European Geosciences Union General Assembly 2011
Vienna, Austria, 03 – 08 April 2011
Session HS7.5/NP6.7 Hydroclimatic stochastics

Theoretical and empirical comparison of stochastic disaggregation and downscaling approaches for rainfall time series

Federico Lombardo, Elena Volpi
Department of Civil Engineering Sciences,
University of Rome "Roma Tre", Italy
(flombardo@uniroma3.it)

Demetris Koutsoyiannis
Department of Water Resources, Faculty of Civil Engineering,
National Technical University of Athens, Greece

Linking scales

- Both disaggregation and downscaling approaches aim at modelling links among different temporal and/or spatial scales of a given process

Rainfall disaggregation and downscaling

- Disaggregation and downscaling models simulate rainfall fields at a specific scale (higher resolution) given a known precipitation field (measured or simulated) at a certain coarser scale (lower resolution)

- Applications (Koutsoyiannis and Langousis, 2011):
  1. Link global-scale weather prediction models to hydrologic impact studies
  2. Use satellite precipitation estimates for hydrologic purposes
  3. Provide hourly or sub-hourly precipitation data (key for many hydrologic applications) consistent with long historical point rainfall records coming from daily raingauges
  4. Couple several stochastic models to reproduce simultaneously the long-term and the short-term stochastic structure of precipitation

Downscaling *vs* Disaggregation

- **Downscaling** aims at producing a finer scale rain field **Y** with the required statistics, being statistically consistent with the given field **X** at the coarser scale

- **Disaggregation** produces a finer scale rain field **Y** that adds up to the given coarse scale total **X**. Thus, an equality constraint is introduced (**C** is a matrix of coefficients)

$$\mathbf{C} \cdot \mathbf{Y} = \mathbf{X}$$

Comparison of two existing approaches

- **Multifractal approach**: based on the empirical detection of multifractal scale invariance of rainfall in a finite but practically important ranges of scales (for a detailed review: Veneziano and Langousis, 2010). A simple procedure to construct discrete multifractal fields is based on the concept of multiplicative cascades

- **Hurst-Kolmogorov approach**: based on the observation that: "Although in random events groups of high or low values do occur, their tendency to occur in natural events is greater" (Hurst, 1951). This can be explained by multiple scales of changes within a stationary setting (Koutsoyiannis, 2002)

Definitions

- A natural process $R(t)$ is usually defined in continuous time t, but we observe or study it in discrete time as $R_i^{(\delta)}$, which is the average of $R(t)$ over a fixed time scale δ in discrete time steps i = 1, 2, …

$$R_i^{(\delta)} := \frac{1}{\delta} \int_{(i-1)\delta}^{i\delta} R(t)\, dt$$

- Let $f\delta$ be a time scale larger than δ where f is a positive integer (for convenience δ will be omitted). Then, we can define the aggregated $Z_i^{(f)}$ and the mean aggregated $R_i^{(f)}$ stochastic processes on that time scale as

$$Z_i^{(f)} := \sum_{l=(i-1)f+1}^{if} R_l = f\, R_i^{(f)}$$

e.g., $Z_1^{(f)} = R_1 + \ldots + R_f$; $Z_2^{(f)} = R_{f+1} + \ldots + R_{2f}$

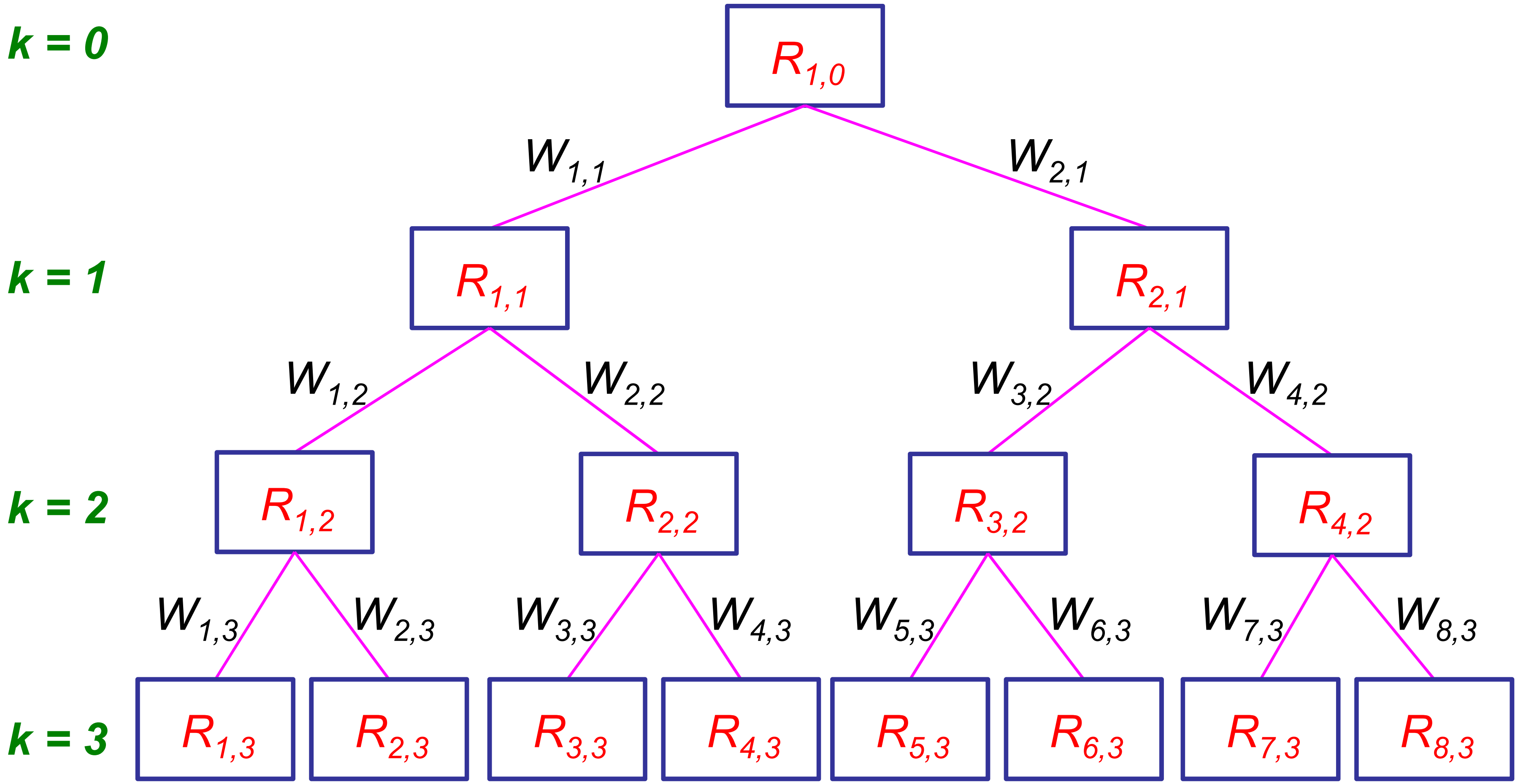
Multiplicative random cascade (MRC)

