

# Critères & indicateurs d'auto-évaluation des modèles

**dans le cadre de l'autosurveillance réglementaire des systèmes  
d'assainissement pour permettre  
au maître d'ouvrage de juger son modèle**

L'objectif de ce document, réalisé avec l'aide des membres du groupe de travail autosurveillance des réseaux du Graie et de leurs expériences, est de proposer une série d'indicateurs de performances afin d'évaluer le pouvoir prédictif d'une simulation à partir d'un modèle mathématique, dans le domaine de l'hydrologie et donc également dans le cadre de l'autosurveillance des systèmes.

Cette évaluation est faite principalement pour comparer les valeurs observées (*par exemple* mesurées) avec les valeurs calculées par un modèle. L'appréciation par le maître d'ouvrage de ces indicateurs dépend de l'objectif du modèle, fixé lors de sa conception (à quelle interrogation précise il doit répondre) ; l'utilisation à minima de deux indicateurs est fortement conseillée.

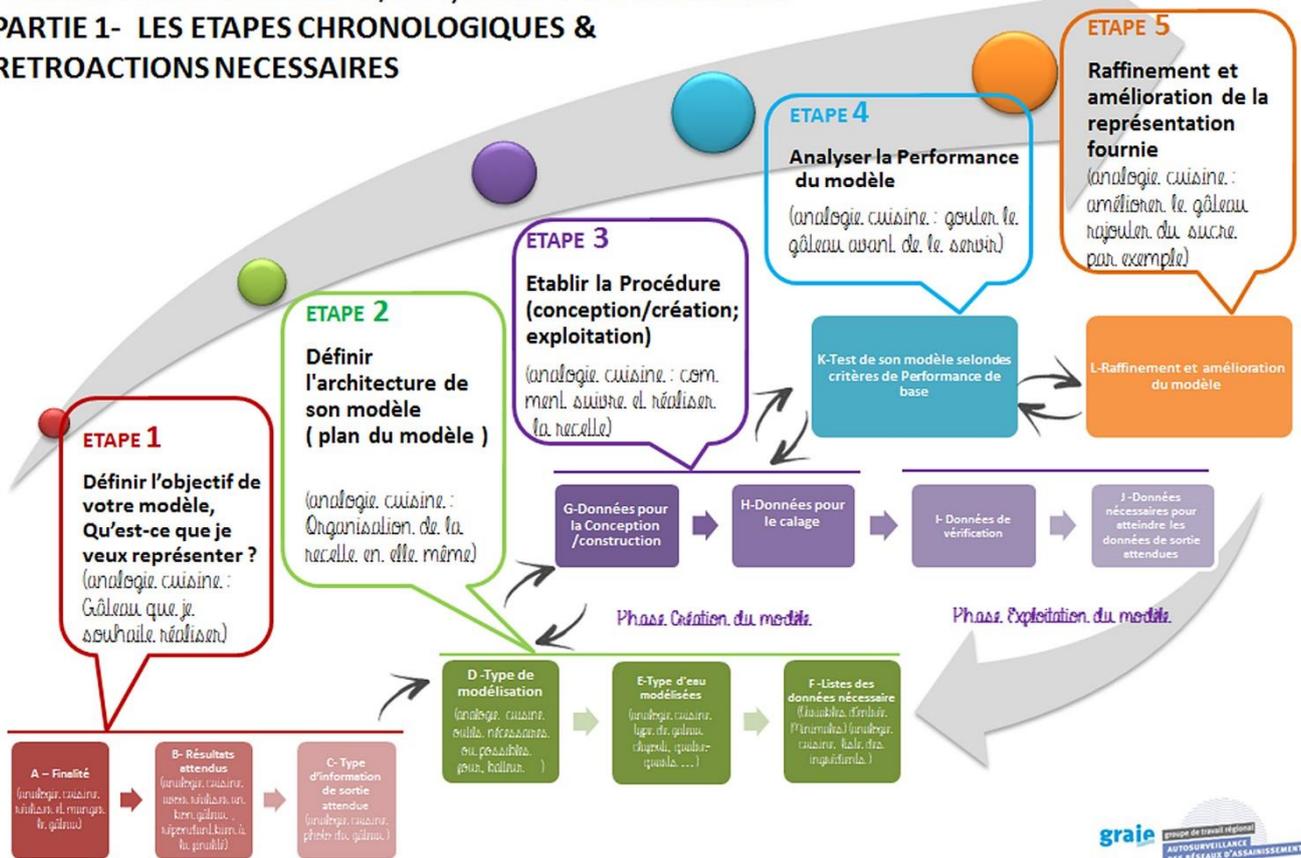
Les indicateurs proposés permettent ainsi d'évaluer : LA FIDÉLITÉ, LA JUSTESSE et L'EXACTITUDE d'un modèle.



Figure 1: Illustration sur des mesurages répétés (Jean-Luc Bertrand Krajewski, INSA Lyon DEEP - 2016)

Ce document vient préciser l'une des étapes décrites dans la méthodologie d'aide à la conception et à l'évaluation d'un modèle pour l'autosurveillance des réseaux et/ou systèmes d'assainissement (ref). Il se focalise sur la phase « analyse de la performance » d'un modèle (cf. Étape 4 et 5). Nous supposons donc que les concepts de conception, d'objectifs du modèle, de données, de calage sont familiers pour le lecteur et que, dans un contexte d'application, ces questionnements et principes ont déjà été considérés dans le projet amont lors de la conception du modèle.

**Aide à la conception et à l'évaluation d'un modèle\* pour l'Autosurveillance réseaux et/ou système d'assainissement**  
**PARTIE 1- LES ETAPES CHRONOLOGIQUES & RETROACTIONS NECESSAIRES**



\*Un modèle en hydrologie est "une représentation mathématique ou physique du cycle de l'eau construite dans le but de mieux comprendre son fonctionnement et de le faire évoluer." Encyclopédie de l'hydrologie urbaine et de l'assainissement - Bernard Chocat, Eurydice, 1136 pages, 1997, N°ISBN: 2-7430-0126-7

Figure 2: Schéma général d'une démarche de modélisation, avec la phase d'exploitation (les étapes 4 et 5), temps de l'autoévaluation de leurs modèles par les maitres d'ouvrages (ref)

Trois types d'indicateurs seront envisagés :

- des indicateurs généraux (Biais, RMSE, MSE),
- des indicateurs normalisés (c'est-à-dire calculés et jugés par rapport à une valeur de référence) (Nash-Sutcliffe, RSR, RVE),
- et enfin des indicateurs visuels (Fonction des résiduels, Graphique des X observés en fonction des X simulés).

Ces trois types d'indicateurs sont décrits puis illustrés grâce à des applications sur des cas réels.

### 1. Indicateurs généraux (Bennett *et al.*, 2013)

Lorsque l'on modélise un comportement réel, la question que l'on pose, une fois le modèle établi, est celle de sa fiabilité, de sa pertinence. Pour évaluer les performances d'un modèle, il est possible de faire appel à des indicateurs de performance<sup>1</sup>.

