

Prévision saisonnière du débit naturel de crue du fleuve Sénégal à Bakel

Adaptation statistique des prévisions saisonnières de pluie produites par Météo France (modèle ARPEGE système 6) sur l'Afrique de l'ouest

Etude réalisée par l'IRD, Météo France et l'OMVS dans le cadre de : 1- l'accord de coopération conclu en 2017 entre Météo France et l'OMVS ; 2- la convention de travaux passée en 2017 entre l'IRD et l'OMVS

Jean-Claude Bader
IRD / UMR G-EAU, OMVS
Montpellier, 5 Juin 2018

Table des matières

Contexte, objectif et résumé.....	3
Données.....	4
Prévisions saisonnières de pluie produites par ARPEGE système 6 sur l'Afrique de l'ouest.....	4
Débit naturel de crue du fleuve Sénégal à Bakel.....	5
Méthode.....	5
Regroupement des prévisions pluviométriques par zones	5
Variables utilisées	6
Années retenues pour le calage et la validation des modèles.....	6
Synthétisation des variables prévisseurs, types de modèles testés	7
Procédure de calage et validation croisés.....	7
Détermination des meilleurs modèles	8
Résultats obtenus.....	9
Prévision début juillet du débit naturel d'août-septembre-octobre (ASO).....	9
Prévision début août du débit naturel de septembre-octobre (SO).....	9
Conclusion.....	10
Références	11
Tableaux et figures	12

Contexte, objectif et résumé

Dans la continuité des travaux effectués de 2000 à 2015 sur le sujet, l'IRD, Météo France et l'OMVS développent depuis 2017 une nouvelle collaboration concernant la prévision saisonnière du débit naturel de crue pour le fleuve Sénégal à Bakel. Pour apporter une aide à la gestion du barrage de Manantali, situé en amont de Bakel, on cherche à prévoir le débit naturel tri-mensuel ASO d'août à octobre (respectivement bimensuel SO de septembre à octobre), à partir des prévisions de précipitations mensuelles sur l'Afrique de l'ouest, produites début juillet pour les mois d'août, septembre et octobre (respectivement début août pour les mois de septembre et octobre) par le modèle ARPEGE de Météo France.

Les premiers travaux développés dans le cadre de cette nouvelle collaboration ont consisté, en 2017, à mettre au point deux modèles de prévision de débit (Bader, 2017a) exploitant les résultats d'ARPEGE système 5 et à utiliser ces modèles en temps réel, début juillet et début août (Bader 2017b et 2017c). Les prévisions réalisées ont surestimé les très faibles débits naturels de crue observés en ASO et SO 2017, mais avec des erreurs acceptables situées dans les intervalles de confiance à 71% et 65 % des modèles respectifs (Bader 2018).

Météo France exploite désormais une nouvelle version de son modèle, ARPEGE système 6, dont les résultats sont par ailleurs fournis sur une nouvelle grille (mailles élémentaires carrées de 1° au lieu de 0.75° pour le système 5). Une nouvelle adaptation statistique des résultats de ce modèle doit donc être réalisée pour mettre au point des modèles de prévision saisonnière de débit utilisables en temps réel à partir de l'hivernage de 2018.

Basée sur les données de la période 1993-2017, cette nouvelle adaptation est réalisée en reproduisant la méthode déjà utilisée en 2017 pour le système 5, à quelques détails près. Cette méthode, qui favorise la robustesse des modèles grâce à une procédure de calage et validation croisés, consiste à effectuer les opérations suivantes pour chaque modèle (ASO et SO) :

- regroupement des prévisions pluviométriques par zones : 95 zones carrées ou rectangulaires (au lieu de 81 pour le système 5), distinctes par leurs dimensions et positions, sont testées dans le but de déterminer celle où les prévisions pluviométriques sont les plus "prédictives" pour le débit du fleuve Sénégal ;
- sur chaque zone, synthétisation des prévisions pluviométriques (prévisseurs de base exprimés uniquement en valeurs brutes de précipitations et non plus en valeurs centrées réduites, concernant chacun une maille élémentaire ou une maille carrée regroupant plusieurs mailles élémentaires) en un nombre restreint de variables : les variables prévisseurs synthétiques sont soit les 8 premiers vecteurs propres V_1 à V_8 des variables prévisseurs de base, déterminés par analyse en composantes principales sur la période 1993-2016, soit les prévisions pluviométriques moyennées sur la zone ;
- dissociation des années 1993 à 2016 en deux sous-échantillons complémentaires imbriqués A et B de 12 années chacun, représentant chacun du mieux possible la diversité des débits naturels réellement observés de 1993 à 2016;
- application sur chaque zone d'une procédure de calage et validation croisés pour déterminer des régressions linéaires multiples entre variables prévisseurs synthétiques de la zone et débit : pour chaque jeu de variables prévisseurs synthétiques (V_1 à V_8 , V_1 à V_7 , ..., V_1 pour 8 modèles de type 1; prévisions pluviométriques moyennées sur la zone pour 1 modèle de type 2), la régression est calée sur l'échantillon A pour estimer ses erreurs en validation sur l'échantillon B, puis calée sur l'échantillon B pour estimer ses erreurs en validation sur l'échantillon A, et enfin calée sur l'échantillon A+B=1993-2016 pour estimer ses erreurs en calage sur cette période ; on conserve pour la zone le modèle de type 1 présentant la plus petite erreur type de validation sur la période 1993-2016, ainsi que le modèle de type 2 ; à chacun de ces deux modèles

sont associés les paramètres suivants, de type "Nash and Sutcliffe coefficient of model efficiency":

- NSE_c caractérise la précision des débits modélisés en calage (calculés sur les années A+B avec le modèle calé sur A+B)
- NSE_v caractérise la précision des débits modélisés en validation (calculés sur les années A avec le modèle calé sur B et calculés sur les années B avec le modèle calé sur A)
- NSE_c caractérise la stabilité des coefficients multiplicatifs de régression du modèle (par rapport aux prévisionneurs de base) en fonction de l'échantillon de calage, en comparant leurs valeurs calées sur les années A et sur les années B avec leurs valeurs calées sur les années A+B ;
- le meilleur des 190 modèles analysés est celui qui présente la plus forte valeur pour le paramètre NSE_g , correspondant à la moyenne des paramètres NSE_c , NSE_v et NSE_f (le même poids est donc accordé à ces trois derniers paramètres, au lieu des différents poids arbitraires utilisés pour l'adaptation du système 5) ;
- on teste enfin le meilleur modèle dans des conditions simulant le "temps réel", en l'appliquant (avec ses coefficients de régression calés sur la période A+B=1993-2016) sur l'année 2017.

Pour la prévision du débit d'août-septembre-octobre (ASO) calculée début juillet, le meilleur modèle, de type 1, exploite les résultats produits par ARPEGE système 6 sur une zone C12 centrée très à l'est/sud-est du bassin du Sénégal. Il est calé sur le premier vecteur propre des prévisions pluviométriques de base sur cette zone. Ses performances sont modestes en calage ($NSE_c = 0.234$) et en validation ($NSE_v = 0.203$), mais ses coefficients de régression sont très stables ($NSE_f = 0.965$) et il donne de bons résultats dans les conditions du "temps réel" pour l'année 2017 (prévision dans l'intervalle de confiance à 58 % du modèle).

Pour la prévision du débit de septembre-octobre (SO) calculée début août, le meilleur modèle, de type 1, exploite les résultats produits par ARPEGE système 6 sur une zone A6 centrée à l'est du bassin du Sénégal. Il est calé sur les 4 premiers vecteurs propres des prévisions pluviométriques de base sur cette zone. Ses performances sont assez bonnes en calage ($NSE_c = 0.399$) et très modestes en validation ($NSE_v = 0.132$). Ses coefficients de régression sont stables ($NSE_f = 0.910$) et il donne des résultats corrects dans les conditions du "temps réel" pour l'année 2017 (prévision dans l'intervalle de confiance à 74% du modèle).

Avec des paramètres NSE_f assez proches de 1, les deux modèles sélectionnés paraissent relativement robustes. Leur utilisation en temps réel peut donc être envisagée, malgré leur faible précision.

Données

Prévisions saisonnières de pluie produites par ARPEGE système 6 sur l'Afrique de l'ouest

Les prévisions pluviométriques sont fournies par Météo-France pour 1122 mailles élémentaires de 1° de côté couvrant une grille située entre les latitudes -3.5° et 18.5° et entre les longitudes -20.5° et 30.5° (22 mailles élémentaires dans la direction nord-sud et 51 dans la direction est-ouest, voir fig. 1). Ces prévisions de précipitations mensuelles, qui couvrent les années 1993 à 2017, concernent :

- le mois d'août, le mois de septembre et le mois d'octobre avec des valeurs calculées début juillet;
- le mois de septembre et le mois d'octobre avec des valeurs calculées début août.

Chaque prévision (concernant une année, une date de calcul, un mois visé et une maille élémentaire donnés) consiste en plusieurs valeurs de lame précipitée mensuelle exprimées en mm (25 valeurs avant l'année 2017 et 51 valeurs à partir de 2017), calculées à partir de conditions initiales légèrement décalées, dont on n'exploite que la moyenne pour cette phase d'adaptation statistique et la mise au point du modèle de prévision saisonnière d'écoulement.

