

Prévision 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

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
Montpellier, 9 Juin 2017

Table des matières

Résumé	3
Contexte	4
Origine et cadre de l'étude	4
Principes du modèle de prévision saisonnière d'écoulement et résultats obtenus précédemment	4
Objectif.....	5
Données	6
Prévisions saisonnières de pluie produites par ARPEGE système 5 sur l'Afrique de l'ouest.....	6
Débit naturel de crue du fleuve Sénégal à Bakel.....	6
Méthode.....	6
Regroupement des prévisions pluviométriques par zones	6
Variables utilisées	7
Années retenues pour le calage et la validation des modèles.....	7
Synthétisation des variables prévisionnaires, types de modèles testés	8
Procédure de calage et validation croisés.....	8
Détermination des meilleurs modèles	9
Résultats obtenus.....	10
Prévision début juillet du débit naturel d'août-septembre-octobre (ASO)	10
Prévision début août du débit naturel de septembre-octobre (SO).....	10
Conclusion.....	11
Références	12
Tableaux et figures	13

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 d'août à octobre (respectivement bimensuel de septembre à octobre), à partir des prévisions pluviométriques 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.

Avec des horizons de prévision de 1 à 4 mois visant la période août-septembre-octobre (ASO) et de 1 à 3 mois visant la période septembre-octobre (SO), les prévisions pluviométriques saisonnières produites par ARPEGE sur des mailles élémentaires de 0.75° de côté sont entachées à l'heure actuelle de grandes incertitudes en valeur et localisation. Leur traduction en prévision de débit, impossible dans ces conditions par modèle "pluie-débit" classique, ne peut être faite que par une adaptation statistique qui les exploite sur une zone pouvant déborder largement du bassin hydrologique concerné.

On présente ici la mise au point de modèles de prévision saisonnière de débit ASO et SO basés sur l'adaptation statistique des résultats d'ARPEGE système 5, exploités en valeurs brutes de lame précipitée ou en valeurs centrées réduites. Ces modèles, destinés à être utilisés en temps réel début juillet et début août dès 2017, sont chacun déterminés par la procédure suivante à partir des données de la période 1991-2016 :

- regroupement des prévisions pluviométriques par zones : 81 zones carrées ou rectangulaires, 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 ;
- synthétisation des prévisions pluviométriques (prévisseurs de base) en un nombre restreint de variables sur chaque zone : les variables prévisseurs synthétiques sont les 8 premiers vecteurs propres V_1 à V_8 des variables prévisseurs de base, déterminés par analyse en composantes principales sur la période 1991-2015, ainsi que les prévisions pluviométriques moyennées sur la zone ;
- dissociation des années 1991 à 2015 en deux sous-échantillons complémentaires A et B de 12 et 13 années respectivement, représentant chacun du mieux possible la diversité des débits naturels réellement observés de 1991 à 2015;
- 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 1991-2015 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 1991-2015, ainsi que le modèle de type 2 ; à chacun de ces deux modèles sont associés les paramètres $NSEc$, $NSEv$ et $NSEf$ (de type "Nash and Sutcliffe coefficient of model efficiency") caractérisant respectivement la précision des débits modélisés en calage, la précision des débits modélisés en validation et la stabilité des coefficients de régression du modèle en fonction de l'échantillon de calage ;

- le meilleur des 162 modèles analysés est déterminé par un paramètre $NSEg$ correspondant à une combinaison linéaire des paramètres $NSEc$, $NSEv$ et $NSEf$, qui donne le plus grand poids aux performances en validation ;
- 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 1991-2015) sur l'année 2016.

Pour la prévision début juillet du débit d'août-septembre-octobre (ASO), le meilleur modèle, de type 1, exploite les résultats produits par ARPEGE sur une zone C5 centrée très à l'est du bassin du Sénégal. Il est calé sur les 3 premiers vecteurs propres des prévisions pluviométriques de base sur cette zone. Ses performances sont modestes en calage ($NSEc = 0.380$) et surtout en validation ($NSEv = 0.188$), mais ses coefficients de régression sont assez stables ($NSEf = 0.895$) et il donne de très bons résultats dans les conditions du "temps réel" pour l'année 2016 (prévision dans l'intervalle de confiance à 11% du modèle).

Pour la prévision début août du débit de septembre-octobre (SO), le meilleur modèle, de type 1, exploite les résultats produits par ARPEGE sur une zone D6 centrée au sud du bassin du Sénégal mais englobant la quasi-totalité de celui-ci. Il est calé sur les 6 premiers vecteurs propres des prévisions pluviométriques de base sur cette zone. Ses performances sont bonnes en calage ($NSEc = 0.656$) et très bonnes en validation ($NSEv = 0.528$). Ses coefficients de régression sont stables ($NSEf = 0.942$) et il donne des résultats corrects dans les conditions du "temps réel" pour l'année 2016 (prévision dans l'intervalle de confiance à 84% du modèle).

Les résultats obtenus, très encourageants pour SO, permettent d'utiliser les deux modèles en temps réel dès 2017 pour la prévision saisonnière de débit, avec des intervalles de confiance très large pour ASO et assez étroit pour SO, directement estimés à partir des incertitudes calculées en validation.

Contexte

Origine et cadre de l'étude

La prévision saisonnière des écoulements naturels de crue du fleuve Sénégal à Bakel peut apporter une aide à la décision pour la gestion du barrage de Manantali, implanté à l'amont au Mali sur le Bafing, principale branche mère du fleuve. Ceci est démontré (Lamagat et al, 1999) pour l'un des objectifs de gestion de l'ouvrage, à savoir le soutien de la crue du fleuve, qui nécessite un volume lâché fortement lié à l'hydraulicité naturelle du fleuve à Bakel durant les mois de septembre et octobre.

C'est pourquoi l'IRD, Météo France et l'OMVS collaborent de 2000 à 2015 pour mettre au point et exploiter un modèle produisant cette prévision saisonnière des écoulements, à partir des prévisions saisonnières de pluie calculées sur l'Afrique de l'ouest par le modèle ARPEGE de Météo France. Cette étude fait l'objet en 2005 d'un accord de coopération entre Météo France et l'OMVS, achevé en 2015.

Les trois organismes décident de poursuivre l'étude à partir de 2017, dans le cadre d'un nouvel accord de coopération entre Météo France et l'OMVS d'une part, et d'une convention de travaux entre l'IRD et l'OMVS d'autre part.

Principes du modèle de prévision saisonnière d'écoulement et résultats obtenus précédemment

Le modèle de prévision saisonnière d'écoulement précédemment mis au point est basé sur une adaptation statistique des prévisions pluviométriques mensuelles, calculées à l'horizon de 1 à 4 mois par le modèle ARPEGE sur une grille dont les mailles couvrent l'Afrique de l'ouest. Cette adaptation, dont les principes sont décrits par Bader et al (2006) et qui doit être refaite

pour chaque nouvelle version d'ARPEGE et chaque horizon de prévision envisagé, consiste principalement à déterminer au sein de la grille une zone élémentaire optimale constituée de mailles contiguës sur lesquelles les prévisions de pluie présentent une combinaison linéaire la mieux reliée possible au débit à prévoir.

En pratique de 2005 à 2015, ce sont les prévisions pluviométriques produites par cinq versions successives d'ARPEGE sur deux maillages différents qui sont utilisées en valeurs centrées réduites pour mettre au point et exploiter en temps réel les modèles de prévision saisonnière d'écoulement suivants :

- débit naturel bimensuel de la période septembre-octobre (SO) calculé à partir des prévisions de pluie mensuelles produites début août pour le mois de septembre et pour le mois d'octobre ;
- débit naturel tri-mensuel de la période août-septembre-octobre (ASO) prévu à partir des prévisions mensuelles de pluie calculées début juillet pour le mois d'août, pour le mois de septembre et pour le mois d'octobre.

