USING STOCHASTIC DISAGGREGATION IN MODELING POORLY MEASURED CATCHMENTS

P.J.J.F. Torfs

Wageningen University, Chair of Hydrology and Quantitative Water, Wageningen, Netherlands
paul.torfs@wur.nl

Introduction

Scale problems are abundant in all kinds of hydrological modeling. In this paper we will discuss a very particular one in the field of rainfall runoff modeling: that of a "poorly measured catchment". This naming will be used here for the case where during a rather short measurement campaign data (both rainfall and runoff) were obtained with a high time resolution (e.g on hour scale) and used to fit a trustworthy rainfall runoff model, proving also that this high resolution is needed to capture the dynamics of the rainfall runoff process. The problem arises when it is asked to use this valuable knowledge to calculate runoff for series where on the input side (rainfall) data are available only on a coarser time scale (e.g. on days). Figure 1 illustrates this problem.

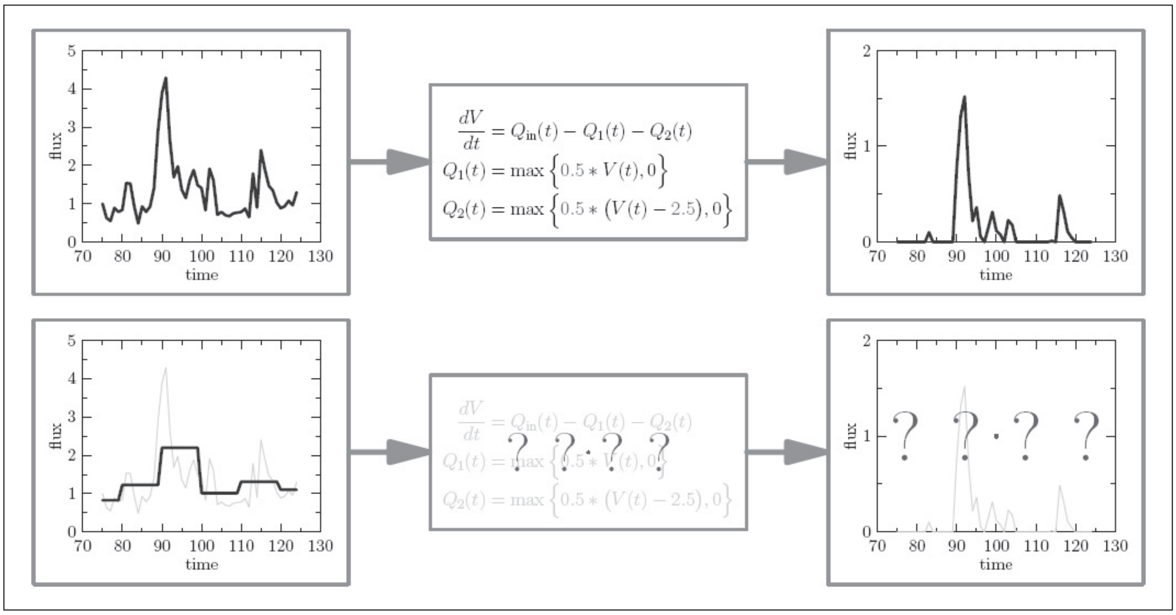

Figure 1. The problem discussed in this paper: top shows the available knowledge: based on a short period but with high resolution data (input left, output right) available, one was able to obtain a good model (here schematically illustrated by some differential equations in the middle box). The problem is how to use this knowledge when only low time resolution data on the input (thick line on the lower left box, the grey background stands unknown reality) are available

In hydrology, this coarser time information usually consists of means: daily rainfall intensity is the mean of the hourly rainfall intensities. On the output side one may want to answer different questions. One of the required outputs may also consist of means: e.g. mean daily runoff. In this contribution we however will ask for more: an assessment of the output on high resolution, e.g. runoff peaks on hourly scale, where one could e.g. ask for the probability that a runoff peak on hourly scale exceeds a certain threshold. Of course these high resolution output can not be "calculated" exactly and can only be answered statistically. This paper we will outline a methodology by which we can make such statistical calculations and estimations. The error on these estimations represent then the lack of knowledge generated by the coarseness of the input.