Pour une modélisation hydraulique et hydrologique, la qualité d'un modèle est normalement jugée à partir de la ressemblance des débits observés (mesurés) et débits simulés par le modèle, obtenus à partir d'une entrée connue (par exemple la pluie mesurée). Comme mentionné précédemment, l'évaluation de cette ressemblance doit être effectuée à partir de données que le modèle n'a jamais vues ou utilisées, c'est-à-dire des données mesurées (pluie et/ou débit) sur une période en dehors de celle utilisée dans la phase de calage. Le schéma ci-dessous illustre la comparaison entre les simulations obtenues à partir d'un modèle et les données mesurées (Figure 3).

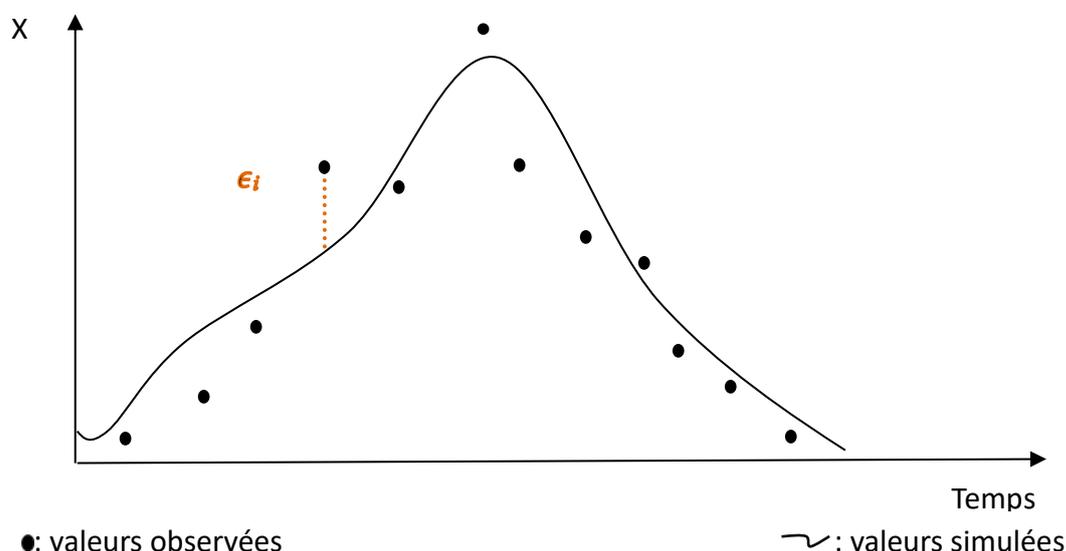


Figure 3: Comparaison au cours du temps entre valeurs observées et simulées par un modèle mathématique

Les valeurs X peuvent être les volumes déversés par événement (en m<sup>3</sup>), le débit en un point du réseau (en L/s), ou encore le nombre de déversements annuel (sans dimension), etc.

Une première différence évidente est la nature continue des simulations du modèle (ligne) par rapport à la nature discrète de données mesurées (points).

Cette différence est due au fait qu'un modèle permet d'établir une valeur à tout instant  $t$  (qui est intrinsèquement continu), alors que la mesure d'une variable physique en continu est en principe infaisable même si elle peut être réalisée à très haute fréquence.

En effet, les débits et la pluie sont mesurés avec un pas de temps qui dépend principalement du principe de mesure et des besoins. Cela peut varier d'une minute à des heures (*par exemple* un hydrographe), même des jours ou des mois (bilans hydrologiques mensuels/annuels), en fonction des besoins du projet et de la vitesse de la dynamique de l'hydrosystème étudié.

<sup>1</sup> REMARQUES : les fonctions objectifs et les indicateurs de performance ne sont pas la même chose.

L'indicateur indique la qualité du modèle, la fonction objectif est ce qu'on minimise ou maximise (selon les algorithmes utilisés) pour caler le modèle. Il est possible, dans certains cas, que les deux coïncident, mais ce n'est pas la même chose.

Cependant, bien que le pas de temps de la modélisation soit continu, pour comparer cette modélisation et de la mesure, un pas de temps commun doit être choisit pour pouvoir comparer les valeurs observées et simulées. **L'écart ( $\epsilon_i$ ) entre chaque observation  $i$  et mesure  $i$  est une valeur positive ou négative, représentant une sur- ou sous-estimation par le modèle, calculé pour chaque instant où il existe une valeur mesurée.**

#### a) Le biais – Ecart entre observations et mesures – Critère de fidélité

Une première condition souhaitée dans la validation d'un modèle est que la moyenne de tous les écarts  $\epsilon_i$  se rapproche le plus possible de zéro, c'est à dire un modèle sans biais.

Le biais peut être calculé comme suit :

$$biais = \frac{1}{n} \sum_{i=1}^n (X_{obs\ i} - X_{sim\ i}) = \frac{1}{n} \sum_{i=1}^n \epsilon_i$$

Avec :

- $biais$  (dimension de  $X$ ),
- $n$ , le nombre d'observations (sans dimension),
- $X_{obs\ i}$ , la  $i$ -ème valeur observée de  $X$  (dimension de  $X$ ),
- $X_{sim\ i}$ , la  $i$ -ème valeur simulée de  $X$  (dimension de  $X$ ),
- $\epsilon_i$ , l'écart entre chaque observation  $i$  et mesure  $i$  (dimensions de  $X$ ),



**Cet indicateur nous décrit la fidélité du modèle : c'est-à-dire si notre modèle surestime ou sous-estime systématiquement les valeurs  $X$ .**

#### b) Critère RMSE (Root Mean Square Error) : la caractérisation de la taille des écarts entre observations et mesures – Critère exactitude

Le biais nous indique des écarts, mais il ne nous donne pas d'information sur l'amplitude de ces écarts, vu que les valeurs positives et négatives de  $\epsilon_i$  se compensent dans la moyenne.

Le critère **RMSE (Root Mean Square Error)** permet de faire ce calcul.

L'amplitude des écarts  $\epsilon_i$  peut se caractériser par la moyenne des carrés des écarts  $\epsilon_i$ , afin de les rendre positifs.

Le calcul est le suivant :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\epsilon_i)^2}$$

Avec :

- $RMSE$ , la racine de l'erreur moyenne quadratique (dimension de  $X$ ),
- $n$ , le nombre d'observations (sans dimension),
- $\epsilon_i$ , l'écart entre chaque observation  $i$  et mesure  $i$  (dimensions de  $X$ ),

La racine carrée utilisée dans le calcul du RMSE a pour objectif de revenir à une unité de valeur identique à celle de la variable comparée (*par exemple*. L/s pour une analyse sur les débits).

Lorsque l'on utilise l'indicateur sans la racine carrée, nous obtenons un autre indicateur, que l'on appelle **MSE (Mean Square Error)**.

$$MSE = \frac{1}{n} \sum_{i=1}^n (\epsilon_i)^2$$

Cette déclinaison du RMSE, exprimé en unité de la variable  $X$  au carré (*par exemple*  $L^2/s^2$ ), est également très utile pour des explications complémentaires de l'exactitude du modèle.

Les valeurs du RMSE et MSE sont toujours positives donc comprises dans l'intervalle  $[0; +\infty[$ .

