Abera, Wuletawu; Formetta, Giuseppe; Borga, Marco; Rigon, Riccardo. 2017. **Estimating the water budget components and their variability in a pre-alpine basin with JGrass-NewAGE**.

**Highlights**

- Separation of snowfall-rainfall using MODIS data
- Estimation of evapotranspiration parameter using Budyko curve concept
- Implementation of a method to assess the separation of water storage and evapotranspiration.
- Estimate the water budget closure at high spatial and temporal resolution in prialpne basin
- Combine various component of JGrass-NewAGE system to close the water budget

# Estimating the water budget components and their variability in a Pre-Alpine basin with JGrass-NewAGE

Wuletawu Abera[a,*], Giuseppe Formetta[b], Marco Borga[c], Riccardo Rigon[a]

[a] *Department of Civil, Environmental and Mechanical Engineering, University of Trento, Italy*
[b] *Centre for Ecology & Hydrology, Crowmarsh Gifford, Wallingford, UK*
[c] *Department of Land and Agroforest Environment, University of Padova, Italy*

**Abstract**

The estimation of water resources at basin scale requires modelling of all components of the hydrological system. Because of the great uncertainties associated with the estimation of each water cycle component and the large error in budget closure that results, water budget is rarely carried out explicitly. This paper fills the gap in providing a methodology for obtaining it routinely at daily and subdaily time scales. In this study, we use various strategies to improve water budget closure in a small basin of Italian Prealps. The specific objectives are: assessing the predictive performances of different Kriging methods to determine the most accurate precipitation estimates; using MODIS imagery data to assist in the separation of snowfall and rainfall; combining the Priestley-Taylor evapotranspiration model with the Budyko hypothesis to estimate at high resolution (in time and space) actual evapotranspiration (ET); using an appropriate calibration-validation strategy to forecast discharge spatially. For this, 18 years of spatial time series of precipitation, snow water equivalent, rainfall-runoff and ET at hourly time steps are simulated for the Posina River basin (Northeast Italy) using the JGrass-NewAGE system. Among the interpolation methods considered, local detrended kriging is seen to give the best performances in forecasting precipitation distribution. However, detrended Kriging gives better results in simulating discharges. The parameters optimized at the basin outlet over a five-year period show acceptable performances during the validation period at the outlet and at interior points of the basin. The use of the Budyko hypothesis to guide the ET estimation shows encouraging results, with less uncertainty than the values reported in literature. Aggregating at a long temporal scale, the mean annual water budget for the Posina River basin is about 1269 ± 372 mm (76.4%) runoff, 503.5 ± 35.5 mm (30%) evapotranspiration, and −50 ± 129 mm (-4.2%) basin storage from basin precipitation of 1730 ± 344 mm. The highest interannual variability is shown for precipitation, followed by discharge. Evapotranspiration shows less interannual variability and is less dependent on precipitation.

*Keywords:* Water Budget, Precipitation, Evapotranspiration, rainfall-runoff, storage, JGrass-NewAGE system

## 1. Introduction

Estimating the terrestrial water balance is one of the main scopes of hydrology (Eagleson, 1994). Nevertheless, the assessment of all the terms of the water budget equation is only rarely pursued. Different techniques and approaches are used to obtain results at various spatial and temporal scales. At the smallest scale (from single trees to a square kilometers catchments) the budget can be assessed with detailed measures which include various estimations of evapotranspiration with modern anemometry, sapflow estimation, groundwater levels determination, and, a variety of models (Dean et al., 2016; Graf et al., 2014; Fang et al., 2015; Obojes et al., 2015; Wilson et al., 2001). Due to the natural heterogeneity of evapotranspiration and storage, water budget closure studies using in-situ observations are restricted at experimental plots (Mazur et al., 2011; Högström, 1968; Scott, 2010). Models, in this case, often include the more modern process-based ones, which usually are data demanding and computationally intensive but produce very detailed results at hourly or sub-hourly time scales (Fatichi et al., 2016b).

At large, continental or global scale (from tens of thousand to millions of square kilometers), measurements are obtained by various remote sensing platforms (Wang et al., 2014a; Sahoo et al., 2011), and water budget estimations exploit atmospheric reanalysis efforts (Pan and Wood, 2006; Sheffield et al., 2009) and use the so called macro-scale hydrological models (Bierkens et al., 2015). Many available models present in literature (e.g LISFLOOD (Van Der Knijff et al., 2010), mHM (Samaniego et al., 2010), VIC (Liang et al., 1994), SWAT (Arnold et al., 1998)) could potentially be used for this scope, but in practice most studies are limited to the use of VIC model (Liang et al., 1994). Generally large-scale analyses have very coarse resolution. One notable exception is Maxwell and Condon (2016), but it is based on a limited description of the spatial heterogeneity of models parameters.

At intermediate scales (i.e. from few square kilometers to some ten thousands of square kilometers), the water budget is performed by using a variety of models and measurements, as shown in the Table 1. Table 1 is the synthesis (one fifth) of

*Corresponding author: wuletawu979@gmail.com

Table 1: A selection of papers estimating the water budget at spatial scale from 1 to $10^5$ km$^2$. This comes from a large selection of papers dealing with all the spatial scales. To enter in this selection the requirement is that all the components of the hydrological budget are estimated and the water budget closure evaluated.

| Study | Method (model) | Spatial Scale [km$^2$] | Resolution (time span) | Location |
|---|---|---|---|---|
| Arnold and Allen (1996) | Modeling with SWAT | 122-248 | Annual (1951-1957) | Illinois, U.S. |
| Batelaan and De Smedt (2007) | Modeling with WetSpass | 694-1677 | Monthly to Yearly | Belgium |
| Bertoldi et al. (2006) | Modeling with GEOtop | 9 and 603 | Hourly | Trento, Italy; Oklahoma, U.S. |
| Claessens et al. (2006) | Modelling | 404 | Monthly to Yearly (1930-2000) | Ipswich River basin |
| Hingerl et al. (2016) | Modeling with GEOtop | 55 | Hourly | Germany |
| Huntington and Niswonger (2012) | Modeling with GSFlow | 54 | Daily | Nevada, U.S. |
| Jothityangkoon et al. (2001) | Modelling | 2545 | Monthly and daily | Semiarid western Australia |
| Mitchell et al. (2003) | Empirical formulas | 27 | Daily (1978-1996) | Curtin Catchment, Camberra, Australia |
| Ogden et al. (2013) | Modeling | 36-175 | Daily (2009-2012) | Panama |
| Schaake et al. (1996) | Modelling | 176-2344 | Daily to monthly | Oklahoma, Mississippi, North Carolina, U.S. |
| Tomasella et al. (2008) | Micrometeorologial tower, ground measurements, neutron probe | 6.58 | Monthly (2001-2004) | Amazonian micro-catchment close to Manaus (BZ) |
| Yang et al. (2007) | Budyko curve | 272 - 94800 | Annual | Tibetan Plateau, River Haihe, Gansu Province |
| Zhang et al. (2008) | Budyko curve | 5- 2000 | Mean Annual, monthly, daily | Various catchments in Australia |

a larger literary review presented in the supplementary material. A couple of the papers in the Table use the Budyko curve concept (Zhang et al., 2008; Yang et al., 2007). Other two use the GEOtop model (Bertoldi et al., 2006; Hingerl et al., 2016), and are supported by measurement campaign and/or facilities which are not commonly available. Two are focused on the groundwater-surface water interactions (Batelaan and De Smedt, 2007; Huntington and Niswonger, 2012). In the remaining group of papers, the most common approach is the use of a rainfall-runoff model and evapotranspiration estimates based on variations of the Penman-Monteith equation (Monteith et al., 1965). The quantity and the quality of ground measurements varies from the interpolation of coarse meteorological data (Claessens et al., 2006) to accurate ad-hoc field campaign (Ogden et al., 2013). The temporal scale of the estimates are equal or longer than a day.

Methodologies for determining routinely the water budget terms at hourly and subdaily time scale, based on the availability of a limited set of ground measures, i.e. rainfall, temperature and discharge in a few locations, are clearly not commonly developed. In this paper we describe a method to fill this gap. We use a non computational demanding modelling framework, called JGrass-NewAGE (Formetta et al., 2014b). The framework offers not only the hydrological modeling components (as described below) but also innovative tools such as automatic calibration methods for parameters estimation, and a geographical information system for input-output data visualization.

Generally, the water budget for an appropriate control volume is:

$$\frac{\partial S_k(t)}{\partial t} = J_k(t) + \sum_i^n (I_{ki}(t)) - ET_k(t) - Q_k(t) \qquad (1)$$

where $S$ [L$^3$] represents the total water storage of the basin, $J$ [L$^3$ T$^{-1}$] is precipitation, $ET$ [L$^3$ T$^{-1}$] is evapotranspiration, and $Q$ [L$^3$ T$^{-1}$] is runoff (surface and groundwater). $I_{ki}$ represents input fluxes to the $k$-th hydrologic response unit (HRU) coming from the set of $n(k)$ HRUs connected to it, these fluxes are of the same nature as $Q$. A HRU represents a part of the basin that can be treated as a single unit (see section 3.1). In Eq.(1) the inputs are precipitation data, $J(t)$, but these must be split into rainfall and snowfall, and $I(t)$. The outputs are evapotranspiration, $ET$, and discharges, $Q$. Any budget term can be further subdivided, by separating it in parts, for instance discharge could be separated in surface, subsurface and groundwater discharge, modelled with various conceptualisations (Clark et al., 2008), which can be the result of a heuristic process of selection (Fenicia et al., 2008).

We take for granted that total precipitation amount can be estimated according to the procedures presented in Garen and Marks (2005); Tobin et al. (2011); Abera (2016). However, we innovate their methods, which have become obsolete in part as computing techniques have progressed (see section 3.2) and use some tools appropriately developed in the JGrass-NewAGE system (Formetta et al., 2014b). Furthermore, we review the choice of the best kriging interpolation method and characterise the rainfall estimation errors. We also discuss how different methods of precipitation interpolation affect the final discharges simulation and the whole budget.

One issue about which many papers are reticent is the separation of rainfall and snowfall. Some interesting efforts present in literature are the psychrometric energy balance (e.g Steinacker, 1983; Harder and Pomeroy, 2013, 2014; Ye et al., 2013) and variations of the phase separation method proposed by the U.S. Corps of Engineers (Army, 1956; Rohrer, 1989). Other solutions can be found in meteorological models like ARPS (Xue et al., 2000, 2001) and WRF (Dudhia et al., 2005; Caldwell et al., 2009). In all of the above methods and procedures there are parameters to calibrate, which is problematic when there are no or few snow related observations. This paper searches for an alternative, obtained by using easily available, remote-sensing data and the modular structure of JGrass-NewAGE (NewAge).

Here we use discharge measurements at three stations to calibrate and validate discharge estimation of the whole basin using a hydrological model. These measurements can be used to calibrate a suitable hydrological model that can in turn be used to predict discharges for any period without data.

In contrast, the other two components of the budget (i.e. ET

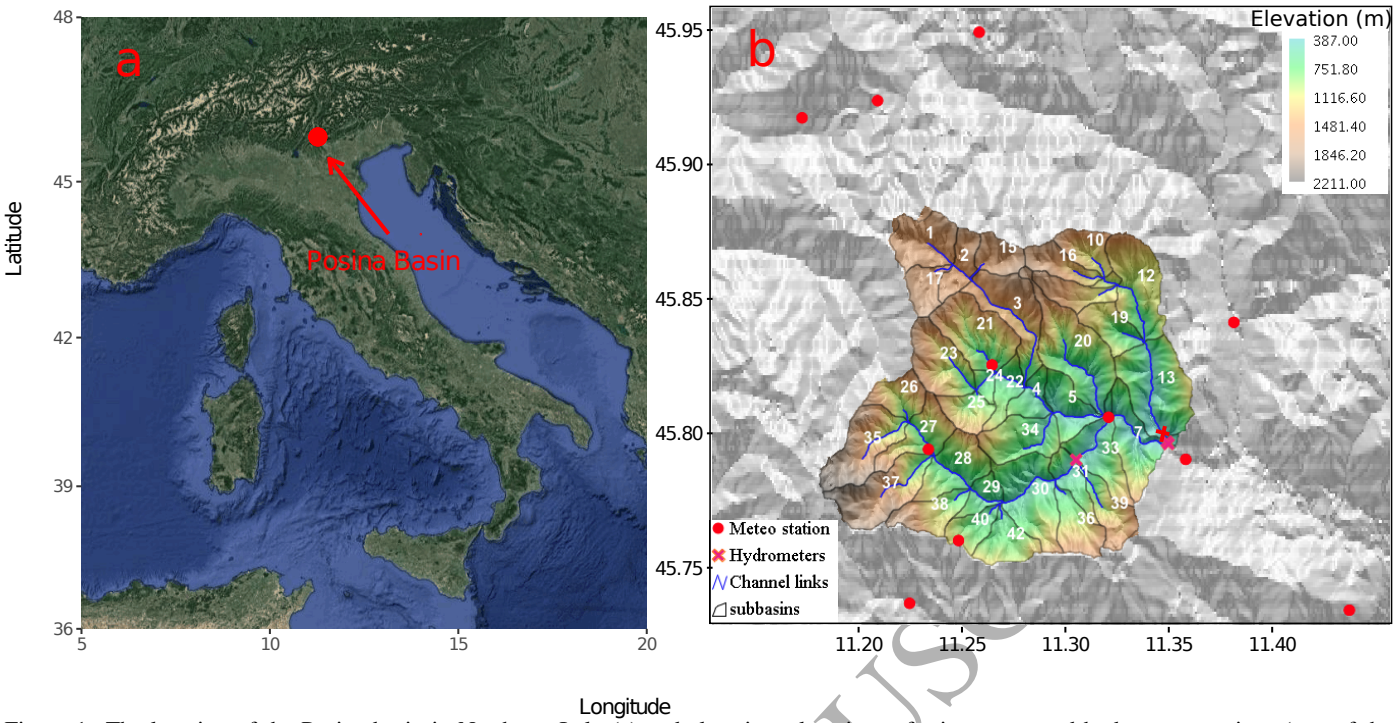


Figure 1: The location of the Posina basin in Northeast Italy (a) and elevations, locations of rain gauges and hydrometer stations (two of the hydrometers are at the outlet: one at the main outlet and the other in the river channel ID 13), and subbasin-channel partitions used in the simulation (b).

and dS/dt) are usually not available. The best approach to estimate ET at high spatial and temporal resolution is based on models that use knowledge of the physical process. However, these models, such as Penman-Monteith (Monteith et al., 1965), require more meteorological data than usually available in most basins, such as radiation, temperature, wind speed, air pressure, and aerodynamic and canopy resistance. Other simplified models, such as the Priestley-Taylor equation (Priestley and Taylor, 1972), on the other hand, need less data (i.e. radiation), but also need measured ET data for parameter calibration, which, if derived from literature, introduces very large uncertainties in the budget (Cristea et al., 2012). In the Methodology part of this paper, we provide a way to close the water budget in the absence of ET and water storage measurements based on the combination of Priestley-Taylor (Priestley and Taylor, 1972) and Budyko hypothesis (Budyko, 1978). Our method reduces uncertainty, but it does not eliminate it. However, relying on mass conservation, it is possible to investigate the variability of the estimates and identify their intervals of confidence, which is overall a gain of knowledge.

