

Stage post-doctoral UPMC / EDF
Rapport intermédiaire n°2

N. Le Moine

11 mai 2009

1 Rappels sur le contexte de l'étude

L'un des objectifs de l'étude est de mener une analyse de sensibilité des paramètres du modèle hydrologique CEQUEAU, utilisé à EDF R&D, de façon à améliorer sa robustesse. Ce type d'analyse, qui nécessite de nombreux tests de calage du modèle avec différentes combinaisons de paramètres fixés, apparaît difficile à mettre en œuvre de façon un peu systématique avec les procédures de calage utilisées actuellement. Pour cette raison, il est apparu utile de consacrer du temps à améliorer la procédure de calage de CEQUEAU avant tout autre chose.

D'autre part, ce modèle hydrologique a de moins en moins vocation à être utilisé pour la seule simulation des débits à l'exutoire d'un bassin versant. À titre d'exemple, *Labadade* (2008) utilise le modèle CEQUEAU pour produire simultanément des chroniques de débit et de prélèvements agricoles, sorties pour lesquelles des observations sont disponibles. Les paramètres contrôlant la simulation de ces différentes grandeurs cibles nécessiteraient dans l'absolu d'être calés simultanément en prenant en compte les différentes catégories d'observations au cours du calage, ce qui n'est pas possible en l'état.

Ainsi, ces premiers mois de travail ont été consacrés au développement et au test d'un algorithme répondant au double besoin d'une méthode de calage à la fois automatique et multi-objectif. Cet algorithme est décrit dans les sections suivantes.

2 Description de l'algorithme développé pour le calage automatique du modèle CEQUEAU

2.1 Généralités sur l'optimisation multi-objectif

2.1.1 Intérêts

Le but d'une calibration multi-objectif d'un modèle est de trouver des jeux de paramètres présentant un compromis entre différents objectifs potentiellement conflictuels (par exemple, pour un modèle hydrologique, simuler correctement les crues *et* les étiages à l'exutoire, ou bien simuler correctement le débit à l'exutoire *et* en un point intérieur, etc.). Cependant, l'optimisation multi-objectif peut aussi être vue comme un moyen de rajouter des contraintes à un problème sous-contraint, tel que l'identification d'un grand nombre de paramètres : c'est un moyen possible de pallier l'équifinalité de plusieurs vecteurs de paramètres.

Le graphique de la Figure 1 illustre le problème dans le cas suivant :

- nous disposons d'un modèle à $n_\theta = 3$ paramètres libres, notés θ_1, θ_2 et θ_3 . La structure du modèle peut donc être représentée de façon univoque par le point (ou vecteur) de coordonnées $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ dans un espace à $n_\theta = 3$ dimensions, que nous nommerons *espace des paramètres* \mathcal{E}_θ ;
- nous supposons que ce modèle est utilisé pour simuler un vecteur \mathbf{y} de n_y grandeurs observées, par exemple des valeurs journalières de débit en deux stations A et B pendant un an : $n_y = 2 \times 365 = 730$. Cette simulation peut donc être positionnée par un vecteur $\hat{\mathbf{y}}(\boldsymbol{\theta})$ dans un espace à n_y dimensions (non-représentable graphiquement!) que nous appelons *espace des grandeurs observables* \mathcal{E}_y ;
- enfin, nous supposons que tout ou partie de ces n_y valeurs sont utilisées pour calculer un vecteur $\mathbf{f}(\boldsymbol{\theta}, \mathbf{y})$ de n_f fonctions objectif, par exemple ici $\mathbf{f} = (f_1, f_2)$ dans un espace à n_f dimensions, appelé *espace des objectifs* \mathcal{E}_f . Par exemple, si f_1 est le critère de Nash sur les débits en A, f_2 pourrait être le critère de Nash sur le logarithme des débits en A (calibration mono-site avec 2 critères différents), ou bien le critère de Nash sur les débits en B (calibration multi-site avec le même critère pour chaque site), etc.

Dans toute la suite, on s'efforcera de respecter les notations suivantes : les vecteurs et matrices sont notés en gras ($\boldsymbol{\theta}, \mathbf{y}, \mathbf{f}, \boldsymbol{\Sigma} \dots$), les composantes des vecteurs et les scalaires en caractères normaux ($\theta_1, \theta_2, \lambda, \dots$), les espaces ou ensembles en lettres cursives ($\mathcal{E}_\theta, \mathcal{F}, \mathcal{A}, \dots$).

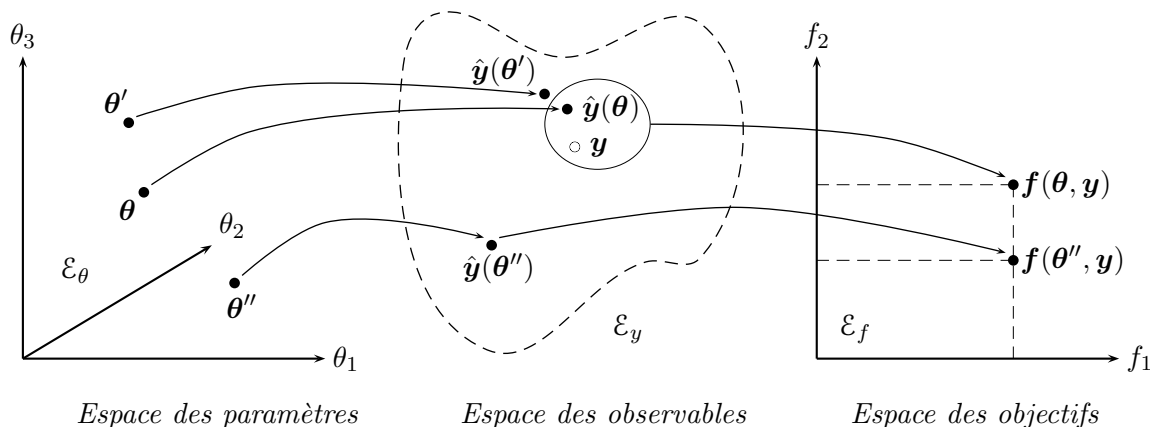


FIGURE 1 – Illustration de l'opération de calibration d'un modèle : passage de l'espace des paramètres (à gauche) à l'espace des observable (centre) puis à l'espace des objectifs (droite).

