

# A strategy for quality controlling hourly rainfall observations and its impact on hourly streamflow simulations

**David E. Robertson<sup>a</sup>, James C. Bennett<sup>a</sup> and Q. J. Wang<sup>a</sup>**

<sup>a</sup> *CSIRO Land and Water, Clayton, Victoria, Australia*  
*Email: David.Robertson@csiro.au*

**Abstract:** Real-time streamflow forecasts are often produced at hourly and shorter time steps. To calibrate the hydrological models used to generate forecasts, archives of streamflow and rainfall observations are essential. The collection and archiving of sub-daily rain gauge observations is typically automated. Errors in rainfall observations can arise for many reasons and may manifest as anomalously high or low values for a single observation or longer periods of time. Quality control of sub-daily rainfall is onerous because of the large volume of data, and sub-daily rainfall data are frequently not quality-controlled when the data are archived in real-time. Errors in rainfall observations used to calibrate hydrological models can lead to poor out-of-sample model simulations and contribute to poor streamflow forecasts.

In this paper we describe a simple automated strategy for quality controlling archives of rain gauge observations. The strategy compares running totals of rain gauge observations with a reference rainfall data set. Where the differences between the observations and the reference dataset exceed a threshold, rain gauge observations are considered to be of poor quality and set to a missing value. The sensitivity of the quality control can be manipulated by adjusting the period over which the running totals are computed, and the definition of the difference threshold.

We apply the quality control strategy to hourly rain gauge observations from five catchments across Australia. For these applications we use the daily Australian Water Availability Project rainfall data set as the reference and quality control the hourly rain gauge observations at daily time steps. We seek only to remove gross errors in the rain gauge observations, where 5-day observed totals are greater than five times the reference or smaller than one fifth of the reference. We demonstrate the efficacy of the quality control strategy for hydrological models run at an hourly time-step with cross-validation experiments. The calibration and validation performance of hydrological simulations forced by the quality-controlled data are vastly superior to those forced by the raw rainfall observations. In some instances, improvements in validation Nash-Sutcliffe Efficiency values greater than 0.7 are achieved by using the quality-controlled rainfall observations. The performance of the hydrological simulations also tends to be more consistent between calibration and validation periods when quality-controlled rainfall observations are used.

The method allows rapid quality-control of large sub-daily rainfall datasets, allowing new streamflow forecasting systems to be established quickly.

**Keywords:** *Rainfall data quality, hydrological modelling and forecasting*

## 1. INTRODUCTION

Rainfall data are essential for calibration and real-time application of continuous hydrological models for forecasting. Calibration of continuous hydrological models often requires many years of historical rainfall data, while real-time forecasting applications require ‘live’ feeds of rainfall data. At daily time steps, archives of quality controlled rain gauge observations are well established and operational methods exist to translate these observations to gridded surfaces, such as the Australian Water Availability Project (AWAP) gridded data set [Jones *et al.*, 2009], from which sub-catchment rainfall can be derived. Flood and short-term streamflow forecasting applications, however, require sub-catchment rainfall at a sub-daily (e.g. hourly) time step. Radar-based gridded rainfall analyses are available for parts of Australia in real-time, however archives of these analyses are temporally inconsistent due to the evolution of algorithms used in real-time, and of insufficient length to calibrate continuous hydrological models. Therefore, only rain gauge observations currently meet both the requirements of having necessary historical archives for hydrological model calibration and data available in real time.

It is critical that rainfall observations used for hydrological modelling and forecasting purposes are as accurate as possible [Oudin *et al.*, 2006]. Errors in rainfall observations can result in poor predictions and forecasts for independent observations. Many factors may contribute to errors in rainfall observations, including sensor recalibration, deterioration and inoperability, incorrect time stamps, changes to the environment surrounding the sensor (e.g. tree growth or location change). The resultant observational errors may persist for only a few time steps or may influence observations for long periods emerging as trends or step-changes in rainfall time series.

Many different quality control methods have been applied to remove errors from rainfall observations at daily and longer time steps, often with different objectives [Allen *et al.*, 1998; Green *et al.*, 2012]. For example some quality control methods seek to remove only gross errors, while others seek to generate statistically homogenous time series. However, all methods typically involve comparing observations for one location to data for neighboring locations or to a reference data set. In many instances, manual intervention is recommended during the quality control process to confirm suspected errors or identify when trends or step-changes in the data commenced. Manual intervention may be practically feasible for daily rainfall observations. However, for quality controlling hourly rainfall observations used for streamflow forecasting, where there are potentially hundreds of thousands of observations at many locations within a catchment, manual intervention in the quality control process becomes impracticable.

