

*"Goodness like beauty lies in the eye of the beholder and  
has many facets"*

*Jolliffe, I. et D. Stephenson, 2003*

### **3. La qualité des prévisions: critères d'évaluation et méthodes de correction de biais**

*Dans ce chapitre, nous présentons les méthodes utilisées pour examiner la qualité des prévisions d'ensemble et ses principaux attributs. Nous discutons également des différentes approches statistiques existantes dans la littérature pour améliorer la qualité des prévisions. Les approches exploitées dans cette thèse sont présentées.*

---



### 3.1 Introduction

Selon Murphy (cité dans Jolliffe et Stephenson, 2003), une "bonne prévision" peut l'être selon deux aspects : i) la qualité, qui examine la correspondance entre observations et prévisions, et ii) la valeur (ou utilité), qui concerne la valeur économique de la prévision pour un utilisateur décideur.

Dans ce chapitre, nous nous focalisons sur l'évaluation de la qualité des prévisions (l'utilité sera abordée au Chapitre 4). Le but est de comprendre les attributs que l'on souhaite retrouver dans les prévisions et d'introduire les critères appropriés qui seront utilisés dans cette thèse.

Une fois les critères d'évaluation définis, des approches statistiques qui cherchent à améliorer la qualité des prévisions sont présentées. Nous nous focalisons également sur les approches qui seront utilisées dans le cadre de cette thèse.

#### **3.1.1 Cadre général de l'évaluation des prévisions d'ensemble**

La qualité d'une prévision est en général jugée par rapport à l'observation. Dans le cadre simple de l'évaluation d'une prévision à scénario unique, on peut directement dire que la prévision est "correcte" ou "fausse" une fois l'événement observé. En revanche, dans le cadre des prévisions probabilistes ou des prévisions d'ensemble certaines particularités sont à prendre en compte dans l'évaluation de la qualité des prévisions. Puisque ces prévisions attribuent une probabilité à l'occurrence et à la magnitude d'un événement, les prévisionnistes doivent également évaluer cette information sur l'incertitude de la prévision émise. Il s'agit de déterminer comment la probabilité des événements prévus dans l'ensemble correspond à la fréquence à laquelle les événements sont observés (évaluation de la *cohérence* statistique des prévisions). Pour cela, il est nécessaire de comparer une longue série de prévisions avec la série correspondante des observations. Selon Jolliffe et Stephenson (2003), cette évaluation de la qualité d'une prévision d'ensemble consiste à trouver où se situe l'observation par rapport à la gamme de valeurs prévues par les différents membres de la prévision d'ensemble.

Ces différents membres de la prévision d'ensemble déterminent la densité de probabilité prédictive. La fonction de répartition correspondante donne les probabilités au non-dépassement pour différentes valeurs de la variable. L'évaluation de la prévision s'effectue en comparant, sur une longue série de paires "prévision-observation", la probabilité de prévision affectée par la prévision d'ensemble à l'observation et la fréquence empirique d'occurrence de l'observation.

---

Murphy et Winkler (1987) ont ainsi proposé une approche d'évaluation des prévisions centrée sur la distribution conjointe des prévisions et des observations. Cette approche est basée sur le principe que, si les prévisions et les observations sont des variables indépendantes et identiquement distribuées, la relation entre prévision et observation est décrite intégralement par leur distribution conjointe.

Deux lectures, dites "factorisations", de la distribution conjointe selon une distribution conditionnelle et une distribution marginale sont possibles :

- Factorisation 1 (en anglais "calibration-refinement factorisation") :

$$p(f, o) = p(o|f)p(f) \quad \text{Eq. 9}$$

, où  $p(o|f)$  : est la distribution de probabilité conditionnelle de l'observation  $o$  étant donnée la prévision  $f$ ,

$p(f)$  : est la distribution de probabilité marginale de la prévision  $f$ .

Les probabilités conditionnelles  $p(o|f)$  sont les probabilités prédictives, qui permettent de répondre, par exemple, à la question "quelle observation est la plus probable pour une prévision donnée ?". Cette distribution de probabilité conditionnelle est également appelée "distribution de probabilité *a posteriori*". Elle donne les valeurs possibles de l'observation en fonction des valeurs prévues par la prévision  $f$ .

- Factorisation 2 (en anglais "likelihood-base rate factorization") :

$$p(f, o) = p(f|o)p(o) \quad \text{Eq. 10}$$

, où  $p(f|o)$  : est la distribution de probabilité conditionnelle de la prévision  $f$  étant donnée l'observation  $o$ ,

$p(o)$  : est la distribution de probabilité marginale de l'observation  $o$ .

La distribution de probabilité conditionnelle  $p(f|o)$  est la fonction de vraisemblance. Elle permet de répondre à la question : "quelle prévision ressemble le plus à l'observation ?". C'est à cette question que nous nous adressons lors du traitement statistique des prévisions pour la correction des erreurs de prévision. Nous cherchons à identifier la distribution prédictive qui capture l'observation. Les erreurs des prévisions brutes sont corrigés de façon à aboutir, par exemple, à un maximum de vraisemblance (les méthodes de traitement statistique de correction de biais sont présentées dans §3.3).

---

Les deux factorisations de la distribution de probabilité conjointe sont équivalentes et reliées par le Théorème de Bayes :

$$p(o|f) = \frac{p(f|o)p(o)}{p(f)} \quad \text{Eq. 11}$$

La distribution de probabilité *a posteriori* (la distribution prédictive)  $p(o|f)$  résume l'incertitude d'estimation de l'observation étant donnée la prévision. Elle est proportionnelle au produit de la connaissance *a priori* de la variable à prévoir (l'incertitude autour de l'observation,  $p(o)$ , sans connaissance de la prévision) et de la fonction de vraisemblance  $p(f|o)$  (la fonction qui traduit combien le système de prévision est vraisemblable par rapport aux observations). A chaque prévision, la connaissance *a priori* est ainsi mise à jour par l'information supplémentaire qui vient des prévisions  $p(f)$  pour fournir la distribution prédictive qui est celle qui intéresse le prévisionniste. C'est à partir de cette distribution prédictive (la distribution *a posteriori*) que le prévisionniste aura l'information sur la probabilité d'occurrence d'un événement donné, par exemple.

### 3.1.2 Attributs de la qualité des prévisions

Les deux factorisations des probabilités conjointes présentées peuvent être utilisées pour déterminer les attributs définissant la qualité des prévisions. Murphy (1997) a montré que les moments de la distribution conjointe, des distributions marginales, ainsi que des distributions conditionnelles décrivent les différents attributs de la qualité des prévisions. La qualité d'une prévision est pourtant un concept multidimensionnel.

Par la factorisation 1, le biais conditionnel des observations  $o$  étant donnée la prévision  $f$  peut être mesuré. Dans la littérature, ce biais est souvent connu comme " Biais du Type I ". Il est une mesure globale de la fiabilité des prévisions (Bradley *et al.*, 2004). Mathématiquement, il est exprimé par :

$$\text{BiaisI} = E_f[(\mu_{o|f} - f)^2] \quad \text{Eq. 12}$$

, où  $E[]$  : est l'espérance,

$\mu_{o|f}$  : est l'espérance de l'observation conditionnée à la prévision.

Les prévisions conditionnellement non-biaisées, c'est-à-dire, fiables, sont désirables. La fiabilité correspond à la ressemblance entre les probabilités associées à l'observation par la prévision et les fréquences relatives d'occurrence des valeurs observées. Un système de prévision fiable est capable de bien représenter les incertitudes de prévision et de fournir des intervalles de confiance qui correspondent bien aux niveaux de confiance réels (ou qui sont proches des niveaux nominaux) des observations. Remarquons que Murphy (1973) a utilisé le terme de " validité " au lieu de " fiabilité ", avec une notion plus générale, liée à la prévision dans sa totalité.

Un autre attribut de la qualité des prévisions extrait par la factorisation 1 est la "résolution". Mathématiquement cette propriété est exprimé par :

$$RES = E_f[(\mu_{o|f} - \mu_o)^2] = \text{var}(\mu_{o|f}) \quad \text{Eq. 13}$$

, où  $E[]$  : est l'espérance,

$\mu_o$  : est l'espérance des observations,

$\mu_{o|f}$  : est l'espérance de l'observation conditionnée à la prévision,

$\text{var}()$  : est la variance.

La résolution est une qualité de la prévision à elle seule, indépendante de la fiabilité des prévisions. Un système de prévision d'ensemble possède une forte " résolution " (souhaitable) si la distribution prédictive est capable de fournir des probabilités distinctes des probabilités climatologiques ou, autrement dit, la résolution mesure quelle part de l'incertitude de l'observation peut être résolue par les prévisions (Weijjs, 2010). Une prévision basée sur la climatologie possède ainsi une bonne fiabilité, mais une très faible résolution, tandis qu'une prévision déterministe ne contiendra aucune information sur la résolution.

Brown et Seo (2012) ont fait une remarque intéressante sur ces deux attributs de la qualité des prévisions, fiabilité et résolution : la probabilité conditionnelle  $P(o|f)$  est à la fois une mesure de la fiabilité des prévisions quand elle est comparée à l'observation et une mesure de la résolution quand nous prenons en compte seulement sa sensibilité à la prévision  $f$ .

Si nous revenons à la factorisation 2, nous pouvons estimer le biais conditionnel des prévisions  $f$  étant donnée l'observation  $o$ . Dans la littérature, ce biais est connu comme " Biais Type II " (Bradley *et al.*, 2004 ; Brown et Seo, 2012). Mathématiquement, ce biais conditionnel est exprimé par :

$$\text{BiaisII} = E_o[(\mu_{f|o} - o)^2] \quad \text{Eq. 14}$$

, où  $E[]$  : est l'espérance,

$\mu_{f|o}$  : est l'espérance de la prévision conditionnée à l'observation.

Idéalement, les prévisions doivent être non-biaisées, c'est-à-dire, la valeur de ce biais doit être la plus petite possible.

L'autre attribut issu de la factorisation 2 est la " discrimination ". Cet attribut représente la capacité du système à discriminer entre deux événements distincts (ex., occurrence ou non d'un dépassement d'un seuil critique) et à attribuer des probabilités fortes à un événement lorsque celui-ci est observé. Par conséquent, les prévisions avec une discrimination plus élevée sont souhaitables, car cela veut dire que des prévisions différentes seront suivies par des observations différentes. L'expression mathématique de cet attribut est donnée par :

$$\text{DIS} = E_o[(\mu_{f|o} - \mu_f)^2] \quad \text{Eq. 15}$$

Indépendamment de la factorisation suivie, la multi-dimensionnalité de la qualité des prévisions est souvent difficile à synthétiser. Pour cette raison, il est souvent essentiel de faire référence à des concepts familiers. La notion de " justesse " en est un exemple.

