 October 2012

Jonas Olsson

# RANDOM CASCADE MODEL: documentation

This report contains a description of two Matlab scripts for analysis and disaggregation of short-term rainfall time series based on the concept of random cascades. The scripts are provided as a supplement to the book *Impacts of Climate Change on Rainfall Extremes and Urban Drainage Systems* (Willems et al. (eds), IWA Publishing, 2012).

# Temporal rainfall disaggregation by random cascade model: micro-canonical approach with beta-distributed generator (= Box 2.2 in the book)

Temporal rainfall disaggregation continues to be important and is required in different projects, e.g. dealing with rainfall erosivity and urban hydrological climate change assessment. One useful application of random cascade models is for disaggregation of continuous rainfall time series. There are several variations on this theme but here we outline one of the main approaches currently used. Generally, a temporal cascade process involve successive branching which describes how some quantity (e.g. rainfall) is redistributed as the time resolution changes. The redistribution implies a successive fine-graining process that starts from an original, large-scale resolution *rL* and continues until a target, small-scale resolution *rS* is reached.

Commonly a branching number of 2 is used, which implies a redistribution of total rainfall in period *i* at resolution *r*, *Ri,r*, between the amount associated with the first and last half of the period, respectively. In a micro-canonical approach, the amount redistribution may be described by multiplicative weights *Wi,r*, 0≤*Wi,r*≤1, that assigns *Wi,r*\**Ri,r* to the first half of the period and (1-*Wi,r*)\**Ri,r* to the last half. In each branching, two principal possibilities exist: (1) *W*1=0 or *W*1=1 and (2) 0<*W*1<1. The former option, where the total amount originates from one half of the period, is related to rainfall intermittency which is one key issue in cascade-based rainfall disaggregation. The occurrence of (1) and (2) may be expressed in terms of probabilities Pr01=Pr(*W*1=0 or *W*1=1) and Pr*xx*=Pr(0<*W*1<1)=1-Pr01. In some applications a distinction has been made between Pr(*W*1=0) and Pr(*W*1=1), respectively. Depending on the range of resolutions considered, Pr01 may either be assumed resolution-independent or parameterised as a scaling law:

$Pr\_{01}\left(r\right)=c\_{1}r^{c\_{2}}$ (1)

where *c*1 and *c*2 are constants.

Another key issue concerns the characteristics of the weights *Wi,r* related to the case 0<*W*1<1, $W\_{i,r}^{xx}$. Generally the probability distribution of $W\_{i,r}^{xx}$ is assumed to follow some theoretical distribution (‘cascade generator’) and the one-parameter beta distribution:

$p(W)=\frac{1}{B(a)}W^{a-1}\left(1-W\right)^{a-1}$

where *a* is a shape parameter and *B* the beta function, is often found to be an acceptable approximation. There is empirical evidence that the shape parameter *a* of the distribution depends on the time resolution *r* and that generally this dependency may be well described by the scaling relationship:

$a\left(r\right)=a\_{S}r^{-H}$ (2)

where *aS* is the shape parameter related to the highest resolution *rS* and *H* an exponent describing how rapidly *a* decreases with decreasing resolution.

In its basic version, the model is thus specified by four parameters: *c*1, *c*2, *aS* and *H*. Parameter estimation may be performed by a coarse-graining procedure, a “reverse cascade” which starts from the highest resolution *rS*. By gradually aggregating adjacent rainfall amounts two by two, breakdown coefficients (BDCs) may be calculated as:

$BDC\_{i,r}=\frac{R\_{i,r}}{\left(R\_{i,r}+R\_{i+1,r}\right)}$ (3)

where *BDCi,r* thus corresponds to the weight *Wi,r* used to redistribute the total rainfall amount in a fine-graining process. After the coarse-graining has proceeded to up to the lowest resolution *rL*, the BDC-values may be used for estimation of all parameters.

Concerning the calibration and application of the disaggregation model to an existing time series at original resolution *rL*, two principal situations may be considered. One is when some representative data at target resolution *rS* are available, e.g. from a temporary measurement campaign at the same location or from a nearby high-resolution gauge. Then BDCs may be extracted and parameters estimated over the actual resolution interval *rL*≤*r*≤*rS*. The other situation is when no representative high-resolution data are available. Then parameter estimation may still be possible by coarse-graining from resolution *rL* to gradually lower resolutions. Then Equations (1) and (2) may be evaluated and if the scaling laws are found to hold they may be extrapolated to estimate Pr01 and *a* for higher resolutions up to *rS*. This is an attractive feature of scaling-based disaggregation but it must be emphasised that extrapolation is always associated with large uncertainties and that it should always be attempted to verify parameters using some surrogate data.

