UNCERTAINTY REDUCTION AND PARAMETERS ESTIMATION OF A DISTRIBUTED HYDROLOGICAL MODEL WITH GROUND AND REMOTE SENSING DATA.

Francesco Silvestro

CIMA Research Foundation, Savona, Italy

Simone Gabellani

CIMA Research Foundation, Savona, Italy

Roberto Rudari

CIMA Research Foundation, Savona, Italy

Fabio Delogu

CIMA Research Foundation, Savona, Italy

Paola Laiolo

CIMA Research Foundation, Savona, Italy

Giorgio Boni

CIMA Research Foundation, Savona, Italy

DIBRIS, University of Genova, Italy

Corresponding author: Francesco Silvestro

mail: francesco.silvestro@cimafoundation.org

CIMA Research Foundation (www.cimafoundation.org)

University Campus, Armando Magliotto, 2. 17100, Savona, Italy

Tel. +39 019230271, fax. +39 01923027240

ABSTRACT

During the last decade the opportunity and usefulness of using remote sensing data in hydrology, hydrometeorology and geomorphology has become even more evident and clear. Satellite based products often provide the advantage of observing hydrologic variables in a distributed way, offering a different view with respect to traditional observations that can help to understand and model the hydrological cycle. Moreover, remote sensing data are fundamental in scarce data environments. The use of satellite derived Digital Elevation Model (DEM), which are globally available now at 30 m resolution (e.g. from Shuttle Radar Topographic Mission, SRTM), have become standard practice in hydrologic model implementation, but other types of satellite derived data are still underutilized. As a consequence there is the need of developing and testing techniques that allow exploiting the opportunities given by remote sensing data to parameterize hydrological model and improving their calibration.

In this work, Meteosat Second Generation Land Surface Temperature (LST) estimates and Surface Soil Moisture (SSM) available from EUMETSAT H-SAF are used together with streamflow observations to calibrate the *Continuum* hydrological model that computes such state variables in a prognostic mode. The first part of the work aims at proving that satellite observations can be exploited to reduce uncertainties in parameters calibration by reducing the parameters equifinality that can become an issue in forecast mode. In the second part four parameter estimation strategies are implemented and tested in a comparative mode: i) a multi-objective approach that includes both satellite and ground observations which is an attempt to use different source of data to add constraints to the parameters; ii and iii) two approaches solely based on remotely sensed data that reproduce the case of scarce data environment where streamflow observation are not available; iv) a standard calibration based on streamflow observations used as a benchmark for the others.

Two Italian catchments are used as test-bed to verify the model capability in reproducing long-term (multi-year) simulations.

The results of the analysis evidence that, as a result of the model structure and the nature itself of the catchment hydrologic processes, some model parameters are only weakly dependent on discharge observations and prove the usefulness of using data from both ground stations and satellite to add constrains to the parameters in the calibration process and reducing the number of equifinal solutions.

Keywords: hydrological models, remote sensing, energy balance, parameters estimation, validation.

1. INTRODUCTION

The estimation of parameters in hydrological models is still a challenge in hydrology. Many works have been devoted to determining the best calibration strategy (Yapo et al 1998, Madesen, 2000; Kim et al., 2007; Singh and Bardossy, 2012, Xu et al., 2013) with some trying to evaluate the uncertainties associated with the parameters estimation process (Beven and Binley, 1992; Vrugt et al., 2003; Carpenter and Georgakakos, 2006; Zappa et al. 2010). This issue has become even more complex with the increasing use of continuous and distributed hydrological models. This trend led to a significant increase of the number of parameters that need calibration. A large number of parameters allow good performance in the calibration phase, but this can lead to a large number of equifinal parameter sets (Beven and Binley, 1992) sometimes hampering the forecast ability of the models.

Traditionally, the calibration of hydrological models requires appropriate series of historical data, particularly of streamflow data, which are not easily available everywhere in the world; this condition aggravates the equifinality problem that is higher the lower the observation capacity. Such issues raised the attention of the scientific community becoming the focus of coordinated scientific initiatives (e.g. Prediction in Ungauged Basin (PUB), a science initiative of the International Association of Hydrological Sciences, that was developed in the period 2003-2012). In a world where the data sharing capacity is increasing, it seems that the problem of data shortage for hydrologic models calibration is not going to disappear; the level gauge stations can be a limited number and in some areas are very rare, additionally in some cases the access to river discharge information have been declining since the 1980s (Vörösmarty et al, 2001).

As a consequence, the use of remote sensing for direct streamflow measurements has received increased attention lately and, even if promising in some cases, it faces various technological, physical and scale limits. The more straightforward approaches use statistical relationships between remotely sensed river widths and in situ measurements (Brakenridge et al, 2005; Pavelsky, 2014) making them suitable for the extension of existing historical data, but unusable for ungauged sites. The limits are mainly due to the fact that accurate estimates of stream flow require the availability of several hydraulic parameters (width, depth, slope, channel morphology), which are difficult to derive entirely from remote sensing. Simplified models that make use of some of these parameters introduce uncertainties that limit their applicability; moreover the detection of changes in hydraulic parameters has to deal with the spatiotemporal resolution of the satellite sensors. These models (see

Bjerklie et al, 2003 for a comprehensive review) are therefore not suitable for detecting changes in discharge for medium to small-scale basins (Brakenridge et al. 2012).

It is therefore compulsory to look at other possibilities offered by satellite sensors. Nowadays, the remote sensing of other meteorological, hydrological and ecological variables is more reliable and widely available at the global scale. Satellite products such as precipitation, Short Wave and Long Wave radiation, atmospheric profiles, vegetation parameters, Land Surface Temperature (LST), evapotranspiration (ET), and Digital Elevation Models (DEM) are now operational and widely used in meteorological and hydrological modelling.

Experiments to understand the accuracy of these products are quite popular (see e.g. Bitew and Gebremichael, 2011; Brocca et al., 2011a; Crow et al., 2012; Göttsche and Hulley 2012; Yu et al, 2012; Murray et al, 2013; Zhang et al, 2013). This kind of data is by now available for a very high percentage of the earth's surface, and covers most of the areas where the density of ground stations is poor. This leads to a panorama in which estimating the parameters of a hydrological model by using only satellite information is a real possibility (Silvestro et al., 2013). However, the ability to calibrate a model using satellite data, even in combination with traditional in-situ data, is still a challenging topic. Scientific work in this field goes in many directions: Rhoads and Dunayah (2001) used satellite-derived LST to validate a land surface model, Caparrini and Castelli (2004) and Sini et al. (2008) assimilated remote sensed measurements into a land-surface model to estimate the surface turbulent fluxes, Brocca et al. (2011a) analysed different remotely sensed soil humidity estimations with the perspective of using them in hydrological modeling, White and Lewis (2011) used satellite imagery to monitor the dynamics of wetlands of the Australian Great Artesian Basin, Khan et al. (2011) have recently proposed a procedure to calibrate a fully distributed hydrological model using satellite-derived flood maps.

The objective of this work is to analyse the calibration skill of a distributed continuous hydrologic model by augmenting the model constraints with satellite-retrieved data. As a first analysis, in the context of a classical uncertainty analysis (Beven and Binley, 1992; Shen et al. 2012) it is shown that using satellite data together with ground stations observations can reduce both parameters' uncertainty and equifinality. The uncertainty analysis is used here also to define parameters sensitivity to different types of observation, with the goal of underpinning the importance of having a plurality of observations that might influence different models parameters in different ways.

After that, three simple calibration methods were applied in order to exploit the advantages of utilizing multi-sensor observations. The first method lies in the family of the multi-objective calibration approaches (Efstratiadis and Koutsoyiannis, 2010) and tries to exploit simultaneously

satellite and streamflow data. The second and third methods are two attempts to use only satellite data without any streamflow measurements, simulating the case of a basin in a scarce data environment. These last experiments are conceived along the lines of Silvestro et al., (2013), but with a more comprehensive approach that exploits both LST and SSM estimates from satellites. The results of the presented methods are then compared with those obtained using a fourth standard calibration methodology based on streamflow data.

The hydrological model used in the study is Continuum (Silvestro et al., 2013). It is a distributed continuous model conceived to satisfy the principle of parsimony in parameterization (Perrin et al., 2001; Coccia et al., 2009; Todini et al., 2009; Efstratiadis and Koutsoyiannis, 2010) and to be balanced between a good representation of physical processes and the simplicity of the schematizations and implementation.

The article is organized as follows: chapter two provides a short overview of the *Continuum* hydrological model, the description of the data set used, the parameters uncertainty analysis and the proposed calibration methods. The application and the analysis of the results are presented in chapter three while chapter four contains discussion and conclusions.

2. MATERIAL AND METHODS

2.1. MODEL OVERVIEW

Continuum is a continuous distributed hydrological model that relies on a morphological approach, based on drainage network components identification (Giannoni et al., 2000; Giannoni et al., 2005). These components are derived from DEMs. The DEM resolution drives the model spatial resolution. Flow in the soil is divided firstly into a sub-surface flow component that is based on a modified Horton schematization (see Gabellani et al., 2008 for details) and that follows the drainage network directions; and secondly, into a deep flow component that moves following the hydraulic head gradient obtained by the water-table modeling. The surface flow schematization distinguishes between channel and hillslope flows. The overland flow (hillslopes) is described by a linear reservoir scheme, while for the channel flow (channel) a schematization derived by the kinematic wave approach (Wooding, 1965; Todini and Ciarapica, 2001) is used. The energy

balance is solved explicitly at cell scale by means of the force-restore equation, that allows having the LST as a distributed state variable of the model (e.g., Lin, 1980; Dickinson, 1988; Sini et al., 2008). For further details on the model please refer to Silvestro et al. (2013).

Various authors highlighted the importance of reducing the model parameterization and maintaining a stable and simple structure (Montaldo et al., 2005; Coccia et al., 2009; Todini, 2009; Brocca et al., 2011b). The design of Continuum follows the philosophy of finding a balance between a detailed description of the physical processes and a robust and parsimonious parameterization (Figure 1).

Leaf Area Index (LAI) is used to parameterize the storage capacity of the vegetation (Kozak et al., 2007)

Continuum has six parameters that need calibration at basin scale: two for the surface flow, two for the sub-surface flow and two for deep flow and the water table. In Table 1 the calibration parameters are listed and linked to the physical processes parameterized.

The hillslope flow motion parameter u_h influences the general shape of the hydrograph, while the impact of u_c on the hydrograph shape depends on the length of the channeled paths. These are two lumped parameters: u_c represents the friction coefficient in the channel motion equation, u_h accounts for the general characteristics of the hillslope that influence the motion (e.g. friction, slope) and is more an empirical parameter (see Figure 1).

The parameter c_t is related to the soil field capacity V_{fc} and identifies the fraction of water volume in the soil that can be extracted only through evapotranspiration. The relationship is:

$$V_{fc} = c_t V_{\max} \qquad (1)$$

Where V_{max} is the maximum capacity of the soil to storage water in the root zone.

The "infiltration capacity" parameter c_f controls the velocity of subsurface flow (i.e, it is related to saturated hydraulic conductivity), defining the asymptotic minimum infiltration rate for saturated soils f_1 with the following equation:

$$f_1 = c_f f_0 \qquad (2)$$

Where f_0 is the maximum infiltration rate for completely dry soil.

The parameters c_t and c_f regulate the dynamics of saturation at cell scale. Since both f_0 and V_{max} are distributed parameters estimated as functions of Curve Number (Gabellani et al., 2008) the pattern of f_1 and V_{fc} is spatially modulated by the pattern of Curve Number maps (Silvestro et al., 2013) which are a synthetic representation of the local soil properties.

The parameters V_{Wmax} and R_f govern the deep flow and the water table dynamic (Silvestro et al., 2013). V_{Wmax} represents the absolute maximum water content of the aquifer on the whole investigated area, the maximum water content on each cell is estimated basing on V_{Wmax} and on the slope (Saulnier et al., 1997). R_f is a multiplicative factor in the Darcy equation used to estimate the

flux per unit area between two contiguous cells and mainly takes care of differentiating the saturated vertical and horizontal conductivity. These two parameters have a reduced influence compared to the other four parameters because of the slow temporal dynamic of the water table. The sensitivity to R_f increases with the total basin drainage area when the effect of the interaction between the water table and the vadose zone becomes crucial in the formation of the recession curve between the rainfall events (Silvestro et al., 2013).

Continuum accounts for LST as an explicit state variable and allows for the estimation of the soil moisture in the root zone as the saturation degree (SD) defined here by the ratio of the actual soil water content V(t) and the maximum storage capacity V_{max} (see Figure 1):

$$SD = \frac{V(t)}{V_{\max}} \qquad (3)$$

Both of these variables are represented at DEM spatial resolution.

The snow melting process was not considered in Silvestro et al. (2013), since multi-year simulation are carried out in this work a simple snow melting model has been introduced and described in Appendix.A.

2.2. DATASET

The first test case is the Orba basin that is located in the Apennine part of the Piemonte region (Italy). It has a total area of approximately 800 km^2 and it is a tributary of the Tanaro River (Figure 2).

The Piemonte and Liguria regions meteorological networks monitor the basin. Data from rain gauges, thermometers, hygrometers, shortwave radiometers and anemometers are available with a temporal resolution of 1 hour. Two stage-gauging stations are working with maintained stage-discharge rating curves; the two stations are located quite far one from each other along the river: Tiglieto in a head catchment (drained area: 75 km^2) and Casalcermelli near the basin outlet (drained area: 800 km^2).

For this application, we extended the data set used in Silvestro et al. (2013). The chosen period starts from June 1st 2006 and ends on December 31st 2011. The first five months of 2006 are used as the model "warm-up" period.

The second test case is the Casentino basin (Figure 2). It is a head catchment of the Arno river basin located in Tuscany. The watershed is located in the Central Apennines with elevation that ranges

between 200 to 1600 m a.s.l.. The mountainous part of the basin is mainly covered by forest, while cultivated fields or zones with low vegetation primarily make-up the flat areas. Urban areas cover a low percentage of territory.

The two basins are only marginally impacted by snowfall and snow cover during winter.

The meteorological network of the Tuscany Region provides rainfall, air temperature, air humidity, solar radiation and wind speed and direction with temporal resolution of 1 hour. Only one stage-gauging station (Subbiano) is working with a maintained stage-discharge rating curve; the gauge is located in the flat area of the basin at about 10 km from the confluence of the Casentino River along the Arno River (drained area 670 km^2). The period of simulation ranges from June 1st 2005 to December 31st 2011. The first five months of 2005 are used to warm-up the model.

In both cases the period has been chosen based on the data availability and in order to have reliable stage-discharge curves to estimate the observed streamflow.

The model temporal resolution is set to 1 hour as the micro-meteorological observations, the surface flow needs a finer time step for computational stability reasons and it was fixed to 30 s.

The remote sensing data employed to implement the model and set additional constraints to the model parameters are:

i) The Istituto Geografico Militare (IGM) DEM used to extract the basin morphological parameters (http://www.igmi.org/prodotti/dati_numerici/dati_matrix.php);

ii) Land Surface Analysis Satellite Applications Facility (LSA-SAF) Land Surface Temperature (LST) product retrieved from Meteosat Second Generation (MSG) observations (landsaf.meteo.pt);

iii) SM-OBS-1 Surface Soil Moisture retrieved from ASCAT (Wagner et al., 2013) and distributed within the EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H-SAF) program used as a benchmark to be compared with the model output (hsaf.meteoam.it);

iv) LSA-SAF Leaf Area Index (LAI) to parameterize the vegetation cover.

The DEM resolution is 0.0011 deg (about 100 m). The model spatial resolution is set equal to the DEM resolution. LST estimations are provided by LSA SAF of EUMETSAT (*EUMETSAT,* 2009). LST data are available every fifteen minutes with a spatial resolution of approximately 0.04 deg (about 4.5 km) since 2009. In order to compare model and satellite data, the approach followed by Silvestro et al. (2013) has been adopted. It allows projecting LST obtained from the model to the same geometry of the satellite observations. In order to carry out comparison at basin scale the mean of the pixel values has been used:

$$\overline{LST} = \frac{1}{N}\sum_{i=1}^{N} LST_i \qquad (4)$$

Where N is the number of cells of the spatial grid, and LST_i the LST of i-th cell. The comparison is carried out at those instants where both model and satellite data are available (1 hour resolution) and if at least 50% of satellite pixels that cover the basin have reliable data (e.g. in case of bad weather no-data values can be found in satellite product).