- Let $R_1^{(f)}$ be the average rainfall intensity averaged over time scale f at the time origin, which is part of a stationary stochastic process in discrete time with mean $\langle R_1^{(f)} \rangle = \mu_0$ and variance $\mathrm{var}[R_1^{(f)}] = \sigma_0^2$ (Gaume et al., 2007)
- $R_1^{(f)}$ (for convenience $R_{1,0}$) is multiplied by b different weights W so as to be distributed over b sub-scale steps of equal size $\Delta s = f/b$ ($R_j^{(\Delta s)}$; $j = 1, \ldots, b$)
- W are identically distributed for all scales with mean μ_W and variance σ_W^2
- After repeating this procedure k times (k cascade levels), the resulting discrete random field at the time scale $\Delta s_k = b^{-k} f$ is of the form

$$R_j^{(\Delta s_k)} = R_{j,k} = R_{1,0} \prod_{i=0}^{k} W_{f(i,j),i}$$

- $j = 1, 2, \ldots, b^k$ is the position in the series at level k
- i is the cascade level
- $f(i,j) = \left\lceil \frac{j}{b^{k-i}} \right\rceil$ denotes a ceiling function which defines the position in the series at the level i
- for $i = 0$, $W_{f(i,j),0} := 1$

Dyadic cascade (b = 2)



$$R_{3,3} = W_{3,3} W_{2,2} W_{1,1} R_{1,0}$$

Kinds of MRCs

- For a *canonical* MRC (<u>downscaling</u>) the expected value of the mean field at the level k equals the expected value of the field at the initial level

$$\left\langle \frac{1}{b^k} \sum_{j=1}^{b^k} R_{j,k} \right\rangle = \langle R_{1,0} \rangle$$

The weights W are statistically independent and satisfy $\langle W \rangle = \mu_W = 1$

- For a *microcanonical* MRC (<u>disaggregation</u>) the mean field at the level k equals the field at the largest scale, which means that for every pair of successive aggregation levels ($k-1$ and k) of the cascade we have

$$\frac{1}{b} \sum_{i=b(j-1)+1}^{b \cdot j} R_{i,k} = R_{j,k-1}$$

Thus, the weights W are dependent random variables which satisfy $\mu_W = 1$ and $W < b$

Summary statistics

- Mean

$$\left\langle R_{j,k}\right\rangle = \left\langle R_{1,0}\prod_{i=0}^{k} W_{f(i,j),i}\right\rangle = \left\langle R_{1,0}\right\rangle \prod_{i=0}^{k}\left\langle W_{f(i,j),i}\right\rangle = \left\langle R_{1,0}\right\rangle\left\langle W\right\rangle^{k} = \mu_0\mu_W^k = \mu_0 = \left\langle R_k\right\rangle$$

where $\mu_W = 1$

- *q*-moment

$$\left\langle R_{j,k}^q\right\rangle = \left\langle R_{1,0}^q\prod_{i=0}^{k} W_{f(i,j),i}^q\right\rangle = \left\langle R_{1,0}^q\right\rangle \prod_{i=0}^{k}\left\langle W_{f(i,j),i}^q\right\rangle = \left\langle R_{1,0}^q\right\rangle\left\langle W^q\right\rangle^{k} = \left\langle R_k^q\right\rangle$$

- Second moment

$$\left\langle R_{j,k}^2\right\rangle = \left\langle R_{1,0}^2\right\rangle\left\langle W^2\right\rangle^{k} = \left(\mu_0^2+\sigma_0^2\right)\left(1+\sigma_W^2\right)^k = \left\langle R_k^2\right\rangle$$

- Variance

$$\sigma_{j,k}^2 = \left\langle R_{j,k}^2\right\rangle - \left\langle R_{j,k}\right\rangle^2 = \left(\mu_0^2+\sigma_0^2\right)\left(1+\sigma_W^2\right)^k - \mu_0^2 = \sigma_k^2$$

Autocorrelation function (b = 2)

- Canonical MRC

$$\rho_{j,k}(t)=\frac{\left\langle R_{j,k}R_{j+t,k}\right\rangle-\left\langle R_{j,k}\right\rangle^2}{\sigma_{j,k}^2}=\frac{\left(\mu_0^2+\sigma_0^2\right)\left(1+\sigma_W^2\right)^{h_{j,k}(t)}-\mu_0^2}{\left(\mu_0^2+\sigma_0^2\right)\left(1+\sigma_W^2\right)^k-\mu_0^2}$$

where t is the lag, and $\left\langle R_{j,k}R_{j+t,k}\right\rangle=\left\langle R_{1,0}^2\right\rangle\left\langle W^2\right\rangle^{h_{j,k}(t)}$

