PADUA UNIVERSITY
ENVIRONMENTAL ENGINEERING

Reconstructing saturated hydraulic conductivity at the LEO-biosphere hillslope via data assimilation

Master degree thesis

Supervisor

Dott. Ing. Mario Putti

Presented by

Eloisa Torresini

Assistant supervisor

Dott. Damiano Pasetto

Academic Year 2014-2015

A chi mi vuole bene:
a chi c'è ancora,
a chi c'è di nuovo,
e a chi non c'è più, ma ci sarà sempre.

Contents

List of Figures iii
List of Tables v

1 Introduction **1**

2 Literature Review **3**
2.1 Coupled hydrological models 3
2.2 Data assimilation in Hydrology 6

3 Landscape Evolution Observatory **11**
3.1 LEO intent and general features 11
3.2 First experiment at LEO 15
3.3 CATHY application on LEO first experiment 16

4 CATHY model **21**
4.1 Catchment hydrology model 21
4.2 Surface runoff routing 22
4.3 Topographic depressions oversight 24
4.4 Subsurface module 25
4.5 Surface-subsurface coupling 29

5 Data assimilation methods **33**
5.1 Data assimilation 33
5.1.1 Data assimilation - mathematical formulation 34
5.1.2 Ensemble Kalman filter 36
5.1.3 Sequential importance resampling 39
5.1.4 Parameter estimation 40
5.2 Data assimilation on LEO experiment - features 42

6 Estimation of the hydraulic conductivity field from distributed soil water content observations at LEO **45**
6.1 Problem setting 45
6.2 Partially heterogeneous hillslope 49
6.2.1 Configuration with 5 zones 49
6.2.2 Configuration with 10 zones 52
6.3 Fully heterogeneous hillslope 58

7 Conclusions **63**

List of Figures

3.1 One of the three slopes composing LEO facility. 12
3.2 Location of the 496 sensors of soil water content and DEM of the surface of the hillslope [Pasetto et al., 2014]. 12
3.3 Comparison between measured and simulated storage, seepage face flow and overland flow. Results for scenarios M1 and M2 [Niu et al., 2014]. . . . 19
3.4 Comparison between measured and simulated storage, seepage face flow and overland flow. Results for scenarios M3 and M4 with heterogeneous K_{sat} [Niu et al., 2014]. 19

4.1 Surface grid subdivision and coarsening for the subsurface solver 31

5.1 Temporal path of the assimilation process 43

6.1 Non-uniform grid representation 46
6.2 Layers in vertical direction 49
6.3 Horizontal zonation in the case with 5 zones 50
6.4 Parameter estimation for the five zones configuration - SIR algorithm . . . 50
6.5 Hydrologic response and error on volumetric water content with 5 zones configuration - SIR algorithm 51
6.6 Horizontal zonation in the case with 10 zones 52
6.7 Parameters estimated in the ten zone configuration - comparison between EnkF and SIR algorithms 53
6.8 Hydrological response with prior and posterior estimated parameter values and error on volumetric water content - Comparison between EnKF and SIR algorithms 54
6.9 Comparison among the overland flow - perturbed sets 56
6.10 Comparison among the seepage face flow - perturbed sets 56
6.11 Comparison among the water storage - perturbed sets 57
6.12 Comparison of errors on volumetric water content - perturbed sets 57

6.13 Parameter estimation - correlation length = 4 m (EnKF algorithm) 59
6.14 Hydrologic response and error on volumetric water content - correlation length = 4 m (EnKF algorithm) 60
6.15 Parameter estimation - correlation length = 8 m (EnKF algorithm) 61
6.16 Hydrologic response and error on volumetric water content - correlation length = 8 m (EnKF algorithm) 61
6.17 Comparison between the errors on volumetric water content of the different attempts 62

List of Tables

3.1 Soil sample characteristics based on column experiment [Gevaert et al., 2014]. 13
3.2 Atmospheric boundary condition values 18

6.1 Depth of nodes along z direction 47
6.2 General assimilation features 48
6.3 Errors on volumetric water content with prior and posterior distribution - discretized case 55
6.4 Errors on volumetric water content with perturbed parameters 58

Summary

This thesis work aims to reconstruct the spatial distribution of the saturated hydraulic conductivity in one of the artificial hillslopes at the Landscape Evolution Observatory (LEO). This is a fundamental task to better describe and forecast flow dynamics in LEO. The advanced characterization of LEO soil properties is motivated by the unexpected hydrological response of the hillslope during the first rainfall experiment, where the generation of overland flow was not forecast by pre-experiment numerical simulations. A possible explanation for this behavior is the development of localized heterogeneity within the hillslope, as indicated by Niu et al. [2014]. Starting from this idea, here we investigate the feasibility of estimating the saturated hydraulic conductivity of LEO soil from the measurement of volumetric water content, which are spatially distributed. The numerical simulations of the rainfall experiment are fulfilled through the physically-based hydrological model CATHY. Using a probabilistic approach, we estimate the posterior probability distribution of the saturated hydraulic conductivity using two data assimilation schemes, ensemble Kalman filter and sequential importance resampling. In particular, we introduce a new restarting process that ensure the physically consistency of the simulations during the assimilation. Different degrees of spatial heterogeneity are considered to model the spatial distribution of the saturated hydraulic conductivity, from zonation pattern to stochastic random field. Our results show that the assimilation process allows us to slightly reduce the error on volumetric water content measurements for almost all the considered configurations. However, the simulated hydrological response with the posterior distribution of hydraulic conductivity is still far from the observed one. These evidences suggest that the problem is ill-conditioned, affected by the measurements uncertainty.

Sommario

Il lavoro di questa tesi è volto a ricostruire la distribuzione spaziale della conducibilità idraulica di uno dei pendii che costituiscono il LEO. Questo obiettivo è fondamentale per descrivere e prevedere le dinamiche idrauliche del LEO in maniera migliore. La caratterizzazione approfondita delle proprietà fisiche del terreno del LEO è motivata dall'inaspettata risposta idrologica del pendio durante il primo esperimento di pioggia, in cui il deflusso superficiale non era stato previsto nelle simulazioni numeriche realizzate prima dell'esperimento stesso. Una possibile spiegazione, proposta da Niu et al. [2014], è l'insorgere di eterogeneità localizzate all'interno del pendio. Partendo da questi presupposti, un processo di data assimilation è stato realizzato, con lo scopo di stimare la conducibilità idraulica del terreno del LEO, incorporando le misure distribuite del contenuto volumetrico d'acqua (VWC) raccolte durante il primo esperimento. Le simulazioni sono realizzate con il modello idrologico CATHY. Usando un approccio probabilistico, stimiamo la distribuzione di probabilità a posteriori della conducibilità idraulica satura usando due schemi di data assimilation, ensemble Kalman filter e sequential importance resampling. In particolare, introduciamo un nuovo processo iterativo che assicura la consistenza sul piano fisico delle simulazioni durante le assimilazioni. Il pendio è stato modellato con diversi gradi di eterogeneità, da una suddivisione in zone a un campo stocastico casuale. I risultati ci mostrano che il processo di assimilazione ci permette di ridurre leggermente l'errore sulle misure di VWC in tutte le configurazioni prese in esame. La risposta idrologica del terreno, tuttavia, rimane molto diversa da quella realmente osservata. Questi fatti suggeriscono che il problema affrontato sia malcondizionato, affetto dall'incertezza sui dati assimilati. La stima dei parametri potrebbe essere migliorata dall'assimilazione di ulteriori misure di altre grandezze geofisiche. italian

Chapter 1

Introduction

A major issue in the hydrological sciences is the study of the response of a basin to atmospheric forcings. The interactions between surface, subsurface and atmosphere are fundamental in determining the partitioning of rainfall and the redistribution of water in the subsoil domain. These aspects are fundamental for agriculture and climate studies, but also for water quality management and prevention and mitigation of natural hazards [McLaughlin, 2002]. The prediction of hydrologic responses is challenged by the spatial variability and temporal dynamics of the physical and biological processes that control water movement in the landscape. Hydrologic systems are interdependent with biota, soils, geomorphology and micro-climate characterizing the soil surface and subsurface. The hydrologic predictions would be improved by the deeper understanding of these interconnections. The difficulty in precisely defining the boundary conditions and the insufficient measurements hurdle the investigation of these aspects in field experiments [Hopp et al., 2009].

These limitations are addressed at the Landscape Evolution Observatory (LEO) rising at Biosphere 2 facility near Tucson, Arizona. The infrastructure, consisting of three identical slanting slopes, offers the possibility to fulfill hydrological experiments under strictly controlled conditions. Sensors and sampling instruments collect different quantities with a spatial density and temporal frequency impossible to be recreated in natural fields. Heterogeneous and homogeneous rainfall patterns may be applied to the soil via an artificial rainfall system. LEO is designed to make possible the study of the evolution of landscape throughout years of experimentation [Gevaert et al., 2014]. A first experiment has been realised in order to test the sensors, to investigate the hydrological dynamics of the landscape and to generate a steady state of soil moisture for further tracer experiments. Numerical simulations were run prior to the experiment in order to estimate the time needed by the system to reach steady state. Unexpectedly, the actual behaviour of

the system has been very different from the predicted one. Steady state was never reached and a saturation excess was developed. A fundamental assumption in the prior simulations was the homogeneity of the soil, derived from the way in which the soil has been built. Niu et al. [2014] fulfilled a modeling study in order to understand the reasons of the differences between forecasted and real response of the system. The results support the hypothesis of the development of localized heterogeneity, due to downslope compaction and sediment transport.

The first experiment at LEO and the work of Niu et al. [2014] traced the path for this thesis work. The addressed issue is the reconstruction of the values of saturated hydraulic conductivity in LEO hillslope. The soil has been considered heterogeneous. The objective is pursued using the coupled physically-based hydrologic model CATHY and using a data assimilation (DA) process to solve the identification problem. DA methods allow to determine the optimal parameters characterizing the soil, using observations collected in a given time and space domain [Yeh, 1986]. The improvement of the spatial characterization of the saturated hydraulic conductivity is expected to increase the accuracy of the response of the model. We tested both ensemble Kalman filter (EnKF) and sequential importance resampling (SIR) DA approaches. The assimilated observations are the volumetric water content measurements, collected during the first experiment at LEO.

The thesis is structured as follows. Chapter 2 contains a review of the literature concerning coupled hydrologic models and the application of data assimilation and parameter estimation to hydrologic issues. In Chapter 3, the LEO facility is described, along with the planning and realisation of the first experiment and the results of the posterior analysis performed by Niu et al. [2014]. Chapter 4 describes the features of CATHY, while Chapter 5 is devoted to the description of the data assimilation schemes applied in this thesis work. Finally, Chapter 6 collects the different simulations performed in order to characterise the spatial distribution of hydraulic conductivity. Different configurations of heterogeneity has been considered and modeled, and the results are reported and compared.

Chapter 2

Literature Review

Data assimilation methods are often applied to hydrologic problems to help correcting the forecast on the basis of updated state observations. In this thesis work, the process-based coupled hydrological model CATHY and the data assimilation methods EnKF and SIR are used to estimate the spatial distribution of saturated hydraulic conductivity of the LEO hillslope. A brief revision of process-based coupled hydrological models is followed by a short introduction to the main data assimilation methods applied to hydrologic problems. Finally, a review of the data assimilation processes implemented to tackle on hydrological issues concludes this Chapter.

2.1 Coupled hydrological models

Surface and subsurface water dybamics are strictly interdependent, but the significant difference in terms of timescales lead to the common approach of considering them as two separated components. This perspective makes the system easier to be understood and to be described in a mathematical way. The resulting models, however, result incomplete. Their accuracy can be enhanced only considering surface and subsurface as a coupled system [Furman, 2007]. Hereinafter a brief review of process-based coupled hydrological models is reported.

Pohll et al. [1996] proposed a coupled surface-subsurface model of a subsidence crater in Nevada (USA). The topographic attributes are individuated by a terrain analysis module, the overland flow is simulated through a kinematic wave equation and the moisture migration through vadose zone is described by Richards equation. The model for surface flow is linked with a two-dimensional axisymmetric vadose zone model. Ad hoc accurate boundary conditions are assigned to the subsurface module.
In the model presented by Singh and Bhallamudi [1998],devoted to the prediction of over-

land flow, surface and subsurface flow are described by the one-dimensional De Saint Venant equation and the two-dimensional Richards equation. VanderKwaak and Loague [2001] developed a fully-coupled model aimed to the quantitative estimation of water flow and solute transport within a spatially variable porous medium. The subsurface flow is considered in a three-dimensional setting, while the surface flow is estimated by the two-dimensional diffusion wave approximation of the depth-integrated shallow water equation. The links between these components is realised by first order flux exchange driven by pressure head gradients.

The two-dimensional approximation of the diffusion wave model is proposed to describe the surface flow also in Morita and Yen [2002]. The subsurface flow is described by the three-dimensional form of Richards equation. The coupling is realised considering the infiltration between surface and subsurface as common internal boundary condition.

In the model proposed by Downer and Ogden [2003], overland flow and channel flow were distinctly considered, respectively by a two-dimensional and one-dimensional model. Analogously, the water flow in saturated and unsaturated flow was described by a two- and one-dimensional version of Richards equation respectively. Snowfall accumulation and melting could be taken into account. Panday and Huyakorn [2004] distinguish overland flow and channel flow with coupling achieved by first order exchange. This model accounts for urban and agricultural features at various scales, e.g., the effects of depression storage may be included in the equations for overland and channel flows and in the surface-subsurface interactions.

Gunduz and Aral [2005] proposed a coupled model to simulate the interactions along a river bed. A two-dimensional vertically averaged saturated groundwater flow model was coupled with a surface channel flow model based on the dynamic wave form of the De Saint Venant equations. The two equations were solved simultaneously, within the same global matrix structure. The first order coupling is used to link the two models.

A modified Dupuit model is presented in Anderson [2005]to simulate the interaction between stream and river. The groundwater flow is steady, the hydraulic head is constant, the incoming and outgoing fluxes are defined. This model is modified to include a second parameter representing the hydraulic conductivity of the aquifer beneath the stream, considered different from the value of hydraulic conductivity in the rest of the aquifer.

The model described in Graham and Butts [2005] distinguished overland flow, channel flow, unsaturated flow and saturated groundwater flow. The unsaturated flow is considered primarily vertical. Infiltration in the unsaturated zone can be described in four different ways: the full Richards equation, a simplified gravity flow procedure, a two-layer water balance method for root zone and the zone between roots and the water table or

the calculation of net recharge by other means. It is solved in parallel with the saturated groundwater flow. The latter is solved with a three-dimensional finite difference scheme. Overland flow is again described by a finite-difference approximation of De Saint Venant equation. Channel flow is again considered one-dimensional.
Kollet and Maxwell [2006] incorporate a two-dimensional overland flow simulator into a parallel three-dimensional variably saturated subsurface flow code. The coupling in this case is realised imposing the continuity of fluxes and of pressure at the ground surface. The surface equation represents a Neumann boundary condition to Richards equation.
A different approach is explored in Weill et al. [2009], where the computational domain is enriched with a fictitious porous layer at the land surface. In this way a continuum is created from saturated zone to land surface. Hydraulic gradient governs the movement of water, and pressure and velocities are continuous. A unique Richards equation is written to describe both surface and subsurface processes. Also Goderniaux et al. [2009] proposed a fully integrated surface-subsurface flow model for the saturated and partially saturated zone. The subsurface flow equation is the three-dimensional Richards equation, while the surface flow is described by a two-dimensional depth-averaged flow equation. These equations are simultaneously solved through a finite element solver.
The model presented by Shen and Phanikumar [2010] addressed a long-term simulations of medium-large spatial domains. Physically-based conservation laws are the basis of this model. The hydrological domain in which calculations take place is subdivided into different compartments. Particular attention is paid to vadose zone, since it controls evapotranspiration and it links groundwater to surface water. Also in this model the overland flow and the channel flow are distinguished. The two-dimensional and one-dimensional form of diffusion wave equation describe respectively the two flows. Another distinction is done for soil water flow and groundwater flow. The first is assumed to be vertical and it is approximated with then one-dimensional Richards equation. The saturated aquifer is conceptualized as a series of vertical layers. In each layer, a two-dimensional equation describes the groundwater flow.

The chosen coupled physical-based model for the simulations fulfilled in this thesis work is CATHY model [Paniconi and Putti, 1994]. It couples a three-dimensional Richards equation describing subsurface flow with a one-dimensional diffusion wave equation for surface flow. As for the described models, it proposes the finite element/finite difference solvers for subsurface and surface equations, the handling of heterogeneous parameters and inputs, and the possibility to define a variety of boundary conditions over the domain. In addition, the drainage network is identified through a digital elevation model analysis. Storage effects and retardation, due to lakes and topographic depressions, are incorpo-

rated. The channel flow and the overland flow are described by the same equation, with a distinction realised through different sets of parameters. Coupling is realised through the assignment of proper nonlinear boundary conditions. The different dynamics of surface and subsurface flow are taken into account. In fact, a nested time stepping is used for the surface module, and a coarsened grid may be generated for the subsurface domain. In the model soil swelling and shrinking can be taken into account, thanks to the formulation of the subsurface storage coefficient with moisture-dependent porosity. Data assimilation is implemented via Newtonian nudging, ensemble Kalman filter and sequential importance resampling.

2.2 Data assimilation in Hydrology

Data assimilation in hydrology is used to forecast the state of the system and also to estimate the parameters starting from observed data. In most cases, the chosen assimilation method is ensemble Kalman filter, since it easy to be implemented and conceptually simple. Other methods, as particle filters (PF), are increasing in importance, since they are flexible and they are more accurate for strongly nonlinear systems. Hereinafter, a brief review of some experimental data assimilation hydrological works using EnKF and PF are reported.

Walker et al. [2002] applied Kalman filter assimilation technique to estimate the soil moisture profile by means of a distributed three-dimensional soil moisture model. The Kalman filter is modified in order to overcome its computational limitations. The assimilation used real data collected in an Australian catchment. As the results showed, if the forecast of the model is far from the observation, then the assimilation would improve the prediction. On the contrary, if the forecast is yet close to the observed value, the assimilation could induce a degradation in the results. In addition, when the model is well calibrated and the forcing data are accurately reported, the updating time interval is relatively unimportant for the correct retrieval of the soil moisture profile.

In Paniconi et al. [2003] the implementation of Newtonian Nudging in the model CATHY is described. This method consists in adding a forcing term to the model equations in order to force the model variables towards the observations. It is computationally less intensive than other methods, and with simple formulation. The assimilation is tested on a test case. Two scenarios are designed, in the first the atmospheric boundary conditions are perturbed, and in the second the initial conditions are perturbed. All the measurements were generated via CATHY model. The results of these simulations show that nudging is simple, efficient and flexible. It allows also the utilization of informations from

different measurement sources. Limitations are implicit in the definition of the interpolation functions used to spatially distribute the effects of the measurements.
A dual state-parameter estimation approach is presented in Moradkhani et al. [2005]. The authors chose the ensemble Kalman filter method. The estimation of both parameters and state variables was achieved through two interactive filters. The updated parameters, at each step, are used to update the model state. Results confirmed that this method is reliable for streamflow forecasting.
Parameter estimation with particle filters is reported in Moradkhani and Hsu [2005]. Particle filters are more suitable than Kalman-filter based recursive procedures when the considered system is strongly nonlinear. In this case, in fact, the first and second order characteristics are not sufficient to fully describe the evolution of the conditional probability of the prediction. Particle filters on the contrary can handle nonlinearities and give a relatively complete representation of posterior distribution. Parameter estimation is obtained with particle filters both for a synthetic scenario and for a real-world problem. The suitability of particle filters for this issue is verified. It is also evidenced the importance of considering as source of uncertainty not only parameter estimation, but also model structural errors and input measurements errors.
In Weerts and El Serafy [2006], Ensemble Kalman filter and two particle filters, sequential importance resampling (SIR) and residual resampling (RR), are applied to a quasi-distributed conceptual hourly rainfall-runoff model. The performances of the different methods are compared. Results showed that when ensemble dimension is small, EnKF achieved better results than particle filters. For intermediate number of realisations, RR method outperformed EnKF, and SIR gave the worse results. The enlargement of the ensemble did not improve the results of RR and EnKF. The simulations showed also the importance in choosing the optimal error models to obtain a good estimate of the system states. In particular, particle filters are more sensitive to this choice than EnKF. EnKF and SIR methods have been compared also in Zhou et al. [2006], in land surface assimilation, and again the validity of EnKF method has been confirmed.
Chen and Zhang [2006] presented a study in which ensemble Kalman filter was used for the assimilation of dynamic and static measurements and for the estimation of model variables and parameters. The validity of EnKF is confirmed both for parameter estimation and for system state prediction. In particular, the assimilation of dynamic measurements improves the predictability of the model. A parameter estimation with ensemble Kalman filter is reported in Liu and Chen [2008] and it concerns hydraulic conductivity and various transport model parameters. Also in this work the validity of EnKF even in approaching nonlinear fluid flow is confirmed.

