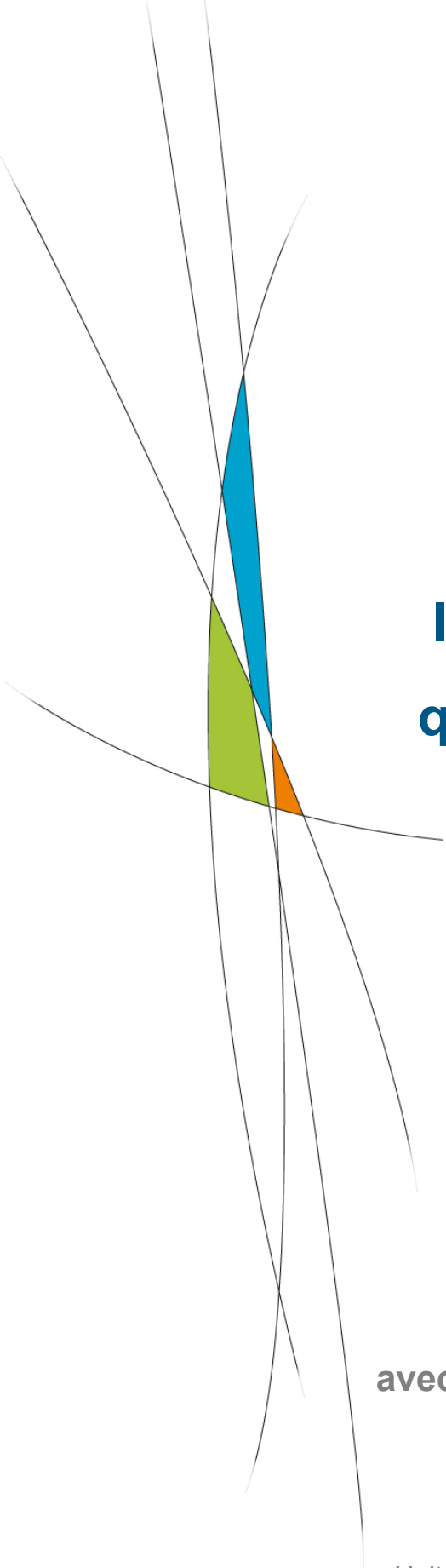




Interpolation des modules : quelles évolutions depuis la note technique de 1987 ?

Note de travail

Septembre 2010

Eric SAUQUET
avec les contributions techniques
de Clotaire CATALOGNE

Unité de Recherche Hydrologie - Hydraulique
Groupement de Lyon
3 bis quai Chauveau – CP 220

Table des matières

1	Introduction	3
2	Le module : définition, représentativités hydrologique et statistique	4
3	Quelles méthodes pour l'estimation en site non jaugés ?	10
3.1	Les méthodes à base de formulation empirique.....	10
3.2	Les méthodes à base d'interpolation.....	15
3.3	Les méthodes à base de modélisation pluie-débit.....	22
4	Conclusion et perspectives	24
	Bibliographie	26

1 Introduction

La note technique du CEMAGREF datant de 1987 intitulée « Guide méthodologique en vue de l'estimation du « module » d'un cours d'eau » avait pour objectif d'aider les services techniques à estimer le module des cours d'eau. Son intérêt réside dans son aspect très pragmatique : est proposée une méthode « jugée la plus performante » selon la disponibilité de données hydrologiques. Elle reflète un état de l'art compte tenu des outils mobilisables et des données facilement disponibles à la date de parution.

L'objectif de ce document est de réactualiser les éléments proposés en 1987, de compléter par d'autres approches possibles et d'illustrer anciens et nouveaux développements par des exemples extraits de projets de recherche ou d'applications réalisées dans le cadre d'études nationales. Nous proposerons quelques recommandations.

2 Le module : définition, représentativités hydrologique et statistique

En préambule, il faut rappeler que le module interannuel recherché doit être relatif à un état naturel. Ceci induit des questionnements sur la nature des débits utilisés sur le secteur étudié. Ces débits sont usuellement extraits de la banque HYDRO (<http://hydro.eaufrance.fr>) qui référence la majorité des stations hydrométriques en France. Au delà des séries temporelles, sont également renseignées différentes meta-données dans les « fiches stations » : surface hydrologique, surface topographique, coordonnées géographiques, qualité globale des mesures... et le caractère influencé du régime (avec plusieurs choix possibles : « pas ou faiblement », « », « fortement » sur l'intégralité du régime, en crue ou en étiage). C'est cette dernière information qui nous intéresse ; mais, malheureusement, elle n'est pas exploitable quantitativement : le qualificatif « faiblement » n'est pas associé à un (pourcentage de) volume sous gestion effective. Le seuil de pression qui fait basculer un bassin dans la catégorie « fortement influencé » est inconnu. En outre, des erreurs sont possibles, sachant que, par défaut, l'appellation « pas ou faiblement » est enregistrée : par exemple, la dernière station hydrométrique de l'Isère, avant confluence avec le Rhône, en aval de l'ensemble des ouvrages hydroélectriques est considérée comme « pas ou faiblement » influencée (interrogation de la base en septembre 2010). Un regard doit également être porté sur la qualité des mesures, sur la base des appréciations du gestionnaire. Cette expertise permettra d'exclure les stations aux données réellement douteuses¹.

