Extended abstract

Introduction

Water resource problems are characterized by the presence of multiple sources of uncertainty. The implementation of Monte Carlo simulation techniques within powerful optimization methods is required, in order to handle these uncertainties. In the framework of the present thesis we investigate how the various sources of uncertainty affect the optimization procedure as well as the various models. Furthermore, we investigate a modified version of the evolutionary annealing-simplex method in global optimization applications, where uncertainty is explicitly considered in terms of stochastic objective functions. We evaluate the algorithm against several benchmark functions, as well as in the stochastic calibration of a lumped rainfall-runoff model (Zygos). In this context, we examine different calibration criteria and different sources of uncertainty, in order to assess not only the robustness of the derived parameters but also the predictive capacity of the models. As one other problem that requires the combined use of optimization and simulation, we examine the applicability of a widely used rainfall model for the case of Athens. Taking advantage of the simulation and optimization functionalities of HyetosR package, we evaluate the performance of two versions of Bartlett-Lewis model in representing the convective and frontal rainfall of Athens. We demonstrate that although these models reproduce the essential statistical characteristics of rainfall at the hourly as well as daily time scales (mean, variance, autocorrelation structure), they fail to preserve important temporal properties, such as the duration and time distance of rainfall events.

Posing the problem of optimization under uncertainty

Uncertainty appears in the majority of the real-world optimization problems, including hydrological. Typical sources are: (a) data uncertainty, due to observation and processing errors; (b) model uncertainty, due to simplified representation of significantly complex systems; (c) parameter uncertainty, due to statistically inconsistent fitting criteria and inefficient calibrations.

The optimization problem under uncertainty can be formulated as:

$$\min f(\mathbf{x}) = \min F(\mathbf{x}, \omega),\ \mathbf{a} < \mathbf{x} < \mathbf{b}$$

where $\mathbf{x}$ is a vector of n control variables, $f(\mathbf{x})$ is the fitness function, ω is a noise component and $F(\mathbf{x}, \omega)$ a random estimate at $\mathbf{x}$. In the case of simulation models, where the system performance f is inferred either from historical or synthetic data samples, ω represents the sampling uncertainty. Uncertainty makes the response surface of the function even rougher, by randomly creating local minima and maxima (Fig. 1).

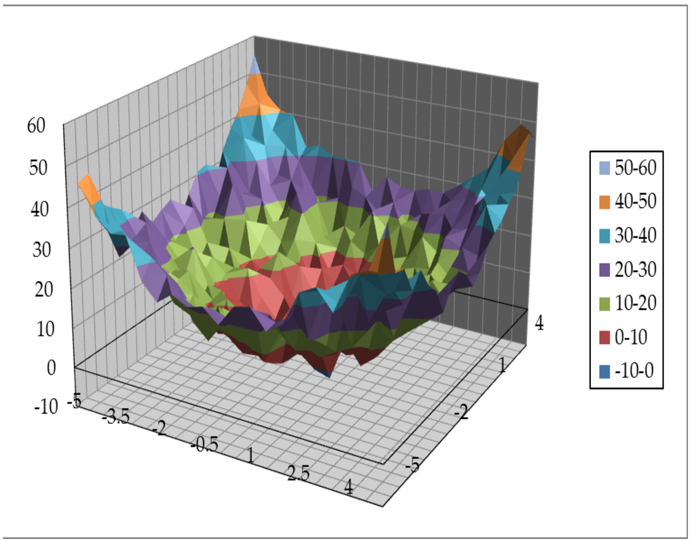

Figure 1: Response surface of noisy sphere function $f(x_1, x_2) = x_1 2 + x_2^2 + N(0, 1)$.

The modified evolutionary annealing-simplex method for stochastic objective function

The evolutionary annealing-simplex method is a heuristic global optimization technique coupling the strength of simulated annealing in rough search spaces with the efficiency of the downhill simplex method (Nelder & Mead, 1965) in smoother spaces (Efstratiadis & Koutsoyiannis, 2002). The key features of the method are:

- an adaptive annealing cooling schedule determines the degree of randomness through the search procedure;
- all transitions are probabilistic, since a stochastic term is added to the objective function, relative to temperature, thus $g(\mathbf{x}) = f(\mathbf{x}) + uT$;
- new points are generated via simplex transformations or mutations;
- all simplex configurations employ quasi-stochastic scale factors;

- multiple expansions and uphill transitions are allowed, in order to accelerate the search and escape from local minima, respectively.