In this paper we describe an automated algorithm for quality controlling hourly rainfall observations prior to their use for the development and evaluation of streamflow forecasting models. We adapt existing methods for quality controlling daily rainfall observations and use the AWAP gridded rainfall analysis as a reference data set. We then demonstrate the value of quality controlling rainfall observations for hydrological modelling through a series of cross-validation experiments.

## 2. METHODS

### 2.1. Quality control of rainfall observations

Our objective for quality controlling rainfall data is to remove spurious observations from the archive of hourly observations that are likely to have a significant impact on the quality of streamflow simulations and forecasts. In operational forecasting applications it is very difficult to identify or remove long-term trends or step changes in data and we do not seek to do so. Based on preliminary data exploration we identified three types of poor quality data that can have significant impacts on estimates of sub-catchment rainfall for hydrological modelling purposes:

- anomalously large values
- anomalously small values or incorrectly recorded zero values
- data associated with inappropriate time stamps, commonly caused by the incorrect time zone being used

Due to the large number of rainfall stations used and the total number of observations (up to 200,000 observations per station) we perform quality control using an algorithm. We base our quality control algorithm on double mass plots [Allen *et al.*, 1998] and correlations between observed and reference data. We also generate several other diagnostic plots that facilitate rapid checking to ensure that the quality controlling has been successful.

Double mass plots are a common tool for diagnosing trends and inconsistencies in rainfall and streamflow data. A double mass plot is produced by plotting the cumulative sum of rainfall observations against the cumulative sum of a reference set of rainfall observations. If the data are consistent and are affected by the same factors, then the plot will be a straight diagonal line. However, where the data are influenced by different trends or have inconsistencies the plot will deviate from the diagonal.

We use double mass plots only to identify gross inconsistencies in station rainfall data. These inconsistencies are where the rain gauges recorded anomalously small or large quantities of rainfall relative to the adopted reference. The gross inconsistencies are identifiable in double mass plots by horizontal or vertical (or near horizontal or vertical) segments.

No reference sets of rainfall observations at the hourly time steps exist. We therefore apply data quality control using double mass plots at daily time steps. We adopt the AWAP rainfall data set [Jones *et al.*, 2009] as the reference for constructing double mass plots. For each rainfall gauge location, we aggregate hourly observations to daily totals ensuring observation times are consistent with the AWAP data set. We compare these observed daily totals to AWAP daily rainfall for the co-located grid cell. We generate double-mass plots for each location separately and visually inspect the plots to understand the nature of the rainfall observations relative to the AWAP reference. In the production and analysis of the double-mass plots, periods where rain gauge observations are missing are removed from both the observed and reference time series.

We use an algorithm to identify segments that suggest inconsistencies exist between the observed and reference (AWAP) data, as follows:

1. For each observed daily total at time  $t$  we calculate the slope,  $s$ , of the double-mass plot for a window from time  $t - n$  to  $t + n$  (i.e., a window width of  $2n + 1$  time steps), where  $n$  is a parameter.
2. We flag observations where  $s > m$  or where  $s < \frac{1}{m}$  as inconsistent data, where  $m$  is a parameter.
3. Periods of inconsistent data in the daily time series are replaced with a missing data value in the hourly rain gauge time series.

Effectively this algorithm identifies observations as inconsistent when the running  $2n + 1$  -day total observed rainfall is  $m$  times larger or smaller than the corresponding  $2n + 1$  -day total AWAP rainfall.

For the window size, we adopt a value of  $n = 2$  (i.e., a 5-day window) through trial-and-error. In some instances, a shorter window appeared to eliminate too many seemingly good observations, while a longer window did not remove all obvious gross errors. For the slope parameter, we adopt  $m = 10$ , again based on trial and error.

The correlation between the daily rain gauge observations and AWAP rainfall is also used for quality control. Where the correlation is less than 0.4 (a value also derived from empirical trial and error), all data for the rain gauge are set to missing. Subsequent investigation found that this problem may often have been caused by an incorrect time stamp for all or some of the record, for example the data were incorrectly marked as local time when they should actually have been labelled UTC, or vice versa. However, as it was not possible to definitively identify the correct timestamp, these data were also replaced by missing values.