The H-SAF SM-OBS-1 product consists of European maps of large scale Surface Soil Moisture (SSM) retrieved from Advanced Scatterometer (ASCAT), the active microwave sensor, which flies on-board two polar-orbiting Meteorological Operational (METOP) satellites. This product gives soil moisture estimates across different test sites in Europe, Americas and Africa (Brocca et al., 2010; Albergel et al., 2012). EUMETSAT makes the product available, from June 3, 2009, in near real-time with a spatial resolution of approximately 25 km and revisit time of twice a day.

The SSM have been re-sampled to the model resolution using the nearest neighbour method. SM-OBS-1 (H07) data with quality flag, provided with the product, greater than 15 were discarded. Since the product is referred to the first centimetres of soil, the Soil Water Index (SWI) method, developed by Wagner et al. (1999), was applied to SSM satellite data to obtain an estimate of the saturation degree in the root zone. This filter allows relating the surface soil moisture estimates to the profile soil moisture content. It relies on the assumption that the variation in time of the average value of the soil moisture profile is linearly related to the difference between the surface and the profile values. In this study a simple recursive formulation of the method was used (Stroud, 1999; Albergel et al., 2008) and the characteristic time length T, which represents the time scale of soil moisture variation (Wagner et al., 1999), was considered equal to 10 days. Since in situ soil moisture measurements are not available and the soil properties are not known quantitatively with high detail, the parameter T has been set to a priori value that has been estimated, as order of magnitude, using the definition of T of Wagner et al. (1999) and used also in Parajka et al. (2006) based on the mean soil characteristics of the considered catchments as described in the model (the average potential soil moisture capacity of the considered basins is around 150-170 mm, assuming a porosity of 0.3, a pseudo diffusivity of 10 days would then translate into a wetting front celerity around 50 mm per day that is a reasonable value for these soils). The SWI has been then rescaled to model climatology using a min-max correction technique (Brocca et al., 2013). After the rescaling the mean SWI at basin scale has been computed as a simple average of the values of pixels that cover the basin and used for comparison with model saturation degree SD. The comparison is carried out at those instants where both model and satellite data are available (twice a day) and if at

least 50% of satellite pixels that cover the basin have reliable data (quality flag greater than 15) LAI maps were produced with temporal update of fifteen days as averaged values of daily LSA-SAF maps at spatial resolution of 0.04 deg. (*EUMETSAT,* 2008) and gridded with nearest neighbour method on model resolution.

2.3. STATISTICS AND SCORES

A series of statistics and scores are used in the following sections to carry out the uncertainty analysis, the calibration and the validation of the Continuum model. They are presented and described in this section. A different subsection is used for each of the three considered observable variables: streamflow, LST and SWI.

2.3.1. Streamflow

The Nash Sutcilffe (NS) coefficient (Nash and Sutcilffe, 1970) was chosen as main likelihood function since it is one of the most widely used measures to evaluate model performances in hydrology:

$$NS = 1 - \sum_{t=1}^{t\max} \frac{\left(Q_m(t) - Q_o(t)\right)^2}{\left(Q_m(t) - \overline{Q_o}\right)^2} \quad (5)$$

Where $Q_m(t)$ and $Q_o(t)$ are the modeled and observed streamflows at time t. $\overline{Q_o}$ is the mean observed streamflow.

Five other scores were evaluated to assess the model performance (Madsen et al., 2000; Batholomes and Todini, 2005).

Chiew McMahon (CM) coefficient (Chiew and McMahon, 1994):

$$CM = 1 - \sum_{t=1}^{t\max} \frac{\left(\sqrt{Q_m(t)} - \sqrt{Q_o(t)}\right)^2}{\left(\sqrt{Q_m(t)} - \sqrt{\overline{Q_o}}\right)^2} \quad (6)$$

Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{t\max} \sum_{t=1}^{t\max} \left(Q_m(t) - Q_o(t)\right)^2} \quad (7)$$

Correlation coefficient (CORR):

$$CORR = \frac{\sum_{t=1}^{t\max}\left(Q_m(t)-\overline{Q_m}\right)\cdot\sum_{t=1}^{t\max}\left(Q_o(t)-\overline{Q_o}\right)}{\sqrt{\left(\sum_{t=1}^{t\max}\left(Q_m(t)-\overline{Q_m}\right)^2\right)\cdot\left(\sum_{t=1}^{t\max}\left(Q_o(t)-\overline{Q_o}\right)^2\right)}} \quad (8)$$

$\overline{Q_m}$ is the mean modeled streamflow.

Relative Error (Rel. Err.):

$$Rel.Err. = \frac{1}{t\max}\sum_{t=1}^{t\max}\frac{\left|Q_m(t)-Q_o(t)\right|}{Q_o(t)} \quad (9)$$

Peak Flow Relative Error (PFRE):

$$PFRE = \frac{1}{Npeaks}\sum_{tp=1}^{Npeaks}\frac{Q_{pm}(tp)-Q_{po}(tp)}{Q_{po}(tp)} \quad (10)$$

Where $Q_{pm}(tp)$ and $Q_{po}(tp)$ are the modeled and observed peak flows. The peak flows are selected considering the values larger of a fixed threshold Q_{th}.

NS, CM and CORR indicate good matching between model and observations when they are next to 1, while RMSE, Rel.Err. and PFRE when they tend to 0.

2.3.2. LST

The Bias between modelled and satellite derived LST was considered as a skill score:

$$BIAS = \frac{1}{t\max}\sum_{t=1}^{t\max}\left|\overline{LST_m(t)}-\overline{LST_s(t)}\right| \quad (11)$$

Where $\overline{LST_m(t)}$ and $\overline{LST_s(t)}$ are the modelled and satellite LST averaged at basin scale at the time t.

We used the BIAS in order to check the capability of the model to reproduce the mean LST on the selected period, more than it would for the overall shape of the time series.

2.3.3. SWI

As for the hydrograph, we considered NS as a score to evaluate the performances of the model in reproducing the SWI derived by the satellite observations. In this case, considering SD and SWI directly comparable (SD≡SWI) NS is defined as:

$$NS = 1 - \sum_{t=1}^{t\max} \frac{\left(\overline{SWI_m(t)} - \overline{SWI_s(t)}\right)^2}{\left(\overline{SWI_m(t)} - \overline{SWI_s}\right)^2} \quad (12)$$

$\overline{\mathrm{SWI_m(t)}}$ and $\overline{SWI_s(t)}$ are the modeled and satellite SWI averaged at basin scale at the time t.

$\overline{SWI_s}$ is the satellite SWI averaged in space and time.

2.4. EXPERIMENTAL SET-UP

2.4.1. Uncertainty analysis

The issue of model parameter uncertainty and sensitivity has been one of the main themes of scientific discussions over the last 30 years. Many authors faced the problem following different approaches (see e.g. Beven and Binley, 1992; Liu et al., 2005; Carpenter and Georgakakos, 2006; Zappa et al., 2010; Rakovec et al., 2014), but it is widely accepted and recognized that parameter uncertainty is inevitable and rarely an optimal set of parameters that allows the best performance of the model in every condition exists; generally, there are multiple sets of parameters able to give similar results and which are therefore equivalent if the final aim is identified, that is the so-called equifinality (Savenije, 2001).

In this work, we did not carry out a full predictive uncertainty analysis, but we analysed the parameter uncertainty based on equifinal realizations obtained by a Monte Carlo experiment; to do that some concepts of the GLUE method (Beven and Binley, 1992) are used, similarly to what was done by Zappa et al. (2010) and Shen et al. (2012). Finally, we made reference to the work of Liu et al. (2005) in order to estimate the probability of parameter couples conditioned to the observations.

The concepts of GLUE approach are applied using objective functions (scores) based on streamflow, LST and SWI in order to analyse how these variables, that are modeled by Continuum and measured through ground based or remote measurement systems, are related to the model parameters. The main objective of the analysis is to study the dependence of each single parameter

from the observed variables in order to pinpoint the importance of the different observations in determining the parameter set performing best in reproducing the full set of observations.

The uncertainty analysis has been done in the Orba basin by considering the four most sensitive parameters of the model as in Liu et al. (2005) (c_t, c_f, u_c, u_h, for Continuum, see Silvestro et al., 2013, for details).

Firstly the analysis based on streamflow is done using NS as likelihood function (equation 5). The sampling space of the four parameters was defined by combining the literature (Beven and Binley, 1991; Liu et al., 2005; Zappa et al., 2010, Shen et al., 2012) with the results of the preliminary sensitivity analysis done by Silvestro et al. (2013) and considerations on the role and physical meaning of the parameters themselves. In Table 2 the range of variability of the parameters is reported.

The other two parameters were set based on physically reasonable values (due to the morphology and the soil type of the basins) assuming that there is no additional information about them. Vw_{max} is set equal to 2000 mm, and R_f is set equal to 1, which indicates a weak anisotropy between vertical and horizontal saturated conductivity. These two parameters, which represent deep soil processes, are only weakly related to the processes that influence LST and SWI observations and hence, they are unlikely to be influenced by the chosen calibration strategy.

The analysis was done by simulating the four parameters c_t, c_f, u_c, u_h, and generating a set of 3000 streamflow simulations for the sub-period 16/8/2006 to 30/9/2006. The parameters have been extracted from a multi-uniform distribution bound in the domain of the parameters. The chosen sub-period includes various streamflow regimes.

As a further method to deepen the parameter uncertainty assessment the original data were transformed into a Gaussian space and ranked in increasing order once standardized (see Liu et al., 2005, for details). Observed and modelled data are then related as follows:

$$\eta_o = \eta_s + \xi \qquad (13)$$

Where η_o and η_s are the normalized vectors of observed and modelled streamflow with 0 mean and unit variance, ξ is the error vector. The Likelihood function L_j for the j-th parameter set after Imax simulation steps can be expressed as (Xu et al., 2013):

$$L_j = \exp(-\frac{1}{2} \cdot \sum_{i=1}^{\mathrm{Im}ax} \left(\xi_{i,j}^2\right) \qquad (14)$$

This function, when properly scaled, can be considered as the posterior parameter probability density.

Similarly to what has been done with streamflow data, by comparing modeled and satellite derived LST and SWI (mean at basin scale) it is possible to carry out the uncertainty analysis to understand how these two variables are related to the model parameters.

In the case of LST the considered analysis period is August-September, 2009. The BIAS between modelled and satellite derived LST was considered as a skill score.

The procedure was then applied to the ASCAT SSM data after their transformation in SWI (Wagner et al., 1999). The considered period is August-October 2011. The model saturation degree SD and the satellite SWI maps have been averaged at basin scale and the resulting time series have been used to compute the NS that, as for the hydrograph, has been considered as a score to evaluate the performances of the model in reproducing the SWI derived by the satellite observations.

2.4.2. Parameters estimation methodologies

The results of the uncertainty analysis presented in section 3.1 show that it is possible to use ground and remote sensing observations in order to reduce equifinality of the parameters defining the hydrological model. The analysis evidences that is possible to identify relatively narrow ranges of parameters' values that optimize the scores based on different observations, each parameter being mainly linked to one or more observed variables and less to others. The uncertainty analysis also highlights that streamflow based calibration offers very limited capability to detect soil parameters and that Satellite based observations deliver a complementary capability in this respect. To exploit this opportunity three parameters estimation methodologies that use satellite data have been designed, tested and compared with a standard calibration method based on streamflow data.

The calibration process is applied, as well as the analysis presented in section 2.4.1, to the most sensitive and impacting parameters of the model (u_c, u_h, c_t, c_f).

2.4.2.1 Calibration based on streamflow observations (S.N.): the benchmark

A methodology based on the maximization of the Nash Sutcliffe coefficient between observed and modeled streamflow time series has been considered as a benchmark in order to compare the methods described in sections 2.4.2.2 and 2.4.2.3 with a more standard approach; hereafter we will call this method S.N.

2.4.2.2 Multi objective calibration (M.O.)

The M.O. approach has been designed in order to exploit the use of all the information available for the calibration process that is generally represented by different observed variables. The method is based on the set up of a multiple objective function such that:

$$Min\{F_1(\theta), F_2(\theta), ...F_n(\theta)\} \text{ with } \theta \in \Theta \quad (15)$$

θ is restricted to the feasible parameter space Θ (Madsen, 2000; Kim et al., 2007; Efstratiadis and Koutsoyiannis, 2010).

This calibration approach is based, in our case, on the comparisons of (ground or satellite) observed vs simulated streamflow, LST and SWI.

The multi-objective function is designed through the following single objective function:

$$F_1 = \left[\sum_{t=1}^{t\max} \frac{(Q_m(t) - Q_o(t))^2}{(Q_m(t) - < Q_o >)^2} \right] \quad (16)$$

$$F_2 = \left[\sum_{t=1}^{t\max} \left| LST_m(t) - LST_s(t) \right| \right] \quad (17)$$

$$F_3 = \left[\sum_{t=1}^{t\max} \frac{\left| Q_m(t) - Q_o(t) \right|}{Q_o} \right]_{Qo>QT} \quad (18)$$

$$F_4 = \left[\sum_{t=1}^{t\max} \frac{(SWI_s(t) - SD_m(t))^2}{(SWI_s(t) - < SD_m >)^2} \right] \quad (19)$$

Where Q is the streamflow, LST the mean Land Surface Temperature at basin scale, SWI and SD are the mean of Soil Water Index and Saturation Degree at basin scale; subscripts *m, o* and *s* indicate model, gauge observations and satellite estimation respectively, *t* is the time, and QT is a discharge threshold. Different periods for LST, SWI and Q could be considered in order to choose the most suitable time window in terms of availability of data and good representativeness of the variable dynamic.

Following Madsen (2000), the n (where n=4) contributors F_i have been combined in the following way:

$$Fadj = \left[(F_1 - A_1)^2 + ... (F_n - A_n)^2 \right]^{0.5} \quad (20)$$

Where the transformation factor A_i is calculated as:

$$A_i = MAX\,(F_{j,\min}; j = 1,2,...n) - F_{i,\min} \quad (21)$$

Fadj (equation 20) consists of four terms and A_i,..A_n have the role of balancing the weights of the different objectives that can have, has in the presented application, different ranges and units.

The term F_1 depends on the streamflow and it is the complement to 1 of the Nash Sutcliffe coefficient, the term F_2 depends on LST and it is the mean of the absolute errors (absolute BIAS) calculated at each time step, F_3 is a relative error estimated on streamflow values larger than a threshold and it is useful to reproduce flow peaks, and F_4 depends on the soil humidity and it is the complement to 1 of the Nash Sutcliffe coefficient. The different terms have also been chosen to be consistent with the results shown in the uncertainty analysis. All the components tend to 0 when simulated and observed variables coincide, so that the calibration process consists in the minimization of the function Fadj. The resulting parameters set is representative of a balance point of multi-dimensional Pareto front due to the different components of the multi-objective function (Madsen, 2000, 2003).

The introduction of remotely sensed LST and SWI in a multi-objective calibration is a new approach since objective functions are usually based on parameters related to observed hydrographs (e.g. total volume as in Yapo et al., 1998; Efstratiadis and Koutsoyiannis, 2010). This approach follows and improves the investigations carried out by other authors (Crow et al., 2003; Santanello et al., 2007; Koren et al., 2008; Flores et al., 2010; Montzka et al., 2011; Ridler et al., 2012; Corbari and Mancini, 2014; Sutanudjaja et al., 2014; Wanders et al., 2014) who attempted to combine, in the calibration process, remote sensed and in situ observations of variables other than streamflow. The target here is not applying and testing quite sophisticated or complex algorithms for calibration, like those described in Yapo et al. (1998) or Vrugt et al. (2003), but rather to assess if the use of satellite observations leads to an advantage in model calibration with respect to its capability of simulating discharge values. Here, a very simple brute-force calibration approach was used.

