European Geosciences Union General Assembly 2015
Vienna, Austria, 12 - 17 April 2015
Session HS3.1: HydroInformatics: computational intelligence , system analysis, optimisation

Handing time-expensive global optimization problems through the Surrogate-Enhanced Evolutionary Annealing-Simplex (SEEAS) algorithm

Ioannis Tsoukalas*, Panagiotis Kossieris, Andreas Efstratiadis, and Christos Makropoulos

School of Civil and Environmental Engineering, National Technical University of Athens, Heroon Polytechneiou 5, GR-157 80, Zographou, Greece

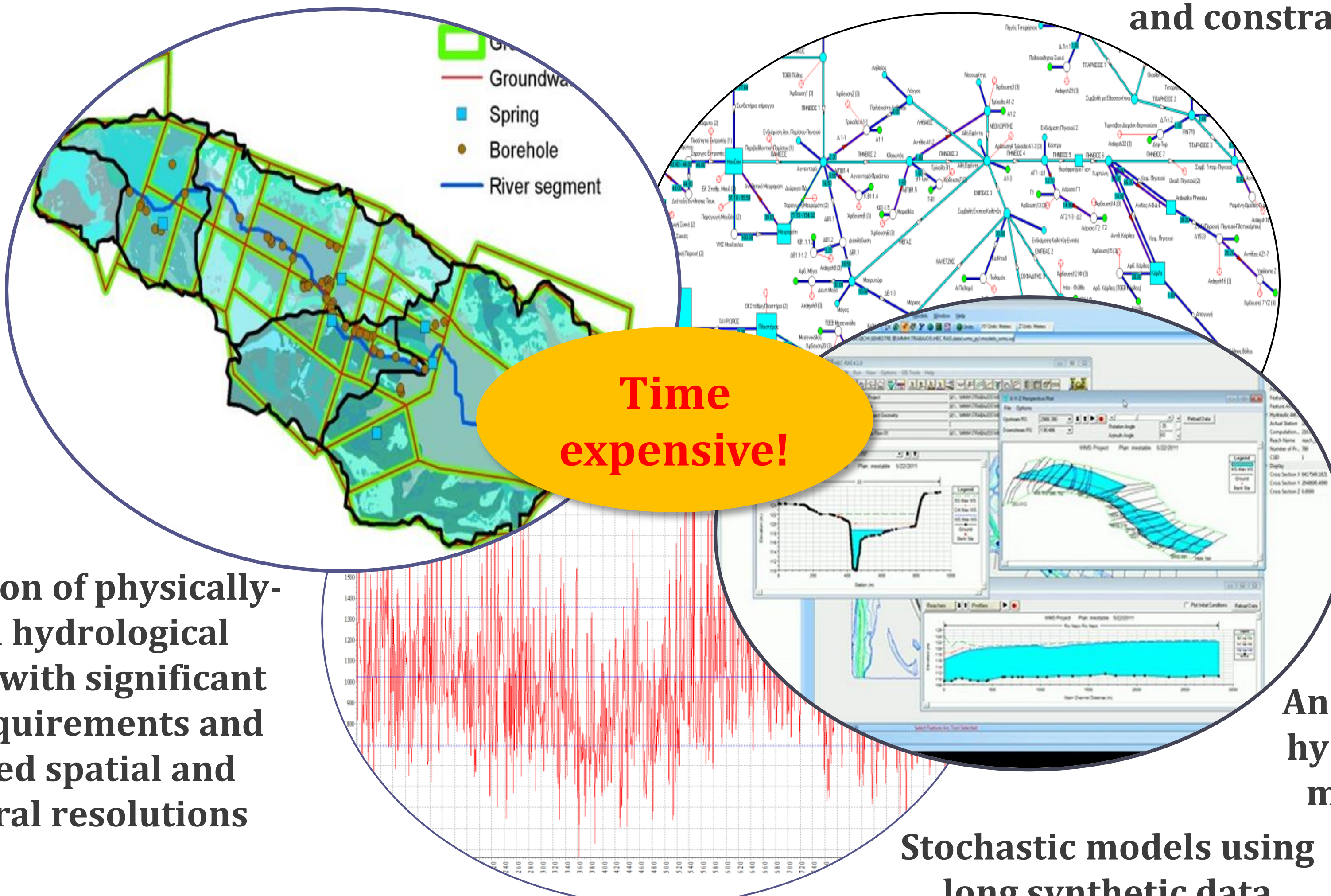
* Corresponding author, Tel. ++302107722886; e-mail: itsoukal@mail.ntua.gr

Typical obstacles to the global optimum

1. **Convergence to local optima**
   - Generally, it is relatively easy to locate a local optimum, but very difficult or even impossible to get out of it.
2. **The curse of dimensionality**
   - The theoretical time to solve a nonlinear optimization problem increases even exponentially vs. the number of control variables.
3. **Extremely large number of trials to locate the global optimum**
   - To avoid getting trapped by local optima, a detailed exploration of the search space may be required.
4. **The practical aspect of real-world applications**
   - In real-world problems, a highly accurate solution is neither **possible**, because of uncertainties and inaccuracies in the underlying model or data, nor **feasible**, due to the prohibitory **high computational effort** required to evaluate the objective function, which in turn requires running a **simulation model**.

Challenging global optimization problems in water resources

Decision-making problems in complex hydrosystems subject to multiple objectives and constraints



Calibration of physically-based hydrological models with significant data requirements and detailed spatial and temporal resolutions

Analytical hydraulic models

Stochastic models using long synthetic data

Surrogate-based optimization (SBO)

Basic concepts of SBO:

- Replace most of the time-expensive simulations with surrogate models of practically negligible (or minor) computational burden (e.g., Kriging, Radial Basis Functions, polynomials).
- Most evaluations are performed through the surrogate model, while the expensive model (i.e. the objective function) is called periodically to improve the accuracy of the results within the search procedure.
- Comprehensive reviews are found in the literature (Forrester and Keane, 2009).