Pour chacune des sept adaptations réalisées (cinq pour SO et deux pour ASO), la zone prédictive optimale est déterminée ainsi à quelques variantes près, en testant différentes zones carrées ou rectangulaires situées à l'intérieur de la grille couvrant l'Afrique de l'ouest, qui se distinguent selon leurs forme, taille et position :

- pour chaque zone testée, constituée de X mailles, on détermine sur un échantillon d'années de calage les N premiers vecteurs propres V_1 à V_N ($N = 5$ ou 8 selon l'adaptation réalisée) des Y variables de prévisions de pluie ($Y = 2X$ pour SO et $Y = 3X$ pour ASO) ;
- toujours pour chaque zone testée, on détermine par régression descendante parmi V_1 à V_N , les n vecteurs propres des prévisions de pluie dont une combinaison linéaire correspond au débit à prévoir avec la plus petite erreur type possible sur l'échantillon de calage. Cette combinaison linéaire déterminée sur les années de calage constitue le modèle de prévision du débit associé aux prévisions de pluie de la zone ;
- on sélectionne parmi les zones testées, celle où le modèle de prévision de débit figure parmi les meilleurs en calage et donne de bons résultats sur un échantillon d'années de validation, toutes distinctes des années de calage. C'est ce modèle qui est retenu pour l'utilisation en temps réel.

Les différents modèles ainsi mis au point sont chacun utilisés en temps réel sur une période qui dure de 1 à 5 ans selon les cas. Les résultats obtenus (Bader 2016) montrent que :

- la zone où les prévisions de pluie sont les plus prédictives pour le débit du Sénégal varie selon la version d'ARPEGE utilisée et selon la prévision visée (SO ou ASO). Par rapport au bassin du Sénégal à Bakel, elle se situe le plus souvent dans un secteur compris entre l'est et le sud-est et à une distance de 1000 à 2000 km environ ;
- les prévisions faites en temps réel montrent des erreurs beaucoup plus fortes qu'en période de calage et tendent globalement à sous-estimer les débits. Ceci montre un certain manque de robustesse des modèles mis au point.

Objectif

Les résultats du modèle ARPEGE système 5, déjà exploités en 2015 pour la mise au point et l'utilisation en temps réel de deux modèles de prévision de débit (SO et ASO), font ici l'objet d'une nouvelle adaptation statistique visant à mettre au point les modèles de prévision de débit utilisables dès l'hivernage de 2017. Cette nouvelle adaptation est justifiée par :

- la possibilité d'utiliser désormais les prévisions mensuelles produites par ARPEGE en valeurs de lame précipitée, à la place de leurs valeurs centrées réduites ;

- l'utilisation des résultats produits par ARPEGE sur un maillage plus fin (mailles de 0.75° de côté au lieu de 1.4°) ;
- l'amélioration de la méthode précédemment utilisée, avec l'utilisation d'une procédure de calage et validation croisés à la place de la procédure de calage / validation simple. Cette nouvelle procédure permet en principe de sélectionner des modèles plus robustes.

Données

Prévisions saisonnières de pluie produites par ARPEGE système 5 sur l'Afrique de l'ouest

Les prévisions pluviométriques sont fournies par Météo-France pour 1972 mailles élémentaires de 0.75° de côté couvrant une grille située entre les latitudes -3.375° et 18.375° et entre les longitudes -20.625° et 30.375° (29 mailles dans la direction nord-sud et 68 dans la direction est-ouest, voir fig. 1). Ces prévisions, qui couvrent les années 1991 à 2016, 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 donnés) consiste en plusieurs valeurs de lame précipitée mensuelle (15 valeurs avant 2015 et 50 valeurs à partir de 2015), 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 1991 à 2016 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, 81 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 5 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, dont chacune correspond soit à une maille élémentaire de 0.75° de côté, soit au regroupement de plusieurs mailles élémentaires (nécessité par certaines limitations imposées par le tableur utilisé pour les analyses) :

- les 28 zones testées de type A sont des carrés constitués de 9×9 mailles carrées, ces dernières correspondant chacune à une maille élémentaire. Les différentes zones de type A se distinguent par des décalages de 3 mailles dans la direction est-ouest et dans la direction nord-sud ;

- les 28 zones testées de type B sont des carrés constitués de 7X7 mailles carrées, ces dernières correspondant chacune au regroupement de 4 mailles élémentaires. Les différentes zones de type B se distinguent par des décalages de 2 mailles dans la direction est-ouest et dans la direction nord-sud ;
- les 12 zones testées de type C sont des carrés constitués de 9X9 mailles carrées, ces dernières correspondant chacune au regroupement de 4 mailles élémentaires. Les différentes zones de type C se distinguent par des décalages de 4 mailles dans la direction est-ouest et dans la direction nord-sud ;
- les 10 zones testées de type D sont des carrés constitués de 8X8 mailles carrées, ces dernières correspondant chacune au regroupement de 9 mailles élémentaires. Les différentes zones de type D se distinguent par des décalages de 3 mailles dans la direction est-ouest et 1 maille dans la direction nord-sud ;
- les 3 zones testées de type E sont des rectangles constitués de 12X7 mailles carrées, allongés dans la direction est-ouest, dont chaque maille correspond au regroupement de 16 mailles élémentaires. Les différentes zones de type E se distinguent par des décalages de 2 mailles dans la direction est-ouest.

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 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 carrées 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 de celles-ci, avec $Y = 2X$ pour les prévisions SO et $Y = 3X$ pour les prévisions ASO :

- pour les zones de type A : $Y = 243$ pour ASO et $Y = 162$ pour SO ;
- pour les zones de type B : $Y = 147$ pour ASO et $Y = 98$ pour SO ;
- pour les zones de type C : $Y = 243$ pour ASO et $Y = 162$ pour SO ;
- pour les zones de type D : $Y = 192$ pour ASO et $Y = 128$ pour SO ;
- pour les zones de type E : $Y = 252$ pour ASO et $Y = 168$ pour SO ;

Années retenues pour le calage et la validation des modèles

La période commune des données disponibles s'étend de 1991 à 2016 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 26 valeurs.

De façon arbitraire, les données de l'année 2016 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 1991-2015, avec un échantillon de 25 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 quasi équivalents (12 années pour A et 13 années pour B). La répartition des années de la période A+B (1991-2015) entre A et B est effectuée de façon que les diverses valeurs réellement observées du débit naturel entre 1991 et 2015

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, 14, 16, 18, 20, 22, 24 ;
- échantillon B : rangs 2, 4, 6, 8, 10, 12, 13, 15, 17, 19, 21, 23, 25.

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 (supérieur ou égal à 98) dépasse largement l'effectif total de l'échantillon disponible (26 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_{sA} . 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 Ev_A ;

- les erreurs de validation E_{vB} et E_{vA} 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 $E_{v_{A+B}}$, 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 :

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

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) et pour un type de valeurs utilisées pour les variables prévisseurs de base (valeurs brutes de lame précipitée mensuelle ou valeur centrée réduite, voir plus bas), l'analyse décrite ci-dessus produit les indicateurs de performance NSE_v , NSE_c et NSE_f pour 162 modèles différents (1 modèle de type 1 et un modèle de type 2 par zone, pour 81 zones testées). Pour comparer ces modèles entre eux et déterminer le plus performant, on utilise un critère global unique NSE_g correspondant à une combinaison linéaire de NSE_v , NSE_c et NSE_f , dont les coefficients arbitrairement choisis attribuent le plus grand poids aux performances en validation :

$$NSE_g = (NSE_v + 0.4NSE_c + 0.1NSE_f) / 1.5 \quad [4]$$

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

Résultats obtenus

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

Pour cette prévision, la recherche du meilleur modèle est faite en utilisant sous deux formes différentes les prévisions pluviométriques relatives aux mailles carrées constituant les zones : les valeurs brutes de lame précipitée mensuelle exprimées en mètre, d'une part, et les valeurs centrées réduites (référencées à la période A+B) de ces valeurs brutes.

Le tableau 2 résume les résultats obtenus pour les 10 meilleurs modèles de chacune de ces deux catégories. On constate que les meilleurs résultats sont obtenus en utilisant les valeurs brutes de prévisions pluviométriques plutôt que leurs valeurs centrées réduites.

Parmi tous les modèles testés, trois présentent des performances nettement meilleures que les autres, basées sur l'utilisation des valeurs brutes de prévisions pluviométriques sur les zones C5, C12 et D10 (fig. 2), situées très à l'est/sud-est du bassin du fleuve Sénégal :

- modèle de type 1 $f(V_1-V_3)$ sur la zone C5 comprise entre les latitudes 4.875 et 18.375 et entre les longitudes 3.375 et 16.875 ;
- modèle de type 2 sur la zone C12 comprise entre les latitudes -1.125 et 12.375 et entre les longitudes 9.375 et 22.875 ;
- modèle de type 2 sur la zone D10 comprise entre les latitudes -1.875 et 16.125 et entre les longitudes 6.375 et 24.375 ;

Les performances modestes de ces trois meilleurs modèles sont très proches entre elles, avec des valeurs respectives de NSE_g de 0.286, 0.285 et 0.280 dont le classement est objectivement très lié aux pondérations arbitraires de NSE_c , NSE_v et NSE_f .