Summarising, the issues dealt with in this paper are: determination of which kriging method generates the best spatial precipitation (and other auxiliary meteorological forcings); separation of snowfall and rainfall, with the aid of satellite data when no snow data are available; determination of the Priestley-Taylor $\alpha$ coefficient (Priestley and Taylor, 1972) with an appropriate hypothesis; and finally, estimate the spatially distributed water budget of the basin with an evaluation of the errors of estimates.

The practical tool we use is the NewAGE modelling system (Formetta et al., 2014b) detailed below. Some of the operations performed in the paper are, in fact, facilitated by modular structure and the various tools that the NewAGE system deploys. These operations would be quite complicated to carry out with other more traditional models.

The paper is organized as follows: first, descriptions of the study area and the experimental setup are given (section 2); then the methodologies (section 3) for control volume discretization (subsection 3.1); followed by comparison of kriging interpolation methods to determine the best one for the spatial characterization of the input components of the water budget (subsection 3.2). Snowfall separation and snow water equivalent (SWE) estimation procedures are discussed in subsection 3.3. Methodological descriptions of discharge and evapotranspiration modelling at each HRU are presented in sections 3.4 and 3.5 respectively. Finally, the results and discussions, and the conclusions of the study are given in sections 4 and 5, respectively.

## 2. Study area and input Data

The study area is Posina River basin, a small catchment (116 km$^2$) located in the Alpine foothills of the Veneto Region in Italy. The basin outlet is at Stancari (figure 1b). The elevation difference of the basin is 1820 meters. The climate is characterised as wet, with annual precipitation of 1740 mm and annual runoff of 1000 mm (Norbiato et al., 2008; Francois et al., Submitted). The long-term (30 years, 1986-2016) snowfall-to-precipitation ratio averages 20% (Francois et al., Submitted), ranging from 5% to 27% depending on years. There are two

relative maxima for monthly precipitation: one in April (163 mm) and one in October (240 mm). A 20 x 20 metres DEM is available for the site. In the Posina area, in and around the basin, there are 11 meteorological stations and three discharge gauges (figure 1b). The meteorological stations provide hourly rainfall and temperature data, and hourly discharge data is collected by means of an ultrasound sensor. The rating curve is updated twice per year by using current meters and Doppler methods for the flow velocity measurements. All of the stations lack measures of other meteorological forcings such as solar radiation, wind speed, relative humidity, and snow depth (or snow water equivalent). Precipitation and temperature data used for the study cover the period 1994 to 2012. Discharge data from all three hydrometers are available for the same period. The first five years of data are used for calibration, and the rest for validation of the rainfall-runoff component.

## 3. Methodology

As stated above, NewAge system (Formetta et al., 2014b) is used for modelling each component of the water budget. NewAge offers a set of model components built according to the Object Modelling System version 3 (OMS3) framework (David et al., 2013). The components cover most hydrological processes and have been discussed in detail in Formetta et al. (2011, 2014c, 2013, 2014b), hence, they will not be fully rediscussed here. In this study, kriging (Formetta, 2013; Abera, 2016), snow melting (Formetta et al., 2014c), shortwave radiation (Formetta et al., 2013), longwave radiation (Formetta et al., 2016), Priestly-Taylor evapotranspiration (Priestley and Taylor, 1972), and the rainfall-runoff (Formetta et al., 2011) components are used. As calibration tools we use Particle swarm (Kennedy et al., 1995) and LUCA (Hay et al., 2006). The following sections provide the methods for modeling each term of the water balance as given in Eq.(1), which also corresponds to the workflow of NewAge system applications.

### 3.1. Watershed partitioning

Our approach uses spatial information at the level of the hydrologic response units (HRUs). A HRU represents a part of the basin that can be treated as a single unit (control volume) on the basis of mathematical, physical or computational arguments. In other words, even if hydrological variables can be calculated at the pixel level, for instance by exploiting a detailed knowledge of topography, it is subsequently coarse-grained to get single values for any HRU. The rationale of this choice is to reduce to a minimum level of spatial heterogeneity in the input data and processes, in an approach similar to those adopted in other models (Lagacherie et al., 2010; Ascough et al., 2012). In this paper, the terms, HRU and subbasin are used synonymously for the same basin partitioning concept.

More specifically, in NewAGE the basin is partitioned into hillslopes and channel links. The hillslopes and links are numbered according to the Pfafstetter numbering scheme (Formetta et al., 2014a), which gives an identifier for subbasins and an order to transverse them in computation (Formetta et al., 2014b; Abera et al., 2014). This approach gives the modeling solution of Eq. (1) for any of the units independently, from the most uphill one to the outlet. Depending on the process, the value of each term in the equation can depend on some sub-HRU analysis. In total, 42 HRUs have been defined for the basin (figure 1b). To illustrate the variability of hydrological quantities across the HRUs, a sample of HRUs (four HRUs: 1, 4, 13, and 37) is systematically selected to represent different elevations (elevation ranges from 656 m to 1616 m) and positions in the basin. Hence, further analyses and results at the subbasin scales are shown at these selected HRU throughout the paper.

### 3.2. Precipitation Interpolation

The accuracy of the raingauge measurements for the study basin has been assessed in the frame of several studies (Borga, 2002; Brocca et al., 2015; Penna et al., 2015), and no systematic over/underestimation was reported. In fact, the stations are used as benchmark for remote sensed estimation of rainfall, either by use of weather radar (Borga, 2002) and satellite (Brocca et al., 2015). Hence, the precipitation data, without systematic bias correction, is interpolated from meteorological stations to points of interest (centroids of each HRU) using the spatial interpolation (NewAGE-SI) tools. According to kriging theory (Goovaerts, 1997, 2000; Ly et al., 2011; Basistha et al., 2008), an experimental semivariogram needs to be fitted with the theoretical semivariogram to estimate its nugget, sill and range, which are kriging model inputs. Four semivariogram models (exponential, spherical, gaussian, and linear) have been selected and implemented in the NewAGE-SI, following (Prudhomme and Reed, 1999) arguments. As methods for geostatistical interpolation, ordinary kriging (OK) and its local version, local ordinary kriging (LOK), are used (Goovaerts, 1997). Because many studies have found that incorporating elevation data into the kriging interpolation improves the performances (Lloyd, 2005; Buytaert et al., 2006; Garen et al., 1994) of their models, we also consider detrended kriging (DK) and local detrended kriging (LDK) (Phillips et al., 1992).

To understand the effects of choice of semivariogram model on kriging, and to determine the best kriging solution, we apply the following procedure for any given time step and using all the weather stations except one. Firstly, we select a single kriging type (e.g. OK) and fit a theoretical semivariogram (e.g. exponential) with the experimental one by minimizing the errors. This is different from other approaches (e.g. Garen and Marks (2005)), where kriging parameters were kept fixed *a priori*. The optimisation is done using the particle swarm algorithm (Kennedy et al., 1995; Formetta et al., 2014b) with the OMS configuration shown in Figure 2a.

Then, measured and estimated precipitation values are compared with goodness-of-fit (GOF) indices (root mean square error, correlation coefficient, and mean error)(Appendix B).

Finally, the error of estimate is evaluated for the weather station not considered. The procedure is then repeated, alternating the weather station left out until done for all, for 18 years of hourly forcings (precipitation and temperature). Each of the above steps is repeated for all four krigings and the four

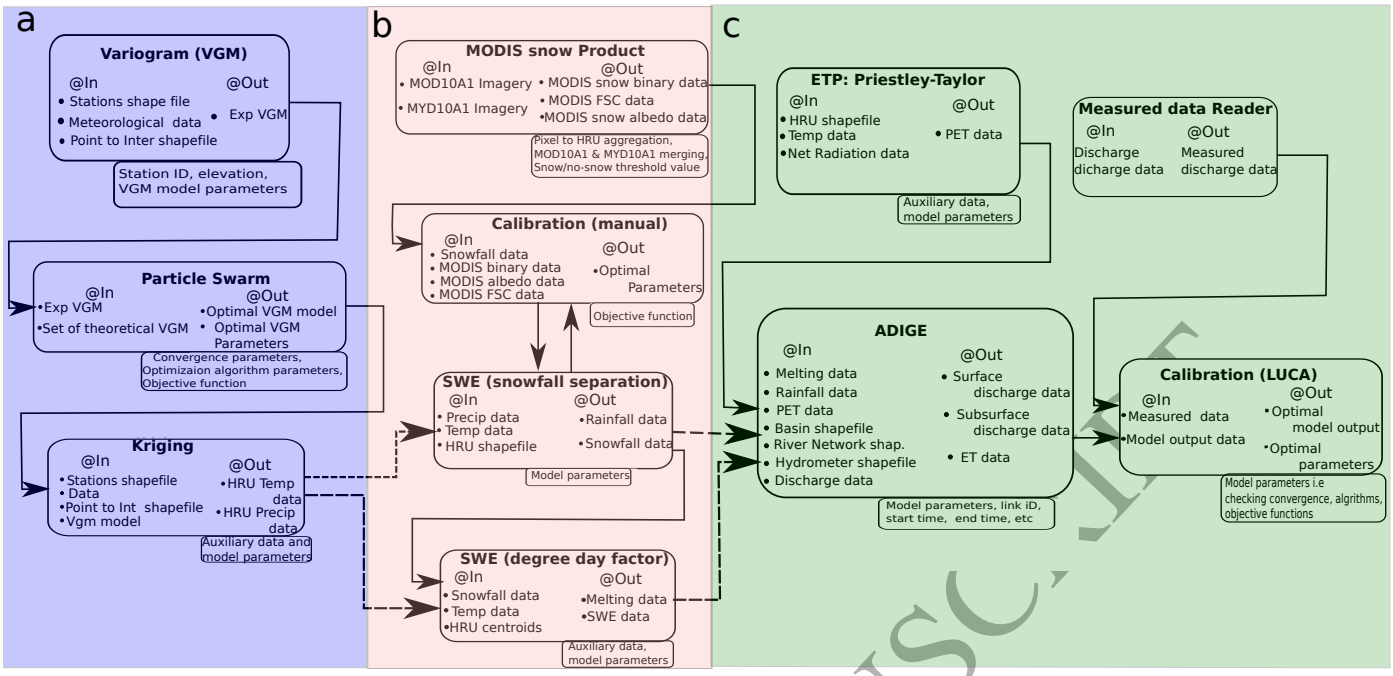


Figure 2: Operational connections of NewAGE components to obtain: a) the spatial interpolation of meteorological data (SI-NewAGE). The "theoretical semivariograms" component uses the "Particle Swarm algorithm to optimize theoretical semivariogram for each time step. Then, "Kriging" uses the best semivariogram model and optimal parameters to estimate the meteorological data; b) the "Snow separation" component, uses MODIS snow products to calibrate the spatial snow accumulation (Eq. 2); c) the estimation of the water budget with the "Priestley-Taylor" and "Adige" components. Data not obtained by interpolation are provided by a suitable reader. Here LUCA calibration component is included. Note that the dashed line, connecting the a, b, and c, are not automatic and need human intervention.

semivariogram models available (for a total of sixteen combinations). In addition to the above, we have developed an iterative procedure that automatically selects the best fitting semivariogram with optimized parameters by choosing the semivariogram model with the best performance for each time step.

### 3.3. Snowfall-rainfall separation and SWE modelling

Precipitation, $J(t)$, at each HRU has to be separated into rainfall $J_R$ and snowfall $J_S$. One option is to use a micro-meteorological model, in which the separation happens automatically, as in WRF (Skamarock et al., 2008). However, the standard procedure in hydrological models is to use simple mathematical expressions based on temperature (e.g. Garen and Marks, 2005). Usually, the separation is based on a threshold temperature, $T_S$: when the average HRU temperature is less than $T_S$, the precipitation is snow ($J_S$) otherwise it is rain ($J_R$). Temperature is interpolated using kriging (as shown in the complimentary material). In NewAGE, a smoothing of the threshold is applied as in Kavetski et al. (2006); Formetta et al. (2014c):

$$\begin{cases} J_R(t) = \alpha_r * \left[\frac{J(t)}{\pi}.\arctan\left(\frac{T(t)-T_s(t)}{m_1}\right) + \frac{J(t)}{2}\right] \\ J_s(t) = \alpha_s * [J(t) - J_R(t)] \end{cases} \quad (2)$$

where the term $J$[mm t$^{-1}$] is measured precipitation, $J_R$[mm t$^{-1}$] is the rainfall, $J_S$[mm t$^{-1}$] is the snowfall, $T_s$ [$^oC$] is the threshold temperature, and $m_1$[$-$] is a parameter controlling the degree of smoothing. The two coefficients ($\alpha_r$, $\alpha_s$) are adjusting parameters for rain and snow measurement errors (Formetta et al., 2014c). $\alpha_r$, $\alpha_s$ and $m_1$[$-$] are dimensionless coefficients that require calibration. The calibration is very problematic due to a complete lack of snow data in the Posina basin. In similar cases, snow parameters are usually calibrated indirectly, using discharge measurements (e.g. Li et al., 2012; He et al., 2014). Instead, we use MODIS snow imagery data (MOD10A1 and MYD10A1). Both the fractional snow cover (FSc) and albedo information, which is available globally at a resolution of 500 m (Hall et al., 2006), are used. The first establishes the fractional area covered by snow pixel by pixel, the second is used for detecting when fresh snow falls on old snow, causing an increase in albedo.

Due to the difference in time steps of the NewAge simulation and MODIS data, a manual optimization procedure is used to determine the parameters of Eq.(2). The objective functions used are the Accuracy index, $AI$, and Spearman's rank correlation coefficient, $\rho_{rank}$. The Accuracy index, $AI$, is given by:

$$AI = \frac{N_a + N_d}{N_a + N_b + N_c + N_d}.100 \quad (3)$$

Where the terms $N_i$ ($i \in \{a, b, c, d\}$) are the number of pixel combinations identified by using the confusion matrix given in Table 2. The optimization procedures, as outlined in figure 2b, maximize true positives and true negatives, while minimizing false positives and false negatives, therefore increasing the overall accuracy. Specifically, this procedure optimizes Eq.(2) to estimate snowfall only at locations where MODIS (MOD10A1 and MYD10A1) shows snow data.

Table 2: Confusion matrix based on the four possible results of the snowfall $J_s$ simulation in comparison with the MODIS snow products. The four possibilities are: true positive (a); false positive (b); false negative (c); and true negative(d).

| | MODIS:Yes | MODIS:No |
|---|---|---|
| Model:Yes | a | b |
| Model:No | c | d |

The binary (snow/no snow) data derived from the FSc does not show snowfall on snowfall. Hence, to include the new snowfall events on pixels already covered with snow, we used the FSc values (i.e new HRU area is covered by snow) and snow albedo. When associated to a precipitation event, an increase in snow albedo can be interpreted as fresh snowfall. Hence, the snow separation equation (Eq.2) is optimized against the FSc and the snow albedo, using Spearman's rank correlation coefficient as follows (Kottegoda and Rosso, 1997):

$$\rho_{rank} = 1 - \frac{6.\sum_{k=1}^{n} D_k^2}{n(n^2-1)} \quad (4)$$

where $D$ is the difference between the rank of the MODIS data (FSc or snow albedo) and snowfall data, $J_s$, at the $K^{th}$ pair, and $n$ is the number of observations.

