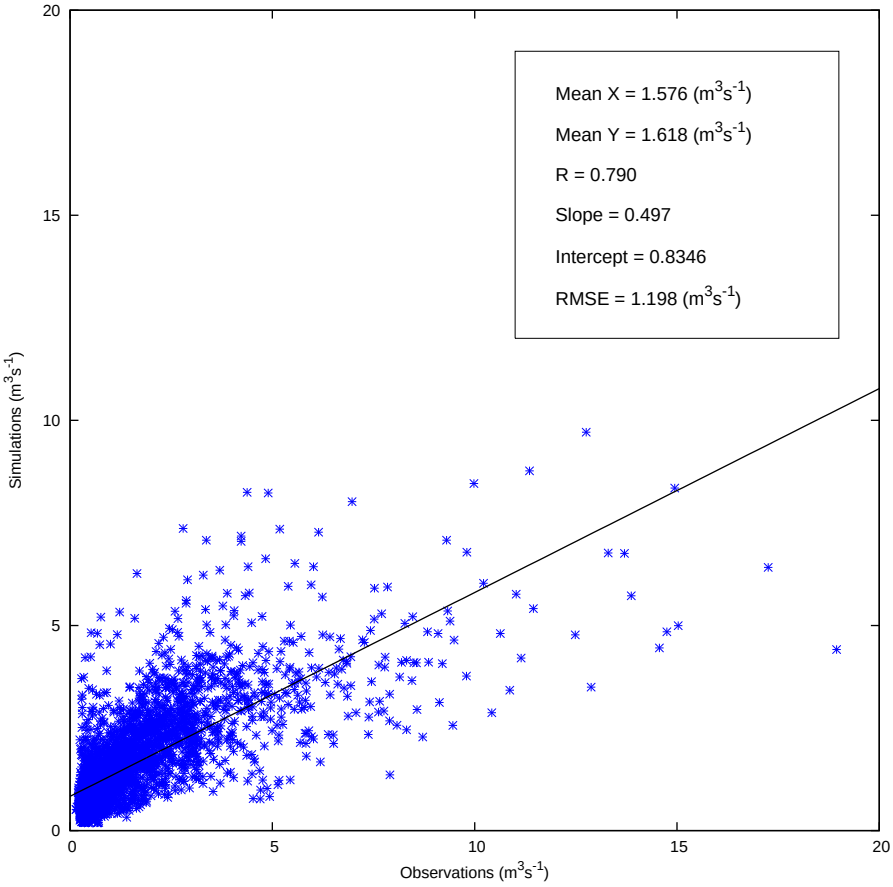


Figure 5.8: Scatter plot and regression line. Simulation period: 1994 - 2002

The coefficient R which has a value of 0.790 indicates a highly linear relationship between the observations and simulations. Moreover, the RMSE index of 1.199 m^3s^{-1}, which is calculated for the entire simulation period of 9 years, indicates a considerable decrease in the estimation error with respect to the RMSE value of 2.147 m^3s^{-1} when HBV model is applied with the initial parameter values. The performance of the model can be improved even more by the application of data assimilation techniques.

5.6 Assimilation of discharge data

The data assimilation methods presented in this work, i.e. EnKF (section 3.4) and particle filters (chapter 4), both perform a point-wise approximation of the posterior distribution. In this sense, the generation of random samples at each assimilation time step plays an important role in the performance of the algorithms. The random samples are refereed to as particles or ensemble forecasts. A brief introduction to the concept of ensemble forecasting was presented in section 3.4.1. In this section, this concept is extended and analyzed in the context of hydrologic data assimilation.

5.6.1 Ensembles in Rainfall-Runoff modeling

An ensemble is a representative sample of the pdf of interest. In weather prediction, ensemble forecasting is adopted as a practical way to estimate the uncertainty of weather forecasts due to the growth of the errors in the initial conditions and/or model structure.

Weather prediction is considered as an initial value forecast problem where dynamic ensembles are generated through the dynamics of the system. In this type of forecast problem, the evolution of the ensemble forecasts is highly dependent on the initial state values. On the other hand, forecasting in hydrologic applications can be described as a boundary-forced problem where the evolution of the ensemble forecasts is affected less by the errors in the initial conditions but more by the errors in the forcings. Moreover, simplifications in the development of hydrologic models may increase the model structure errors thus the uncertainty in the model structure is represented by considering errors in the model parameters and/or errors in the state vector.

In this study, the uncertainty of the hydrologic forecasts is estimated by considering errors in the initial conditions of the states, errors in the forcings, errors in the model parameters and errors in the model structure. Initial conditions are perturbed using additive Gaussian noise with zero mean and a predefined variance magnitude which remains constant (i.e., homoscedastic assumption). The perturbation is performed as follows:

$$\begin{aligned} \mathbf{x}_{0,i} &= \mathbf{x}_0 + w_{t,i} \\ w_{t,i} &\sim \mathcal{N}(0, (\xi_{x0}\mathbf{x}_0)^2) \quad \{i = 1, \ldots, N\} \end{aligned} \tag{5.23}$$

with $\mathbf{x}_0$ the initial conditions state vector, w_t is a white Gaussian noise and ξ_{x0} is the variance scale factor.

In order to account for forcing data measurement error, the evapotranspiration data are perturbed by additive Gaussian noise with a heteroscedastic

assumption, i.e. the variance of the noise is scaled by the magnitude of the evapotranspiration, as follows:

$$\begin{aligned} E_{tp_{t,i}} &= E_{tp_t} + e_{t,i} \\ e_{t,i} &\sim \mathcal{N}(0, (\xi_E E_{tp_t})^2) \quad \{i = 1, \ldots, N\} \end{aligned} \tag{5.24}$$

with $e_{t,i}$ a white Gaussian noise and ξ_E the variance scale factor.

Precipitation is considered to be affected by multiplicative error by following the approach presented in Leisenring and Moradkhani (2011) where a log-normally distributed noise is utilized in the perturbation of the precipitation R_{tot} as follows:

$$\begin{aligned} \mu_R &= \ln \left[\frac{R_{tot_t}^2}{\sqrt{R_{tot_t}^2 + (\xi_R R_{tot_t})^2}} \right] \\ \sigma_R &= \sqrt{\ln \left[\frac{(\xi_R R_{tot_t})^2}{R_{tot_t}^2} + 1 \right]} \\ R_{tot_{t,i}} &= \exp \left[\mu_R + w_{R_{t,i}} \frac{\sigma_R^2}{2} \right] \qquad w_{R_{t,i}} \sim \mathcal{N}(0, 1) \end{aligned} \tag{5.25}$$

with $R_{tot-true_t}$ the perturbed precipitation at time t, ξ_R is a variance scaling factor for precipitation data which is set to 0.30 and w_R is white Gaussian noise with zero mean and standard deviation equal to 1.

The uncertainty in the model parameter values is represented by the perturbation of the calibrated parameters with additive white Gaussian noise:

$$\begin{aligned} \Theta_i &= \Theta + w_{\Theta_i} \\ w_{\Theta_i} &\sim \mathcal{N}(0, (\xi_\Theta \Theta)^2) \quad \{i = 1, \ldots, N\} \end{aligned} \tag{5.26}$$

with Θ representing the vector of identified parameter values (Table 5.2), w_Θ is a white Gaussian noise and ξ_Θ is the variance scale factor.

Additionally, the model structure error is taken into consideration by the perturbation of the state vector at every assimilation time step with white Gaussian noise:

$$\begin{aligned} \mathbf{x}_{t,i} &= \mathbf{x}_t + w_{x_{t,i}} \\ w_{x_{t,i}} &\sim \mathcal{N}(0, (\xi_x \mathbf{x}_t)^2) \quad \{i = 1, \ldots, N\} \end{aligned} \tag{5.27}$$

with $\mathbf{x}_t$ the state vector at time step t, w_{x_t} is a white Gaussian noise and ξ_x is the variance scale factor. Table 5.3 lists all the variance scale factors (a.k.a. hyper-parameters) for the different error structures considered in this work.

Table 5.3: Variance scale factors and error structures

Factor	Description (assumed error distribution)
ξ_{x0}	Initial conditions variance scale factor (normal, homoscedastic)
ξ_E	Evapotranspiration variance scale factor (normal, heteroscedastic)
ξ_R	Precipitation variance scale factor (lognormal, heteroscedastic)
ξ_Θ	Model parameter variance scale factors (normal, homoscedastic)
ξ_x	State vector variance scale factors (normal, heteroscedastic)

At this point, the structure of the errors have been presented. However, the performance of the assimilation algorithms strongly depend on the magnitude of these errors. Two approaches for the identification of realistic error magnitudes are represented in the following section.

5.6.2 Ensemble quality control

It is clear that the performance of any assimilation method depends upon a realistic generation of the state ensemble. In this sense, two approaches are used in this study aiming at a correct representation of the forcing, parameter and model structure errors. The first approach concerns the identification of error magnitudes which remain constant along the simulation period (De Lannoy et al., 2006a), while the second approach is based on a dynamic update of the error magnitudes (Leisenring and Moradkhani, 2012).

Constant error magnitudes

The quality of the discharge ensemble is verified according to De Lannoy et al. (2006a) where the ensemble spread ($ensp_t$), the ensemble mean square error (mse_t), and the ensemble skill ($ensk_t$) have to be computed first and at each time step t:

$$\begin{aligned} ensp_t &= \frac{1}{N}\sum_{i=1}^{N}\left(Q^i_{dis_t} - \overline{Q_{dis_t}}\right)^2 \\ mse_t &= \frac{1}{N}\sum_{i=1}^{N}\left(Q^i_{dis_t} - Q_{obs_t}\right)^2 \\ ensk_t &= \left(\overline{Q_{dis_t}} - Q_{obs_t}\right)^2 \end{aligned} \tag{5.28}$$

In equation 5.28, $Q^i_{dis_t}$ is the modeled discharge (m^3/s) for particle i at time t and Q_{obs_t} is the corresponding observation of the discharge in m^3/s at time step t. In order to have a large enough ensemble spread, on average

the ensemble mean differs from the observation by a value that is equal to the time average of the ensemble spread. Therefore, the following expression should be true:

$$\frac{< ensk >}{< ensp >} \approx 1 \tag{5.29}$$

where $< . >$ indicates an average over the simulation period. Furthermore, if the truth is statistically indistinguishable from a member of the ensemble, the following expression should be true:

$$\frac{< \sqrt{ensk} >}{< \sqrt{mse} >} \approx \sqrt{\frac{N+1}{2N}} \tag{5.30}$$

Dynamic update of the error magnitudes

A procedure to update the error magnitudes during the assimilation cycles was introduced by Leisenring and Moradkhani (2012). More specifically, ensemble spread is updated by varying the variance multipliers ξ (a.k.a. variance scale factors) at every assimilation time step. ξ is increased when the absolute bias is larger than the outer 95th percent uncertainty bound and it is reduced when the bias is smaller than the outer 95th percent uncertainty bound. The procedure is indicated as follows:

$$\hat{e}_t = |\overline{Q_{dis_t}} - Q_{obs_t}|, \tag{5.31}$$

$$ub_t = \begin{cases} \overline{Q_{dis_t}} - Q_{dis_t}^{L95} & \text{if } Q_{obs_t} < \overline{Q_{dis_t}}, \\ Q_{dis_t}^{U95} - \overline{Q_{dis_t}} & \text{if } Q_{obs_t} > \overline{Q_{dis_t}}, \end{cases} \tag{5.32}$$

$$er_t = \frac{\hat{e}_t}{ub_t}, \tag{5.33}$$

where $\hat{e}_t$ is the absolute value of the mean model error, ub_t is the partial uncertainty bound, $Q_{dis_t}^{L95}$ is the lower 95th percent uncertainty bound of the predicted observation, $Q_{dis_t}^{U95}$ is the upper 95th percent uncertainty bound of the predicted observation, and er_t is the ratio of the model error to the partial uncertainty bound. Finally variance scaling factors are corrected according to er_t at each time step as follows:

$$\xi_t = er_t \times \xi. \tag{5.34}$$

The assimilation methods are assessed in the remaining sections of this chapter. For this, discharge data are generated and used in an assimilation experiment.

5.6.3 Experiment with synthetic data

A synthetic discharge data assimilation study is performed. The experimental setup consists of the artificial generation of true discharge records through the application of additive and multiplicative Gaussian noise to the initial conditions, forcings, model parameters and model structure. A true discharge data ($Q_{dis-true}$) record is calculated based on this artificial true state vector ($s_{soil-true}$, $s_{fast-true}$, $s_{slow-true}$).

For the generation of the truth, the initial conditions of the three water storages were estimated by using the in-situ observed discharge data, the identified initial values are: $S_{soil_{t=0}} = 1.11 \times 10^{07}$ (m^3), $S_{fast_{t=0}} = 1.36 \times 10^{07}$ (m^3) and $S_{slow_{t=0}} = 8.10 \times 10^{05}$ (m^3). With respect to the error structure, initial state conditions were perturbed according to equation 5.24 with the value of ξ_{x0} set to 0.5.

The errors which might have been introduced in the derivation of the input evapotranspiration are considered in this study through the perturbation of the evaporation time series by white Gaussian noise according to equation 5.25 with ξ_E equal to 0.3. The precipitation error structure is considered to be multiplicative with a log-normally distributed noise affecting the precipitation measurements. The perturbed time series of precipitation is obtained by the application of equation 5.25 where the value of ξ_R is fixed to 0.30.

Additionally, the identified values of the model parameters shown in Table 5.2 are perturbed with Gaussian noise with zero mean and standard deviation set to ξ_Θ times the nominal value for each parameter, respectively. Large uncertainty is considered for the errors of the identified model parameter values with ξ_Θ equal to 0.5.

The true states obtained from the process described above are used in the generation of the true discharge. The discharge error, which is used in the generation of the synthetic observations, is different from additive Gaussian noise. Additive Gaussian noise can lead to better performances of the EnKF and its variant: EnGPF. In order to perform a realistic experiment and a fair comparison of the methodologies, the discharge error is represented by multiplicative noise (Leisenring and Moradkhani, 2011). The synthetic observations are obtained by the perturbation of the true discharge with log-normally distributed noise according to equation 5.25 with the variance scaling factor (ξ_Q) set to 0.25. Figure 5.9 shows both an ensemble forecast (see below) and the true states while Figure 5.10 shows the forecasted and true discharge.

The aim of the synthetic study is to assess the performance of the filtering techniques when retrieving the true states and true discharge. Synthetic observations $Q_{dis-obs_t}$ are assimilated by the filters at every daily model

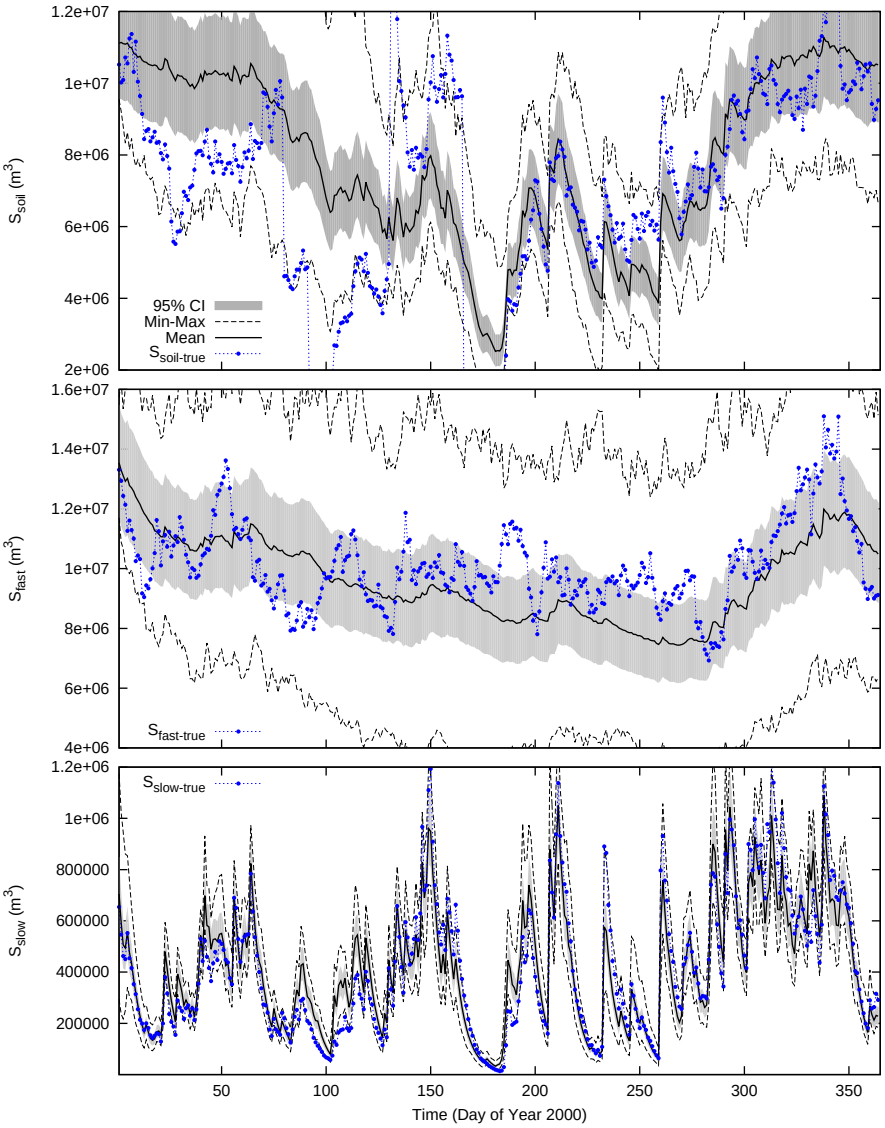


Figure 5.9: Ensembles of the forecasted and synthetic-generated true states: black solid line corresponds to the ensemble mean, dashed lines corresponds to the maximum and minimum ensemble members, dots correspond to the synthetic-generated true states and the gray shaded area shows the 95% confidence interval. The same symbols are used in the remaining figures.

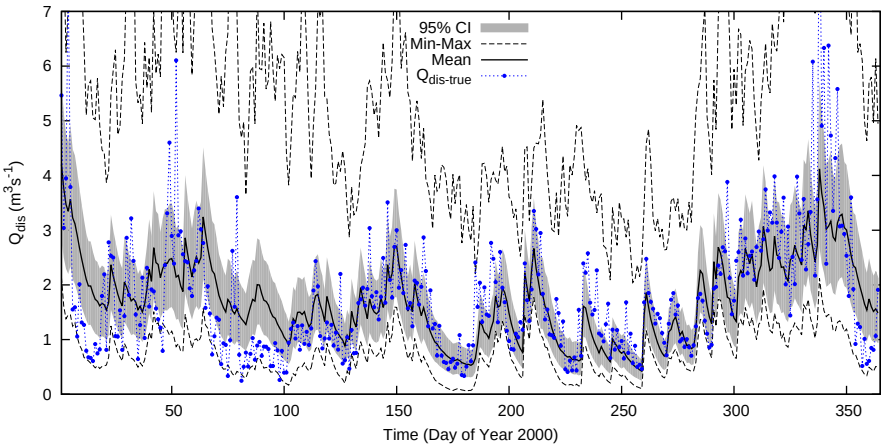


Figure 5.10: Same as Figure 5.9, but for the true and forecasted discharge.

time step during the year 2000. The standard deviation considered in the measurement error is set to $0.2 \times Q_{dis-obs_t}$ (m^3/s).