Globalement, une prévision est considérée comme bonne si elle diffère peu de l'observation. Plus particulièrement pour une prévision d'ensemble, une prévision sera qualifiée comme bonne si tous les membres sont proches de l'observation, c'est-à-dire, si le faisceau de prévision est serré autour de l'observation (Figure 16).

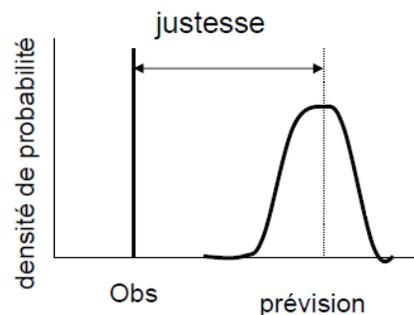


Figure 16 : Représentation schématique de la justesse d'une prévision.

La justesse est facilement calculée par une différence quadratique entre la prévision et l'observation. L'Eq. 16 exprime mathématiquement la justesse.

$$JUST = E[(f - o)^2] \quad \text{Eq. 16}$$

Dans le cas de prévisions déterministes, cette équation correspond à l'erreur quadratique moyenne des prévisions. En décomposant l'Eq. 16, nous obtenons :

$$JUST = (\sigma_f^2 + \mu_f^2) + \mu_o(1 - 2\mu_{f|o}) \quad \text{Eq. 17}$$

, où  $\mu_f$  : est l'espérance de la prévision,

$\sigma_f$  : est la variance de la prévision,

$\mu_o$  : est l'espérance de l'observation,

$\mu_{f|o}$  : est l'espérance de la prévision conditionnée à l'observation.

En suivant la factorisation 1 et en prenant en compte les Eq. 12 et Eq. 13, nous obtenons :

$$JUST = \text{BiaisI} - \text{RES} + \sigma_o^2 \quad \text{Eq. 18}$$

, où  $\sigma_o$  : est la variance de l'observation.

De même en suivant la factorisation 2 et en combinant les Eq. 14 et Eq. 15, la justesse s'exprime par :

$$JUST = \text{BiaisII} - \text{DIS} + \sigma_f^2 \quad \text{Eq. 19}$$

, où  $\sigma_f$  : est la variance de la prévision.

Les Eq. 18 et Eq. 19 montrent alors la composition de la justesse par les autres attributs de la qualité des prévisions. Nous remarquons que les deux variances,  $\sigma_o$  et  $\sigma_f$ , sont des propriétés statistiques des observations et des prévisions, respectivement. La variance des observations  $\sigma_o$  correspond à l'incertitude intrinsèque de la variable à prévoir (température, précipitation, débit, etc.). De l'autre côté, la variance des prévisions  $\sigma_f$  correspond à la dispersion des membres de la prévision d'ensemble entre eux, autrement appelée "*finesse*" (en anglais *sharpness*). Pour une prévision parfaitement fiable, la distance moyenne entre la moyenne des ensembles et l'observation est égale à la distance moyenne entre la moyenne des membres et chaque membre individuel. Dans ce cas la dispersion des ensembles est une mesure de la résolution (Jolliffe et Stephenson, 2003).

Dans le cas de l'évaluation des prévisions d'ensemble, le but est d'évaluer, d'une part, la justesse de l'ensemble (et ses attributs) et, d'autre part, la précision des membres de l'ensemble. Nous souhaitons, par exemple, évaluer si les valeurs prévues sont proches ou non des valeurs observées ainsi que si la probabilité d'occurrence de l'événement est respectée (c'est-à-dire que si un système de prévision d'ensemble donne, en moyenne, une probabilité d'occurrence de 30% pour un événement donné, cet événement est, en moyenne, observé dans 30% des cas). Graphiquement, ces propriétés sont schématisées dans la Figure 17.

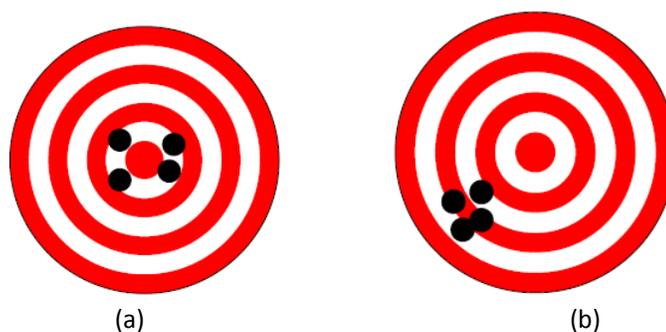


Figure 17 : Illustration de : (a) la justesse et (b) la précision d'une prévision pour attendre sa cible, l'observation.

### 3.2 Les critères d'évaluation des prévisions

Il existe deux principales démarches pour l'évaluation de la qualité des prévisions probabilistes : celles qui évaluent les propriétés intrinsèques à l'ensemble et qui, strictement parlant, sont indépendantes des observations (ex., la dispersion des membres) et celles qui utilisent les scores ou critères de performance pour évaluer des paires "prévision-observation" sur une longue série de données (Jolliffe et Stephenson, 2003 ; Wilks, 2011).

Parmi les critères les plus utilisés pour l'évaluation de la qualité de la dispersion de l'ensemble et de son rapport à l'observation, il ressort : le diagramme de rang ou diagramme de Talagrand (Talagrand *et al.*, 1997), le diagramme de PIT (Hamil, 2001 ; Gneiting *et al.*, 2007), la dispersion (ou écart-type de l'ensemble) et le diagramme de fiabilité (Wilks, 2011).

En ce qui concerne les scores probabilistes d'évaluation des paires "prévision-observation", le Score de Brier ou *BS* (Brier, 1950 ; Epstein et Murphy, 1965), le Score de Probabilité Ordonnée ou *RPS* (en anglais *Ranked Probability Score*, Epstein, 1969 ; Murphy, 1971), et sa forme continue, le Score Continu de Probabilité Ordonnée ou *CRPS* (en anglais *Continuous Ranked Probability Score*, Matheson et Wilkler, 1976 ; Hersbach., 2000), le Score de Divergence (Weijs, 2010) et les courbes ROC (en anglais *Receiver operating characteristic*, Mason, 1982 ; Ahrens et Jaun, 2007) sont parmi les scores les plus largement utilisés pour les

prévisions probabilistes (Murphy et Epstein, 1967 ; Epstein, 1969 ; Mason, 1982 ; Parker et Fordham, 1996 ; Wood *et al.*, 2002 ; Jolliffe et Stephenson, 2003 ; Laio et Tamea, 2007 ; Roulin, 2007 ; Zappa *et al.*, 2008 ; Casati *et al.*, 2008 ; Cloke et Pappenberger, 2009a ; Demargne *et al.*, 2009 ; Jaun et Ahrens, 2009 ; Brown *et al.*, 2010a ; Wilks, 2011 ; etc.).

De plus, des scores traditionnellement utilisés en prévision déterministe (tableau de contingence, Taux de Bonne Détection *POD*, Taux de Fausses Alertes *FAR*, *RMSE*, etc.) peuvent également être utilisés dans l'évaluation des prévisions probabilistes (ou d'ensemble). Dans ce cas, c'est soit la moyenne des membres (ou moyenne d'ensemble) qui est souvent considérée, soit nous évaluons le tableau de contingence pour différentes valeurs de probabilité (pourcentage de membres de l'ensemble qui prévoient l'événement en question).

Souvent les scores probabilistes sont exprimés par rapport à une référence, donnant lieu aux scores de performance (en anglais, *Skill Scores*). Ils indiquent la qualité d'une prévision par rapport à une référence, qui est définie par l'utilisateur. Il s'agit souvent d'une prévision naïve que l'utilisateur aurait utilisé s'il ne disposait pas d'un système hydrométéorologique de prévision. Les références les plus courantes sont : la climatologie, qui peut correspondre à la valeur moyenne des observations historiques ou à un ensemble de valeurs tirées de la distribution climatologique, la climatologie conditionnelle, qui recherche dans le passé des périodes analogues à celle qui a été observée à l'instant de la prévision (Wilson, 1999), et la persistance, qui correspond au résultat d'un modèle naïf pour lequel la quantité prévue est égale à la dernière valeur observée.

Dans cette pléthore de critères d'évaluation, il est souvent difficile de choisir le score le plus adapté à chaque cas. Une façon de réduire les options possibles est d'examiner si un score d'évaluation est propre ou pas. Un score est strictement propre si, et seulement si, la prévision  $F$  correspond au meilleur jugement  $Q$  du prévisionniste (ex., Score de Brier, Score de Divergence, etc.) et, par conséquent, elle ne peut pas être améliorée en fournissant une prévision différente. Ce fait est en contraste avec les cas où le prévisionniste sait à l'avance que le score sera amélioré si la prévision diffère de son meilleur jugement. Mathématiquement, la définition d'un score propre est exprimée par :

$$S(Q, Q) \geq S(F, Q) \text{ quand } Q \geq F \quad \text{Eq. 20}$$

, où  $S(F, O)$  : est la récomparance du prévisionniste pour sa prévision  $F$  quand l'observation  $O$  se réalise.

Puisque chaque critère apporte des informations supplémentaires il est essentiel d'examiner plusieurs critères d'évaluation afin d'obtenir une vision complète de la qualité des prévisions selon ses différents attributs. Dans cette thèse, après avoir synthétisé les principaux scores rencontrés dans la littérature - tout en précisant la question à laquelle chaque score cherche à répondre, ainsi que les attributs qu'il identifie - nous avons sélectionné quelques mesures que nous jugeons le mieux adaptées à l'évaluation de la qualité des prévisions hydrologiques d'ensemble. Cette sélection est une combinaison de mesures graphiques et de scores numériques. L'objectif est de rassembler les informations nécessaires pour identifier la qualité des faisceaux de nos prévisions, et, par conséquent, pour évaluer l'apport des méthodes de post-traitement adoptées. Ci-après, nous présentons avec plus de détails les critères choisis.

