

Cartographie des débits de référence – Interpolation des caractéristiques d'étiage

Rapport final

*Clotaire CATALOGNE & Eric SAUQUET
Irstea, Lyon*

Janvier 2012

Contexte de programmation et de réalisation

Ce rapport fait état des travaux de recherche sur l'estimation des variables d'étiage en sites peu ou pas jaugés réalisé au sein l'Unité de Recherche Hydrologie-Hydraulique du Cemagref (centre de Lyon).

L'étude présentée ici fait suite aux travaux d'estimation du QMNA5 réalisés dès 2008 sur le secteur pilote du bassin Seine-Normandie par Chopart et Sauquet (2008), puis sur les bassins de la Loire amont et de la Durance par pour une ébauche de la cartographie du QMNA5 à l'échelle nationale [Plasse et Sauquet (2009)]. Ces travaux, qui ont permis de mieux cibler les pistes de recherche en matière d'estimation des variables d'étiage en sites peu ou pas jaugés, ont été poursuivis au cours de ces deux dernières années avec pour objectif la mise en œuvre d'une démarche se voulant exhaustive, passant par la valorisation d'un maximum de données de débit disponible sur le réseau hydrographique national (chroniques courtes, jaugeages épisodiques).

L'effort porté en 2010 s'est concentré sur :

- l'analyse et la critique des stations hydrométriques disponibles dans la banque de données HYDRO pour la sélection d'un échantillon de référence réactualisé par rapport à celui de 2009 ;
- la caractérisation des régimes d'étiages en sites instrumentés au travers de descripteurs statistiques extraits des chroniques de débits ;
- la réactualisation des travaux de régionalisation du quantile de débit mensuel minimal annuel sec de période de retour 5 ans QMNA5, et du quantile de débit moyen sur 10 jours minimal annuel sec de période de retour 5 ans VCN10(5).

L'année 2011 a permis de conclure les travaux sur l'intercomparaison de différentes méthodes de définition de voisinages hydrologiques dans le cadre de l'estimation par régression multiple des variables QMNA5 et VCN10(5) en sites non jaugés.

Les travaux destinés à l'estimation en site peu jaugé, en particulier ceux dédiés à valoriser les données de jaugeage épisodiques, sont en voie d'achèvement. Ils feront l'objet d'une note ultérieure.

Les auteurs

Clotaire Catalogne
Doctorant
clotaire.catalogne@irstea.fr
Irstea Lyon, HHLY

Eric Sauquet
Chargé de Recherche
eric.sauquet@irstea.fr
Irstea Lyon, HHLY

Les correspondants

Pascal Maugis, DAST, pascal.maugis@onema.fr

Eric Sauquet, HHLY, eric.sauquet@irstea.fr

Droits d'usage :	<i>accès libre</i>
Couverture géographique :	<i>France</i>
Niveau géographique [un seul choix] :	<i>National</i>
Niveau de lecture [plusieurs choix possibles] :	<i>Professionnels, experts</i>
Nature de la ressource :	<i>Document</i>

Interpolation des débits de référence d'étiage

Rapport d'avancement

C. Catalogne, E. Sauquet

Introduction	5
1 Constitution du jeu de stations de référence	6
2 Analyse et interprétation des interrelations entre indices hydrologiques d'étiage pour la définition de similitudes entre bassins instrumentés	15
2.1 Analyse préliminaire des interrelations entre indices d'étiage	16
2.2 Définition d'un espace hydrologique par analyse en composante principale	18
2.3 Classification des régimes d'étiage	19
3 Comparaison des méthodes de régionalisation des débits caractéristiques QMNA5 et VCN10(5) spécifiques.....	22
3.1 Protocole de définition des modèles régressifs, variables explicatives et mesure des performances d'estimation	23
3.2 Construction d'un modèle régressif national	25
3.3 Intérêt de la définition de voisinage : démonstration dans le cas de voisinages idéalement connus	27
3.4 Estimation sur la base de l'appartenance à une HER de niveau I	33
3.5 Couplage CART-RLM	36
3.6 Couplage ROI-RLM	41
3.7 Conclusion	49
4 Cartographie des variables QMNA5 et VCN10(5) spécifiques et de leurs incertitudes....	52
5 Conclusion	56
Bibliographie	58

Interpolation des débits de référence d'étiage

Rapport d'avancement

C. Catalogne, E. Sauquet

Résumé

Ce rapport fait état des avancées en matière d'interpolation des débits caractéristiques d'étiage QMNA5 (quantile de débit mensuel minimal annuel sec de période de retour 5 ans) et VCN10(5) (quantile de débit moyen sur 10 jours minimal annuel sec de période de retour 5 ans) en France.

Il contient tout d'abord des éléments techniques sur l'élaboration d'un jeu de stations de référence (stations hydrométriques considérées comme peu influencées par des perturbations d'origine anthropiques, dotées de chroniques suffisamment longues pour garantir une bonne estimation des statistiques d'étiage, mesures de bonne qualité et représentatives de la diversité des régimes hydrologiques rencontrés en France).

A partir de ces données, une analyse statistique de 11 indices représentatifs de différents aspects du régime d'étiage est ensuite proposée puis interprétée pour fournir une caractérisation relativement grossière des régimes d'étiage à l'échelle nationale. Celle-ci a permis de disposer d'éléments pertinents pour une meilleure description des similitudes hydrologiques entre bassins instrumentés et a été exploitée pour la définition de voisinages adaptés à l'estimation par régression multiple des variables QMNA5 et VCN10(5) en site non jaugé.

Ce travail d'estimation, passant par la comparaison de différentes approches, est présenté en détail et occupe une large partie de ce rapport.

Finalement une cartographie des débits caractéristiques QMNA5 et VCN10(5) a été établie sur la base de la méthode fournissant les meilleurs résultats en validation croisée.

Mots clés (thématique et géographique)

Etiage, formulation empirique, région hydrologique, typologie, arbre de régression

Introduction

Ce rapport d'avancement s'inscrit dans la poursuite des actions inscrites dans la convention cadre entre le Cemagref et l'ONEMA portant sur la « Synthèse des régimes hydrologique d'étiage en France ». Il fait suite et complète le rapport précédent, daté de février 2010 [**Catalogne et Sauquet (2010)**], et présente la démarche finalement retenue par l'équipe de Lyon pour l'estimation des débits caractéristiques QMNA5 (quantile de débit mensuel minimal annuel sec de période de retour 5 ans) et VCN10(5) (quantile de débit moyen sur 10 jours minimal annuel sec de période de retour 5 ans) spécifiques (i.e. exprimés en l/s/km²) en sites non jaugés.

Il reprend en premier lieu les éléments se rapportant à la constitution d'un jeu de chroniques de débit journalier de référence adapté à l'étude des basses eaux. Sont ensuite présentés les éléments nouveaux de la démarche adoptée avec :

1. une caractérisation des étiages à l'échelle nationale à partir de 11 descripteurs du régime d'étiage extraits des chroniques de débit disponibles ;
2. la définition de critères pour l'identification de similitudes entre bassins sur la base de caractéristiques hydrologiques ;
3. la comparaison de différentes méthodes de définition de voisinages dédiées à l'estimation en sites non jaugés par régression multiple ;
4. la cartographie des variables QMNA5 et VCN10(5) spécifiques et de l'incertitude associée en tout point d'un réseau hydrographique théorique pour des bassins drainant une surface supérieure à 50 km².

1 Constitution du jeu de stations de référence

En 2009, un précédent travail avait abouti à la constitution d'un premier jeu de données pour les besoins de cartographie du quantile sec de période de retour 5 ans QMNA5 [Plasse et Sauquet (2009)]. L'extension envisagée des travaux d'interpolation à d'autre type de variable, notamment de débits d'étiage représentatifs de durées inférieures à un mois, a nécessité de revoir ce jeu de données sur la base d'une analyse plus fine des séries chronologiques de débit et de reconstruire un nouvel échantillon de référence à partir des stations recensées dans la Banque HYDRO (<http://www.hydro.eaufrance.fr>).

Il a donc été procédé à un nouveau travail de vérification et de sélection des données disponibles en tentant d'adopter une démarche aussi rigoureuse que possible compte tenu des informations parfois insuffisantes sur les caractéristiques des stations hydrométriques. Ainsi s'il s'agit généralement de disposer d'un jeu de données suffisamment étoffé pour répondre aux exigences de représentativité nécessaire pour assurer l'efficacité des méthodes d'interpolation, il est également indispensable d'exclure les données ne présentant pas une fiabilité suffisante ou influencées par les activités anthropiques (*i.e.* régime non naturel) afin de ne pas introduire de biais artificiels dans les traitements réalisés.

La notion de représentativité se réfère ici à trois aspects importants qui ont été examinés en détail au cours de la sélection :

- la représentativité spatiale, qui sous-entend une répartition homogène des sites de référence ;
- la représentativité « comportementale » permettant de rendre compte de la diversité des réponses hydrologiques ;
- la représentativité statistique, qui requiert l'emploi de chroniques suffisamment longues et concomitantes pour assurer la robustesse des statistiques calculées (notamment par l'absence de la sensibilité aux variations climatiques).

Toute la difficulté consiste alors à trouver le meilleur compromis entre représentativité spatiale/comportementale et représentativité statistique qui constitue généralement le facteur le plus contraignant en terme de sélection des données.

Comme précédemment, la période 1970-2005 a été choisie comme période de référence commune pour l'ensemble des sites participant à l'échantillon de référence. Elle représente en effet la période pour laquelle le nombre de stations en fonctionnement est le plus élevé (Figure 1). Ce choix permet par ailleurs de travailler sur des données récentes (étude de l'état « actuel »), généralement validées par les gestionnaires de stations, et d'assurer un maximum de concomitance entre les chroniques sélectionnées.

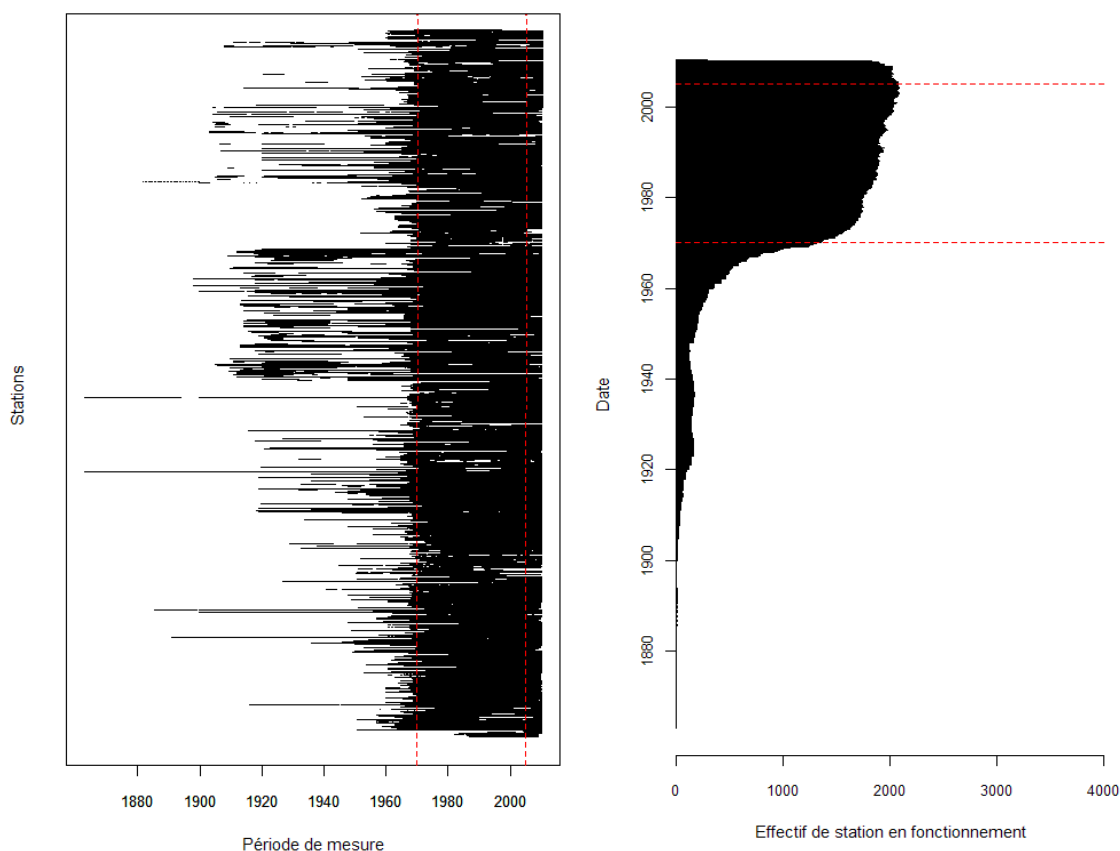


Figure 1 : Disponibilité des données pour l'ensemble des stations de la banque HYDRO et effectif de stations en fonctionnement sur la période 1860-2010

Un certain nombre d'informations mises à disposition par la banque HYDRO, parallèlement aux données de débit, ont été examinées afin de réaliser une première sélection visant à dégrossir notre échantillon pour ne conserver que les sites répondant aux critères suivants : mesures à signification hydrologique, caractère peu ou pas influencé du régime hydrologique et bonne qualité des mesures en basses eaux¹. Après croisement de l'ensemble de ces critères un échantillon préliminaire de 2375 stations a été retenu (parmi 3782 stations initialement disponibles). Il convient toutefois de préciser que, parmi les informations examinées ici, celles relatives à la qualité des données et l'existence d'influence anthropique doivent être considérées avec précaution (l'existence d'une influence anthropique significative mais non signalée dans la banque HYDRO s'avère par exemple courante).

Etant donné que, au sein de cet échantillon, le nombre de sites disposant de données sur la totalité de la période de référence est pratiquement nul, il a été examiné la possibilité d'employer des chroniques plus courtes sans affecter significativement la détermination des statistiques hydrologiques.

On trouve dans la littérature quelques recommandations sur la quantité d'observations appropriée pour assurer une estimation fiable des statistiques de débit aux sites de référence. Cependant ces recommandations restent souvent subjectives et diffèrent, parfois fortement, selon les auteurs : **Tallaksen**

¹ Selon la terminologie adoptée dans la banque HYDRO

et **Van Lanen (2004)** recommandent par exemple de travailler avec des chroniques disposant d'au moins 15 à 20 années d'observation tandis que **Dyck (1976)** juge les chroniques de plus de 3 années « déjà » exploitables en contexte climatique humide. En pratique les besoins doivent s'exprimer en fonction de la variable d'intérêt (par exemple pour les variables se rapportant à la description des crues, il est admis que la fiabilité des quantiles de période de retour T est assurée si la chronique comprend au moins $T/3$ années).

Ainsi, afin de déterminer objectivement la longueur de chronique requise pour la détermination fiable des statistiques d'étiage sur la période de référence 1970-2005, il a été choisi de réaliser une analyse de sensibilité par ré-échantillonnage au sein de chroniques suffisamment longues. Plusieurs statistiques hydrologiques ont ainsi été calculées successivement pour des longueurs de chronique comprises entre 5 et 30 années d'enregistrement consécutives (*i.e.* échantillonnage par blocs) : le module interannuel (QA), les quantiles Q90 et Q95 extraits de la courbe de débits classés interannuelle (valeurs de débits dépassées en moyenne 90 et 95% du temps), la date moyenne des jours où $Q < Q90$ (pris à titre d'indicateur de saisonnalité) et le « Base Flow Index » (noté BFI, proportion de l'écoulement de base par rapport à l'écoulement total). Un écart à la valeur de référence obtenue sur la totalité de la chronique disponible (*i.e.* jusqu'à 36 années) est ensuite calculé à l'issue de chaque ré-échantillonnage.

Pour chaque site examiné, la longueur de chronique appropriée a été prise par défaut comme la longueur pour laquelle 75 % des valeurs obtenues en ré-échantillonnage montrent un écart à la valeur de référence inférieur à 10 % dans le cas des variables de débit QA, Q90 et Q95, inférieur à 10 jours dans le cas de l'indice de saisonnalité et inférieur à 0.1 pour le BFI. Ces paramètres quoique arbitraires nous paraissent suffisamment exigeants pour assurer dans la majorité des cas la bonne représentativité statistique des indices calculés.

Pour mener à bien cette analyse, un échantillon pilote constituée de 274 sites² disposant de chroniques réputées fiables, suffisamment longues (*i.e.* au moins 30 années d'enregistrements sans lacunes sur la période de référence), relativement bien répartis sur l'ensemble du territoire et jugés suffisamment représentatifs des différents comportements hydrologiques pouvant exister à l'échelle nationale a été exploité.

Dans l'exemple donné ci-dessous pour la station P3234010 (Figure 2) la longueur de chronique appropriée à la détermination du Q95 approche 16 ans (graphique de gauche). Cette figure permet également d'illustrer l'impact de la variabilité climatique sur la détermination des statistiques hydrologiques (graphique de droite) et, par extension, la sensibilité des variables étudiées à la période échantillonnée. Le choix d'une longueur de chronique permettant d'obtenir 75 % de valeurs dans un intervalle de moins de 10 % autour de la valeur de référence assure que la sensibilité à la période échantillonnée reste limitée et permettra de s'affranchir d'une stricte concomitance entre les chroniques.

² Croisement des échantillons de référence employés dans l'étude menée à l'échelle nationale par Giuntoli et Renard (2009 et 2010) comptant 262 sites (ayant finalement évolué vers 236 sites) et dans l'étude menée sur le secteur Seine-Normandie par Plasse et Sauquet (2009) comptant 144 sites.

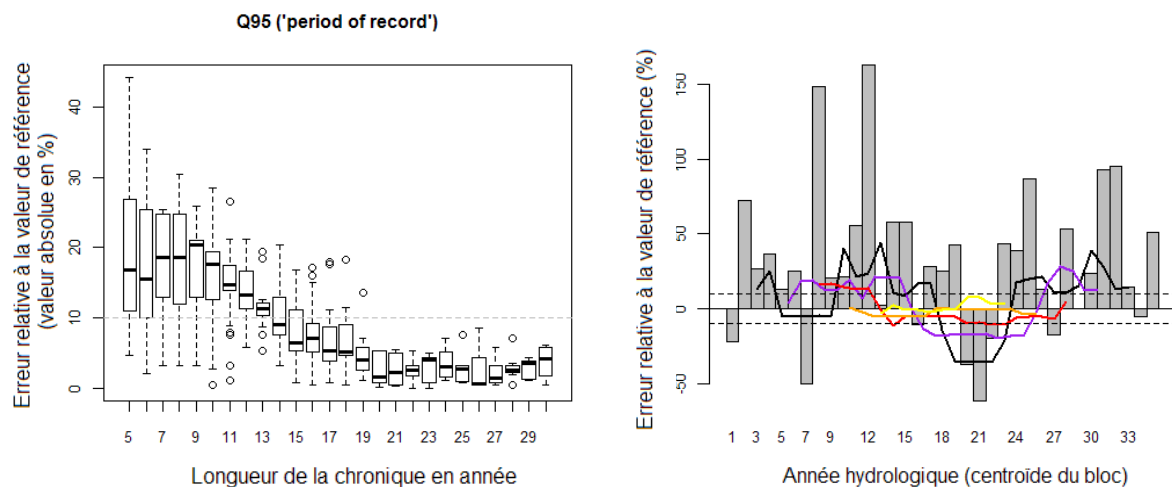


Figure 2 : Erreurs relatives (%) de détermination du quantile Q_{95} de la courbe de débits classés interannuelle (écart à la valeur de référence sur 36 ans) pour l'ensemble des ré-échantillonnages par blocs de 5 à 30 ans (graphique de gauche) et évolution de l'erreur relative de détermination selon la période échantillonnée (graphique de droite, du noir au jaune : ré-échantillonnage sur 5, 10, 15, 20 et 25 ans) pour la station P3234010

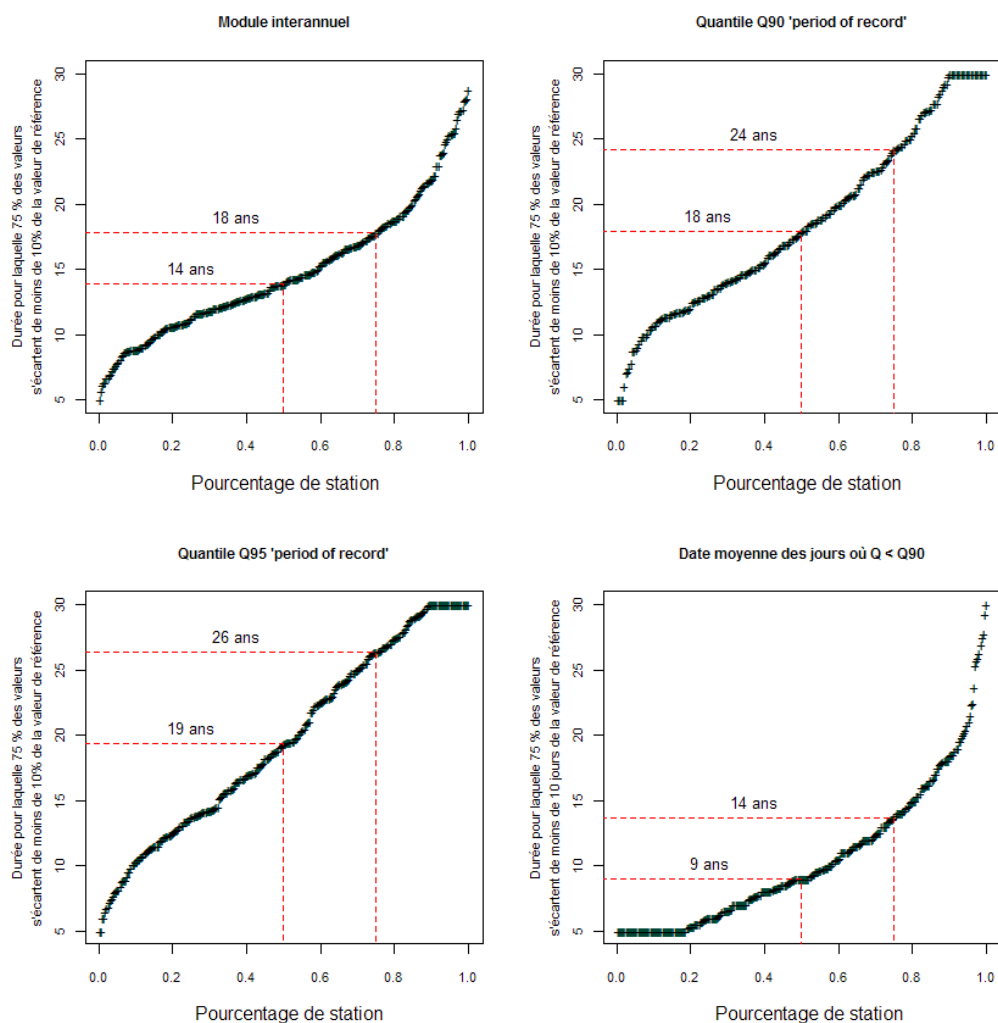


Figure 3 : Distributions cumulées des longueurs de chroniques jugées adéquates pour les 274 sites de l'échantillon pilote.