5.6.4 Generation of the discharge ensemble

A synthetic-generated discharge data record was obtained in the former section. The aim of the ensemble quality control is to identify the magnitudes of the errors in order to obtain a proper ensemble of forecasted discharge. For this, the magnitude of the noises utilized in the generation of the synthetic observations were increased. More specific, the standard deviation of the noise used in the perturbation of the initial state values was increased from 50% to 60% ($\xi_{x0} = 0.6$) of the nominal values for the ensemble forecast. For the perturbation of the evapotranspiration, the standard deviation of the white Gaussian noise is set to $0.50 \times E_{tp_t}$ and for the precipitation, ξ_R is set to 0.50. The variance scaling factor ξ_Θ of the model parameters is equal to 0.7.

Additionally, the states of the system were perturbed by additive Gaussian noise with zero mean and standard deviation equal to $\xi_x \times \mathbf{x}_t$, where ξ_x is the variance scaling factor for the state vector error and it is set to 0.10. The magnitude of the state errors is considerably lower than the magnitudes for the error in the initial conditions, parameters and forcings. This, partly assures that the Gaussian component in the structure of the forecast error is not dominant enough as to lead to biased performances in favor of the Gaussian filters. The corresponding ensemble verification measures for

discharge are:

$$\frac{< ensk >}{< ensp >} = 1.01 \qquad \frac{< \sqrt{ensk} >}{< \sqrt{mse} >} = 0.71$$

Figures 5.9 and 5.10 show the ensemble mean, the 95% Confidence Interval (CI) and the maximum and minimum ensemble members for the states and the synthetic-generated true discharge respectively.

5.7 Performance measures

The skill of the assimilation methods in the estimation of the model states and output flow is assessed by the comparison of performance metrics related to the ensemble mean prediction and the ensemble spread prediction. The accuracy of the filters is verified by the Root Mean Square Error (RMSE) and the Nash-Sutcliffe Efficiency (NSE) index (Nash and Sutcliffe, 1970). A detailed description of these metrics is presented as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^{N_{data}} (\overline{yp}_t - y_t)^2}{N_{data}}} \tag{5.35}$$

where $Ndata$ indicates the number of model time steps, $\overline{yp}_t$ is the ensemble mean of the predicted variable and y_t is the corresponding observation at time step t.

$$\text{NSE} = 1 - \frac{\sum_{t=1}^{N_{data}} (\overline{yp}_t - y_t)^2}{\sum_{t=1}^{N_{data}} (\overline{yp}_t - \bar{y})^2} \tag{5.36}$$

where $\bar{y}$ is the mean of the observations from $t = 1$ to $t = N_{data}$. The range of the RMSE is 0 to ∞ with a perfect score of 0. The range of the NSE index is $-\infty$ to 1 with a perfect score of 1.

The Percent Bias (%BIAS) is used as a measure of precision. The range of the %BIAS index is -100 to 100% with a perfect score of 0.

$$\%\text{BIAS} = \left(\frac{\sum_{t=1}^{N_{data}} (\overline{yp}_t - y_t)}{\sum_{t=1}^{N_{data}} y_t} \right) * 100 \tag{5.37}$$

Moreover, the spread of the predicted ensemble is supervised by the Normalized Root Mean Square Error Ratio (NRR) index (Moradkhani et al., 2005a) with NRR < 1 indicating excessive spread and NRR > 1 indicating insufficient spread.

$$\text{NRR} = \frac{\sqrt{\frac{\sum_{t=1}^{N_{data}} (\overline{yp}_t - y_t)^2}{N_{data}}}}{\frac{1}{N} \left[\sum_{i=1}^{N} \sqrt{\frac{\sum_{t=1}^{N_{data}} (\overline{yp}_{t,i} - y_t)^2}{N_{data}}} \right] \sqrt{\frac{N+1}{2N}}} \tag{5.38}$$

The range of the NRR index is 0 to ∞ with a perfect score of 1.

5.8 Comparative study with artificially generated discharge observations

In this section, the assimilation methodologies are applied to the hydrologic estimation problem introduced in section 5.5. The setup of the synthetic experiment was described in section 5.6.3. Synthetic-generated discharge records $Q_{dis-obs_t}$ are assimilated at every time step with a predefined standard deviation of the measurement error equal to $0.2 \times Q_{dis-obs_t}$ (m^3/s). The performance metrics: RMSE, NSE, %BIAS and NRR are used in order to quantify the performance of the filter and for posterior comparison.

5.8.1 Ensemble size and computational time demand of the EnKF

The performance of the EnKF applied to nonlinear estimation problems relies on the correct MC approximation of the covariance matrix and a proper Gaussian representation of the true posterior distribution. In this context, the ensemble size plays an important role in the skill of the filter. Therefore, a sensitivity study of the performance of the filter concerning different ensemble sizes is carried out in this section. Specifically, 50 MC simulations are performed and the RMSE is computed between time series of the predicted and true discharge records. Moreover, five ensemble sizes (32, 64, 128, 256 and 528) are considered in the experiment.

Table 5.4 presents the mean and standard deviation (over 50 MC runs) of the discharge RMSE and the time demanded by each ensemble size. The data assimilation experiment is implemented over a computer with the Intel Core i3, 1.4 GHz processor and the software used in the simulations is MATLAB 7.9. The RMSE values listed in Table 5.4 indicates an slight improvement in the performance when the ensemble size is increased. This improvement in the performance was expected since an increase in the ensemble size should lead to a better approximation of the sample covariance error matrix. However, the improvement becomes marginal above an ensemble size equal to 128. With respect to the computational time demand, the time demand of the filter increases proportional with the increment of the ensemble size. Finally, the ensemble size equal to 128 members is a good trade off by considering proper accuracy with low time demand.

Table 5.4: RMSE [m^3/s] of the simulated and true discharge, averaged (μ) over 50 MC runs, with indication of 1 standard deviation (σ) and averaged computational time demand (μ_{time}) [s]

Ensemble size	μ	σ	μ_{time}
32	0.679	0.015	0.68
64	0.675	0.009	1.21
128	0.671	0.006	2.28
256	0.669	0.005	4.48
528	0.668	0.003	9.01

5.8.2 Resampling strategies and number of particles in the particle filter

The application of a resampling strategy is important in order to overcome problems related to the degeneracy of the particles. A sensitivity test of the performance of the particle filter with different resampling strategies is conducted. Specifically, the performance of the SPF along with the MulR, ResR, SysR and StrR strategies is quantified through the computation of the discharge RMSE, averaged over 50 Monte Carlo (MC) runs. The RMSE is computed between time series of the modeled discharge and the true discharge. The generation of the true discharge is explained in section 5.6.3. An observation noise variance of $(0.2 \times Q_{dis-obs_t})^2$ $(m^3/s)^2$ and five particle sets (32, 64, 128, 256 and 528) are considered in the experiment.

Table 5.5 presents the mean and standard deviation (over 50 MC runs) of the discharge RMSE and the computational time demand of each resampling strategy. According to Table 5.5, the performance of the particle filter improves when more particles are used in the approximation of the posterior. However, beyond 128 particles, the improvement with more particles becomes marginal. A set of 128 particles was selected as a good trade off between accuracy and computational time demand.

Moreover, the RMSE values are close to each other when comparing the different resampling strategies, especially when the number of particles is above 128. The StrR approach performs slightly better in terms of the RMSE mean and computational time demand thus the StrR was selected as the strategy to be used within the standard particle filter in this study.

5.8.3 Estimation of the true states

In this section, the performance of the assimilation methodologies in the estimation of the three water reservoirs is analyzed. The aim of the tracking of the system states is to check for consistency in the functioning of the

Table 5.5: RMSE [m^3/s] of the simulated and true discharge, averaged (μ) over 50 MC runs, with indication of 1 standard deviation (σ) and averaged computational time demand (μ_{time}) [s]

Particles	**SPF-MulR**			**SPF-RR**		
	μ	σ	μ_{time}	μ	σ	μ_{time}
32	0.660	0.019	0.70	0.657	0.019	0.72
64	0.652	0.018	1.26	0.648	0.015	1.26
128	0.636	0.010	2.41	0.637	0.010	2.42
256	0.632	0.009	4.70	0.633	0.009	4.73
528	0.628	0.007	9.41	0.628	0.008	9.69
Particles	**SPF-SysR**			**SPF-StrR**		
	μ	σ	μ_{time}	μ	σ	μ_{time}
32	0.654	0.022	0.69	0.656	0.017	0.67
64	0.642	0.012	1.31	0.644	0.015	1.22
128	0.635	0.010	2.39	0.634	0.009	2.37
256	0.631	0.009	4.56	0.631	0.009	4.55
528	0.627	0.007	9.52	0.625	0.008	9.39

filters, mainly the sophisticated filters such as SPF-RM and EnGPF. Table 5.6 presents the performance metrics between the true and predicted states. In table 5.6, the open loop ensemble (w/o data assimilation) is used as a baseline with the purpose of comparison. First the performance of the filters is compared to the baseline run while an overall analysis of filters performance is carried out at the end of this section.

EnKF performance

The upper part of Table 5.6 lists the performance measures for the water storage in the soil reservoir S_{soil}. The measures of accuracy and precision indicates a slight worse performance of the filter compared to the baseline run with a RMSE value greater than the baseline, the NSE is negative indicating the mean of the observations is better than the model predictions and the %BIAS value greater than the baseline. In general, the water storage in the soil reservoir is poorly estimated due to a weak influence of this state on the total output flow (see equation 5.21). The ensemble spread is analyzed at the end of the section since the same trend is observed for the performance metrics of the three water storages.

The middle and lower parts of Table 5.6 show the metrics for the water storage in the fast S_{fast} and slow S_{slow} reacting reservoirs respectively. Different results are observed between the accuracy and precision metrics. The accuracy measures (RMSE and NSE) indicate a correction of the states while

Table 5.6: Comparison of the performance metrics (RMSE [m^3] %BIAS[-] NSE[-] and NRR[-]) between the modeled and true states

Filter				
	s_{soil}			
	RMSE	**%BIAS**	**NSE**	**NRR**
Baseline	2.38×10^{06}	10.78	0.10	1.27
EnKF	2.45×10^{06}	11.75	-0.02	1.28
SPF	2.33×10^{06}	8.87	0.19	1.30
SPF-RM	2.31×10^{06}	9.09	0.05	1.28
EnGPF	2.06×10^{06}	5.62	0.32	1.27
	s_{fast}			
Baseline	1.41×10^{06}	-4.84	0.10	0.83
EnKF	1.33×10^{06}	-6.90	0.48	1.26
SPF	1.07×10^{06}	-5.25	0.55	1.19
SPF-RM	1.12×10^{06}	-5.70	0.56	1.20
EnGPF	1.01×10^{06}	-3.31	0.57	1.26
	s_{slow}			
Baseline	8.93×10^{04}	0.17	0.85	1.18
EnKF	8.81×10^{04}	1.30	0.86	1.21
SPF	8.43×10^{04}	-1.19	0.87	1.20
SPF-RM	8.63×10^{04}	-0.32	0.86	1.22
EnGPF	7.65×10^{04}	-1.13	0.90	1.22

the precision is worsened by an increase in the %BIAS. The inconsistencies in the performance metrics are due to the fact that the observation model is highly nonlinear and the noise, which is used in the generation of the synthetic observations, is different from additive Gaussian noise.

Finally, the ensemble spread is considerably reduced during the assimilation steps. The reduction of the ensemble spread can be observed by comparing the values of the NRR in the last column of Table 5.6. The same trend is observed for all the water storages in the three reservoirs.

SPF performance

Table 5.6 lists the values of the metrics between the predicted and observed states regarding the SPF performance. With respect to the accuracy of the filter, the SPF performs a correction of the three states of the water reservoirs. This is indicated by the reduction in the RMSE values and the increase in the NSE values compared to the baseline run. The precision of the state estimation is considerably affected according to the %BIAS values for the water storages in the fast and slow reacting reservoirs. While a slight correction of the bias is observed for the water storage in the soil reservoir.

The water storage in the soil reservoir is less affected by the presence of bias due to the low observability of this state.

SPF-RM performance

A similar performance between SPF and SPF-RM is observed in Table 5.6. SPF-RM performs an improvement in the estimation of the three water storages. With respect to the accuracy of the filter, the application of RM steps allows for a correction of the three states indicated by the reduction in the RMSE values and the increase in the NSE values when comparing to the baseline run. The precision of the state estimation is affected. This is indicated by the increment of the values corresponding to the %BIAS index for the water storages in the fast and reacting reservoirs. On the other hand, the state of the soil reservoir is less affected by the presence of bias.

EnGPF performance

The EnGPF performs the best among the filters presented in Table 5.6. The approximation of the EnKF to the optimal proposal distribution in the Gaussian particle filter allows for a correction of the three states. The RMSE values are reduced indicating an improvement in accuracy. Moreover, the improvement is corroborated by the increase in the values of the NSE index. With respect to the precision of the filter, EnGPF is able to reduce the %BIAS for the water storages in the soil reservoir and fast reservoir but the value of the precision metric is increased in magnitude when the state of the slow reacting reservoir is estimated.

Overall comparison

EnKF is the filter with the worst performance. This result is somehow expected since the setup of the synthetic experiment involved the use of forms of errors different from additive Gaussian. The better performance of the EnGPF compared to the EnKF and mainly to the SPF and SPF-RM demonstrates that EnGPF is able to perform a proper estimation of the hidden states of a non-Gaussian or close to non-Gaussian system with a non-linear observation equation. Nevertheless, the slight improvement obtained by the EnGPF compared to the SPF and the SPF-RM indicates that the outperformance of the EnGPF cannot be generalized to any kind of non-Gaussian system. Thus, further research in this context is recommended.

A marginal outperformance of the SPF over the SPF-RM is observed in Table 5.6. The proper performance of the SPF is due to the magnitude of the

errors considered in the setup of the experiment. For instance, the observation error is inflated which might avoid the impoverishment of particles in the resampling step. In the assimilation problem presented in this section, the SPF-RM is only able to perform similar to the SPF. However, Moradkhani et al. (2012) reported a study-case where the SPF-RM outperformed the SPF. The main difference in the setup carried out in (Moradkhani et al., 2012) and the setup in this dissertation is the join estimation of states and parameters. Therefore, a further analysis of the impact of a join estimation on the performance of the EnGPF is recommended.

Finally, the results listed in Table 5.6 show that the performance of the filters are affected to different extents by bias because of the strong perturbation of the model parameters. Additionally, the ensemble spread of the predicted states, which is tracked by the NRR performance metric, is highly reduced for all the three water storages indicating a strong correction of the states.

5.8.4 Estimation of the true discharge

EnKF

Table 5.7 shows the performance measures between the predicted discharge Q_{dis} and the true discharge records $Q_{dis-true}$. In terms of accuracy and precision, the ENKF shows a correction of the predicted discharge with lower RMSE and %BIAS values and a higher NSE value. However, the ensemble spread is strongly affected as can be seen through the comparison of the NRR values in Table 5.7.

Table 5.7: Comparison of the performance metrics (RMSE [m^3/s] %BIAS[-] NSE[-] and NRR[-]) between the modeled and true discharge

Filter	RMSE	%BIAS	NSE	NRR
Baseline	0.86	-16.04	-0.28	1.02
EnKF	0.66	-14.02	0.50	1.32
SPF	0.64	-13.88	0.52	1.32
SPF-RM	0.63	-13.49	0.53	1.32
EnGPF	0.55	-10.85	0.67	1.30

SPF

The predicted discharge, which is obtained by the application of the SPF methodology, is corrected in terms of accuracy (RMSE and NSE) and precision (%BIAS). The performance metrics computed over the predicted discharge (Table 5.7) show consistency with the results of the state estimation

displayed in Table 5.6. With respect to the ensemble spread, the NRR value shows a significant increase indicating a reduction of the ensemble spread.

SPF-RM

Table 5.7 shows that the SPF-RM performs very similarly to the SPF. In general, the performance metrics in Table 5.7 indicate a correction of the true discharge in terms of accuracy and precision. On the other hand, the NRR value indicates a reduction of the ensemble spread as a result of the discharge estimation compared to the baseline run.

EnGPF

The comparison of the performance metrics listed in Table 5.7 between the baseline run and the model run with the EnGPF indicates, in general, that the predicted discharge is considerably corrected. While, the ensemble spread monitored by the NRR value shows a significant increase indicating a reduction of the ensemble spread.

Overall comparison

The comparison of the filter performance metrics indicated in Table 5.7 lead us to the conclusion that all the filters perform a correction of the modeled discharge. Moreover, the similarity in the performance metric values listed in Table 5.7 does not show clearly the best performance. Nevertheless, the proper state estimation performance of the EnGPF analyzed in section 5.8.3 allows to conclude that the EnGPF performs the best. The reduction of the ensemble spread can be mitigated by updating the magnitudes of the errors dynamically (see section 5.6.2). The latter is adopted and discussed in the next section.

5.9 Comparative study with in-situ observed discharge data

In this section, a time series from the year 2000 corresponding to in-situ observed discharge data (Q_{obs_t}) is used in the assimilation experiment. The aim of the experiment is to compare the performances of the different filtering techniques used in this chapter. A predefined observation error with a standard deviation equal to $0.1 \times Q_{obs_t}$ (m^3/s) is considered in the study. The variance scale fator of 0.1 is a realistic assumption regarding the observation

error. The ensemble generation is performed according to a combination of the methodologies described in section 5.6.2 and further discussed below.

5.9.1 Ensemble generation

The generation of the discharge ensemble is performed by the identification of an optimal initial ensemble spread (open loop ensemble) which is carried out by using all the time series of discharge data (constant error magnitudes) and the posterior correction of the noise magnitudes at every assimilation time step (dynamic update of the error magnitudes). This twofold adaptive update of the error magnitudes is performed aiming to improve the performance of the data assimilation methods concerning a proper ensemble spread.

First, the variance scale factors involved in the perturbation of the model variables are set to the values indicated in section 5.6.4. Further, the noise levels ξ_θ and ξ_x were readjusted in order to obtain the optimal spread with values of 0.05 for ξ_θ and 0.5 for ξ_x. The corresponding statistic measures,

$$\frac{< ensk >}{< ensp >} = 1.01 \qquad \frac{< \sqrt{ensk} >}{< \sqrt{mse} >} = 0.71,$$

indicates an optimal ensemble spread which will be further altered at every assimilation time step.

Second, the noise levels are sequentially adjusted according to the variable variance multipliers approach. Specifically, the variance scaling factors ξ_θ and ξ_x are dynamically updated at every time step according to equation 5.34. The upper bound of the ratio of the error er_t (equation 5.33) is set to 2.

5.9.2 Results

Figures 5.11 and 5.12 show the discharge ensemble mean, the 95% Confidence Interval (CI) and the maximum and minimum ensemble members for the EnKF in figure 5.11 and for the particle filters in figure 5.12. It is difficult to determine the best performance solely by visual inspection since all the figures show similar performance. However, a small difference is observed by checking the peak around time step 210. The EnKF, SPF and SPF-RM performances show insufficient ensemble spread as to cover this peak flow. On the contrary, the ensemble corresponding to the EnGPF performance shows sufficient spread as to cover the peak flow around time step 170. The figures also shows that the filters perform better for the low flows than for

Table 5.8: Discharge estimation performance metrics when using variable variance multipliers. Real case scenario

Filter	RMSE	%BIAS	NSE	NRR	CTD
EnKF	0.61	-14.41	0.55	1.17	4.36
SPF	0.58	-11.96	0.61	1.23	4.35
SPF-RM	0.56	-11.20	0.63	1.24	6.84
EnGPF	0.46	-6.79	0.79	1.11	4.60

the high flows. This is consistent with the considered variance of the observation errors which depends on the magnitude of the observation at each time step.