Spearman's rank correlation is used because it provides a means to quantify the monotonic relationship between two variables with no frequency distribution assumption (nonparametric). The higher the value of $\rho_{rank}$, the higher the correlation between $J_s$ and snow albedo. Those parameters producing the highest $\rho_{rank}$ and $AI$ are used to model the hourly time steps of snowfall for each HRU. In principle, derivation of snow separation parameters for each HRU is possible, however, as is pertinent to the overall analysis of other components of the study, single, global and optimized values of Eq.(2) parameters are derived in this study.

NewAGE offers possibilities to simulate the snow metamorphism by using the radiation budget components also utilised for evapotranspiration. However, to keep the modelling simple, in this paper, we use the degree-day model implemented in the snow water equivalent (SWE) component as described in Formetta et al. (2014c). Due to the lack of SWE data, the model parameters are calibrated using discharge data together with the parameters of the ADIGE rainfall-runoff component (described below), which is in line with a consolidated approach (e.g. Li et al. (2012); Mou et al. (2008); He et al. (2014)). However, in our approach we deviate from the cited literature, given that the snowfall parameters are already calibrated using MODIS snow imagery and discharge data are used only for calibrating the parameters of the degree-day model.

### 3.4. Runoff Estimation

The NewAGE component that estimates runoff is called ADIGE (after the second largest river in Italy). It is the assembly of many Hymod models (Moore, 1985, 2007; Formetta et al., 2011), one for each HRU, whose outputs are collected at the appropriate point of the channel network, and routed to the outlet. The inputs of the ADIGE model are rainfall, evapotranspiration, and melting snow. The output is discharge at each channel-link. Hymod is a conceptual rainfall-runof model that separates a quick (shallow) flux from a slower (deeper) one. The quick flux accounts for surface runoff by using three linear reservoirs, while the slower flux accounts for subsurface storm flow with a single linear reservoir (Moore, 1985; Formetta et al., 2011; Formetta, 2013). Groundwater is described with further storage. Essential to the Hymod calculation is the estimation of the water "losses" by ET, which are modelled as detailed in the next section. For readers' convenience, details of Hymod are presented in the Supplementary material of the paper.

We conduct two groups of rainfall-runoff simulation experiments. The first is to evaluate the effects of different precipitation data sets, generated using different krigings, on the runoff calibration and modelling results. The Hymod parameters are calibrated for various precipitation data sets for a five-year period (1994-1999), using LUCA (Hay et al., 2006) as the optimization tool. The simulation for the period 2000-2012 is used for validation of the data set, by comparison. Based on the validation results, the second simulation is set up to estimate discharge at each link, providing data that are then used for spatially distributed water budget closure. The objective function used as standard in this paper is the KGE (Kling et al., 2012) which was considered more robust than the Nash-Shutcliffe efficiency. However, the goodness of fit was estimated by two other indicators, Pearsons correlation coefficient and Percentage bias (PBIAS), all of which are described in Appendix B. Other statistical indicators of performances are specified case by case. The configuration of the OMS-components in the workflow of Hymod simulation is shown in Figure 2c.

### 3.5. Evapotranspiration (ET) estimation, and residual storage

In the specific case study, ET measurements were completely missing and, therefore, need to be modelled. The following modification to the Priestley-Taylor (PT) Formula (Priestley and Taylor, 1972) for each HRU is considered:

$$ET(t) = \alpha \frac{S(t)}{C_{max}} \frac{\Delta}{\Delta + \gamma} (Rn - G) \quad (5)$$

where: $ET$ is the actual evapotranspiration [$LT^{-1}$]; $\alpha$ is the so-called Priestley-Taylor coefficient; $S(t)$ is the water storage in the root zone; $C_{max}$ the maximum storage of water in the HRU (a parameter to be calibrated by using measured discharges); $Rn$[$L^2T^{-1}$] is net radiation; $G$ [$L^2L^{-1}$] is soil heat flux; $\Delta$ is the slope of the Clausius-Clapeyron relation, which is given as a function of air temperature (Murray, 1967); and $\gamma$ is the psychometric constant [in $K$ $T^{-1}$]. The net radiation is modulated according to the two-radiation component included in the NewAGE system (Formetta et al., 2013, 2016). The effect of cloud climatology and atmospheric absorption were included through calibration of $\alpha$ to obtain the actual evapotranspiration. The formula is further simplified by assuming $G \propto R_n$, and including it in $\alpha$ (Clothier et al., 1982) too.

Temperature is interpolated from the meteorological stations to the centroids of each HRU using kriging (i.e LDK). In the

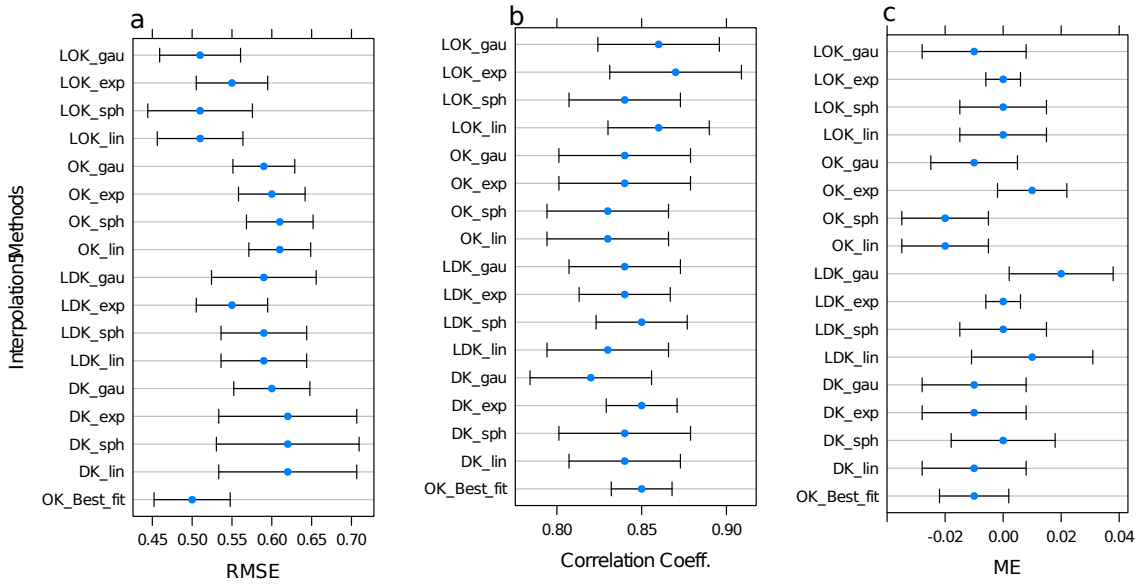


Figure 3: The figure contains the mean performance values, in terms of RMSE (a), correlation coefficient (b) and mean error (c), of all the rainfall interpolation methods considered for this study. sph = spherical semivariogram model; gau = gaussian semivariogram; exp = exponential semivariogram; lin = linear spherical semivariogram. The blue dot in the middle is the mean value, while the length of the line represents the 95% confidence interval with respect to the mean.

trade-off between accuracy and speed of execution, this simplification has been deemed acceptable, since the HRUs are of quite small extension (and the internal temperature variability within HRU has been checked to be less than half a degree Celsius on average).

The choice of reducing $ET$ with the relative water content is similar to the one pursued in several studies, e.g. Porporato et al. (2002) and Rodríguez-Iturbe and Porporato (2004). Cristea et al. (2012) reports that the value of $\alpha$ varies from 0.6 to 2.4 in previous studies, depending on land cover and site conditions, thus making literature almost useless in determining the conditions specific to a basin. In our modelling approach, we estimate $\alpha$ using the water budget (Appendix C), obtaining the actual evapotranspiration, instead of the potential evapotranspiration.

The whole process, which involves an iterative procedure between the PT and runoff estimation, is detailed in Appendix C. Our procedure assumes that the water storage is null after a specified number of years, said Budyko's time, $T_B$, (after Budyko (1978)). We actually use two algorithms: the first neglects soil moisture variability (i.e. $S(t)/C_{max} \equiv 1$) and generates $\alpha$ from the observed data using Eq C.3; the second uses variable storages within the HRUs ( Appendix C) and consists in simultaneously optimising the parameters of the ADIGE component and requiring that the water storage be null after $T_B$ years.

In the first procedures, if the first year of the simulation implies a negative storage, this is assumed to be present at the initial time, which, in turn, implies that the $\alpha$ coefficient must be recalculated to obtain null storage after $T_B$ years. Therefore the whole procedure is repeated until the initial storage and estimated $ET$ are consistent. Once $ET$ is estimated with the above procedure (and with varying $T_B$), the mean $\alpha$ coefficient is taken as the most reliable and used. Varying $T_B$, obviously implies different estimates of $ET$, which, however, remain confined within certain range. We interpret this as being representative of the epistemic uncertainty of our approach. Being assessed with this method, $ET$ does not balance Eq (1).

The second algorithm is highly demanding computationally, and only a single optimization procedure was actually performed for the years 1995 to 1999, with $T_B = 5$ years, after observing the trends of $dS(t)/dt$ in the case of the first procedure. Because in these simulations $ET$ is storage-limited (according to equation 5) the initial storage is taken to be null.

In this approach, the uncertainty of the $ET$ estimation has two components. One comes from the errors of estimation of the other components. Using the standard theory of errors (Rodell et al., 2004), the standard deviation of $ET$ estimates ($\sigma_{ET}$) can be derived from rainfall estimation error ($\sigma_J$, in this case the kriging error), discharge estimation (model) error ( $\sigma_Q$). The other contribution to the uncertainty of the $ET$ estimation comes from the error made in estimating $\alpha$ ($\sigma_\alpha$). As representative of the total mean error, therefore, we use $\sigma_{ET} = \sqrt{\sigma_J^2 + \sigma_Q^2 + \sigma_\alpha^2}$.

The primary verification of our $ET$ estimation is its consistency with the rest of the water budget. However, for further evaluation of the results, at the whole basin scale, we used the Global Land Evaporation Amsterdam Methodology (GLEAM) (Miralles et al., 2011a), a global, satellite-based, $ET$ data set. The performance of GLEAM is assessed positively in different studies (McCabe et al., 2016; Miralles et al., 2011b) and it is available at $0.25^o$ spatial resolution and daily temporal resolution. For comparison, therefore, we aggregated NewAge ET

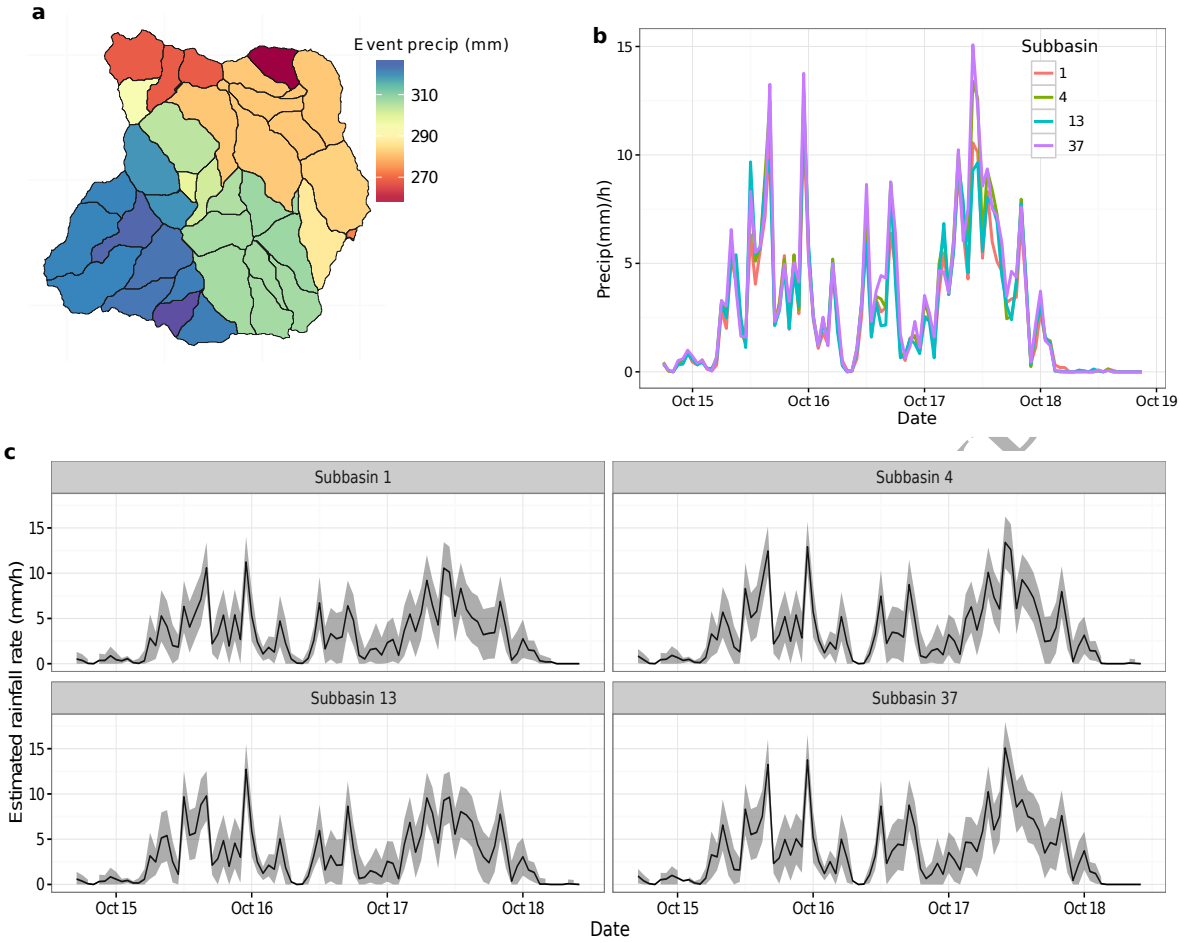


Figure 4: Spatial rainfall variability in an aggregated subbasin approach: (a) variability in the estimated total rainfall (the code number in the subbasin identifies the subbasin, while the colour shows the total rainfall distribution), (b) comparison of four selected time series of subbasin rainfall estimates, and (c) further analysis on the kriging estimation error, used to estimate the confidence interval of the estimates for some selected subbasins. The analysis is based on the event of 16 Oct. 1996.

estimation to daily time scale for the whole basin.

## 4. Results and Discussion

The results of the study are presented as follows: first, we report briefly the results of the performance of kriging analysis for precipitation; second comes the rainfall-snowfall separation and estimation of SWE; the third and fourth subsections contain the runoff simulation results and the spatio-temporal estimation of ET and storage respectively; last, following proper temporal and spatial characterization of all the water budget components, the results of water budget closure analysis are presented.