Xie and Zhang [2010] studied a combined state-parameter estimation via EnKF, based on a physically-based hydrological model. Runoff data and other measurements are assimilated. Results showed that the estimation of hydrologic states is accurate even if the prior knowledge of the parameter is not precise. The estimation of hydrologic parameters, on the other hand, is harder due to the ill-posedness of the problem in distributed models. The estimation of parameters may be improved assimilating other measurements besides runoff data. Also in Franssen et al. [2011] the ensemble Kalman filter is used for a state-parameter estimation. In particular, the model is updated in real time with the collected measurements.

Ensemble Kalman filter is compared with Newtonian Nudging in the assimilation of synthetic observations in Camporese et al. [2009]. The model used to fulfill this comparison is CATHY. Measurements were extracted from a synthetic true simulation, set up from realistic data of rainfall and evaporation. The results evidenced that newtonian nudging performances depend strongly on the number of spatial observations. The accuracy of predicted streamflow hydrograph is sufficiently high only when there is a large number of available measurements. EnKF method, on the other hand, well reproduces the true subsurface state when streamflow is assimilated along with pressure head and soil moisture. The produced streamflow hydrograph is more accurate.

In Camporese et al. [2012], hydraulic conductivity and specific storage coefficients are estimated in a real-world case in Padua. Piezometric data are assimilated with EnKF method.

Pasetto et al. [2012] compared the performances of EnKF and SIR, implemented in CATHY model. The physical, statistical and numerical consistency of the state variables during the entire assimilation process were assessed. Results showed that both methods helped in improving the system forecast, with respect to the simulations without data assimilation. When only streamflow was assimilated, then EnKF did not reproduce adequately the pressure heads during evaporation periods. Streamflows and pressure heads, in fact, are poorly correlated and the unsaturated zone has highly non-linear dynamics. Another conclusion of this experiment is that the update step of EnKF may return a physically inconsistent updated state. This circumstance increase the difference in the non-linear solver, and some realisations which do not reach convergence may get lost. SIR updates, on the other hand, are affected by numerical degradation of the ensemble when the ensemble size is small and the initial conditions are not accurate. The authors proposed a modified version of the algorithm to avoid the degradation. Overall, however, the experiments demonstrated that both methods may be used as data assimilation algorithms for hydrological simulations with CATHY model, with SIR the method of choice

if non-Gaussian distributions arise.

This overview evidences the reliability of EnKF, both for state prediction and parameter estimation, even if the updated state is not always physically consistent with the model. Particle filters, and SIR in particular, offer analogous performances with the advantage of being able to handle also strong nonlinearities of the considered system and to assure the physical consistency of the updated states. These considerations led to the choice of EnKF and SIR methods for the parameter estimation performed in this thesis work.

Chapter 3

Landscape Evolution Observatory

In the Earth surface and outer subsurface, physical, chemical, biological, geological processes take place. The quantitative understanding of the complex interaction between these processes is a difficult but essential task to describe, and eventually control, the Earth surface dynamics. The LEO was built in order to improve the knowledge about the interactions among these processes and to have a better instrument to investigate:

- how water, carbon and energy move through landscape;
- the consequences of climate changes on Earth's landscapes;
- the effects of biological systems as vegetation and microbes on landscapes;
- the alteration of terrestrial water resources due to climate changes.

This thesis work aims in estimating the hydraulic conductivity of the LEO soil. This chapter describes the characteristics of the facility and the features of the experiment.

3.1 LEO intent and general features

The LEO is located inside the facility of Biosphere 2 that rises near Tucson, Arizona. LEO was built by the University of Arizona with the objective of better quantifying the interactions among hydrologic partitioning, geochemical weathering, ecology, microbiology, atmospheric processes and geomorphic changes associated with incipient hillslope development.

LEO is composed of three identical slanting slopes that are 11.25 m wide and 29.60 m long, with an average declination equal to 10°. Figure 3.1 shows the image of one of the three slopes. Along with the main inclination, also the transverse sections of the

Figure 3.1: One of the three slopes composing LEO facility.

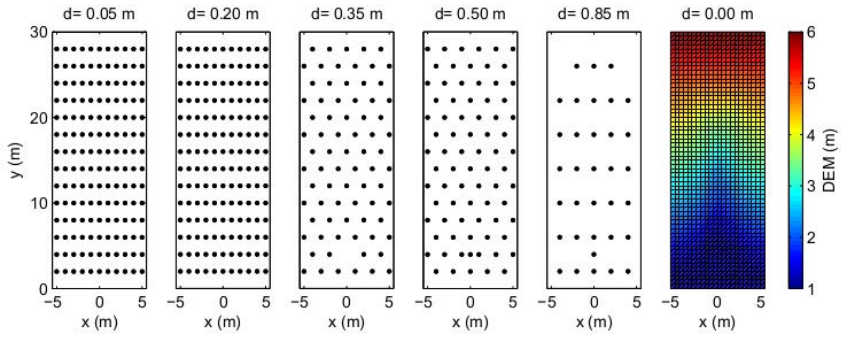


Figure 3.2: Location of the 496 sensors of soil water content and DEM of the surface of the hillslope [Pasetto et al., 2014].

Table 3.1: Soil sample characteristics based on column experiment [Gevaert et al., 2014].

Characteristic	Value
Porosity	0.39
Bulk density	1.59 g/cm^3
Capillary fringe	30 cm
Saturated hydraulic conductivity	2.2×10^{-5} m/s

hillslope slightly converge toward the center with an average angle of 7°, forming a V-shaped profile. Basaltic ground and loamy sand are the components of the landslides. Seeking for a spatially homogeneous landscape, the soil was spread as loose material on the basement, for several horizontal strips and vertical layers [Gevaert et al., 2014]. Each layer was compacted to a specific depth before the deposition of new soil. The final thickness of the hillslope is equal to 1 m. In order to ease the downslope flow, a gravel section is placed at the lower end of each slope. This is 50 cm wide and it is kept in place by a plastic plate perforated with 2 mm holes [Niu et al., 2014]. The initial soil hydraulic properties were estimated analysing a sample of soil of the same kind of the one used to build the hillslope. In order to limit the disturbance, the sample was not extracted directly from the hillslope [Gevaert et al., 2014]. The initial sample soil characteristics are listed in table 3.1. In a further investigation based on the particle size distribution, the value of the hydraulic conductivity was determined to be 7.8×10^{-6} m/s [Gevaert et al., 2014].

A large number of sensors have been installed in LEO for the monitoring of many physical, chemical and biogeochemical features. The spatial density of sensors in LEO is unlikely to be achievable in natural field settings. In detail, each hillslope has the following instrumentation:

- 496 sensors of soil water content;
- 992 sensors of soil temperature;
- 496 sensors of soil water potential;
- 496 samplers of soil water;
- 141 samplers of soil gas;
- 48 sensors of carbon dioxide concentration;
- 24 surface heatflux plates;

- 120 custom electrical resistivity tomography probes;
- 34 piezometers;
- 10 load cells;
- 15 magflow and tipping bucket flow meters in rain and drainage system;
- 1 3D laser scanner;
- a suite of permanent and deployable atmospheric and ecological instrumentation.

Figure 3.2 shows the location of the 496 sensors measuring soil water content, which are placed at five different depths. Note that the majority of the sensors is located at the superficial layers, with the goal of capture in detail the interactions between the unsaturated soil and the atmosphere. The last panel of Figure 3.2 shows the DEM of the LEO surface.

The artificial rainfall system allows to design both homogeneous and heterogeneous rain patterns. Rainfall rate may be kept constant in time or also vary on any of the three slopes. The range of possible rates is 0.003-0.045 m/h. The sensors allow to collect measurements at short time intervals and in a wide spatial range.

The LEO high verifiability and monitoring capacity of real-scale hydrological processes make the facility suitable to validate and, eventually, develop hydrological models. In turn, the models may be used to interpret measured data and, in future, to forecast the outcomes of LEO experiments. An example is given in Niu et al. [2014], where the coupled surface-subsurface model CATHY (see Chapter 4) is used first to predict the behaviour of the LEO hillslope during the first experiment, and then to better understand LEO hydrological response.

In the first three years from the construction, the slopes will have no vegetation. Many aspects will be investigated:

- hydrologic processes;
- surface modification due to rainsplash and overland flow;
- hillslope-scale fluid transit times;
- evolution of moisture distribution;
- rates and patterns of geochemical processes;
- emergent non-vascular and microbial ecology;

- development of carbon and energy cycles within the shallow subsurface.

In a second phase of the experiment, vascular plants able to tolerate both heat and droughts will be introduced and the consequences on water carbon and energy circulation in the slopes will be investigated.

3.2 First experiment at LEO

The hydrological dynamics of the landscape are investigated in the first experiment at LEO-1, the first of the three identical landscapes. The experiment was aimed to test the functionality of the sensor network, investigate the landscape hydrological response under a heavy rainfall, and generate a steady state of soil moisture for further tracer experiments.

Experiment planning

The experiment planning comprised the following successive steps [Niu et al., 2014]:

1. a continuous and constant rainfall rate is applied to the landscape, in order to make it reach the hydrologic steady state;

2. at the reaching of steady state, the artificial rainfall event is interrupted and the slope drains for a week;

3. another continuous and constant rainfall event is applied to the slope, the water being labeled with deuterium.

The planning included the automatic sampling of rainfall and seepage water outflow at every 15 min and the manual measuring from a subset of the soil suction lysimeters array at each 3 h. Initial water storage was equal to 36.13 m^3, which corresponds to a volumetric water content equal to 8-11% in most of the hillslope. The bottom of soil near to the central trough however was wetter due to rainfall system deliver testing several weeks prior to the described experiment [Gevaert et al., 2014]. Rainfall started at 10:00 LT, 18 February 2013 and ended at 8:00 LT, 19 February 2013 [Niu et al., 2014]. Rainfall intensity was constant and equal to 12 mm/h [Gevaert et al., 2014].

Experimental actual outcome

Prior to the experiment, numerical simulations were run in order to estimate the time requested by LEO-1 to reach steady state under the chosen experimental conditions. The

hydraulic homogeneity of the slope was a fundamental assumption. According to the simulation, seepage flow would have been equal to the precipitation rate 36 h after the beginning of the rainfall. On the contrary, in the actual process steady state was never reached, while saturation excess was developed. Overland flow occurred about 15 h after the start of the rainfall, which caused the removal of a small fraction of soil and the formation of a slight channel in the central part of the slope [Niu et al., 2014]. Overland flow was estimated in an indirect way. During rainfall, it has been calculated on the basis of a water mass balance using measured precipitation. After the end of the rainfall, overland flow rate was calculated basing on the time needed to fill a fixed volume. This procedure was repeated at every half hour [Gevaert et al., 2014].

From the analysis of the measured data, Gevaert et al. [2014] discover that the saturation process had been step-wise. In the first phase, the soil was in dry conditions. In the second phase, an abrupt passage from dry to wetter unsaturated conditions occurred. This circumstance corresponded to the passage of the infiltration front. Once it passed, the soil moisture content remained constant in time, while it decreased with increasing depth. The third phase started with the passage of the saturation front. Soil passed from wet unsaturated conditions to wet saturated conditions. Shallower sensors perceived the passage of the infiltration phase first. The second phase then started before in the shallower part of the hillslope and then propagated towards the bottom of the soil profile. On the contrary, the third phase started from the deeper layers and then propagated towards the surface. The latter phase did not ever reached neither the top of the hillslope nor the far sides.

Numerical simulations with the hydrological model CATHY may help to better understand the water dynamics in the artificial hillslope. The model is described in Chapter 4, while the analysis on LEO first experiment, which is the core of this thesis work, is described in the following section.

3.3 CATHY application on LEO first experiment

In the first LEO experiment, results given by the prior numerical simulations were relevantly different from the actual behaviour of the slope during the experiment (see 3.2). A possible explication of the sizeable disparity between the numerical simulation and the actual behaviour of the hillslope under the same conditions is the development of heterogeneity due to transport of fine sediments in the downstream direction. The gravel section was partially removed and analysed shortly after the experiment. The fraction of fine particles for volume of gravel was equal to 2%, and the holes in the plastic plate

were partially clogged with fines. The importance of these two facts in determining the difference between predicted and actual behaviour is not known. A modeling study has been fulfilled in order to better understand the causes of the yet mentioned differences between hydrological model response and actual behaviour of the hillslope [Niu et al., 2014].

Hillslope Modeling and boundary conditions

For the numerical simulation of the hydrological experiment, a discrete representation of the hillslope is necessary. Niu et al. [2014] discretize the LEO-1 slope into a grid of 60×24 cells in the horizontal direction and 8 layers in the vertical direction. The resulting number of superficial nodes is 61×25. In the vertical direction the 9 nodes are not uniformly distributed, being thicker at the top and the bottom of soil surface. In this way the infiltration at soil surface and seepage flow at the bottom nodes is better characterised. The following boundary conditions (BC) are assigned to the nodes on the boundaries of the slope:

- seepage face BC to the 25×8 nodes modeling the downslope boundary;
- atmospheric BC to the nodes on hillslope surface;
- zero flux conditions to every other LEO boundary node.

The atmospheric BC, Q_{atm}, was indirectly estimated, since evaporation E was not directly measured. The estimate differs depending on the time period [Niu et al., 2014]. From 08:00 LT to 20:00 LT, 18 February 2013 the rate of change in water storage dS/dt is used to estimate Q_{atm}. Prior to the beginning of seepage and overland flow, in fact, evaporation can be seen as the difference between dS/dt and the rainfall rate P. Evaporation is considered negligible in the successive time period, from 20:00, 18 February to 08:00, 19 February. Atmospheric boundary condition in this phase is considered equal to the sprinkler rainfall rate. In the following and last time period, rainfall was absent and evaporation rate is set equal to the average evaporation rate in Arizona in the actual conditions. Table 3.2 summarizes the values of Q_{atm} in the different phases.

In the simulations time stepping varies from 0.1 to 180 s, basing on the convergence of the iterative linearisation scheme of Richards equation (see 4.4). Different numerical simulations are fulfilled with different parameter configurations for soil properties [Niu et al., 2014]. Involved soil parameters are, among the others, the van Genuchten curve fitting parameter (n), the porosity (θ_{sat}) and the saturated hydraulic conductivity (K_{sat}).

Table 3.2: Atmospheric boundary condition values

Phase	Time interval (LT)	Q_{atm}
1	08:00 18/02 - 20:00 18/02	dS/dt
2	20:00 18/02 - 08:00 19/02	12 mm/h
3	08:00 19/02 - end	-0.083 mm/h

Scenarios overview

The numerical simulations reported by Niu et al. [2014] have been fulfilled both before and after the real experiment. Two scenarios, M1 and M2, refer to two modelizations of LEO-1 soil prior to the experiment. They both assume homogeneous hydraulic conductivity, K_{sat}, with different values. In M1, K_{sat} is modeled with the value obtained from an analysis of soil particle size distribution while the K_{sat} calibrated based on a previous LEO test is used in M2. The results using these models do not forecast any overland flow as shown in Figure 3.3. This implies that the value of porosity considered in M1 and M2 is too high, and/or the homogeneity assumption is not correct. The following attempts, M3 and M4, explore the eventual heterogeneity of saturated hydraulic conductivity. Scenarios M3 and M4 differ in the value of van Genuchten parameter n. In the configuration M3, n is derived from laboratory tests on LEO soil samples, in such a way to minimize the error with the soil retention data (n=1.72). In configuration M4, n is obtained from a preexperiment analysis of particle size distribution, and it is equal to the one of scenarios M1 and M2 (n=2.26). Being fixed the other soil parameters, both in M3 and M4 configurations two spatial distributions of K_{sat} are explored. One is under the hypothesis of homogeneous soil, the other assumes that the LEO hillslope surface has a lower hydraulic conductivity than the rest of the slope soil. The best fitting set of parameters is evaluated minimizing the error between measured and simulated data. The mean relative error is computed as the mean between the error on water storage variation $e_{\Delta S}$ and seepage face flow e_{QS}.

$$e = \frac{1}{2}(e_{\Delta S} + e_{QS}) \tag{3.1}$$

$$e_{\Delta S} = \frac{\int_0^T |\Delta S_m - \Delta S_s| \, dt}{\int_0^T \Delta S_m \, dt} \tag{3.2}$$

where the subscripts m and s indicate respectively the measured and simulated data sets. The relative error for seepage face flow e_{Qs} is computed similarly to $e_{\Delta S}$.

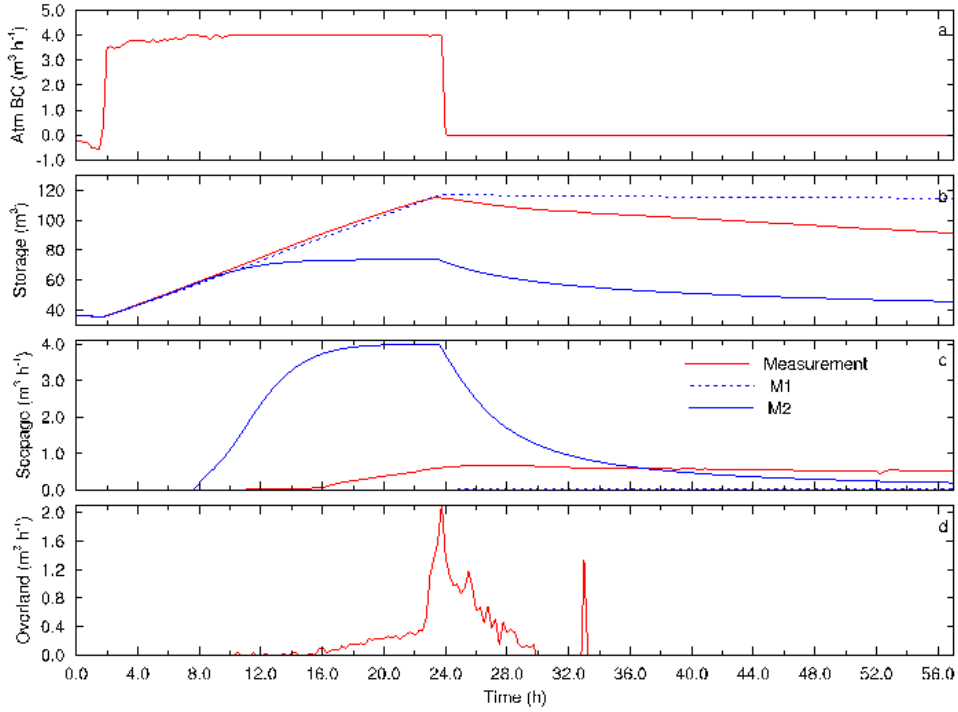


Figure 3.3: Comparison between measured and simulated storage, seepage face flow and overland flow. Results for scenarios M1 and M2 [Niu et al., 2014].

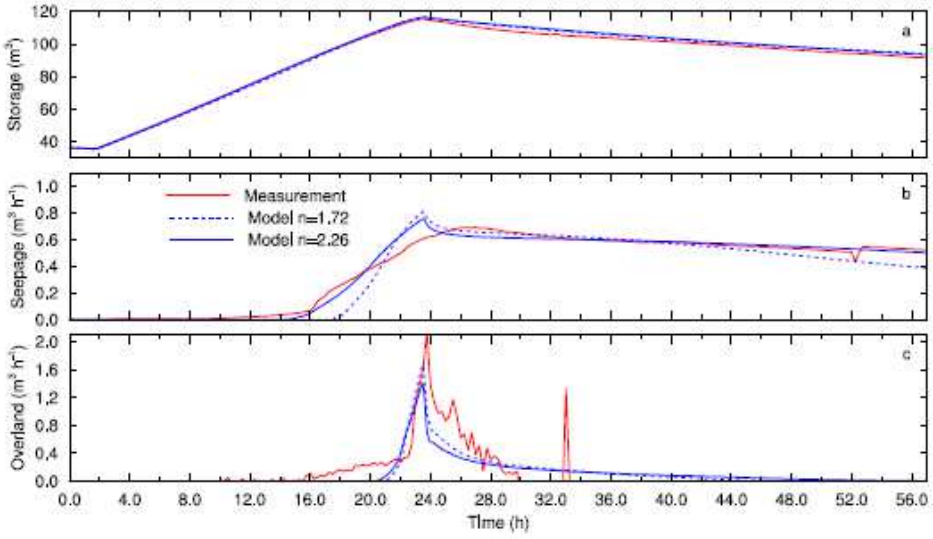


Figure 3.4: Comparison between measured and simulated storage, seepage face flow and overland flow. Results for scenarios M3 and M4 with heterogeneous K_{sat} [Niu et al., 2014].

Results

The results of the pre-experimental simulations (scenarios M1 and M2) are plotted along with the actual measurements in Figure 3.3. Both configurations do not give overland flow, while the answer in terms of seepage flow is diametrically opposite and far from the measurements. M1 has a smaller value for saturated hydraulic conductivity, then seepage face flow is negligible. Water storage remains constant after having reached the peak. On the other hand, the higher K_{sat} value of M2 gives an overestimation of the seepage face flow. Being overland flow null in both scenarios, the only variation of saturated hydraulic conductivity value is not sufficient to reproduce the observed overland flow. Both scenarios M3 and M4 consider a heterogeneous and a homogeneous configuration. In scenarios M3, the homogeneous configuration gives a higher number of best simulations in terms of water storage, while the heterogeneous configuration gives a higher number of best simulations concerning seepage flow. The quality of the simulations is established on the basis of the relative error. However, considering the mean error normalized on the number of simulations, the M3 heterogeneous configuration has more best simulations than the homogeneous. M4 scenario has a higher value for van Genuchten parameter n. The results are analogous to those of M3 scenario. In this case, the heterogeneous configuration has a higher number of best simulations, in particular for seepage flow. In addition, the heterogeneous configurations of M3 and M4 scenarios show that the latter has a best performance in terms of mean relative errors. The fraction of simulations with a mean relative error lower than 10% is higher in the latter configuration than in the former. The same is valid for an error threshold of 15%. Both the best heterogeneous realisations of M3 and M4 scenarios give a good response also in terms of overland flow, that almost match the estimated overland flow occurred during the experiment as shown in Figure 3.4.

