This poster participates in OSP
Outstanding Student Poster Contest

European Geosciences Union General Assembly 2011 Vienna, Austria, 3 - 8 April 2011 Session: HS7.5/NP6.7 Hydroclimatic stochastics

# A computer system for the stochastic disaggregation of monthly into daily hydrological time series as part of a three–level multivariate scheme

Y. Dialynas, S. Kozanis and D. Koutsoyiannis,
Department of Water Resources and Environmental Engineering, School of Civil Engineering, National Technical University of Athens, Greece (www.itia.ntua.gr)


## 1. Abstract

Castalia is a software package (Koutsoyiannis, D., and A. Efstratiadis, A stochastic hydrology framework for the management of multiple reservoir systems, Geophysical Research Abstracts, Vol. 3, European Geophysical Society, 2001) that uses an original two-level multivariate scheme (from annual to monthly time scale) appropriate for preserving the most important statistics of the historical time series and reproducing characteristic peculiarities of hydrological processes such as long-term persistence, periodicity and skewness. A module was developed as an expansion of Castalia, which implements a methodology for the multivariate stochastic simulation and disaggregation of monthly hydrological time series into daily series. This upgraded version of Castalia uses a three-level multivariate scheme that simultaneously preserves the above characteristics for the annual, monthly and daily time scale. Moreover, this module efficiently handles additional difficulties due to peculiarities which frequently appear in daily hydrological series, such as high variation coefficients, high values of skewness and intermittency (preservation of probability dry in rainfall). The computer system was applied for the generation of synthetic hydrological time series within simulation models that are components of a decision support system for hydrosystem management.


## 2. Introduction

- *Castalia* is a software package (Koutsoyiannis and Efstratiadis, 2001) for the stochastic simulation of hydrological series. It uses a two-level multivariate scheme which:

  - Preserves the most important statistics of the historical time series for the annual and monthly time scales:
    1. Mean
    2. Standard deviation
    3. Skewness
    4. Autocorrelation coefficient (lag one)
    5. Cross-correlation coefficient (lag zero)

    (1–3: Marginal statistics; 4–5: Joint statistics)
  - Reproduces characteristic peculiarities of hydrological processes such as:
    - Long-term persistence
    - Periodicity

- The software is used by the Athens Water Supply and Sewage Company (EYDAP) for the optimization of the water supply system in Athens. It has also been used for the management of several hydrosystems in Greece.
- Here we upgraded *Castalia* to implement a three-level multivariate scheme that simultaneously preserves the above characteristics for the annual, monthly and daily time scales.
- Furthermore, this upgraded version of *Castalia* efficiently handles additional difficulties due to peculiarities which appear in daily hydrological series, such as high variation coefficients, high values of skewness and intermittency (preservation of probability dry).

## 3. Generation of daily time series

- The periodic autoregressive model of order 1 (PAR (1)) is used for the generation of initial values of daily series:

$$\mathbf{Y}_{s,\tau} = \mathbf{a}_s \mathbf{Y}_{s,\tau-1} + \mathbf{b}_s \mathbf{V}_{s,\tau}$$

where:

$\mathbf{Y}_{s,\tau} = (Y_{s,\tau}^1, \ldots, Y_{s,\tau}^m)^T$: vector of daily variables with size $m$ (subscripts $s$ and $\tau$ are integer time indices that stand for period (month) and subperiod (day) respectively),

$\mathbf{a}_s$, $\mathbf{b}_s$: ($m \times m$) matrices of parameters (for every period),

$\mathbf{V}_{s,\tau} = (V_{s,\tau}^1, \ldots, V_{s,\tau}^m)^T$: vector of innovations (independent, both in time and space, random variables) with size $m$.

- The estimation of the multivariate PAR(1) model parameters requires the decomposition of covariance matrices $\mathbf{c}_s = \mathbf{b}_s\, \mathbf{b}_s^T$:

$$\mathbf{b}_s\, \mathbf{b}_s^T = \mathrm{Cov}\,[\mathbf{Y}_{s,\tau}, \mathbf{Y}_{s,\tau}] - \mathbf{a}_s\, \mathrm{Cov}\,[\mathbf{Y}_{s,\tau-1}, \mathbf{Y}_{s,\tau-1}]\, \mathbf{a}_s^T$$

A generalized method (Koutsoyiannis, 1999) is implemented for the decomposition of $\mathbf{c}_s$, which is applicable to both positive definite and indefinite matrices.