Pour les trois modèles, la figure 3 compare les débits réels aux débits modélisés avec le calage effectué sur l'échantillon A+B (1991-2015). 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 2016 dans les conditions du "temps réel" est très bon dans les trois cas, se situant respectivement dans l'intervalle de confiance à 11 %, 44% et 39%.

La figure 4 renseigne sur la robustesse du meilleur modèle ($f(V_1..V_3)$ sur zone C5), 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 243 valeurs (=9×9×3) des coefficients multiplicatifs de régression C_A et C_B calés respectivement sur les échantillons A et B. Ces coefficients sont reliés entre eux par une régression linéaire proche de la première bissectrice.

La figure 6 montre que pour le meilleur modèle ASO, les erreurs de modélisation se répartissent quasiment suivant une loi normale de moyenne nulle et d'écart type égal à $ET_{C_{A+B}}$ (= 420 m³/s) en calage et égal à $ET_{V_{A+B}}$ (= 480 m³/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 une variabilité en fonction de la longitude qui est forte pour les basses latitudes et faible pour les hautes latitudes.

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

Pour cette prévision, la recherche du meilleur modèle est faite en utilisant seulement les valeurs brutes des prévisions pluviométriques, qui sont des lames précipitées mensuelles exprimées en mètre sur chacune des mailles carrées constituant les zones testées.

Parmi tous les modèles SO testés, les dix meilleurs (tab. 2) présentent tous des résultats de meilleure qualité que le meilleur modèle ASO déterminé plus haut, avec NSE_C toujours supérieur à 0.408 et NSE_V toujours supérieur à 0.303. Les trois meilleurs modèles utilisent les prévisions pluviométriques sur les zones D6, A12 et A5 (fig.8), qui incluent le bassin du fleuve Sénégal à Bakel en partie ou en quasi-totalité :

- modèle de type 1 $f(V_1-V_6)$ sur la zone D6 comprise entre les latitudes -1.875 et 16.125 et entre les longitudes -20.625 et -2.625 ;
- modèle de type 2 sur la zone A12 comprise entre les latitudes 9.375 et 16.125 et entre les longitudes -10.875 et -4.125 ;
- modèle de type 2 sur la zone A5 comprise entre les latitudes 11.625 et 18.375 et entre les longitudes -10.875 et -4.125 ;

Pour ces trois modèles, la figure 9 compare les débits réels aux débits modélisés avec le calage effectué sur l'échantillon A+B (1991-2015). On constate que la totalité des débits modélisés sur la période A+B se situe pour chaque modèle 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). Mais le premier modèle, déjà classé au premier rang de tous les modèles sur la période A+B à la fois en calage ($NSE_C = 0.656$) et en validation ($NSE_V=0.528$, très bon résultat), est largement meilleur que ses deux suivants dans les conditions du temps réel, avec un débit modélisé en 2016 se situant dans l'intervalle de confiance à 84% au lieu de 99.9 % et 99.8% pour les deux autres.

La figure 10 renseigne sur la robustesse du meilleur modèle ($f(V_1..V_6)$ sur zone D6), 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 128 valeurs ($=8 \times 8 \times 2$) des coefficients multiplicatifs de régression C_A et C_B calés respectivement sur les échantillons A et B. Ces coefficients sont étroitement reliés par une régression linéaire très proche de la première bissectrice.

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 $ET_{C_{A+B}}$ ($= 310 \text{ m}^3/\text{s}$) en calage et égal à $ET_{V_{A+B}}$ ($= 364 \text{ m}^3/\text{s}$) 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 du sud-ouest vers le nord-est en août et de l'ouest vers l'est en octobre.

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 2017 sont basés sur les valeurs brutes de lame précipitée prévues par ARPEGE système 5. Ces modèles sont les suivants :

- pour ASO : $f(V_1-V_3)$ utilisant les prévisions pluviométriques sur une zone C5 située très à l'est du bassin du fleuve Sénégal, entre les latitudes 4.875 et 18.375 et entre les longitudes 3.375 et 16.875 ;
- pour SO : $f(V_1-V_6)$ utilisant les prévisions pluviométriques sur une zone D6 limitée par les latitudes -1.875 et 16.125 et les longitudes -20.625 et -2.625, qui englobe la quasi-totalité du bassin du fleuve Sénégal.

Grâce à la procédure très exigeante de calage et validation croisés utilisée pour sélectionner ces modèles, ceux-ci sont caractérisés par une assez bonne stabilité ($NSE_f = 0.895$ et 0.942 resp.) de leurs paramètres de traduction des prévisions pluviométriques en débit.

Cette stabilité montre une certaine robustesse des modèles et permet raisonnablement d'estimer leurs intervalles de confiance en temps réel à partir de leurs incertitudes déterminées en validation (répartition quasi normale des erreurs ; $ET_{v_{A+B}} = 480 \text{ m}^3/\text{s}$ et $NSE_v = 0.188$ pour ASO ; $ET_{v_{A+B}} = 364 \text{ m}^3/\text{s}$ et $NSE_v = 0.528$ pour SO). La simulation d'une utilisation en temps réel des deux modèles pour l'année 2016 donne des résultats cohérents avec ces incertitudes, avec des erreurs situées dans les intervalles de confiance à 11% pour ASO et 84% pour SO.

Au vu des résultats précédemment obtenus entre 2000 et 2015 et compte tenu de la méthode utilisée ici pour la sélection des modèles, les performances modestes du modèle ASO ne sont pas très étonnantes, alors que celles du modèle SO paraissent étonnamment bonnes. Ces différences de performances s'expliquent peut-être par l'horizon de prévision, beaucoup plus long pour ASO que pour SO.

Références

-Bader J.C., Piedelievre J.P., Lamagat J.P., 2006 : Prévision saisonnière du volume de crue du fleuve Sénégal: utilisation des résultats du modèle ARPEGE climat. Hydrol. Sci. J. 51(3) : 406-417.

-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., 2016 : Prévision saisonnière du débit naturel de crue du fleuve Sénégal à partir des prévisions pluviométriques du modèle de circulation générale de l'atmosphère ARPEGE. Evaluation des prévisions effectuées en temps réel de 2005 à 2015. IRD / G-EAU, Montpellier, 15 p.

-Lamagat J.P., Bader J.C., Vauchel P., 1999 : Programme d'optimisation de la gestion des réservoirs, phase 2, Synthèse (annexe 3.3), IRD/ OMVS, Dakar, Montpellier, 118p.

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
1991	1601	11	A	1513	10	B
1992	1059	1	A	1164	5	A
1993	1108	2	B	1136	1	A
1994	2368	20	A	2415	21	B
1995	1984	15	B	1859	15	B
1996	1352	5	A	1421	7	A
1997	1451	8	B	1459	9	A
1998	1984	16	A	2020	16	A
1999	2727	24	A	2679	24	A
2000	1418	7	A	1333	6	B
2001	1513	9	A	1555	11	A
2002	1360	6	B	1148	2	B
2003	2643	23	B	2408	20	A
2004	1236	4	B	1154	3	A
2005	1698	13	B	1656	12	B
2006	1138	3	A	1160	4	B
2007	2190	19	B	2153	17	B
2008	1820	14	A	1808	14	A
2009	2126	17	B	2312	18	A
2010	2140	18	A	2434	22	A
2011	1616	12	B	1436	8	B
2012	2884	25	B	2775	25	B
2013	2457	21	B	2407	19	B
2014	1545	10	B	1663	13	B
2015	2520	22	A	2474	23	B
2016	2454			2306		