The values of K_{sat} calibrated in these analysis are used to define the prior distribution of K_{sat} in the data assimilation process fulfilled in this thesis work. Chapter 6 contains the detailed description of the procedures and of the results.

Chapter 4

CATHY model

The numerical simulation of LEO experiments by means of a hydrological model is a fundamental step for a deeper description and, eventually, the forecast of LEO hydrologic dynamics. In this thesis work, the numerical simulations are conducted with the model CATHY (CATchment HYdrology). This chapter is devoted to the description of the model.

4.1 Catchment hydrology model

Atmosphere, surface and subsurface represent a continuous system in which water circulates under different phases [Furman, 2007]. Focusing on the surface-subsurface subsystem, water passes from surface to vadose or saturated zones through infiltration. In turn, groundwater nourishes surface streams through springs. Both surface and vadose zone receive water from atmosphere via rain precipitation. Water fluxes in the different system elements are interdependent. In order to reproduce accurately these processes, the interconnections between surface and subsurface must be taken into account, despite of the significant difference in the timescales of surface and subsurface water flow. Coupled hydrological models like CATHY try to recreate the continuum between surface and subsurface modules. In these models the constraints that each element of the system exerts on the others are expressed by boundary conditions.

CATHY is a physically-based model [Camporese et al., 2010] that couples the partial differential equations describing the surface and subsurface flows. CATHY automatically subdivides atmospheric fluxes into infiltration, exfiltration and change in surface water storage [Sulis et al., 2010]. The model consists of two partial differential equations, describing the surface and subsurface flows. The surface module is approximated by a rill flow along a one-dimensional drainage pattern, while the subsurface module is treated

with a three-dimensional approach. The numerical solutions of the two equations are described in Subsections 4.2 and 4.4, respectively. In addition to the usual atmospheric specified flux and specified head boundary conditions, the model allows for the definition of seepage face boundary conditions and non-atmospheric Dirichlet and Neumann conditions, which may vary in time and space. Different levels of soil heterogeneity can be handled by CATHY, by assigning different parameter values to different zones of the domain.

4.2 Surface runoff routing

CATHY surface module subdivides the drainage basin in hillslope and channel flows, which determine the two main processes governing the surface hydrologic response [Camporese et al., 2010]. At the beginning of the rainfall event, in fact, water flows over hillslopes or agricultural watersheds as sheet flow. In a second time, soil erodibility and topographic irregularities lead to the concentration of flow into small channels [Orlandini and Rosso, 1996]. However, it is assumed that also hillslope flow can be described by an accurate surface flow path. The whole surface flow then is described by a system of one-dimensional rivulet flow routing along a network of surface flow paths. The eventual presence of lakes, streams and pools that may cause retardation in the water flow can be taken into consideration (see Subsection 4.3). Available DEM data allow us to accurately represent the surface topography, and, as a consequence, the hillslope flow paths. In fact, drainage network and flow direction are individuated starting from the catchment DEM and following a precise procedure. First, the DEM cells are sorted into descending elevation order. After that, the cells in flat or depressed areas are individuated and their elevation is raised in order to have always a small positive slope downward. Finally, the cells are again sorted into descending elevation order. The flow direction is then assigned following mainly three schemes:

- D8 scheme;
- Multiple direction D∞ method;
- Nondispersive D8-LTD scheme.

Depending on the chosen scheme, each cell will discharge water to one or two downward cells (respectively for the D8 and D8-LTD schemes and for the D∞ scheme). By means of this methods it is also possible to define the drainage area of each cell. Three ways are available in order to distinguish hillslope flow and channel flow: posing a threshold on

the upstream drainage area A, on the function $A \times S^k$, where S is the local terrain slope and k an exponent, or on the ratio between land surface curvature and the mean terrain slope. Notwithstanding the chosen criterion, if the imposed threshold is not exceeded, then hillslope rill flow occurs. Otherwise channel flow takes place. This approach allows the model to define the channel head locations with a physically realistic description of the surface flow. It is assumed that the bed slope and length of each part of the extracted drainage network are dependent on their location in the transport network. Cross section resistance and conductance vary with the discharge, along with the water-surface width. In this way the surface flow in the model may range from a system of tiny rivulets to a network of wide rivulets, with the formation of a sheet flow.

In CATHY a preprocessor determines the surface drainage channel network from the DEM data. Water flux in the channel network is described by the diffusion wave approximation of the Saint-Venant equation:

$$\frac{\partial Q}{\partial t} + c_k \frac{\partial Q}{\partial s} = D_h \frac{\partial^2 Q}{\partial s^2} + c_k q_s \tag{4.1}$$

where s is a spatial coordinate $[L]$, Q is the discharge along the rivulet/stream channel $[L^3/T]$, c_k is the kinematic celerity $[L/T]$, D_h is the hydraulic diffusivity $[L^2]$, and q_s is the inflow (positive) or outflow (negative) rate from the subsurface to the surface $[L^3/LT]$. Muskingum-Cunge method is the basis for the surface routing scheme used in CATHY model. Hillslope rill and channel flows have different characteristics. These distinctions are taken into account by considering different distributions of Gauckler-Strickler roughness coefficient [Orlandini and Rosso, 1996]. Each cell receives water both from the upward cells and the subsurface. The latter contribution includes both the water actually exfiltrating from the subsurface and the water that cannot infiltrate. The subsurface contribution is expressed as total flux of water per unit length of channel link $[L^3/LT]$. Also lateral inflow q_L has the same unit measure, since it is calculated as described in Equation (4.2) [Orlandini and Rosso, 1996]:

$$q_L = q \frac{\Delta x \Delta y}{\Delta s} \tag{4.2}$$

where q $[L/T]$ is the local contribution to infiltration excess runoff, Δx and Δy are DEM cell sizes in the direction of the horizontal cohordinates x and y and Δs is the channel length within the cell. Water depth on each cell, i.e., the ponding head h, is determined starting from the incoming and outgoing discharges of that cell. Equation (4.3) [Camporese et al., 2010] allows the routing of lateral inflows and inflow hydrographs onto each individual channel:

$$Q_{i+1}^{k+1} = C_1 Q_i^{k+1} + C_2 Q_i^k + C_3 Q_{i+1}^k + C_4 q_{s_{i+1}}^k \tag{4.3}$$

where the term Q_{i+1}^{k+1} ($[L^3/T]$) is the discharge at network point $(i+1)\Delta s$ and time $(k+1)\Delta t$, $q_{s_{i+1}}^k$ is the subsurface contribution at the $(i+1)th$ space interval and time $k\Delta t$. The coefficients C_1, C_2, C_3 [-] and C_4 [L] depend on the kinematic celerity c_k, on the temporal interval Δt, on the channel length Δs and on a weighting factor X.

4.3 Topographic depressions oversight

The artificial modification of the elevation of isolated depressed DEM cells is applied in order to correct DEM errors and regularise the grid. It is a suitable approach also if the basin is steep, where the main driving force of the water flow is slope. In quite flat areas, on the contrary, this approach may lead to incorrect reproduction of storage and retardation effects on the catchment response. In fact, the topographic depressions may represent DEM errors but also pools and lakes. Then it is necessary to warily procede in modifying cell elevation. In CATHY model the real nature of the depressions is individuated basing both on DEM informations and on prior field observations. The correct extraction of the drainage network may be compromised by these depressions. The matter is avoided applying a "lake boundary-following" procedure. The cells around the pit ("buffer cells") collect the water of all the cells which draw off water to the depression. The flow direction in these cells is forced to address water to a single cell, representing the lake outlet. After that the modified DEM is again corrected, excluding from the analysis the central cells of the depression. The lake outlet cell, or reservoir cell, assumes all the geometrical and physical characteristics of the depression. The buffer cells transfer with infinite celerity all the water they drain to the outlet cell, expressing in this way the storage and retardation effects of the depression. Solving the continuity equation of the reservoir gives the outflow from the reservoir cell.

$$\frac{dV}{dt} = I(t) - O(h^*) \tag{4.4}$$

where V is the storage volume of the reservoir, I and O are the incoming and outgoing discharges, respectively dependent on time t and water elevation h^* above a reference level in the reservoir. The calculated outflow is used in the surface routing as outgoing flux from the cell. Reservoir water elevation is assigned to all the lake cells. The boundary conditions of subsurface solver include also these values corresponding to the ponding head.

4.4 Subsurface module

The subsurface flow is modeled by the Richards equation:

$$S_w(\psi)S_s\frac{\partial\psi}{\partial t} + \phi\frac{\partial S_w}{\partial t} = \text{div}\left[K_s K_r(S_w(\psi))\left(\nabla\psi + \eta_z\right)\right] + q_{ss}(h_p) \tag{4.5}$$

where S_w is water saturation $[-]$, i.e., the ratio between the volumetric moisture content θ and the saturated moisture content θ_s (equal to the porosity, ϕ), S_s is the aquifer specific storage coefficient $[L^{-1}]$, ψ is pressure head $[L]$, t is time $[T]$, ∇ is the gradient operator $[L^{-1}]$, K_s is the saturated hydraulic conductivity tensor $[L/T]$, $K_r(\psi)$ is the relative hydraulic conductivity function $[-]$, $\eta_z = (0, 0, 1)'$, z is the vertical coordinate directed upward $[L]$, and q_{ss} represents distributed source or sink terms (respectively positive and negative) $[L^3/L^3T]$. The nonlinear relations between ψ and the relative hydraulic conductivity K_r and the water saturation S_w can be explicitly described by the curves of Brooks and Corey [1964], Van Genuchten and Nielsen [1985] or Huyakorn et al. [1984]. The equation needs boundary conditions. In this case, at the surface-subsurface interface they are derived from the numerical solution of the surface module (see 4.5). In the model, heterogeneity is expressed by assigning different initial values of hydrological parameters in different zones of the domain. The zones may include many cells or correspond even to a single cell. In the latter case, there is full heterogeneity.

Galerkin discretization

A finite element Galerkin method is used to numerically solve the Richards equation (4.5) [Camporese et al., 2010]. The spatial discretization of the domain is based on tetrahedral elements, since the problem is three-dimensional [Paniconi and Putti, 1994]. The soil parameter that characterize the nonlinear equation are estimated in element centroids. Time discretization is fulfilled through a weighted finite difference scheme [Paniconi and Putti, 1994]. Galerkin method is finite element procedure usable in continuum problems [Huyakorn and Pinder, 1983]. In general, the finite element approach can be described with five steps. In the first step, the physical system is discretized into finite elements. A discrete number of nodal points assures the link among the elements. In the second step, nodal variables of each element are connected via a matrix expression, the so-called *element matrix*. In the third step, a description of the entire global system is built via the assemblage and combination of the element matrices into algebraic equations. A *global matrix* collects the coefficients of the final set of equations. The nodes shared by different elements are characterised by the assignment of compatibility conditions which

must be respected in the assembling phase. In the fourth step, the global matrix equation is completed with the prescribed boundary conditions. In the fifth and last step, the obtained system of algebraic equations is solved. Galerkin method describes a particular way to compute the element matrix, i.e., step two. It belongs to the group of finite element methods based on the weighted residuals (MWR) [Huyakorn and Pinder, 1983]. These methods are applied on continuum problems governed by a differential equations like the following:

$$\mathcal{L}(u) - f = 0 \tag{4.6}$$

where u is the unknown function, $\mathcal{L}$ the differential operator and f the known forcing term. Equation (4.6) is defined in a domain Ω, which is a compact subset of $\mathbb{R}^3$. In general, finite element methods approximate the solution throughout three steps.

1. In the first step the unknown function u is approximated with a trial function defined as a linear combination of n basis functions

$$\widehat{u} = \sum_{i=1}^{n} v_i c_i \tag{4.7}$$

where v_i are linearly independent basis functions defined over the domain Ω, c_i the unknown coefficients to be determined, and n is the number of unknowns in the discretized system. The substitution of the trial function $\widehat{u}$ in Equation (4.6) produces a residual ε:

$$\varepsilon = \mathcal{L}(\widehat{u}) - f \tag{4.8}$$

2. MWR methods try to minimize the residual ε. To achieve this minimum, a weighted integral on the entire domain Ω is realised and then placed equal to zero. In the second step a number n of weighting functions w_i is selected. MWR methods compute c_i imposing the following condition:

$$\int_\Omega w_i \varepsilon \, d\Omega = \int_\Omega w_i [L(\widehat{u}) - f] \, d\Omega = 0 \quad \text{for} \quad i = 1, 2, ..., n \tag{4.9}$$

Once the weighting functions are specified, then Equations (4.7) and (4.9) can be combined. The result is a set of equations in the n unknowns c_i, $i = 1, ..., n$.

3. In the final step the set of equations is solved with respect to c_i and then the unknown function u can be approximated by means of Equation (4.7).

In the Galerkin method the weighting functions and the basis functions are the same, i.e., $w_i = v_i$. Then, the weighted residual equations are:

$$\int_\Omega v_i \varepsilon \, d\Omega = \int_\Omega v_i [L(\widehat{u}) - f] \, d\Omega = 0 \quad \text{for} \quad i = 1, 2, ..., n \tag{4.10}$$

First, in Galerkin finite element method the domain is subdivided in elements, e.g., tetrahedral, which cover all the domain without overlapping. The unknown function is approximated over each element T^e by the following trial function:

$$\widehat{u}\Big|_{T_e} = \sum_{i=1}^{n_e} v_i u_i \tag{4.11}$$

where n_e represents the number of nodes on the element T_e,v_i are the interpolating functions defined on the element T_e (which means that, for each node on T_e there is a basis function that is equal to one over that node, and zero over the other nodes), and the coefficients u_i become the unknown values of the function u over the nodes. The integral in Equation (4.10) is equal to the sum of the integrals realised over each element, and thus it can be easily computed. The local integrals Q_i^e governing the behaviour of each element can be written as:

$$Q_i^e = \int_{T_e} v_i [\mathcal{L}(\widehat{u}) - f] \, dT_e, \quad i = 1, 2, ..., n \tag{4.12}$$

Then, the global matrix equation is obtained assembling the local contributions Q_i^e.

In CATHY model, the Galerkin method is applied to Richards equation 4.5. The final system of equations is the following:

$$f(\Psi^{k+1}) = A(\Psi^{k+1})\Psi^{k+1} + F(\Psi^{k+1})\frac{\Psi^{k+1} - \Psi^k}{\Delta t^{k+1}} + b(\Psi^{k+1}) - q(t^{k+1}) = 0 \tag{4.13}$$

where Ψ is the vector of nodal pressure heads, superscript k denotes the time step, A is the stiffness matrix, F is the storage or mass matrix, b contains the gravitational gradient component of Richards equation, and q contains the specified Darcy flux boundary conditions and forcing terms. Note that the temporal integration in (4.13) is achieved with the implicit Euler scheme with a variable time step $\Delta t^{k+1} = t^{k+1} - t^k$. The nonlinear equation (4.13) is then solved via Newton or Picard methods [Paniconi and Putti, 1994].

Newton and Picard nonlinear solvers

Newton scheme is an iterative method to compute the solution of a nonlinear equation. A sequence $\Psi^{k+1,(1)}, \Psi^{k+1,(2)}, \ldots, \Psi^{k+1,(m)}$ is construct until the satisfaction of a convergence criteria for the solution of Equation 4.13. Newton scheme can be written as follows:

$$f'(\Psi^{k+1,(m)})h = -f(\Psi^{k+1,(m)}) \tag{4.14}$$

where $h = \Psi^{k+1,(m+1)} - \Psi^{k+1,(m)}$, superscript (m) is an iteration index and $f'(\Psi^{k+1})$ is the Jacobian matrix. The $ij-th$ component of the Jacobian matrix is written as:

$$f'_{ij} = \lambda A_{ij} + \frac{1}{\Delta t^{k+1}} F_{ij} + \sum_s \frac{\partial A_{is}}{\partial \psi_j^{k+1}} \psi_s^{k+\lambda} + \frac{1}{\Delta t^{k+1}} \sum_s \frac{\partial F_{is}}{\partial \psi_j^{k+1}} (\psi_s^{k+1} - \psi_s^k) + \frac{\partial b_i}{\partial \psi_j^{k+1}} \tag{4.15}$$

The initial guess for the new iteration is typically set to $\Psi^{k+1,(0)} = \Psi^k$.
Picard scheme can be seen as an approximation of the Newton scheme. In fact, it is:

$$\left[\lambda A(\Psi^{k+1,(m)}) + \frac{1}{\Delta t^{k+1}} F(\Psi^{k+\lambda,(m)})\right] h = -f(\Psi^{k+1,(m)}) \tag{4.16}$$

where the meanings of the terms are the same of the ones in Newton scheme. Picard scheme converges only linearly, while Newton scheme converges quadratically. The presence of the three derivative terms in Newton scheme makes it computationally expensive. The initial estimate of the solution affects sensitively the convergence of both schemes. In CATHY model the time evolution of the simulation is controlled by the subsurface solver. In fact, the length of its time steps is not fixed, but it may be adjusted [Camporese et al., 2010]. The length of the time steps is changed throughout the simulation, depending on the convergence behaviour of the nonlinear scheme [Paniconi and Putti, 1994]. A tolerance and a maximum number of iterations allowed for each time step are defined. Simulation ends at time T_{max}, that corresponds to the final time of the experiment. The initial time step is Δt_0, with a maximum number of iteration equal to $iter^0_{max}$. For the following iteration, Δt_0 is increased of a fixed quantity if the system converged in less than the allowed iterations. It remains equal if the convergence needed a number of iterations included between $iter^1_{max}$ and $iter^2_{max}$. If the convergence is reached in more than the allowed iterations, Δt_0 is reduced of a fixed quantity. Finally, if convergence is not achieved then back-stepping occurs, i.e., the solution for the actual time level is recomputed with a time step lower than the previous one. In the first iteration a first solution estimate is necessary for the beginning of the iteration procedure, and it corresponds to the initial

conditions.

4.5 Surface-subsurface coupling

The coupling between surface and subsurface equations is forced by ensuring the mass balance at the surface/subsurface interface. The water exchange between surface and subsurface flow regimes is represented by q_{ss} (from surface to subsurface) in the subsurface flow equation (4.5) and by q_s (from subsurface to surface) in the surface flow equation (4.1). Every time step starts with the solution of the surface flow equation allowing the model to compute the surface-to-subsurface contribution. This is given as an input to the subsurface flow equation at the subsequent time. In turn, the solution of the subsurface flow equation gives the value of subsurface-to-surface contribution, which will be an input for surface flow equation. Surface-to-subsurface flux may be also equivalently expressed as ponding head. Time step size, surface area attributions to cells and nodes allow to easily pass from a volumetric flux $[L^3/T]$ to a specific flux $[L/T]$ and to ponding head $[L]$. The coupling process is predominantly governed by the subsurface element. This is due to the fact that the subsurface equation handles atmospheric forcing, which is fundamental in surface-subsurface partitioning. The boundary conditions assigned to the surface nodes are different depending on the saturation of the considered node. The condition may be a Neumann or a Dirichlet condition. The former implies the assignment of a specified flux to the node, the latter the assignment of a specified head. When precipitation and evaporation rates are governed by atmospheric conditions, these fluxes correspond to the Neumann condition assigned to the surface nodes. The situation changes in prolonged periods of rainfall or droughts, when the threshold values respectively of saturation or soil moisture deficit are reached. Infiltration and exfiltration processes are no more governed by the atmospheric status, but they are driven by soil conditions. Boundary conditions are switched from a Neumann sort to a Dirichlet (specified head) one. CATHY considers also the distinction between saturation and ponding status. In order to make this differentiation possible, each surface node is marked with a parameter representing the threshold pressure head value beyond which the water is routed by the surface flow model. During rainfall events, a fixed head boundary condition is assigned to the unsaturated surface nodes which became saturated or ponded during the event. The next time step of the subsurface flow equation will give the return flow rate. When the computed flux exceeds the input potential rate, the Dirichlet condition is switched to a Neumann one. There is no switching if the flux is negative (i.e., if the flux is going backward from the subsurface to the surface). In dry periods, when evaporation takes place, the assigned condition

changes from the Neumann type to the Dirichlet type when soil moisture drops below the moisture deficit threshold, correspondent to a specific pressure head value ψ_{min}. The condition returns a Neumann BC when a rainfall event starts or, even if the dry period endures, when the absolute value of the computed flux becomes larger than the absolute value of the input potential rate. The available informations are then:

- the updated pressure head of surface nodes;
- whether the atmospheric forcing is positive or negative, i.e., either if it is raining or evaporating;
- in presence of Dirichlet condition, the absolute value of the computed flux respecting the potential flux and whether the former represents infiltration or exfiltration.