Débit naturel de crue du fleuve Sénégal à Bakel

On exploite ici les moyennes bimensuelle de septembre-octobre et tri-mensuelle d'août-septembre-octobre du débit naturel du fleuve Sénégal à Bakel (tab. 1), qui est reconstitué au pas de temps journalier pour les années 1993 à 2017 par la méthode décrite par Bader (2015), à partir des observations hydrométriques figurant dans la base de l'OMVS et relatives au régime influencé par le barrage de Manantali.

Méthode

Regroupement des prévisions pluviométriques par zones

Au total, 95 zones distinctes sont considérées pour déterminer celles où les prévisions pluviométriques sont les plus "prédictives" pour le débit naturel de crue du fleuve Sénégal. Ces zones, de forme carrée ou rectangulaire, correspondent à 6 types différents caractérisés chacun par sa forme et ses dimensions (fig. 1). Pour un type donné, les différentes zones ne se distinguent que par leur position.

Toute zone est formée de X mailles carrées étendues, dont chacune correspond soit à une maille élémentaire de 1° de côté, soit au regroupement de plusieurs mailles élémentaires (nécessité par certaines limitations imposées par le tableur utilisé pour les analyses) :

- les 30 zones testées de type A sont des carrés de 7° de côté. Chacune est constituée de $7 \times 7 = 49$ mailles étendues, ces dernières correspondant chacune à une maille élémentaire. Les différentes zones de type A se distinguent par des décalages de 3° (3 mailles étendues) dans la direction est-ouest et 2° (2 mailles étendues) dans la direction nord-sud ;
- les 24 zones testées de type B sont des carrés de 10° de côté. Chacune est constituée de $5 \times 5 = 25$ mailles étendues de 2° de côté, ces dernières correspondant chacune au regroupement de 4 mailles élémentaires. Les différentes zones de type B se distinguent par des décalages de 4° (2 mailles étendues) dans la direction est-ouest et dans la direction nord-sud ;
- les 12 zones testées de type C sont des carrés de 14° de côté. Chacune est constituée de $7 \times 7 = 49$ mailles étendues de 2° de côté, ces dernières correspondant chacune au regroupement de 4 mailles élémentaires. Les différentes zones de type C se distinguent par des décalages de 6° (3 mailles étendues) dans la direction est-ouest et dans la direction nord-sud ;
- les 12 zones testées de type D sont des carrés de 18° de côté. Chacune est constituée de $6 \times 6 = 36$ mailles étendues de 3° de côté, ces dernières correspondant chacune au regroupement de 9 mailles élémentaires. Les différentes zones de type D se distinguent par des décalages de 6° (2 mailles étendues) dans la direction est-ouest et 3° (1 maille étendue) dans la direction nord-sud ;
- les 3 zones testées de type E sont des rectangles de 36° de côté dans la direction est-ouest et 21° de côté dans la direction nord-sud. Chacune est constituée de $12 \times 7 = 84$ mailles étendues de 3° de côté, ces dernières correspondant chacune au regroupement

de 9 mailles élémentaires. Les différentes zones de type E se distinguent par des décalages de 6° (2 mailles étendues) dans la direction est-ouest.

- les 14 zones testées de type F sont des rectangles de 12° de côté dans la direction est-ouest et 18° de côté dans la direction nord-sud. Chacune est constituée de $6 \times 9 = 54$ mailles étendues de 2° de côté, ces dernières correspondant chacune au regroupement de 4 mailles élémentaires. Les différentes zones de type F se distinguent par des décalages de 4° (2 mailles étendues) dans la direction est-ouest et dans la direction nord-sud.

Remarque : le changement de taille des mailles élémentaires de la grille d'Arpege intervenu entre le système 5 (0.75° de côté) et le système 6 (1° de côté) ne permettait pas de tester des zones exactement identiques à celles qui avaient été testées en 2017 pour l'adaptation statistiques des résultats d'Arpege système 5. Les dénominations de zones utilisées ici ne correspondent donc pas aux dénominations utilisées en 2017.

Variables utilisées

La variable à prévoir (prévisande) est bien sûr le débit moyen Q du fleuve Sénégal à Bakel, soit sur la période d'août-septembre-octobre (modèles ASO), soit sur la période de septembre-octobre (modèles SO).

Pour une zone donnée, les Y variables prévisseurs de base sont quant à elles les prévisions pluviométriques relatives aux X mailles étendues constituant la zone, calculées soit début juillet pour les mois d'août, septembre et octobre (modèles ASO), soit début août pour les mois de septembre et octobre (modèles SO). Pour les mailles étendues constituées de plusieurs mailles élémentaires, chaque lame précipitée prévue correspond à la moyenne des lames précipitées prévues sur ces mailles élémentaires.

Le nombre Y des variables prévisseurs de base des zones testées dépend du nombre X de mailles étendues de celles-ci, avec $Y = 2X$ pour les prévisions SO et $Y = 3X$ pour les prévisions ASO :

- pour les zones de type A : $Y = 147$ pour ASO et $Y = 98$ pour SO ;
- pour les zones de type B : $Y = 75$ pour ASO et $Y = 50$ pour SO ;
- pour les zones de type C : $Y = 147$ pour ASO et $Y = 98$ pour SO ;
- pour les zones de type D : $Y = 108$ pour ASO et $Y = 72$ pour SO ;
- pour les zones de type E : $Y = 252$ pour ASO et $Y = 168$ pour SO ;
- pour les zones de type F : $Y = 162$ pour ASO et $Y = 108$ pour SO ;

Remarque : les calculs réalisés ici n'utilisent que les valeurs brutes des prévisions pluviométriques, exprimées en mètre, pour définir les variables prévisseurs de base. Les valeurs centrées réduites de ces prévisions ne sont pas utilisées.

Années retenues pour le calage et la validation des modèles

La période commune des données disponibles s'étend de 1993 à 2017 pour les prévisions pluviométriques et les débits naturels. Avec une seule valeur par année, chacune des variables prévisande ou prévisseur de base présente donc un échantillon total de 25 valeurs.

De façon arbitraire, les données de l'année 2017 sont isolées pour n'intervenir ni dans le calage ni dans la validation des modèles testés. Elles peuvent ainsi être exploitées pour effectuer un test ultime des modèles retenus, dans des conditions simulant l'utilisation en temps réel de ces modèles.

Le calage et la validation des modèles testés sont effectués sur la période 1993-2016, avec un échantillon de 24 valeurs pour chacune des variables. Pour les besoins de la procédure de calage et validation croisés, cet échantillon est scindé en deux sous échantillons

complémentaires nommés A et B, d'effectifs équivalents (12 années pour A et 12 années pour B). La répartition des années de la période A+B (1993-2016) entre A et B est effectuée de façon que les diverses valeurs réellement observées du débit naturel entre 1993 et 2016 soient représentées du mieux possible par chacun des échantillons A et B. À cette fin, les valeurs du débit de la période A+B sont d'abord classées par ordre croissant. Elles sont ensuite réparties pour constituer les échantillons A et B (ce qui détermine les années associées à ces deux sous-échantillons) en fonction du rang du débit :

- échantillon A : rangs 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23 ;
- échantillon B : rangs 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24.

Les années relatives aux échantillons A et B, qui diffèrent pour les modèles ASO et SO, sont indiquées dans le tableau 1.

Synthétisation des variables prévisseurs, types de modèles testés

Sur toutes les zones testées, et quelle que soit la période visée par la prévision saisonnière de débit (ASO ou SO), les prévisions pluviométriques utilisées définissent des variables prévisseurs de base dont le nombre Y (toujours supérieur ou égal à 50) dépasse largement l'effectif total de l'échantillon disponible (25 années). Il est donc impératif de réduire considérablement le nombre des variables prévisseurs pour envisager leur adaptation statistique, visant à les transformer en prévisions de débit. Deux méthodes sont utilisées à cette fin :

- la première méthode consiste à remplacer les Y variables prévisseurs de base par leurs N premiers vecteurs propres V_1 à V_N déterminés sur la période A+B par analyse en composantes principales, avec N compris entre 1 et 8. Pour rappel, ces vecteurs propres sont des combinaisons linéaires des variables de base, qui ont la particularité d'être tous indépendants entre eux sur la période d'analyse et d'être classés par variance décroissante. Les premiers vecteurs propres sont donc ceux qui regroupent la majeure partie de l'information contenue dans les variables de base. Avec cette méthode, on teste pour chaque zone 8 modèles différents, de type 1, permettant de prévoir le débit Q par une combinaison linéaire des N variables prévisseurs synthétiques V_1 à V_N , avec N compris entre 1 et 8 ;
- la seconde méthode consiste à remplacer les Y variables prévisseurs de base par leurs valeurs mensuelles moyennées sur toutes les mailles de la zone. Avec cette méthode, on teste pour chaque zone un unique modèle, de type 2, permettant de prévoir le débit Q par une combinaison linéaire de N variables prévisseurs synthétiques, avec $N=3$ pour ASO et $N=2$ pour SO.

