Improving *Bowen-ratio* estimates of evaporation using a rejection criterion and multiple-point statistics

Alessandro Comunian[a], Mauro Giudici[a], Luca Landoni[b], Sergio Pugnaghi[c]

[a] *Università degli Studi di Milano, Dipartimento di Scienze della Terra "A. Desio", via Cicognara 7, 20129 Milano, Italy*
[b] *Università degli Studi di Milano, Dipartimento di Fisica, Via Celoria, 16 - 20133 Milan, Italy*
[c] *ex Università degli Studi di Modena e Reggio Emilia, Dipartimento di Scienze Chimiche e Geologiche, Via G.Campi 103 - 41125 Modena, Italy*

Abstract

The application of the Bowen ratio method to estimate evaporation is heavily affected by uncertainties on the measured quantities. Time series collected with a hydro-meteorological monitoring station often contain measurements for which a reliable estimate of evaporation cannot be computed. Such measurements can be identified with standard error propagation methods. However, simply discarding some values might introduce a bias in the cumulative evaporation for long time intervals, also depending on the threshold of acceptance. In this paper, we propose the use of multiple-point statistics simulation to integrate the time series of reliable evaporation estimates. A test conducted on a two-year-long time series of data collected with a hydro-meteorological station in the Po plain (Italy) shows that the usage of a rejection criteria in conjunction with multiple-point statistics simulation is a promising and useful tool for the reconstruction of reliable evaporation time series. In particular, it is shown that if the rejected values are not replaced by simulation, then the cumulative evaporation curves are estimated with a

bias comparable with estimates of cumulative annual evaporation. Moreover, the test gives some insights for the selection of the best rejection threshold.
Keywords: evaporation, Bowen ratio, multiple-point statistics, time series reconstruction, direct sampling

1. Introduction

Evaporation and transpiration are key factors in the water balance at any temporal and spatial scale and their estimate is of paramount importance in several disciplines, from hydrology to soil science, climatology, etc. (Allen et al., 1998; Eagleson, 2003). Unfortunately, means of direct measurement of evaporation are not available; therefore, the estimate of such a quantity always relies on models of variable complexity or on parameterization.

Among physically-based methods, i.e., those which derive from basic physical principles, the *Bowen ratio* method (BRM, Bowen, 1926) uses quantities which can be measured with an hydro-meteorological monitoring station. However, the uncertainties on the measured quantities can affect the Bowen ratio (B), namely the ratio between sensible and latent heat fluxes, in such a way as to yield an unrealistic value of real evaporation (E).

Very often, sensible and latent heat fluxes are computed from quantities measured at two heights only. Therefore, some authors proposed to improve the estimation of B by increasing the spatial resolution of the measurements required to compute the aforementioned fluxes (Euser et al., 2014).

Another solution to cope with unrealistic E values is data rejection, and the literature contains a number of approaches to handle it. Some authors (Tanner et al., 1987; Ortega-Farias et al., 1996; Cellier and Olioso, 1993) proposed to reject data on the basis of the value of B. Other authors, like for example Ohmura (1982) and Perez et al. (1999), proposed criteria for data rejection based on the analysis of the limits related to the instrument resolution and physical considerations. Many of the aforementioned works were summarized and integrated by Payero et al. (2003). A different approach was proposed by Romano and Giudici (2009), by taking into account the measurement errors and their propagation through the formula to estimate the evaporation with the BRM.

No matter which method is used to select the unreliable samples from a data set, the simplest approach of excluding the physically inconsistent data from the time series of evaporation introduces a bias in the estimate of cumulative evaporation. In fact, this can alter the results for long time periods, e.g., if the BRM is used to perform climatological analyses. It is therefore important to develop an approach to integrate the rejected data samples.

Some authors proposed to integrate the missing or rejected values of B using estimates based on an exchange coefficient, computed from quantities like wind speed or temperature-variance (Savage et al., 2009). Here an alternative approach is proposed, where the rejected values of E are replaced by a stochastic simulation method.

In this paper, the use of multiple-point statistic simulation (MPS) for the replacement of the rejected values of E is proposed and tested. Among the multiple-point simulation paradigms, the Direct Sampling (DS, Mariethoz et al., 2010) method is considered for its flexibility in handling the simulation of continuous variables and the possibility to incorporate secondary information. Moreover, the DS was already tested with success for the reconstruction of incomplete flow rate time-series in a karstic network by Oriani et al. (2016). In the present work, the methodology is tested on a real case study with a two-year-long time series of hydro-meteorological data, whose length can help to evidence particular features, strengths and weaknesses of the method. To our knowledge, this is the first time that this algorithm is tested on the reconstruction of evaporation, and in conjunction with a rejection criteria.

In particular, the following questions are addressed. As the method of Romano and Giudici (2009) requires to define a rejection threshold ε, what is the impact of the selection of ε on the cumulative evaporation? How much the estimates of cumulative evaporation are improved when the rejected values of E are replaced by the simulated ones? Is it possible to improve the DS simulation of the rejected values of E by including one or more measured quantities as covariates?

The field data used to demonstrate the proposed approach, the method used to compute the evaporation E, and the two main steps of the approach (rejection and simulation) are described in Section 2. Section 3 briefly reports the results, which are then discussed in detail in Section 4. The conclusion are reported in Section 5.

2. Materials and Methods

This section illustrates first the field data used to demonstrate the proposed approach. Then, the Bowen ratio method used to compute the evaporation E is briefly recalled. Finally, the two main steps of the proposed approach are described: (1) the criterion used to reject the estimates of E that are not reliable, and (2) the direct sampling method, used to replace the rejected values of E.

[Table 1 about here.]

2.1. Field data

The data set used to test the procedure proposed in this paper were acquired at an hydro-meteorological monitoring station installed in 2006 at Roncopascolo, in the valley of the Taro river, within the Po plain, at about 6 km NW from the city of Parma (Italy).

The position of the station was chosen on the basis of some constraints: the ground has not being subject to human activities for a long time, the area is far from buildings or other obstructing bodies, the installed instruments are protected against thieve or damages.

The meteorological sensors of the monitoring station are installed on a five-meter-tall pole. Two couples of humidity and temperature sensors are installed at 2 m (h_1) and 4 m (h_2) from the ground surface, a pressure sensor is installed in the box containing the data logger, an anemometer is installed at the top of the pole, i.e. 5 m above the ground, and a sensor of net radiation is installed at an height of about 2 m. Moreover, a rain gauge is installed at a distance of 2 m from the principal pole at an height of 1.5 m from the ground and a sensor to measure the heat flux is immersed in the soil, at a depth of few centimeters, a couple of meters far from the pole. All the data have been collected with a sampling interval of 20 minutes.