These acquired facts lead to the definition of the term q_s, given as an input to the surface flow routing module for the next step. The potential fluxes are separated into infiltration, exfiltration as evaporation and exfiltration as return flux by means of the boundary condition switching procedure.

Time stepping

The time evolution of the simulation is governed by the subsurface solver by means of adaptation and back stepping (see 4.4). The chosen time step in subsurface solver, Δt_{ss}, is usually larger than the optimal time step for the surface solver, Δt_s. The consequence is that for each subsurface time step a major number n_s of surface time steps is fulfilled (n_s = int ($\Delta t_{ss}/\Delta t_s$)). The value of n_s changes at each new outer time level, and it depends on the surface flow discharges of the previous time level. The surface and subsurface time steps are then nested one into the other. The constraints imposed to the subsurface solver are intended to avoid an excessive number of surface solver time steps per subsurface solver time step.

Being Muskingum-Cunge method explicit in time, it is then possible to solve equations (4.5) and (4.1) with a non-iterative sequential algorithm, starting at $t^k = 0$ and ending at $t^k = T_{max}$:

1. n_s is determined starting from Δt_{ss} and Δt_s and the equation (4.1) is solved n_s times, where q_s^k is used as input and the outputs are Q^{k+1} and consequently the distribution of ponding heads h^{k+1};
2. boundary conditions for the subsurface solver are set basing on the values h^{k+1} along with the atmospheric input at time t^{k+1}, allowing the solution of equation (4.5) for pressure heads ψ^{k+1};

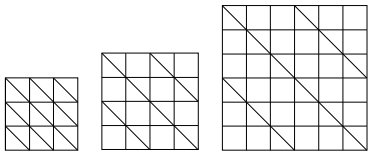


Figure 4.1: Surface grid subdivision and coarsening for the subsurface solver

3. the inflow term of overland flux q_s^{k+1} is calculated via the subsurface solver using ψ^{k+1} and the balance between atmospheric inputs and actual fluxes.

Flow discharges are set as initial condition to the surface equation (4.1) in order to initialise the algorithm. When the needed values are not known, the subsurface solver is run. The output gives an initial guess for the overland flow based on the actual atmospheric input and on the initial distribution of the pressure heads ψ (initially uniformly distributed, hydrostatically distributed along vertical direction or deduced from a steady state water balance simulation).

Surface and subsurface mesh definition

Surface and subsurface modules differ in some not negligible aspects. The former has a one dimensional nature, a cell-based spatial discretization and it ignores the pit cells. The latter, on the other hand, has a three dimensional nature, a node-based spatial discretization, and is forced to consider also the topographic depression cells. Surface and subsurface grids differ for cell numbering, since the latter takes into consideration the pit cells, while the former not. The surface grid is the basis of the tetrahedral three-dimensional mesh. In fact, each cell is subdivided in two triangles, which are then replicated along the vertical direction. The subsurface grid may be coarsened with respect to the surface one, since the subsurface processes are typically slower and computationally more expensive than the surface one. Instead of two triangles per cell, it may consists of two triangles per group of four cells, or per group of nine cells, and so on. Figure 4.1 shows this escamotage. It is finalised to the reduction of computational effort in the subsurface solver, keeping the grid resolution, then the accuracy, of the surface solver. The exchange of informations between cells and nodes needs special attention, due to the differences between surface and subsurface modules described above. A linear interpolation process is necessary since the subsurface grid nodes are at the corners of the cells. Ponding head values in output of the surface solver are transmitted to the node after a linear interpolation. In turn, overland flow contributions, computed by the subsurface solver, are transferred to the

surface cells after a linear interpolation. In defining the boundary conditions, the value assigned to the surface nodes of the subsurface grid corresponds to the computed water depth averaged on the cell area.

Chapter 5

Data assimilation methods

The application of hydrological models to real scenarios may result in large discrepancies between the simulated and the observed processes, which are mainly due to uncertainties on the input data and model parameters, or to model structural errors. Thus, there is the need of correcting the model output with the observations. The idea of data assimilation (DA) methods is to update the model state variables at the observation times in order to obtain a better forecast of the model. The hydrologic model CATHY is integrated with two DA techniques: ensemble Kalman filter (EnKF) and sequential importance resampling (SIR) particle filter. These two methods have been used in this thesis work. In this Chapter we present the mathematical framework of DA methods, with particular attention to EnKF and SIR, and their use for parameter estimation.

5.1 Data assimilation

Hydrologic measurements are often local both in time and space, and are hardly measurable in a direct way. As a consequence, available data are usually scarce, and this condition affects hydrologic studies. The advent of remote sensing data and automated ground-based sensors are improving the data availability. However, the traditional manual approaches and computer-based estimation procedures may be not able to manage the large volume of data yielded by the new measuring technologies. DA techniques have been developed to supply the lack of suitable instruments, with the idea that, in addition to the advanced measuring instruments, also mathematical models describing the environmental systems are a source of information for the DA process [McLaughlin, 2002].

5.1.1 Data assimilation - mathematical formulation

Characterising the state of the environmental system is the principal aim of DA. In hydrologic applications, the relevant variables usually vary both in time and space. These variables may represent a forcing on the system (*input variables*), describe the system behaviour (*state variables*), or be observable variables depending on the state of the system (*output variables*). In dynamic applications, being the system spatially discretized by means of a numerical scheme, a set of state equations describe the temporal evolution of the system state [McLaughlin, 2002]. These equations may be written as follows:

$$x(t) = \mathcal{A}\left(x(\tau), \alpha, u(t), t, \tau\right), \quad t > \tau \geq 0 \tag{5.1}$$

$$x(t_0) = \mathcal{A}_0(\alpha) \tag{5.2}$$

where the vector $x(t)$ collects the discretized system state variables defined at time t, the vector α collects the time-invariant inputs (model parameters), the vector $u(t)$ collects the time-dependent inputs, and $x(t_0)$ is the initial state of the system. The state at time τ is related to the state at time t by means of a nonlinear operator $\mathcal{A}$. The system state variables, which are represented by the vector $x(t)$, are not necessarily measurable in a direct way. It is useful to define a vector $y(t)$ of measurable output variables, depending on the system states:

$$y(t) = \mathcal{H}(x(t), t) \tag{5.3}$$

where $\mathcal{H}$ represents the nonlinear relation between observations and system state. The knowledge of the variables of the state equation is never perfect, and the variation scale is often smaller than the grid resolution. A convenient approach is to consider the state variables as random variables characterized by the joint probability density functions (pdf) $p(u(t_1), ..., u(t_{nx}), \alpha)$ and $p(x(t))$, where t_{nx} is the last time step, $t_{nx} = t_{max}$. In theory, if the former is specified, the latter can be derived through the state equation 5.1. In most cases, the high dimensionality of the spatially discretized state vector and the nonlinearity of the state equation make the derivation unfeasible. Instead of considering the full probability density function $p(x(t))$, DA procedures focus on the mean, mode and variance which characterize it. The mean represents an estimate of the state variable, while the variance measures the uncertainties [McLaughlin, 2002]. Hydrological models compute a forecast of the behaviour of the system, starting from given initial and boundary conditions. Input, output and state variables are affected by significant uncertainties, which may be limited by DA processes. This aim is reached through the incorporation of measurements of the output variables represented by the vector $y(t)$. The measurements

are not perfect, due to sensor errors, extraneous noises, and differences in temporal and/or spatial scales between measurements and output variables. These uncertainties are taken into account in the definition of the output measurement vector $y(t)$ which is slightly different from 5.3:

$$y(t) = \mathcal{H}(x(t), \omega(t), t) \tag{5.4}$$

where $\omega(t)$ is a random variable accounting for the measurement errors (described by the prior pdf $p(\omega(t))$), and the other elements have the meaning described above. Now the available informations are the following:

- state and measurement equations (Equations 5.1 and 5.4);
- probability density functions of system inputs and measurements errors;
- the vector $y_{1:nz} = (y(t'_1), ..., y(t'_{n_z}))$, collecting the measurements at n_z output times which are a subset of the discrete times $t_0, t_1, ..., t_{nx}$.

The conditional probability density $p(x(t)|y_{1:nz})$ is the objective of DA. When the full pdf is not obtainable, then its mean and covariance are investigated [McLaughlin, 2002]. DA in hydrologic systems is highly challenging since the spatial and temporal variability of many hydrological processes and state variables as soil moisture, rainfall, evapotranspiration and hydraulic conductivity are unknown.

Filtering problem

A variation of the DA problem consists in the filtering problem. In this case, the aim is to characterize the pdf of the system state $x(t)$ at a given measurement time, $t = t'_j$, given the set of measurements collected in the interval (t_0, t'_j) [Jazwinski, 1970]. For this reason we will use the notation $x_j = x(t'_j)$ and $y_j = y(t'_j)$. This kind of issue is solved with a sequential approach [McLaughlin, 2002]. Given the knowledge of the pdf at the previous observation time t'_{j-1}, $p(x_{j-1}|y_{1:j-1})$, and the new measurement vector at time t'_j, y_j, the filtering pdf of the system state $p(x_t|y_{1:j})$ is evaluated in two steps:

- the propagation step that propagates the state pdf until the new assimilation time t_j, computing the forecast pdf $p(x_j|y_{1:j-1})$.
- the update step that combines the new measurement y_j with the forecast pdf.

The starting point for the recursion of these two steps is the definition of the initial pdf, $p(x_0)$ at time t_0, which may be derived from the input pdf of the parameters $p(\alpha)$. Then, this pdf is propagated to the first measurement time t'_1. The forecast pdf $p(x_1)$ is then

updated with the measurement y_1 using the update step. The result is the conditional density $p(x_1)|y_{1:1})$, being $y_{1:1} = (y(t'_1))$. The procedure is repeated until the final assimilation time t'_{nz}. At the end, the pdf will be conditioned to the whole set of measurements available in the interval (t_0, t'_{nz}).

The analytical solution for the filtering problem exists for linear models and is the well known Kalman filter (KF). In linear models the forecast and update pdfs are Gaussian, and thus they can be simply described by their first and second order moments [Arulampalam et al., 2002]. The KF method recursively computes the exact mean and variance of these distributions. For nonlinear problems, however, it is necessary to use nonlinear filters. Nonlinear dynamics, in fact, cannot be fully described by the mean and the covariance, they need further moments. Two nonlinear assimilation methods are ensemble Kalman filter (EnKF) or the sequential importance resampling (SIR) filter. These methods are described hereinafter.

5.1.2 Ensemble Kalman filter

The EnKF is an assimilation algorithm based on the Monte Carlo (MC) method. An ensemble of model predictions is used to evaluate error covariance information. The ensemble is then updated with the observations through the KF. Each state realisation in the ensemble is propagated in time through the model dynamics (5.1) [Das and Mohanty, 2006]. Then, the update step is essentially based on KF but with a different calculation of the error covariance matrix. KF computes and propagates in time the error covariance matrix, which implies a high computational effort, especially in case of large nonlinear problems. In EnKF, on the contrary, the error covariance matrix is calculated on the basis of the ensemble of realisations [Chen and Zhang, 2006]. The EnKF method, similarly to KF is defined by three principal components:

- the state vector $x(t)$, which collects model parameters and dependent variables;
- the forecast model, such as Equation (5.1);
- the observation model that links the available observations with the state vector.

Each member of the ensemble has a state vector x_j^k, where k is the realisation index, $k = 1, \ldots, N$. The forecast model is performed on each ensemble member using Equation 5.1:

$$x_j^{k,f} = \mathcal{A}\left(x_j^{k,a}, \rho_j, \alpha^k, t'_j, t'_{j-1}\right) \tag{5.5}$$

where the postscript k indicates the ensemble member, α^k is the vector of parameters associated to the $k-th$ realisation, the superscript f and a indicate the realization state

value after the forecast and the assimilation respectively, j indicates the assimilation time t'_j, $\mathcal{A}$ is the nonlinear operator, the vector x_j^k collects the discretized system state variables defined at time t'_j and ρ represents a realization of the model error.

An observation vector y^k for each ensemble member is computed at each assimilation time t_j:

$$y_j^k = \mathcal{H}(x_j^{k,f}, \omega_j) \tag{5.6}$$

where y_j^k represents the observation data simulated with realisation k and ω_j is a random realisation of the measurement error. In standard EnKF framework, Equation (5.6) is linear ($\mathcal{H}$ is a matrix, H) with an additive noise ω modeled as Gaussian with 0 mean and variance R_j: $y_j^k = Hx_j^{k,f} + w_j$. The assimilation procedure starts when the first set of observation data is available. The Kalman gain, which is necessary for the update step, is calculated by means of the following equations:

$$\left\langle x_j^f \right\rangle \approx \frac{1}{N} \sum_{k=1}^{N} x_j^{k,f} \tag{5.7}$$

$$P_j^f \approx \frac{1}{N-1} \sum_{k=1}^{N} (x_j^{k,f} - \left\langle x_j^f \right\rangle)(x_j^{k,f} - \left\langle x_j^f \right\rangle \tag{5.8}$$

$$K_j = P_j^f H^T (H P^f(t'_j) H^T + R_j)^{-1} \tag{5.9}$$

where $\left\langle x_j^f \right\rangle$ is the mean state vector, P_j^f is the state covariance matrix, N is the dimension of the ensemble, K_j is the Kalman gain, the superscript T indicates transpose, H is the linear observation operator of Equation 5.6, R_j is the covariance matrix of the observation error. The ensemble is updated using the formulas of the Kalman gain, i.e.:

$$x_j^{k,a} = x_j^{k,f} + K_j(y_j^k - \mathcal{H}x_j^{k,f}) \tag{5.10}$$

where $x_j^{k,a}$ is the updated ensemble member. The EnKF method is widely used since it is conceptually simple and easy to be implemented [Evensen, 2003]. In addition, even if the forecast realisations are far from the measurements, EnKF method leads each ensemble realisation toward the observations [Pasetto et al., 2012]. A principal assumption in EnKF is that model errors are Gaussian distributed [Clark et al., 2008]. The Gaussian pdf is fully characterised by the first and second order moments. The method uses the Kalman gain even if the Gaussian nature of forecast and measurement errors pdfs is not assessed. However, in nonlinear systems the Gaussian approximation may introduce errors in the update step and higher order DA methods would be necessary to properly manage the assimilation. In fact, the simply update using only the second order moment may lead to

divergence and instability in nonlinear systems [Moradkhani and Hsu, 2005]. In addition, the physical consistency of the updated ensemble is not verified with EnKF, with the meaning that the updated state variables of the model may be not consistent with the dynamics of the model. The consequences are numerical difficulties and inaccuracies in the computation of the system solution in the time step following the update [Pasetto et al., 2012].

In CATHY model, EnKF algorithm is implemented through the following three equations [Camporese et al., 2009].

Model equation

$$x(t) = A(x(\tau), \alpha, u(t), t, \tau) \quad for \quad t > \tau \geq 0 \tag{5.11}$$

$$x(0) = x_0(\alpha) \tag{5.12}$$

where the vectors $x(t)$ collect the hydrologic unknowns of the problem, in this case the pressure head at each node of the subsurface grid and inflow and outflow discharge at each cell of the surface discretization. The vector α collects the time-invariant set of soil parameters. The vector $u(t)$ collects the time-dependent atmospheric forcing variables. The initial state of the system is $x_0(\alpha)$. The state at time τ is related to the state at time t by means of the nonlinear operator A. The latter operator is linked to the numerical solution of the Richards equation, as described in Equation (4.13).

Measurement equation

$$y_j = \mathcal{H}_j(x_j, \omega_j, t'_j) \quad \text{for} \quad j = 1, ..., nz \tag{5.13}$$

where the vector y_j collects the measurements of volumetric water content, H is the nonlinear operator between observations and system states, ω_i is a random term taking into account the measurement errors.

Update equation

$$x_j^{k,a} = x_j^{k,f} + K_j \left(y_j - \mathcal{H}(x_j^{k,f}, \omega_j^k, t'_j) \right) \tag{5.14}$$

where the vectors $x_j^{k,a}$ collects the update system variables, the superscript k indicates realisation of the ensemble, the term K_j represents the Kalman gain, dependent on the system state and the measurement error covariance matrices [Camporese et al., 2009].

The nominal mean values of soil parameters, initial conditions and atmospheric forcing are perturbed in order to obtain the ensemble components. The ensemble of measurements is realised in an analogous way, to simulate measurement errors.

5.1.3 Sequential importance resampling

SIR is a DA technique belonging to the family of particle filters. Particle filter implement recursive Bayesian filter by MC simulations. In this approach, the required posterior density function is represented by a set of random samples with associated weights. The samples and the respective weights are the basis for the computation of the estimated ensemble [Arulampalam et al., 2002]. Particle filters can be applied also to non-linear and non-Gaussian state-space models. Each particle represents a set of spatially distributed state variables with an associated weight that somehow describes the probability of that particular state. The spatial distribution of the realisation is not changed during the update step. In fact, the update is realised on the particle weights, rather than on state variables directly [Salamon and Feyen, 2009]. Basically, the filtering pdf is approximated by a set of random samples x_j^k associated to a set of weights π_j^k. The importance sampling technique is the origin of the weighted representation of the filtering pdf, which is approximated as follows:

$$p(x_j|y_{1:j}) \approx \sum_{k=1}^{N} \pi_j^k \delta(x_j - x_j^k) \tag{5.15}$$

where δ is the Dirac delta function. In SIR procedure, at the beginning of the simulation, uniform weights are assigned to each independent realisation of the initial pdf. Being the pdf $p(x_0)$, and being N the number of independent realisations of $p(x_0)$, the initial weights are $\pi_0^k = \frac{1}{N}$. In the forecast step, each realisation is propagated in time using the model dynamic equations (5.1.2) without changing the weights. For the update step, new weights are recursively computed with the following equation:

$$\widetilde{\pi}_j^k = \pi_{j-1}^k \; p(y_j|x_j) \tag{5.16}$$

$$\pi_j^k = \frac{\widetilde{\pi}_j^k}{\Sigma_{i=1}^N \widetilde{\pi}_j^k} \tag{5.17}$$

where $p(y_j|x_j)$ is the likelihood function of the observations, which can be obtained from the pdf of the measurement error and Equation (5.13). The deterioration of the ensemble is a possible problem, due to the degeneration of the weights to negligible values after few steps. In this case, the state estimation may be inaccurate and the propagation of particles with negligible importance is a waste of computational effort. This problem

frequently happen when the space dimension is big with respect to the sample size, or when the variance of the pdf of the observations is small [Arulampalam et al., 2002]. The number of realisations effectively contributing to the determination of the empirical pdf is estimated by the mean effective sample size $\widetilde{N}_{eff}(\leq N)$:

$$\widetilde{N}_{eff} = \frac{1}{\Sigma_{j=1}^{N}(\pi_{t'_j}^{k})^2} \tag{5.18}$$

Typically, deterioration occurs for small $\widetilde{N}_{eff}$. In order to avoid this chance, a resampling step is added after the update step. This step consists of removing the particles with small weights and focus on particles with more relevant weights. A new set of samples $(x_j^{k*})_{k=1}^{N}$ is generated by resampling the filtering pdf (5.15). The new samples are typically copies of the realisations with larger weights, and thus are closer to the measurements. The resampling avoid the degeneracy of the ensemble, but on the other hand it may induce a loss in the diversity among the particles. The ones with higher weights, in fact, are statistically selected many times. The resultant sample will have many repeated points. This problem is enhanced if the noise of the model is small.

5.1.4 Parameter estimation

The described DA techniques are used to correct the pdf of the system state with the available data, in order to reduce the uncertainties on the state estimation during the forecast step. DA methods are used also to approach inverse problems, where also the pdf of the model parameters has to be determined. The traditional inverse problem aims to the determination of the optimal parameter set which minimizes the error with the observations [Yeh, 1986]. Here, using a probabilistic approach, we are interested in approximating the pdf of the parameters.

In EnKF the parameter estimation is performed in the update step, by correcting simultaneously both model state variables and parameters. The used approach is called state augmentation, in which state and parameter vectors are combined in a unique state vector [Moradkhani et al., 2005]. The parameters then are updated at each assimilation step together with the state variables.

In SIR, the resampling step is used to obtain a new distribution of the parameter values. In the resampling step is not advantageous to simply copy the most probable parameters, otherwise, after the resample, most of the parameter values $\{\widehat{\alpha}_t^{(i)}\}, i = 1, ..., N$ are equal. In this case, the number of different parameter values is equal to the number of realisations with non-negligible weights. If the consecutive update is realised with

these values, the ensemble collapses on one parameter value after few updates, since the parameters do not change in the forecast step. Thus, the posterior distribution is not adequately explored and parameter estimations may be erroneous. The perturbation of the duplicated parameters may solve this problem. Pasetto et al. [2014] propose to sample new parameter values for the duplicated realisations in the resample step. The analytical form of the parameter distribution is maintained during the update step, while its expected value and variance are corrected according to the updated realisations weights. Being the initial distribution lognormal, it is fully described by the first and second order moments. Prior and resampled parameters, respectively $\{\alpha_j^{(k-)}\}$ and $\{\widehat{\alpha}_j^{(k)}\}$, are the basis for the update of the expected mean value μ_{α_j} and coefficient of variation cv_{α_j} at each assimilation step. The mean of the updated distribution is imposed to be equal to the weighted mean of the parameters, while the coefficient of variation is imposed to be equal to the maximum coefficient of variation between forecasted and updated parameters:

$$\mu_{\alpha_j} = E\big(\widehat{\alpha}_j^{(k)}\big) \tag{5.19}$$

$$cv_{\alpha_j} = s \cdot max\big(cv_{\alpha_j}^{(-)}, cv_{\widehat{\alpha}_j}\big) \tag{5.20}$$

where, in the second equation, s is a coefficient which reduces gradually the variance of the distribution. The use of the maximum value for the coefficient of variation prevents the fast collapse of the filter when few realisations are resampled.