**La variance du terme  $\epsilon_i$  au cours de tout l'intervalle temporel de simulation sera définie comme la « justesse » de la modélisation.**

On peut calculer la justesse  $\sigma_\epsilon^2$  grâce à l'équation suivante (Gy, 1998) :

$$\sigma_\epsilon^2 = RMSE^2 - biais^2 \rightarrow RMSE = \sqrt{\sigma_\epsilon^2 + biais^2} \rightarrow MSE = \sigma_\epsilon^2 + biais^2$$

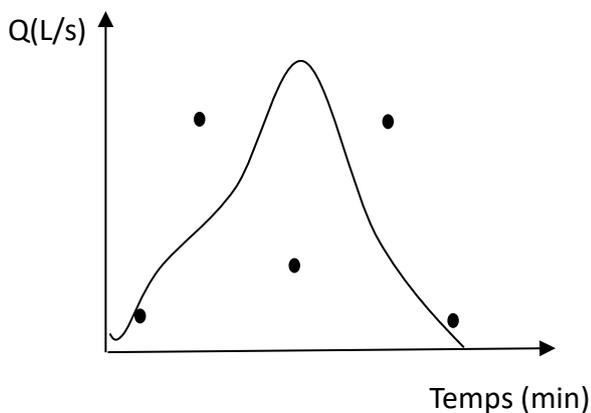
Cette équation nous permet également de visualiser que le RMSE et le MSE, en tant que mesure globale de l'exactitude du modèle, prennent en compte à la fois le biais (fidélité) et sa variance (justesse).

Ainsi, un modèle jugé fidèle au travers du biais (proche de zéro) peut être très inexact (valeurs de RMSE et le MSE élevées) du fait de la variabilité des écarts ou de la justesse (valeurs de  $\sigma_\epsilon^2$  élevées) (Gy, 1998).

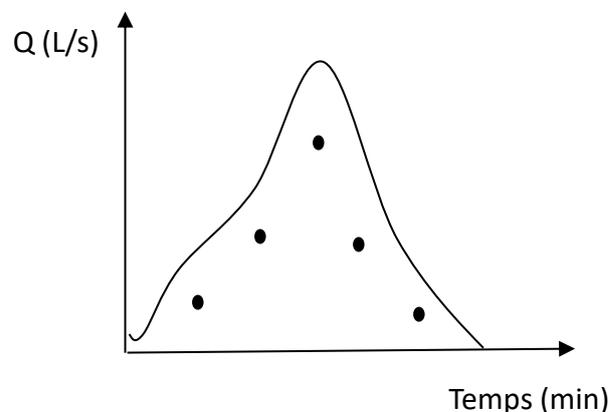


**Plus la valeur des critères RMSE ou MSE est proche de zéro, meilleur est le modèle évalué en terme d'exactitude.** Néanmoins, ces indicateurs ne donnent pas d'information sur la contribution de chaque composant d'erreur (justesse et fidélité) séparément. Donc, le biais (témoin de la fidélité du modèle) est complémentaire au RMSE et au MSE.

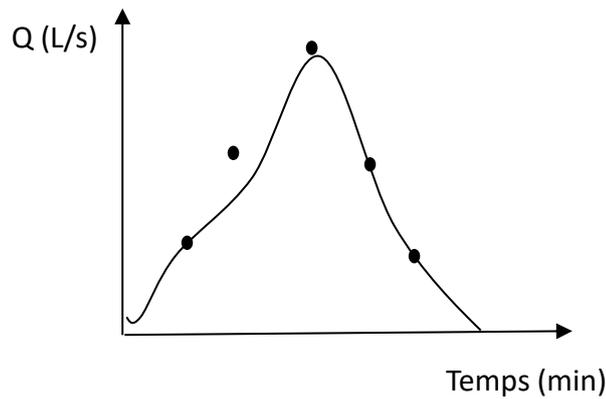
Les exemples suivant montrent différents scénarios pour illustrer les concepts décrits précédemment avec  $X$  égal au débit de sortie  $Q$  dans un bassin versant (en L/s) (Figure 4) :



a) mauvais RMSE (exactitude)/ biais faible (fidélité)



b) mauvais RMSE (exactitude)/ biais élevé (fidélité)



### c) bon RMSE (exactitude)/ biais faible (fidélité)

Figure 4: Scenarios de comparaison entre des simulations avec a) mauvais RMSE (exactitude)/ biais faible (fidélité), b) mauvais RMSE (exactitude)/ biaisé (fidélité), c) bon RMSE (exactitude)/ biais faible (fidélité).

Sur l'exemple a), on peut apprécier un modèle où le biais est potentiellement faible (les erreurs se compensent) mais le RMSE peut être élevé (mauvaise exactitude).

Dans le deuxième cas b), même si la simulation suit de manière appropriée la dynamique du débit, il y a un écart systématique entre les valeurs observées et simulées, ce qui rend l'indicateur de biais plus élevé (positif ou négatif) et donc le RMSE aussi. Dans ce cas la variance peut nous aider à juger de la justesse du modèle.

Le troisième cas c) représente un bon ajustement donc un RMSE ainsi qu'un biais assez bas. Si l'on calcule le MSE, il aura tendance à pénaliser, plus que le RMSE, les simulations qui ne représentent pas les pics de l'hydrographe.

## 2. Indicateurs normalisés (avec un point de référence) (Bennett *et al.*, 2013)

Néanmoins, juger acceptable un modèle et ses paramètres sur la base du RMSE ou MSE ainsi qu'avec le biais reste difficile, vu que la grandeur de chaque indicateur sera forcément dépendant des données utilisées et de chaque cas particulier. La question de la détermination d'un plafond ou point de référence semble intéressante.

Ainsi dans les indicateurs normalisés, une valeur de performance de référence ou une performance relative dans chaque indicateur est établie pour pouvoir normaliser l'évaluation du modèle.

Effectivement, la grande force des critères normalisés est d'être sans dimension, ce qui permet notamment de comparer des modèles entre eux. Dans la suite du document, nous allons présenter des indicateurs normalisés qui permettront d'amener plus d'informations sur la pertinence d'un modèle.

N.B. : Pour les cas du MSE (ou RMSE), cette valeur de performance de référence est définie comme la variance (ou l'écart type) de valeurs mesurées de  $X$ , par exemple  $\sigma_x^2$ .

Cela, vu que  $\sigma_x^2$  (ou l'écart type) est une représentation du MSE (ou RMSE) commise par un modèle ou l'on simule la sortie  $X$  comme la moyenne des observations  $\bar{X}_{obs}$ .

### a) Critère de Nash-Sutcliffe

Le critère de Nash-Sutcliffe est un indicateur de performance construit à partir de la normalisation du MSE, dont les valeurs sont comprises dans l'intervalle  $]-\infty; 1]$ .



**Il permet d'estimer la capacité d'un modèle à reproduire un comportement observé.**

Il se calcule comme suit :

$$NS = 1 - \frac{MSE}{\sigma_x^2} \rightarrow NS = 1 - \frac{\sum_{i=1}^n (X_{obs\ i} - X_{sim\ i})^2}{\sum_{i=1}^n (X_{obs\ i} - \bar{X}_{obs})^2}$$

- Avec :
- NS, le critère de Nash-Sutcliffe (sans dimension),
  - MSE, le *Mean Square Error* (dimensions de  $X^2$ ),
  - n, le nombre d'observations (sans dimension),
  - $X_{obs\ i}$ , la  $i$ -ème valeur observée de la grandeur  $X$  (dimensions de  $X$ ),
  - $X_{sim\ i}$ , la  $i$ -ème valeur simulée de la grandeur  $X$  (dimensions de  $X$ ),
  - $\bar{X}_{obs}$ , la moyenne des valeurs observées (dimensions de  $X$ ),
  - $\sigma_x^2$ , la variance des valeurs observées (dimensions de  $X^2$ ).