### *4.1. Precipitation estimation, model performances and uncertainty*

The RMSE and correlation coefficient results presented in figure 3 indicate that the LDK and LOK outperform the other two kriging groups (DK and OK). For each station, the LDK and LOK analyses are based only on the nearest five stations. Both LDK and LOK show lower RMSE and higher correlation coefficients following cross-validation analyses (figure 3a). The idea that DK improves performance is not clearly visible in this experiment. This may be due to the small number of stations from which to draw the elevation trend. The results of the mean error value are found to be inconsistent with the reports from the previous two performance indicators. The report shows slight differences between the semivariogram fittings within a kriging method (figure 3c). However, comparing across all 16 interpolations, the main difference is observed between the kriging methods, not between the semivariogram fitting within a single model (Haberlandt, 2007).

To use the precipitation estimate in conjunction with the runoff component, we must aggregate it at the subbasin level, by assuming the precipitation at the centroid of a HRU as representative of the entire subbasin. This operation is repeated for each time step and, for instance, figure 4 shows the spatial distribution of precipitation for a time instant event (16 October, 1996). The accumulated event rainfall shows a relevant spatial distribution difference, which is more than 70 mm (22% of the basin total accumulated rainfall) for the particular event presented in figure 4a. It is evident that rainfall is higher in the south-western part of the basin (figure 4a). Figure 4b shows the rainfall variability for four selected subbasins (ID 1, 4, 13,

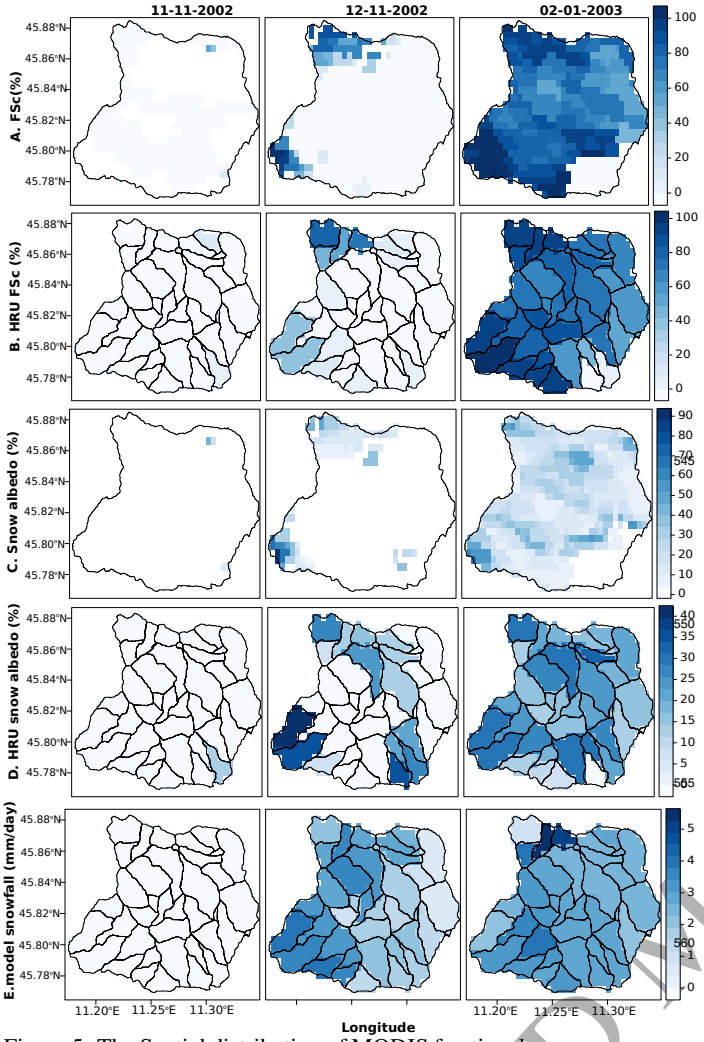


Figure 5: The Spatial distribution of MODIS fractional snow cover extracted from both Terra and Aqua products (A), fractional snow cover aggregated into the HRUs (B), MODIS snow cover albedo derived from both Terra and Aqua products (C), snow albedo aggregated into HRUs (D), and HRU snowfall estimated using the separation algorithms (E) for three days (11-11-2002, 12-11-2002, 02-01-2003).

37) along a limited time-window. As a product of the kriging procedure, the time-varying errors in rainfall can be estimated (figure 4c). In general, the cross-validation results show that LOK and LDK (particularly LOK) outperform OK and DK.

### 4.2. *Rainfall-snowfall separation and SWE estimations*

A sample of MODIS maps and spatial snowfall is shown in figure 5. For instance, in the first column of maps (11-11-2002) both the MODIS FSc and albedo show that the surface is snow free, although there is precipitation on this particular day. On the next day (12-11-2002), the MODIS imagery data shows that, in some parts of the basin, the surface is covered with some level of snow. The model separates the precipitation into snowfall and rainfall and, on this day, the spatial distribution is consistent with the MODIS data. Note that where MODIS shows no snow, the model estimates very light snowfall that could be approximated to zero (figure 5, first column). It is also

Table 3: $\rho^1_{rank}$ and $\rho^2_{rank}$ are the rank correlation coefficients between the model snowfall and the MODIS Albedo and FSc, respectively. The calibration period covers the 2002/2003, 2003/2004,and 2004/2005 snow seasons, while the validation period is for the 2005/2006 snow season.

| Period | $AI$ | $\rho^1_{rank}$ | $\rho^2_{rank}$ |
|---|---|---|---|
| Calibration | 60% | 0.41 | 0.52 |
| Validation | 45% | 0.34 | 0.37 |

important to note that after the fresh snowfall that is captured by MODIS, fresh snow on existing snow surface is depicted by the increase in FSc/albedo. Figure 5 (third column) shows snowfall in the middle of the snow season, on 01-02-2003, as depicted by the FSc and albedo. Clearly, the basin is covered by snow, and the model also estimates spatially consistent snowfall.

The accuracy index between the snowfall estimate and the MODIS snow binary data is 60% during the calibration and 45% during the validation period (table 3). These accuracy values can also be seen from the perspective of the 85% global accuracy of the MODIS snow product itself (Parajka et al., 2012). Hence, the 60% binary mapping accuracy in the calibration period can be considered acceptable for the long term water balance analysis in this study. The rank correlation result of modelled snowfall with MODIS snow albedo and with FSc, maintaining the 60% binary accuracy, is 0.41 and 0.52 respectively (table 3). This correlation value is considered to be of medium performance (Kottegoda and Rosso, 1997). The performance of the rank correlation decreases during the validation period. For many reasons, such as differences between MODIS and model time steps and spatial units, the correlation and spatial consistency between MODIS and the snowfall model is not very high. At basin scale, however, the approach is helpful to maintain some level of spatial consistency. Looking at the results, the estimated snowfall values mimic the spatial distribution of the FSc and the snow albedo data. Clearly, similar spatial variability cannot be obtained using only discharge data. Due to slight decreases in temperature at higher elevation HRUs, there is more snowfall and higher spatial variability with elevation in NewAGE estimation. However, the relationship of the MODIS snow product with topography is neither linear nor strong. Topographical complexities (such as slope, aspect, wind speed, shading, and vegetation) might be suggested as responsible factors for this non-linearity.

Finally, the time series of precipitation separated into rainfall and snowfall is obtained for the whole data set. A sample for the event of 21-22 Feb. 2004 is shown in figure 6 for the four selected HRUs (1,4,13,37). It is interesting to observe the variability of the snowfall/rainfall partition from hour to hour during the same event. The spatial variability in snowfall (or snowfall/rainfall ratio) between the different HRUs is also appreciable.

Uncertainty is introduced in snowfall estimates from both the precipitation and the temperature data. In this study, the best performing temperature (LDK) and precipitation (DK) estimation at each HRU is used to predict the snowfall estimate.

The hourly time-series analysis of SWE for some selected

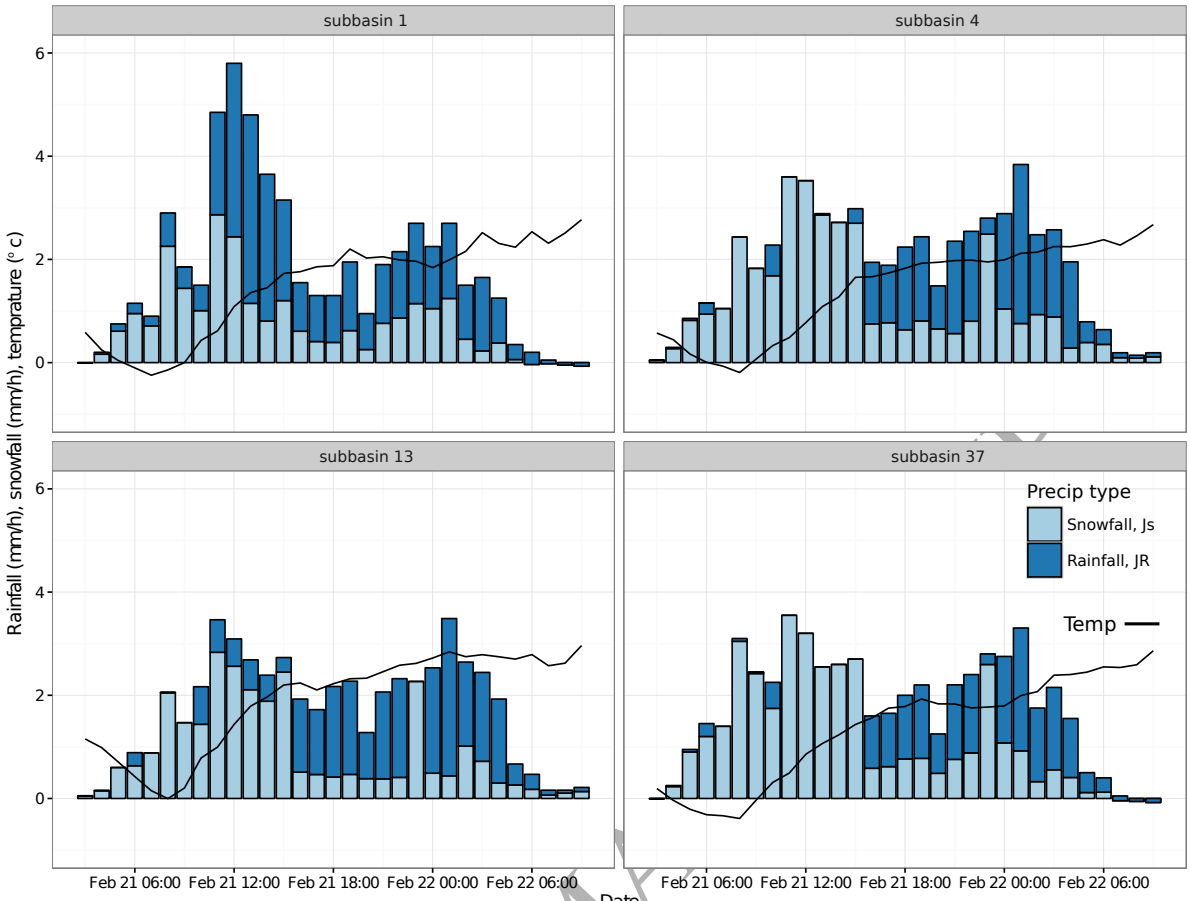


Figure 6: Comparison of time series of snowfall separation estimates during the 21-22 Feb. 2004 event for four selected HRUs (HRU 1, 4, 13, 37). Modelling at HRU level, which is the aggregation of each point within the HRU that can be characterized by pure snowfall or pure rainfall or snow-water mix event, the water-snow mixing is more physically and statistically meaningful. The line is the time series temperature data.

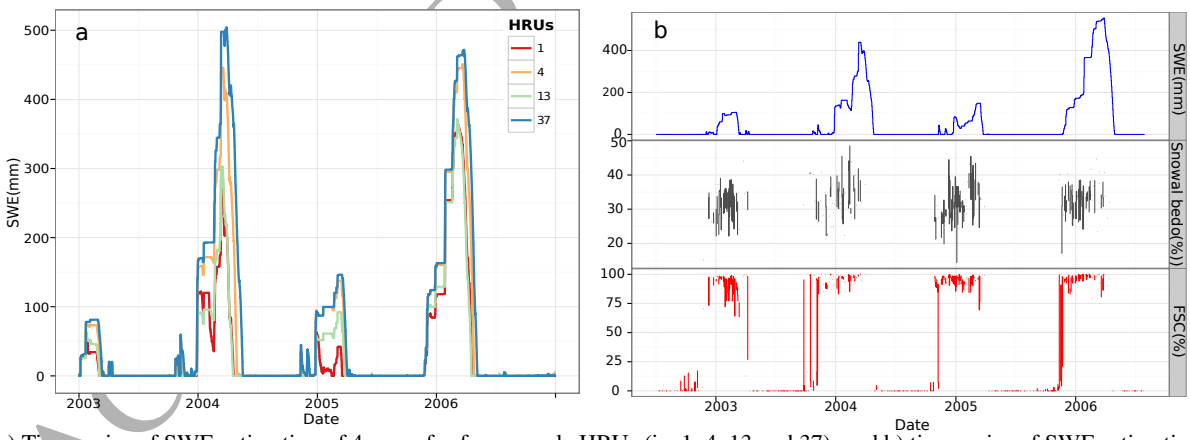


Figure 7: a) Time series of SWE estimation of 4 years for four sample HRUs (i.e 1, 4, 13,and 37); and b) time series of SWE estimation for HRU 1 (above) along with MODIS snow albedo (middle) and fractional snow cover (bottom)

HRUs is shown in figure 7 a. The 18 years of simulation shows that snow forms mostly in the period from October to March/April of the following year. Although, the actual timing in the formation period varies across years. The mean annual snow accumulation on the basin was 192± 88 mm. The annual variability of snow accumulation is appreciable. The winters of 1997/1998, 2000/2001, and 2006/2007 show small snow accumulations, while the winters of 2003/2004, 2005/2006, 2008/2009, and 2010/2011 show higher snow accumulations in the basin. Considering all the winter seasons over the 18 years, the winter of 2008/2009 has the highest accumulation (515 mm) and the winter of 2006/2007 the lowest (99.5 mm). The SWE estimation comparison between HRUs (figure 7a) shows that, as expected, the HRUs at higher elevation (HRU 1 and 37) always have higher estimated values than the HRUs at lower elevations (HRUs 4 and 13). The plot of SWE against the MODIS FSc and

snow albedo (figure 7 b) for a single HRU (HRU1) shows some level of consistency between the SWE dynamics and MODIS imagery data. The correlation of SWE values with snow fractional area coverage is high (0.65), while with snow albedo is low (0.15).

### 4.3. Streamflow modelling and impacts of rainfall interpolations

Four precipitation data sets generated using four types of kriging are used as input for the discharge model. To investigate the effect of precipitations generated using different kriging methods, all four kriging interpolation data sets are used to calibrate the rainfall-runoff model, even though LOK was actually found to perform better in precipitation reproduction. The different kriging methods cause different performances in reproducing runoff (as shown in Table 4 and figure 8). Surprisingly, the result using LOK (relatively the best in reproducing measured precipitation) is actually the lowest in reproducing the observed discharge, with KGE=0.78 during the calibration period and KGE=0.40 during validation. This is also confirmed by the Percentage bias (PBIAS) indicator (Appendix B) results. In general, the simulation results using LDK and DK are acceptable (Moriasi et al., 2007) in both calibration and validation periods, while OK and LOK return very low goodnes-of-fit (GOF) indices values during the validation period (table 4).