Les résultats obtenus pour l'ensemble de l'échantillon sont illustrés sur la Figure 3 sous la forme de distributions cumulées. Les résultats pour le BFI ne sont pas fournis en raison de la relative stabilité des valeurs obtenues quelle que soit la période échantillonnée (*i.e.* l'écart relatif à la valeur de référence reste toujours inférieur à 0.1) et pour lequel 5 années semblent déjà suffisantes pour obtenir une valeur proche de la valeur obtenue sur le long terme. On remarquera que ce dernier résultat ne s'avère pas surprenant, en effet le BFI peut être considéré comme une caractérisation des propriétés intrinsèques du bassin (considéré comme un filtre) et ne présente par conséquent qu'une faible sensibilité aux variations climatiques à moyen et long terme.

Pour les autres descripteurs hydrologiques, si l'on fixe la longueur de chronique acceptable à l'échelle nationale comme celle obtenue pour 75% de l'échantillon, celle-ci s'élève à 18 ans pour le module interannuel, 24 et 26 ans pour les quantiles Q90 et Q95 extraits de la courbe des débits classés et 14 ans pour l'indice de saisonnalité. On constate ainsi que les débits caractéristiques d'étiage se révèlent beaucoup plus sensibles à la longueur de chronique disponible que la statistique représentative du comportement moyen (inexact car dépend de T : si T = 100 ans il faut une chronique d'au moins 33 années), cette sensibilité étant par ailleurs souvent accrue pour les plus faibles valeurs (probablement en raison de leur plus forte sensibilité aux variations climatiques locales). A l'inverse une variable telle que la saisonnalité, se rapportant davantage à une caractéristique intrinsèque du climat régional, s'avère moins sensible à la durée de chronique disponible..

Pour le cas le plus exigeant dégagé ci-dessus, c'est-à-dire pour une longueur de chronique d'au moins 26 années, le nombre de sites retenu parmi les 2735 stations déjà présélectionnées est ramené à 796. Afin de rendre compte de la représentativité spatiale de ces 796 sites, il a été choisi de calculer et de représenter sur une carte la distance de chaque point du réseau hydrographique drainant une surface supérieure à 100 km² à la station la plus proche (distance entre centres de gravités des bassins versants, Figure 4). Le résultat obtenu montre qu'un peu plus de 80 % des points du réseau se situent à moins de 20 km d'une station de l'échantillon, néanmoins la répartition des sites n'est pas homogène avec une très faible représentativité en Charente, dans le Nord, à l'extrémité du Cotentin et dans une moindre mesure en Camargue et dans le nord de la Corse.

Un choix moins exigeant sur la longueur de chronique (20 ans, correspondant à une proportion de l'échantillon pilote montrant un écart inférieur à 10 % du Q95 de référence proche de 50 %) permet d'accroître le nombre de stations sélectionnées de presque 300 et d'amener à 86 % le nombre de points du réseau éloignés de moins de 20 km d'une station de l'échantillon. Toutefois si cet accroissement permet d'améliorer la représentativité en Corse (8 stations supplémentaires) et à l'extrémité du Cotentin (deux stations supplémentaires), les secteurs du Nord³, de la Charente⁴ et de la Camargue restent en revanche

³ Les stations de ce secteur, lorsqu'elles ont suffisamment de données, sont souvent qualifiées de « douteuse en basses eaux » selon la terminologie de la Banque HYDRO

⁴ Les stations de ce secteur, lorsqu'elles ont suffisamment de données, sont souvent affectées par une influence anthropique des débits

dépourvus de site de référence. Ces résultats montrent l'intérêt limité à accroître le nombre de sites sélectionnés eu égard à la perte de fiabilité sur la détermination des indices d'étiage.

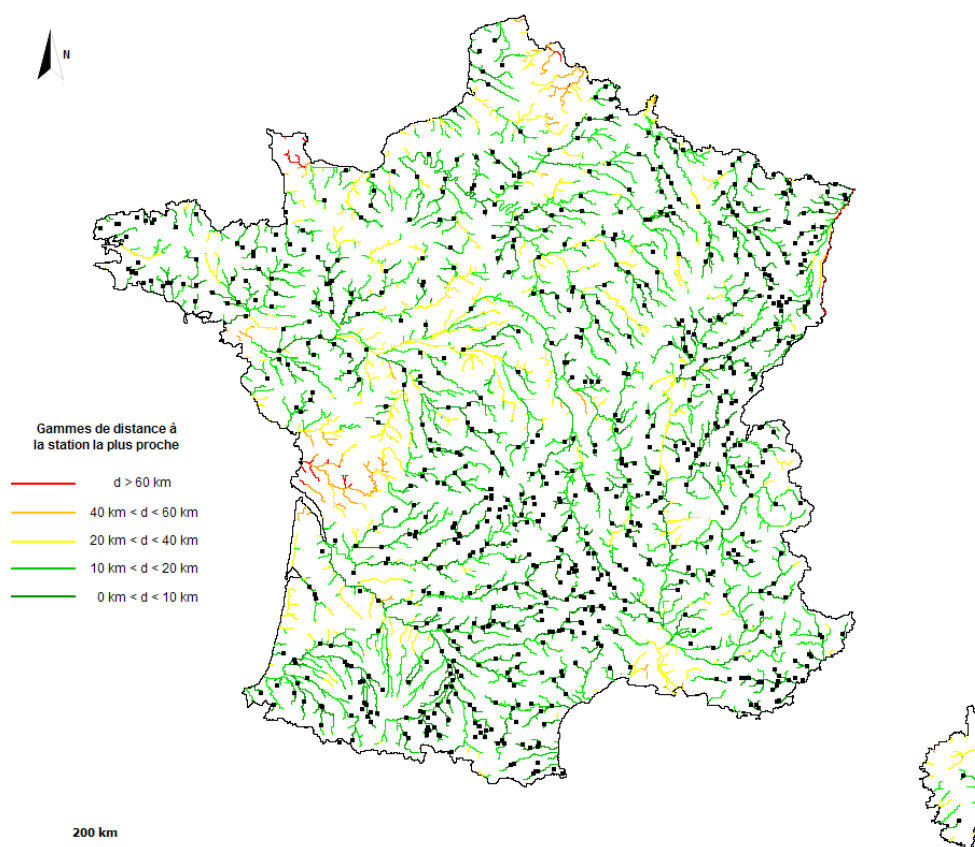


Figure 4 : Distance de chaque point du réseau hydrographique drainant une surface supérieure à 100 km² à la station la plus proche (distance entre centres de gravité des bassins versants) et localisation des 796 stations disposant d'au moins 26 années de données

Une critique approfondie des 796 sites obtenus à ce stade de la sélection a finalement été entreprise afin de valider notre échantillon de station, notamment vis-à-vis des problématiques de fiabilité des données et d'influence anthropique potentiellement omises dans la Banque HYDRO.

La principale phase de cette critique a consisté à identifier les sites les plus douteux, en premier lieu grâce à l'examen visuel des chroniques puis à l'aide de tests statistiques de rupture et de tendance sur les séries annuelles de débit moyen, de débits moyens mensuels minimaux (QMNA) et de débits moyens minimaux sur 10 jours (VCN10) (pour les outils de cette analyse on se reportera à **[Plasse et Sauquet (2009)]**). Parallèlement, le recueil d'informations sur l'existence de barrages, retenues, dérivations et/ou prélèvements susceptibles d'impacter significativement le régime d'écoulement en amont des stations de l'échantillon, a permis de disposer d'arguments pour justifier le rejet de stations douteuses ou indéniablement soumises à une influence anthropique non mentionnée dans la Banque HYDRO.

Les exemples ci-dessous illustrent deux types de problèmes rencontrés sur certaines des chroniques et les causes identifiées (Figure 5 et Figure 6):

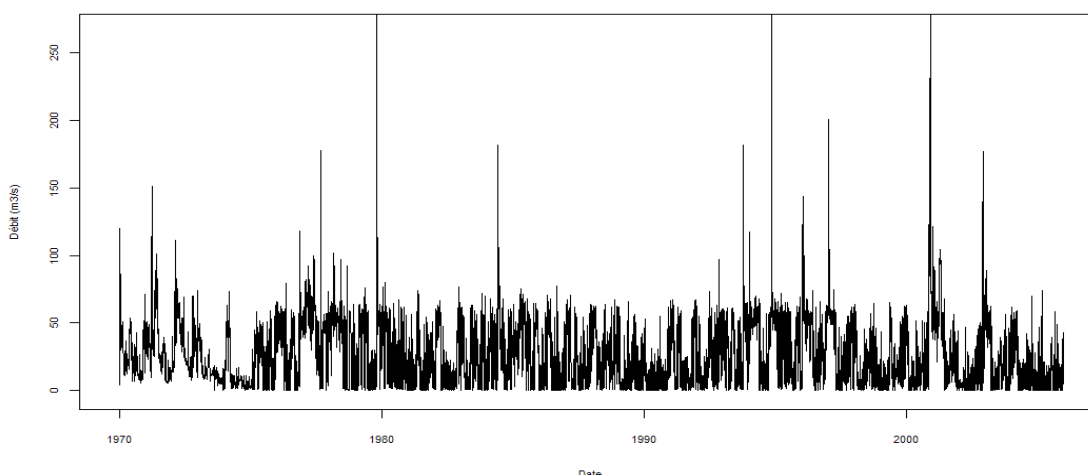


Figure 5 : Chronique du Verdon à Vinon-sur-Verdon (X2812010) sur la période 1970-2005 montrant l'effet des aménagements hydroélectriques situés en amont (station qualifiée de « pas ou faiblement influencée » dans la Banque HYDRO)

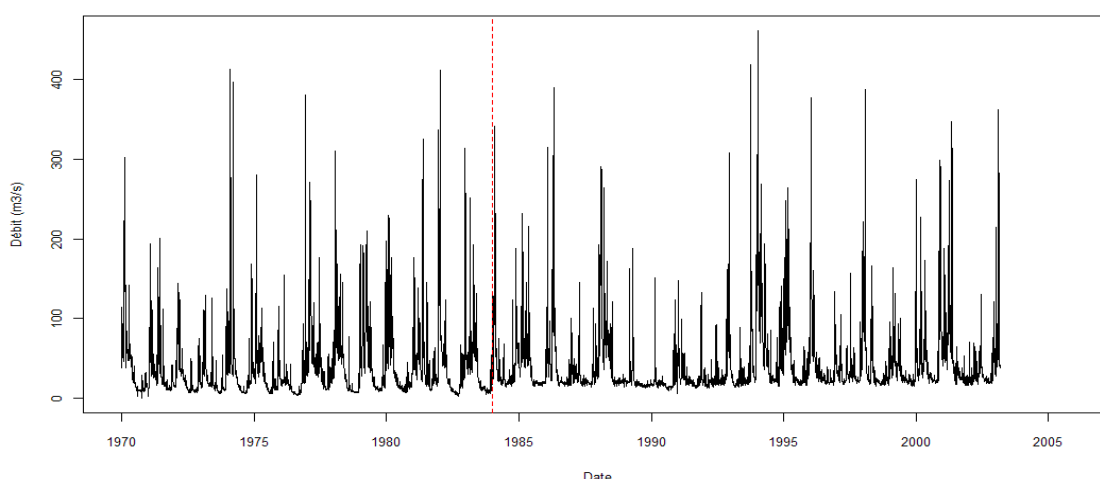


Figure 6 : Chronique de l'Isle à Saint-Laurent-des-Hommes (P7181510) sur la période 1970-2005 montrant une rupture nette à la fin de l'année 1983

Les stations « suspectes » suite à cette critique s'avèrent malheureusement nombreuses, il en résulte que l'échantillon est fortement réduit et se trouve actuellement ramené à 632 sites. Une nouvelle analyse, similaire à celle présentée précédemment, permet de constater que la représentativité spatiale se trouve globalement peu affectée par l'exclusion des sites jugés douteux (la proportion de points du réseau hydrographique éloignés de moins de 20 km atteint environ 77 % contre 80 % avec l'échantillon

précédent). En complément, la représentativité des régimes hydrologiques observés au sein de l'échantillon a été comparée aux résultats extraits de la cartographie des régimes hydrologiques réalisés par **Sauquet et al. (2008)** à l'échelle nationale (Figure 7). Cette comparaison montre que notre échantillon satisfait relativement bien nos exigences en termes de représentativité « comportementale » (Tableau 1) en restituant correctement la diversité des régimes rencontrés en France. On signalera par ailleurs la présence au sein de notre échantillon de sites présentant périodiquement des débits « nuls » (débits inférieurs à 1 l/s correspondant à la précision des valeurs de débits fournies par la Banque HYDRO), représentatifs de régimes intermittents.

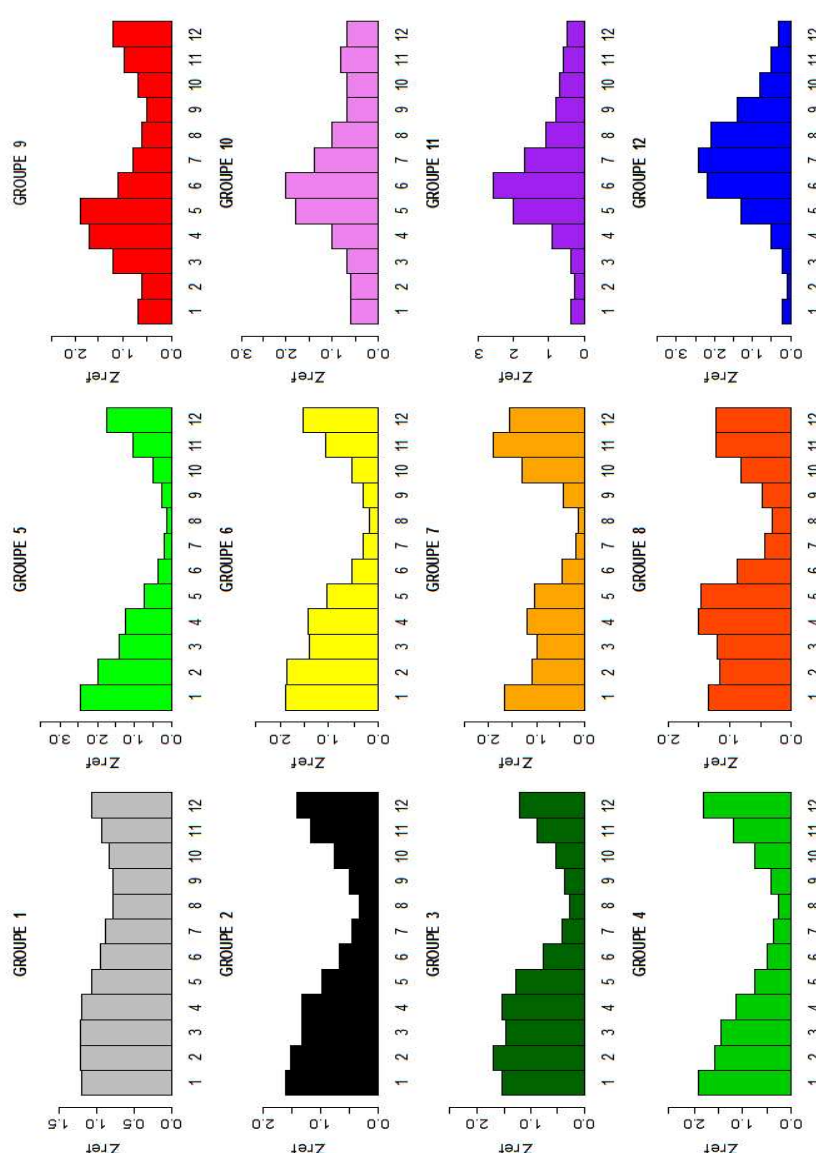


Figure 7 : Classification en régimes hydrologiques proposée par Sauquet et al. (2008).
Zref est le rapport du débit mensuel moyen (mm/mois) divisé par le débit annuel moyen (mm/an)

Classes de régimes définies par Sauquet et al. (2008)	Proportion (%) sur le réseau hydrographique (S > 50 km ²) ⁵	Proportion (%) dans l'échantillon de référence
1 – Uniforme (nappe)	7	6
2 - Pluvial type A	23	18
3- Pluvial type B	11	15
4 - Pluvial type C	7	5
5- Pluvial type D	12	15
6 - Pluvial type E	22	23
7 - Méditerranéen	3	4
8 - Pluvio-nival	8	8
9 - Nivo-Pluvial	2	2
10 - Nival de transition	3	1
11 - Nival	1	2
12 - Glaciaire	< 1	< 1

Tableau 1 : Comparaison des proportions de régimes hydrologiques représentés sur le réseau hydrographique à l'échelle nationale et dans l'échantillon de référence

Pour conclure, notons que, malgré des exigences relativement contraignantes qui aboutissent à limiter dangereusement la représentativité spatiale de notre échantillon, la validité d'un certain nombre de sites pourrait être encore remise en cause, notamment vis-à-vis de la fiabilité des mesures qui constitue probablement l'une des principales faiblesses des données sélectionnées. L'hydrométrie des débits de basses eaux se révèle en effet particulièrement délicate en raison des nombreux facteurs susceptibles d'impacter la mesure (sans parler de l'imprécision de la mesure elle-même)⁶. La connaissance de l'incertitude résultant des problèmes hydrométriques reste difficilement accessible, elle n'est jamais mentionnée dans la Banque HYDRO et ne peut être quantifiée sans une enquête minutieuse auprès des gestionnaires⁷.

⁵ Données extraites des travaux de cartographie des régimes hydrologiques réalisés par **Sauquet et al. (2008)** à l'échelle nationale.

⁶ En revanche, contrairement aux crues, la mesure en continu des débits d'étiage autorise un pas de temps d'enregistrement relativement grossier sans répercussion importante en termes de perte d'information.

⁷ Il pourrait être envisagé de réaliser ce type d'enquête dans un secteur précis afin de mettre en évidence l'éventuel impact de ces incertitudes dans les démarches d'estimation en site peu ou pas jaugé.

2 Analyse et interprétation des interrelations entre indices hydrologiques d'étiage pour la définition de similitudes entre bassins instrumentés

On se propose ici d'analyser et caractériser les régimes d'étiage à l'échelle nationale au travers d'un certain nombre de descripteurs extraits des chroniques de débits journaliers des stations de l'échantillon de référence. Cette analyse aura pour finalité l'identification de critères pertinents vis-à-vis de la discrimination des comportements hydrologiques et la sélection de voisinage en sites instrumentés.

La description du régime d'étiage peut être appréhendée au travers d'un certain nombre de caractéristiques différentes (sévérité, durée, saisonnalité, fréquence...) chacune étant susceptible de répondre à différentes problématiques de gestion. Ainsi, au delà des variables réglementaires habituellement employées en France, nous avons souhaité élargir au maximum les possibilités de description en recensant dans la littérature d'autres variables caractéristiques d'étiage couramment utilisées (pour une revue relativement exhaustive à ce sujet on se reportera par exemple à **Smakthin (2001)**). Un important travail a été mené pour définir les modalités les plus appropriées pour l'extraction de ces différents descripteurs à partir des séries chronologiques de débit moyen journalier (non présenté dans le cadre de ce rapport). Finalement, 11 indices jugés représentatifs des différents aspects du régime de basses eaux ont été conservés (Tableau 2).

Avant de procéder à l'analyse on notera que, vis-à-vis d'une éventuelle utilisation dans le cadre opérationnel, toutes ces variables n'ont pas la même utilité. Nous distinguerons ainsi les variables d'intérêt, cibles pour la régionalisation (variables de sévérité dont le QMNA5 et le VCN10(5) considérés dans ce rapport, de durée ou de déficit) et les variables annexes, se rapportant plus spécifiquement à la description des processus hydrologiques d'étiage et susceptibles d'être employées comme variables d'appui pour l'analyse, la compréhension et la classification des comportements hydrologiques.

Catégorie	Indice	
Sévérité	Q90	Quantile de fréquence au dépassement 90% de la CDC interannuelle
	QMNA5	Débit moyen mensuel minimal annuel de période de retour 5 ans
	VCN1(5)	Débit moyen minimal annuel ⁽¹⁾ sur 1 jours de période de retour 5 ans
	VCN10(5)	Débit moyen minimal annuel ⁽¹⁾ sur 10 jours de période de retour 5 ans
Déficit	DefXA(5)	Déficit maximal annuel ⁽¹⁾ de période de retour 5 ans (défini vis-à-vis du Q90)
Durée	dXA(5)	Durée maximale annuelle ⁽¹⁾ de période de retour 5 ans (définie vis-à-vis du Q90)
Saisonnalité	SR	Rapport des Q90 hivernaux et estivaux des CDC saisonnières interannuelle
Débit de base	BFI	Rapport du débit de base moyen interannuel sur le module (algorithme de séparation : Lyne et Hollick (1979))
	Ic	Indice de concavité : $(Q1 - Q10)/(Q10 - Q99)$ représentatif du comportement plus ou moins réactif du bassin
	QB	Rapport Q50 / Q90 représentatif de la proportion de débit de base par rapport au débit médian
Résilience	α_{rec}	Coefficient de tarissement caractéristique

⁽¹⁾ sur la base de l'année hydrologique

Tableau 2 : Variables hydrologiques pré-sélectionnées pour caractériser les régimes d'étiage

2.1 Analyse préliminaire des interrelations entre indices d'étiage

L'analyse préliminaire proposée ici vise simplement à mettre en évidence, au moyen de la matrice des coefficients de corrélation linéaire (corrélation de Pearson), les interrelations existant entre les différents descripteurs du régime d'étiage. On notera que cette analyse a été réalisée en normalisant les variables de sévérité et déficit par le débit moyen interannuel (QA). Cette opération a pour effet d'éliminer la composante de variabilité se rapportant directement à la productivité du bassin expliquée en grande partie par la surface drainée et permet ainsi de renforcer le lien entre variables de processus (i.e. descriptives du comportement du bassin) et certaines des variables cibles pour la régionalisation (sévérité, déficit) .

La représentation donnée ci-dessous a été conçue pour permettre de visualiser rapidement les relations entre les variables examinées (Figure 8). Afin de simplifier notre analyse il a été jugé approprié de regrouper les différentes variables examinées en 4 familles principales regroupant :

- les variables de sévérité (groupe 1) ;
- les variables de déficit normé⁸ et durée (groupe 2) ;
- la variable de saisonnalités (groupe 3) ;
- les variables de BFI, Ic, QB et la variable de résilience Alpha qui caractérisent plus ou moins directement l'existence de réserves permettant d'assurer un débit de base (groupe 4).