The data used in this work correspond to the period from June 2009 to July 2011, for which a rather complete series of data is available. Later, it was not possible to perform a regular maintenance of the monitoring station, which has newly been working since spring 2016.

A preliminary accurate analysis of the recorded data already shows some anomalous measurements. In winter, negative values of net radiation, with high absolute values, have been measured and interpreted as an effect of intense snow, as supported from meteorological bulletins of the surrounding area. In those cases snow could cover the upper part of the net radiation sensor and filter out the direct solar radiation, whereas the high albedo of snow on the ground could enhance the reflected radiation, thus producing values as low as $-150\,\mathrm{W/m^2}$. In summer and spring, some spikes appear in the time series of different quantities, but they seem to be due to interference of lighting with the instrumentation, again as confirmed by the inspection of meteorological bulletins of the surrounding area. These evident anomalous measurements were removed from the time series prior to the application of the proposed work-flow.

2.2. Bowen ratio method

Using the data collected at the hydro-meteorological station, evaporation E can be computed using the BRM. The latter is based on the computation of the Bowen ratio, i.e., the ratio between sensible and latent heat flux, which is estimated from measurements of temperature and vapor partial pressure at two different heights as

$$B = \frac{C_a P_a}{0.622\lambda_v} \frac{T_2 - T_1}{e_2 - e_1} \tag{1}$$

where C_a is the specific heat of air at constant pressure per unit mass, P_a is the atmospheric pressure, λ_v is the latent heat of evaporation per unit mass,

T_i and e_i, with $i = 1, 2$, are, respectively, air temperature and vapor partial pressure at two different heights h_i above the ground surface. Given the air temperatures, the vapor partial pressures can be converted, using some empirical relation (Dingman, 2015), into the corresponding relative humidity RH_1 and RH_2.

The energy balance at the soil, by neglecting the advective contribution and energy storage, yields the following expression for the evaporation E, i.e., the volume of liquid water evaporating from the surface per unit time and unit surface:

$$E = \frac{R_n - G}{\rho_w \lambda_v (1 + B)} \tag{2}$$

where R_n is the net radiation, G is the geothermal heat flow and ρ_w is the water density.

2.3. Rejection of unreliable estimates of evaporation

Equations (1) and (2) show that (i) B can be computed only if $e_1 \neq e_2$ and (ii) E can be computed with (2) only if $B \neq -1$. These conditions are not always met when dealing with field monitoring data. Moreover, even if they are satisfied, the propagation of measurement errors could yield unrealistic values of E. For example, if $e_2 - e_1 \to 0$, i.e. the vapor partial pressure is constant along the vertical, and $B \neq 1$, then $E \to 0$, namely the evaporation is negligible. Instead, if $e_2 \neq e_1$, then $B \to -1$ implies $E \to \pm\infty$. In other words, when B is close to -1, the estimated evaporation rate would achieve unrealistic values.

The criterion used to reject data is taken from Romano and Giudici (2009) and operates according to the following procedure. For every physical quantity appearing in equations (1) and (2) an estimate of its uncertainty is given, based on the accuracy of the measurement instrument. From these estimates and the law of error propagation (Bevington and Robinson, 2003), the uncertainty on the estimate of the evaporation rate, δ_E, is computed. If $\delta_E/E > \varepsilon$, where ε is a prescribed threshold, then the value of E is considered to be unreliable and it is discarded. Romano and Giudici (2009) tested values of $\varepsilon \in [0.1, 50]$ for a data set collected in the suburbs of the city of Milan and suggest the value of 5 acceptable for relative errors of the cumulative evaporation lower than 20%. In the following sections we report and discuss the results obtained with $\varepsilon \in \{0.5, 1, 5, 10\}$.

2.4. Reconstruction of rejected estimates of evaporation with MPS

Starting from a time series of meteorological data, the procedure described in Section 2.2 and Section 2.3 can be applied. The values of evaporation that were rejected because considered not reliable according to the criterion described in Section 2.3 should be replaced by reliable estimates. If these values are not replaced, then the cumulative evaporation assessed for a long time period could be strongly underestimated.

This problem can be limited with a proper simulation of the missing values in the series of evaporation. In this paper, this is obtained with the application of MPS. In particular, a direct sampling (DS) algorithm (Mariethoz et al., 2010) is used. Our approach is similar to that applied by Oriani et al. (2014) to model rainfall time series and the flow rate of two karstic springs in the Jura Mountains, Swiss Alps (Oriani et al., 2016).

For our application, a training image (TI) is given by the time series of the acceptable values of E (E_{TI}). The simulation grid (SG) is the array which contains the whole time series of E, including both the acceptable values estimated from (2) and the values simulated with the DS algorithm to replace the rejected ones. Hereinafter we briefly outline the working principle of the DS, as applied to our case study.

1. Let $\boldsymbol{t} = \{t_1, t_2, \ldots, t_n\}$ be the array of the times for which the SG has to be built, let $\boldsymbol{\tau} = \{\tau_1, \tau_2, \ldots, \tau_m\}$ be the array of the times for which acceptable values of E were found and let $\widetilde{E}$ be the evaporation rate, normalized with a linear scaling in such a way that it is comprised between -1 and 1. This step is required to homogenize the distance computations and the comparisons between variables and covariates.
2. Randomly select an empty cell of the SG, i.e. a time t_i. The set $\boldsymbol{\tau}_i$ of times $\tau_j \in \boldsymbol{\tau}$, such that $|t_i - \tau_j| < R$, where R is a prescribed search radius, and such that the cardinality of τ_i is smaller than a prescribed maximum number N, is used to define a data event, i.e., a set of couples of time lags and corresponding values of $\widetilde{E}$ such that

$$\boldsymbol{d}_i = \{(t_i, \widetilde{E}(\tau_j)) \, \text{with} \, |\tau_j - t_i| < R, \tau_j \neq t_i, \text{card}\,(\tau_j) \leq N\}. \quad (3)$$

   The number of values τ_j is limited by the user provided parameter N, that is the maximum number of nodes in the search neighborhood. This parameter allows to dynamically define the radius R by considering only the N values of τ_j closest to t_i.
3. The TI is scanned until a data event, for time $t_k \in \boldsymbol{\tau}$, similar to $\boldsymbol{d}_i$ is found, i.e., when $|\boldsymbol{d}_i - \boldsymbol{d}_k| < \sigma$, where σ is a prescribed threshold of acceptance.
4. The time t_i is added to $\boldsymbol{\tau}$ and $\widetilde{E}(t_i) = \widetilde{E}(t_k)$. The procedure continues from point 2 above, until the whole SG is filled with simulated values (E_{sim}).