The convergence of posterior parameters pdf, in this way, occurs after several updates. When the repetition of the assimilation interval does not give any improvement of the ensemble, then the process is stopped.

The parameters distribution computed with this procedure may change at every assimilation time during the simulation. However, the model parameter α are typically time independent, and should not change in time. For this reason it is useful to repeat the assimilation procedure until that the parameter distribution converges. Thus Pasetto et al. [2014] introduces an external loop that repeats the DA simulation. At each new external iteration, i.e., the repetition of the filtering process, the mean and variance of the parameter prior distribution is updated by the average values of the updated mean and variance of the parameters computed in the previous filtering process:

$$\mu_{\alpha_0}^{l+1} = \frac{1}{n_z}\sum_{j=1}^{n_z}\mu_{\alpha_j}^{l} \tag{5.21}$$

$$cv_{\lambda_0}^{l+1} = \frac{1}{n_z}\sum_{j=1}^{n_z} cv_{\alpha_j}^{l} \tag{5.22}$$

where n_z is the number of updates in each sampling DA cycle and l is the index of the external iterations. In practice the prior distribution at the $(l+1)^{th}$ filtering process is the *mean posterior distribution* at the l_{th} filtering process. This procedure reduces the effect of the initial bias on the parameter estimation.

5.2 Data assimilation on LEO experiment - features

In the DA process fulfilled in this thesis work, the hillslope is represented as a three-dimensional domain 1 m deep with a DEM as that reported in Figure 3.2. The two side boundaries, the bottom of the hillslope and the upslope boundary are impermeable, while the downslope boundary is modeled as a seepage face boundary condition. Given an initial spatial distribution of soil water content θ, rainfall and evaporation boundary conditions are assigned at the surface as the one estimated during the experiment. The soil water content responds according to this forcing term and to the soil hydraulic properties. The DA procedure is aimed to infer the spatial distribution of soil hydraulic conductivity of the LEO hillslope. The estimated K_{sat} is then used in order to simulate the behaviour of the hillslope. Both EnKF and SIR algorithms are applied. The problem of the physical consistency of the EnKF updated states becomes critical in the simulation of LEO dynamics, in particular if the considered domain is not completely saturated. In this situation, in fact, there are strong nonlinearities which contrast with the Gaussian approximation. In CATHY model, the back-stepping is forced, in the time steps immediately after an EnKF update, in order to ease the convergence [Pasetto et al., 2012].

In this thesis work, we propose to overcome this problem by considering a different temporal path for the assimilation updates. According to this new path, the simulations restart from the initial time after every update step, i.e., the updated parameter pdf is used as prior distribution and the process is restarted from time $t = 0$. The new update step occurs at the following assimilation time. Again the simulation is stopped at the update time, the updated parameter pdf is considered as prior pdf at time $t = 0$ and the update time is adjourned to the following assimilation time. Figure 5.1 shows the iterative path. This procedure is followed until the end of the simulation.

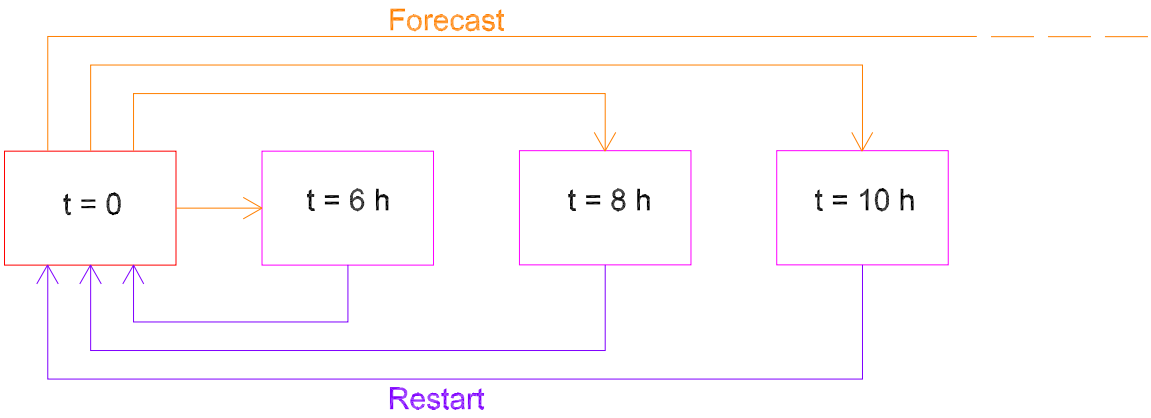


Figure 5.1: Temporal path of the assimilation process

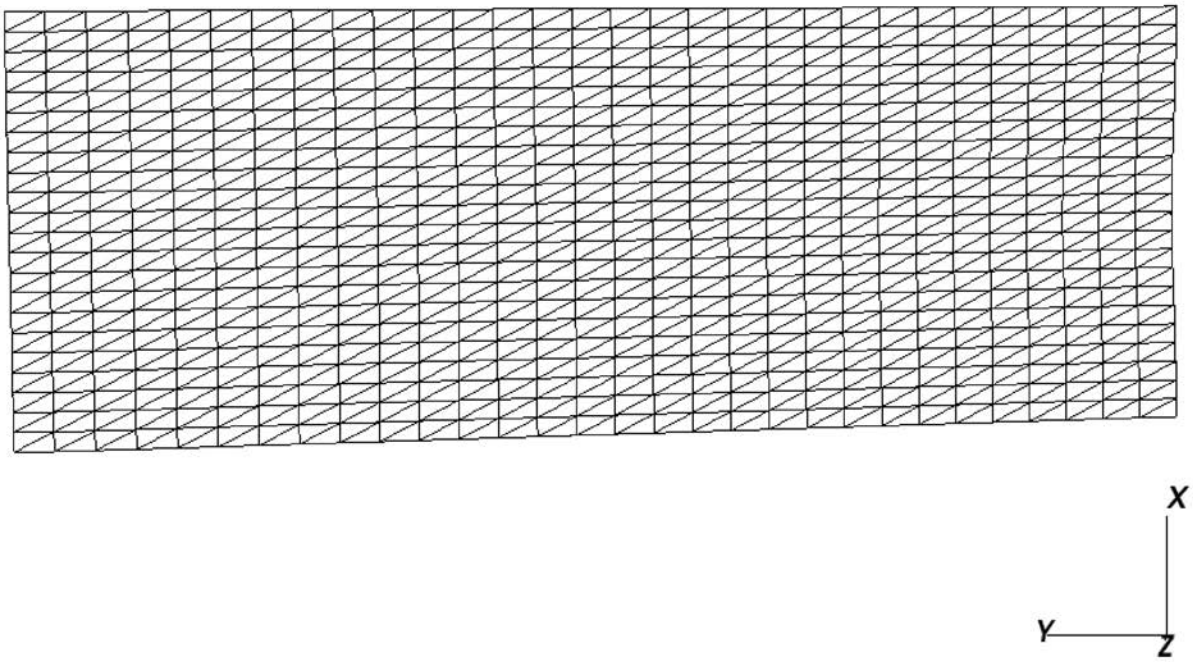
Chapter 6

Estimation of the hydraulic conductivity field from distributed soil water content observations at LEO

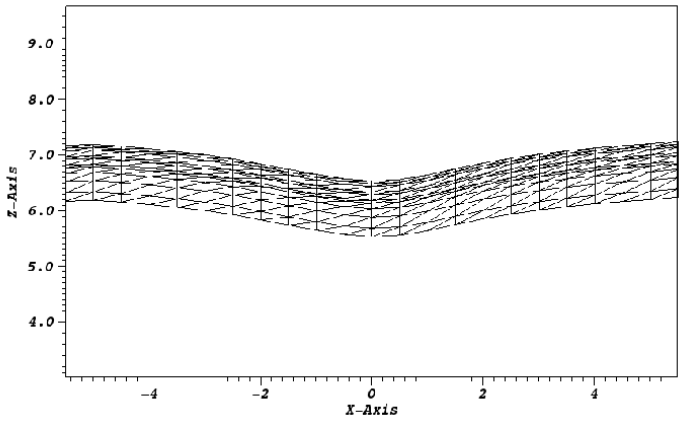
The first experiment at LEO-biosphere (see Chapter 3) is the focus of this thesis work. The gap between the actual hydrologic response of the hillslope and the results of the initial numerical simulations shows that the assumption of homogeneity is not verified. The most accredited hypothesis is that the water flow may have changed the internal particle distribution, inducing a variation in the value of saturated hydraulic conductivity in different zones of the hillslope. Here, a data assimilation process is employed to reconstruct the value of K_{sat} from the observations of soil water content obtained at the 496 sensors and to try to find the spatial distribution that can reproduce, through the model, the actual hydrologic response of the hillslope.

6.1 Problem setting

The LEO-1 hillslope is the domain of interest. As told in 3 it is a sloping hillslope 11.25 m wide and 29.6 m long. Soil thickness is equal to 1 m. The average declination is equal to 10°. The transverse section of the hillslope converges toward the center with an average angle of 7°, forming a v-shaped profile. Many samplers and sensors have been installed in the hillslope. In particular, there are 496 sensors for volumetric water content. The data collected by these sensors are used in this DA process. The sensors are placed at depths equal to 0.05, 0.2, 0.35, 0.5, 0.85 and 1 m (see Figure 3.2).



(a) *Horizontal grid.*



(b) *Vertical Section of the mesh.*

Figure 6.1: Non-uniform grid representation

Table 6.1: Depth of nodes along z direction

# node	1	2	3	4	5	6	7	8	9	10	11
depth [m]	0	0.05	0.1	0.2	0.25	0.35	0.4	0.5	0.65	0.85	1

The hillslope is discretized by means of a three-dimensional grid. This grid is obtained by replicating a surface triangulation along the vertical direction in order to define the tetrahedral elements needed for the solution of the Richards equation (see Subsection 4.4). The surface grid, shown in Figure 6.1, has uniform node spacing of 0.5 and 1 m in the x and y direction, respectively. This choice derives from the intent to have a node in correspondence of each sensor, allowing a direct comparison between the measurements of a sensor and the values computed by the model without any interpolation. Its vertical replications form 10 computational layers, spaced as shown in Table 6.1.

Since the considered problem is nonlinear, in order to achieve the convergence of the DA algorithm for the identification problem an iterative scheme is used, as described in Section 5.2. The first assimilation update is fixed 6 h after the beginning of the measurements, i.e. 4 h after the beginning of the rainfall. The assimilation process ends 36 h after the starting of data collection, with update every 2 h. Since the variance of the ensemble reduces as the simulation proceeds, in order to avoid a too fast reduction a slight increase of variance is assigned at the beginning of each external DA iteration. Time step size in CATHY is adapted to the convergence behaviour of the nonlinear solver, which can vary with each realisation(see Subsection 4.4). A minimum and maximum value for time step length is fixed. If the convergence is not reached even with the smallest time step size, the correspondent realisation is discarded. A check in the model makes the entire simulation stop when the number of active realisations fall below a fixed threshold.
Since the DA algorithm is essentially an iterative procedure, we need an initial guess for the parameters to be identified. For this, we start from the M4 scenario of Niu et al. [2014] (see 3.3). We consider as constant the soil porosity θ_{sat}, the Van Genuchten fitting parameter n and the elastic storage coefficient S_s, and try to identify the spatial distribution of saturated hydraulic conductivity K_{sat}.

During the assimilation, K_{sat} is allowed to vary between 7.0e-06 and 7.0e-04 m/s. Table 6.2 lists all the assimilation characteristics.

The hillslope may be considered as partially or fully heterogeneous, according on whether hydraulic conductivity value is assigned by zone (a group of neighbouring cells) or varies at each point of the domain. In the first case, the spatial distribution of K_{sat} is defined by ‘zones‘ (subsets of connected tetrahedra) of constant K_{sat}. In the second case

Table 6.2: General assimilation features

Feature	Value
K_{mean} [m/s]	1.0e-04
K_{min} [m/s]	7.0e-06
K_{max} [m/s]	7.0e-04
θ_{sat} $[m^3/m^3]$	0.37
n [-]	2.26
Number of realisations	50
Number of updates	16
Time of first update [h]	6
Step length [h]	2
Time of last update [h]	36
Time step for non linear solver [s]	0.1
Minimum time step for nonlinear solver [s]	0.05
Maximum time step for nonlinear solver [s]	900

K_{sat} is considered a second order stationary random field with given mean, variance and correlation structure [Dagan, 1989].

At the end of the assimilation process, the final ensemble is given in input to the model CATHY, which simulates the experiment given the identified parameters and the time-dependent atmospheric boundary conditions described in Section 3.2. CATHY model is run also with the initial set of parameters. The resulting values of overland flow, seepage face flow and water storage are compared against the observed values. In both cases, the error on volumetric water content is calculated. The error takes into account the measured and the assimilated values of volumetric water content in the nodes corresponding to the sensors. The error is calculated by means of Equation 6.1:

$$e_i(t_j) = \sqrt{\frac{\sum_{l=1}^{NOBS}\left(\theta_l^{obs}(t_j) - \theta_l^{ass}(t_j)\right)^2}{NOBS}} \qquad \begin{matrix} i = 1, ..., NENS \\ j = 1, ..., nz \end{matrix} \tag{6.1}$$

where $e_i(t_j)$ is the error referred to the $i-th$ realisation at the $j-th$ time of assimilation, $NOBS$ is the number of available measures, θ_l^{obs} is the volumetric water content value measured at the $l-th$ sensor (node) and θ_l^{ass} is the value obtained from the assimilated ensemble correspondent to the same sensor node. $NENS$ is the number of final realizations of the ensemble and T_{max} is the time of last update of the assimilation process. The error is then calculated on every realization belonging to the ensemble. The mean error of the ensemble is calculated and is plotted together with the minimum and maximum error values.

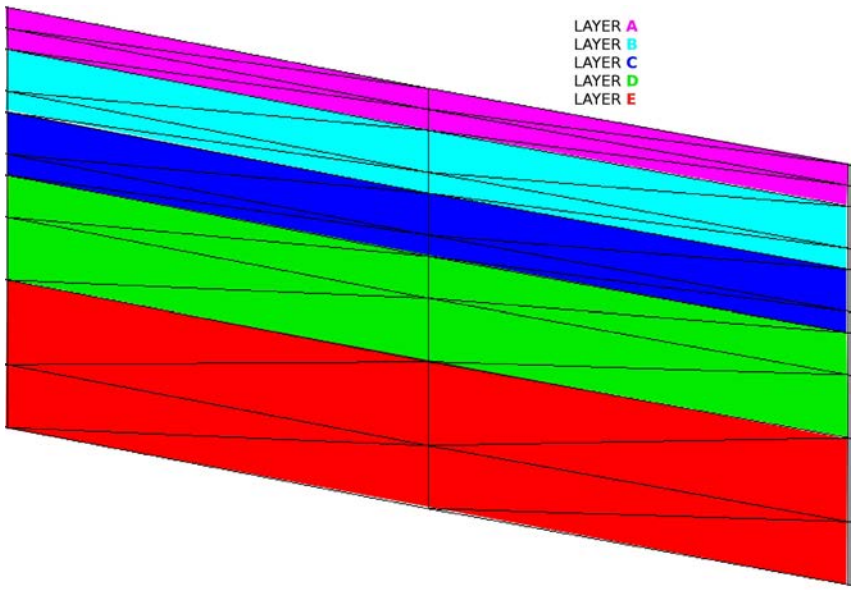


Figure 6.2: Layers in vertical direction

6.2 Partially heterogeneous hillslope

In this case, the spatial distribution of K_{sat} is defined by zones of constant K_{sat}. Initially, the hillslope has been subdivided in 5 layers. The layering has been designed in such a way to have a sensor in the vertical midsection of each layer. In a second time, a further subdivision has been implemented in order to account for the seepage face, i.e., the section where, most probably, the fine particles accumulated during the first LEO experiment. As referred in Section 6.1, the hydrologic response in terms of overland flow, seepage face flow and water storage and the error on volumetric water content are plotted for each attempt. In addition, the evolution of the ensemble throughout the assimilation time steps is plotted for each zone, individually.

6.2.1 Configuration with 5 zones

Figure 6.3 and Figure 6.2 show the horizontal and vertical dissection of the 5 zone configuration. The zones correspoond to the layers in Figure 6.2 and are labeled with capital letters. The chosen assimilation method for this configuration is SIR (see 5.1.3 in Chapter 5). The evolution of the ensemble in time is plotted Figure 6.4. The vertical axis is logarithmic, and each plot represents a zone of the 5 in which the hillslope is subdivided. In layers A and B, the final value is about $1.0e-05m/s$. In the remaining layers the final value is higher than $1.0e-04m/s$.

The assimilated ensemble has been given in input to the model CATHY,producing

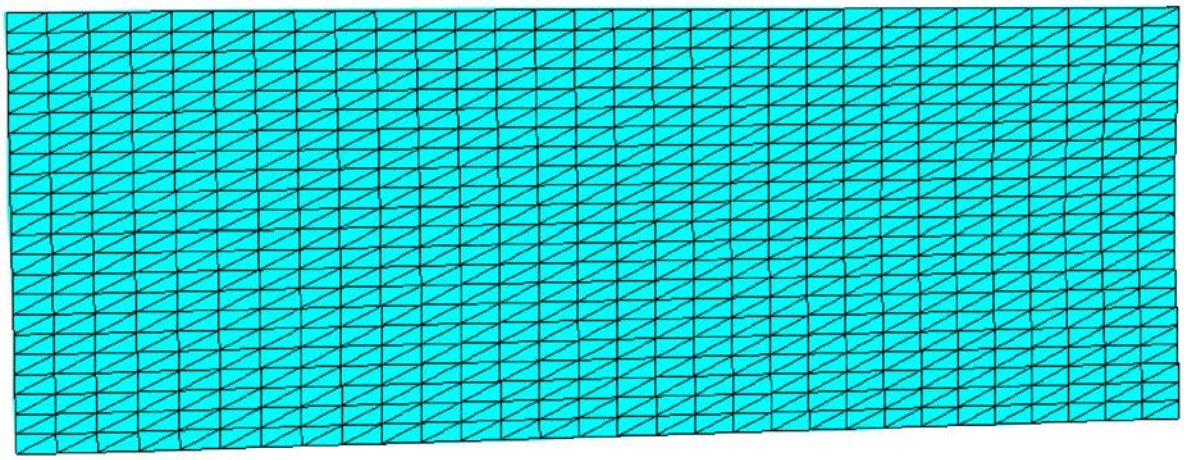


Figure 6.3: Horizontal zonation in the case with 5 zones

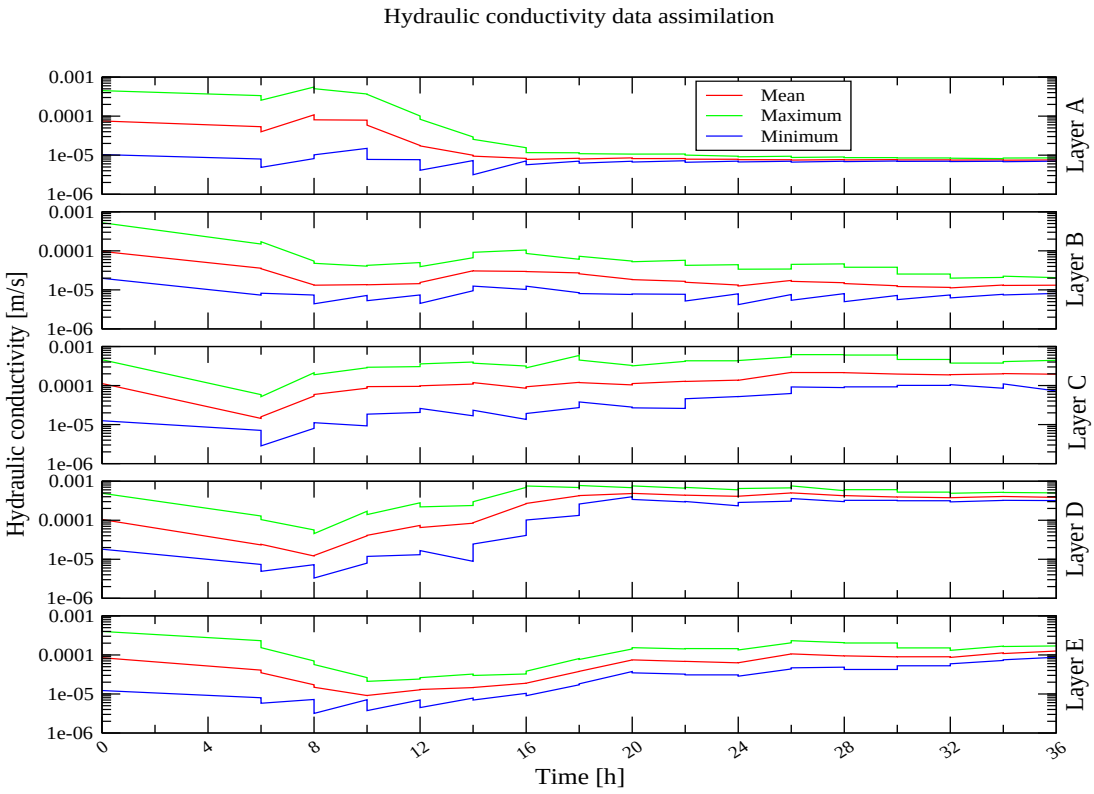


Figure 6.4: Parameter estimation for the five zones configuration - SIR algorithm