Cette figure et ces notations permettent de mieux comprendre l'intérêt d'une calibration multi-objectif, en distinguant deux types d'équifinalité :

- I. une équifinalité liée à structure, lorsque deux points $\boldsymbol{\theta}$ et $\boldsymbol{\theta}'$ assez éloignés dans l'espace des paramètres \mathcal{E}_θ peuvent se projeter en des points très proches $\hat{\mathbf{y}}(\boldsymbol{\theta})$ et $\hat{\mathbf{y}}(\boldsymbol{\theta}')$ dans l'espace des observables \mathcal{E}_y , et par suite sur un même point dans l'espace des objectifs \mathcal{E}_f . Il s'agit dans ce cas d'un problème d'écriture du modèle ou d'un défaut de parcimonie : certains paramètres sont certainement redondants ou liés analytiquement (cf. le cas des paramètres *CeqoNPas* et *CeqoTransfert* dans *CEQUEAU* par exemple),

- II. une équi-finalité liée au(x) critère(s) utilisé(s) pour la calibration : en fonction de la dimension de l'espace \mathcal{E}_f des objectifs, et particulièrement dans le cas où $n_f = 1$ (calibration mono-objectif), il est tout à fait possible que des points distincts de l'espace des observables \mathcal{E}_y se projettent sur un même point de \mathcal{E}_f . Sur la Figure 1, les vecteurs $\boldsymbol{\theta}$ et $\boldsymbol{\theta}'$ apparaissent équi-finaux du point de vue de la composante f_1 de la fonction objectif, mais l'ajout de dimensions supplémentaires (ici l'axe f_2) permet de lever cette équi-finalité.

Ainsi, nous pouvons résumer ces deux points en disant qu'and dans l'équi-finalité de type I, c'est la trop grande dimension n_θ de l'espace des paramètres \mathcal{E}_θ qui pose problème, tandis que dans l'équi-finalité de type II c'est la trop faible dimension n_f de l'espace des objectifs \mathcal{E}_f qui est à incriminer.

Bien sûr, on peut argumenter qu'un bon modélisateur doit savoir adapter la complexité de son modèle (i.e., la dimension n_f) à la fois à la quantité de données de validation disponibles (i.e. la dimension n_y de l'espace des grandeurs observables) et aux objectifs de modélisation (i.e. les critères constituant les axes de \mathcal{E}_f). Cependant, rien ne justifie de se limiter systématiquement à une unique fonction objectif, ce qui supposerait d'ailleurs de savoir écrire rigoureusement dans tous les cas la vraisemblance $\mathcal{L}(\boldsymbol{\theta}|\mathbf{y})$ du modèle, ou plus généralement de disposer d'une statistique exhaustive au sens de *Fisher* (1922) résumant toute l'information apportée par les données de calibration concernant la valeur des paramètres. À défaut, on peut recourir à une définition plus faible de l'optimalité et rechercher des jeux de paramètres offrant un compromis acceptable entre plusieurs objectifs non-exhaustifs.

2.1.2 Terminologie

Soient $\boldsymbol{\theta} = (\theta_1, \dots, \theta_{n_\theta})$ et $\boldsymbol{\theta}' = (\theta'_1, \dots, \theta'_{n_\theta})$ deux vecteurs de paramètres de \mathcal{E}_θ , $\mathbf{f}(\boldsymbol{\theta}) = (f_1(\boldsymbol{\theta}), \dots, f_{n_f}(\boldsymbol{\theta}))$ et $\mathbf{f}(\boldsymbol{\theta}') = (f_1(\boldsymbol{\theta}'), \dots, f_{n_f}(\boldsymbol{\theta}'))$ leurs projections dans l'espace des objectifs \mathcal{E}_f à n_f dimensions¹.

On dira que $\mathbf{f}(\boldsymbol{\theta}')$ domine $\mathbf{f}(\boldsymbol{\theta})$ (et par abus de langage que $\boldsymbol{\theta}'$ domine $\boldsymbol{\theta}$) si et seulement si on vérifie les conditions :

$$\begin{cases} \forall i \in [1, n_f], & f_i(\boldsymbol{\theta}) \leq f_i(\boldsymbol{\theta}') \\ \exists j \in [1, n_f] / & f_j(\boldsymbol{\theta}) < f_j(\boldsymbol{\theta}') \end{cases}$$

Si $\boldsymbol{\theta}$ est un vecteur de paramètres pour lequel il n'existe aucun $\boldsymbol{\theta}' \in \mathcal{E}_\theta$ tel que $\mathbf{f}(\boldsymbol{\theta}')$ domine $\mathbf{f}(\boldsymbol{\theta})$, alors cette solution est dite *optimale au sens de Pareto* (ou *Pareto-optimale*) : aucune amélioration ne peut être apportée à quelque fonction objectif que ce soit sans dégrader (au sens strict) au moins l'une des autres.

Ces solutions non-dominées forment, dans l'espace des objectifs, un ensemble appelé *front de Pareto* \mathcal{F} . Il s'agit d'une hypersurface de \mathcal{E}_f i.e. un ensemble de dimension $(n_f - 1)$. Il est illustré sur la Figure 2 en dimension $n_f = 2$ (il s'agit alors d'une ligne, qui peut être discontinue). Cet ensemble est l'antécédent par \mathbf{f} d'un sous-ensemble \mathcal{A} de \mathcal{E}_θ : $\mathcal{A} = \{\boldsymbol{\theta} \in \mathcal{E}_\theta / \mathbf{f}(\boldsymbol{\theta}) \in \mathcal{F}\}$. Cet ensemble \mathcal{A} n'a, dans le cas général, aucune raison d'être convexe.