### 3.2.1 Biais inconditionnel

La comparaison des prévisions à l'observation permet l'identification des biais éventuels (écarts à l'observation) et plus particulièrement du biais inconditionnel. Ce biais se réfère au degré de correspondance entre les valeurs moyennes (ou la moyenne des valeurs) de la prévision et de l'observation. Par exemple, les graphiques de type " nuages de points " affichant la relation entre les prévisions (en abscisses) et les observations (en ordonnées) fournissent une première estimation visuelle des biais (inconditionnels) existants. Cette première estimation peut être confirmée par des graphiques montrant les erreurs (ou résidus) de prévision en fonction des valeurs prévues. De tels graphiques montrent aussi la tendance à la sur ou sous-estimation des observations et ce pour différentes catégories (classes) de débits.

Une mesure souvent utilisée pour évaluer le biais inconditionnel des prévisions est la racine carrée de l'erreur quadratique moyenne RMSE (ou l'erreur absolue moyenne, MAE). Dans cette thèse, nous utiliserons l'erreur RMSE, calculée par la formule :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2} \quad \text{Eq. 21}$$

, où  $o_i$  : est la valeur observée à l'instant de prévision  $i$ ,

$f_i$  : est la valeur prévue à l'instant de prévision  $i$ ,

$N$  : est le nombre total d'instant de prévisions utilisées pour calculer le score.

La RMSE varie entre 0 et  $\infty$  ; score parfait = 0. Elle est exprimée en unités de la variable évaluée.

### 3.2.2 Diagramme de PIT

L'identification du biais inconditionnel est essentielle pour la correction des prévisions, mais elle doit être accompagnée d'une estimation du biais conditionnel. Les deux principaux types de biais conditionnels sont le Type I et le Type II (Bradley *et al.*, 2004 ; Brown et Seo, 2012). Ils sont liés aux factorisations 1 et 2 présentées ci-dessus. Nous rappelons que :

- le biais conditionnel du Type I est lié à la probabilité de l'observation  $o$  conditionnée à la prévision  $f$  :  $p(o|f)$ ,
- le biais du Type II est liée à la probabilité de prévision  $f$  conditionnée à l'observation  $o$  :  $p(f|o)$ .

Les mesures qui sont souvent utilisées pour l'identification du biais du Type I sont le diagramme de fiabilité, le diagramme de rang (au diagramme de Talagrand ; Talagrand, 1997) et l'histogramme de PIT (en anglais, *Probability Integral Transform* ; Diebolt *et al.*, 1998). La principale différence entre ces mesures est que le diagramme de fiabilité est construit pour un événement donné, et donc répond à un problème particulier de prise de décision (dépassement ou non d'un seuil critique), tandis que l'histogramme de PIT et le diagramme de rang évaluent la fiabilité des prévisions sur l'ensemble des observations. L'histogramme de PIT est visuellement similaire au diagramme de rang (Glahn, 2009).

Dans cette thèse, nous utiliserons l'histogramme de PIT ou sa version cumulée, le diagramme de PIT, pour évaluer la fiabilité des prévisions. L'histogramme de PIT considère la localisation de l'observation dans la fonction de répartition des prévisions (Hamill, 2001). A chaque observation, nous évaluons la probabilité de prévision associée à cette valeur observée. L'ensemble de probabilités peut ensuite être analysé selon un histogramme de fréquence (histogramme de PIT) ou une fonction de répartition (diagramme de probabilités cumulées).

La Figure 18 présente les différentes formes que l'on peut obtenir à partir de l'analyse des PIT. Un histogramme uniforme ou un diagramme de probabilité cumulée aligné à la diagonale est une indication de prévisions fiables. Comme pour le diagramme de rang, un histogramme sous forme d'arche correspond à des prévisions plutôt sur-dispersées, tandis qu'un diagramme sous forme de "U" est indicatif de prévisions sous-dispersées (Hamill, 2001). Ainsi, quand les fréquences des observations sont supérieures aux probabilités cumulées des prévisions le système de prévision a tendance à sous-estimer l'observation et vice versa. Des diagrammes asymétriques illustrent la présence de biais particuliers dans les prévisions. Le diagramme de probabilités cumulées représente l'accord entre les fréquences d'observation et la probabilité cumulée des prévisions probabilistes correspondantes. La correspondance idéale est donnée par la diagonale.

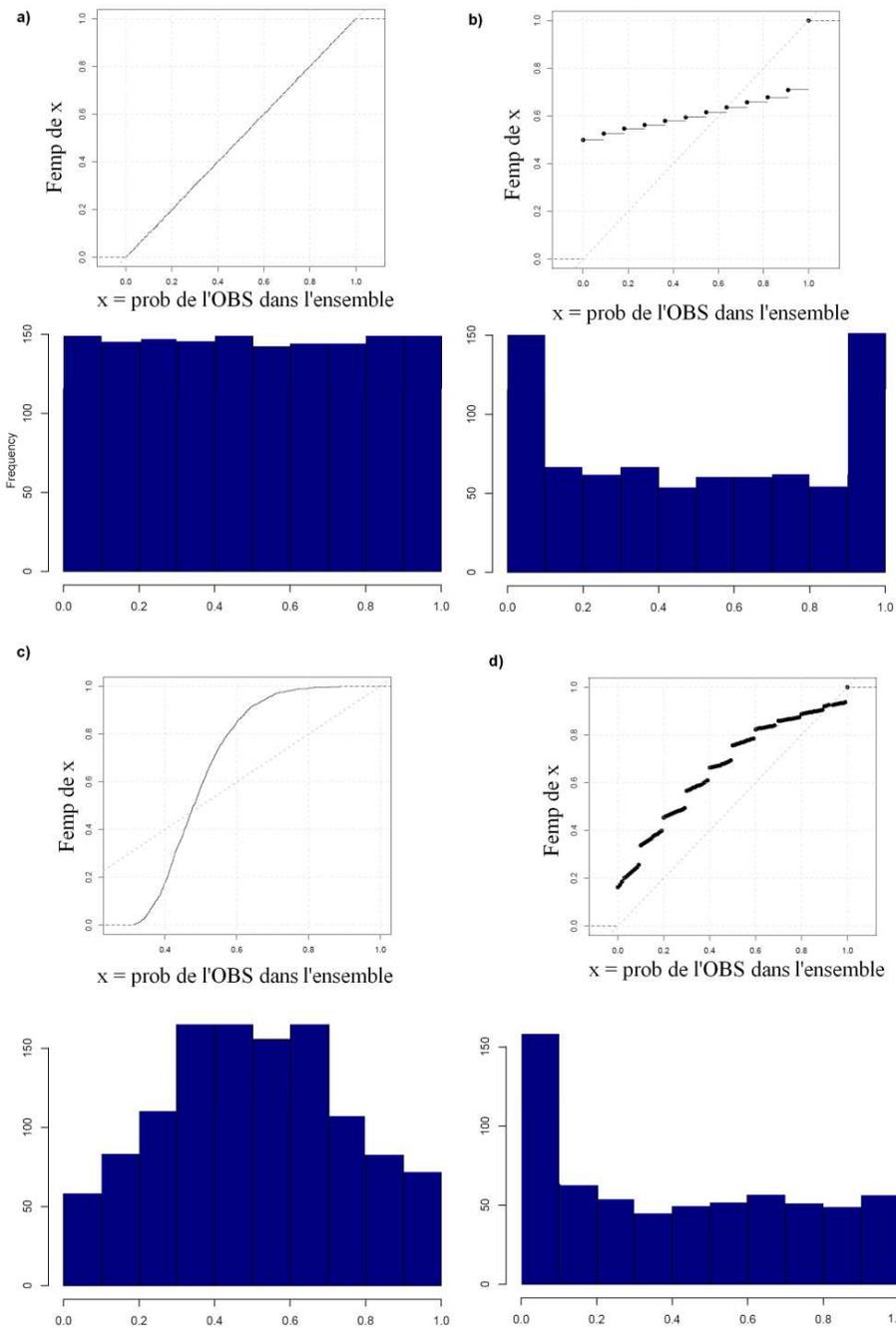


Figure 18 : Interprétation du critère " PIT ", a) prévision fiable, b) prévision sous-dispersée, c) prévision sur-dispersée, d) prévision biaisée. Diagramme de PIT en haut et histogramme de " PIT " en bas.

### 3.2.3 Dispersion des membres

Souvent la sur- ou sous- dispersion des membres est responsable des événements manqués ou des fausses alarmes. La dispersion de l'ensemble représente l'écart moyen des membres de l'ensemble par rapport à la moyenne de l'ensemble. Elle est une mesure de la précision de l'ensemble.

La sous-dispersion en particulier constitue souvent un défaut important des systèmes de prévision d'ensemble, surtout pour les prévisions aux plus courtes échéances. Dans ce cas-là, la dispersion est trop faible et l'ensemble risque de ne pas capturer suffisamment la réalité. D'un autre côté quand un ensemble est sur-dispersif l'observation est en moyenne "noyée" entre les membres de l'ensemble et ainsi la prévision fournie par l'ensemble n'est pas pertinente.

La mesure de la dispersion des membres de l'ensemble la plus utilisée est l'écart-type  $\sigma$  :

$$\sigma_i = \sqrt{\frac{1}{K} \sum_j^K (f_{ij} - \bar{f}_i)^2} \quad \text{Eq. 22}$$

, où  $K$  : le nombre de membres de l'ensemble,

$f_{ij}$  : la valeur prévue par le membre  $j$  à l'instant  $i$ ,

$\bar{f}_i$  : la moyenne des valeurs prévues par les membres à l'instant  $i$ .

Souvent la dispersion des membres de la prévision d'ensemble est confrontée à l'erreur quadratique moyenne (RMSE), calculée à partir de la moyenne des membres de l'ensemble. Pour un bassin versant et pour chaque échéance, le graphique " précision-dispersion " montre ainsi la relation entre la RMSE (en abscisses) et la dispersion (en ordonnées) pour un instant de prévision  $i$ . La diagonale représente une erreur quadratique moyenne égale à la dispersion des ensembles, ce qui correspond à des erreurs de prévision expliquées totalement par la dispersion de l'ensemble.

Il faut souligner que l'interprétation de la relation entre la dispersion et l'erreur de la prévision n'est pas toujours évidente (Mason et Stephenson, 2008). Plusieurs études ont montré une faible corrélation entre la dispersion et l'erreur de prévision (Atger, 1999 ; Hamill *et al.*, 2004). Ce fait est partiellement expliqué par la faible variabilité de la dispersion (Whitaker et Loughe, 2003) rencontrée souvent dans les prévisions d'ensemble.