We apply the above analysis separately to each rainfall station. To validate our quality control method, we use the common (manual) technique of visually inspecting plots of cumulative rainfall against time for all stations and the entire period of record on a single plot. We also compare double mass plots of the raw and quality controlled data to double check that our algorithm removes gross errors.

## 2.2. Impact of rainfall data quality on performance of hydrological models

To understand the impact of quality controlling rainfall observations on the performance of hydrological model we perform a series of cross-validation experiments. Semi-distributed hydrological models used for hourly streamflow forecasting applications [Bennett *et al.*, 2014] are used for these cross-validation experiments. Hourly sub-catchment rainfall is derived from rain gauge observations using inverse-distance-squared weighting. (We note that more sophisticated algorithms that take into account, e.g., elevation, may result in better estimates of sub-catchment rainfall, but inverse-distance-squared weighting is sufficient for this study.) Sub-catchment rainfall is derived using both the raw rain gauge observations and the cleaned data that is quality controlled.

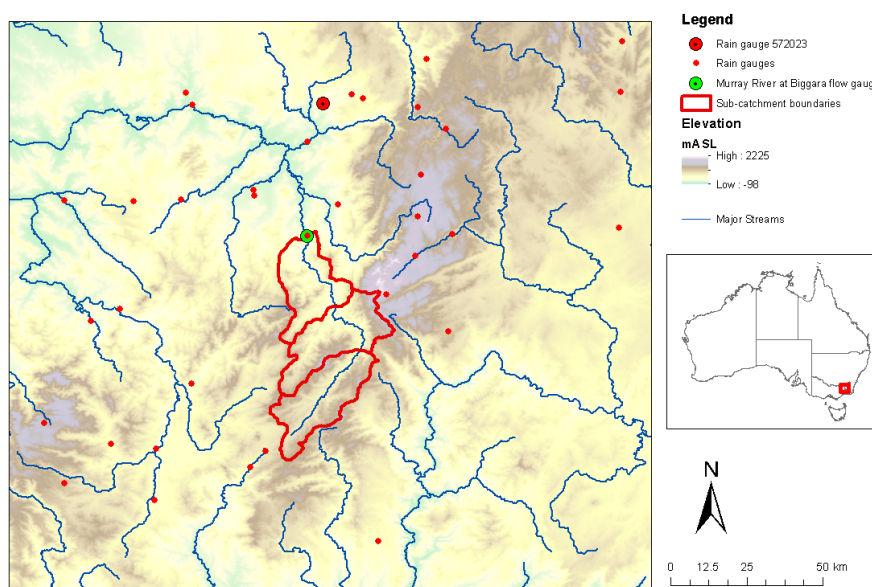
Hydrological model parameters are calibrated using the shuffled complex evolution algorithm to maximize an objective function that is a weighted average of the Nash-Sutcliffe efficiency of untransformed and log-transformed streamflow and a measure of bias [Bennett *et al.*, 2014]. A 3-fold cross-validation approach is used to assess the performance of the hydrological model. Practical application of the cross-validation approach involves running the hydrological model for the entire record during calibration with the observed streamflow for the validation period of interest omitted in the computation of the objective function.

In this study we assess the performance of the hydrological model using the Nash-Sutcliffe efficiency of both untransformed and log-transformed streamflows and also using flow-duration curves.

### 3. RESULTS

The quality control algorithm was applied to a range of catchments across Australia with contrasting hydrological and climatic conditions. In some instances, the observed rainfall record contained few gross errors and therefore application of the quality control algorithm had an inconsequential impact on the observed rainfall time series and streamflow simulations. Here we present results for the catchment of the Murray River at Biggara streamflow gauge, located on the border between Victoria and New South Wales, where quality control of rainfall observations was found to be important.

Figure 1 shows the location of the rain gauge network, sub-catchment boundaries used in hydrological modelling and the catchment of the Murray River at Biggara streamflow gauge. For this analysis, we use available rainfall and streamflow records for the period 1990 – 2012.