Procédure de calage et validation croisés

Pour chaque zone testée, la procédure de calage et validation croisés est appliquée pour chacun des 8 modèles de type 1 et pour le modèle de type 2 mentionnés ci-dessus. Cette procédure consiste à effectuer les opérations suivantes :

- le modèle est d'abord calé par régression linéaire multiple entre les N variables prévisseurs synthétiques et le débit Q sur l'échantillon A, avec $N+1$ coefficients de régression notés C_{S_A} . En exprimant les variables prévisseurs synthétiques en fonction des Y variables prévisseurs de base, cette régression s'exprime finalement sous forme d'une combinaison linéaire des variables prévisseurs de base dont les $Y+1$ coefficients de régression sont notés C_A (un coefficient multiplicatif pour chacune des Y variables prévisseurs de base, plus un coefficient additif). L'application sur l'échantillon B du modèle ainsi calé sur A permet d'évaluer ses erreurs de validation Ev_B sur B, par différence entre les débits modélisés Q_m et les débits réels Q ;

- les mêmes opérations sont reproduites après avoir inversé le rôle des échantillons A et B : le modèle alors calé sur B donne des coefficients de régression notés C_B permettant d'exprimer le débit en fonction des variables prévisseurs de base ; appliqué sur l'échantillon A, il donne des erreurs de validation sur A notées E_{v_A} ;
- les erreurs de validation E_{v_B} et E_{v_A} constituent un ensemble d'erreurs de validation $E_{v_{A+B}}$ sur l'échantillon A+B, dont la moyenne quadratique correspond à une erreur type de validation globale sur A+B, notée $ET_{v_{A+B}}$. Avec l'écart type $\sigma_{A+B}(Q)$ des débits Q réels sur A+B, ceci permet de calculer un coefficient d'efficacité de modèle en validation, noté NSE_v par référence à la notation habituelle du "Nash and Sutcliffe coefficient of model efficiency" :

$$NSE_v = 1 - (ET_{v_{A+B}} / \sigma_{A+B}(Q))^2 \quad [1]$$

- en reproduisant la procédure précédemment utilisée pour les calages sur A et sur B, on cale ensuite le modèle sur l'échantillon A+B, ce qui détermine des coefficients de régression notés C_{A+B} reliant le débit aux variables prévisseurs de base. L'application sur l'échantillon A+B du modèle ainsi calé sur A+B permet d'évaluer ses erreurs de calage, notées $E_{c_{A+B}}$, dont la moyenne quadratique $ET_{c_{A+B}}$ permet de calculer un coefficient d'efficacité de modèle en calage, noté NSE_c :

$$NSE_c = 1 - (ET_{c_{A+B}} / \sigma_{A+B}(Q))^2 \quad [2]$$

- la stabilité des coefficients de régression du modèle en fonction de l'échantillon de calage est enfin évaluée par un coefficient NSE_f . En notant $\sigma_{A+B}(C_{A+B})$ l'écart type des Y coefficients multiplicatifs de régression du modèle calé sur A+B (la constante additive de l'équation de régression n'est pas considérée), on calcule ainsi NSE_f , qui est égal à 1 en cas de stabilité parfaite des coefficients de régression :

$$NSE_f = 1 - (\sum_{i..j} ((C_B - C_{A+B})^2 + (C_A - C_{A+B})^2) / (\sigma_{A+B}(C_{A+B}))^2) / (2N) \quad [3]$$

Pour chaque zone testée, on conserve finalement la version calée sur A+B du modèle de type 1 présentant, parmi les huit évalués, les meilleurs résultats en validation (c'est-à-dire avec la plus forte valeur de NSE_v), ainsi que la version également calée sur A+B du modèle de type 2. Pour la comparaison des modèles obtenus sur les différentes zones, chacun de ces deux modèles est caractérisé par les coefficients NSE_v , NSE_c et NSE_f .

Détermination des meilleurs modèles

Pour une période visée par la prévision de débit (ASO ou SO), l'analyse décrite ci-dessus produit les indicateurs de performance NSE_v , NSE_c et NSE_f pour 190 modèles différents (un modèle de type 1 et un modèle de type 2 par zone, pour 95 zones testées). Pour comparer ces modèles entre eux et déterminer le plus performant, on utilise un critère global unique NSE_g correspondant à la moyenne de NSE_v , NSE_c et NSE_f :

$$NSE_g = (NSE_v + NSE_c + NSE_f) / 3 \quad [4]$$

Les meilleurs modèles sont caractérisés par les plus fortes valeurs obtenues pour NSE_g .

Remarque : l'adaptation statistique des résultats d'Arpege système 5 réalisée en 2017 utilisait un critère NSE_g calculé par moyenne pondérée de NSE_v , NSE_c et NSE_f , avec des poids arbitrairement choisis favorisant les performances des modèles en validation. L'utilisation

d'une moyenne non pondérée vise à obtenir ici un critère moins arbitraire pour déterminer les meilleurs modèles, avec une importance non négligeable accordée à la stabilité de leurs coefficients de régression.

Résultats obtenus

Prévision début juillet du débit naturel d'août-septembre-octobre (ASO)

Le tableau 2 résume les résultats obtenus pour les 10 meilleurs modèles ASO, qui présentent tous des résultats modestes en calage ($NSE_c \leq 0.276$) et en validation ($NSE_v \leq 0.203$), malgré une bonne stabilité de leurs paramètres ($NSE_f \geq 0.889$). Parmi les trois meilleurs modèles testés, les deux premiers utilisent les prévisions pluviométriques sur les zones C12 et C6 (fig. 2), situées très à l'est/sud-est du bassin du fleuve Sénégal :

- modèle de type 1 $f(V_1)$ sur la zone C12 comprise entre les latitudes -1.5° et 12.5° et entre les longitudes 9.5° et 23.5° ;
- modèle de type 1 $f(V_1-V_2)$ sur la zone C6 comprise entre les latitudes 4.5° et 18.5° et entre les longitudes 9.5° et 23.5° ;
- modèle de type 2 sur la zone B15 comprise entre les latitudes 0.5° et 10.5° et entre les longitudes -12.5° et -2.5° ;

Pour les trois modèles, la figure 3 compare les débits réels aux débits modélisés avec le calage effectué sur l'échantillon A+B (1993-2016). Pour chaque modèle, la quasi-totalité des débits modélisés se situe dans l'intervalle de confiance à 90%, évalué à partir de l'erreur type $ET_{V_{A+B}}$ de validation avec l'hypothèse de répartition normale des erreurs (voir plus bas). Le débit modélisé pour l'année 2017 dans les conditions du "temps réel" est assez bon pour les deux premiers modèles et acceptable pour le troisième, se situant respectivement dans l'intervalle de confiance à 58 %, 49% et 86%.

La figure 4 renseigne sur la robustesse du meilleur modèle ($f(V_1)$ sur zone C12), en comparant sur l'échantillon A+B le débit réel au débit modélisé en validation (calculé sur A (respectivement B) avec le modèle calé sur B (respectivement A)). On constate là encore que la quasi-totalité des points modélisés se situe dans l'intervalle de confiance à 90%. La robustesse du modèle est également renseignée par la figure 5, qui compare les 147 valeurs ($=7 \times 7 \times 3$) des coefficients multiplicatifs de régression C_A et C_B calés respectivement sur les échantillons A et B avec celles des coefficients C_{A+B} calés sur A+B. Du fait que le modèle est une fonction affine du seul premier vecteur propre des 147 variables de prévisions pluviométriques, lui-même combinaison linéaire de ces variables, les coefficients C_A et C_B sont liés de façon linéaire à C_{A+B} . Ces relations sont assez proches de la première bissectrice.

La figure 6 montre que pour le meilleur modèle ASO, les erreurs de modélisation se répartissent à peu près suivant une loi normale de moyenne nulle et d'écart type égal à $ET_{C_{A+B}}$ ($= 462 \text{ m}^3/\text{s}$) en calage et égal à $ET_{V_{A+B}}$ ($= 472 \text{ m}^3/\text{s}$) en validation.

Enfin, les coefficients de régression C_{A+B} du meilleur modèle ASO présentent une certaine organisation en fonction des coordonnées géographiques des mailles (fig. 7), avec des valeurs globalement croissantes en fonction de la latitude pour les mois d'août et septembre.

Prévision début août du débit naturel de septembre-octobre (SO)

Parmi tous les modèles SO testés, les dix meilleurs (tab. 2) présentent des résultats généralement meilleurs que pour ASO en calage ($NSE_c > 0.3$ à 80%), moins bons en validation (NSE_v inférieur ou égal à 0.205 et négatif à 40%) et assez bons pour la stabilité des paramètres ($NSE_f \geq 0.728$). Les trois meilleurs modèles utilisent les prévisions pluviométriques sur les zones A6, A5 et B21 (fig.8), décalées à l'est du bassin du fleuve Sénégal à Bakel pour les deux premières et très au sud pour la troisième :

- modèle de type 1 $f(V_1-V_4)$ sur la zone A6 comprise entre les latitudes 10.5° et 17.5° et entre les longitudes -4.5° et 2.5° ;
- modèle de type 1 $f(V_1-V_2)$ sur la zone A5 comprise entre les latitudes 10.5° et 17.5° et entre les longitudes -7.5° et -0.5° ;
- modèle de type 1 $f(V_1-V_5)$ sur la zone B21 comprise entre les latitudes -3.5° et 6.5° et entre les longitudes -12.5° et -2.5° .