⁸ La normalisation a eu ici pour effet de renforcer très sensiblement la corrélation entre ces deux variables

Afin de permettre d'affiner l'analyse une distinction entre 2 classes de saisonnalité a également été introduite et mise en évidence sur les graphiques bi-variés par des couleurs différentes (en rouge pour les régimes nivaux ($SR > 1$) et en noir pour les régimes pluviaux ($SR < 1$))

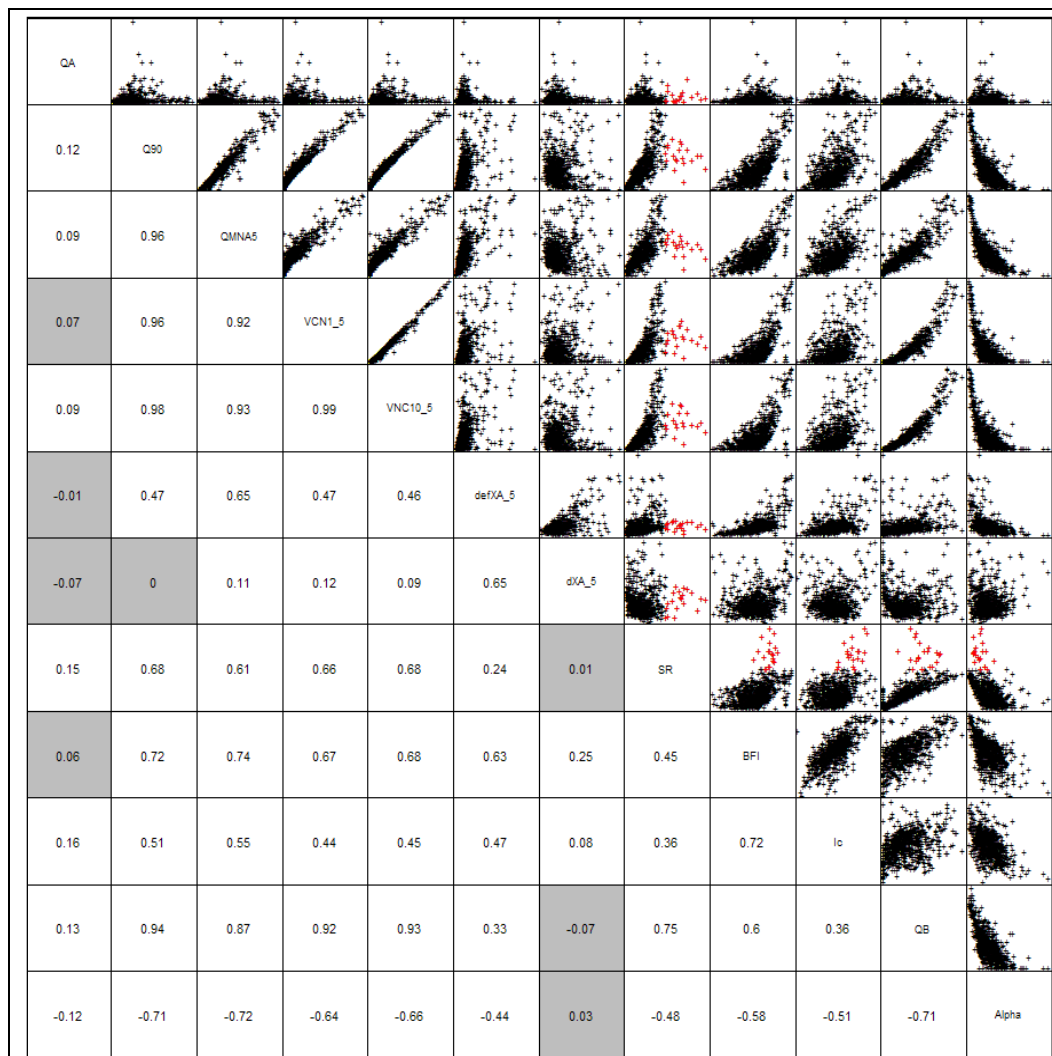


Figure 8 : Matrice des corrélations de Pearson pour le débit moyen interannuel (QA) et les 11 indices d'étiage examinés (en gris : corrélations non significatives au niveau de confiance 5 %). Les variables de sévérité et de déficit sont normalisées par le QA

On constate en premier lieu que la normalisation par le QA a pour effet de décorrélérer complètement cette dernière variable du reste des indices. Quoique de manière indirecte, il apparaît ainsi que la taille du bassin versant n'intervient que de façon mineure dans la variabilité des comportements hydrologiques d'étiage.

Si l'on examine maintenant les relations entre les différents groupes il ressort clairement un lien très prononcé entre le groupe 1, le groupe 4 et la variable de saisonnalité. Ce lien (non linéaire) démontre l'importance des propriétés de vidange et/ou la présence de réserves susceptible d'assurer un débit de base

pour expliquer la variabilité des régimes d'étiage à l'échelle nationale. Il s'agira par conséquent du facteur le plus discriminant pour regrouper les bassins similaires. Quoique de manière atténuée, on note également que ce même lien existe avec la variable de déficit $\text{defXA}(5)$. Cependant cette dernière présente plus de dépendance avec la variable de durée $\text{dXA}(5)$ qui apparaît comme indépendante du reste des indices examinés

Si l'on revient plus spécifiquement à l'information sur la saisonnalité on constate que la distinction opérée entre régimes nivaux et régimes pluviaux s'avère pertinente. En effet si l'indice SR représentatif des régimes pluviaux ($\text{SR} < 1$) présente assez clairement une certaine dépendance aux processus de stockage/vidange (i.e. lorsque SR tend vers 1, les variations saisonnières de débit sont amorties par la présence d'échanges importants avec les aquifères), il semble que cette relation n'apparaisse plus dans le cas des régimes nivaux ($\text{SR} > 1$) si ce n'est qu'ils sont globalement associés à des bassins dotés d'une fonction d'inertie relativement importante comme le démontre les valeurs de BFI plutôt élevées (il s'agit ici essentiellement du processus de stockage/restitution des précipitations sous forme solide). Ainsi il apparaît d'ores et déjà pertinent de différencier ces bassins du reste de l'échantillon et d'en faire un groupe à part entière (ce groupe comprend 21 individus).

2.2 Définition d'un espace hydrologique par analyse en composante principale

Compte tenu de la forte redondance d'information dans les descripteurs du régime d'étiage analysés plus haut et afin de faciliter l'identification des similitudes entre bassins instrumentés, il a été choisi d'appliquer à nos données une analyse en composante principale (ACP). Cette technique présente en effet l'avantage de permettre de résumer et hiérarchiser l'information contenue dans le jeu de données au travers d'un nombre limité de dimensions (i.e. sous la forme de combinaisons linéaires de variables) orthogonales (i.e. mutuellement décorrélées deux à deux, indépendantes). Les premières dimensions portant le plus de variance (i.e. les plus discriminantes) pourront ainsi se substituer aux 11 descripteurs pour définir un nouveau référentiel (i.e. un espace « hydrologique ») pertinent vis-à-vis de l'identification des similitudes et des voisinages entre bassins instrumentés.

L'ACP proposée a été réalisée ici après standardisation des 11 descripteurs du régime d'étiage et exclusion des 21 bassins à régime nival. L'interprétation des résultats obtenus confirme bien entendu les observations précédentes. Elle permet toutefois de préciser certaines de nos conclusions en établissant explicitement une hiérarchie entre les caractéristiques les mieux exprimées dans les données. Ainsi il apparaît essentiellement un axe factoriel significatif (qui porte de loin la plus grande partie de la variance : 35 %), porteur comme attendu de la relation entre l'existence d'une inertie permettant de maintenir un débit de base en période d'étiage et la productivité du bassin versant affranchie de l'effet de taille (s'exprimant au travers des variables de débit normées). Par contraste, les axes suivants s'avèrent beaucoup moins informatifs en terme de pouvoir discriminant avec 17 %, 13 %, 9 %... de la variance expliquée. On note cependant que le second axe factoriel apparaît clairement comme porteur du lien entre durée et déficit (tandis que les axes suivants sont plus difficilement interprétables), par conséquent il semble tout de même pertinent de le conserver pour une description plus fine des similitudes entre bassins. Ainsi dans la suite de ces travaux seuls les deux premiers axes factoriels (portants au total 52 % de la variance expliquée)

ont été considérés comme appropriés (Figure 9) pour constituer l'espace hydrologique, ils seront désignés dans la suite de ce document par les notations *ACP1* et *ACP2*.

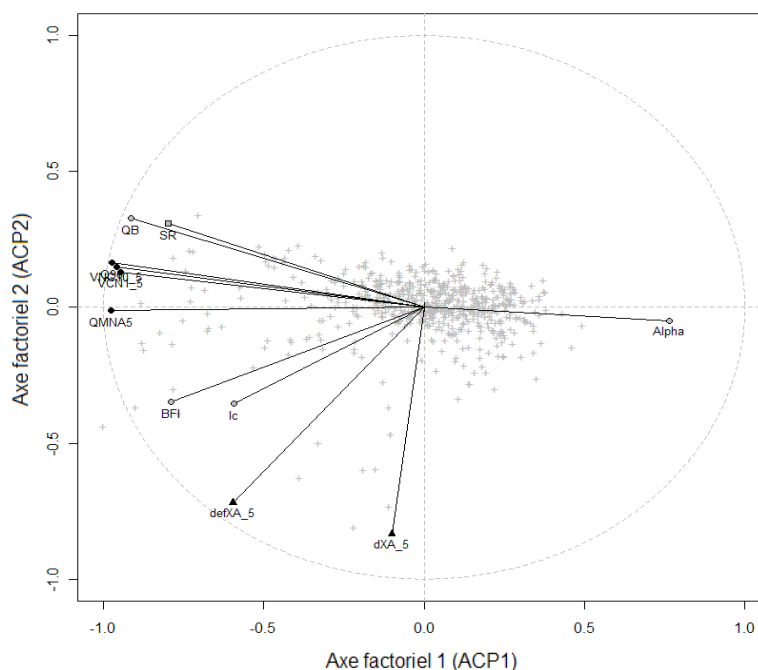


Figure 9 : Projection des indices d'étiage et des sites de l'échantillon (en gris) dans le premier plan factoriel obtenu par ACP

2.3 Classification des régimes d'étiage

Suite à l'ACP, une classification ascendante hiérarchique (CAH) a été mise en œuvre sur la base de la position de chaque individu dans le premier plan factoriel obtenu précédemment (excluant toujours les 21 bassins à régime nival, considérés comme une classe à part entière). Cette classification vise à constituer empiriquement un ensemble de regroupements permettant de visualiser grossièrement la répartition des régimes hydrologiques d'étiage à l'échelle de la France. L'arbre de classification obtenu est reporté ci-dessous (Figure 10). Au regard des valeurs du critère de dissemblance (distance de Ward) on constate que 3 classes principales se dégagent assez nettement (les effectifs sont de 88, 224 et 295 respectivement pour les classes 1 à 3).

Si l'on examine, pour chaque classe, les distributions de chacune des deux variables hydrologiques de substitution (Figure 11), on remarque sans surprise que celles-ci se distinguent de manière presque exclusive sur la base de la première dimension *ACP1*, c'est-à-dire de l'information sur l'existence de réserves susceptibles de maintenir le débit de base. La seconde dimension *ACP2*, qui porte l'information de durée, apparaît comme peu discriminante si ce n'est que la classe 1 est caractérisée par une forte variabilité avec la présence de sites caractérisés par des étiages particulièrement longs ($V2 \ll 0$). Il s'agit de bassins dont le régime est calqué sur la dynamique interannuelle du niveau des nappes avec lesquels ils communiquent. Les périodes d'étiage correspondent alors aux périodes de basses eaux de l'aquifère (qui peuvent s'étaler sur plusieurs années).

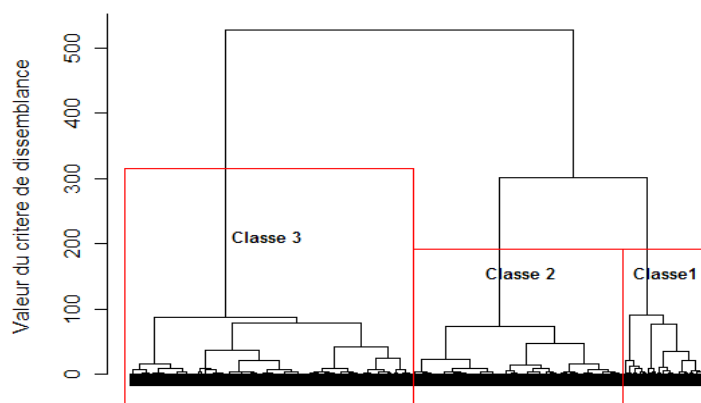


Figure 10 : Arbre de classification obtenu à l'issue de la CAH

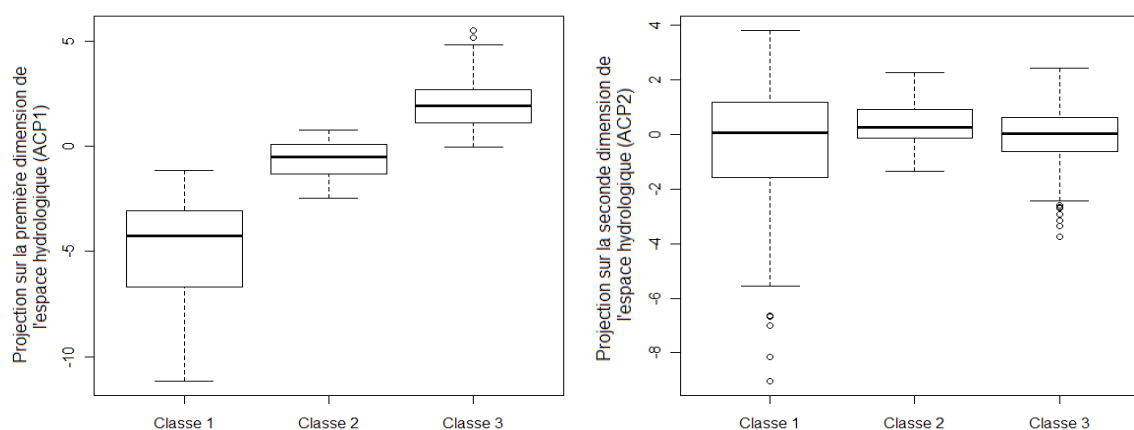


Figure 11 : Distributions des deux variables hydrologiques de substitution en fonction des trois classes définies au moyen de la CAH

La carte ci-dessous (Figure 12) illustre la répartition géographique des différentes classes de régime d'étiage identifiées précédemment (incluant la classe spécifique aux régimes nivaux). La connaissance des grandes caractéristiques climatiques, géologiques ou topographiques régionales permet de donner l'interprétation suivante :

- la classe 1, principalement représentée dans le bassin parisien, correspond aux bassins dotés de réserves importantes (contributions d'aquifères étendus),
- la classe 2 représente un régime intermédiaire dans les zones de moyenne altitude et/ou à climat océanique très marqué (précipitations régulières et abondantes, contraste saisonnier peu marqué, *e.g.* Finistère) ;
- la classe 3 correspond aux bassins sans fonction de stockage importante (substratum imperméable, *e.g.* Vendée) et/ou en climat contrasté et saison estivale sèche (*e.g.* pourtour méditerranéen).

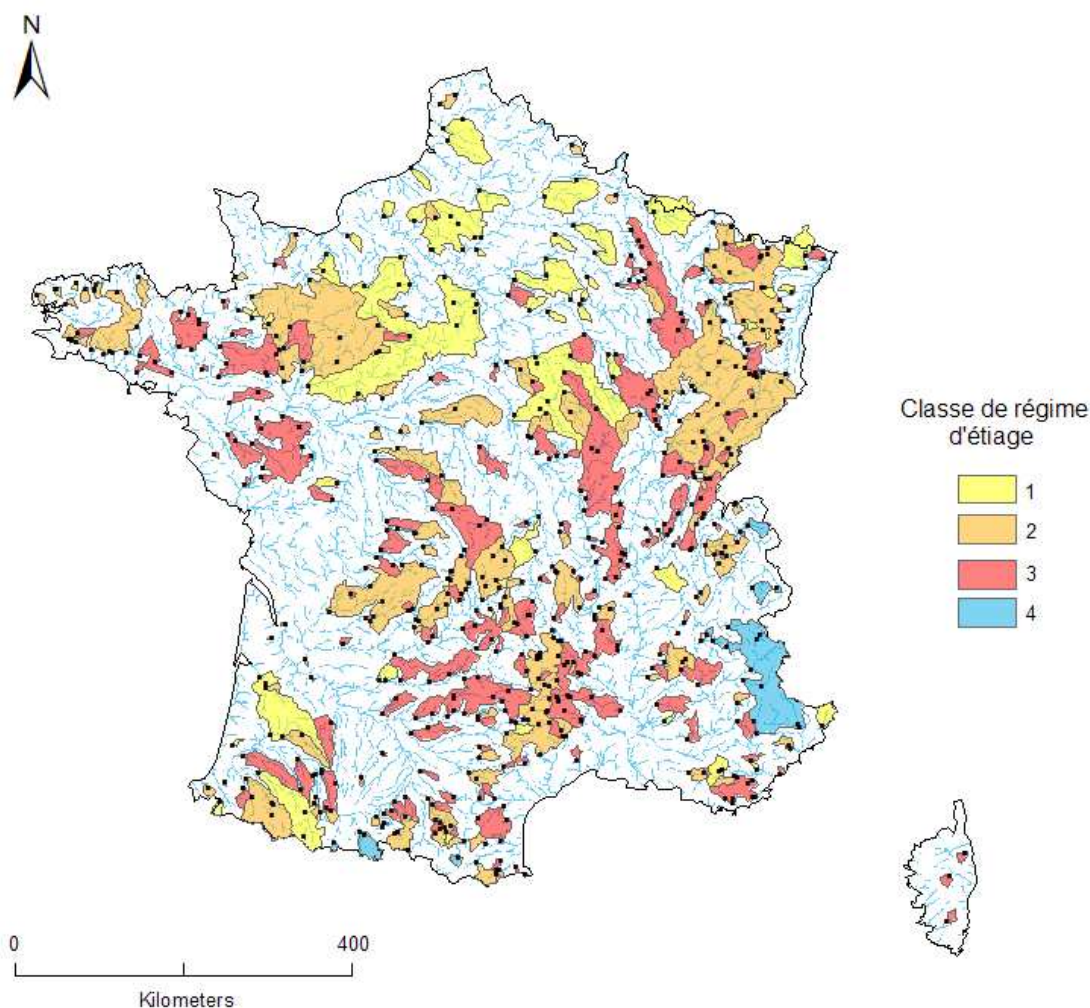


Figure 12 : Répartition des différentes classe de régime hydrologique d'étiage à l'échelle de la France (incluant ici les bassins à régime nival : classe 4 en bleu). Les sites sont représentés aux coordonnées des centres de gravité des bassins versants

Bien entendu cette classification reste relativement grossière et ne donne qu'une vision qualitative des similitudes entre bassins. Elle permet toutefois de constater que certains secteurs présentent une certaine homogénéité et permet de distinguer quelques ensembles régionaux cohérents, assez proches par ailleurs de ceux définis dans la typologie en HER proposée par **Wasson et al. (2001)**. On reconnaît ainsi une nette différenciation entre la partie occidentale de la Bretagne (plus arrosée) et la partie orientale qui se prolonge vers le sud jusqu'en Charente (sur substratum imperméable et périodes estivales généralement sèches). Le secteur du bassin parisien, caractérisé par la présence de nappes étendues susceptibles d'alimenter les cours d'eau en période sèche, semble également bien s'individualiser avec une nette prédominance de la classe 1. En revanche l'homogénéité s'avère beaucoup plus incertaine (à cette échelle) sur une vaste partie centrale de la France, ceci semble témoigner de l'absence de facteurs régionaux réellement dominants (climat, géologie, relief) à l'origine d'un comportement plus ou moins uniforme des bassins. Enfin il convient de préciser que le manque de station dans certains secteurs rend difficile toute interprétation (*e.g.* sur le pourtour méditerranéen).

3 Comparaison des méthodes de régionalisation des débits caractéristiques QMNA5 et VCN10(5) spécifiques

La méthode d'estimation considérée ici s'appuiera sur l'élaboration de modèles régressifs entre les deux variables d'intérêt et un ensemble de descripteurs du bassin. Cette méthode constitue probablement la technique d'estimation en site non jaugeé la plus répandue en hydrologie et a donné lieu à une riche littérature scientifique, trop importante pour être traitée ici de manière exhaustive. Dans le cadre de l'estimation des descripteurs du régime d'étiage, la revue proposée par **Smakhtin (2001)** fournit déjà un aperçu relativement complet des différents travaux sur le sujet.

Dans cette approche la dépendance de la variabilité spatiale du phénomène étudié vis-à-vis des variations régionales de géologie, climat, altitude, etc. est décrite sous la forme de relations mathématiques simples. Celles-ci sont établies empiriquement (*i.e.* sur les observations disponibles) par ajustement de régressions linéaires multiples (« RLM », procédure décrite dans de nombreux ouvrages, *e.g.* **Yevjevich (1972)**) entre la variable étudiée et différents descripteurs du bassin, nécessairement connus en tout point du domaine d'étude. L'estimation en sites non jaugeés s'appuie alors sur la relation ajustée à partir des observations d'un ensemble de sites de référence donnés pour calculer la variable d'intérêt à partir des descripteurs du bassin locaux (*i.e.* propre aux bassins ciblés).

Afin d'assurer de bonnes performances d'estimation, la principale problématique abordée ici sera liée à l'étape de définition de régions/voisinages hydrologiques, étape souvent indissociable de la procédure d'estimation par régression multiple. A l'échelle d'un territoire aussi diversifié que la France la construction d'un modèle régressif unique et performant s'avère en effet illusoire en raison des multiples facteurs de variabilité qui interfèrent. L'idée est alors d'éliminer au moins partiellement une partie de cette variabilité en regroupant les bassins au comportement similaire en un ensemble de voisinages/régions plus homogènes, pour lesquels seront définis un ensemble de modèles régressifs autonomes.

Les travaux réalisés au cours de l'année 2009 [**Plasse et Sauquet (2009)**] se sont concentrés sur la plus-value de découpages en entités figées contiguës : bassins hydrographiques, zones hydro-éco-régions de différents niveaux pour l'interpolation du débit caractéristiques QMNA5. Or comme le montre le résultat de la classification des régimes d'étiage présentée dans la partie précédente, la contiguïté (*i.e.* la proximité géographique) ne semble pas constituer un facteur de similitude/discrimination suffisant à l'égard des différents types de comportements hydrologiques. Ainsi, dans la continuité des travaux

précédents, nous avons souhaité explorer deux nouvelles approches permettant la construction de région/voisinages non-contiguës dans l'espace.