2.4.2.3 Remote sensing data calibration approach (R.S.)

When no streamflow data are available, we can still calibrate the model on satellite data, LST or SWI (derived by SSM), and on the morphologic characteristics of the basin extracted from the DEM. This methodology was presented in Silvestro et al. (2013) with respect to LST and it investigates the possibility of calibrating a sub-set of model parameters in an ungauged basin. Since it can be applied using LST or SWI we have two calibration methods to be tested.

The morphologic characteristics mainly influence the surface flow while LST and SWI are more related to subsurface flow.

The estimation of the overland and channel flow parameters is carried out by using geomorphological information derived from the DEM. The methodology is described in Silvestro et al. (2013), and we synthetically report its description in the following steps and in Figure 3:

Step 1. Identify a formulation to estimate the typical lag time (t_{lo}: temporal distance between the centre of mass of the hydrograph and the centre of mass of the mean hyetograph) of a basin based on its main morphologic characteristics. The soil was considered to be completely impermeable so that the subsurface and deep flow parameters (c_t, c_f, R_f and Vw_{max}) therefore become irrelevant.

Step 2. Identify two sections along the streamline of the basin, one at the head of the basin and the other downstream. Estimate the lag-time, t_{lo} based on the DEM and geographical information.

Step 3. Generate a set of synthetic events with constant intensity in space and in time having duration equal to the typical response time of the basin closed at the above mentioned sections (See Table 3).

Step 4. Set a first estimate for the value of u_c and calibrate u_h for each value of P_{cum} referring to the upstream section, using the objective function to minimize:

$$Of = |t_{lo} - t_{lm}| \qquad (22)$$

Where t_{lo} is the t_l derived by the geomorphologic characteristics of the basin while t_{lm} is the t_l obtained from the model simulations. Calculate the average of the u_h values.

Step 5. Fix u_h and calibrate u_c as in Steps 3 and 4, referring this time to the downstream section.

Step 6. Iterate the process until it converges.

According to Silvestro et al. (2013), it is possible to separate the calibration of the two surface flow parameters. In the case of head section with reduced paths in channelized network, the influence of u_c is scarce; as a consequence, an average value of u_c can be set and the calibration can be done only for u_h. The value of u_c is then calibrated based on data from a downstream section with a longer channelized network. This procedure is iterated as shown in Figure 3. Usually 3-4 iterations are sufficient for a good convergence of the process.

Once the surface flow parameters are estimated, the subsurface soil parameters can be evaluated optimizing a proper score between satellite derived and modelled LST (Silvestro et al., 2013) or SWI. In this case we considered the BIAS and Nash Sutcliffe coefficient at basin scale as scores for LST for SWI respectively.

2.4.3. Calibration settings

In both test cases a calibration period has been chosen for each variable (Streamflow, LST, SWI). Using a same period for all the variables is not always the best option, in fact they describe components of the hydrological cycle that need to be sampled in different periods of time, moreover in certain cases data are available on different periods that not always overlap. Furthermore it is interesting to understand what kind of results can be achieved reducing the length of calibration periods but augmenting the number of observed variables. Alternatively, in the case data are available, one could work on longer periods of time that ensure to catch the seasonality of the hydrological processes but with the disadvantage of lengthening the calculation time.

The calibration strategies that have been compared are four:

- The standard approach S.N. described in section 2.4.2.1.
- The M.O. strategy described in section 2.4.2.2 (in this case the calibration periods of the different observations are merged)
- The R.S. approach described in section 2.4.2.3 using as comparison data the satellite LST. Hereafter we will call this strategy **R.S. (LST)**
- The R.S. approach described in section 2.4.2.3 using as comparison data SWI estimation derived from satellite SSM. Hereafter we will call this strategy **R.S. (SWI)**

The periods used for calibration are chosen in order to balance the following characteristics:

- have the presence of different streamflow regimes and soil moisture conditions
- presence of extreme conditions: e.g. flood and drought periods
- presence of reliable data, especially for SWI and LST comparison
- having periods' length manageable in terms of computational time

In the case of Orba basin the calibration was carried out considering the period July – October, 2009 for LST comparison, July – November, 2011 for SWI comparison and August – October, 2006 for streamflow comparison (in this latter period, two intense events preceded by periods of droughts occurred; as a result, the model is forced to work under extreme conditions). The streamflow threshold used in the third component of the M.O. function is $Q_T = 200$ m^3/s. Validation of multi-annual simulations were carried out using the parameters calibrated with the proposed methodologies. The validation period is from January 1st, 2006 to December 31th, 2011; the first five months were used for the model warm up.

In the case of Casentino basin the calibration was carried out considering the July – October, 2009 for LST comparison, July – November, 2011 for SWI comparison and September – December,

2005 for streamflow comparison. The periods were chosen based on the same constraints presented for the Orba case study (see previous paragraph). The streamflow threshold used in the third component of the M.O. function was $Q_T = 200$ m^3/s. The validation period is from January 1st, 2005 to December 31th, 2011; the first five months were used for the model warm up.

Table 4 summarizes the calibration periods for the two test cases.

3. RESULTS

3.1. UNCERTAINTY ANALYSIS

Each score based on streamflow data, and presented in section 2.3.1, can be influenced differently by different flow regimes and hydrograph characteristics, therefore for each simulation the NS was plotted against the other scores (Zappa et. al, 2010); the results are reported in Figure 4 and the graphs show that in all cases there are sets of behavioral parameters (Beven and Binley, 1992) that give similarly good values of the scores, indicating good simulation of the observed streamflow series.

In Figure 5, the dotty plots of the four parameters are reported. Each graph shows the NS value as a function of the parameter values. The variability of NS for a single parameter is quite high. In the case of the two surface parameters u_c and u_h (upper subplots in the figure) a maximum for NS can be identified, while for ct and c_f the behavior of NS is quite homogeneous for all the values in the physically acceptable range. This indicates that u_c and u_h are closely linked to the streamflow simulation in the model regardless of the other parameters value, while the impact of c_f and c_t in the discharge follows more complex paths, and it is hard to identify such parameters by matching the streamflow time series alone. For values of u_c greater than 30÷35 m$^{0.5}$s^{-1} and values of u_h greater than 7÷9 s^{-1} the values of the NS coefficient seem to be uniformly distributed over a large range, this indicates that different combinations of the two parameters can lead to very different performances. For lower values of u_c and u_h the NS coefficient converges to high values highlighting a minor variability of the score and general better performances.

Simulations with NS lower than a fixed threshold (NS=0.4) are considered "non-behavioural" according to Shen et al., (2012). By sorting the discharge time series according to NS values it is possible to evaluate the percentiles at each time step and show the uncertainty in terms of confidence intervals. In Figure 6, the 10% and 90% confidence limits are reported for two time windows across the main streamflow events, which occurred in the considered period. The results

show that the observed streamflow lays in the 90% limit, therefore a parameter configuration that allows reproducing the flow observations exists at any time. Most of the observed hydrographs, and specifically the peak flow, lay in the 10% limits. Part of the receding curve is not included showing some limited ability of the model in the representation of the processes related to the drainage of the soil and aquifers.

The results of the analysis done using the Likelihood function represented by equation 14 are presented in Figures 7 and 8 where the probability density is plotted considering two parameters at a time. In this case, a more evident concentration of the Likelihood function appears when compared to the dotty plots representation of the NS score presented in Figure 5. This is again valid, especially if the case of parameters u_c versus u_h is considered (Figure 7). Anyway various relative maximums of the Likelihood function are present.

The uncertainty analysis using skill scores based on streamflow provides evidence of the presence of equifinal sets of model parameters, this behaviour can be found in other continuous and distributed models; nevertheless, there is a reduced number of parameter sets that generate evidently better performances among all the possible configurations randomly generated. This raises the necessity of finding additional constraints to improve the estimation of the parameters. The focus was then placed on two meteo-hydrological variables whose observations are now widely available from remote sensing techniques: LST and SWI.

By comparing modelled and satellite derived LST (mean at basin scale) it is possible to build dotty-plot representations similar to those presented in Figure 5 but using the BIAS score (equation 11). Figure 9 shows that it is almost impossible to find a well-defined or unique set of surface parameters that minimize BIAS on LST. Same considerations can be done for c_f. The c_t shows an evident trend: this is reasonable since this parameter strongly influences the time of permanence of water in the soil and the LST diurnal dynamics (Caparrini et al., 2004; Sini et al., 2008; Silvestro et al., 2013).

Finally the time series obtained by averaging at basin scale the model saturation degree SD and the satellite SWI maps have been used to build a dotty-plot graph using NS described by equation 12 as score (Figure 10). The maximum of NS lies in the range 0.45-0.55 of the parameter c_t; and a weak, but quite evident independence of c_f arises with optimal values around 0.015-0.025 (close to the lower limit of the parameter range). In both cases the NS values are in a quite narrow range in correspondence of the aforementioned parameters range.

The parameters values individuated by the LST and SWI analysis are consistent with some of the best equifinal parameters combinations on the basis of the streamflow analysis.

3.2. CALIBRATION AND VALIDATION ON MULTI-YEARLY SIMULATIONS

3.2.1. Orba basin

The sets of parameters obtained by the four calibration strategies are reported in Table 5.

The surface flow parameters obtained with the R.S. and M.O. calibration methodologies are similar, while the S.N. method produces slightly lower u_c and higher u_h values. In the case of sub-surface flow, the values are a little bit different for the three considered cases probably because they are more sensitive to the different adopted approaches (Efstratiadis and Koutsoyiannis, 2010).

In the case of R.S. (SWI) the calibrated c_t and c_f values confirm the results of section 3.1, even if the dotty plot in Figure 10 does not show a really strong independence of c_t and c_f from the other parameters, but more a range of the two parameters that shows good simulations of SWI.

In the case of R.S. (LST) results are different in respect to the results of section 3.1 (Figure 9). The c_t optimal parameter has a different value respect to the optimal range found in uncertainty analysis; this is probably related to the fact that the simulations show a complex inter-dependency between the parameter c_t and the parameter c_f both closely related to LST. This ends up increasing the equifinal parameter sets when only LST is used, on the other hand c_t and c_f influence in a complex way the different terms of the multi objective function during the calibration process driving to the direction of a reduction of equifinality. When using the R.S (LST) strategy the benefit of exploiting LST data seems more related to opportunity of doing a calibration in case of lack of streamflow data than in reducing equifinality as already noted in Silvestro et al. (2013).

Figure 11 reports values of the Nash Sutcliffe coefficient (that depends only on the streamflow) versus values of the objective function of the M.O. approach. The parameter values that optimize the single score (S.N.) are not the same that optimize the M.O. function.

M.O. approach appears to be a good way to reduce equifinalty. Looking at the dotty plot representations in section 3.1 built with streamflow, LST and SWI data, it is evident that even when graphs show an independence of a parameter from the others, this independence is not very pronounced. In other cases there is no evident independence. Combining the different objectives showed in section 2.4.2.2 we should eliminate those solutions (parameters combinations) that give good values for a certain metric (for example NS on streamflow) but not optimal values for another one (for example BIAS on LST), thus obtaining an overall better calibration.

Obviously, the choice of the single components of the M.O. function influences the way the various variables impact on the final results, but the applied methodology proposed by Madsen (2000) helps to normalize and balance the weights of the components.

In Table 6 the values of the scores for the validation period are reported while Figures 12 and 13 show the comparison between modeled and observed streamflow. The calibration period belongs to the validation period but it is considerable shorter, we thus decided to estimate the scores on the entire time window. In each figure some significant sub-periods are reported on small panels while in the bottom panel the entire simulation period using a logarithmic scale is shown.

The values of the scores are good in all the cases. The Casalcermelli section performances are better than those for Tiglieto, this may be due to the fact that the first section corresponds to a larger drainage area and therefore the integration effects smooth the uncertainties of the rainfall fields. The M.O. approach leads to score values on the streamflow similar to the S.N. method in the validation period, while the R.S. approach produced poorer performance with respect to the other two approaches. In the case of PFRE the two best sets of parameters are S.N. and R.S.(SWI). Notwithstanding, all the parameters sets led to good results in terms of the modelled hydrographs.

The M.O. calibration strategy leads to good performances in reproducing the observed streamflows despite the fact that these latter measurements are not the only ones used in the calibration process; there are good performances over long periods of simulation for both of the considered outlet sections. The peak flows and the time of the peak flows are generally well reproduced as well as the periods of flow recession and drought between the most relevant events. In general M.O. produces better results with respect to S.N. when it comes to Correlation and CM, especially in the Tiglieto section where the uncertainty in rainfall input hampers the S.N. calibration strategy performance and the advantage of having more sources of observation is more evident.

The series of mean LST at basin scale were compared with LST satellite estimation. A similar comparison was made between the satellite SWI and modelled SD. Tables 7 and 8 show the values of the scores for the four considered parameters sets.

The classic calibration obtained with the maximization of the Nash Sutcliffe coefficient of the streamflow (S.N.) allows obtaining good performances in terms of streamflow simulation, but it produces higher values of LST BIAS, while the M.O. approach balances between the different components. The reproduction of SWI is quite good for both the M.O. and S.N. cases.

The accumulated discharge volume simulated by the model was compared with that derived from the streamflow observations in order to verify the behavior of the model in terms of total runoff volumes. The results are reported in Figure 14. The model reproduces with fine approximation the observed volumes; the error on the entire period is approximately -1.9 %, -1.3 %, -3.0% and -2.5 % for M.O, R.S.(LST), R.S.(SWI) and S.N., respectively. These errors are probably lower than the uncertainties introduced by the level-discharge transformation.

3.2.2. Casentino basin

The sets of parameters obtained by the four calibration strategies are reported in Table 9.

In Table 10 the scores are reported, while Figures 15 and 16 show the comparison between modeled and observed streamflow. The Figures report on small panels different significant events, while in the bottom panel the entire simulation period is shown using a logarithmic scale.

Tables 11 and 12 shows the values of the scores for the variables LST and SWI. The values of the scores are good in all the cases; they are better for M.O. with respect to S.N. for both LST and SWI variables.

The model delivers good performances over long periods for all the four considered calibration strategies. As in the case of Orba basin, M.O. finds a compromise parameter set that allow to obtain a good modeling of all the three variables involved in the objective functions, while the two R.S: method shows that is possible to obtain a reasonable calibration even in the case of ungauged basins. The peak flows are better reproduced by the configurations S.N. and R.S.(SWI) which show PFRE values smaller than 0.1 (10%).

The accumulated volume over the 7 years of simulation is generally well simulated (Figure 17); in this case, there is a larger difference between the total volumes obtained with the four different parameter sets, the errors are in fact of the order of 9%, 8.8%, 5.3% and 4% for M.O., R.S.(LST), R.S.(SWI) and N.S., respectively. The errors on the total volume are a larger than the case of Orba basin.

4. DISCUSSION AND CONCLUSIONS

This paper shows that satellite data are useful in reducing the uncertainty of the parameterization of a distributed hydrological model and that they can be used in calibration strategy to improve model representation of hydrological processes.

The uncertainty analysis (Zappa et al, 2010; Shen et al., 2012) of the most sensitive parameters shows that the equifinality can be reduced using Land Surface Temperature and Soil Water Index satellite estimations. The independence of surface flow parameters seem to be linked in a clear way to streamflow, while soil parameters are more directly linked to LST and SWI.

Three methodologies to estimate a subset of the parameters of the model by exploiting remote sensing were applied. The first methodology consists of the minimization of a multi-objective function that depends on streamflow, LST and SWI and it is inspired by the work of Sutanudjaja et al. (2014) and Wanders et al. (2014). The second and third methodologies simulate the case when no streamflow data are available and the calibration is carried out based only on LST and SWI retrieved from satellite data and information derived from a DEM. A multi-year period validation was done in terms of reproduction of both streamflow time series and total volume over the considered period. A comparison with a fourth standard calibration strategy based on streamflow data was also carried out.

The skill scores on streamflow show good performances when satellite data are involved in the calibration process (M.O. and R.S. methods), comparable with values obtained using only the streamflow in the objective function (S.N. method); even if the observed and simulated streamflow are in some cases quite different, the general trend is good and there are not large biases in terms of runoff volumes over long simulation periods. The largest errors seem to be more related to the uncertainties of the input rainfall fields rather than on the model parameterization. Moreover, the skill scores on LST and SWI have generally better values in the case of parameter sets derived by the multi objective approach with respect to those obtained by the streamflow based calibration strategy.

