Webinaire sur l'utilisation des méthodes de Machine Learning en hydrologie, Mars 18th 2022

Utilisation de la technique KNN pour améliorer l'efficience de calibration de modèles pluie-débit

Hamouda DAKHLAOUI [1,2]

[1]LMHE, Ecole Nationale d'Ingénieurs de Tunis, Université Tunis El Manar, Tunisie

[2]Ecole Nationale d'Architecture et d'Urbanisme, Université de Carthage, Sidi Bou Said, Tunisie.

Introduction

- Rainfall-Runoff models are used widely in different fields such as water quantity and quality management, flood forecasting, and climatic changes studies
- The successful use of these models depends on the quality of the estimation of their parameters.
- Generally, the calibration of such models could be considered as an optimisation problem that maximise the agreement between observed and simulated historical runoff.
- This optimisation problem is well known for carrying some difficulties such as interdependence between model parameters, the indifference of objective function to some parameters, difficulty of definition of gradient direction, the existence of local optimums and scale problems related to parameters.
- The stochastic global optimisation methods such as genetic algorithm (Goldberg, 1989), Shuffled Complex Evolution (Duan et al., 1993) or simplex simulated annealing algorithm (Bates, 1994) are found to be the most powerful methods to overcome such optimisation difficulties and to give best results with calibration of rainfall-runoff models.
- SCE-UA needs a great number of objective function evaluations to reach the optimal solution and could be time consuming especially in the case of a huge number of parameters to be calibrated, semi-distributed/distributed model, long time period, or fine time step. This problem becomes more and more annoying particularly when a large number of calibration exercices is needed (e.g. DSST)

- Since the objective function estimation by model simulation is the most time consumer in the optimisation process, the reduction in the number of evaluations of objective function needed by calibration algorithm has been proposed as an important way to improve its efficiency.
- The implementation of the approximation functions is one of the main techniques adopted for this purpose, This method involves estimating the objective function from the data sets of parameters explored by the optimization algorithm instead of estimating by direct simulation using the rainfall-runoff model.
- The approximation methods proposed in the literature are polynomial interpolation, geostatistical interpolation, the K-Nearest Neighbours method, the Radial Basis functions.
- However, applying function approximations may affect the ability of the optimisation algorithm to find the optimal solution. The estimating error of the approximation function could perturb the algorithm and direct its search mechanism in the wrong direction.
- Thus, the introduction of approximation function reduces the computation time but at the expense of the objective function accuracy. To reduce the number of simulations required for convergence without affecting the ability of the algorithm to identify the optimal solution represents a challenge.

Objective of the study:

- To improve the efficiency of SCE-UA optimisation algorith used to calibrate HBV model using the K-Nearest Neighbours (KNN) to approximate the objective function, without affecting its effectiveness

- Evaluate the effect of KNN technique ob effectivness and efficiency of SCE-UA

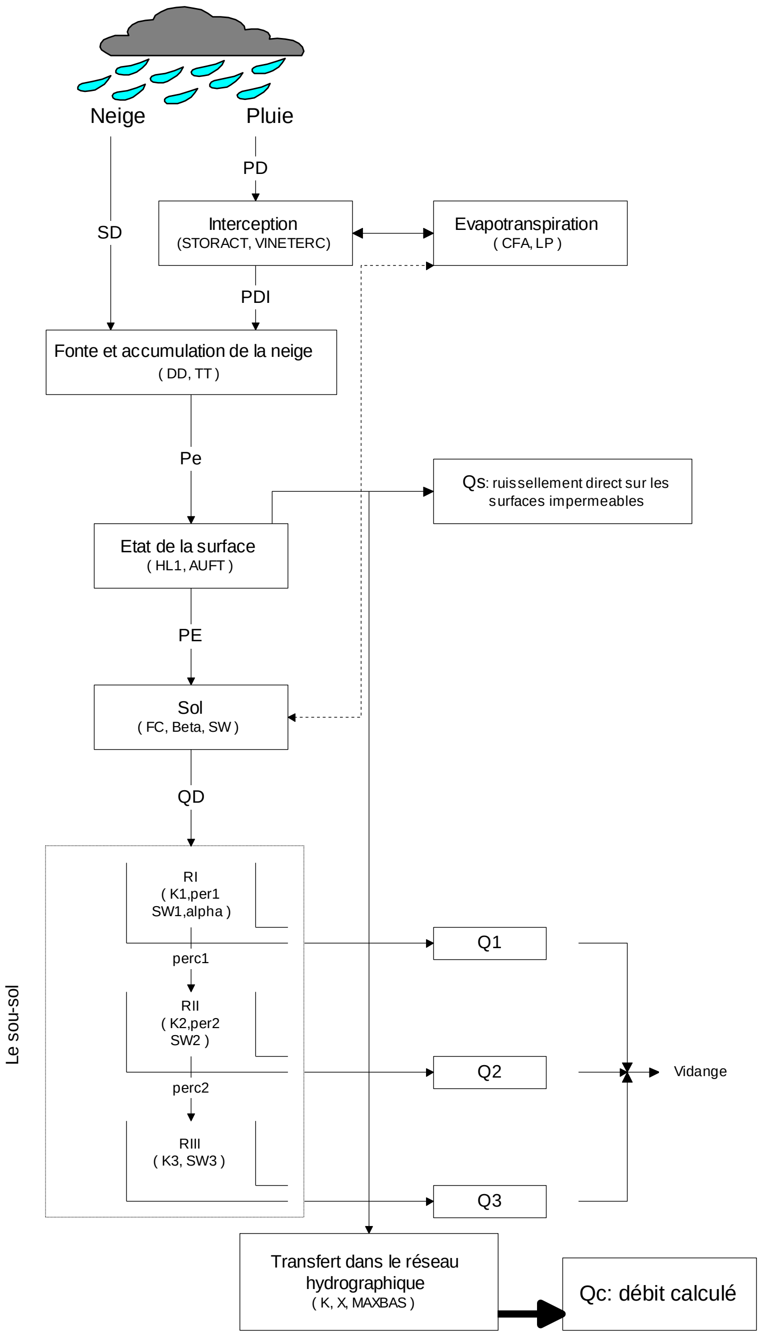
- We evaluate the hydbrid SCE-UA for the calibration of HBV model over several catchment in northern Tunisia