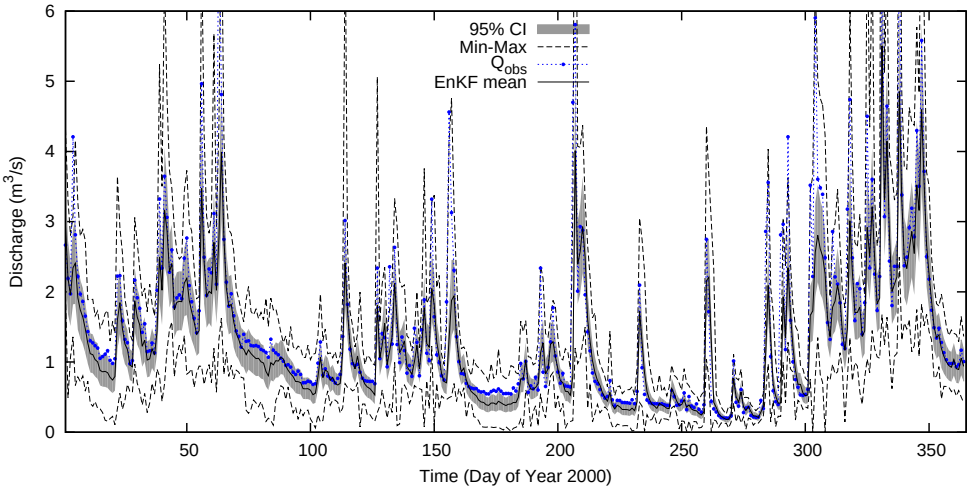


Figure 5.11: EnKF performance for observation error with standard deviation equal to $0.1 \times Q_{obs}$.

Table 5.8 presents the performance metrics between the predicted discharge and the discharge observations along with the computational time demanded (CTD) by each assimilation method. Although all the values in Table 5.8 are close in magnitude, the performance metrics indicate that the data assimilation method with the least skill is the EnKF. With respect to particle filter performances, the SPF outperforms the EnKF and the SPF-RM outperforms the SPF and the EnKF. The lower noise levels considered in the real experiment allows for a correct state estimation performance in the SPF and SPF-RM. EnGPF has the best performance with the lowest value for the RMSE index and % Bias index along with the highest value of NSE. Moreover, the NRR value is the lowest indicating an improvement also in the ensemble spread.

Overall, the results of the experiment with in-situ discharge data indicates

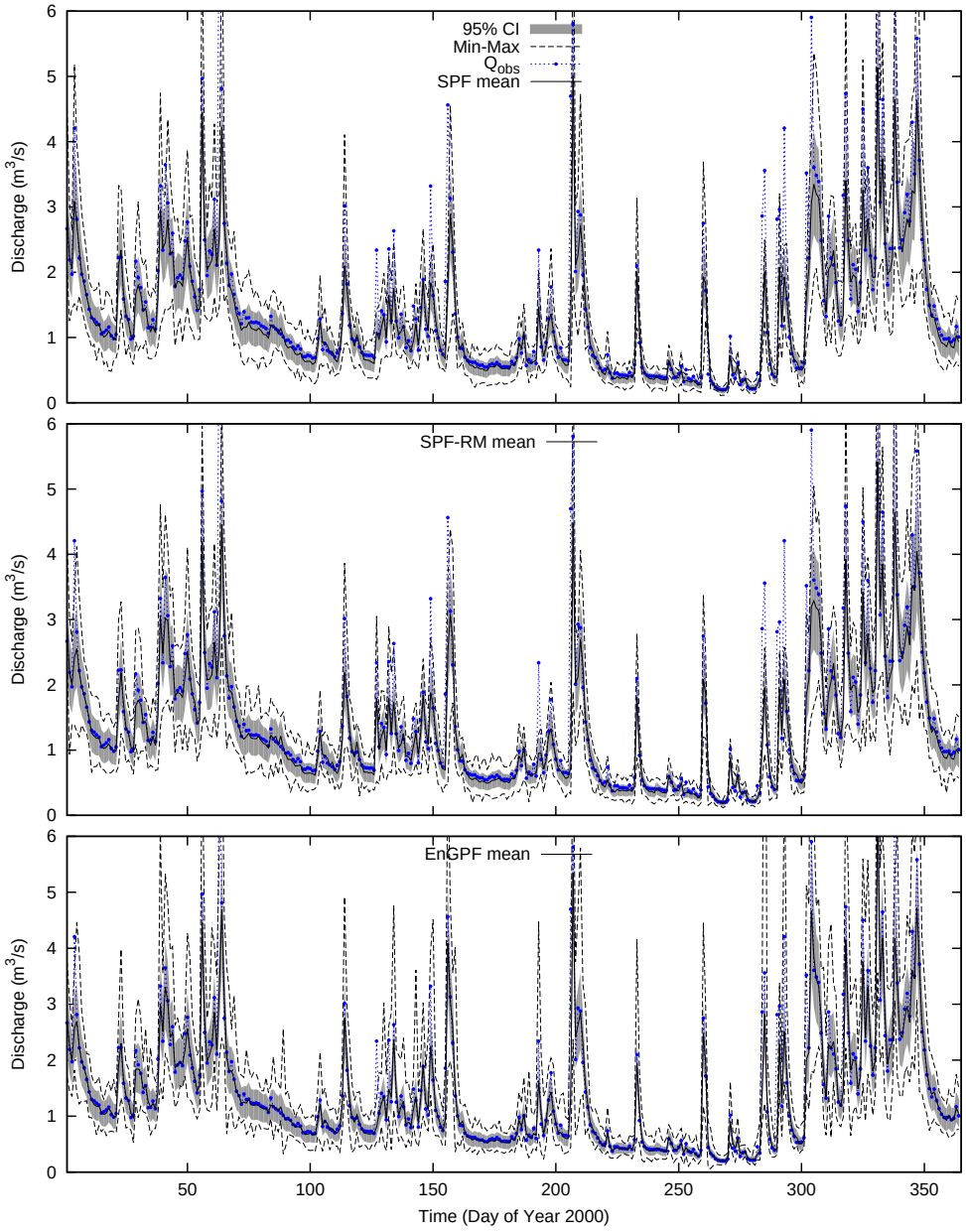


Figure 5.12: Performances corresponding to the particle filters for observation error with standard deviation equal to $0.1 \times Q_{obs}$.

that the EnGPF can be applied to state estimation problems with certain degree of non-Gaussian noise. Nevertheless, further research is needed in identifying to what extent the EnGF is able to perform better or similar to natural non-Gaussian filters such as the SPF and the SPF-RM.

Computational demand

The last column of Table 5.8 shows the computer time demanded by each algorithm. The application of the EnKF involves the computation of matrix operations while in the SPF the computation of the particle weights along with the resampling of particles is required. Although the EnKF and the SPF are based on different theoretical foundations and the corresponding implementations, both filters perform with a similar computational efficiency.

Moreover, the SPF-RM demands more computer time and the EnGPF slightly more compared to SPF and EnKF. This can be explained by the complexity of these filters. For the SPF-RM, the additional computer time demanded by the implementation of the RM step, which involves the generation and selection of a new set of particles, decreases the efficiency of the filter.

The implementation of the EnGPF consist in the application of the EnKF and the GPF. According to Kotecha and Djuric (2003a), the GPF demands less computer time when comparing to the SPF since the application of the resampling step is not required in the GPF. The benefit obtained from this fact is that the EnGPF is computationally more efficient than the SPF-RM with a marginal increase in the time demand when compared to EnKF and SPF. The efficiency of the SPF-RM can be increased by a selective application of the RM step as reported in Moradkhani et al. (2012).

5.10 Summary

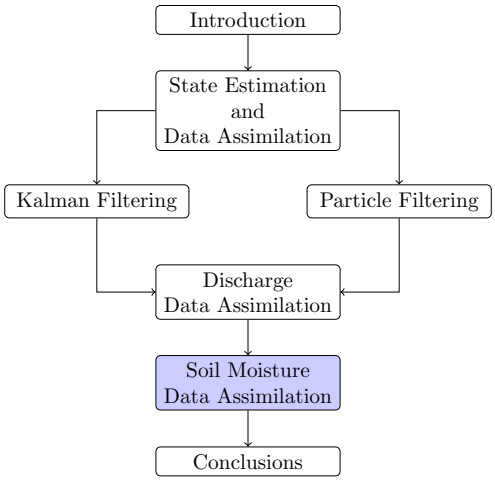
The modified HVB model was introduced in this chapter. The modification allowed for a lumped implementation of the model. The values of the parameters were identified by using an automatic calibration method. Although more advanced calibration techniques are available nowadays, the application of data assimilation methods is still a key component in the hydrologic modeling framework. Thus the remaining part of the chapter focused on the exploration of advanced assimilation methods.

The performance of particle filter based methods was assessed within two comparative studies: a study with artificial generated observations and a second study with in-situ observations. In the synthetic study, the water

storages of the three model reservoirs are retrieved along with the estimation of the true discharge while in the study with in-situ data, the discharge is estimated. The main contribution of this chapter is the analysis of the performance of the ensemble Gaussian particle filter which is based on the theoretical foundations of the ensemble Kalman filter and the Gaussian particle filter. The results showed an outperformance of the proposed filter when compared to the standard implementations of the ensemble Kalman filter and particle filter in a specific hydrologic estimation problem.

Chapter 6

Assimilation of soil moisture data



6.1 Introduction

The ensemble Gaussian particle filter was assessed with a conceptual rainfall-runoff model and the discharge as the variable to be assimilated in the previous chapter. In this chapter, the particle filter with state and parameter resampling is assessed with a physical based model and with the volumetric soil moisture as the variable of interest to be assimilated. Moreover, not only the model skill with respect to soil moisture assimilation is assessed but also with respect to the predicted baseflow.

6.2 The Community Land Model CLM2.0

6.2.1 Introduction

The Community Land Model (CLM) can be classified as a descriptive hydrologic model. In the descriptive modeling approach, all the physical processes occurring both at the land surface and in the soil are modeled. These processes includes for example infiltration, evapotranspiration, surface runoff and ground water flow just to mention some of them. The fundamental principles of a descriptive model are the conservation of energy and the conservation of water. In this context, CLM is able to describe the interaction between the land surface, the vegetation in the land surface and the interaction with the atmosphere. This type of model is also referred to as Soil-Vegetation-Atmosphere Transfer Scheme (SVATS) or Land Surface Model (LSM). Figure 6.1 shows a schematic diagram of CLM with all the processes considered in the model. The upper part of figure 6.1 shows the processes related to the conservation of energy (Biogeophysics) while the lower part indicates the processes related to the conservation of mass (Hydrology).

The development of CLM is the result of the collaboration between experts from different institutes. The aim of this effort is to provide the community with a free model that captures most of the best science currently available for land surface modeling. The CLM is the land model used in the Community Climate System Model (CCSM) and the Community Atmosphere Model (CAM) of the US National Center for Atmospheric Research (NCAR). CCSM is a global climate model with coupled sea ice, land, ocean and atmospheric components communicating through a coupler module. CLM may be run in stand-alone mode, coupled to an atmospheric model, or fully coupled with CCSM. In this work, CLM is used in stand-alone mode since the objective of the research is to apply assimilation with soil moisture data in order to correct the outflow of the catchment. The version of CLM used in this work is 2.0. This version was released in May 2002 and the current version is 4.0. The main difference between these two versions is the addition of biogeochemistry processes and dynamic vegetation processes in version 4.0. The CLM can be applied to a large scale, even up to a global scale. The spatial and temporal structure of CLM is presented below.

6.2.2 CLM structure

The overall structure of the model includes: the core single-point soil-snow vegetation biophysical process code, the land boundary data and the scaling

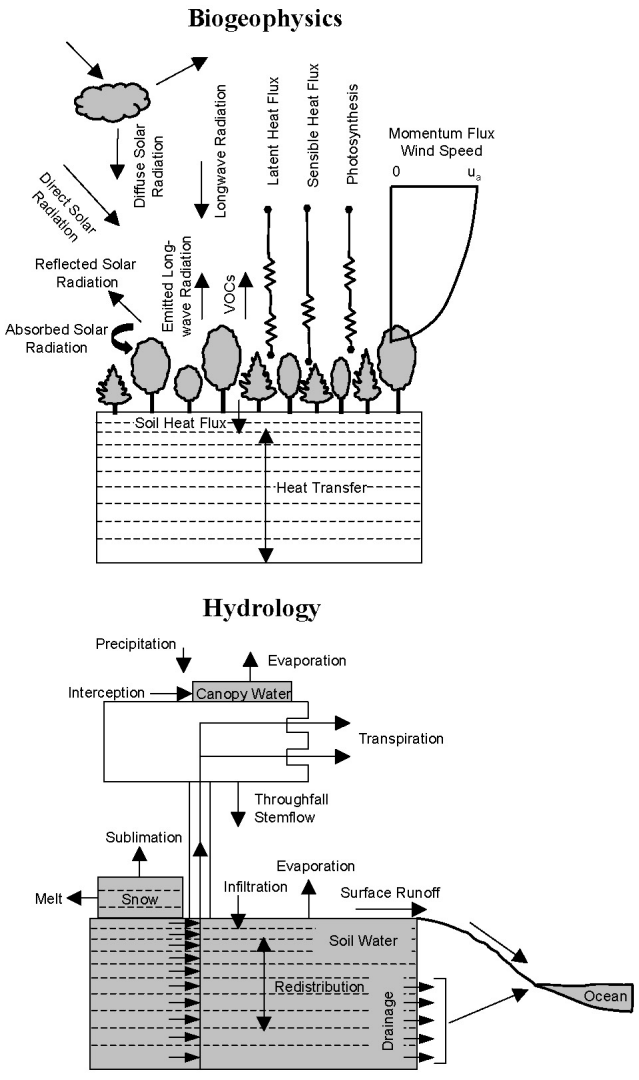


Figure 6.1: Processes related to the Energy and water balance in CLM.

procedures required to interface atmospheric model grid-square inputs to land single-point processes. These components are explained below.

Horizontal structure

The model represents biogeophysical and other processes over a predefined grid by calculating water and heat fluxes and states for every grid cell separately, without any interaction between cells. Each grid cell can be subdivided into several patches, containing a single land cover type. These land cover types are each covering a fraction of the grid cell: vegetation, bare-soil, wetland, lake, urban and glacier. The vegetated portion is further divided into patches of plant functional types.

At every model time step, heat and water flux and state calculations are performed over each patch individually. There is no interaction between the patches, or in other words each patch maintains its own prognostic variables. By default, all patches within a grid cell have the same soil texture, soil color and corresponding physical properties. They respond to the same mean conditions (forcings) of the overlying atmospheric grid cell, and the model grid cell responds to a really-weighted fluxes from the underlying patches.

Vertical structure

The model has one vegetation layer, 10 vertical soil layers that can be changed to the required number by the user, and up to 5 snow layers. For the soil layers, a CCM-like vertical discretization is used, with mesh points specified and interfaces located half way between 2 neighboring layers.

Time-integration scheme

The model is integrated forward with a constant time step. Soil moisture fluxes may depend on the time steps used. After the model initialization, a time loop starts for the calculation of surface fluxes and the update of the state variables for all patches. In the time loop first atmospheric forcings are processed and then the solution is split into an energy balance and a water balance phase in a very modular structure.

Physical processes in CLM

The physics related to soil-water processes are partly described in this section. A complete description is presented in De Lannoy (2006b) and in the Technical Description of CLM Oleson et al. (2004). The focus of this section

is on the description of the moisture content in the soil layers. Moreover, the aim is to clarify where the parameters, which will be calibrated, occur within the model structure. Tunable parameters and constants are typed in bold fonts. Some constants in the default code were turned into tunable parameters according to De Lannoy (2006b). The new parameters are identified with a $^{\diamond}$.

The core single-point code is applied to each land surface patch. Each land surface patch has six state variables which are associated with the energy balance equations and the water balance equations: canopy temperature (Tc), temperature at each node of soil/snow (Tj), canopy interception water store (w_{can}), mass of water within each layer of soil/snow ($w_{liq,j}$), mass of ice within each layer of soil/snow ($w_{ice,j}$) and the snow layer thickness (Δz_j).

Soil moisture can be expressed in several ways. The liquid mass in layer j is given by $w_{liq,j}$. The liquid mass is related to volumetric soil moisture $\theta_{liq,j}$ by:

$$\theta_j = \frac{w_{liq,j}}{\Delta z_j \cdot \boldsymbol{\rho_{liq}}} \; [\mathrm{m}^3.\mathrm{m}^{-3}] \tag{6.1}$$

with

$\boldsymbol{\rho_{liq}}$ = density of liquid [kg.m^{-3}]
Δz_j = thickness of soil layer j [m], with the node index $j = 1, ..., N$
= $z_{h,j} - z_{h,j-1}$, with $Z_{h.j}$ the interface depths (see figure 6.2)

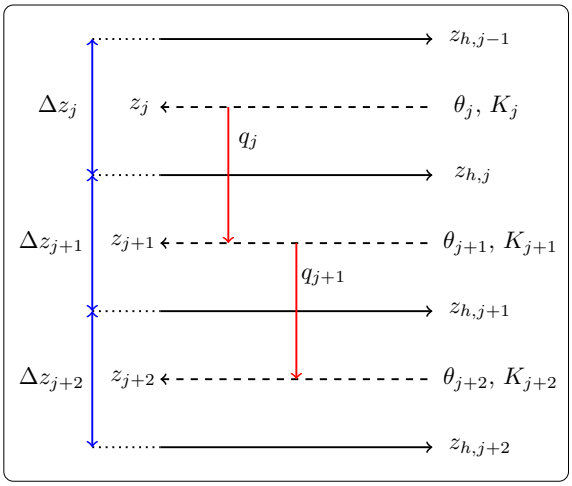


Figure 6.2: Interfaces and soil water flow computation in CLM

The soil wetness is defined as:

$$s_j = \frac{\theta_j}{\boldsymbol{\theta_{sat_j}}} \; [-] \tag{6.2}$$

with

$$\boldsymbol{\theta_{sat_j}} = \text{volumetric soil water at saturation in layer } j \text{ [m}^3\text{.m}^{-3}\text{]}$$
$$\text{parameter } \boldsymbol{watsat_j}$$

In CLM2.0 the vertical water flow is calculated through a combination of the continuity equation for water and Darcy's law. This approach approximates a 1D Richards equation. The vertical water transport is governed by infiltration, surface and subsurface runoff, gradient diffusion, gravity and root extraction through canopy transpiration. Integration of the liquid water mass balance over a soil control volume (soil layer j of thickness Δz_j)) yields:

$$\frac{\partial w_{liq,j}}{\partial t} = (q_{j-1} - q_j) - f_{root,j} E_{tr} + (M_{ice,liq} \Delta z)_j \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \tag{6.3}$$

with

q_j = water flow between neighboring layers, positive downward [kg.m^{-2}.s^{-1}]

$f_{root,j} E_{tr}$ = water extracted by plant roots, transpiration [kg.m^{-2}.s^{-1}]

$M_{ice,liq}$ = rate of phase change (ice to liquid) [kg.m^{-3}.s^{-1}]

- q_0: at the surface, the flow for bare soil is given by the infiltration:

$$q_0 = G_w - R_s - E_{g,Surf} \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \tag{6.4}$$

with

G_w = sum of effective precipitation and snowmelt [kg.m^{-2}.s^{-1}]

R_s = surface runoff [kg.m^{-2}.s^{-1}]

$E_{g,Surf}$ = soil surface evaporation [kg.m^{-2}.s^{-1}]

  - G_w: The total amount of liquid water that really reaches the ground is composed by throughfall, drip from canopy and snowmelt.

$$G_w = D_d + D_r + S_m \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \tag{6.5}$$

with

$$S_m = \text{snowmelt [kg.m}^{-2}\text{.s}^{-1}\text{]}$$

For snow covered soil, the infiltration is given by $q_0 = S_m - R_s$.