Both the results of calibration, especially M.O. approach, and uncertainty analysis confirm that a way to reduce equifinality and to augment the parameter constraints is related to the increase of model state variables and model output variables that can be derived from both gauge and remote sensing data. This helps to reduce the possibility of obtaining similar results with a large number of parameter sets. We can thus state that the presented work explored the direction proposed by Seibert and McDonnel (2002) and Efstratiadis and KoutsoYiannis (2010), which consisted of obtaining a better overall performance of the model and ensure consistency across its various aspects.

In addition, remote sensing data (in this specific case the LST and SSM) offer alternative ways to carry out parameter calibration in cases where no streamflow data might be available. Satellite derived data such as DEM, SSM and LST are generally universally available.

The described methodologies can be adapted and applied to other hydrological models that have characteristics similar to Continuum and that can simulate LST and soil moisture as state variables in a prognostic way; moreover, these methods can be extended by referring to other remote sensing data, and in general observed data, that can be reproduced by the model. More the model has the

capability of reproducing observable quantities (e.g. evapotranspiration, soil humidity) more the constraints that can be imposed to the model can increase.

Finally, the results of the presented work can be read from two different points of view. On one hand, they highlight the advantages of using distributed hydrological models that allow for the reproducing with some degree of detail the physical processes, such models in fact, simulate a larger number of variables which can also be observed. On the other hand, similarly to what demonstrated by other authors (Corbari and Mancini (2014); Montzka et al., 2011; Sutanudjaja et al., 2014; Wanders et al., 2014), they highlight the opportunities given by remote sensing and the necessity of augmenting the number (and the quality) of these data. Remotely sensed data can in fact be used to parameterize hydrological models and to set up constraints to the parameters in the calibration process, while offering an alternative way of calibrating these models where standard observation are lacking.

5. ACKNOWLEDGMENTS

This work is supported by the Italian Civil Protection Department, and by the Italian Regions of Valle d'Aosta and Liguria. We acknowledge the Italian Civil Protection Department for providing us with the data from the regional meteorological observation networks. We also thank the H-SAF project for the availability of Surface Soil Moisture data.

6. APPENDIX A

The snow accumulation-melting module was introduced in order to carry out multi-year simulations in alpine climates. It is a simple model that is derived from commonly used equations (Maidment, 1992) and it is forced by meteorological observations.

The equations that describe the snow mass conservation and its melting are the following:

$$\frac{\Delta SWE}{\Delta t} = S_f - SM \qquad \text{(A1)}$$

where SWE is the snow water equivalent, Sf is the solid precipitation and SM is the snow melting estimated as:

$$SM = \frac{R_n}{\rho_w \lambda_f} + m_c \cdot (T_a - T_0) \qquad \text{(A2)}$$

Where Rn is the net radiation, ρw the water density, λf the latent heat of melting, Ta the air temperature. T0 and mc are two parameters that represent the temperature at which the melting starts and the melting coefficient, respectively. These two parameters are estimated using values from the literature (Maidment, 1992), T_0=0 °C and mc=4 mm/day.

The mass balance is applied at cell scale for the entire domain of the model, so that a snow cover map can be generated with the same resolution of the DEM. The energy balance and, as a consequence, the evapotranspiration are inhibited for those cells where snow cover is present.

The applied approach is very simple and neglects the heat exchanges between the soil and the snow cover, but it is generally sufficient if the goal is the estimation of the snow contribution to the runoff, especially when the regime of the basin is not strongly influenced by snow melting.

The precipitation is partitioned into solid or liquid if the air temperature is below or above a fixed threshold.

7. REFERENCES

Albergel, C., Rüdiger, C., Pellarin, T., Calvet, J. C., Fritz, N., Froissard, F., Suquia, D., Petitpa, A., Piguet, B., & Martin, E., (2008), From near-surface to root-zone soil moisture using an exponential filter: An assessment of the method based on in-situ observations and model simulations. Hydrology and Earth System Sciences, 12, 1323–1337, doi:10.5194/hess-12-1323..

Albergel C., de Rosnay P., Gruhierand C., J.Munoz-Sabater, Hasenauer S., Isaksen L., Kerr Y., Wagner W. (2012), Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. Remote Sensing of Environment, 118, 215-226

Bartalis, Z., Wagner, W., Naeimi, V., Hasenauer, S., Scipal, K., Bonekamp, H. Figa, J., c. Anderson, C., (2007), Initial soil moisture retrievals from the METOP-A Advanced Scatterometer (ASCAT). – Geophys. Res. Lett., 34, l20401.

Bartholomes, J., Todini, E. (2005), Coupling meteorological and hydrological models for flood forecasting. *Hydrology and Earth System Science*, 9, 333 – 346.

Beven, K., Binley, A. M., (1992), The future of distributed models: model calibration and uncertainty prediction. *Hydrol. Proc.*, 24, 43-69.

Bitew, M. M., & Gebremichael, M. (2011), Evaluation of satellite rainfall products through hydrologic simulation in a fully distributed hydrologic model. *Water Resources Research*, *47*(6).

Bjerklie, D. M., Dingman, S. L., Vorosmarty, C. J., Bolster, C. H., & Congalton, R. G. (2003). Evaluating the potential for measuring river discharge from space. *Journal of Hydrology*, *278*(1), 17-38.

Brakenridge G. R., Nghiem S. V., Anderson E. A., Chien S. (2005), Space-based Measurement of River Runoff. *EOS Trans.* AGU 86(19): 185.

Brakenridge, G. R., Cohen, S., Kettner, A. J., De Groeve, T., Nghiem, S. V., Syvitski, J. P., & Fekete, B. M. (2012). Calibration of satellite measurements of river discharge using a global hydrology model. *Journal of Hydrology*, *475*, 123-136.

Brocca, L., Melone, F., Moramarco, T., Wagner, W., & Hasenauer, S. (2010)., ASCAT soil wetness index validation through in situ and modelled soil moisture data in cen- tral Italy. Remote Sensing of Environment, 114 (11), 2745–2755, doi: 10.1016/j.rse.2010.06.009.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., Bittelli, M., (2011a), Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe, *Remote Sensing of Environment*,115, 3390–3408.

Brocca, L., Melone, F., Moramarco, T., (2011b), Distributed rainfall-runoff modelling for flood frequency estimation and flood forecasting, *Hydrol. Process*., 25, 2801–2813.

Caparrini, F., Castelli F., Entekhabi, D., (2004), Estimation of surface turbulent fluxes through assimilation of radiometric surface temperature sequences, *Journal of Hydrometeorology*, 5, 145-159.

Carpenter, T. M. and Georgakakos, K. P., (2006), Intercomparison of lumped versus distributed hydrologic model ensemble simulations on operational forecast scales. *Journal of Hydrology*, 329, 174– 185.

Coccia G, Mazzetti C, Ortiz EA, Todini E., (2009), Application of the TOPKAPI model within the DMIP 2 project. Proceedings of the International Conference, 23rd Conference on Hydrology, Phoenix, Arizona.

Corbari, C., Mancini, M., (2014), Calibration and Validation of a Distributed Energy–Water Balance Model Using Satellite Data of Land Surface Temperature and Ground Discharge Measurements, *Journal of Hydrometeorology*, 15, 376–392. doi: http://dx.doi.org/10.1175/JHM-D-12-0173.1.

Crow, W. T., Wood E.F., Pan, M., (2003) Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals," *Journal of Geophysical Research - Atmospheres*, 108(D23), doi:10.1029/2002JD003292.

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., and Walker, J. P. (2012). Upscaling sparse ground based soil moisture observations for the validation of coarse resolution satellite soil moisture products. *Reviews of Geophysics*, *50*(2).

Chiew, F., Mc Mahon, T., (1994), Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments, *Journal of Hydrology*, 153, 383-416.

Dickinson, R., (1988), The force-restore method for surface temperature and its generalization, *Journal of Climate*, 1:1086-1097.

EUMETSAT, (2008), Product user manual, vegetation parameters. http://landsaf.meteo.pt/

EUMETSAT, (2009), Algorithm Theoretical Basis Document for Land Surface Temperature (LST). http://landsaf.meteo.pt/

Efstratiadis, A., Koustoyiannis, D., (2010), One decade of Multi-Objective calibration approaches in hydrological modelling: a review. *Hydrological Science Journal*, 55(1), 58-78.

Flores, A. N., Entekhabi, D., Bras, R. L., (2010), Reproducibility of soil moisture ensembles when representing soil parameter uncertainty using a Latin Hypercube–based approach with correlation control, *Water Resour. Res*., 46, W04506, doi:10.1029/2009WR008155.

Gabellani, S., Silvestro, F., Rudari, R., Boni, G., (2008), General calibration methodology for a combined Horton-SCS infiltration scheme in flash flood modeling, *Nat. Hazards Earth Science*., 8, 1317 - 1327.

Giannoni, F., Roth., G., Rudari, R., (2000), A Semi – Distributed Rainfall – Runoff Model Based on a Geomorphologic Approach, *Physics and Chemistry of the Earth*, 25/7-8, 665-671.

Giannoni, F., Roth., G., Rudari, R., (2005), A procedure for drainage network identification from geomorphology and its application to the prediction of the hydrologic response, *Advances in Water Resources*, 28, 6, 567-581.

Göttsche, F. M., & Hulley, G. C. (2012). Validation of six satellite-retrieved land surface emissivity products over two land cover types in a hyper-arid region. *Remote Sensing of Environment*, *124*, 149-158.

Khan, S. I., Hong, Y., Wang, J., Yilmaz, K. K., Gourley, J. J., Adler, R. F., Brakenridge, G. R., Policelli, F., Habib, S., and Irwin, D. (2011). Satellite remote sensing and hydrologic modeling for flood inundation mapping in Lake Victoria Basin: Implications for hydrologic prediction in ungauged basins. *Geoscience and Remote Sensing, IEEE Transactions on*, *49*(1), 85-95.

Kim, S. M., Benham, B. L. , Brannan, K. M., Zeckoski, R. W. ,Doherty, J., (2007), Comparison of hydrologic calibration of HSPF using automatic and manual methods, *Water Resour. Res*., 43, W01402, doi:10.1029/2006WR004883.

Koren, V., Moreda, F., Smith M., (2008): Use of soil moisture observations to improve parameter consistency , *J. Phys. Chem. Earth* , doi:10.1016/j.pce.2008.01.003

Kozak, J. A., Ahuja, L. R., Green, T. R., Ma, L., (2007), Modelling crop canopy and residue rainfall interception effects on soil hydrological components for semi-arid agriculture, *Hydrol. Process*. 21, 229–241.

Kunstmann, H., Krause, J. & Mayr, S., (2006), Inverse distributed hydrological modelling of Alpine catchments, *Hydrol. Earth System Sci.* ,10, 395–412.

Lin, J., D., (1980), On the Force-Restore method for prediction of ground surface temperature, *Journal of Geophysical Research*, 85, 3251-3254.

Liu., Z., Martina, M. L. V., Todini, E., (2005), Flood forecasting using a fully distributed model: application of the TOPKAPI model to Upper Xixian Catchment, *Hydrology and Earth System Science,* 9, 347–364.

Madsen, H., (2000), Automatic calibration of a conceptual rainfall–runoff model using multiple objectives, *Journal of Hydrology*, 235, 276–288.

Madsen, H., (2003), Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. *Adv. Water Resour*. 26, 205–216.

Maidment, D., (1992), Handbook of Hydrology. McGraw-Hill, Inc.

Montaldo, N., Rondena. R., Albertson, J.D., Mancini. M., (2005), Parsimonious modelling of vegetation dynamics for ecohydrological studies of waterlimited ecosystems. *Water Resources Research*, 41: W10416.

Montzka, C., Moradkhani, H., Weihermuller, L., Franssen, H.-J. H., Canty, M., Vereecken, H. (2011), Hydraulic parameter estimation by remotely-sensed top soil moisture observations with the particle filter, *Journal of Hydrology*, 399 (34), 410 – 421, doi:10.1016/j.jhydrol.2011.01.020

Murray, S. J., Watson, I. M., & Prentice, I. C. (2013), The use of dynamic global vegetation models for simulating hydrology and the potential integration of satellite observations. *Progress in Physical Geography, 37*(1), 63-97. doi:http://dx.doi.org/10.1177/0309133312460072

Naeimi, V., Scipal, K., Bartalis, Z., Hasenauer, S., Wagner, W., (2009), An improved soil moisture retrieval algorithm for ERS and METOP scatterometer observations. IEEE Transaction on Geoscience and Remote Sensing 47, 1999–2013.

Nash, J. E., Sutcliffe, J. V, (1970), River flood forecasting through conceptual models I: a discussion of principles, *Journal of Hydrology*, 10, 282-290.

Parajka, J., Naeimi, V., Bloschl, G., Wagner, W., Merz, R., and Scipal, K., (2006), Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale, Hydrol. Earth Syst. Sci., 10, 353–368, doi:10.5194/hess-10-353-2006.

Pavelsky, T. M. (2014). Using width-based rating curves from spatially discontinuous satellite imagery to monitor river discharge. *Hydrological Processes* 28, 3035–3040.

Perrin, C., Michel, C., Andréassian, V., (2001), Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *J. Hydrol.* 242(3-4), 275–301.

Ridler, M. E., Sandholt, I., Butts , M., Lerer, S., Mougin, E., Timouk, F., Kergoat, L., Madsen, E. (2012), Calibrating a soil–vegetation–atmosphere transfer model with remote sensing estimates of surface temperature and soil surface moisture in a semi arid environment, *Journal of Hydrology* 436–437 (2012) 1–12.

Rhoads, J., Dubayah, R., Lettenmaier, D., O'Donnell, G., Lakshmi, V., (2001), Validation of land surface models using satellite derived surface temperature, *Journal of Geophysical Research*, 106 (D17), 20,085-20,099.

Rakovec, O., Hill, M. C. ,Clark, M. P. , Weerts, A. H., Teuling, A. J., Uijlenhoet, R., (2014), Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models*, Water Resour. Res.*, 50, doi:10.1002/2013WR014063.

Santanello, J., Peters-Lidard, C. D., Garcia, M. E., Mocko, D. M., Tischler, M. A., Moran, M. S., Thoma, D., (2007), Using remotely-sensed estimates of soil moisture to infer soil texture and hydraulic properties across a semi-arid watershed, *Remote Sensing of Environment*, 110 (1), 79 – 97, doi:10.1016/j.rse.2007.02.007

Savenije, H. H.. G., (2001), Equifinality, a blessing in disguise?, *Hydrol. Proc.,* 15, 2835-2838.

Schmugge, T. J., Kustas, W. P., Ritchie, J. C., Jackson, T. J., Rango, A., (2002), Remote sensing in hydrology, *Advances in Water Resources*, 25, 1367–1385.

Shen, Z. Y., Chen, L., and Chen, T., (2012), Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: a case study of SWAT model applied to Three Gorges Reservoir Region, China, *Hydrol. Earth Syst. Sci.*, 16, 121–132, doi:10.5194/hess-16-121-2012.

Silvestro, F., Gabellani, S., Delogu, F., Rudari, R., Boni, G., (2013), Exploiting remote sensing land surface temperature in distributed hydrological modelling: the example of the Continuum model. Hydrol. Earth Syst. Sci., 17, 39-62, doi:10.5194/hess-17-39-2013.

Singh, S.K., Bárdossy, A., (2012), Calibration of hydrological models on hydrologically unusual events, *Advances in Water Resources*, 38, 81–91. doi:10.1016/j.advwatres.2011.12.006

Sini, F., Boni, G., Caparrini, F., & Entekhabi, D. (2008). Estimation of large-scale evaporation fields based on assimilation of remotely sensed land temperature. *Water resources research*, *44*(6), DOI: 10.1029/2006WR005574.

Stroud, P. D., (1999), A recursive exponential filter for time-sensitive data, Los Alamos national Laboratory, LAUR-99-5573, available at: public.lanl.gov/stroud/ExpFilter/ExpFilter995573.pdf, (last access July 2008).