### 3.2.4 Courbe ROC

La courbe ROC (Mason, 1982) consiste à tracer les taux de fausses alertes *POFD* (Eq. 26) contre les probabilités de bonne détection *POD* (Eq. 27), les deux mesures issues de l'analyse des tableaux de contingence (Figure 19) pour un seuil critique donné qui définit un événement observé et pour différents seuils de probabilités de prévision de l'événement.

		Événement observé		
		Ouï	Non	
Événement prévu	Ouï	Succès (a)	Fausse Alerte (b)	Prévu
	Non	Manqué (c)	Rejet Correct (d)	Non prévu
		Observé	Non Observé	Total

Figure 19 : Tableau de contingence pour une prévision.

En pratique, pour les prévisions d'ensemble, nous calculons la probabilité de dépasser le seuil critique par au moins 1, 2, 3... ou  $K$  membres de la prévision d'ensemble. Pour chaque probabilité de dépassement du seuil, on calcule un tableau de contingence pour déduire les scores *POFD* et *POD*. Par conséquent, si on a  $K$  membres de prévision, l'ensemble des  $K$  couples (*POFD*, *POD*) constitue la courbe ROC (Atger, 2001). Plusieurs systèmes (d'ensemble ou déterministe) peuvent être graphiquement comparés par rapport à leur position dans le graphique de la courbe ROC. Une prévision déterministe figurera comme un point dans le graphique, tandis que pour les prévisions d'ensemble une courbe peut être construite en fonction d'un ensemble croissant de probabilités de prévision (Figure 20).

Dans le graphique ROC, une prévision parfaite donnerait un taux de fausses alertes nul (*POFD*=0), alors que le taux de bonnes prévisions serait de 100% (*POD*=1). Ainsi, plus le point ou la courbe ROC sera proche du couple (0,1), c'est-à-dire, proche du coin supérieur gauche du graphique, meilleures seront les prévisions. Une courbe qui approche la diagonale montre une capacité nulle du système de prévision à discriminer entre bonne détection et fausses alertes de l'événement en question.

$$POD = \frac{a}{a + c} \quad \text{Eq. 23}$$

$$POFD = \frac{b}{b + d} \quad \text{Eq. 24}$$

L'aire sous la courbe ROC est aussi fréquemment utilisée pour comparer différents systèmes de prévision. Une courbe ROC confondue avec la 1ère bissectrice indique que la prévision n'a pas de performance et son aire est ainsi égale à 0.5. Dans ce cas, les valeurs de POFD et de POD sont confondues ; le système de prévision serait fait au hasard puisqu'il présente des probabilités de détection et de fausses alertes égales (le pourcentage d'événements prévus quand ils sont observés est le même des événements prévus quand ils ne sont pas observés).

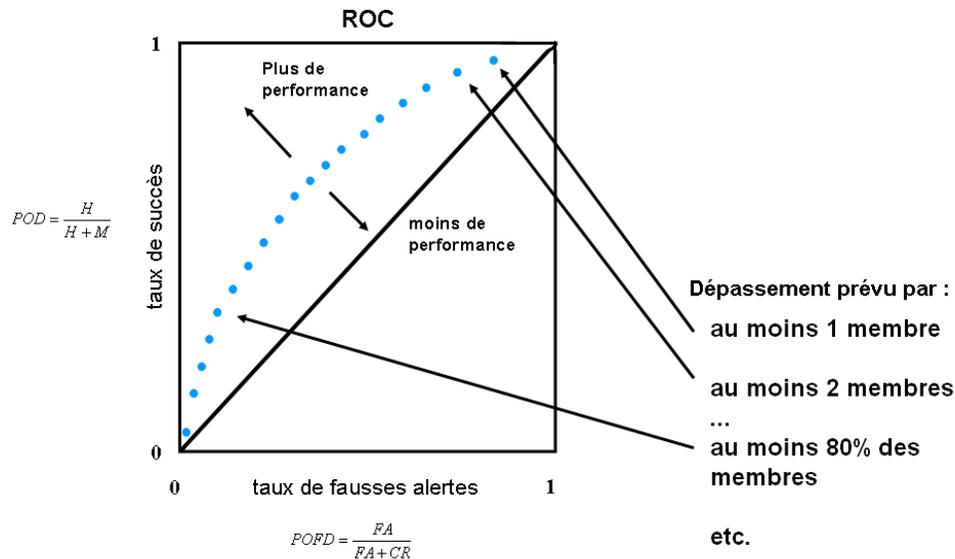


Figure 20 : Schéma représentant la construction d'une courbe ROC ( $H=a$ ,  $M=c$ ,  $FA=b$ ,  $CR=d$  du Tableau de Contingence).

Les courbes ROC permettent d'évaluer la capacité des prévisions à discriminer entre les événements (discrimination). Cette mesure n'évalue pas la fiabilité des prévisions : par exemple, une prévision qui n'est pas fiable peut donner une bonne courbe ROC, tandis qu'une prévision basée sur la climatologie, donc fiable, peut donner une courbe ROC proche de la diagonale. Pour cette raison, la courbe ROC doit être utilisée en combinaison avec d'autres mesures de fiabilité.

### 3.2.5 Score de Brier

Si on s'intéresse à une catégorie spécifique (ex., la prévision du dépassement d'un seuil d'alerte donné), un autre score adapté aux prévisions probabilistes est le Score de Brier, qui a été introduit par Brier en 1950. Ce score est simplement l'erreur quadratique moyenne des probabilités de prévision et d'observation de l'événement en question. Il y a deux approches principales de présentation du Score de Brier : en considérant la prévision probabiliste

directement comme une quantité scalaire (une probabilité) ou comme un vecteur (un ensemble de valeurs de prévision). Selon Murphy *et al.* (1972a, 1972b, 1973), du point de vue scientifique et économique, la définition sous la forme de vecteur est avantageuse. Leur argumentation est basée sur le fait que si on s'intéresse principalement à la prévision, étant donné que la prévision est un vecteur, il est préférable de la traiter sous la forme de vecteur. Par contre, si on s'intéresse principalement aux probabilités fournies par la prévision, étant donné que la probabilité est une quantité scalaire, il est plus simple de la traiter sous la forme scalaire. De ce fait, le Score de Brier s'exprime souvent sous sa forme scalaire.

Pour un événement donné, le Score de Brier est calculé pour deux catégories (occurrence ou non de l'événement) sur un échantillon de  $N$  paires prévision-observation par :

$$BS = \frac{1}{N} \sum_{i=1}^N (f_i - o_i)^2 \quad \text{Eq. 25}$$

, où  $f_i$  : est la probabilité associée à la paire prévision-observation à l'instant de prévision  $i$ ,

$o_i$  : = 1, si l'événement a été observé à l'instant de prévision  $i$ .  
= 0, sinon.

Le Score de Brier est donc estimé en évaluant la somme des différences au carré entre les probabilités de prévision (pourcentage des membres de l'ensemble qui dépassent le seuil critique) et l'occurrence de l'observation. Pour améliorer ce critère, il faut alors augmenter la probabilité de prévision quand l'événement se produit et la diminuer quand l'événement ne se produit pas.

Le Score de Brier ( $BS$ ), comme les erreurs du type erreur moyenne absolue ( $MAE$ ) ou erreur quadratique moyenne ( $RMSE$ ), est négativement orienté : plus petite est sa valeur, meilleure est la performance du système évalué. Le score optimal ( $BS = 0$ ) est atteint pour une prévision déterministe " parfaite ", c'est-à-dire, toujours correcte.

### 3.2.6 Score RPS

Le score de probabilité ordonnée  $RPS$  a été introduit par Epstein (1969), suivi par une modification majeure proposée par Murphy (1971). Le principe de ce score consiste à comparer la fonction de répartition de la prévision (la distribution prédictive donnée par l'ensemble) à la fonction de répartition de l'observation (qui, en fait, est une fonction cumulée d'une densité de probabilité qui vaut 1 au point de l'observation et 0 ailleurs). Le score RPS est équivalent à la somme des scores de Brier, évalués sur différentes catégories  $m$ .

Pour un échantillon de  $N$  paires prévision-observation et  $M$  états mutuellement exclusifs et exhaustifs (composés des sous-catégories, définies, par exemple, par les percentiles 10% à 90% des observations), le score est donné par :

$$RPS = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{M} \sum_{m=1}^M (f_{i,m} - o_{i,m})^2 \right) \quad \text{Eq. 26}$$

, où  $f_{i,m}$  : la probabilité prédictive pour la paire prévision-observation à l'instant de prévision  $i$  et, pour la sous-catégorie  $m$ ,

$o_{i,m} = 1$ , si l'observation appartient à la sous-catégorie  $m$  à l'instant de prévision  $i$ .

$= 0$ , si l'observation n'appartient pas à la sous-catégorie  $m$ .

Le score parfait est égal à 0 et le score 1 est attribué à la plus mauvaise performance du système. Le  $RPS$  est sensible à la distance entre la probabilité prédictive et la fréquence d'observation.

L'avantage du score  $RPS$  est sa formulation adaptée au cas des prévisions exprimées en catégories ordonnées : par exemple, le cas où nous nous intéressons à la capacité d'un système à prévoir : 1) les débits faibles (inférieurs au quantile 20%), 2) les débits moyens (entre les quantiles 20% et 80%) et 3) les débits forts (supérieurs au quantile 80%). Néanmoins, un inconvénient majeur de ce score est le fait que sa valeur peut varier fortement selon le découpage en catégories (le choix des catégories et de la largeur de chaque catégorie).

### 3.2.7 Score CRPS

Si on considère un nombre de catégories infinies avec une largeur négligeable, le  $RPS$ , un score qui s'exprime sous la forme discrète, tend vers la forme continue donnée par le  $CRPS$  (Matheson et Winkler, 1976 ; Hersbach, 2000). De plus, le  $CRPS$  a un caractère dimensionné : il s'exprime dans les mêmes unités de la variable évaluée (ici,  $m^3/s$  pour les débits et mm pour les pluies).