Pour ces trois modèles, la figure 9 compare les débits réels aux débits modélisés avec le calage effectué sur l'échantillon A+B (1993-2016). On constate que les débits modélisés sur la période A+B se situent (en totalité pour le premier modèle et en quasi totalité pour les deux autres) dans l'intervalle de confiance à 90%, évalué à partir de l'erreur type $ET_{V_{A+B}}$ de validation avec l'hypothèse de répartition normale des erreurs (voir plus bas). Le débit modélisé pour l'année 2017 dans les conditions du "temps réel" est assez bon pour les deux premiers modèles et mauvais pour le troisième, se situant respectivement dans l'intervalle de confiance à 74 %, 89% et 99.3%.

La figure 10 renseigne sur la robustesse du meilleur modèle ($f(V_1..V_4)$ sur zone A6), en comparant sur l'échantillon A+B le débit réel au débit modélisé en validation (calculé sur A (respectivement B) avec le modèle calé sur B (respectivement A)). On constate que la quasi-totalité des points modélisés se situe dans l'intervalle de confiance à 90%. La robustesse du modèle apparaît également sur la figure 11, qui compare les 98 valeurs ($=7 \times 7 \times 2$) des coefficients multiplicatifs de régression C_A et C_B calés respectivement sur les échantillons A et B avec les valeurs C_{A+B} calées sur A+B. La proximité des points avec la première bissectrice montre que C_A et C_B sont très proches de C_{A+B} .

La figure 12 montre que pour le meilleur modèle SO, la répartition des erreurs de modélisation est assez proche d'une loi normale de moyenne nulle et d'écart type égal à $ET_{C_{A+B}}$ ($= 407 \text{ m}^3/\text{s}$) en calage et égal à $ET_{V_{A+B}}$ ($= 489 \text{ m}^3/\text{s}$) en validation.

Enfin, les coefficients de régression C_{A+B} du meilleur modèle SO présentent une certaine organisation en fonction des coordonnées géographiques des mailles (fig. 13), avec une tendance relative à l'augmentation vers l'est en août (sauf extrême nord-est) et vers le nord en septembre.

Conclusion

Les deux modèles de prévision saisonnière du débit naturel de crue du fleuve Sénégal déterminés ici et proposés pour l'utilisation en temps réel à partir de 2018 sont les suivants :

- pour ASO : $f(V_1)$ utilisant les prévisions pluviométriques d'ARPEGE système 6 calculées début juillet sur une zone C12 située très à l'est/sud-est du bassin du fleuve Sénégal, entre les latitudes -1.5° et 12.5° et entre les longitudes 9.5° et 23.5° ;
- pour SO : $f(V_1-V_4)$ utilisant les prévisions pluviométriques d'ARPEGE système 6 calculées début août sur une zone A6 située à l'est du bassin du fleuve Sénégal, entre les latitudes 10.5° et 17.5° et entre les longitudes -4.5° et 2.5° .

Les paramètres de traduction des prévisions pluviométriques en débit sont relativement robustes pour les deux modèles : $NSE_f = 0.965$ pour ASO et $NSE_f = 0.910$ pour SO. Ceci permet d'estimer les intervalles de confiance des modèles à partir de leurs incertitudes déterminées en validation (répartition quasi normale des erreurs ; $ET_{V_{A+B}} = 472 \text{ m}^3/\text{s}$ et $NSE_v = 0.203$ pour ASO ; $ET_{V_{A+B}} = 489 \text{ m}^3/\text{s}$ et $NSE_v = 0.132$ pour SO). Ces intervalles de confiance sont malheureusement très larges :

- intervalle de confiance à 90 % : $\pm 776 \text{ m}^3/\text{s}$ pour ASO et $\pm 805 \text{ m}^3/\text{s}$ pour SO;
- intervalle de confiance à 80 % : $\pm 605 \text{ m}^3/\text{s}$ pour ASO et $\pm 627 \text{ m}^3/\text{s}$ pour SO.

Pour l'année 2017, les modèles donnent des erreurs de +380 m³/s pour ASO et +552 m³/s pour SO, situées respectivement dans les intervalles de confiance à 58 % et 74 %. Ces résultats, obtenus en simulant l'utilisation des modèles en temps réel avec des données non utilisées pour leur mise au point, sont relativement satisfaisants.

Références

-Bader J.C. (ed.), 2015 : Monographie hydrologique du fleuve Sénégal : de l'origine des mesures jusqu'en 2011, 79 p. (livret) + CD-Rom (920 p.), IRD, Marseille, ISBN 978-2-7099-1885-5

- Bader J.C., 2017a : Prévion saisonnière du débit naturel de crue du fleuve Sénégal à Bakel. Nouvelle adaptation statistique des prévisions saisonnières de pluie produites par Météo France (modèle ARPEGE système 5) sur l'Afrique de l'ouest. IRD/ UMR G-EAU, Montpellier, 5 juin 2017, 21 p.

- Bader J.C., 2017b : Prévion saisonnière du débit naturel de crue du fleuve Sénégal à Bakel. Prévion en temps réel du débit d'août-septembre-octobre 2017, calculée le 1^{er} juillet 2017. IRD/ UMR G-EAU, Montpellier, 3 juillet 2017, 2 p.

- Bader J.C., 2017c : Prévion saisonnière du débit naturel de crue du fleuve Sénégal à Bakel. Prévion en temps réel du débit de septembtre-octobre 2017, calculée le 1^{er} août 2017. IRD/ UMR G-EAU, Montpellier, 2 août 2017, 2 p.

- Bader J.C., 2018 : Prévion saisonnière du débit naturel de crue du fleuve Sénégal à Bakel. Evaluation a posteriori des prévisions rélisées en temps réel en 2017. IRD/ UMR G-EAU, Montpellier, 15 janvier 2018, 5 p.

Tableaux et figures

année	août-septembre-octobre (ASO)			septembre-octobre (SO)		
	débit (m ³ /s)	rang (A+B)	échantillon	débit (m ³ /s)	rang (A+B)	échantillon
1993	1108	1	A	1136	1	A
1994	2368	18	B	2415	20	B
1995	1984	13	A	1859	13	A
1996	1352	4	B	1421	6	B
1997	1451	7	A	1459	8	B
1998	1984	14	B	2020	14	B
1999	2727	23	A	2679	23	A
2000	1418	6	B	1333	5	A
2001	1513	8	B	1555	9	A
2002	1360	5	A	1148	2	B
2003	2643	22	B	2408	19	A
2004	1236	3	A	1154	3	A
2005	1698	11	A	1656	10	B
2006	1138	2	B	1160	4	B
2007	2190	17	A	2153	15	A
2008	1820	12	B	1808	12	B
2009	2126	15	A	2312	17	A
2010	2140	16	B	2434	21	A
2011	1616	10	B	1436	7	A
2012	2884	24	B	2775	24	B
2013	2457	20	B	2407	18	B
2014	1545	9	A	1663	11	A
2015	2520	21	A	2474	22	B
2016	2446	19	A	2297	16	B
2017	1206			1151		

Tableau 1 : valeurs moyennes d'août-septembre-octobre et de septembre-octobre du débit du fleuve Sénégal à Bakel (régime naturel reconstitué) pour les années 1993 à 2017

mois visés	nom	zone utilisée caractéristiques	modèle	valeur				rang (sur 190)				hors A+B
				<i>NSEc</i>	<i>NSEv</i>	<i>NSEf</i>	<i>NSEg</i>	<i>NSEc</i>	<i>NSEv</i>	<i>NSEf</i>	<i>NSEg</i>	<i>fdep(EAmax)</i>
ASO	C12	Z(2.000X2.000;12.500;9.500;-1.500;23.500)	f(V1)	0.234	0.203	0.965	0.467	17	1	10	1	0.421
	C6	Z(2.000X2.000;18.500;9.500;4.500;23.500)	f(V1-V2)	0.234	0.183	0.979	0.465	18	2	5	2	0.506
	B15	Z(2.000X2.000;10.500;-12.500;0.500;-2.500)	f(prev moy)	0.256	0.166	0.939	0.454	10	4	17	3	0.138
	A15	Z(1.000X1.000;13.500;-13.500;6.500;-6.500)	f(V1-V2)	0.202	0.166	0.980	0.449	31	3	4	4	0.311
	A22	Z(1.000X1.000;11.500;-10.500;4.500;-3.500)	f(prev moy)	0.265	0.112	0.953	0.443	8	20	14	5	0.443
	F14	Z(2.000X2.000;14.500;3.500;-3.500;15.500)	f(V1)	0.180	0.154	0.965	0.433	52	6	11	6	0.314
	A9	Z(1.000X1.000;15.500;-13.500;8.500;-6.500)	f(V1-V2)	0.269	0.135	0.889	0.431	7	11	39	7	0.292
	A28	Z(1.000X1.000;9.500;-10.500;2.500;-3.500)	f(prev moy)	0.276	0.098	0.910	0.428	6	27	33	8	0.156
	D12	Z(3.000X3.000;15.500;9.500;-2.500;27.500)	f(V1)	0.200	0.152	0.919	0.424	32	7	27	9	0.403
	B24	Z(2.000X2.000;6.500;-0.500;-3.500;9.500)	f(prev moy)	0.178	0.133	0.959	0.423	54	13	13	10	0.251
SO	A6	Z(1.000X1.000;17.500;-4.500;10.500;2.500)	f(V1-V4)	0.399	0.132	0.910	0.480	9	3	9	1	0.259
	A5	Z(1.000X1.000;17.500;-7.500;10.500;-0.500)	f(V1-V2)	0.241	0.205	0.993	0.480	17	1	2	2	0.107
	B21	Z(2.000X2.000;6.500;-12.500;-3.500;-2.500)	f(V1-V5)	0.446	0.085	0.903	0.478	7	7	11	3	0.007
	C5	Z(2.000X2.000;18.500;3.500;4.500;17.500)	f(V1-V4)	0.429	0.164	0.820	0.471	8	2	29	4	0.045
	C4	Z(2.000X2.000;18.500;-2.500;4.500;11.500)	f(V1-V5)	0.489	0.092	0.786	0.456	3	6	38	5	0.072
	B6	Z(2.000X2.000;18.500;-0.500;8.500;9.500)	f(V1-V3)	0.302	0.126	0.906	0.445	11	4	10	6	0.150
	D3	Z(3.000X3.000;18.500;-8.500;0.500;9.500)	f(V1-V7)	0.567	-0.026	0.759	0.433	1	44	44	7	0.129
	F14	Z(2.000X2.000;14.500;3.500;-3.500;15.500)	f(V1-V5)	0.461	-0.016	0.789	0.411	4	38	37	8	0.063
	D9	Z(3.000X3.000;15.500;-8.500;-2.500;9.500)	f(V1-V7)	0.540	-0.047	0.728	0.407	2	51	54	9	0.159
	D6	Z(3.000X3.000;18.500;9.500;0.500;27.500)	f(V1-V2)	0.125	0.111	0.958	0.398	55	5	6	10	0.271

