

**Université Pierre et Marie Curie, Université Paris-Sud,  
École des Mines de Paris  
& École Nationale du Génie Rural des Eaux et des Forêts**

---

**DEA Hydrologie, Hydrogéologie, Géostatistique et Géochimie  
Filière Hydrologie et Hydrogéologie Quantitatives**

Apport des Multi-Modèles pour la modélisation sur des bassins  
versants non jaugés

**Marie-Perrine Miossec**

**Directeurs de recherche : Vazken Andréassian (CEMAGREF)**

**Marie-Pierre Etienne et Eric Parent (ENGREF)**



**CEMAGREF Antony**



**Février-Septembre 2004**

**UNIVERSITE DE PARIS SUD**

## REMERCIEMENTS

Je tiens à remercier chaleureusement toute l'unité Qualité et Fonctionnement Hydrologique des Systèmes Aquatiques du CEMAGREF Antony, et notamment l'ensemble de l'équipe « hydrologie ». Je remercie tout particulièrement Vazken Andreassian qui m'a encadrée au cours de ce stage et qui a fait preuve d'une grande disponibilité et d'un fort enthousiasme. Je tiens par ailleurs à exprimer mes sincères remerciements à Marie-Pierre Etienne et Eric Parent du laboratoire GRESE de l'ENGREF pour leurs conseils et leurs interventions. Enfin je souhaiterais remercier mes « collègues stagiaires » Majida et Timothée pour les moments partagés au cours de ce stage.

## RESUME

*If you can look into the seeds of time, and say which grain will grow and which will not, speak then unto me. Shakespeare (Macbeth, I,3)*

La modélisation Pluie-Débit vise à simuler le débit des cours d'eau à partir des observations de pluies tombées sur le bassin versant (BV). Les modèles actuels ne parviennent pas à reproduire parfaitement la réalité, si bien que les modélisateurs cherchent des méthodes alternatives pour améliorer les performances de leurs simulations. Cela est particulièrement vrai dans le cas de BV non jaugés (BV où il n'y a pas de mesure de débit et donc de possibilité d'optimisation des paramètres), pour lesquels les simulations par les modèles sont très mauvaises. L'objectif de ce stage a été d'étudier les méthodes dites Multi-Modèles (M-M) susceptibles d'améliorer les performances de ces simulations. Deux méthodes M-M ont été développées basées sur un même principe : la combinaison linéaire des sorties de différents modèles. La première méthode (SAM) est une combinaison simple (moyenne arithmétique) tandis que la deuxième (WAM) est une combinaison où l'on affecte des poids spécifiques à chacun des modèles. Les résultats ont été assez surprenants puisque si chaque M-M a permis d'améliorer les performances des modèles seuls, la SAM a donné des résultats équivalents à ceux de la WAM. En ce qui concerne les BV non jaugés, la méthode SAM a pu être appliquée et a permis d'améliorer la qualité des simulations. La WAM a servi de base au développement d'une méthode dite d'ensemble qui pourrait permettre d'estimer de manière assez efficace les débits sur ce type de BV.

## ABSTRACT

*If you can look into the seeds of time, and say which grain will grow and which will not, speak then unto me. Shakespeare (Macbeth, I,3)*

Rainfall-Runoff models aim at simulating streamflows on the basis of rainfall measurements over the watershed. Current models do not perfectly reproduce the real runoff so that modellers try to find alternative methods to improve model simulations. This is particularly important in the case of ungauged basins (where no discharge measurement devices are available) for which model's predictions are weak. This study aims at developing a Multi-Model approach to improve prediction's quality. Two methods have been developed based on the linear combination of the outputs of different single models. The first method is the simplest and is known as Simple Average Method (SAM), and the second method is known as Weighted Average Method where specific weights are assigned for each model. The results of the study are quite surprising, Multi-Models methods do generate improvements in the quality of the predictions; but the SAM seems to be as efficient as the WAM. Finally, for ungauged basins, the SAM has been applied and predictions have been improved; as for the WAM, it has served as a basis for the development of an "Ensemble Method" to estimate the discharges of this type of catchment.

<b><u>1</u></b>	<b><u>Introduction</u></b>	<b>7</b>
<b><u>2</u></b>	<b><u>Constat d'une insuffisance des modèles Pluie-Débit</u></b>	<b>8</b>
<b>2.1</b>	<b>Principes et classification des modèles Pluie-Débit</b>	<b>8</b>
2.1.1	Principes de base de la modélisation Pluie-Débit	8
2.1.2	Diversité et classification des modèles	8
<b>2.2</b>	<b>Défis à relever en modélisation Pluie-Débit</b>	<b>9</b>
2.2.1	Les performances des modèles	9
2.2.2	Les simulations en mode non jaugé	10
2.2.3	Les incertitudes	10
<b>2.3</b>	<b>Possible complémentarité des modèles Pluie-Débit</b>	<b>11</b>
<b><u>3</u></b>	<b><u>Revue des principales méthodes Multi-Modèles</u></b>	<b>12</b>
<b>3.1</b>	<b>Principes généraux et revue des différentes méthodes</b>	<b>12</b>
3.1.1	Principe général	12
3.1.2	Trois méthodes principales	12
3.1.3	Différentes méthodes d'estimation des poids pour la méthode WAM	13
<b>3.2</b>	<b>Les principales méthodes M-M utilisées en hydrologie</b>	<b>13</b>
3.2.1	M-M pour la prévision en temps réel par Roche et Torterotot (1988)	13
3.2.2	M-M pour la prévision en temps réel par Coulibaly et al. (2003)	14
3.2.3	M-M en simulation et en prévision par Shamseldin et O'Connor (1996, 1999)	15
<b><u>4</u></b>	<b><u>Développement d'une méthode Multi-Modèle</u></b>	<b>17</b>
<b>4.1</b>	<b>Elaboration de la méthode</b>	<b>17</b>
4.1.1	Théorie de Bayes et mise à jour de la simulation à partir des informations des modèles	17
4.1.2	Méthode théorique pour le cas d'un seul modèle	17
4.1.3	Méthode théorique pour le cas de m modèles	18
<b>4.2</b>	<b>Test de la méthode M-M sur des données synthétiques</b>	<b>19</b>
4.2.1	Principe du test sur les données synthétiques	19
4.2.2	Résultats et intérêts du test sur les données synthétiques	20
4.2.3	Conclusions	21

<b><u>5</u></b>	<b><u>Application à la modélisation des bassins versants jaugés</u></b>	<b><u>22</u></b>
<b>5.1</b>	<b>Contexte du stage : l'initiative MOPEX (Model Parameter Estimation Experiment) et le 4<sup>ème</sup> atelier de travail MOPEX.</b>	<b>22</b>
5.1.1	L'initiative MOPEX	22
5.1.2	Intérêt de MOPEX pour tester la méthode M-M	22
<b>5.2</b>	<b>Description de la base de données</b>	<b>23</b>
5.2.1	Données disponibles	23
5.2.2	Descripteurs des BV	24
<b>5.3</b>	<b>Test du M-M en mode jaugé</b>	<b>25</b>
5.3.1	Principe	25
5.3.2	Modèles choisis	26
5.3.3	Résultats des MM sur la base de données	27
<b>5.4</b>	<b>Comparaison avec les résultats de Shamseldin et O'Connor</b>	<b>38</b>
<b><u>6</u></b>	<b><u>Application à la modélisation des bassins versants non jaugés</u></b>	<b><u>39</u></b>
<b>6.1</b>	<b>Utilisation du MM en mode non jaugé sur les BV de MOPEX</b>	<b>39</b>
6.1.1	Principe d'utilisation de la WAM sur des BV non jaugés	39
6.1.2	Limites de la WAM sur les BV non jaugés	40
<b>6.2</b>	<b>Une méthode alternative : la simulation par ensemble</b>	<b>40</b>
6.2.1	Principe	40
6.2.2	Tests sur les BV MOPEX	41
<b><u>7</u></b>	<b><u>Conclusions générales et perspectives</u></b>	<b><u>43</u></b>
<b>7.1</b>	<b>Conclusions</b>	<b>43</b>
<b>7.2</b>	<b>Perspectives</b>	<b>43</b>

## TABLE DES ILLUSTRATIONS

### Liste des Figures

Figure 1 : Principe de fonctionnement des modèles Pluie-Débit.....	8
Figure 2 : Localisation des 40 BV MOPEX en France.....	23
Figure 3: Débit annuel en fonction de la pluie annuelle pour les 40BV .....	25
Figure 4: Structure des résidus du modèle GR4J pour le bassin du Lunain à Episy (Seine et Marne).....	28
Figure 5: QQPLOT des résidus des modèles pour le bassin du Lunain à Episy .....	29
Figure 6: Hydrogrammes en période 2 pour le bassin du Lunain à Episy.....	30
Figure 7: Comparaison des performances des M-M et des modèles seuls en période 2, bassin par bassin en période 3.....	31
Figure 8 : Comparaison des critères de Nash de la WAM et des modèles seuls bassin par bassin.....	33
Figure 9 : Hydrogrammes en période 3 pour le bassin du Lunain à Episy.....	34
Figure 10: Représentation des critères de Nash de la SAM en fonction de ceux de la WAM bassin par bassin .....	35
Figure 11: Simulation par ensemble sur le BV $i$ .....	41
Figure 12 : Fréquences cumulées des critères de Nash transformés pour la simulation par ensemble .....	42

### Liste des Tableaux

Tableau 1 : Comparaison des critères de Nash pour des données synthétiques indépendantes .....	20
Tableau 2 : Comparaison des critères de Nash pour des données synthétiques corrélées.....	20
Tableau 3: Comparaison des critères de Nash pour des données synthétiques corrélées en imposant une matrice de variance-covariance diagonale.....	21
Tableau 4: Matrice de complémentarité entre les structures d'après Perrin (2000) .....	27
Tableau 5 : Comparaison des critères de Nash moyens en période 2.....	31
Tableau 6: Comparaison des critères de Nash moyens en période 3.....	32
Tableau 7: Nombre de bassins où le M-M est meilleur que le modèle seul .....	33
Tableau 8: Comparaison en critères de Nash de la WAM et de la SAM.....	35
Tableau 9 : Test réalisés pour la méthode M-M .....	36
Tableau 10 : Influence des modèles inclus dans le M-M.....	36
Tableau 11: Influence de la structure de la matrice de variance-covariance .....	37
Tableau 12 : Dégradation des performances pour un M-M basé sur $Y = Q + e$ .....	37
Tableau 13: Résultats des méthodes de Shamseldin et O'Connor en mode jaugé.....	38

# 1 Introduction

Les modèles Pluie-Débit sont des modèles hydrologiques particuliers qui permettent de simuler le débit des rivières à partir de l'observation des pluies tombées sur le bassin versant. Les objectifs de cette modélisation sont variés et on peut notamment citer :

1. **La simulation des débits** (dans le cas de lacunes de débits, cela permet d'étendre les chroniques à partir de celles de pluie) ;
2. **La prévision en temps réel** des débits des cours d'eau (notamment pour la prévision des crues ou des étiages) ;
3. **La pré-détermination des débits de crue ou d'étiage** (par exemple les valeurs de débits d'une crue de période de retour de 100 ans<sup>1</sup>) notamment dans la perspective de dimensionner des ouvrages et des aménagements dans le lit des cours d'eau ;
4. **La détermination des conséquences d'aménagements anthropiques** (barrages, occupation des sols...) sur l'hydrologie des rivières.

Dans notre étude, nous nous limiterons à l'étude des modèles Pluie-Débit dans le contexte de la simulation (point1).

Le nombre de modèles Pluie-Débit différents qui ont été développés à ce jour, atteint presque la centaine. Un certain nombre d'études comparatives ont été menées<sup>2</sup>, et il n'en ressort pas de tendance nette : si certains modèles sont en moyenne plus performants que les autres, aucun ne l'est de façon systématique. Par ailleurs, chaque modèle reflète la diversité des approches que l'on peut emprunter pour modéliser le comportement d'un BV. De ces constatations a pu émerger l'idée de combiner plusieurs de ces modèles au sein d'un Multi-Modèle (M-M) dans la perspective d'augmenter l'information prise en compte, d'exploiter les complémentarités éventuelles des modèles, et au final, d'améliorer les performances des simulations.

Cette problématique est d'autant plus importante pour des bassins versants (BV) dits non jaugés (c'est à dire où il n'y a pas d'appareils de mesure des débits). En effet, c'est quand aucune information de débit n'est disponible pour optimiser les paramètres des modèles, qu'une simulation par un modèle Pluie-Débit présente le plus d'incertitudes. Or, les BV non jaugés constituent la majeure partie des BV de la planète car les appareils de mesure coûtent très chers à l'achat comme à l'entretien. Il est donc vraiment nécessaire de tester des méthodes alternatives qui permettraient d'accroître les performances des simulations sur ce type de bassins.

Dans le contexte que nous venons de développer, ce rapport s'articulera autour de cinq points essentiels :

- Le constat des insuffisances des modèles Pluie-Débit actuels et donc l'intérêt potentiel de développer un Multi-Modèle.
- Une revue des principaux M-M utilisés en prévision ;
- Le développement de notre méthode M-M pour la simulation ;
- Les résultats de cette méthode sur des données réelles pour des bassins versants jaugés ;
- Les perspectives pour la simulation sur les bassins non jaugés.

<sup>1</sup> Le débit a une probabilité de 1/100 d'être atteint chaque année

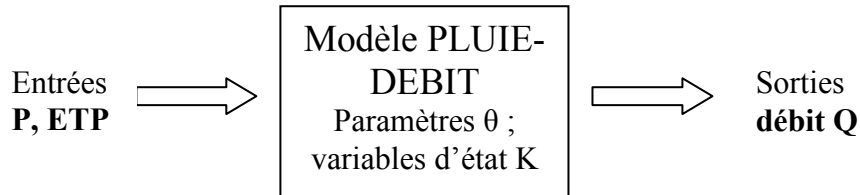
<sup>2</sup> Travaux qui ont débuté par l'intercomparaison de l'Organisation Mondiale de Météorologie (WMO) en 1975, les plus récents ont été effectués par Perrin (2000, 2001, 2003)

## 2 Constat d'une insuffisance des modèles Pluie-Débit

### 2.1 Principes et classification des modèles Pluie-Débit

#### 2.1.1 Principes de base de la modélisation Pluie-Débit

La majorité des modèles Pluie-Débit fonctionne selon le principe décrit dans la *Figure 1*.



*Figure 1 : Principe de fonctionnement des modèles Pluie-Débit*

Les **variables d'entrée** du modèle, encore appelées variables de forçage sont le plus souvent des mesures de la pluie (P) et des estimations de l'évapotranspiration potentielle (ETP). Les **variables de sortie** sont les débits (Q). Le modèle dépend de **variables d'état** qui traduisent les caractéristiques du système modélisé (par exemple le taux de saturation en eau du sol ; la profondeur de sol ; la pente...) et de **paramètres** représentant des caractéristiques du bassin qui n'ont pas toujours d'interprétation physique (Gaume, 2003-2004).

Une **étape de calage** permet d'estimer de façon empirique la valeur des paramètres en optimisant une fonction objectif du calage, encore appelée fonction coût. Cela consiste par exemple à minimiser l'erreur quadratique de modélisation :  $E = \sum_t (Q_{t, \text{simulé}} - Q_{t, \text{mesuré}})^2$ .

Une fois les paramètres calés<sup>3</sup>, une **étape de validation ou contrôle** est nécessaire pour évaluer les performances du modèle sur un jeu de données qui n'a pas été utilisé pour le calage.

#### 2.1.2 Diversité et classification des modèles

Il existe plusieurs dizaines de modèles Pluie-Débit et de nombreuses caractérisations et classifications de ces modèles ont été proposées. Ainsi, on peut distinguer les modèles stochastiques (qui fournissent les distributions des variables de sortie) et les modèles déterministes (qui fournissent des valeurs ponctuelles en sortie).

Parmi les modèles les plus courants, Perrin (2000) a proposé dans sa thèse la classification suivante :

- Les **modèles boîtes noires** : ils établissent un lien mathématique linéaire ou non linéaire entre les variables d'entrée et les variables de sortie (exemples : Processus Auto-Régressif ; Réseaux de Neurones) ;
- Les **modèles à réservoirs** : ils assurent la transformation de la pluie en débit au moyen de réservoirs interconnectés qui se vident et se remplissent au cours du

<sup>3</sup> D'après l'expression de l'erreur quadratique, on voit qu'il faut nécessairement connaître les séries de pluie et de débit sur la période considérée pour le calage, ce qui peut poser des problèmes dans le cas des bassins dits non jaugés pour lesquels on ne dispose pas de ce type de données.



temps. Les paramètres de ces modèles n'ont pas d'interprétation physique directe et sont estimés par les procédures de calage définies précédemment;

- Les **modèles fondés sur la physique** : ils sont basés sur la résolution numérique des équations de Saint Venant et de Richards et leurs paramètres sont des données physiques mesurables sur le terrain.