## 4. Disaggregation methodology (1)

Let $\mathbf{X} = (X^1, \ldots, X^m)^T$ and $\mathbf{Y}_\tau = (Y_\tau^1, \ldots, Y_\tau^m)^T$ be the higher-level (monthly) and lower-level (daily) discrete time processes, respectively ($m$ denotes the number of variables used in the simulation scheme). The higher- and lower-level processes are related by the *additive property*:

$$\sum_{\tau=1}^{s} \mathbf{Y}_\tau = \mathbf{X}$$

where $s$ is the number of fine-scale time steps within each coarse-scale time step.

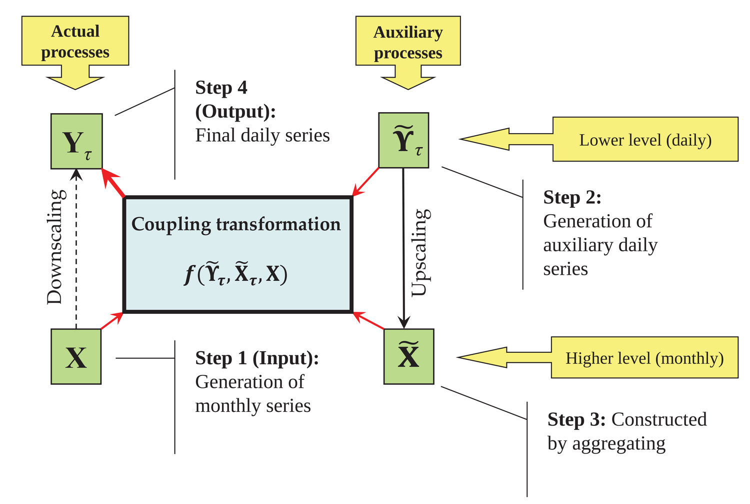



**Figure 1.** Schematic representation of the processes (actual and auxiliary), their links and the steps followed to construct the actual daily-level rainfall series from the actual monthly-level rainfall series (Koutsoyiannis, 2001).

## 5. Disaggregation methodology (2)

- For each period (month), the following *coupling transformation* (Koutsoyiannis, 2001) is applied in the disaggregation scheme:

$$Y_\tau = \tilde{Y}_\tau + \lambda_\tau (X - \tilde{X})$$

where:

$X$ denotes the corresponding synthetic value of the higher-level process,

$\tilde{X} = \sum_{\tau=1}^{s} \tilde{Y}_\tau$, where $s$ is the number of fine-scale time steps within each coarse-scale time step,

$\tilde{Y}_\tau$ stands for the initial daily value generated by the PAR(1) model,

$\lambda_\tau = \frac{1}{n}$ for each subperiod, where $n$ is the number of wet subperiods in each period.
Also, for each period: $\sum_{\tau=1}^{s} \lambda_\tau = 1$

This method does not affect the preservation of the probability dry, as the coupling transformation is only applied to wet subperiods in each period.

- Moreover, a Monte Carlo repetition scheme is applied in order to increase the efficiency of the simulation: for every period several sequences of subperiods are generated via the daily PAR(1) model, until the following value is minimized (Koutsoyiannis and Manetas, 1996):

$$\Delta X = \frac{1}{m} \sum_{l=1}^{m} \frac{\left| X^l - \tilde{X}^l \right|}{\sqrt{Var[X^l]}}$$

where $m$ denotes the number of variables used in the simulation scheme.

## 6. Preservation of skewness

- Hydrological processes at fine time scales have asymmetric distribution functions.
- While the methodology applied in *Castalia* is appropriate for preserving the first and the second moments of the processes, it cannot preserve high values of skewness.
- Therefore, a power transformation is applied to the daily variables for the preservation of skewness (Koutsoyiannis et al., 2003):

$$\mathbf{Y}_t := \mathbf{Y}_t^{(m)}$$

where the symbol ($m$) means that all items of the vector $\mathbf{Y}_t$ are raised to the power $m$ ($0 < m < 1$).

- The preservation of the statistics of the untransformed variables does not necessarily lead to the preservation of the statistics of the transformed variables. However, these discrepancies are expected to be insignificant for high values of $m$ (e.g., for $m \geq 0.5$).

