The quest for consistent representation of rainfall and realistic simulation of process interactions in flood risk assessment (1)

EGU General Assembly 2010, Vienna, Austria, 2 – 7 May 2010

Session HS5.1/AS1.20/NH1.11/NP3.6: *Precipitation: from measurement to modelling and application in catchment hydrology*

A. Efstratiadis and S. M. Papalexiou

Department of Water Resources and Environmental Engineering, National Technical University of Athens, Greece

1. Abstract

We present a methodological framework for flood risk assessment in medium and large-scale basins, characterized by the complexity of processes (both physical and anthropogenic) and the lack of detailed hydrological and hydraulic data. Evidently, to evaluate the probability of extreme floods, it is necessary to provide both a statistically consistent description of forcing (precipitation) and a realistic simulation of the runoff mechanisms. The proposed methodology comprises (a) an original multivariate stochastic rainfall model to simulate the point precipitation in daily basis, and (b) the conjunctive modelling scheme HYDROGEIOS, linking a hydrological, a hydrogeological and a water management module, which represents the allocation of flows and abstractions (surface and groundwater) across hydrosystems of any complexity. The aforementioned tools were successfully tested in the Boeoticos Kephisos river basin, in Greece, for the generation of synthetic discharge series at characteristic cross-sections, and the evaluation of the corresponding flood risk.

2. Shortcomings of typical flood modelling practices

Most engineering studies are based on outdated or simplistic rules-of-thumb, where flood modelling is addressed through deterministic design storms and "event-based" hydrological tools, which ignore significant processes and interactions. Evidently, there are several shortcomings in such approaches, especially when employed to large-scale hydrosystems.

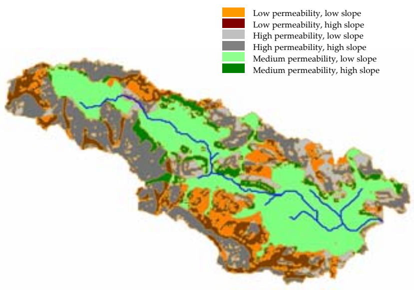
In rainfall modelling, the conventional methodology for constructing design storms fails to properly represent the variability of precipitation, given that the **temporal and spatial correlations** of the historical records are not represented. For instance, it is generally assumed that the input storms to all sub-basins derive from a "representative" intensity-duration-frequency relationship (ombrian curve), where the partial depths for all time resolutions correspond to the same return period. Besides, the spatial variability of precipitation, which is a key characteristic of the flood regime of a river basin, is represented through over-simplified approaches, such as the areal reduction factor, which reduces the estimated rainfall depths by a constant ratio, depending on the area of the upstream basin.

On the other hand, most of the widely-used flood models do not simulate (in space and time) the entire hydrological cycle but rather transform rainfall to runoff in a black-box way, using semi-empirical tools such as the SCS curve number method, synthetic unit hydrographs, etc. However, this prohibits for **interpreting flood risk as joint probabilities of all hydrological variables** that interrelate in runoff generation (rainfall, stream-aquifer interactions, soil moisture accounting). In addition, the model parameters are either estimated on the basis of regional formulas of limited reliability (since they derive from experimental basins) or calibrated against normally few historical flood events, which is at least questionable, given that fitting data is little representative of the overall catchment mechanisms.

3. Study area and input data

General information

The Boeoticos Kephisos river basin drains a formerly closed area of 1850 km^2, whose flows are conducted to the neighbouring Lake Yliki. Due to the dominance of highly-permeable geologic formations, almost half of the runoff derives from karst springs, which rapidly contribute to the streamflow, in contrast to the unusually low contribution of direct (flood) runoff. In addition, due to the combined abstractions from surface and groundwater recourses and the existence of an artificial drainage network in the lower part of the basin (where slopes are noticeably low), the system is heavily modified.



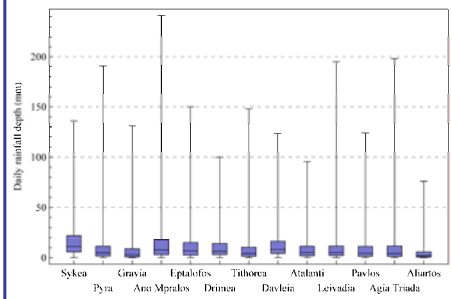
River network and hydrological response units (HRUs), classified according to permeability and slope

Data for flood modelling

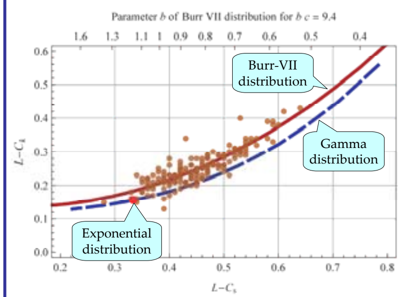
- Point rainfall data at 13 stations (records starting from 1955 to 1969 and ending at 2006, in daily basis);
- Mean daily discharge series (derived from once-a-day stage data) at the basin outlet, from 1977 to 2003;
- Sparse albeit systematic (1-2 per month) flow measurements along the river course and downstream of the main karst springs (dense sample during 1984-1990);
- Lack of detailed hydraulic data (cross-section geometry, stage-discharge curves), apart from the basin outlet.