Sutanudjaja, E. H., van Beek, L. P. H. , de Jong, S. M., van Geer, F. C. , Bierkens, M. F. P., (2014), Calibrating a large-extent high-resolution coupled groundwater-land surface model using soil moisture and discharge data, *Water Resources Research*, 50, 687-705, doi:10.1002/2013WR013807.

Todini, E., Ciarapica, L., (2001), The TOPKAPI Model. Mathematical Models of Large Watershed Hydrology. In: Singh, V. P. et al. (eds) Water Resources Publications, Littleton, Colorado, Chapter 12.

Todini, E., (2009), History and perspective of hydrological catchment modelling. In Water, Environment, Energy and Society, Jain SK, Singh VP, Kumar V, Kumar R, Singh RD, Sharma KD (eds). *Proc. Of WEES-2009*, Allied Publishers Pvt. Ltd: New Delhi (India), 512–523.

Vörösmarty, C., Askew, A., Grabs, W., Barry, R. G., Birkett, C., Döll, P., Goodison, B., Hall, A., Jenne, R., Kitaev, L., Landwehr, J., Keeler, M., Leavesley, G., Schaake, J., Strzepek, K., Sundarvel, S. S., Takeuch K. and Webster, F., (2001), Global water data: A newly endangered species. *Eos, Transactions American Geophysical Union*, *82*(5), 54-58.

Vrugt, J. A., Gupta, H. V., Bouten, W. & Sorooshian, S., (2003), A Shuffled Complex Evolution Metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resour. Res.* 39(8), doi:10.1029/2002WR001642.

Wagner W., Lemoine G., Rott H., (1999), A method for estimating soil moisture from ERS scatterometer and soil data. Remote Sensing Environment 70: 191-207.

Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Zuger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Bloeschl, G., Eitzinger, J., Steinnocher, K., Zeil, P., Rubel, F., (2013), The ASCAT soil moisture product: specifications, validation results, and emerging applications. Meteorologische Zeitschrift, 22(1), 5-33, doi:10.1127/0941-2948/2013/0399. http://dx.doi.org/10.1127/0941-2948/2013/0399.

Wanders, N., Bierkens, M. F. P., Jong, S. M., Roo, A., Karssenberg, D., (2014), The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models, Water Resour. Res., 50, 6874–6891, doi:10.1002/2013WR014639.

Ward, A. D., Trimble, W., *Environmental Hydrology*, Second Edition Published: by CRC Press - 504 Pages, December 2003.

White, C. D., Lewis, M: M., (2011), A new approach to monitoring spatial distribution and dynamics of wetlands and associated flows of Australian Great Artesian Basin springs using QuickBird satellite imagery. *Journal of Hydrology,* 408, 140–152.

Wooding, R. A., (1965), A hydraulic modeling of the catchment-stream problem. 1. Kinematic wave theory, *Journal of Hydrology*, 3, 254–267.

Xu, D., Wang, W., Chau, K., Cheng C., Chen, S., (2013), Comparison of three global optimization algorithms for calibration of the Xinanjiang model parameters. Jourlan of Hydroinformatics, 15.1, 174-193.

Yapo, P. O., Gupta, H. V., Sorooshian, S., (1998), Multi-Objective global optimization for hydrological models. *Journal of Hydrology,* 204, 83-97.

Yu, Y., Tarpley, D., Privette, J. L., Flynn, L. E., Xu, H., Chen, M., ... & Tian, Y. (2012). Validation of GOES-R satellite land surface temperature algorithm using SURFRAD ground measurements and statistical estimates of error properties. *Geoscience and Remote Sensing, IEEE Transactions on*, *50*(3), 704-713.

Zhang, T., Stackhouse Jr, P. W., Gupta, S. K., Cox, S. J., Colleen Mikovitz, J., & Hinkelman, L. M. (2013). The validation of the GEWEX SRB surface shortwave flux data products using BSRN measurements: A systematic quality control, production and application approach. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *122*, 127-140.

Zappa, M., Jaun, S., Germann, U., Walser, A. and Fundel, F. (2011), Superposition of three sources of uncertainties in operational flood forecasting chains. *Atmos. Res.*, doi:10.1016/j.atmosres.2010.12.005.

8. TABLES

Parameter	Physical process parameterized
u_h [s^{-1}]	Flow motion in hillslopes
u_c [$m^{0.5}s^{-1}$]	Friction in channels
c_f [-]	Infiltration capacity at saturation
c_t [-]	Mean field capacity
R_f [-]	Anisotropy between the vertical and horizontal saturated conductivity and soil porosity
V_{Wmax} [mm]	Maximum storage capacity of the aquifer

Table 1:the parameters of Continuum model that need calibration at basin scale are reported with a brief description.

Parameter	Unit	Min	Max
c_t	[-]	0.15	0.65
c_f	[-]	0.015	0.1
u_c	$m^{0.5}/s$	15	55
u_h	1/s	0.0002	0.0015

Table 2 range of variability of the parameters used in the calibration and uncertainty analysis.

Parameter	P_{cum} [mm]	Reference Section	Area [km^2]	t_{lo} [hours]	Duration [hours]
u_h [s^{-1}]	10, 20,…, 60, 70	Tiglieto	75	4.5	4
u_c [$m^{0.5}s^{-1}$]	10, 20,…, 60, 70	Casalcermelli	800	11.6	10
u_h [s^{-1}]	10, 20,…, 60, 70	Upstream (no Gauge)	58	2.98	4
u_c [$m^{0.5}s^{-1}$]	10, 20,…, 60, 70	Subbiano	670	8.4	9

Table 3: characteristics of the synthetic rainfall events used for the estimation of the two surface parameters in the R.S. approach. The event length is set depending on the considered basins and sections. Sections with smaller upstream drainage area are used to estimate u_h while sections with larger area to estimate u_c.

Basin	Q	LST	SWI
Orba	08-10/2006	07-10/2009	07-11/2011
Casentino	09-12/2005	07-10/2009	07-11/2011

Table 4: summary of the periods considered for the calibration process for each variable. In the case of M.O. approach the 3 periods are merged.

Basin	Orba			
Parameter	S.N.	R.S.(LST)	R.S.(SWI)	M.O.
u_c [$m^{0.5}$/s]	29.0	29.42	29.42	29.92
u_h [1/s]	0.00052	0.000458	0.000458	0.00041
c_t [-]	0.52	0.56	0.58	0.40
c_f [-]	0.020	0.030	0.018	0.020

Table 5: The Orba basin parameters were calibrated following the different approaches: using the multi-objective function (M.O.), remote sensing data and morphologic characteristics (R.S.), and the standard hydrographs comparison (S.N.). The table reports the values obtained by the 4 different calibration strategies.

Basin	Section	Parameter set	N.S.	CM	RMSE	CORR	PFRE
Orba	Casalcermelli	M.O.	0.82	0.81	1.67	0.91	-0.162
		R.S.(LST)	0.81	0.83	1.41	0.90	-0.133
		R.S. (SWI)	0.82	0.82	1.35	0.90	-0.063
		S.N.	0.83	0.82	1.31	0.91	-0.095
	Tiglieto	M.O.	0.69	0.65	0.80	0.87	-0.251
		R.S. (LST)	0.67	0.62	0.78	0.83	-0.245
		R.S. (SWI)	0.66	0.63	0.76	0.81	-0.143
		S.N.	0.66	0.62	0.70	0.80	-0.213

Table 6: Orba basin skill scores on hydrographs for the entire validation period (2006 - 2011) and for the different parameters sets. The scores value is calculated for the two sections with available streamflow observations. PFRE is calculated with Q_{th}=200 m^3/s for Casalcermelli section and with Q_{th}=50 m^3/s for Tiglieto section

Variable	Parameter Set	BIAS
LST	M.O.	0.89
	R.S. (LST)	0.58
	R.S.(SWI)	0.92
	S.N.	1.02

Table 7: Orba basin. Values of Bias on LST series obtained with the different parameters sets. The basin scale mean has been considered.

Variable	Parameter Set	NS
SWI	M.O.	0.74
	R.S.(LST)	0.56
	R.S.(SWI)	0.80
	S.N.	0.77

Table 8: Orba basin. Values of NS on SWI series obtained with the different parameters sets. The basin scale mean has been considered.

Basin	Casentino			
Parameter	S.N.	R.S. (LST)	R.S. (SWI)	M.O.
u_c [$m^{0.5}$/s]	28.51	43.02	43.02	47.43
u_h [1/s]	0.00052	0.00047	0.00047	0.00043
c_t [-]	0.44	0.49	0.52	0.51
c_f [-]	0.018	0.032	0.018	0.029

Table 9: The Casentino basin parameters were calibrated following the different approaches: using the multi-objective function (M.O.), remote sensing data and morphologic characteristics (R.S.), and the standard hydrographs comparison (S.N.). The table reports the values obtained by the 4 different calibration strategies.

Basin	Section	Parameter set	NS	CM	RMSE	CORR	PFRE
Casentino	Subbiano	M.O.	0.80	0.77	2.37	0.89	-0.251
		R.S. (LST)	0.78	0.74	2.39	0.88	-0.311
		R.S. (SWI)	0.79	0.75	2.25	0.88	0.092
		S.N.	0.81	0.75	2.23	0.89	0.075

Table 10: Casentino basin skill scores on hydrographs on the entire validation period (2005-2011) for the different parameters sets. The scores value is calculated for the Subbiano section. PFRE is calculated with Q_{th}=200 m^3/s.

Variable	Parameter Set	BIAS
LST	M.O.	1.83
	R.S. (LST)	1.79
	R.S. (SWI)	1.87
	S.N.	1.88

Table 11: Casentino basin. Values of BIAS on LST series obtained with the different parameters sets. The mean basin scale has been considered

Variable	Parameter Set	NS
SWI	M.O.	0.80
	R.S. (LST)	0.63
	R.S. (SWI)	0.83
	S.N.	0.73

Table 12: Casentino basin. Values of NS on SWI series obtained with the different parameters sets. The mean basin scale has been considered

9. FIGURES

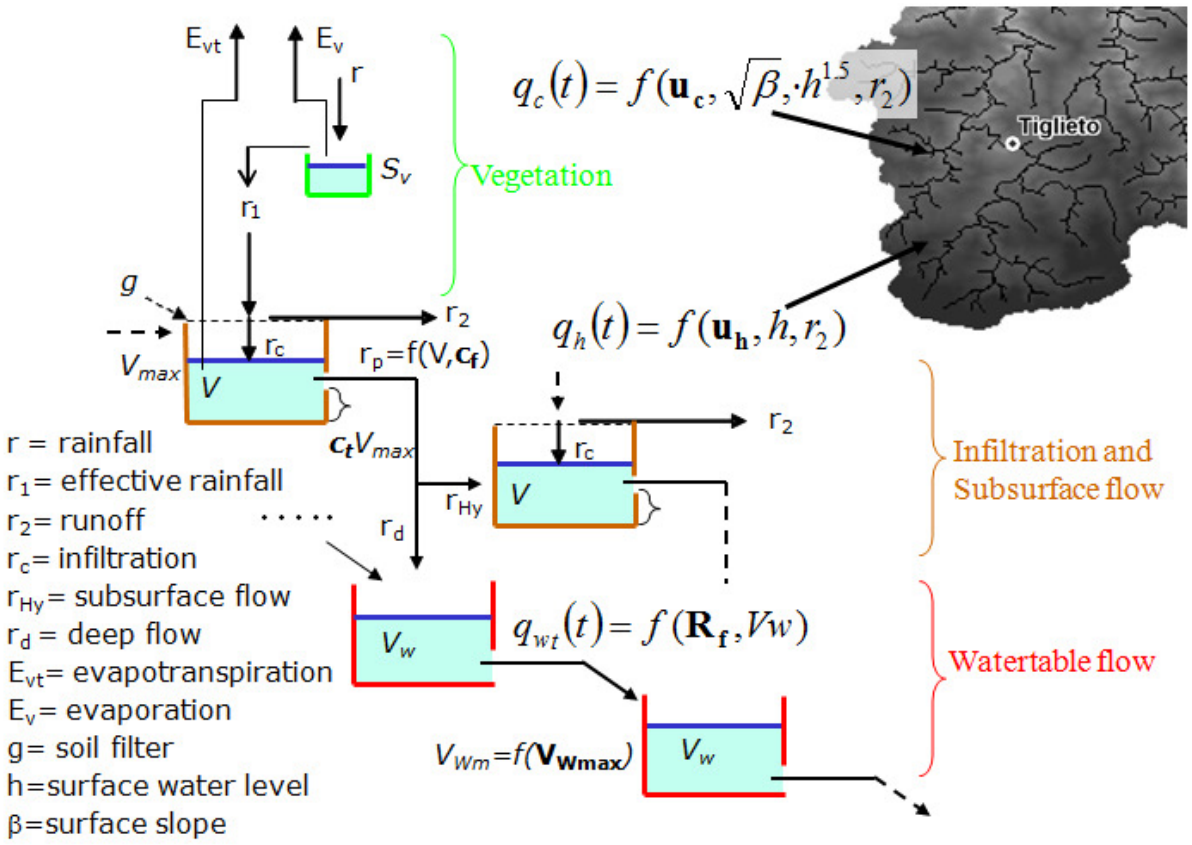


Figure 1. Representation of the different processes described in Continuum model and how different cells are connected. Surface flow is described by non-linear and linear motion equations respectively on channels (q_c) and hillslopes (q_h). Two consecutive cells are represented, each cell has the x and y dimensions that correspond with the spatial resolution of the model.

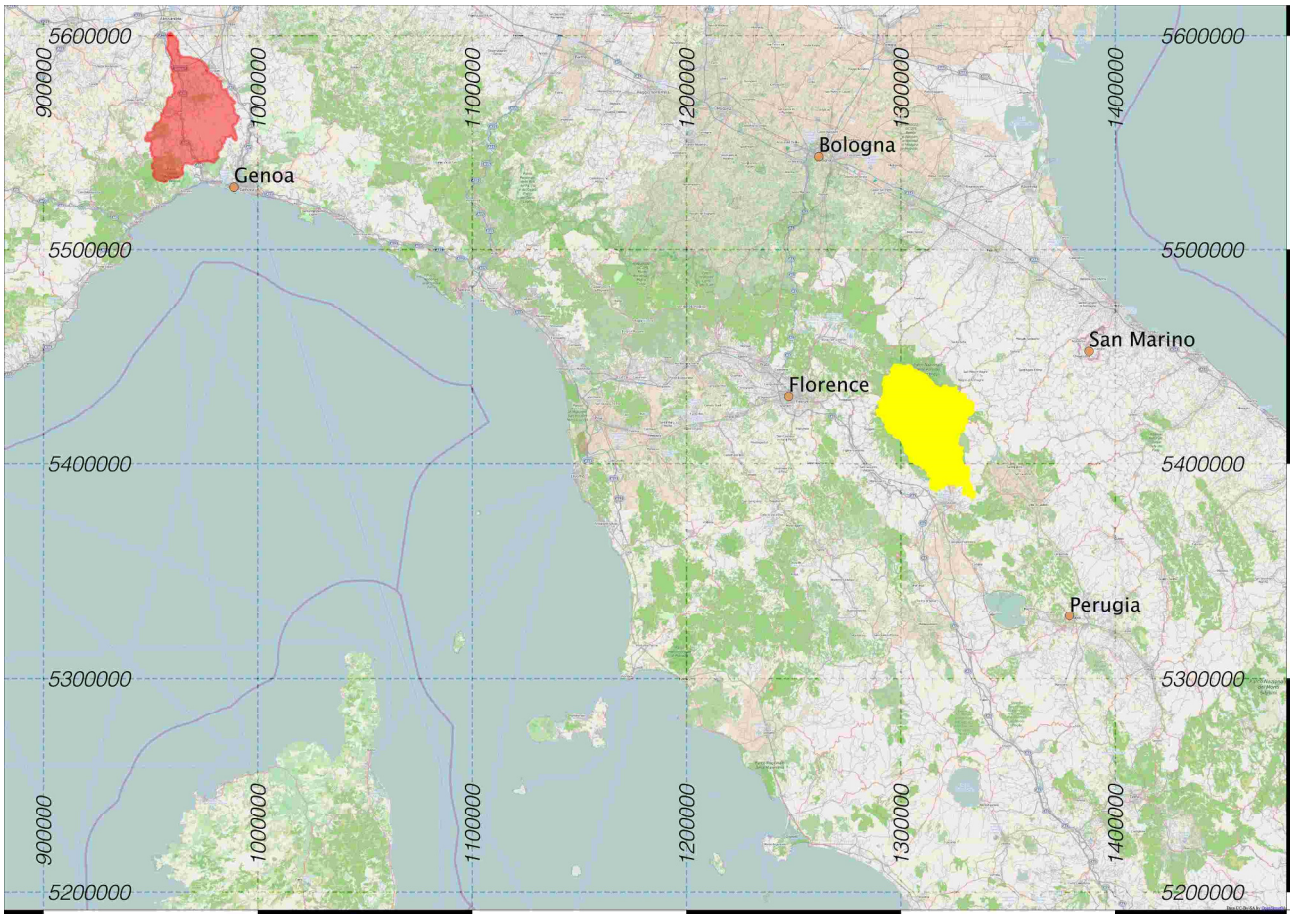


Figure 2. Location of the study areas. The Orba basin (in red) in north west of Italy and the Casentino basin (in yellow) in central Italy.