## 7. Preservation of probability dry (1)

- Simulation of hydrological processes (such as rainfall and runoff) at fine time scales requires the preservation of the probability dry.
- The proportion of dry intervals cannot be preserved by linear stochastic models in an explicit manner. Therefore, in *Castalia* the total number of dry periods results according to the following methodology:

1) The negative values, unavoidably generated by any linear stochastic model when the coefficient of variation is high, are truncated to zero (Koutsoyiannis et al., 2003).

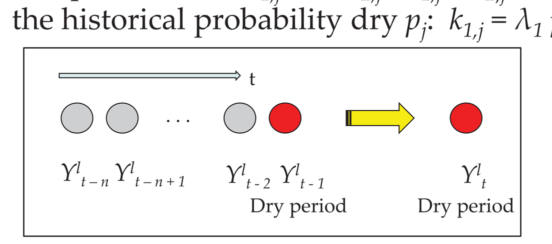
2) A proportion $\pi_0$ of the very small positive values, chosen at random among the generated values that are smaller than a threshold $l_0$ (e.g., 0.1– 0.3 mm), are set to zero (Koutsoyiannis et al., 2003).

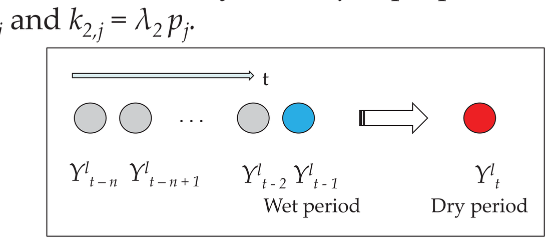
3) An additional method is applied which respects the *Markov property* and takes into consideration dry days in space.

## 8. Preservation of probability dry (2)

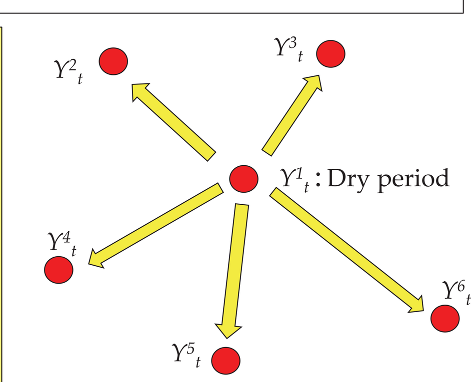
- If a dry period is generated at a time step ($Y^l_{t-1}$), then there is a probability $k_{1,j}$ that it will be followed by another dry period ($Y^l_t$) (*Markov property*).
- On the contrary, if a wet period is generated at a time step ($Y^l_{t-1}$), then there is a smaller probability $k_{2,j}$ that it will be followed by a dry period ($Y^l_t$).

The probabilities $k_{1,j}$ and $k_{2,j}$ ($k_{1,j} > k_{2,j}$) are defined for every month $j$ in proportion of the historical probability dry $p_j$: $k_{1,j} = \lambda_1 p_j$ and $k_{2,j} = \lambda_2 p_j$.




Moreover, for every dry period resulted by this method ($Y^1_t$), the following technique is applied: There is a probability ($k_3$) that dry periods are forced to the rest of the locations at the same time step (e.g., $Y^2_t$ to $Y^n_t$). Thus, the rainfall simulation scheme takes into consideration dry days in space. An overprediction of the cross-correlations is expected, but it should be insignificant for low values of $k_3$ (e.g., for $k_3 \leq 0.5$).




## 9. Methodology implementation

*Castalia* was applied for the generation of daily synthetic series (length: 1000 years), using 43 years (01/01/1964 – 31/12/2006) of daily data series from three rain gauges (Tithorea, Pavlos and Drimea).