- Microcanonical MRC

$$\rho_{j,k}(t)=\frac{\left\langle R_{j,k}R_{j+t,k}\right\rangle-\left\langle R_{j,k}\right\rangle^2}{\sigma_{j,k}^2}=\frac{\left(\mu_0^2+\sigma_0^2\right)\left(1+\sigma_W^2\right)^{h_{j,k}(t)}\left(1-\sigma_W^2\right)-\mu_0^2}{\left(\mu_0^2+\sigma_0^2\right)\left(1+\sigma_W^2\right)^k-\mu_0^2}$$

where, $\left\langle R_{j,k}R_{j+t,k}\right\rangle=\left\langle R_{1,0}^2\right\rangle\left\langle W^2\right\rangle^{h_{j,k}(t)}\left(2-\left\langle W^2\right\rangle\right)$

The exponent $h_{j,k}(t)$

- The exponent $h_{j,k}(t)$ (at the position $j = 1, \ldots, 2^k - t$ in the cascade at level k) denotes the number of vertices of the tree belonging to both paths leading to the vertices $R_{j,k}$ and $R_{j+t,k}$ from the start vertex $R_{1,0}$ (not included).

$$h_{j,k}(t) = \begin{cases} \left(h_{j,k-1}(t)+1\right)\Theta\left[2^{k-1} - j - t\right], & j \leq 2^{k-1}, t > 0 \\ h_{2^k - j - t + 1, k}(t), & j > 2^{k-1}, t > 0 \\ h_{2^k - j + 1, k}(|t|), & j > |t|, t < 0 \end{cases}$$

where $\Theta[n]$ is the discrete form of the Heaviside step function, which is defined as a function of a discrete variable n (integer)

$$\Theta[n] = \begin{cases} 0, & n < 0 \\ 1, & n \geq 0 \end{cases}$$

- The exponent $h_{j,k}(t)$ is bounded in $\left[0, k - 1 - \lfloor \log_2 t \rfloor\right]$ if $0 < t \leq 2^k - 1$, while $h_{j,k}(t=0) = k$, given any j and k

Example for a canonical MRC

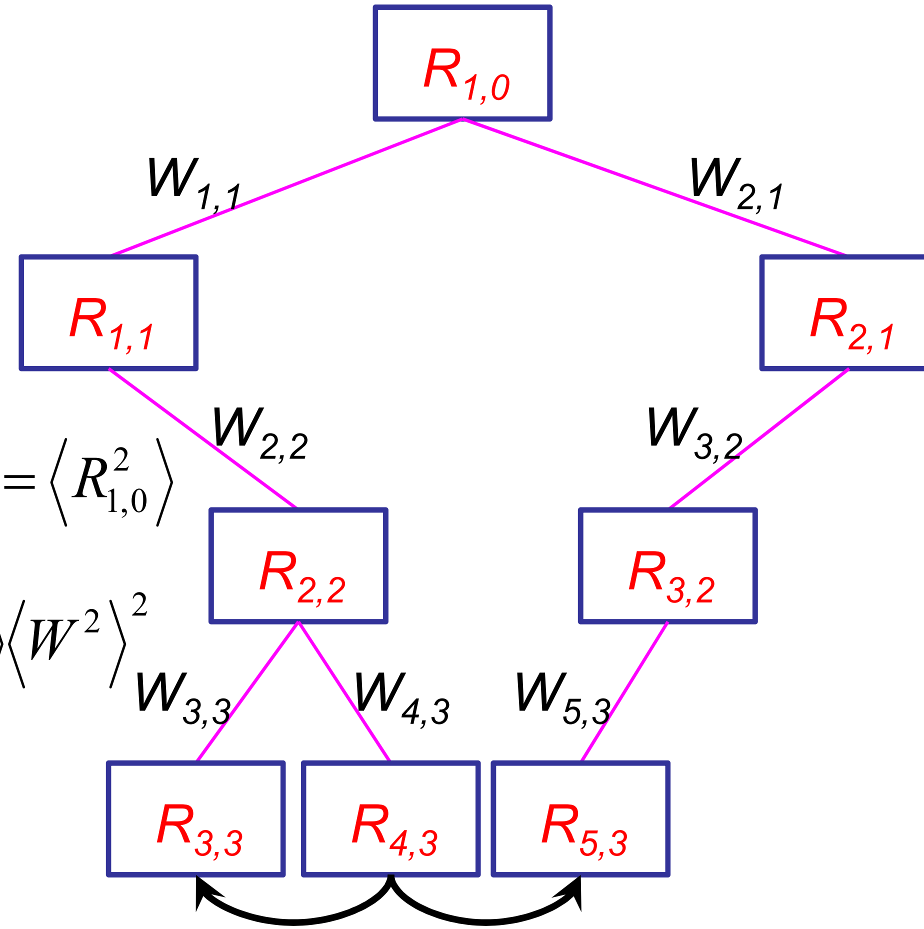
$$\boxed{\left\langle R_{j,k} R_{j+t,k} \right\rangle = \left\langle R_{1,0}^2 \right\rangle \left\langle W^2 \right\rangle^{h_{j,k}(t)}}$$

$h_{4,3}(t=1) = 0$

$$\left\langle R_{4,3} R_{5,3} \right\rangle = \left\langle W_{4,3} W_{2,2} W_{1,1} W_{5,3} W_{3,2} W_{2,1} R_{1,0}^2 \right\rangle = \left\langle R_{1,0}^2 \right\rangle$$

$$\left\langle R_{4,3} R_{3,3} \right\rangle = \left\langle W_{4,3} W_{2,2}^2 W_{1,1}^2 W_{3,3} R_{1,0}^2 \right\rangle = \left\langle R_{1,0}^2 \right\rangle \left\langle W^2 \right\rangle^2$$