The original version of the above described method was modified to handle noisy objective functions and avoid early convergence to false optima or local minima, due to the dominance of noise. The modifications include:

- **Dynamic adjustment of shrinkage coefficient**, based on the current temperature of the system, *T*, which protects the algorithm from an early degeneration of the simplex (Fig. 2).
- **Re-evaluation of the current best point in the population** after *n* subsequent transformations that reduce the size of the simplex; this ensures that search will not be guided by a point, in which has assigned an erroneously low value, due to noise.
- **Re-annealing of the system** when the temperature, *T*, becomes lower than a specific value, to enhance the search procedure with sufficient randomness.

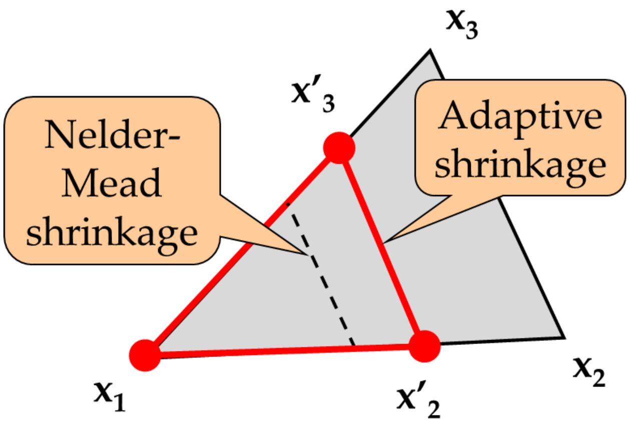

Figure 2: Shrinkage of the simplex around the current best vertex x1 according to the Nelder-Mead formula, i.e. $\mathbf{x}'_i = 0.5\ (\mathbf{x}_i + \mathbf{x}_1)$ and the dynamic adjustment formula, given by $\mathbf{x}'_i = \delta\ \mathbf{xi} + (1 - \delta)\ \mathbf{x}_1$, where $\delta = 1 - 0.5\ (T - T_0)$, T is the current temperature and T0 is the initial temperature.

Test new evolutionary annealing-simplex algorithm in mathematical functions

We tested six benchmark functions of ranging complexity in deterministic and stochastic setting, assuming three levels of Gaussian noise by changing the standard deviation of the stochastic term, $N(0.0,\ 0.75)$, $N(0.0,\ 1.00)$ and $N(0.0,\ 1.25)$. In all cases the global minimum lies in the origin ($\mathbf{x^*} = \mathbf{0}$). For each test function we carried out 100 independent runs of the algorithm, for n = 2 and n = 10 variables, as well as three population sizes (n+1, 2n+1, 8n+1). The results are summarized in the next figure (Fig.3).