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 1991 à 2015

prévisseurs de base	mois visés	zone utilisée		modèle	valeur				rang (sur 162)			
		nom	caractéristiques		<i>NSEc</i>	<i>NSEv</i>	<i>NSEf</i>	<i>NSEg</i>	<i>NSEc</i>	<i>NSEv</i>	<i>NSEf</i>	<i>NSEg</i>
valeurs brutes (m)	ASO	C5	Z(1.500X1.500;18.375;3.375;4.875;16.875)	f(V ₁ -V ₃)	0.380	0.188	0.895	0.286	1	3	21	1
		C12	Z(1.500X1.500;12.375;9.375;-1.125;22.875)	f(prev moy)	0.304	0.210	0.955	0.285	3	2	14	2
		D10	Z(2.250X2.250;16.125;6.375;-1.875;24.375)	f(prev moy)	0.261	0.217	0.994	0.280	4	1	3	3
		D5	Z(2.250X2.250;18.375;6.375;0.375;24.375)	f(prev moy)	0.241	0.173	0.990	0.245	7	4	4	4
		C6	Z(1.500X1.500;18.375;9.375;4.875;22.875)	f(V ₁)	0.212	0.116	0.988	0.199	8	5	6	5
		C5	Z(1.500X1.500;18.375;3.375;4.875;16.875)	f(prev moy)	0.256	0.070	0.924	0.176	5	8	19	6
		D10	Z(2.250X2.250;16.125;6.375;-1.875;24.375)	f(V ₁)	0.141	0.084	0.988	0.160	34	6	5	7
		D5	Z(2.250X2.250;18.375;6.375;0.375;24.375)	f(V ₁)	0.154	0.083	0.927	0.158	19	7	17	8
		E3	Z(3.000X3.000;18.375;-8.625;-2.625;27.375)	f(V ₁ -V ₂)	0.122	0.061	0.966	0.138	50	11	11	9
		C12	Z(1.500X1.500;12.375;9.375;-1.125;22.875)	f(V ₁)	0.139	0.052	0.978	0.137	35	13	8	10
valeurs centrées réduites	ASO	C6	Z(1.500X1.500;18.375;9.375;4.875;22.875)	f(prev moy)	0.277	0.160	0.742	0.230	2	2	34	1
		D5	Z(2.250X2.250;18.375;6.375;0.375;24.375)	f(prev moy)	0.181	0.160	0.990	0.221	7	1	2	2
		C12	Z(1.500X1.500;12.375;9.375;-1.125;22.875)	f(prev moy)	0.278	0.131	0.874	0.220	1	4	17	3
		C11	Z(1.500X1.500;12.375;3.375;-1.125;16.875)	f(prev moy)	0.167	0.149	0.977	0.209	14	3	5	4
		D10	Z(2.250X2.250;16.125;6.375;-1.875;24.375)	f(prev moy)	0.173	0.111	0.864	0.178	10	5	18	5
		D4	Z(2.250X2.250;18.375;-0.375;0.375;17.625)	f(V ₁ -V ₃)	0.205	0.067	0.795	0.152	4	7	26	6
		C6	Z(1.500X1.500;18.375;9.375;4.875;22.875)	f(V ₁)	0.105	0.080	0.940	0.144	72	6	7	7
		B28	Z(1.500X1.500;9.375;-2.625;-1.125;7.875)	f(V ₁ -V ₂)	0.079	0.063	0.947	0.126	85	8	6	8
		E2	Z(3.000X3.000;18.375;-14.625;-2.625;21.375)	f(V ₁ -V ₃)	0.197	0.033	0.750	0.125	5	14	33	9
		D10	Z(2.250X2.250;16.125;6.375;-1.875;24.375)	f(V ₁)	0.079	0.049	0.986	0.119	86	11	3	10
valeurs brutes (m)	SO	D6	Z(2.250X2.250;16.125;-20.625;-1.875;-2.625)	f(V ₁ -V ₆)	0.656	0.528	0.942	0.590	1	1	49	1
		A12	Z(0.750X0.750;16.125;-10.875;9.375;-4.125)	f(prev moy)	0.502	0.443	0.998	0.496	4	2	1	2
		A5	Z(0.750X0.750;18.375;-10.875;11.625;-4.125)	f(prev moy)	0.496	0.390	0.949	0.456	5	3	45	3
		D1	Z(2.250X2.250;18.375;-20.625;0.375;-2.625)	f(V ₁ -V ₄)	0.489	0.369	0.979	0.442	6	4	14	4
		B11	Z(1.500X1.500;15.375;-11.625;4.875;-1.125)	f(V ₁ -V ₃)	0.423	0.369	0.994	0.425	14	5	3	5
		A12	Z(0.750X0.750;16.125;-10.875;9.375;-4.125)	f(V ₁ -V ₃)	0.482	0.344	0.954	0.421	7	6	37	6
		C1	Z(1.500X1.500;18.375;-20.625;4.875;-7.125)	f(V ₁ -V ₆)	0.575	0.314	0.875	0.421	2	9	91	7
		A4	Z(0.750X0.750;18.375;-13.125;11.625;-6.375)	f(V ₁ -V ₃)	0.469	0.342	0.903	0.413	9	7	77	8
		B4	Z(1.500X1.500;18.375;-11.625;7.875;-1.125)	f(prev moy)	0.408	0.335	0.989	0.398	16	8	5	9
		A5	Z(0.750X0.750;18.375;-10.875;11.625;-4.125)	f(V ₁ -V ₂)	0.482	0.303	0.897	0.390	8	11	82	10

Tableau 2 : résultats obtenus avec les 10 meilleurs modèles par catégorie (caractéristiques des zones = dimensions de maille carrée en degrés, latitude et longitude de l'angle nord-est, latitude et longitude de l'angle sud-est ; modèle : f(V₁-V_N) pour type 1 et f(prev moy) pour type 2)

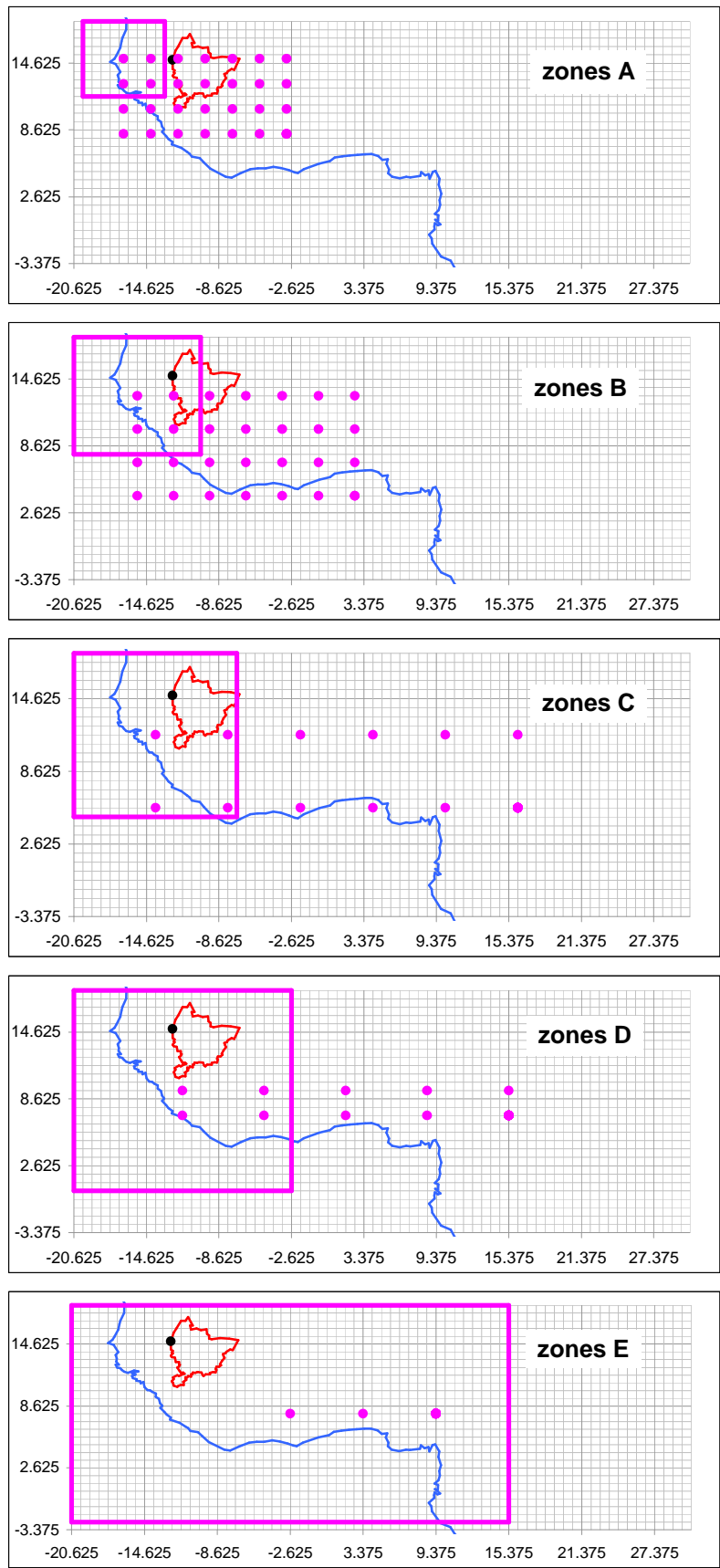


Figure 1 : types de zones testées sur la carte d'Afrique de l'ouest (gris : mailles élémentaires ; bleu : côte atlantique ; rouge : bassin versant du fleuve Sénégal à Bakel ; noir : Bakel ; points mauve : centres des zones ; traits mauve : zone située la plus au nord-est pour chaque type)