Tableau 2 : résultats obtenus avec les 10 meilleurs modèles par catégorie

- caractéristiques des zones : dimensions de maille étendue en degrés, latitude et longitude (angle nord-ouest), latitude et longitude (angle sud-est) ;
- modèle : f(V₁-V_N) pour type 1 et f(prev moy) pour type 2);
- *NSE* : indice de performance (*NSEc* : calage sur A+B ; *NSEv* : validation croisée sur A et B ; *NSEf* : stabilité des coefficients ; *NSEg* : global) ;
- *fdep(EAmax)* : probabilité de dépassement de l'erreur absolue maximale constatée hors de l'échantillon A+B (année 2017)

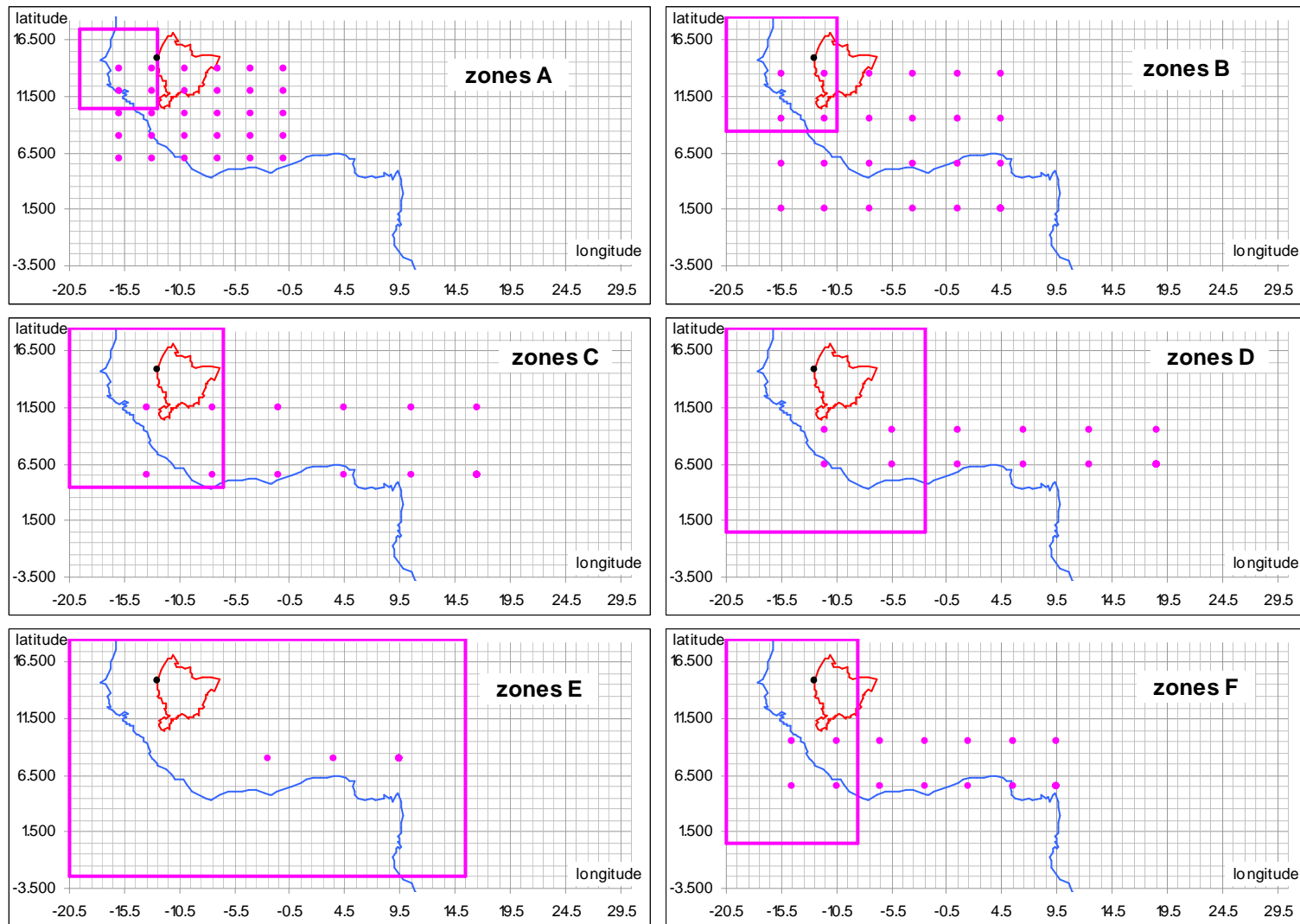


Figure 1 : types de zones testées sur la carte d'Afrique de l'ouest (gris : mailles élémentaires ; bleu : côte atlantique ; rouge : bassin versant du fleuve Sénégal à Bakel ; noir : Bakel ; points mauve : centres des zones ; traits mauve : zone située la plus au nord-est pour chaque type)

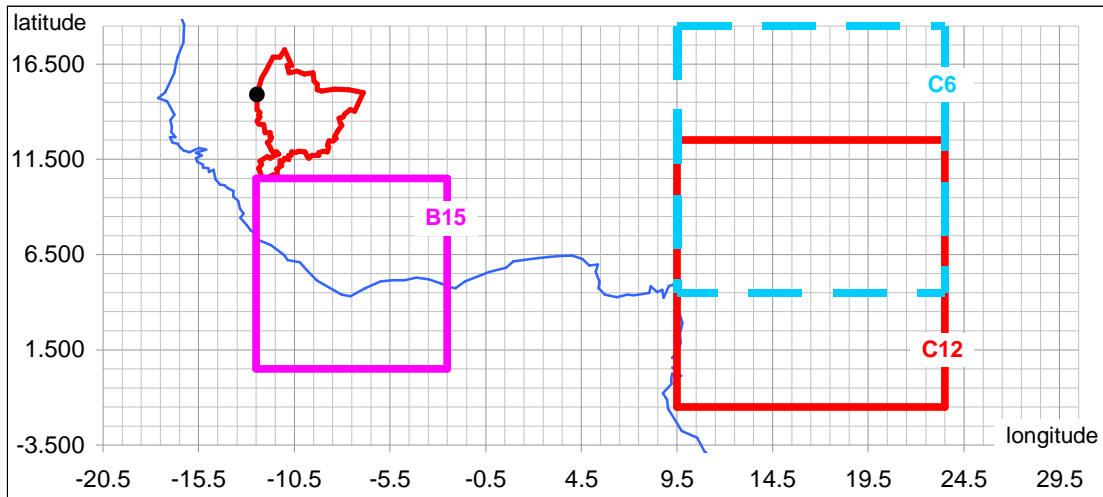


Figure 2 : zones C12, C6 et B15 associées aux trois meilleurs modèles de prévision ASO