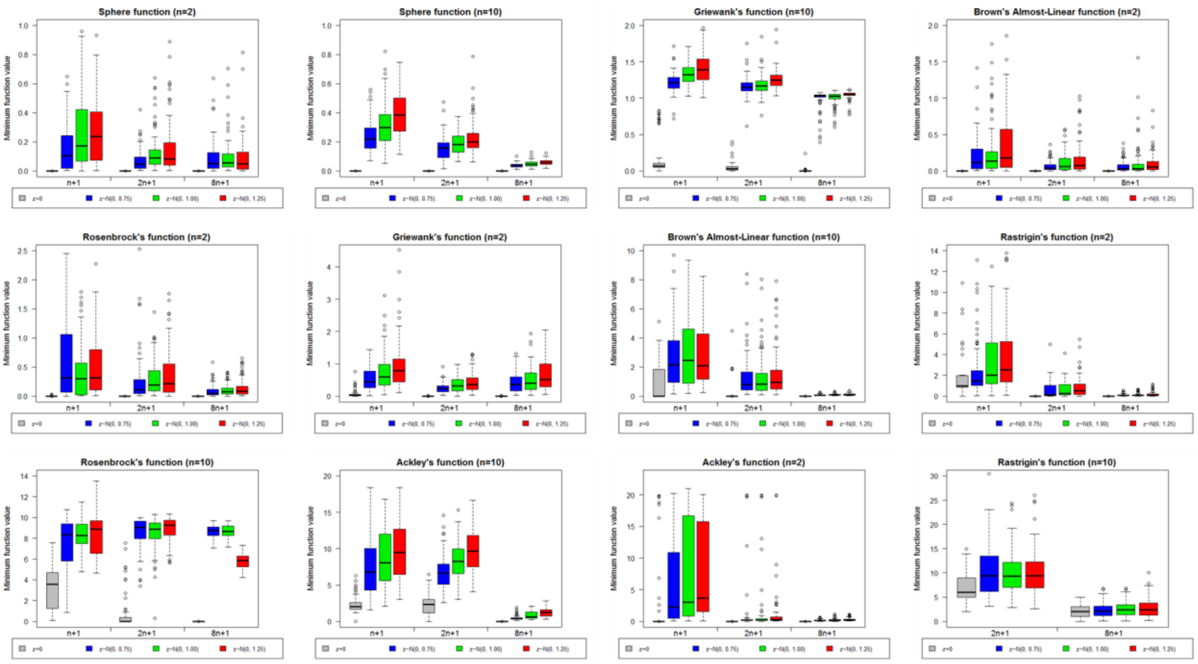

Figure 3: Box plots with optimization results, derived from 100 runs of each problem

Stochastic calibration of hydrological models

It is well-known that the parameters of conceptual hydrological models may vary substantially across different calibration periods. This questions model transposability in time, which is key requirement for ensuring a satisfactory predictive capacity (Gharari et al., 2013). In this context, we propose a stochastic calibration procedure, in which the fitting criterion (e.g. Nash-Sutcliffe efficiency, NSE) is estimated from randomly changing samples that are determined by means of (typically short) moving windows across the full series of the observed responses. The above strategy was tested in three large-scale river basins of Greece (Acheloos, Aliakmon and Boeoticos Kephisos) that exhibit different hydrological behaviour, where we fitted the conceptual model Zygos against the observed runoff. The software supports various parameterizations, according to the complexity of each basin and the available data, and its full structure uses nine parameters (http://itia.ntua.gr/en/softinfo/22). We applied the EAS algorithm to provide 100 independent stochastic calibrations at each basin, with different moving windows. As shown in the next Figures (Fig. 4-6), even when using very short windows (i.e. from 1 to 5 years), the NSE values are close to the ones estimated from the full sample of observed runoff.