En réalité, il ne devrait pas y avoir de « conflits » entre ces différents modèles qui ont été développés à des fins différentes (Bergström, 1991). Ainsi, les modèles physiques sont plutôt utilisés pour une étude fine des processus hydrologiques tandis que les modèles conceptuels correspondent à une approche quasi statistique, où les paramètres correspondent à une moyenne sur l'ensemble du bassin. En conséquence, les modèles physiques sont souvent appliqués à des échelles petites où les caractéristiques physiques sont accessibles et mesurables. Des modèles physiques peuvent néanmoins être utilisés sur de très grands bassins et inversement. Dans notre étude, nous utiliserons essentiellement des modèles conceptuels à réservoirs.

## 2.2 Défis à relever en modélisation Pluie-Débit

### 2.2.1 Les performances des modèles

Pour quantifier les performances des modèles, il n'y a pas de critère d'évaluation universel. Le principe général est de comparer les débits calculés/simulés aux débits observés. De nombreux critères sont utilisés, notamment :

**Le critère de Nash-Sutcliffe (ou critère de Nash) :** 
$$R = 1 - \frac{\sum_i (Q_{i,obs} - Q_{calc,i})^2}{\sum_i (Q_{obs,i} - \overline{Q_{obs}})^2},$$

Il varie entre  $-\infty$  et 1. Il vaut 1 (ou 100%) si le modèle est parfait. Il prend des valeurs élevées (>80%) même si le modèle n'est pas très bon (40% d'erreur relative dans un critère de Nash de 84%). L'asymétrie de la distribution du critère de Nash-Sutcliffe peut poser des problèmes, par exemple si le modèle est très mauvais, il y aura des valeurs négatives très grandes en valeur absolue qui constituent un obstacle au calcul d'une moyenne. De plus, le critère de Nash-Sutcliffe accorde plus d'importance aux erreurs sur les forts débits. En pratique, il convient de garder cette propriété à l'esprit pour l'interprétation des performances des modèles ou de transformer les variables sur lesquels on calcul le critère (par exemple en calculant le critère sur les racines carrées des débits, on accordera la même importance à toutes les classes de débit).

On peut interpréter les critères en terme de **qualité** (ajustement du modèle à la réalité), de **robustesse** (conservation des performances d'un modèle de la phase de calage à la phase de contrôle), et de **fiabilité** (conservation des performances d'un modèle d'un bassin à un autre).

L'utilisation de ces critères suppose la connaissance des débits observés : ils ne sont donc applicables que pour des BV jaugés. En fait, dans le contexte de notre travail sur les BV non jaugés, nous supposerons les débits réels inconnus pour le calage, mais nous les utiliserons pour la validation afin de pouvoir calculer les performances de notre méthode.

Dans la bibliographie, la comparaison des performances moyennes des modèles Pluie-Débit, montre **qu'aucun modèle n'est meilleur que les autres de manière systématique**, c'est-à-dire dans toutes les circonstances possibles. D'une manière générale, l'objectif constant des modélisateurs est donc de chercher à améliorer les performances de leurs simulations.

### 2.2.2 Les simulations en mode non jaugé

Sur les bassins versants dits non jaugés, les simulations par les modèles Pluie-Débit sont souvent très médiocres : il s'agit de bassins pour lesquels les séries de mesures et d'observations hydrologiques ne sont pas suffisantes pour pouvoir utiliser les méthodes d'optimisation qui déterminent les paramètres des modèles. Un bassin est dit **non jaugé par rapport à une variable** donnée, par exemple la pluie, le débit, la concentration en sédiment, le taux d'érosion... En modélisation Pluie-Débit, on utilise le plus souvent le terme de bassin non jaugé pour des bassins qui présentent des lacunes importantes de débit et pour lesquels la procédure de calage des modèles paramétriques s'avère impossible.

Pour Wheeler et al. (1993), la simulation sur des BV non jaugés est possible, en théorie, pour des modèles à bases physiques car leurs paramètres sont mesurables. Il n'en est pas de même dans le cas des modèles conceptuels, pour lesquels une méthode d'estimation des paramètres doit être développée empiriquement. La méthode la plus courante est la **régionalisation** des paramètres. Il s'agit d'acquérir de l'expérience sur beaucoup de bassins afin de définir des relations entre les paramètres et les caractéristiques physiographiques et climatologiques de ces bassins.

Soit  $\theta_1, \dots, \theta_n$  les paramètres du modèle,  
 $\lambda_1, \dots, \lambda_k$  les variables physiographiques et climatologiques du bassin  
 On veut obtenir  $\forall i, 1 < i < n \quad \theta_i = f(\lambda_1, \dots, \lambda_k)$

Mais les résultats des méthodes de régionalisation ne sont pas très satisfaisants et des efforts continus doivent être portés sur l'amélioration de ces méthodes. Dans ce cadre, l'Association Internationale des Sciences Hydrologiques (IAHS) (Sivapalan, 2003) a lancé pour la décennie 2003-2012 une initiative pour la modélisation sur les bassins non jaugés (Prediction in Ungauged Basins, PUB). Les objectifs principaux de cette initiative sont **d'améliorer les performances** et de **réduire les incertitudes** des modélisations sur ces BV.

### 2.2.3 Les incertitudes

Les fortes incertitudes associées au fonctionnement des modèles Pluie-Débit sont aussi responsables de la faible confiance généralement accordée à leurs simulations.

Les principales sources d'incertitude concernent la structure même du modèle et la difficulté à estimer le jeu de paramètres optimal au cours de la procédure de calage. Désireux de représenter au mieux la réalité, on a observé parmi les modélisateurs une tendance à la **sur-paramétrisation et à la sur-complexité**. Pourtant, c'est pour ce type de modèles qu'il est le plus difficile d'estimer les valeurs optimales des paramètres ; à l'issue du calage, plusieurs jeux de paramètres optimaux sont obtenus, c'est ce qu'on appelle le phénomène d'équifinalité.

Pour éviter ce type de problèmes, un certain nombre de modélisateurs ont décidé de simplifier autant que possible leurs modèles. Ainsi, le modèle HBV développé par

Bergström est resté relativement simple, car rapidement, aucune amélioration des performances du modèle n'était observée lorsqu'on le complexifiait. Et même en suivant ces principes, selon Bergström, les modèles ont une complexité telle que personne n'en maîtrise entièrement le fonctionnement. Leurs résultats dépendent donc fortement du **jugement du modélisateur**.

D'autres **sources d'incertitude relèvent des données d'entrée** du modèle (mesures ou prévisions des pluies, évaluation de l'ETP, mesure des débits pour le calage...). Ainsi, même le modèle le plus performant ne réussira jamais à reproduire parfaitement la réalité (Bergström, 1991).

Dans la majeure partie des cas, les incertitudes ne sont pas quantifiées. Ainsi, un des défis majeurs à relever pour les modélisateurs ou les utilisateurs des modèles consiste à tenter de **caractériser les incertitudes** et à savoir **quelle confiance accorder aux résultats des modèles** (Bergström, 1991). Ensuite, l'objectif est de diminuer ces incertitudes. De manière générale pour accroître la confiance en un modèle, il faut le tester sur le maximum de données possibles.

### *2.3 Possible complémentarité des modèles Pluie-Débit*

Il existe une grande diversité des modèles Pluie-Débit. S'il existait une complémentarité entre certains de ces modèles, celle-ci pourrait être exploitée pour essayer d'améliorer les performances et diminuer les incertitudes des simulations. Dans sa thèse, Perrin (2000) a proposé une **analyse empirique** de la complémentarité entre modèles. Il a fait tourner 38 modèles conceptuels différents sur 429 BV et il a observé les complémentarités éventuelles entre ces modèles pris 2 à 2. Concernant la fiabilité (conservation des performances d'un BV à l'autre) Il a observé que certaines associations par paires permettent d'obtenir des résultats nettement plus performants que pour les modèles pris seuls à seuls. Ceci indique qu'il y a dans chaque modèle des composants spécifiques dont l'efficacité peut être complétée par d'autres et que les **structures des modèles étudiés sont en partie complémentaires**.

*La constatation des difficultés pour un modèle unique à prendre en compte la totalité de l'information et l'existence de potentielles complémentarités entre modèles a conduit un certain nombre d'hydrologues à s'intéresser aux méthodes Multi-Modèles (basées sur la combinaison de plusieurs modèles) dans le but d'augmenter les performances et de réduire les incertitudes des simulations.*

### 3 Revue des principales méthodes Multi-Modèles

#### 3.1 Principes généraux et revue des différentes méthodes

##### 3.1.1 Principe général

Les méthodes Multi-Modèles (M-M) sont toutes basées sur un même principe : **combinaison des sorties (pour nous, les séries temporelles de débit) de différents modèles.**

$$Y = F(Y_i),$$

avec Y la simulation par le M-M,  
 $Y_i$ , les simulations par les modèles individuels ( $0 < i < m$ )  
 F : fonction linéaire ou non linéaire

Les méthodes M-M ont tout d'abord été développées pour la prévision économique. L'un des textes fondateurs est l'article de Newbold et Granger (1974) qui a établi les premières équations de combinaison. Ces méthodes se sont ensuite étendues aux autres domaines de prévision tels que la climatologie avec notamment Raftery et al. (2003) ou l'hydrologie notamment dans le contexte de la prévision des crues en temps réel. Notre étude se place dans un contexte de simulation, mais nous nous inspirerons essentiellement des résultats développés en prévision.

##### 3.1.2 Trois méthodes principales

Un grand nombre de méthodes M-M ont été développées. Plusieurs revues de ces méthodes ont été proposées dans la littérature (Clemen, 1989; De Menezes et al., 1998). On y distingue notamment:

- The **Simple Average Method** (SAM): on combine les sorties des modèles par une pondération où tous les poids sont égaux et de somme égale à 1.
- The **Weighted Average Method** (WAM) ou méthode de la moyenne pondérée : les poids accordés à chaque modèle individuel varient en fonction des performances passées de ce modèle.
- Les **réseaux de neurones** qui traduisent une relation non linéaire entre le M-M et les modèles pris seuls à seuls.

La WAM est la méthode la plus utilisée dans le contexte hydrologique, mais il existe beaucoup de variations autour du principe initial. Clemen (1989), dans sa bibliographie annotée des méthodes M-M, énonce deux résultats principaux, l'un attendu, l'autre plus surprenant. Le premier souligne que l'ensemble des méthodes de combinaison engendre une diminution de l'erreur de prévision, tandis que le second montre que la méthode du Simple Average donne souvent des résultats aussi bons que des méthodes plus sophistiquées. Il montre aussi que plus le nombre de modèles pris en compte est élevé, plus la qualité de la prévision augmente. Mais ce résultat ne semble valable que jusqu'à 5 modèles et à partir de 6 modèles combinés, il n'y a plus de gain en précision.

### 3.1.3 Différentes méthodes d'estimation des poids pour la méthode WAM

Diebold et Lopez (1995) séparent les méthodes de combinaison en deux groupes selon la méthode utilisée pour estimer les poids de la combinaison linéaire: les méthodes bayésiennes de **variance-covariance** et les méthodes de **régression par les moindres carrés**. Les deux méthodes sont équivalentes si à l'estimation par les moindres carrés on ajoute les contraintes suivantes : terme constant nul et somme des coefficients de la régression (somme des poids) égale à 1 (Granger et Ramanathan, 1984).

De nombreuses variations existent autour des principes de base de la combinaison. On distingue notamment les méthodes où :

- Les **poids varient en fonction du temps** : on augmente le poids accordé à un modèle qui a des performances élevées aux pas de temps précédents l'estimation. Cette méthode est notamment très utilisée pour la prévision en temps réel.
- Les poids correspondent à la probabilité que les modèles effectuent la meilleure estimation. Dans ce cas, **un raisonnement bayésien** affecte aux modèles une probabilité à posteriori qui correspond à une probabilité à priori multipliée par les poids obtenus par moindres carrés (Bayesian Model Averaging développée par Raftery et al. (2003)).

## 3.2 Les principales méthodes M-M utilisées en hydrologie

Dans un but d'uniformisation, les notations dans ce document ne sont pas identiques à celles utilisées dans les articles d'origine.

Notations :

Q : le débit recherché ;

Y : la prévision du débit par le Multi-Modèle ;

$Y_i$  : la prévision du débit par le modèle i (i variant de 1 à m) ;

$e_i = Q - Y_i$  : l'erreur du modèle i ;

$w_i$  : le poids accordé au modèle i.

### 3.2.1 M-M pour la prévision en temps réel par Roche et Torterotot (1988)

Contexte :

Il s'agit d'améliorer la **prévision des crues en temps réel** sur deux cas d'application : la Vézère et la Garonne Amont. Le M-M comprend **cinq modèles** de type «Débit-Débit».

Méthode :

Le principe de base de leur méthode est une mise à jour bayésienne de la prévision par le Multi-Modèle à partir des informations apportées par chacune des prévisions des modèles individuels. La méthode employée est de type WAM et les poids sont estimés par la méthode de variance-covariance.

Après les calculs, les auteurs obtiennent la distribution de Y (la prévision par le M-M) qui suit une loi normale  $N(M,U)$ .

$$\text{Avec } M = \sum_i \frac{\sum_j \alpha_{i,j}}{\sum_{i,j} \alpha_{i,j}} y_i \quad \text{et} \quad U = \frac{1}{\sum_{i,j} \alpha_{i,j}}$$

où  $\{\alpha_{i,j}\} = V^{-1}$  et  $V$  est la matrice de variance-covariance<sup>4</sup> des erreurs  $e_i$  des modèles.

Etant donné qu'une estimation de  $Y$  est la moyenne sa distribution, on a donc :

$$Y = M = \sum_i w_i Y_i$$

avec  $w_i = \frac{\sum_j \alpha_{i,j}}{\sum_{i,j} \alpha_{i,j}}$ , les poids accordés aux différents modèles.

On remarque que les poids dépendent de la matrice de variance-covariance des erreurs des modèles.

Roche et Torterotot (1988) et Tamin (1986) proposent une alternative à cette méthode par une **pondération dite évolutive**. Le principe est d'augmenter le poids accordé aux modèles dont la qualité (performance) s'améliore sur les derniers pas de temps.

Ils se basent sur une méthode développée par Newbold et Granger (1974) qui permet d'estimer par une procédure empirique les poids accordés aux modèles. Ils obtiennent la formule suivante en faisant l'hypothèse que la matrice  $V$  est diagonale (c'est à dire que les modèles sont indépendants entre eux)<sup>5</sup>:

$$W_{i,t} = \frac{\left( \sum_{s=t-v}^{t-1} e_{s,i}^2 \right)^{-1}}{\sum_{j=1}^m \left( \sum_{s=t-v}^{t-1} e_{s,j}^2 \right)^{-1}}$$

modèles. Où  $v$  est le pas de temps considéré pour la mise à jour,  $m$  le nombre de

#### Conclusions de leurs études :

Il existe une grande variété de méthodes utilisables en prévision en temps réel. Que ce soit avec la pondération classique ou avec la pondération évolutive, les résultats semblent meilleurs qu'avec les modèles seuls, puisqu'une diminution de l'erreur quadratique moyenne est observée. La variance de la combinaison est toujours inférieure aux variances de chaque modèle, on gagne donc en certitude dans tous les cas.

### 3.2.2 M-M pour la prévision en temps réel par Coulibaly et al. (2003)

#### Contexte :

Il s'agit d'effectuer des **prévisions à 1, 2, 3, et 4 jours** sur deux BV québécois. Leur M-M est formé de **trois modèles** Pluie-Débit : deux modèles boîte noire (modèle du voisin le plus proche et réseau de neurone artificiel) et un modèle à réservoirs (HSAMI).

<sup>4</sup> L'élément diagonal  $V_{ii}$  de la matrice  $V$  représente la variance  $s_i^2$  de la variable  $e_i$  ( $s_i$  étant l'écart-type)  
L'élément non diagonal  $V_{ij}$  de la matrice  $V$  représente la covariance des variables  $e_i$  et  $e_j$

<sup>5</sup> Newbold et Granger (1974) avaient montré que la prise en compte des corrélations (éléments non diagonaux) n'entraînait pas d'amélioration des performances des prévisions

Méthode :

Comme dans le cas précédent, il s'agit d'une WAM dont les poids sont estimés par une mise à jour bayésienne à partir de la variance et des corrélations des erreurs de chaque modèle.

$$\text{On a alors: } Y = \sum_i w_i Y_i \text{ avec } w_i = \frac{\sum_j \alpha_{i,j}}{\sum_{i,j} \alpha_{i,j}} \text{ où } \{\alpha_{i,j}\} = V^{-1}$$

Avec V la matrice de covariance des erreurs des modèles, estimée pendant la période de calage.