Methods

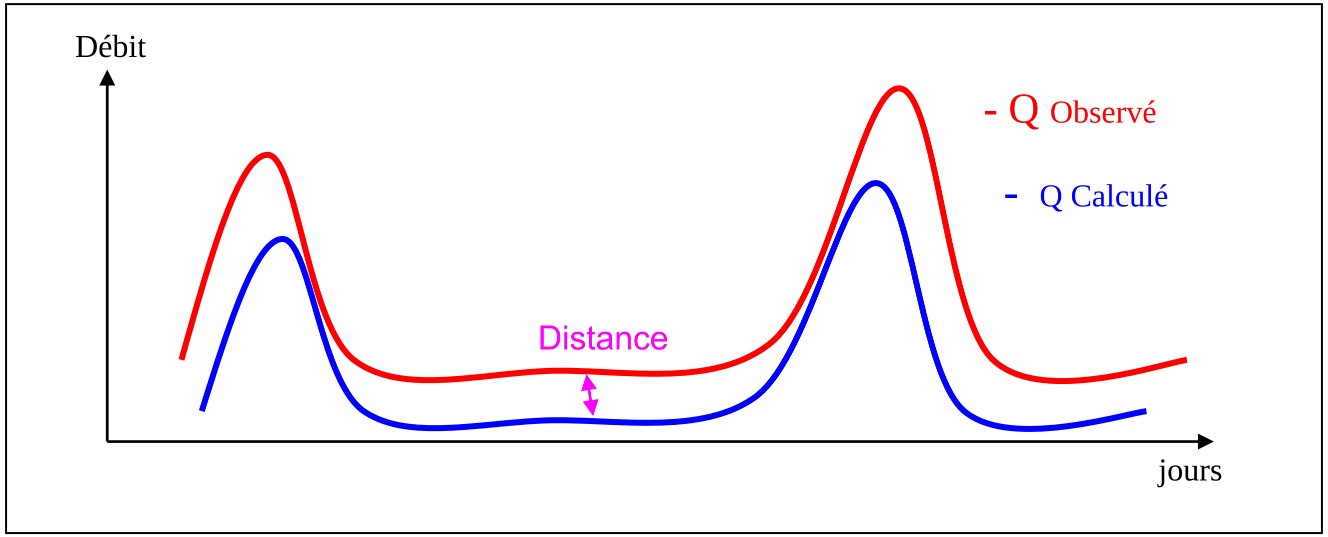
HBV Model

Models parameters

AUFT	: Parameter of direct Runoff
HL1	: Parameter of direct Runoff
Alpha	: non linear transfer reservoir parameter
Beta	: Non linear soil parameter
CEF	: Evapotranspiration Correction Coefficient
Degd (Cmelt)	: Degree day parameter
FC	: Field capacity of soil
K1, K2, K3	: Recession Coefficients of reservoirs
MAXBAS	: Unit hydrograph time parameter
Per1, Per2	: Percolation Coefficients
PWP (LP)	: Threshold of potential evapotranspiration
Vinterc	: capacity of interception
TT	: Threshold temperature
X, K	: Muskingum Parameters



Calibration of HBV model



Min FO (Qio,Qis) , i=1,n

With

FO : Objective function estimating simulation errors
Qic : observed runoff
Qis : simulatd runoff
n : number of time step simulated

Calibration criteria

$$RV = NASH - \omega\,|RD|$$

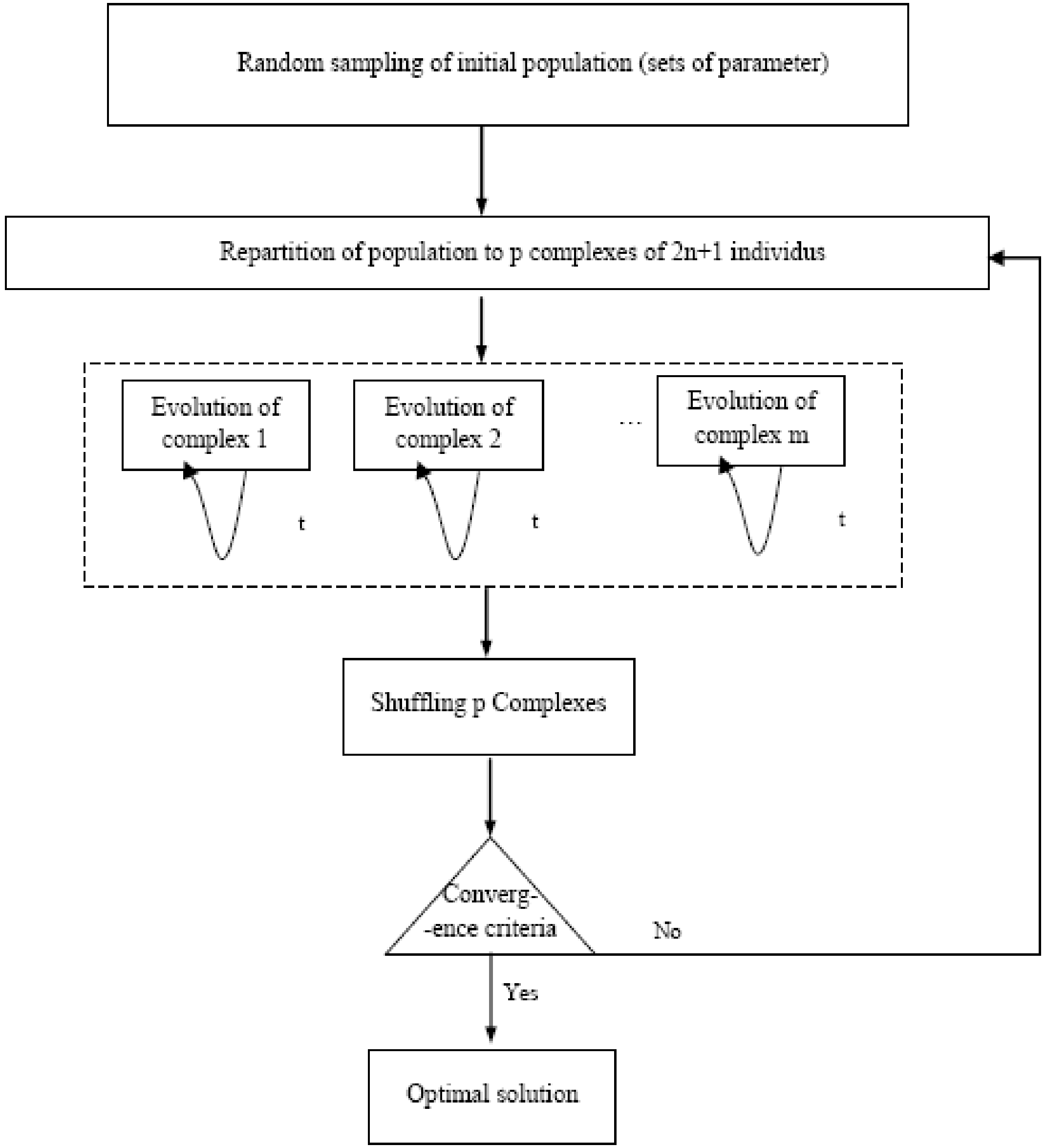
with

$|RD|$: relative error (bias),

w : weight,

Nash : $NASH = 1 - \dfrac{\sum_{i=1}^{n}(q_{ci} - q_{oi})^2}{\sum_{i=1}^{n}(q_{oi} - \bar{q}_o)^2}$

Optimisation algorithm : SCE-UA



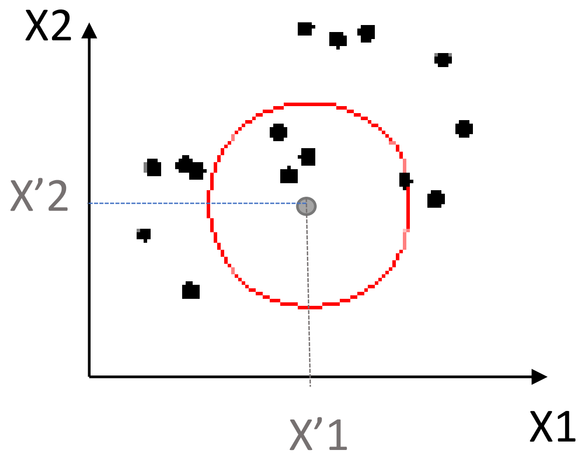
The K-Nearest Neighbour Technique

The K-Nearest Neighbours (KNN) (Fix and Hodges, 1952; Skellam, 1952) may be considered as a learning technique based on classifying objects according to the closeness to their K neighbours in a feature space, where K is the fixed number of neighbours.