If, for the time series, measurements of other variables supposedly correlated with the simulated variable are available, then the approach can be extended in a straightforward manner to include them in a co-simulation framework, where the training image becomes a multi-variate training image and different thresholds and search radius can be defined, one for each variable. For more details refer to Mariethoz et al. (2010) and to Oriani et al. (2014, 2016).

When using the DS simulation technique, the choice of several simulation parameters can have an important impact on the final results. In this work, a number of preliminary tests were performed to select the suitable simulation parameters, also following the guidelines presented by Meerschman et al. (2013) and the parameterization adopted by Oriani et al. (2014, 2016). A good balance between CPU requirements (that anyhow remained below the order of few seconds per realization) and quality of the simulation were obtained with a search radius R of 28 days, a threshold $\sigma = 0.001$, and $N = 20$. To smooth the simulated values of E, the average over 10 equiprobable realization is considered. All the direct sampling simulations were performed with the `deesse` simulation code (Mariethoz et al., 2010; Straubhaar, 2017).

An important remark has to be made here: in a standard MPS simulation setting, TI and SG are separated entities which can have, in general, a different size and represent different time (or space) windows. In the simulation setting presented here, TI and SG share the same grid and the same time window. In fact, the TI is incomplete (rejected values of E) and the simulation procedure aims at inserting the missing E values.

2.5. Validation

A validation step was performed to support the results obtained by the reconstruction. In practice, a given percentage of the E values considered reliable according to the adopted rejection criterion is randomly selected and excluded both from the training and the conditioning data set, but is used for cross-validation (E_{val}). The validation is performed for different values of the rejection threshold ε in terms of Q-Q plots, and also in terms of a coefficient inspired by the Nash-Sutcliffe efficiency (NSE, Nash and Sutcliffe, 1970)

$$NSE = 1 - \frac{\sum_{t_i \in \boldsymbol{\tau}_{\text{val}}} (E_{\text{val}}(t_i) - E_{\text{sim}}(t_i))^2}{\sum_{t_i \in \boldsymbol{\tau}_{\text{val}}} (E_{\text{val}}(t_i) - \overline{E}_{\text{val}}(t_i))^2} \tag{4}$$

Here $\boldsymbol{\tau}_{\text{val}}$ contains the time steps t_i for which a validation value of E is selected, and $\overline{E}_{\text{val}}$ represents the average of these values over a given time window. In brief, $NSE \simeq 1$ indicates that simulation has better performances if compared to simple approaches where the average value of E is considered; $NSE \simeq 0$ indicates that simulation and simple approaches are equivalent; $NSE < 0$ indicates that simple approaches outperform simulation.

3. Results

First of all the deleterious effect of B is analyzed. By looking at Figure 1 it is clear that for values of $B \simeq -1$ $|E|$ reaches completely unreliable values.

[Figure 1 about here.]

Then, in the two following section, the results obtained by investigating the impact of a different rejection threshold ε and the usage of one or more covariates in the DS simulation are briefly illustrated.

To illustrate the impact of the rejection threshold, the criterion proposed by Romano and Giudici (2009) was applied using four different thresholds to the evaporation computed with the Bowen-ratio method. The values of E that were not rejected are used both as training image and as conditioning data in the DS simulation. For each value of ε, 10 DS realizations are performed and the rejected values of E are replaced by the arithmetic average of the 10 realizations. The results are compared in terms of Q-Q plots, visual inspection of time series, cumulative E curves, and also using diverse statistical indicators to average the realizations obtained for each ε. To further support the results, a validation test is performed by randomly selecting the 25% of the non-rejected E. The validation values are then compared with values simulated for the same time steps using the 75% of the non-rejected data as training and conditioning data.

Then, an intermediate value of $\varepsilon = 5$ was selected to illustrate the effect of adding a covariate in the simulation process. Seven different covariates were selected and used in the DS to simulate the rejected E values, including T_1, RH_1, P_a, precipitation, G, R_n, and v.

3.1. The impact of the rejection threshold ε

The basic problem to be solved when applying the method by Romano and Giudici (2009) is the choice of the threshold ε for the rejection criterion. Here we investigate the effects that four different thresholds have on the reconstructed time series of E and the corresponding cumulative time series. The main impact is evident on the number of rejected values of E, which are reported in Table 2. The percentage of rejected values is also listed on a seasonal basis to illustrate its variability, for each of the investigated years and for the complete time series.

[Table 2 about here.]

Another tool useful to compare the results obtained by changing the rejection threshold ε is the Q-Q plot. In Figure 2 are reported, for different values of ε, the quantiles of the time series completed with the simulated data on abscissa, and the quantiles of the time series containing only the reliable (non-rejected) values of E on ordinate. The orange line represents the case when the quantiles computed for the simulation coincide with the training data.

[Figure 2 about here.]

Q-Q are not sufficient to discern if the missing values were correctly replaced by the simulated ones. A visual inspection of the time series can reveal some details which are not put in evidence by the Q-Q plot. In Figure 3, for example, we compare the E time series obtained with $\varepsilon = 1$ and $\varepsilon = 10$ for a time window with a high density of simulated data (second half of January 2011)

[Figure 3 about here.]

It is also important to check the impact of applying different rejection thresholds ε on the cumulative E curves (Figure 4).

[Figure 4 about here.]

One of the main goals of this research was to estimate the impact of neglecting the contribution of the rejected values of E on the cumulative curves. This aspect is illustrated in Figure 5 for $\varepsilon = 1$, a parameter that provides a good balance between the number of rejected values of E and the reliability of the time series. In Figure 5 the cumulative E curves obtained by replacing the rejected data with simulated values of E (continuous line) are compared against the curves obtained without replacing the rejected values of E (dashed lines). Comparable results were obtained with the other values of ε, which are not shown here for the sake of brevity. Nevertheless, the differences between the cumulated E are reported for each year and for each value of ε in Table 3.

[Figure 5 about here.]

[Table 3 about here.]

As mentioned in Sect. 2.4, the simulated values of E are presented as the arithmetic mean over 10 DS realizations. Fig. 6 illustrates how the cumulative E curves behave when different statistical indicators are used to aggregate the 10 realizations. In particular, Fig. 6 reports, for year 2010 and for the different values of the threshold ε, the cumulative E curves computed using the arithmetic mean (continuous blue line), the 1-st quartile (Q_1, dashed orange line), the median (Q_2, dash-dotted green line), and the 3-rd quartile (Q_3, dashed red line).

[Figure 6 about here.]