Based on the overall ranking of performances using the KGE index, the DK interpolation method is found to be the best performing, followed by LDK, OK and LOK respectively. Hence, we decide to use DK (instead of LOK) for further discharge analyses. For the impacts of different kriging generated precipitation data sets on long-term runoff estimation, figure 8 shows that, in general, all the precipitation data sets generate higher long-term annual runoff than really observed and that simulations using DK input are relatively better than the others.

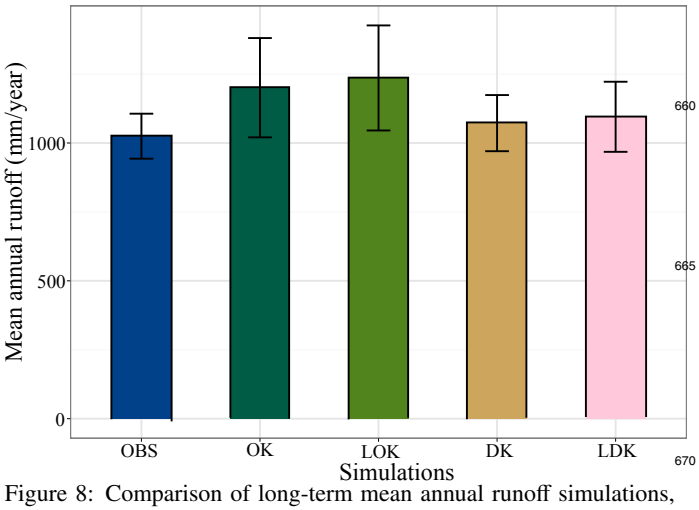


Figure 8: Comparison of long-term mean annual runoff simulations, using different kriging interpolation precipitation inputs, with observed discharge at basin outlet.

The higher performances of the DK methods would indicate that this method, even if not visible in the calibration phase, could actually better capture the physics of the processes. A

Table 4: The model performance statistics of the rainfall-runoff model, based on the four types of kriging interpolation methods used for input data. Performances of the model during both the calibration and validation periods. Percent bias (PBIAS) measures range form $-\infty$ to $+\infty$, with an optimal value of 0.0. Positive values indicate model underestimation bias, and negative values indicate model overestimation bias. DK*, which is used for the water budget analysis, is based on DK precipitation and is optimized for both discharge and Budyko assumption of five years.

| | Calibration | | | Validation | | |
|---|---|---|---|---|---|---|
| Methods | KGE | PBIAS | r | KGE | PBIAS | r |
| OK | 0.80 | 1.80 | 0.80 | 0.40 | 66.8 | 0.50 |
| LOK | 0.78 | 5.2 | 0.79 | 0.35 | 40.1 | 0.49 |
| DK | 0.85 | -0.8 | 0.85 | 0.56 | 14.30 | 0.68 |
| LDK | 0.83 | 1.30 | 0.83 | 0.56 | 15.50 | 0.66 |
| DK* | 0.71 | 14.6 | 0.81 | 0.63 | -14.30 | 0.82 |

sample calibration and validation hydrograph from the four precipitation data sets is available in the supplementary material.

In the simulations above, the ADIGE and PT models are coupled to optimize for both the discharge and the Budyko assumption of water budget: the performances are reported in table 4 (last row). This simulation is used to simulate discharge at each link of the basin. Observed discharges at two interior points, not used in the calibration process (for their locations see figure 1b, they drain areas of 22.2 km$^2$ and 38.8 km$^2$ respectively), are used to evaluate the potential of the model for estimation of runoff at each link (figure 9). On the basis of all the performance statistics, the calibrated model solution provides acceptable results at the two interior links (Moriasi et al., 2007). In fact, a KGE=0.73 and a PBIAS=3.30% at Valoje (ID 201) and a KGE= 0.62 and a PBIAS=2.50% at Bazzoni (ID 203) are obtained. While there are some studies that show a degradation of model performance when applying basin outlet calibration parameters to interior sites (Moussa et al., 2007; Feyen et al., 2008; Boscarello et al., 2013), the study of Lerat et al. (2012), using a large number of basins, has demonstrated that a semi-distributed rainfall-runoff model using a single site (such as a basin outlet) calibration could give acceptable estimates of ungauged internal points. Our case seems to confirm the latter case, but we are not able to discern if this is due to the characteristics of our modelling solutions or, perhaps, to the study basin size, dimension and location, although size and dimension certainly play a role.

### 4.4. Evapotranspiration estimation and uncertainty

ET is estimated at hourly time steps for each HRU, by optimizing PT's $\alpha'$ against water budget steady-state hypothesis (stationarity). The calibration procedure for parameter $\alpha$ returns the results shown in Figure 10a.

Varying $T_B$ from one year to seventeen years, as described in section 3.5, the value of PT's $\alpha$ varies between 0.38 and 0.8, with the mean $\bar{\alpha} = 0.56$ (figure 10a). This result represents an aftermath verification of the chosen method. The variability of $\alpha$ about the mean, while still large, is one half of the one found in literature. Besides, the $\alpha$ itself, which here assumes

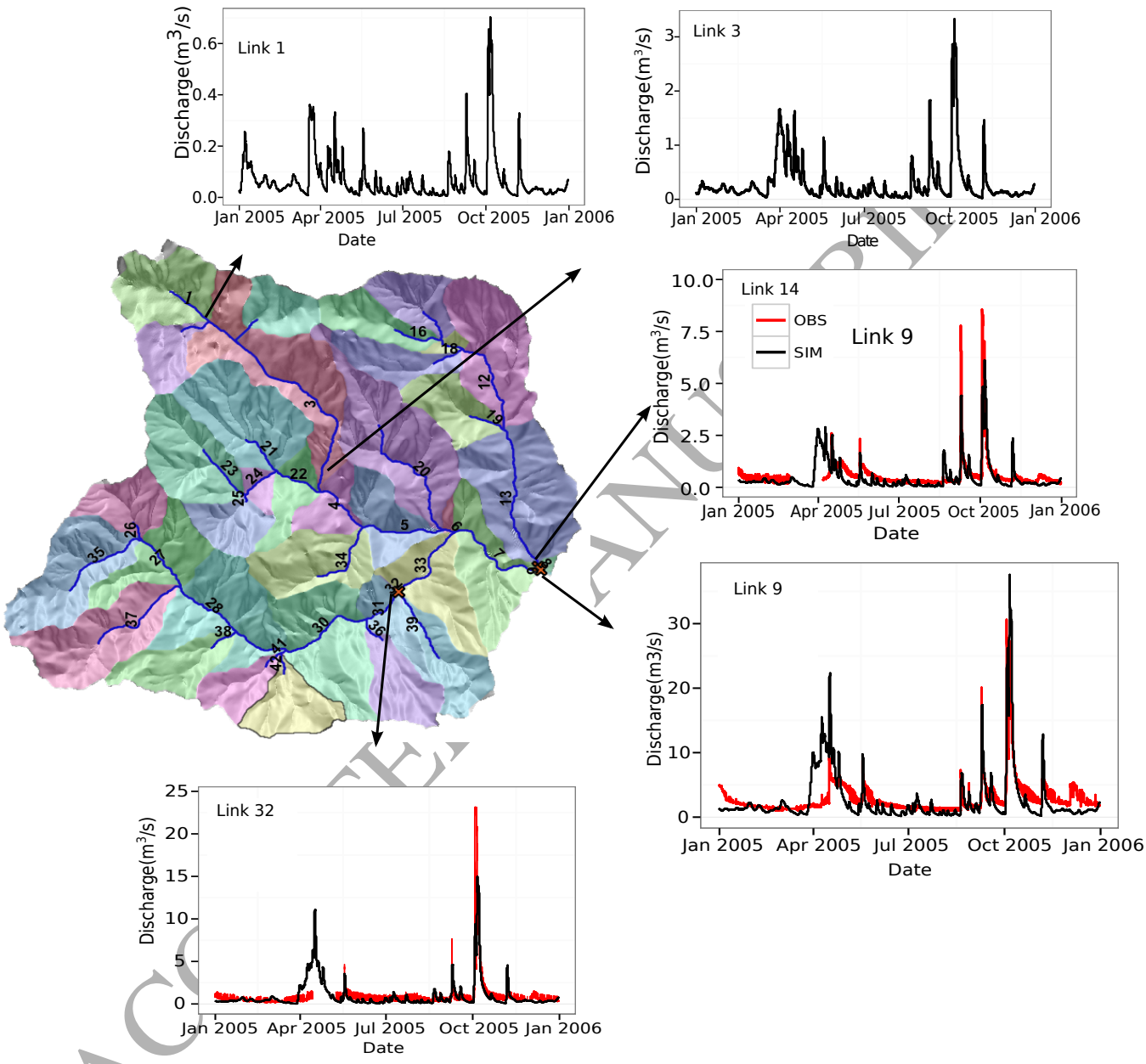


Figure 9: NewAGE model forecasting validation at internal links. Discharge is estimated for all links, here plotted for links 1,3, 9 (the outlet links), and 14, 27, and 32 as samples. When data is available at any internal point, model performances can be evaluated by comparison (e.g. for links 14 and 32).

a different meaning being related to the actual evapotranspiration and not to the potential one, is at the lower end of values used in literature (Aminzadeh and Or, 2014; Cristea et al., 2012; Cho et al., 2012; Viswanadham et al., 1991; Carmona et al., 2013). Gradually increasing the number of years for the water balance closure assumption, the $\alpha$ is highly variable for the first ten years of simulation, subsequently becoming relatively stable around the mean for the remaining seven years (figure 10a), suggesting that taking the $T_B \sim 10$ years would be a reasonable choice.

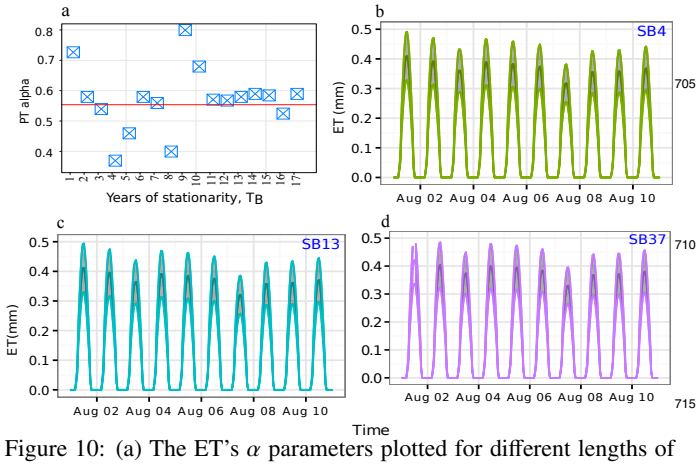


Figure 10: (a) The ET's $\alpha$ parameters plotted for different lengths of assumed stationarity; (b, c, and d) estimated ET at subbasin 4, 13, and 37 for 10 days of hourly simulation in August 2005, respectively.The grey band is the uncertainty band associated with the water balance approach to ET estimation.

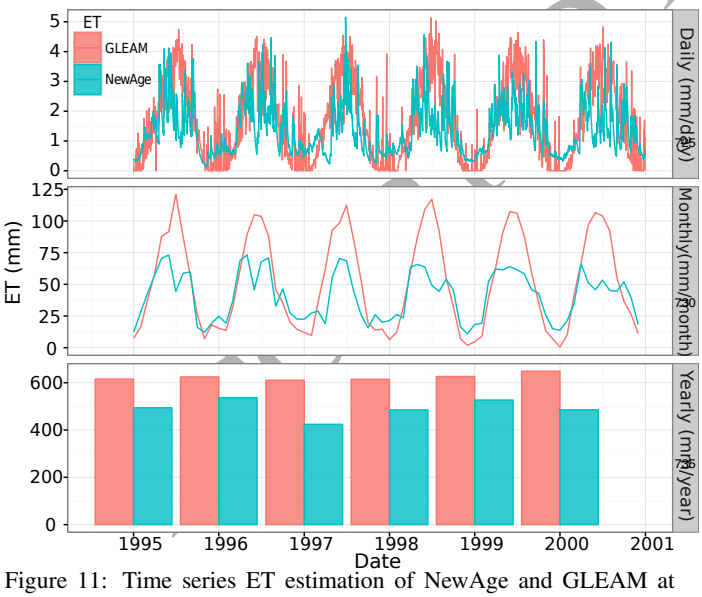


Figure 11: Time series ET estimation of NewAge and GLEAM at the whole basin scale at daily (above), monthly (middle) and yearly (lower) time steps.

Following proper determination of the $\alpha$ values of the PT equation, the ET of each HRU is estimated with its uncertainty errors. Figure 10 is a sample of time-series of ET during August 2005 for four HRUs (1, 4, 13, 37). According to our procedures, the resulting annual ET ranges from a maximum of 550 mm in 2001/2 to a minimum of 424 mm in 2002/2003. The errors of estimation can vary as much as 20%. A comparison of time series (aggregated at three time resolutions i.e. daily, monthly, and yearly) of NewAge *ET* with GLEAM estimates for whole basin is shown in figure 11 for 6 years (1995-2000). At daily time step, the correlation and the PBIAS between GLEAM and modelled ET are 0.82 and -35.7 respectively, whereas at monthly time step they are 0.9 and -55.5, respectively (figure 11). The comparison result shows that NewAGE *ET* is more variable than GLEAM ET. GLEAM clearly returns higher estimated values. From 1995-2000, the maximum annual difference between the two estimations is about 180 mm in 1997, while the minimum is 50 mm in 1996. The basin scale difference is accumulating and gives a difference of 715 mm after the six years and is not compatible with the ground measurements, i.e. precipitation and discharge, we have.

In accordance with the hypothesis of consistency we make, the relative water storage varies from a negative maximum of 410 mm to a positive maximum of 87 mm. While the quantitative assessments of both ET and relative storage can be thought of as imprecise, their overall internal variability cannot certainly be neglected and must be considered a product of this modeling effort.

### *4.5. Closing the water budget at basin scale*

Figure 12 shows the mean estimate of the water budget obtained with the method described in the previous sections and in the Appendix. The annual water balance is based on the hydrological year, in this case from October to September of the following year.

The components of the water balance of the basin are estimated for each year and the relative share of each component can be observed in figure 12. Overall, in the same figure 12, one can see that years with higher precipitation are accompanied by higher Q, which indicates that increases in J tend to contribute directly to Q with minor effects on S and ET. $Q$ and $ET$ exchange their roles over the years and there are clearly years where Q is larger than ET and others where it is vice-versa, defining, therefore, a climatic characteristic of the basin.

The hypothesis that the budget is stationary after $T_B$ years implies major interannual variability with both negative and positive storages. This variability is very pronounced and, in the catchment under study, $ET$ varies from about 19% to 35%, $Q$ from 64% to 95% and $S$ from -19% to 5% of the whole yearly budget. The negative storage changes, at the particular years, are most likely attributed to: 1) the amount, seasonal variability and nature of precipitation and/or 2) high atmospheric demand causing high evapotranspiration. Essentially this means that outflow (ET plus Q) is higher than the inflow (J), which is possible because storage was accumulated in the previous years. The main source of variability in the budget is clearly in the rainfall input as in Fatichi et al. (2016a) and Wang et al. (2014c). ET tends to be pretty smooth (figure 12), since the main driver of ET is radiation which is relatively consistent across years. It is possible that simulated ET is smoother than in reality, since it

has been estimated on average by assuming a single PT parameter over the whole eighteen years of simulation. However, also other studies in similar catchments in forested and humid areas have shown that ET has very low interannual variability (Lewis et al., 2000; Yoshiyukiishii and Nakamura, 2004; Oishi et al., 2010; Möller and Stanhill, 2007).