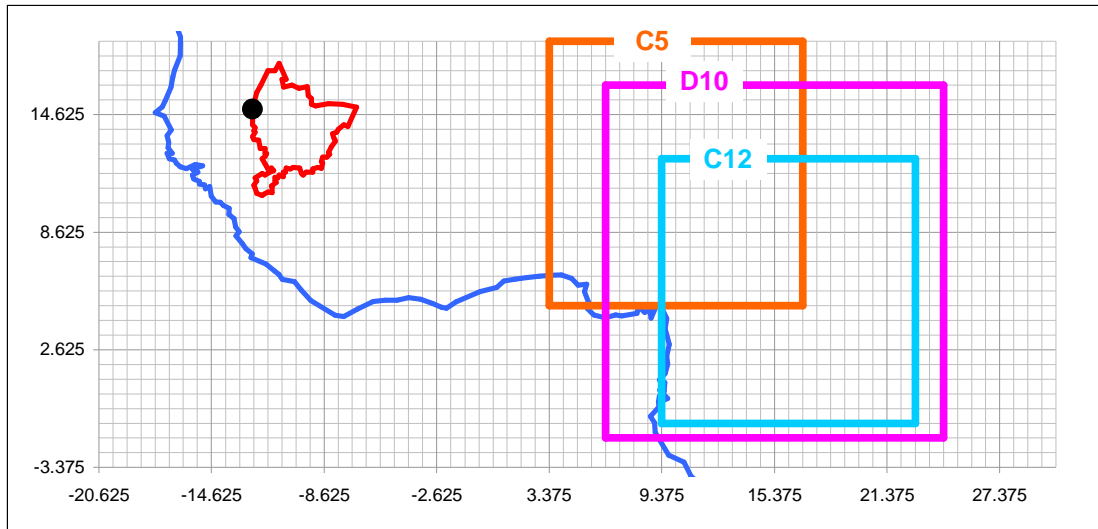


Figure 2 : zones C5, C12 et D10 associées aux trois meilleurs modèles de prévision ASO

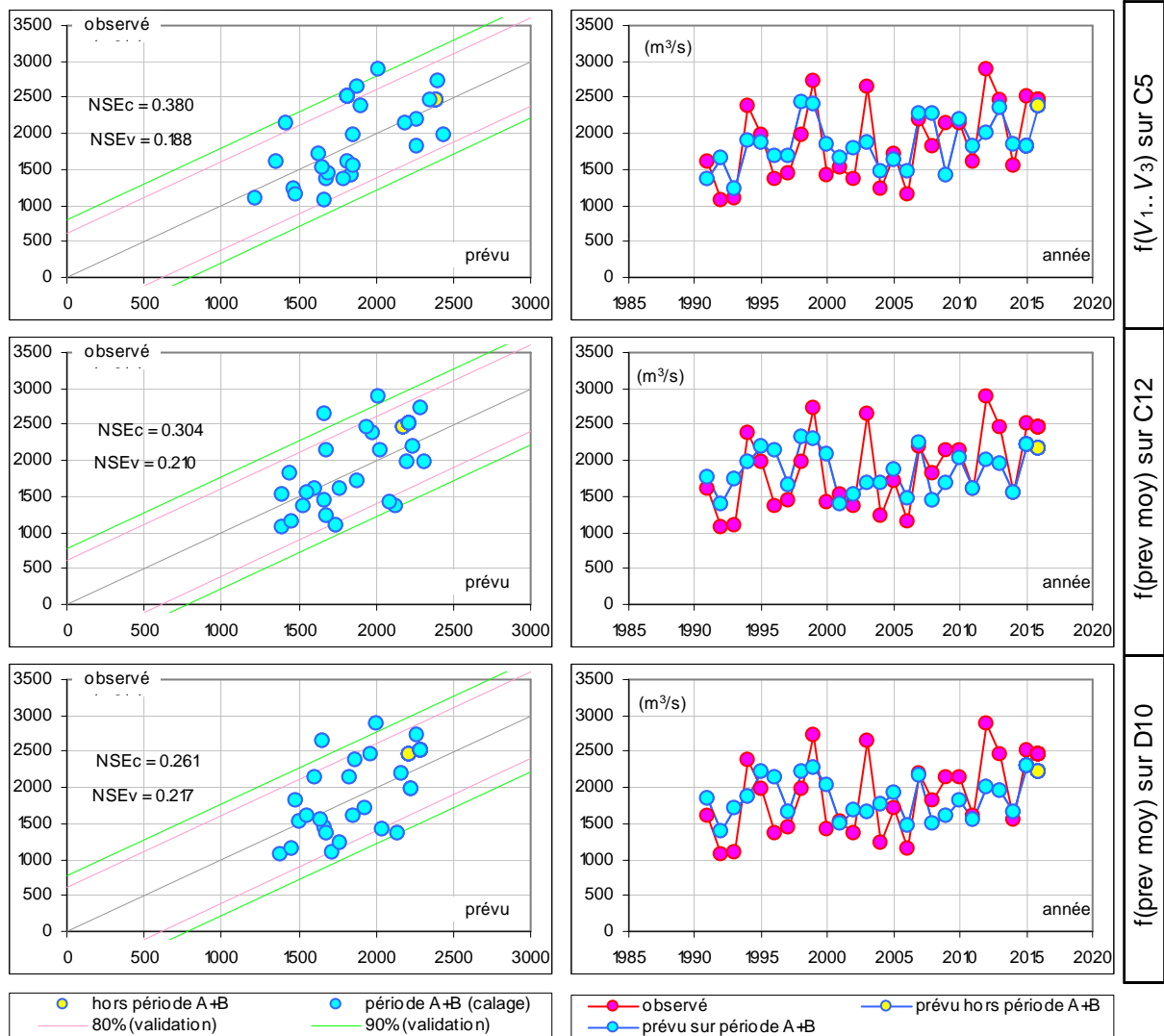


Figure 3 : résultats des trois meilleurs modèles de prévision de débit ASO : comparaison des débits réels et modélisés (modèles calés sur période A+B (1991-2015))

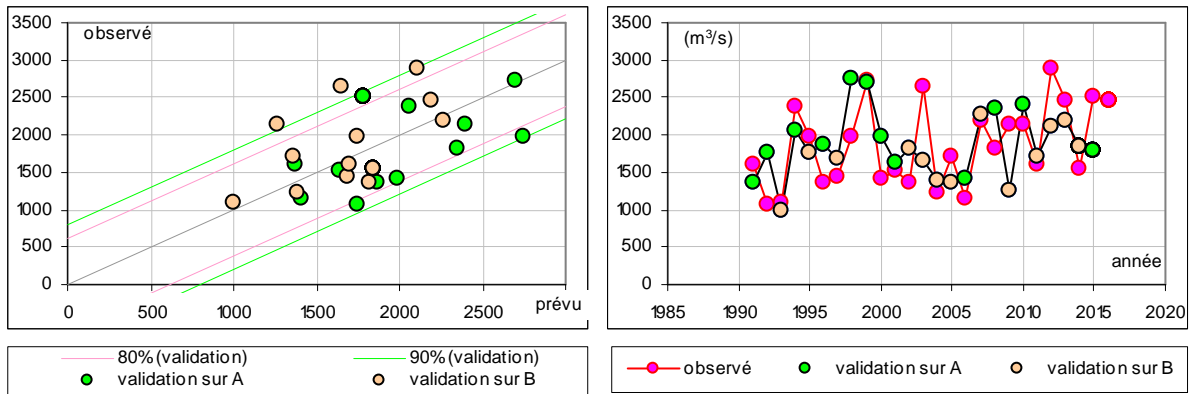


Figure 4 : résultats du meilleur modèle de prévision du débit ASO ($f(V_1..V_3)$) sur C5 : comparaison du débit réel avec le débit modélisé en validation (sur échantillon A avec le modèle calé sur B et sur échantillon B avec le modèle calé sur A)