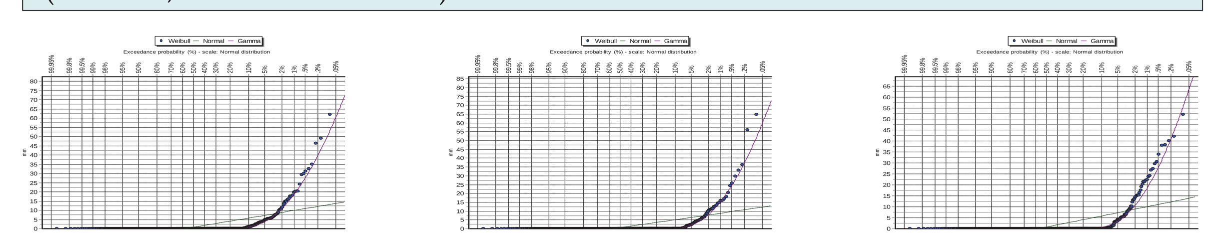



**Figure 2:** Comparison of empirical (blue), theoretical Gamma (purple) and theoretical Normal (green) distribution functions for the historical series at the three gauges (Tithorea, Pavlos and Drimea) for the month of September.

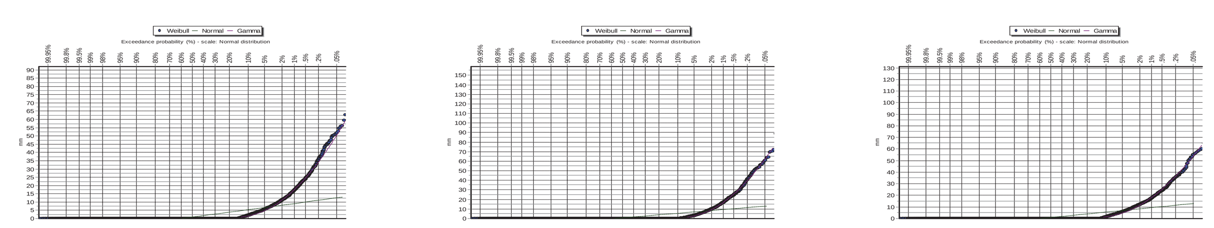
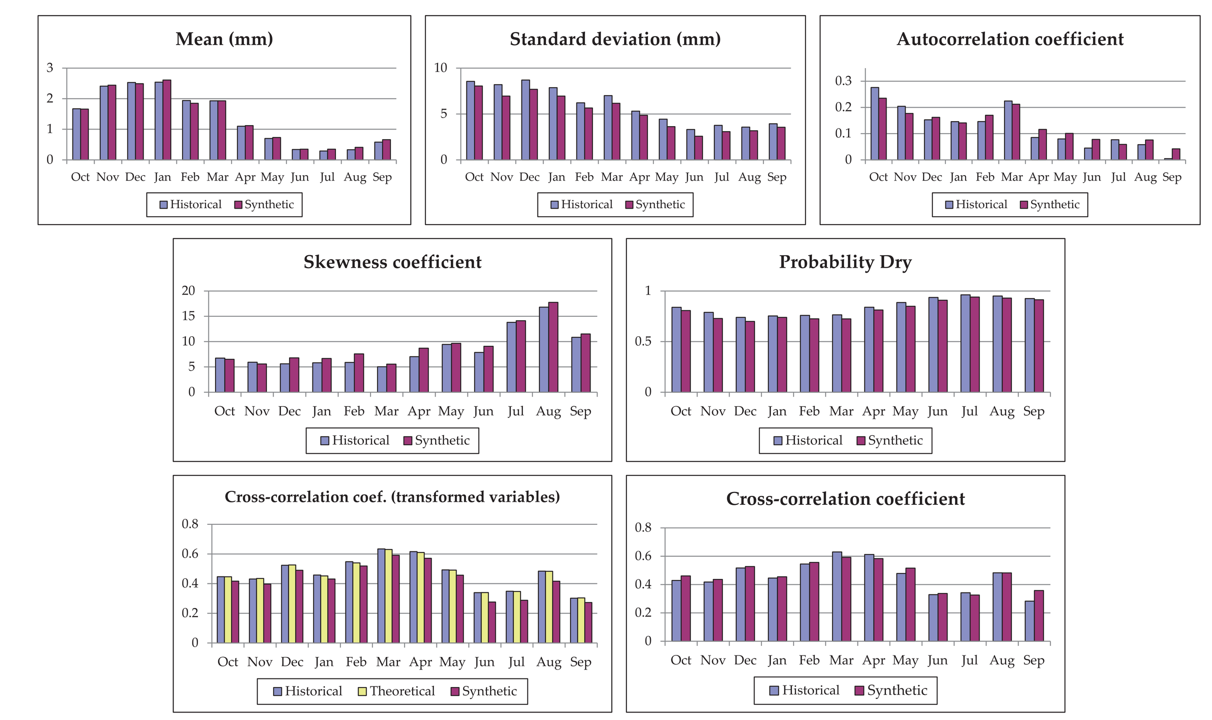



**Figure 3:** Comparison of empirical (blue), theoretical Gamma (purple) and theoretical Normal (green) distribution functions for the synthetic series at the three gauges (Tithorea, Pavlos and Drimea) for the month of September.