Plus la valeur obtenue pour ce critère est proche de 1, meilleure est l'adéquation du modèle aux valeurs observées. **Il est généralement admis que le critère de Nash-Sutcliffe doit être supérieur à 0,7 pour pouvoir affirmer qu'un modèle est satisfaisant**, c'est à dire que le modèle et les valeurs observées sont cohérents. On peut nuancer cette valeur qui dépend aussi du domaine dans lequel on travaille et de la qualité des modèles utilisés.

Il est par exemple possible de viser des valeurs plus élevées (0,7 - 0,8) pour la modélisation pluie-débit que pour la modélisation des flux polluants par exemple car ce dernier est beaucoup moins performant. Un critère de Nash-Sutcliffe égal à 0 indique que le modèle ne fait pas mieux que simuler la sortie du modèle avec la moyenne des valeurs observées. Lorsque que le critère devient négatif, cela témoigne d'un modèle plus mauvais encore.

Une autre interprétation du Nash-Sutcliffe est : combien de pourcentage de la variance de la sortie ( $\sigma_x^2$ ) l'on explique avec le modèle mathématique. Un Nash-Sutcliffe de 0.8 voudrait dire que notre modèle explique au moins 80 % de la variance de  $X$ . Si votre Nash est inférieur à 0.5 cela indique que le modèle reproduit inexactement le comportement observé, il est donc nécessaire de rechercher les causes. Un Nash en dessous environ de 0.6 démontre une mauvaise adéquation du modèle aux valeurs observées.

## b) Critère RSR (*RMSE standard deviation ratio*)

Le RSR est un critère similaire au Nash-Sutcliffe, néanmoins moins utilisé, basé sur la normalisation du RMSE, au lieu du MSE.

Il peut s'exprimer comme suit (Moriassi *et al.*, 2007):

$$RSR = \frac{RMSE}{\sigma_x} \rightarrow RSR = \sqrt{\frac{\sum_{i=1}^n (X_{obs\ i} - X_{sim\ i})^2}{\sum_{i=1}^n (X_{obs\ i} - \bar{X}_{obs})^2}}$$

- Avec :
- RSR, RMSE standard deviation ratio (sans dimension),
  - RMSE, le *Root Mean Square Error* (dimensions de  $X$ ),
  - n, le nombre d'observations (sans dimension),
  - $X_{obs\ i}$ , la  $i$ -ème valeur observée de la grandeur  $X$  (dimensions de  $X$ ),
  - $X_{sim\ i}$ , la  $i$ -ème valeur simulée de la grandeur  $X$  (dimensions de  $X$ ),
  - $\bar{X}_{obs}$ , la moyenne des valeurs observées (dimensions de  $X$ ),
  - $\sigma_x$ , la variance des valeurs observées (dimensions de  $X$ ).



**Plus la valeur obtenue pour ce critère est proche de 0 , meilleure est l'adéquation du modèle aux valeurs observées.**

Une valeur qui indique une simulation acceptable devrait être inférieure à 0.2. La différence principale entre le RSR et le Nash-Sutcliffe est, par exemple, que le deuxième pénalise encore plus les différences de pics de l'hydrographe. Ce critère est interprétable comme le pourcentage de l'écart type  $\sigma_x$  non-expliqué par le modèle. Un RSR de 0.2 voudrait dire qu'on n'explique pas au moins 20 % de la variance des valeurs observées ( $\sigma_x$ ) avec le modèle testé.

### c) Critère d'Erreur de Volume Relatif (RVE)

Toujours dans le cadre d'indicateurs de performance relative, on peut normaliser l'indicateur de biais présenté dans la section précédente par une simulation avec  $\bar{X}_{obs}$  égal à zéro.

La manière la plus courante d'établir cet indicateur normalisé, est le RVE .

L'erreur de volume relatif est la somme des erreurs liées à la somme des valeurs observées, exprimées en valeur relative ou en pourcentage.

Pour cela on divise le biais par le volume total de simulation comme suit :

$$RVE = \frac{\text{biais}}{\sum_{i=1}^n X_{obs\ i}} = \frac{\sum_{i=1}^n \epsilon_i}{\sum_{i=1}^n X_{obs\ i}}$$

Avec :

- RVE (sans dimensions),
- $n$ , le nombre d'observations (sans dimension),
- $X_{obs\ i}$ , la  $i$ -ème valeur observée de  $X$  (dimension de  $X$ ),
- $\epsilon_i$ , l'écart entre chaque observation  $i$  et mesure  $i$  (dimensions de  $X$ ),



**L'indicateur de RVE peut être interprété comme l'erreur sur le volume modélisé par rapport au volume total observé (en pourcentage, si on le souhaite).**

Remarque : le volume dans le cas de cet indicateur est utilisé au sens « quantité globale de quelque chose » et non pas comme le volume mathématique - mesuré en  $m^3$ .

**Plus le RVE est faible, meilleure est l'adéquation globale entre volume modélisé et observé.**

Ce critère mesure une adéquation globale, mais les erreurs peuvent être compensées.

De façon similaire au biais, une valeur faible de RVE ne signifie pas une bonne performance, mais celui-ci peut nous donner une idée de la nature des erreurs de modélisation (justesse ou fidélité).

Si nous reprenons les exemples de la Figure 4, le critère de Nash-Sutcliffe (ou le RSR) pénalisera le cas a et b. À la différence du RVE qui peut avoir une valeur assez faible (donc bonne) lorsque les erreurs individuelles se compensent en moyenne.

## 3. Indicateurs de performances complémentaires

**Les critères visuels** restent très utilisés pour vérifier les conclusions envisagées à partir des indicateurs de performances numériques et pouvoir diagnostiquer plus en détail les points faibles d'un modèle mathématique.

Ils sont des indicateurs intéressants pour l'aide à la décision sur la performance du modèle.

### a) Fonction des résiduels

En effet, une première manière de pouvoir diagnostiquer notre modèle peut consister à illustrer la chronologie des valeurs  $\epsilon_i$  (il s'agit d'une fonction des résiduels), voir ci-dessous (Figure 5) :

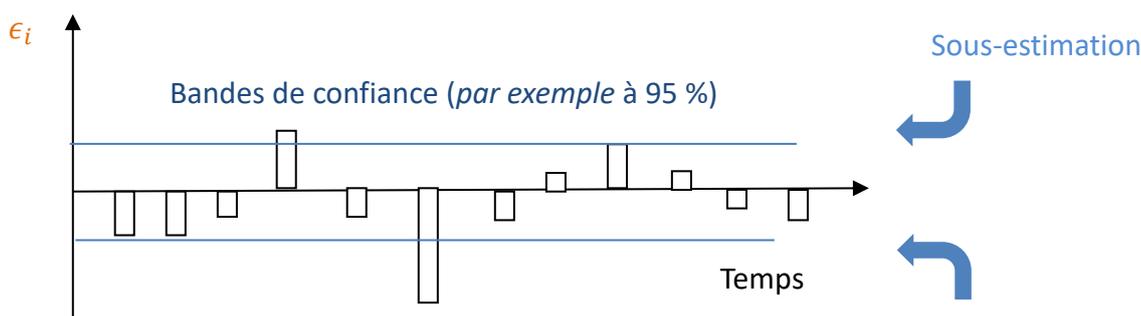


Figure 5: Fonction des résiduels



Cette représentation permet d'analyser l'évolution temporelle de l'erreur de modélisation de manière graphique et identifier les fenêtres temporelles où le modèle a représenté les valeurs d'une manière moins performante (*par exemple en début ou fin de l'hydrographe*).