The great part (88%, $R^2$=0.88) of Q variance is explained by the variance of $J$. As mentioned above, the variability in $ET$ is smaller than other components, and only 38% can be explained by $J$ (i.e. $R^2$=0.38). The error of estimation is shown by the error bars in figure 12, which is highest and lowest for $S$ and $Q$, respectively. This is expected because $Q$ is the more reliable of the measured data sets in the water budget equation at the basin scale (Wang et al., 2014b). Since the budget was actually simulated at hourly time steps and at small units (HRU), plots analogous to figure 12 can be produced for any hour of any day of the year and for any HRU of the basin. These can be obtained with increasing uncertainty with decreasing time steps, due to the ways in which the $ET$ parameter and storage were assessed.

As an example, figure 13 shows the monthly budget estimated for the year 2011/2012. It was obtained by using the interpolated rainfalls (with DK), the simulated discharges, using the parameter calibrated between 1995-1999, and the PT calibrated for $T_B$ = 5 years. The highest variability is mainly in $J$ and $S$. During the summer, all the components show high magnitude (high $J$, high $Q$). The variability in monthly $S$ is governed by $J$, though non-linearly. $ET$ is evidently connected with the annual cycle of solar radiation. It is highest in June and July and lowest from November to February, as expected, but it is less smooth than in the annual budget. In 2012, from late spring to August, evapotranspiration is sustained by the water storage more than from direct precipitation, indicating that, without winter rainfalls, the vegetation of the catchment could have undergone considerable stress in these months.

To show the variability between HRUs, the monthly means of the water budget components from the 18 years of simulations have been analyzed, and the mean monthly estimates for four months (January, April, July, and October, one from each season) are presented in figure 14. The result confirms the monthly analysis given for the year 2012 (figure 13). The trend in $Q$ follows the trend in $J$, but it is not linearly proportional.

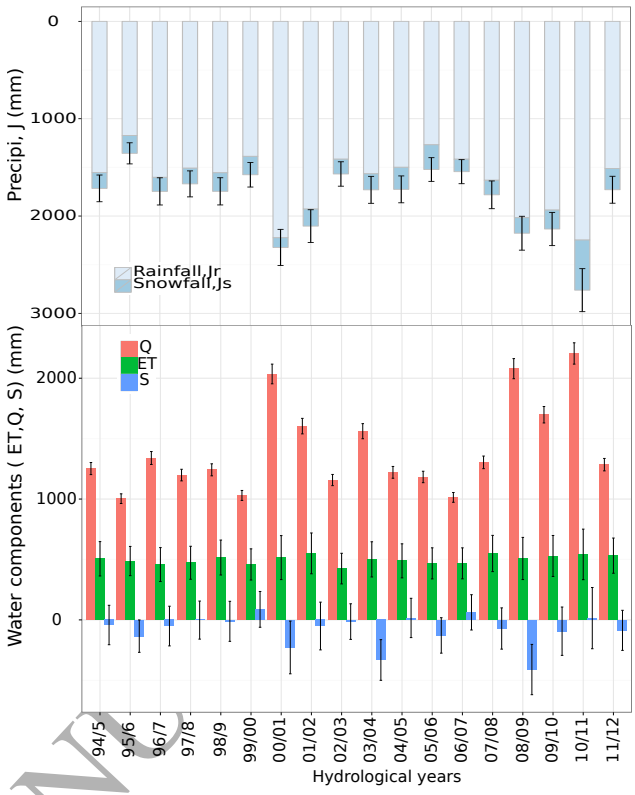


Figure 12: Water budget components of the basin and their annual variability from 1994/95 to 2011/2012. The lower graph shows the relative share (the size of the bars) of the three components (Q, ET and S). The upper graph shows the total available water, $J$, divided in its snow and rainfall parts.

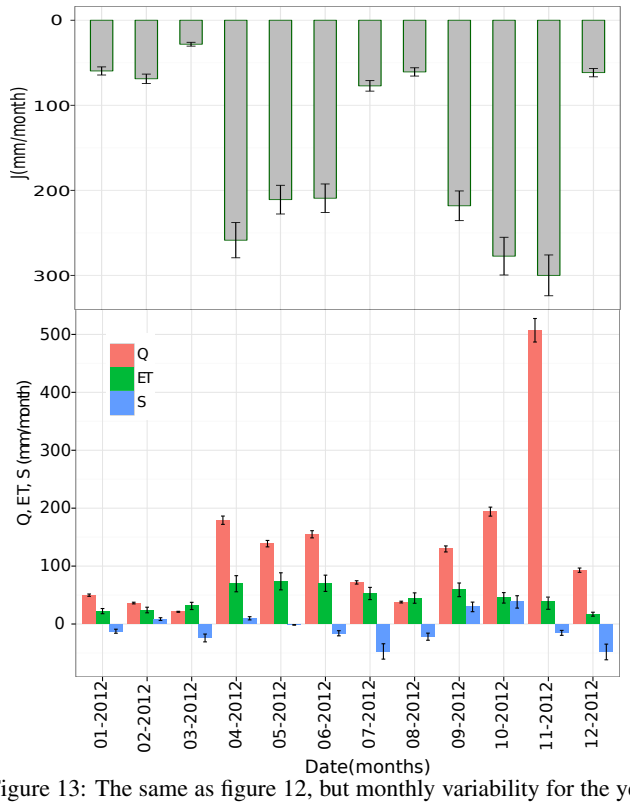


Figure 13: The same as figure 12, but monthly variability for the year 2012.

## 5. Conclusions

The water budget of the Posina River basin has been analyzed with the NewAGE system at hourly time-steps, using 18 years of meteorological data (rainfall and temperature) and discharge data. The analyses include estimations of the four components of the water budget (precipitation, discharge, relative storage, and evapotranspiration) under the hypothesis of stationarity (i.e. null storage) for one of the years where measurements are available. NewAGE system components are used to capture basin behaviour and forecast the water cycle. The procedure implemented is general and can be transposed to all basins where the same type of data are available. Whilst precipitation was measured in twelve locations, part of the work was to interpolate the data and analyse when they were liquid

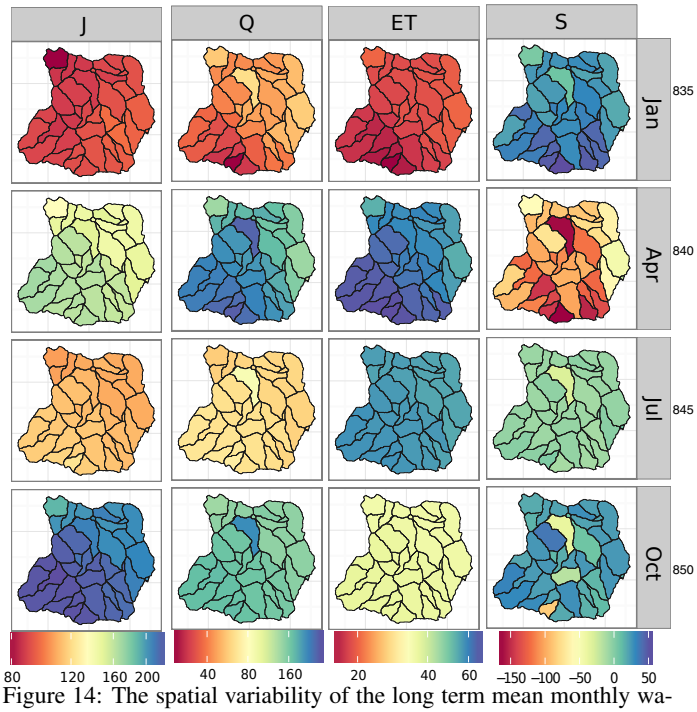


Figure 14: The spatial variability of the long term mean monthly water budget components (J, ET, Q, S). For clarity, the colour scale is separate for each component.

(rainfall) or solid (snow). To separate snowfall and rainfall, a new method based on the use of MODIS satellite data was used. This allowed us to decouple this problem from the one of simulating discharges. The spatial distribution of precipitation was obtained by the use of kriging, almost a standard methodology, which we improved by using the calibration methods permitted by the NewAGE system: a real time calibration method, in which the model and parameters of the kriging semiovariogram are chosen automatically time step by time step.

To obtain accurate spatial time series of precipitation data for water budget modeling, the kriging procedure is validated using cross-validation methods. It emerges that OK and LOK performed better than DK and LDK in detecting the total mass of water that fell in the basin.

The variability of total annual rainfall is high, with the lowest annual J of 1355 mm in 1995/1996 and greatest of 2700 mm in 2010/2011, with a mean of 1730±344 mm. For use in the NewAGE ADIGE rainfall-runoff component, precipitation and temperature data are produced for each HRU. To asses the impact of precipitation interpolation and its coarse graining at hillslope scale, we analyze the discharge forecast by using all four kriging methods, independently of their performances in assessing the precipitation volume. As anticipated, the DK and LDK performances were found to give better results than the OK and LOK, reversing the previous results.

The GOF index of the simulated discharge against observed discharge shows that the model performances are acceptable (especially considering that we are simulating hourly discharges). Using discharge measures within the basin, it is possible to quantify the reliability of internal discharge estimations by assuming the validity of model parameters calibrated at the furthermost downstream outlet. The model maintains similar performances at the interior sites, which is probably due to the small size of the catchment considered. This is indeed a good result, which contrasts what obtained several times in past literature. The inter annual variability in Q is high, with the minimum annual Q of 1003 mm in 1995/96 and the maximum of 2072 mm in 2010/2011, with estimated errors of 372 mm.

Finally, the Priestley-Taylor method is used to estimate the evapotranspiration components of the water budget and to infer the relative storage of water. To obtain this, the hypothesis of null storage after an assigned number of years, named Budyko's time, is used. By moving the null storage hypothesis time incrementally along the first 17 years of data, we obtain different values of the PT $\alpha$. The variability of $\alpha$ can be interpreted as an estimate of the epistemic errors in ET (to be accumulated with the uncertainty produced in precipitation and discharge estimates). We actually use two models of $ET$, one using the standard PT formulation, and a second assuming a reduction in $ET$ proportional to the water storage in each HRU.

In the first case, over the 17 years, the mean was $\alpha = 0.56 \pm 0.1$, which is considerably lower than what is proposed by literature (where, however, it is relative to the potential evapotranspiration). In the second case, the PT $\alpha$ was considered as an optimization parameter to add to the ADIGE component set. In this case, due to the length of the procedure, just the case with $T_B = 5$ years was considered, obtaining $\alpha = 1.89$ (which reconciles with the values found in literature). Remarkably, this procedure of simultaneous calibration is made easier by the modular structure of the NewAGE system. In both the cases, the contribution of ET to the water budget and its uncertainty are very high. While ET accounts 19-38% of J, its annual uncertainty is about 20% (148mm on average). ET is certainly variable in space and in seasonal time scale but the annual totals tends to be similar over the years due to constant annual atmospheric demand.

With all the disclaimers of the case, once setup on a basin, the NewAGE system can produce data which are of real interest in many practical cases of water management. It represents a considerable improvement in the analysis of fluxes of the hydrological cycle within catchments where evapotranspiration is usually evaluated without reference to mass conservation and the water budget is not clearly closed.

The capabilities of JGrass-NewAge system are not fully exploited in this study, but it produces estimates of the above quantities for any hour and for any subcatchment (HRU), with errors that depend on where the parameters of the model chain can be calibrated. This study would benefit of further investigations on: 1) the effect of basin partitioning level; 2) the use of surface energy budget closure; and 3) an effective treatment of vegetation heterogeneity and dynamics.

## Code Availability and reproducibility

To reproduce this paper's findings:

- Catchment DEM can be found at https://zenodo.org/badge/DOI/10.5281/zenodo.215011.svg

- The source code of JGrass-NewAGE is available on GitHub at https://github.com/geoframecomponents.
- Documentation of the code and modelling solutions for practice are at http://geoframe.blogspot.it/
- The supplementary material and a lot of other material is posted at the AboutHydrology Blog (http://abouthydrology.blogspot.it/2015/08/estimating-hydrological-budgets-with.html).

## Acknowledgments

The authors acknowledge the Trento University project CLIMAWARE (/http://abouthydrology.blogspot.it/search/label/CLIMAWARE) and the European Union FP7 Collaborative Project GLOBAQUA (Managing the effects of multiple stressors on aquatic ecosystems under water scarcity, grant no. 603629-ENV-2013.6.2.1) that partially financed this research. They also thank three anonymous reviewers for their work that helped to enhance the initial manuscript with their comments.

## 6. References

## References

Abera, W., 2016. Modelling water budget at a basin scale using jgrass-newage system. Ph.D. thesis, University of Trento.

Abera, W., Antonello, A., Franceschi, S., Formetta, G., Rigon, R., 2014. The uDig Spatial Toolbox for hydro-geomorphic analysis, in: clarke & nield (eds.) geomorphological techniques (online edition) Edition. British Society for Geomorphology, London, UK.

Aminzadeh, M., Or, D., 2014. Energy partitioning dynamics of drying terrestrial surfaces. Journal of Hydrology 519, 1257–1270.

Army, U., 1956. Snow hydrology. summary report of the snow investigations. north pacific division of corps of engineers. US-Army, Portland, Oregon.

Arnold, J., Allen, P., 1996. Estimating hydrologic budgets for three illinois watersheds. Journal of hydrology 176 (1), 57–77.

Arnold, J. G., Srinivasan, R., Muttiah, R. S., Williams, J. R., 1998. Large area hydrologic modeling and assessment part i: Model development1.

Ascough, J., David, O., Krause, P., Heathman, G., Kralisch, S., Larose, M., Ahuja, L., Kipka, H., 2012. Development and application of a modular watershed-scale hydrologic model using the object modeling system: runoff response evaluation. Transactions of the ASABE 55 (1), 117–135.

Ashraf, M., Loftis, J. C., Hubbard, K., 1997. Application of geostatistics to evaluate partial weather station networks. Agricultural and forest meteorology 84 (3), 255–271.

Basistha, A., Arya, D., Goel, N., 2008. Spatial distribution of rainfall in indian himalayas–a case study of uttarakhand region. Water Resources Management 22 (10), 1325–1346.

Batelaan, O., De Smedt, F., 2007. Gis-based recharge estimation by coupling surface–subsurface water balances. Journal of Hydrology 337 (3), 337–355.

Bertoldi, G., Rigon, R., Over, T. M., 2006. Impact of watershed geomorphic characteristics on the energy and water budgets. Journal of Hydrometeorology 7 (3), 389–403.

Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., et al., 2015. Hyper-resolution global hydrological modelling: what is next?everywhere and locally relevant. Hydrol Process 29 (2), 310–320.

Borga, M., 2002. Accuracy of radar rainfall estimates for streamflow simulation. Journal of Hydrology 267 (1), 26–39.