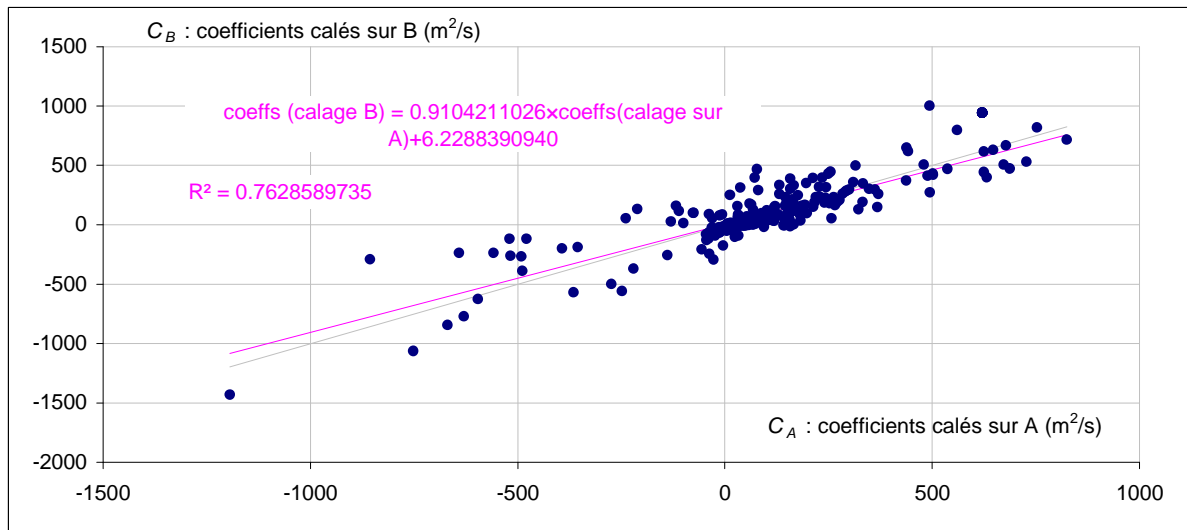


Figure 5 : comparaison des 243 coefficients multiplicatifs de régression C_A (calés sur échantillon A) et C_B (calés sur échantillon B) reliant le débit aux prévisions pluviométriques de base, pour le meilleur modèle ASO ($f(V_1..V_3)$) sur la zone C5)

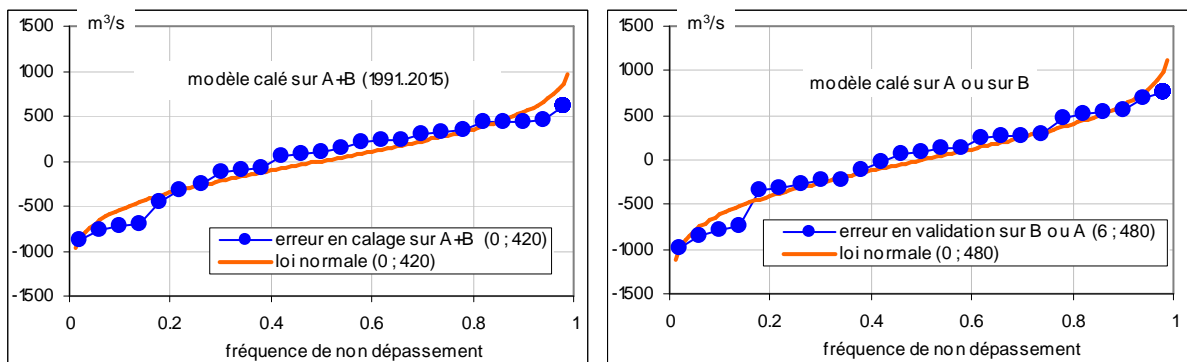


Figure 6 : fonction de répartition des erreurs de modélisation en calage sur l'échantillon A+B (1991-2015) et en validation (sur A avec le modèle calé sur B et sur B avec le modèle calé sur A), pour le meilleur modèle ASO ($f(V_1..V_3)$) sur zone C5)

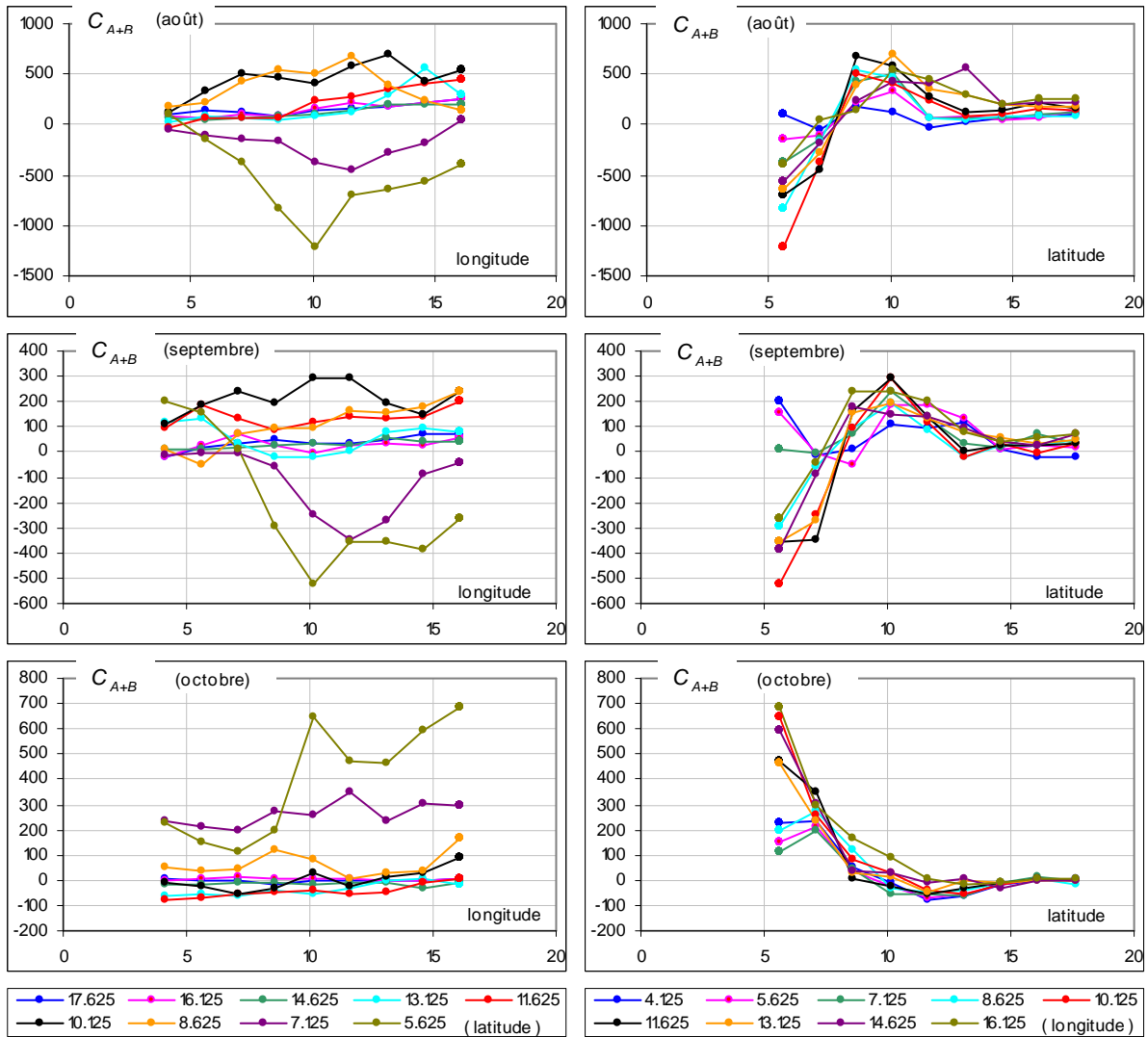


Figure 7 : Valeurs des coefficients de régression C_{A+B} (m^2/s) calés sur la période A+B en fonction des coordonnées géographiques centrales des 81 mailles carrées de la zone C5 pour le meilleur modèle ASO ($f(V_1..V_3$ sur zone C5)