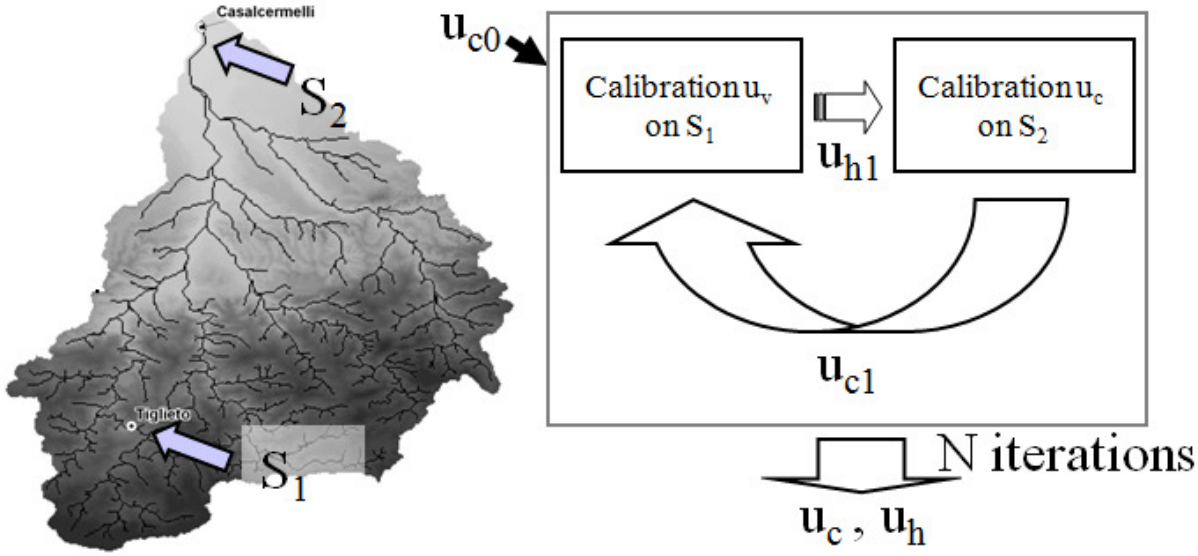
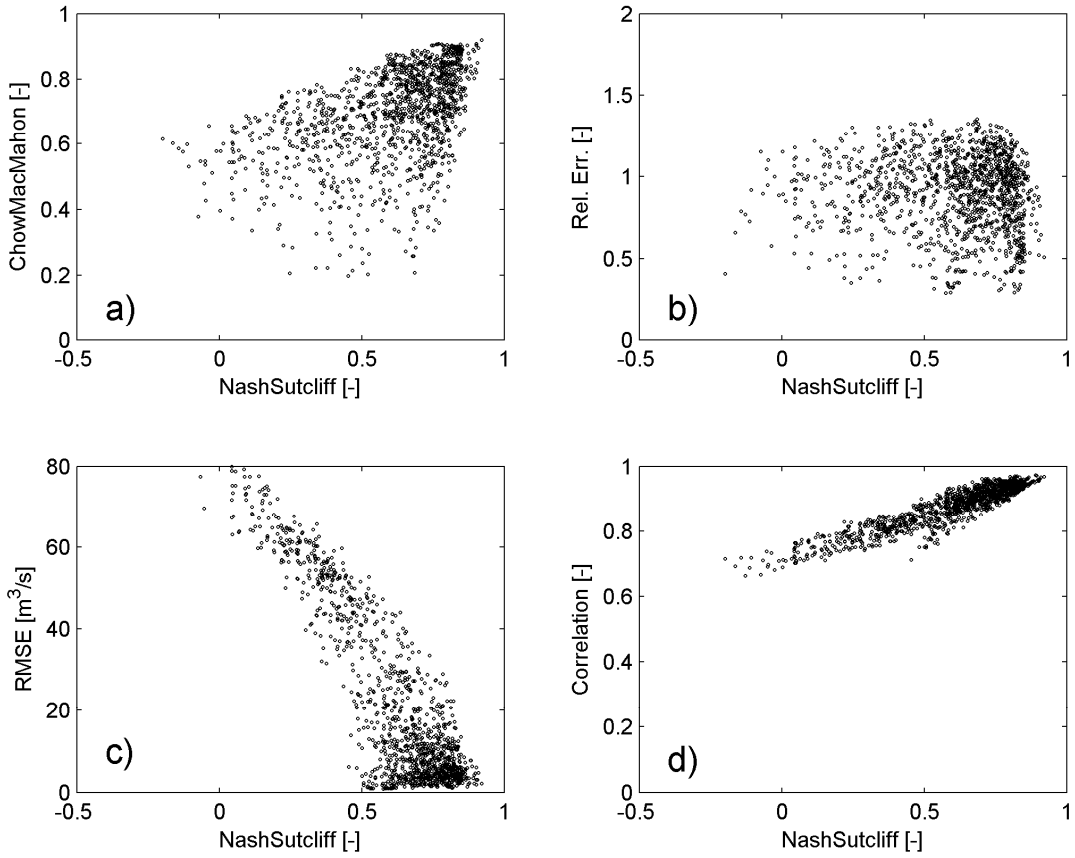


Figure 3: Scheme of calibration of the surface parameters in R.S. approach based on the reproduction of the basin lag time at two different closure sections using synthetic rainfall events. A first guess u_c is used to calibrate u_h on section S_1, this latter is used to calibrate u_c on S_2, then the procedure is iterated for N times.



1

2 Figure 4: Orba basin, August-September 2006. NS on the hydrographs (X axis) versus other
3 statistics (Y axis): Chiew McMahon coefficient (CM), Root Mean Square Error (RMSE), the
4 correlation coefficient (CORR) and the relative error (Rel. Err.). The graphs show that in all cases
5 there are sets of parameters that give similarly good values of the compared scores.

6

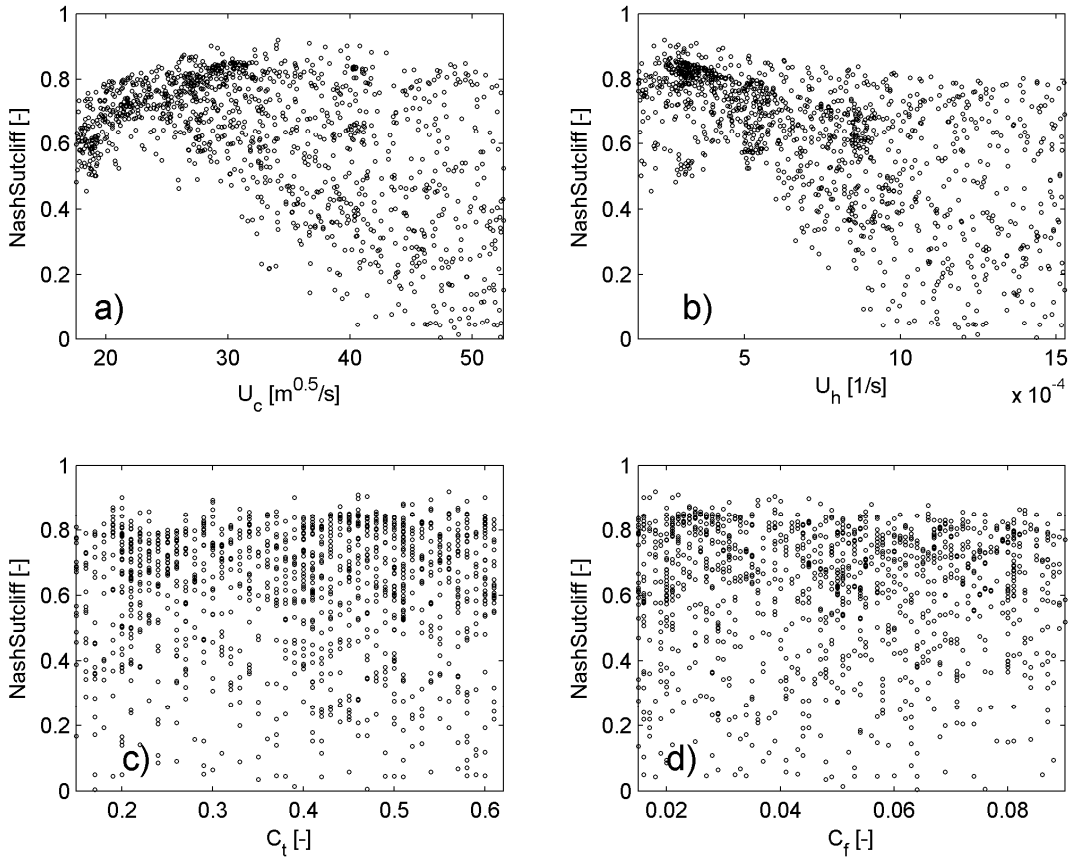
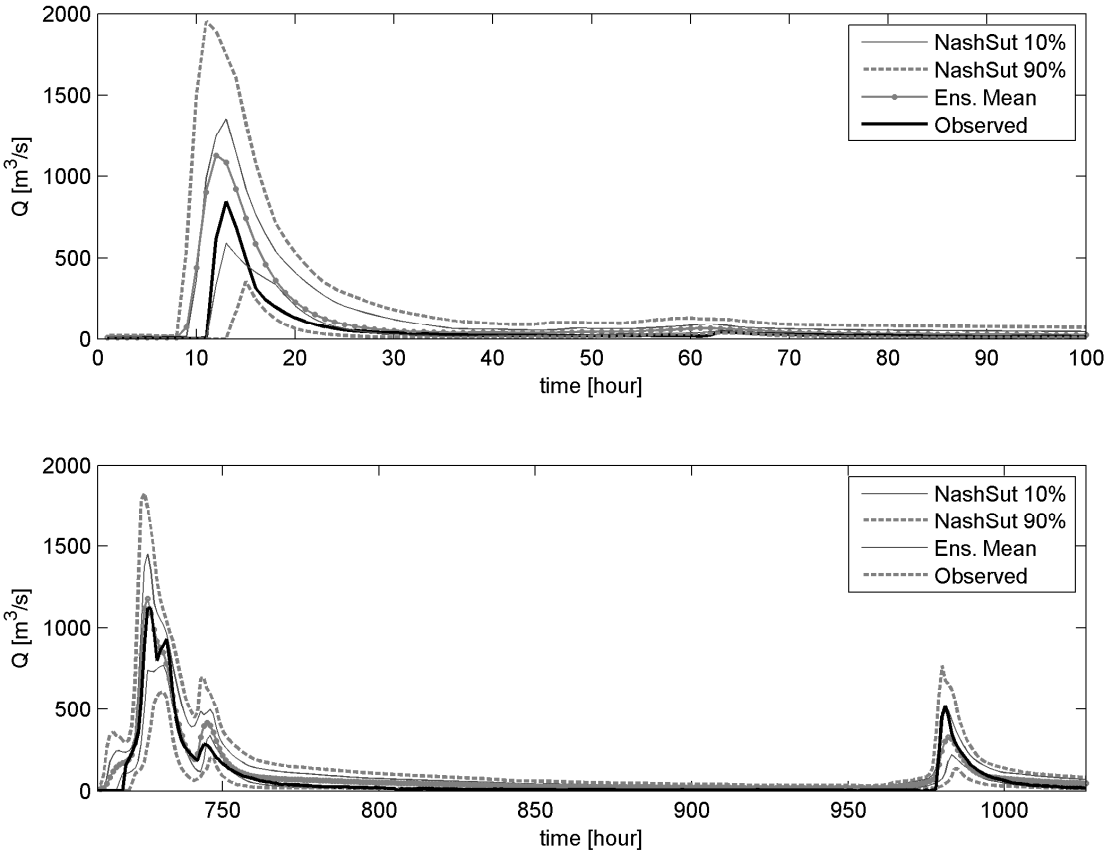


Figure 5: Orba basin, August-September 2006. Dotty plot representation of the NS on the hydrographs. Each subplot reports the parameter value on X axis and the value of NS statistic on Y axis. The surface parameters, subplots a) and b), have quite clear ranges where NS reaches maximum values independently from other parameters. This not occur for sub-surface flow parameters, subplots c) and d).



1

2 Figure 6: Orba basin, August-September 2006. Confidence intervals compared with observed
3 streamflow. Two representative periods are shown. X axis reports time while Y axis reports
4 streamflow. Observed streamflow is compared with the confidence intervals based on NS statistic
5 and the ensemble mean.

6

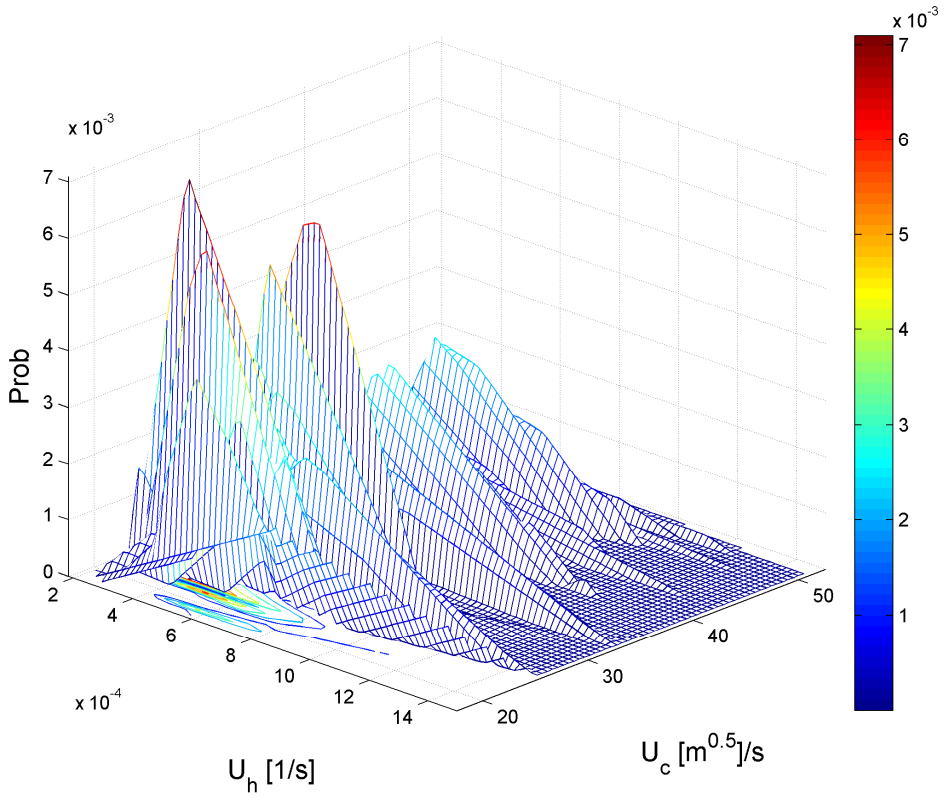
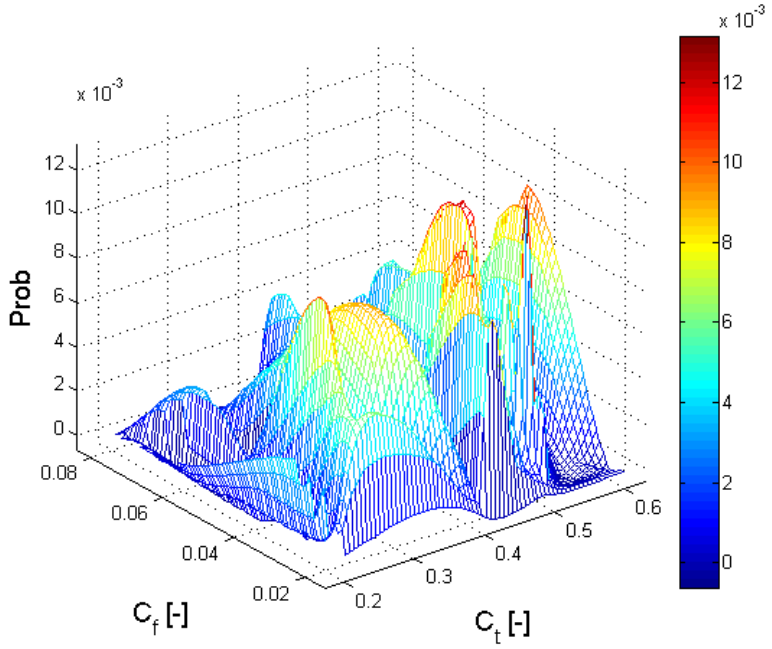


Figure 7: Orba basin, August-September 2006. Example of a two by two sensitivity analysis: Surface parameters, u_c versus u_h. On X and Y axis the values of the two parameters are reported while on Z axis the parameter likelihood is shown.