$h_{4,3}(t=-1) = 2$



$R_{3,3} = W_{3,3} W_{2,2} W_{1,1} R_{1,0}$ $\quad$ $R_{4,3} = W_{4,3} W_{2,2} W_{1,1} R_{1,0}$ $\quad$ $R_{5,3} = W_{5,3} W_{3,2} W_{2,1} R_{1,0}$

Monte Carlo simulations

- Numerical simulations are used to explore the behaviour of the ensemble statistics of the random field generated by a canonical dyadic MRC

- We assumed unit mean and variance $\sigma_0^2 = \mu_0 = 1$ of the rainrate $R_{1,0}$ at the largest scale

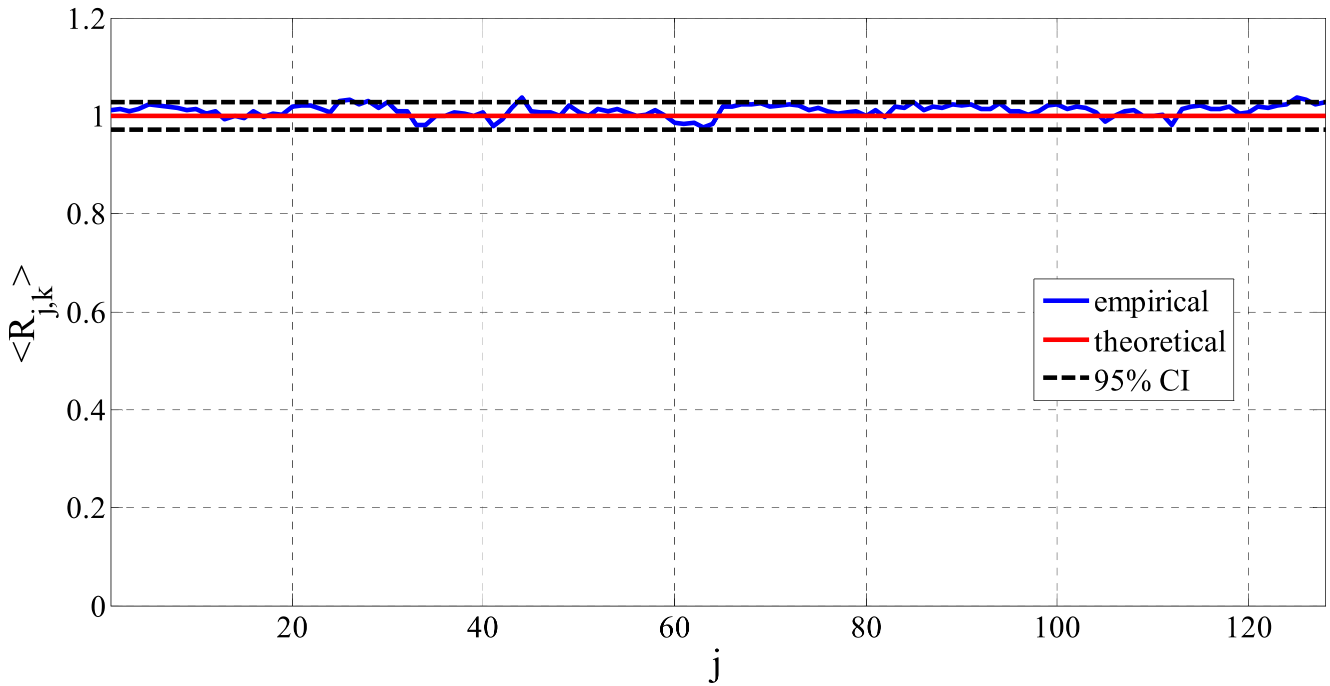
- iid weights W are log-normally distributed (Molnar and Burlando, 2005)

$$W = e^{\alpha Y - \frac{\alpha^2}{2}}$$

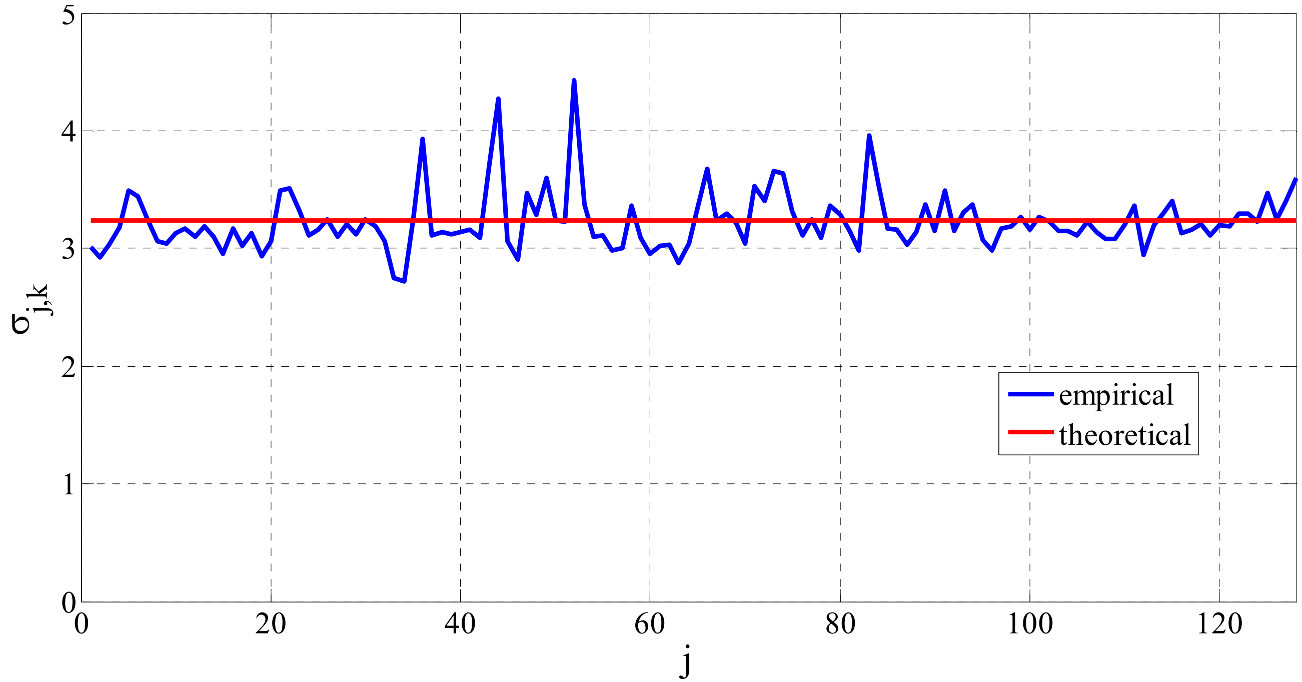
  where Y is a normal $N(0,1)$ random variable and $\alpha = 0.5$ is a parameter