1. Dans les notations, on "court-circuitera" à présent l'espace des observables en raisonnant à \mathbf{y} fixé.

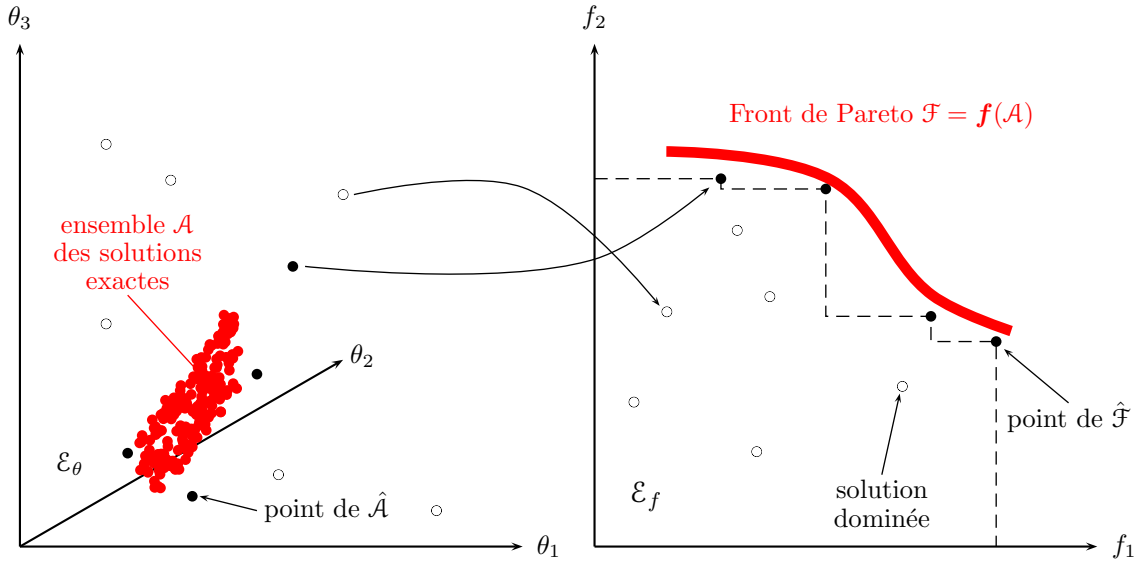


FIGURE 2 – Illustration des concepts de dominance et de Pareto-optimalité.

Le but d'un algorithme d'optimisation multi-objectif est d'approcher le front de Pareto par un ensemble fini de points de \mathcal{E}_θ , noté $\hat{\mathcal{F}}$ (front de Pareto approximé). Dans la suite, nous nommerons *archive* $\hat{\mathcal{A}}$ l'antécédent par \mathbf{f} de $\hat{\mathcal{F}}$: $\hat{\mathcal{A}} = \{\boldsymbol{\theta} \in \mathcal{E}_\theta / \mathbf{f}(\boldsymbol{\theta}) \in \hat{\mathcal{F}}\}$.

2.2 Description détaillée de l'algorithme

L'algorithme d'optimisation multi-objectif présenté ici appartient à la famille des algorithmes génétiques. L'idée de ces algorithmes est de partir d'un ensemble initial de solutions (appelé "population"), que l'on fait évoluer en suivant des règles de génération de nouvelles solutions et en ne conservant que les meilleures (les plus "adaptées") à chaque génération.

Cet algorithme fonctionne de façon optimale lorsque la variance des paramètres dans la population initiale est assez proche de leur variance a priori (i.e., la population initiale est assez représentative de la diversité des paramétrisations possibles du modèle). Idéalement, on pourra prendre comme population initiale une librairie de jeux de paramètres issus de calages précédents.

Dans cette subsection, on décrit en détail la façon dont sont générés les nouveaux jeux de paramètres à chaque génération, et la façon dont on définit le degré "d'adaptation" dans un contexte multi-objectif. L'algorithme est largement inspiré de deux algorithmes existants :

- l'algorithme MEAS (*Efstratiadis et Koutsoyiannis, 2005*), dont il reprend la méthode de recherche directionnelle, basée sur les simplexes de l'espace des objectifs.
- l'algorithme ε -NSGA-II (*Reed et Devireddy, 2004*), dont il reprend le mode de classement des vecteurs de paramètres et la gestion de la précision par ε -dominance,

2.2.1 Règles de génération de nouvelles solutions


L'algorithme mis au point possède quatre règles permettant de générer de nouvelles solutions à chaque génération (le nombre total de nouvelles solutions et la proportion générée selon chaque règle peuvent être modifiés par l'utilisateur).

Les deux premières règles (interpolation et extrapolation) s'appuient sur la triangulation des points dans l'espace des objectifs. Elles partent du principe que deux points voisins dans l'espace \mathcal{E}_f des objectifs ont pour antécédents deux points voisins dans l'espace \mathcal{E}_θ des paramètres, i.e. que la fonction

$$\begin{aligned} f : \quad \mathcal{E}_\theta &\longrightarrow \mathcal{E}_f \\ \theta &\longmapsto f(\theta) \end{aligned}$$

est continue et que par conséquent, on peut essayer de “deviner” les directions d'amélioration dans l'espace des paramètres à partir des directions d'amélioration (au sens de Pareto) dans l'espace des objectifs, au moins à proximité de la zone optimale.

Les deux dernières règles de génération créent des nouveaux points en explorant l'espace des paramètres de façon non-directionnelle et moins locale, en tirant de nouveaux paramètres soit indépendamment et avec leur variance a priori (règle n°3), soit en respectant la structure de corrélation entre les solutions situées au niveau du front de Pareto approximé (règle n°4).

 Règle n°1 : Interpolation

Considérons, comme sur la Figure 3 à droite, une triangulation de l'ensemble des solutions dans l'espace des objectifs \mathcal{E}_f i.e. l'ensemble des simplexes² construits avec les points $f(\theta_i)$ et formant une partition de la zone explorée dans cet espace (voir *Efstratiadis et Koutsoyiannis, 2005*).