#### b) Graphique des $X$ observés en fonction des $X$ simulés

Une autre alternative est de représenter la même information avec un graphique des  $X$  observés en fonction des  $X$  simulés comme suit (*Figure 6 et Figure 7*).

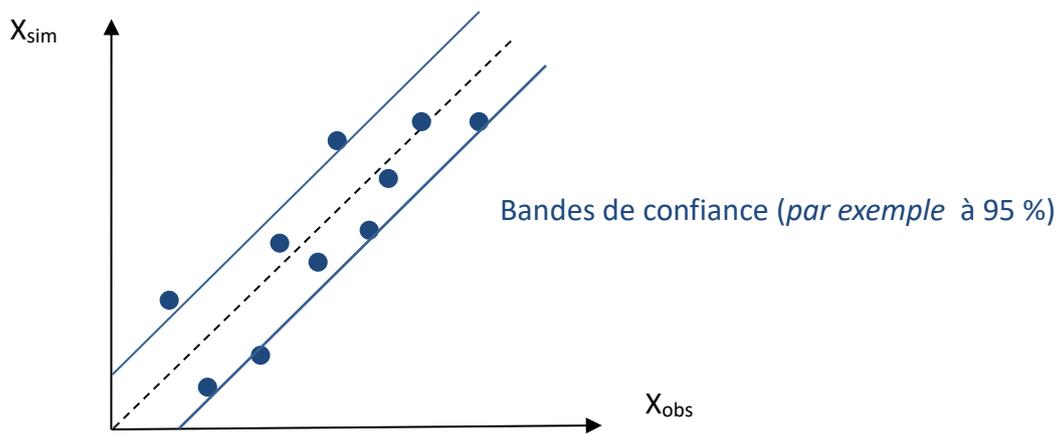


Figure 6: Graphique de valeurs observées versus simulées.

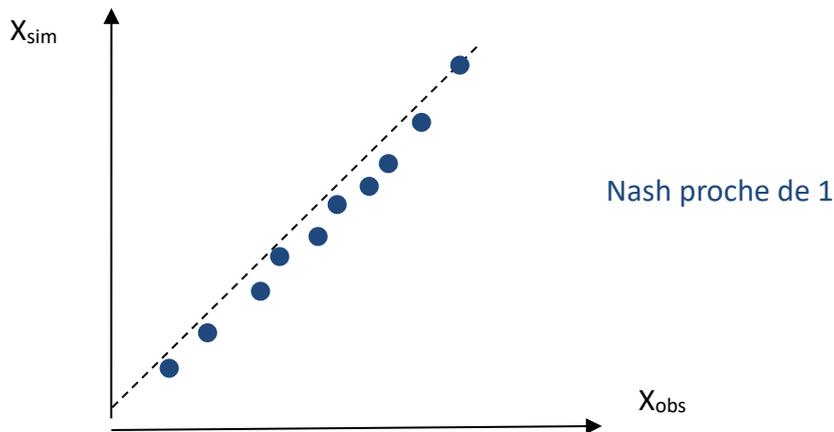


Figure 7: Graphique de valeurs observées versus simulées.  
Pour un modèle avec un critère de Nash très proche à 1

Cette représentation permet d'analyser s'il y a une tendance à mieux représenter les observations en fonction de la taille de  $X$  (par exemple le modèle représente moins bien les débits forts).

En effet, des critères de performances quantitatifs plus rigoureux peuvent être aussi développés avec les représentations proposées dans les Figure 5 et Figure 6. L'estimation de tendance dans les erreurs au cours du temps où en fonction de la taille de  $X$  peut être vérifiée de manière quantitative.

Des bandes de confiance peuvent être estimées à partir des hypothèses statistiques, en ayant pour objectif d'identifier les simulations qui ne sont pas égales à la valeur observée, avec une probabilité donnée (points en dehors des bandes de confiance à 95 % par exemple) (Figure 5 et Figure 6).

Couramment, ces bandes de confiance sont calculées comme  $\pm 2 \cdot \text{RMSE}$ , en faisant l'hypothèse des résiduelles symétriques, pas corrélés et distribués avec une loi normale de probabilité de l'écart type constante et égale à RMSE.

Néanmoins, leur calcul peut changer en fonction des hypothèses statistiques faites sur la représentation de l'erreur choisie entre la Figure 5 ou Figure 6 (voir Métadier, 2011).

## B- APPLICATION CONCRÈTE

### 4. Évaluation de deux modèles : estimation de volumes de déversement et estimation de débit à la sortie d'un bassin versant

Cette partie montre l'application sur des données réelles et souligne l'importance d'utiliser plusieurs critères cités précédemment afin de tirer les bonnes conclusions sur l'exactitude de notre modèle et son adéquation avec les observations.

#### a) Exemples d'auto-évaluation de modèles d'estimation de volumes de déversement

**Dans le premier cas**, on s'intéresse à la capacité du modèle hydrologique utilisé par ces deux villes à bien représenter les volumes instantanés déversés au cours d'un évènement pluvieux par les déversoirs d'orage. Le Tableau 1 et 2 recensent les résultats pour la première et deuxième ville, respectivement.

Tableau 1: Évaluation d'un modèle de volume déversé pour plusieurs déversoirs de la ville 1 à l'aide de plusieurs indicateurs.  
Calcul pour un évènement pluvieux.

|  | DO101 | DO22   | DO91     | DO107   | DO104   |
|--|-------|--------|----------|---------|---------|
| Volume total mesuré (m <sup>3</sup> )    | 1548  | 24611  | 214104   | 60634   | 14455   |
| Volume total modèle (m <sup>3</sup> )    | 42    | 24827  | 175086   | 47730   | 10485   |
| Biais (m <sup>3</sup> )                  | 1506  | -216   | 39017    | 12905   | 3970    |
| RMSE (m <sup>3</sup> )                   | 166   | 628    | 4433     | 2686    | 1345    |
| MSE ((m <sup>3</sup> ) <sup>2</sup> )    | 27392 | 394786 | 19651176 | 7214558 | 1809073 |
| RSR                                      | 1,01  | 0,68   | 0,50     | 0,09    | 0,69    |
| Nash-Sutcliffe                           | -0,01 | 0,54   | 0,75     | 0,99    | 0,52    |
| RVE                                      | 97%   | - 1%   | 18%      | 21%     | 27%     |
| Biais relatif en % (  Biais /vol.mesuré) | 97%   | 1%     | 18%      | 21%     | 27%     |

