HyetosR: An R package for temporal stochastic simulation of rainfall at fine time scales

P. Kossiris, D. Koutsoyiannis, C. Onof, H. Tyralis, A. Efstratiadis

1. Abstract

A complete software package for the temporal stochastic simulation of rainfall at fine time scales is developed in the R programming environment. This includes several functions for sequential simulation or disaggregation. Specifically, it uses the Bartlett-Lewis rectangular pulse model for rainfall generation and proven disaggregation techniques which adjust the finer scale (hourly) values to obtain the required, coarser scale (daily) values, without affecting the stochastic structure implied by the model. Additionally, a replication scheme is incorporated in order to improve the Bartlett-Lewis model performance, without significant decrease of computational time. Finally, the package includes an enhanced version of the evolutionary annealing-simplex optimization method for the estimation of Bartlett-Lewis parameters. Multiple calibration criteria are introduced, in order to reproduce the statistical characteristics of rainfall at all time scales. This upgraded version of the original Hyetos program (Koutsoyiannis & Onof, 2005) operates on several modes and combinations thereof (depending on data availability), with many options and graphical capabilities. The package, under the name HyetosR, is available free in [http://sites.uoa.gr/soft/hyetos/](http://sites.uoa.gr/soft/hyetos/). This presentation is available online at [http://sites.uoa.gr/soft/hyetos/](http://sites.uoa.gr/soft/hyetos/).

2. Rainfall generation by Bartlett-Lewis model

- Model assumptions (original version, Rodriguez-Iiturbe et al., 1987):
  - Storm origins, \( i \), occur in a Poisson process, with rate \( \lambda \)
  - Cell origins, \( j \), occur in a Poisson process, with rate \( \lambda \)
  - Cell arrivals terminate at \( x_i \) exponentially distributed (parameter \( w \))
  - Cell durations, \( x_i \), exponentially distributed (parameter \( \mu \))
  - Cell intensities, \( a_i \), either exponentially or gamma distributed

- In the modified version (Rodriguez-Iiturbe et al., 1987; Onof & Whetton, 1994), \( n \) is assumed gamma distributed, with scale parameter \( \alpha \) and shape parameter \( \beta \), and varies for each storm event, such that the ratios \( f(t) / \beta \) remain constant.

3. Disaggregation model

- Different sequences of wet days separated by at least one dry day can be assumed independent. This reduces computational time rapidly.
- The BL model runs separately for each cluster of wet days, forming a sequence of storms and cells.
- For each cluster, a departure is calculated by:
  \[ d = \left( \frac{Z_i}{Z} \right)^{\frac{1}{\beta}} \]
  where \( Z \) is the daily sum of simulated fine-scale data and \( Z_i \) is the known total rainfall depth.
- Several runs are performed with initial values lower than an acceptable limit \( d_i \).
- The chosen sequence is adjusted to become fully consistent with the given sequence, according to the proportional adjusting procedure:
  \[ x_i = \left( \frac{i}{\sum x_j} \right) d_i \]
- In the case of very long clusters, the cluster is subdivided into sub-clusters, each treated independently, for details see Koutsoyiannis & Onof, 2001.

4. Parameter estimation

- The modified version of the BL model contains 6 or 7 parameters:
  - For given parameter values, some of the most important theoretical statistical characteristics of rainfall (mean, variance, covariance, probability density) can be analytically computed.
  - On the other hand, the inverse procedure (i.e. the estimation of model parameters for given statistical characteristics) has no analytical solution.
  - This is handled through calibration, seeking to minimize the "distance" between the theoretical and the observed statistics.
  - The calibration problem is inherently multivariate, although it is typically treated as single-objective.
  - In the formulation of the objective function, a number of questions arise:
    - Which are the statistical characteristics to preserve and for which time scales?
    - Which distance metric is the most convenient, for the specific statistical parameter?
    - How are the different metrics combined (e.g. in terms of weighting coefficients) to provide a unique performance measure, i.e. a scalar objective function?
    - Additional constraints are due to the highly non-convex response surface, which makes essential to use advanced optimization algorithms to obtain a robust parameter set, with reasonable computational effort.
    - Usually, a huge number of almost equivalent local optima exist.