**Figure 1.** Map of the catchment area of the Murray River at Biggara streamflow gauge.

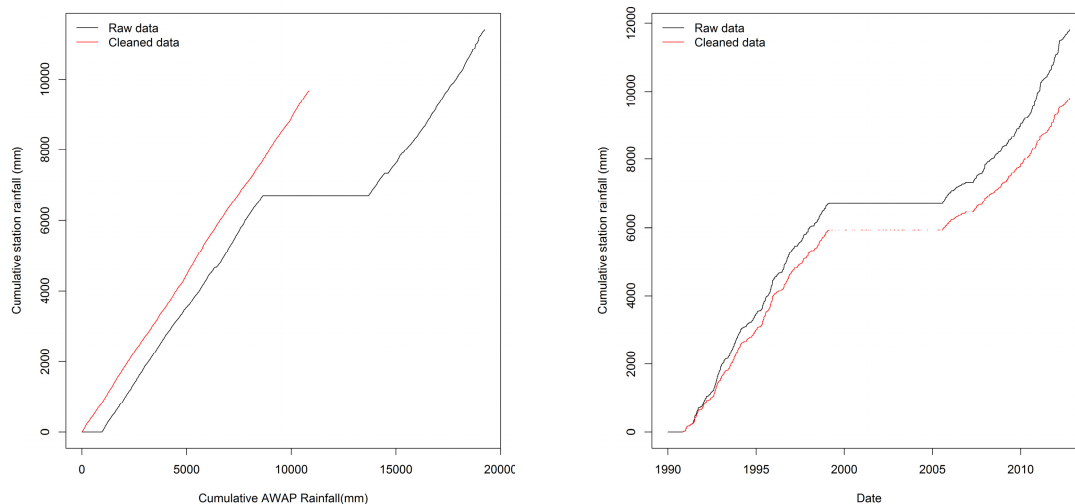
#### 3.1. Quality controlling rainfall observations

Figure 2 presents the double-mass plot and cumulative rainfall plotted against time for one station in the northern part of the Murray River at Biggara catchment. The double-mass plot of the raw station rainfall data (black line) clearly shows that there are potentially significant data quality issues with the one very long horizontal line segment, and multiple other shorter ones. Closer examination also reveals there are also some (near) vertical line segments particularly near the top left hand corner of the plot. The horizontal line segments are related to occasions where the rain gauge recorded zero rainfall, while the reference rainfall recorded non-zero values. The vertical segments indicate that the recorded station observations are considerably larger than corresponding reference rainfall data. The plot of cumulative rainfall against time indicates that the available rainfall time series contains only zero values for the period between 1998 and 2008.

The red lines in Figure 2 are for cleaned data after application of the quality control algorithm. The double mass plot is a straight diagonal line with no obvious vertical or horizontal segments, indicating the quality controlled data contain few gross errors. Not only did the quality control algorithm remove the incorrectly recorded zero values, but it also removed many non-zero rain gauge observations that were not consistent with

the reference time series. The net effect of removing the non-zero observations was to reduce the total rainfall observed over the period between 1990 and 2012 by approximately 2000 mm, or ~60 mm/year.

The red line in the plot of cumulative rainfall against time shows that not all the zero values that were recorded during the period between 1998 and 2008 were removed by the quality control algorithm. The zero values that are not removed correspond to occasions when the total rainfall for the five days either side is near zero for both the observed and reference (AWAP) time series.



**Figure 2.** Cumulative rainfall gauged at site 572023 in the Northern Murray River catchment plotted against cumulative AWAP rainfall (left) and against time (right), before and after quality control is applied.

### 3.2. Impact on performance of hydrological models

Quality controlling the rain gauge observations can lead to substantial improvements in the cross-validation performance of hydrological models. Tables 1 and 2 show that the calibration performance of the models calibrated using quality controlled rainfall data are better, and often substantially better, than the performance of the model calibrated using the raw rainfall observations. The improvements in the model calibration performance are evident in the NSE of both untransformed and log-transform flow indicating that improvements occur for both high and low flows. Improvements in validation performance also occur, although for one period (Period 2) the NSE validation performance declines somewhat. When improvements in the validation performance do occur they are very large, with the NSE increasing from values less than zero to levels that are more comparable to the calibration performance.

The performance of the models calibrated using the quality controlled rainfall data is also much more consistent than for models calibrated using the raw rainfall observations. This improved consistency exists between the different cross-validation periods and between calibration and validation results. For hydrological forecast evaluation and real-time forecasting applications, consistent model performance is important. Hydrological forecasting applications usually use bias-correction and error updating methods to improve forecast accuracy and to ensure that forecasts appear to be a natural extension of observed streamflows. Consistent hydrological model performance between calibration and validation periods means that error models derived are likely to perform well for a range of conditions.