The K-Nearest Neighbours estimator is a nonparametric estimator of an n dimensional scalar function F(X), with X a vector of n attributes (x_1 , x_2 ,.., x_n).

Given a data set of p realizations of this function, {(x_{i1} , x_{i2} ,.., x_{in}), F(x_{i1} , x_{i2} ,.., x_{in}); i=1,p}, this technique gives an estimation of F(X') for X' that could take a different value from sample values.

The process is to choose the K-Nearest Neighbours in the feature space {(x_{1l} , x_{2l} ,.., x_{nl}), l=1,K} where K is the number of neighbors, and to compute a KNN estimation of F(X') as the average of the F(X_l) values.



Application de la technique KNN

- We propose to use the KNN technique to estimlate the objective function instead of HBV simulation

- **FO= ∑ 1/di * FOi**

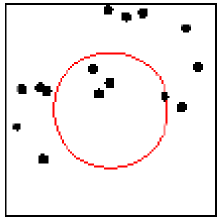
  FO : objective function estimated by KNN at the new set of parameter explored by SCE-UA,

  di : distance betwee the new set of parameter and its neighbour i,

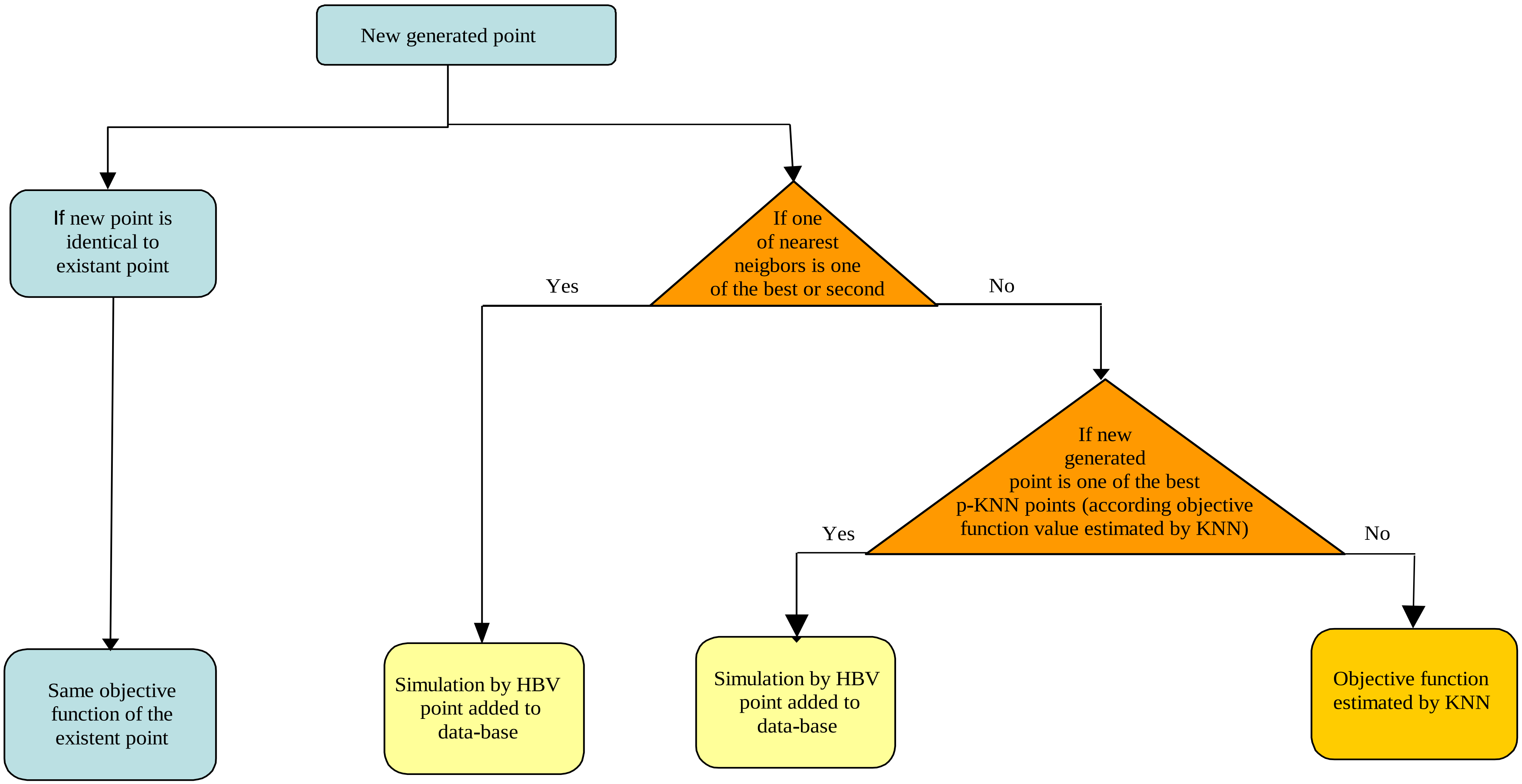
  FOi : objective function evaluated by HBV simulation of neighbours