Methodology

On can in general distinguish between several methodologies to answer to the problem raised above:

1. Use the old model (including its fitted parameters) on the coarse resolution. If the model would have been linear, this would yield acceptable results: mean input generates mean output. As rainfall runoff models are usually non-linear, this does in general produce acceptable results for catchments.
2. Use the structure of the old model, but adjust the parameters. The new parameters are often called "effective". These new parameters values should then replace the effects due to the fluctuations of the input on the high resolution scale. By that, these parameters become less physical and more statistical and are difficult to transfer to other catchments.
3. Build a new model designed to work on the coarse time scale. This approach neglects completely the available the high-resolution knowledge. All the approaches use as principle: adapt the model to the data. Moreover, none of these approaches allow any conclusion (not even statistical) on the high resolution time scale.

Here a fundamentally different approach is proposed: the data will be adapted to the existing model. For this, artificial disaggregated input series with the given low resolution means are constructed, fed to the original model resulting in the high resolution output. There are of course many possible disaggregations: a large number of them should be generated, and this ensemble should statistically be representative for the high resolution input observed during the measurement campaign. All these disaggregations are used as input for the model to produce a statistically representative ensemble of outputs. By this Monte-Carlo approach,not only the means of the runoff can be estimated, but also their uncertainty and even the statistics of the runoff on a high resolution scale (Figure. 2).

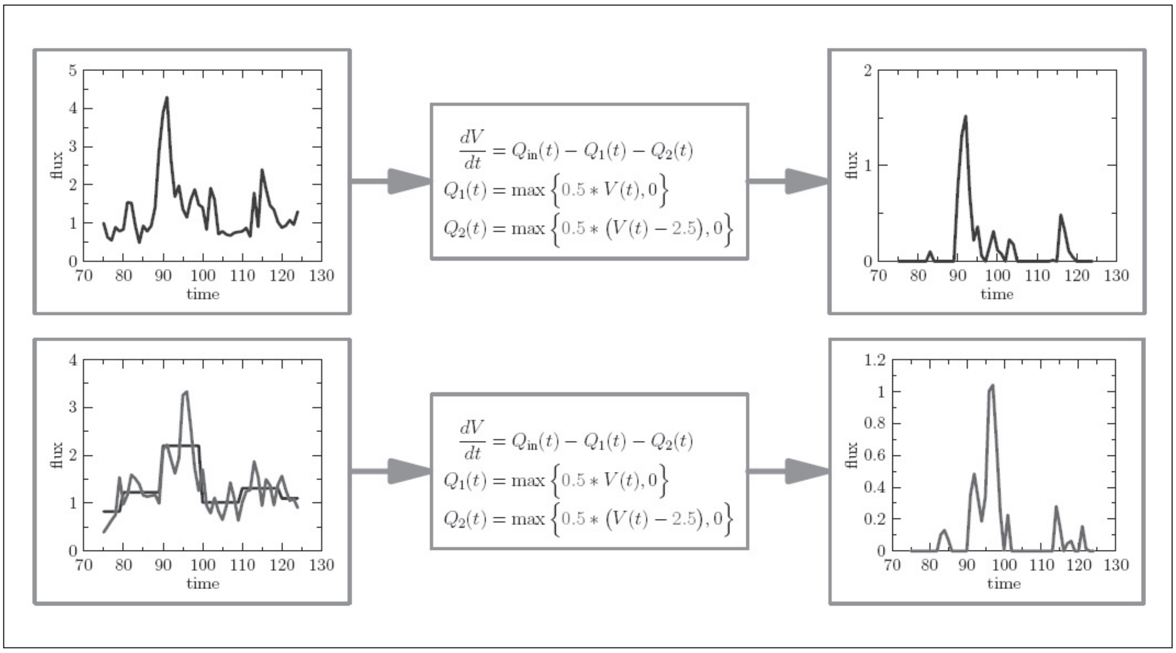

Figure 2. At the top the available knowledge is shown. At the bottom left, the available low resolution input means are stochastically disaggregated into high resolution input and can as such be used for existing model to produce high resolution output. This procedure should be used in a Monte-Carlo framework: many inputs should be simulated and the corresponding outputs should be analyzed statistically