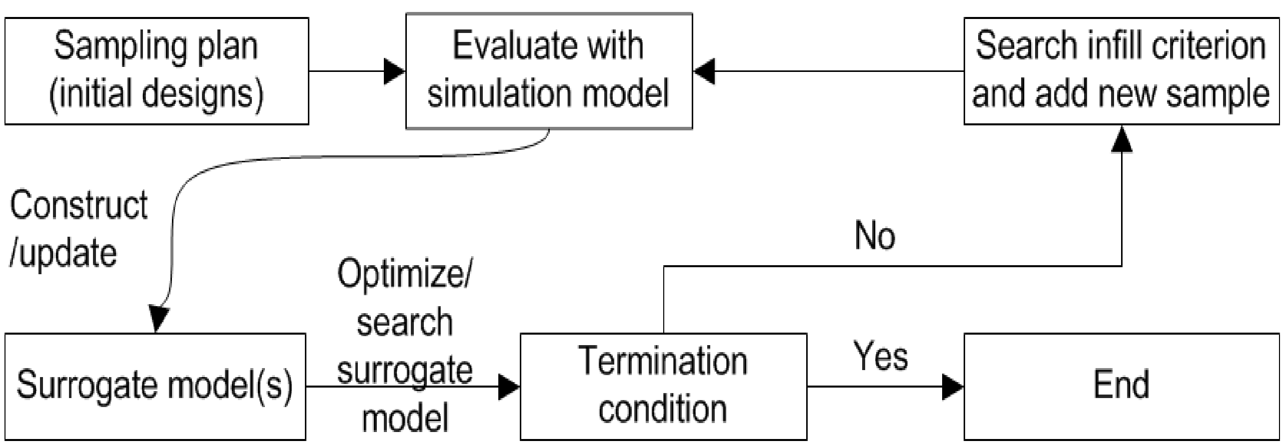


Figure 1: Flowchart of a typical surrogate-based optimization procedure.

The Evolutionary Annealing-Simplex method

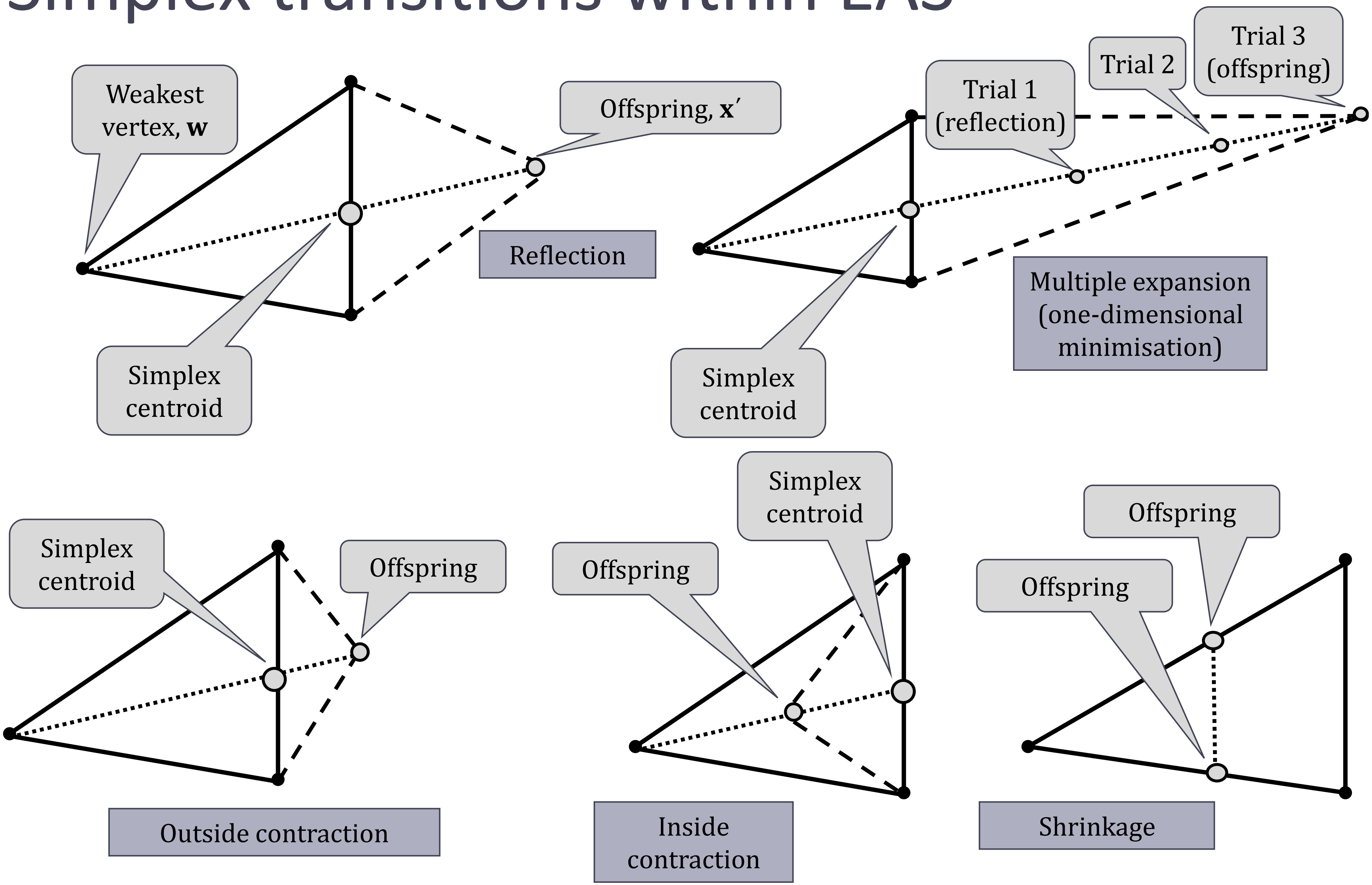
- Heuristic global optimization technique coupling the strength of **simulated annealing** in rough search spaces with the efficiency of the **downhill simplex method** (Nelder & Mead, 1965) in smoother spaces (Efstratiadis & Koutsoyiannis, 2002).

Key features

- 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 $g(\mathrm{x}) = f(\mathrm{x}) + u\, T$;
- 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.