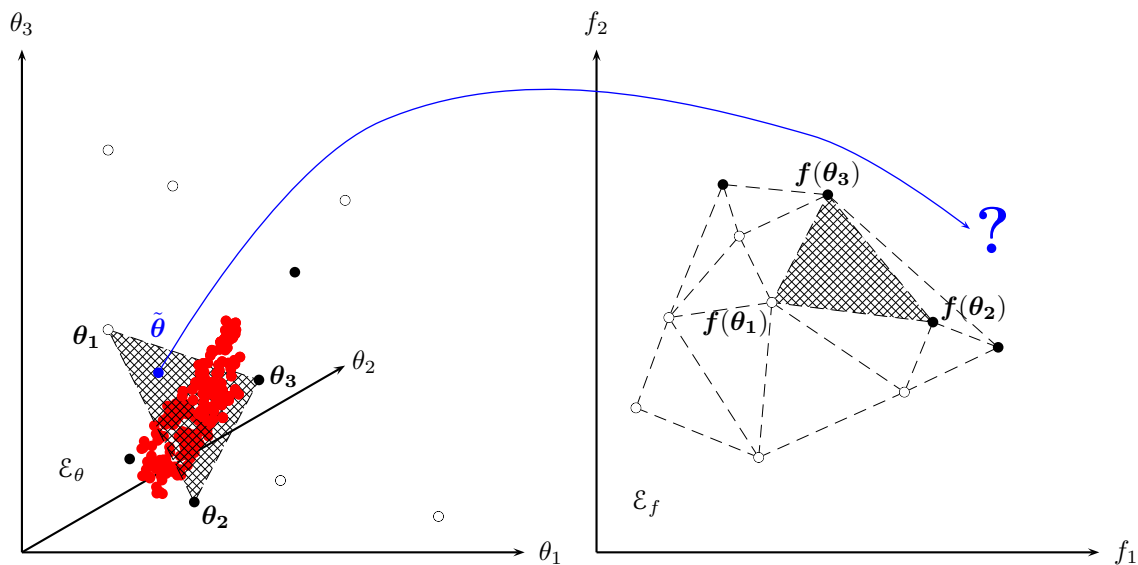


FIGURE 3 – Illustration de la règle n°1 d'interpolation. On commence par construire (figure de droite) une triangulation de l'espace des objectifs sur la base des solutions déjà évaluées puis, pour chaque simplexe possédant au moins un sommet non-dominé (i.e. appartenant au front de Pareto approximé), on interpole un vecteur $\tilde{\theta}$ dans l'espace des paramètres (avec une probabilité proportionnelle à son volume).

2. Dans un espace à n dimensions, un n -simplexe est constitué de $(n + 1)$ points : un 1-simplexe est un segment, un 2-simplexe est un triangle, un 3-simplexe est un tétraèdre, et au-delà il est impossible de se le représenter !

Isolons maintenant un simplexe ayant au moins l'un de ses sommets sur le front de Pareto approximé à la génération courante. Ce simplexe est l'image par la fonction \mathbf{f} d'un ensemble de $(n_f + 1)$ points de l'espace des paramètres \mathcal{E}_θ , qui est lui de dimension n_θ . Sous l'hypothèse de continuité de \mathbf{f} , il est possible qu'une combinaison linéaire "interpolante", de forme

$$\tilde{\boldsymbol{\theta}} = w_1 \boldsymbol{\theta}_1 + \dots + w_{(n_f+1)} \boldsymbol{\theta}_{(n_f+1)}$$

avec $w_i \geq 0 \forall i$ et $\sum_i w_i = 1$ (coordonnées barycentriques),

donne une nouvelle solution $\mathbf{f}(\tilde{\boldsymbol{\theta}})$ Pareto-optimale à l'intérieure de cette zone non-explorée.

En pratique, la triangulation est obtenue par l'appel du programme externe Qhull³. Ce programme renvoie la liste des simplexes, dont on calcule ensuite le volume. La probabilité de générer un point à partir d'un simplexe avec la règle d'interpolation est proportionnelle à son volume s'il possède au moins un sommet dans le front de Pareto approximé, et nulle sinon. Pour cette raison, il est plus pratique d'utiliser des fonctions objectifs bornées, de façon à ce que le volume des simplexes ait une limite dans l'espace des objectifs.

Si le simplexe est sélectionné, alors on tire un jeu de coordonnées barycentriques en tirant d'abord $(n_f + 1)$ valeurs ε_i dans une loi uniforme sur $[0,1]$ puis en calculant

$$w_i = \frac{\varepsilon_i}{\sum_{j=1}^{(n_f+1)} \varepsilon_j}$$

Règle n°2 : Extrapolation

La seconde règle de génération part de la même hypothèse de continuité mais en testant cette fois si une amélioration peut être obtenue par *extrapolation* le long de certaines directions. Pour choisir ces directions, on repart de la triangulation précédente et on sélectionne les arêtes qui vérifient les deux propriétés suivantes :

- l'un des 2 sommets appartient au front de Pareto approximé : notons ce sommet $\mathbf{f}(\boldsymbol{\theta}_1)$,
- l'autre sommet n'appartient pas au front et est dominé par le premier : notons le $\mathbf{f}(\boldsymbol{\theta}_2)$.

Nous voyons sur la Figure 4, schéma de droite, que ces arêtes ainsi orientées correspondent à des directions de recherche dans l'espace des paramètres qui améliorent simultanément toutes les composantes de la fonction objectif.

3. Implémentation libre de l'algorithme Quickhull, pour le calcul d'enveloppes convexes. Voir <http://www.qhull.org>

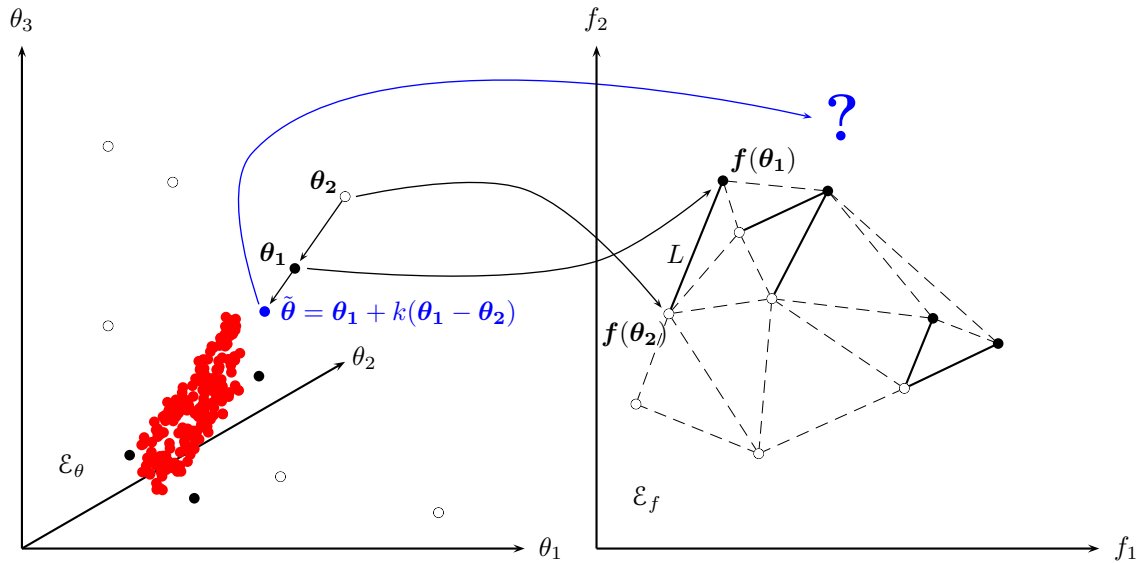


FIGURE 4 – Illustration de la règle n°2 d’extrapolation. On repart de la triangulation de l’espace des objectifs (schéma de droite) en identifiant les directions d’amélioration et en les extrapolant dans l’espace des paramètres (figure de gauche).