After calibration, Monte Carlo simulations are performed to gradually fine-grain the data and generate realisations at resolution *rS*. In case *rS* is not reached exactly by resolution doubling from *rL*, a higher target resolution may be used followed by interpolation to *rS*.

Alternative options include using a different generator, i.e. theoretical distribution for $W\_{i,r}^{xx}$, and e.g. normal and beta-normal distributions have been suggested. Further, a dependency of Pr01 on rainfall amount *Ri,r* is empirically supported. Also a dependency on the position within the rainfall sequence may be considered. It should finally be mentioned that the intensity resolution of the measurement device may have a strong impact on the BDC characteristics which requires careful analysis.

For applications of random cascade micro-canonical disaggregation with particular focus on urban hydrological processes and time scales, see e.g. Molnar and Burlando (2005), Hingray and Ben Haha (2005) and Licznar et al. (2011).

# Matlab scripts

## General model structure

The simple model of Olsson (1998) (further developed and tested by Güntner et al. (2001)) has proved to be promising but also limited. It assumes a very strict form of scaling with constant properties over all scales. This often seems to work reasonably well approximately between time scales 1 week and 1 hour, but not below one hour. In many applications it is however essential to reach time scales of 5-10 min. To improve the model applicability it was decided to try to modify the model, mainly towards less strict scaling but also in other respects to better reproduce observed dependencies. Another purpose of this work is to implement the model in reasonably flexible Matlab routines, which is convenient in light of the statistical and graphical capabilities of Matlab.

As in Olsson (1998), the basis is a random cascade process. In this process, time periods are successively halved and the total rainfall volume redistributed between the halves according to the probabilities P(0/1), P(1/0) and P(x/x). These probabilities are assumed to depend on both the position of the period in the rainfall sequence and the rainfall volume during the period. If P(x/x), the volume in each half is determined based on weights *w* drawn from a theoretical probability distribution (uniform in Olsson (1998)).

Based on tests using rainfall time series from Lund (see further below), the following modifications of the model in terms of probabilities and x/x-distributions appear likely to improve its flexibility and accuracy.

### Probabilities

- More volume classes. In Olsson (1998), two volume classes were used, above and below the mean volume of the position type and cascade step, which is a rather coarse division. In this evaluation it is tried to use three classes instead, separated by percentiles 33 and 67. Figure 1 shows how the probabilities vary with volume class and it is clear that there is often a substantial difference between the classes. Thus a more detailed volume dependence appears justified.

- P-dependence on volume. Instead of having separate P-values for the four volume classes, it appears possible to reasonably well describe the variation of P with volume as a linear function P=*int*+*slo*\*vc where vc is volume class (1-4), *slo* is the slope of a linear regression and *int* is the intercept at vc=0 (Figure 1).

- P-dependence on cascade step (i.e. time scale). In Figure 1 it is indicated that the linear relationship between P and vc changes with cascade step. Overall the slope *slo* seems to remain relatively constant but the intercept *int* seems to change rather regularly. A possible model for this observed variation is to use a fixed slope and then let the intercept vary with cascade step. As the fixed slope, the mean slope *slom* obtained from all cascade steps is used (the mean regression line *intm*+ *slom*\*vc is shown as a dashed line with squares in Figure 1). Figure 2 shows how the fitted intercept *int* varies with cascade step. It appears possible to model the dependence of *int* on cascade step cs using a linear model *int*=*c1*+*c2*\*cs.