EAS package is available in R: www.itia.ntua.gr/el/softinfo/29/

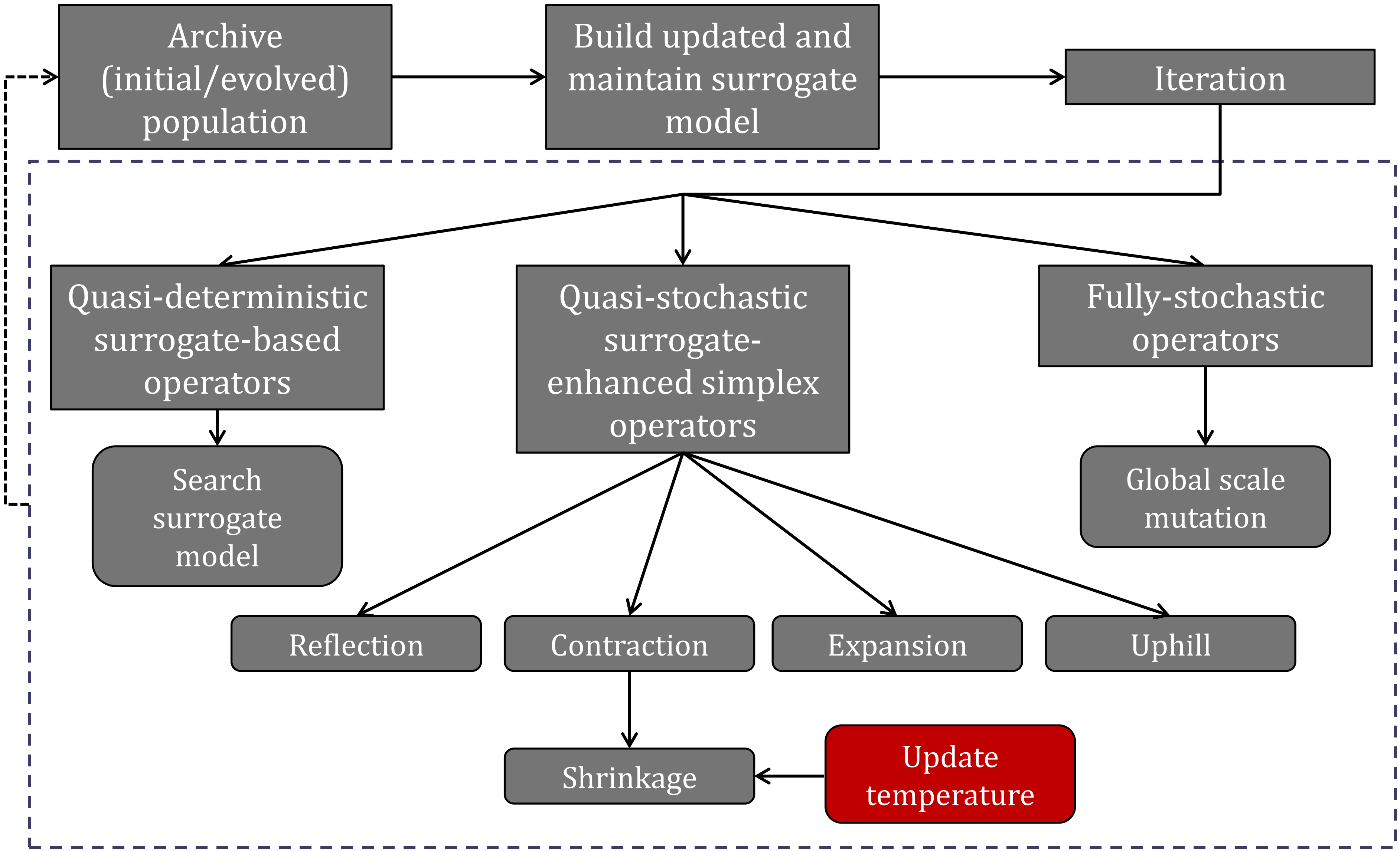
Simplex transitions within EAS



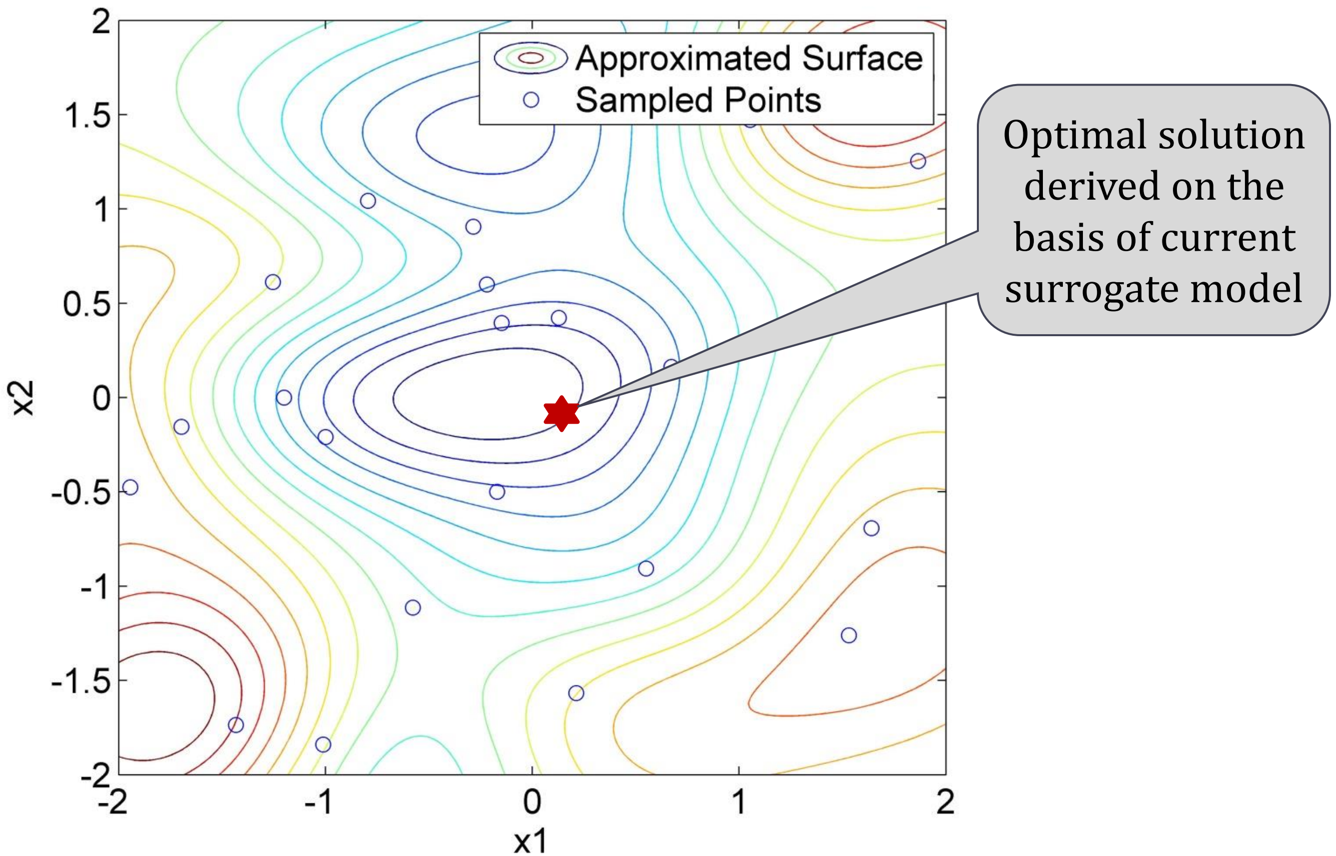
Key aspects of SEEAS

- Surrogate-Enhanced Evolutionary Simplex-Annealing approach (SEEAS) is a novel global optimization algorithm for **time-expensive** functions.
- The algorithm is a surrogate-enhanced extension of EAS incorporating Surrogate Modelling (SM) techniques to build, maintain and exploit approximations of real response surfaces, aiming to **support transitions** and **accelerate search** towards favorable areas of the response surface.
- The role of SM in searching procedure is twofold:
  - Providing new promising points that are directly embedded in the current population (similarly to SBO);
  - Assisting specific transitions of the **simplex-based** evolutionary operator of EAS.