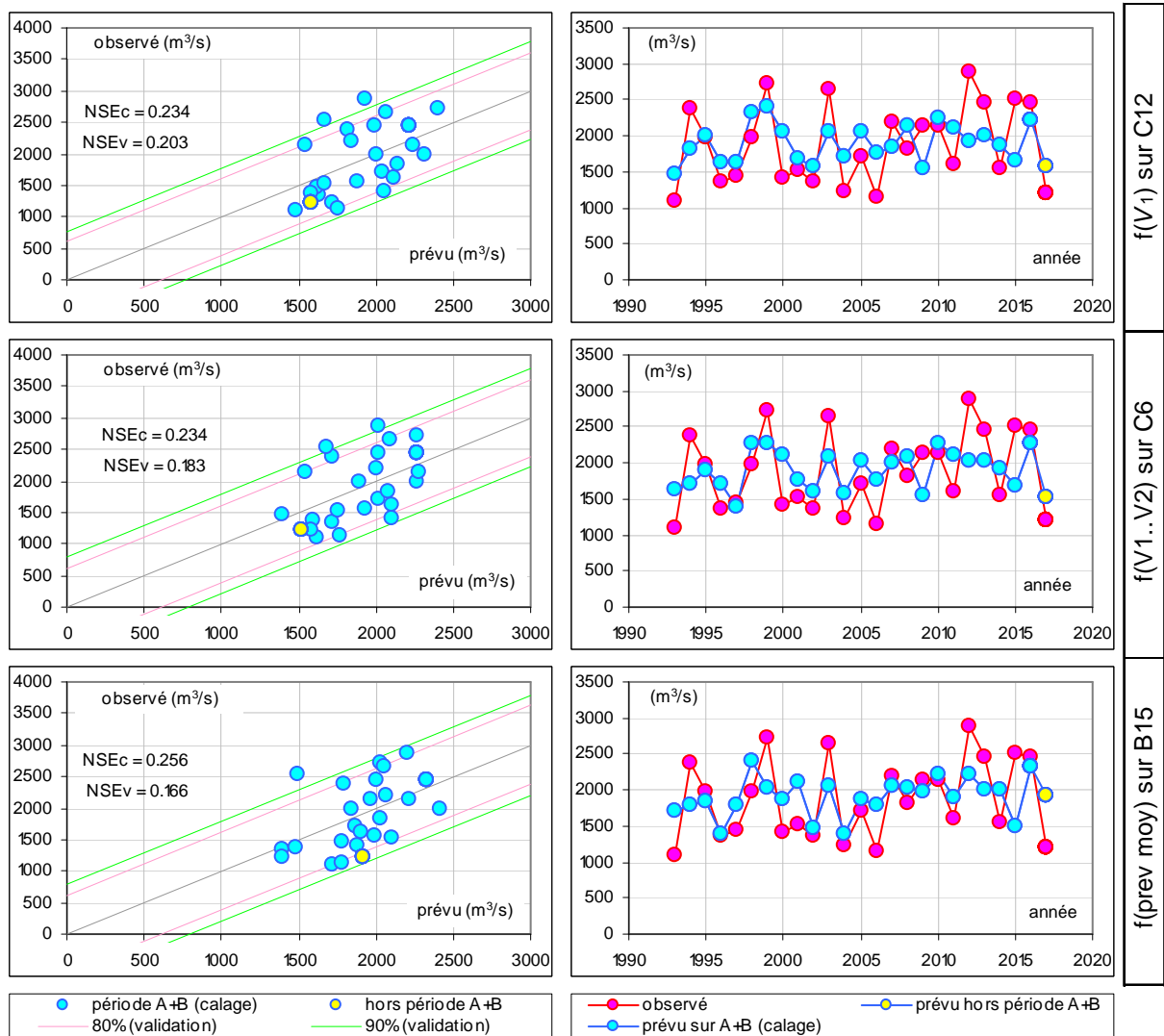


Figure 3 : résultats des trois meilleurs modèles de prévision de débit ASO : comparaison des débits réels et modélisés (modèles calés sur période A+B (1993-2016))

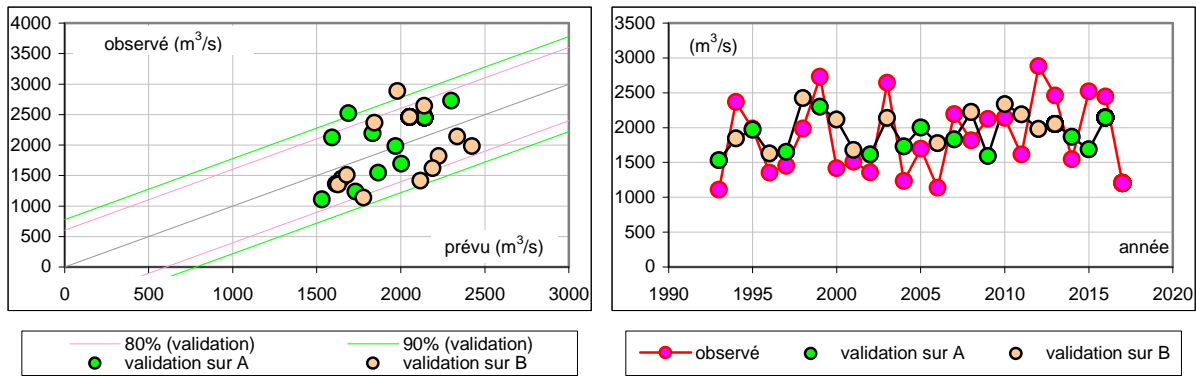


Figure 4 : résultats du meilleur modèle de prévision du débit ASO ($f(V_1)$ sur C12) : comparaison du débit réel avec le débit modélisé en validation (sur échantillon A avec le modèle calé sur B et sur échantillon B avec le modèle calé sur A)

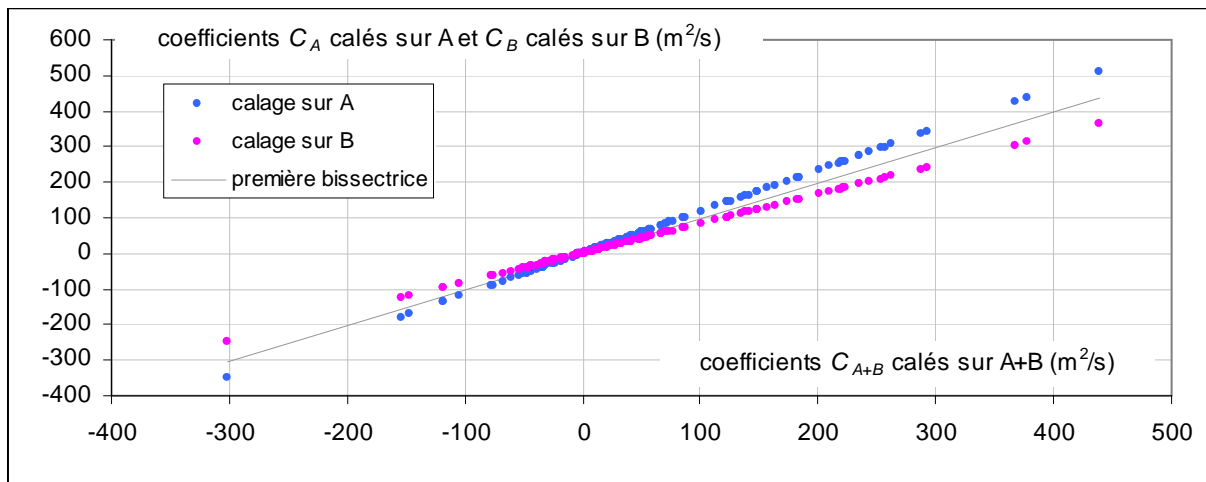


Figure 5 : coefficients multiplicatifs de régression reliant le débit aux prévisions pluviométriques de base, pour le meilleur modèle ASO ($f(V_1)$ sur la zone C12) : comparaison de C_A (147 valeurs calées sur échantillon A) et C_B (147 valeurs calées sur échantillon B) avec C_{A+B} (147 valeurs calées sur échantillon A+B)

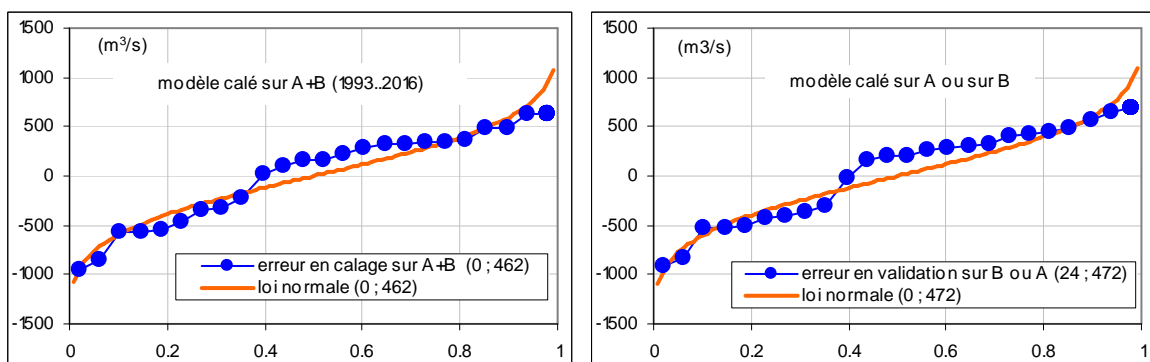


Figure 6 : fonction de répartition des erreurs de modélisation en calage sur l'échantillon A+B (1993-2016) et en validation (sur A avec le modèle calé sur B et sur B avec le modèle calé sur A), pour le meilleur modèle ASO ($f(V_1)$ sur zone C12)