Figure 7 illustrates the results of the validation step, where for the different values of ε considered in this work the 25% of the reliable E is randomly selected and kept for validation purposes, and compared with the values simulated for the same time step. The NSE for the corresponding value of ε is reported in the lower right corner of each sub-plot.

[Figure 7 about here.]

3.2. Simulating the rejected values of E using a covariate

Another aspect explored by this research is the influence of considering a covariate in the simulation of the rejected values of E. Here the considered covariates are a number of quantities measured at the hydro-meteorological station of Roncopascolo including T_1, RH_1, P_a, precipitation, G, R_n, and the wind speed v. All the parameters already used in Equations (1) and (2) for the computation of E are correlated with E itself. Nevertheless, for the time step where E is rejected, the aforementioned parameters have reliable values and therefore they can potentially improve the simulation of E. Here the impact of considering one covariate in the DS simulations is illustrated via Q-Q plots in Figure 9. Figure 9a illustrates the reference case when no covariates are used, while the remainders sub-plots (Figure 9b-h) represent the results obtained considering one of the aforementioned covariates. Also, the same results are presented in terms of cumulative E (Figure 8).

Only the results for the rejection threshold $\varepsilon = 5$ are shown here, because they provide a situation where many E values are rejected and there is room for improving the estimates of the missing E values obtained without the use of a covariate.

[Figure 8 about here.]

[Figure 9 about here.]

4. Discussion

As anticipated in Section 2.3 and as expected from Equation (2), the results show that when B is close to -1 the computed values of $|E|$ become more and more high and unreliable (Figure 1). From Figure 1 it is evident that those E values can have a deleterious effect when considered in cumulative E curves. It becomes therefore crucial to reject unreliable values of E.

With the rejection thresholds ε tested in this work, for the same time series the percentage of rejected values varies from 6.1 % to 70.6 % (Table 2). The percentages of rejected data regrouped by season (Table 2) suggest that spring is the season where most data are incorrectly determined, and this is thought to be related to the fact that this is the season with the highest atmospheric instability.

Q-Q plots and visual inspection of the time series were used to evaluate the reconstructed time series (Figure 2). From the Q-Q plots, $\varepsilon = 1$ appears to provide the best results. At the same time, a restrictive value of ε (i.e. $\varepsilon = 1$ or $\varepsilon = 0.5$) reduces considerably the number of data and the number of extreme events in the incomplete data series that are used as training data in the DS. This has a clear impact on the reconstructed time series variability (Figure 2). A visual inspection of the time series integrates the results of the Q-Q plots, showing that a quite restrictive rejection threshold ($\varepsilon = 1$) provides a reliable temporal variability of E (Figure 3a) and filters out some spikes that instead appear for less restrictive ε ($\varepsilon = 10$, Figure 3b). Note that in Figure 3 we deliberately selected a time window where many values of E were rejected to illustrate the DS simulation capabilities.

Figure 4 illustrates the impact of different rejection thresholds ε on the cumulated E, when the rejected values of E are replaced with simulated values, for years 2009, 2010 and 2011. The features of the cumulated E curves are useful for the selection of the optimal ε. For example, for years 2010 and 2011 (Figure 4b and Figure 4c), only with the values $\varepsilon = 0.5$ or $\varepsilon = 1$ the anomalous features of the cumulative curves around March 2010 and March-May 2011 are filtered out. Also, the difference between the growth rates of the curves for $\varepsilon = 0.5$ and $\varepsilon = 1$ show that the two rejection thresholds have a different impact depending on the sign of the rejected E values. As a consequence, the rejection procedure has a different impact depending on the season and on the prevailing physical process. However, if we exclude year 2011, the final plateau reached by using diverse values of ε has an impact of few tenths of millimeters on the yearly cumulated E. The values of the cumulated precipitation reported on the top of each figure (Fig. 4) are also useful to check the reliability of the cumulative E curves computed with different ε. Note also that while for year 2010 a more complete data set is available, years 2009 and 2011 are incomplete for a rather different time period. This justified the noticeable differences between the cumulative E curves (Fig. 4).

Probably, the most important result is illustrated in Figure 5. When the rejected values of E are not replaced by simulation, the cumulative E obtained from instantaneous values which were not rejected is strongly underestimated (dashed lines, Figure 5). The cumulative E could be somehow corrected by considering the percentage of rejected values. However, here it is possible to provide a more precise estimate: the yearly cumulative E generally is underestimated by an amount that has its same order of magnitude.

For example, in our case, for years 2009 and 2010 the yearly cumulative E is underestimated by more that 100 mm (Figure 5a and Figure 5b), while (for the available period) of the 2011 by more than 300 mm (Figure 5c). The numerical values of the differences are reported in Table 3. Here it is important to remark that for years 2009 and 2010 the differences decrease with the increase of ε, whereas for year 2011 the trend is quite peculiar, with a peak of difference for the value of $\varepsilon = 1$.

Figure 6 illustrates the impact of the statistical indicator used to aggregate the results of the simulation over many realizations on the cumulative E. With a relative low rejection threshold ($\varepsilon = 0.5$, Fig. 6a), many values are rejected, many are simulated and few conditioning data are kept; this has a clear effect on the spreading of the cumulated E, and implies that arithmetic mean and median (Q_2) at the end of the year accumulate more than 50 mm of difference. Nevertheless, it is sufficient to rise the rejection threshold above 1 to reduce the cumulated difference between mean and median to few millimeters per year (Fig. 6b, c, and d). In addition, the curves reported for Q_1 and Q_3 show not only the uncertainty on the simulated cumulative E, but also the effectiveness of the arithmetic mean in smoothing extreme values for high values of ϵ, when the number of rejected and simulated values of E is small.

To further support the results, one depletion test for each rejection threshold was performed (Fig. 7). When many values of E are rejected, the statistics of the simulated values are coherent with those of the validation data (Fig. 7a and b). Nevertheless, the slight deterioration of the statistics when many data are rejected (Fig. 7a) suggests that a too high fraction of rejected data entails a pauperization of the training data. Differently, and in particular when many unreliable values are kept, the statistics of the simulated values (Fig. 7d) depart from the validation data, for example for $E < 0$ mm/h and $E >$0.6 mm/h. The NSE indices reported in Fig. 7 illustrate the efficiency of the proposed work-flow against a naive approach where the missing values are reconstructed using the weekly averaged values of E. Also, its variability against ε provides a useful guide for the selection of this rejection threshold.