- Balance between **exploration** (i.e., detailed sampling) and **exploitation** (i.e., blind use of predictions) is achieved through a **dynamically adjusted weighted prediction-distance metric**, termed **acquisition function** (AF).
- An **external archive** of already evaluated points is updated per iteration thus ensuring systematically more accurate approximations of the region of interest (i.e., the area around the current best point).

SEEAS conceptual flow chart

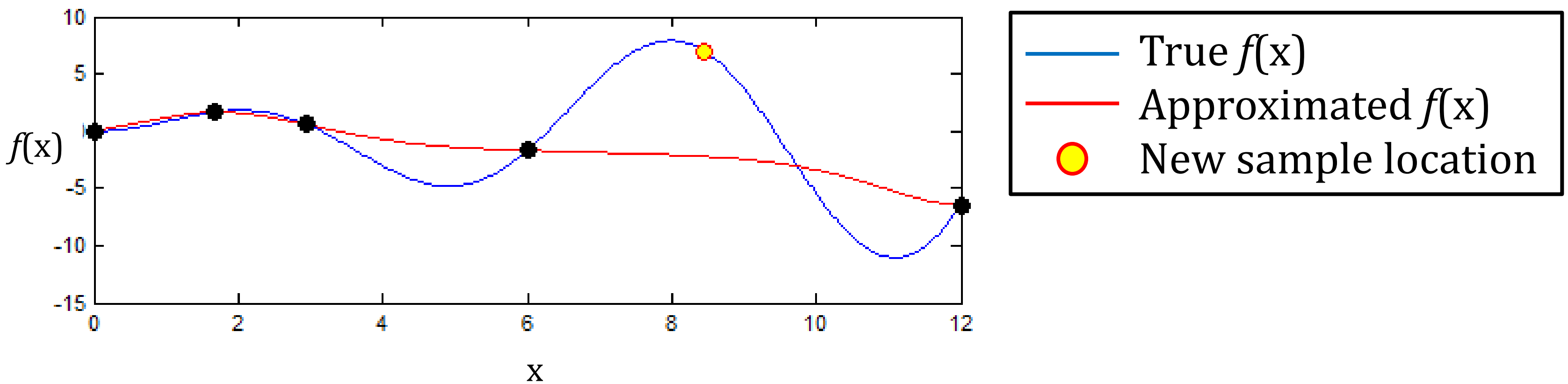


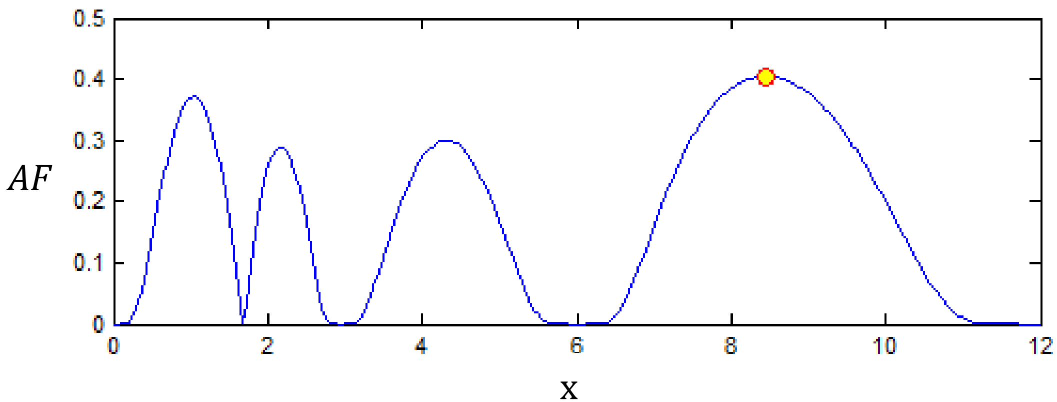
An illustrative example: Quasi-deterministic surrogate-based operators



Remarks:
In the beginning of each iteration, an initial search (internal optimization) is performed using only the surrogate model, to locate potential sample points for evaluation with the expensive function.

An illustrative example: Balancing exploration-exploitation with acquisition function (AF)