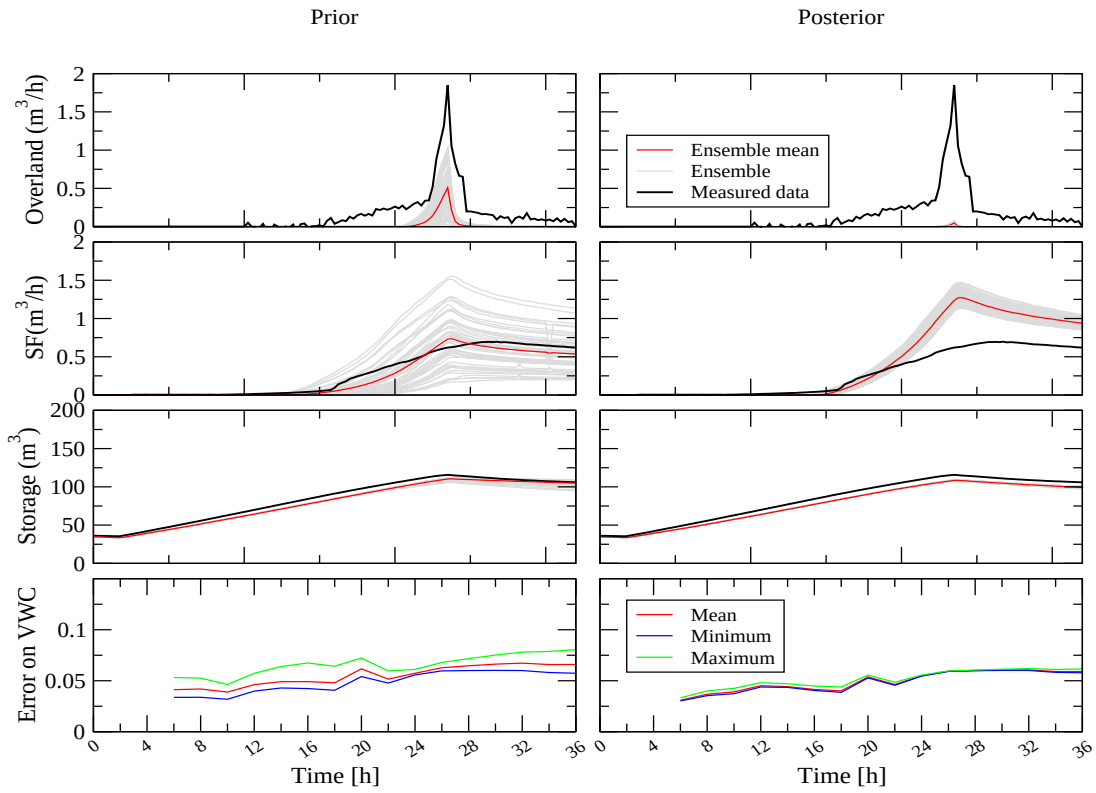


Figure 6.5: Hydrologic response and error on volumetric water content with 5 zones configuration - SIR algorithm

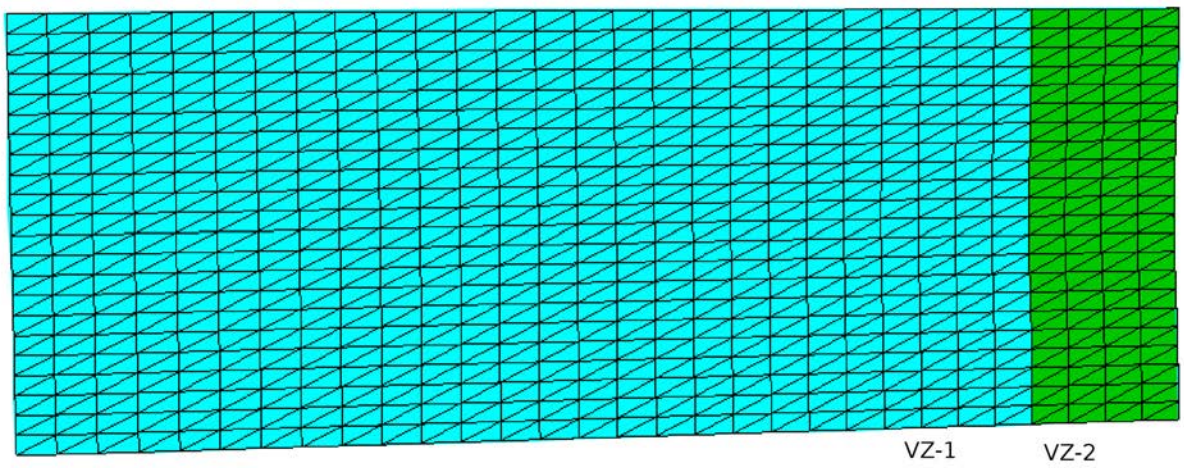


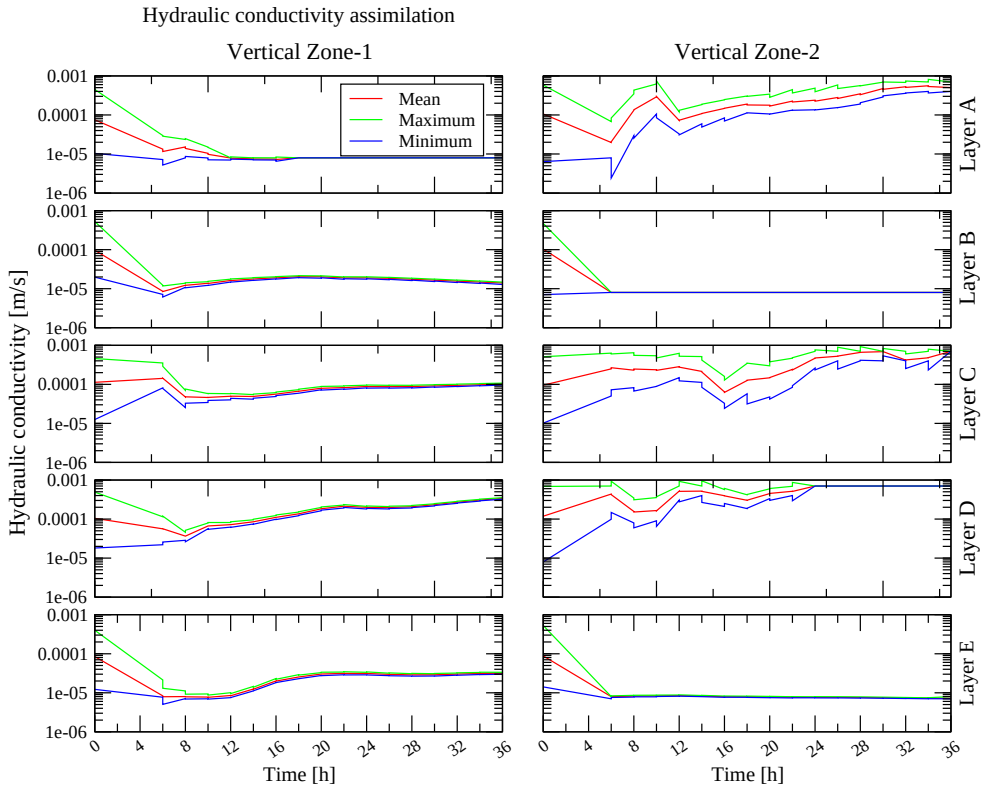
Figure 6.6: Horizontal zonation in the case with 10 zones

forecast results for each realisation of the ensemble. These results are plotted in the same figure to form a band of lines representing the response distribution. In Figure 6.5 the hydrologic response is compared against the measured value, in the right columns of plots. A plot is dedicated to the mean, maximum and minimum error values. The same quantities have been computed for the prior distribution of parameters. These results are plotted in the left column of plots. The overland flow is underestimated with respect to the measured one. It is necessary to remember, however, that the measures of overland flow are affected by large uncertainty. Seepage flow is initially overestimated, while the error is slightly reduced at the end of the assimilation process.

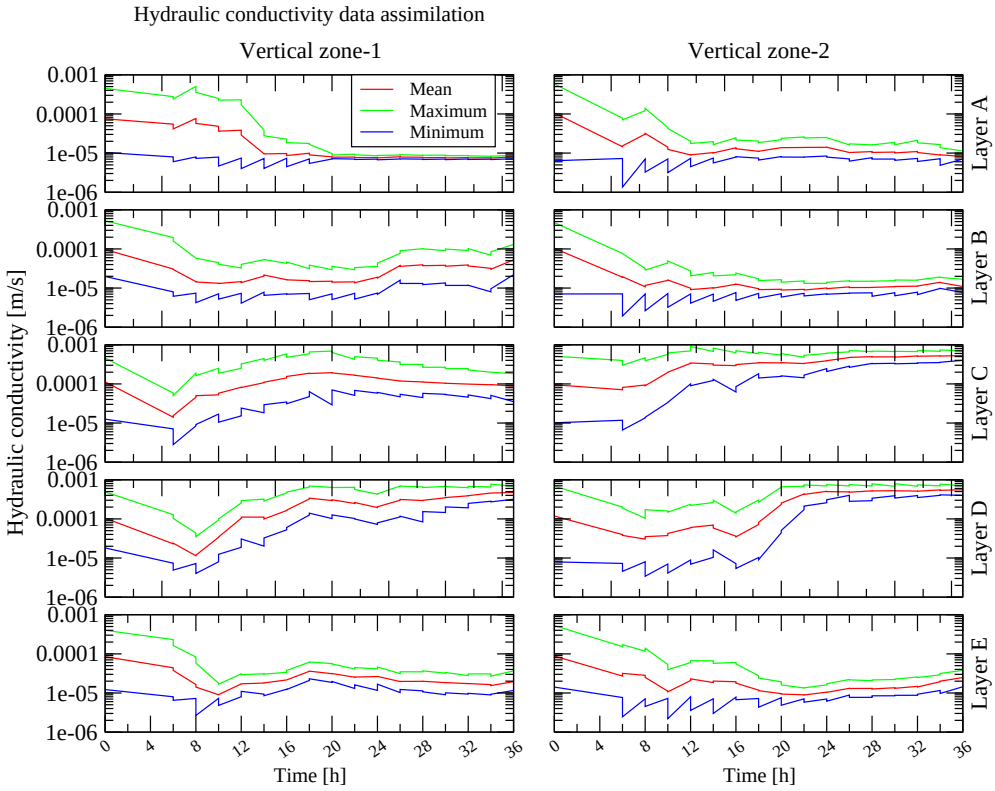
6.2.2 Configuration with 10 zones

In this configuration, a 2 m wide section has been defined in the lower part of the hillslope, as shown in Figure 6.6. The vertical sections are labeled with numbers. The layers are the same of the 5 zone configuration. In this case, both SIR and EnKF methods (see subsections 5.1.2) are used for the assimilation, and the results are compared.

In Figure 6.7 the time evolution of the ensemble in time is plotted in semilogarithmic scale both for EnkF algorithm (top figure) and for the SIR algorithm (bottom figure). Each box of figures represents a different material zone. The left column includes the zones of Section 1 (turquoise area in Figure 6.6) while the right column includes the zones of Section 2 (green area in Figure 6.6).
The plot for the ensemble Kalman filter is now considered. In all the zones except Zones 2-A and 2-C, the maximum, minimum and mean values collapse to a single line. This implies that all the ensemble realisations converge to the same value. In Zone 1-A, the hydraulic

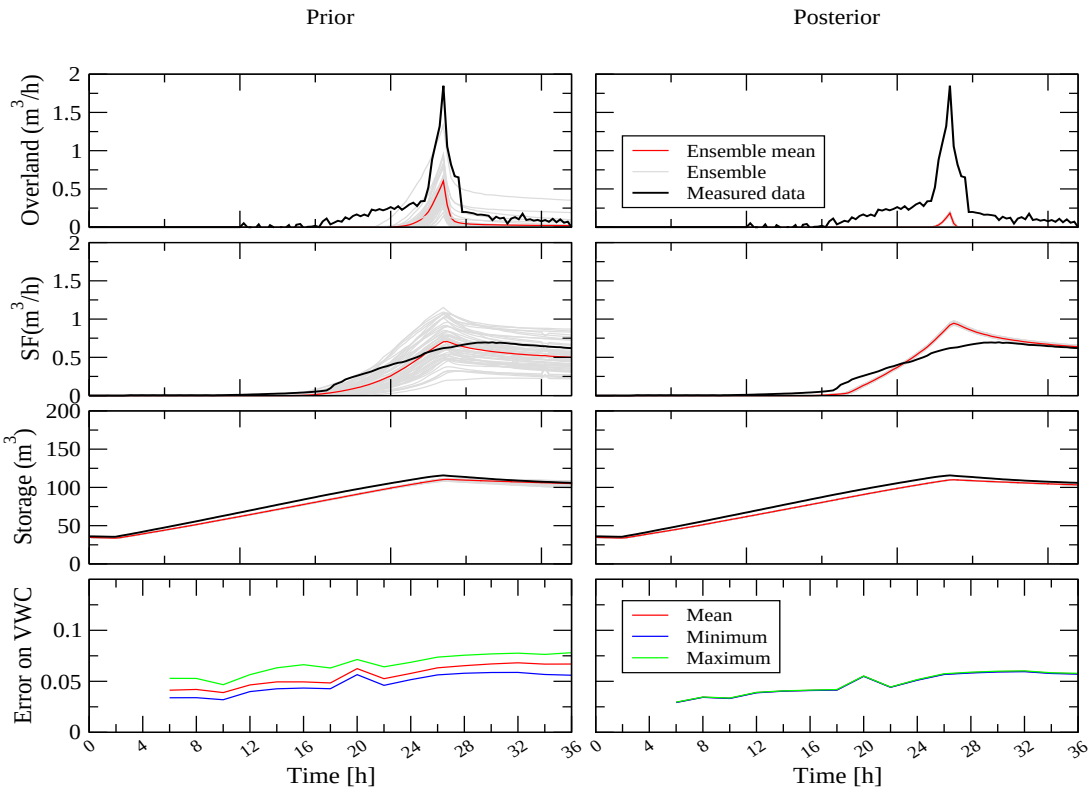


(a) *Assimilation with EnkF algorithm.*

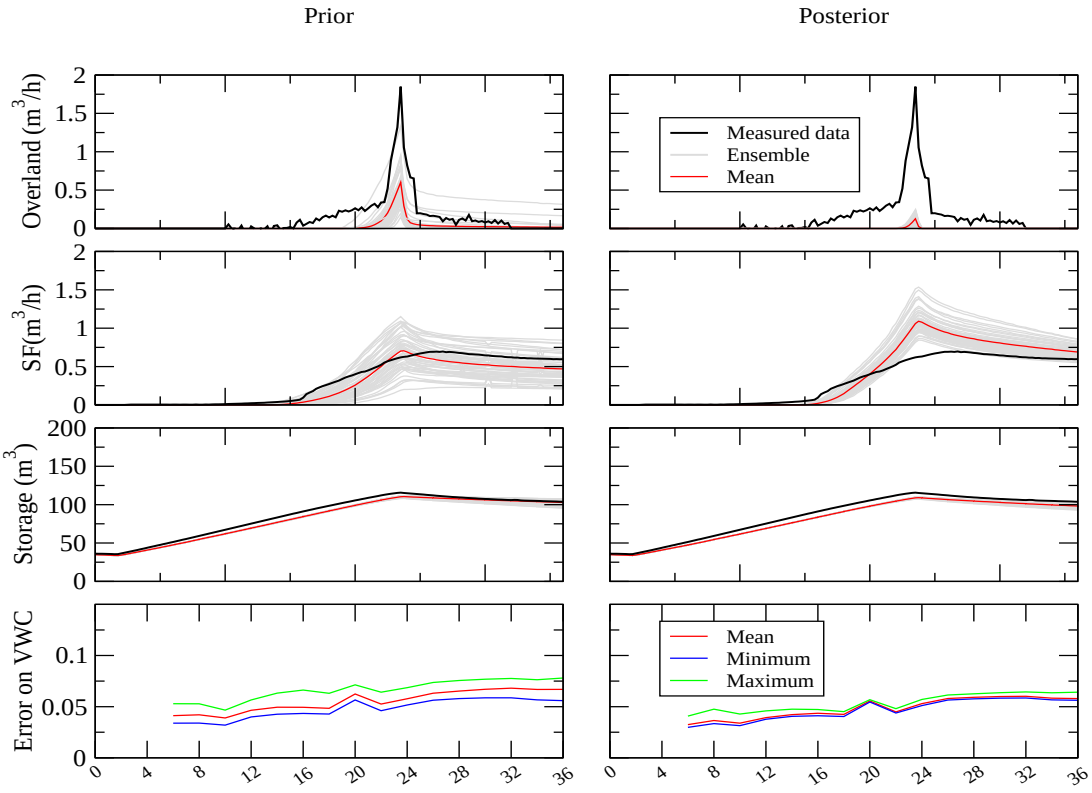


(b) *Assimilation with SIR algorithm.*

Figure 6.7: Parameters estimated in the ten zone configuration - comparison between EnkF and SIR algorithms



(a) *EnKF algorithm*



(b) *SIR algorithm*

Figure 6.8: Hydrological response with prior and posterior estimated parameter values and error on volumetric water content - Comparison between EnKF and SIR algorithms

Table 6.3: Errors on volumetric water content with prior and posterior distribution - discretized case

Attempt	Final mean error	
	Prior distribution	Posterior distribution
10 zones - EnKF	0.055371	0.047519
10 zones - SIR	0.055371	0.048573
5 zones - SIR	0.054857	0.049435

conductivity decreases throughout the assimilation until reaching the minimum value after few steps. In zone 2-A, on the contrary, the hydraulic conductivity increases significantly. In Zones 1-B and 2-B the final value is lower than that of the zones in layers C and D. Looking at the plot for the SIR algorithm, the situation is different. The maximum, minimum and mean values do not collapse to the same value. As iterations progress, the gap between maximum and minimum values shrinks, due to the forced reduction of variance throughout the iterations. Starting from the initial value of $1.0e-04$ m/s, the hydraulic conductivity evolves in different ways. In the surface zones, 1-A and 2-A, the value drops to the lower limit of $7.0e-06$ m/s. The behaviour of the surface layer 2-A is in open contrast with that observed in EnkF assimilation. In the lower zones, both in sections 1 and 2, the hydraulic conductivity increases. The deepest layers of both vertical sections have a smaller K value.

Concerning the hydrologic responses, in the case of EnKF the outputs are concentrated around the mean. This behaviour is a typical phenomenon occurring in EnKF and is called 'filter imbreeding': in many zones the ensemble collapses on a unique value and subsequent assimilation has no effect. Both EnKF and SIR ensembles show an overestimation of measured seepage flow and a underestimation of the overland flow. Looking at the error plots, in both cases the errors of the posterior distribution are smaller than those of the prior distribution, and achieve similar values.

Table 6.3 collects the final errors of all attempts realised with the current configuration. As it was expected, they are smaller than those obtained with the prior distribution, even if the hydrologic response is far from the observed behaviour of the hillslope. Even if the configurations are different, the errors on volumetric water content do not change significantly.