Another interesting aspect investigated here is the influence of a covariate in the simulation of the rejected E values. The Q-Q plots (Figure 9) already reveal that the impact is much less evident than changes in the value of the rejection threshold ε (Figure 2). Nevertheless, some variables provide a better representation of the quantiles. This is for example the case of relative humidity (RH_1, Figure 9c), atmospheric pressure (P_a, Figure 9d), and wind speed (v, Figure 9h), where the scattered quantiles (blue dots) are closer to the ideal case (orange line) than the results obtained without the use of any covariate in the DS simulation (Figure 9a). For the covariates that provide a better representation of the data in terms of Q-Q plots, the visual inspection of the E time series (not shown here) reveals a smoothed and spike-free trend if compared to the E time series simulated without a covariate. Clearly, taking into account a covariate in the simulation of the rejected E has an impact on the cumulative E curves (Figure 8). For the considered years, the difference between the reference black curve with markers (simulation without covariate) and the colored curves (simulation with one covariate), has a maximum of about 30 mm. One interesting aspect is that the sign of this impact depends, for the time period investigated, not only on the considered covariate but also on the time window considered. For year 2009, for example, all the cumulative curves obtained using a covariate are above the one obtained with no covariate (Figure 8a), while for years 2010 and 2011 diverse covariates have a diverse impact on the cumulative curves (Figure 8b and Figure 8c).

5. Conclusions

In this technical note, a straightforward work-flow to improve the reliability of cumulative time series of evaporation E is presented. The work-flow is made of two main steps: firstly, the values of E that are deemed unreliable according to a threshold defined by error propagation techniques are rejected; then, the rejected values are replaced by multiple-point statistics simulation using a direct sampling algorithm. The applicability of the work-flow is demonstrated on a data-set collected by a hydro-meteorological station located in the Po plain (Italy), from May 2009 to July 2011. This data-set allows to test the work-flow on values of E estimated with the Bowen ratio method. However, the proposed work-flow has a general validity and can be applied in different contexts, like for example where E is estimated from eddy covariance measurements.

It is shown that the proposed work-flow can be used to integrate incomplete time series in a straightforward way. In particular, when applied to the reconstruction of evaporation time series, the results demonstrate that if the rejected values of E are not replaced by simulated values, then the cumulative E can be underestimated by quantities comparable to its total per annum. Focusing on the data set considered in this study, the annual underestimation of E can easily exceed 100 mm/year.

Unfortunately, direct measurements of E are not available for the same time period and region. However, although a direct comparison with reference values cannot be performed, the cumulative E time series obtained integrating the rejected values with simulated values allows to quantify the approximation made when the missing values are not properly replaced.

In addition, this study provides useful insights for the selection of ε, the threshold used to reject unreliable values of E based on the error propagation theory. Here, simple tools like Q-Q plots and visual inspection of time series allowed to select the value of ε that provided a good compromise between number of rejected samples and a reliable reconstruction of the E time series.

Another aspect investigated in this research is the potential improvement of the simulation results provided by the usage of covariates. This study shows that including a covariate in the simulation process has an impact on the final results, which of course depends on the covariate considered, but also on the part of the considered year. Further research is required to investigate the effects of taking into account, in the simulation process, of the combination of two or more covariates. Besides different parameterizations of the direct sampling algorithm, other covariates derived from the variables measured at the hydro-meteorological station can be considered, like for example a moving average of the temperature, that could provide a seasonal trend useful to improve the simulation.

6. Acknowledgments

The authors kindly acknowledge M.Adorni and V.Piramide from IRETI S.p.A. for hosting the hydro-meteorological measurement station at the Roncopascolo site, F.Oriani, P.Renard, J.Straubhaar for the fruitful discussions, R.Poli, A.Tartaglia and S.Zoia for their help, the two anonymous reviewers for their constructive comments, and the University of Neuchâtel for providing the `deesse` simulation code.

References

Allen R, Pereira L, Raes D, Smith M (1998) Crop evapotranspiration - guidelines for computing crop water requirements. FAO Irrigation and drainage paper 56

Bevington P, Robinson D (2003) Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill, Inc.

Bowen IS (1926) The ratio of heat losses by conduction and by evaporation from any water surface. Phys Rev 27:779–787, DOI 10.1103/PhysRev.27.779

Cellier P, Olioso A (1993) A simple system for automated long-term bowen ratio measurement. Agricultural and Forest Meteorology 66(1):81–92, DOI 10.1016/0168-1923(93)90083-T

Dingman S (2015) Physical Hydrology. Waveland Press

Eagleson PS (2003) Dynamic Hydrology. EGU

Euser T, Luxemburg WMJ, Everson CS, Mengistu MG, Clulow AD, Bastiaanssen WGM (2014) A new method to measure bowen ratios using high-resolution vertical dry and wet bulb temperature profiles. Hydrology and Earth System Sciences 18(6):2021–2032, DOI 10.5194/hess-18-2021-2014

Mariethoz G, Renard P, Straubhaar J (2010) The direct sampling method to perform multiple-point geostatistical simulations. Water Resour Res 46(11):W11,536, DOI 10.1029/2008WR007621

Meerschman E, Pirot G, Mariethoz G, Straubhaar J, Meirvenne MV, Renard P (2013) A practical guide to performing multiple-point statistical simulations with the direct sampling algorithm. Computers & Geosciences 52(0):307 – 324, DOI 10.1016/j.cageo.2012.09.019

Nash J, Sutcliffe J (1970) River flow forecasting through conceptual models part i - a discussion of principles. Journal of Hydrology 10(3):282 – 290, DOI https://doi.org/10.1016/0022-1694(70)90255-6, URL http://www.sciencedirect.com/science/article/pii/0022169470902556

Ohmura A (1982) Objective criteria for rejecting data for bowen ratio flux calculations. J Appl Meteor 21(4):595–598, DOI 10.1175/1520-0450(1982)021<0595:OCFRDF>2.0.CO;2

Oriani F, Straubhaar J, Renard P, Mariethoz G (2014) Simulation of rainfall time-series from different climatic regions using the direct sampling technique. Hydrology and Earth System Sciences Discussions 11(3):3213–3247, DOI 10.5194/hessd-11-3213-2014

Oriani F, Borghi A, Straubhaar J, Mariethoz G, Renard P (2016) Missing data simulation inside flow rate time-series using multiple-point statistics. Environmental Modelling & Software 86:264 – 276, DOI 10.1016/j.envsoft.2016.10.002

Ortega-Farias SO, Cuenca RH, Ek M (1996) Daytime variation of sensible heat flux estimated by the bulk aerodynamic method over a grass canopy. Agricultural and Forest Meteorology 81(1):131–143, DOI 10.1016/0168-1923(95)02278-3

Payero O, Neale C, Wright L, Allen G (2003) Guidelines for validating Bowen ratio data. Transactions of the ASAE 46(4):1051–1060