Le  $CRPS$  a été interprété comme la version probabiliste de l'erreur absolue moyenne  $MAE$  (Gneiting et Raftery, 2007) et utilisé pour faire des comparaisons entre la performance d'un système de prévision d'ensemble (exprimée par le  $CRPS$ ) et celle d'un système déterministe (exprimée par la  $MAE$ ), pour une même variable évaluée (ex., Velázquez *et al.*, 2011). Enfin, nous notons que le  $CRPS$  peut être aussi compris comme l'intégrale du Score de Brier sur toutes les sous-catégories et il est exprimé en mm.

$$\overline{CRPS} = \int_{-\infty}^{+\infty} BS(x) dx \quad \text{Eq. 27}$$

$$CRPS = \frac{1}{N} \sum_{i=1}^N \int_{x=-\infty}^{x=+\infty} (F_i^f(x) - F_i^o(x))^2 dx \quad \text{Eq. 28}$$

, où  $F_i^f(x)$  : est la fonction de répartition des prévisions à l'instant de prévision  $i$ ,

$F_i^o(x)$  : est l'observation, exprimée comme une fonction de répartition, à l'instant de prévision  $i$ ,

$N$  : est le nombre total d'instants de prévision utilisé pour calculer le score.

Si l'observation est une valeur unique, alors sa fonction de répartition est une fonction du type "marche d'escalier" ou indicatrice (fonction Heaviside), prenant la valeur 0 pour tous les réels inférieurs à la valeur observée et la valeur 1 partout ailleurs :

Afin de calculer numériquement le  $CRPS$ , Gneiting et Raftery (2007) proposent l'adoption d'une loi de distribution gaussienne pour l'estimation de la fonction de probabilité. Une telle hypothèse peut être satisfaite pour les prévisions de température ou de pression atmosphérique, par contre ceci n'est pas le cas des séries hydrologiques. Dans ce cas, le calcul du  $CRPS$  est souvent fait par une approximation discrète de l'intégrale de l'Eq. 28 (Hersbach, 2000). Dans cette thèse nous avons appliqué une loi empirique pour l'estimation de la densité de probabilité cumulée et une discrétisation suffisamment importante de cette distribution prédictive.

### 3.2.8 Score de Divergence

Sous un principe similaire à celui du Score de Brier et du Score RPS, est fondé le Score de Divergence (en anglais "*Kullback-Leibler Divergence Score*"), qui quantifie le gain en information supplémentaire obtenu par la prévision d'ensemble par rapport à l'observation (Weijs et al., 2010). Ce score  $DS$  calcule la divergence entre la loi de fréquence de l'observation  $o_i$  et la distribution prédictive  $f_i$  à l'instant de prévision  $i$  (Eq. 29 et Eq. 30). Pour un événement binaire (ex., dépassement ou non d'un seuil critique) et  $N$  paires "prévision-observation", le score de divergence est défini par les équations suivantes :

$$DS(o_i|f_i) = \sum_{i=1}^N [o_i] \log \left( \frac{[o_i]}{[f_i]} \right) \quad \text{Eq. 29}$$

$$DS = \frac{1}{N} \sum_{i=1}^N DS(o_i|f_i) \quad \text{Eq. 30}$$

, où  $f_i$  : est la probabilité associée à la paire prévision-observation à l'instant de prévision  $i$ ,

$o_i$  : = 1, si l'événement a été observé à l'instant de prévision  $i$ .

= 0, sinon.

Ce score peut être interprété comme le gain d'information en passant de la distribution prédictive à la fréquence de l'observation. Si la prévision contient déjà toute l'information existante dans l'observation, le gain d'information est 0 et le score est parfait. Pour le cas des événements multi-catégoriques ( $M$  étant le nombre de catégories), le Score de Divergence Ordonnée RDS (Weijis *et al.*, 2010) est défini par :

$$RDS = \frac{1}{M} \sum_{m=1}^M \left( \frac{1}{N} \sum_{i=1}^N DS(o_{i,m}|f_{i,m}) \right) \quad \text{Eq. 31}$$

, où  $f_{i,m}$  : la probabilité prédictive pour la paire prévision-observation à l'instant de prévision  $i$  et, pour la sous-catégorie  $m$ ,

$o_{i,m}$  : = 1, si l'observation appartient à la sous-catégorie  $m$  à l'instant de prévision  $i$ .

= 0, si l'observation n'appartient pas à la sous-catégorie  $m$ .

### 3.2.9 Remarques générales sur l'évaluation de la qualité des prévisions

Un des problèmes qui émerge pendant l'évaluation d'un système de prévision est lié à la taille de l'échantillon. Souvent les scores statistiques utilisés demandent une série de données suffisamment longue pour la vérification des prévisions. Le recueil d'un grand nombre de paires " prévision-observations " n'est cependant pas toujours faisable pour plusieurs raisons : le coût important des simulations des systèmes de prévision, l'indisponibilité ou bien l'hétérogénéité des données, les erreurs et lacunes dans les séries des données observées, etc.

Candille (2003) a calculé le nombre théorique de réalisations nécessaires pour la vérification des prévisions d'ensemble, en prenant en compte le nombre des membres de l'ensemble. L'auteur a indiqué que, afin de réaliser une évaluation fiable des prévisions pour un ensemble de 10 membres, il faudrait plus de 5000 prévisions indépendantes. En pratique cette taille de séries n'est pas disponible et l'évaluation des prévisions s'effectue sur des échantillons d'une longueur de quelques années.

En particulier, dans le cas des événements rares, quand la prévision et sa vérification deviennent cruciales, les échantillons de données sont encore plus courts à cause du nombre limité d'occurrences de ces phénomènes. Dans ce cas, l'incertitude liée à l'échantillonnage devient significativement plus importante. Il est néanmoins difficile d'associer une valeur d'incertitude à un score moyen calculé, comme le score de Brier *BS* et le *CRPS*, sur un échantillon. Dans la littérature, des ré-échantillonnages du type Monte Carlo sont proposés (Brown et Seo, 2010), mais la procédure reste coûteuse en temps de calcul et dépendante de la longueur des séries disponibles.

Un deuxième aspect important affectant l'évaluation de la qualité des prévisions d'ensemble est lié au nombre de membres de l'ensemble. La performance de plusieurs mesures de vérification est fortement liée au nombre de membres de l'ensemble. Comme indiqué par Muller *et al.* (2004) et Weigel *et al.* (2006), pour un nombre de membres d'ensemble limité (inférieur à 40 membres), le Score de Brier (*BS*), ainsi que le Score de Probabilité Ordonnée (*RPS*) sont négativement biaisés. Dans ce cas, une correction est proposée dans la littérature (Stephenson *et al.*, 2007).

Enfin, remarquons que si parfois les prévisions sont bonnes et capturent bien les événements à prévoir, nous sommes souvent en présence d'un biais dans le système de prévision qu'il faut corriger. Cette procédure est connue sous le terme *traitement statistique* de correction des sorties d'un modèle, nécessaire afin de débiaiser les prévisions et les rendre bonnes. Les principales familles de méthodes de traitements statistiques sont présentées ci-après, et les méthodes qui seront utilisées dans cette thèse y sont décrites.

### 3.3 Approches de traitement statistique des prévisions

Pour définir la notion de " traitement statistique des prévisions ", il faut d'abord revenir aux deux factorisations de la probabilité conjointe des prévisions-observations (Eq. 9 et Eq. 10). Rappelons que la factorisation 2 fait appel à la fonction de vraisemblance et cherche à répondre à la question "quelle prévision ressemble le plus à une observation donnée ?".

---

C'est à cette question que nous nous adressons lors de la procédure de traitement statistique des prévisions. Nous cherchons à identifier la distribution prédictive ayant engendré une observation donnée. Sous l'hypothèse que, à n'importe quel instant  $i$  la fréquence d'une observation vient d'une fonction de répartition  $F_i^o$  et, en même temps, la prévision probabiliste vient d'une fonction de répartition  $F_i^f$ , la prévision idéale vérifierait :

$$F_i^f = F_i^o \quad \text{Eq. 32}$$

Néanmoins, la fonction de répartition des observations n'est, au mieux, qu'une hypothèse, tandis que la fonction de répartition prédictive des prévisions reste souvent une estimation d'une loi de probabilité, basée sur un nombre limité de membres d'un ensemble.

L'objectif du traitement statistique des prévisions est de minimiser la différence entre la fonction de répartition des observations et la fonction de répartition prédictive des prévisions. La démarche souvent entreprise est celle qui fait appel aux techniques de correction statistique avant (pré-) ou après (post-) le modèle hydrologique, avec pour but d'améliorer la qualité de prévisions en termes de volume, temps et distribution.

De manière générale, il s'agit d'examiner les erreurs qui se réfèrent au degré de correspondance entre l'observation et la prévision, ainsi que les biais provenant d'une mauvaise dispersion des membres. Pour les prévisions météorologiques, la dispersion des membres est surtout liée à la génération des ensembles et ainsi aux perturbations des conditions initiales considérées. Si la variation initiale introduite par les perturbations n'est pas suffisante, cela conduit à des membres de l'ensemble peu dispersés entre eux et à des prévisions manifestant un excès de confiance (Buizza *et al.*, 2005 ; Wang et Bishop, 2005 ; Zhu, 2005). En même temps, une dispersion insuffisante des membres peut aussi être liée à une incapacité du modèle numérique à représenter proprement l'incertitude du climat et des processus physiques.

Pour ce qui est des prévisions hydrologiques, la dispersion des membres peut être perturbée par la procédure de mise à jour qui, selon l'approche adoptée - même si elle n'introduit pas un biais, elle peut "écraser" l'ensemble. Les approches statistiques de post-traitement visent particulièrement à s'attaquer à ces problèmes, de façon à corriger les biais de la prévision et à mieux ajuster la dispersion des membres de l'ensemble de façon à les rendre plus fiables.

---

### 3.3.1 Familles de méthodes de traitement statistique

Un nombre considérable d'approches statistiques pour la correction des prévisions est présenté dans la littérature. En faisant une généralisation, nous pouvons séparer ces approches en deux familles principales :

- les approches qui visent à corriger l'erreur totale de la prévision,
- les approches qui traitent les sources d'erreur séparément.

Parmi les approches qui traitent les sources d'erreur séparément se trouvent : les approches de mise à jour du modèle hydrologique par une technique d'assimilation de données (Seo *et al.*, 2003 ; Goswani *et al.*, 2005 ; Andreatis et Lettenmaier, 2006 ; Liu *et al.*, 2012), les approches d'équiprobabilité qui utilisent des jeux de paramètres du modèle (Kuczera et Parent, 1998 ; Vrugt *et al.*, 2003 ; Beven et Binley, 1992 ; Pappenberger *et al.*, 2005), l'habillage des simulations par la prise en compte exclusivement des erreurs passées du modèle hydrologique (Bogner et Kalas, 2008 ; Zhao *et al.*, 2011 ; Pagano *et al.*, 2012).