The following parameters were used at this implementation of *Castalia*: $m$=0.95 (power transformation), $\pi_0$=0.99 and $l_0$=0.3 mm (threshold method), $\lambda_1$=0.28 και $\lambda_2$=0 and $k_3$=0.60 (Markovian method).

## 10. Results

The following figures show for each month the preservation of the most important daily statistics (mean, standard deviation, autocorrelation coefficient (lag one), skewness coefficient and probability dry) for Pavlos gauge, in addition to the cross-correlation coefficient (lag zero) between Pavlos and Tithorea gauges for the transformed and the untransformed variables.




## 11. Conclusions

- *Castalia* (Koutsoyiannis and Efstratiadis, 2001) uses an original two-level multivariate scheme (from annual to monthly time scale) appropriate for preserving the most important statistics and reproducing characteristic peculiarities of hydrological processes such as long-term persistence, periodicity and skewness.
- Here we developed a module as an expansion of *Castalia*, which implements a methodology for the multivariate stochastic simulation and disaggregation of monthly into daily hydrological series.
- The updated version of *Castalia*:
  1) Simultaneously reproduces the above characteristics for the annual, monthly and daily time scale.
  2) Efficiently handles additional difficulties due to peculiarities which frequently appear in daily hydrological series, such as high variation coefficients, high values of skewness and intermittency.
  3) Can be used in the framework of decision support systems for hydrosystem management.

## 12. References


Efstratiadis, A., Stochastic simulation of hydrological process - The software "Castalia", 18 pages, Department of Water Resources and Environmental Engineering – National Technical University of Athens, January 2011.

Efstratiadis, A., and D. Koutsoyiannis, Castalia (version 2.0) - A system for stochastic simulation of hydrological variables, Modernisation of the supervision and management of the water resource system of Athens, Report 23, 103 pages, Department of Water Resources, Hydraulic and Maritime Engineering – National Technical University of Athens, Athens, January 2004.

Bras, R. L. and Rodriguez-Iturbe, I., *Random functions and hydrology, Addison-Wesley, USA*, 1985.

Koutsoyiannis, D., A generalized mathematical framework for stochastic simulation and forecast of hydrologic time series, *Water Resources Research, 36(6), 1519-1533, 2000.*

Koutsoyiannis, D., Coupling stochastic models of different time scales, *Water Resources Research, 37(2), 379*-392, 2001.

Koutsoyiannis, D., Optimal decomposition of covariance matrices for multivariate stochastic models in hydrology, *Water Resources Research 35(4), 1219-1229, 1999.*

Koutsoyiannis, D., and A. Efstratiadis, A stochastic hydrology framework for the management of multiple reservoir systems, Geophysical Research Abstracts, Vol. 3, European Geophysical Society, 2001.

Koutsoyiannis, D., and A. Manetas, Simple disaggregation by accurate adjusting procedures, *Water Resources Research, 32(7) 2105-2117, 1996.*

Koutsoyiannis, D., C. Onof, and H. S. Wheater, Multivariate rainfall disaggregation at a fine timescale, Water Resources Research, 39 (7), 1173, doi:10.1029/2002WR001600, 2003.

Matalas, N.C. and Wallis, J.R., Generation of synthetic flow sequences, in *Systems approach to water management, A.K. Biswas editor, McGraw Hill, 1976.*

Salas, J. D., Analysis and modeling of hydrologic time series, Chapter 19, *Handbook of Hydrology, edited by D. Maidment, McGraw-Hill, New York, 1993.*

Salas, J. D., Delleur, J. W., Yevjevich, V., and Lane, W. L., *Applied Modelling of Hydrologic Time Series, Water Resources Publications, Littleton, Co., USA, 1988.*


Castalia on the web: http://itia.ntua.gr/en/softinfo/2/