**Table 1.** Calibration and validation Nash-Sutcliffe efficiency of untransformed flows for Murray at Biggara.

Cross-validation period	Calibration start*	Calibration end*	Validation start	Validation end	Calibration NSE		Validation NSE	
					Raw rainfall data	Cleaned rainfall data	Raw rainfall data	Cleaned rainfall data
1	1/02/1992	30/07/2012	1/02/1992	1/12/1998	0.71	0.72	-0.07	0.65
2	1/02/1992	30/07/2012	1/12/1998	30/09/2005	0.29	0.77	0.77	0.71
3	1/02/1992	30/07/2012	30/09/2005	30/07/2012	0.42	0.79	-0.33	0.62

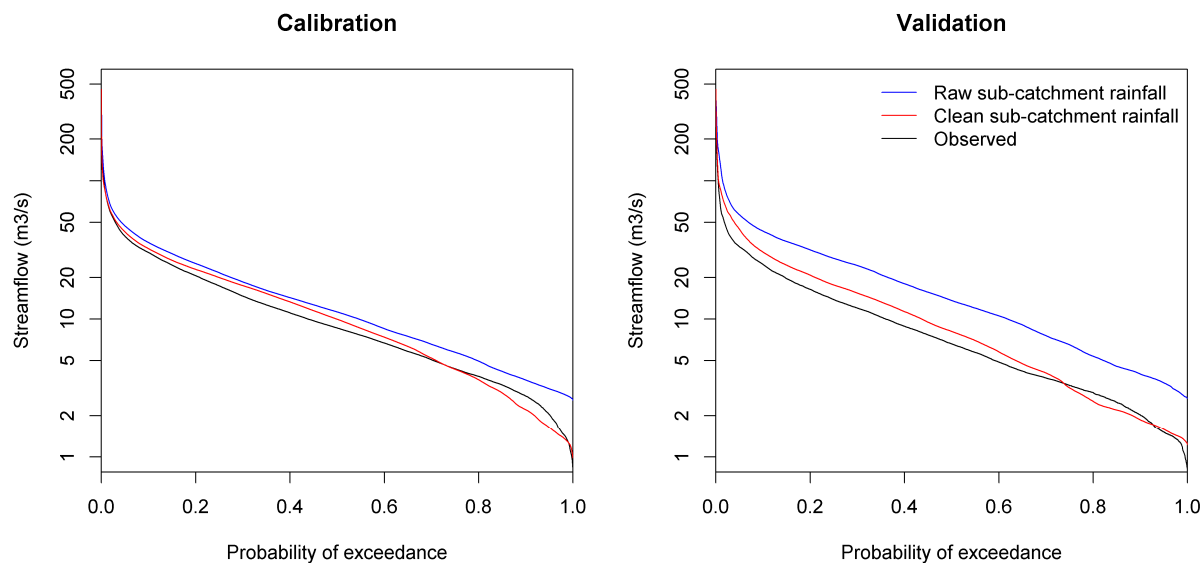
\*Streamflow data for the validation period are marked as missing during calibration

**Table 2.** Calibration and validation Nash-Sutcliffe efficiency of log-transformed flows for Murray at Biggara.

Expt. Num.	Calibration start*	Calibration end*	Validation start	Validation end	Calibration NSE(log)		Validation NSE(log)	
					Raw rainfall data	Cleaned rainfall data	Raw rainfall data	Cleaned rainfall data
1	1/02/1992	30/07/2012	1/02/1992	1/12/1998	0.80	0.84	-0.38	0.80
2	1/02/1992	30/07/2012	1/12/1998	30/09/2005	0.46	0.86	0.82	0.73
3	1/02/1992	30/07/2012	30/09/2005	30/07/2012	0.63	0.83	0.44	0.83

\*Streamflow data for the validation period are marked as missing during calibration

Figure 3 presents calibration and validation flow-duration curves for the first cross-validation period. Differences between the calibration flow-duration curves of the models calibrated using the raw and quality controlled rainfall data are relatively small but tend to be most evident for the highest 5% and the lowest 20% of flows. The validation flow-duration curves show much larger differences than the calibration flow-duration curves. In general, the flow-duration curves produced using the quality controlled rainfall data are much closer to the observed curve than those produced using the raw rainfall data. The validation flow-duration curve shown in Figure 3 is a typical example where the curve produced using the raw rainfall data is offset from the observed curve, suggesting that there is a clear bias in the validation simulations for all exceedance probabilities. As this problem is largely not present in the streamflow simulated using the quality controlled rainfall data we conclude that these errors result from inconsistencies in the observed rainfall between the calibration and validation periods which the quality control algorithm was able to effectively remove.