  - R_s: the total surface runoff is the sum of the surface runoff from saturated and unsaturated regions:

$$R_s = (1 - f_{sat}) \bar{w}_s^4 G_w + f_{sat} G_w \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \tag{6.6}$$

with

$$
\begin{aligned}
f_{sat} &= \text{partial contributing area, saturated fraction [-]} \\
&= \boldsymbol{w_{fact}} \cdot \exp(-z_w) \\
w_{fact} &= \text{determined by the distribution of the topographic index, parameter } \boldsymbol{wtfact} \text{ [-]} \\
z_w &= \text{mean water table depth} \\
&= \boldsymbol{f_z}(z_{h,N} - \Sigma_{j=1,N} s_j \Delta z_j) \text{ [-]} \\
z_{h,N} &= \text{bottom depth of the lowest soil layer [m]} \\
\boldsymbol{f_z} &= \text{water table depth scale parameter [m}^{-1}\text{]} \\
&\quad \text{parameter } \boldsymbol{fz}^{\diamond} \\
s_j &= \text{soil wetness [-]} \\
\bar{w}_s &= \text{soil layer thickness weighted soil wetness [-] in the top } \boldsymbol{NwRs} \text{ layers} \\
&= \frac{\Sigma_{j=1,NwRs} s_j \Delta z_j}{\Sigma_{j=1,NwRs} \Delta z_j} \\
\boldsymbol{NwRs} &= \text{last top layer contributing to the calculation of surface runoff [-],parameter } \boldsymbol{NwRs}^{\diamond}
\end{aligned}
$$

Note that the top soil layer is impermeable, if the effective porosity $(\theta_{sat_j} - \theta_{ice,j})$ is less than a predefined parameter $\boldsymbol{wimp}$ [-]. In that case, $\bar{w}_s$ reduces to 1 and all the water reaching the soil surface runs off.

- $E_{g,Surf}$: the evaporation from bare soil is determined by the profile of the humidity of the air and by the amount of water in the upper soil/snow layer.

- q_N: at the bottom of the soil column, the flow is simply given by the hydraulic conductivity, i.e. assuming no change of soil matrix potential with depth, as for free drainage.

- q_j: the water flow from one soil layer to another is determined by Darcy's law:

$$q_j = -K_j \frac{(\psi_{j+1} - \psi_j) - (z_{j+1} - z_j)}{z_{j+1} - z_j} \text{ [kg.m}^{-2}\text{.s}^{-1}\text{]} \qquad (6.7)$$

with

z_j = node depth of layer j, positive, increasing downward [mm]
$\psi_j - z_j$ = hydraulic height [mm], with $-z_j$ the gravitational potential, with the reference elevation at the soil surface
K_j = unsaturated hydraulic conductivity in layer j [mm.s^{-1}]
ψ_j = matrix potential of soil layer j [mm] (< 0 in the unsaturated zone)

- $f_{root,j}E_{tr}$: the water extracted by plant roots is removed from soil by transpiration. For each layer the effective root fraction $f_{root,j}$ is based on plant physiological characteristics (root distribution) and the soil matrix potential:

$$f_{root,j} = \frac{f_{r,j}w_{LT}(j)}{10^{-10} + \Sigma_{j=1,N} f_{r,j}w_{LT}(j)} \tag{6.8}$$

$$w_{LT}(j) = \frac{\boldsymbol{\psi_{max}} - \psi_j}{\boldsymbol{\psi_{max}} + \boldsymbol{\psi_{sat}}} \tag{6.9}$$

with

$f_{r,j}$ = root fraction within soil layer j [-], calculated based on a pft-parameter for root distribution ***roota_par***
$W_{LT}(j)$ = transpiration restricting factor [-], ranging from 0 at the permanent wilting point to 1 at saturation
$\boldsymbol{\psi_{max}}$ = maximum value of negative of leaves potential before dessication or wilting point potential [mm], parameter ***smpmax*** (negative value)
$\boldsymbol{\psi_{sat}}$ = saturated soil water potential [mm], parameter ***sucsat*** (positive value)
ψ_j = matrix potential limited by $\boldsymbol{\psi_{max}}$ [mm]

The denominator of Equation (6.8) is defined as the soil water transpiration factor (0 to 1) [-].

The numerical scheme is obtained by rewriting Equation (6.3) for liquid water, which yields an expression for the soil moisture in soil layer j (of thickness Δz_j, not to be confused with the possible notation for spatial discretization) at time step $n + 1$:

$$\frac{\Delta\theta_j^{n+1}\Delta z_j^{n+1}}{\Delta t} = [q_{j-1}^{n+1} - q_j^{n+1}] - f_{root,j}E_{tr}\,[\text{mm.s}^{-1}] \tag{6.10}$$

with

$\Delta\theta_j^{n+1}$ = $(\theta_j^{n+1} - \theta_j^n)$ [-], change in volumetric soil liquid water in layer j during time interval Δt.

Δz_j = soil layer thickness expressed in mm, since $\boldsymbol{\rho_{liq,j}}$ is not included

which can be further numerically expanded as:

$$\frac{\Delta\theta_j^{n+1}\Delta z_j^{n+1}}{\Delta t} = q_{j-1}^n + \frac{\partial q_{j-1}}{\partial\theta_{j-1}}\Delta\theta_{j-1}^{n+1} + \frac{\partial q_{j-1}}{\partial\theta_j}\Delta\theta_j^{n+1} - q_j^n - \frac{\partial q_j}{\partial\theta_j}\Delta\theta_j^{n+1} - \frac{\partial q_j}{\partial\theta_{j+1}}\Delta\theta_{j+1}^{n+1} - f_{root,j}E_{tr} \quad (6.11)$$

In general Equation (6.11) can be written as:

$$a_j\Delta_{j-1}^{n+1} + b_j\Delta\theta_j^{n+1} + c_j\Delta\theta_{j+1}^{n+1} = r_j \quad (6.12)$$

with

$$\begin{aligned} a_j &= -\left[\frac{\partial q_{j-1}}{\partial\theta_{j-1}}\right] \\ b_j &= \left[\frac{\Delta z_j}{\Delta t} - \frac{\partial q_{j-1}}{\partial\theta_j} + \frac{\partial q_j}{\partial\theta_j}\right] \\ c_j &= \left[\frac{\partial q_j}{\partial\theta_{j+1}}\right] \\ r_j &= [q_{j-1}^n - q_j^n] + f_{root,j}E_{tr} \end{aligned}$$

and

$$\begin{aligned} a_j &= 0, \qquad \text{if } j = 1 \\ c_j &= 0, \qquad \text{if } j = N \end{aligned}$$

Application of this equation to each node results in a tridiagonal matrix:

$$\begin{bmatrix} b_1 & c_1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ a_2 & b_2 & c_2 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & a_3 & b_3 & c_3 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \dots & a_{N-1} & b_{N-1} & c_{N-1} \\ 0 & 0 & 0 & 0 & 0 & \dots & 0 & a_N & b_N \end{bmatrix} \begin{bmatrix} \Delta\theta_1^{n+1} \\ \Delta\theta_2^{n+1} \\ \Delta\theta_3^{n+1} \\ \vdots \\ \Delta\theta_{N-1}^{n+1} \\ \Delta\theta_N^{n+1} \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \vdots \\ r_{N-1} \\ r_N \end{bmatrix} \quad (6.13)$$

Once the terms $\Delta\theta_j^{n+1}$ are known, $w_{liq,j}^n$ can be updated by:

$$w_{liq,j}^{n+1} = w_{liq,j}^n + \Delta\theta_j^{n+1}\Delta z_j \quad (6.14)$$

The terms in the coefficients, $\frac{\partial q_j}{\partial \theta_j}$ and $\frac{\partial q_j}{\partial \theta_{j+1}}$ are found by derivation of the flow q through the interfaces at the surface, the interior interfaces and through the bottom interface:

$$q_0^{n+1} = q_0^n + \frac{\partial q_0}{\partial \theta_1}\Delta\theta_1 = \text{infiltration} \tag{6.15}$$

$$q_j^{n+1} = q_j^n + \frac{\partial q_j}{\partial \theta_j}\Delta\theta_j + \frac{\partial q_j}{\partial \theta_{j+1}}\Delta\theta_{j+1} \tag{6.16}$$

$$\text{with } q_j = -K_j \frac{(\psi_{j+1-\psi_j})-(z_{j+1}-z_j)}{z_{j+1}-z_j} \tag{6.17}$$

$$q_N^{n+1} = q_N^n + \frac{\partial q_N}{\partial \theta_N}\Delta\theta_N \tag{6.18}$$

$$\text{with } q_N = -K_N(0-1) = K_N \tag{6.19}$$

The respective derivatives are given by:

$$\frac{\partial q_0}{\partial \theta_1} = -\frac{\partial E_{g,Surf}}{\partial \theta_1} = S_d \Delta z_1 \tag{6.20}$$

$$\frac{\partial q_j}{\partial \theta_j} = -\left[\frac{K_j}{z_{j+1}-z_j}\frac{\partial(-\psi_j)}{\partial \theta_j} + \frac{\partial K_j}{\partial \theta_j}\frac{(\psi_{j+1}-\psi_j)-(z_{j+1}-z_j)}{z_{j+1}-z_j}\right] \tag{6.21}$$

$$\frac{\partial q_j}{\partial \theta_{j+1}} = -\left[\frac{K_j}{z_{j+1}-z_j}\frac{\partial(\psi_{j+1})}{\partial \theta_{j+1}} + \frac{\partial K_j}{\partial \theta_{j+1}}\frac{(\psi_{j+1}-\psi_j)-(z_{j+1}-z_j)}{z_{j+1}-z_j}\right] \tag{6.22}$$

$$\frac{\partial q_N}{\partial \theta_N} = \frac{\partial K_N}{\partial \theta_N} \tag{6.23}$$

with

$E_{g,Surf}$ = ground evaporation [mm.s^{-1}]

S_d = extrapolates soil water dependence of ground evaporation, currently not used in the code and set to 0.

- ψ_j: the soil matrix potential [mm] and its partial derivative at node depth z_j are given by:

$$\psi_j = \max\left[\text{-}\boldsymbol{\psi_{sat}}\left[\frac{\theta_j}{\boldsymbol{\theta_{sat_j}}}\right]^{-B}, \boldsymbol{\psi_{min}}\right] \text{ [mm]} \tag{6.24}$$

$$\frac{\partial \psi_j}{\partial \theta_j} = -\psi_j\left[\frac{\boldsymbol{B}}{\theta_j}\right] \text{ [mm]} \tag{6.25}$$

with

$\boldsymbol{\psi_{sat}}$ = saturated soil water potential [mm]
parameter ***sucsat*** (positive value)

$\boldsymbol{\psi_{min}}$ = restriction for minimal soil water potential [mm],
parameter ***smpmin*** (negative value)

$\boldsymbol{B}$ = parameter ***bws*** [-], defined by Clapp and Hornberger [1978]

- K_j: the unsaturated hydraulic conductivity and its derivative at layer j are calculated based on the saturated hydraulic conductivity,$\boldsymbol{K_{sat_j}}$, which is assumed to decrease exponentially with depth and calculated at the interface depth, $z_{h,j}$:

$$\boldsymbol{K_{sat_j}} = \boldsymbol{K_{sat_0}} . \exp\left(- \frac{z_{h,j}}{\boldsymbol{z^*}} \right) [\text{mm.s}^{-1}] \tag{6.26}$$

with

$\boldsymbol{K_{sat_0}}$ = saturated hydraulic conductivity at the surface, parameter $\boldsymbol{xksat_0}$ [mm.s^{-1}]

$\boldsymbol{z^*}$ = length scale for K_{sat_j} decrease, parameter $\boldsymbol{hkdepth}$ [-]

The unsaturated hydraulic conductivity at layer j is then given by:

$$K_j = \boldsymbol{K_{sat_j}} . s_j^{(2B+3)} \ [\text{mm.s}^{-1}] \tag{6.27}$$

Consequently, the hydraulic conductivity and its partial derivative at interface depth $z_{h,j}$ are given by:

$$k_j = \boldsymbol{K_{sat_j}} \left[\frac{0.5.(\theta_j + \theta_{j+1})}{0.5.(\boldsymbol{\theta_{sat_j}} + \boldsymbol{\theta_{sat_{j+1}}})} \right]^{2B+3} \quad \text{if } 1 \leq j \leq N-1 \tag{6.28}$$

$$= \boldsymbol{K_{sat}} \left[\frac{\theta_j}{\boldsymbol{\theta_{sat_j}}} \right]^{2B+3} \quad \text{if } j = N \tag{6.29}$$

$$\frac{\partial K_j}{\partial \theta_j} = \frac{\partial K_j}{\partial \theta_{j+1}} \tag{6.30}$$

$$= (2B+3) \boldsymbol{K_{sat_j}} \left[\frac{0.5.(\theta_j + \theta_{j+1})}{0.5.(\boldsymbol{\theta_{sat_j}} + \boldsymbol{\theta_{sat_{j+1}}})} \right]^{2B+2} \left(\frac{0.5}{\boldsymbol{\theta_{sat_j}}} \right) \quad \text{if } 1 \leq j \leq N-1 \tag{6.31}$$

$$= (2B+3) \boldsymbol{K_{sat_j}} \left[\frac{\theta_j}{\boldsymbol{\theta_{sat_j}}} \right]^{2B+2} \left(\frac{1}{\boldsymbol{\theta_{sat_j}}} \right) \quad \text{if } j = N \tag{6.32}$$

as the difference between the interface depth and the upper and lower node is per definition equal. If the effective porosity of either layer j or $j+1$ is less than $\boldsymbol{wimp}$, then $K_j = 0$ and $\frac{\partial K_j}{\partial \theta_j} = 0$.

Note that the interface depth is by definition exactly halfway in between two subsequent nodes and therefore soil moisture at both nodes contribute equally to the hydraulic conductivity and the flow through one interface.

Two subsequent interfaces are not situated equally far from a node point, but the flows through 2 subsequent interfaces are assumed to contribute equally to the calculation of soil moisture at a node. After the determination of the surface runoff and the infiltration, the water content in the soil layers is calculated using Darcy's equation. Next, the soil moisture profile is updated for baseflow runoff and water in excess in the soil column. The base-flow drainage, R_b, is obtained as the sum of the baseflows for unsaturated and saturated regions:

$$R_b = (1 - f_{sat}).\boldsymbol{K_d}.\bar{w}_b^{(2B+3)} + f_{sat}.\boldsymbol{l_d}.\exp(-z_w)\ [\text{kg.m}^{-2}.\text{s}^{-1}] \tag{6.33}$$

with

$$\begin{aligned}
R_b &= \text{baseflow drainage [kg.m}^{-2}\text{.s}^{-1}] \\
\boldsymbol{K_d} &= \text{saturated soil hydraulic conductivity at the bottom,} \\
&\quad \text{parameter } \boldsymbol{k_d^\diamond}\ [\text{mm.s}^{-1}] \\
\boldsymbol{l_d} &= \text{base flow parameter for saturated fraction,} \\
&\quad \text{parameter } \boldsymbol{ld^\diamond}\ [\text{mm.s}^{-1}] \\
\bar{w}_b &= \text{weighted soil wetness for bottom } \boldsymbol{NwRb} \text{ layers [m.mm.s}^{-1}] \\
&= \frac{\sum_{j=NwRb,N} s_j.K_j\Delta z_j}{\sum_{j=NwRb,N} K_j\Delta z_j} \\
\boldsymbol{NwRb} &= \text{first bottom layer contributing to the calculation of} \\
&\quad \text{surface runoff [-], parameter } \boldsymbol{NwRb^\diamond}
\end{aligned}$$

The bottom soil layers (from $\boldsymbol{NwRb}$ through N) are updated by:

$$\begin{aligned}
w_{liq,j} &= w_{liq,j} - R_b \frac{\Delta z_j K_j}{\sum_{\boldsymbol{j=nNwRb,N}} K_j\Delta z_j}.\Delta t\ [\text{kg.m}^{-2}] \\
&\quad \text{for } \boldsymbol{NwRb} < j < N-1
\end{aligned} \tag{6.34}$$

All soil layers are finally updated by compensating shortages of water in one layer by moving water from the soil layer below. Shortage in the bottom layer is filled by water from R_b, resulting in R_b^1. Water in excess of saturation, R_{excess} for each depth is added to the remaining R_b^1 and removed from the soil column. From the bottom layer, an outflux of $q_N = K_N$ is contributing to the total baseflow drainage, so that the total runoff, R_{tot}, is given by:

$$R_{tot} = R_b + R_{excess} + K_N + \frac{\partial K_N}{\partial \theta_N}\Delta\theta_N\ [\text{mm.s}^-1] \tag{6.35}$$

Water is allowed to pond on the surface soil layer so that the maximum amount of water for this layer is defined as:

$$w_{liq,j} = \boldsymbol{w_{pond}} + \boldsymbol{\theta_{sat_1}}\Delta z_1 \tag{6.36}$$

with w_{pond} the ponding depth [mm], parameter $\boldsymbol{pondmx.}$

6.3 Site Description

The area (Figure 6.3) to be studied is located in the Grand Duchy of Luxembourg and includes the drainage area expanded from the head of the Alzette River basin, 4 km south of the French-Luxembourg border, to the stream gauge located in Pfaffenthal (Luxembourg City).

The discharge area covers a surface of 356 km^2 and consists of about 50% cultivated land, 22% urban centers and 28% woodland. The topography of the floodplain is characterized by a natural sandstone bottleneck which is located near Luxembourg city. The valley located upstream of the bottleneck is up to 2.5 km wide, while in the Luxembourg sandstone the valley is only 75 m wide. The geological substratum is dominated by marls on the left bank and by limestone and sandstones deposits on the right bank. Sand and gravel, as well as marls and clay alternate in the alluvial deposits covering the stratum. A gauging station, operated since 1996, is located around the village of Livange providing accumulated precipitation amounts with a sampled frequency of 15 min. The meteorological station at Findel Airport is operated in the vicinity of the catchment.

6.4 Model Description

The Community Land Model (CLM2.0) is the hydrologic model used in this study. CLM2.0 simulates land surface processes by calculating water and heat fluxes for each grid cell separately, without any interaction between cells. Each grid cell can be subdivided into several patches, containing one single land cover type such us urban, vegetated, wetlands, glacier and lake. The vegetated fraction is further subdivided into patches of plant functional types, which maintain their own prognostic variables (i.e., a vegetated land cover with 4 patches representing 4 different plant functional types). In this study, CLM2.0 was adapted in order to be able to use the individual patches as ensemble members according to De Lannoy et al. (2006a).

The meteorological forcings required by the model are the air temperature, wind speed, specific humidity, incoming solar radiation, and precipitation. The meteorological forcings were assumed to be spatially uniform over the complete study area. Vertical layers in CLM2.0 embody one vegetation layer, up to ten soil layers and up to 5 snow layers. In this application, soil layers depths were set to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100 cm. CLM2.0 computes the surface runoff and the baseflow for every grid cell. The discharge is routed to the basin outlet using the linear unit hydrograph approach of Troch et al. (1994).