Résultats :

Leurs résultats sont très bons puisque les performances (en critère de Nash) de la WAM sont supérieures à celles du meilleur des trois modèles sur les 2 bassins, quelle que soit la période considérée (calage ou validation) et pour chaque prévision (à 1, 2, 3 ou 4 jours).

Les conclusions de ces deux travaux montrent que la combinaison semble fournir de très bons résultats, que ce soit pour des modèles Débit-Débit ou pour de modèles Pluie-Débit.

**3.2.3 M-M en simulation et en prévision par Shamseldin et O'Connor (1996, 1999)**

Shamseldin et O'Connor ont travaillé dans deux contextes différents de modélisation : la simulation et la prévision en temps réel.

Contexte 1 (Shamseldin et O'Connor (1996)) :

Il s'agit de **simuler des débits** pour 11 BV. Le M-M a été appliqué à **5 modèles** de types différents (4 modèles boîte noire : modèle linéaire simple, modèle de perturbation linéaire, modèle de facteur de gain linéaire, système linéaire contraint avec un seul seuil ; et 1 modèle conceptuel : SMAR).

Méthode 1 :

Shamseldin et O'Connor (1996) ont développé **trois méthodes de « pondération »** :

- The Simple Average Method (SAM),
- The Weighted Average Method (WAM),
- The Neural Network Method (NNM).

Pour la WAM, la combinaison s'écrit :

$$Q = \sum_i w_i Y_i + e, \text{ où } e \text{ est le terme d'erreur de la combinaison ;}$$

Ils écrivent alors ces relations (vraies pour tout t) sous forme matricielle :

$$Q = Y W + E$$

Où Q est le vecteur (de taille T) des débits recherchés aux T temps

Y est la matrice (de taille T,m) des débits estimés par les m modèles aux T temps

W est le vecteur (de taille m) des poids accordés aux modèles

E est le vecteur (de taille T) des erreurs aux différents pas de temps

Une estimation par les moindres carrés leur permet d'évaluer les valeurs des différents poids W pendant la période de calage. Ces poids sont alors réutilisés pour les calculs des simulations par le M-M en contrôle.

Résultats 1:

Les résultats obtenus montrent qu'en calage le critère de Nash est supérieur pour la WAM par rapport aux modèles seuls pour les 11 BV. Pendant la période de contrôle, les résultats sont un peu moins bons : pour la WAM, la prévision en combinaison est meilleure que celle du meilleur modèle pour 6 bassins sur 11.

Contexte 2 : Shamseldin et O'Connor (1999):

Il s'agit d'effectuer une **prévision en temps réel** sur 5 BV. Le M-M comprend trois modèles individuels (2 modèles boîte noire et un modèle conceptuel SMAR).

Méthode 2 :

Ils ont mis en place une **procédure de mise à jour des pondérations** en utilisant les débits des séries temporelles précédant l'estimation pour chaque modèle individuel. Comme pour la pondération évolutive de Roche et Torterotot (1988), leur but est d'accorder plus de poids aux modèles qui ont été très performants sur les pas de temps antérieurs et proches du temps de l'estimation. Ils nomment leur méthode **RTMOCM** (Real-Time Model Output Combination Method). Elle est basée sur un processus Auto-Régressif pour les erreurs d'estimations, soit un modèle de fonction de transfert linéaire (d'ordre p,q) pour le débit estimé :

$$Q_t = \sum_{k=1}^p \Psi_k Q_{t-k} + \sum_{k=0}^q \xi_k Y_{t-k} + a_t$$
 Où  $a_t$  est un bruit blanc,  $\xi$  et  $\psi$  sont les paramètres du modèle de fonction de transfert linéaire et  $Q$  est le débit recherché.

A ce processus de transfert linéaire ils ajoutent une combinaison M-M de type WAM, et obtiennent comme formule pour la RTMOCM :

$$Q_t = \sum_{k=1}^p \Psi_k Q_{t-k} + \sum_{k=0}^q \sum_{j=1}^m \xi_k^{wj} Y_{j,t-k} + a_t$$
 où  $w_j$  est le poids du modèle  $j$ .

Les paramètres de cette équation sont alors estimés par la méthode des moindres carrés.

Résultats 2 :

Les résultats avec cette méthode de pondération montrent que :

- La prévision par la RTMOCM est de meilleure qualité qu'une mise à jour de chaque modèle pris individuellement ;
- Les résultats de la RTMOCM sont meilleurs que ceux d'une WAM sans mise à jour ;
- La RTMOCM (évaluation des coefficients de mise à jour et évaluation des pondérations synchrones) et la WAM suivie d'une mise à jour, ont des performances similaires.

*En conclusion, nous avons pu observer qu'il existait plusieurs méthodes Multi-Modèles utilisables. Dans l'ensemble, les résultats de ces méthodes sont très encourageants, notamment pour la prévision en temps réel. En conséquence, il nous a paru important de développer une méthode similaire dans un contexte de simulation et de l'appliquer à des données réelles.*



## 4 Développement d'une méthode Multi-Modèle

Nous avons vu au cours de l'analyse de la bibliographie que de nombreuses méthodes de combinaison ont déjà été développées dans un contexte de prévision. Dans notre cas, nous allons développer une méthode équivalente dans un contexte de simulation. La différence principale est que les informations récentes de débit ne pourront pas être utilisées pour mettre à jour les poids des modèles comme pour la pondération évolutive. C'est aussi la difficulté majeure rencontrée par cette technique.

### 4.1 Elaboration de la méthode

#### 4.1.1 Théorie de Bayes et mise à jour de la simulation à partir des informations des modèles

Le principe de la méthode est basé sur la statistique bayésienne, dans le but d'effectuer une prévision prenant en compte l'ensemble des informations apportées par chacun des modèles individuels.

Soit  $Q$  le débit inconnu,

Soit  $Y_i$  ( $i$  variant de 1 à  $m$ ) une prévision de  $Q$  par le modèle  $i$  ( $Y_i =$  informations)

Soit  $Y$  la prévision par le M-M

On veut déterminer la loi de  $Q$  obtenue à partir des informations  $\{Y_i\}$  c'est-à-dire la loi a posteriori de  $Q$ , qui donnera l'estimation du débit recherché par le M-M,  $Y$ .

On choisit une **loi a priori pour  $[Q]$  par exemple une loi normale  $N(\mu_0, \nu_0)$**

On suppose que les estimations  $Y_i$  varient linéairement avec  $Q$  :

Pour tout  $i$  :  $Y_i = a_i Q + b_i + e_i$       **(Équation 1)**

**On suppose de plus que les erreurs ont une distribution normale :  $e_i$  suit  $N(0, \sigma_i)$**

D'après l'équation 1, la loi de vraisemblance  $[Y_i/Q]$  suit donc une loi normale  $N(a_i Q + b_i, \sigma_i)$

D'après la **loi de Bayes**, la loi a posteriori est proportionnelle à la loi a priori multipliée par la vraisemblance :  $[Q/Y]_{\text{posteriori}} \propto [Q]_{\text{priori}} \cdot [Y/Q]$       **(Équation 2).**

L'équation 2 nous permet donc de déterminer la **loi a posteriori de  $Q$**  c'est-à-dire l'estimation par le Multi-Modèles.

#### 4.1.2 Méthode théorique pour le cas d'un seul modèle

Dans ce cas  $Y_1 = a_1 Q + b_1 + e_1$ ,      **(Équation 3)**

avec  $e_1$  qui suit  $N(0, \sigma_1)$  ; et  $Q_{\text{priori}}$  suit  $N(\mu_0, \nu_0)$

Par la formule de Bayes, on sait que :

$[Q/Y_1]_{\text{posteriori}} \propto [Q]_{\text{priori}} \cdot [Y_1/Q]$       **(Équation 5)**

Après les calculs,  $Q_{\text{posteriori}}$  suit une loi  $N(\mu_1, \nu_1)$  telle que

$$\frac{1}{\nu_1^2} = \frac{a_1}{\sigma_1^2} + \frac{1}{\nu_0^2} \quad \text{(Équation 6)}$$

$$\frac{\mu_1}{v_1} = \frac{\mu_0}{v_0} + \frac{a_1(Y_1 - b_1)}{\sigma_1^2} \quad \text{(Équation 7)}$$

On observe que la variance à posteriori est toujours plus faible que la variance a priori et que la variance de la vraisemblance, c'est-à-dire la variance de l'erreur du modèle 1,  $\sigma_1$ . La méthode bayésienne permet donc de diminuer les incertitudes.

### 4.1.3 Méthode théorique pour le cas de m modèles

On peut généraliser cette méthode au cas de m modèles afin de déterminer la loi a posteriori de Q, mise à jour à partir des prévisions des m modèles  $[Q/Y_1, Y_2, \dots, Y_m]$ .

On suppose que  $Y = Q.A + B + E$  (Equation 8)

Avec  $Y = (Y_1, Y_2, \dots, Y_m)^t$ , les simulations par les modèles ;

Q est un scalaire (le débit inconnu) ;

$A = (a_1, a_2, \dots, a_m)^t$ ,  $B = (b_1, b_2, \dots, b_m)^t$  ;

$E = (e_1, e_2, \dots, e_m)^t$  et suit une loi **multinormale**<sup>6</sup> d'ordre m:  $N^{(m)}(0, V)$

où, V est la matrice de covariance des  $e_i$ .

Comme précédemment, on applique la formule de Bayes (Equation 2), en choisissant pour [Q] une loi a priori normale  $N(\mu_0, v_0)$ .

D'après les expressions ci-dessus,  $[Y/Q]$ , la vraisemblance, suit une loi multinormale d'ordre m,  $N^{(m)}(Q.A+B, V)$

D'après l'équation 8 et les hypothèses ci-dessus, on obtient que  $[Q/Y]_{\text{posteriori}}$  suit une loi normale  $N(\mu, v)$  tels que :

$$\frac{1}{v} = A' V^{-1} A + \frac{1}{v_0}$$

$$\frac{\mu}{v} = \frac{\mu_0}{v_0} + A' V^{-1} (Y - B)$$

$$\text{D'où, } v^2 = \frac{v_0^2}{1 + v_0^2 \sum_{i,j} a_i a_j \alpha_{i,j}}, \text{ où } \{\alpha_{i,j}\} = V^{-1} \quad \text{(Equation 9)}$$

$$\text{Et } \mu = \frac{\mu_0}{1 + v_0^2 \sum_{i,j} a_i a_j \alpha_{i,j}} + \frac{v_0^2 \sum_{i,j} a_i \alpha_{i,j} (y_j - b_j)}{1 + v_0^2 \sum_{i,j} a_i a_j \alpha_{i,j}} \quad \text{(Equation 10)}$$

On a donc obtenu la loi de Q connaissant les prévisions des différents modèles.

L'espérance  $\mu$  de cette distribution correspond alors à l'estimation de Q par le M-M et on voit que c'est bien une combinaison linéaire des prévisions  $Y_i$  :

<sup>6</sup> La loi multinormale correspond à l'extension de la loi normale au cas de n variables  $x_1, x_2, \dots, x_n$ . Elle est caractérisée par un vecteur de moyennes  $\mathbf{m}$  et une matrice de variance-covariance  $\mathbf{V}$ . On la note  $N(\mathbf{m}, \mathbf{V})$ .

$$Y = \mu = K + \sum_i w_i Y_i \quad \text{avec} \quad w_i = \frac{v_0 \sum_j a_j \alpha_{i,j}}{1 + v_0 \sum_{i,j} a_i a_j \alpha_{i,j}} \quad \text{(Equation 11)}$$

Conclusions :

Nous avons donc élaboré une méthode M-M qui peut être classée parmi les méthodes **de type WAM** où les poids dépendent de la variance des erreurs et des corrélations entre les erreurs des modèles (Equation 11). Comme précédemment la variance a posteriori est inférieure aux variances a priori et aux variances de vraisemblance. On a donc **diminué les incertitudes**.

**Trois hypothèses majeures** ont été admises dans cette méthode concernant la structure des erreurs ou résidus des modèles :

- l'indépendance temporelle des résidus ;
- l'homogénéité des variances des résidus ;
- la normalité de la distribution des résidus.

4.2 *Test de la méthode M-M sur des données synthétiques*

Avant d'appliquer le M-M à des données réelles, on souhaite vérifier la validité de notre méthode sur des données que l'on génère.

4.2.1 **Principe du test sur les données synthétiques**

On sélectionne une série de débits Q dans un bassin versant « pilote », représentant les débits réels : Q<sub>obs</sub>. Puis on génère des séries de Y<sub>i</sub> (représentant les prévisions par les modèles) tels que Y<sub>i</sub> = a<sub>i</sub> Q<sub>obs</sub> + b<sub>i</sub> + e<sub>i</sub>, en choisissant les coefficients a et b ainsi que la structure des erreurs des modèles (corrélées ou non corrélées.).

On peut alors calculer Y une estimation de Q<sub>obs</sub> par la méthode M-M, et comparer les critères de Nash-Sutcliffe (N-S) pour les modèles individuels et pour le M-M.

Critère de N-S pour le Multi-Modèles : 
$$R^2 = 1 - \frac{\sum_t (Q_{t,obs} - Y_t)^2}{\sum_t (Q_{t,obs} - \overline{Q_{obs}})^2}$$

Critère de N-S pour chaque modèle i : 
$$R_i^2 = 1 - \frac{\sum_t (Q_{t,obs} - Y_{i,t})^2}{\sum_t (Q_{t,obs} - \overline{Q_{obs}})^2}$$

### 4.2.2 Résultats et intérêts du test sur les données synthétiques

On a travaillé sur des chroniques de 2 ans soit 729 données de  $Q_{obs}$ . On a testé deux cas :

- i. les erreurs des chroniques générées sont indépendantes,
- ii. les erreurs des chroniques générées sont corrélées.

Pour chaque cas, on a testé deux situations :

- (a) on génère 4 chroniques proches du  $Q_{obs}$  ce qui se traduit par un vecteur  $A = [0.85, 0.9, 1.1, 0.9]$  et un vecteur  $B = [0.1, 0.2, 0.1, 0.2]$ ,
- (b) on génère les 3 premières chroniques identiques par rapport au premier test, et la dernière très mauvaise par rapport à  $Q_{obs}$ . Cela se traduit par un vecteur  $B$  identique et un vecteur  $A$  dont le dernier coefficient est très différent de 1 :  $A = [0.85, 0.9, 1.1, 0.1]$ .

#### i. Génération de 4 modèles, dont les erreurs sont indépendantes

La génération des chroniques, le calcul de la chronique du M-M et l'estimation des performances de chacun des modèles par le critère de Nash (N) ont été réalisés avec MATLAB.

Les résultats donnent (avec  $N_i$  le critère de Nash du modèle  $i$  ;  $i \in [1,4]$ ) :

	$N_1$	$N_2$	$N_3$	$N_4$	$N_{wam}$
$A=[0.85,0.9,1.1,0.9]$	0.43	0.1	-0.1	0.45	0.83
$A=[0.85,0.9,1.1,0.1]$	0.43	0.1	-0.1	-0.36	0.78

*Tableau 1 : Comparaison des critères de Nash pour des données synthétiques indépendantes*

Comme « prévu », on observe (*Tableau 1*) que la chronique obtenue avec le M-M est bien meilleure que toutes les autres, son critère de Nash étant presque deux fois plus élevé que celui de la meilleure chronique.

#### ii. Générations de 4 modèles, dont les erreurs sont corrélées

Dans ce cas, on a distingué deux situations :

- Dans un premier temps, on a calculé les poids affecté à chaque modèle en utilisant la matrice de covariance des erreurs et on a obtenu les résultats suivants :

	$N_1$	$N_2$	$N_3$	$N_4$	$N_{wam}$
$A=[0.85,0.9,1.1,0.9]$	0.12	0.19	0.21	0.09	0.64
$A=[0.85,0.9,1.1,0.1]$	0.12	0.19	0.21	-0.72	0.89

*Tableau 2 : Comparaison des critères de Nash pour des données synthétiques corrélées*

De même, on observe (*Tableau 2*) que la chronique obtenue avec le M-M est bien meilleure que toutes les autres.

- Dans un deuxième temps, on a effectué la même opération en contraignant la matrice de corrélation à être diagonale (comme si les erreurs étaient non corrélées). On obtient alors :

	$N_1$	$N_2$	$N_3$	$N_4$	$N_{wam}$
$A=[0.85,0.9,1.1,0.9]$	0.12	0.19	0.21	0.09	0.48
$A=[0.85,0.9,1.1,0.1]$	0.12	0.19	0.21	-0.72	0.50

*Tableau 3: Comparaison des critères de Nash pour des données synthétiques corrélées en imposant une matrice de variance-covariance diagonale*

Ici encore (*Tableau 3*), les résultats du M-M sont meilleurs que ceux des modèles seuls, cependant ils sont nettement moins bons que lorsqu'on prenait en compte les covariances entre les différentes chroniques.