Perez P, Castellvi F, Ibañez M, Rosell J (1999) Assessment of reliability of bowen ratio method for partitioning fluxes. Agricultural and Forest Meteorology 97(3):141 – 150, DOI 10.1016/S0168-1923(99)00080-5

Romano E, Giudici M (2009) On the use of meteorological data to assess the evaporation from a bare soil. Journal of Hydrology 372(1):30–40, DOI 10.1016/j.jhydrol.2009.04.003

Savage M, Everson C, Metelerkamp B (2009) Bowen ratio evaporation measurement in a remote montane grassland: Data integrity and fluxes. Journal of Hydrology 376(1):249 – 260, DOI 10.1016/j.jhydrol.2009.07.038

Straubhaar J (2017) DeesSe User's Guide. University of Neuchâtel

Tanner B, Greene J, Bingham G (1987) A bowen-ratio design for long term measurements. In: ASAE (ed) Proc. ASAE 1987 International Winter Meeting. St. Joseph, Mich, 87-2503

List of Figures

1 Scatter plot of evaporation E vs Bowen ratio B 26

2 Q-Q plots of simulated E vs training E for different values of rejection threshold ε. For each value of ε it is also reported the percentage of rejected values (in brackets) 27

3 Comparison of two E time series obtained using two different values of rejection threshold ε for a heavily simulated time period (second half of January 2011) 28

4 Cumulative E computed with the BRM for years a) 2009, b) 2010 and c) 2011. The different colors correspond to a different value of the rejection threshold ε. Here the rejected values are replaced by DS simulation 29

5 Cumulative E computed with the BRM for years a) 2009, b) 2010 and c) 2011. The continuous lines correspond to time series where the rejected values are replaced by DS simulation (data and sim.), while the dashed lines corresponds to time series where the rejected values are not replaced (data only) . 30

6 Comparison of the cumulative E computed with the BRM for year 2010 obtained using different statistical indicators (mean and $Q_{1,2,3}$) for different values of the rejection threshold: a) $\varepsilon = 0.5$, b) $\varepsilon = 1$, c) $\varepsilon = 5$, d) $\varepsilon = 10$ 31

7 Q-Q plots of simulated E vs validation E for different values of rejection threshold ε. For each value of ε it is also reported the percentage of rejected values (in brackets) and on the lower right corner the NSE 32

8 Q-Q plots of simulated E vs training E using different covariates. On the lower right corner Pearson's correlation coefficients 33

9 Cumulative E computed with the BRM for years a) 2009, b) 2010 and c) 2011. Unreliable values of E are rejected using a threshold $\varepsilon = 5$. The different colors represent a different covariate used in the DS simulation of the rejected values of E; the black curve with markers represents the time series simulated taking into account for the variable E only 34

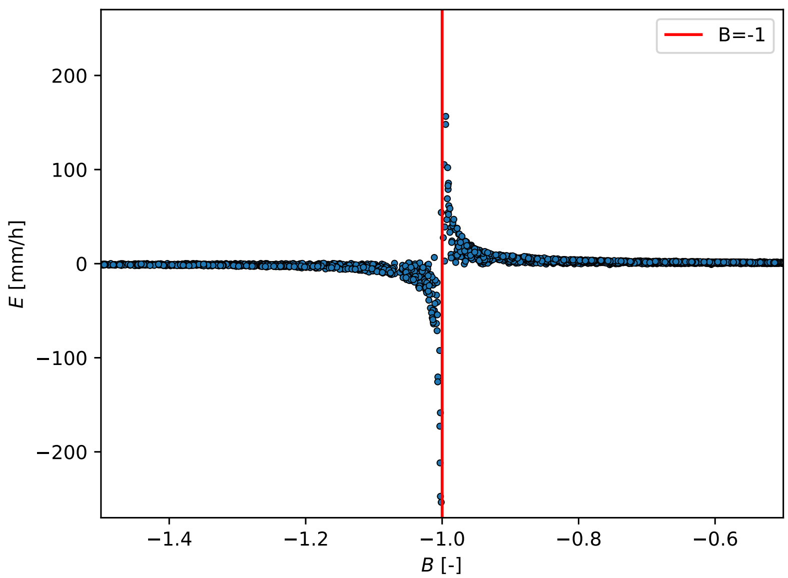

Figure 1: Scatter plot of evaporation E vs Bowen ratio B

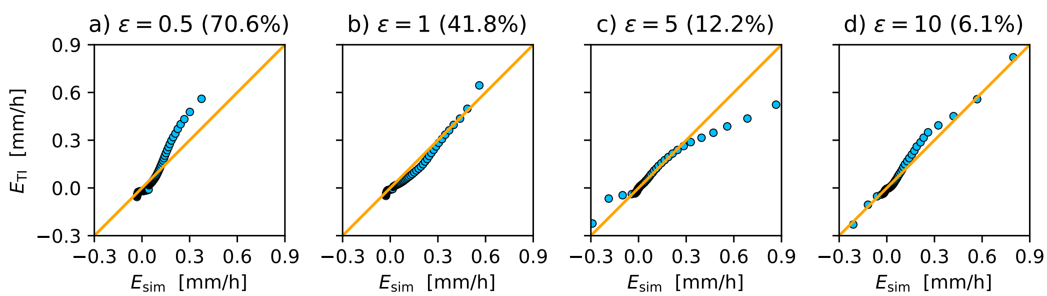

Figure 2: Q-Q plots of simulated E vs training E for different values of rejection threshold ε. For each value of ε it is also reported the percentage of rejected values (in brackets)

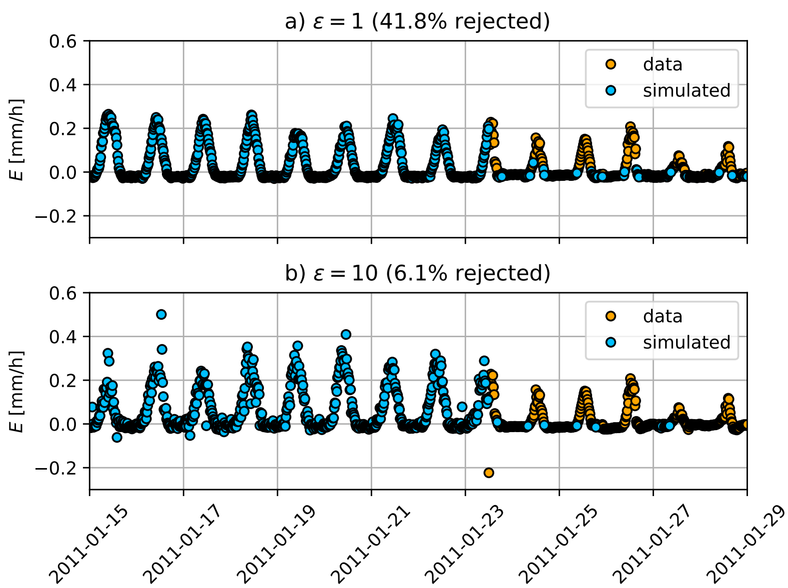

Figure 3: Comparison of two E time series obtained using two different values of rejection threshold ε for a heavily simulated time period (second half of January 2011)