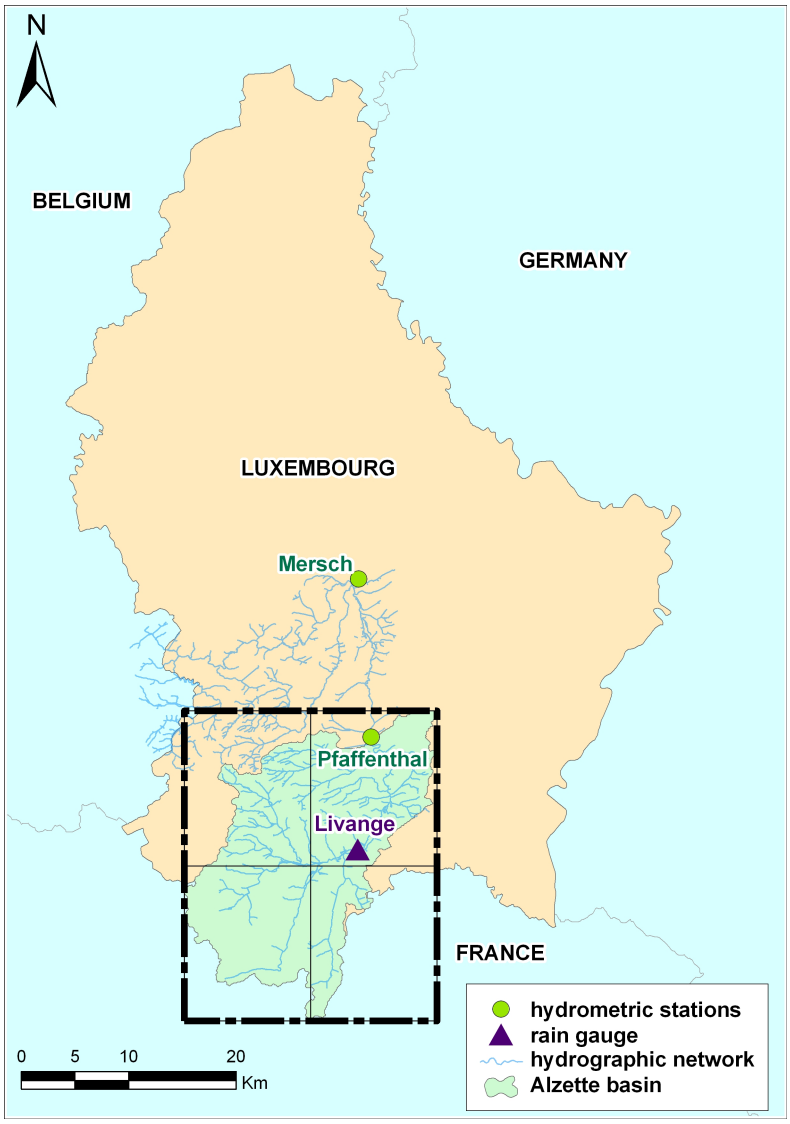


Figure 6.3: The study area: the discharge area in the Alzette river basin is indicated by the green patch.

In CLM2.0, each grid cell contains around 30 model parameters related to the different physical processes represented by the model such as the canopy water balance, the soil water balance, and the energy balance. From these 30 parameters, 10 parameters are related to the soil water balance. The reduced parameter set allows for the application of automatic calibration algorithms, such as the Shuffled complex evolution approach (Duan et al., 1993b) which was used in this study. Table 6.1 presents the description of the selected parameters and three corresponding sets of optimal parameter values (set 1, set 2, set 3) which yield a similar good model performance. Due to the size of the parameter space and the complexity of the model, the system is prone to the equifinality phenomenon. The optimal values were identified by minimizing the Root Mean Square Error (RMSE) between observed and simulated discharge during the year 2006. The three parameter sets will be used to validate the data assimilation methodology.

The model is applied using a constant hourly time step and the study area is represented by 4 grid cells at a 10 km x 10 km resolution which is consistent with the resolution of large scale models. For the sake of clarity in the presentation of the algorithm performances, results corresponding to the cell located in the lower left quadrant in figure 6.3 are presented.

6.5 Experimental Setup

A synthetic soil moisture data assimilation study is performed to assess the performance of the filters. Soil moisture assimilation has received a lot of attention during the last decades, but insight in the impact of soil moisture assimilation on dependent variables, for instance discharge, has been limited (Pauwels et al., 2002; De Lannoy et al., 2007b; Brocca et al., 2010).

For each model grid cell, synthetic volumetric soil moisture observations, corresponding to the top 10 cm soil layer, are generated with the CLM2.0. The generation of the observation consists of the perturbation of the model parameters presented in table 1 (set 4) and the perturbation of the forcings. Parameters and forcings were perturbed by white Gaussian noise with zero mean and the standard deviation set to 1% of the parameter value and 1% of the maximum forcing value. Furthermore, it is worth mentioning the following two facts: first, the small level of noise used in the perturbation of the parameters and forcings can limit the representation of a real-case model error and second, the use of a different parameter set (set 4) in the generation of the synthetic observations introduces bias in the observations themselves. Therefore, the synthetic experiment is focused more on the study of the performances of the filters in the removal of bias and may not represent most of the real-world situations. However, the way how the experiment is

Table 6.1: Optimal parameter sets: **NwRb** and **NwRs** were converted into tunable parameters (De Lannoy, 2006b), k is the soil layer index. **NwRb** and **NwRs** are not considered in the parameter resampling step.

Description	**set 1**	**set 2**	**set 3**	**set 4**
Fraction of model area with high water table (**wtfact**[fraction])	0.280	0.704	0.742	0.7174
Water table depth scale parameter (**fz** [m^{-1}])	49.173	3.423	3.475	3.523
Saturated soil hydraulic conductivity (**kd** [$mm \cdot s^{-1}$])	0.827	0.095	0.099	0.098
Base flow parameter for saturated fraction of watershed (**ld** [$mm \cdot s^{-1}$])	0.0071	0.0034	0.0027	0.0038
First bottom layer contributing to the calculation of base flow (**NwRb** [-])	5	5	6	5
Last top layer contributing to the calculation of the surface runoff (**NwRs** [-])	3	4	4	4
Clapp and Hornberger constant (**bsw**$_k$ [-])	5.487	4.659	4.623	5.919
Volumetric soil water at saturation (**watsat**$_k$ [-])	0.638	0.597	0.600	0.617
Hydraulic conductivity at saturation (**hksat**$_k$ [$mm \cdot s^{-1}$])	0.047	0.011	0.010	0.024
Minimum soil suction (**sucsat**$_k$ [mm])	284.76	557.17	606	497.16

carried out allows to demonstrate the applicability of the particle filter in this study-case.

The forecast uncertainty is introduced through the generation of soil moisture random samples, which is referred to ensemble generation. The meteorological forcings and the model parameters were disturbed with an additive zero mean white Gaussian noise in order to obtain the soil moisture ensemble (De Lannoy et al., 2006a). The standard deviation of this random number for the parameters was set to a predefined fraction of the parameter value. In order to check for the correctness of the ensemble, two different ensemble verification measures were used (De Lannoy et al., 2006a). The ensemble spread ($ensp_t$), the ensemble mean square error (mse_t), and the ensemble skill ($ensk_t$) have to be computed first and at each time step t:

$$\begin{aligned} ensp_t &= \frac{1}{N}\sum_{i=1}^{N}(\hat{z}_{t,i} - \bar{\hat{z}}_t)^2 \\ mse_t &= \frac{1}{N}\sum_{i=1}^{N}(\hat{z}_{t,i} - z_t)^2 \\ ensk_t &= (\bar{\hat{z}}_t - z_t)^2 \end{aligned} \tag{6.37}$$

where $\hat{z}_t$ is the variable to be estimated and z_t is the corresponding observation of the estimated variable at time step t. In order to have a large enough ensemble spread, on average the ensemble mean differs from the observation by a value that is equal to the time average of the ensemble spread. Therefore, the following expression should be true:

$$\frac{< ensk >}{< ensp >} \approx 1 \tag{6.38}$$

where $< . >$ indicates an average over the simulation period. Furthermore, if the truth is statistically indistinguishable from a member of the ensemble, the following expression should be true:

$$\frac{< \sqrt{ensk} >}{< \sqrt{mse} >} \approx \sqrt{\frac{N+1}{2N}} \tag{6.39}$$

For the selection of the ensemble/particle size, the assimilation algorithms were evaluated using three different ensemble sizes: 64, 128 and 256. Figure 6.4 shows that the improvement obtained when increasing the size is not very significant while the increase in the computational time demand is very significant. Therefore, an ensemble size of 64 is used in this study.

The standard deviation of the perturbation noises corresponds to 10% of the nominal values for the model parameters, and 1% for the meteorological forcings. These fractions have been calibrated in order to balance the different

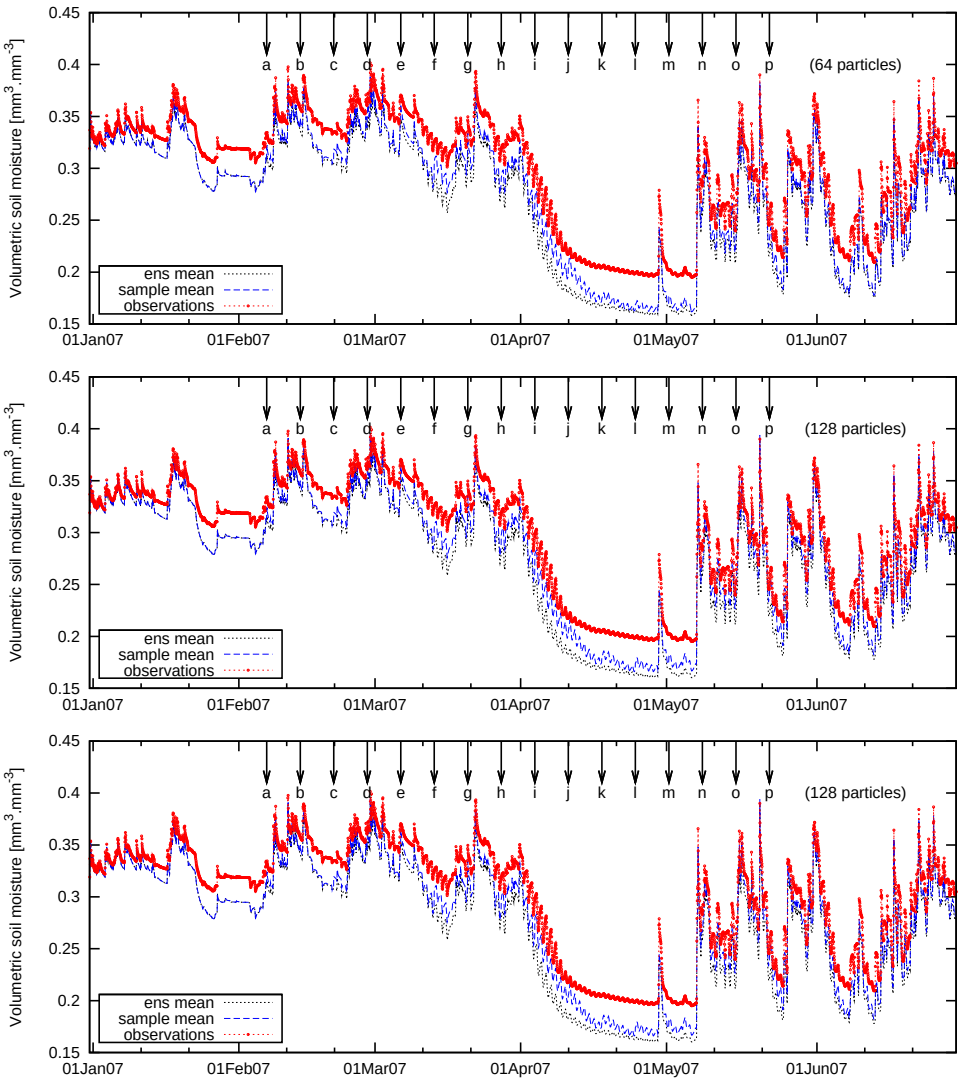


Figure 6.4: Performance of the particle filter concerning 3 different particle sizes.

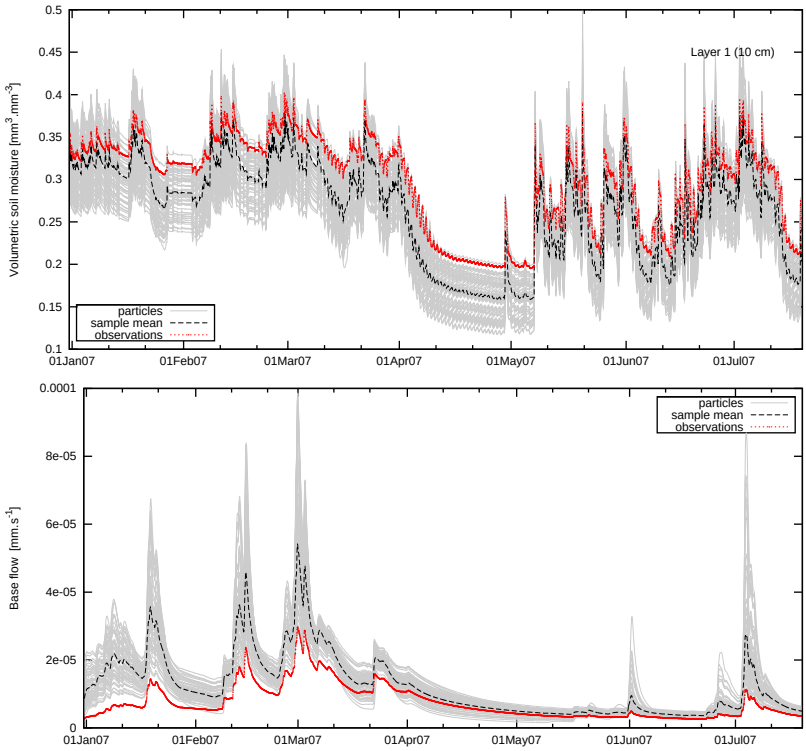


Figure 6.5: Soil moisture and baseflow ensembles: The upper plot corresponds to the generation of the volumetric soil moisture ensemble with the ensemble members in gray line, the ensemble mean in black line and the synthetic soil moisture observations in red dotted line. The lower plot corresponds to the baseflow ensemble.

sensitivities of model parameters and meteorological forcings in the generation of an adequate ensemble. Figure 6.5 shows the soil moisture ensemble and the corresponding baseflow ensemble, the ratio $< ensk > / < ensp >$ is equal to 1.09 which approximates 1 and the ratio $< \sqrt{ensk} > / < \sqrt{mse} >$ is equal to 0.72 which approximates the value of $\sqrt{1/2}$ with the simulation period corresponding to year 2007.

A robustness test of the assimilation algorithms will be performed by considering the impact of the data assimilation frequency and of different optimal parameter values for the model integration. Discussion on the filter performances for these scenarios will be extended in the results section.

6.6 Particle filter based assimilation of soil moisture

6.6.1 Introduction

In this section, the standard implementation of the particle filter with the additional resampling of the model parameters is presented. Although the resampling of the model parameters involve a joint state and parameter estimation, the goal of the application of this method in this study case (soil moisture assimilation with a distributed model) is to improve model flow predictions rather than an explicit identification of model parameter values.

A preliminary comparative study between the EnKF and the SPF showed a deterioration of the predicted baseflow for both filters. Therefore, the addition of the parameter correction in the SPF algorithm was adopted as a way to improve model flows which depend on the soil moisture. The functioning of the EnKF and SPF differs in the formulation and the corresponding update of each of the ensemble members/particles. In the EnKF, all the ensemble members at time step t are updated using the same approximated Kalman gain and the innovation factor which depends on the perturbed measurements and the observation model. On the other hand, in the SIR filter all the particles at time step t are weighted and resampled. The replication and suppression of particles decrease the particles variance limiting the state space representation and possibly affecting the behavior of variables (surface and sub-surface model flows) related to the assimilated state. The results of the application of the EnKF are not shown within this chapter but they are presented in Appendix 7.2.2 since the scope of this section is limited to the application of particle filter-based methods in this study case. The results of the SPF application are presented further with a discussion section concerning the degree of affectance of the baseflow.

6.6.2 The standard particle filter with parameter resampling (SPF-PR)

In particle filtering theory, the uncertainty in the model is represented through samples referred to as particles. These samples are drawn from the importance density function (equal to the prior density function for the standard particle filter).

The uncertainty in the model is caused by uncertainty in the meteorological forcings, initial conditions and parameters. Thus, the generation of ensembles, presented in the experiment setup section, is fundamental since the ensemble should represent this model uncertainty. The perturbation of the

parameters plays an important role in the generation of the ensemble due to the contribution of the parameters to the modeling errors.

The state estimation method aims at finding the optimal state value based on the information from the measurements. The estimated state value can positively or negatively affect the behavior of other variables in the model. In this study, soil moisture is the state variable that will affect the baseflow.

Moradkhani et al. (2005) presented the SPF-PR as a potential methodology to assess the uncertainty in the states and parameters of a hydrologic model. In this study we adopted the same methodology aiming at the correction of model flows after the assimilation of the states. The hydrologic literature on Data Assimilation with the Particle Filter focuses either on direct assimilation of discharge or on an evaluation of the improvement in the assimilated variable itself. In this paper, we assess whether the resampling of the parameters along with the states improves the behavior of the model flows due to a proper combination between states and parameters.

The operation of the parameter resampling step is the following: after the resampling of the states, the same vector/matrix containing the particle indices to be resampled is used to resample the parameter set. The last action leads to a selection (replication or suppression) of parameters that are tied to a particular state realization.

An extreme replication of the parameter values poses the same problem as in the case of the state replication. Moreover, the ensemble will fail in the representation of the model uncertainty since the spread of the ensemble is decreased after the parameter resampling. In order to overcome this problem, the resampled parameter values are perturbed with the addition of white Gaussian noise (Moradkhani et al., 2005) and the variance (var) of the noise is set to a fraction of the optimal parameter value.

The SPF-PR filter applied in this study is summarized in the following algorithm:

6.7 Results and Discussion

The data assimilation experiments are validated by comparing soil moisture and baseflow assimilation results against synthetic observed soil moisture and baseflow values. The reference model integration without data assimilation is performed with parameter set 2, while the assimilation integrations are performed with the different parameter sets 1,2 and 3.

Data assimilation is performed every week, with the first DA event at 08 February 2007 and the last at 24 May 2007. Every DA event is indicated

Algorithm 12 Standard Particle Filter with Parameter Resampling

For $t = 1$ to the number of DA time steps

1. Forecast step:
   $\mathbf{x}_{t,i} \sim p(\mathbf{x}_t|\mathbf{x}^*_{t-1,i}), \quad i = 1, 2, \ldots, N$
2. Analysis step:
   $w_{t,i} = p(\mathbf{y}_t|\mathbf{x}_{t,i}) \quad i = 1, 2, \ldots, N$
   $\tilde{w}_{t,i} = \dfrac{w_t(\mathbf{x}_{t,i})}{\sum_{i=1}^{N} w_t(\mathbf{x}_{t,i})}$
   Obtain the resampling index vector according to StrR.
   Resample the state vector $\{\mathbf{x}_{t,i}\} \Rightarrow \{\mathbf{x}^*_{t,i}\}$
   Resample the model parameters θ: $\{\theta_i\} \Rightarrow \{\theta^*_i\}$
   Perturb the resampled parameter set $\theta_i + \mathcal{N}(0, var)$
   Obtain the resample set $\mathbf{x}^*_{t,i}$ according to StrR.
   Assign $\tilde{w}_{t,i} = \frac{1}{N}$
   $\overline{E[\mathbf{x}_t]} = \dfrac{1}{N} \sum_{i=1}^{N} \mathbf{x}^*_{t,i}$

by a black arrow in the figures and the simulation period corresponds to the first half of year 2007 (01 January-01 July). The Root Mean Square Error (RMSE), between the synthetic observed and modeled soil moisture and baseflow, is computed over the time period starting 1 day before the first DA event and 1 day after the last DA event (i.e., from 7 February to 25 May 2007).