### 4.2.3 Conclusions

**Théoriquement la méthode choisie et les programmes réalisés sur le logiciel Matlab semblent corrects** puisque les critères de Nash du Multi-Modèles sont supérieurs à ceux des modèles seuls dans tous les cas. Il semble qu'il soit plus intéressant de calculer les poids en utilisant les corrélations entre les différents modèles car alors le M-M est plus performant.

Il est intéressant de remarquer que lorsqu'on introduit un modèle très mauvais, les résultats du M-M ne sont pas ou peu dégradés. Dans le cas du *Tableau 2*, les performances sont même améliorées par rapport au cas où tous les modèles sont de qualité équivalente. Cela peut s'interpréter comme le fait qu'un modèle toujours mauvais apporte aussi une information qui peut être prise en compte par le M-M. Dans le cas réel cela peut s'illustrer par un exemple : si un modèle fournit à chaque pas de temps une prévision correspondant à la moitié du débit réel, c'est un très mauvais modèle dans l'absolu, mais les informations qu'il apporte sont très intéressantes.

*Conformément aux résultats publiés dans la bibliographie sur les Multi-Modèles, et compte tenu des bons résultats théoriques de la méthode M-M, il semblait pertinent de tester cette méthode sur des données de BV réels.*

## 5 Application à la modélisation des bassins versants jaugés

### 5.1 Contexte du stage : l'initiative MOPEX (Model Parameter Estimation Experiment) et le 4<sup>ème</sup> atelier de travail MOPEX.

#### 5.1.1 L'initiative MOPEX

Le programme MOPEX a démarré en 1996. Son objectif premier est de développer des **techniques d'estimation a priori** (sans calage) pour les paramètres des modèles hydrologiques notamment pour la prévision sur les bassins versants non jaugés. Les méthodes d'estimation a priori sont le plus souvent fondées sur l'établissement de relations entre les paramètres des modèles et les caractéristiques physiques des bassins (méthodes de type régression). Or pour l'instant, ces relations sont encore ténues. Il y a donc une grande incertitude inhérente aux paramètres déterminés par les procédures a priori et donc dans les prévisions qui leur sont associées. L'objectif de MOPEX est de déterminer des procédures d'estimation a priori qui permettent de réduire cette source d'incertitude.

Le programme a aussi pour objectif de former une base de données la plus importante possible contenant les séries temporelles de pluie et de débit et les caractéristiques physiques pour des bassins du monde entier. L'intérêt du programme MOPEX est alors de favoriser les échanges d'idées et d'expériences entre les équipes de recherche internationales, en leur permettant de travailler sur les mêmes bases de données. Dans ce cadre, quatre ateliers internationaux (Workshops) ont déjà eu lieu, dont le quatrième s'est déroulé en **juillet 2004 à Paris**.

Les objectifs principaux de cet atelier étaient :

- la prévision sur les bassins non jaugés en régions humides,
- le développement de la base de données avec l'intégration de bassins français.

Les équipes de modélisation pouvaient choisir entre présenter leurs résultats selon une voie intensive (**40 bassins français**) ou extensive (3 ou 12 bassins parmi ces 40 BV français ou parmi ceux élaborés pour le 2<sup>ème</sup> Workshop de MOPEX à Tucson aux Etats-Unis). Selon ces différentes voies, une comparaison des méthodes d'estimation des paramètres a pu être réalisée. Concernant les BV non jaugés, la majorité des participants ont travaillé sur 3 ou 12 bassins.

Au cours de ce Workshop, **deux présentations des méthodes M-M** ont été réalisées : l'une par Shamseldin et O'Connor (Université de Galway, Irlande), l'autre dans le cadre de ce stage de DEA.

#### 5.1.2 Intérêt de MOPEX pour tester la méthode M-M

Les intérêts de participer à l'atelier MOPEX dans le cadre de ce stage étaient multiples. L'objectif initial était de combiner les simulations des différents participants afin d'illustrer la **plus-value du Multi-Modèle** par rapport aux modèles seuls. Pour des contraintes pratiques, les tests ont d'abord été réalisés sur des simulations obtenues avec des modèles dont les codes étaient disponibles au CEMAGREF. Néanmoins, quelques jours avant l'atelier, quand les simulations ont été disponibles, des tests ont pu être réalisés sur les simulations des participants. Ensuite, l'avantage du programme MOPEX était de pouvoir **comparer notre méthode M-M** à d'autres méthodes M-M présentées au cours de l'atelier notamment celles de Shamseldin et O'Connor. Enfin, la possibilité **d'utiliser la base de données** de MOPEX était très intéressante au vu de la diversité et de la qualité de ces données.

## 5.2 Description de la base de données

La base de données du 4<sup>ème</sup> Workshop de MOPEX a été fournie par METEO FRANCE pour sa composante « pluies » et « ETP » et par la BANQUE HYDRO pour sa composante « débits ». Elle contient 40 BV français (*Figure 2*).

### Figure Manquante

*Figure 2 : Localisation des 40 BV MOPEX en France*

#### 5.2.1 Données disponibles

Pour chaque BV, des données hydroclimatiques sont disponibles à un **pas de temps journalier et à un pas de temps horaire** :

- H1 : les séries temporelles de pluie,
- H2 : les séries temporelles de débit
- H3 : les séries temporelles d'ETP.

Les séries temporelles s'étendent sur 7 ans du 1/08/1995 au 31/07/2002.

Un certain nombre de données physiques sont aussi disponibles en plus des données hydroclimatiques :

- P1 : la localisation (avec en plus la localisation des stations pluviométriques, et la localisation des stations hydrologiques) ;
- P2 : la surface ;
- P3 : le relief ;
- P4 : l'occupation du sol.

Grâce à ces données, des descripteurs des bassins ont pu être calculés dans la perspective de pouvoir les caractériser et les comparer.

### 5.2.2 Descripteurs des BV

Les descripteurs des BV sont très utiles d'une part pour les **modèles à base physique**, d'autre part pour tester des **relations de régionalisation** pour la modélisation sur des BV non jaugés. Par ailleurs, ils facilitent l'interprétation des résultats des simulations par les modèles, que ce soit en mode jaugé ou non. Dans le cas des 40 BV MOPEX, les descripteurs suivants ont pu être calculés à partir des données précédentes :

- La surface (P2);
- L'écart type des altitudes (d'après P3);
- Le pourcentage de forêt (d'après P4);
- La Pluie moyenne annuelle (d'après H1);
- L'ETP moyenne annuelle (d'après H3);
- Le coefficient d'irrégularité des pluies (d'après H1);

La superficie des BV varie de 11 km<sup>2</sup> (pour le bassin de La Denante à Davayé, Saône et Loire) à 3234 km<sup>2</sup> (pour le bassin de La Loire Bas-en-Basset, Haute Loire) et la surface moyenne des BV est de 280 km<sup>2</sup>.

L'écart type des altitudes (plus l'écart type est élevé plus la pente moyenne du bassin est grande) varie entre 12m (bassin du ru d'Ancoeur à Blandy, Seine et Marne) et 439m (bassin du Loup à Villeneuve-Loubet, Alpes Maritimes) pour une moyenne de 128m.

Le pourcentage de forêts sur le bassin varie de 0.6% (pour le bassin du Guillec à Trézilidé, Finistère) à 77% (pour le bassin de la Lauch à Guebwiller, Haut Rhin) avec une moyenne de 30%.

Le coefficient d'irrégularité des pluies, traduit la variabilité saisonnière du climat et est mesuré en pourcentage :

$$CIP = 100 \times \left( \frac{P_{mx} - P_{mn}}{P_{moy}} \right) \text{ où } P_{mx} \text{ est la pluie moyenne du mois le plus pluvieux (en mm/mois) ;}$$

$P_{mn}$  la pluie moyenne du mois le moins pluvieux (en mm/mois) et  $P_{moy}$  la pluie moyenne (en mm/mois).

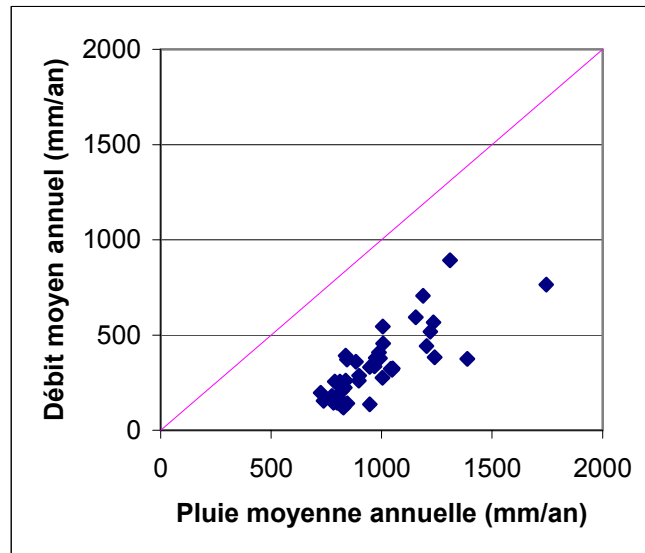
Il varie entre 41% (sur le bassin de l'Orgeval à Boissy le Châtel, Seine et Marne) et 162% (sur le bassin de La Cagne à Cagnes sur Mer, Alpes Maritimes) avec une moyenne de 83%.

Les pluies moyennes annuelles varient entre 709 mm/an (bassin de La Mauldre à Aulnay-sur-Mauldre, Yvelines) et 1721 mm/an (bassin de La Lauch à Guebwiller, Haut Rhin) avec une moyenne de 970 mm/an.

L'ETP moyenne annuelle varie entre 642 mm/an (bassin de L'Ingressin à Toul, Meurthe et Moselle) et 1236 mm/an (bassin de L'Artuby à la Bastide [Taulane], Var) avec une moyenne de 790 mm/an.

Le débit moyen ainsi que le rendement des BV (le rapport  $Q_{moyen}/P_{moyen}$ , visualisé en *Figure 3*) ont pu être calculés, mais ils ne peuvent pas être utilisés pour établir des équations de régionalisation pour les BV non jaugés où les débits sont théoriquement inconnus.





*Figure 3: Débit annuel en fonction de la pluie annuelle pour les 40BV*

Sur notre échantillon, les rendements variaient de 14% (pour le bassin du Lunain à Paley, Seine et Marne) à 68% (pour le bassin de l'Yonne à Corancy, Nièvre) avec une moyenne de 33%.

### 5.3 Test du M-M en mode jaugé

La méthode M-M que nous avons présentée dans la partie 4 a donc pu être testée sur cette base de données des 40 BV MOPEX à la fois sur données horaires et journalières.

#### 5.3.1 Principe

Il y a deux types de paramètres différents à caler : les paramètres des modèles hydrologiques et les paramètres du M-M ou poids. Dans ce cadre, il nous a semblé préférable de séparer la période de calage des paramètres hydrologiques et celle des paramètres du M-M. En effet, les poids du M-M dépendent de la variance des erreurs des modèles individuels. Donc si on cale les poids pendant la période de calage de paramètres hydrologiques, la variance des erreurs sera sous-estimée et les poids ne seront pas représentatifs. Cela nous a donc conduit à diviser les chroniques de pluie, d'ETP et de débit en trois périodes, plus une année de mise en route pour les modèles hydrologiques:

- **mise en route** des modèles hydrologiques : du 01/08/1995 au 31/07/1996 ;
- **période 1** : calage des modèles hydrologiques : détermination du jeu de paramètres optimal des modèles P-Q. Elle s'étend du 01/08/1996 au 31/07/1999 ;
- **période 2** : calage du Multi-Modèles : détermination des poids affectés à chacun des modèles. Elle s'étend du 01/08/1999 au 31/12/2000 ;
- **période 3** : période de contrôle ou validation: application du Multi-Modèle et des modèles seuls sur une série temporelle n'ayant pas servie aux calages. Elle s'étend du 01/01/2001 au 31/07/2002.

L'inconvénient de diviser les chroniques en trois est l'utilisation de périodes plus courtes donc un calage moins performant et surtout un contrôle moins significatif. Nous croyons néanmoins que c'est la méthode la plus correcte pour estimer les différents paramètres.

**Sur la période de calage 1**, on applique une méthode automatique d'optimisation des paramètres des modèles : la méthode « pas à pas » décrite dans la thèse de Perrin (2000).

**Sur la période de calage 2**, on doit caler les poids du Multi-Modèle.

Conformément aux équations établies dans le chapitre précédent, il est nécessaire pour calculer les poids de chaque modèle dans le M-M, d'évaluer un certain nombre de coefficients.

- Pour la loi à priori de  $Q$  (une loi normale  $N(\mu_0, v_0)$ ), on peut définir  $\mu_0$  et  $v_0$  à partir d'un histogramme des  $Q_{\text{obs}}$  sur la période 2 dont on calcule la moyenne et la variance.
- On fait tourner les modèles avec les jeux de paramètres définis en période 1 ce qui nous donne les estimations  $Y_i$ .
- On détermine alors les relations linéaires qui lient  $Y_i$  à  $Q_{\text{obs}}$  en effectuant une régression linéaire pour toute la période 2 avec MATLAB. Cela permet de déterminer les coefficients  $a_i$  et  $b_i$ .
- Enfin, après avoir estimé les erreurs de chaque modèle ( $e_i = Y_i - a_i Q_{\text{obs}} - b_i$ ), on peut déterminer la matrice de corrélation de ces erreurs avec MATLAB.

Une fois l'ensemble de ces coefficients quantifiés, on peut calculer la distribution de  $Y$  c'est à dire  $\mu$  (l'estimation de  $Q_{\text{obs}}$  par le M-M) et  $v$  (la variance de cette estimation) pour chaque pas de temps en utilisant les formules établies dans la partie 4.

**Sur la période de contrôle**, on fait tourner les modèles choisis avec le jeu de paramètres déterminé en période 1 et les poids affectés aux différents modèles qui ont été calculés en période 2.

**Pour évaluer les performances de notre Multi-Modèle** sur chacune des périodes (calage, période 2 et contrôle, période 3), on compare tout d'abord, les résultats de celui-ci aux résultats des modèles seuls. La comparaison en période 2 permet de valider la méthode (en effet le M-M doit nécessairement être le meilleur sur la période de calage). La comparaison sur la période 3 permet de véritablement évaluer la plus-value apportée par le M-M. Dans un second temps, on peut comparer la méthode (de type WAM) à une méthode M-M naïve comme la SAM ce qui permet alors de situer l'apport de notre méthode au sein des méthodes de la même famille.

### 5.3.2 Modèles choisis

Afin de tester notre méthode, nous avons tout d'abord effectué une combinaison sur des modèles dont les codes étaient disponibles au CEMAGREF. Il s'agit de 4 modèles conceptuels à réservoirs, modifiés pour la thèse de Perrin (2000):

- ✓ le modèle **GR4J** (GR4) développé au CEMAGREF Antony, France (Perrin, 2000).
- ✓ le modèle **HBV** (HBV) initialement développé au SMHI (Swedish Meteorological and Hydrological Institute), Suède (Bergström and Forsman, 1973).
- ✓ le modèle **IHACRES** (IHAC) initialement développé à l'Institute of Hydrology, Wallingford, Oxfordshire, Royaume Uni et à l'Australian National University, Canberra, Australie (Jakeman et al., 1990).
- ✓ Le modèle **TOPMODEL** (TOPM), initialement développé à l'Institute of Environmental and Biological Sciences, University of Lancaster, Royaume Uni, et à School of Geography, University of Leeds, Royaume Uni (Beven and Kirkby, 1979)

Ces modèles diffèrent notamment par le nombre de paramètres qu'il faut ajuster pour les caler. Pour GR4J, il y a 4 paramètres, pour HBV et IHACRES, 6 paramètres et 8 pour TOPMODEL. Bien qu'il s'agisse de 4 modèles conceptuels à réservoirs, on peut penser qu'il existe une certaine complémentarité dans leur fonctionnement. Ainsi, d'après l'étude de Perrin (2000) réalisée sur 429 BV avec 38 structures de modèles, une association par paire de ces modèles augmente leur fiabilité. Le *Tableau 4* représente les paires de modèles et le nombre de bassins pour lesquels ces modèles ou paires de modèles se classent parmi les 4 meilleurs. Pour les modèles choisis, on obtient le tableau suivant :

	GR4J	HBV	TOPMODEL	IHACRES
GR4J	185			
HBV	264	131		
TOPMODEL	261	246	157	
IHACRES	239	183	216	67

*Tableau 4: Matrice de complémentarité entre les structures d'après Perrin (2000)*

Dans tous les cas, on observe que l'association par paires de modèles augmente le nombre de bassins où les performances de ces modèles sont parmi les meilleures. Ce tableau semble donc bien indiquer qu'il existe une certaine complémentarité entre ces 4 modèles pris 2 à 2.

Il semble donc raisonnable d'effectuer les tests du M-M avec ces 4 modèles.