The influence of hydraulic conductivity variation on the estimation of volumetric water content has been further investigated by running CATHY using the K_{sat} distribution identified in the first assimilation of the SIR algorithm. Ten simulations are then run, each one with the K_{sat} value in the zone replaced by K_{sat}=$1.0e-04$ m/s. Figures 6.9, 6.10

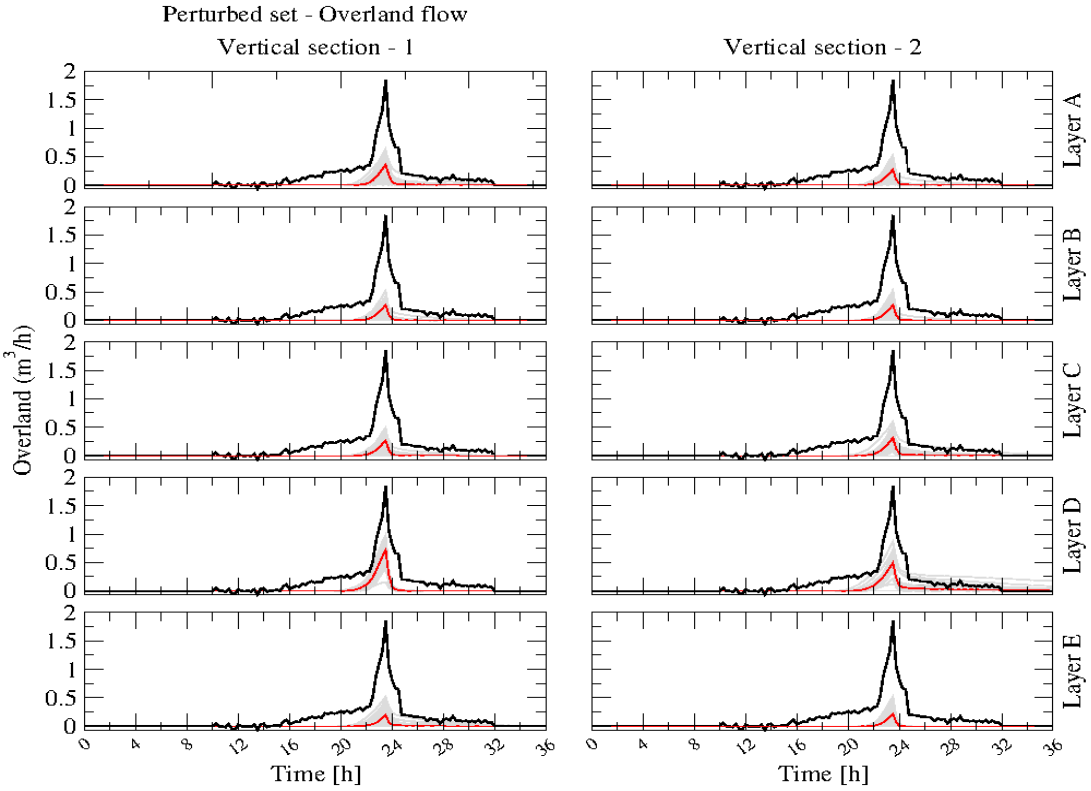


Figure 6.9: Comparison among the overland flow - perturbed sets

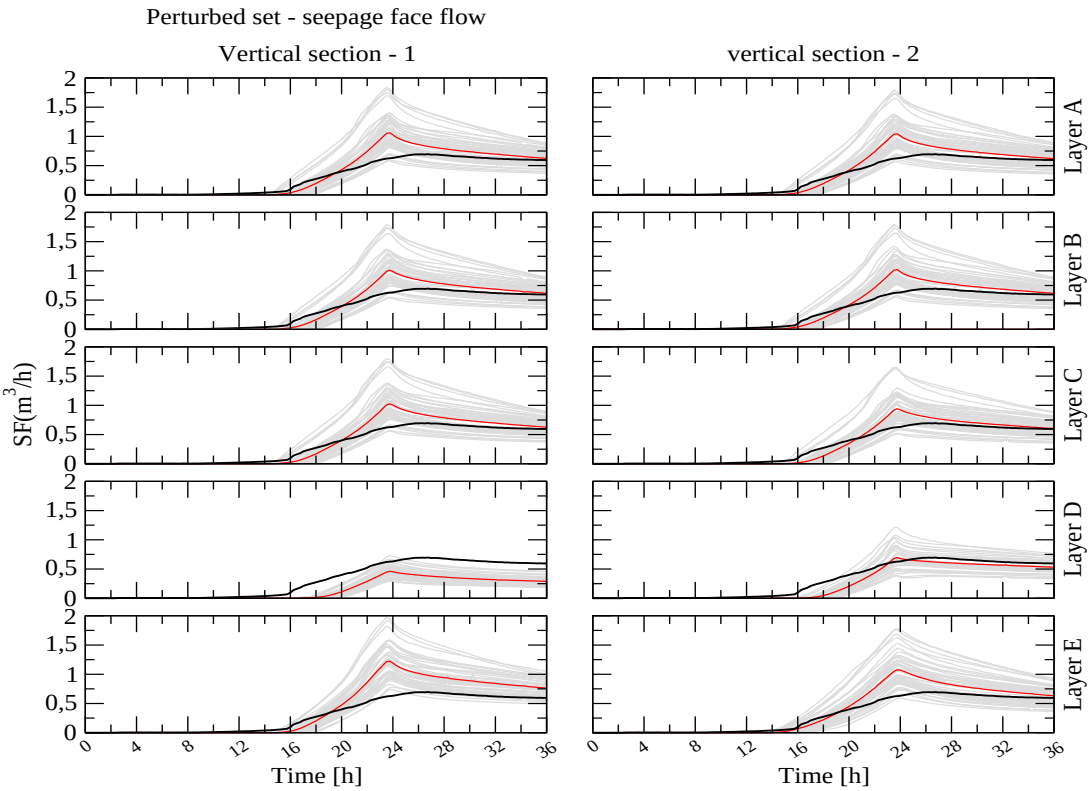


Figure 6.10: Comparison among the seepage face flow - perturbed sets

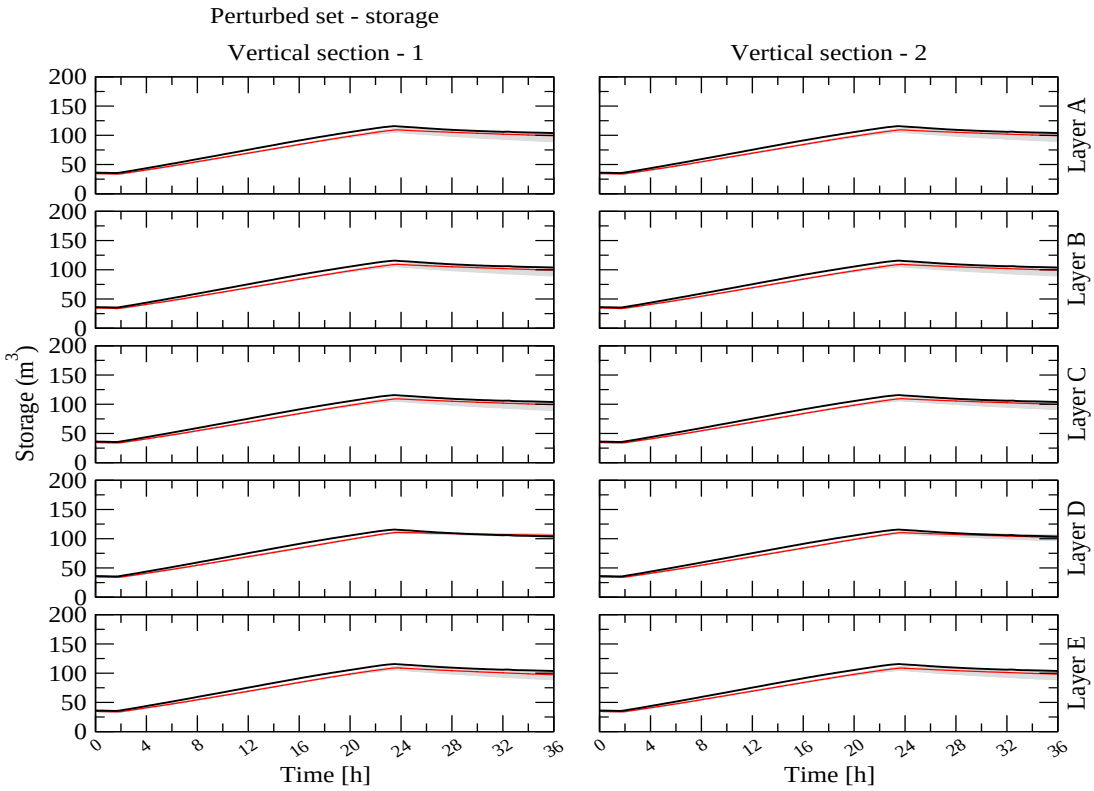


Figure 6.11: Comparison among the water storage - perturbed sets

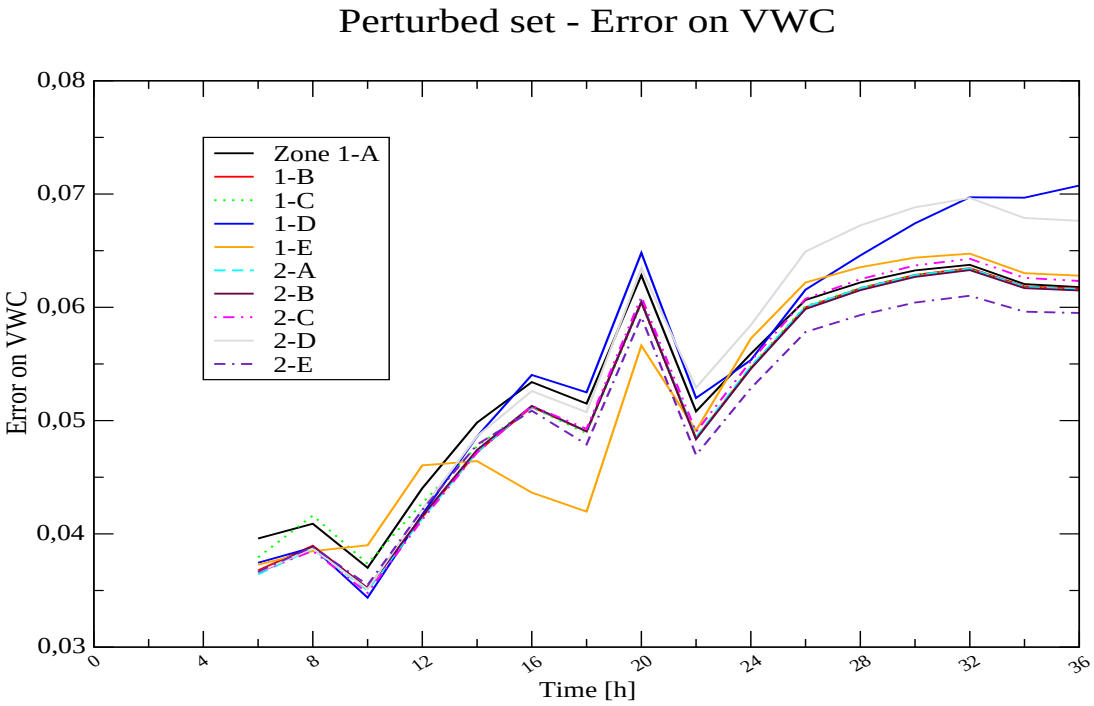


Figure 6.12: Comparison of errors on volumetric water content - perturbed sets

Table 6.4: Errors on volumetric water content with perturbed parameters

Perturbed zone	Error on volumetric water content
1-A	0.053717
1-B	0.052182
1-C	0.052645
1-D	0.055206
1-E	0.052280
2-A	0.052139
2-B	0.052120
2-C	0.052516
2-D	0.055347
2-E	0.051027

and 6.11 show the hydrologic response of the hillslope with the ten differently perturbed sets of hydraulic conductivity. The hydrologic responses change evidently when Zone 1-D and 2-D are perturbed. The overland flow in all the cases is less underestimated than in the hydrologic response obtained with the unperturbed set of parameters. Even if this difference is visible, the errors on volumetric water content are similar in the perturbed and unperturbed cases (Figure 6.12). Table 6.4 summarizes the final average errors on volumetric water content.

In summary, the results of the partially heterogeneous configurations are all similar to each other, even if the configurations and the initial set of parameters change. In addition, the comparison between the hydrologic responses of prior and posterior distributions shows that storage is well reproduced, volumetric water content estimation is improved but overland flow and seepage face flow are not well reproduced. The uncertainty in the overland flow measurements but also the ill-posedness of the problem may be the causes of this result.

6.3 Fully heterogeneous hillslope

In the fully homogeneous hillslope configuration, the grid described in Section 6.1 is uniformly refined in order to better describe the problem and to ease the convergence of the nonlinear problem. In this case, the spatial distribution of K_{sat} is considered a random field with defined mean, variance and exponential correlation structure. The correlation length, which defines the distance beyond which two points are statistically uncorrelated, is different in the vertical and horizontal directions. Two different cases are considered. In the first case, the correlation length is set equal to 4 m in horizontal direction and 0.25

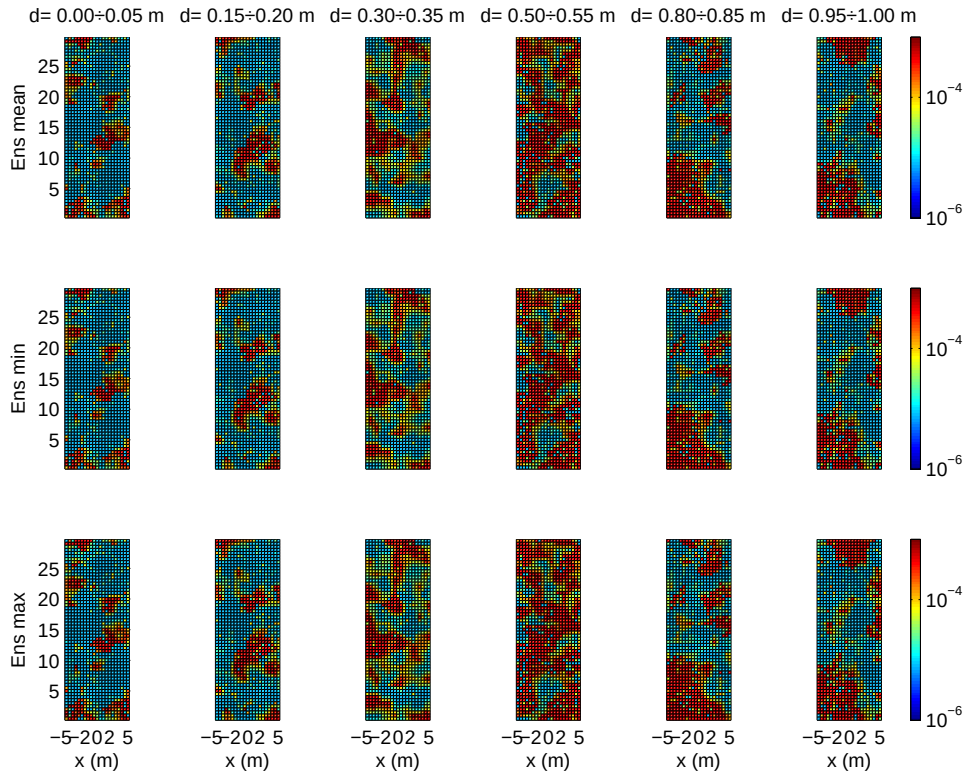


Figure 6.13: Parameter estimation - correlation length = 4 m (EnKF algorithm)

m in vertical direction. In the second case, the horizontal correlation length is equal to 8 m and the vertical correlation length is set equal to 0.5 m. In the fully heterogeneous case, the SIR method is not suitable since it does not account for the spatial correlations. The EnKF method must be chosen notwithstanding its intrinsic limitation related to the Gaussian assumption. The first attempts, realised with the usual iterative assimilation scheme, did not reach convergence. For this reason, the modified temporal path suggested in section 5.2 has been implemented. The initial parameter values assigned to each cell are randomly generated, given mean, variance and correlation length.

Concerning the configuration with a correlation length equal to 4 m, thanks to the iterations the problem converged and reached the end of the simulation. In Figure 6.13 the assimilated ensemble of hydraulic conductivity has been plotted. Due to the large number of cells, only the last ensemble has been plotted on a grid, selecting the layers where the sensors are located. In particular, the mean, maximum and minimum values are plotted. The identified distribution of saturated hydraulic conductivity has a mean that is lower for the surface and the deeper layers, with respect to the intermediate layers. The average hydrologic response is again different from the actual behaviour of the hillslope in terms of overland flow and seepage face flow. The average storage, on the contrary, is well

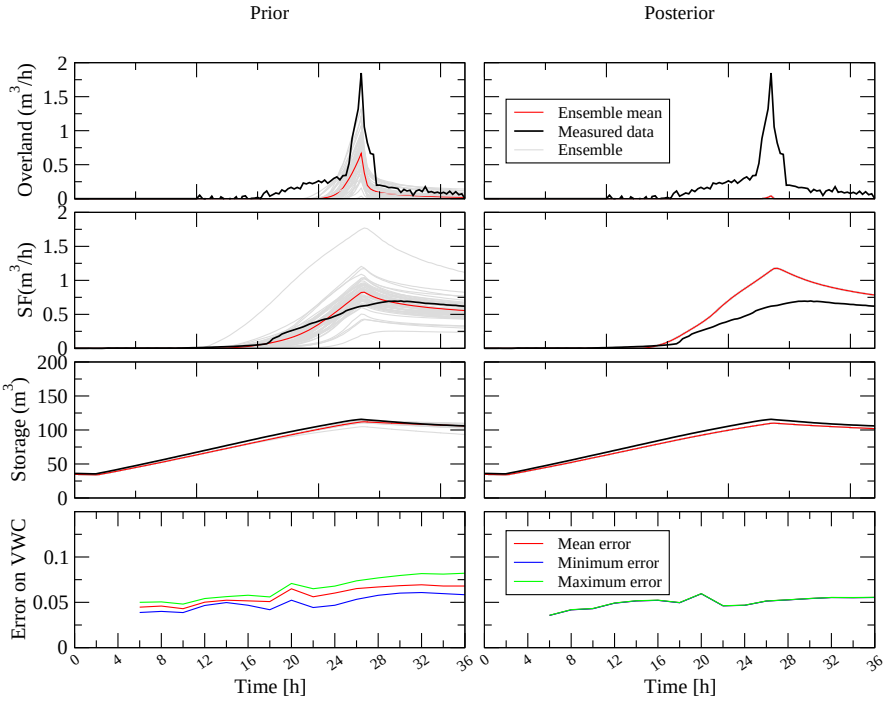


Figure 6.14: Hydrologic response and error on volumetric water content - correlation length = 4 m (EnKF algorithm)

reproduced. The error is still lower than the one calculated with the prior distribution.

The assimilated ensemble of hydraulic conductivity obtained in the configuration with horizontal correlation length equal to 8 m is reported in Figure 6.15. The surface layer presents a low hydraulic conductivity value, which increases with increasing depth. This is a relevant difference compared to the case with horizontal correlation length equal to 4 m. In that case, in fact, the mean saturated hydraulic conductivity was low also in the deeper layers. This difference is particularly evident in the deepest layer. Despite of this difference, the hydrologic responses (see Figure 6.16) are very similar, with negligible overland flow, excessive seepage face flow, and well reproduced storage.