6.7.1 SPF filter performance

Figure 6.6 shows the performance of the SPF filter for soil moisture assimilation and the corresponding impact of the assimilation on the baseflow. According to the RMSE values: 3.65 vol% without assimilation and 2.19 vol% after assimilation, the improvement obtained from the SPF filter application is significant. However, when looking at the assimilation impact on the baseflow (lower part of figure 6.6) a different performance is observed. The filter performs negatively according to the RMSE value ($3.61 \cdot 10^{-5}$ mm·s^{-1} for the SPF filter) when compared to the model run without assimilation ($6.56 \cdot 10^{-6}$ mm·s^{-1}).

The behavior of the baseflow during the assimilation of soil moisture is the result of an inconsistent combination between resampled states and perturbed parameters mainly in the bottom soil layers which contribute to the generation of the baseflow. The replication of those state particles with higher weight in combination with the parameter values affect the baseflow behavior negatively. More specifically, the rearrangement of the soil mois-

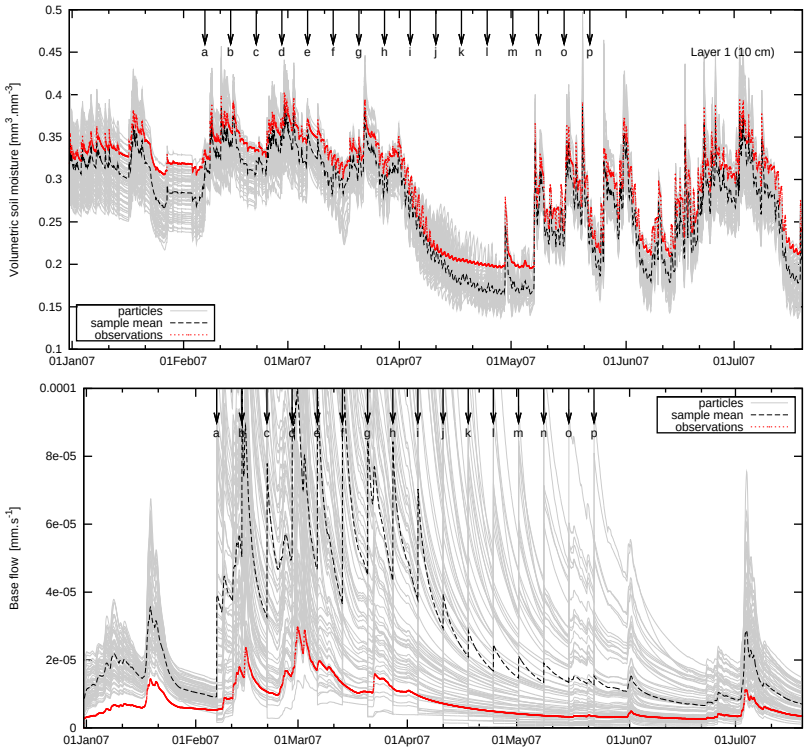


Figure 6.6: SPF filter performance and assimilation impact on the baseflow. The soil moisture and baseflow time series correspond to the DA study performed with set 2 as the initial parameter set.

ture particles in combination with parameter values for wet or dry moisture conditions generates huge baseflow peaks as can be seen in the lower part of figure 6.6.

In order to assign to each resampled state particle a consistent parameter value, the application of the parameter resampling is important and it is evaluated as an alternative to improve the filter performance and to have a positive impact on the baseflow.

6.7.2 SPF-PR performance

SPF with parameter resampling aims at a combination of estimated state values with consistent parameter values. This procedure should result in a positive impact on the land surface variables that dynamically depend (through the model, including the parameter configuration) on the assimilated soil moisture state variable. The parameters involved in the resampling step are listed in Table 1, parameters NwRb and NwRs are not considered

for resampling.

Figure 6.7 shows the performance of the soil moisture assimilation and the impact of the assimilation on the baseflow for the SPF-PR filter without the perturbation of the resampled parameters. Looking at figures 6.5 and 6.7, the decrease in the dispersion of the soil moisture and baseflow particles is noticeable when the parameter resampling is performed. This reduction is indicated by the time-averaged ensemble spread $< ensp >$ (Eq. 6.37), calculated over the entire validation period with inclusion of the DA time steps, with values of $6.16{\cdot}10^{-4}$ and $5.92{\cdot}10^{-5}$ $(\mathrm{mm}^3{\cdot}\mathrm{mm}^{-3})^2$ for soil moisture and values of $1.34{\cdot}10^{-11}$ and $9.27{\cdot}10^{-12}$ $(\mathrm{mm}{\cdot}\mathrm{s}^{-1})^2$ for the baseflow.

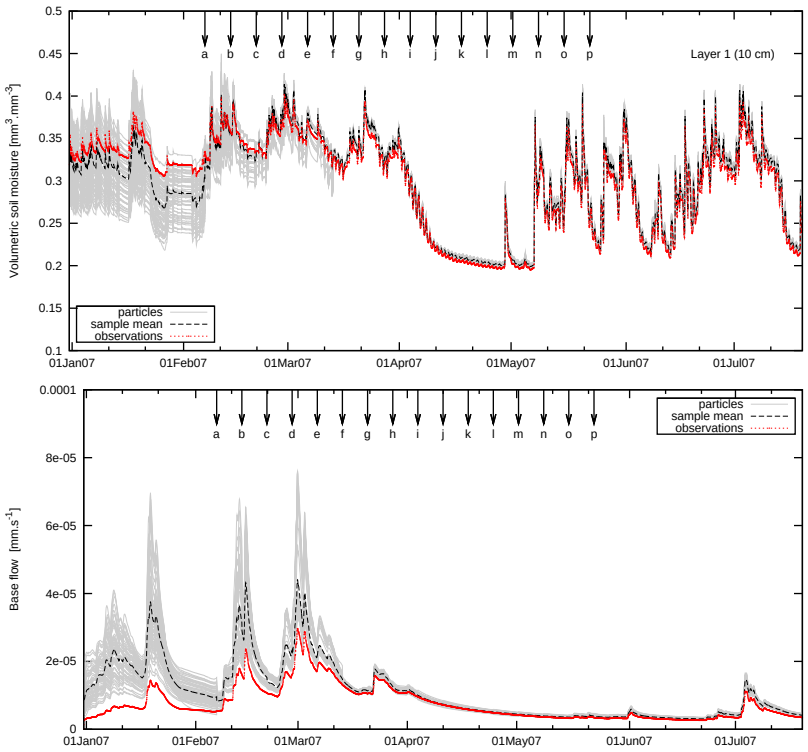


Figure 6.7: SPF-PR filter performance and assimilation impact on the baseflow without parameter perturbation. The soil moisture and baseflow time series correspond to the DA study performed with set 2 as the initial parameter set.

Resampling the parameters along with the state, the filter causes a reduction of the analysis error (the ensemble spread represents the uncertainty at the analysis step). An extreme reduction of the ensemble spread due to an extreme state and parameter particles replication needs to be avoided. Here, we propose the perturbation of the resampled parameters by using additive

white Gaussian noise as the solution to the particles collapse problem. The predefined standard deviation of the noise is set to a fraction of the optimal parameter values, for the results presented in figure 6.8 the fraction is set to 0.01 of parameter set 2. This fraction was obtained based on a proper representation of the baseflow ensemble through the calibration of the ensemble spread measure.

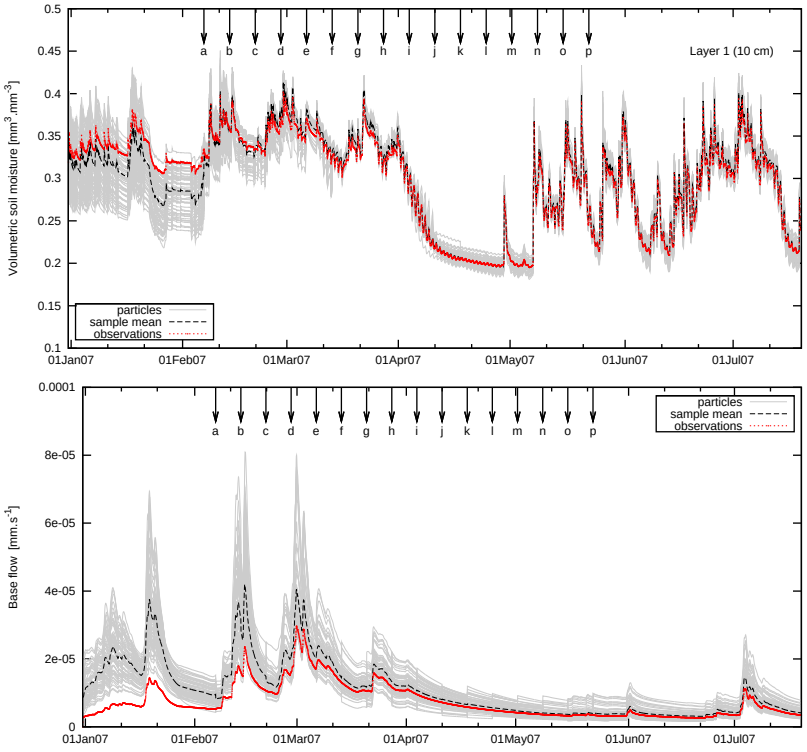


Figure 6.8: SPF-PR filter performance and assimilation impact on the baseflow with parameter perturbation. The soil moisture and baseflow time series correspond to the DA study performed with set 2 as the initial parameter set.

Figure 6.8 shows the SPF-PR filter performance with the perturbation of the resampled parameters. The upper part of this figure presents the performance for the soil moisture assimilation. The dynamics of the state ensemble is positively affected by the parameter resampling improving the overall performance of the filter and keeping the benefit of the state updating for a long time after the DA events. The benefit is quantified by the RMSE values corresponding to 3.65 vol% without assimilation and 0.51 vol% when the SPF-PR is applied. Moreover, the perturbation of the resampled parameters increases the ensemble spread from $5.91{\cdot}10^{-5}$ $(\mathrm{mm}^3{\cdot}\mathrm{mm}^{-3})^2$ to $7.97{\cdot}10^{-5}$ $(\mathrm{mm}^3{\cdot}\mathrm{mm}^{-3})^2$.

Additionally, the plot of the baseflow (see lower part of figure 6.8) shows graphically a considerable improvement on the behavior when comparing to the assimilation effects of the SPF filter application. Figure 6.9 shows the impact of the assimilation on the bottom soil layers. The benefit of the parameter resampling in the top soil layer is propagated trough the bottom soil layers. As a consequence, the impact on the baseflow is positively affected. This improvement can be corroborated with the reduction in the RMSE values from $6.56 \cdot 10^{-6}$ mm·s^{-1} when no assimilation is performed to $3.40 \cdot 10^{-6}$ mm·s^{-1} when soil moisture DA is performed. The baseflow ensemble spread can be increased by the parameter perturbation. The ensemble spread values indicate an increase from $9.26 \cdot 10^{-12}$ (mm·s^{-1})^2 to $1.51 \cdot 10^{-11}$ (mm·s^{-1})^2.

Figure 6.10 shows the evolution in time of the soil hydraulic model parameters for the top soil moisture layer. Parameters $bsw_1, watsat_1, hksat_1$ converge to the "truth" (parameter values used in the generation of the synthetic observations) and parameter $sucsat_1$ converges to a different value. The correction of the soil moisture in the bottom layers and of the parameters during the first DA time steps allows for a correction of the baseflow even in the case when the baseflow measurements are not covered by the baseflow ensemble as can be seen in figures 6.7 and 6.8.

Figure 6.11 shows the modeled ensemble discharge with and without the application of data assimilation. The top panel corresponds to the discharge time series in open loop. While in the bottom panel the corrected discharge is showed when SPF-PR is applied. The comparison of both panels by visual inspection indicates a correction of the total discharge. The results demonstrate that the baseflow and the surface runoff are corrected. In other words, the positive impact on the discharge is due to the overall correction of the modeled flows through the resampling of the model parameters.

An overall conclusion based on the good RMSE values obtained for soil moisture and baseflow is that the addition of the parameter resampling to the SPF filter is effective in removing the bias through an indirect calibration of the modeled particles.

6.7.3 Sensitivity study

The performance of the SPF and SPF-PR filters with parameter perturbation are further analyzed for 3 different initial parameter sets, each identified by the automatic calibration algorithm with a similar optimization index value. The filter performance is analyzed through the comparison of the RMSE values. The parameter sets 1,2 and 3 represent 3 different local minima in the parameter space, the idea behind this is to check the robustness

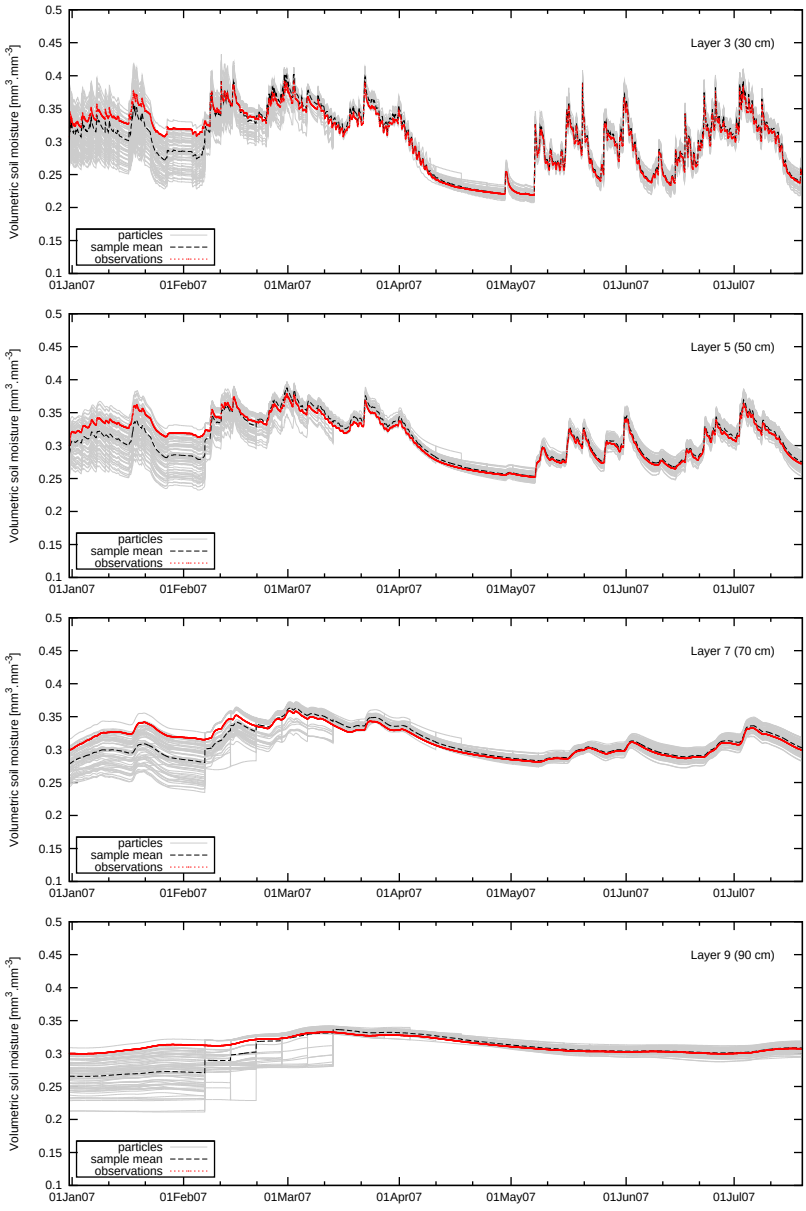


Figure 6.9: Impact of the assimilation of soil moisture in the top layer on the bottom layers when the SPF-PR filter is applied with parameter perturbation.

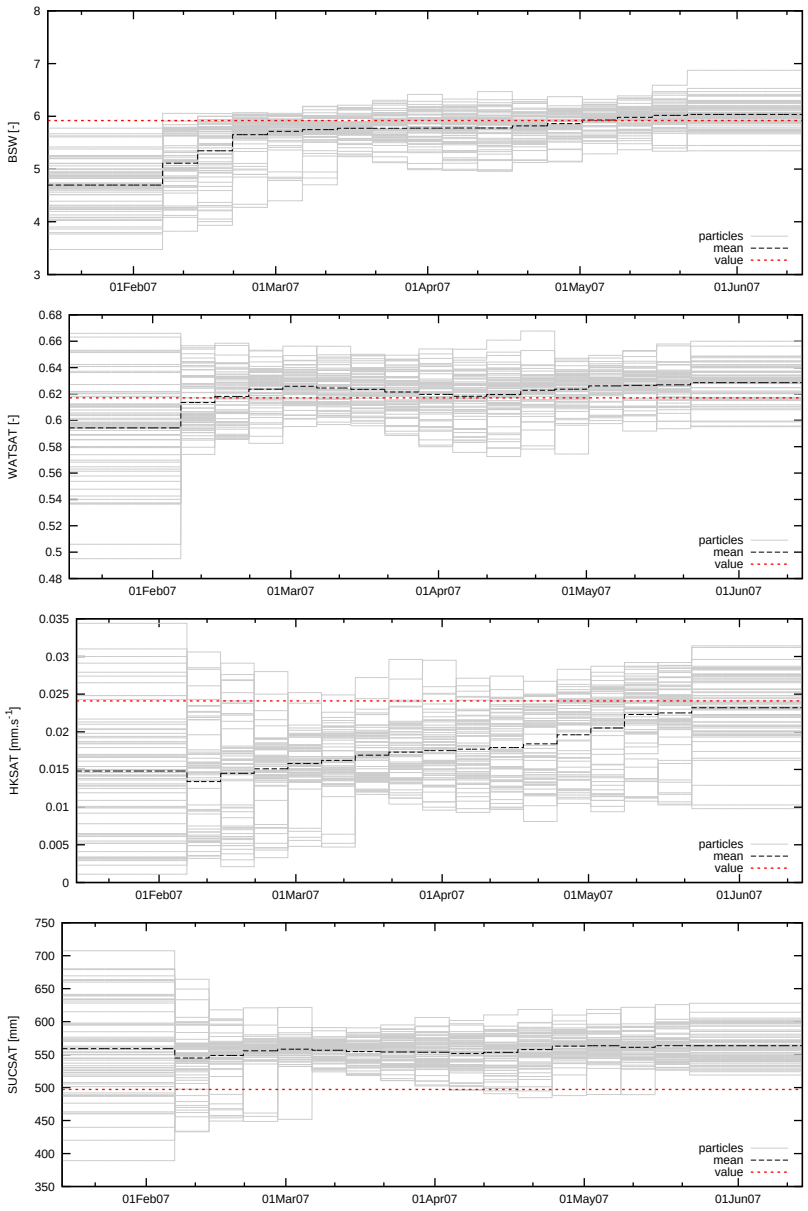


Figure 6.10: Evolution in time of the soil parameters in the top layer when SPF-PR filter is applied with parameter perturbation.