### 5.3.3 Résultats des MM sur la base de données

Nous avons comparé les résultats des différents modèles et méthodes M-M pour les données MOPEX au pas de temps horaire :

- GR4J ;
- TOPMODEL ;
- IHACRES ;
- HBV ;
- WAM (incluant les 4 modèles ci-dessus) ;
- SAM (incluant les 4 modèles ci-dessus) ;

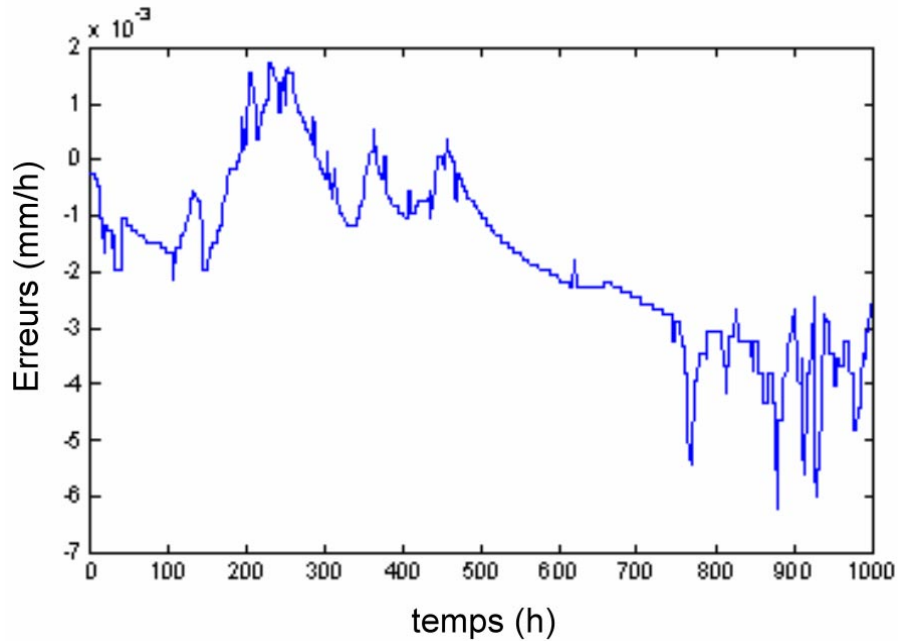
#### ➤ Vérification des hypothèses de la méthodes et Approximations

La méthode WAM présentée au chapitre 4 comporte trois hypothèses qui concerne la structure des erreurs:

- L'indépendance temporelle de ces erreurs ;
- L'homogénéité des variances ;
- La normalité.

Dans un premier temps, il convenait donc de vérifier si ces hypothèses étaient raisonnables dans le cas des données réelles que l'on s'appropriait à tester:

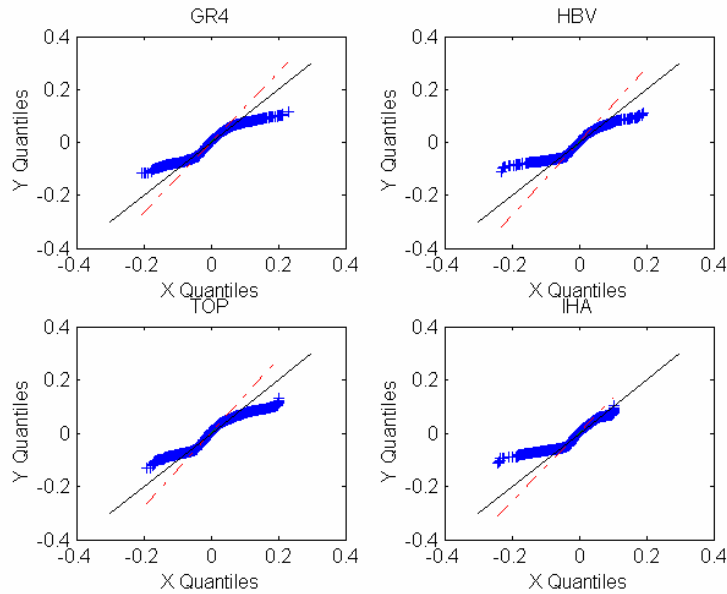
Concernant **l'indépendance des résidus**, on a tracé les erreurs pour chaque bassin et pour chaque modèle. Visuellement (*Figure 4*), on remarque immédiatement que les résidus ne sont pas indépendants. Or on souhaite que ces résidus constituent un bruit ; aucune structure ne devrait subsister. De plus, on obtient la même figure pour les 40 BV Il s'agit donc là d'une **approximation importante** dans notre méthode



*Figure 4: Structure des résidus du modèle GR4J pour le bassin du Lunain à Episy (Seine et Marne)*

Concernant l'**homogénéité des variances**, on a étudié la distribution de ces variances pour chaque modèle parmi les 40 BV. Par exemple pour la variance des erreurs de GR4, la moyenne sur les 40 BV est  $1.10^{-3}$  et la variance  $3.2.10^{-6}$ . Donc on peut considérer que l'hypothèse d'homogénéité est vérifiée et les résultats sont identiques pour les autres modèles individuels.

Concernant la **normalité des résidus**, on a testé celle-ci en traçant au moyen du logiciel Matlab des QQPLOT des distributions. Par exemple pour GR4J, on détermine la moyenne et la variance des erreurs et on compare la distribution des erreurs avec celle de la loi normale de même moyenne et de même variance. Si la distribution est normale, le QQPLOT donne une droite. Plus on s'écarte de cette droite (la première bissectrice) moins la structure des résidus sera normale. Dans notre cas, comme on peut le voir sur la *Figure 5* pour le bassin du Lunain à Episy, les résidus s'alignent presque sur la droite pour chaque modèle donc l'hypothèse de normalité est grossièrement vérifiée (et c'est aussi le cas pour les 39 autres BV).



*Figure 5: QQPLOT des résidus des modèles pour le bassin du Lunain à Episy*

**Conclusion :**

Dans notre méthode, nous avons fait une approximation importante, celle de l'indépendance des résidus. Pour être plus correct, il aurait fallu, éliminer la structure qui détermine encore ces résidus, par exemple en ajoutant un modèle de type Auto-Régressif (AR). Nous n'avons pas testé cette alternative dans la mesure où l'objectif de notre étude était d'utiliser la méthode WAM sur des BV non jaugés et où l'application d'un modèle AR sur BV non jaugé n'est pas envisageable. Cette approximation peut être la cause d'éventuelles mauvaises performances du M-M.

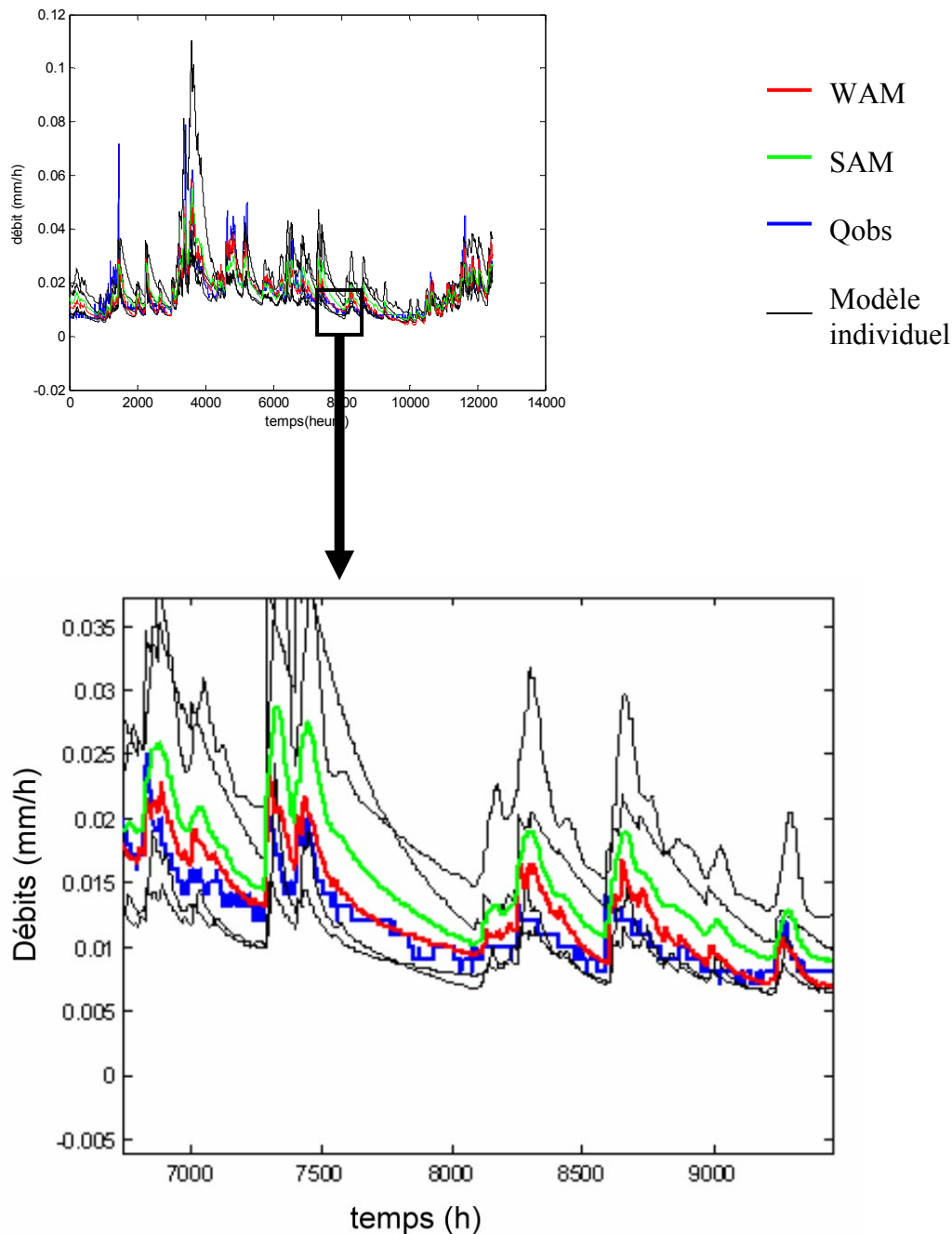
➤ Résultats en calage (période 2)

Pour comparer les performances relatives des modèles, on a recours à plusieurs indicateurs :

- Les hydrogrammes pour une approche visuelle,
- Les critères de Nash calculés en débit.

Les résultats en calage n'apportent pas vraiment d'information sur l'efficacité de la méthode. Ils doivent « forcément » être très bons pour les méthodes M-M. Il s'agit simplement d'un moyen de vérifier que les M-M sont meilleurs que les modèles seuls pour tester la validité de notre méthode et de nos programmes pour le logiciel Matlab (1<sup>er</sup> test : données générées ; 2<sup>ème</sup> test : valeurs en calage).

Si on s'intéresse tout d'abord aux hydrogrammes (*Figure 6*) ; cela permet de se rendre compte visuellement de la plus-value apportée par les M-M par rapport aux modèles seuls.

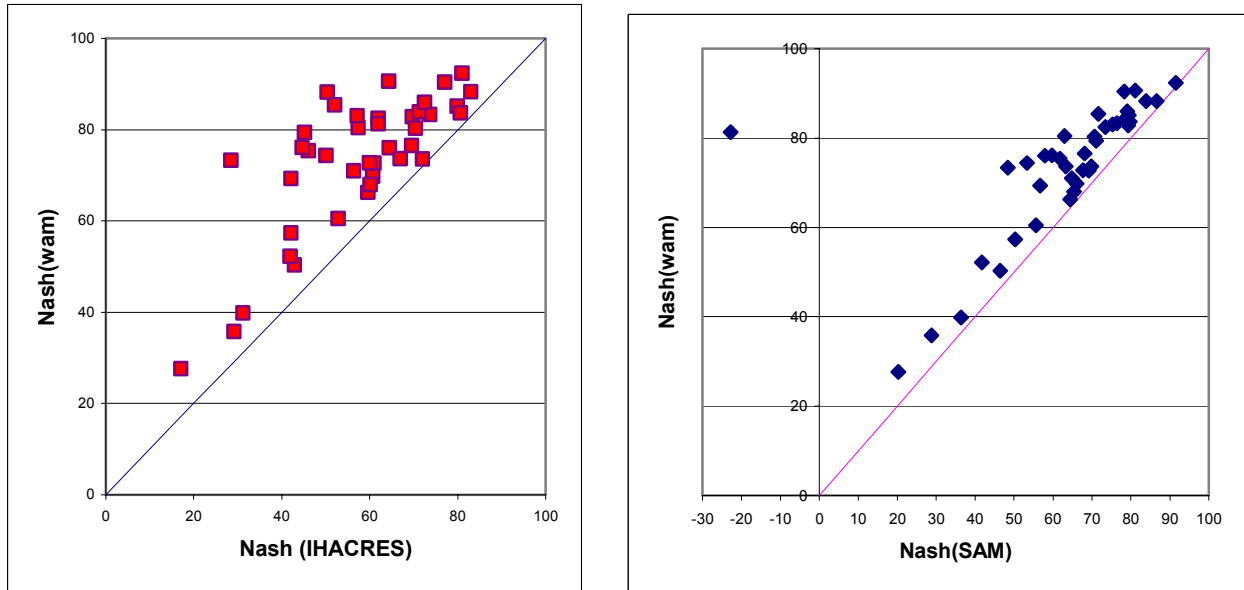


*Figure 6: Hydrogrammes en période 2 pour le bassin du Lunain à Episy*

D'après ces hydrogrammes, on observe bien que la courbe bleue (Qobs) est approchée au mieux par la courbe rouge (la WAM). Cette observation peut être confirmée de manière quantitative par la comparaison des critères de Nash obtenus selon les différentes méthodes :

- Comparaison bassin par bassin ;
- Comparaison des moyennes des critères de Nash sur la totalité des bassins.

Deux comparaisons sont réalisées : la méthode WAM par rapport aux modèles individuels et la méthode WAM par rapport à la méthode SAM.



*Figure 7: Comparaison des performances des M-M et des modèles seuls en période 2, bassin par bassin en période 3*

L'observation du premier graphique (*Figure 7*) nous indique que les critères de Nash en calage pour le M-M (WAM) sont supérieurs à ceux d'un modèle seul quelque soit le bassin considéré (ici IHACRES, mais les résultats sont identiques pour les trois autres modèles). De même, d'après le deuxième graphique (*Figure 7*), les performances de la WAM sont supérieures à celles de la SAM pour tous les bassins. On peut remarquer que pour un des bassins l'estimation par la SAM donne un critère de Nash négatif d'environ -22 (sur ce bassin à la fois GR4J et TOPMODEL ont des critères de Nash négatifs) alors que le critère pour la WAM sur le même bassin est de 81.

N (GR4)	N (HBV)	N (TOPM)	N (IHAC)	N (WAM)	N (SAM)
55.64	56.01	48.44	57.72	73.50	62.79

*Tableau 5 : Comparaison des critères de Nash moyens en période 2*

Les observations précédentes sont confirmées par la comparaison des critères de Nash moyens sur les bassins en période 2 (*Tableau 5*). La moyenne des critères pour la méthode WAM est de 15 points supérieure à celle du meilleur modèle (IHACRES) et de plus de 10 points par rapport à la SAM.

Conclusions :

On observe bien qu'en période 2 le M-M est meilleur que les modèles seuls et que le M-M simple (SAM) sur la totalité des bassins. Ces résultats tendent à prouver que la méthode développée (WAM) est théoriquement correcte. Cependant on ne pourra conclure à un avantage des méthodes M-M que si ces bons résultats sont conservés en période 3.

➤ Résultats en validation (période 3)

**Comparaison M-M et modèles seuls**

L'observation du Tableau 6 montre que le M-M en période 3 est plus performant que chacun des modèles seuls. En moyenne, le critère de Nash gagne 7 points par la méthode M-M. De même, la méthode qui est la plus efficace sur le plus grand nombre de bassin est aussi la WAM.

	GR4	HBV	TOPM	IHAC	WAM
Nash (moyen)	54.7	58.2	56.10	53.9	<b>65.3</b>
Nombre de bassins ou le modèle est le meilleur	2	1	14	0	<b>14</b>

Tableau 6: Comparaison des critères de Nash moyens en période 3

Il convient cependant de noter que la valeur moyenne du critère de Nash pour TOPMODEL n'est sans doute pas la plus représentative. En effet, sur un bassin, le critère de Nash pour TOPMODEL est très négatif (-188) et cela affecte la représentativité de la moyenne du critère de Nash. Ainsi l'observation de la Figure 8 ci-dessous montre que si l'avantage apporté par le M-M apparaît assez clairement pour IHACRES, GR4J et HBV, celui-ci est moins net pour TOPMODEL (Figure volontairement coupée pour la lisibilité, ne faisant pas apparaître le bassin dont le critère vaut -188).