5. Software implementation in R

- The model is implemented in R programming language, under the name HyetosR.
- The R is an open source programming language and software environment for interactive data analysis, statistical computation and graphics.
- HyetosR operates on several modes and combinations thereof (depending on data availability), and includes the following functions:
  - `DisagSimList`: with hourly input. The daily initial sequences either is generated by the BL model or is read from a file. The program `DisagSimList` produces the corresponding synthetic hourly series (the entire model performance is tested).
  - `DisagSimCL`: with daily input. Similar to `DisagSimList` but for test function but the input file contains only daily data (no tests for weather). The program `DisagSimList` produces the corresponding synthetic hourly series (the entire model performance is tested).

6. Characteristic features

- The platform allows the user to formulate all aspects of the calibration problem (objective function, parameter bounds, population size, etc.)

7. Evolutionary annealing-simplex (version 3.0)

- Evaporation is based on simplex transformations or mutations.
- All evaluations are based on probabilistic criteria, since a stochastic term is added to the objective function, relative to a "temperature" metric.
- "Temperature" is gradually decreased, on the basis of an adaptive annealing cooling schedule, which controls the degree of randomness within evolution.
- "Utile" transitions are also allowed to escape from local minima.
- The major difference to the version by Koutsoyiannis & Onof (2002) involves the reflection step, which is now implemented through a weighted centroid (from the gradient) instead of the geometrical one.
- This change made the algorithm even an order of magnitude faster.
- The EAS package is also implemented in R ([http://sites.uoa.gr/soft/hyetos/](http://sites.uoa.gr/soft/hyetos/))

8. Case study: Simulation of Athens rainfall

- As a test case, we use the hourly data sets from the National Observatory of Athens (1940-95), for two months with different characteristics (January, June).
- Model parameters were calibrated on theoretical mean, standard deviation and probability dry for 1 and 24 h, and theoretical autocorrelation coefficient of 1 hour.
- The model runs for 1000 years, to generate synthetic hourly rainfall data.
- The statistical characteristics of the synthetic time series were extracted and compared to the historical ones.

9. Reproduction of autocorrelation functions and characteristic probabilities

- The statistical autocorrelation function and characteristic probabilities of the hourly rainfall data were reproduced by the model.

10. Reproduction of distributions of extremes

- Reflective of the samples through its potential, according to the original model (Modus operandi procedure).
- Reflective of the samples through its potential, according to the original model (Modus operandi procedure).

11. Characteristic synthetic rainfall events

- Rainfall events are simulated according to the model parameters and the characteristic probabilities.
- The model reproduces almost all the essential statistical characteristics of the observed data, at both the daily and hourly time scales, and it reproduced the hourly frequency of the highest events.
- The stochastic approach is validated by comparing the empirical pdf of the synthetic rainfall events with the theoretical distribution (i.e. the GEV, with \( k = 0.35 \).
- The empirical pdf of disaggregated extremes fits well the GEV model; thus HyetosR is appropriate for generating design storms in flood studies.
- Further improvement of the parameter estimation problem, towards a multivariate calibration framework, is possible.

12. Conclusions

- The HyetosR package is a fully operational program, used for reconstructing past hourly rainfall on the basis of known daily data (through disaggregation) or for generating synthetic rainfall data at fine time scales.
- In our case study, the model reproduced almost all the essential statistical characteristics of the observed data, at both the daily and hourly time scales, and it reproduced the hourly frequency of the highest events.
- The stochastic approach is validated by comparing the empirical pdf of the synthetic rainfall events with the theoretical distribution (i.e. the GEV, with \( k = 0.35 \).
- The empirical pdf of disaggregated extremes fits well the GEV model; thus HyetosR is appropriate for generating design storms in flood studies.
- Further improvement of the parameter estimation problem, towards a multivariate calibration framework, is possible.