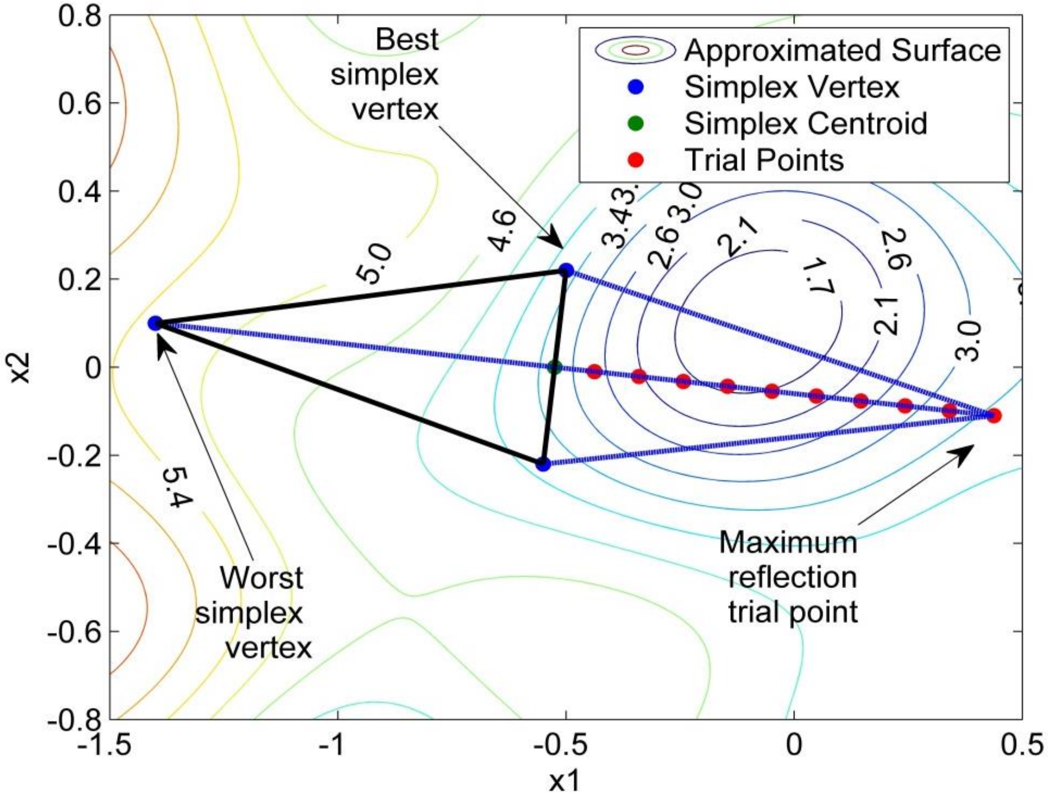




Example
$f(x) = x \sin(x)$
s.t. $0 \leq x \leq 4\pi$

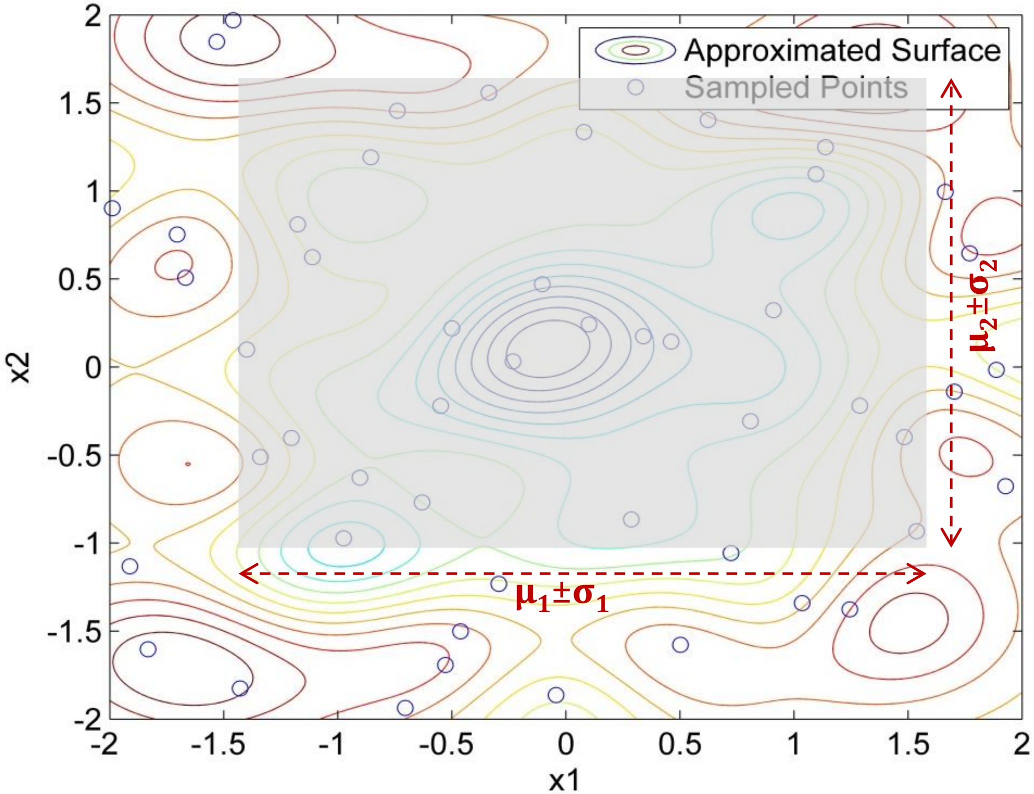
Remarks:
The acquisition function (AF) is used to **balance** exploration-exploitation of surrogate model, i.e., balance between surrogate model prediction estimates and search space densification.

An illustrative example: Quasi-stochastic surrogate-enhanced simplex operators



Remarks:
Demonstration of a randomly selected simplex and the modified surrogate-enhanced reflection movement using candidate points on the line formed from the simplex centroid and the maximum reflection point; the simplex is reflected at the candidate point with the minimum function value.

An illustrative example: Fully-stochastic operators (mutation)



Remarks:
Demonstration of a global scale mutation by generating a random point out of the range (**μ-σ**, **μ+σ**) of the current population (where **μ** and **σ** are the average and standard deviation of the coordinates of all population members).

Benchmarking methodology

- The performance of SEEAS was compared against five state-of-the-art methods:
  - DYCORS (Regis and Shoemaker, 2013) and MLMSRBF (Regis and Shoemaker, 2007), which are surrogate-enabled;
  - EAS and DDS (Tolson and Shoemaker, 2007) that do not employ surrogate models through search.
- A variety of test problems were examined, theoretical as well as real-world, considering two alternative computational budgets, which are quantified in terms of function evaluations (FE; 500 and 1000).
  - six mathematical test functions, formulated with 30 control variables;
  - a hydrological calibration problem with real and synthetic runoff data;
  - a time-expensive multi-reservoir management problem.
- In all problems we employed multiple independent runs, considering:
  - the same population size;
  - the same random generation technique for the initial population (LHS).
- The three surrogate-assisted algorithms (SEEAS, DYCORS, MLMSRBF) use the same metamodel, i.e. RBF with cubic basis functions and linear polynomial tail.

Mathematical applications: Test functions

The mathematical benchmark suite includes:

- Six mathematical problems (test functions) with 30 control variables;
- Two alternative computational budget (500 and 1000 FE);
- 30 independent optimization trials (i.e. with different initial populations that are randomly generated).

Problem	Test function	Response surface properties
OF1	Sphere	Unimodal and convex
OF2	Ackley	Multimodal with many local minima
OF3	Griewank	Multimodal with many regularly distributed local minima
OF4	Zakharov	Unimodal with a plate-shaped valley
OF5	Rastrigin	Multimodal with many local minima
OF6	Levy	Multimodal with many local minima and parabolic valleys

<u>Remarks:</u>
In all cases, the location and thus the value of the global optimum are known.

Mathematical applications: Results

Table: Mean and standard deviation of best function values for 30-D obtained from all algorithms.