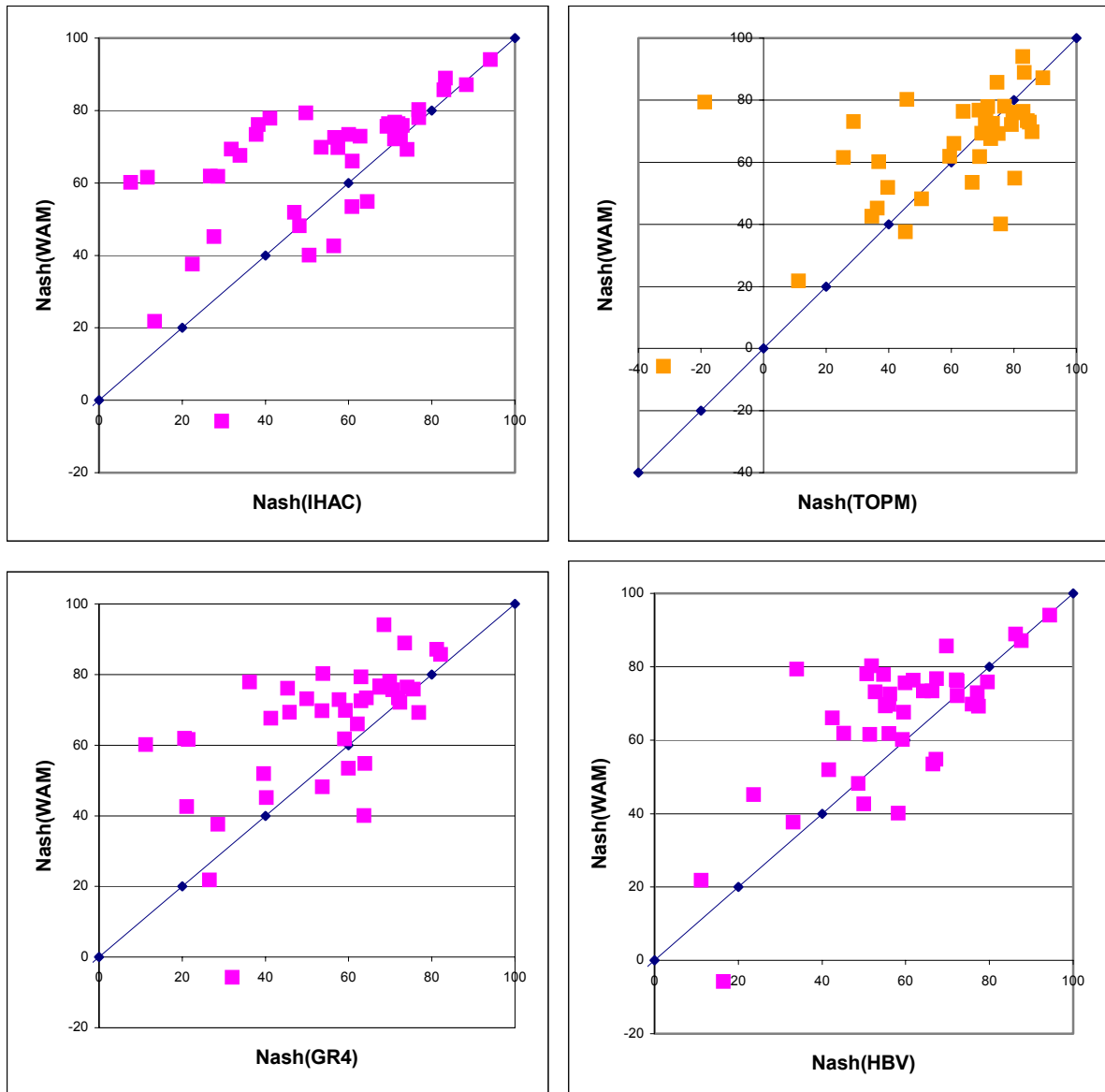


Figure 8 : Comparaison des critères de Nash de la WAM et des modèles seuls bassin par bassin

Ces observations sont d'ailleurs confirmées par le *Tableau 7* qui montre que pour GR4J, IHACRES et HBV, le critère de Nash de la WAM est supérieur à celui des modèles seuls pour environ ¾ des bassins et seulement pour la moitié des bassins dans le cas de TOPMODEL.

	GR4	HBV	TOPM	IHAC
SAM	38	33	24	36
WAM	33	28	22	33

Tableau 7: Nombre de bassins où le M-M est meilleur que le modèle seul

**Conclusions :**

L'avantage apporté par le M-M est certain pour 3 des modèles seuls : IHACRES, GR4J et HBV. Pour TOPMODEL, qui apparaît comme le plus performant des 4 modèles pour les données horaires, le M-M semble aussi apporter une légère amélioration.

Résultats comparés de la WAM et de la SAM

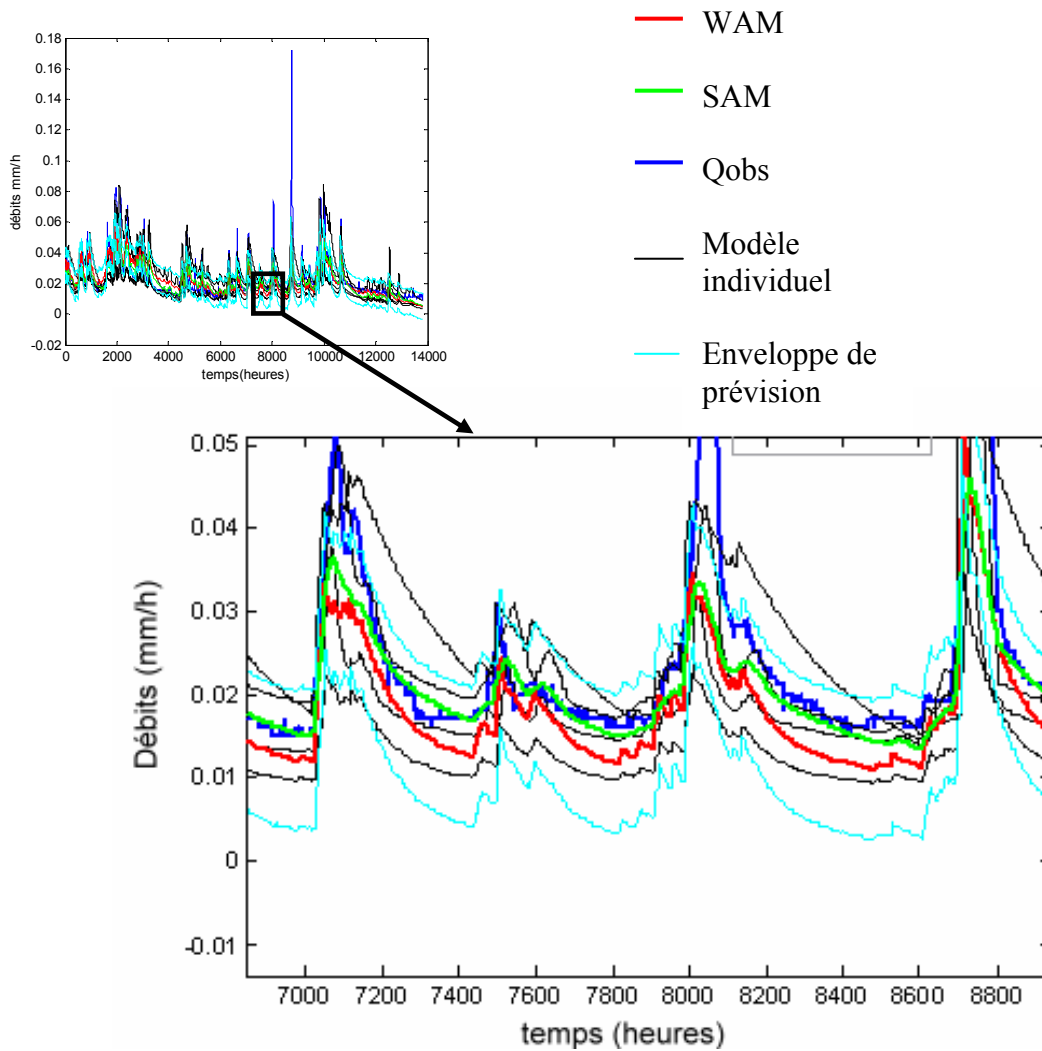
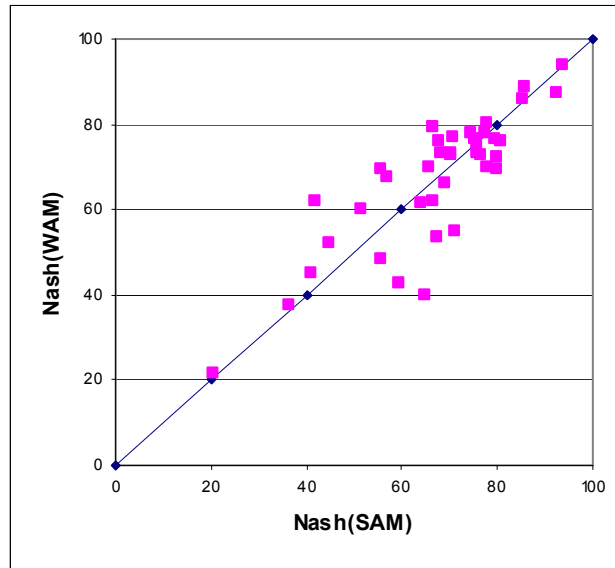


Figure 9 : Hydrogrammes en période 3 pour le bassin du Lunain à Episy

Contrairement aux hydrogrammes de la période 2, il semble qu'en période 3 (Figure 9), la courbe verte (SAM) se rapproche plus de la courbe bleue ( $Q_{obs}$ ) que la courbe rouge (WAM). Pour confirmer cette observation, on a tracé les critères de Nash bassin par bassin pour les deux méthodes (Figure 10).



*Figure 10: Représentation des critères de Nash de la SAM en fonction de ceux de la WAM bassin par bassin*

Au total, le nombre de bassins où la WAM est plus performante que la SAM n'est que de 21/40. Par ailleurs, si on reprend le *Tableau 7*, on remarque que la SAM est plus performante que les modèles seuls sur un plus grand nombre de bassins que la WAM. Au vu de ces résultats, il semble que les deux méthodes WAM et SAM donnent des résultats équivalents. C'est aussi ce qui apparaît dans le *Tableau 8* ci-dessous, où le critère de Nash moyen de la SAM est supérieur à celui de la WAM mais où le nombre de bassin où la méthode est la plus efficace est plus élevé pour la WAM.

	GR4	HBV	TOPM	IHAC	<b>SAM</b>	<b>WAM</b>
Nash (moyen)	54.7	58.2	56.10	53.9	<b>66.6</b>	<b>65.3</b>
Nombre de bassins où le modèle est le meilleur	2	1	14	0	<b>9</b>	<b>14</b>

*Tableau 8: Comparaison en critères de Nash de la WAM et de la SAM*

**Conclusions :**

Les 2 méthodes Multi-Modèles fournissent des résultats meilleurs que les modèles hydrologiques individuels utilisés. Mais le résultat le plus surprenant est que la méthode simple, la SAM donne des résultats presque équivalents à ceux de la méthode « élaborée », la WAM. Cela semble indiquer que la méthode WAM manque de robustesse (très bonnes performances en calage qui ne sont pas conservées en validation). Cela peut être dû à un certain nombre de facteurs, notamment :

- Les approximations réalisées quant à la structure des résidus et à leur indépendance ;
- la sur-paramétrisation due à la combinaison de différents modèles qui constitue un obstacle à la détermination des paramètres (poids).

Ces résultats sont néanmoins cohérents avec les observations de Clemen dans sa bibliographie annotée des méthodes M-M (Clemen, 1989) qui remarquaient que la SAM fournissait des résultats équivalents à ceux de méthodes plus élaborées. Les raisons avancées par certains modélisateurs résident dans la non stationnarité des phénomènes modélisés et l'instabilité des poids évalués. Les questions qui restent à être débattues sont :

- pourquoi la SAM fonctionne-t-elle si bien ?
- et sous quelles conditions les autres méthodes peuvent-elles mieux fonctionner ?

➤ Détermination des conditions de « meilleur » fonctionnement de la WAM

Malgré les nécessaires modifications qui restent à porter à la WAM pour améliorer ses performances, nous avons effectué plusieurs tests pour déterminer sous quelles conditions la méthode actuelle était la plus performante.

Le Tableau 9 ci-dessous représente les différents tests réalisés. Le test 1 correspond à l'expérience témoin. Dans les tests 4 et 5 on a ajouté un modèle aux 4 modèles initiaux : soit un modèle très bon (test 4 : ajout du modèle MORDOR développé par EDF) soit un modèle très mauvais (test 5 : ajout d'un modèle naïf : distribution normale connaissant la moyenne et la variance de  $Q_{obs}$ ). Dans les tests 2 et 3, on a modifié le mode de calcul des poids du M-M.

TEST WAM	Nombre de modèles inclus	Modèles inclus	Relation utilisée pour le calcul des erreurs	Corrélations
1	4	<b>GR4 ; HBV ; TOPM ; IHAC</b>	$Y = aQ + b + e$	<b>calculées</b>
2	4	GR4 ; HBV ; TOPM ; IHAC	$Y = Q + e$	calculées
3	4	GR4 ; HBV ; TOPM ; IHAC	$Y = aQ + b + e$	<b>Imposées à 0</b>
4	5	GR4 ; HBV ; TOPM ; IHAC; <b>MORDOR</b>	$Y = aQ + b + e$	calculées
5	5	GR4 ; HBV ; TOPM ; IHAC; <b>NAIF</b>	$Y = aQ + b + e$	calculées

Tableau 9 : Test réalisés pour la méthode M-M

**1<sup>er</sup> test : Variation du nombre et du type de modèles**

On va comparer les résultats des tests 1, 4 et 5.

	N(GR)	N(HBV)	N(TOPM)	N(IHA)	N(modèle5)	N(SAM)	N(WAM)
1	54.74	58.17	56.10	53.86		<b>66.63</b>	<b>65.33</b>
4	54.74	58.17	56.10	53.86	<b>63.97</b>	<b>67.21</b>	<b>66.12</b>
5	54.74	58.17	56.10	53.86	<b>-0.31</b>	<b>61.94</b>	<b>65.35</b>

Tableau 10 : Influence des modèles inclus dans le M-M sur la moyenne des critères de Nash

On peut noter que l'ajout d'un modèle meilleur (MORDOR) entraîne une augmentation des performances des deux types de M-M : SAM et WAM. La SAM conserve son léger avantage sur la WAM en terme de moyenne sur les critères de Nash.

Quand on rajoute un « mauvais » modèle, en revanche, la WAM devient plus performante que la SAM, car les informations apportées par le mauvais modèle sont « interprétées » dans le cas de la méthode WAM alors que la méthode SAM se contente de faire la moyenne arithmétique des prévisions. L'ajout du mauvais modèle entraîne une diminution des performances moyennes d'environ 5 points pour la SAM, alors que les performances restent équivalentes pour la WAM (*Tableau 10*).

**2<sup>ème</sup> test : Variations dans le mode de calcul des poids**

Tout d'abord on a testé les performances d'une WAM où la matrice de variance-covariance a été imposée diagonale (test 3) pour le calcul des poids des modèles. On observe que la WAM est moins performante si on annule les corrélations entre modèles (*Tableau 11*).

	N(GR)	N(HBV)	N(TOPM)	N(IHA)	N(SAM)	N(WAM)
1	54.74	58.17	56.10	53.86	66.63	<b>65.33</b>
3	54.74	58.17	56.10	53.86	66.63	<b>63.54</b>

*Tableau 11: Influence de la structure de la matrice de variance-covariance*

Conclusion 1 :

Conformément à ce que l'on avait obtenu pour le test sur données synthétiques, il semble plus correct de prendre en compte les corrélations entre modèles.

Ensuite on a comparé les résultats des tests 1 et 2, c'est-à-dire si on fait les calculs en supposant que les simulations  $Y_i$  varient de manière directe avec  $Q$  :  $Y_i = Q + e$  ; sans passer par l'estimation de coefficients linéaires  $a_i$  et  $b_i$ . On a alors observé une nette dégradation, en calage comme en contrôle. Par exemple, dans le *Tableau 12*, on a comparé le nombre de bassins où la simulation la plus performante est celle donnée par la WAM. Pour le test 2, la WAM n'est la méthode la plus performante que pour 3 bassins ce qui est très mauvais par rapport aux modèles seuls ou à la méthode SAM.

	Nombre de bassins où le M-M est 1 <sup>er</sup> en Nash
$Y = aQ + b + e$	14
$Y = Q + e$	3

*Tableau 12 : Dégradation des performances pour un M-M basé sur  $Y = Q + e$*

Conclusions 2 :

Il est préférable de conserver une structure basée sur des relations linéaires entre  $Y_i$  (les simulations par les modèles seuls) et  $Q$  (le débit recherché) telles que  $Y_i = a_i Q + b_i + e_i$ . Le M-M est alors beaucoup plus performant. L'inconvénient réside alors dans le nombre de paramètres à déterminer qui se trouve alors considérablement augmenté par les paramètres  $a_i$  et  $b_i$  et le risque de sur-paramétrisation de la méthode est d'autant plus important.