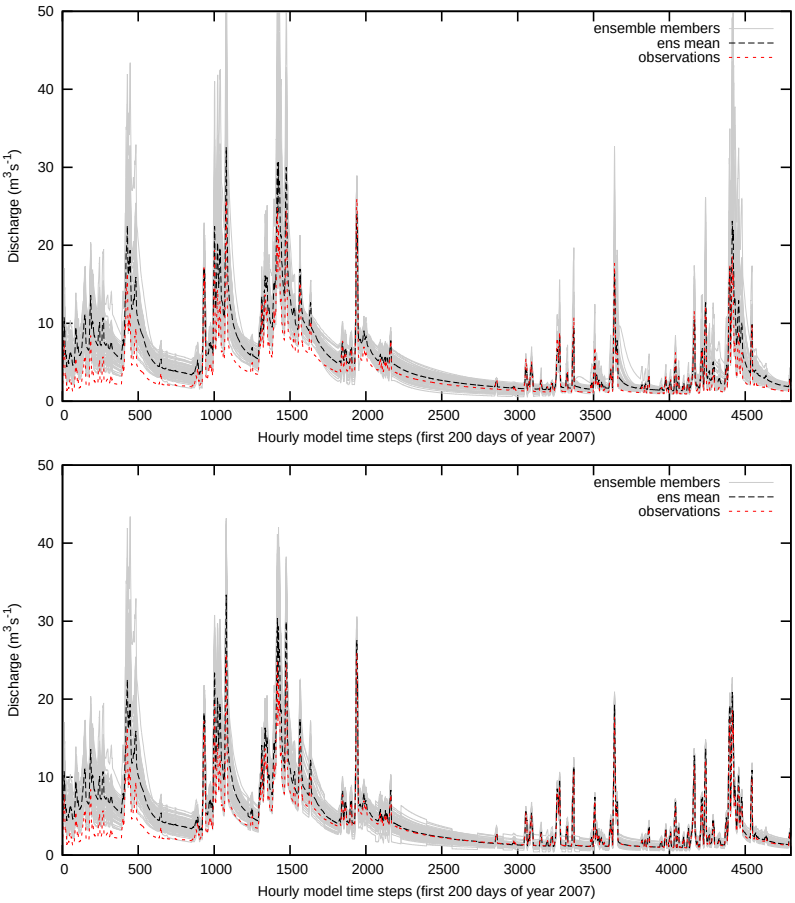


Figure 6.11: Comparison of the discharge without and with DA. Top panel shows the discharge without the application of data assimilation methods. Bottom panel corresponds to SPF-PR filter assimilation impact on the discharge with parameter perturbation. The discharge time series correspond to the DA study performed with set 2 as the initial parameter set.

Table 6.2: RMSE [$mm^3 \cdot mm^{-3}$] between the observed and simulated soil moisture for 3 parameter sets.

Filter	**set 1**	**set 2**	**set 3**
Ensemble	2.89	3.65	3.38
SPF	2.33	2.19	1.61
SPF-PR	1.85	0.51	0.74

Table 6.3: RMSE [$mm \cdot s^{-1}$] between the observed and simulated baseflow for 3 parameter sets. SPF-PR is applied with parameter perturbation.

Filter	**set 1**	**set 2**	**set 3**
Ensemble	$1.39 \cdot 10^{-5}$	$6.56 \cdot 10^{-6}$	$5.89 \cdot 10^{-6}$
SPF	$2.71 \cdot 10^{-5}$	$3.61 \cdot 10^{-5}$	$4.08 \cdot 10^{-5}$
SPF-PR	$1.16 \cdot 10^{-5}$	$3.40 \cdot 10^{-6}$	$2.39 \cdot 10^{-6}$

of the parameter resampling algorithm.

Table 6.2 presents the RMSE values between the estimated and observed volumetric soil moisture at the surface for every filter and for every parameter set. Although the SPF-PR RMSE values are different, due to different system dynamics the SPF-PR filter outperforms the rest of the filters indicating robustness of the algorithm. Additionally, according to table 6.3 the positive impact on the baseflow persists among the three cases.

Considering the assimilation of remote sensed soil moisture data, the availability of data is of main importance in the application of the assimilation algorithm. Therefore, the SPF-PR performance is tested for 3 DA frequencies. Additionally to the DA frequency corresponding to 16 DA events, the methodology is evaluated for 8 DA events with 1 event every 2 weeks and 4 DA steps with 1 event every four weeks.

Table 6.4 and table 6.5 show the RMSE values for the 3 DA frequencies for soil moisture and baseflow respectively. The values indicate a notorious improvement when using the SPF-PR and the positive impact on the baseflow is maintained for the 3 DA frequencies. An additional sensitivity test is recommended concerning the impact of the noise level and the magnitude of the truth parameter set on the behavior of the baseflow since different performances have been noted when using different parameter values. In table 6.5, the RMSE values corresponding to the performance of the SPF filter for the 3 DA frequencies indicate the degeneracy of the baseflow due to an inconsistent combination of states and parameters.

Table 6.4: RMSE [$mm^3 \cdot mm^{-3}$] between the observed and simulated soil moisture for 3 DA frequencies.

Filter	**DA Freq1**[a]	**DA Freq2**[b]	**DA Freq3**[c]
Ensemble	3.65	3.65	3.65
SPF	2.19	2.34	2.60
SPF-PR	0.51	0.76	1.22

[a] - Freq1: indicates DA every week. [b] - Freq2: indicates DA every 2 weeks. [c] - Freq3: indicates DA every 4 weeks.

Table 6.5: RMSE [$mm \cdot s^{-1}$] between the observed and simulated baseflow for 3 DA frequencies. SPF-PR is applied with parameter perturbation.

Filter	**DA Freq1**[*]	**DA Freq2**[*]	**DA Freq3**[*]
Ensemble	$6.56 \cdot 10^{-6}$	$6.56 \cdot 10^{-6}$	$6.56 \cdot 10^{-6}$
SPF	$3.61 \cdot 10^{-5}$	$3.26 \cdot 10^{-5}$	$2.51 \cdot 10^{-5}$
SPF-PR	$3.40 \cdot 10^{-6}$	$4.23 \cdot 10^{-6}$	$4.37 \cdot 10^{-6}$

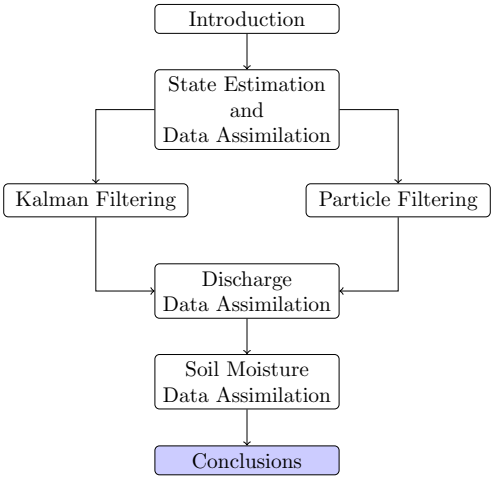
[*] - Same as in table 6.4.

6.8 Summary

Both, the EnKF and the SPF filters perform a decent estimation of soil moisture. However, the predicted baseflow is affected. Therefore, we proposed the correction of the model parameters to tackle this limitation. Overall, the results indicated that the particle filter is a promising tool for hydrologic modeling purposes in distributed applications, but an additional parameter resampling may be necessary to consistently update all state variables and fluxes within the model.

Chapter 7

Conclusions and future directions



This dissertation deals with the application of particle filters to hydrologic systems either for improving flood forecasting systems or for the assimilation of sparse soil moisture observations into a distributed hydrologic model. The summary and the corresponding conclusions are presented in this section along with the possible directions for further research.

7.1 Assimilation of discharge data

7.1.1 Conclusions

In this work two alternatives to improve the performance of the particle filter have been considered. The first approach consists of implementing a resample move step in the standard particle filter structure, while the second approach consists of the combination of two nonlinear/Gaussian filters which are the ensemble Kalman filter and the Gaussian particle filter. The performances of the ensemble Kalman filter and the particle filters are assessed through experiments with synthetic discharge observations and in-situ discharge data.

In the synthetic experiment, the errors assumed in the control set-up allow for an evaluation of the data assimilation methods in a non-Gaussian scenario or close to this scenario. In non-Gaussian scenario, the standard particle filter should outperform the ensemble Kalman filter. The results showed an improvement of the SPF over the EnKF. However, the obtained improvement is relatively small. The reason of the slight improvement in the synthetic study is attributed to the recombination of the model states and model parameters performed by the resampling step. This finding leads to the recommendation that state-parameter estimation needs to be considered in further studies. The results obtained from the experiment with real data and concerning the performances of the EnKF and the SPF indicates an outperformance of the SPF.

In general, the standard particle filter with resample-move step shows a consistent performance. SPF-RM outperforms the standard implementation of the particle filter by dispersing the particle set after the resampling step. The additional RM step increases the computation time since extra particles are obtained from a second run of the rainfall-runoff model.

The variant of the Gaussian particle filter, EnGPF outperformed the EnKF and the SPF in general, but EnGPF performed slightly better than the particle filter with resample-move steps in the real experiment. The good results corresponding to the EnGPF performance are attributed to the use of a better importance density function compared to the SPF and its variant. Additionally, the importance sampling step in the EnGPF does not involve resampling but the sampling of Gaussian distributed particles. The latter could lead to a divergence of the filter performance when the real posterior density distribution is different from Gaussian. However, the results of this study show that the EnGPF is able to deal with non-Gaussian error structure. The model used in this study corresponds to a parsimonious rainfall-runoff model and the concentration time (time needed for water to flow from the most remote point to the outlet) of the study area is smaller

than the model time step allowing for a simplification in the computation of the output flow. Further research is needed to extend the potential use of the EnGPF methodology to complex hydrologic models. In this context, the absence of the resampling step in the EnGPF methodology allows for a straightforward parallel implementation of the algorithm which can be useful in the application to spatially distributed hydrologic models.

Finally, the dynamic adjustment of the noise levels based on the accuracy of the mean prediction relative to the ensemble spread demonstrated the increase in the effectiveness of data assimilation methods. In this study, the initial ensemble spread before assimilation was optimized by the identification of the noise levels in order to assure enough ensemble spread as to cover the observations during the entire simulation period.

7.1.2 Future research

The ensemble Gaussian particle filter is a promising tool for improving flood forecasting systems. Additional research is needed to extend the application of this filter to complex distributed hydrologic systems. However, the filter is prone to suffer from divergence in extreme non-Gaussian conditions. In order to generalize the application of the filter to all kind of non-Gaussian scenarios, the Gaussian sum approach can be adopted as a methodology to increase the robustness of the EnGPF.

7.2 Assimilation of soil moisture data

7.2.1 Conclusions

The standard particle filter with parameter resampling has been evaluated for the performance in assimilation of soil moisture and the impact thereof on baseflow fluxes. The filter performs relatively good for the correction of the modeled soil moisture, although it should be noted the presence of bias. The impact of the soil moisture assimilation on the baseflow results indicates a strong negative effect. The SPF-PR approach is presented as a solution to this shortcoming in the SPF performance.

The SPF-PR methodology strives on the correction of the consistency between parameters and soil moisture states replicating the consistent parameters and rejecting the erratic parameter values. Results indicate a notorious improvement of the performance not only in the estimation of the soil moisture but also in the influence on the baseflow.

Yet, a severe replication affects the parameter diversity and leads to an improper representation of the posterior pdf when assimilating data. The perturbation of the resampled parameter set by a white Gaussian noise with zero mean and predefined standard deviation mitigates the side-effects of the replication.

The robustness of the SPF-PR has been tested through the evaluation of the SPF-PR filter for different parameter sets and different assimilation frequencies. An overall conclusion is that the addition of parameter resampling is effective in removing the bias.

7.2.2 Future research

The application of the particle filter in this work corresponds to a proof of concept study. Therefore, further research should involve the application of the methodology presented in this manuscript to a real-life scenario with in-situ or remote sensing based soil moisture observations.

Bayesian state inference

The aim of this section is to present a detailed formulation of the equations presented in Chapter 2.

In Chapter 2, the state transition distribution and the measurement distribution are defined. Here, these distributions are generalized in order to consider the complete sequence of the state and measurements up to time t; that is:

$$p(\mathbf{x}_{1:t}) = p_1(\mathbf{x}_1) \prod_{k=2}^{t} f(\mathbf{x}_k|\mathbf{x}_{k-1}) \tag{1}$$

and

$$p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t}) = \prod_{k=1}^{t} h(\mathbf{y}_k|\mathbf{x}_k) \tag{2}$$

Based on the definition of the former distributions, the goal of state inference is to find the smoothing distribution which can be written as a posterior distribution as follows:

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t})}{p(\mathbf{y}_{1:t})} \tag{3}$$

where $p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t})$ is the unnormalized posterior distribution. According to (3) the unnormalized posterior can be written as:

$$p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})p(\mathbf{y}_{1:t}) \tag{4}$$

by applying the Bayes's rule to the posterior distribution in (4), we obtain:

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t})p(\mathbf{x}_{1:t})}{p(\mathbf{y}_{1:t})} \tag{5}$$

If we insert (5) into (4), we obtain an expression for the unnormalized posterior which depends on the distributions presented in (1) and (2); that is:

$$p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = p(\mathbf{y}_{1:t}|\mathbf{x}_{1:t})p(\mathbf{x}_{1:t}) \tag{6}$$

The normalizing constant in (3) is given by:

$$p(\mathbf{y}_{1:t}) = \int p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) d\mathbf{x}_{1:t} \tag{7}$$

A recursive expression of the unnormalized distribution (4) is given by:

$$p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}, \mathbf{x}_t, \mathbf{y}_t) \tag{8}$$

Considering independent events, (8) can be written as:

$$p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_t, \mathbf{y}_t) \tag{9}$$

where according to (6):

$$p(\mathbf{x}_t, \mathbf{y}_t) = p(\mathbf{y}_t|\mathbf{x}_t) p(\mathbf{x}_t) \tag{10}$$

By inserting (10) into (9), the posterior can be written as:

$$p(\mathbf{x}_{1:t}, \mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t-1}, \mathbf{y}_{1:t-1}) h(\mathbf{y}_t|\mathbf{x}_t) f(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{11}$$

The joint distributions in (11) can be written as conditional distributions according to (4)

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) p(\mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}) p(\mathbf{y}_{1:t-1}) h(\mathbf{y}_t|\mathbf{x}_t) f(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{12}$$

In (12), the distributions $p(\mathbf{y}_{1:n})$, $p(\mathbf{y}_{1:n-1})$ can be written as:

$$\frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t, \mathbf{y}_{1:t-1})} = \frac{p(\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1}) p(\mathbf{y}_{1:t-1})} = \frac{1}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \tag{13}$$

Finally, by replacing (13) in (12), the smoothing distribution is given by:

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t}) = p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}) \frac{h(\mathbf{y}_t|\mathbf{x}_t) f(\mathbf{x}_t|\mathbf{x}_{t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})} \tag{14}$$

where

$$p(\mathbf{y}_n|\mathbf{y}_{1:n-1}) = \int p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}) h(\mathbf{y}_t|\mathbf{x}_t) f(\mathbf{x}_t|\mathbf{x}_{t-1}) d\mathbf{x}_{t-1:t} \tag{15}$$

The prediction step can be written from (14) as follows:

$$p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t-1}) = p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}) f(\mathbf{x}_t|\mathbf{x}_{t-1}) \tag{16}$$

The marginal distribution (filtering distribution) is obtained from (16) by integrating out (marginalizing) $\mathbf{x}_{1:t-1}$

$$\begin{aligned} p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \int p(\mathbf{x}_{1:t-1}, \mathbf{x}_t|\mathbf{y}_{1:t-1}) d\mathbf{x}_{1:t-1} \\ p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) &= \int p(\mathbf{x}_{1:t-1}|\mathbf{y}_{1:t-1}) p(\mathbf{x}_t|\mathbf{x}_{1:t-1}) d\mathbf{x}_{1:t-1} \end{aligned} \tag{17}$$

In (17), the Markov property is applied to the dynamics, as a result the equations of the Bayes's filter are obtained. The prediction step:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int f(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \tag{18}$$

and the correction step:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{h(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{\int h(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})d\mathbf{x}_t} \tag{19}$$

Assimilation of soil moisture data with the EnKF

The EnKF has been widely used and accepted as a sequential data assimilation method in Geosciences. Thus, the results from the EnKF are presented in this appendix.

Data assimilation is performed every week with the first DA event at 08-February-2007 and the last at 24-May-2007. All the 16 DA events are indicated by black arrows in the figures and the simulation period corresponds to the first half of year 2007 (01/January-01/July). In order to quantify the filter performances, the Root Mean Square Error (RMSE) is computed around the time period corresponding to 1 day before the first DA event and 1 day after the last DA event.

The upper part of figure 1 shows the evolution over time of the 64 soil moisture ensemble members in light gray color, the ensemble mean in a black dashed line corresponding to the EnKF performance and the synthetic observations in red dotted line. From the figure, it is noticeable the correction of the soil moisture values at every DA assimilation event. Moreover, the corresponding correction is not persistent since after some simulation time steps the effect of the assimilation disappears and the RMSE between the estimated and the observed soil moisture with a value of 2.50 vol%, indicates a low improvement when contrasting against the RMSE between the ensemble mean and the observations 3.07 vol%.

The influence of the soil moisture assimilation on the baseflow can be seen in the lower part of figure 1, every arrow indicates the corresponding soil moisture assimilation event and at first glance it is noticeable the presence of small peaks in the behavior of the baseflow as a consequence of the assimilation. The RMSE values of $4.786 \cdot 10^{-6}$ [mm·s^{-1}] for the baseflow ensemble mean and $6.908 \cdot 10^{-6}$ [mm·s^{-1}] for the affected baseflow corroborates the negative impact on the baseflow after the assimilation.

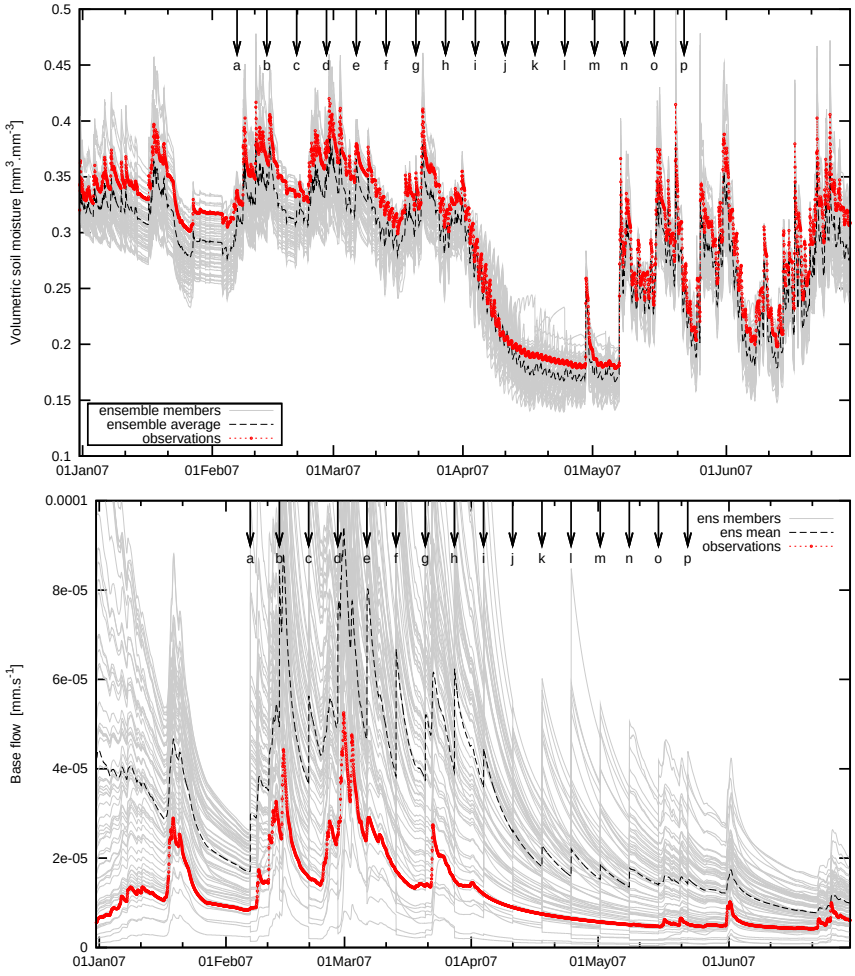


Figure 1: Kalman filter performance and assimilation impact on the base-flow.