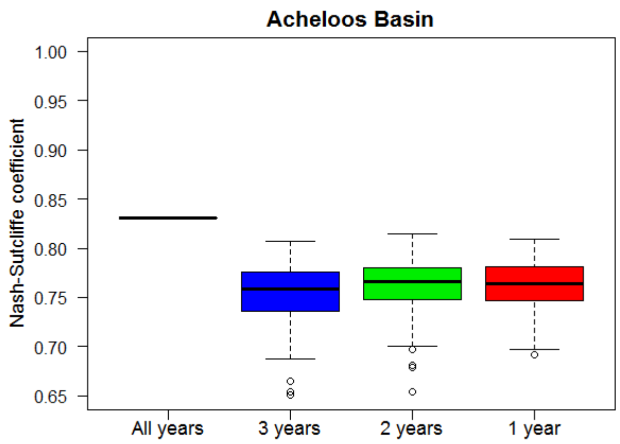

Figure 4: Boxplot of NSE at Acheloos Basin for different moving windows

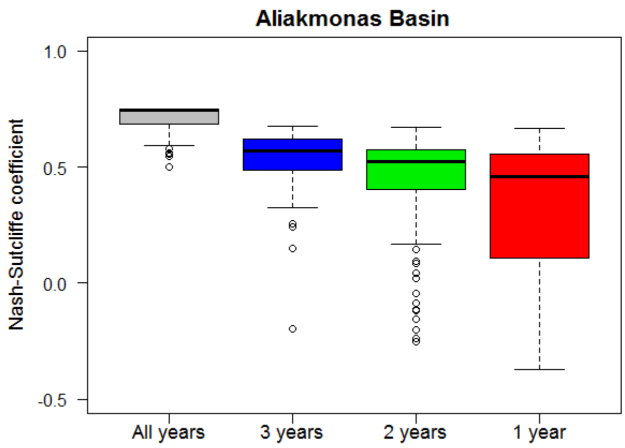

Figure 5: Boxplot of NSE at Aliakmonas Basin for different moving windows

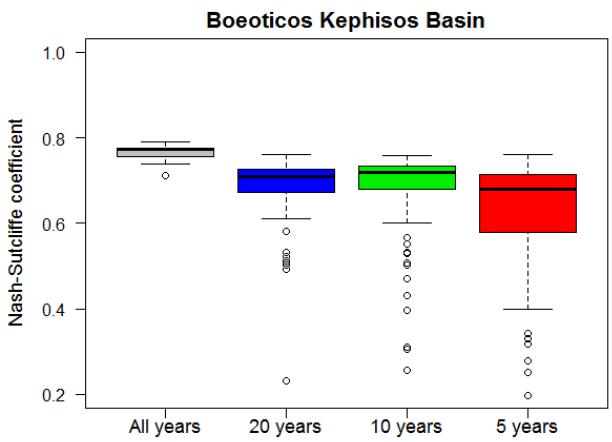

Figure 6: Boxplot of NSE at Boeotikos Kephisos Basin for different moving windows

Applicability of Bartlett-Lewis model in Athens Rainfall

In the framework of present study, we investigate the applicability of rectangular pulse Bartlett-Lewis model for the simulation of Athens rainfall. The main assumptions of the model are (Fig. 7):

- **Storm origins** t_i occur in a Poisson process, with rate λ
- **Cell origins** t_{ij} occur in a Poisson process, with rate β
- **Cell arrivals** terminate after time v_i, which is exponentially distributed (parameter γ)
- **Cell durations** w_{ij} are exponentially distributed (parameter η)
- **Cell intensities** x_{ij} are either exponentially or gamma distributed.

In the modified version (Rodriguez-Iturbe *et al.*, 1988), parameter η is assumed gamma distributed, with scale parameter v and shape parameter a, and varies for each event, such as β/η and γ/η remain constant. Model parameters are estimated via calibration, seeking to minimize the departures between the key theoretical and observed statistics.

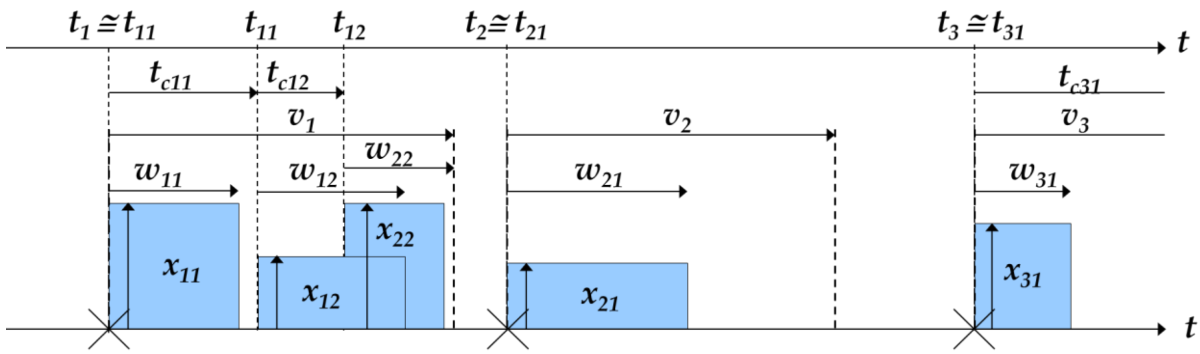

Figure 7: The Bartlett-Lewis model

We examined the performance of the original (BL) and modified (MBL) Bartlett-Lewis model using hourly rainfall data from the National Observatory of Athens (1927-1996), for two months with different meteorological behaviour (January, June). Model parameters were calibrated against the theoretical statistics (mean, standard deviation, autocovariance and probability dry), for 1 and 24 h. The simulated statistics were estimated from a synthetic series of 1000 years length. For the generation of synthetic series we used HyetosR package (Kossieris et al., 2011). The next table represents a comparison between the historical, modelled and simulated statistics for the two versions of Bartlett-Lewis model.