FE	Test function	EAS		DDS		SEEAS		DYCORS		MLMSRBF	
		Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev	Mean	StDev
500	OF1	4.3053	1.1633	9.5161	2.7373	**0.0185**	0.0058	0.0832	0.0337	0.7393	0.7078
	OF2	9.9234	1.1601	12.8723	1.3289	**1.8775**	0.3012	4.2965	3.7206	6.1931	4.3622
	OF3	17.8661	3.4553	38.3984	12.0501	**0.7823**	0.1177	1.2647	0.0785	3.4586	1.9269
	OF4	**117.821**	28.7568	562.1452	113.2297	173.240	44.1854	472.8149	90.8972	575.4238	174.0726
	OF5	228.6932	18.4416	132.1485	24.5665	122.657	19.4271	**112.046**	23.0760	165.4371	46.8463
	OF6	6.3378	2.6523	15.8232	5.4805	**0.6587**	0.1843	3.4072	2.5399	7.3257	10.9444
1000	OF1	2.5294	0.9330	2.1121	0.7910	**0.0058**	0.0036	0.0105	0.0042	0.3577	0.1771
	OF2	6.5163	0.8448	7.6695	0.9239	1.2064	0.2965	**1.0846**	0.1682	3.6428	1.1026
	OF3	8.8362	2.6165	8.2729	2.6788	**0.5492**	0.0930	1.0200	0.0264	2.4197	0.7127
	OF4	**94.5982**	20.3174	412.2383	118.5726	151.471	54.0970	403.8120	93.0812	491.4248	146.0970
	OF5	198.3349	16.5873	**71.5980**	15.0283	98.370	19.5046	85.2669	22.9556	134.8636	39.1934
	OF6	2.6831	0.7357	3.9213	2.2148	**0.4426**	0.1263	4.2130	5.4396	2.8652	4.5828
Wins		2/12		1/12		7/12		2/12		0/12	

Hydrological model calibration

- **Study area and data**: Boeticos Kephisos basin, Eastern Greece (1850 km^2)
  - The basin extends over a heavily-modified karst system with multiple peculiarities, as result of complex interactions between surface and groundwater processes as well as human interventions, by means of surface and groundwater abstractions;
  - Monthly time series of precipitation, PET, groundwater abstractions and runoff are available for a 77-year period (1907-1984; 924 months).
- **Model**: Lumped version of Hydrogeios (Efstratiadis et al., 2008)
  - The basin is vertically subdivided into three storage elements that represent interception, soil moisture and groundwater accounting processes;
  - The model estimates the main responses of the basin, i.e. actual evapotranspiration, surface and groundwater runoff and underground losses, using nine parameters;
  - Initial conditions are the water levels of soil and groundwater tanks at the beginning of simulation.
- Formulation of **calibration** problem:
  - Two problems were formulated, one with real (measured) and one with artificially generated runoff data, considering arbitrary parameter values (toy model);
  - Objective function = Nash-Sutcliffe efficiency (NSE).

Results for model calibration with actual runoff data (unknown parameter values)

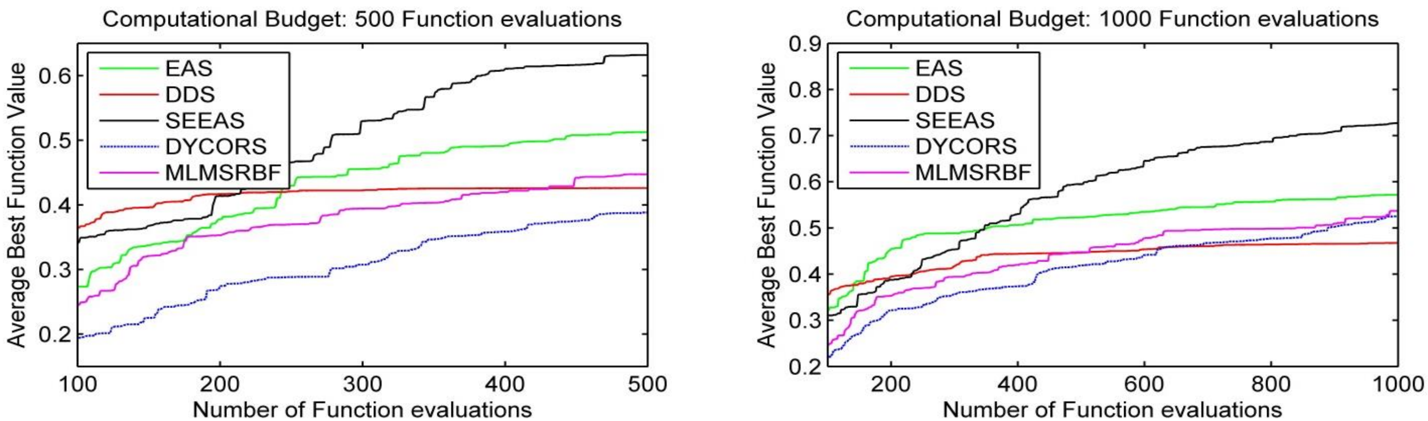


Figure: Convergence curves for FE = 500 (left) and FE = 1000 (right).

Remarks:

- SEEAS clearly outperforms the other methods for 500 and 1000 FE.
- EAS is the next most promising algorithm.

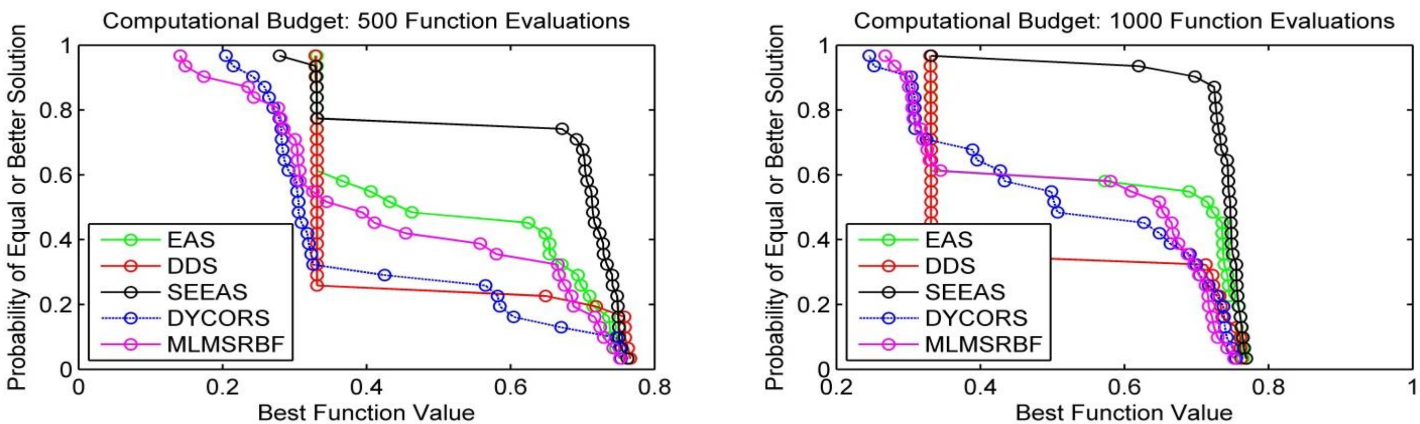


Figure: Empirical CDFs of best NSE values for FE= 500 (left) and FE = 1000 (right).