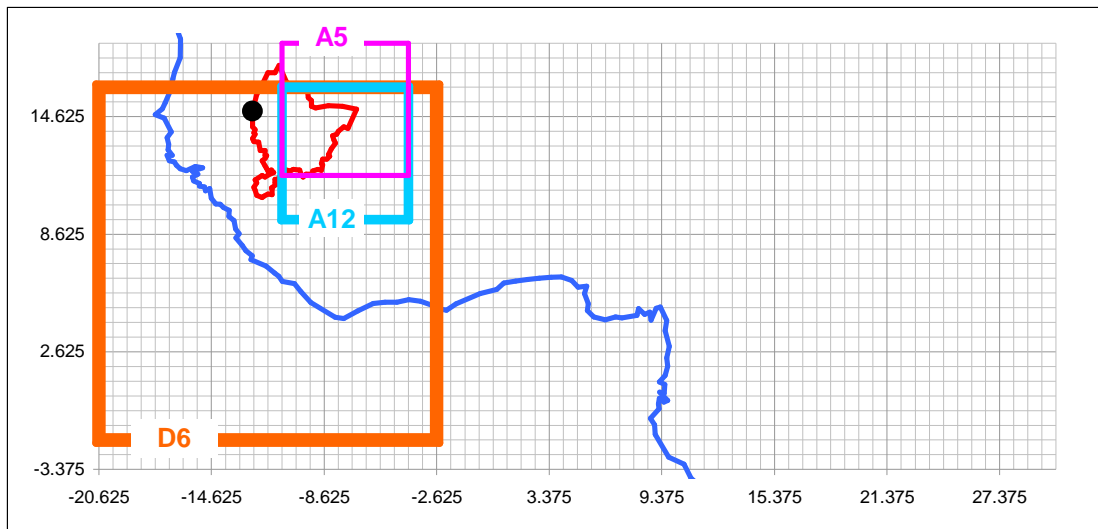


Figure 8 : zones D6, A12 et A5 associées aux trois meilleurs modèles de prévision SO

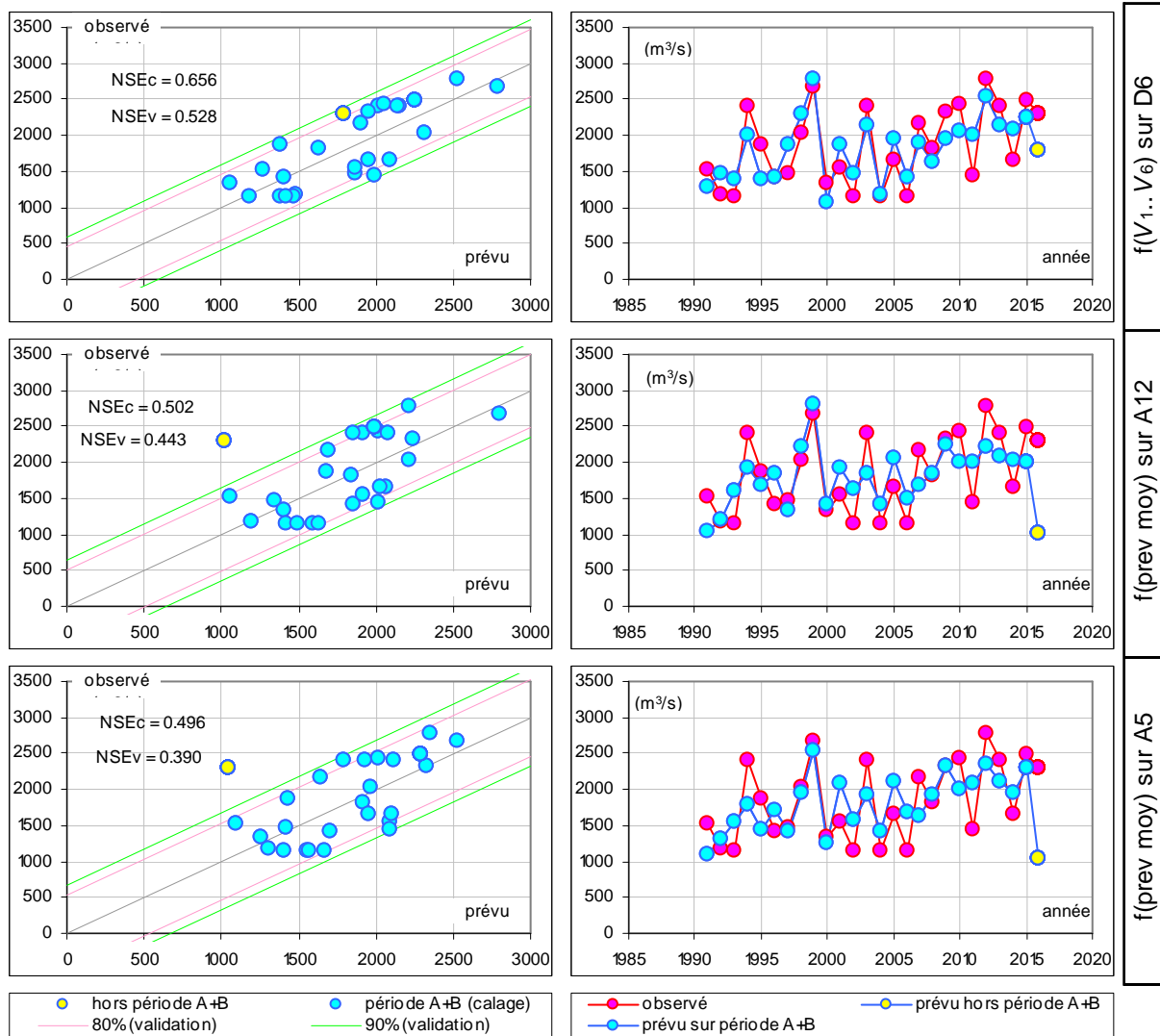


Figure 9 : résultats des trois meilleurs modèles de prévision de débit SO : comparaison des débits réels et modélisés (modèles calés sur période A+B (1991-2015))

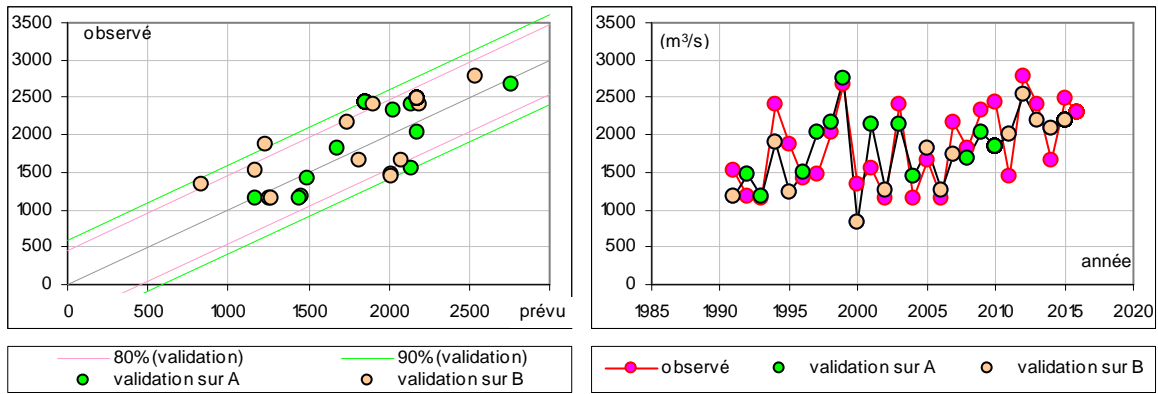


Figure 10 : résultats du meilleur modèle de prévision du débit SO ($f(V_1..V_6)$) sur D6 : comparaison du débit réel avec le débit modélisé en validation (sur échantillon A avec le modèle calé sur B et sur échantillon B avec le modèle calé sur A)