We use the eucleadian distance in the normalised parameter space

$$D(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}$$

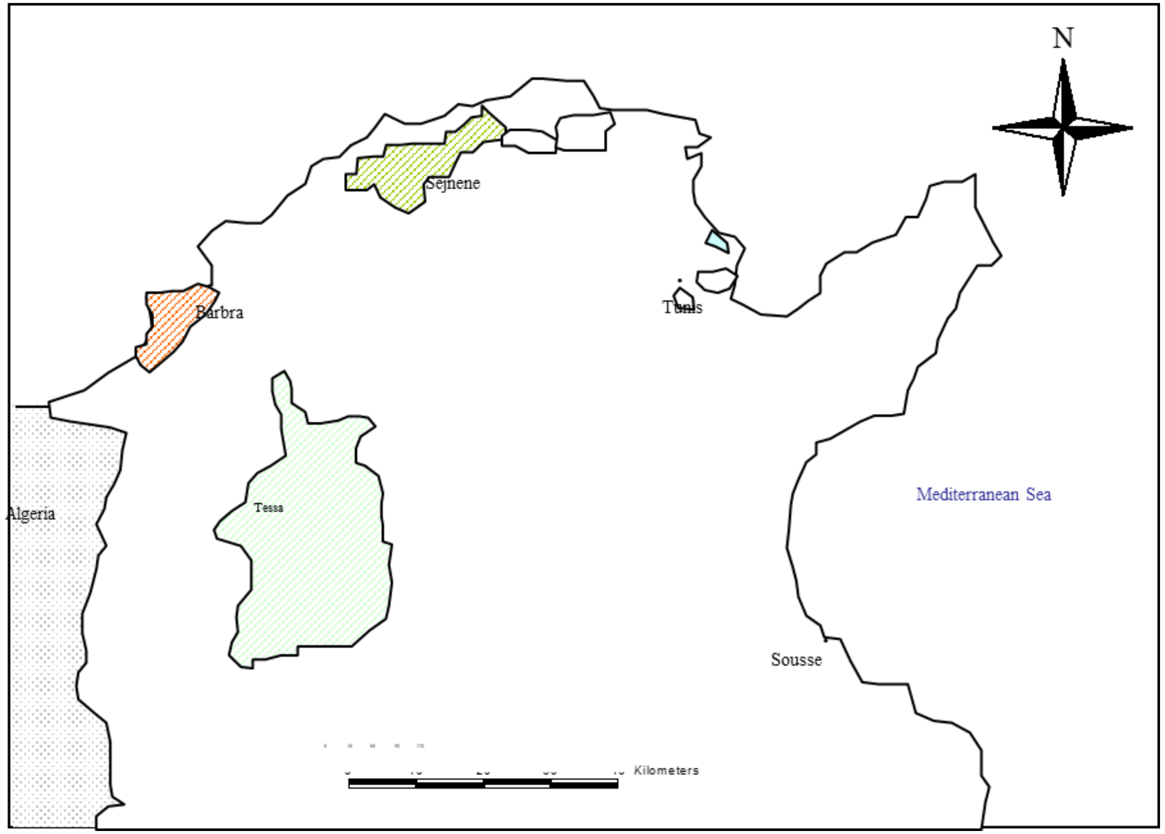


Voisin le plus proche dans l‘espace de paramètre



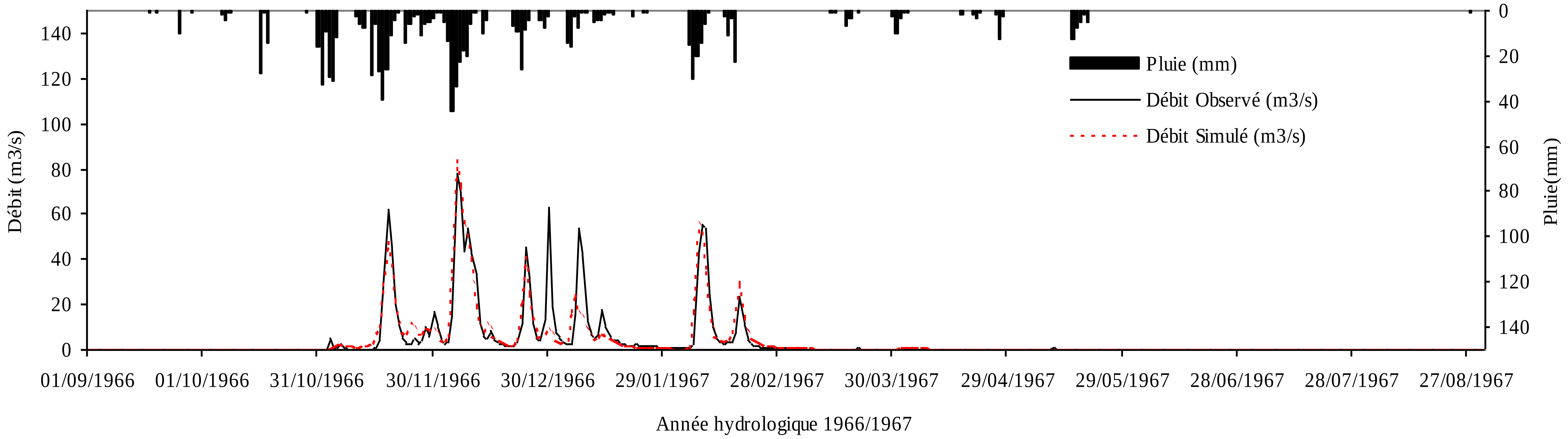
Cases of application of KNN for the estimation of objective function in a new generated point.