Boscarello, L., Ravazzani, G., Mancini, M., 2013. Catchment multisite discharge measurements for hydrological model calibration. Procedia Environmental Sciences 19, 158–167.

Brocca, L., Massari, C., Ciabatta, L., Moramarco, T., Penna, D., Zuecco, G., Pianezzola, L., Borga, M., Matgen, P., Martínez-Fernández, J., 2015. Rainfall estimation from in situ soil moisture observations at several sites in europe: an evaluation of the sm2rain algorithm. Journal of Hydrology and Hydromechanics 63 (3), 201–209.

Budyko, M., 1978. 1., 1974: Climate and life. International Geophysics Series 18.

Buytaert, W., Celleri, R., Willems, P., Bièvre, B. D., Wyseure, G., 2006. Spatial and temporal rainfall variability in mountainous areas: A case study from the south ecuadorian andes. Journal of Hydrology 329 (3), 413–421.

Caldwell, P., Chin, H.-N. S., Bader, D. C., Bala, G., 2009. Evaluation of a wrf dynamical downscaling simulation over california. Climatic Change 95 (3-4), 499–521.

Carmona, F., Rivas, R., Caselles, V., 2013. Estimate of the alpha parameter in an oat crop under rain-fed conditions. Hydrological Processes 27 (19), 2834–2839.

Cho, J., Oki, T., Yeh, P. J.-F., Kim, W., Kanae, S., Otsuki, K., 2012. On the relationship between the bowen ratio and the near-surface air temperature. Theoretical and Applied Climatology 108 (1-2), 135–145.

Chu, T., Shirmohammadi, A., 2004. Evaluation of the swat model's hydrology component in the piedmont physiographic region of maryland. Transactions of the ASAE 47 (4), 1057–1073.

Claessens, L., Hopkinson, C., Rastetter, E., Vallino, J., 2006. Effect of historical changes in land use and climate on the water budget of an urbanizing watershed. Water Resources Research 42 (3).

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., Hay, L. E., 2008. Framework for understanding structural errors (fuse): A modular framework to diagnose differences between hydrological models. Water Resources Research 44 (12).

Clothier, B., Kerr, J., Talbot, J., Scotter, D., 1982. Measured and estimated evapotranspiration from well-watered crops. New Zealand journal of agricultural research 25 (3), 301–307.

Cristea, N. C., Kampf, S. K., Burges, S. J., 2012. Revised coefficients for priestley-taylor and makkink-hansen equations for estimating daily reference evapotranspiration. Journal of Hydrologic Engineering 18 (10), 1289–1300.

David, O., Ascough Ii, J., Lloyd, W., Green, T., Rojas, K., Leavesley, G., Ahuja, L., 2013. A software engineering perspective on environmental modeling framework design: The object modeling system. Environmental Modelling & Software 39, 201–213.

Dean, J., Camporese, M., Webb, J., Grover, S., Dresel, P., Daly, E., 2016. Water balance complexities in ephemeral catchments with different land uses: Insights from monitoring and distributed hydrologic modeling. Water Resources Research.

Dudhia, J., Gill, D., Henderson, T., Klemp, J., Skamarock, W., Wang, W., 2005. The weather research and forecast model: software architecture and performance. In: Proceedings of the Eleventh ECMWF Workshop on the Use of High Performance Computing in Meteorology. World Scientific, pp. 156–168.

Eagleson, P. S., 1994. The evolution of modern hydrology (from watershed to continent in 30 years). Advances in water resources 17 (1), 3–18.

Fang, Z., Bogena, H., Kollet, S., Koch, J., Vereecken, H., 2015. Spatio-temporal validation of long-term 3d hydrological simulations of a forested catchment using empirical orthogonal functions and wavelet coherence analysis. Journal of hydrology 529, 1754–1767.

Fatichi, S., Ivanov, V. Y., Paschalis, A., Peleg, N., Molnar, P., Rimkus, S., Kim, J., Burlando, P., Caporali, E., 2016a. Uncertainty partition challenges the predictability of vital details of climate change. Earth's Future 4 (5), 240–251.

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., Downer, C. W., Camporese, M., Davison, J. H., Ebel, B., et al., 2016b. An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. Journal of Hydrology 537, 45–60.

Fenicia, F., McDonnell, J. J., Savenije, H. H., 2008. Learning from model improvement: On the contribution of complementary data to process understanding. Water Resources Research 44 (6), W06419.

Feyen, L., Kalas, M., Vrugt, J. A., 2008. Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization/optimisation de paramètres semi-distribués et évaluation de l'incertitude pour la simulation de débits à grande échelle par l'utilisation d'une optimisation globale. Hydrological Sciences Journal

53 (2), 293–308.

Formetta, G., 2013. Hydrological modelling with components: the oms3 newage-jgrass system. Ph.D. thesis, University of Trento.

Formetta, G., Antonello, A., Franceschi, S., David, O., R., R., 2014a. The basin delineation and the built of a digital watershed model within the jgrass-newage system. Bolet?n Geol?gico y Minero: Special Issue "Advanced GIS terrain analysis for geophysical applications.

Formetta, G., Antonello, A., Franceschi, S., David, O., Rigon, R., 2014b. Hydrological modelling with components: A gis-based open-source framework. Environmental Modelling & Software 55, 190–200.

Formetta, G., Bancheri, M., David, O., Rigon, R., 2016. Site specific parameterizations of longwave radiation. Hydrology and Earth System Sciences Discussions 2016, 1–22.

Formetta, G., Kampf, S. K., David, O., Rigon, R., 2014c. Snow water equivalent modeling components in newage-jgrass. Geoscientific Model Development 7 (3), 725–736.

Formetta, G., Mantilla, R., Franceschi, S., Antonello, A., Rigon, R., 2011. The jgrass-newage system for forecasting and managing the hydrological budgets at the basin scale: models of flow generation and propagation/routing. Geoscientific Model Development 4 (4), 943–955.

Formetta, G., Rigon, R., Chávez, J., David, O., 2013. Modeling shortwave solar radiation using the jgrass-newage system. Geoscientific Model Development 6 (4), 915–928.

Francois, B., Zoccatelli, D., Borga, M., Submitted. Assessing small hydro/solar power complementarity in ungauged mountainous areas: a crash test study for hydrological prediction methods. Energy.

Garen, D. C., Johnson, G. L., Hanson, C. L., 1994. Mean areal precipitation for daily hydrologic modeling in mountainous regions1.

Garen, D. C., Marks, D., 2005. Spatially distributed energy balance snowmelt modelling in a mountainous river basin: estimation of meteorological inputs and verification of model results. Journal of Hydrology 315 (1), 126–153.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford university press.

Goovaerts, P., 2000. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. Journal of hydrology 228 (1), 113–129.

Graf, A., Bogena, H. R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., Vereecken, H., 2014. Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment. Water resources research 50 (6), 4837–4857.

Gupta, H. V., Kling, H., Yilmaz, K. K., Martinez, G. F., 2009. Decomposition of the mean squared error and nse performance criteria: Implications for improving hydrological modelling. Journal of Hydrology 377 (1), 80–91.

Gupta, H. V., Sorooshian, S., Yapo, P. O., 1999. Status of automatic calibration for hydrologic models: Comparison with multilevel expert calibration. Journal of Hydrologic Engineering 4 (2), 135–143.

Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. Journal of Hydrology 332 (1), 144–157.

Hall, D. K., Riggs, G. A., Salomonson, V. V., 2006. Modis snow and sea ice products. In: Earth science satellite remote sensing. Springer, pp. 154–181.

Harder, P., Pomeroy, J., 2013. Estimating precipitation phase using a psychrometric energy balance method. Hydrological Processes 27 (13), 1901–1914.

Harder, P., Pomeroy, J. W., 2014. Hydrological model uncertainty due to precipitation-phase partitioning methods. Hydrological Processes 28 (14), 4311–4327.

Hay, L. E., Leavesley, G. H., Clark, M. P., Markstrom, S. L., Viger, R. J., Umemoto, M., 2006. Step wise, multiple objective calibration of a hydrologic model for a snowmelt dominated basin1.

He, Z., Parajka, J., Tian, F., Blöschl, G., 2014. Estimating degree-day factors from modis for snowmelt runoff modeling. Hydrology and Earth System Sciences 18 (12), 4773–4789.

Hingerl, L., Kunstmann, H., Wagner, S., Mauder, M., Bliefernicht, J., Rigon, R., 2016. Spatiotemporal variability of water and energy fluxes-a case study for a meso-scale catchment in pre-alpine environment. Hydrological Processes.

Högström, U., 1968. Studies on the water balance of a small natural catchment area in southern sweden. Tellus 20 (4), 633–641.

Huntington, J. L., Niswonger, R. G., 2012. Role of surface-water and groundwater interactions on projected summertime streamflow in snow dominated regions: An integrated modeling approach. Water Resources Research 48 (11).

Jothityangkoon, C., Sivapalan, M., Farmer, D., 2001. Process controls of water balance variability in a large semi-arid catchment: downward approach to hydrological model development. Journal of Hydrology 254 (1), 174–198.

Kavetski, D., Kuczera, G., Franks, S. W., 2006. Calibration of conceptual hydrological models revisited: 1. overcoming numerical artefacts. Journal of Hydrology 320 (1), 173–186.

Kennedy, J., Eberhart, R., et al., 1995. Particle swarm optimization. In: Proceedings of IEEE international conference on neural networks. Vol. 4. Perth, Australia, pp. 1942–1948.

Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper danube basin under an ensemble of climate change scenarios. Journal of Hydrology 424, 264–277.

Kottegoda, N. T., Rosso, R., 1997. Probability, statistics, and reliability for civil and environmental engineers.

Lagacherie, P., Rabotin, M., Colin, F., Moussa, R., Voltz, M., 2010. Geomhydas: a landscape discretization tool for distributed hydrological modeling of cultivated areas. Computers & Geosciences 36 (8), 1021–1032.

Lerat, J., Andreassian, V., Perrin, C., Vaze, J., Perraud, J.-M., Ribstein, P., Loumagne, C., 2012. Do internal flow measurements improve the calibration of rainfall-runoff models? Water Resources Research 48 (2).

Lewis, D., Singer, M., Dahlgren, R., Tate, K., 2000. Hydrology in a california oak woodland watershed: a 17-year study. Journal of Hydrology 240 (1), 106–117.

Li, H., Sivapalan, M., Tian, F., 2012. Comparative diagnostic analysis of runoff generation processes in oklahoma dmip2 basins: The blue river and the illinois river. Journal of Hydrology 418, 90–109.

Liang, X., Lettenmaier, D. P., Wood, E. F., Burges, S. J., 1994. A simple hydrologically based model of land surface water and energy fluxes for general circulation models. JOURNAL OF GEOPHYSICAL RESEARCH-ALL SERIES- 99, 14–415.

Lloyd, C., 2005. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in great britain. Journal of Hydrology 308 (1), 128–150.

Ly, S., Charles, C., Degre, A., 2011. Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the ourthe and ambleve catchments, belgium. Hydrology & Earth System Sciences 15 (7).

Maxwell, R. M., Condon, L., Jul. 2016. Conenctions between groundwater flow and transpiration partitioning. Science 353 (6297), 377–379.

Mazur, K., Schoenheinz, D., Biemelt, D., Schaaf, W., Grünewald, U., 2011. Observation of hydrological processes and structures in the artificial chicken creek catchment. Physics and Chemistry of the Earth, Parts A/B/C 36 (1), 74–86.

McCabe, M. F., Ershadi, A., Jimenez, C., Miralles, D., Michel, D., Wood, E. F., 2016. The gewex landflux project: evaluation of model evaporation using tower-based and globally gridded forcing data. Geoscientific Model Development 9 (1), 283–305.

Miralles, D., Holmes, T., De Jeu, R., Gash, J., Meesters, A., Dolman, A., 2011a. Global land-surface evaporation estimated from satellite-based observations. Hydrology and Earth System Sciences 15 (2), 453–469.

Miralles, D. G., De Jeu, R. A. M., Gash, J. H., Holmes, T. R. H., Dolman, A. J., 2011b. Magnitude and variability of land evaporation and its components at the global scale. Hydrology and Earth System Sciences 15 (3), 967–981. URL http://www.hydrol-earth-syst-sci.net/15/967/2011/

Mitchell, V. G., McMahon, T. A., Mein, R. G., 2003. Components of the total water balance of an urban catchment. Environmental Management 32 (6), 735–746.

Möller, M., Stanhill, G., 2007. Hydrological impacts of changes in evapotranspiration and precipitation: two case studies in semi-arid and humid climates/impacts hydrologiques de changements dans l'évapotranspiration et les précipitations: deux cas d'études en climats semi-aride et humide. Hydrological Sciences Journal/Journal des Sciences Hydrologiques 52 (6), 1216–1231.

Monteith, J., et al., 1965. Evaporation and environment. In: Symp. Soc. Exp. Biol. Vol. 19. p. 4.

Moore, R., 1985. The probability-distributed principle and runoff production at point and basin scales. Hydrological Sciences Journal 30 (2), 273–297.

Moore, R., 2007. The pdm rainfall-runoff model. Hydrology and Earth System Sciences 11, 483–499.

Moriasi, D., Arnold, J., Van Liew, M., Bingner, R., Harmel, R., Veith, T., 2007. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Trans. ASABE 50 (3), 885–900.

Mou, L., Tian, F., Hu, H., Sivapalan, M., 2008. Extension of the representative elementary watershed approach for cold regions: constitutive relationships and an application. Hydrology and Earth System Sciences 12 (2), 565–585.

Moussa, R., Chahinian, N., Bocquillon, C., 2007. Distributed hydrological modelling of a mediterranean mountainous catchment–model construction and multi-site validation. Journal of Hydrology 337 (1), 35–51.

Murray, F. W., 1967. On the computation of saturation vapor pressure. Journal of Applied Meteorology 6 (1), 203–204.

Nalder, I. A., Wein, R. W., 1998. Spatial interpolation of climatic normals: test of a new method in the canadian boreal forest. Agricultural and forest meteorology 92 (4), 211–225.

Norbiato, D., Borga, M., Degli Esposti, S., Gaume, E., Anquetin, S., 2008. Flash flood warning based on rainfall thresholds and soil moisture conditions: An assessment for gauged and ungauged basins. Journal of Hydrology 362 (3), 274–290.

Obojes, N., Bahn, M., Tasser, E., Walde, J., Inauen, N., Hiltbrunner, E., Saccone, P., Lochet, J., Clément, J., Lavorel, S., et al., 2015. Vegetation effects on the water balance of mountain grasslands depend on climatic conditions. Ecohydrology 8 (4), 552–569.

Ogden, F. L., Crouch, T. D., Stallard, R. F., Hall, J. S., 2013. Effect of land cover and use on dry season river runoff, runoff efficiency, and peak storm runoff in the seasonal tropics of central panama. Water Resources Research 49 (12), 8443–8462.

Oishi, A. C., Oren, R., Novick, K. A., Palmroth, S., Katul, G. G., 2010. Interannual invariability of forest evapotranspiration and its consequence to water flow downstream. Ecosystems 13 (3), 421–436.

Pan, M., Wood, E. F., 2006. Data assimilation for estimating the terrestrial water budget using a constrained ensemble kalman filter. Journal of Hydrometeorology 7 (3), 534–547.