Hydrometric (cycles) and rainfall (diamonds) stations

4. Analysis of historical daily rainfall time series



Box-plots of wet-day daily rainfall in the 13 stations of the basin. The analysis revealed some similarities regarding the major statistics, e.g., the mean value fluctuates around 10 mm ± 2 mm, while the standard deviation fluctuates around 13 mm ± 2 mm. Of course, the variability of those statistics on the monthly basis is radically increased.



Empirical sample (brown dots) of the wet-day daily rainfall per month (13 × 12 points in total), depicted in *L*-ratio diagram. The Gamma distribution (blue dashed line) exhibits an exponential tail, and is the most common probability model for describing daily rainfall. The solid brown line corresponds to a power-type distribution function (Burr type VII), of the form:

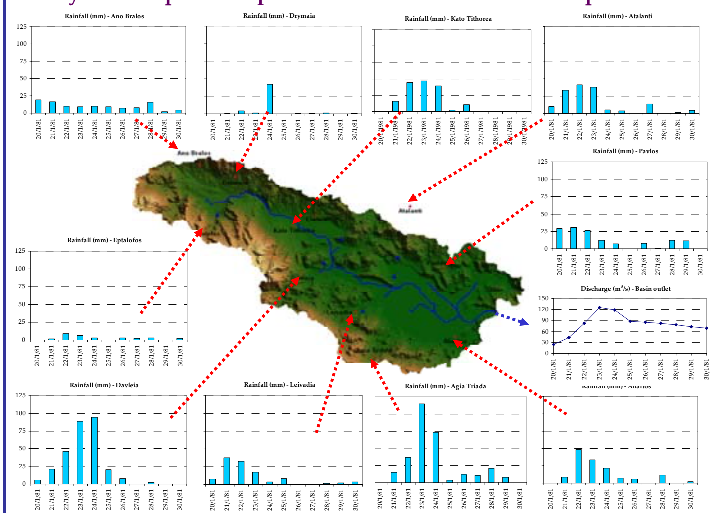
$$F_X(x) = 1 - [1 + (x/a)^b]^c$$

that is able to generate frequently and severe rainfalls (compared to an exponential-tail model). The distribution fits well the empirical points, while is by far superior than the Gamma model.

The table presents the monthly average values, resulted from all stations, of the probability dry P_d, and the auto- and cross-correlation coefficients, $\rho_A(\tau)$ and $\rho_C(\tau)$ respectively, for lag τ. We remark that the point values vary significantly from station to station.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
$\overline{P}_d$	0.66	0.64	0.68	0.76	0.82	0.89	0.93	0.92	0.88	0.77	0.70	0.63
$\overline{\rho}_A(1)$	0.28	0.21	0.24	0.21	0.18	0.13	0.08	0.05	0.13	0.35	0.28	0.25
$\overline{\rho}_A(2)$	0.08	0.06	0.02	0.03	0.03	0.04	0.02	0.02	0.01	0.09	0.08	0.05
$\overline{\rho}_C(0)$	0.44	0.43	0.46	0.49	0.42	0.37	0.35	0.45	0.37	0.55	0.52	0.45
$\overline{\rho}_C(1)$	0.24	0.2	0.22	0.23	0.22	0.15	0.12	0.11	0.17	0.33	0.25	0.21

5. Why are the spatio-temporal correlations of rainfall so important?



Observed rainfall depths and outlet runoff during 20-30 January 1981

The quest for consistent representation of rainfall and realistic simulation of process interactions in flood risk assessment (2)

EGU General Assembly 2010, Vienna, Austria, 2 – 7 May 2010

Session HS5.1/AS1.20/NH1.11/NP3.6: *Precipitation: from measurement to modelling and application in catchment hydrology*

A. Efstratiadis and S. M. Papalexiou

Department of Water Resources and Environmental Engineering, National Technical University of Athens, Greece

6. Stochastic rainfall modelling

A consistent multivariate stochastic rainfall model should reproduce: (a) the marginal distribution at every location and every month, which includes not only the wet-day daily rainfall distribution (continuous) but also the probability dry, and (b) the auto- and cross-correlation structure. Briefly, our method comprises: (a) a Multivariate Auto Regressive (MAR) model to generate normal multivariate cyclostationary random variables in order to preserve the auto- and cross-correlation structure as close as possible, and (b) a general technique for transforming the normal variables to variables exhibiting the rainfall characteristics at each station and month.