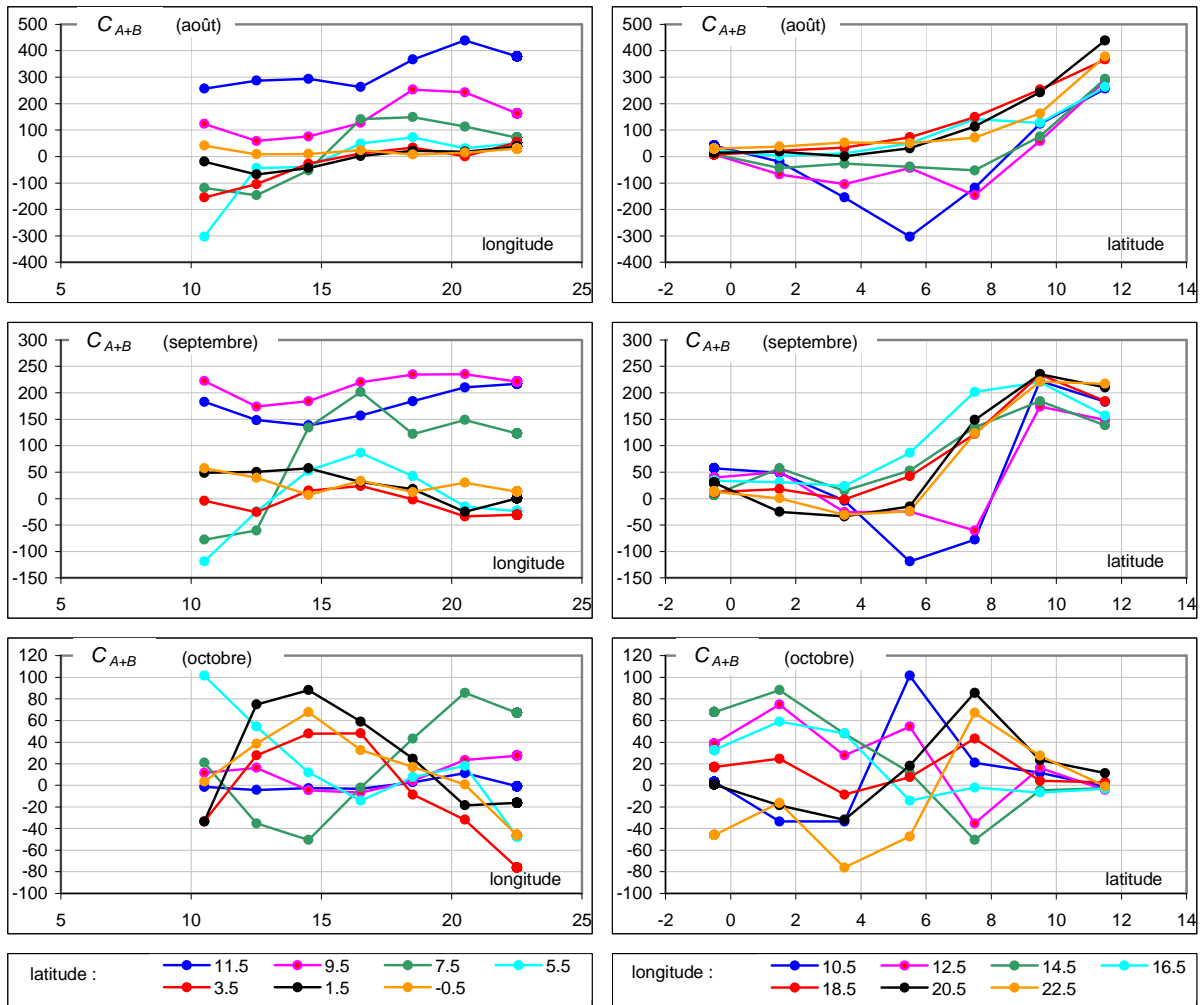


Figure 7 : Valeurs des coefficients de régression C_{A+B} (m^2/s) calés sur la période A+B (1993-2016) en fonction des coordonnées géographiques centrales des 49 mailles étendues de la zone C12 pour le meilleur modèle ASO ($f(V_1)$) sur zone C12)

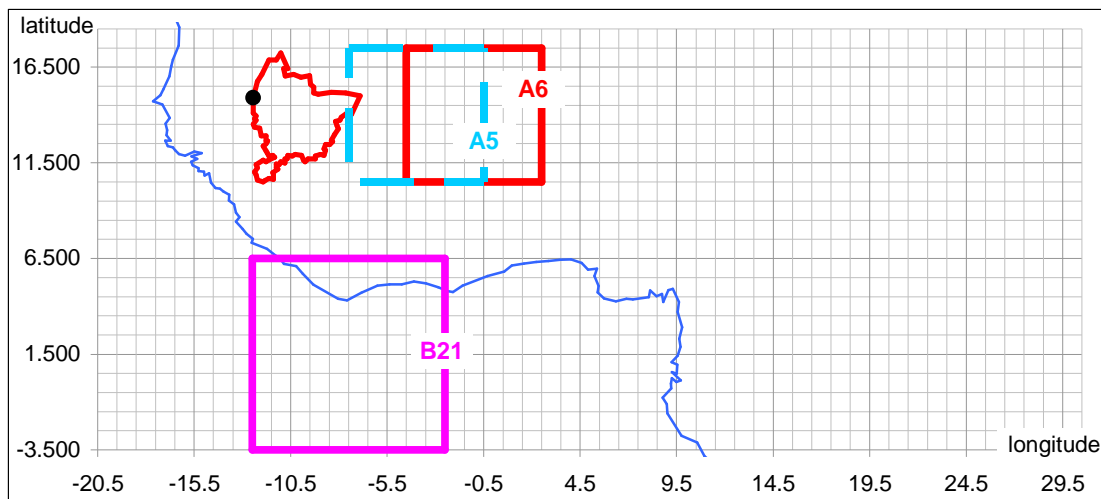


Figure 8 : zones A6, A5 et B21 associées aux trois meilleurs modèles de prévision SO

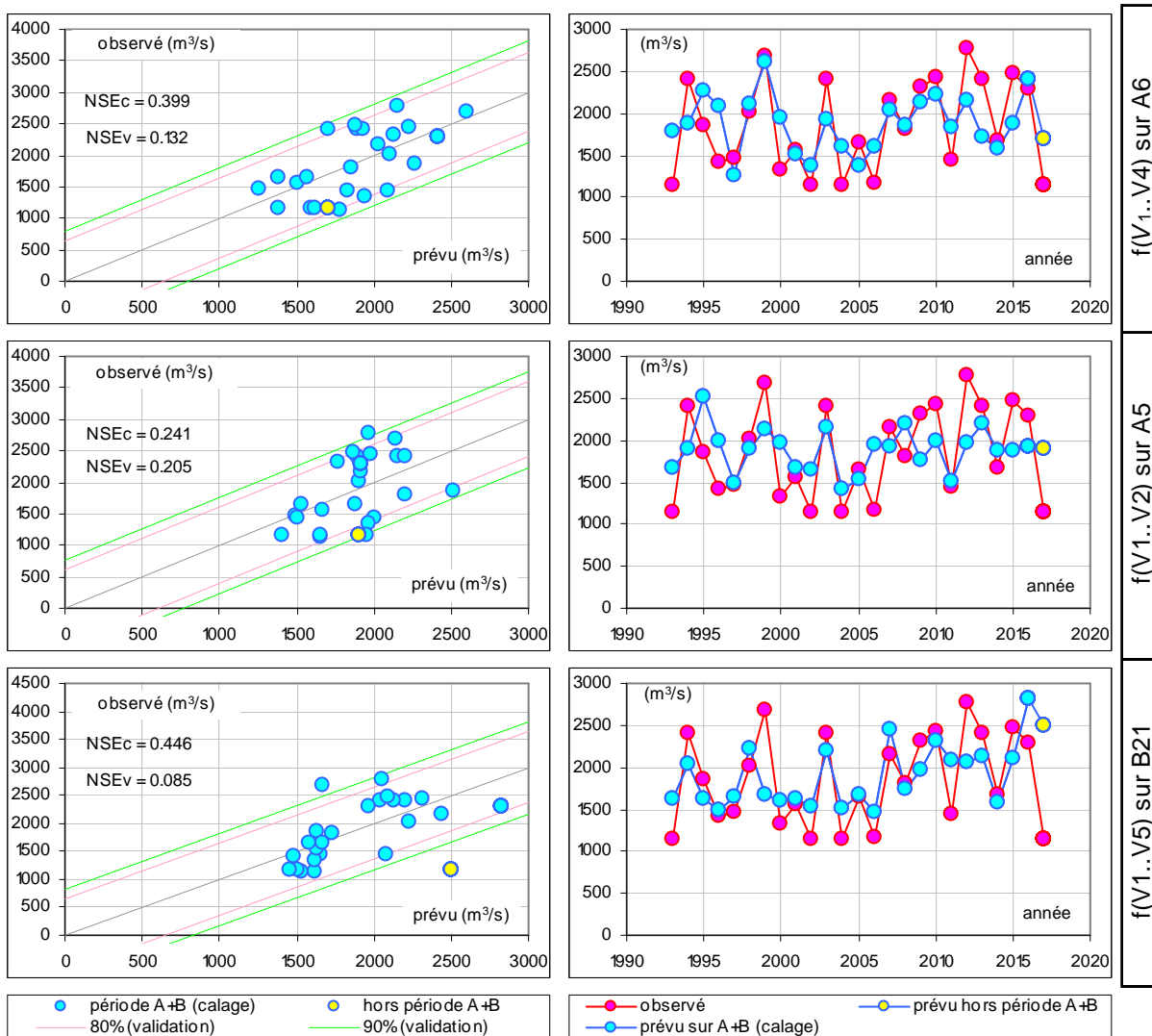


Figure 9 : résultats des trois meilleurs modèles de prévision de débit SO : comparaison des débits réels et modélisés (modèles calés sur période A+B (1993-2016))