1

2 Figure 8: Orba basin, August-September 2006. Example of a two by two sensitivity analysis:
3 Subsurface parameters, c_t versus c_f. On X and Y axis the values of the two parameters are reported
4 while on Z axis the parameter likelihood is shown.

5

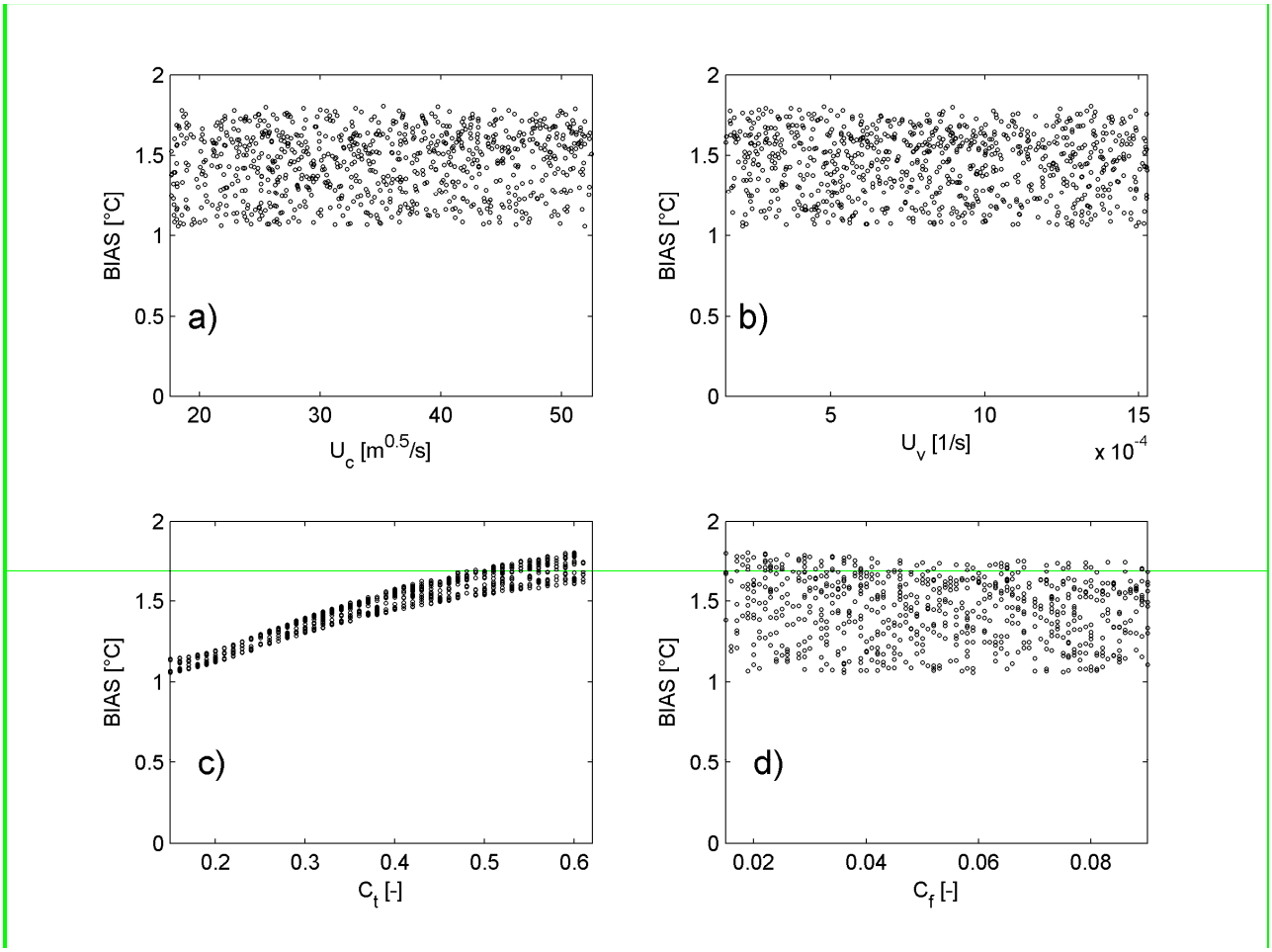


Figure 9: Orba basin, August-September 2009. Dotty plot representation of the BIAS on mean LST at basin scale. Each subplot reports the parameter value on X axis and the value of BIAS statistic on Y axis. The parameter c_t (subplot c) has a quite clear range where BIAS reaches minimum values independently from the other parameters, This not occur for the other parameters (subplots a, b and d)

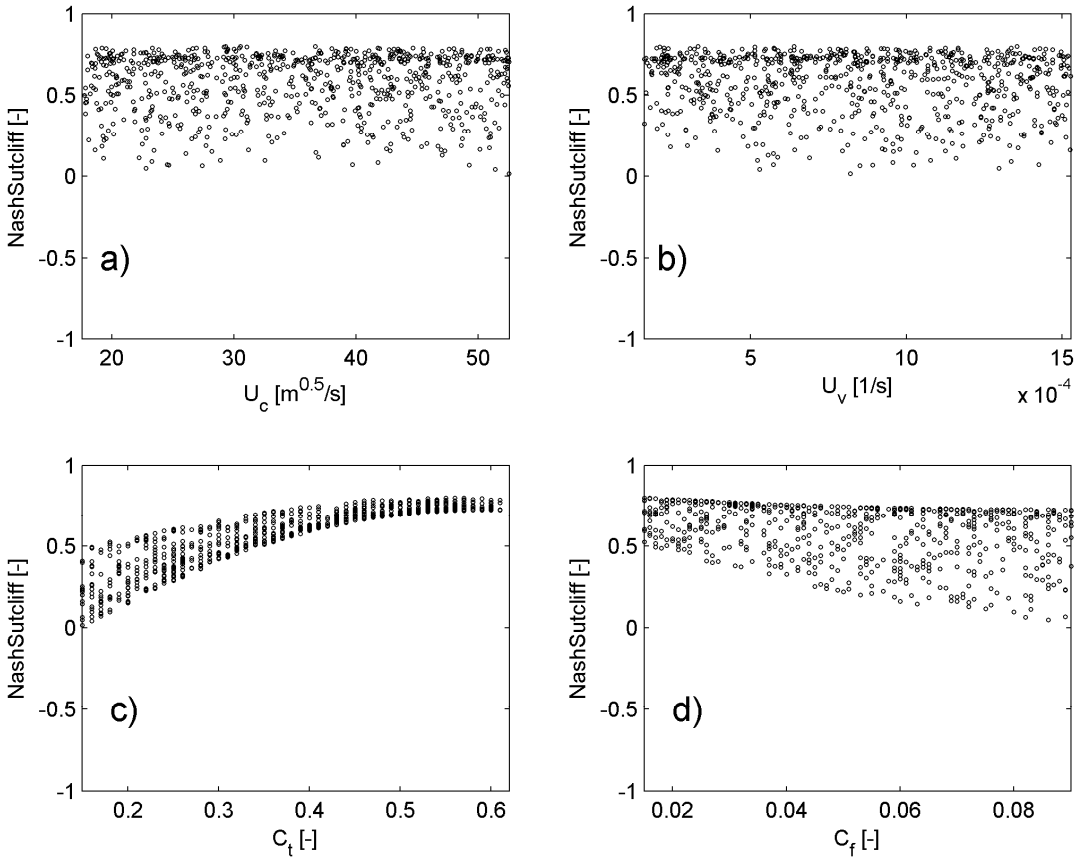


Figure 10: Orba basin, August-October 2011. Dotty plot representation of the NS on mean SWI at basin scale. Each subplot reports the parameter value on X axis and the value of NS statistic on Y axis. The parameter c_t (subplot c) has a quite clear range where NS reaches maximum values independently from the other parameters. This occurs even for parameter c_f although the independence is weaker (subplot d). No independence is evident for the surface parameters (subplots a and b)

1

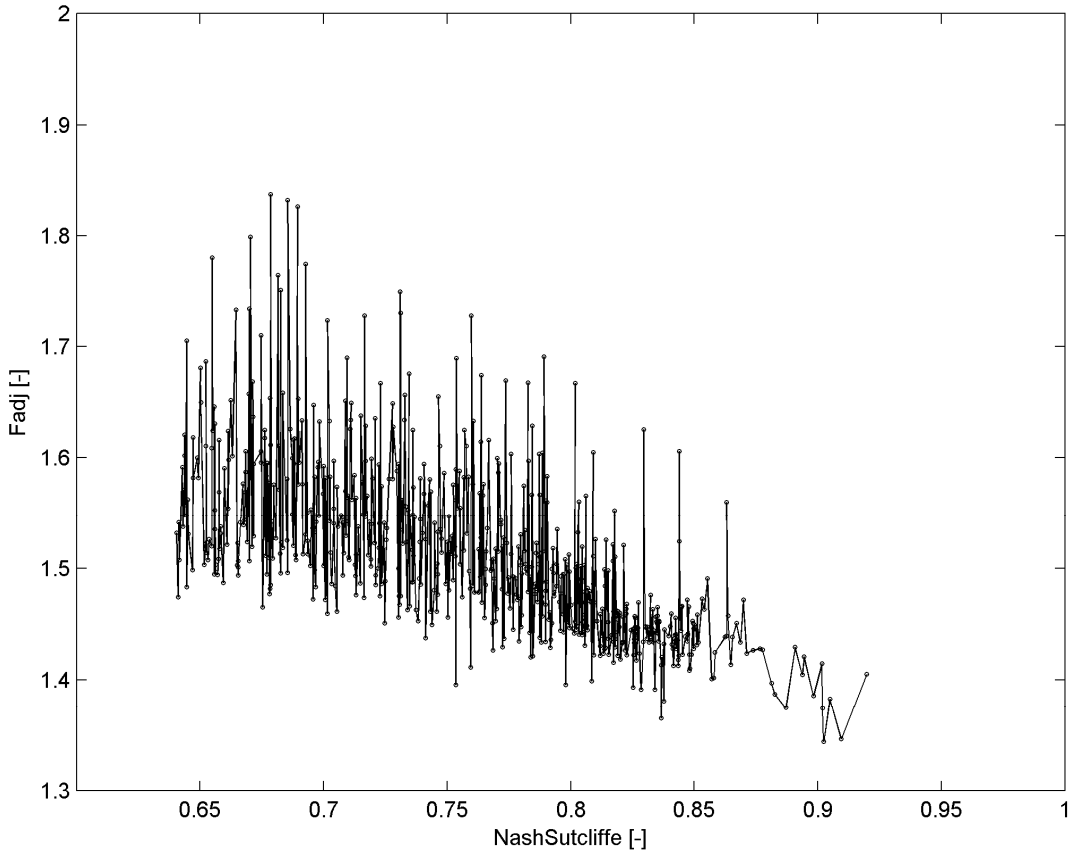


Figure 11: Nash Sutcliffe coefficient of hydrographs (X axis) versus the Multi Objective Function (Fadj on Y axis). The maximum of NS does not correspond with the minimum of the Fadj because this latter is influenced by the component that depends on LST ad SWI. Anyway in correspondence of the minimum of Fadj the NS value is high.

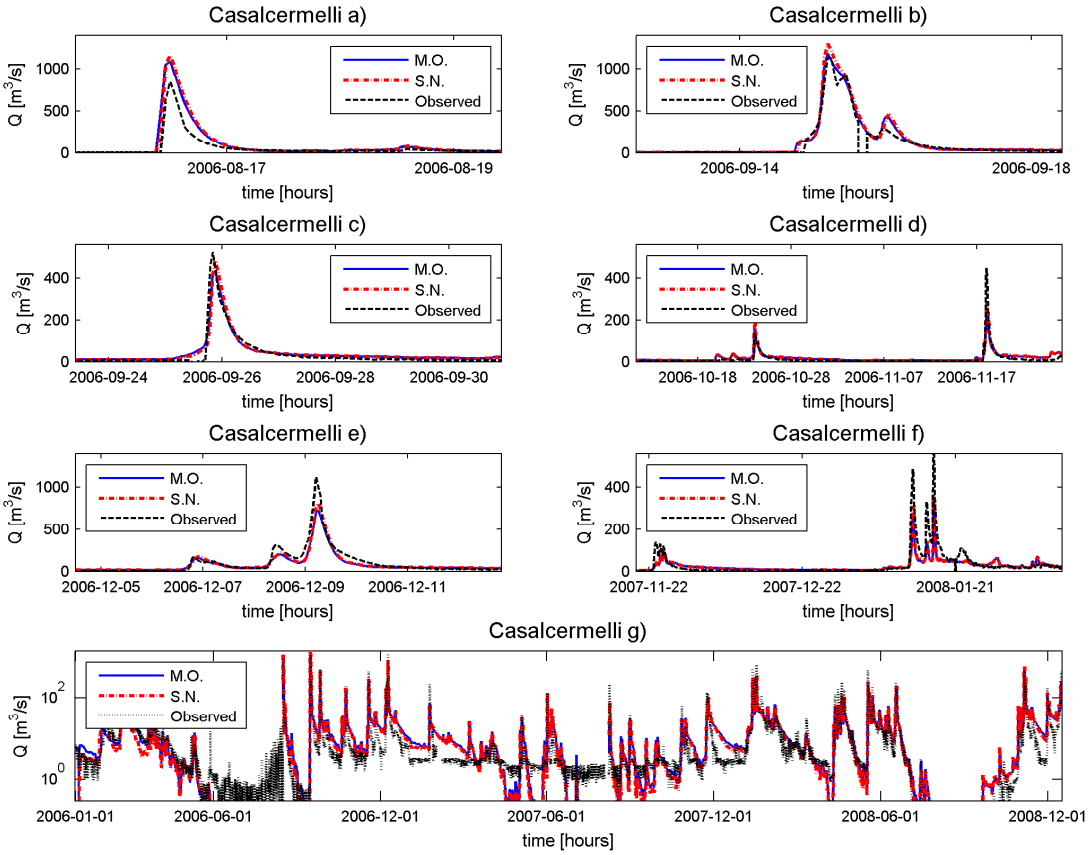


Figure 12: Observed streamflow for the Orba basin compared with simulations obtained using the M.O. and S.N. parameter sets for the period 2006 to 2008. Time is reported on X axis and streamflow on Y axis. The bottom subpanel shows the entire period with log scale on Y axis. The other subpanels show the main flood event with linear scale on Y axis.