Néanmoins, Krzysztofowicz (2002) a souligné l'importance de prendre en compte l'incertitude totale lors du traitement statistique des prévisions. Dans cette direction sont dirigées les approches proposées par Seo *et al.* (2006) et Regianni *et al.* (2009), qui ont utilisé l'inférence bayésienne afin de calculer la probabilité d'occurrence d'un événement basé sur l'expérience du passé. Raftery *et al.* (2005) ont également proposé une méthodologie – dite *Bayesian Model Averaging* (BMA) – qui traite l'incertitude totale en utilisant le cadre bayésien afin de former un ensemble de prévisions de bonne qualité à partir de prévisions de différentes sources. Cette approche est également appliquée par Slougher *et al.* (2007), Duan *et al.* (2007), Vrugt *et al.* (2008), Bao *et al.* (2010) sur des prévisions hydrométéorologiques d'ensemble. L'hypothèse principale derrière la BMA est que la distribution de probabilité conditionnelle des prévisions sachant les observations suit la loi normale ou gamma. Sur ce point Rings *et al.* (2012) ont récemment proposé une modification à la BMA afin de rendre la représentation de la distribution de probabilité conditionnelle plus flexible en utilisant un filtre particulière dans un contexte gaussien. Une autre sous-catégorie de méthodes qui visent le traitement statistique de l'incertitude totale des prévisions est la famille de méthodes de régression linéaires et non-linéaires (Gneiting *et al.*, 2005 ; Hamill *et al.*, 2008 ; Wilks, 2009 ; Weerts *et al.*, 2011).

Dans la plupart des méthodes, quelle que soit l'approche adoptée, la loi de probabilité utilisée pour décrire les erreurs de prévision n'est pas toujours connue ou facile à estimer. Les transformations " normalisatrices " des données sont souvent appliquées pour ajuster une loi normale, plus facile à manipuler, aux erreurs de prévisions. Des techniques alternatives (paramétrées et non-paramétrées) sont proposées à partir : d'une estimation bayésienne

linéaire optimale des variables indicatrices (Brown et Seo, 2010), des fonctions copules (Madadgar *et al.*, 2012), des transformations dite -" *wavelet* " - (Bogner et Pappenberger, 2011), ou de " méthodes de noyaux " (Roulston et Smith, 2003 ; Wang et Bishop, 2005 ; Fortin *et al.*, 2006 ; Boucher *et al.*, 2012).

Naturellement, l'objectif des deux approches principales est de rendre les prévisions plus précises et fiables en utilisant l'information contenue dans les observations. Pour chaque approche, il y a plusieurs groupes de méthodes de traitement statistique qui ont été proposés dans la littérature. Pour les méthodes déjà testées en hydrologie, celles-ci ont été appliquées soit sur un petit échantillonnage de bassins versants (Seo *et al.*, 2006 ; Hashino *et al.*, 2006 ; Olsson et Lindström, 2008 ; Reggiani *et al.*, 2009 ; Fortin *et al.*, 2006 ; Brown and Seo, 2010, 2012 ; Weerts *et al.*, 2011 ; Boucher *et al.*, 2012), soit pour évaluer la capacité d'un système de prévision lors d'événements particuliers (Pappenberger *et al.*, 2008).

Parmi les techniques existantes, nous avons choisi d'appliquer des approches empiriques dites " méthodes d'habillage " ou " méthodes de noyaux ", qui, de manière générale, font appel à la loi de distribution des erreurs de prévision observées dans le passé pour corriger l'incertitude prédictive (Smith *et al.*, 1992 ; Roulston et Smith, 2003 ; Wang et Bishop, 2005 ; Fortin *et al.*, 2006 ; Hashino *et al.*, 2006 ; Olsson and Lindström, 2008 ; Kang *et al.*, 2010 ; Weerts *et al.*, 2011 ; Boucher *et al.*, 2012 ; Pagano *et al.*, 2012).

Ces approches peuvent corriger soit chaque membre de l'ensemble séparément, soit la distribution prédictive de l'ensemble. La stratégie de ces méthodologies se résume néanmoins au même principe : la correction des prévisions en utilisant une série historique des erreurs observées dans une période passée (séries historiques de prévisions archivées ou de prévisions rétrospectives). Leur avantage principal se résume à la simplicité de leur application. Même si les approches bayésiennes proposent un cadre théorique solide pour l'estimation de l'incertitude prédictive, néanmoins leur implémentation n'est pas toujours facile.

Dans notre étude, nous avons également appliqué une technique d'habillage empirique traitant spécifiquement les erreurs du modèle hydrologique. Enfin, des méthodes d'habillage spécifiquement destinées à traiter les prévisions de précipitation qui seront utilisées en entrée des modèles hydrologiques ont également été examinées. L'ensemble des méthodes de traitement statistique que nous avons étudiées est présentée en détail ci-après.

---

### 3.3.2 Méthodes appliquées aux prévisions de précipitation

#### 3.3.2.1 Méthode empirique par facteur de correction

Les méthodes empiriques par facteur de correction sont des méthodes simples, souvent utilisées intuitivement par les services opérationnels. Ces approches de débiaisage prennent la forme d'une correction additive ou multiplicative, appliquée à chaque prévision. Le facteur de correction est en général estimé à partir de la moyenne de l'erreur calculée sur une série de paires "prévisions-observations" disponible. La correction multiplicative est la plus courante, car elle permet d'éviter la génération de valeurs négatives. Mathématiquement, elle est exprimée par :

$$b = E_f \left[ \frac{o}{f} \right] \quad \text{Eq. 33}$$

, où  $E []$  : correspond à l'espérance mathématique,  
 $b$  : est le facteur de correction multiplicatif,  
 $o$  : est la valeur observée,  
 $f$  : est la valeur prévue.

Le facteur de correction multiplicatif peut être évalué de manière générale ou selon des catégories prédéfinies (saisons, types de temps, classes de valeurs observées, etc.). Lors de la prévision, il sera appliqué à chaque nouvelle prévision :

$$\hat{f} = b * f \quad \text{Eq. 34}$$

, où  $f$  : est la prévision brute (non-corrigée),  
 $\hat{f}$  : est la prévision corrigée.

Un exemple d'une simple correction multiplicative, proposée à l'origine par Smith *et al.* (1992), est présenté par Hashino *et al.* (2006) dans un cadre d'application à des prévisions mensuelles de volumes écoulés dans le bassin Des Moines à Iowa. Dans leur étude, l'hypothèse principale est que toutes les prévisions ont le même biais multiplicatif à chaque fois que la même séquence climatique est utilisée pour générer les ensembles. En conséquence, le facteur de correction appliqué à la prévision brute dépend seulement de la séquence climatique (c'est-à-dire, du mois et de l'année en question).

Hashino *et al.* (2006) proposent également une approche équivalente. Il s'agit de corriger (de façon multiplicative ou additive) soit les moments de la distribution prédictive, et plus particulièrement la moyenne et l'écart-type (moments de 1<sup>er</sup> et 2<sup>ème</sup> ordre, respectivement), soit les quantiles de la distribution prédictive. Cette méthode consiste à remplacer la variable prévue par une variable observée avec la même probabilité au non-dépassement.

D'autres auteurs comme Olsson et Lindström (2008) proposent une méthodologie semblable qui consiste à ajuster les quantiles 25% et 75% de la distribution prédictive de l'ensemble, ainsi que les valeurs minimales et maximales des prévisions de l'ensemble, en les multipliant par des coefficients correctifs calculés sur une période de calage pour chaque bassin versant pris individuellement.

Dans cette thèse, nous avons examiné la méthode de correction par facteur multiplicatif pour corriger des biais des précipitations prévues (voir Chapitre 6, §6.3). Plus particulièrement, nous appliquons une correction multiplicative sur chaque membre de la prévision d'ensemble calculée sur une période de calage.

L'équation Eq. 35 décrit mathématiquement l'erreur de la moyenne de la prévision d'ensemble  $\varepsilon$  par bassin versant  $ibv$  en fonction du jour  $i$  et du membre  $k$  pour la première échéance de prévision.

$$\varepsilon_{ibv,i,1} = \left( \frac{1}{K} \sum_{k=1}^K f_{ibv,i,1,k} \right) - o_{ibv,i,1} \quad \text{Eq. 35}$$

Une fois les erreurs de la moyenne de la prévision d'ensemble calculées par jour  $i$ , nous constituons une archive de ces erreurs pour chaque bassin versant. Le facteur correctif s'exprime en pourcentage d'augmentation ou de réduction de la valeur prévue. Le calage de ces facteurs correctifs se fait année par année sur l'ensemble de la période 2005-2008 et leurs valeurs moyennes sont retenues à la fin. Pour l'instant le même facteur correctif est appliqué pour toutes les échéances  $j$ .

Les principaux avantages de la méthode de correction par facteur multiplicatif sont :

- sa simplicité,
- l'élimination du biais inconditionnel,
- la bonne performance par rapport au biais conditionnel et l'amélioration de la dispersion des membres quand la méthode est appliquée aux moments ou aux quantiles de la distribution prédictive.

La principale limitation de la méthode est :

- la dépendance à la longueur de la période de calage, notamment quand les facteurs de correction doivent être évalués sur des sous-échantillons de l'échantillon total (manque de robustesse des estimations si la longueur n'est pas suffisante).

### 3.3.2.2 Méthode d'habillage à l'aide des situations analogues

Dans cette thèse, nous avons pris en main l'archive disponible à EDF pour la prévision par analogie. L'information extraite par les archives des analogues est utilisée afin de produire des prévisions de précipitation. Les variables météorologiques, prises dans la recherche de situations analogues, sont les champs d'altitude des niveaux de pression 700hPa et 1000hPa à 0h00 et 24h00, qui correspondent en moyenne à une altitude de 3000m et 100m, respectivement. Afin de définir la ressemblance entre la situation prévue et les situations observées dans le passé, un critère de distance est utilisé (Score de Teweles-Wobus, 1954) et les 50 situations analogues qui ont les meilleurs scores sont retenues.

La méthode par analogie fonctionne de la même façon dans le cadre des prévisions déterministes et probabilistes. La seule différence repose sur le nombre de scénarios initiaux à la base desquels se fait la recherche des situations analogues. Dans cette thèse, nous avons suivi la procédure utilisée à EDF, qui prévoit la sélection de 50 situations analogues par scénario initial de prévision. La méthode a été appliquée à la prévision d'ensemble du CEPMMT. Chaque scénario initial est formé par un des 50 membres de la prévision d'ensemble. L'application de la méthode des analogues a ainsi généré un total de 2500 membres ou scénarios de prévision (50 membres x 50 situations analogues) par jour de prévision.