*Figure 1: Variation of probabilities with volume class (x: volume class (1: small, 2: medium, 3: large); y: probability). In the diagram titles, P denotes position type (1: isolated, 2: starting, 3: enclosed, 4: ending), D denotes division type (1: 0/1, 2: 1/0, 3: x/x), and N denotes the total number of periods for this position and division type. The different-coloured solid lines represent different cascade steps. The dashed line with squares represent the mean of all cascade steps.*



*Figure 2: Variation of intercept int with cascade step (x: cascade step (step 1 represents the “cascading” from 16 to 8 min, step 2 from 32 to 16 min, etc.); y: value of int). In the diagram titles, P denotes position type (1: isolated, 2: starting, 3: enclosed, 4: ending), D denotes division type (1: 0/1, 2: 1/0, 3: x/x), and Slo denotes the mean slope for this position and division type, estimated from the fitted mean lines in Figure 1. The diagrams in the third column shows the fraction of 0/1-divisions of all “non-x/x-divisions”, summed over all volume classes.*

- Simplifications and suggested P model. As discussed also in Güntner et al. (2001), it appears possible to simplify the P model somewhat. Concerning starting and ending boxes, these exhibit a similar variation of P(x/x) and a similar but reversed variation of P(0/1) and P(1/0). Thus these may be considered as the same edge-of-rainfall-sequence type of box. For this type, it appears suitable to model P(x/x) as P(x/x)=*int*+*slom*\*vc with *int* estimated from *int*=*c1*+*c2*\*cs, i.e. the most flexible P model with the three parameters *slom*, *c1* and *c2*. Then P(0/1) for starting (P(1/0) for ending) is relatively independent of cascade step, thus it can be modeled using only *intm*+ *slom*\*vc. Finally P(1/0) for starting (P(0/1) for ending) is estimated as P(1/0)=1-(P(x/x)+P(0/1) (P(0/1)=1-(P(x/x)+P(1/0)). Concerning isolated and enclosed boxes, these also exhibit similar properties and judged from these results they may in fact be also considered as one type of box. If modelling them separately, as for the edge-of-rainfall-sequence type of box, P(x/x) should be estimated as P(x/x)=*int*+*slom*\*vc with *int* estimated from *int*=*c1*+*c2*\*cs. Then P(0/1) and P(1/0) are always approximately equal (see third column in Figure 2), i.e. they can both be estimated as P(0/1)=P(1/0)=(1-P(x/x))/2.

### x/x-distributions

- One-parameter beta distribution. The uniform distribution is clearly not suitable for enclosed boxes at small time scales in the case of Lund (see Figure 2 in Olsson (1998)). An alternative may be to use a one-parameter (called *a*) beta distribution, which has uniform distribution as a special case (*a*=1). Figure 3 shows fitted beta distributions (lines) to the empirical histograms (bars) and overall the fit is good (at high cascade steps, i.e. large time scales, the number of values is small and the histograms as well as the fits naturally very uncertain).



*Figure 3: Variation of empirical x/x-distribution with cascade step (bars) and fitted beta distribution (lines) (x: weight w; y: probability). In the diagram titles, P denotes position type (1: isolated, 2: starting, 3: enclosed, 4: ending), CS denotes cascade step (for Lund, step 1 represents the “cascading” from 16 to 8 min, step 2 from 32 to 16 min, etc.), and N denotes the total number of x/x-divisions for this position type and cascade step.*

- Dependence of *a* on cascade step and suggested model. Figure 4 shows how *a* varies with cascade step. For starting and ending boxes, *a* is relatively constant (although there is a weak decrease of a with cascade step for starting boxes) and the joint mean value appears to be a satisfactory approximation. Also in the case of isolated boxes a constant approximation is probably sufficient. All these position types would thus have distributions rather close to uniform but with slightly higher probability in the centre of the distribution (i.e. where *w*=0.5). (In fact it may be possible to use uniform distributions with little loss of accuracy.) For enclosed boxes it seems possible to model *a* as a log-log linear function of the time scale *s*, i.e. log(*a*)=*c3*+*c4*\*log(*s*).



*Figure 4: Variation of beta-parameter a with cascade step (x: cascade step (step 1 represents the “cascading” from 16 to 8 min, step 2 from 32 to 16 min, etc.); y: value of a). In the diagram titles, P denotes position type (1: isolated, 2: starting, 3: enclosed, 4: ending). The left column has linear diagrams; the right column log-log diagrams.*

## Analysis (model evaluation and calibration)

The analysis script is called *Random\_Cascade\_Model\_analysis.m*.