Les sous-parties suivantes sont destinées à présenter successivement le protocole retenu pour la construction des modèles régressifs, les résultats obtenus pour l'établissement d'un modèle national (à titre de référence), le potentiel d'estimation dans le cas de voisinages idéalement connus et différentes applications dédiées à l'estimation en site non jaugé. Une conclusion générale est enfin proposée pour synthétiser les principaux résultats obtenus.

3.1 Protocole de définition des modèles régressifs, variables explicatives et mesure des performances d'estimation

Pour l'ensemble des procédures d'estimation abordées dans cette partie (couplage entre une méthode de définition de voisinage et la régression linéaire multiple), un même protocole de définition des modèles régressifs a été appliqué de manière à obtenir des relations performantes, robustes et respectant, dans la mesure du possible, les hypothèses statistiques propres à la méthode employée. En résumé ce protocole est le suivant :

- 1) Ajustement des modèles de forme linéaire et puissance par minimisation du critère des moindres carrés pour l'ensemble des combinaisons comprenant de 1 à 4 variables explicatives ;
- 2) Elimination de toutes les combinaisons pour lesquels l'une des variables n'est pas significative (test de Student au niveau de significativité 10 %) comme suggéré par **Aschwanden et Kan (1999)** ;
- 3) Sélection des 20 meilleures combinaisons restantes sur la base du coefficient de détermination R^2 ajusté. Puis pour chaque combinaison sélectionnée :
 - vérification de la normalité et de l'homoscédasticité des résidus (test de Shapiro-Wilk et test de Breush-Pagan au niveau de significativité 10 %) ;
 - élimination récursive des horsains (test sur les résidus normalisés avec un seuil de rejet fixé à 3) comme suggéré par **Laaha et Blöschl (2006)** ;
 - ajustements et comparaison des performances des différents modèles en validation croisée (incluant les horsains qui ne participent pas à l'ajustement des relations) avec remplacement des valeurs non réalistes (*i.e.* valeurs négatives) par zéro.

Le choix du meilleur modèle est finalement réalisé sur la base de la racine carrée de l'écart quadratique moyen RMSE (en anglais, Root Mean Square Error), plus pénalisant à l'égard de la présence de horsains que le coefficient de détermination R^2 .

Les descripteurs du bassin disponible à l'échelle nationale et employés à titre de variables explicatives sont résumés dans le tableau suivant (Tableau 3). Leur extraction en tout point du réseau hydrographique et le calcul de moyennes de bassin se sont appuyés sur le traitement d'un modèle numérique de terrain (MNT) de résolution kilométrique disponible à l'échelle nationale. Ce traitement, réalisé à l'aide du logiciel « Hydrodem⁹ », a permis d'identifier les contours des bassins versants associés à chaque point d'un réseau

⁹ Logiciel développé par E. Leblois l'unité de recherche Hydrologie-Hydraulique du Cemagref de Lyon

hydrographique théorique par l'intermédiaire d'un plan de drainage conçu pour décrire la direction d'écoulement en chaque maille dans un plan horizontal.

Catégorie	Indice	
Localisation	Xcg	Coordonnées du centre de gravité du bassin
	Ycg	
Morphologie (extrait du MNT de résolution kilométrique)	S	Surface du bassin (surface topographique)
	Pente	Pente moyenne du bassin (indice de Roche)
	Z25	Quantile d'altitude à 25, 50 et 75 % de la courbe hypsométrique
	Z50	
	Z75	
	IF	Indice de compacité de Gravelius (déduit du rectangle équivalent)
Climat (Extrait de la ré-analyse SAFRAN Sur la période 1970-2005)	PA	Pluies de bassin moyennes interannuelles
	Pwinter	Pluies de bassin saisonnières moyennes interannuelles
	Pspring	
	Psummer	
	Pautumn	
	VarP	Variabilité (écart-type) des précipitations moyennes mensuelles
	nP0	Nombre de jours sans pluie moyen interannuel
	dP0_max	Durée sans précipitation maximale annuelle moyenne
	TA	Températures moyennes interannuelles
	Twinter	Températures saisonnières moyennes interannuelles
	Tspring	
	Tsummer	
	Tautumn	
	VarT	Variabilité (écart-type) des températures moyennes mensuelles
	ETPsummer	Evapotranspiration potentielle moyenne sur Juillet et Aout
	ETRA	Evapotranspiration réelle interannuelle (Formule de TURC)
Géologie	%imp	Pourcentage du bassin versant à substratum imperméable
	DD	Densité de drainage déterminé sur la base du réseau hydrographique CARTHAGE (longueur totale du réseau ramenée à la surface du bassin)
Occupation des sols (extrait du Corine Land Cover 1990)	%forest	Pourcentage du bassin occupé par la forêt
	%grass	Pourcentage du bassin occupé par la prairie et/ou végétation basse
	%agro	Pourcentage du bassin occupé par les surfaces agricoles
	%urban	Pourcentage du bassin en zone urbaine
	%rock	Pourcentage du bassin sans végétation
	%water	Pourcentage du bassin occupé par les plans d'eau et/ou marais

Tableau 3 : Descripteurs du bassin versant disponibles à l'échelle nationale

Pour chaque modalité de définition de voisinage/région homogène, la mesure des performances sera réalisée de manière globale, c'est-à-dire toutes stations confondues (une analyse plus détaillée des résultats par région pourra être cependant proposée dans certains cas) de manière à permettre une comparaison entre les différentes approches. Les principaux critères retenus pour cela comprennent le critère de NASH (dans ce cas il est utilisé pour quantifier si les valeurs observées et prédites s'alignent sur une droite de prédiction parfaite, un NASH = 1 correspond à une prédiction parfaite, un NASH < 0 témoigne de performances moins efficaces qu'une simple moyenne des valeurs observées), le RMSE (racine de la moyenne des écarts entre observations et prédictions au carré) et le BIAS (moyenne des écarts entre valeurs prédites et valeurs observées, témoignant d'une sous-estimation lorsque BIAS < 0 et d'une surestimation lorsque BIAS > 0). Pour affiner la description des incertitudes en termes de distribution, les percentiles d'erreur absolue pour 10 %, 50 % (médiane) et 90 % de l'échantillon seront également fournis.

En complément des critères numériques, une représentation graphique des valeurs prédites en validation croisée en fonction des valeurs observées sera également proposée pour chaque variable estimée. Afin de pouvoir apprécier qualitativement la variation des erreurs en fonction de la gamme de valeur observée (*i.e.* le degré d'hétéroscédasticité), un tracé supplémentaire donnant l'erreur absolue moyenne pour différentes gammes de valeurs observées (définies telles que chaque gamme de valeur possède un effectif égal à 2 % de l'échantillon total soit environ 10 individus) a été superposé en rouge sur ces graphiques autour de la droite de prédiction parfaite.

3.2 Construction d'un modèle régressif national

On se propose en premier lieu de donner une estimation grossière du QMNA5 et du VCN10(5) à partir d'un modèle régressif national, *i.e.* ajusté sur l'ensemble des sites de référence disponibles. Cette estimation de référence permettra ainsi de mieux évaluer le gain à travailler à l'aide de voisinages/régions pour l'établissement de modèles régressif. Les résultats obtenus pour chacune des variables sont reportés ci-dessous (Tableau 4 et Figure 13).

Modèles ajustés :

$$\text{VCN10(5)} = 5,683 + \text{PA}.1,542.10^{-3} - \text{ETPsummer}.1,554.10^{-2} - \%imp.0,825 + \%rock.7,878$$

Critère		QMNA5VCN 10(5)	
NASH		0.483	0.357
BIAS (l/s/km ²)		-0.146	-0.165
RMSE (l/s/km ²)		1.166	1.086
Erreur absolue (l/s/km ²)	10%	0.116	0.137
	50%	0.780	0.694
	90%	2.411	2.212

Tableau 4 : Performances d'estimation en validation croisée des variables QMNA5 et VCN10(5) spécifiques au moyen d'un modèle régressif national

Compte tenu de la forte corrélation entre les deux variables considérées ici, les modèles ajustés font assez logiquement appel aux mêmes variables explicatives. Ces variables, qui se rapportent au climat (PA, ETPsummer) et à la nature plus ou moins perméable du substratum (%imp et %rock), peuvent être considérées comme une première indication sur les caractéristiques les plus discriminantes vis-à-vis des régimes de basses eaux à l'échelle nationale (parmi les descripteurs de bassin disponibles). En revanche, il apparaît clairement que celles-ci ne suffisent pas à expliquer correctement la variabilité des deux variables étudiées. En effet on constate comme attendu que les performances globales s'avèrent médiocres avec des critères de NASH relativement faibles. D'autre part et malgré un biais relativement faible en moyenne on remarque sur les graphiques une sous-estimation parfois importante des fortes valeurs et, à l'inverse, une sur-estimation presque systématique des faibles valeurs. Parallèlement les graphiques montrent l'existence d'une forte hétéroscédasticité avec un accroissement important de la variance d'erreur pour les valeurs élevées. Tout ceci confirme la difficulté à établir un modèle régressif unique dans un contexte où la variabilité des réponses hydrologiques fait intervenir un trop grand nombre de facteurs différents et illustre aussi le manque de variables de bassin réellement pertinentes à l'échelle nationale.

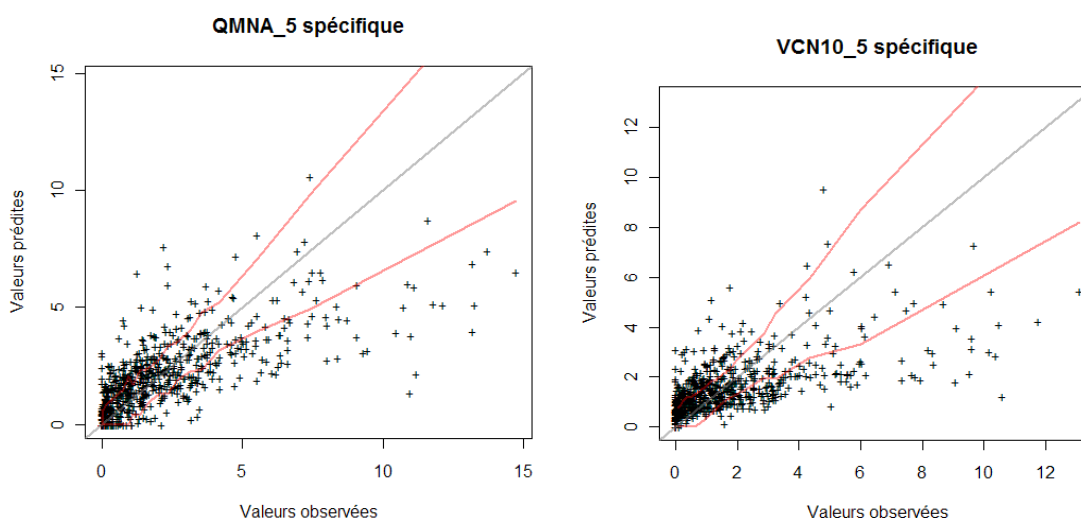


Figure 13 : Valeurs de QMNA5 et VCN10(5) spécifiques (l/s/km²) observées et prédites en validation croisée au moyen d'un modèle régressif national et erreur moyenne par gamme de valeur observée (en rouge)

Bien entendu, eu égard au nombre d'individus de l'échantillon, la limite imposée de 4 variables explicatives semble pénalisante. Toutefois des essais complémentaires avec l'introduction d'un plus grand nombre de variables explicatives n'ont pas montrés de gain significatif sur les performances d'estimation, notamment en raison de la multi-colinéarité existant entre les différents descripteurs physiographiques disponibles.

3.3 Intérêt de la définition de voisinage : démonstration dans le cas de voisinages idéalement connus

On se propose ici de vérifier si la discrimination des bassins au travers de l'espace hydrologique et de la classification des régimes d'étiage construits précédemment (cf. section 3) s'avère efficace vis-à-vis de la construction de modèles régressifs. Autrement dit il s'agit d'examiner le potentiel d'estimation dans le cas de voisinages idéalement connus (*i.e.* pour les bassins instrumentés) avant de procéder à la définition de voisinages dédiés à l'estimation en site non jaugé. Cette étape, généralement négligée dans la littérature, apparaît pourtant comme un préalable important pour :

- servir de point de comparaison aux méthodes de définition de voisinages dédiés à l'estimation en site non jaugé en définissant la borne haute des performances attendues (au même titre que le modèle national qui constitue la borne inférieure) ;
- valider l'utilisation d'un critère hydrologique pour une éventuelle utilisation ultérieure dans les méthodes de définition de voisinage dédiées à l'estimation en site non jaugé (*i.e.* critère d'homogénéité des voisinages).

La première situation envisagée a consisté à établir un modèle régressif pour chacune des 4 classes de régimes hydrologiques d'étiage définies en section 2.3. Les résultats obtenus sont détaillés ci-dessous (Tableau 5 (en gris sont rappelés les résultats de référence obtenus au moyen du modèle national) et Figure 14).

Modèles ajustés pour le QMNA5 :

Classe 1 $\log(\text{QMNA5}) = 18,105 - \log(\text{Ycg}).1,167 + \log(\text{PA}).1,609 - \log(\text{nP0}).2,148 - \log(\%\text{agro}).0,152$

Classe 2 $\text{QMNA5} = 17,437 + \text{PA}.6,639.10^{-3} - \text{TA}.1,017 - \text{VarT}.1,449 - \%\text{agro}.2,394$

Classe 3 $\text{QMNA5} = 2,270 + \text{XG}.7,784.10^{-7} + \text{PA}.2,309.10^{-3} - \text{VarT}.0,390 - \text{ETRA}.3,996.10^{-3}$

Classe 4 $\log(\text{QMNA5}) = -5,860 + \log(\text{S}).0,206 + \log(\text{PA}).1,241 - \log(\text{VarP}).1,159 - \log(\%\text{agro}).0,267$

Modèles ajustés pour le VCN10(5) :

Classe 1 $\log(\text{VCN10(5)}) = 24,513 - \log(\text{Ycg}).1,425 + \log(\text{PA}).1,433 - \log(\text{nP0}).2,456 - \log(\%\text{agro}).0,143$

Classe 2 $\text{VCN10(5)} = 12,172 + \text{PA}.7,355.10^{-3} - \text{TA}^*.0,793 - \text{VarT}.1,251 - \%\text{imp}.1,710$

Classe 3 $\text{VCN10(5)} = 0,0961 + \text{S}.1,841.10^{-4} + \text{PA}.9,995.10^{-4} + \text{VarP}.7,322.10^{-3} - \text{TA}.7,748.10^{-2}$

Classe 4 $\log(\text{VCN10(5)}) = 8,065 + \log(\text{IF}).1,748 - \log(\text{VarP}).1,138 - \log(\%\text{agro}).0,516 + \log(\%\text{urban}).1,314$

Critère		QMNA5VCN10(5)			
NASH		0.785	0.483	0.757	0.357
BIAS (l/s/km ²)		-0.118	-0.146	-0.103	-0.165
RMSE (l/s/km ²)		0.700	1.166	0.615	1.086
Erreur absolue (l/s/km ²)	10%	0.055	0.116	0.057	0.137
	50%	0.432	0.780	0.335	0.694
	90%	1.667	2.411	1.570	2.212

Tableau 5 : Performances d'estimation des variables QMNA5 et VCN10(5) spécifiques obtenues en validation croisée au moyen de modèles régressifs établis pour chaque classe de régime d'étiage (en noir) et résultats de référence obtenus au moyen du modèle national (en gris)

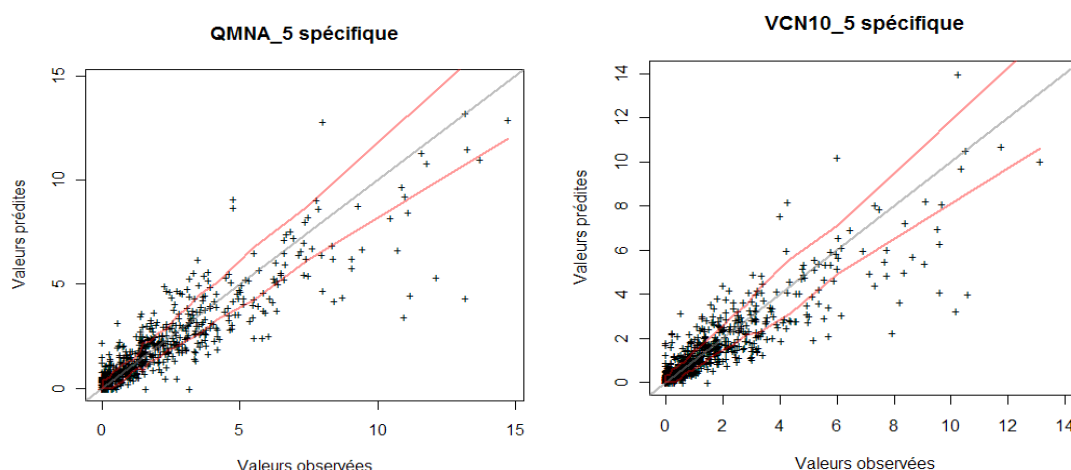


Figure 14 : Valeurs de QMNA5 et VCN10(5) spécifiques (l/s/km²) observées et prédites en validation croisée au moyen d'un modèle régressif établi pour chaque classe de régime d'étiage et erreur moyenne par gamme de valeur observée (en rouge)

Au vu de ces résultats on constate une amélioration très significative des performances de prédiction par rapport aux modèles nationaux. Il semble ainsi que la discrimination réalisée ici au moyen de seulement 4 classes permet bien d'éliminer une part de variabilité difficile à exprimer autrement (*i.e* au travers d'un unique modèle national) et d'expliquer une grande partie de la variabilité résiduelle au moyen des descripteurs de bassin disponibles (caractéristiques de type climatique ou d'occupation des sols essentiellement).

Dans un second temps il a été examiné le résultat des performances d'estimation dans le cas de la sélection de voisinages glissants définis sur la base de la proximité dans l'espace hydrologique (approche dite « ROI » pour « region of influence », **Burn (1990)**) : pour chaque site de l'échantillon, un voisinage de N stations de référence a été défini sur la base de la distance euclidienne dans l'espace hydrologique (*i.e.* constitué par les axes factoriel ACP1, ACP2) puis un modèle régressif a été ajusté sur les données considérées pour fournir une estimation au site cible. Afin de tenir compte de leur importance respective dans la discrimination des régimes d'étiages et le calcul des distances entre bassins, on remarquera que chacune des deux dimensions ACP1 et ACP2 a préalablement été pondérée par la valeur propre

(pourcentage de variance expliquée) qui lui était attribuée lors de l'ACP. Les résultats ci-dessous donnent les performances obtenues pour différentes tailles de voisinage N (Tableau 6 et Figure 15).

Sur ce type de voisinage on constate que le gain vis-à-vis des performances d'estimation est significatif pour la majorité de l'échantillon avec une nette diminution de l'erreur médiane par rapport aux regroupements par classes de régime d'étiage et des RMSE pouvant être équivalents. A l'inverse le coefficient de détermination et le risque de fortes erreurs (percentile d'erreur à 90%) s'avèrent moins satisfaisants. Comme le montrent nettement les graphiques présentés ci-dessous (Figure 15), ces observations résultent de la forte dépendance des résultats à la gamme de valeurs observées (*i.e.* forte hétéroscédasticité). On remarque en effet que les estimations données pour les bassins les moins productifs (correspondants grossièrement aux classes 2 et 3 qui sont aussi les plus représentées au sein de l'échantillon) sont globalement très satisfaisantes tandis que celles obtenues pour les bassins les plus productifs (*i.e.* classe 1 et 4) s'écartent parfois beaucoup des valeurs observées, ce qui aboutit à dégrader sensiblement le NASH. Les bassins concernés qui, pour certains, présentaient déjà des écarts aux observations importants dans l'approche par classe de régime d'étiage, ont été identifiés et examinés en détail. Il s'agit d'une part de bassins karstiques (U2022010, U2604030, V1015030 et W3335210) dont la forte spécificité locale rend inadéquate la définition de similitudes dans l'espace hydrologique pour l'identification de voisinages (il en résulte des régressions inadaptées ou insuffisamment robustes pour assurer une bonne estimation). D'autre part, un certain nombre de mauvaises estimations sont associées à des bassins influencés mais conservés dans l'échantillon dans la mesure où les débits fournis dans la Banque HYDRO étaient qualifiés de « naturels reconstitués » (O0105110, O6332510, W0224010, W2714010 et Y5615020). Ces stations n'ont pas été remises en cause pour le moment¹⁰.