Stochastic disaggregation

The proposed methodology depends on the stochastic disaggregation procedure, which is a technically difficult problem. This problem can be stated in general as: given the means of a time series on a coarse scale, simulate a statistically representative ensemble of high resolution time series which have exactly the given means.

For the statistics to be simulated the variation over different time scales is used, as given y (see also Figure 3):

$$\mathtt{V}[T] = \mathtt{VAR}[X_T(t)] = \mathtt{VAR}\left[\frac{1}{T}\int_{nT}^{(n+1)T} X(\tau)\, d\tau\right] \tag{1}$$

This function should be estimated using high resolution data, e.g. by the data obtained during the measurement campaign. It can be proven by the well known Jaynes theorem (see e.g. MacKay, 2003) that description above defines a unique time series, which can be simulated even conditionally given the means by using standard Metropolis-Hastings Markov Chain (see e.g.: Bistinas, Torfs, Warmeredam, 2008) methods. A particular difficult problem for the application to rainfall series (see also: Granville, Smith, 1996) is the choice of measuring variation. Most success was obtained by measuring this twofold:

1. by the amount of zeroness over different time scales
2. by measuring a form of percentage of variance for the wet part.

Results and conclusions

Artifical results were obtained by using a form of the Wageningen model (see also: Jaynes, 1990). High resolution data were available on 3 hour scale. They were artificially aggregated to daily scale to test the proposed methodology. The first results obtained show that the methodology works. The quality of the results depend however on the quality of the disaggregations. A particular hard problem proved to be the estimation of the parameters in the Jaynes model. However the Monte Carlo approach proposed seems to be the best way to treat the poorly measured catchment problem. It uses optimally the knowledge obtained from the measuring campaign and translated the lack of high resolution information in standard statistically treatable uncertainty. But most important, it produces results (although with uncertainty) in runoff on the high resolution scale, important for all kind of hydrological studies.

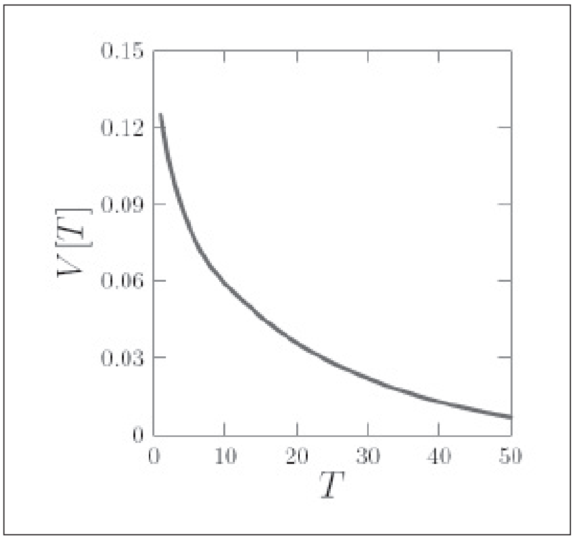

Figure 3. A typical variation reduction function, showing how the variation in time series diminishes as one averages out over large time intervals

References

Bistinas I., Torfs P., Warmeredam P., 2008: *Impact assessment of different time scales on model parameters for the small catchment fo the Hupselse Beek.*

Granville V., Smith R.L, 1996: *Disaggregation of rainfall time series via gibbs sampling*. Preprint, Statistics Department, University of North Carolina at Chapell Hill.

Jaynes E. T., 1990: *Probability Theory: The Logic of Science*. Cambridge University Press.

MacKay D.J.C., 2003: *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.