On calcule la longueur $L = \|\mathbf{f}(\boldsymbol{\theta}_1) - \mathbf{f}(\boldsymbol{\theta}_2)\|$ de chaque arête ainsi sélectionnée dans l’espace des objectifs, et leur longueur moyenne \bar{L} . Cette longueur moyenne représente donc l’amélioration moyenne que l’on obtient entre un point juste derrière le front de Pareto approximé, et le point du front le plus proche : on va l’utiliser pour mettre à l’échelle le vecteur de recherche dans l’espace des paramètres.

La probabilité d’utiliser un arête comme support de recherche est proportionnelle à sa longueur L . Si elle est sélectionnée, le vecteur de recherche dans l’espace des paramètres est alors défini par :

$$\mathbf{U} = \frac{L}{\bar{L}}(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$$

et on génère un nouveau vecteur de paramètres par :

$$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta}_1 + \lambda \mathbf{U}$$

où λ est un scalaire tiré dans une loi exponentielle de moyenne 1.

📎 Règle n°3 : Echantillonnage indépendant avec variance a priori des paramètres

Avec les deux premières règles, l’inconvénient est que l’on génère des nouveaux vecteurs par combinaisons linéaires d’un petit nombre de vecteurs existants ($n_f + 1$ avec la règle n°1, 2 avec la règle n°2) : par conséquent, on cherche les nouvelles solutions en se déplaçant localement dans des sous-espaces de \mathcal{E}_θ dont la dimension est très inférieure à la dimension de cet espace si, comme c’est souvent le cas, $n_\theta \gg n_f$. Pour compenser cette recherche par gradient qui peut, le cas échéant, converger vers un optimum secondaire, on ajoute une troisième règle de génération permettant :

- de faire varier les paramètres avec un pas plus large qu’avec les règles locales précédentes,
- de les faire varier indépendamment les uns des autres, de façon à casser les corrélations factices.

La troisième règle est simple à décrire : de temps à autre, on choisit un vecteur $\boldsymbol{\theta}$ de l'archive $\hat{\mathcal{A}}$ et on génère n_θ nouveaux vecteurs en faisant varier individuellement chacune des composantes de $\boldsymbol{\theta}$:

$$\text{pour } k \text{ de } 1 \text{ à } n_\theta, \quad \tilde{\boldsymbol{\theta}}_k = \boldsymbol{\theta} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \sigma_k \varepsilon_k \\ 0 \\ \vdots \\ 0 \end{pmatrix} \begin{array}{l} \text{composante } k \\ \\ \\ \text{composante } n_\theta \end{array}$$

où σ_k^2 est la variance a priori du k -ième paramètre, et ε_k une v.a. tirée dans une loi normale de moyenne nulle et de variance unité.

Dans l'algorithme, on choisit des points particuliers de l'archive pour effectuer cette opération : les vecteurs de paramètres qui maximisent individuellement chaque fonction objectif, soit n_f vecteurs. Ainsi, quand on décide d'appliquer la règle n°3, on génère $(n_f \times n_\theta)$ nouveaux vecteurs d'un seul coup. Pour cette raison, on n'applique pas cette règle à chaque génération, mais seulement toutes les K générations, la valeur de K étant laissée au choix de l'utilisateur.

📎 Règle n°4 : Echantillonnage avec respect d'une structure de corrélation

Soit $\boldsymbol{\theta}$ un vecteur de l'ensemble \mathcal{A} précédemment défini (ensemble des solutions Pareto-optimales, que nous cherchons à approximer par $\hat{\mathcal{A}}$). Définissons $\boldsymbol{\mu} = \mathbb{E}_{\boldsymbol{\theta} \in \mathcal{A}} [\boldsymbol{\theta}]$ le barycentre de \mathcal{A} , et $\boldsymbol{\Sigma}$ la matrice de variance-covariance (symétrique définie positive) :

$$\boldsymbol{\Sigma} = \mathbb{E}_{\boldsymbol{\theta} \in \mathcal{A}} [(\boldsymbol{\theta} - \boldsymbol{\mu})^t (\boldsymbol{\theta} - \boldsymbol{\mu})]$$

où $\mathbb{E}[X]$ désigne l'espérance mathématique d'une variable aléatoire X (scalaire, vecteur ou matrice) et ${}^t M$ la transposée de la matrice M .

Sur la diagonale de $\boldsymbol{\Sigma}$, on trouve la variance de chaque paramètre, et en ligne i , colonne j , la covariance $\sigma_{i,j} = \sigma_{j,i}$ entre θ_i et θ_j :

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n_\theta} \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{2,n_\theta} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,n_\theta} & \sigma_{2,n_\theta} & \cdots & \sigma_{n_\theta}^2 \end{pmatrix}$$

Cette matrice reflète la structure de corrélation entre les jeux de paramètres qui sont solutions du problème : en effet, dans le cas d'un modèle hydrologique, il semble logique que les paramètres ne puissent pas prendre n'importe quelle valeur indépendamment les uns des autres. Ainsi, si l'on pouvait obtenir une estimation $\hat{\boldsymbol{\Sigma}}$ de $\boldsymbol{\Sigma}$ et une estimation $\hat{\boldsymbol{\mu}}$ de $\boldsymbol{\mu}$, alors on pourrait générer de nouveaux vecteurs de paramètres respectant la "bonne" structure de corrélation : cela limiterait le risque de générer des jeux "non-fonctionnels" qui sont autant d'évaluations inutiles de la fonction objectif f .