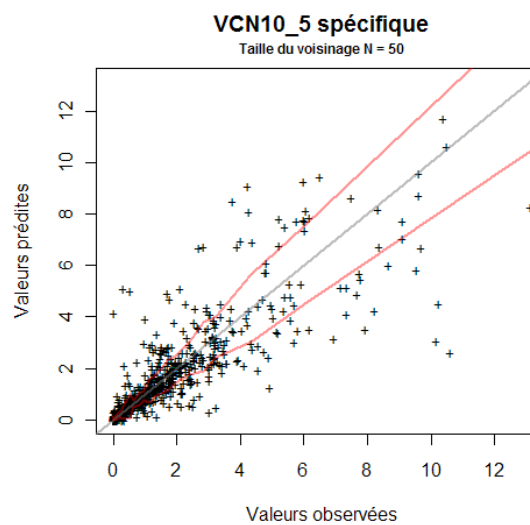
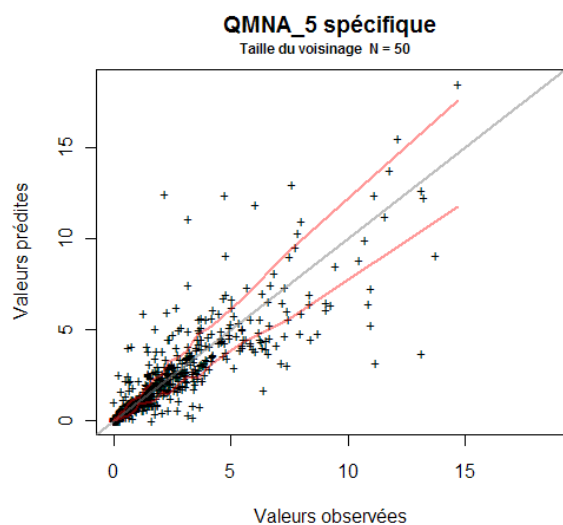
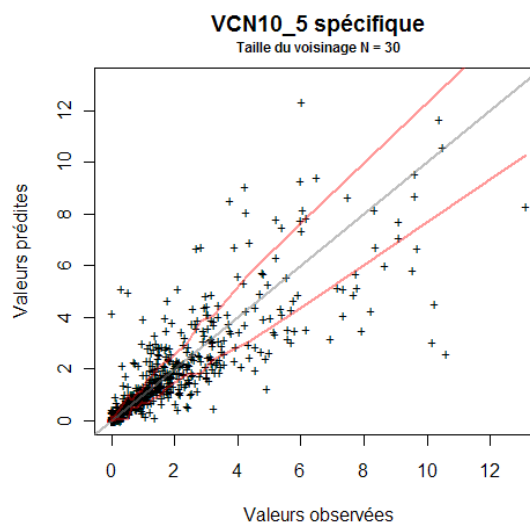
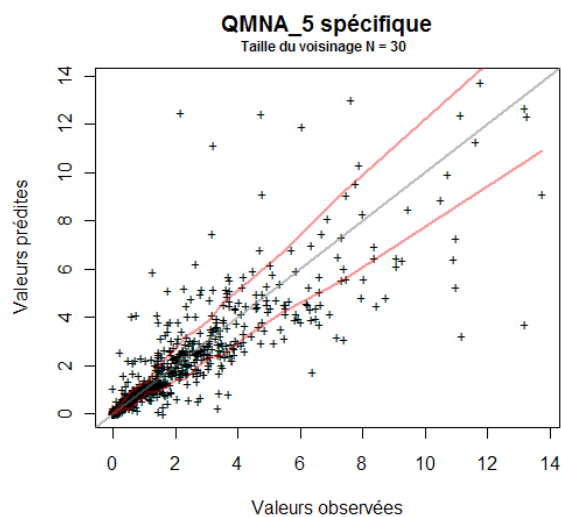
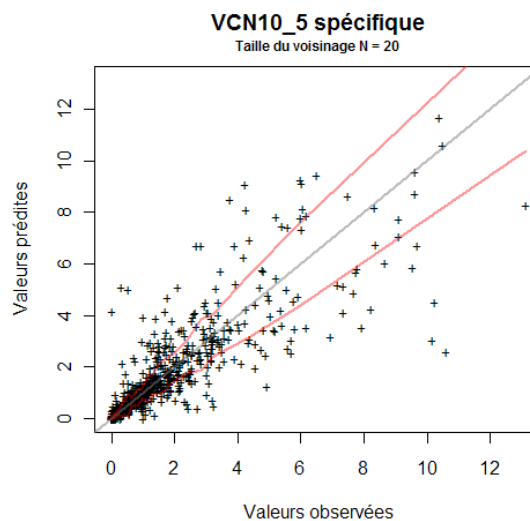
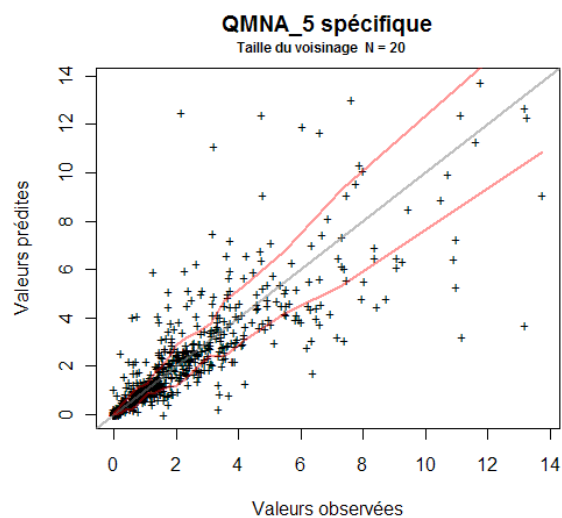
En termes de taille de voisinage, on remarque que les scores évoluent peu et qu'il n'apparaît pas de lien particulièrement explicite entre les performances et le nombre de sites voisins employés pour ajuster les modèles régressifs. Néanmoins lorsque l'on considère tous les critères confondus, les résultats suggèrent qu'une taille de voisinage comprise entre 50 et 70 est susceptible de fournir les estimations les plus satisfaisantes.

¹⁰ Une discussion avec les équipes d'Aix-en-Provence et d'Antony sera engagée pour déterminer si des problèmes analogues ont été mis en évidence dans leurs travaux respectifs

Critère		QMNA5					
Taille du voisinage (N)		20	30	50	70	100	National
NASH		0.661	0.678	0.693	0.680	0.663	0.483
BIAS (l/s/km ²)		-0.016	-0.071	-0.024	-0.064	-0.107	-0.146
RMSE (l/s/km ²)		0.737	0.719	0.711	0.700	0.731	1.166
Erreur absolue (l/s/km ²)	10%	0.030	0.031	0.023	0.030	0.030	0.116
	50%	0.310	0.283	0.281	0.288	0.274	0.780
	90%	1.982	1.823	1.905	1.753	1.982	2.411

Critère		VCN10(5)					
Taille du voisinage		20	30	50	70	100	National
NASH		0.644	0.636	0.656	0.642	0.626	0.357
BIAS (l/s/km ²)		-0.026	-0.039	-0.047	-0.064	-0.099	-0.165
RMSE (l/s/km ²)		0.627	0.634	0.605	0.600	0.606	1.086
Erreur absolue (l/s/km ²)	10%	0.021	0.018	0.016	0.020	0.021	0.137
	50%	0.236	0.238	0.221	0.238	0.228	0.694
	90%	1.748	1.706	1.695	1.573	1.678	2.212

Tableau 6 : Performances d'estimation du QMNA5 et du VNC10(5) spécifiques obtenus en validation croisée au moyen de modèles régressifs établis pour des voisinages de différentes tailles sur la base de la proximité dans l'espace hydrologique (en noir) et résultats de référence obtenus au moyen du modèle national (en gris)



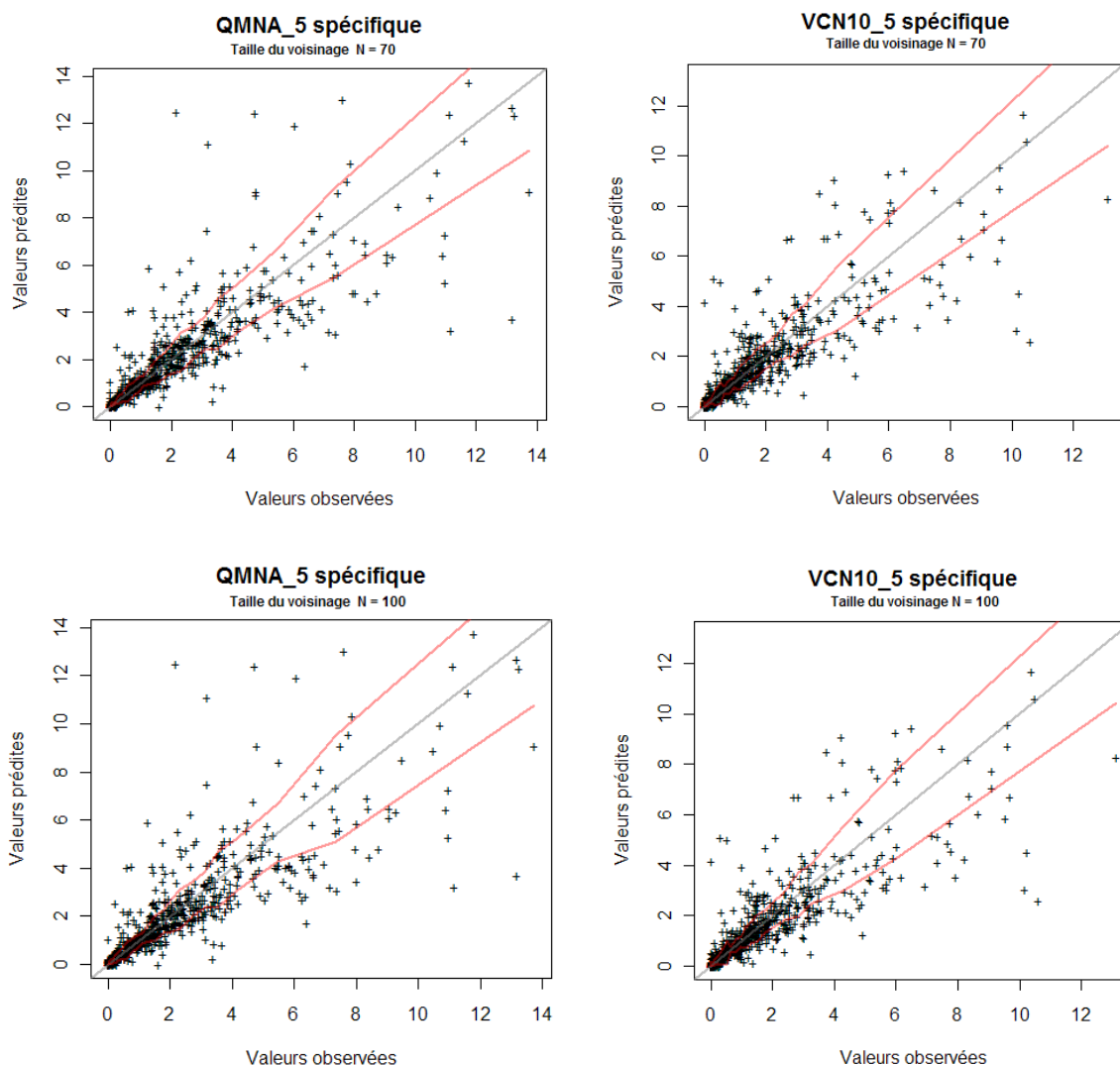


Figure 15 : Valeurs de $QMNA_5$ et $VCN10(5)$ spécifiques ($l/s/km^2$) observées et prédites en validation croisée au moyen de modèles régressifs établis pour des voisinages de différentes tailles déterminés sur la base de la proximité dans l'espace hydrologique et erreur moyenne par gamme de valeur observée (en rouge)

Pour résumer, compte tenu des éléments rapportés ici, il semble que la définition de similitudes entre bassins sur la base de critères hydrologiques (sous la forme de classe de régime ou de proximité dans l'espace hydrologique) s'avère globalement favorable à une estimation au moyen de modèles régressifs et fournit ainsi un résultat de référence pertinent avant l'identification de voisinages dédiés à l'estimation en sites non jaugés. Bien entendu pour ces derniers, l'absence d'observations à caractère hydrologique ne permet pas d'exploiter directement les résultats précédents pour parvenir à la définition de voisinages. Il est alors nécessaire de faire appel à un certain nombre de variables descriptives des bassins, qualitatives ou quantitatives (e.g. climat, relief, géologie...), pour établir des similitudes entre bassins et identifier des voisinages. Pour ce faire plusieurs méthodes ont été envisagées avec la volonté d'explorer les trois grandes catégories de voisinage (région « figées » à forte cohérence géographique, régions « figées » éclatées dans l'espace et région d'influence spécifique à chaque site cible (ROI)). Celles-ci sont présentées dans les parties suivantes.

3.4 Estimation sur la base de l'appartenance à une HER de niveau I

Afin de disposer d'un point de comparaison avant la mise en œuvre de méthodes plus élaborées, il a été tout d'abord choisi d'appliquer la procédure de RLM sur un ensemble de 21 régions prédéfinies par la typologie en Hydro-eco-regions (HER) de niveau I [Wasson et al. (2001)]. Cette typologie désigne un ensemble d'entités géographiques définies comme ayant des caractéristiques communes, principalement en termes de relief, de géologie et de climat, à l'origine d'un fonctionnement biologique des milieux aquatiques similaire (Figure 16). Dans cette application il est fait l'hypothèse que ces ensembles géographiques permettent également de discriminer les comportements hydrologiques d'étiage.

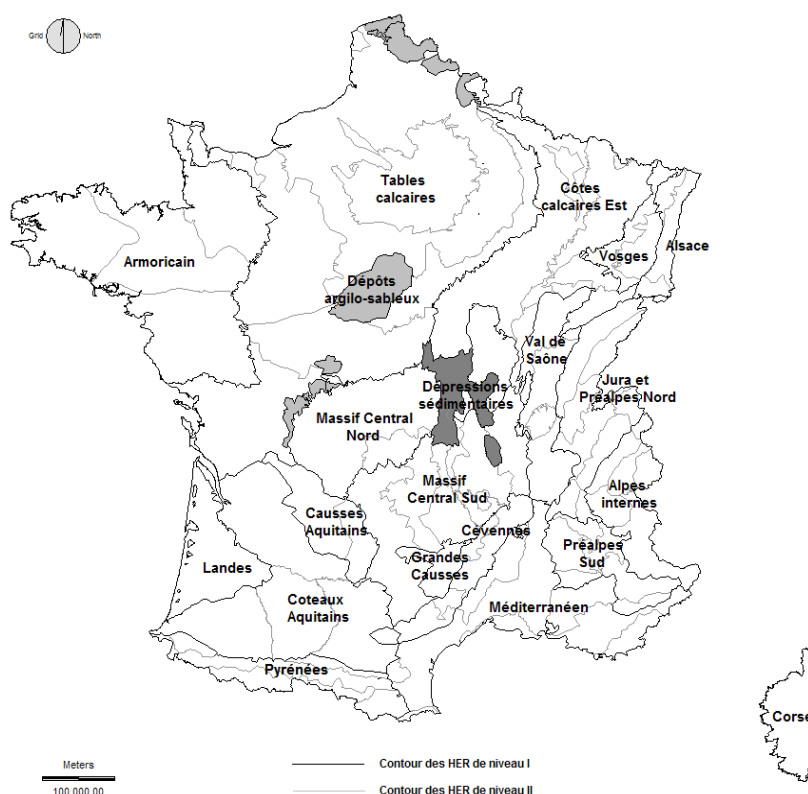


Figure 16 : Délimitation des HER de niveau I et II proposée par [Wasson et al. (2001)]

L'approche adoptée est identique à l'approche désignée par « variante P2 » dans les travaux de **Plasse et Sauquet (2010)** : pour chacune des HER, un modèle régressif est ajusté sur l'ensemble de l'échantillon en donnant un poids plus ou moins important aux individus selon la proportion de surface de bassin recoupant l'HER considérée (on se réfère ici à la technique de régression pondérée « WLS »)¹¹. On notera cependant que, pour certaines HER, le nombre de stations participant effectivement à l'ajustement s'avère très faible (< 10), voire insuffisant pour permettre cet ajustement (< 4). Il a donc été nécessaire de procéder à des regroupements de manière à obtenir un minimum de 20 sites par région¹². Finalement 16

¹¹ Un bassin totalement inclus dans l'HER possède un poids maximum tandis qu'un bassin hors de l'HER a un poids nul et n'intervient pas dans la procédure d'ajustement. L'ensemble des critères de performance par région tiennent également compte des pondérations associées à chaque individu.

¹² Ces regroupements sont les suivants : Landes/Coteaux aquitains, Corse/Méditerranéen, Dépressions sédimentaires/Val de Saône, Alsace/Vosges et Dépôts argilo-sableux/Tables calcaires

entités distinctes (notées HERm pour HER modifiées) sont utilisées ici avec un effectif moyen de 52 individus.

A l'issue des ajustements, une valeur peut être fournie pour chaque site de l'échantillon de référence à partir d'une combinaison de toutes les estimations obtenues en validation croisée pour chaque HER, pondérées par les proportions de surface correspondantes. Les résultats obtenus sont reportés ci dessous (Tableau 7 et Figure 17) en rappelant dans le Tableau 7 les performances obtenues au moyen d'un unique modèle national (en gris) et au moyen des modèles établis pour les 4 classes de régime d'étiage (en rouge).

Critère		QMNA5VCN10(5)				
NASH		0.684	0.785	0.483	0.665	0.357
BIAS (l/s/km ²)		-0.304	-0.118	-0.146	-0.261	-0.165
RMSE (l/s/km ²)		0.803	0.700	1.166	0.698	1.086
Erreur absolue (l/s/km ²)	10%	0.061	0.055	0.116	0.036	0.137
	50%	0.419	0.432	0.780	0.381	0.694
	90%	1.923	1.667	2.411	1.630	2.212

Tableau 7 : Performances d'estimation des variables QMNA5 et VCN10(5) spécifiques obtenues en validation croisée au moyen d'un modèle régressif établi pour chaque HERm (en noir) et résultats de référence obtenus au moyen de modèles établis pour chaque classe de régime d'étiage (en rouge) et à l'échelle nationale (en gris).

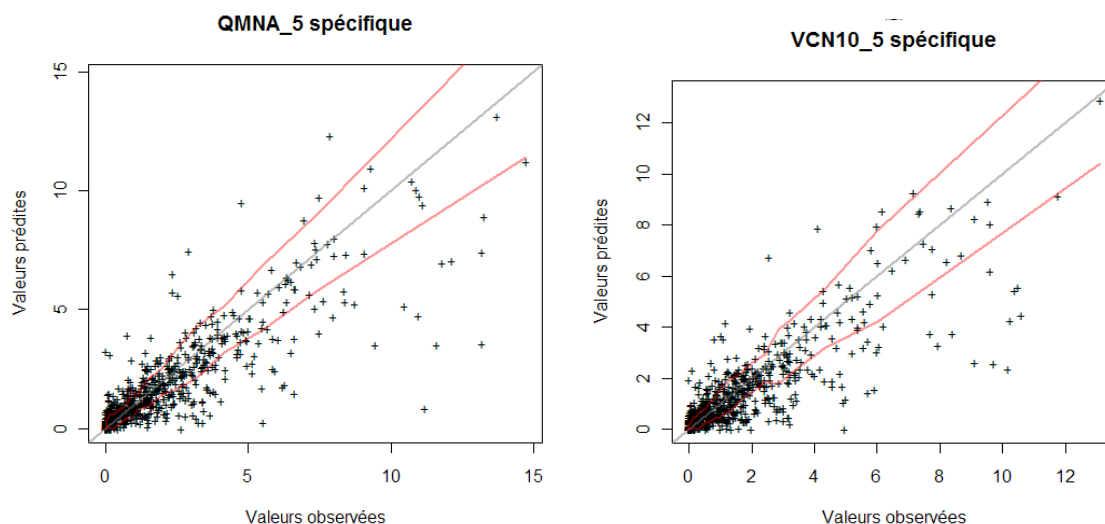


Figure 17 : Valeurs de QMNA5 et VCN10(5) spécifiques (l/s/km²) observées et prédites en validation croisée au moyen d'un modèle régressif établi pour chaque HERm et erreur moyenne par gamme de valeur observée (en rouge)

Bien que n'atteignant pas les scores issus de l'estimation réalisée à partir des voisinages idéalement connus, les performances ici obtenues s'avèrent plutôt satisfaisantes par rapport au modèle national et confirment la pertinence des regroupements par HERm vis-à-vis de l'élimination d'une partie de la variabilité des comportements hydrologiques à grande échelle. On constate cependant que de forts écarts

aux valeurs observées peuvent à nouveau se produire pour certains sites. Il s'agit en grande partie des sites déjà identifiés précédemment comme karstiques ou ayant des données de débit re-naturalisées.

3.5 Couplage CART-RLM

Afin d'approfondir la problématique de définition de voisinage/région il a été choisi d'examiner une méthode de classification automatique permettant la construction de régions homogènes non contiguës dans l'espace. Parmi les différentes méthodes relevées dans la littérature la méthode de classification par arbre de régression (CART, voir **Laaha et Blöschl (2006)**, **Vezza et al. (2010)** pour une application aux variables d'étiage) nous est apparu comme la plus avantageuse et la plus souple. Elle permet en effet l'introduction d'une information hydrologique explicite dans la procédure de regroupement des bassins et facilite grandement l'affectation des sites non jaugés dans les différents groupes.

Dans cette méthode la classification s'effectue par partitionnement récursif binaire de l'échantillon en séparant successivement les individus de manière conditionnelle en groupes non recouvrants de plus en plus homogènes vis-à-vis d'une ou plusieurs variables hydrologiques (minimisation de la somme des écarts quadratiques à la moyenne). En termes plus pratiques : l'algorithme explore par étapes successives toutes les partitions possibles d'un espace constitué de N variables et produit une série de règles de décision du type « si $P(x) < 534$ mm alors l'individu i appartient au groupe A sinon il appartient au groupe B » permettant de maximiser l'homogénéité dans chaque groupe vis-à-vis du terme cible choisi. Chaque étape conduit à l'individualisation d'un nouveau groupe à partir des précédents. Ce mode de construction permet de placer aux premiers niveaux de décision les variables auxiliaires les plus discriminantes à grande échelle et fournit ainsi une hiérarchie explicite des facteurs les plus importants vers les moins importants.

Dans notre application, le package R proposé par **Therneau et al. (2010)** a été utilisé. Il a été laissé à l'algorithme une totale liberté dans le choix des variables de décision parmi l'ensemble de descripteurs du bassin disponibles, à l'exception des variables de localisation du bassin (X_{cg} et Y_{cg}) afin de ne pas contraindre la construction des régions sur une base explicite de proximité géographique. D'autre part, afin de limiter de manière automatique le nombre de groupes/régions créés mais aussi d'obtenir dans chacun d'entre eux un effectif suffisant pour permettre d'assurer la robustesse des modèles régressifs régionaux, l'algorithme a été contraint de s'arrêter dès la création d'un groupe contenant un effectif minimal de 20 individus (soit au moins 5 fois plus d'individus que de variables explicatives). On notera enfin que les 21 bassins de la classe 4 n'ont pas été introduits dans la procédure et ont été considérés comme une classe à part entière. Pour ces bassins, une brève analyse des données a logiquement révélé une forte dépendance de la variabilité de l'indice SR à la température moyenne annuelle de bassin (corrélation de -0.72) et indirectement avec les variables d'altitude, de proportion de précipitation solide et la proportion de bassins sans végétation. Pour discriminer les bassins non jaugés susceptibles d'appartenir à cette classe il a donc été choisi de se baser sur un critère (arbitraire mais relativement robuste) d'altitude moyenne Z50 fixé à 1400 m (les bassins instrumentés de la classe concernée s'échelonnent entre 1420 et 2615 m d'altitude) qui constituera le premier niveau de décision de l'arbre de classification.