If we symbolize $Z_{i,j}(t)$ the normalized random variable at day t, at the i-th station for month j, where $i = 1,\dots, m$ and $j = 1,\dots,12$, then the MAR1 model is defined as $\mathbf{Z}^T_j(t) = \mathbf{A}_j\, \mathbf{Z}_j(t-1) + \mathbf{B}_j\, \boldsymbol{\varepsilon}_j(t)$, where $\mathbf{Z}^T_j(t) = [Z_{1,j}(t), \dots, Z_{m,j}(t)]$, with $Z_{i,j}(t) \sim N(0, 1)$, $\boldsymbol{\varepsilon}^T_j(t) = [\varepsilon_{1,j}(t), \dots, \varepsilon_{m,j}(t)]$ is the vector of innovations, with $\varepsilon_i(t) \sim N(0, 1)$, and A_j and B_j are the $m \times m$ parameter matrices (here $m = 13$ stations), with j depending on the month that day t belongs. Clearly, 12 different parameter matrices $\mathbf{A}_j$ and $\mathbf{B}_j$ are to be estimated.

The normal random variables were denormalized using the transformation:

$$x = g^{-1}_{Z_{i,j}}(z) = \begin{cases} 0, & z \le z_{p_{i,j}} \\ a\left\{\left[\exp\left[(z-d)^2\right]-1\right]^{\frac{1}{c}}-1\right\}^{\frac{1}{b}}, & z > z_{p_{i,j}} \end{cases} \qquad g^{-1}: \left(d+\sqrt{\ln 2}, \infty\right) \to \mathrm{R}^+$$

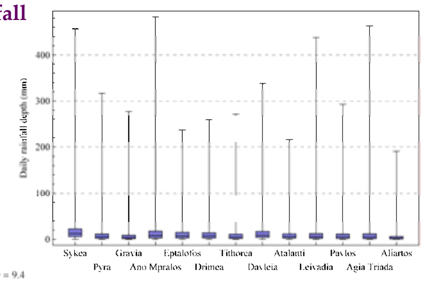
where $g^{-1}_{Z_{i,j}}(z)$ is the inverse normalizing transformation related to the i-th station of month j, and $p_{i,j}$ is the corresponding probability dry. The parameters of the transformation were estimated for every station and month using the corresponding normalizing transformation

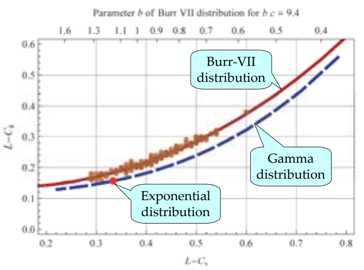
$$z = g(x) = d + \sqrt{\ln\left[1+\left(1+\frac{x}{a}\right)^{b}\right]^{c}} \qquad g: \mathrm{R}^+ \to \left(d+\sqrt{\ln 2}, \infty\right)$$

so that the historical records were transformed to standardized normal variables.

7. Analysis of synthetic rainfall

Box-plots of 1000-year synthetic wet-day daily rainfall in the 13 stations of the basin. The graphs are similar to the historical ones (panel 4), except for the magnitude of the extremes. However, this was expected, since the length of the historical records is only 43 years, while the length of the simulated ones is much greater (1000 years).





Empirical sample (brown dots) of the wet-day daily synthetic rainfall per month, depicted in L-ratio diagram. The empirical points fit almost perfectly the theoretical curve of the Burr type VII distribution. The lower variance around the theoretical curve, compared to the observed data, is also due to the length of the simulated time series, which is 20 times greater than the length of the historical ones.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
$\bar{P}_d$	0.66	0.64	0.68	0.77	0.82	0.89	0.93	0.92	0.87	0.77	0.70	0.62
$\bar{\rho}_A(1)$	0.22	0.24	0.2	0.19	0.18	0.12	0.05	0.05	0.11	0.25	0.3	0.27
$\bar{\rho}_A(2)$	0.07	0.08	0.06	0.07	0.08	0.05	0.04	0.02	0.05	0.1	0.11	0.08
$\bar{\rho}_C(0)$	0.42	0.45	0.46	0.45	0.44	0.36	0.34	0.34	0.41	0.48	0.51	0.48
$\bar{\rho}_C(1)$	0.19	0.21	0.2	0.21	0.2	0.14	0.1	0.1	0.16	0.25	0.26	0.23

The model reproduces almost perfectly the probability dry and very satisfactorily the auto- and cross-correlations. Clearly, these are mean values, while the individual values for each station exhibit significant variability.