La question est alors : à partir de quel sous-ensemble de vecteurs de paramètres *déjà testés* faut-il estimer la matrice de covariance pour générer les nouveaux points à la génération courante ? Par définition, nous ne connaissons ni \mathcal{A} ni un même un échantillon issu de \mathcal{A} , puisque c'est ce que nous cherchons... Plusieurs solutions sont défendables, comme par exemple :

- la matrice de covariance a priori, construite à partir de la librairie de vecteurs de paramètres “historiques” : elle a l'avantage de permettre une exploration large de l'espace des paramètres, mais l'inconvénient de ne pas tenir compte des progrès que l'on accomplit à chaque génération dans l'estimation de l'ensemble \mathcal{A} des solutions par $\hat{\mathcal{A}}$;
- la matrice de covariance des vecteurs de l'archive $\hat{\mathcal{A}}$ à la génération courante (i.e., la meilleure estimation possible de l'ensemble des solutions jusqu'à présent). Testée en pratique, cette solution génère effectivement des jeux qui viennent presque à coup sûr enrichir la zone du front de Pareto approximé. Cependant, elle a l'inconvénient de son avantage dans la mesure où elle ne crée pas de jeux radicalement nouveaux, la variance des paramètres du front étant en général bien plus faible que la variance des paramètres de la population complète. Ainsi, elle ne permet pas plus que les règles locales 1 et 2 de sortir d'un éventuel optimum local. De plus, l'archive $\hat{\mathcal{A}}$ ne contient pas toujours un nombre de vecteurs suffisant pour calculer une matrice de covariance significative ;
- la matrice de covariance des vecteurs de la population complète à la génération courante (i.e., vecteurs dominés et non-dominés confondus). Ce choix semble un compromis acceptable entre maintien de la diversité et optimalité des nouveaux vecteurs générés. Cependant il s'avère, là encore d'après la pratique, que son coût calculatoire reste assez élevé : pour une dizaine de vecteurs générés avec cette matrice à chaque génération, il est rare d'en observer un qui vienne enrichir le front de Pareto approximé.

Ces conclusions, tirées après essais et erreurs, conduisent à préférer une solution intermédiaire entre les points deux et trois : nous repartons de la triangulation utilisée à la règle n°1, en sélectionnant tous les simplexes qui ont au moins un sommet dans le front de Pareto approximé. Nous choisissons alors comme vecteurs de référence pour le calcul de la matrice de covariance l'ensemble \mathcal{G} des antécédents par \mathbf{f} de tous les sommets de ces simplexes “frontaux” :

$$\mathcal{G} = \{\boldsymbol{\theta} \in \mathcal{E}_\theta \text{ tels que } \mathbf{f}(\boldsymbol{\theta}) \text{ est sommet d'un simplexe "frontal"}\}$$

On prend donc :

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \mathbb{E}_{\boldsymbol{\theta} \in \mathcal{G}} [\boldsymbol{\theta}] \\ \hat{\boldsymbol{\Sigma}} &= \mathbb{E}_{\boldsymbol{\theta} \in \mathcal{G}} [(\boldsymbol{\theta} - \hat{\boldsymbol{\mu}})^t (\boldsymbol{\theta} - \hat{\boldsymbol{\mu}})]\end{aligned}$$

Ce choix a le mérite d'augmenter sensiblement le nombre de points sélectionnés pour calculer les moyennes. Cependant, le problème mentionné au deuxième point, à savoir le risque d'avoir une variance trop faible, subsiste. Pour le pallier, nous utilisons un artifice consistant à augmenter légèrement, d'un même facteur, la variance de tous les paramètres, en utilisant la matrice :

$$\hat{\hat{\boldsymbol{\Sigma}}} = (1 + \alpha) \hat{\boldsymbol{\Sigma}}, \quad \alpha > 0$$

Empiriquement, la valeur $\alpha=1$ paraît satisfaisante (échantillonnage avec $\sqrt{1+\alpha} \simeq 1.4$ fois l'écart-type des paramètres dans \mathcal{G}). Les nouveaux vecteurs sont alors obtenus en suivant une procédure classique de génération multivariée :

1. Tout d'abord, on calcule la matrice \mathbf{T} triangulaire supérieure telle que ${}^t\mathbf{T}\mathbf{T} = \hat{\hat{\boldsymbol{\Sigma}}}$, par factorisation de Cholesky (fonction `chol` dans Scilab) ;

2. Ensuite, on génère des vecteurs de forme $\tilde{\theta} = \hat{\mu} + {}^t\mathbf{T} \cdot \varepsilon$, où ε est un vecteur à n_θ composantes indépendantes normalement distribuées, de moyenne 0 et de variance 1. Les vecteurs ainsi générés ont donc bien la structure de corrélation désirée :

$$\begin{aligned}
\mathbb{E} \left[\left(\tilde{\theta} - \mathbb{E}[\tilde{\theta}] \right) \cdot {}^t \left(\tilde{\theta} - \mathbb{E}[\tilde{\theta}] \right) \right] &= \mathbb{E} \left[({}^t\mathbf{T}\varepsilon) \cdot {}^t({}^t\mathbf{T}\varepsilon) \right] \\
&= \mathbb{E} \left[{}^t\mathbf{T} \varepsilon \varepsilon {}^t\mathbf{T} \right] \\
&= {}^t\mathbf{T} \mathbb{E} \left[\varepsilon \varepsilon {}^t \right] \mathbf{T} \\
&= {}^t\mathbf{T} \mathbf{I} \mathbf{T} \\
&= {}^t\mathbf{T} \mathbf{T} \\
&= \hat{\hat{\Sigma}}
\end{aligned}$$

L'ajout de cette quatrième règle, qui permet d'explorer aléatoirement des zones de l'espace \mathcal{E}_θ tout en réduisant implicitement sa dimension par le biais des corrélations observées entre paramètres, diminue assez nettement le nombre d'évaluations de la fonction objectif.