Bibliography

Anderson, J. (1996). A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9(7):1518–1530.

Anderson, J. (2001). An ensemble adjustment Kalman filter for data assimilation. *Monthly Weather Review*, 129(12):2884–2903.

Andersson, L. (1992). Improvement of runoff models, what way to go? *Nordic Hydroly*, 23(5):315–332.

Andreas, S., Karlsen, H. A., Naevdal, G., Skaug, H. J., and Valles, B. (2011). Bridging the ensemble Kalman filter and particle filters: the adaptive Gaussian mixture filter. *Computational Geosciences*, 15(2):293–305.

Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo methods. *Journal of Royal Statistic Society B*, 72(3):269–342.

Arulampalam, M. S., Maskell, S., and Gordon, N. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50:174–188.

Bellantoni, J. and Dodge, K. (1967). A square root formulation of the kalman-schmidt filter. *AIAA Journal*, 5:1309 – 1314.

Bergström, S. (1976). Development and Application of a Conceptual Runoff Model for Scandinavian Catchments. Report RHO 7. Swedish Meteorological and Hydrological Institute, Norrköping, Sweden.

Berzuini, C., Best, N. G., Gilks, W. R., and Larizza, C. (1997). Dynamic conditional independence models and Markov Chain Monte Carlo methods. *Journal of the American Statistical Association*, 92:1403–1412.

Bishop, C. H., Etherton, B., and Majumdar, S. (2001). Adaptive sampling with the ensemble transform Kalman filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129(3):420–436.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Naeimi, V., Bartalis, Z., and Hasenauer, S. (2010). Improving runoff prediction through the assimilation of the ascat soil moisture product. *Hydroly and Earth System Sciences*, 14(10):1881–1893.

Burgers, G., van Leeuwen, P. J., and Evensen, G. (1998). On the Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, 126:1719–1724.

Caparrini, F., Castelli, F., and Entekhabi, D. (2004). Variational estimation of soil and vegetation turbulent transfer and heat flux parameters from sequences of multisensor imagery. *Water Resources Research*, 40.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, 87(418):493–500.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999). An Improved Particle Filter for Non-linear Problems. In *IEE Proceedings - Radar, Sonar and Navigation*, volume 146(1), pages 2–7.

Crassidis, J. (2006). Sigma-point kalman filtering for integrated gps and inertial navigation. *Aerospace and Electronic Systems, IEEE Transactions on*, 42(2):750–756.

Crisan, D. (2000). Particle filters – a theoretical perspective. In *Sequential Monte Carlo Methods in Practice*. Springer – Verlag.

De Lannoy, G. J. M. (2006b). *Assimilation of soil moisture observations into a spatially distributed hydrologic model*. PhD thesis, Ghent University.

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., and Verhoest, N. E. C. (2006a). Assessment of model uncertainty for soil moisture through ensemble verification. *Journal of Geophysical Research*, 111(D10).

De Lannoy, G. J. M., Houser, P. R., Pauwels, V. R. N., and Verhoest, N. E. C. (2007a). State and bias estimation for soil moisture profiles by an ensemble Kalman filter: Effect of assimilation depth and frequency. *Water Resources Research*, 43(6).

De Lannoy, G. J. M., Reichle, R. H., Houser, P. R., Pauwels, V. R. N., and Verhoest, N. E. C. (2007b). Correcting for forecast bias in soil moisture assimilation with the ensemble kalman filter. *Water Resources Research*, 43(9).

DeChant, C. M. and Moradkhani, H. (2012). Examining the effectiveness and robustness of sequential data assimilation methods for quantification of uncertainty in hydrologic forecasting. *Water Resources Research*, 48.

Dooge, J. C. I., Costin, A. B., and Finkel, H. J. (1973). Man's influence on the hydrological cycle. In *FAO irrigation and drainage paper*, number 17. Food and Agriculture Organization (FAO), Rome, Italy.

Douc, R., Cappe, O., and Moulines, E. (2005). Comparison of resampling schemes for particle filtering. In *ISPA 2005: Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64–69.

Doucet, A. (1998). On sequential simulation-based methods for Bayesian filtering. Technical Report CUED/F-INFENG/TR. 310, Cambridge University Department of Engineering.

Doucet, A., Godsill, S., and Andrieu, C. (2000). On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208.

Doucet, A., Gordon, N., and Krishnamurthy, V. (2001). Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624.

Doucet, A. and Johansen, A. M. (2009). A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. In *Handbook of Nonlinear Filtering*. Oxford University Press.

Duan, Q. (2003). Global optimization for watershed model calibration. In *Calibration of Watershed Models, Water Sci. Appl.*, volume 6, pages 89–104. AGU, Washington, DC.

Duan, Q., Sorooshian, S., and Gupta, V. K. (1994). Optimal use of the sceua global optimization method for calibrating watershed models. *Journal of Hydrology*, 158(34):265 – 284.

Duan, Q. Y., Gupta, V. K., and Sorooshian, S. (1993a). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

Duan, Q. Y., Gupta, V. K., and Sorooshian, S. (1993b). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, 76(3):501–521.

Dunne, S. and Entekhabi, D. (2005). An ensemble-based reanalysis approach to land data assimilation. *Water Resources Research*, 41.

Entekhabi, D., Nakamura, H., and Njoku, E. G. (1994). Solving the inverse-problem for soil moisture and temperature profiles by sequential assimilation of multifrequency remotely sensed observations. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2):438–448.

Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, 99(C5):10143–10162.

Fearnhead, P. (2002). Markov chain Monte Carlo, sufficient statistics, and particle filters. *Journal of Computational and Graphical Statistics*, 11(4):848–862.

Franssen, H. J. H. and Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, 44.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica*, 57(6):1317–1339.

Gilks, W. R. and Berzuini, C. (2001). Following a moving target-Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society*, 63(1):pp. 127–146.

Giustarini, L., Matgen, P., Hostache, R., Montanari, M., Plaza, D., Pauwels, V. R. N., De Lannoy, G. J. M., De Keyser, R., Pfister, L., Hoffmann, L., and Savenije, H. H. G. (2011). Assimilating SAR-derived water level data into a hydraulic model: a case study. *Hydroly and Earth Systems Sciences*, 15(7):2349–2365.

Godsill, S. J., Doucet, A., and West, M. (2004). Monte Carlo smoothing for non-linear time series. *Journal of The American Statistical Association*, 99:156–168.

Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107 –113.

Grewal, M. S. and Andrews, A. P. (2001). *Kalman Filtering : Theory and Practice Using MATLAB*. Wiley-Interscience, 2 edition.

Heathman, G. C., Starks, P. J., Ahuja, L. R., and Jackson, T. J. (2003). Assimilation of surface soil moisture to estimate profile soil water content. *Journal of Hydrology*, 279:1–17.

Higuchi, T. (1997). Monte Carlo filter using the genetic algorithm operators. *Journal of Statistical Computation and Simulation*, 59(1).

Hoeben, R. and Troch, P. A. (2000). Assimilation of active microwave observation data for soil moisture profile estimation. *Water Resources Research*, 36(10):2805–2819.

Hoteit, I., Luo, X., and Pham, D.-T. (2012). Particle Kalman Filtering: A Nonlinear Bayesian Framework for Ensemble Kalman Filters. *Monthly Weather Review*, 140(2):528–542.

Hoteit, I., Pham, D. T., Triantafyllou, G., and Korres, G. (2008). A new approximate solution of the optimal nonlinear filter for data assimilation in meteorology and oceanography. *Monthly Weather Review*, 136(1):317–334.

Houser, P. R., Shuttleworth, W. J., Gupta, H. V., Famiglietti, J. S., Syed, K. H., and Goodrich, D. C. (1998). Integration of soil moisture remote sensing and hydrologic modeling using data assimilation. *Water Resources Research*, 34(12):3405–3420.

Houtekamer, P. and Mitchell, H. (2001). A sequential ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137.

Ito, K. and Xiong, K. (2000). Gaussian filters for nonlinear filtering problems. *Automatic Control, IEEE Transactions on*, 45(5):910 –927.

Jazwinski, A. H., editor (1970). *Stochastic Processes and Filtering Theory*, volume 64. Elsevier.

Julier, S. (2002). The scaled unscented transformation. In *American Control Conference, 2002. Proceedings of the 2002*, volume 6, pages 4555 – 4559 vol.6.

Julier, S. and Uhlmann, J. (2004). Unscented filtering and nonlinear estimation. *Proceedings of the IEEE*, 92(3):401 – 422.

Julier, S., Uhlmann, J., and Durrant-Whyte, H. (2000). A new method for the nonlinear transformation of means and covariances in filters and estimators. *Automatic Control, IEEE Transactions on*, 45(3):477 – 482.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Eng-T ASME*, 82(Series D):35–45.

Kavetski, D., Kuczera, G., and Franks, S. W. (2006). Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. *Journal of Hydrology*, 320:173–186.

Kitagawa, G. (1996). Monte Carlo filter and smoother for non-Gaussian nonlinear state space models. *Journal of Computational and Graphical Statistics*, 5(1):1–25.

Kong, A., Liu, J. S., and Wong, W. H. (1994). Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288.

Kotecha, J. and Djuric, P. (2003a). Gaussian particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2592–2601.

Kotecha, J. and Djuric, P. (2003b). Gaussian sum particle filtering. *IEEE Transactions on Signal Processing*, 51(10):2602–2612.

Lahoz, W. A., Khattatov, B., and Ménard, R., editors (2010). *Data Assimilation: Making Sense of Observations*. Springer, 1st edition.

Leisenring, M. and Moradkhani, H. (2011). Snow water equivalent prediction using bayesian data assimilation methods. *Stochastic Environmental Research and Risk Assessment*, 25(2):253–270.

Leisenring, M. and Moradkhani, H. (2012). Analyzing the uncertainty of suspended sediment load prediction using sequential data assimilation. *Journal of Hydrology*, 468469(0):268 – 282.

Lindström, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S. (1997). Development and test of the distributed HBV-96 hydrological model. *Journal of Hydrology*, 201(1-4):272 – 288.

Liu, J. S. and Chen, R. (1995). Blind Deconvolution via Sequential Imputations. *Journal of the American Statistical Association*, 90(430):567–576.

Liu, J. S. and Chen, R. (1998). Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044.

Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P. (2012). Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences*, 16(10):3863–3887.

Matgen, P., Henry, J., Hoffmann, L., and Pfister, L. (2006). Assimilation of remotely sensed soil saturation levels in conceptual rainfall-runoff models. *IAHS book, prediction in ungauged basins: Promise and progress*, IAHS Publ. 303.

Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Plaza, D., Pauwels, V. R. N., Lannoy, G. J. M. D., Keyser, R. D., and Savenije, H. H. G. (2010). Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: proof of concept. *Hydrology and Earth System Sciences*, 14:1773–1785.

Montzka, C., Moradkhani, H., Weihermller, L., Franssen, H.-J. H., Canty, M., and Vereecken, H. (2011). Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter. *Journal of Hydrology*, 399:410–421.

Moore, R. J. (2007). The PDM rainfall-runoff model. *Hydrology and Earth System Sciences*, 11(1):483–499.

Moradkhani, H., DeChant, C. M., and Sorooshian, S. (2012). Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-markov chain monte carlo method. *Water Resources Research*, 48.

Moradkhani, H., Hsu, K.-L., Gupta, H., and Sorooshian, S. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. *Water Resources Research*, 41(W05012).

Moradkhani, H. and Sorooshian, S. (2008a). General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis. In Sorooshian, S. et al., editors, *Hydrological Modelling and the Water Cycle*, volume 63 of *Water Science and Technology Library*, pages 1–24. Springer Berlin Heidelberg. 10.1007/978-3-540-77843-1_1.

Moradkhani, H. and Sorooshian, S. (2008b). General Review of Rainfall-Runoff Modeling: Model Calibration, Data Assimilation, and Uncertainty Analysis. In Sorooshian, S and Hsu, KL and Coppola, E and Tomassetti, B and Verdecchia, M and Visconti, G, editor, *Hydrological Modelling and the Water Cycle: Coupling the Atmosheric and Hydrological Models*, volume 63 of *Water Science and Technology Library*, pages 1–24.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005a). Dual state-parameter estimation of hydrological models using ensemble kalman filter. *Advances in Water Resources*, 28(2):135 – 147.

Moral, P. D. (1996). Measure valued processes and interacting particle systems. Application to non linear filtering problems. *Annals of Applied Probability*, 8:438–495.

Musso, C., Oudjane, N., and Le Gland, F. (2001). Improving Regularised Particle Filters. In Doucet, A., Freitas, N. D., and Gordon, N., editors, *Sequential Monte Carlo Methods in Practice*, pages 247–271. Springer.

Nagarajan, K., Judge, J., Graham, W. D., and Monsivais-Huertero, A. (2010). Particle Filter-based assimilation algorithms for improved estimation of root-zone soil moisture under dynamic vegetation conditions. *Advances in Water Resources*, 34(4):433–477.

Nagasaki, M., Yamaguchi, R., Yoshida, R., Imoto, S., Doi, A., Tamada, Y., Matsuno, H., Miyano, S., and Higuchi, T. (2006). Genomic data assimilation for estimating hybrid functional petri net from time-course gene expression data. *Genome Informatics Series*, 17(1):46 – 61.

Nash, J. and Sutcliffe, J. (1970). River flow forecasting through conceptual models part i a discussion of principles. *Journal of Hydrology*, 10(3):282 – 290.

Norgaard, M., Poulsen, N. K., and Ravn, O. (2000). New developments in state estimation for nonlinear systems. *Automatica*, 36(11):1627 – 1638.

Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A. (2010). Ocean data assimilation: a case for ensemble optimal interpolation. *Australian Meteorological And Oceanographic Journal*, 59(SI):67–76. Workshop on Ensemble Prediction and Data Assimilation, Melbourne, AUSTRALIA, 2009.

Oleson, K., Dai, Y., Bonan, G., Bosilovichm, M., Dickinson, R., Dirmeyer, P., Hoffman, F., Houser, P., Levis, S., Niu, G.-Y., Thornton, P., Vertenstein, M., Yang, Z.-L., and Zeng, X. (2004). Technical description of the community land model (clm). Technical report.

Paniconi, C., Maroccu, M., Putti, M., and Verbunt, M. (2002). Newtonian nudging for a richards equation-based distributed hydrological mode. *Advances in Water Resources*, 26:161–178.

Papadakis, N., Memin, E., Cuzol, A., and Gengembre, N. (2010). Data assimilation with the weighted ensemble Kalman filter. *Tellus A-dyn. Meteorol. Oceanog.*, 62(5):673–697.

Parrish, M. A., Moradkhani, H., and DeChant, C. M. (2012). Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resources Research*, 48(W03519).

Pauwels, V. R. N. (2010). Hydrologic modelling. Course notes. Ghent University, Faculty of Bioscience Engineering.

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., and De Troch, F. P. (2001). The importance of the spatial patterns of remotely sensed soil moisture in the improvement of discharge predictions for small-scale basins through data assimilation. *Journal of Hydrology*, 251(1-2):88–102.

Pauwels, V. R. N., Hoeben, R., Verhoest, N. E. C., De Troch, F. P., and Troch, P. A. (2002). Improvement of toplats-based discharge predictions through assimilation of ers-based remotely sensed soil moisture values. *Hydrological Processes*, 16:995–1013.

Pechlivanidis, I. G., Jackson, B. M., Mcintyre, N. R., and Wheater, H. S. (2011). Catchment Hydrological Modelling: A Review of model types, calibration approaches and uncertainty analysis methods in the context of recent developments in technology and applications. *Global Nest Journal*, 13(3):193–214.

Pitt, M. and Shephard, N. (1999). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association*, 94(446):590–599.

Plaza, D. A., De Keyser, R., De Lannoy, G. J. M., Giustarini, L., Matgen, P., and Pauwels, V. R. N. (2012). The importance of parameter resampling for soil moisture data assimilation into hydrologic models using the particle filter. *Hydrology and Earth System Sciences*, 16(2):375–390.

Reichle, R., McLaughlin, D., and Entekhabi, D. (2002). Hydrologic data assimilation with the ensemble Kalman filter. *Monthly Weather Review*, 130(1):103–114.

Rings, J., Vrugt, J. A., Schoups, G., Huisman, J. A., and Vereecken, H. (2012). Bayesian model averaging using particle filtering and Gaussian mixture modeling: Theory, concepts, and simulation experiments. *Water Resources Research*, 48(W05520).

Robert, C. P. and Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag, 1 edition.

Schei, T. S. (1997). A finite-difference method for linearization in nonlinear estimation algorithms. *Automatica*, 33(11):2053–2058.

Seuffert, G., Wilker, H., Viterbo, P., Drusch, M., and Mahfouf, J. F. (2004). The usage of screen-level parameters and microwave brightness temperature for soil moisture analysis. *Journal of Hydrometeorology*, 5:516–531.

Simon, D. (2006). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley-Interscience, 1st edition.

Troch, P. A., Smith, J. A., Wood, E. F., and de Troch, F. P. (1994). Hydrologic controls of large floods in a small basin: central appalachian case study. *Journal of Hydrology*, 156(1-4):285 – 309.

Van Der Merwe, R., de Freitas, N., Doucet, A., and Wan, E. (2001). The unscented particle filter. In *Advances in Neural Information Processing Systems 13*.

Van Der Merwe, R., Doucet, A., De Freitas, N., and Wan, E. (2001). The Unscented Particle Filter. In *Advances in Neural Information Processing Systems*. MIT; 1998.

Vrugt, J., Gupta, H., and Nuallain, B. (2006). Real-time data assimilation for operational ensemble streamflow forecasting. *Journal of Hydrometeorology*, 7(3):548–565.

Vrugt, J. A., ter Braak, C. J., Diks, C. G., and Schoups, G. (2012). Hydrologic Data Assimilation Using Particle Markov Chain Monte Carlo Simulation: Theory, Concepts and Applications. *Advances in Water Resources*, (0).

Šimandl, M., KráLovec, J., and SöDerströM, T. (2006). Advanced point-mass method for nonlinear state estimation. *Automatica*, 42(7):1133–1145.

Walker, J. P., Willgoose, G. R., and Kalma, J. D. (2002). Three-dimensional soil moisture profile retrieval by assimilation of near-surface measurements: Simplified kalman filter covariance forecasting and field application. *Water Resources Research*, 38(12).

Wan, E. A. and Van Der Merwe, R. (2000). The unscented Kalman filter for nonlinear estimation. In *IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium*, pages 153–158.

Wang, D., Chen, Y., and Cai, X. (2009). State and parameter estimation of hydrologic models using the constrained ensemble kalman filter. *Water Resources Research*, 45.

Weerts, A. H. and El Serafy, G. Y. H. (2006). Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, 42(W09403).

Wells, C. (1996). *The Kalman Filter in Finance*. Springer.

Whitaker, J. and Hamill, T. (2002). Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924.

Zhang, J., Wang, W., and Leung, L. (2008). Contribution of land-atmosphere coupling to summer climate variability over the contiguous united states. *Journal of Geophysical Research*, 113(D22109).