Tableau 2: Évaluation d'un modèle de volume déversé pour plusieurs déversoirs de la ville 2 à l'aide de plusieurs indicateurs.  
Calcul pour un évènement pluvieux.

|  | DO1    | DO2    | DO34     | DO9   | DO15    |
|--|--------|--------|----------|-------|---------|
| Volume mesuré (m <sup>3</sup> )          | 257185 | 96365  | 357605   | 12114 | 59519   |
| Volume modèle (m <sup>3</sup> )          | 287975 | 93406  | 505009   | 14664 | 88894   |
| Biais (m <sup>3</sup> )                  | -30790 | 2958   | -147404  | -2551 | -29375  |
| RMSE (m <sup>3</sup> )                   | 755    | 530    | 3759     | 177   | 3056    |
| MSE ((m <sup>3</sup> ) <sup>2</sup> )    | 569353 | 281126 | 14129495 | 31239 | 9341701 |
| RSR                                      | 0,22   | 0,24   | 0,50     | 0,09  | 0,69    |
| Nash-Sutcliffe                           | 0,95   | 0,94   | 0,75     | 0,99  | 0,52    |
| RVE                                      | -12%   | 3%     | -41%     | -21%  | -49%    |
| Biais relatif en % (  Biais /vol.mesuré) | 12%    | 3%     | 41%      | 21%   | 49%     |

#### Légendes

|                      |  |
|----------------------|--|
| Critère satisfaisant |  |
| Critère Moyen        |  |
| Critère Insuffisant  |  |

#### COMMENTAIRES

**Si l'on regarde les critères Biais, MSE et RMSE**, on visualise la complexité de conclure juste à partir de ces indicateurs, notamment pour les DO91 et DO107 de la ville 1 qui ont des valeurs beaucoup plus élevées que pour d'autres déversoirs (entre 6 et 8 fois).

**En regardant les critères relatifs**, spécifiquement le RVE (en admettant 30 % d'erreur en biais par exemple), on peut valider de manière préliminaire le modèle pour tous les déversoirs sauf le D0101 de la ville 1 et le DO34 et le DO15 de la ville 2 .

Cela nous indique également que si le but recherché est simplement le volume déversé annuel (évalué à travers du RVE), les modèles ne sont a priori pas mauvais, sauf pour les 3 déversoirs cités ci-avant.

Le signe de l'estimation de RVE nous indiquera si le modèle a plus de tendance à sur (négatif) ou sous-estimer (positif) les volumes déversés.

Enfin, si on ajoute encore le critère relatif de Nash-Sutcliffe ou RSR, on invalide le modèle pour les déversoirs DO22 et DO104 de la ville 1. En effet, un score de l'ordre de 0,5 pour le critère de Nash-Sutcliffe est trop faible, les modèles ne peuvent donc pas être considérés satisfaisants.

On se retrouve ici dans un cas où le biais ou le RVE sont acceptables, mais la représentation de la dynamique de l'événement n'est pas correcte. Pour illustrer ce cas, tout se passe comme si le pic et la durée du déversement étaient parfaitement représentés, mais avec un décalage dans le temps de 3-4h.

On trouvera qu'un critère Nash-Sutcliffe assez élevé (e.g. > 0,9 pour le DO1, DO2 et DO9 de la ville 2) rend des valeurs de RVE assez performants (erreurs dans le volume total de < 20 %). Le contraire n'est pas forcément le cas comme cela a été démontré : un RVE de 1 % peut avoir un Nash-Sutcliffe égal à zéro (ville 1, DO101). C'est pour cela qu'il est recommandé d'accompagner : un indicateur de biais (fidélité) par un indicateur d'exactitude (e.g. Nash-Sutcliffe). Dans les deux cas, il est aussi conseillé d'employer des indicateurs relatifs (e.g. RVE pour la fidélité et Nash-Sutcliffe pour l'exactitude).

Le Nash-Sutcliffe utilisé seul a l'inconvénient de ne pas pouvoir identifier l'apport du biais (fidélité) dans l'erreur totale de modélisation et permet d'établir si le biais est dû à une sur ou sous-estimation par le modèle (signe dans Biais ou RVE).

Le Nash-Sutcliffe et le RSR donnent des informations sur l'erreur totale de modélisation, sur l'exactitude. Néanmoins, ils peuvent donner des résultats différents.

Par exemple, dans le D034 en ville 2, on pourrait accepter le modèle à partir du Nash-Sutcliffe (0,75) mais on pourrait le rejeter à partir d'une analyse du RSR (0,5). Les différences dans les deux résultats sont principalement explicables du fait que le RSR punit de façon moins importante les erreurs de pics que le Nash-Sutcliffe (dû au terme au carré dans le calcul du Nash-Sutcliffe).

Le Nash-Sutcliffe reste le critère plus utilisé et cité dans la littérature, afin d'évaluer l'exactitude de la modélisation.

## **b) Exemple d'auto-évaluation d'un modèle d'estimation de débit à la sortie d'un bassin versant**

**Un deuxième exemple d'utilisation de ces critères est proposé dans le cas d'une modélisation mathématique du débit en fonction de la pluie avec les données de l'OTHU<sup>2</sup>.** Les données ont été enregistrées lors d'un événement pluvieux dans le bassin versant de Chassieu, France en 2007. Le modèle est un réservoir linéaire simple qui permet d'obtenir un hydrographe simulé à partir du hétérogramme de pluie, les entrées ainsi que les sorties sont mesurées avec un pas de temps de deux minutes. La simulation de vérification a été faite avec un jeu de paramètres déjà déterminé à partir d'un calage avec des autres données disponibles (Sandoval, 2017).

La simulation obtenue est montrée dans la *Figure 8*, ainsi que les résultats des indicateurs de performance proposés, pour l'événement pluvieux analysé.

---

<sup>2</sup> [L'OTHU](http://othu.org) (Observatoire de Terrain en Hydrologie Urbaine) est un laboratoire de recherche hors murs, constitué par un ensemble d'appareils de mesure installés sur le système d'assainissement de la Communauté Urbaine de Lyon et sur les milieux récepteurs recevant les effluents issus de ce système d'assainissement. Il fédère 12 équipes de recherche lyonnaises venant de 9 organismes différents (INSA, Lyon I, Lyon II, Lyon III, ENTPE, Ecole Centrale de Lyon, VetAgro Sup, BRGM, IRSTEA). Cet observatoire est destiné à acquérir des données pérennes sur les rejets urbains, notamment de temps de pluie, et leurs impacts sur les milieux récepteurs, pour proposer de nouvelles solutions de conception et de gestion de l'assainissement. Plus d'informations [othu.org](http://othu.org)