2.2.2 Réduction de la taille de la population à la fin de chaque génération

À la fin de chaque génération, on ne souhaite conserver qu'un nombre N_{\max} de points, à la fois pour des raisons de mémoire (il est inutile de conserver les résultats relatifs à des vecteurs de paramètres peu performants) et de temps de calcul de la triangulation dans l'espace des objectifs (qui est effectuée à chaque génération). La réduction de la taille de la population s'effectue en trois étapes :

1. On commence par ordonner les points en fonction du front de Pareto auquel ils appartiennent (Figure 5) : le front d'ordre 1 est l'ensemble des solutions non dominées ; le front d'ordre 2 est l'ensemble des solutions qui ne sont dominées que par des solutions du front d'ordre 1, etc. ; de façon générale, le front d'ordre i est l'ensemble des solutions qui ne sont dominées que par des solutions appartenant aux fronts d'ordres strictement inférieurs à i (soit de 1 à $i - 1$),
2. On découpe ensuite l'espace des objectifs avec une précision δ_i désirée pour chaque fonction objectif, comme illustré sur la Figure 6. Tous les points situés dans un même hypercube sont considérés comme équifinaux à la précision près, on n'en conserve donc qu'un seul. Le point retenu est celui appartenant au front d'ordre le plus faible ; en cas d'égalité, on tire au hasard parmi les points appartenant à ce front,
3. Si, à l'issue de l'étape précédente, la taille N de la population excède encore N_{\max} , on supprime les $(N - N_{\max})$ points situés dans les fronts de plus grands ordres.

Ces méthodes sont reprises de l'algorithme ε -NSGA-II (*Reed et Devireddy, 2004*).

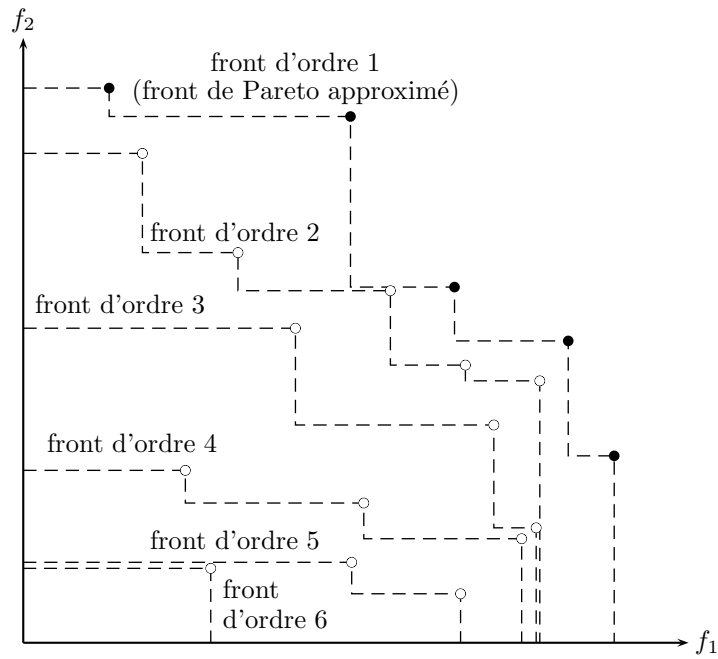


FIGURE 5 – Méthode de classement des points en fronts successifs dans l'espace des objectifs.

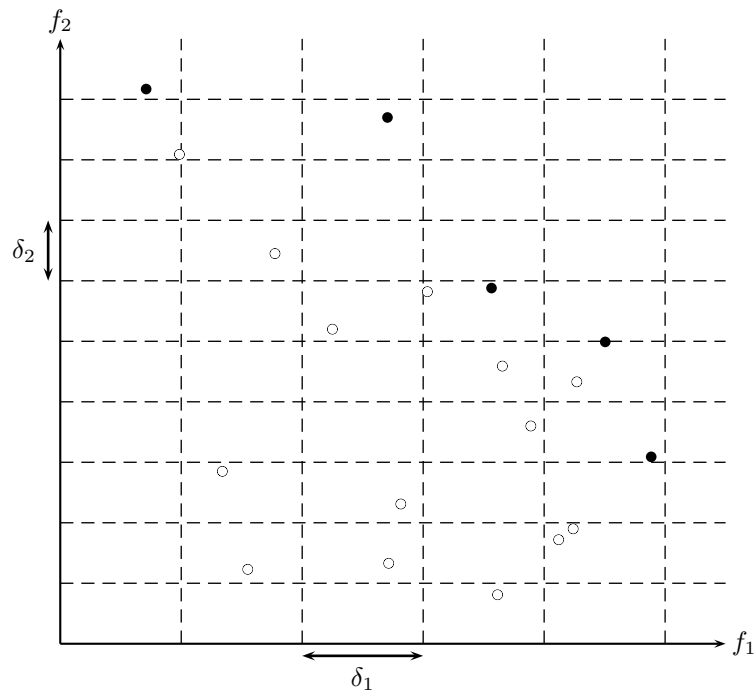


FIGURE 6 – Méthode de découpage de l'espace des objectifs en fonction de la précision δ_i requise pour chaque composante f_i .

2.3 Tests de l’algorithme

L’algorithme a été utilisé pour caler les 27 paramètres du modèle CEQUEAU sur plusieurs bassins versants de la Garonne et du Rhône. Les fonctions objectif utilisées sont les critères C_{MM} sur le débit et sur le logarithme des débits (le critère C_{MM} étant une version bornée entre -1 et 1 du critère de Nash ; voir *Mathevet et al.*, 2006).

Partant d’une librairie comportant une centaine de jeux issus de calages antérieurs, l’algorithme converge de façon satisfaisante en quelques milliers de “runs” du modèle (une valeur de 5000 peut être retenue comme critère d’arrêt). La Figure 7 illustre l’évolution du front de Pareto approximé pour le bassin de la Saône au Chatelet (11700 km²) sur la période 1969–1982. L’algorithme n’étant pas purement déterministe, cette séquence n’est qu’une des “trajectoires” possibles avec ce jeu de données.