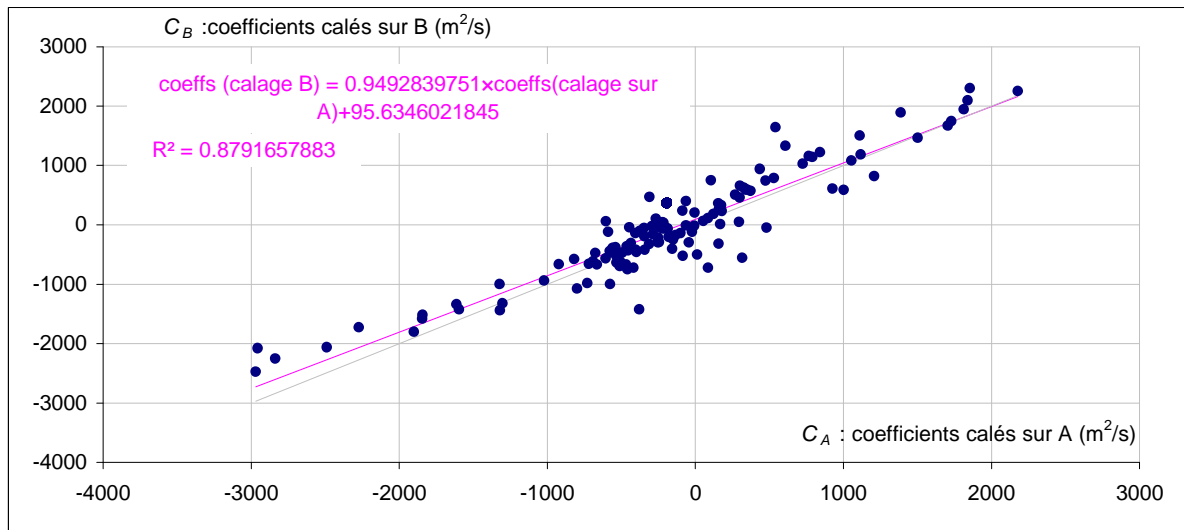


Figure 11 : comparaison des 128 coefficients multiplicatifs de régression C_A (calés sur échantillon A) et C_B (calés sur échantillon B) reliant le débit aux prévisions pluviométriques de base, pour le meilleur modèle SO ($f(V_1..V_6)$) sur la zone D6)

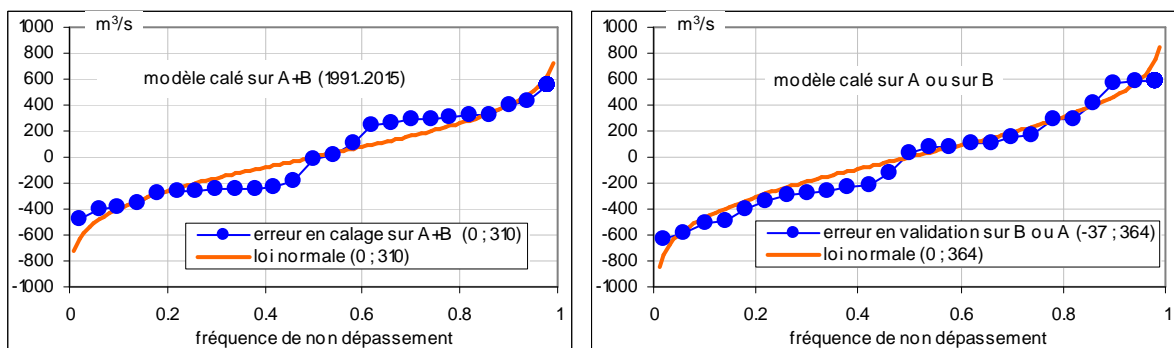


Figure 12 : fonction de répartition des erreurs de modélisation en calage sur l'échantillon A+B (1991-2015) et en validation (sur A avec le modèle calé sur B et sur B avec le modèle calé sur A), pour le meilleur modèle SO ($f(V_1..V_6)$) sur zone D6)

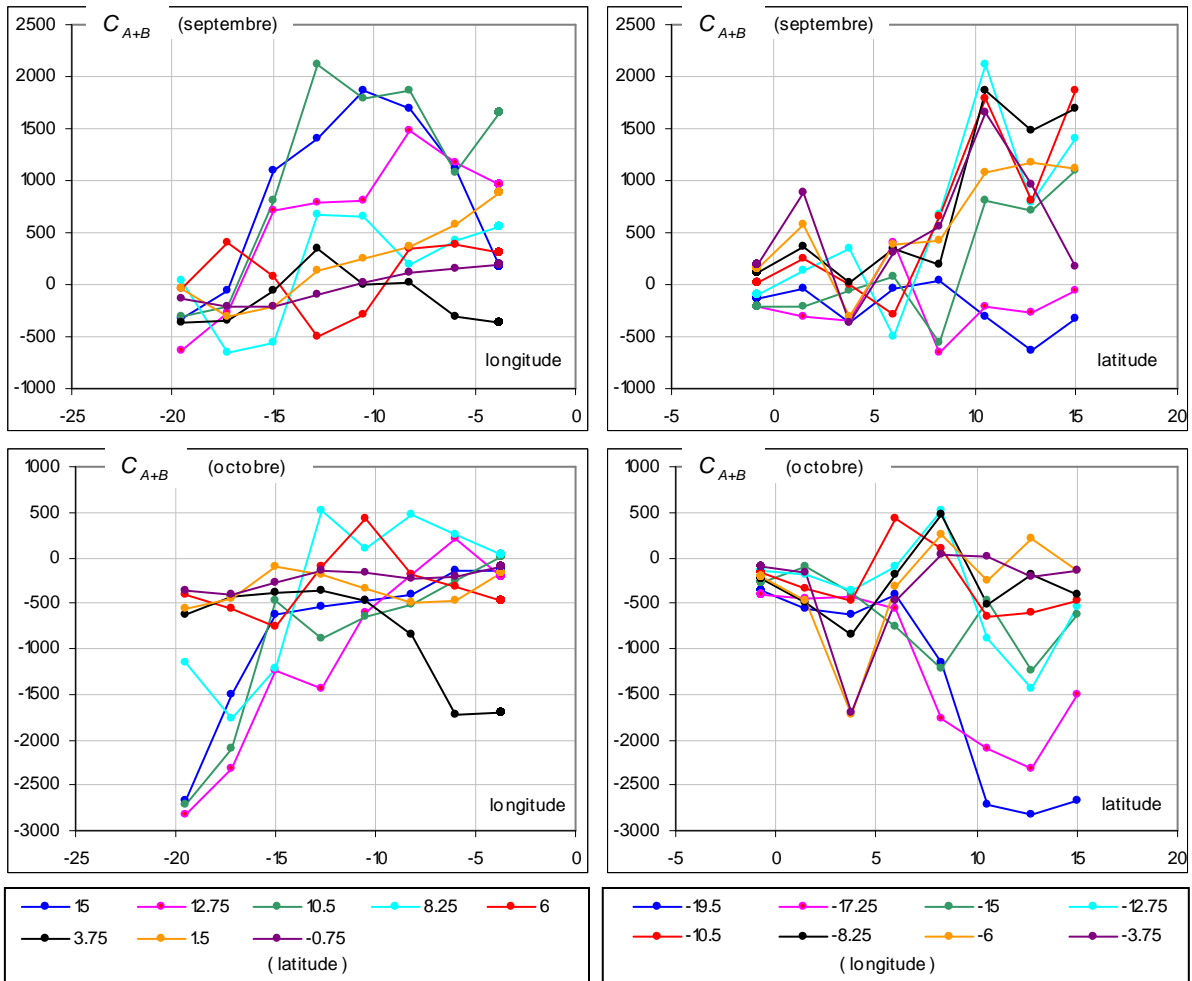


Figure 13 : Valeurs des coefficients de régression C_{A+B} calés sur la période A+B en fonction des coordonnées géographiques centrales des 64 mailles carrées de la zone D6 pour le meilleur modèle SO ($f(V_1..V_6$ sur zone D6)