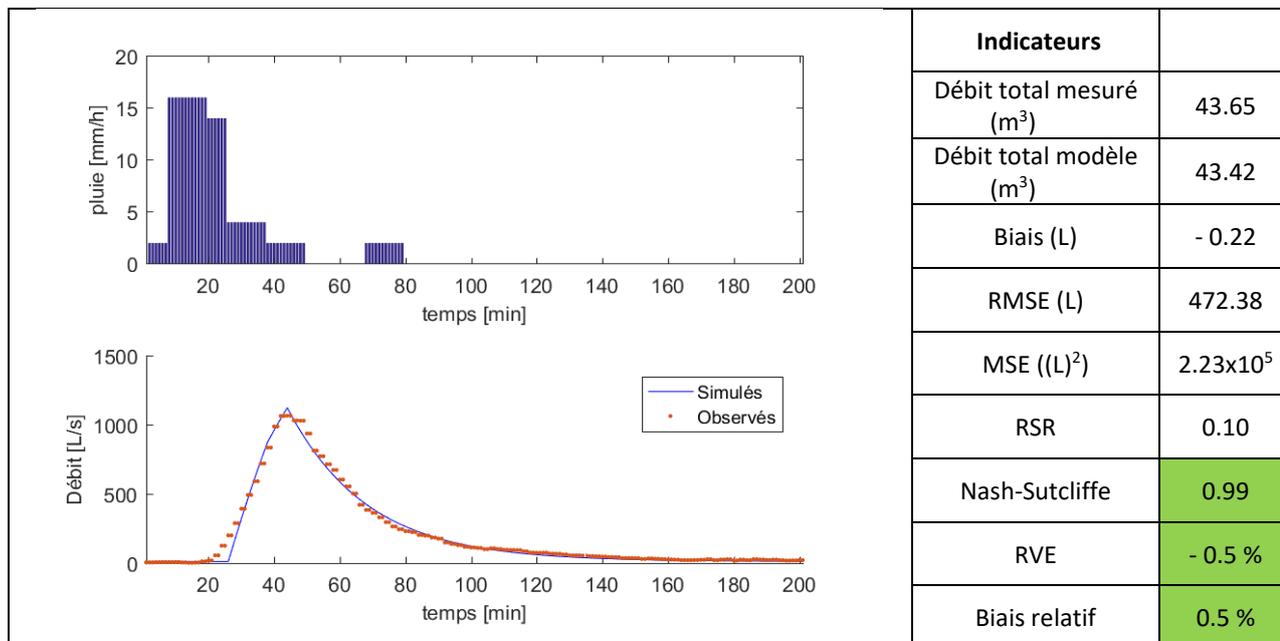


Figure 8: Graphique de valeurs observées (points oranges) versus simulées (ligne bleue) au cours du temps pour un évènement pluvieux à Chassieu.

Dans ce cas, on peut constater une très bonne performance du modèle pour cet évènement, en termes de reproduction du débit à la sortie du bassin versant de Chassieu à partir de la pluie (*Par exemple* Nash-Sutcliffe  $\approx 1$ ).

Effectivement, les indicateurs de fidélité (biais) montrent de valeurs de surestimations très faibles (*par exemple* RVE = - 0.5 %), ce qui nous permet de conclure qu'il n'existe pas une surestimation systématique significative de la part de notre modèle. Des indicateurs visuels confirment l'acceptation de ce modèle, au moins pour cet évènement de vérification analysé.

## 5. Quelques éléments pour conclure

- **Avantages de ces critères d'autoévaluation**

Ces critères :

- sont simples à comprendre et à appliquer pour l'utilisateur d'un modèle.
- peuvent permettre d'alerter l'utilisateur sur la cohérence des valeurs mesurées/simulées et vice-versa.
- permettent de critiquer le modèle ou de le conforter (d'avoir plus de « confiance » dans les simulations).

- **Difficultés**

Ces critères s'adressent aux utilisateurs du modèle, et non pas à un utilisateur lambda, car ils nécessitent une très bonne connaissance de l'outil afin d'analyser finement chaque hydrogramme et chaque point de mesure testés.

En effet, un modèle est un outil créé pour répondre à un objectif. L'objectif du modèle est fixé lors de sa conception (à quelle interrogation précise il doit répondre) ; c'est pourquoi il ne peut pas répondre aux questions pour lesquelles il n'a pas été conçu.

## 6. Références

Bennett, N.D., Croke, B.F.W., Guariso, G., Guillaume, J.H.A., Hamilton, S.H., Jakeman, A.J., Marsili-Libelli, S., Newham, L.T.H., Norton, J.P., Perrin, C., Pierce, S.A., Robson, B., Seppelt, R., Voinov, A.A., Fath, B.D., Andreassian, V., (2013). Characterising performance of environmental models. *Environmental Modelling & Software*, 40, 1-20.

Gy, P.M., (1998), *Sampling for Analytical Purposes*, John Wiley and Sons Ltd., Chichester.

Métadier, M., (2011). *Traitement et analyse de séries chronologiques continues de turbidité pour la formulation et le test de modèles des rejets urbains par temps de pluie*. Thèse, Université de INSA, Lyon, France, 339 p. + appendices.

Sandoval, S., (2017). *Revisiting stormwater quality conceptual models in a large urban catchment: online measurements, uncertainties in data and models*. Thèse, Université de INSA, Lyon, France, 168 p.

## 7. Pour en savoir plus

**Rédacteurs : Santiago SANDOVAL- INSA Lyon DEEP, Adrien MOMPLOT – AEGIR, Laëtitia BACOT – GRAIE – [laetitia.bacot@graie.org](mailto:laetitia.bacot@graie.org)**

en appui sur le sous-groupe modélisation du groupe de travail autosurveillance des réseaux d'assainissement : Pascal BRET, La Métropole de Lyon – Direction de l'eau | Alejandro CLARO BARRETO, INSA Lyon DEEP - Valence Romans Agglo | Virginie DANIEL, Valence Romans Agglo | Nicolas DRUT, Veolia Eau Grand Est | Sandrine FRAYSSE Valence Romans Agglo | Marine GIRES, SUEZ Eau France | Boumediene Ghoulam, Altereo - G2C Ingénierie | Gael LORINI, Agglomération Villefranche Beaujolais (AVB) | Benoît MARDUEL, Altereo - G2C Ingénierie | Clément OYON, AGGLO VILLEFRANCHE BEAUJOLAIS - Services Techniques | Ronan PHILIPPE, La Métropole de Lyon – Direction de l'eau ...

COURANT 2019, le groupe de travail autosurveillance des réseaux d'assainissement proposera un utilitaire de calcul excel pour faciliter le calcul de ces critères d'auto-évaluation.

---

## ANNEXE 1 : Rappel de quelques notions de base

---

### Définition d'un modèle

Un modèle en hydrologie est "une représentation mathématique ou physique du cycle de l'eau construite dans le but de mieux comprendre son fonctionnement et de le faire évoluer." Encyclopédie de l'hydrologie urbaine et de l'assainissement - Bernard Chocat, Eurydice, 1136 pages, 1997, N°ISBN: 2-7430-0126-7

De façon imagée "Un modèle est une photographie de son système d'assainissement à l'instant T qu'il est nécessaire de régulièrement de réévaluer, vérifier, tester " Pascal Bret, Métropole de Lyon

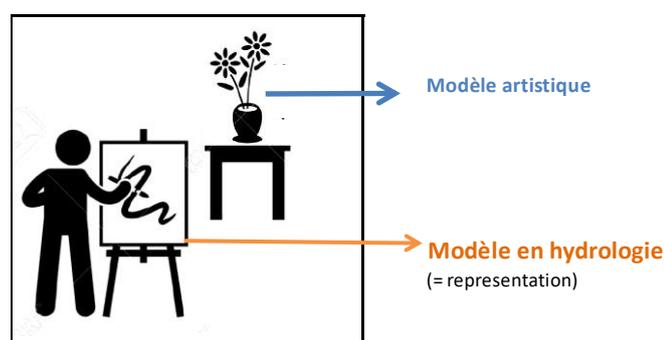


Figure 9: représentation schématique de la différence entre un modèle artistique et un modèle en hydrologie

### Conception du modèle :

Cette phase s'agit de définir le modèle mathématique qui pourra satisfaire au mieux les objectifs du modélisateur, en dépendant de la disponibilité de l'outil, les temps de calcul, l'interprétation physique du modèle et de l'information disponible.