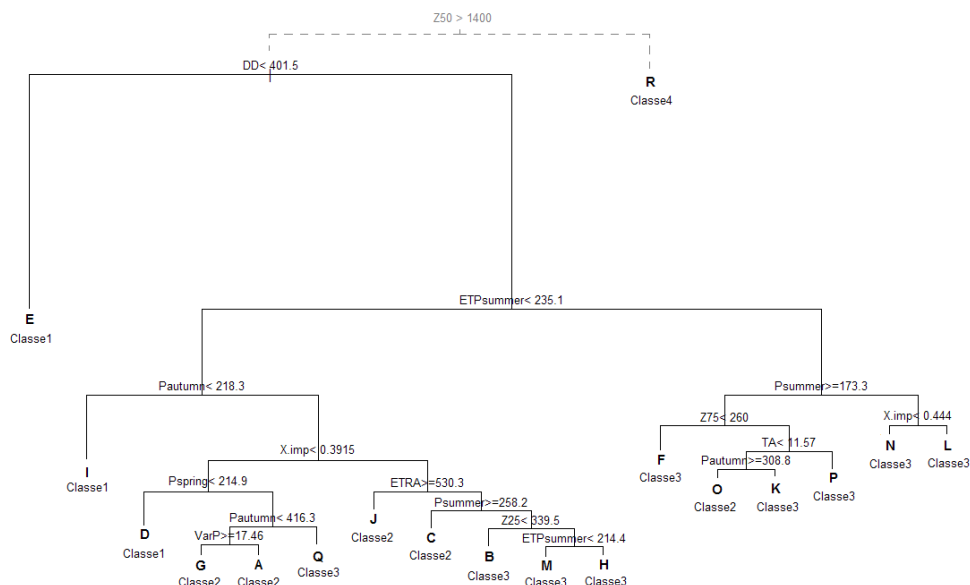


Figure 18 : Arbre de classification obtenu en considérant la première dimension de l'espace hydrologique comme critère d'homogénéité (la classe dominante est indiquée sous chaque groupe)

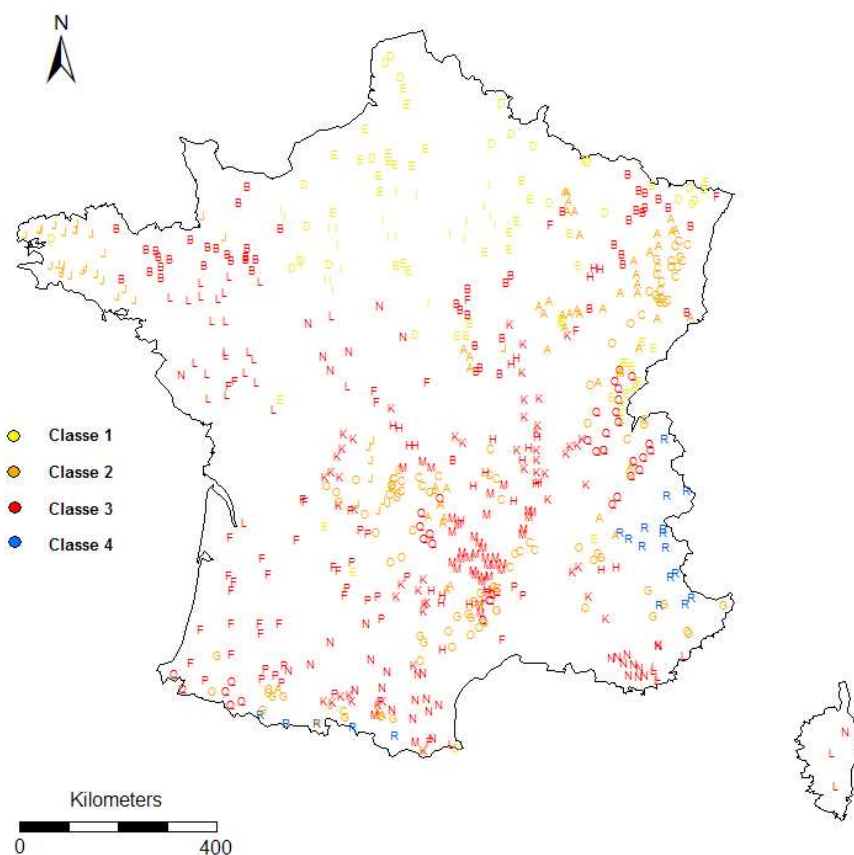


Figure 19 : Répartition des régions établies par CART en considérant la première dimension de l'espace hydrologique comme critère d'homogénéité (la classe dominante de chaque groupe est indiquée à l'aide de couleurs différentes)

Ci-dessus est présenté l'arbre de classification obtenu en proposant comme critère d'homogénéité pour la constitution des groupes la première dimension de l'espace hydrologique *ACP1* (Figure 18). Cette classification a abouti à la constitution de 18 groupes (si l'on inclue le groupe de bassins de la classe 4) avec un effectif moyen de 35 individus. D'après la cartographie donnée sur la Figure 19, il est intéressant de noter que, si la plupart des groupes sont fortement éclatés dans l'espace, la classification permet aussi de retrouver certains ensembles cohérents avec ceux décrits dans la section 3.3. Ainsi la partie occidentale de la Bretagne (région J), les différentes auréoles du bassin parisien (régions I E et D, calquées sur la géologie), la Vendée (région L) semblent bien s'individualiser. En termes de variables de décision, on retrouve aux premiers niveaux les variables se rapportant à la nature plus ou moins perméable du substratum (*i.e.* densité de drainage) à l'évapotranspiration et aux précipitations. Il est intéressant de noter la forte similitude entre la nature de ces variables et celles utilisées dans le cadre du modèle régressif national (cf. section 3.2). Ceci confirme l'importance de ces variables vis-à-vis de la discrimination des bassins en termes de régimes d'étiage.

Un modèle régressif autonome a été ajusté pour chaque groupe/région identifié au moyen de cette classification. Une estimation est alors obtenue en validation croisée pour tous les individus d'un même groupe puis ces estimations sont rassemblées pour calculer les performances globales à l'échelle nationale. Les résultats obtenus sont reportés ci-dessous (Tableau 8 et Figure 20). L'analyse de ceux-ci indique que les estimations s'avèrent relativement satisfaisantes avec un gain significatif par rapport au modèle national mais des performances comparables à celles obtenues dans le cas du regroupement par HERm, voire un peu moins bonnes. Il apparaît donc que, malgré sa souplesse en termes de définition de voisinage (voisinages non contraints par la proximité géographique), sa nature empirique (classification basée sur les données) et son objectivité, l'utilisation de la procédure CART n'apporte pas ici de gain réellement significatif par rapport à une typologie d'expert.

Critère		QMNA5VCN10(5)					
NASH		0.733	0.785	0.483	0.699	0.757	0.357
BIAS (l/s/km ²)		-0.146	-0.118	-0.146	-0.131	-0.103	-0.165
RMSE (l/s/km ²)		0.801	0.700	1.166	0.725	0.615	1.086
Erreur absolue (l/s/km ²)	10%	0.048	0.055	0.116	0.053	0.057	0.137
	50%	0.454	0.432	0.780	0.414	0.335	0.694
	90%	1.960	1.667	2.411	1.761	1.570	2.212

Tableau 8 : Performances d'estimation des variables *QMNA5* et *VCN10(5)* spécifiques obtenues en validation croisée au moyen de modèles régressifs établis pour chaque régions identifiées par la procédure CART (en noir), et résultats de référence obtenus au moyen de modèles établis pour chaque classe de régime d'étiage (en rouge) et à l'échelle nationale (en gris)

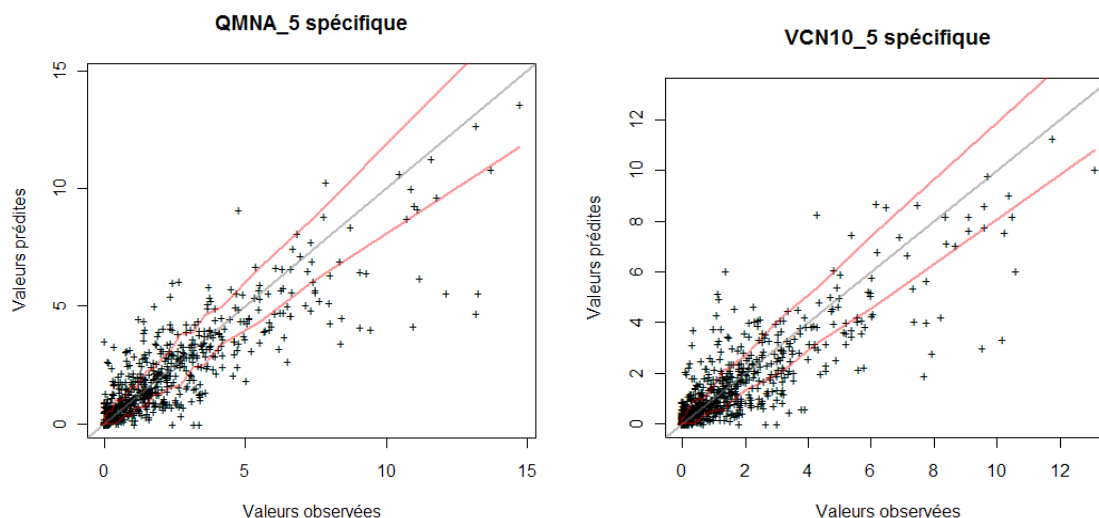


Figure 20 : Valeurs de *QMNA5* et *VCN10(5)* spécifiques ($l/s/km^2$) observées et prédites en validation croisée au moyen de modèles régressifs établis pour chaque régions identifiées par la procédure CART et erreur moyenne par gamme de valeur observée (en rouge)

A titre complémentaire, il a été choisi d'examiner la possibilité d'exploiter la cartographie des régimes hydrologiques mensuels, réalisée par **Sauquet (2006)** (cf. section 2), en considérant cette information comme une variable susceptible d'aider à l'identification de régions homogènes (il s'agit d'une variable qualitative) malgré qu'elle soit issue d'une étape d'estimation préalable et s'accompagne d'un certain degré d'incertitude. On remarquera que, pour l'échantillon de référence, un examen visuel de l'adéquation entre les régimes observés et prédits montre une bonne concordance, au moins pour les grands types de régimes (*i.e.* nappe, pluvial, pluvio-nival, nival et méditerranéen). Pour l'application envisagée ici les individus appartenant aux classes de régime Nivo-Pluvial, Nival de transition, Nival et Nivo-glaciaire ont été regroupés pour ne former qu'une seule classe contenant un effectif suffisant (22 bassins correspondants, à une exception près, aux 21 bassins à régime nival individualisés précédemment sur la base de l'indice SR). Le nombre de classes est ainsi ramené à 9 au lieu de 12 dont un régime « uniforme » (ou régime de nappe, noté Parde1), cinq régimes pluviaux *stricto sensu* (notés Parde 2 à Parde 6), un régime « méditerranéen » (noté Parde 7), un régime pluvio-nival (noté Parde 8) et le régime Nival *lato sensu* (noté Parde 9).

Pour valoriser cette classification, l'information sur le régime hydrologique mensuel a été introduite dans la procédure CART (*i.e.* en guise de variable de décision) en utilisant la même procédure que précédemment. Lorsque cette variable est proposée à l'algorithme, trois des classes de régime sont retenues aux deux premiers niveaux de décision, ce qui semble démontrer, dans une certaine mesure, qu'elles permettent une bonne discrimination de certains régimes hydrologiques d'étiage¹³ (Figure 21). La classification qui en résulte comprend 20 groupes possédant un effectif moyen de 32 individus. Les

¹³ On note qu'au premier niveau, la procédure CART distingue les individus de la classe de régime uniforme (régime de nappe, du bassin parisien essentiellement). Sur le reste de l'échantillon il est ensuite fait une distinction entre deux groupes sur la base de l'appartenance ou non à deux régimes pluviaux (régimes pluviaux type A et B) au sein desquels les regroupements se poursuivent à l'aide des autres descripteurs de bassins.

Critère		QMNA5VCN10(5)					
NASH		0.751	0.785	0.483	0.742	0.757	0.357
BIAS (l/s/km²)		-0.098	-0.118	-0.146	-0.076	-0.103	-0.165
RMSE (l/s/km²)		0.751	0.700	1.166	0.677	0.615	1.086
Erreur absolue (l/s/km²)	10%	0.059	0.055	0.116	0.048	0.057	0.137
	50%	0.459	0.432	0.780	0.395	0.335	0.694
	90%	1.880	1.667	2.411	1.630	1.570	2.212

Janvier 2012

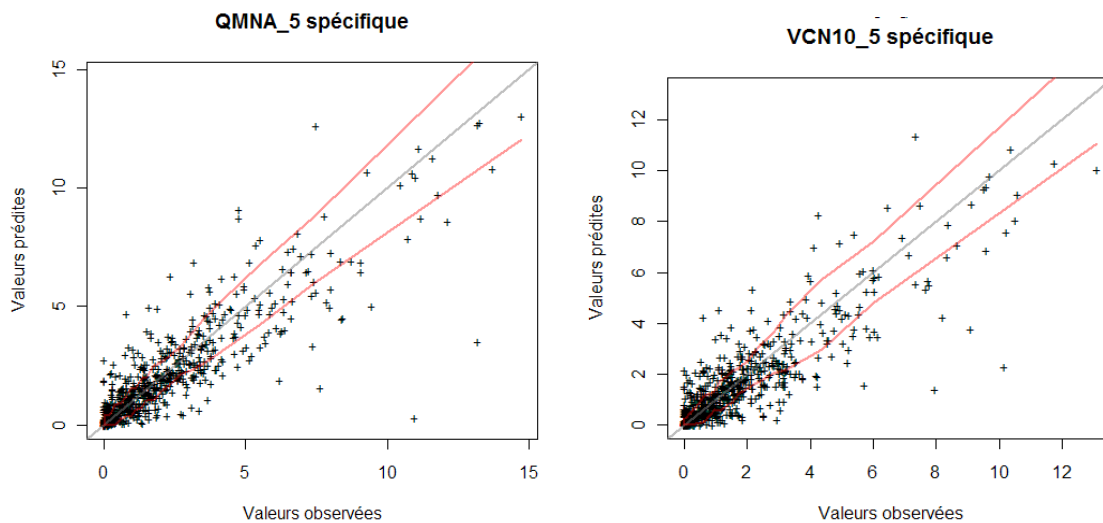


Figure 22 : Valeurs de $QMNA5$ et $VCN10(5)$ spécifiques ($l/s/km^2$) observées et prédites en validation croisée au moyen de modèles régressifs établis pour chaque régions identifiées par la procédure CART avec introduction de l'information sur le régime hydrologique mensuel dans les variables de décision et erreur moyenne par gamme de valeur observée (en rouge)

Il est possible de considérer qu'il s'agit là des meilleurs résultats obtenus jusqu'à présent pour l'estimation du $QMNA5$ et du $VCN10(5)$ en site non jaugés dans la mesure où la procédure utilisée ici surpasse les autres méthodes pour l'ensemble des critères de performance examinés, à l'exception d'une faible dégradation du critère d'erreur médiane par rapport à l'estimation réalisée sur la base des HERm. On précisera que ce gain pourrait en partie être expliqué par le nombre de régions plus élevé que dans les autres approches, cependant il apparaît plus probable que l'utilisation d'une information qualitative à caractère hydrologique (*i.e.* le régime hydrologique mensuel) pour réaliser la classification par arbre de régression permet de mieux discriminer les régimes d'étiage.

3.6 Couplage ROI-RLM

La troisième et dernière approche examinée ici s'appuie sur l'identification de voisinages spécifiques au site cible (approche dite « region of influence » (ROI)) : pour chaque bassin non jaugé est déterminé un ensemble de sites de référence voisins sur la base de la distance euclidienne mesurée dans un espace approprié ; un modèle régressif est ensuite construit sur ce voisinage pour fournir une estimation au site cible considéré (*i.e.* un modèle autonome est donc construit pour chaque bassin considéré).

Afin de disposer d'un point de comparaison, la même approche, appliquée sur un voisinage idéalement connu (*i.e.* dans l'espace hydrologique) a été abordé en section 3.3, permettant au passage de définir une taille de voisinage optimale comprise entre $N = 50$ et $N = 70$. Pour l'application réalisée ici il a été choisi de fixer ce critère à $N = 50$ de manière à s'approcher des tailles de voisinages employées dans le cas des regroupements par HERm ou issus de la procédure CART. Il a également été choisi de montrer les résultats obtenus dans le cas d'un voisinage naïf identifié sur la base de la proximité géographique (distance entre bassins au sens de **Gottschalk (1993)**).

3.6.1 Estimation à partir de voisinages déterminés sur la base de la proximité géographique

Les résultats obtenus toutes stations confondues sont reportés ci-dessous de la même manière que précédemment (Tableau 10 et Figure 23). Au vu de ces résultats on constate que l'application des méthodes de régression multiple sur un voisinage glissant déterminé sur la base de la proximité géographique est loin de permettre l'obtention de performances satisfaisantes et ne surpasse que de peu les résultats obtenus au moyen d'un modèle national (en gris dans le Tableau 10). Ceci confirme la forte hétérogénéité des comportements hydrologiques pouvant exister sur des bassins pourtant proches et plaide en faveur la nécessité de construire un espace plus approprié pour décrire la similitude entre bassins.

Critère		QMNA5VCN10(5)					
NASH		0.474	0.693	0.483	0.388	0.656	0.357
BIAS (l/s/km ²)		-0.148	-0.024	-0.146	-0.116	-0.047	-0.165
RMSE (l/s/km ²)		1.109	0.711	1.166	0.990	0.605	1.086
Erreur absolue (l/s/km ²)	10%	0.088	0.023	0.116	0.068	0.016	0.137
	50%	0.628	0.281	0.780	0.537	0.221	0.694
	90%	2.549	1.905	2.411	2.344	1.695	2.212

Tableau 10 : Performances d'estimation des variables QMNA5 et VCN10(5) spécifiques obtenus en validation croisée au moyen de modèles régressifs établis pour des voisinages déterminés sur la base de la proximité dans l'espace géographique (noir), résultats de référence obtenus au moyen de modèles établis pour des voisinages déterminés sur la base de la proximité dans l'espace hydrologique (en rouge) et à l'échelle nationale (en gris)

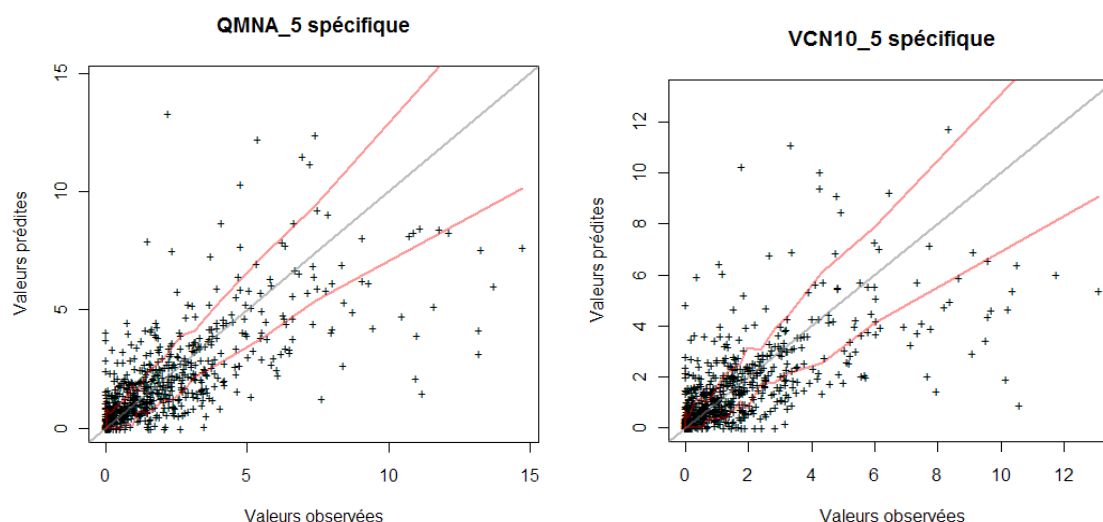


Figure 23 : Valeurs de QMNA5 et VCN10(5) spécifiques (l/s/km²) observées et prédites en validation croisée au moyen de modèles régressifs établis pour des voisinages déterminés sur la base de la proximité dans l'espace géographique et erreur moyenne par gamme de valeur observée (en rouge)

3.6.2 Estimation à partir de voisinages déterminés sur la base de la proximité dans un espace physiographique

L'analyse réalisée ci-dessous vise à identifier, parmi les différents descripteurs du bassin versant disponibles (Tableau 3), les variables susceptibles de permettre la bonne discrimination des régimes hydrologiques d'étiage avec, pour finalité, la construction d'un espace physiographique permettant une bonne identification des voisinages en site non jaugés. On rappellera que cette analyse exclut les 21 bassins à régime nival, toujours discriminés sur la base d'un seuil d'altitude ($Z_{50} > 1400$ m, cf. section 3.5). Dans cette application le voisinage de ces bassins sera donc par défaut toujours constitué des sites de référence appartenant à la même classe.

Pour les bassins restants, les deux représentations données ci-dessous fournissent une vision plus ou moins synthétique mais équivalente des relations entre variables physiographique, climatiques, géologiques... et les deux premières dimensions de l'espace hydrologique issu de l'ACP des variables caractéristiques d'étiage (Figure 24 et Figure 25).

Hormis la forte redondance d'information entre certaines des variables descriptives des bassins, il apparaît qu'aucun lien de corrélation réellement significatif ne ressort vis-à-vis des deux dimensions de l'espace hydrologique *ACP1* et *ACP2* (aucune corrélation ne dépasse 0.4), témoignant ainsi de la difficulté à identifier un (ou plusieurs) descripteur(s) réellement pertinent(s) en termes de discrimination des régimes hydrologiques d'étiage à l'échelle nationale. On constate néanmoins que les variables se rapportant à l'évapotranspiration (et indirectement à la température) et à la nature plus ou moins perméable du substratum (%imp, DD et %urban) sont sans surprise les plus informatives vis-à-vis de *ACP1*. On note également l'importance de la variable *Ycg* qui permet probablement d'expliquer une part de variabilité n'apparaissant pas au sein des autres variables. Pour la seconde dimension *ACP2* une relation apparaît préférentiellement avec les caractéristiques se rapportant aux précipitations, la proportion de bassin occupé par les surfaces agricoles ou forestières et la densité de drainage.