	Historical	Theoretical – BL	Simulated – BL	Theoretical – RBL	Simulated – RBL	
Average (mm)	0.065	0.065	0.065	0.065	0.065	January,
Standard deviation (mm)	0.458	0.458	0.458	0.458	0.458	Hourly
Coefficient of skewness	16.957	-	11.884	-	12.663	statistics
Average (mm)	1.555	1.555	1.557	1.563	1.563	January,
Standard deviation (mm)	4.532	4.532	4.535	4.053	4.083	Daily
Coefficient of skewness	5.301	-	4.235	-	5.289	statistics
Average (mm)	0.015	0.015	0.015	0.015	0.015	June,
Standard deviation (mm)	0.370	0.370	0.375	0.370	0.374	Hourly
Coefficient of skewness	50.578	-	47.684	-	49.428	statistics
Average (mm)	0.365	0.365	0.360	0.365	0.365	June,
Standard deviation (mm)	2.694	2.694	2.822	2.692	2.638	Daily
Coefficient of skewness	11.881	-		-		statistics

Table 1: Historical, modeled and simulated characteristics of BL model for daily and hourly rainfall of January and June

Apart from the main statistics, we also examined the performance of the two models on the temporal characteristics of Athens rainfall. Both versions of the BL model fail to reproduce the significant variability of rainfall events, due to the overclustering of pulses. This also results to an over-estimation of probability dry, at the hourly and daily time scales (Fig. 8) as well as the generation of rainfall events of shorter duration, and thus longer dry intervals (Figs. 9 and 10).

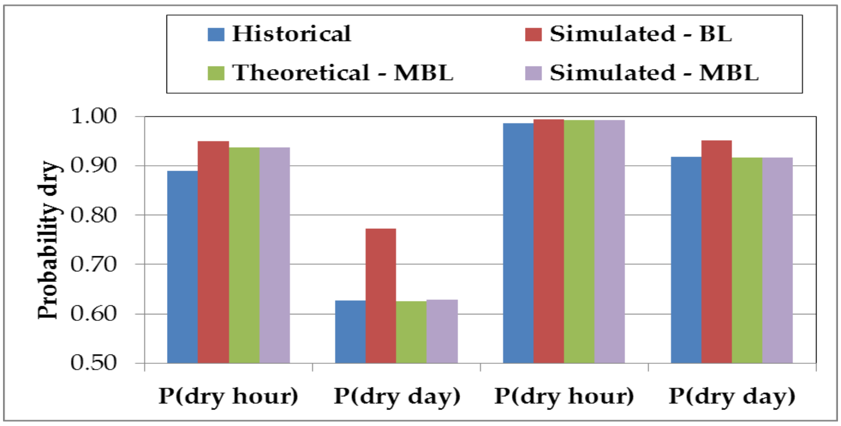

Figure 8: Probabilities of dry hour and dry day for January (left) and June (right).