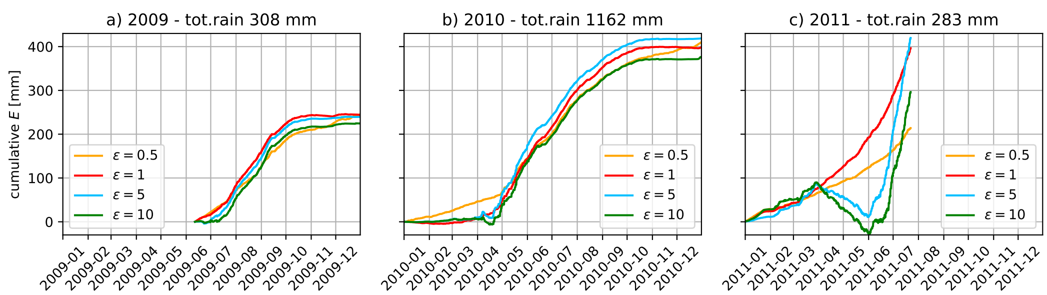

Figure 4: Cumulative E computed with the BRM for years a) 2009, b) 2010 and c) 2011. The different colors correspond to a different value of the rejection threshold ε. Here the rejected values are replaced by DS simulation

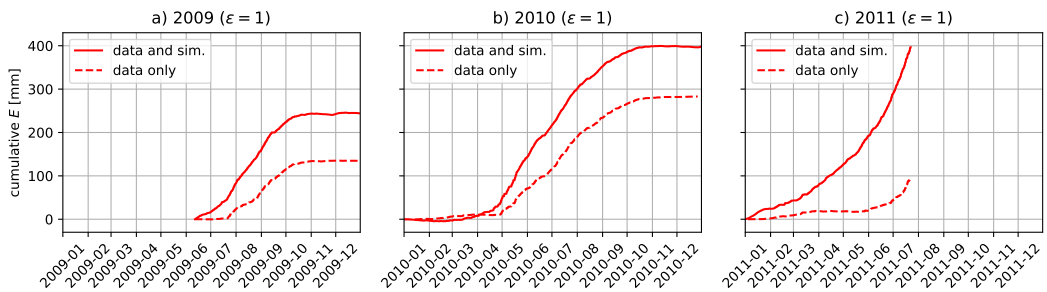

Figure 5: Cumulative E computed with the BRM for years a) 2009, b) 2010 and c) 2011. The continuous lines correspond to time series where the rejected values are replaced by DS simulation (data and sim.), while the dashed lines corresponds to time series where the rejected values are not replaced (data only)

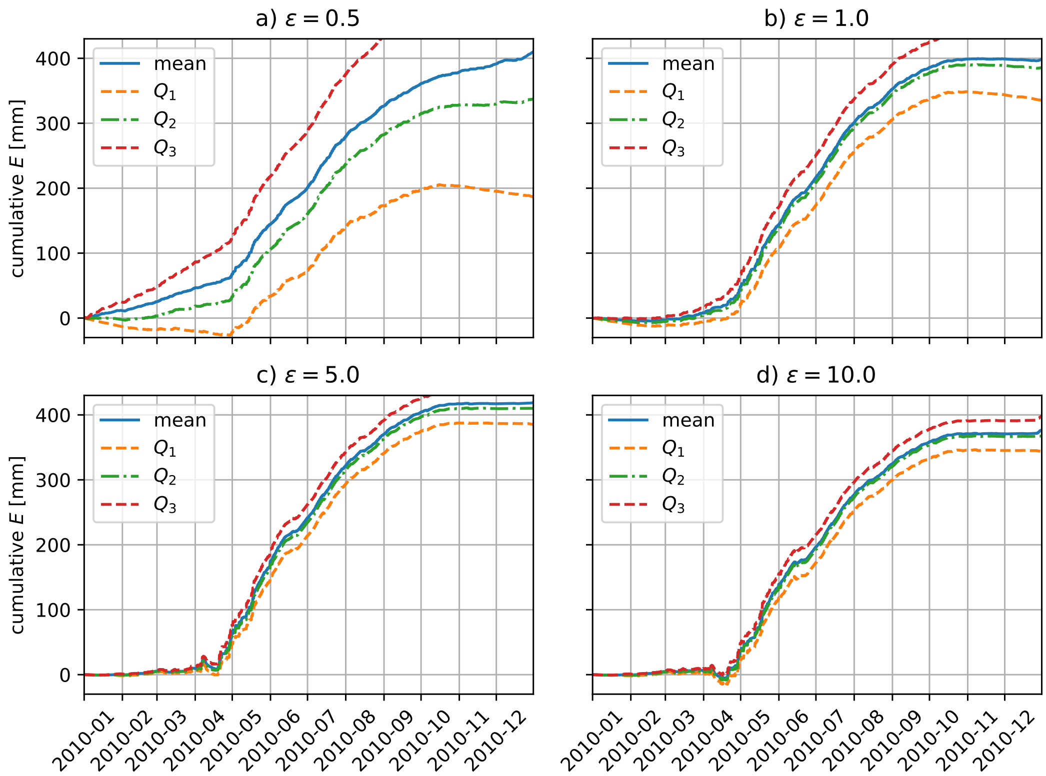

Figure 6: Comparison of the cumulative E computed with the BRM for year 2010 obtained using different statistical indicators (mean and $Q_{1,2,3}$) for different values of the rejection threshold: a) $\varepsilon = 0.5$, b) $\varepsilon = 1$, c) $\varepsilon = 5$, d) $\varepsilon = 10$