### Les données :

Les données sont généralement les chroniques d'entrée et sortie utilisées dans un modèle mathématique. Celles-ci sont employées pour l'estimation des paramètres dans la phase de calage ainsi que pour l'évaluation de la performance du modèle avec les paramètres estimés dans la phase de vérification. Normalement, une partie des données est destinée à déterminer les paramètres du modèle (phase de calage) et une autre pour la vérification. Un critère courant de division de données est 70 % dans le calage et 30 % dans la vérification.

Les données doivent être traitées afin de mettre en œuvre le calage et la vérification d'un modèle mathématique. Néanmoins que la validation et le traitement des données est un sujet en dehors de la porte de ce document, les étapes suivantes de validation de données peuvent être conseillées :

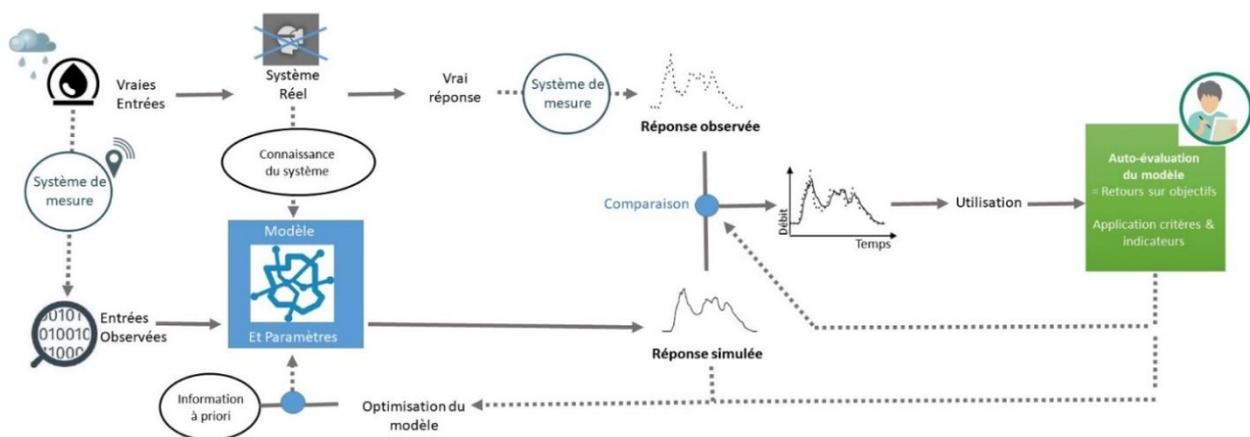
- Détecter et éventuellement remplacer les valeurs aberrantes (entrée et sortie) qui ne correspondent pas au comportement attendu d'un point de vu physique ou statistique. Ceci peut s'atteindre avec des techniques de filtrage des signaux de débit et/ou pluie (Métadier, 2011).
- Effectuer un lissage des données (entrée et sortie) pour éviter l'interférence du bruit des signaux

dans les estimations de paramètres (calage) et les indicateurs de performance dans la vérification. Ceci peut être atteint avec des techniques de lissage des signaux de débit et/ou pluie (Métadier, 2011).

- Une agrégation ou désagrégation spatial/temporelle des signaux et conseillée lorsque la résolution des données est plus fine ou plus grosse que la résolution de travail. Par exemple, un signal de débit peut être agrégé d'un pas de temps d'une minute à un pas de temps de 5 minutes, en prenant la moyenne des 5 données enregistrées dans la fenêtre temporelle correspondante aux 5 minutes. Les processus de désagrégation est beaucoup plus complexe, et s'il y a besoin d'en faire, il est plutôt conseillé de répéter les valeurs enregistrées à l'échelle désirée. Par exemple, si on souhaite désagréger un signal mesuré toutes les 5 minutes, on peut plutôt faire l'hypothèse que la même valeur a été enregistrée toutes les minutes, dans l'intervalle correspondant aux 5 minutes du signal original.
- Sélectionner les données correspondantes à une période représentative de modélisation. Il est important de considérer que les chroniques avec une grosse quantité de valeurs de zéros ont plus de tendance à rapporter des résultats positifs dans l'évaluation de la performance d'un modèle mathématique. Donc, si l'intérêt de modélisation est plutôt de représenter la dynamique du signal, l'on pourrait sélectionner des périodes sans beaucoup de données égales à zéro (très important pour les modèles de déversements).

### Calage :

La procédure de calage s'agit de l'estimation des valeurs optimales des paramètres d'un modèle mathématique sélectionné, permettant de satisfaire au mieux les objectifs du modélisateur. Ce sont ces derniers qui conditionnent la définition du critère d'optimisation, qualifié de fonction objectif. En général, les valeurs paramètres sont ajustées lors du calage, par rapport aux données et simulations du modèle. La mise en œuvre du calage nécessite donc le choix de la fonction objectif de calage et de l'algorithme mathématique d'optimisation pour la recherche des paramètres (Métadier, 2011). Ces paramètres sont le jeu de valeurs, parmi toutes les valeurs possibles, qui rendent les simulations de sortie d'un modèle données par quelques entrées connues le plus proche possible aux données observées de la variable de sortie (aussi appelé le problème inverse). L'estimation des paramètres obtenus par résoudre le problème inverse qui donne un jeu optimal de paramètres, peut être aussi appelée *identification des paramètres*. Lorsque les paramètres d'un modèle sont directement établis par l'expérience du modélisateur ou par de mesures physiques ou de terrain, on dira que le modèle est *paramétré*.



GRAIE - Inspiré de Métadier et al. (2011)

Les différents choix possibles d'algorithmes d'optimisation et des fonctions objectives pour le calage, ainsi que l'estimation des incertitudes de paramètres et des simulations, sont hors de la portée de ce document.

## **Vérification :**

L'acceptation d'un modèle mathématique ainsi que les valeurs de ses paramètres est absolument dépendant de la phase vérification. Un modèle sans vérification est un outil inutile qui ne permet pas délivrer des conclusions valides pour supporter l'aide à la décision. La vérification s'agit donc de la comparaison entre les résultats obtenus de la simulation avec un modèle mathématique et ceux qui ont été observés dans la réalité mais ne jamais connus par le modèle dans la phase de calage. Le control visuel des résultats est toujours conseillé en tant qu'une première approche de vérification. Néanmoins, il existe des autres critères plus objectifs pour évaluer la justesse des données de vérification simulées à partir des entrées connues par rapport aux données observées. (Métadier, 2011)