Parajka, J., Holko, L., Kostka, Z., Blöschl, G., 2012. Modis snow cover mapping accuracy in a small mountain catchment–comparison between open and forest sites. Hydrology and Earth System Sciences 16 (7), 2365–2377.

Penna, D., Meerveld, H., Oliviero, O., Zuecco, G., Assendelft, R., Dalla Fontana, G., Borga, M., 2015. Seasonal changes in runoff generation in a small forested mountain catchment. Hydrological Processes 29 (8), 2027–2042.

Phillips, D. L., Dolph, J., Marks, D., 1992. A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. Agricultural and Forest Meteorology 58 (1), 119–141.

Porporato, A., D?odorico, P., Laio, F., Ridolfi, L., Rodriguez-Iturbe, I., 2002. Ecohydrology of water-controlled ecosystems. Advances in Water Resources 25 (8), 1335–1348.

Priestley, C., Taylor, R., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Monthly weather review 100 (2), 81–92.

Prudhomme, C., Reed, D. W., 1999. Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in scotland. International Journal of Climatology 19 (12), 1337–1356.

Rodell, M., Famiglietti, J., Chen, J., Seneviratne, S., Viterbo, P., Holl, S., Wilson, C., 2004. Basin scale estimates of evapotranspiration using grace and other observations. Geophysical Research Letters 31 (20).

Rodríguez-Iturbe, I., Porporato, A., 2004. Ecohydrology of water-controlled ecosystems. Soil Moisture and Plant Dynamics.

Rohrer, M., 1989. Determination of the transition air temperature from snow to rain and intensity of precipitation. In: WMO IASH ETH International Workshop on Precipitation Measurement. pp. 475–582.

Sahoo, A. K., Pan, M., Troy, T. J., Vinukollu, R. K., Sheffield, J., Wood, E. F., 2011. Reconciling the global terrestrial water budget using satellite remote sensing. Remote Sensing of Environment 115 (8), 1850–1865.

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research 46 (5).

Schaake, J. C., Koren, V. I., Duan, Q.-Y., Mitchell, K., Chen, F., 1996. Simple water balance model for estimating runoff at different spatial and temporal scales. Journal of Geophysical Research: Atmospheres 101 (D3), 7461–7475.

Scott, R. L., 2010. Using watershed water balance to evaluate the accuracy of eddy covariance evaporation measurements for three semiarid ecosystems. Agricultural and Forest Meteorology 150 (2), 219–225.

Sheffield, J., Ferguson, C. R., Troy, T. J., Wood, E. F., McCabe, M. F., 2009. Closing the terrestrial water budget from satellite remote sensing. Geophysical Research Letters 36 (7).

Singh, J., Knapp, H. V., Arnold, J., Demissie, M., 2005. Hydrological modeling of the iroquois river watershed using hspf and swat1.

Skamarock, W., Klemp, J., Dudhia, J., Gill, D., Barker, D., Duda, M., Huang, X., Wang, W., Powers, J., 2008. A description of the advanced research wrf version 3, ncar, tech. note, mesoscale and microscale meteorology division. National Center for Atmospheric Research, Boulder, Colorado, USA.

Steinacker, R., 1983. Diagnose und prognose der schneefallgrenze. Wetter und Leben 35 (81-90), 120.

Tobin, C., Nicotina, L., Parlange, M. B., Berne, A., Rinaldo, A., 2011. Improved interpolation of meteorological forcings for hydrologic applications in a swiss alpine region. Journal of Hydrology 401 (1), 77–89.

Tomasella, J., Hodnett, M. G., Cuartas, L. A., Nobre, A. D., Waterloo, M. J., Oliveira, S. M., 2008. The water balance of an amazonian micro-catchment: the effect of interannual variability of rainfall on hydrological behaviour. Hydrological Processes 22 (13), 2133–2147.

Van Der Knijff, J., Younis, J., De Roo, A., 2010. Lisflood: a gis-based distributed model for river basin scale water balance and flood simulation. International Journal of Geographical Information Science 24 (2), 189–212.

Viswanadham, Y., Silva Filho, V., Andre, R., 1991. The priestley-taylor parameter $\alpha$ for the amazon forest. Forest Ecology and Management 38 (3), 211–225.

Wang, S., Huang, J., Li, J., Rivera, A., McKenney, D. W., Sheffield, J., 2014a. Assessment of water budget for sixteen large drainage basins in canada. Journal of Hydrology 512, 1–15.

Wang, S., Huang, J., Yang, D., Pavlic, G., Li, J., 2014b. Long-term water budget imbalances and error sources for cold region drainage basins. Hydrological Processes.

Wang, S., McKenney, D. W., Shang, J., Li, J., 2014c. A national-scale assessment of long-term water budget closures for canada's watersheds. Journal of Geophysical Research: Atmospheres 119 (14), 8712–8725.

Wilson, K. B., Hanson, P. J., Mulholland, P. J., Baldocchi, D. D., Wullschleger, S. D., 2001. A comparison of methods for determining forest evapotranspiration and its components: sap-flow, soil water budget, eddy covariance and catchment water balance. Agricoltural and Forest Meteorology 106, 153–168.

Xue, M., Droegemeier, K. K., Wong, V., 2000. The advanced regional prediction system (arps)–a multi-scale nonhydrostatic atmospheric simulation and prediction model. part i: Model dynamics and verification. Meteorology and atmospheric physics 75 (3-4), 161–193.

Xue, M., Droegemeier, K. K., Wong, V., Shapiro, A., Brewster, K., Carr, F., Weber, D., Liu, Y., Wang, D., 2001. The advanced regional prediction system (arps)–a multi-scale nonhydrostatic atmospheric simulation and prediction tool. part ii: Model physics and applications. Meteorology and atmospheric physics 76 (3-4), 143–165.

Yang, D., Sun, F., Liu, Z., Cong, Z., Ni, G., Lei, Z., 2007. Analyzing spatial and temporal variability of annual water-energy balance in nonhumid regions of china using the budyko hypothesis. Water Resources Research 43 (4).

Ye, H., Cohen, J., Rawlins, M., 2013. Discrimination of solid from liquid precipitation over northern eurasia using surface atmospheric conditions*. Journal of Hydrometeorology 14 (4), 1345–1355.

Yoshiyukiishii, Y. K., Nakamura, R., 2004. Water balance of a snowy watershed in hokkaido, japan. Northern Research Basins Water Balance (290), 13.

Zambrano-Bigiarini, M., 2013. hydrogof: Goodness-of-fit functions for comparison of simulated and observed hydrological time series. R package version 0.3-7.

Zhang, L., Potter, N., Hickel, K., Zhang, Y., Shao, Q., 2008. Water balance modeling over variable time scales based on the budyko framework–model development and testing. Journal of Hydrology 360 (1), 117–131.

## Appendix A. List of Symbols, Acronyms, and Notation

## Appendix B. Model performance criteria

The model evaluation statistics used in the paper are the goodness-of-fit (GOF) indices. Theses evaluation statistics are used for cross-validation of rainfall interpolation in kriging and rainfall-runoff simulation performances. The GOF indices used in this paper are defined by the following equations:

Table A.5

| Symbol | Name | Units |
|---|---|---|
| $J$ | Precipitation per unit area | L |
| $J_s$ | Snowfall per unit area | L |
| $J_R$ | Rainfall per unit area | L |
| HRU | Hydrologic response unit | - |
| S (t) | Volume of water stored in HRU per unit area | $LT^{-1}$ |
| k | Index denoting a HRU | - |
| $t$ | Time | T |
| $ET$ | Actual evapotranspiration | $LT^{-1}$ |
| $Q_{ki}$ | incoming discharge to HRU | $L^3T^{-1}$ |
| $Q_k$ | discharge from HRU | $L^3T^{-1}$ |
| $\alpha_r$, $\alpha_s$ | adjusting parameters for rain and snow measurement errors | - |
| $m_1$ | snow-separation parameter controlling the degree of smoothing | - |
| $T_s$ | Threshold temperature | K |
| $T$ | Temperature | K |
| $AI$ | Accuracy index | % |
| $\epsilon_{rank}$ | rank correlation coefficient | - |
| $D$ | difference between simulated and observed data | - |
| $C_{max}$ | maximum water storage capacity of HRU per unit area | $LT^{-1}$ |
| $R_n$ | Total net radiation | $ET^{-1}T^{-2}$ |
| $G$ | soil heat flux | $ET^{-1}T^{-2}$ |
| $\alpha$ | Priestly-Taylor coefficient | |
| $Q$ | Discharge | $L^3T^{-1}$ |
| $\Delta$ | the slope of the Clausius-Clapeyron relation | $KT^{-1}$ |
| $\gamma$ | the psychometric constant | $KT^{-1}$ |
| $\sigma_J$ | standard deviation of precipitation | L |
| $\sigma_Q$ | Standard deviation of discharge | L |
| $\sigma_{\alpha'}$ | standard deviation of due to variable $\alpha'$ | L |
| $\sigma_{ET}$ | Standard deviation of ET | L |
| $R_H(t)$ | Direct runoff | L |
| $R(t)$ | Residual runoff | L |
| $PT$ | Priestly-Taylor ET model | - |
| OK | Ordinary Kriging | - |
| LOK | Local ordinary kriging | - |
| DK | Detrended Kriging | - |
| LDK | Local detrended kringing | |

1. Mean Error (ME) is calculated as

$$ME = \frac{1}{n}\sum_{i=1}^{n} P_i - O_i \tag{B.1}$$

where $P_i$ is the predicted value and $O_i$ is the observed value of the rainfall data at a given time step. The optimal value is 0, and the negative and positive values are underestimation and overestimation respectively.

2. The Root Mean Square Error (RMSE)(Chu and Shirmohammadi, 2004; Singh et al., 2005). The lower the RMSE the better the model performance is. It is given by

$$RMSE = \sum_{i=1}^{n} \sqrt{\frac{1}{n}(P_i - O_i)^2} \tag{B.2}$$

The RMSE is a joint measure of bias in the mean and in the variance, as the square of individual differences between estimated and observed values puts the emphasis on the errors in outliers or higher differences (Ashraf et al., 1997; Nalder and Wein, 1998).

3. Pearson correlation coefficient (r): please refer to Moriasi et al. (2007).

4. PBIAS: is a measure of the average tendency of estimated values to be large or smaller than the corresponding measured values. A value near to zero indicates high estimation, a positive value indicates overestimation and a negative value indicates underestimation (Moriasi et al., 2007; Gupta et al., 1999).

$$PBIAS = \frac{\sum_{i=1}^{n}(P_i - O_i)}{\sum_{i=1}^{n} O_i} * 100 \tag{B.3}$$

This value can be used to assess the systematic under/over estimation of the model.

5. Kling-Gupta efficiency (KGE) has been developed by Gupta et al. (2009) to provide a diagnostically interesting decomposition of the Nash-Sutcliffe efficiency. It facilitates the analysis of the relative importance of the different components (correlation, bias and variability) in the context of hydrological modelling. Kling et al. (2012) has proposed a revised version of this index. It is given by

$$KGE = 1 - ED \tag{B.4}$$

$$ED = \sqrt{(r-1)^2 + (v-1)^2 + (\beta - 1)^2} \tag{B.5}$$

where: ED is the Euclidian distance from the ideal point; $\beta$ is the ratio between the mean simulated and mean observed flows; $r$ is the Pearson product-moment correlation coefficient; and $v$ is the ratio between observed ($\sigma_o$) and modelled ($\sigma_s$) standard deviations of the time series and takes account of the relative variability (Zambrano-Bigiarini, 2013).

## Appendix C. Evapotranspiration Estimation: Procedure details

A gross estimation of the evapotranspiration coefficient $\alpha$ (e.g. Eq. (5) can be obtained directly from data available, under what we call the Budyko Hypothesis. This hypothesis considers that water storage oscillates and, after a number of days/years, which we call Budyko's time, $T_B$, it is back to the same level it was at the initial time. This implies that the water budget is not very far from equilibrium, even in times of climate change. In this case, considering each HRU or the whole basin as a unique control volume (as allowed by data), the water budget can be written, by *not considering any dependence on water storage*, as:

$$S(t) - S(0) = \int_0^{T_B} (J(t) - Q_m(t) - \alpha\, ET(t))dt = 0 \tag{C.1}$$

where $Q_m(t)$ is the measured discharge, we define

$$ET(t) := \frac{\Delta}{\Delta + \gamma} R_n \tag{C.2}$$

and where the dependence on water storage has necessarily been neglected (as it would bring into the process the required knowledge of the parameter $C_{max}$, the maximum allowable storage, a Hymod parameter that is object of calibration).

Therefore, from data only, the maximum we can obtain is:

$$\tilde{\alpha}(T_B) = \frac{\int_0^{T_B}(J(t) - Q_m(t))dt}{\int_0^{T_B} ET(t))\, dt} \tag{C.3}$$

This averaged Priestley-Taylor coefficient is clearly a function of Budyko's time, which we do not know, unless we can perform appropriate groundwater level measurements.

However, through modelling, an estimation of ET *dependent on soil water storage* can be obtained. This estimation process is not that smooth because the estimation of $\alpha$ is intertwined with the process of calibration of ADIGE/Hymod parameters. ADIGE/Hymod, in fact, requires knowledge of the storage $S_g(t)$, which, in turn, depends on how much water is withdrawn by evapotranspiration. In this case, the Budyko's hypothesis says that:

$$S_g(T) - S_g(0) = \int_0^{T_B} (J(t) - R(t) - R_H(t)) - \alpha\, ET(t) dt = 0 \quad \text{(C.4)}$$

where, $R_H(t)$ and $R(t)$ are direct and residual runoff according to Hymod procedure (supplementary material), and

$$ET(t) = \frac{S(t)}{C_{max}} \frac{\Delta}{\Delta + \gamma} R_n \quad \text{(C.5)}$$

and implies that we can estimate an average $\alpha$ as:

$$\bar{\alpha}(T_B) = \frac{\int_0^{T_B} (J(t) - R(t) - R_H(t)) dt}{\int_0^{T_B} ET(t) dt} \quad \text{(C.6)}$$

Therefore, in a set of $n$ years of data, we can generate $n$ values of $\alpha(T_B)$ in the range [1, n] years to setup the model.

In the ADIGE/Hymod calibration phase, knowledge of $\alpha$ is necessary to update $C(t)$ (supplementary material). Therefore, the calibration procedure must simultaneously estimate ET (i.e. $\alpha$) and the runoff parameters. An iterative procedure for this can be:

- for an assigned $T_B$,
- assign a first trial value for $\alpha$, say $\hat{\alpha}^0$, as in eq. (C.3) ;
- assign a tolerance, $\epsilon > 0$ for $\alpha$ estimates.
- Estimate through calibration the Hymod parameters, which are obviously $\alpha^0$ dependent;
- estimate $\hat{\alpha}^1$;
- Repeat the calibration procedure until $|\hat{\alpha}^m - \hat{\alpha}^{m-1}| < \epsilon$, being $m$ the iteration index.

The value of $\alpha$ obtained accounts for variable storage in soil. It has to be remarked that, depending on available data, $\hat{\alpha}$ can be a global or local (to HRU) parameter, while the storage fractions $S(t)/C_{max}$ are always estimated locally for each HRU.