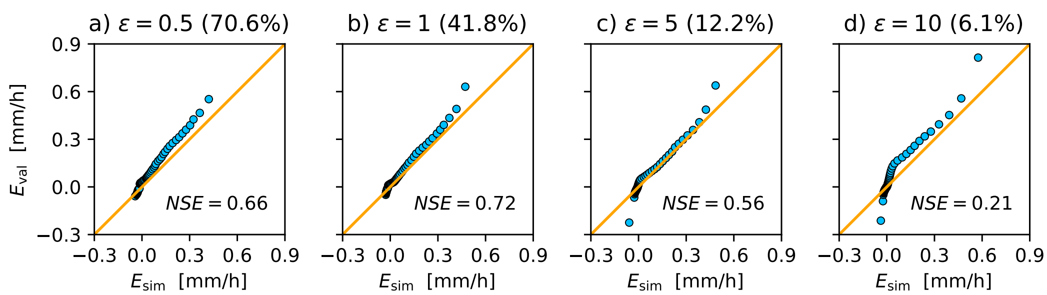

Figure 7: Q-Q plots of simulated E vs validation E for different values of rejection threshold ε. For each value of ε it is also reported the percentage of rejected values (in brackets) and on the lower right corner the NSE

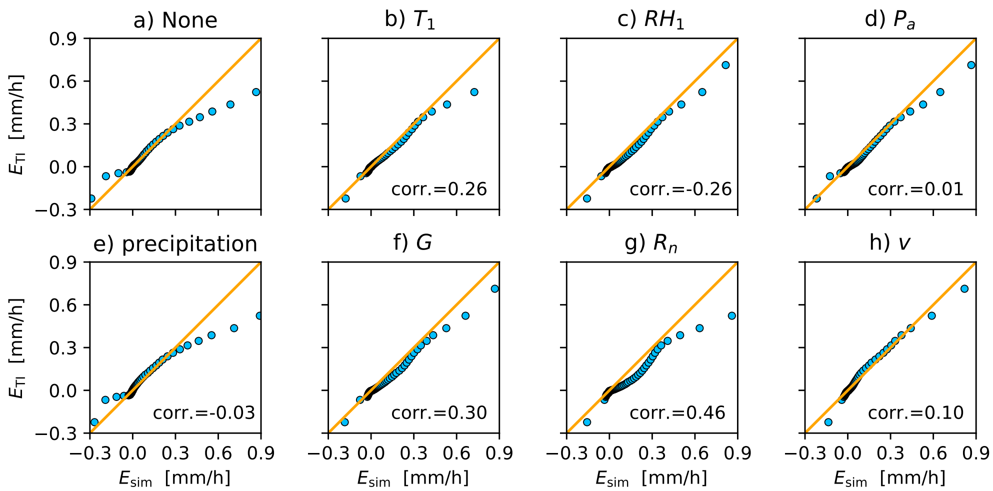

Figure 8: Q-Q plots of simulated E vs training E using different covariates. On the lower right corner Pearson's correlation coefficients

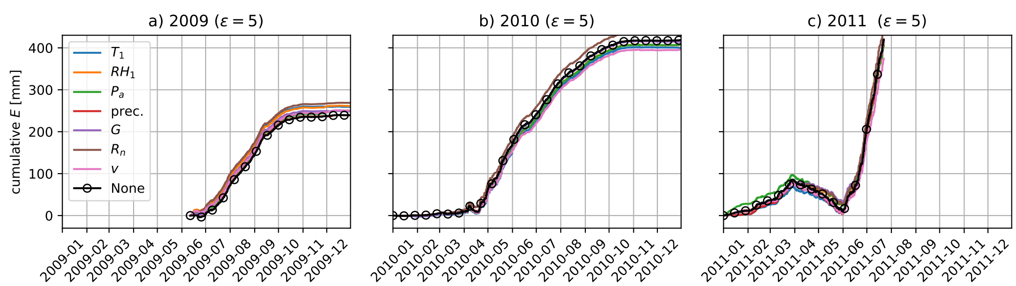

Figure 9: Cumulative E computed with the BRM for years a) 2009, b) 2010 and c) 2011. Unreliable values of E are rejected using a threshold $\varepsilon = 5$. The different colors represent a different covariate used in the DS simulation of the rejected values of E; the black curve with markers represents the time series simulated taking into account for the variable E only

List of Tables

1 List of symbols and variables. $i \in \{1, 2\}$. For symbols related to the DS please refer to the text 36

2 Percentage of data rejected for each season of the time series . 37

3 Differences between the cumulated E computed by replacing the rejected data with the DS simulation and without replacing the rejected data. Units are in mm. 38

Table 1: List of symbols and variables. $i \in \{1, 2\}$. For symbols related to the DS please refer to the text

symbol	units	description
B	$-$	Bowen ratio (computed)
E	mm/h	evaporation rate (computed/simulated)
ε	$-$	rejection threshold (user defined)
h_i	m	height above the ground of the sensor
T_i	K	temperature at h_i (measured)
RH_i	%	relative humidity at h_i (measured)
P_a	Pa	atm. pressure (measured)
v	m/s	wind speed (measured)
R_n	W/m^2	net radiation (measured)
G	W/m^2	soil heat flux (measured)
C_a	J/kgK	specific heat of air at constant pressure per unit mass
λ_v	J/kg	latent heat of evaporation per unit mass
e_i	Pa	vapor partial pressure at h_i (derived from RH_i)
ρ_w	kg/m^3	water density
δ_E	mm/h	uncertainty on E (computed)
E_{sim}	mm/h	simulated values of E
E_{TI}	mm/h	training values of E
E_{val}	mm/h	values of E kept for validation
NSE	$-$	Nash-Sutcliffe efficiency coefficient (computed)

Table 2: Percentage of data rejected for each season of the time series

		ε			
		0.5	1	5	10
Summer	2009	89.9 %	52.8 %	15.5 %	8.7 %
Autumn	2009	63.2 %	31.2 %	7.2 %	3.5 %
Winter	2009/2010	75.8 %	54.4 %	15.5 %	7.6 %
Spring	2010	74.7 %	46.8 %	14.0 %	7.4 %
Summer	2010	69.1 %	26.4 %	6.4 %	3.2 %
Autumn	2010	61.3 %	33.9 %	8.4 %	3.9 %
Winter	2010/2011	68.1 %	48.6 %	16.0 %	7.5 %
Spring	2011	70.4 %	46.8 %	15.8 %	7.7 %
Summer	2011	61.4 %	38.5 %	13.9 %	6.8 %
Total	2009	75.7 %	42.7 %	11.7 %	6.0 %
Total	2010	70.0 %	40.5 %	11.3 %	5.8 %
Total	2011	66.3 %	43.4 %	14.6 %	7.0 %
Total	2009, 2010, 2011	70.6 %	41.8 %	12.2 %	6.1 %

Table 3: Differences between the cumulated E computed by replacing the rejected data with the DS simulation and without replacing the rejected data. Units are in mm.

ε	2009	2010	2011
0.5	176	244	192
1.0	109	117	301
5.0	31	66	144
10.0	14	23	72