### 3<sup>ème</sup> test : simulations au pas de temps journalier

Les résultats obtenus au pas de temps journalier sont similaires à ceux obtenus au pas de temps horaire sauf que l'ordre de performance des modèles seuls est modifié (par exemple, au pas de temps journalier, c'est GR4J qui est le plus performant). Les poids accordés aux différents modèles sont donc différents, mais les résultats généraux de la méthode sont inchangés.

#### 5.4 Comparaison avec les résultats de Shamseldin et O'Connor

La présentation de Shamseldin et O'Connor à l'atelier MOPEX a montré les résultats qu'ils avaient obtenus sur 12 des BV disponibles<sup>7</sup>. Ils ont appliqué les méthodes présentées dans la partie 3 sur la revue des méthodes M-M. Ils comparent les résultats obtenus en combinant 7 modèles individuels (6 modèles boîte noire et un modèle conceptuel : SMAR) puis en combinant les 6 meilleurs etc... jusqu'aux 2 meilleurs modèles. Au total, ils ont donc testé 7 méthodes M-M. En mode jaugé, le meilleur des modèles simples est le modèle SMAR pour 8 bassins sur 12. Pour les 12 BV, les simulations M-M sont toujours plus performantes que le meilleur des modèles simples. Leur méthode semble donc être très efficace. Pour 8 bassins sur 12 c'est la combinaison des 7 modèles simples qui est la plus performante, mais d'une manière générale les performances des M-M restent assez proches (*Tableau 13*).

BV	1	2	3	4	5	6	7	8	9	10	11	12
Nash (meilleur modèle)	90.8	81.6	74.5	89.4	84.5	93.4	60.8	63.5	64.3	65.5	65.3	72.6
Min (Nash (M-M))	90.9	82.1	86	89.4	85.9	93.6	83	86.1	87.2	78.9	86.5	90.3
Max (Nash (M-M))	92.3	90.5	88.7	95.0	90.2	94.7	89	89.1	93.6	85.8	89	93.9

*Tableau 13: Résultats des méthodes de Shamseldin et O'Connor en mode jaugé*

Dans le cadre de l'atelier MOPEX, O'Connor a ensuite présenté les résultats qu'il avait obtenus en appliquant cette méthode pour des BV non jaugés. Les résultats sont plus mitigés puisque la méthode M-M ne fournit d'amélioration que pour 5 BV sur 11.

*Dans notre cas, nous allons aussi nous inspirer de la méthode M-M pour développer une méthode pour essayer d'améliorer les performances des simulations en mode non jaugé.*

<sup>7</sup> Les résultats ne sont pas tout à fait comparables dans la mesure où ils ont travaillé au pas de temps journalier

## 6 Application à la modélisation des bassins versants non jaugés

### 6.1 Utilisation du MM en mode non jaugé sur les BV de MOPEX

#### 6.1.1 Principe d'utilisation de la WAM sur des BV non jaugés

Sur les bassins non jaugés, on ne dispose pas d'informations de débit pour caler les paramètres de modèles hydrologiques (période 1) et ceux de la WAM (période 2).

Pour évaluer les paramètres des modèles hydrologiques, on peut utiliser des équations de régionalisation (reliant les valeurs des paramètres à des caractéristiques physiques des BV telles que la superficie ou l'épaisseur du sol...) qui sont établies par les concepteurs/utilisateurs des modèles.

Pour évaluer les poids qui seront accordés à chacun des modèles dans le M-M, une méthode (analogue à celle de la régionalisation des paramètres des modèles hydrologiques) consisterait à déterminer des relations reliant les poids des modèles aux descripteurs physiques et hydroclimatiques des BV.

Pour établir ces relations, on ne peut pas utiliser les résultats du mode jaugé, car les performances relatives des modèles varient en fonction de la situation dans laquelle on se trouve (mode jaugé ou non jaugé). En conséquence, il est nécessaire d'utiliser des simulations en mode non jaugé.

Une méthode consisterait à caler les poids des modèles sur 39 des BV en utilisant les 39 simulations en mode non jaugé pour chaque modèle. Dans ce cas, on peut parler d'un mode « **pseudo jaugé** » puisqu'on considère les 39 BV comme non jaugés pour les paramètres des modèles hydrologiques, puis on utilise de nouveaux les débits sur ces bassins versants pour caler le poids des modèles. Pour ces 39 BV, seraient donc disponibles à la fois les poids des modèles en mode « pseudo-jaugé » et les descripteurs des BV.

Soit  $BV_1, \dots, BV_{40}$  les 40 BV ;

Soit  $BV_j$  le bassin versant pour lequel on souhaite la simulation M-M non jaugée ;

$\forall k \in [1, 40], k \neq j$ , on a  $Y_1^k, \dots, Y_i^k, \dots, Y_m^k$  les simulations non jaugées par les  $m$  modèles sur le bassin  $k$  ;

On en déduit par calage :

$\forall k \in [1, 40], k \neq j$  les poids des  $m$  modèles sur le bassins  $k$  :  $w_1^k, \dots, w_i^k, \dots, w_m^k$

Par ailleurs, on a :  $\forall k \in [1, 40], k \neq j$   $D_1^k, \dots, D_x^k$  les descripteurs du BV  $k$  ;

Par régression multiple, on obtient :

$\forall i \in [1, m] \quad w_i = f(D_1, \dots, D_x) + \text{erreur} \quad \text{(Equation 12)}$

**Avec  $w_i$ , le poids accordé au modèle  $i$ .**

Et on applique cette équation au  $j^{\text{ème}}$  BV et on obtient les poids pour le  $j^{\text{ème}}$  BV non jaugé.

On répèterait alors ce processus 40 fois afin d'obtenir des simulations M-M non jaugées pour chacun des BV.

### 6.1.2 Limites de la WAM sur les BV non jaugés

Concrètement cette méthode n'est réalisable que si l'on dispose de simulations non jaugées pour les 40 BV. Or, dans le cas de MOPEX, aucun des participants n'a fourni de résultats non jaugés pour plus de 12BV. D'autre part, au CEMAGREF, des équations de régionalisation sont disponibles pour le modèle GR4J mais pas pour les autres modèles utilisés au cours de ce stage pour la simulation en mode jaugé (IHACRES, TOPMODEL et HBV). L'application de la WAM en mode non jaugée n'a donc pas pu être testée, faute de simulations pour un nombre suffisant de modèles et de BV.

De plus, on peut se demander s'il est vraiment raisonnable de vouloir **rattacher les poids des modèles à des descripteurs physiques du bassin**. En effet, il n'a pour l'instant jamais été démontré que tel type de modèle fonctionnait mieux sur tel type de BV.

#### Conclusion :

Si la méthode WAM n'a pu être appliquée sur des données non jaugées, le principe a néanmoins permis d'élaborer une méthode alternative : la méthode dite par ensemble.

## 6.2 Une méthode alternative : la simulation par ensemble

### 6.2.1 Principe

Le principe de la simulation par ensemble se rapproche du principe du M-M dans le sens où les simulations recherchées s'obtiennent aussi par une méthode de combinaison linéaire. Dans ce cas, ce ne sont pas les sorties des différents modèles que l'on combine, mais les sorties d'un même modèle sur d'autres BV de la région. En clair pour les données de MOPEX et avec le modèle GR4J :

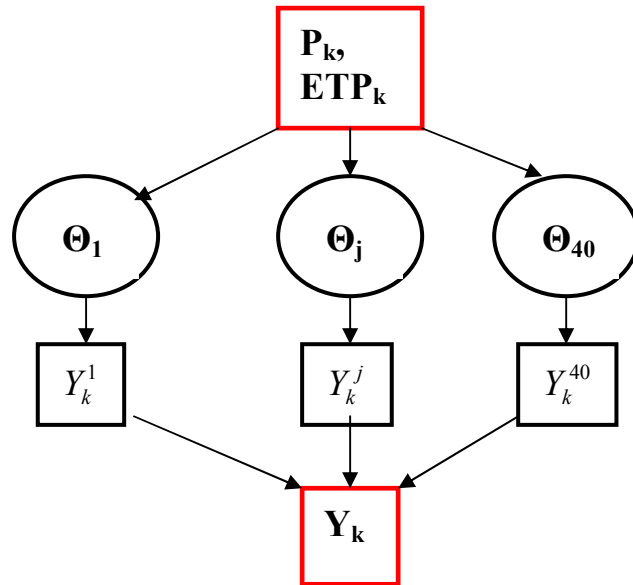
Soit  $Y_k$  la simulation recherchée sur le bassin  $k$  non jaugé, et  $Y_k^i$  les simulations obtenues sur les 39 autres BV avec les paramètres caractéristiques du BV  $i$  mais les entrées (pluie, ETP) de  $k$ .

$$Y_k = \sum_{i=1}^{40} w_i Y_k^i \text{ avec } i \neq k \quad \text{(Equation 13)}$$

Sur ces 39 bassins, le modèle GR4J a été calé au préalable, on a donc 39 jeux de paramètres différents  $\theta_i$  (un pour chaque bassin).

Sur la période considérée pour la simulation par ensemble, les données d'entrée injectées dans GR4J pour chaque bassin sont celles du bassin  $k$  et les paramètres utilisés sont ceux obtenus lors de la période de calage. On obtient alors 39 simulations que l'on combine d'après l'équation 13 pour donner la simulation sur le bassin  $k$  (*Figure 11*). On réitère alors ce processus pour l'ensemble des 40 BV.





*Figure 11: Simulation par ensemble sur le BV i*

Sur la base de ce principe, une simulation naïve revient à effectuer la moyenne arithmétique des simulations obtenues sur les autres bassins.

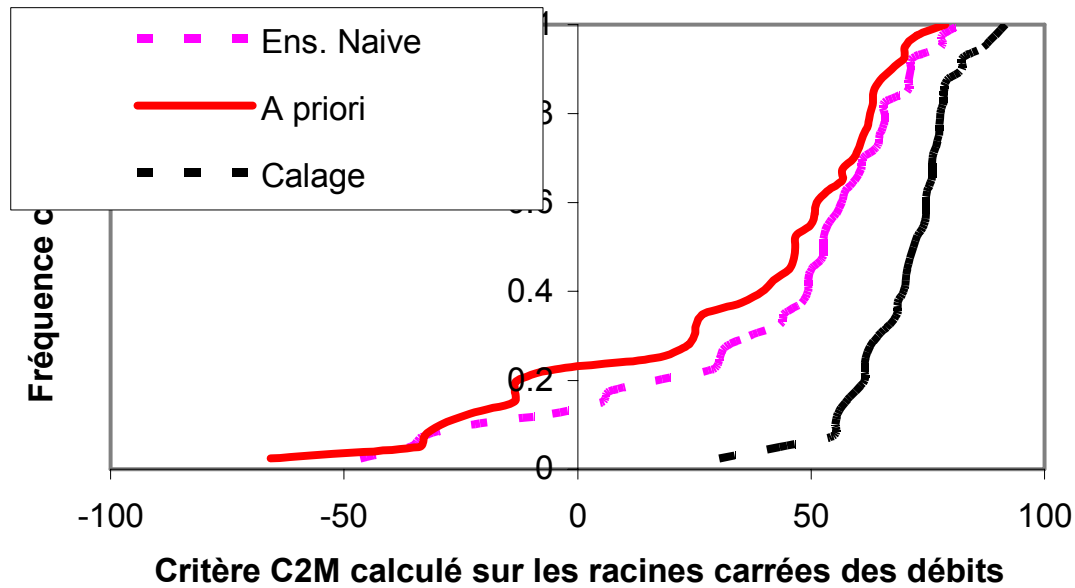
$$Y_k = \sum_{i=1}^{40} \frac{1}{39} Y_k^i \quad \text{avec } i \neq k \quad \text{(Equation 14)}$$

Pour améliorer cette simulation naïve, il s'agit comme pour la WAM, d'estimer les poids de chaque modèle de manière à optimiser les simulations (voir Chapitre suivant sur les Perspectives)

### 6.2.2 Tests sur les BV MOPEX

Un test automatisé de la méthode a été réalisé au moyen d'un programme en langage FORTRAN pour les 40 BV.

Les résultats peuvent être représentés graphiquement (*Figure 12*). On observe que les simulations obtenues se situent au sein d'une enveloppe délimitée pour sa partie inférieure par les simulations a priori et pour sa partie supérieure par les simulations en mode jaugé. Les simulations ont été évaluées au moyen du critère C2M qui est une version normée du critère de Nash et Sutcliffe symétrique (comprise entre -100 et +100).



*Figure 12 : Fréquences cumulées des critères de Nash transformés pour la simulation par ensemble*

Conclusions :

On observe que les résultats par la simulation naïve sont compris entre la simulation a priori et la simulation en mode jaugé. Cela signifie tout d'abord que la simulation par ensemble semble bien être une méthode efficace pour améliorer les simulations sur les bassins versants non jaugés. Cependant, cette méthode présuppose l'existence de BV jaugés « proches » du BV non jaugé auquel on s'intéresse (sinon il ne serait pas raisonnable de mixer les données d'entrée et les paramètres utilisés).

Si les résultats de cette simulation naïve sont encourageants, ils dressent aussi un grand nombre de perspectives à la fois pour mieux comprendre le fonctionnement de la méthode (par exemple on pourrait tester l'influence du nombre et du type de BV utilisés) et pour tenter d'améliorer les performances afin de se rapprocher au maximum des performances en mode jaugé.

*Que ce soit pour la WAM ou pour la méthode par ensemble, les principaux résultats obtenus dans le cadre de la simulation en mode non jaugé correspondent à des simulations naïves (moyennes arithmétiques). Les principaux défis qui restent à relever consistent donc à améliorer et affiner ces simulations en affectant des poids adaptés à chaque membre de la combinaison.*

## 7 Conclusions générales et perspectives

### 7.1 Conclusions

L'amélioration des techniques de prévision des débits sur les bassins versants non jaugés est l'un des objectifs majeurs que s'est fixée l'Association Internationale des Sciences Hydrologiques pour la décennie à venir. Une méthode qui a été proposée pour améliorer ces prévisions est l'**approche Multi-Modèle** qui consiste à combiner (combinaison linéaire ou non linéaire) les sorties de différents modèles Pluie-Débit supposés complémentaires.

La revue des différentes méthodes de combinaison de modèles qui ont été développées dans les contextes de prévision économique, atmosphérique ou hydrologique nous a permis de développer deux types de Multi-Modèle, l'un basique correspondant à une simple moyenne arithmétique (SAM), l'autre basé sur l'attribution de poids spécifiques à chacun des modèles pris en compte dans la combinaison (WAM). Ces méthodes ont été testées sur les données de 40 BV définis pour le quatrième Workshop du programme MOPEX (Model Parameter Estimation Experiment).

Dans un premier temps les M-M ont été testés sur un mode jaugé (c'est-à-dire lorsque le calage des paramètres des modèles est possible et efficace) à partir de quatre modèles disponibles au CEMAGREF. Les résultats ont montré que les méthodes M-M permettent bien d'augmenter les performances des simulations. Par ailleurs, les tests ont aussi montré que si la méthode WAM était plus efficace que la SAM en période de calage, les deux méthodes fournissent des résultats similaires en période de contrôle. Or, la méthode SAM est une méthode naïve de référence, cela signifie que la méthode **WAM doit être modifiée et complétée pour permettre l'amélioration de ses performances.**

Dans un second temps, on a souhaité appliquer ces méthodes à la simulation en mode non jaugé. Les méthodes M-M se sont avérées difficiles à mettre en œuvre, essentiellement parce qu'elles sont dépendantes de l'existence de simulations par des modèles simples en mode non jaugé qui sont très difficiles à établir. Dans ce contexte, une méthode alternative a été développée, il s'agit de la **simulation par ensemble** qui détermine les simulations sur le bassin non jaugé à partir des sorties des modèles sur des bassins jaugés de la région. Cette méthode semble prometteuse pour améliorer les simulations en mode non jaugé.

### 7.2 Perspectives

#### Sur les méthodes Multi-Modèles :

La première perspective qui peut être dressée à la lumière des résultats de cette étude est la nécessaire amélioration de la méthode WAM. En effet, dans le cas des BV de MOPEX on observe que la WAM donne des résultats similaires à ceux de la SAM. Or si on compare avec les résultats de Shamseldin et O'Connor pour les mêmes données, on s'aperçoit que leur méthode M-M fournit des résultats bien meilleurs qu'une simple moyenne arithmétique. Il faut donc améliorer notre méthode théorique de calcul du M-M. Ici deux perspectives différentes se présentent :

- La première consiste à rendre le M-M plus juste en corrigeant l'approximation relative à l'indépendance temporelle des erreurs des modèles individuels. Cette

correction pourrait notamment prendre la forme d'un modèle Auto-Regressif combiné au M-M déjà déterminé.