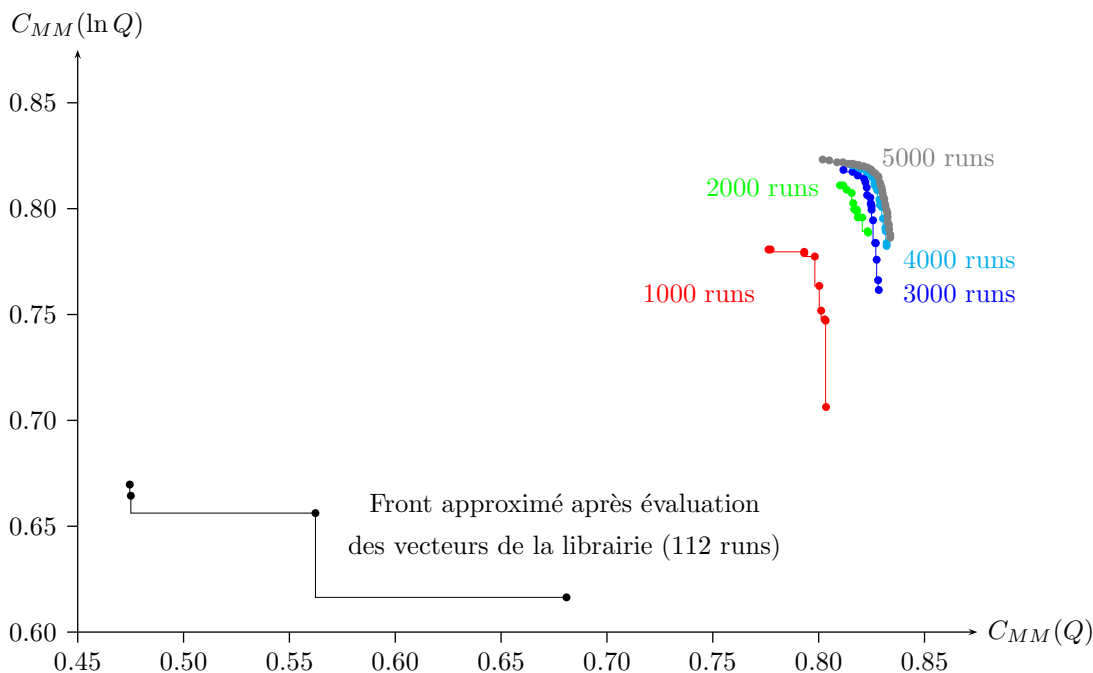


FIGURE 7 – Evolution du front de Pareto approximé dans l’espace des objectifs : exemple avec le modèle CEQUEAU calé sur la Saône au Chatelet, période 1969–1982.

Graphiquement, on visualise l’intérêt d’utiliser plusieurs fonctions objectif : lorsque l’optimisation progresse lentement voire stagne sur l’un des axes, elle peut continuer à progresser sur l’autre. Par exemple, entre 3000 et 4000 runs on a peu progressé sur l’axe $C_{MM}(Q)$ mais le front s’est déplacé vers des valeurs supérieures en $C_{MM}(\ln Q)$.

Sur la Figure 8 on montre, pour quelques paramètres, la distribution au sein du front de Pareto approximé. Notons d’ailleurs que le rapport entre la variance de la distribution finale d’un paramètre, et la variance de sa distribution a priori, fournit un élément de diagnostic pour évaluer la sensibilité de ce paramètre : si ce rapport est très petit devant 1, cela signifie que le paramètre est bien identifié ; si il reste proche de 1, alors le paramètre est certainement peu sensible ou compensé par d’autres.

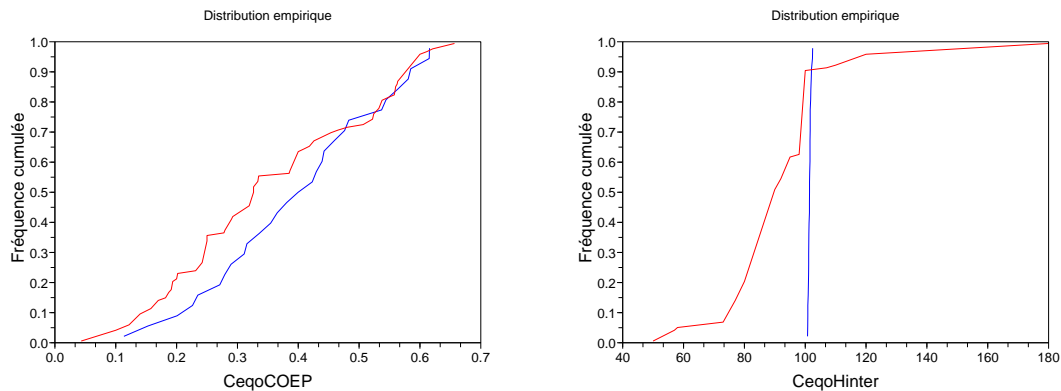


FIGURE 8 – Distributions a priori (en rouge) et après calage (dans le front de Pareto) pour deux paramètres de CEQUEAU. À gauche : paramètre de gradient pluviométrique orographique $CeqoCOEP$; À droite : paramètre “hauteur d’infiltration” $CeqoHInfil$.

Références

- [1] Efstratiadis, A., et D. Koutsoyiannis (2005), The multiobjective evolutionary annealing-simplex method and its application in calibrating hydrological models, in *EGU General Assembly 2005, Geophysical Research Abstracts, Vol. 7*, Vienna, 04593, European Geophysical Union.
- [2] Fisher, R. A. (1922), On the mathematical foundation of theoretical statistics, *Philos. Trans. Royal Soc. London, A 222*, 309–368.
- [3] Labedade, R. (2008), Naturalisation des débits influencés par l’irrigation et modélisation hydrologique sur des sous-bassins versants de la Garonne à Lamagistère, Rapport de stage, ENSEEIHT / EDF.
- [4] Mathevet, T., C. Michel, V. Andréassian, et C. Perrin (2006), A bounded version of the Nash-Sutcliffe criterion for better model assessment on large sets of basins, in *Large sample basin experiments for hydrological model parameterization : Results of the MOdel Parameter EXperiment — MOPEX*, pp. 117–125, Edited by V. Andréassian, A. Hall, N. Chahinian, and J. Schaake, IAHS Publ. n° 307.
- [5] Reed, P., et D. Devireddy (2004), Groundwater monitoring design : a case study combining epsilon-dominance archiving and automatic parameterization for the NSGA-II, in *Coello-Coello C, editor. Applications of multi-objective evolutionary algorithms, Advances in natural computation series, vol. 1*, pp. 79–100, World Scientific, New York.