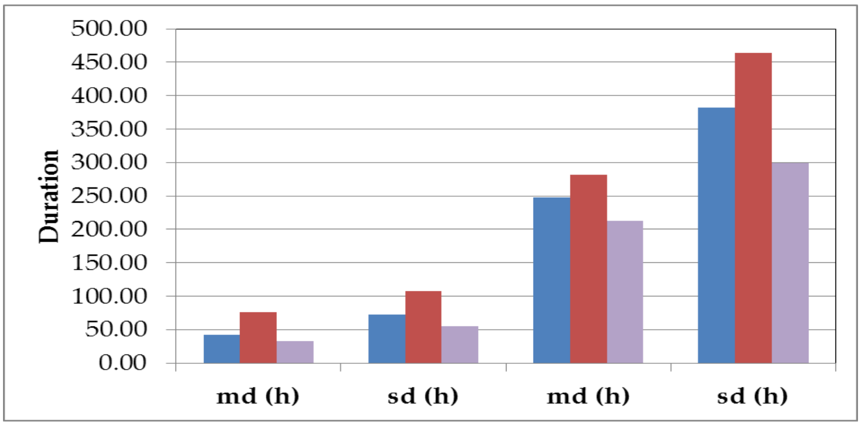

Figure 9: Mean (md) and standard deviation (sd) of dry time intervals for January (left) and June (right).

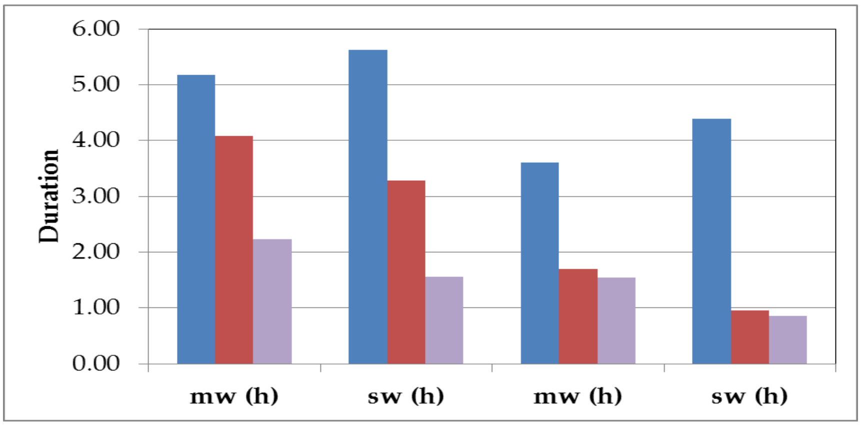

Figure 10: Mean (mw) and standard deviation (sw) of duration of rainfall events for January (left) and June (right).

Conclusions

- The various sources of uncertainty, involved either directly or indirectly, in real system optimization problems pose particular difficulties in the search of optimal solutions and decision-making. Uncertainties transferred to optimization problems creating response surfaces which are strongly disordered, rough and non-convex.
- The evolutionary annealing-simplex algorithm was tested on a series of mathematical functions when the response surface is disrupted by the addition of a stochastic term by normal distribution. As revealed by the increase in the intensity of noise, which is controlled through the standard deviation of the distribution, the performance of the algorithm is deteriorated. However, it seemed that the use of large populations for efficient exploration of the feasible search space allows the detection of global optimal point with relatively high accuracy. This finding is

consistent with the general literature which stated that the algorithms that make use of large populations (e.g. genetic algorithms) supersede those based on a small number of points to identify any area attraction.

- The presence of noise in the response of objective function changes drastically the search path of the algorithm against deterministic problems. Specifically, at each iteration the probability of executed move that reduces the volume of the simplex is twice the probability that the simplex become larger. This has as a consequence the strong disturbance of the shape and size of the simplex and fast convergence of the algorithm suboptimal spots.
- The integration of original annealing-simplex algorithm with new techniques and mechanisms has a positive effect on the overall performance of the algorithm. These techniques aim at preservation of simplex from premature degeneration and incorrect convergence to a sub-optimal point. The performance improvement was more pronounced in the use of smaller populations, while the performance of the two algorithms is almost identical for large populations.
- From the stochastic calibration of the conceptual hydrological model Zygos was proven both the robustness of the solutions found by the algorithm and the adequate predictive capacity of the model. Even with the use of very small sample length, model succeeds to yield sufficiently large amplitude responses that produce three basins with completely different characteristics.
- The Bartlett-Lewis model reproduces with high accuracy the basic statistical characteristics of Athens rainfall at different scales for a single set of parameters. However, the model fails to maintain the temporal properties of rainy episodes and dry periods so that its use is limited to the field of study of flooding.