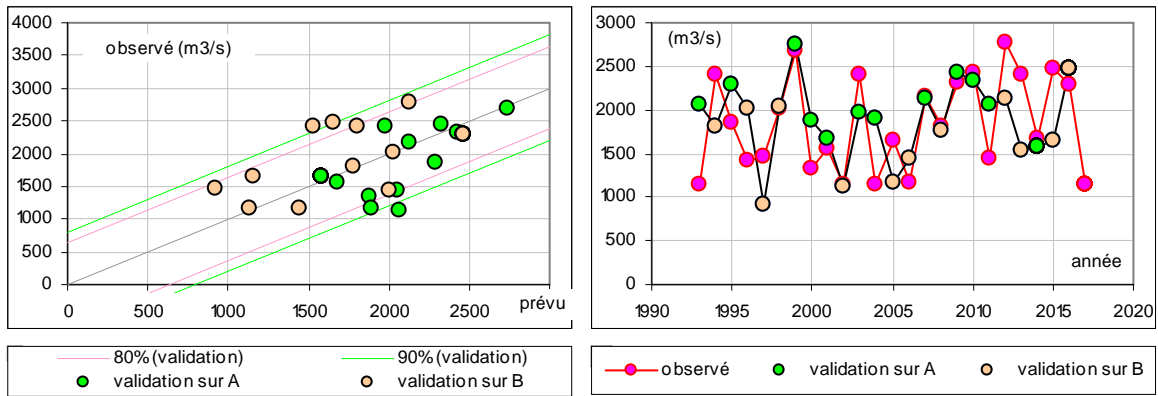


Figure 10 : résultats du meilleur modèle de prévision du débit SO ($f(V_1..V_4)$) sur A6) : comparaison du débit réel avec le débit modélisé en validation (sur échantillon A avec le modèle calé sur B et sur échantillon B avec le modèle calé sur A)

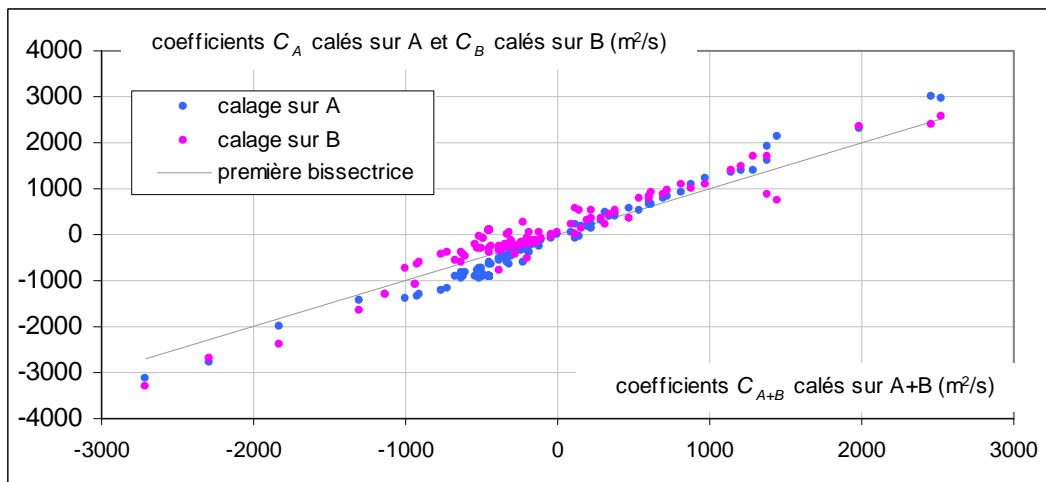


Figure 11 : coefficients multiplicatifs de régression reliant le débit aux prévisions pluviométriques de base, pour le meilleur modèle SO ($f(V_1..V_4)$) sur la zone A6) : comparaison de C_A (98 valeurs calées sur échantillon A) et C_B (98 valeurs calées sur échantillon B) avec C_{A+B} (98 valeurs calées sur échantillon A+B)

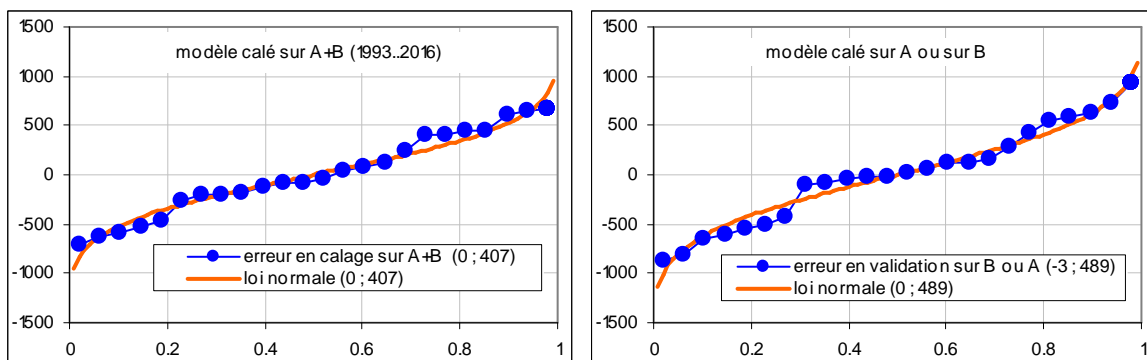


Figure 12 : fonction de répartition des erreurs de modélisation en calage sur l'échantillon A+B (1993-2016) et en validation (sur A avec le modèle calé sur B et sur B avec le modèle calé sur A), pour le meilleur modèle SO ($f(V_1..V_4)$) sur zone A6)

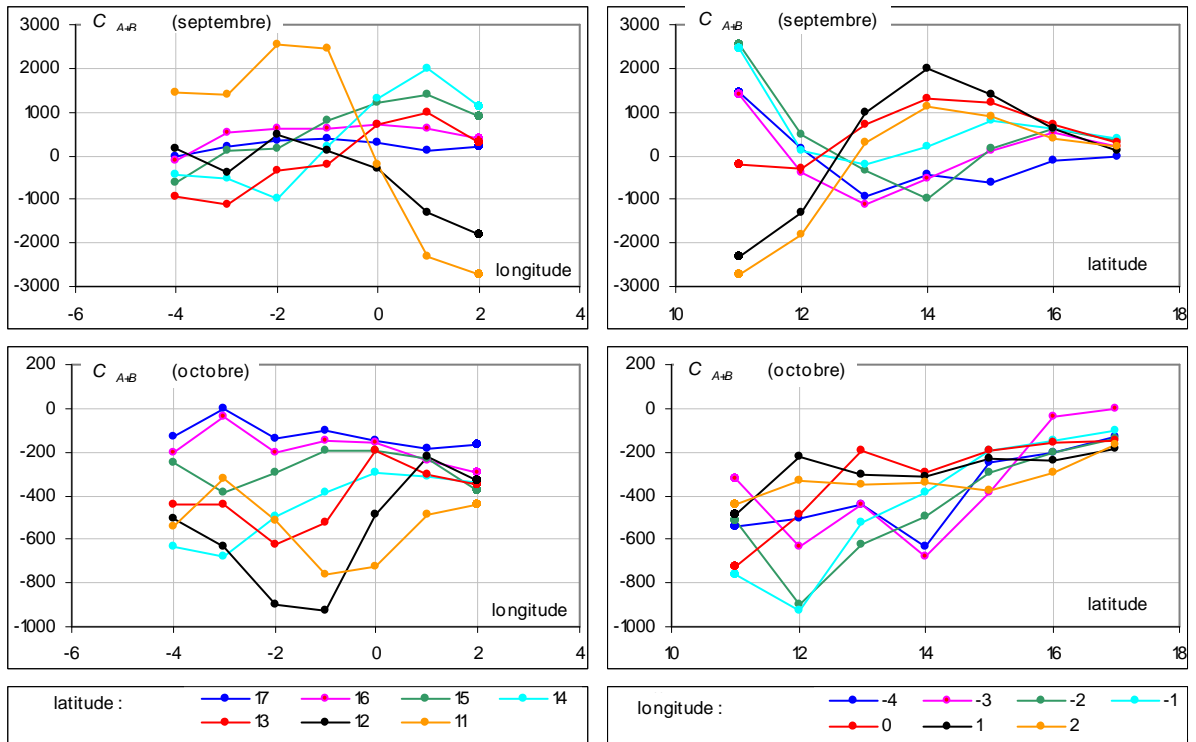


Figure 13 : Valeurs des coefficients de régression C_{A+B} calés sur la période A+B en fonction des coordonnées géographiques centrales des 49 mailles étendues de la zone A6 pour le meilleur modèle SO ($f(V_1..V_4)$ sur zone A6)