The study catchments

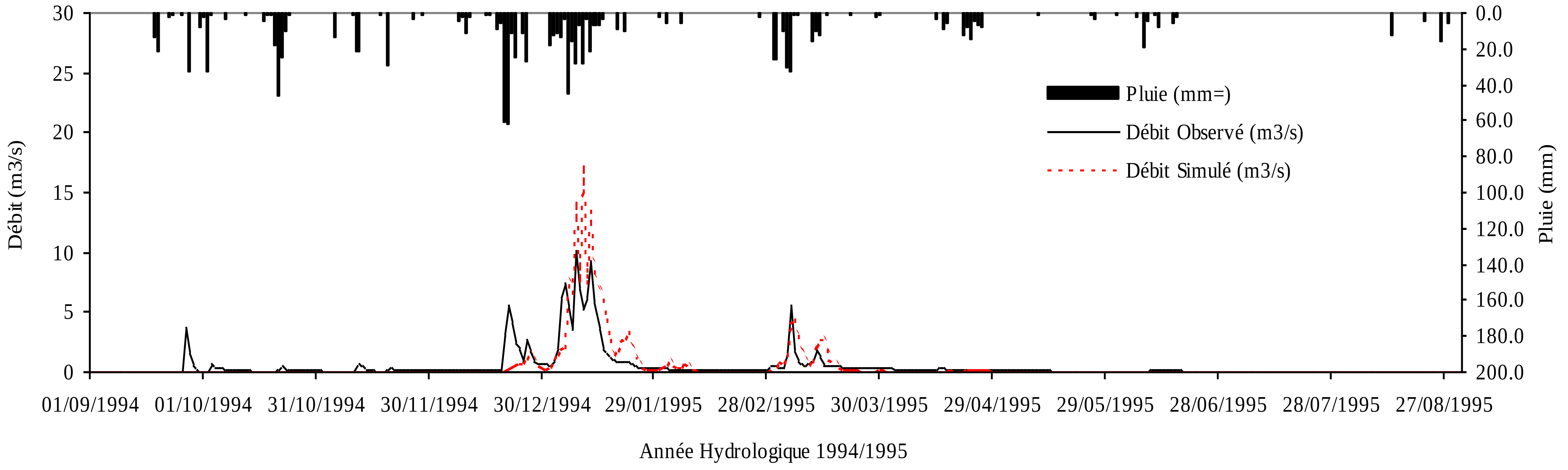


Study Catchments

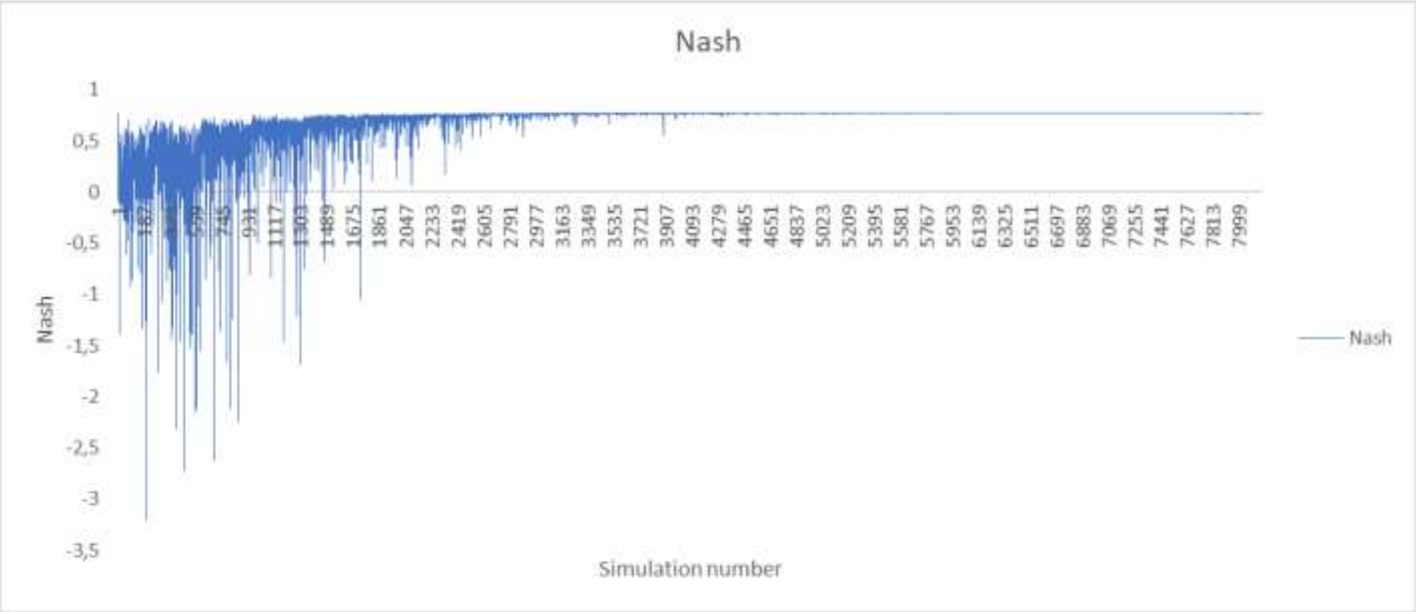
Résultats

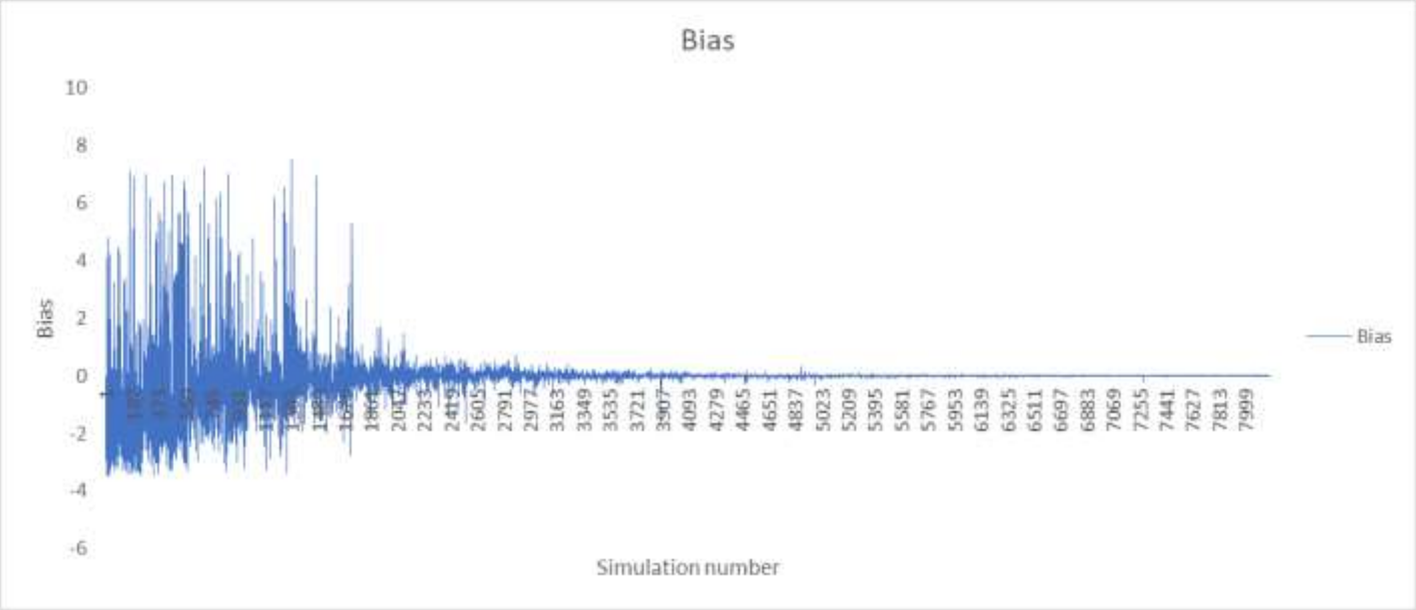


Hydrogrammes des débits observés et calculés (Bassin de Sejnene)