Remarks:

- SEEAS achieves NSE > 0.65 in **23 out of 30** trials.
- SEEAS achieves NSE > 0.70 in **28 out of 30** trials.

Results for toy model calibration (a priori known parameter values)

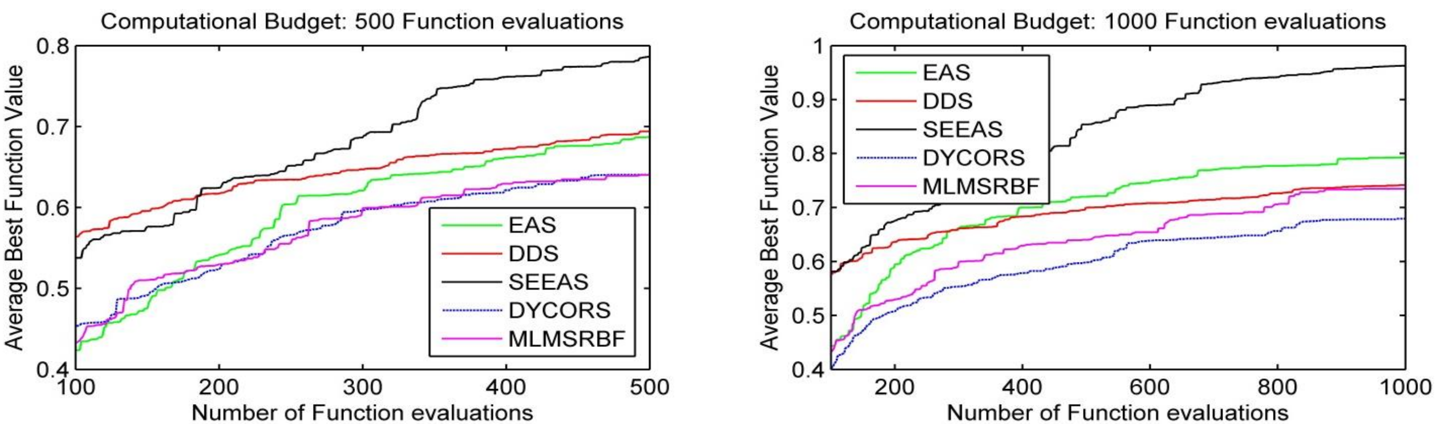


Figure: Convergence curves for FE = 500 (left) and FE = 1000 (right).

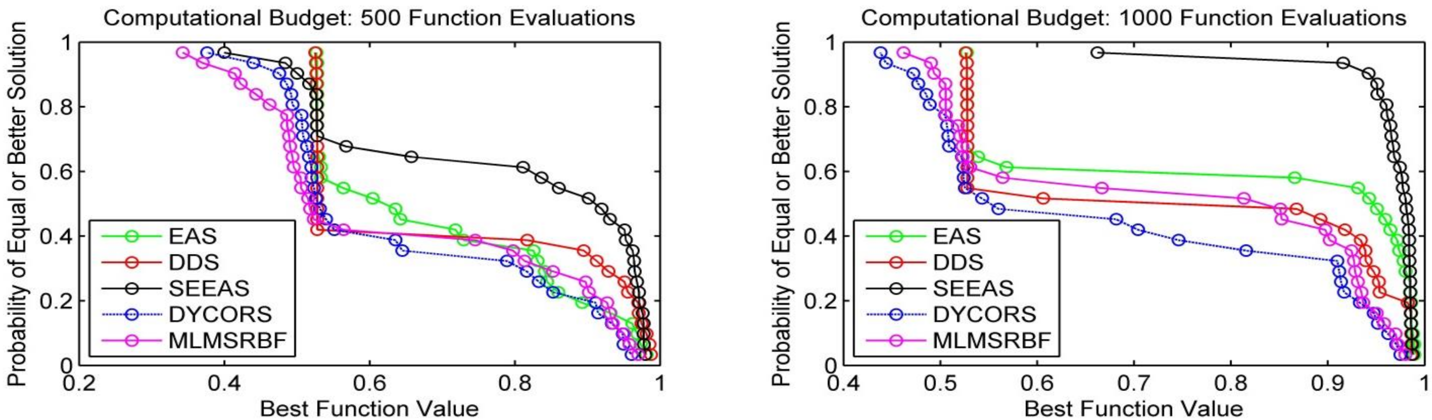


Figure: Empirical CDFs of best NSE values for FE= 500 (left) and FE = 1000 (right).

Remarks:

- SEEAS clearly outperforms the other algorithms for both 500 and 1000 FE.
- EAS and DDS are the more suitable alternatives for 500 FE.
- EAS is the best alternative for 1000 FE.

Remarks:

- SEEAS achieves NSE>0.80 in **20 out of 30** trials.
- SEEAS achieves NSE>0.95 in **27 out of 30** trials.

Multi-reservoir management problem

- **Case study:** Nestos hydrosystem, NE Greece
  - Four reservoirs, serially connected.
  - The first three are hydropower reservoirs, with the first two of them reversible (pumped storage).
  - The forth (lowest) reservoir is small irrigation reservoir.
- **Problem statement:** Development of uncertainty-aware operational rules of the multi-reservoir management model deployed in WEAP21 (Yates et al, 2005) and coupled with MATLAB mathematical environment (initially introduced by Tsoukalas and Makropoulos, 2014).
- **Control Variables:** Expressed in terms of seasonally (considering 4 seasons per year) varying energy targets, which are assigned to the associated system components; $(3 + 2) \times 4 = 20$ control variables in total.
- **Objective:** Maximization of mean annual benefit from hydropower production under hydrological uncertainty, inherited by long stochastically generated time-series using CASTALIA stochastic model (Efstratiadis et al., 2014).