- We assumed k = 7 downscaling levels thus generating time series of length $n = 2^7 = 128$

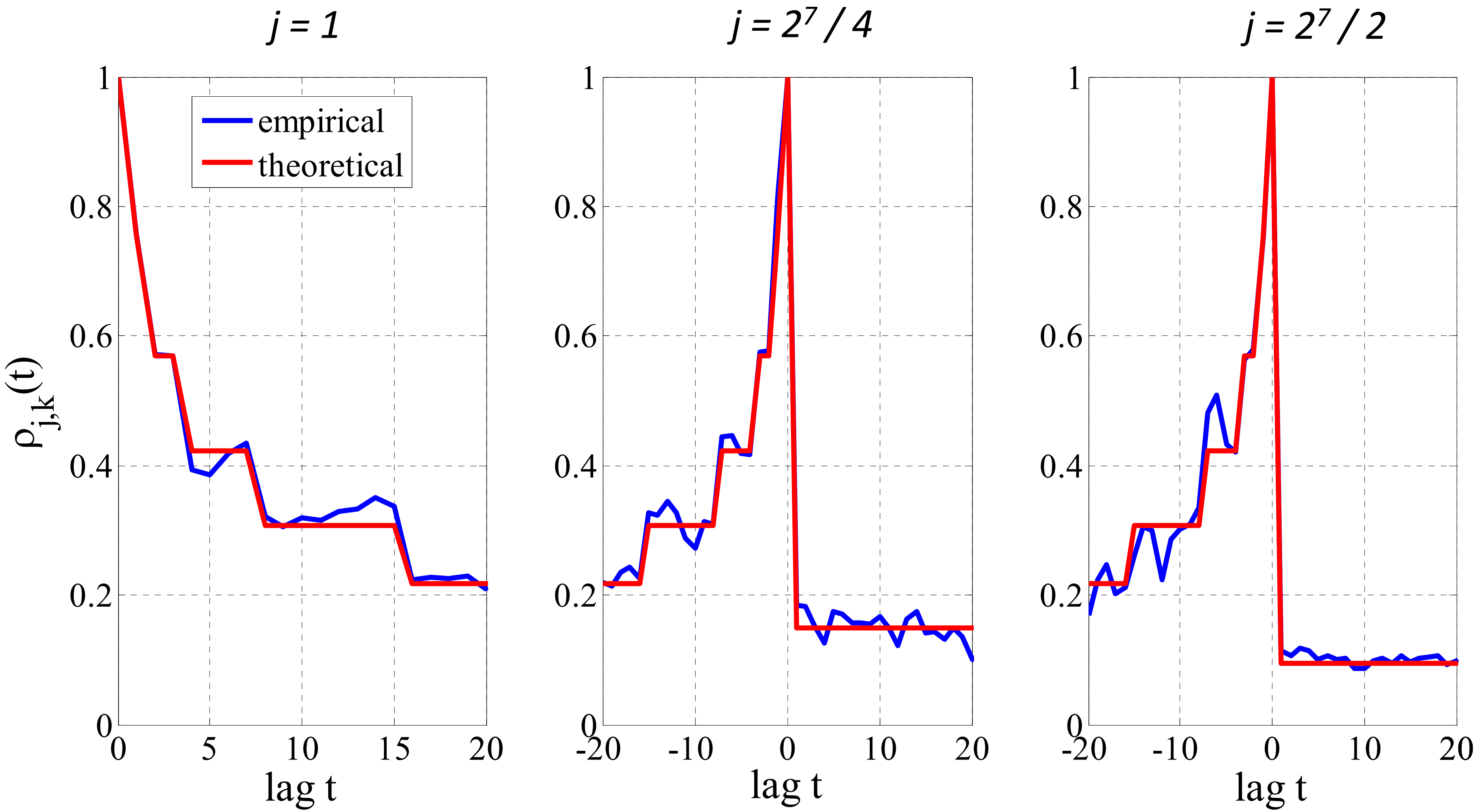
Ensemble Mean



Ensemble Standard Deviation



Ensemble Autocorrelation Function depends on the position *j* in the cascade level *k*



Hurst-Kolmogorov process (HKp)

- The HKp can be defined as a stochastic process which, for any integer *i* and *j* and any time scales *f* and *l*, has the property

$$\left(R_j^{(f)} - \mu\right) =_{\mathrm{d}} \left(\frac{f}{l}\right)^{H\text{-}1} \left(R_i^{(l)} - \mu\right)$$

where $0 < H < 1$ is the Hurst coefficient and R_j is Gaussian

- For the relevant process $Z_j^{(f)}$ the following holds

$$\operatorname{var}\left[Z_j^{(f)}\right] = \left(\sigma^{(f)}\right)^2 = f^{2H}\sigma^2 = f^2 \operatorname{var}\left[R_j^{(f)}\right]$$