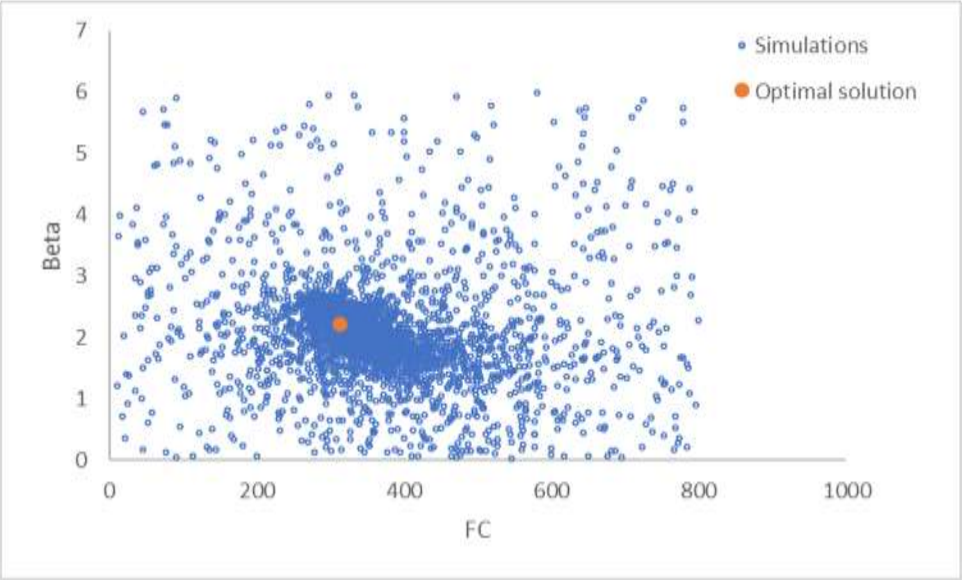


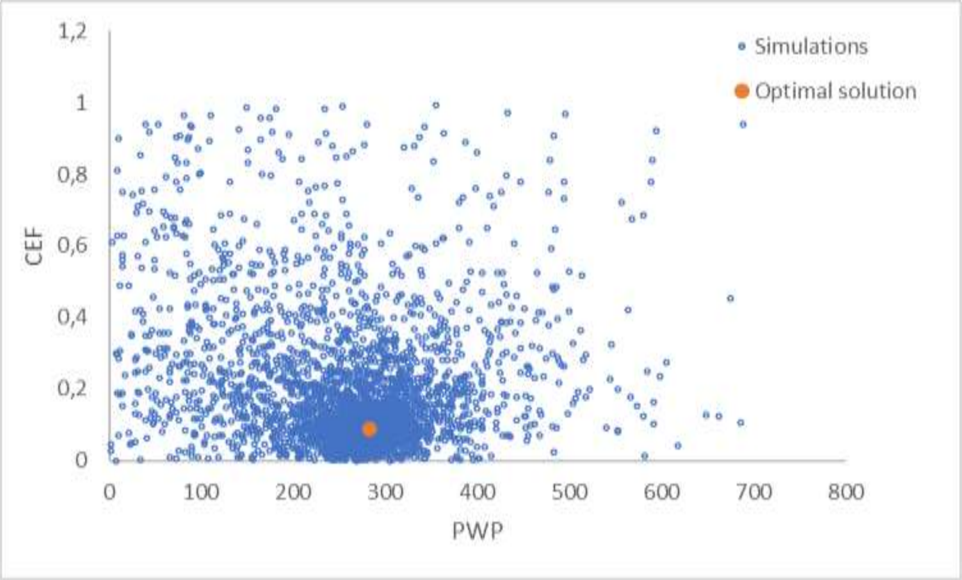
Hydrogrammes des débits observés et calculés (Barbra Validation)



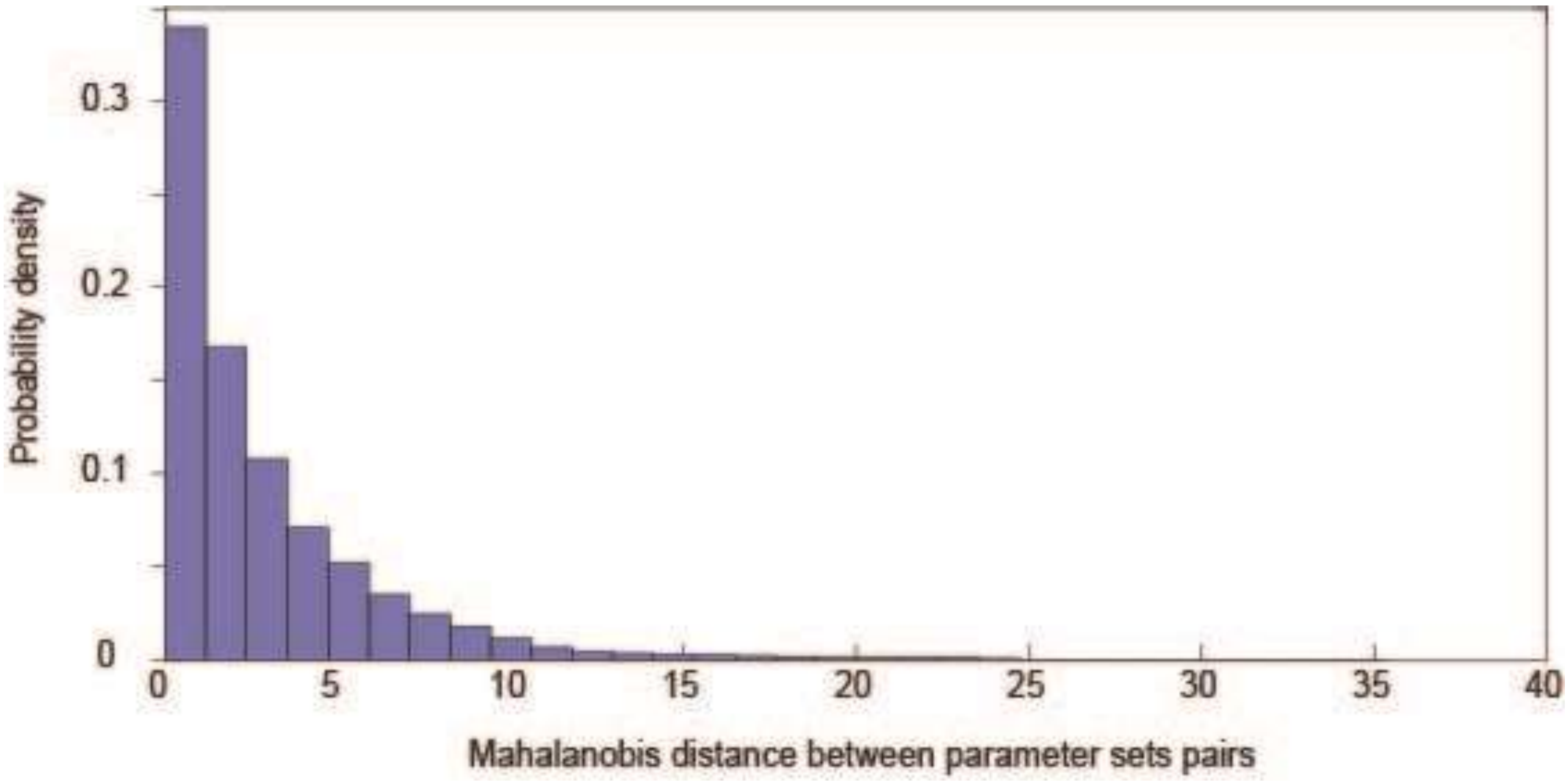


Best objective function value at each iteration during SCE-UA running





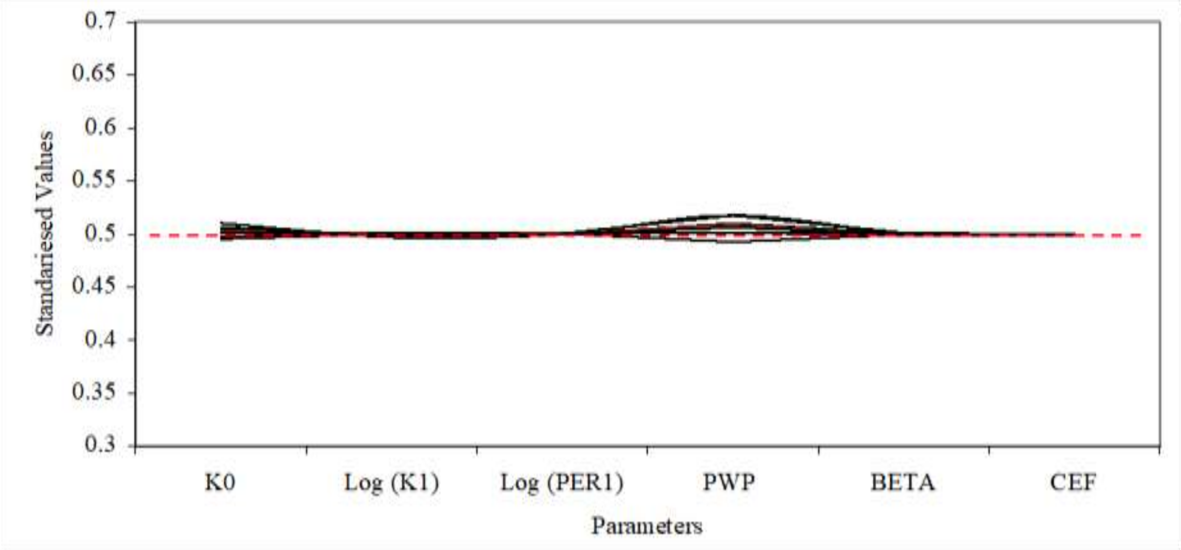
Parameter sets generated by SCE-UA during run (Sejnène Catchment)