Derrière chaque scénario nous avons ainsi une paire composée d'une situation analogue et du membre correspondant de la prévision d'ensemble qui a déclenché la recherche de la situation analogue. Une correction est ensuite appliquée à chaque membre de la prévision d'ensemble, et ce pour chaque échéance de prévision. Elle se traduit par :

---

$$\hat{f} = \beta * f^{ECMWF} (f^{ANALOG} / \bar{f}^{ANALOG})^\alpha \quad \text{Eq. 36}$$

, où  $\hat{f}$  : est la prévision corrigée du membre de la prévision d'ensemble,  
 $f^{ECMWF}$  : est la valeur de la prévision brute (avant correction) du membre de la prévision d'ensemble,  
 $f^{ANALOG}$  : est la valeur de la variable de prévision donnée par la situation analogue,  
 $\bar{f}^{ANALOG}$  : est la moyenne des 50 valeurs de la variable de prévision, données par les 50 situations analogues associées au membre de la prévision d'ensemble qui est corrigé,  
 $\alpha, \beta$  : sont des paramètres libres de la correction (paramètres à caler).

Les valeurs de paramètres  $\alpha$  et  $\beta$  peuvent varier selon le bassin versant et l'échéance de prévision. Dans cette thèse, nous avons utilisé les valeurs qui étaient utilisées par le système de prévision opérationnelle à EDF. Il s'agit de valeurs fixes et égales à 0.3 et 1.2, pour les paramètres  $\alpha$  et  $\beta$ , respectivement.

Etant donné le grand nombre de scénarios (2500) générés par la méthode, ce qui correspond à un temps de calcul proportionnellement important, un ré-échantillonnage est souvent appliqué dans les procédures opérationnelles à EDF. La technique de ré-échantillonnage consiste à trier les 2500 scénarios obtenus en ordre croissant et, parmi eux, choisir 50 scénarios correspondant à des quantiles équidistants dans la distribution empirique obtenue. Ces 50 scénarios de prévision de précipitation sont ensuite utilisés en entrée du modèle hydrologique.

Les principaux avantages de l'habillage des prévisions avec l'information de situations analogues sont :

- sa simplicité d'application, notamment quand la méthode est déjà mise en place,
- la bonne performance par rapport à la correction des biais conditionnels et non-conditionnels,
- la correction de la dispersion des membres.

Les principales limitations se résument aux points suivants :

- les scénarios originaux de la prévision d'ensemble (les membres de la prévision d'ensemble) ne sont pas forcément tous gardés par la méthode,
- le pré-requis d'avoir une archive la plus longue possible disponible pour la recherche des situations analogues.

### 3.3.3 Méthodes appliquées aux prévisions de débits

#### 3.3.3.1 Méthode du meilleur membre et ses variantes

Une approche typique de la famille des techniques d'habillage est la méthode du meilleur membre. Développée initialement par Roulston et Smith (2003), elle a subi des modifications proposées par Wang et Bishop (2005) et par Fortin *et al.* (2006). Cette méthode propose d'assigner à chaque membre de l'ensemble une distribution de probabilité d'erreur, estimée à l'aide des erreurs du meilleur membre de l'ensemble. Le meilleur membre est celui qui a la plus petite erreur absolue par rapport à l'observation. Lors du calage de la méthode, ces erreurs minimales sont évaluées et une distribution de probabilité est ajustée aux valeurs obtenues. Cette distribution servira pour le tirage aléatoire des erreurs qui iront corriger les prévisions d'ensemble brutes lors de l'application de la méthode. L'hypothèse souvent adoptée est que les erreurs du meilleur membre suivent une loi de probabilité normale. Afin d'assurer que cette hypothèse soit respectée, une transformation logarithmique des valeurs prévues et observées est appliquée lors du calcul de l'erreur du meilleur membre.

L'ensemble corrigé est alors obtenu par :

$$\hat{f}_{i,k,n} = f_{i,k} * \exp(\varepsilon_{i,k,n}) \quad \text{Eq. 37}$$

, où  $f_{i,k}$  : est la valeur de la prévision brute (avant correction) du membre  $k$  de la prévision d'ensemble à l'instant  $i$ ,

$\hat{f}_{i,k,n}$  : est la valeur de la prévision corrigée du membre  $k$  de la prévision d'ensemble à l'instant  $i$ , habillé avec l'erreur, correspondant au tirage  $n$  aléatoire de la distribution d'erreur estimée en calage,

$\varepsilon_{i,k,n}$  : est l'erreur qui est attribuée à la prévision brute, correspondant au tirage  $n$  aléatoire de la distribution d'erreur estimée en calage.

Une des limites de l'approche originalement proposée par Roulston et Smith (2003) est qu'elle présente parfois une efficacité limitée dans la correction des ensembles sous-dispersés, car elle n'arrive pas à bien capturer la variance des observations. Wang et Bishop (2005) ont proposé une première modification pour augmenter la correction de la dispersion dans le cas des ensembles sous-dispersés. Ils ont proposé d'habiller les ensembles avec une contrainte posée sur le moment de deuxième ordre (variance ou covariance dans le cas multi-varié).

En pratique, il s'agit de multiplier la distribution de probabilité d'erreur par un facteur de réduction  $\omega$ . L'objectif est qu'en multipliant par le facteur de réduction  $\omega$  nous arrivons à une distribution d'erreur avec une moyenne égale à zéro et une variance égale à la covariance entre la moyenne des membres de l'ensemble et l'observation. Le facteur de réduction est donné par :

$$\omega = \sqrt{\frac{s^2}{s_e^2}} \quad \text{Eq. 38}$$

, où  $s^2$  est la covariance entre la moyenne d'ensemble et l'observation,  
 $s_e^2$  est la variance des membres de l'ensemble.

Si l'ensemble à corriger est sur-dispersé,  $\omega$  prend des valeurs inférieures à 1. Par contre, si l'ensemble est sous-dispersé à l'origine la valeur de  $\omega$  sera supérieure à 1.

Fortin *et al.* (2006) ont proposé une deuxième modification de la méthode du meilleur membre, en introduisant le facteur de réduction comme un paramètre de calage qui prendra des valeurs différentes pour chaque membre de la prévision d'ensemble et selon sa position (ou rang) par rapport au meilleur membre. Les erreurs du meilleur membre sont séparées en sous-groupes selon la position du meilleur membre au sein de l'ensemble. Par la suite un vecteur d'erreurs est formé pour chaque rang. Le but était de pouvoir ainsi donner un poids de correction différent  $\omega_k$  à chaque membre  $k$  de la prévision d'ensemble, en fonction de sa position par rapport au membre le plus proche de l'observation. Cette approche utilise une fonction de paramétrage avec un seul paramètre  $\alpha$ , qui est calé avec l'objectif de mieux rapprocher la variance des prévisions à la variance des observations et de mieux estimer les probabilités des événements extrêmes.

$$\omega_k \propto \int_{\frac{k-1}{K}}^{\frac{k}{K}} u^{\alpha-1} (1-u)^{\alpha-1} du \quad \text{Eq. 39}$$

Dans cette thèse, nous avons examiné la méthode originale proposée par Roulston et Smith (2003) et ses modifications proposées par Wang et Bishop (2005), Fortin *et al.* (2006), en les appliquant à la correction des prévisions de débits. Ces approches sont d'abord appliquées en météorologie (températures et précipitations) et ensuite Boucher *et al.* (2012) les ont appliquées en hydrologie sur des bassins versants québécois avec des prévisions d'ensemble fournies par le centre météorologique canadien ou avec des données synthétiques.

Les principaux avantages de la méthode du meilleur membre et ses modifications sont :

- la simplicité d'application,
- la bonne correction de la dispersion dans les cas des ensembles sous- et sur-dispersés (ceci dépendant de l'approche appliquée),
- la correction du biais conditionnel de Type I.

Les principales limitations sont :

- la faible correction du biais inconditionnel,
- la faible correction de la capacité des prévisions à discriminer entre les événements,
- la dépendance à la longueur de la période du calage pour l'estimation des paramètres de la distribution de probabilité de l'erreur, notamment lorsque des paramètres doivent être estimés à partir de sous-échantillons de la série disponible.

#### 3.3.3.2 *Modifications méthodologiques de la méthode du meilleur membre*

Nous avons apporté trois modifications méthodologiques supplémentaires à la méthode du meilleur membre originale et nous les avons testées sur nos données.

Dans un premier temps, une discrétisation par quantile des débits observés a été introduite. Au lieu d'avoir une distribution unique des erreurs de prévision, une distribution est estimée pour chaque gamme de débits étudiée (ici, les quantiles 10%, 20%, ..., 90% des débits observés). De cette façon, nous avons espéré améliorer la performance de la méthode en prenant en compte les différences entre les propriétés statistiques des erreurs commises en, par exemple, bas, moyen ou hauts débits. Toutefois, il faut noter qu'une importante limitation de cette modification par quantiles est la longueur restreinte de la période de calage, en particulier pour les quantiles extrêmes pour lesquels il y a peu de données disponibles. Par conséquent, étant donné que nous avons à notre disposition que 4 ans de prévisions, le calage des paramètres de la distribution des erreurs par gamme de débits n'était pas toujours évident et l'amélioration apportée par rapport à la méthode originale restait limitée.

Dans un deuxième temps, nous avons voulu répondre à un problème fréquent dans une grande majorité de méthodes de correction statistique de biais : le maintien de la cohérence temporelle entre les échéances de prévision. L'application d'un traitement statistique par habillage empirique ne garantit pas que les trajectoires des membres seront conservées au long des échéances. L'habillage est en effet souvent calé et appliqué à des échéances fixes, prises séparément. Néanmoins, quelques études ont essayé de répondre à cette

---

problématique dans un contexte de prévision (Krzysztofowicz et Maranzano, 2004 ; Coccia et Todini, 2011 ; Zhao *et al.*, 2011).