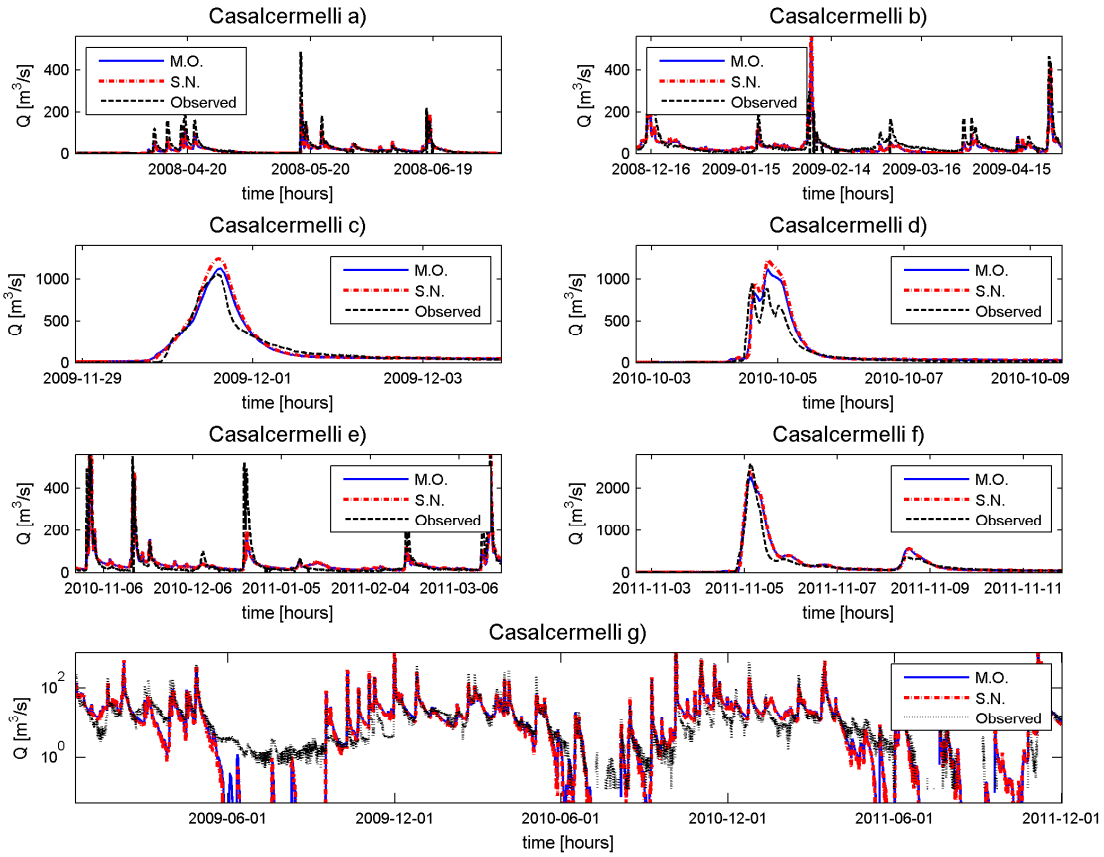


Figure 13: Observed streamflow for the Orba basin compared with simulations obtained using the M.O. and S.N. parameter sets for the period 2009 to 2011. Time is reported on X axis and streamflow on Y axis. The bottom subpanel shows the entire period with log scale on Y axis. The other subpanels show the main flood event with linear scale on Y axis.

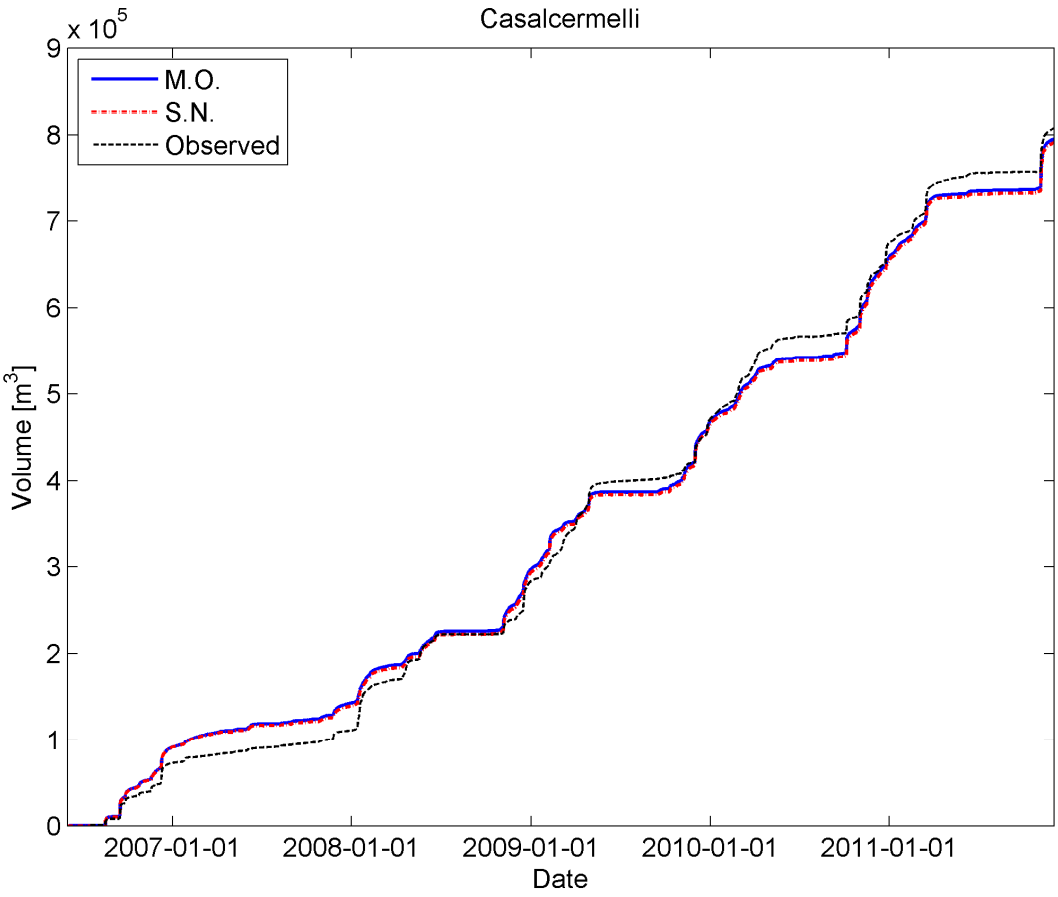
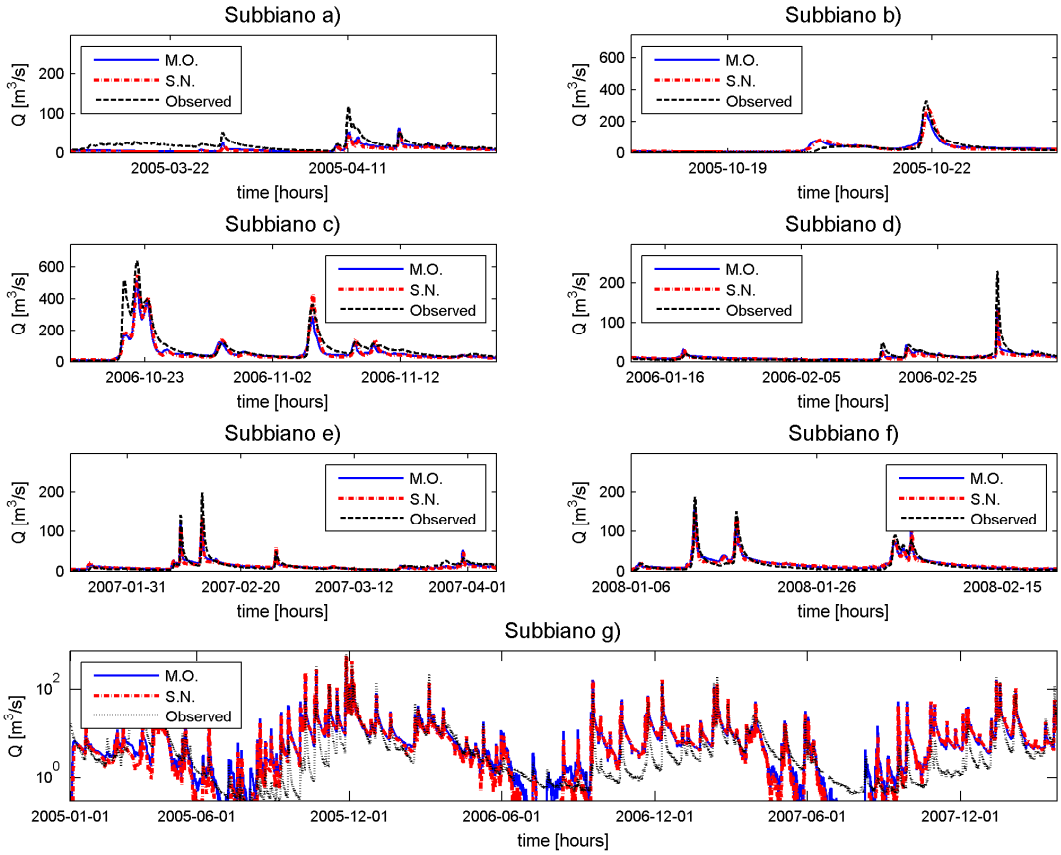


Figure 14: Accumulated runoff volumes for the Orba basin obtained with the M.O. and S.N. parameters sets for the period 2006-2011. Time is reported on X axis and total accumulated volume on Y axis.



1

2 Figure 15: Observed streamflow for the Casentino basin compared with simulations obtained using
3 the M.O. and S.N. parameter sets for the period 2005 to 2007. Time is reported on X axis and
4 streamflow on Y axis. The bottom subpanel shows the entire period with log scale on Y axis. The
5 other subpanels show the main flood event with linear scale on Y axis.

6

7

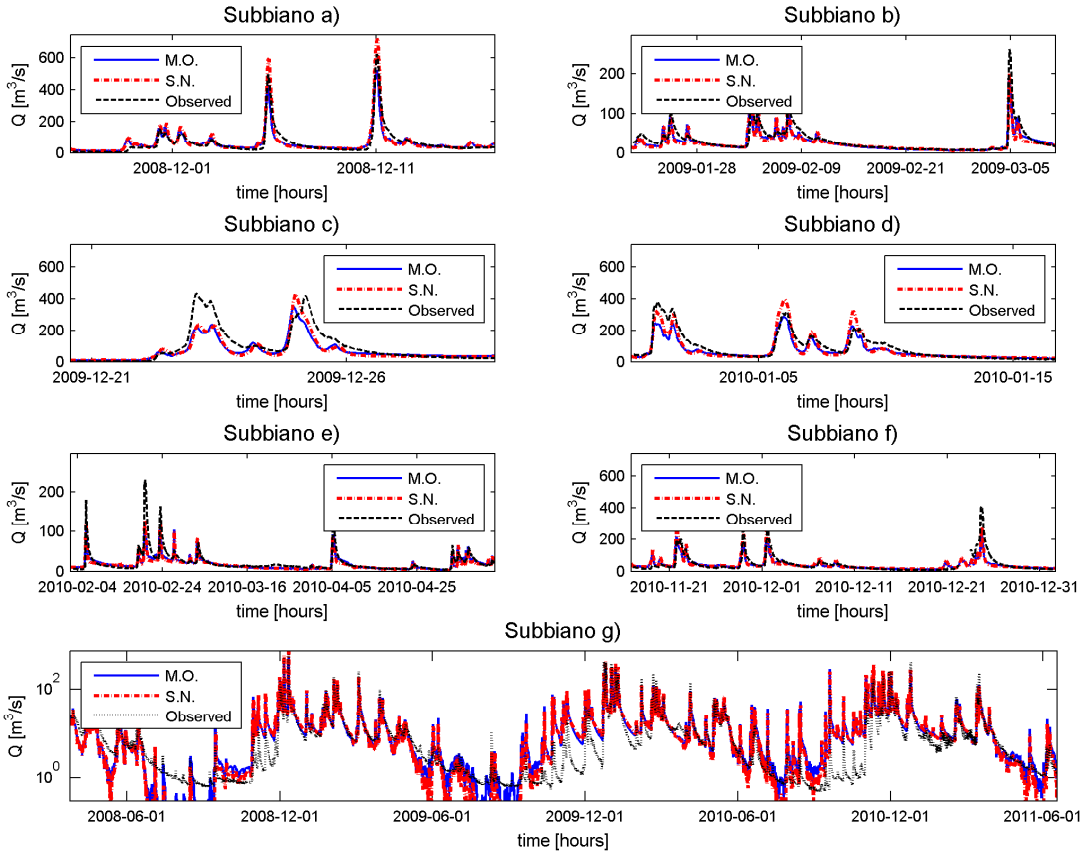


Figure 16: Observed streamflow for the Casentino basin compared with simulations obtained using M.O. and S.N. parameter sets for the period 2008 to 2011. The bottom subpanel shows the entire period with log scale on Y axis. Time is reported on X axis and streamflow on Y axis. The other subpanels show the main flood event with linear scale on Y axis.

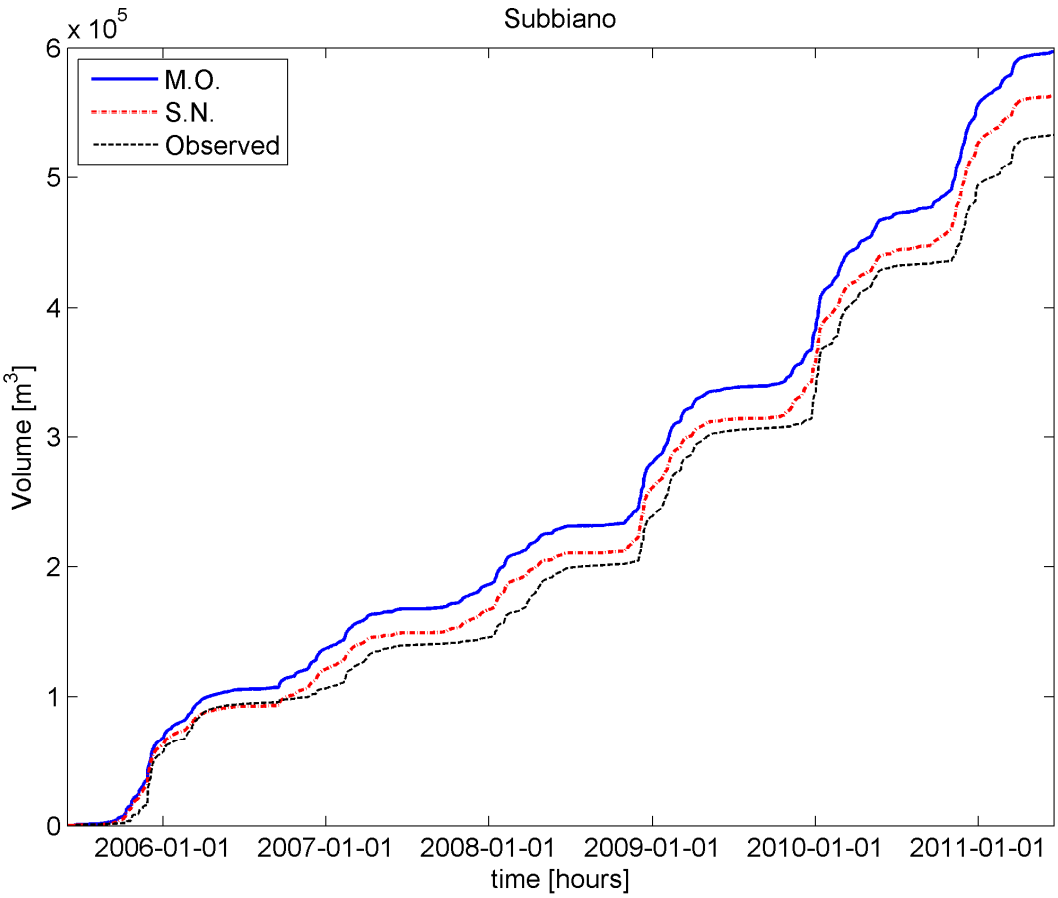


Figure 17: Accumulated runoff volumes of the Casentino basin obtained with the M.O. and S.N. parameters sets for the period 2005-2011. Time is reported on X axis and total accumulated volume on Y axis.