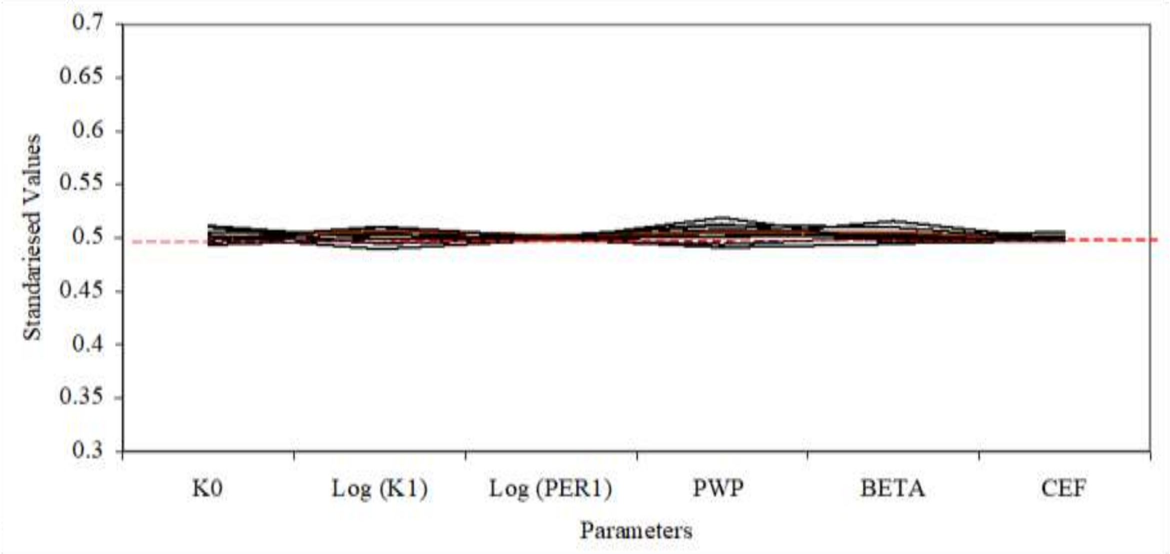
Density function of distance between all pairs of parameter sets generated by SCE-UA for Sejnène catchment(all pairs are considered)