Pour répondre à cette question dans cette thèse, tout en gardant une approche simple d'habillage par la méthode du meilleur membre, nous avons remplacé dans la méthode modifiée par Wang et Bishop (2005) la variance des erreurs du meilleur membre  $\sigma^2$  à l'échéance  $j$  par la covariance des erreurs du meilleur membre entre les échéances  $j$  et  $j-1$ . Nous avons également introduit une modification similaire à la méthode originale (Roulston et Smith, 2003). L'expression mathématique de la modification proposée est donnée dans l'équation suivante :

$$\sigma^2 = \overline{(\bar{f}_j - o_j)(\bar{f}_{j-1} - o_{j-1})} - \left(1 - \frac{1}{N}\right) \overline{\left(\sum (\bar{f}_j - f_j)(f_{j-1} - f_{j-1})\right)} \quad \text{Eq. 40}$$

Dans un troisième temps, nous avons examiné une modification méthodologique simple de la méthode du meilleur membre, consistant à prendre en compte l'erreur de plusieurs membres et non pas seulement du meilleur membre. En effet, parfois la méthode du meilleur membre n'arrive pas à cerner l'incertitude totale des prévisions et les prévisions d'ensemble corrigées restent sous-dispersées. Ceci est notamment le cas quand le meilleur membre est très proche de l'observation, conduisant à des erreurs faibles de prévision.

Pour cette raison, et afin d'élargir le faisceau des ensembles corrigés, nous avons introduit dans la construction de la distribution des erreurs de prévision, l'erreur de plusieurs membres. Pour définir le nombre optimal de "meilleurs membres" à considérer, nous avons au préalable conduit quelques tests préliminaires. De ces tests, l'utilisation des 30% des "meilleurs membres" (c'est-à-dire, 15 membres sur un ensemble de 50 membres) a été retenue. Similairement à la méthode du meilleur membre originale, décrite dans le paragraphe précédent, une distribution de probabilité est ajustée aux valeurs des erreurs obtenues. Cette distribution servira à nouveau pour le tirage aléatoire des erreurs sous l'hypothèse que les erreurs de 30% des meilleurs membres suivent une loi de probabilité normale.

En plus des principaux avantages de la méthode du meilleur membre présentés précédemment, nous espérons que les modifications méthodologiques proposées apporteront :

- une cohérence temporelle améliorée,
- une amélioration en termes de biais inconditionnel,

- une correction améliorée du biais conditionnel de Type I,
- une correction de la capacité de discrimination des prévisions,
- une correction améliorée de la dispersion des membres.

Néanmoins, nous sommes conscients que les principales limitations restent les mêmes, notamment celle liée à la forte dépendance à la longueur de la période de calage.

### 3.3.3.3 *Méthode empirique d'habillage avec les erreurs de modélisation*

Nous avons également examiné une approche de traitement statistique qui prend en compte exclusivement les erreurs liées au modèle hydrologique de prévision. Ce type d'approche est une méthode qui peut être assimilée à la famille des méthodes empiriques d'habillage des prévisions (Mathevet, 2010 ; Pagano *et al.*, 2012), ayant pour but spécifique la prise en compte de l'incertitude de modélisation hydrologique dans les prévisions de débits.

La particularité de cette technique réside dans le fait que, au lieu d'évaluer les erreurs commises en situation de prévision, c'est-à-dire, à partir d'une archive de paires " prévisions-observations ", nous évaluons la distribution de probabilité de l'erreur de simulation du modèle hydrologique, c'est-à-dire, quand ce dernier est alimenté par des prévisions dites " parfaites " (la pluie prévue est égale à la pluie observée).

Le post-traitement des prévisions de débits est donc basé sur la modélisation de l'erreur empirique du modèle hydrologique et cette modélisation peut être faite par gamme (ou catégorie) de débits et par échéance de prévision. Dans cette thèse, nous avons suivi la méthode appliquée dans Mathevet (2010), utilisée déjà opérationnellement à EDF. Les gammes de débits ont été définies par des quantiles choisis par un pas de 8% pour les quantiles entre 10%-90% et par un pas de 2% pour le restant des quantiles. Ce choix de découpage par gamme de débits vise à obtenir un nombre limité (20) de sous-populations pour lesquelles les caractéristiques statistiques de l'erreur doivent être estimées.

Pour chaque gamme de débits et pour chaque échéance, les débits simulés par le modèle sur une longue série de simulations passées sont comparés aux débits observés. Les erreurs commises par le modèle hydrologique sont normalisées à partir d'une transformation logarithmique des ratios entre débits observés et débits prévus. Chaque erreur est ainsi calculée par la formule :

$$\varepsilon = \ln\left(\frac{Q}{\hat{Q}}\right) \quad \text{Eq. 41}$$

, où  $\varepsilon$  : est l'erreur du modèle,

$Q$  : est le débit observé,

$\hat{Q}$  : est le débit simulé.

La moyenne et l'écart-type des erreurs obtenues sont ensuite utilisés pour caler une loi normale à la distribution de probabilité de l'erreur. Le tirage aléatoire de plusieurs valeurs d'erreurs permet ensuite de corriger chaque membre de la prévision hydrologique d'ensemble. Cette correction se fait de la même manière que celle présentée dans l'Eq. 45. La correction est appliquée pour chaque gamme de débit et pour chaque échéance de prévision.

Les principaux avantages de cette méthode sont :

- la simplicité d'application,
- la correction du biais inconditionnel,
- la correction du biais conditionnel de Type I.

Les principales limitations sont :

- la faible correction de la capacité des prévisions à discriminer entre les événements,
- l'introduction parfois d'une dispersion non-appropriée (trop importante),
- la nécessité d'une longue archive des erreurs passées du modèle hydrologique.

### **3.3.4 Approches pour le calage des paramètres des méthodes appliquées**

La plupart des approches testées lors de cette thèse sont paramétriques – surtout bi-paramétriques quand on considère l'estimation de la moyenne et de l'écart-type des distributions normales des erreurs de prévision. La procédure habituelle pour estimer les paramètres libres d'une méthode consiste à diviser les séries temporelles disponibles en deux périodes indépendantes : période de calage et période de validation (ou de contrôle). Les études en hydrologie s'appuient souvent sur cette méthodologie, connue par le nom de " split-sample test " et introduite en hydrologie par Klemeš (1986). Par cette approche, les périodes de calage et de validation sont souvent de même longueur pour que les résultats provenant de ces deux sous-échantillons puissent être comparables. La faiblesse de toute procédure de calage/ contrôle est notamment liée à la taille de l'échantillon : plus la série de données disponible est longue, plus performante sera le calage des paramètres.

---

De plus, si les deux périodes sont très contrastées climatologiquement, la performance en validation peut s'avérer très faible.

Une façon de contourner ce problème consiste à effectuer une validation croisée. Lors des travaux de cette thèse nous avons testé l'approche de validation croisée dite "leave one out" (Kearns et Ron, 1999). Le principe de cette approche est d'enlever de la période entière disponible une sous-période sur laquelle nous effectuons la validation de l'approche. Le calage est ainsi effectué sur le reste de la période.

Dans un premier temps nous avons divisé la période entière en deux sous périodes égales de 24 mois. Nous avons effectué le calage sur la première période et la validation sur la deuxième, suivie du calage sur la deuxième période et de la validation sur la première. De ce fait, nous disposons d'une période plus longue de résultats de validation (24+24=48 mois).

Nous avons également appliqué la technique de validation croisée en divisant la période entière (48 mois) en sous-périodes de trois mois. Nous avons calé les paramètres de différentes méthodes sur les 45 mois restants et nous avons fait la validation sur les sous-périodes de 3 mois. Nous avons appliqué cette technique en répétition pour toutes les sous-périodes (fenêtres glissantes) de façon à ce que la validation et le calage soient faits sur la période entière.

Cependant nous avons constaté que pour notre jeu de données la différence en termes d'estimation de paramètres entre les deux divisions étudiés (24 et 3 mois) était presque négligeable. De ce fait, dans le Chapitre 6, nous présentons les résultats de la validation croisée appliquée à deux sous-périodes de 24 mois.

Le Tableau 5 illustre les différentes périodes utilisées selon la méthode de correction statistique appliquée. Pour la famille de méthodes du meilleur membre – étant donné que ces approches utilisent la série temporelle des erreurs de prévision, qui est bien plus courte que les séries historiques d'observations et simulations des modèles – la validation croisée est appliquée avec des périodes de 24 mois. Pour l'habillage avec les analogues et avec l'erreur du modèle hydrologique, le calage s'est effectué sur des périodes plus longues grâce à la disponibilité des archives respectives.

---

Tableau 5 : Périodes de calage et de validation pour chaque approche de traitement statistique testée.

Approche de traitement statistique	Période de calage	Période de validation
Meilleur membre et ses modifications ; autres approches empiriques	01/01/2005-31/12/2006	01/01/2007-26/12/2008
	01/01/2007-26/12/2008	01/01/2005-31/12/2006
Habillage avec les analogues	01/01/1953*-31/12/2007	01/01/2005*-31/12/2008
Habillage avec l'erreur du modèle hydrologique	01/01/1958*-31/12/2004	01/01/2005-26/12/2008

\* selon le bassin versant

### 3.4 Synthèse

Ce chapitre a présenté les principaux concepts autour de l'évaluation de la qualité des prévisions et des méthodes de traitement statistique pour la correction des biais de prévision. Nous avons détaillé les critères et méthodes qui seront utilisées dans les prochains chapitres pour évaluer et corriger les prévisions hydrologiques d'ensemble des échantillons de bassins versants étudiés.

La présentation faite n'est sûrement pas exhaustive, car des nombreux critères et méthodes existent dans la littérature et sont continuellement développés pour améliorer les systèmes de prévision hydrométéorologiques. Nous retiendrons que quels que soient les avantages et les faiblesses de chaque approche, il est essentiel de garder à l'esprit que l'objectif d'un traitement statistique des prévisions est de rendre les prévisions plus fiables, plus justes et précises en termes de volumes prévus, ou de temps et de pointe des crues. La question clé reste à savoir s'il existe une méthode plus performante que toutes les autres et si ceci est valable pour tous les bassins versants, toutes les saisons et tous les événements de prévision. Bien évidemment, il se peut que cette question reste rhétorique, néanmoins c'est cette quête qui inspire plusieurs travaux d'application et de développement de méthodes. Dans cette thèse, nous nous intéressons également à cette question et, plus particulièrement, nous cherchons à comprendre l'impact que l'amélioration de la qualité des prévisions peut avoir sur sa valeur économique, face à une utilité définie par un utilisateur-décideur.

Dans le prochain chapitre (Chapitre 4), nous nous penchons sur les notions d'utilité des prévisions et sur les méthodes existantes pour mesurer la valeur des prévisions. Nos développements et résultats numériques sont ensuite présentés dans le Chapitre 5, Chapitre 6 et Chapitre 7 pour ce qui est, respectivement, de la qualité, du traitement statistique et la valorisation économique des prévisions.

---