Figure 24 : Matrice des corrélations entre variables hydrologiques transformées issue de l'ACP des variables caractéristiques d'étiage et descripteurs du bassin

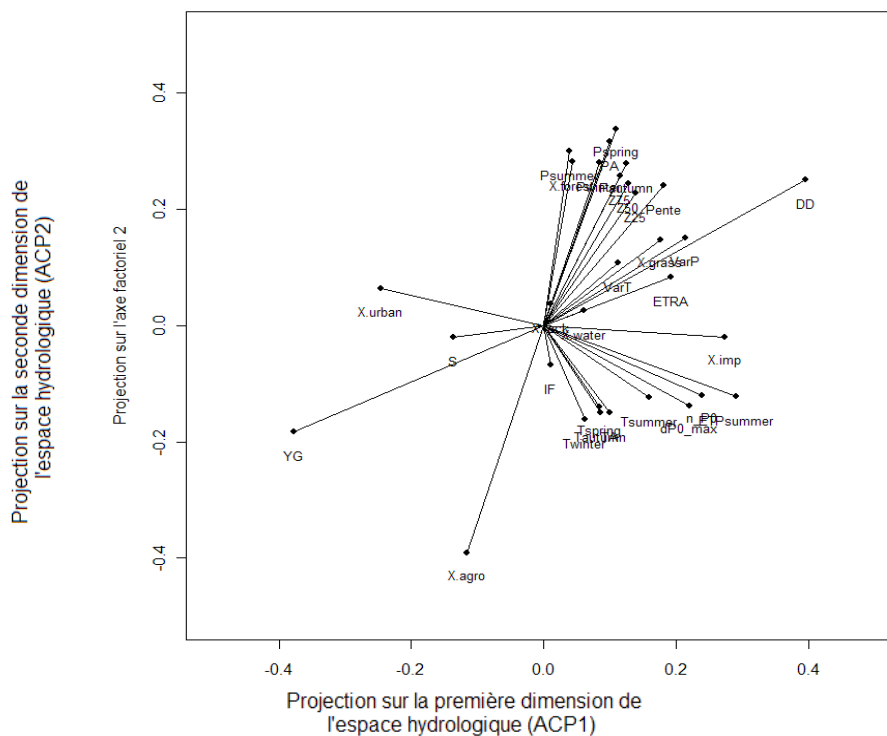


Figure 25 : Projection des variables descriptives du bassin dans l'espace hydrologique issu de l'ACP des variables caractéristiques d'étiage

Afin d'exploiter au mieux ces données il peut-être envisagé de construire un ensemble de combinaisons linéaires de variables pouvant se substituer aux différents descripteurs du bassin de manière à définir un nouveau référentiel (espace physiographique) s'approchant le plus possible de l'espace hydrologique utilisé comme référence pour la définition de voisinage (*cf.* section 2.2). Pour ce faire un certain nombre de techniques ont été déjà explorées dans la littérature (régression multiple classique ou pas-à-pas, ACP des variables physiographiques [Nathan et McMahon (1990)], analyse canonique des corrélations (ACC) [Ouarda et al. (2001) et Haché et al. (2002)]).

Les résultats obtenus en établissant un simple modèle régressif national n'étant pas apparus satisfaisants pour l'élaboration de l'espace physiographique (résultats non présentés ici), il a été choisi ici de mettre en œuvre une procédure plus perfectionnée de régression dite « PLS » (« partial least square », Wold et al. (1984)), technique qui peut être assimilée à la réalisation d'une ACP des variables physiographiques mais sous contrainte d'optimiser l'explication de la variance d'un ou plusieurs terme(s) cible(s), ici représentés par les deux premières dimensions de l'espace hydrologique *ACP1* et *ACP2*. Les combinaisons linéaires (*i.e.* les variables explicatives de substitution) ainsi construites possèdent la propriété d'être indépendantes deux à deux et d'éliminer ainsi le risque de multi-colinéarité. D'autre part le choix des variables explicatives est ici substitué par le choix du nombre de dimensions significatives.

On remarquera que, dans sa finalité, la procédure adoptée ici (ACP des variables hydrologiques couplée à la régression PLS) s'avère en apparence très voisine de la technique d'ACC appliquée par **Ouarda et al. (2001)** et **Haché et al. (2002)** dans le cadre de l'estimation des quantiles de crue. Ainsi dans les deux cas sont construits deux jeux de variables transformées, définis de manière à renforcer les liens de dépendances entre les deux ensembles et minimiser les interactions entre variables d'un même ensemble (*i.e.* orthogonalité des variables d'un même ensemble). Néanmoins alors que l'ACC peut conduire à une déformation des deux espaces d'origines (à divers degrés), la procédure PLS préserve l'espace hydrologique d'origine (et des voisinages qui en découlent) ce qui nous a semblé plus favorable pour l'application que l'on se propose de mener.

D'après le graphique ci-dessous (Figure 26), qui donne les niveaux de corrélation atteints grâce à la procédure PLS (en validation croisée) entre chacune des deux dimensions des espaces physiographique et hydrologique, on constate que les corrélations se renforcent progressivement avec l'accroissement du nombre de composantes pour atteindre un maximum de 0.66 et 0.59 respectivement. Cependant l'accroissement des corrélations n'apparaît plus significatif au-delà de la troisième composante (corrélation de 0.6 et 0.5 respectivement avec les deux premières dimensions de l'espace hydrologique), témoignant indirectement de la forte redondance entre les descripteurs de bassin disponibles. Si ces niveaux de corrélation s'avèrent supérieurs à ceux obtenus à l'issue de la régression classique, ils restent relativement faibles vis-à-vis des objectifs de définition de voisinages en site non jaugé, ainsi, si l'on ne conserve que les trois premières composantes (Figure 27), la concordance entre voisinages attendus (dans l'espace hydrologique) et voisinages prédits (dans l'espace physiographique) s'avère très médiocre (de l'ordre de 20% seulement en médiane pour une taille de voisinage $N = 50$).

On précisera qu'en termes de variables explicatives les poids attribués aux différents descripteurs de bassins dans les combinaisons linéaires fournies par la régression PLS sont en accord avec les analyses précédentes. Ainsi pour *ACP1* les variables se rapportant à la densité de drainage (DD), la longitude du centre de gravité du bassin (Y_{cg}), le pourcentage de surface imperméable (%imp) et l'ETP estivale (ETPsummer) sont les plus explicatives tandis que pour *ACP2* on retrouve essentiellement la variable se rapportant aux pourcentages de surface agricole (%agro) et forestière (%forest), la densité de drainage (DD) et les variables de précipitations moyenne annuelle et saisonnière. (PA, Pwinter, Pspring...).

Au vu des résultats obtenus ici, la définition de voisinages glissants pour les sites non jaugés apparaît d'ores et déjà comme relativement incertaine. La principale difficulté réside dans l'absence de descripteurs du bassin susceptible de permettre une bonne discrimination des comportements hydrologiques à l'échelle nationale. Une solution serait de compartimenter préalablement l'échantillon en régions plus homogènes puis d'appliquer à nouveau la procédure décrite ici. Toutefois cette solution limite fortement l'intérêt de l'approche « ROI » par rapport aux méthodes reposant sur l'établissement de régressions régionales telles qu'elles ont été mises en œuvre précédemment dans les sections 3.4 et 3.5.

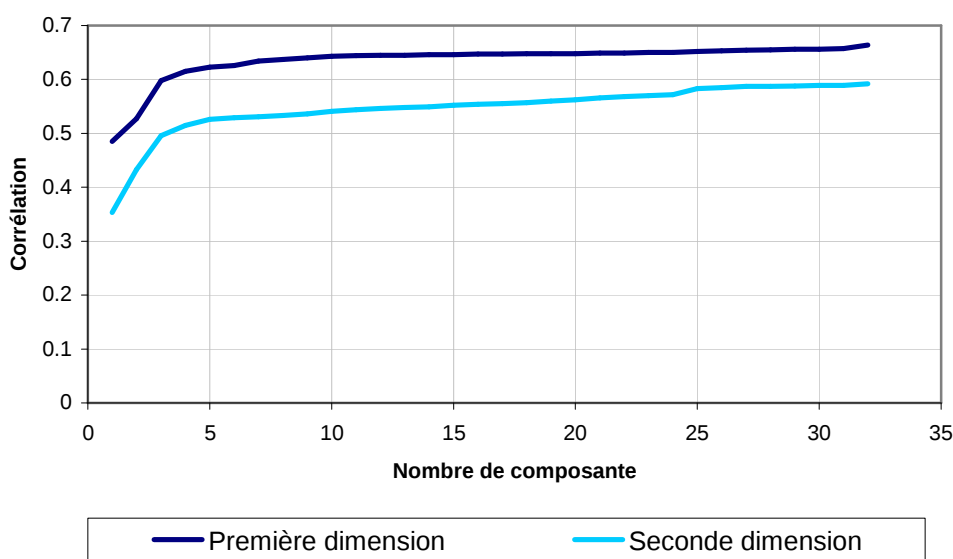


Figure 26 : Evolution de la corrélation entre dimensions des espaces hydrologique et physiographique en fonction du nombre de composantes retenues à l'issue de la régression PLS

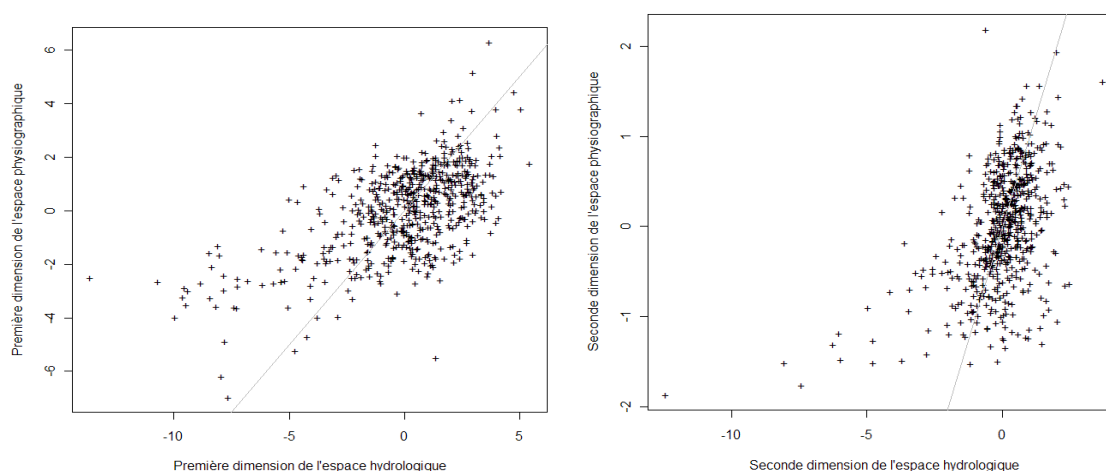


Figure 27 : Concordance pour chaque dimension entre l'espace hydrologique obtenu par ACP des variables caractéristiques d'étiage et l'espace physiographique obtenu par régression PLS (en gris : droite de prédiction parfaite)

Malgré la difficulté à définir un espace physiographique convenable pour une bonne définition des voisinages, la procédure de régression multiple a été tout de même appliquée de manière à vérifier si l'information valorisée au travers de cet espace permettait un gain sur l'estimation des variables QMNA5 et VCN10(5) par rapport aux voisinages définis sur la base de la proximité géographique. On rappellera que, comme dans l'application proposée en section 3.3, chaque dimension de l'espace physiographique a préalablement fait l'objet d'une pondération pour respecter leur importance relative dans le calcul des distances entre individus. Les résultats obtenus pour une taille de voisinage $N = 50$ sont donnés ci-dessous (Tableau 11 et Figure 28).

Critère		QMNA5VCN10(5)					
NASH		0.330	0.693	0.483	0.337	0.656	0.357
BIAS (l/s/km ²)		-0.117	-0.024	-0.146	-0.106	-0.047	-0.165
RMSE (l/s/km ²)		1.201	0.711	1.166	1.016	0.605	1.086
Erreur absolue (l/s/km ²)	10%	0.099	0.023	0.116	0.069	0.016	0.137
	50%	0.682	0.281	0.780	0.566	0.221	0.694
	90%	2.763	1.905	2.411	2.207	1.695	2.212

Tableau 11 : Performances d'estimation en validation croisée des variables QMNA5 et VCN10(5) spécifiques au moyen de modèles régressifs établis pour des voisinages déterminés sur la base de la proximité dans l'espace physiographique

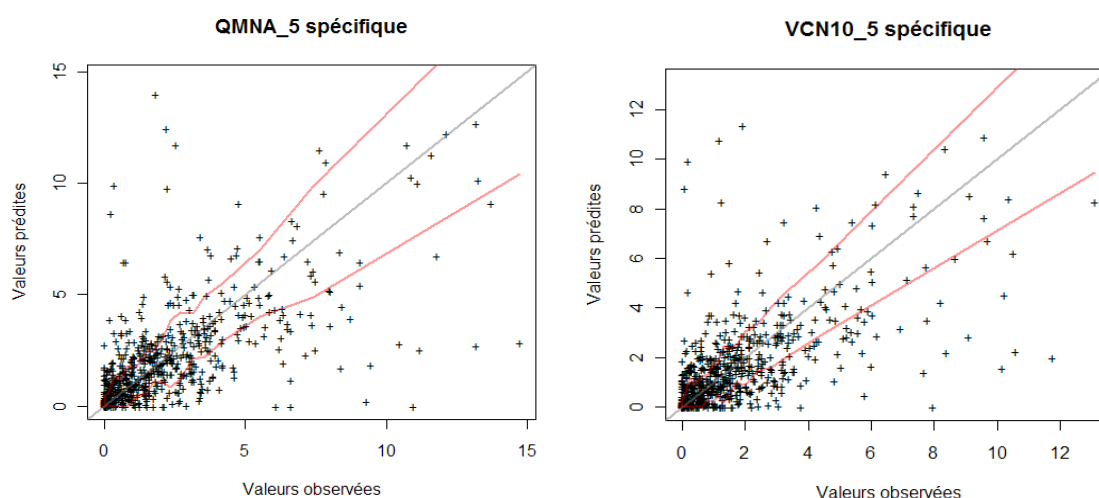


Figure 28 : Valeurs de QMNA5 et VCN10(5) spécifiques observées et prédites en validation croisée au moyen de modèles régressifs établis pour des voisinages déterminés sur la base de la proximité dans l'espace physiographique et erreur moyenne par gamme de valeur observée (en rouge)

Les critères de performances obtenus indiquent, comme attendu, que l'estimation réalisée ici s'avère très médiocre. On note même que ces résultats sont globalement moins bons que ceux obtenus en définissant les voisinages sur la base de la proximité géographique. Ceci confirme que la trop faible concordance entre l'espace hydrologique et physiographique ne permet pas de parvenir à la définition de voisinages efficaces vis-à-vis de la construction de modèles régressifs et que l'information valorisée au travers de cet espace est pratiquement inexistante.

3.7 Conclusion

Les conclusions données ici reprennent l'ensemble des résultats précédents pour faciliter la comparaison entre les performances de chaque méthode de définition de voisinage/région vis-à-vis de la construction de modèles régressifs destinés à estimer les débits caractéristiques QMNA5 et VCN10(5) spécifiques.

Les différentes modalités testées sont rappelées dans le tableau suivant avec les notations simplifiées utilisées par la suite :

Modalité	Notation	Section correspondante
Modèle national (sans définition de voisinage)	<i>Nat</i>	Section 4.2
Regroupements par classes de régime d'étiage	<i>Classe</i>	Section 4.3
Voisinages glissants de taille $N = 50$ identifiés sur la base de la proximité dans l'espace hydrologique	<i>ROI_hydro</i>	Section 4.3
Regroupements par HERm (régressions pondérées par la proportion de surface de chaque bassin dans chaque HERm)	<i>HERm</i>	Section 3.4
Regroupements issus de la procédure CART (section 4.5) avec et sans introduction de l'information sur la classe de régime hydrologique mensuel prédite par Sauquet (2006) dans les variables de décision	<i>CART1</i> <i>CART2</i>	Section 4.5
Voisinages glissants de taille $N = 50$ identifiés sur la base de la proximité géographique	<i>ROI_geo</i>	Section 3.6.1
Voisinages glissants de taille $N = 50$ identifiés sur la base de la proximité dans l'espace physiographique	<i>ROI_physio</i>	Section 3.6.2

Les résultats de synthèse pour chacune des deux variables estimées sont détaillés ci-dessous sous la forme de boxplot (Figure 29) donnant un aperçu de la distribution des erreurs absolues¹⁴. Sur ces deux représentations les valeurs d'erreur au delà des « moustaches » n'ont pas été représentées pour permettre une meilleure lisibilité. On remarquera que ces valeurs élevées sont de manière générale à l'origine de la forte différence observée entre l'erreur médiane et le RMSE. Ce dernier critère sera donc jugé suffisant pour différencier les méthodes sur la base des erreurs extrêmes. A ce titre on remarquera qu'il apparaît difficile d'affirmer si ces erreurs se produisent de manière systématique pour certains bassins (*i.e.* indépendamment de la méthode de regroupement considérée). Un certain nombre d'exceptions concernent cependant des bassins très productifs, correspondant pour la plupart à des bassins karstiques (*e.g.* Le Guiers Mort à Saint-Laurent-du-Pont (V1504010), L'Adouin à Saint-Martin-en-Vercors (W3335210), La Valserine à Chézery-Forens (V1015030), La Loue à Vuillafans (U2604030), L'Hers à Bélesta [source de Fontestorbes] (O1432930)) et pour lesquels tous les modèles aboutissent presque systématiquement à une sous-estimation très marquée. Ce résultat n'est pas surprenant, il démontre la

¹⁴ La médiane est représentée en trait gras, la moyenne (RMSE) est représenté par une croix rouge, la boîte donne les quartiles d'erreur à 25 % et 75 %, les moustaches donnent un équivalent de l'intervalle de confiance à 95 % (+/- 1.58 fois la distance interquartile normalisée par la racine carrée de l'effectif).

difficulté à traiter correctement le cas de bassins très atypiques (*i.e.* présentant une forte singularité locale) au travers de modèles régressifs régionaux. On notera qu'ils sont d'ailleurs généralement traités comme des horsains et ne participent pas à l'ajustement des modèles. Un certain nombre d'estimations médiocres sont également associées à des bassins sur lesquels persiste un doute quant à la qualité des données. Il s'agit essentiellement de stations influencées (présence d'ouvrages hydroélectriques à l'amont) mais conservées dans l'échantillon dans la mesure où les débits fournis dans la Banque HYDRO sont qualifiés de « naturels reconstitués » (O0105110, O6332510, W0224010, W2714010 et Y5615020).

Pour le reste des stations les figures fournissent une vision synthétique des observations déjà reportées en détail dans les parties précédentes. Ainsi pour résumer on constate que, parmi les estimations de référence, les approches visant à constituer des regroupements de bassins sur la base de similitudes hydrologiques (*Classe* et *ROI_hydro*) s'avèrent très largement favorables à l'estimation des variables QMNA5 et VCN10(5) spécifiques par rapport à un simple modèle national (*Nat*), confirmant ainsi la pertinence des critères de similitudes choisis pour regrouper les bassins (classe de régime d'étiage ou proximité dans l'espace hydrologique défini par ACP des principaux descripteurs du régime d'étiage). Pour ces deux approches les résultats diffèrent essentiellement sur la première moitié de la distribution des erreurs pour laquelle la méthode *ROI_hydro* apporte un gain significatif. Cet avantage provient essentiellement de la plus grande souplesse de l'approche par voisinages glissants qui permet, par rapport à l'identification de regroupements « figés », d'obtenir un plus fort degré d'homogénéité en fonction du site cible considéré. Néanmoins cette méthode peut également conduire à des estimations médiocres, en particulier lorsque le voisinage n'apparaît pas comme suffisamment approprié pour parvenir à un ajustement robuste et permettre l'extrapolation à des bassins atypiques.

Dans le cadre de l'application en site non jaugé, la distinction est nette entre les trois approches de regroupement en régions (*HERm*, *CART1* et *CART2*) et celles basées sur la définition de voisinages glissants (*ROI_geo* et *ROI_physio*). Pour ces dernières, la définition de similitudes entre bassins se montre clairement insatisfaisante par rapports aux voisinages de références exploités dans l'approche *ROI_hydro* et conduit par conséquent à des performances médiocres, même si l'on observe un gain pour une majorité d'individus (*i.e.* gain en médiane) par rapport au modèle national. Il est intéressant d'insister sur le fait que ce résultat démontre que la proximité géographique ne constitue pas un critère suffisamment pertinent en termes de similitude des comportements hydrologiques d'étiage. D'autre part les tentatives visant à restituer ces similitudes au travers d'un espace physiographique (*i.e.* élaboré à partir des descripteurs de bassin) se sont avérées décevantes et n'ont pas permis d'apporter de gain par rapport au voisinage géographique.

Pour les trois méthodes restantes (*HERm*, *CART1* et *CART2*), les performances obtenues sont relativement voisines et s'approchent des résultats de références obtenus au moyen de l'approche *Classe*. La discrimination réalisée par le biais des *HERm* ou, plus empiriquement au moyen de la procédure *CART* s'avèrent donc efficaces pour permettre l'ajustement de modèles régressifs performants et robustes. Dans le détail, on remarque que la modalité *CART2*, qui intègre une information qualitative à caractère hydrologique dans les variables de décision (*i.e.* les classes de régime hydrologique mensuel), permet globalement d'aboutir aux résultats les plus satisfaisants même si elle ne surpasse l'approche *HER* qu'en

terme de RMSE et de NASH (*i.e.* elle est la plus appropriée pour limiter l'apparition de forts écarts aux observations).