**Figure 3.** Calibration and validation flow-duration curves produced using hydrological models forced with raw and quality controlled rainfall data for the Murray catchment for the first cross-validation period.

#### 4. DISCUSSION AND CONCLUSIONS

In this paper we have described a relatively simple, but robust, algorithm for automated quality control of hourly rainfall observations. The method makes use of now commonly available gridded daily rainfall data products. It removes gross errors in rainfall time-series using an adaption of a double-mass plot to ensure consistency between five-day running totals of rain gauge observations and a reference rainfall data set. We show that the quality control method effectively removed gross errors in rain gauge observations and that the removal of these gross errors can significantly improve the cross-validation performance of hydrological models.

For this study we adopted the gridded AWAP rainfall analysis product as our reference climate data set. AWAP rainfall is derived from quality controlled daily observations and is available across Australia. However, even where no such gridded rainfall product is available, the method can be, and has been, applied. In these instances, we have constructed a reference data set using the mean rainfall from a neighbouring set of rainfall gauges, but it may also be possible to use a global reanalysis such as ERA-Interim (<http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>). However, using either a global reanalysis or average of neighbouring stations as the reference data set can highlight one of the limitations of the approach described in this manuscript.

The thresholds used to flag inconsistent data in this study assume that the slope of the double-mass plot is close to one. In some circumstances the expected slope of a double-mass plot may be different to one, such as when taking the average of neighbouring stations that are far away and experience different climates or when using gridded products which characterise rainfall at large spatial scales. In these circumstances, the parameters adopted may require adjustment, for example to ensure that (correctly) measured large rainfalls are not mistakenly removed from the data set if  $m$  is too small.

There are several ways that the method can be tailored to different applications in addition to adjusting the  $m$  parameter used to flag inconsistent data. In our applications, we found that adjusting the  $n$  parameter that controls the window size over which the running totals are compared influences the sensitivity of the algorithm. Ultimately,  $n$  and  $m$  can be adjusted so that the quality control algorithm best suits the purposes of the user.

The quality control method described in this paper has been applied retrospectively to archives of rainfall data. This is highly valuable for the development of forecasting systems and their retrospective evaluation. However, real-time quality controlling of rainfall observations is also required for operational forecasting applications. AWAP data are not presently available in real-time, meaning that AWAP cannot be used as a reference to quality control rainfall data in real-time. However, the method has potential to be adapted for real-time applications, particularly if the average of neighbouring rain gauges is used as a reference data set.

## REFERENCES

- Allen, R. G., L. S. Pereira, D. Raes, and M. Smith (1998), *Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements*, 1-326 pp., FAO, Rome.
- Bennett, J. C., D. E. Robertson, D. L. Shrestha, Q. J. Wang, D. Enever, P. Hapuarachchi, and N. K. Tuteja (2014), A System for Continuous Hydrological Ensemble Forecasting (SCHEF) to lead times of 9 days, *Journal of Hydrology*, 519, Part D(0), 2832-2846.
- Green, J., F. Johnson, D. McKay, S. Podger, M. Sugiyanto, and L. W. Siriwardena (2012), Quality Controlling Daily Read Rainfall Data for the Intensity-Frequency-Duration (IFD) Revision Project, in *2012 Hydrology and Water Resources Symposium*, edited, Engineers Australia, Sydney, NSW Australia.
- Jones, D. A., W. Wang, and R. Fawcett (2009), High-quality spatial climate data-sets for Australia, *Aust. Meteorol. Oceanogr. J.*, 58(4), 233-248.
- Oudin, L., C. Perrin, T. Mathevet, V. Andréassian, and C. Michel (2006), Impact of biased and randomly corrupted inputs on the efficiency and the parameters of watershed models, *Journal of Hydrology*, 320(1-2), 62-83.