Acknowledgments – Contact info

The research of this paper was performed within the project *Flood risk estimation and forecast using hydrological models and probabilistic methods*. The HYDROGEIOS model (version 2.0) was developed within the project *Development of Database and software applications in a web platform for the "National Databank for Hydrological and Meteorological Information"*.

E-mail contact: andreas@itia.ntua.gr Software web page: http://www.hydroscope.gr/software/hydrogeios.htm/

The presentation is available online at http://www.itia.ntua.gr/en/docinfo/961/

8. The HYDROGEIOS modelling framework

Surface hydrology module (daily time step)

- Semi-distributed schematization;
- Conceptualization through 3 interconnected tanks, representing the surface hydrological processes;
- Model inputs: daily precipitation and potential precipitation (PET) data, varying per sub-basin;
- Model parameters: 7 per hydrological response unit;
- Model outputs: evapotranspiration, percolation and runoff (directly transferred to the sub-basin outlet).

Groundwater module (daily time step)

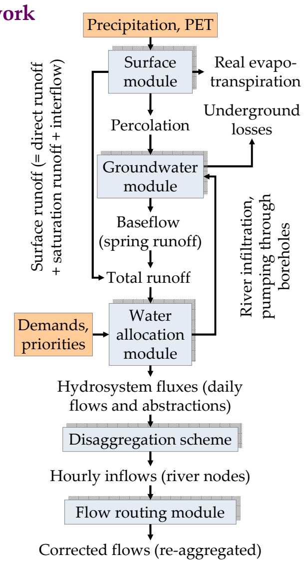
- Finite-volume approach, aquifer discretization to a limited number of polygonal cells of flexible shape;
- Darcian representation of flow field;
- Stress data: percolation, infiltration, pumping;

Water allocation module (daily time step)

- Representation of water uses and main hydraulic structures (aqueducts, boreholes, diversion projects);
- Step-by-step estimation of unknown flows and abstractions through a linear optimization approach, where artificial capacities and unit costs are imposed to preserve constraints and water use priorities.

Flow routing module (hourly time step)

- Construction of hourly-resolved inflow hydrographs, through an empirical disaggregation scheme;
- Flow routing through a kinematic-wave or a Muskingum diffusive-wave model.



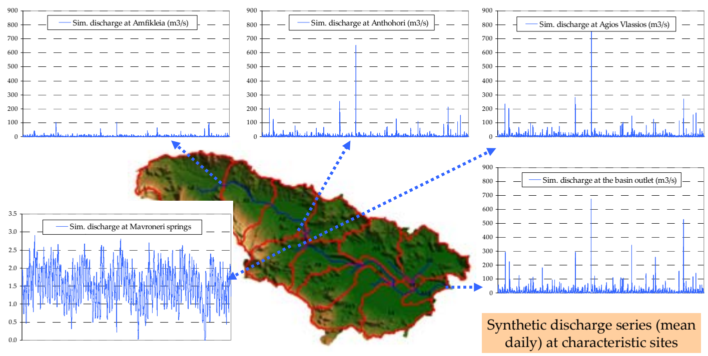
9. Model application for calibration and stochastic simulation

Model calibration (1984-1990)

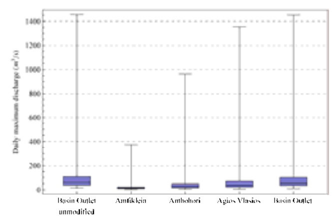
- Optimized parameters ~70 (assigned to river segments, HRUs and groundwater cells);
- Multi-site model fitting (13 flow stations), based on multiple statistical and empirical criteria;
- Hybrid calibration, driven by the evolutionary annealing-simplex method.

Stochastic simulation of hydrosystem fluxes

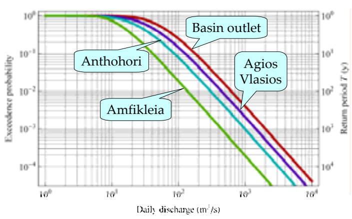
- Spatial integration of 1000-year synthetic rainfall data (daily input data to 13 sub-basins);
- Assumption of the mean monthly potential evapotranspiration for each sub-basin;
- Assumption of constant agricultural demand, assigned to seven conceptual irrigation nodes;



Synthetic discharge series (mean daily) at characteristic sites



Box-plots of simulated discharge (annual maxima)



Theoretical distributions for mean daily discharge