- La deuxième consiste, au contraire à simplifier l'ensemble de la méthode dans le but d'éviter une éventuelle sur-paramétrisation qui constituerait un obstacle à la bonne détermination des poids de la combinaison.

D'autres efforts pourraient être fournis pour appliquer la méthode M-M au cas des BV non jaugés par exemple en développant des équations de régionalisation pour différents modèles Pluie-Débit (les équation existent pour GR4J) et en essayant de combiner au mieux les simulations non jaugées fournies par ces modèles.

### Sur la méthode par ensemble

Un des objectifs majeurs consiste à essayer d'améliorer les performances obtenues au moyen de la simulation par ensemble naïve. Pour cela, une méthode consisterait à affecter aux BV des poids différents en fonction de leur degré de similarité avec le bassin auquel on s'intéresse. Le défi réside dans la **quantification de cette similarité**. Pour cela, on peut utiliser les descripteurs de bassins qui ont été calculés à partir des données disponibles :

- a : La surface du bassin ;
- b : L'écart type des altitudes ;
- c : Le pourcentage de forêt ;
- d : La Pluie moyenne annuelle ;
- e : L'ETP moyenne annuelle ;
- f : Le coefficient d'irrégularité des pluies ;

Dans le cas le plus simple, on s'intéresse à **un seul de ces descripteurs** pour évaluer le degré de similarité :

Soit  $a_i$  un descripteur sur le bassin  $i$  ; on recherche le poids  $w_i$  de la simulation sur le bassin  $i$  relativement au bassin non jaugé  $k$ .

Un indice de similarité peut être : 
$$\omega_i = \frac{1}{\left( \frac{a_i}{a_k} + \frac{a_k}{a_i} \right)} \quad \text{(Equation 15)}$$

et on en déduit directement le poids accordé au BV  $i$  en normalisant cet indice :

$$w_i = \frac{\omega_i}{\sum \omega_i}$$

A priori, on ne peut pas savoir quel descripteur permet de quantifier au mieux la similarité entre deux BV. L'idée est donc de tester la méthode par ensemble pour chaque descripteur (6 simulations par ensemble).

Une autre idée est de **combinaison plusieurs de ces descripteurs** pour calculer des poids qui prendraient en compte une information plus importante. Par exemple pour les descripteurs  $a$  et  $b$  **(Equation 16):**

On aurait :  $\omega_i = \left( \frac{1}{\frac{a_i + a_k}{a_k} + \frac{a_k}{a_i}} \right) \times \left( \frac{1}{\frac{b_i + b_k}{b_k} + \frac{b_k}{b_i}} \right)$  **(Equation 16)** et de même,  $w_i = \frac{\omega_i}{\sum \omega_i}$

On peut encore **affiner le calcul des poids** affectés à chaque simulation par BV, par exemple, en essayant d'optimiser les puissances affectées à chacun des descripteurs en fonction de leur « importance » pour quantifier la similarité entre deux BV :

$$\omega_i = \left( \frac{a_i + a_k}{a_k} + \frac{a_k}{a_i} \right)^{-\alpha} \left( \frac{b_i + b_k}{b_k} + \frac{b_k}{b_i} \right)^{-\beta} \dots \dots \text{(Equation 17)}$$

Sur les deux méthodes...

Enfin, il pourrait être intéressant de comparer les performances relatives d'une méthode par ensemble et d'une méthode Multi-Modèles sur un BV non jaugé.

Il pourrait d'ailleurs être envisageable de combiner les deux méthodes par exemple en effectuant des simulations M-M sur 39 BV et une simulation par ensemble pour le 40<sup>ème</sup> BV ! Mais le principal risque de cette méthode est la sur-paramétrisation puisque chaque combinaison conduirait à multiplier les paramètres introduits dans la modélisation.

## BIBLIOGRAPHIE

1. **Bergström, S.**, 1991. Principles and Confidence in Hydrological Modelling. *Nordic hydrology*, 22: 123-136.
2. **Bergström, S. and Forsman, A.**, 1973. Development of a conceptual deterministic rainfall-runoff model. *Nordic hydrology*, 14: 147-170.
3. **Beven, K.J. and Kirkby, M.J.**, 1979. A physically based, variable contributing area model of basin hydrology. *Hydrological Sciences Bulletin*, 24(1): 43-69.
4. **Clemen, R.T.**, 1989. Combining Forecasts: a review and annotated bibliography. *International Journal of Forecasting*, 5: 559-581.
5. **Coulibaly, P., Hache, M., Fortin, V. and Bobee, B.**, 2003. Improving daily reservoir inflow forecasts with model combination. *ASCE Journal of Water Resources Planning and Management*. Submitted for review.
6. **De Menezes, L.M., Bunn, D.W. and Taylor, J.W.**, 1998. Review of guidelines for the Use of Combined Forecasts. *European Journal of Operational research*, 120: 190-204.
7. **Diebold, F.X. and Lopez, J.A.**, 1995. Forecast evaluation and combination. Prepared for G.S. MADDALA and C.R. RAO., *Handbook of Statistics*.
8. **Fenelon, J.P.**, 1981. *Qu'est ce que l'analyse de données?* Lefonen, Paris, 311 pp.
9. **Gaume, E.**, 2003-2004. Hydrologie de versants et de bassins versants et modèles Pluie-Débit, Cours d'Hydrologie. ENPC, DEA Sciences et Techniques de l'Environnement.
10. **Granger, C.W.J. and Ramanathan, R.**, 1984. Improved methods of forecastings. *Journal of Forecasting*, 3: 197-204.
11. **Jakeman, A.J., Littlewood, I.G. and Whitehead, P.G.**, 1990. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology*, 117: 275-300.
12. **Nash, J.E. and Sutcliffe, J.V.**, 1970. River flow forecasting through conceptual models. Part I - a discussion of principles. *Journal of Hydrology*, 10: 282-290.
13. **Newbold, P. and Granger, C.W.J.**, 1974. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society*, 137: 131-146.
14. **Perrin, C.**, 2000. Vers une amélioration d'un modèle global pluie-débit au travers d'une approche comparative. PhD Thesis, INPG, Grenoble, 530 pp.
15. **Raftery, A.E., Balabdaoui, F., Gneiting, T. and Polakowski, M.**, 2003. Using Bayesian Model Averaging to calibrate forecasts ensembles. Technical Report, n°440, Department of Statistics. University of Washington.
16. **Roche, P.A. and Torterotot, J.P.**, 1988. Les procédures de décision en temps réel pour la prévision des crues. *La Houille Blanche*(5/6).
17. **Shamseldin, A.Y. and O'Connor, K.M.**, 1996. Methods for combining the outputs of different rainfall-runoff models. *Journal of Hydrology*, 197: 189-202.
18. **Shamseldin, A.Y. and O'Connor, K.M.**, 1999. A real-time combination method for the Outputs of different rainfall-runoff models. *Hydrological sciences*, 44.
19. **Sivapalan, M.**, 2003. IHAS decade on Prediction in Ungauged Basins (PUB) 2003-2012: shaping an exciting future for the hydrological sciences. *Hydrological sciences*, 8.
20. **Tamin, R.Z.**, 1986. Procédures de décision multi-modèles applicables à la prévision des crues en temps réel. Ph.D. Thesis.
21. **Wheater, H.S., Jakeman, A.J. and Beven, K.J.**, 1993. Progress and directions in rainfall-runoff modelling - Chapter 5. In: A.J. Jakeman, M.B. Beck and M.J. McAleer (Editors), *Modelling change in environmental systems*. John Wiley & sons, Chichester, pp. 101-132.

# ANNEXES : PRESENTATION AU WORKSHOP MOPEX



## INTRODUCTION

A major issue in water management is the need of accurate and reliable hydrologic predictions.

- ➔ Good Predictions lead to good decisions.
  - ◆ Need to improve existing Rainfall-Runoff Models simulations
  - ◆ Possible solution: “a Multi-Model approach” (Coulibaly et al. (2003) (submitted for review); Shamseldin and O’Connor (1996, 1999))

2nd July 2004 M.P. MIOSSEC. MOPEX 2

## Outlines of the talk

1. Why combine simulations?
2. Methods for combining
  - ◆ Different Methods
  - ◆ Our Methodology: a Bayesian Approach
3. Application to the MOPEX data
  - ◆ For Gauged Basins
  - ◆ For Ungauged Basins

2nd July 2004 M.P. MIOSSEC. MOPEX 3

## 1 Why a Multi-Model Approach?

- Existence of many Rainfall-Runoff Models
- BUT no single model identified to perform better for all types of catchments, under all circumstances.
- Consequences: uncertainties and potential danger when relying on a single model

*Can competing predictions be fruitfully combined, to produce a composite prediction superior to all single ones?*

2nd July 2004 M.P. MIOSSEC. MOPEX 4

## 2.1 Different methods to combine models’ outputs

- **The Simple Average Method (SAM):**

$$Y_{multi} = \frac{1}{m} \sum_i Y_i$$
- **The Weighted Average Method (WAM):**

$$Y_{multi} = w_0 + w_1 Y_1 + \dots + w_m Y_m + error$$
  - ◆ Weights are high for well-performing models
  - ◆ Weights are low for poorly-performing models
- **Non Linear Relations:** Neural Network Method

2nd July 2004 M.P. MIOSSEC. MOPEX 6

## 2.2 Discussions about combination methods (de Menezes et al. (1998))

- Agreement concerning the advantages:
  - ◆ Errors will be diversified and could be compensated
  - ◆ Based on more information.
- No more than 4 or 5 models in a combination
- How to estimate the weights? (WAM):
  - ◆ Regression methods (Least Squares, CLS...)
  - ◆ Variance-Covariance methods

2nd July 2004 M.P. MIOSSEC. MOPEX 7

## 2.3 Our Methodology: The Bayesian Theory

- Objective: find a posterior distribution for the runoff, R:  $[R]_{posterior}$
- Bayes Theory:
 
$$[R/Y]_{posterior} \propto [R]_{prior} \cdot [Y/R]$$
  - ◆  $[R]_{prior}$ : the prior distribution for R
  - ◆  $[Y/R]$ : the likelihood, (Y: single simulation)

2nd July 2004 M.P. MIOSSEC. MOPEX 8

## 2.3 Our Methodology: Assumptions

- $[R]_{prior} \longrightarrow N(\mu_0, v_0)$
- **Likelihood  $[Y/R]$ :**
  - ◆  $Y = R \cdot A + B + E$ : linearity
  - ◆  $Y = (Y_1, Y_2, \dots, Y_m)^T$ : single models simulations
  - ◆  $A = (a_1, \dots, a_m)^T$ ;  $B = (b_1, \dots, b_m)^T$ : coefficients
  - ◆  $E = (e_1, e_2, \dots, e_m)^T$ : errors  $\longrightarrow N^{(m)}(0, V)$
  - ◆ Normality and independence of the errors

➔  $[Y/R] \longrightarrow N^{(m)}(AR+B, V)$

2nd July 2004 M.P. MIOSSEC. MOPEX 9

### 2.3 Our Methodology: Equations of the Multi-Model

■  $[R/Y]_{\text{posteriori}} \longrightarrow N(\mu, \nu)$

$$\mu = K + \sum_i w_i Y_i$$

■ Weights depend on : A, B and V (Variance-Covariance Matrix)

$$w_i = \frac{v_0^2 \sum_j a_j \alpha_{ij}}{1 + v_0^2 \sum_{i,j} a_i a_j \alpha_{ij}}$$

2nd July 2004

M.P. MIOSEC. MOPEX

10

### 3.1 Applications to gauged and ungauged catchments

■ Final Objective: improve Rainfall-Runoff simulations specially on ungauged basins

■ Methodology :

- ◆ apply the multi-model methods on gauged basins
- ◆ adapt the multi-model structure
- ◆ test on ungauged basins

2nd July 2004

M.P. MIOSEC. MOPEX

12

### 3.1 Application to MOPEX Data for gauged basins

- Hourly simulations on the 40 catchments of the French MOPEX database:
- Individual models: GR4J, HBV0, IHAC, TOPMO

Period 1 (1995-98)	Period 2 (1999-00)	Period 3 (2001-02)
Calibration of the R-R models' parameters	Calibration of the Multi-model's parameters (weights)	Validation

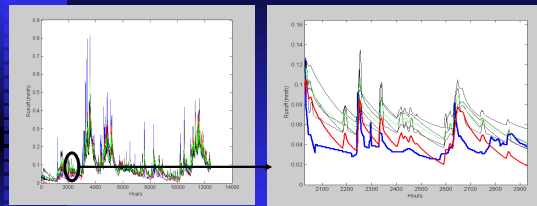
➡ 2 types of parameters to be calibrated

2nd July 2004

M.P. MIOSEC. MOPEX

13

### 3.1 Results in Calibration (1/2)



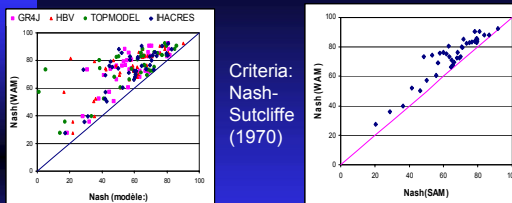
As expected in calibration, WAM (Red curve) is the best approximation of the real Runoff (Blue curve). SAM (green curve) is less accurate

2nd July 2004

M.P. MIOSEC. MOPEX

14

### 3.1 Results in Calibration (2/2)



Nash (multi) > Nash (models)

N(WAM) > N(SAM)

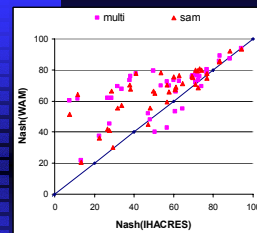
➡ Will these « expected results » in calibration be conserved in validation?

2nd July 2004

M.P. MIOSEC. MOPEX

15

### 3.1 Results in validation (1/2)



➡ Multi-Models are still better than single models for almost all catchments

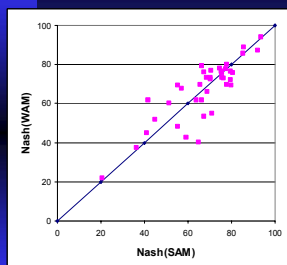
$N_{GR4J}$	$N_{HBV0}$	$N_{TOPMO}$	$N_{IHAC}$	$N_{sam}$	$N_{wam}$
54.7	58.2	60.3	53.9	66.6	65.3

2nd July 2004

M.P. MIOSEC. MOPEX

16

### 3.1 Results in validation (2/2)



Average of the Nash criteria

N(SAM)	N(WAM)
66.3	65.3

➡ WAM and SAM give equivalent results

2nd July 2004

M.P. MIOSEC. MOPEX

17

### 3.1 Conclusions for Gauged Catchments

■ Surprising (disappointing!) results:

- ◆ 1/5 of the catchments where a single model over performed.
  - ◆ WAM and SAM have similar performances.
- Causes:
- ◆ Approximations made for the WAM?
  - ◆ Ranking changes between calibration and validation?
  - ◆ Number of parameters? Complexity of the WAM?

■ BUT:

- ◆ Multi-Models still perform better than single Models.
- ➡ Good perspectives for ungauged basins

2nd July 2004

M.P. MIOSEC. MOPEX

18



### 3.2 Application of the Multi-Model to ungauged basins

- First step: Test the SAM with « ungauged simulations » provided for the 3 « main catchments » : (SACRAMENTO and VIC)
- Results: Nash-Sutcliffe criteria

	N(SACRAMENTO)	N(VIC)	N(SAM)
Le Guillec	48.8	66.3	75.6
LeToulourenc	34.6	42.4	61.7
Le Loup	32.8	77.4	68.1

 encouraging results for the SAM

2nd July 2004

M.P. MIOSEC. MOPEX

19

### 3.2 Perspectives for ungauged basins

- Test the SAM :
  - ◆ with other ungauged simulations
  - ◆ on other basins (12, 40)
- For the WAM:
  - ◆ Get ungauged simulations for 40 basins
  - ◆ calibrate weights for 39 basins
  - ◆ determine an optimal weights' set
  - ◆ test this set on the 40th basin

2nd July 2004

M.P. MIOSEC. MOPEX

20

### CONCLUSIONS and PERSPECTIVES

- **CONCLUSIONS:** WAM and SAM perform better than single models, but are quite similar.
- **APPROXIMATIONS:** independence and normality of the errors
- **PERSPECTIVES:**
  - ◆ **improve the wam:** Add an autoregressive structure?
  - ◆ **On the contrary: simplify the wam.** Weights depending only on the variance of the models?
- **Use other criteria to assess the Models' efficiencies**

2nd July 2004

M.P. MIOSEC. MOPEX

21