Input is supposed to be a high-resolution rainfall time series with a regular time step (e.g. 5 or 10 min). The values, in [mm/time step] (including 0-values), should be in one column (see file *sample\_data\_analysis.txt*). Note that no seasonal sub-division is done but such separation (which is recommended) requires that different seasons are pooled into separate time series which are then analysed.

In the script, the only decisions to be made is (1) the number of volume classes used (three is recommended) and (2) the number of aggregation steps, where in each step the temporal resolution is halved.

As output, the script generates graphs as shown in Figures 1-4 which illustrate how well the assumed parameterisation is supported by the data. Further, a parameter file called *RCM\_parameters.mat* is generated with estimated values of the parameters described in Section 2.1 above.

## Disaggregation

The disaggregation script is called *Random\_Cascade\_Model\_disaggregation.m*.

Input is supposed to be rainfall time series with a regular time step (e.g. 1 day). The values should be given as *multiples of the volume resolution* of the gauge used in the model calibration. Thus, if the daily total is 1.5 mm and the volume resolution is 0.1 mm, the value used should be 1.5/0.1=15 (this corresponds to the number of tips of a tipping bucket gauge with resolution 0.1 mm). The values (including 0-values) should be in one column (see file *sample\_data\_disaggregation.txt*). Note that no seasonal sub-division is done but such separation (which is recommended) requires that different seasons are pooled into separate time series which are then analysed.

In the script, the only decisions to be made is (1) the number of volume classes used (three is recommended), (2) the number of realisations to be generated and (3) the number of disaggregation steps, where in each step the temporal resolution is doubled.

As output, the script generates a file *RCM\_disaggregated\_data.txt* with the desired number of stochastic realisations of time series at the desired higher temporal resolution.

## Applications

Concerning application of the disaggregation model to an existing time series at original resolution *rL*, two principal situations may be considered. One is when some representative data at target resolution *rS* are available, e.g. from a temporary measurement campaign at the same location or from a nearby high-resolution gauge. Then BDCs may be extracted and parameters estimated over the actual resolution interval *rL*≤*r*≤*rS*. The other situation is when no representative high-resolution data are available. Then parameter estimation may still be possible by coarse-graining from resolution *rL* to gradually lower resolutions. Then Equations (1) and (2) may be evaluated and if the scaling laws are found to hold they may be extrapolated to estimate Pr01 and *a* for higher resolutions up to *rS*. This is an attractive feature of scaling-based disaggregation but it must be emphasised that extrapolation is always associated with large uncertainties and that it should always be attempted to verify parameters using some surrogate data.

# Concluding remarks

* Although the scripts are designed to be in principle directly applicable to real data, they are mainly intended as a pedagogical tool and a starting point from which to develop suitable models for specific data sets.
* The approach has proved to work reasonably well in different climates (e.g. Jebari et al., 2012) but the model structure must always be verified for new data sets. Modification may be required to reflect other underlying assumptions about dependencies between variables.
* It is very important to verify that there are enough data for meaningful parameter estimation in all classes and at all scales considered. If not, less parameterised versions can be developed and tested.
* The scripts have been rather carefully tested and evaluated but there is no guarantee that they are totally free from bugs and errors. No responsibility is taken for real-world applications of the scripts.
* In case of questions, problems, bugs, etc., please contact jonas.olsson@smhi.se

# References

Güntner, A., Olsson, J., Calver, A., and B. Gannon (2001) Cascade-based disaggregation of continuous rainfall time series: the influence of climate, Hydrol. Earth System Sci., 5, 145-164.

Jebari, S., Berndtsson, R., Olsson, J., and A. Bahri (2012), Soil erosion estimation based on rainfall disaggregation, J. Hydrol., 436-437, 102-110, doi: 10.1016/j.jhydrol.2012.03.001.

Hingray, B., Ben Haha, M., 2005. Statistical performances of various deterministic and stochastic models for rainfall disaggregation. Atmos. Res., 77, 152-175.

Licznar, P., Łomotowski, J., Rupp, D.E., 2011. Random cascade driven rainfall disaggregation for urban hydrology: An evaluation of six models and a new generator. Atmos. Res., 99(3–4), 563–578.

Molnar, P., Burlando, P., 2005. Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model. Atmos. Res., 77, 137-151.

Olsson, J. (1998) Evaluation of a scaling cascade model for temporal rainfall disaggregation, Hydrol. Earth System Sci., 2, 19-30.