- The autocorrelation function of either of $R_i^{(f)}$ and $Z_i^{(f)}$, for any aggregated time scale *f*, is a only function of the lag *t* and *H* (Koutsoyiannis, 2002)

$$\operatorname{corr}\left[Z_j^{(f)}, Z_{j+t}^{(f)}\right] = \rho_t^{(f)} = \rho_t = \frac{|t+1|^{2H}}{2} + \frac{|t-1|^{2H}}{2} - |t|^{2H}$$

Disaggregation approach

- Let $Z_1^{(f)}$ be a Gaussian random variable of the aggregated HKp at the largest scale of interest f, which is to be disaggregated by a dyadic cascade

- $Z_1^{(f)}$ (for convenience $Z_{1,0}$) is partitioned into $b = 2$ variables on the time scale $\Delta s = f/2$, i.e. the first cascade level ($k = 1$)

$$Z_{1,1} + Z_{2,1} = Z_{1,0}$$

- Likewise, at the cascade level k we have

$$Z_{2j-1,k} + Z_{2j,k} = Z_{j,k-1}$$

- Thus, it suffices to generate $Z_{2j-1,k}$ and then obtain $Z_{2j,k}$ from the equation above

- A linear generation scheme is used, which preserves autocorrelations with two earlier lower-level variables (level k) and one later higher-lever variable (level k–1) (Koutsoyiannis, 2002)

Generation procedure

- In each disaggregation step the first lower-level variable $Z_{2j-1,k}$ is generated

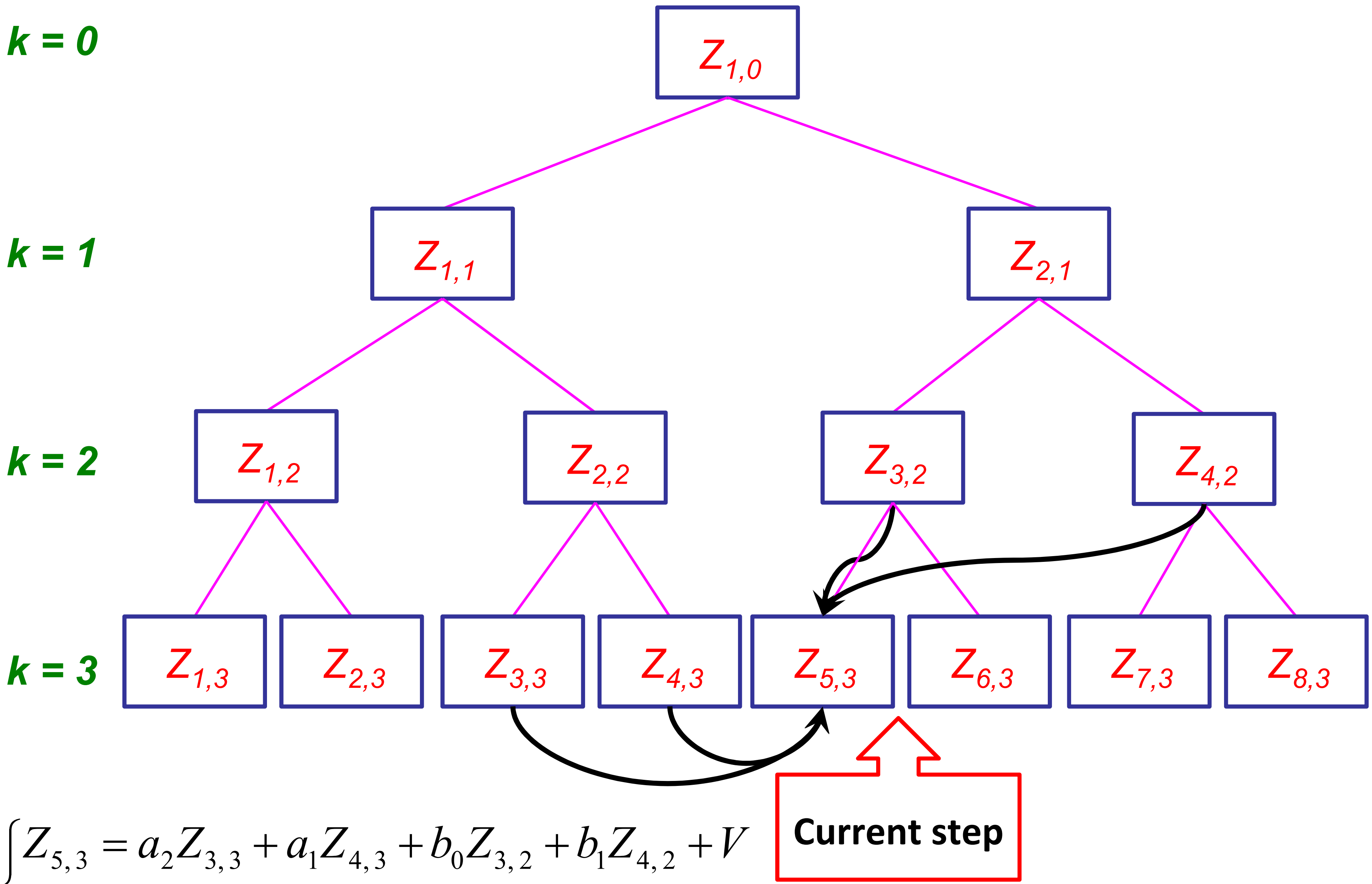
$$Z_{2j-1,k} = a_2 Z_{2j-3,k} + a_1 Z_{2j-2,k} + b_0 Z_{j,k-1} + b_1 Z_{j+1,k-1} + V$$