Results for multi-reservoir optimization

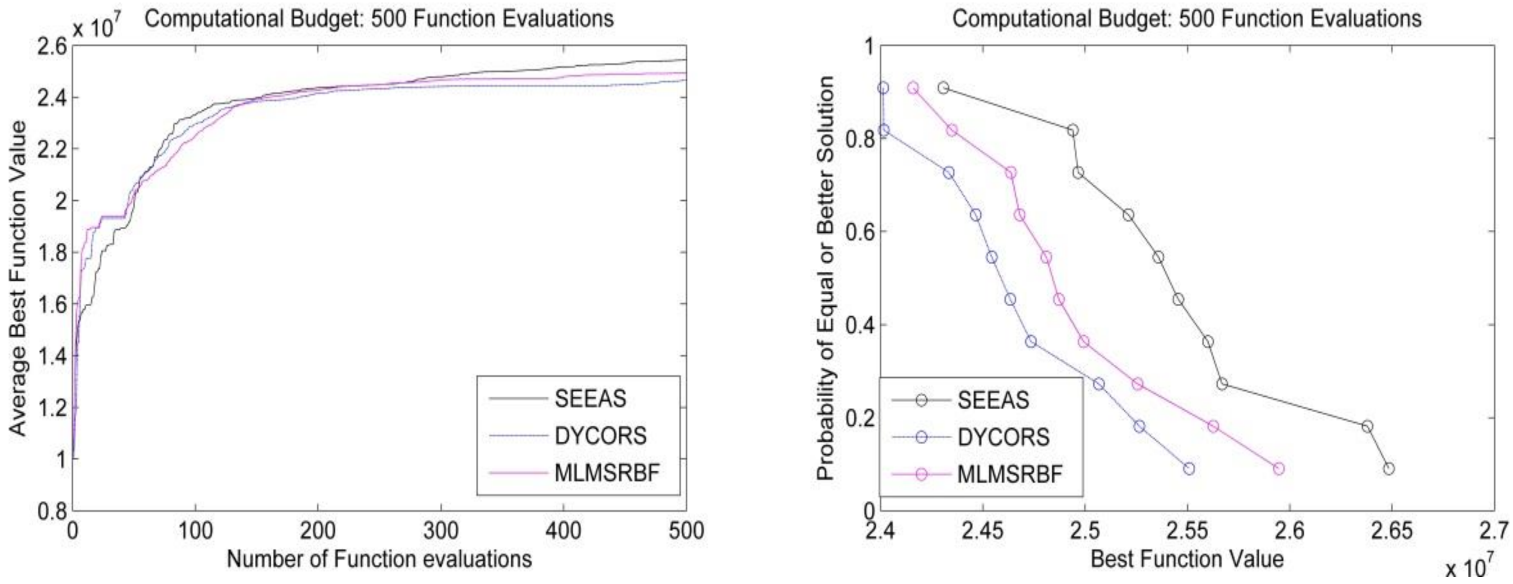


Figure: Convergence curves (left) and empirical CDFs (right) for FE = 500.

- ❑ SEEAS outperforms both DYCORS and MLMSRBF, considering the budget of 500 FE.
- ❑ Algorithms seem performing equally until ~300 FE, but then SEEAS evolves faster.
- ❑ In terms of CDFs, SEEAS stochastically dominates MLMSRBF, which dominates DYCORS.

<u>Remark</u>: The two figures reveal the key peculiarity of reservoir optimization problems, which is the formulation of **flat response surfaces**, due to the existence of **numerous constraints**, physical and operational, which significantly restrict the flexibility of decisions, thus resulting to **low sensitivity** of the system performance against the associated parameters.

Conclusive remarks

- The novel Surrogate-Enhanced Evolutionary Annealing Simplex algorithm (SEEAS) is introduced.
- The surrogate model is employed efficiently (in terms of balancing exploration and exploitation) and it is used for global search and also identifies the most promising positions to perform the simplex movements using trial samples (i.e., candidate locations).
- SEEAS outperforms alternative algorithms (DDS, DYCORS*, MLMSRBF*) in 9/12 mathematical problems (i.e. six test functions with 30 control variables) for two computational budgets (500 and 1000 MFE).
- SEEAS outperforms all alternative algorithms in the examined real-world problems, including two hydrological model calibration (11 parameters) and a multi-reservoir management problem (20 decision variables).

*Surrogate-assisted algorithm

Thank you!

SEEAS was recently submitted to **Environmental Modelling & Software** under the title:

Tsoukalas I., Kossieris P., Efstratiadis A. and Makropoulos C., **Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget**, *Environmental Modelling & Software*, 2015.

Acknowledgements

The proposed methodology was developed for the conduction of the research project entitled “Investigation of climate change in Greece and its impact on the sustainability of projects dealing with hydroelectric power and the agricultural economy: Application in the Nestos river basin – KLIMENESTOS” which was financed by the Ministry of Education, Lifelong Learning and Religious Affairs, General Secretariat for Research and Technology, Greece through the National Strategic Reference Framework (NSRF) 2007-2013 and under the operational programmes “Competitiveness and Entrepreneurship and Regions in Transition” and within the National Action “Cooperation 2009”

References

- Efstratiadis, A., Koutsoyiannis, D., 2002. An evolutionary annealing-simplex algorithm for global optimisation of water resource systems, *5th International Conference on Hydroinformatics*, Cardiff, UK, 1423–1428, IWA.
- Efstratiadis, A., Nalbantis, I., Koukouvinos, A., Rozos, E., Koutsoyiannis, D., 2008. HYDROGEIOS: a semi-distributed GIS-based hydrological model for modified river basins. *Hydrology and Earth System Sciences,* 12(4) 989-1006.
- Efstratiadis, A., Dialynas, Y., Kozanis, S., Koutsoyiannis, D., 2014. A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence. *Environmental Modelling & Software* 62(0) 139-152.
- Forrester, A., Keane, A., 2009. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1–3), 50-79.
- Jones, D., Schonlau, M., Welch, W., 1998. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization,* 13, 455-492.
- Nelder, J.A., Mead, R., 1965. A Simplex method for function minimization. *The Computer Journal*, 7(4), 308-313.
- Regis, R.G., Shoemaker, C.A., 2013. Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. *Engineering Optimization*, 45(5), 529-555.
- Regis, R.G., Shoemaker, C.A., 2007b. A stochastic radial basis function method for the global optimization of expensive functions. *Informs Journal on Computing*, 19(4), 497-509.
- Tolson, B.A., Shoemaker, C.A., 2007. Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, 43(1).
- Tsoukalas, I., Makropoulos, C., 2014. Multiobjective optimisation on a budget: Exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. *Environmental Modelling & Software* (in press).
- Yates D., Sieber J., Purkey, D., Huber-Lee, A., 2005. WEAP21: A Demand, priority, and preference driver water planning model. Part 1: Model Characteristics. *Water International*, 30, 487-500.