Cases	**p-KNN value**
1	Without KNN
2	10
3	20
4	30
5	40
6	90

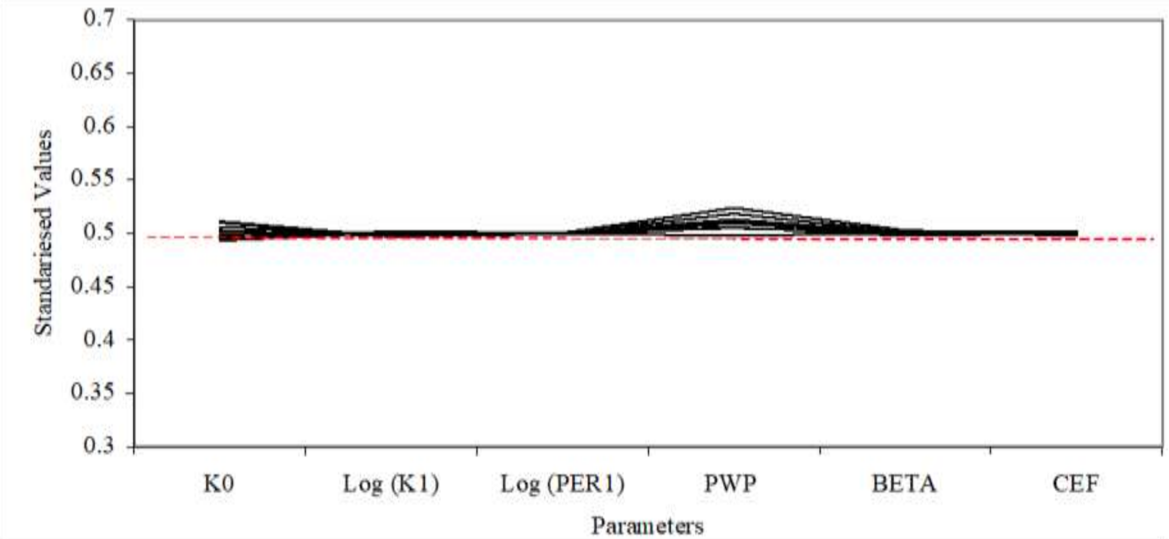
Studied Cases



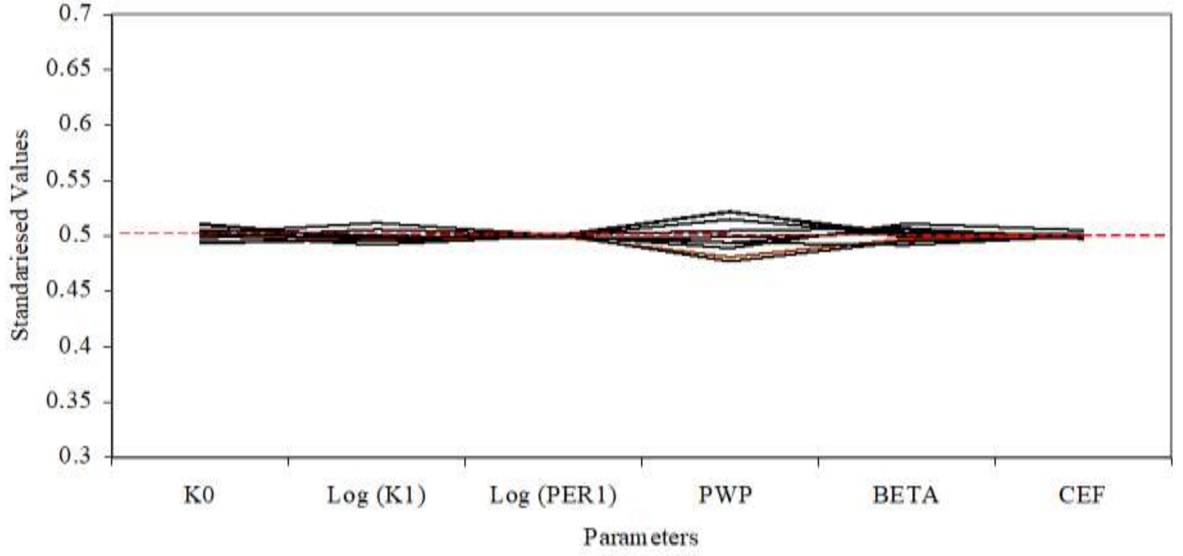
a) Case 1



b) Case 4

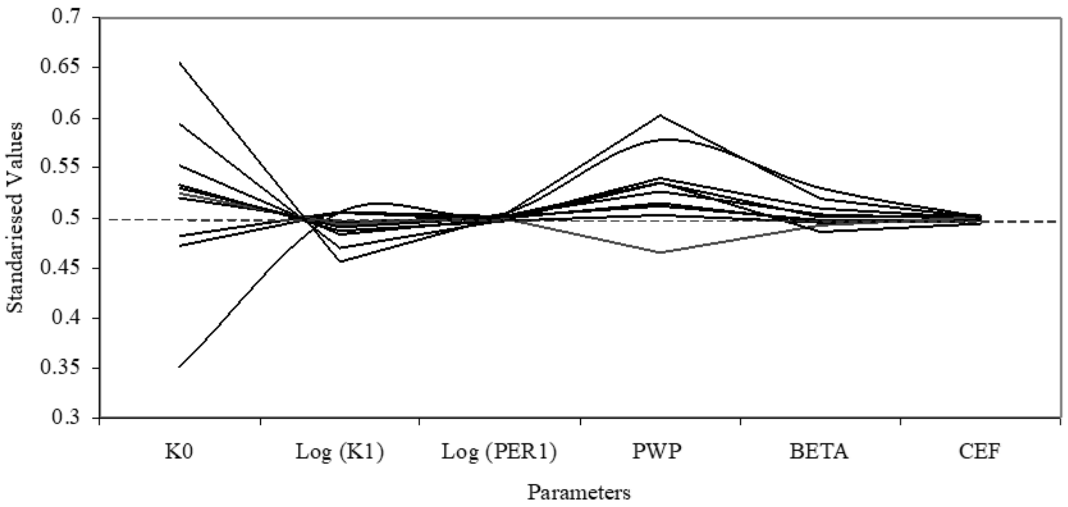


c) Case 2

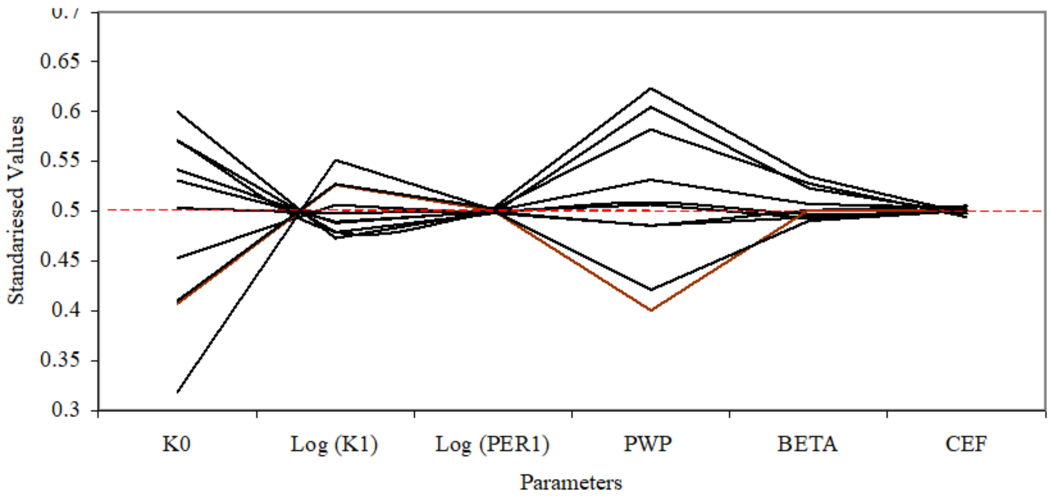


e) Case 3

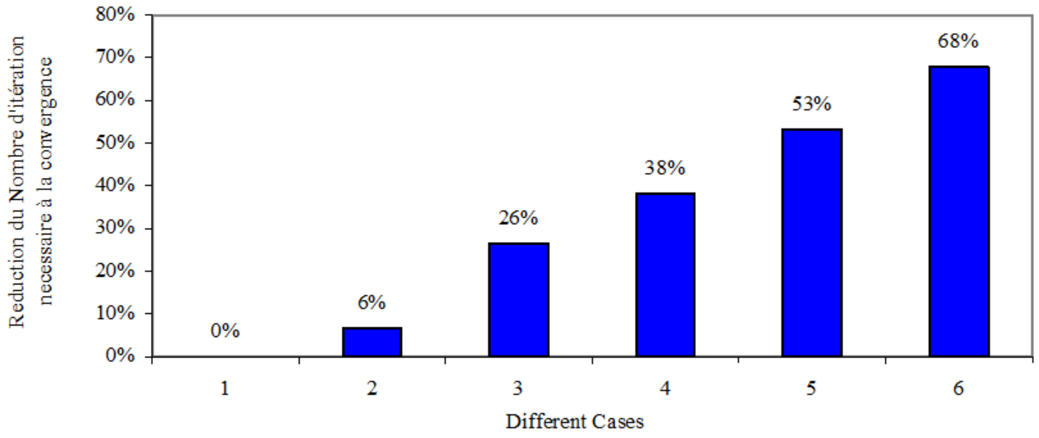
Resultats of application of KNN to SCE-UA



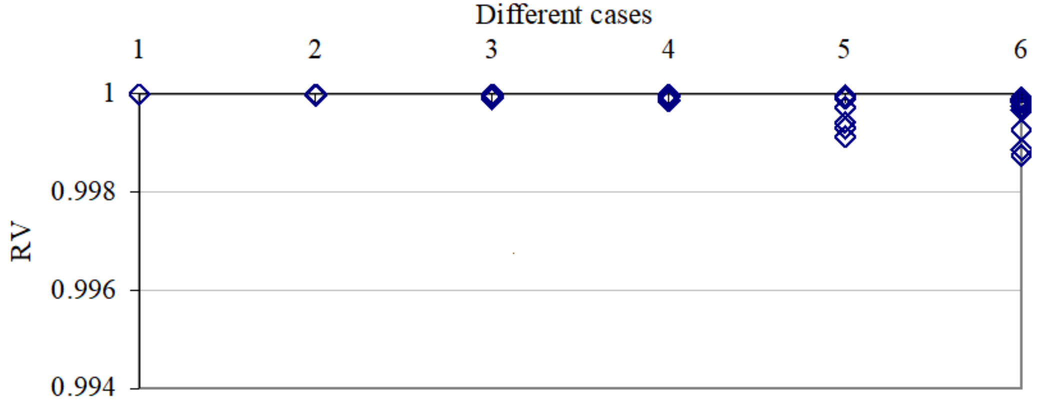
d) Case5



f) Case 6



g)



h)

Resultats of application of KNN to SCE-UA

Conclusion

- In this study we have tried to continue our previous works on improving the efficiency of SCE-UA by using the KNN technique as an alternative estimator for the objective function during the optimization process.
- The implementation of KNN to help estimate the objective function in specific optimization stages allows the possibility to SCE-UA to make use of all the information about the explored points by prior model iterations.
- The adoption of the objective function computation by KNN interpolation from previous estimations was tested in the present. However, to reduce the effect of KNN estimation error, the assessment of objective function by model simulation was restricted according a discrimination of new generated point among a ranking of the explored population by their objective function values.
- On the basis of synthetic discharge data adopted using HBV rainfall-runoff model , the conclusion can be drawn that the efficiency of the optimization algorithm could be improved by about 40 % compared to the initial one.