hence, the second one is $Z_{2j,k} = Z_{j,k-1} - Z_{2j-1,k}$

- Parameters a_2, a_1, b_0 and b_1 and the variance of the innovation term V are estimated in terms of correlations ρ_t, which are independent of j and k, and the variance of the HKp at the level k, σ_k^2 (Koutsoyiannis, 2001)

$$\begin{bmatrix} a_2 \\ a_1 \\ b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_1 + \rho_3 & \rho_4 + \rho_5 \\ \rho_1 & 1 & \rho_1 + \rho_2 & \rho_3 + \rho_4 \\ \rho_1 + \rho_3 & \rho_1 + \rho_2 & 2(1+\rho_1) & \rho_1 + 2\rho_2 + \rho_3 \\ \rho_4 + \rho_5 & \rho_3 + \rho_4 & \rho_1 + 2\rho_2 + \rho_3 & 2(1+\rho_1) \end{bmatrix}^{-1} \begin{bmatrix} \rho_2 \\ \rho_1 \\ 1 + \rho_1 \\ \rho_2 + \rho_3 \end{bmatrix}$$

$$\mathrm{var}[V] = \sigma_k^2 \left(1 - [\rho_2, \rho_1, 1+\rho_1, \rho_2+\rho_3][a_2, a_1, b_0, b_1]^T\right)$$



$$\begin{cases} Z_{5,3} = a_2 Z_{3,3} + a_1 Z_{4,3} + b_0 Z_{3,2} + b_1 Z_{4,2} + V \\ Z_{6,3} = Z_{3,2} - Z_{5,3} \end{cases}$$

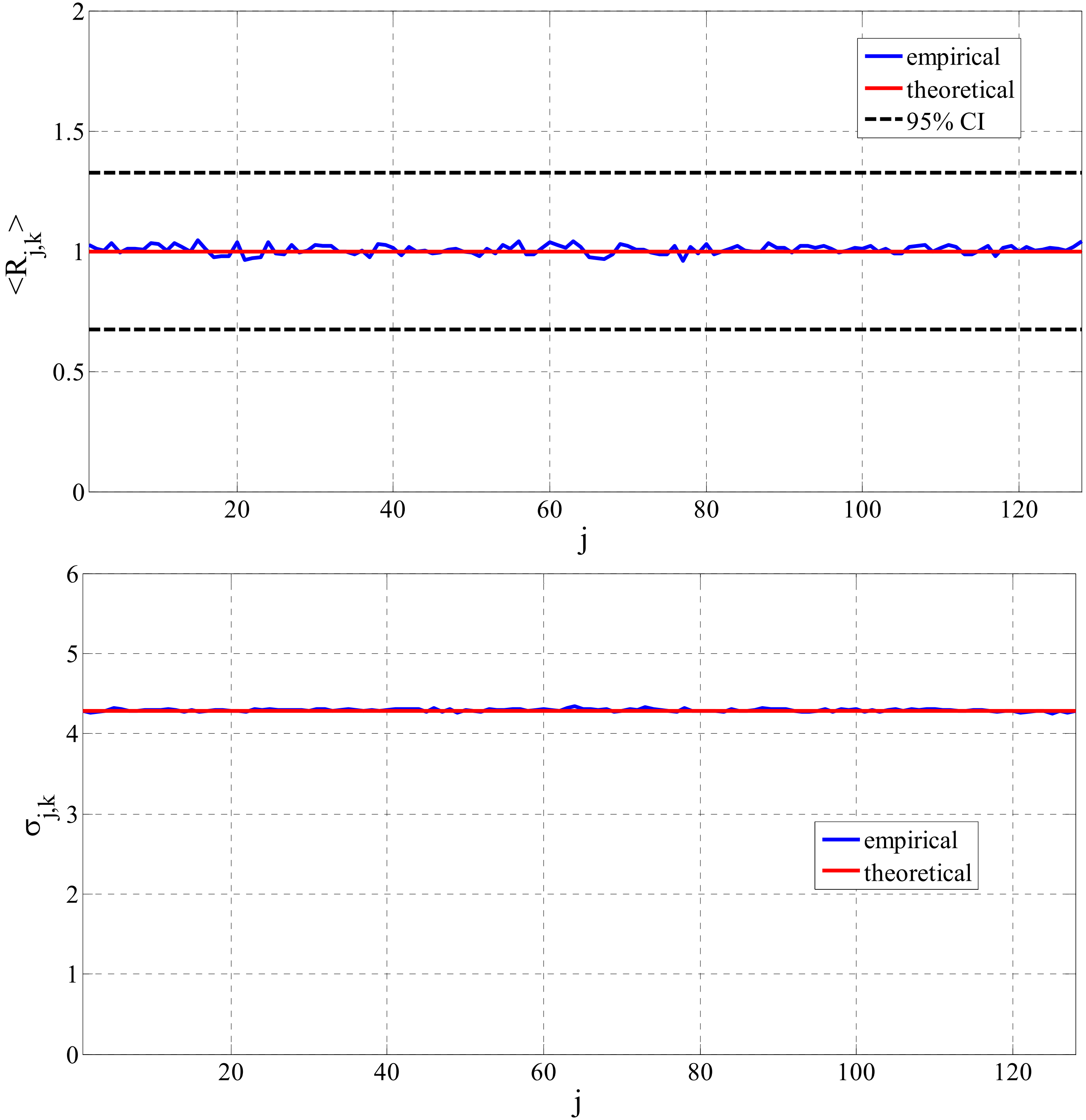
Ensemble Mean

Simulations:

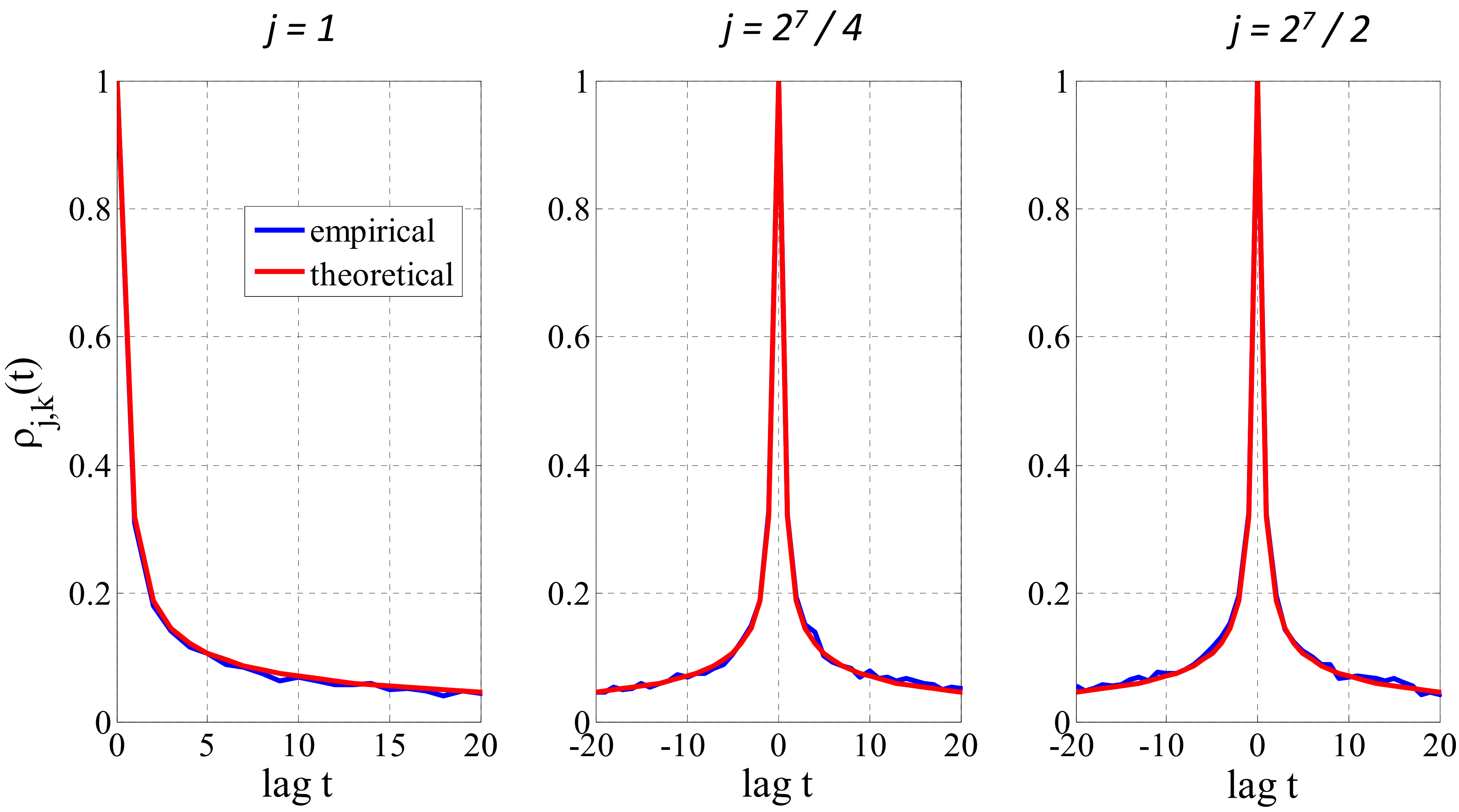
***H* = 0.7 and**

$\mu_0 = \sigma_0 = 1$

Ensemble Standard Deviation



Ensemble Autocorrelation Function independent of the position *j* in the cascade level *k*, and symmetric



Conclusions

- Structures of two simple and widely used approaches of stochastic disaggregation and downscaling for rainfall time series are compared by means of theoretical reasoning and Monte Carlo experiments
- Autocorrelograms produced by multiplicative random cascade (MRC) models seem to have a physically unrealistic attitude to the rainfall process
- We started assuming a stationary setting of the entire process at the largest scale, then we concluded with a downscaled process that we demonstrated to be non-stationary
- The other stepwise disaggregation approach effectively generates Gaussian time series that respect the Hurst phenomenon
- However, observed rainfall time series (especially at the resolution needed for hydrologic applications) are not Gaussian; thus they must be normalized to estimate the model parameters. This could be not always easy especially for high resolution data sets with high intermittency.

References

- Gaume, E., N. Mouhous, and H. Andrieu, Rainfall stochastic disaggregation models: Calibration and validation of a multiplicative cascade model. *Advances in Water Resources*, 30, 1301–1319, 2007.
- Hurst, H.E., Long term storage capacities of reservoirs, *Trans. ASCE*, 116, 776–808, 1951.
- Koutsoyiannis, D., Coupling stochastic models of different time scales, *Water Resources Research*, 37 (2), 379–392, 2001.
- Koutsoyiannis, D., The Hurst phenomenon and fractional Gaussian noise made easy, *Hydrological Sciences Journal*, 47 (4), 573–595, 2002.
- Koutsoyiannis, D., and A. Langousis, Precipitation, *Treatise on Water Science*, edited by P. Wilderer and S. Uhlenbrook, 2, 27–78, Academic Press, Oxford, 2011.
- Molnar, P., and P. Burlando, Preservation of rainfall properties in stochastic disaggregation by a simple random cascade model, *Atmospheric Research*, 77, 137–151, 2005.
- Veneziano, D., and A. Langousis, Scaling and fractals in hydrology, *Advances in Data-based Approaches for Hydrologic Modeling and Forecasting*, edited by B. Sivakumar, World Scientific, Chapter 4, 145p., 2010.