In general, the assimilated ensembles present the lower hydraulic conductivity in the surface layer, and higher values in the deeper layers. Even if the error on volumetric water content are always lower for the posterior distribution with respect to the prior one, the hydrologic response of the system is more distant from the true behaviour when the posterior distribution is used. In addition, the difference among errors referring to the different configurations is small, notwithstanding the difference in grid resolution, spatial configuration and assimilation algorithm. Figure 6.17 compares the mean errors of the various attempts, compared with a reference error calculated for the homogeneous hillslope having hydraulic conductivity equal to $1.0e-04$ m/s. The figure shows that the errors are

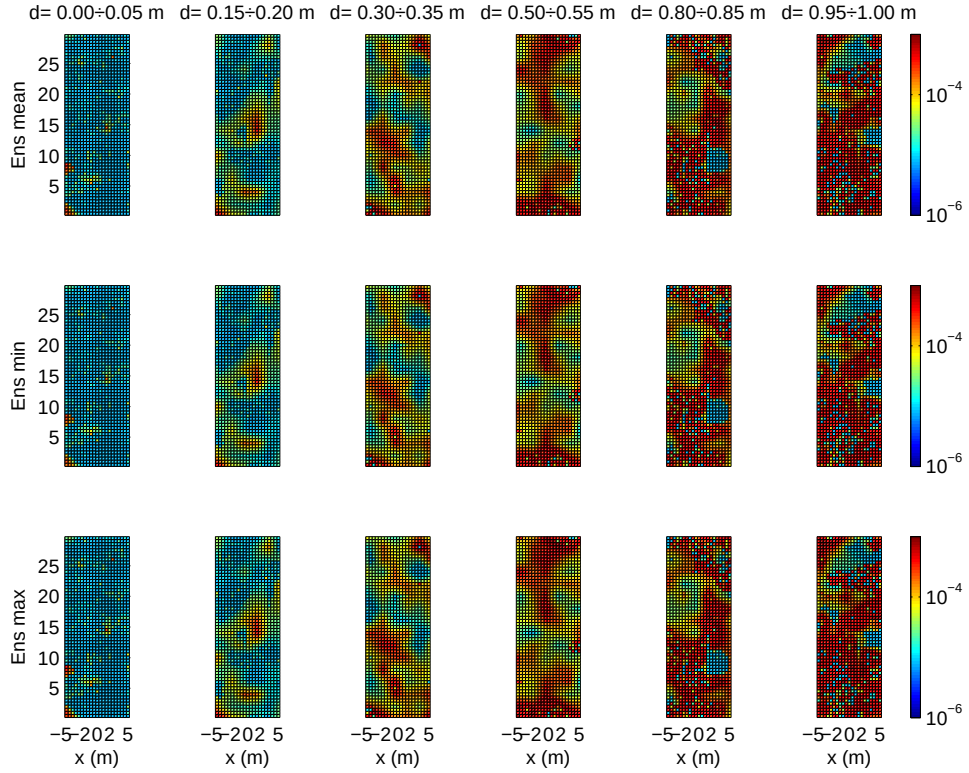


Figure 6.15: Parameter estimation - correlation length = 8 m (EnKF algorithm)

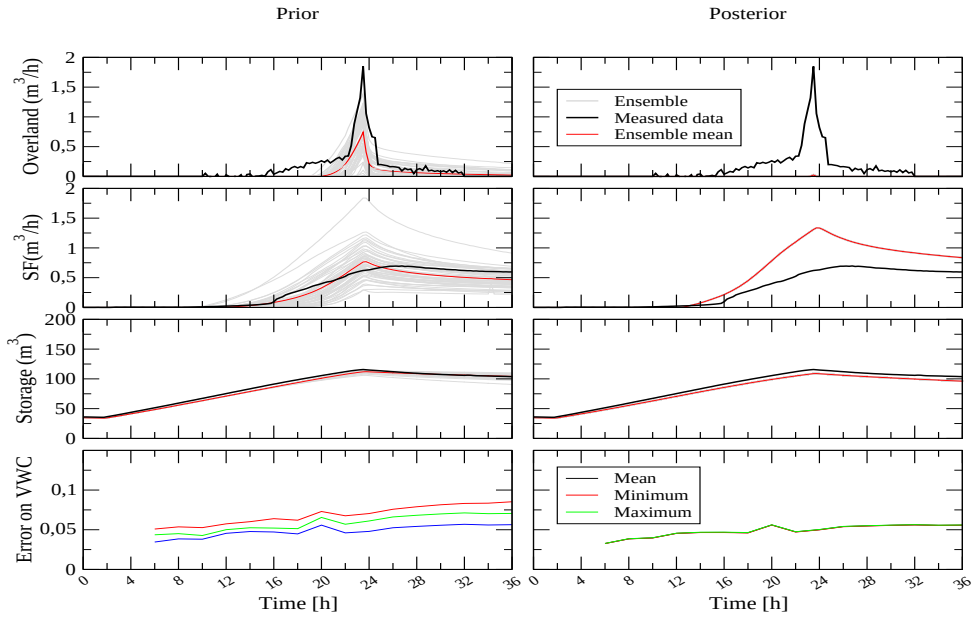


Figure 6.16: Hydrologic response and error on volumetric water content - correlation length = 8 m (EnKF algorithm)

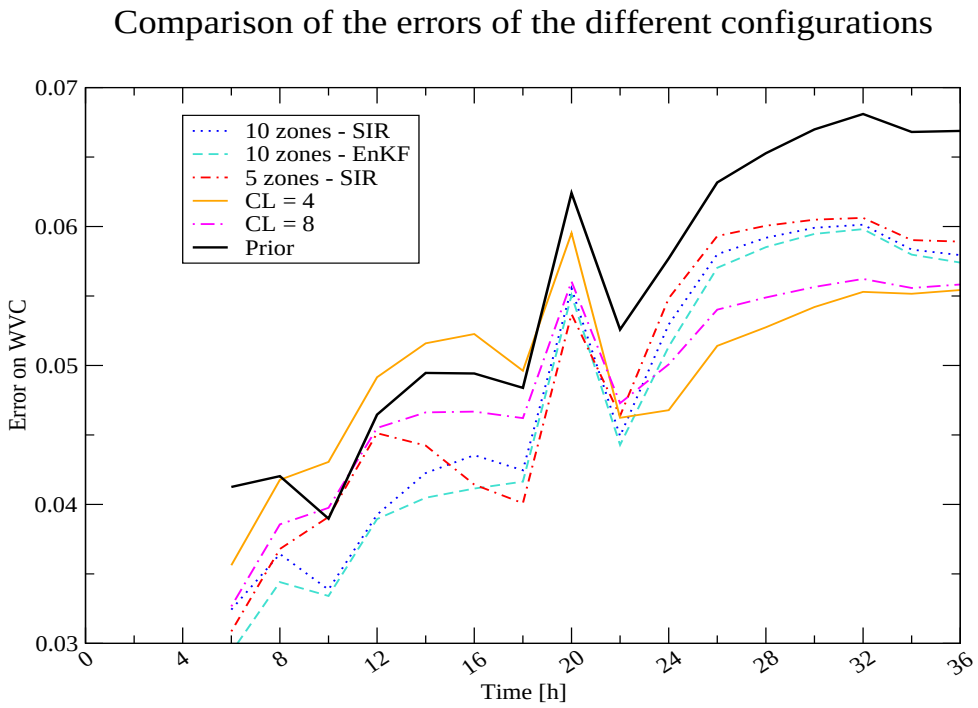


Figure 6.17: Comparison between the errors on volumetric water content of the different attempts

very similar despite of the different spatial configuration and distribution of parameters. The simulations fulfilled with perturbed parameters confirm this fact. Even if the spatial distribution was forcedly changed, the hydrologic response was similar and especially the volumetric water content estimation was not affected by the imposed perturbation. These circumstances suggest that the reconstruction of hydraulic conductivity spatial distribution is a highly ill-conditioned problem. When a problem is well-conditioned, then a unique solution exists and it depends on the input data in a continuous way, i.e., small variations in the input data imply small variations in the results. The ill-conditioning, on the contrary, implies that small variations in the initial values cause large variations in the results. The unknown uncertainty of the available data is a cause. The reconstruction would be improved if other kinds of data, such as geophysical measurements were available.

Chapter 7

Conclusions

The LEO is an 'open access' experimental hillslope, where laboratory scale experiments that are 'close' to the natural scale can be performed. The facility offers a great opportunity to improve the knowledge about hydrologic processes and landscape evolution. During the first experiment run in 2013, the response of the hillslope proved to be surprisingly different from pre-experimental predictions of numerical simulations. Niu et al. [2014] found a possible explanation of the occurred discrepancy, suggesting the development of a localized heterogeneity. Starting from these conclusions, this thesis work aims to reconstruct the spatial distribution of hydraulic conductivity in the LEO soil. The objective is pursued via data assimilation of the measurements collected during the first experiment at LEO. Ensemble Kalman filter and sequential particle filtering are the chosen assimilation methods. The parameter estimation is carried out by means of the physical-based coupled hydrological model CATHY. Many simulations have been implemented considering different degrees of soil heterogeneity. A novel assimilation iteration scheme has been developed in order to force the convergence of the model. The hillslope, modeled by means of a surface grid replicated in vertical direction, has been subdivided into ten zones, then into five. In each zone, the soil has been considered homogeneous. In the ten zones configuration, the assimilation has been carried out both with EnKF and SIR configuration, while in the five zones configuration only the SIR method has been applied. A second experiment considers the hillslope as fully heterogeneous, modeling the hydraulic conductivity as a second order stationary random field. In this case, only EnKF algorithm has been used for the assimilation.

For each spatial configuration, the values of the parameters identified by the various assimilation procedures was used in CATHY to simulate the hydrologic response of the hillslope under the real atmospheric conditions of the first experiment at LEO. In this way, both the assimilated spatial distributions of hydraulic conductivity and the state

predictions could be compared against measurements. In addition, a simulation has been carried out also with the prior distribution of parameters. The results, computed in terms of overland flow, seepage face flow and water storage, do not compare favourably with the observations showing an error that is larger than the error of the simulations reported in Niu et al. [2014]. The error on volumetric water content is improved by the assimilation process while the total runoff is still very far from the observed values. The influence of hydraulic conductivity variation on the estimation of volumetric water content has been further investigated. Ten perturbed sets of parameters have been given in input to CATHY, and the results again are similar in the ten cases, with little variations of the error on volumetric water content. These circumstances induce to consider the reconstruction of saturated hydraulic conductivity as an ill-conditioned problem. When a problem is ill-conditioned, a small variation of the input parameters corresponds to a great variation on the results. Our results suggest that the use of distributed soil water content measurements to identify a distributed parameter with a physically based catchment model seems to be insufficient to constrain and render well-conditioned the identification problem. Possible explanations for this behaviour are mainly two.

On one hand, the uncertainty related to the accuracy of the used sensors is large. In fact, the raw data had to be preconditioned to exclude a number of non-physical data values and the entire story of surface-most measurements were also excluded because unrealistically noisy. It is here suggested that smoothing techniques could be used here to improve the data quality without excluding data values.

A second source of ill-conditioning raised by the use of moisture content data to constrain the identification problem in the fact that this type of measurement does not give any information in the saturated areas, where saturation is constantly unitary (moisture content equal to saturated porosity). The overland and seepage fluxes are controlled principally by the values of the saturated conductivity in the saturated zone. For this reason, a better and more accurate strategy, not investigated in this thesis, would be to include flow (overland and seepage) measurements in the assimilation process. This possibility is not explored at the moment because the time series of the overland flow rates measured during the first experiment is strongly underestimated, and corrections are being investigated. Once more accurate data will be available, the experiments reported in this thesis will be repeated including also flux data information during the assimilation.

Acknowledgements

Today an important journey has come to an end, and the moment for acknoledgements has arrived.

First I would like to thank Professor Putti. The fascinating thesis topic he proposed led me to learn many new things. I thank him too for the kindness and courtesy with which he has supervised me over these past months.

I would like to thank Damiano Pasetto for instructing me the codes necessary for the fulfillment of the thesis work, for the patience with which he listened to my questions and responded to my many doubts. His diligent revision was fundamental to the completion of this thesis.

I would like to thank my Computer Room colleagues and all those working in the CD passage at the third floor of Torre Archimede. They made my thesis period the highlight of my master degree.

My sister Lavinia deserves special thanks. Although far away, she never left me deprived of her affection and support, and she never stopped spurring me on.

Big thanks go to my parents, Nadia and Luciano, for supporting me in all these years of hard work. Thank you for accepting my choices and for motivating me to do my best.

Thanks to Grandma Lea, for having taught me many things, and for her affection.

Thanks to my uncles Danilo and Antonio. They made me happy by accepting to be my uncles, and since then they never ceased being interested in my life and believing in me. Particular thanks go to my Aunt Isabella, for being my ant and for telling me something that still reassures me in the hard moments.

Another special 'thank you' is reserved for Andrea. He accepted me as I am, without ever trying to change me. He helps, supports and reassures me.

Thanks to Mari and thanks to Fabio, since they never tired of being my friends, even though many years have passed since we met!

Thanks to ELena and Elison, since they too are great friends and it's always fun to go out together!

Thanks to Stress and to athletics, for helping me understand that it's not only where

you arrive that is important, but also the dedication you apply on the way there.

Thanks to Anna, for making the bachelor a fun and cheerful period, and for being here with me still!

Thanks to Tatiana for her help, without which I would not yet be here!

Thanks to all my friends for the many enjoyable evenings, for the support and attentive ears!

Finally, thanks to you who are here with me today to celebrate the successful completion of a hard, but satisfying, journey!

Ringraziamenti

Un importante percorso giunge oggi a compimento, ed è quindi il momento dei ringraziamenti.

Ringrazio innanzitutto il professor Putti per avermi proposto una tesi interessante e che mi ha permesso di imparare tante cose nuove. Lo ringrazio anche per la cortesia e la disponibilitá con cui mi ha seguito in questi mesi.

Ringrazio Damiano Pasetto per avermi istruito nell'uso dei codici necessari, per aver risposto a tutti i miei (molti) dubbi e per la pazienza con cui ha accolto tutte le mie domande. Il suo lavoro di revisione è stato fondamentale per la redazione di questa tesi.

Ringrazio anche i compagni della Sala Calcolo e tutti quelli che lavorano nel corridoio CD al terzo piano della Torre Archimede. Grazie a loro il periodo di tesi è stato divertente e, senza dubbio, il migliore della magistrale.

Un grazie speciale va a Lavinia, mia sorella, che pur se distante non mi ha mai fatto mancare il suo affetto e il suo sostegno e non si è mai stancata di pungolarmi e spronarmi.

Un grandissimo grazie va ai miei genitori Nadia e Luciano, per avermi supportato (e sopportato) in tutti questi anni di studio matto e dispera(n)tissimo. Grazie perchè avete sempre accettato le mie scelte e mi avete sempre spinto a fare meglio.

Grazie alla nonna Lea, perchè mi ha insegnato tante cose e mi ha sempre voluto bene.

Grazie ai miei zii Danilo e Antonio, che mi hanno reso molto felice accettando di diventare i miei zii, e che da allora non hanno mai smesso di interessarsi della mia vita e credere in me. Un grazie particolare va alla zia Isabella, per essere mia zia e per avermi detto una cosa che ancora mi rassicura nei momenti difficili.

Un altro grazie speciale è riservato ad Andrea, perchè mi ha sempre accettato per come sono, senza volermi cambiare. Perchè mi aiuta, mi sostiene e mi rasserena.

Grazie alla Mari e grazie a Fabio, perchè nonostante mi conoscano da una vita non si sono ancora stancati di essere miei amici!

Grazie a Elena ed Elison, perchè sono ottime amiche ed è sempre un divertimento uscire insieme!

Grazie a Stress e all'atletica, per avermi aiutato a capire che non conta solo il risultato che ottieni, ma anche l'impegno che ci metti per raggiungerlo.

Grazie ad Anna per aver reso gli anni della triennale divertenti e spensierati, e per essere ancora qui con me oggi!

Grazie a Tatiana per avermi aiutato in tutti questi anni, perchè diciamocelo, senza di lei non sarei già qui!

Grazie a tutti i miei amici e amiche, che in questi anni mi hanno regalato serate di divertimento e spensieratezza ma anche sostegno e ascolto!

E infine grazie a tutti voi che siete qui con me oggi a festeggiare insieme il coronamento di un percorso faticoso ma soddisfacente!

Bibliography

Anderson, E. I. (2005). Modeling groundwatersurface water interactions using the dupuit approximation. 28:315–327.

Arulampalam, M. S., Maskell, S., Gordon, N., and Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. 50.

Brooks, R. H. and Corey, A. T. (1964). Hydraulic properties of porous media.

Camporese, M. nad Paniconi, C., Putti, M., and Salandin, P. (2009). Comparison of data assimilation techniques for a coupled model of surface and subsurface flow. 8:837–845.

Camporese, M., Crestani, E., and Salandin, P. (2012). Assessment of local hydraulic parameters by enkf data assimilation in real aquifers: a case study in downtown padova (italy). In of Illinois at Urbana Champaign, U., editor, *XIX International Conference on Water Resources*.

Camporese, M., Paniconi, C., Putti, M., and Orlandini, S. (2010). Surface-subsurface flow modeling with path-based runoff routing, boundary condition-based coupling, and assimilation of multisource observation data. 46:W02512.

Chen, Y. and Zhang, D. (2006). Data assimilation for transient flow in geologic formations via ensemble kalman filter. 29:11071122.

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbit, R. P., Slater, A. G., Schmidt, J., and Uddstrom, M. J. (2008). Hydrological data assimilation with the ensemble kalman filter: Use of streamflow observations to update states in a distributed hydrological model. 31:1309–1324.

Dagan, G. (1989). *Flow and transport in porous formations*. Springer-Verlag.

Das, N. N. and Mohanty, B. P. (2006). Root zone soil moisture assessment using remote sensing and vadose zone modeling. 5:296–307.

Downer, C. W. and Ogden, F. L. (2003). Prediction of runoff and soil moistures at the watershed scale: Effects of model complexity and parameter assignment. 39(3).

Evensen, G. (2003). The ensemble kalman filter: theoretical formulation and practical implementation. 53:343367.

Franssen, H. J. H., Kaiser, H. P., Kuhlmann, U., Bauser, G., Stauffer, F., Müller, R., and Kinzelbacj, W. (2011). Operational real-time modeling with ensemble kalman filter of variably saturated subsurface flow including stream-aquifer interaction and parameter updating. 47.

Furman, A. (2007). Modeling Coupled SurfaceSubsurface Flow Processes: A Review. 7:741–756.

Gevaert, A. I., Teuling, A. J., Uijlenhoet, R., DeLong, S., Huxman, T. E., Pangle, L. A., Breshears, D. D., Chorover, J., Pelletier, J. D., Saleska, S. R., Zeng, X., and Troch, P. (2014). Hillslope-scale experiment demonstrates the role of convergence during two-step saturation. 18:3681–3692.

Goderniaux, P., Brouyre, S., Fowler, H. J., Blenkinsop, S., Therrien, R., Orban, P., and Dassargues, A. (2009). Large scale surfacesubsurface hydrological model to assess climate change impacts on groundwater reserves. 373:122–138.

Graham, D. N. and Butts, M. B. (2005). Flexible integrated watershed modeling with MIKE SHE. In Singh, V. P. and Frevert, D. K., editors, *Watershed Models*, chapter 10.

Gunduz, O. and Aral, M. M. (2005). River networks and groundwater flow: a simultaneous solution of a coupled system. 301:216–234.

Hopp, L., Harman, C., Desilets, S. L. E., Graham, C. B., McDonnel, J. J., and Troch, P. A. (2009). Hillslope hydrology under glass: confronting fundamental questions of soil-water-biota co-evolution at biosphere 2. 13:2105–2118.

Huyakorn, P. S. and Pinder, G. (1983). *Computational mathods in subsurface flow*. Academic Press.

Huyakorn, P. S., Thomas, S. D., and Thompson, B. M. (1984). Techniques for making finite elements competitve in modeling flow in variably saturated porous media. 8:1099–1115.

Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press.

Kollet, S. J. and Maxwell, R. M. (2006). Integrated surfacegroundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. 29:945–958.

Liu, G. and Chen, Y. an Zhang, D. (2008). Investigation of flow and transport processes at the made site using ensemble kalman filter. 31:975–986.

McLaughlin, D. (2002). An integrated approach to hydrologic data assimilation: interpolation, smoothing, and filtering. 25:1275–1286.

Moradkhani, H. and Hsu, K. (2005). Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter. 41.

Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P. R. (2005). Dual stateparameter estimation of hydrological models using ensemble kalman filter. 28:135–147.

Morita, M. and Yen, B. C. (2002). Modeling of conjunctive two-dimensional surface-three-dimensional subsurface flows. pages 184–200.

Niu, G.-Y., Pasetto, D., Scudeler, C., Paniconi, C., Putti, M., Troch, P. A., DeLong, S. B., Dontsova, K., Pangle, L., Breshears, D. D., Chorover, J., Huxman, T., Pelletier, J., Saleska, S. R., and Zeng, X. (2014). Incipient subsurface heterogeneity and its effect on overland flow generation-insight from a modeling study of the first experiment at the Biosphere 2 Landscape Evolution Observatory. 18:1873–1883.

Orlandini, S. and Rosso, R. (1996). Diffusion wave modeling of distributed catchment dynamics. 1:103–113.

Panday, S. and Huyakorn, P. S. (2004). A fully coupled physically-based spatially-distributed model for evaluating surface/subsurface flow. 27:361–382.

Paniconi, C., Marrocu, M., Putti, M., and Verbunt, M. (2003). Newtonian nudging for a richards equation-based distributed hydrological model. 26:161–178.

Paniconi, C. and Putti, M. (1994). A comparison of picard and newton iteration in the numerical solution of multidimensional variably saturated flow problems. 12:3357–3374.

Pasetto, D., Camporese, M., and Putti, M. (2012). Ensemble kalman filter versus particle filter for a physically-based coupled surfacesubsurface model. 47:1–13.

Pasetto, D., Niu, G.-Y., Pangle, L., Paniconi, C., Putti, M., and Troch, P. (2014). Observability and sensor failure analysis for the biosphere 2 landscape evolution observatory based on distributed volumetric water content data.

Pohll, G. M., Warwick, J. J., and Tyler, S. W. (1996). Coupled surface-subsurface hydrologic model of a nuclear subsidence crater at the nevada test site. 186:43–62.

Salamon, P. and Feyen, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. 376:428–442.

Shen, C. and Phanikumar, M. S. (2010). A process-based, distributed hydrologic model based on a large scale method for surfacesubsurface coupling. 33:1524–1541.

Singh, V. and Bhallamudi, S. M. (1998). Conjunctive surface-subsurface modeling of overland flow. 21:567–579.

Sulis, M., Meyerhoff, S., Paniconi, C., Maxwell, R., and Putti, M. (2010). A comparison of two physics-based numerical models for simulating surface watergroundwater interactions. 33:456–467.

Van Genuchten, M. T. and Nielsen, D. R. (1985). On describing and predicting the hydraluic properties of unsaturated soils. pages 615–628.

VanderKwaak, J. E. and Loague, K. (2001). Hydrologic-response simulations for the r-5 catchment with a comprehensive physics-based model. 37(4):999–1013.

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. 38(12).

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. 42.

Weill, S., Mouche, E., and Patin, J. (2009). A generalized richards equation for surface/subsurface flow modelling. 366:9–20.

Xie, X. and Zhang, D. (2010). Data assimilation for distributed hydrological catchment modeling via ensemble kalman filter. 33:678–690.

Yeh, W. W.-G. (1986). Review of parameter identification procedures in groundwater hydrology: the inverse problem. 2:95–108.

Zhou, Y., McLaughlin, D., and Entekhabi, D. (2006). Assessing the performance of the ensemble kalman filter for land surface data assimilation. 134:2128–2142.