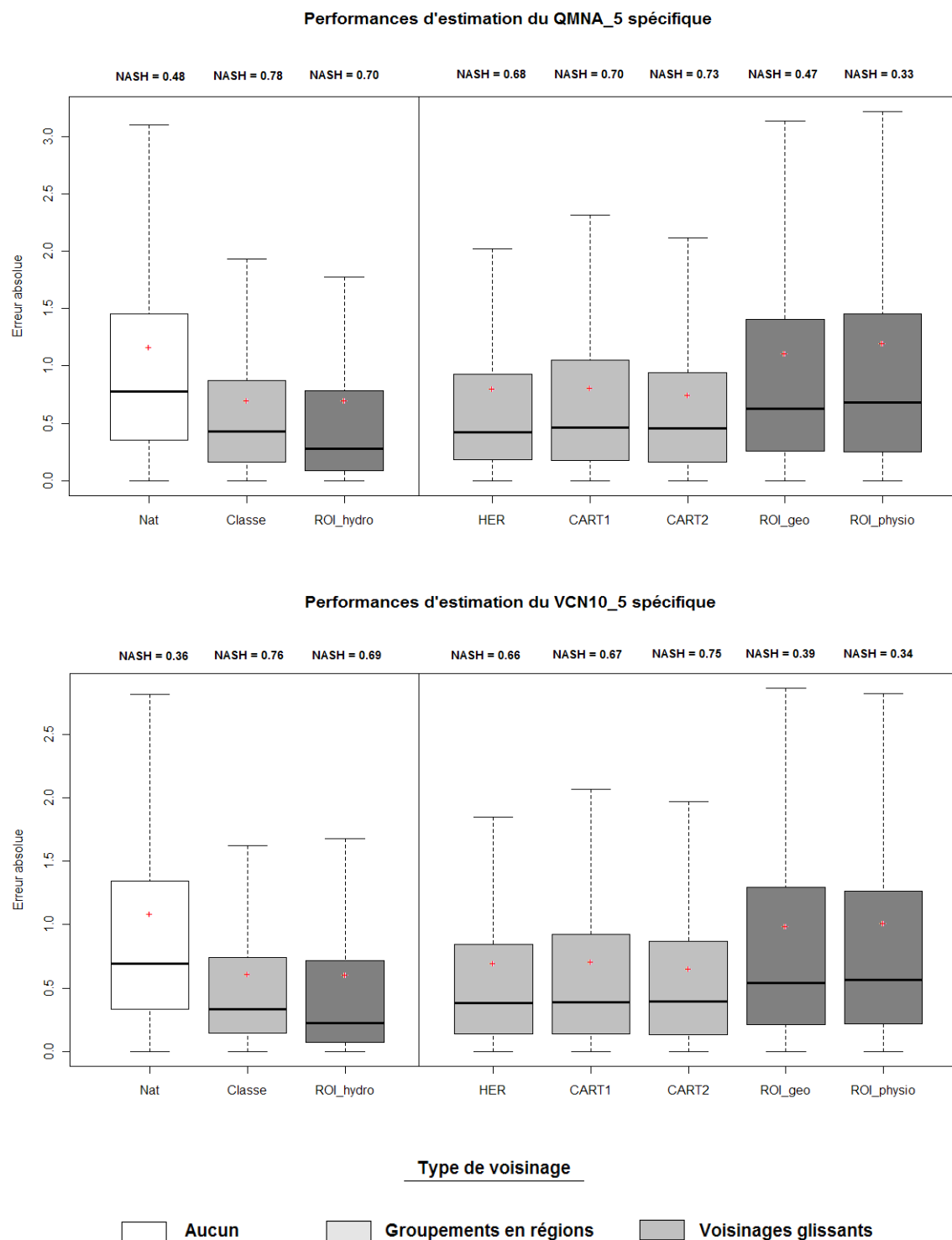


Figure 29 : Distributions des erreurs absolues obtenues par régression multiple (en validation croisée) pour l'ensemble de l'échantillon et différentes modalités de regroupement de bassin (la croix rouge indique l'erreur moyenne)

4 Cartographie des variables QMNA5 et VCN10(5) spécifiques et de leurs incertitudes

Sont présentées ici les résultats d'estimation des deux variables QMNA5 et VCN10(5) exprimées en débits spécifiques (l/s/km²) obtenus à l'aide de la méthode *CART2* sur l'ensemble des points du réseau hydrographique drainant une surface supérieure à 50 km².

Le réseau proposé correspond à un réseau théorique issu du traitement d'un MNT de résolution kilométrique disponible à l'échelle nationale. Ce traitement, réalisé à l'aide du logiciel « Hydrodem¹⁵ », s'appuie sur la construction d'un plan de drainage conçu pour décrire en chaque maille la direction d'écoulement (dans un plan horizontal) puis identifier la surface contributive correspondante. Le réseau hydrographique obtenu a fait l'objet d'une vérification quant à la bonne restitution des surfaces de bassin fournies dans la Banque HYDRO pour un échantillon de 2728 stations hydrométriques. Le critère retenu pour cette validation a été défini tel que l'écart entre surface reconstituée et surface officielle fournie par la Banque HYDRO doit être inférieur à la surface donnée par le produit :

$$|\Delta S| = |S_{ref} - S_{optim}| \leq a \cdot 2\pi r \quad \text{où} \quad r = \sqrt{\frac{S_{ref}}{\pi}}$$

S_{ref} est la surface annoncée par la Banque HYDRO

S_{optim} est la surface déterminée à l'aide du plan de drainage

r le rayon du disque de surface équivalente au bassin versant

a est la résolution du plan de drainage (1 km²)

En cas de non-respect de ce critère, il a été procédé à une vérification des coordonnées géographiques fournies par la banque HYDRO (permettant au passage de remplacer certaines stations mal positionnées). Pour les autres sites, le plan de drainage a pu faire l'objet de corrections manuelles pour ajuster le réseau hydrographique reconstitué au réseau du référentiel CARTHAGE.

Si, après vérification, la concordance avec le réseau de référence CARTHAGE peut être jugée satisfaisante pour les bassins de grande ou moyenne dimension, elle reste beaucoup plus incertaine pour les petits bassins en raison de la résolution du MNT utilisé mais aussi par la difficulté à restituer l'effet de

¹⁵ Logiciel développé par E. Leblois de l'unité de recherche Hydrologie-Hydraulique du Cemagref de Lyon

la géologie sur l'existence ou non d'écoulement en surface (indépendamment du relief). Il s'agit là de la principale raison qui nous a conduit à limiter nos estimations aux points du réseau drainant une surface supérieure à 50 km². On signalera d'autre part que les valeurs données pour certains cours d'eau frontaliers (en particulier pour le Rhin, le Rhône et le Doubs) doivent être considérées comme invalides. Les variables de bassin utilisées pour réaliser l'estimation ne sont en effet pas toujours disponibles sur la partie amont de ces bassins puisque située en dehors du territoire français.

En complément des valeurs estimées, ont été calculées deux valeurs donnant les bornes inférieure et supérieure d'un intervalle de prédiction à 90% (la probabilité que la valeur réellement observée au point *i* se trouve dans cet intervalle est de 90%) permettant de donner une idée de l'incertitude associée à chaque estimation (valeurs exprimées elles aussi en l/s/km²). Cet intervalle correspond à l'incertitude inhérente à la détermination des paramètres du modèle régressif qui croît lorsque l'on s'éloigne de la moyenne des observations ayant servi au calage de la relation à laquelle vient s'ajouter l'incertitude liée à la variance des observations (i.e. incertitude découlant de la dispersion du nuage d'observation autour de la droite de régression). A ce jour, l'ensemble de ces résultats est résumé dans un fichier autonome donnant les coordonnées de chaque point du réseau (coordonnées des centres de maille de résolution kilométrique), la surface de bassin théorique correspondante, l'estimation et les bornes de l'intervalle de confiance à 90%.

Les résultats d'estimation à l'échelle nationale sont donnés ci-dessous sous forme de cartes (Figure 30 et Figure 31). Une échelle de couleur et une transformation adaptée des valeurs estimées (la carte représente la racine cubique des valeurs estimées) ont été choisies afin de permettre une meilleure visualisation des contrastes régionaux de QMNA5 et de VCN10(5) spécifiques (exprimés en l/s/km²). Les cartes correspondant aux intervalles de prédiction ne sont pas fournies ici car peu informatives à cette échelle, cependant l'analyse détaillée de ces intervalles montre bien la différence de fiabilité des estimations en fonction de la gamme de valeur estimée et de la relation régionale utilisée pour réaliser l'estimation (i.e. l'incertitude dépend la région d'affectation du point) et en particulier selon le type de modèle employé (i.e. linéaire ou puissance).

A cette échelle il apparaît que les estimations s'avèrent relativement cohérentes avec les grandes caractéristiques hydro-climatiques régionales identifiées précédemment à partir des observations locales disponibles aux stations hydrométriques. On retrouve ainsi des valeurs de débit spécifique particulièrement élevées (> 10 l/s/km²) dans les zones de reliefs connaissant des phénomènes de stockage/fonte de neige (Alpes, Pyrénées), des valeurs assez fortes (> 5 l/s/km²) dans les zones d'altitude moyenne (Jura, Vosges, Massif Central) ou dans les régions présentant des échanges nappes-rivière généralisés et importants (Bassin parisien, cours d'eau côtiers normands, Landes), des valeurs moyennes (autour de 1 à 2 l/s/km²) dans les zones présentant une influence océanique marquée (Bretagne occidentale, Cotentin) et enfin des valeurs faibles ou nulles (< 0.1 l/s/km²) dans les secteurs à substratum fortement imperméable et/ou à saison estivale particulièrement sèche (Vendée, Coteau de Gascogne, pourtour méditerranéen).

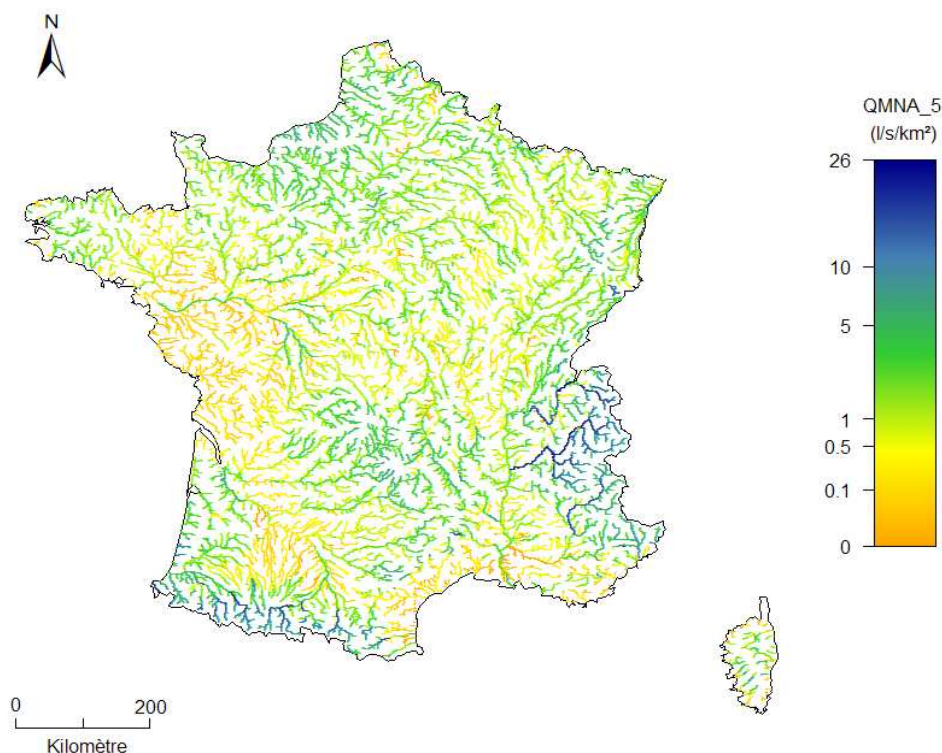


Figure 30 : Cartographie du QMNA5 spécifique ($l/s/km^2$) estimé par régression multiple au moyen de l'approche CART2 sur l'ensemble des points du réseau hydrographique théorique drainant une surface supérieure à $50 km^2$

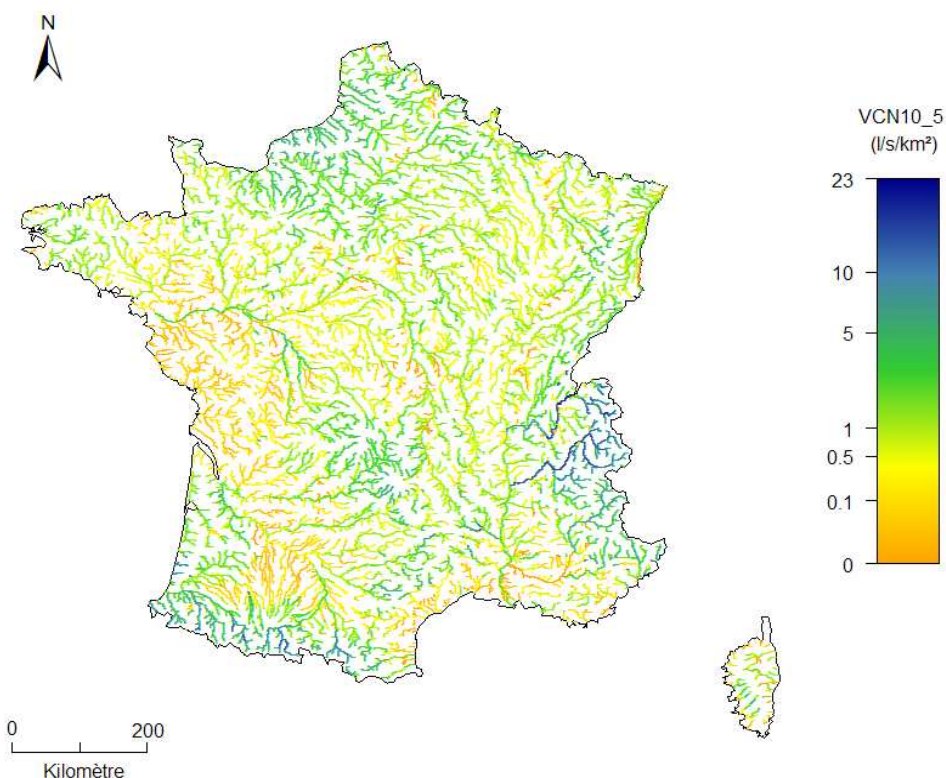


Figure 31 : Cartographie du VCN10(5) spécifique ($l/s/km^2$) estimé par régression multiple au moyen de l'approche CART2 sur l'ensemble des points du réseau hydrographique théorique drainant une surface supérieure à $50 km^2$

En revanche l'analyse détaillée de ces cartes permet de déceler de nombreux défauts, essentiellement liés à la présence de discontinuités brutales dans les estimations fournies sur des points du réseau hydrographique voisins. Ces discontinuités, qu'elles apparaissent sur un même tronçon ou au niveau d'une confluence, sont dans la grande majorité des cas liées à un saut entre deux régions distinctes (ces sauts s'avèrent inhérents à la méthode de regroupement CART en raison du mode de classification binaire qui ne permet pas d'assurer la continuité entre deux régions). On signalera toutefois que, au niveau des confluences, il est particulièrement délicat de déterminer si ces discontinuités sont justifiées ou non. Une des solutions envisagées pour l'étape suivante inter-modèles consistera à appliquer un lissage le long du réseau pour atténuer la présence de ces sauts brutaux.

A ces problèmes de discontinuité viennent s'ajouter quelques interrogations quant au réalisme de certaines valeurs. On remarque par exemple que les valeurs données sur les grands cours d'eau alpins tels que l'Isère s'avèrent probablement surestimées (toutefois les intervalles de confiance associés à ces points sont assez larges, ce qui signifie que l'incertitude sur ces estimations est élevée). A l'inverse on remarque que les estimations fournies sur la partie aval de la Durance sont extrêmement faibles (voire nulles) ce qui semble peu réaliste pour un cours d'eau de cette dimension. Ce type de critique reste malheureusement difficile à réaliser pour l'ensemble des valeurs fournies sans une connaissance approfondie du terrain. A ce stade l'utilisation des cartes de QMNA5 et VCN10(5) proposées ici devra donc faire l'objet de nombreuses précautions.

5 Conclusion

Ce rapport fait état des avancées en matière d'interpolation des débits caractéristiques d'étiage QMNA5 et VCN10(5) spécifiques en France. Il contient en premier lieu des éléments techniques sur l'élaboration d'un nouveau jeu de stations de référence. L'échantillon constitué sur la base de critères relativement exigeants est aujourd'hui stabilisé. Il comprend 632 stations hydrométriques réparties sur l'ensemble du territoire avec néanmoins quelques secteurs mal représentés (en Charente, dans le Nord, à l'extrémité du Cotentin et dans une moindre mesure en Camargue et dans le nord de la Corse).

A partir de ces données, différents indicateurs représentatifs du régime d'étiage (sévérité, déficit, durée, saisonnalité, débit de base) ont été extraits des chroniques hydrométriques. Ils ont fait l'objet d'une analyse préliminaire qui a permis de déterminer que les processus liés à l'existence ou non de stocks sur le bassin (aquifères, stock neigeux) susceptibles d'expliquer l'inertie de la réponse hydrologique étaient les plus déterminants pour la différenciation des régimes d'étiages à l'échelle nationale. Suite à cette analyse, deux méthodes statistiques ont été mises en œuvre successivement afin 1) de construire un référentiel approprié pour la définition de similitude entre bassin du point de vue de l'hydrologie des basses eaux (espace hydrologique) puis 2) d'établir une classification simplifiée des régimes d'étiage en quatre grandes familles.

Une importante partie du travail a ensuite été consacrée à tester et confronter (en validation croisée) différentes approches pour la régionalisation des variables VCN10(5) et QMNA5 par régression multiple. Nos travaux se sont essentiellement concentrés sur la problématique de définition de régions/voisinages homogènes en vérifiant tout d'abord le potentiel d'estimation dans le cas de voisinages idéalement connus (définis sur la base de similitudes à caractère hydrologique) par rapport à un unique modèle national. Compte tenu des résultats encourageants (malgré la mise en évidence de problèmes spécifiques sur certains sites atypiques ou douteux), la démarche a été prolongée pour la définition de voisinages dédiés à l'estimation en sites non jaugés. Pour ce faire trois méthodes distinctes, exploitant de différentes manières les variables descriptives du climat, de la géologie, du relief... pour l'identification de similitudes entre bassins, ont été testées.

L'analyse des résultats indique que deux de ces méthodes sont relativement équivalentes pour l'établissement de modèles régressifs performants en permettant de parvenir à un gain significatif par rapport à l'emploi d'un unique modèle national. Il s'agit de l'approche visant à constituer des regroupements basés sur la typologie en Hydro-éco-région proposée par **Wasson et al.** (2001) et de la procédure de classification par arbre de régression (CART) qui s'avère être une technique relativement récente dans ce type d'application.

La troisième approche examinée, qui faisait appel à l'identification de voisinages glissants (i.e. voisinages spécifiques autour du site cible considéré), s'est révélée réellement décevante par rapport aux succès qu'elle a pu rencontrer dans la littérature. Nos résultats montrent en effet que l'identification d'un espace approprié pour l'identification de voisinage en sites non jaugés s'avère particulièrement délicate en l'absence de descripteur(s) de bassin réellement susceptible(s) de permettre une bonne discrimination des régimes d'étiage à l'échelle nationale. Ainsi, dans notre cas, le recours à des approches par regroupement en région reste le moyen le plus pertinent pour parvenir à éliminer une partie de la variabilité difficile à exprimer au travers de modèles régressifs nationaux.

Pour conclure ce rapport sont finalement proposés deux cartographies (une pour chaque variable) représentant les estimations fournies en tout point d'un réseau hydrographique théorique (bassins drainant une surface supérieure à 50 km²) par la meilleure des méthodes de régionalisation. En complément une évaluation de l'incertitude est également proposée sous la forme d'intervalles de confiance des estimations. Cette incertitude mesure la plus ou moins grande fiabilité des estimations fournies selon les secteurs et doit rappeler aux futurs utilisateurs que ces estimations doivent toujours faire l'objet d'une critique avant une éventuelle utilisation.

Bibliographie

- Burn, D.H. (1990). Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resour. Res.*, Vol. 26(10), pp. 2257-2265
- Catalogne, C., Sauquet, E. (2010). Interpolation des courbes des débits classés. Rapport d'avancement. Rapport d'étude Cemagref-ONEMA.
- Chopart, S., Sauquet, E. (2008). Usage des jaugeages volants en régionalisation des débits d'étiage. *Revue des Sciences de l'Eau*, Vol. 21, 3, pp. 267 - 281
- Dyck, S. (1976). *Applied Hydrology, part I. Calculation and regulation of the discharge of streams*. VEB Verlag für Bauwesen, Berlin.
- Giuntoli, I., Renard, B. (2010). Identification des impacts hydrologiques du changement climatique : constitution d'un réseau de référence pour la surveillance des étiages. Rapport d'avancement ONEMA-Cemagref
- Giuntoli, I., Renard, B. (2009). Identification des impacts hydrologiques du changement climatique : vers un réseau de référence pour la surveillance des étiages. Rapport d'avancement ONEMA-Cemagref.
- Gottschalk, L., (1993). Interpolation of runoff applying objective methods. *Stochastic Hydrology and Hydraulics*, Vol. 7, pp. 269-281.
- Haché, M., Ouarda, T.B.J.M, Bruneau, P., Bobée, B. (2002). Estimation régionale par la méthode de l'analyse canonique des corrélations : comparaison des types de variables hydrologiques. *Can. J. Civ. Eng.*, Vol. 29, pp. 899-910
- Laaha, G., Blöschl, G. (2006). A comparison of low flow regionalisation methods - catchment grouping. *Journal of Hydrology*, Vol. 323(1-4), pp. 193-214.
- Nathan, R.J., McMahon, T.A. (1990). Identification of homogeneous regions for the purpose of regionalization. *J. Hydrol.* Vol. 121, pp. 217-238.
- Ouarda, T., Girard, C., Cavadias, G.S., Bobée, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal of Hydrology*, Vol. 254(1-4), pp. 157-173.
- Plasse, J., Sauquet, E. (2009) Interpolation des débits de référence d'étiage. Rapport d'étude, Cemagref-ONEMA, décembre 2009.

- Plasse, J., Sauquet, E. (2009). Cartographie des débits de référence sur le bassin de la Seine Normandie. Rapport d'avancement Phase 2.
- Sauquet, E. (2006). Mapping mean annual river discharges: geostatistical developments for incorporating river network dependencies. *Journal of Hydrology*, Vol. 331(1-2), pp. 300-314.
- Sauquet, E., Gottschalk, L., Krasovskaia, I. (2008). Estimating mean monthly runoff at ungauged locations: an application to France, *Hydrology Research*, 39(5-6), 403 – 423.
- Smakhtin, V.Y. (2001). Low flow hydrology : a review. *Journal of hydrology*, Vol. 240 (3-4), pp.147-186
- Tallaksen, L. M., Van Lanen, H.A.J. (2004). (Eds) *Hydrological Drought – Processes and Estimation Methods for Streamflow and Groundwater*. *Developments in Water Sciences* 48. Elsevier B.V., the Netherlands, 580 pp.
- Therneau, T. M., Atkinson, B. (2010). *rpart*: Recursive Partitioning, R package. <http://cran.r-project.org/web/packages/rpart/>
- Veza, P., Comoglio, C., Rosso, M., Viglione, A. (2010). Low Flows Regionalization in North-Western Italy. *Water Resources Management*, Vol. 24, pp. 4049–4074
- Wasson J.G., Chandresis A., Pella H., Souchon Y. (2001). Définition des hydroécorégions françaises. Méthodologie de détermination des conditions de référence au sens de la Directive Cadre pour la gestion des eaux. Rapport d'étude (Cemagref).
- Wold, S., Ruhe, A., Wold, H., Dunn, W.J. (1984). The collinearity problem in linear regression – The partial least squares (PLS) approach to generalized inverses. *Journal on scientific and statistical computing*, Vol. 5(3), pp. 735-743.
- Yevjevich, V. (1972). *Probability and Statistics in Hydrology*. Water Resources Publications, Fort Collins, CO, USA, 302 pp.



Partenariat 2011
*Domaine Hydro-morphologie et altérations
physiques des hydrosystèmes continentaux
Action Prédétermination des étiages*



Onema
Hall C – Le Nadar
305 square Félix Nadar
94300 Vincennes
01 45 14 36 00
www.onema.fr

IRSTEA
1, rue Pierre-Gilles de Gennes CS 100
92761 Antony cedex
01 40 96 61 21
www.irstea.fr