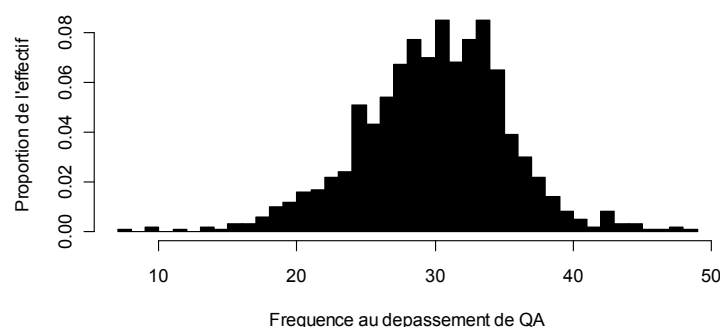
Le diagnostic sur les pressions auxquelles le bassin est soumis et sur la qualité des enregistrements étant réalisé, on peut exploiter les chroniques des stations réputées fiables sans influence notable ou, dans le cas contraire, engager une opération de re-naturalisation. Dans ce dernier cas, des modèles statistiques à base de bilan ou plus complexes, physiques ou comportementaux, doivent être élaborés. Pour ce qui concerne l'irrigation, des outils pour quantifier le besoin en eau de la plante, représenter le comportement des agriculteurs, traduire dans le temps la soustraction à la ressource (représentation par des réservoirs simulant le remplissage/la vidange des stocks collinaires différés)... sont à construire et des données d'usage de l'eau doivent être collectées pour les caler et en vérifier la pertinence. Les

¹ A ce jour, il n'est pas possible de connaître *a priori* les erreurs métrologiques. On peut tolérer la présence d'une station douteuse, dès lors qu'elle apporte de l'information et que sa présence ne compromet pas l'efficacité des méthodes d'interpolation en sites non jaugés. Le principe de précaution veut que par défaut, les stations douteuses soient exclues.

développements à réaliser sont parfois conséquents (fonctions du nombre de sources de perturbation), des hypothèses simplificatrices peuvent être proposées pour obtenir des débits naturels ; elles doivent être clairement énoncées (par exemple : hiérarchisation des pressions pour ne corriger que les influences les plus significatives). A l'issue de ces opérations, nous supposons disposer de chronique de débits naturels fiables. Le calcul du module, noté QA , est possible.

Le module QA est défini comme le débit annuel moyen ; il s'agit donc de la moyenne des débits annuels. Le débit plancher QP , dans son acception la plus répandue, est fixé au 10^{ème} du module. Il convient de rappeler quelques propriétés statistiques de ces variables.

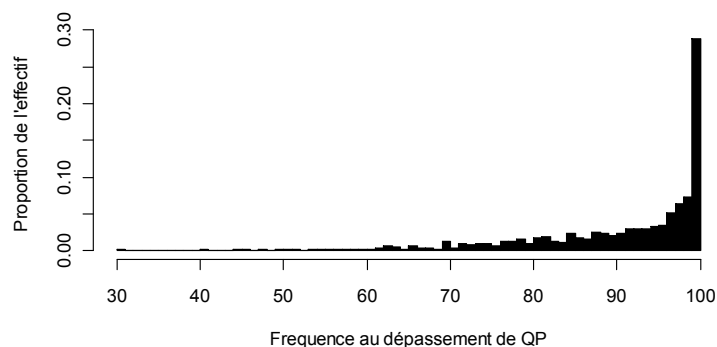
Tout d'abord, le module n'est pas égal à la médiane des débits journaliers. La distribution statistique des débits journaliers n'est pas symétrique (en particulier, elle n'est pas gaussienne). De ce fait, les statistiques moyenne et médiane ne sont pas confondues. Nous avons essayé de placer les modules de dans la courbe des débits classés en exploitant le jeu de données national de Catalogne et Sauquet (2010). Pour chaque station, sont calculés le module QA et la fréquence au dépassement de QA , proportion de jours dans l'année pendant lesquels le débit journalier est supérieur au module QA . La distribution des fréquences de dépassement de QA collectée sur un ensemble de près de 1000 stations est donnée dans la figure suivante :



Il apparaît que le module est une valeur qui est dépassée entre 25 et 35% du temps. Le module peut s'écarter de manière significative de la médiane.

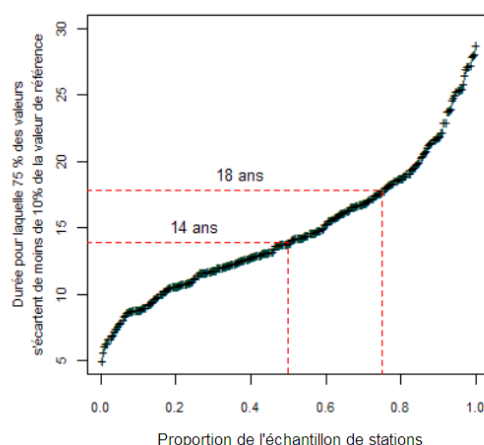
Nous avons effectué le même travail sur le débit plancher, QP , fixé au 10^{ème} du module.

La valeur correspondante au QP est une valeur faible excessivement rare, proche du débit minimal enregistré, puisque, pour près de 30% des stations, cette valeur est dépassée 99% du temps. En d'autres termes, le 10^{ème} du module n'est que très rarement sous-passé. Ce résultat est cohérent avec l'analyse réalisée sur 74 stations de (Baran *et al.*, 2008) qui concluait « *Les débits caractéristiques d'étiage inférieurs à 10% [$QMNA5$, $Q90$, $Q95$] du module ne sont pas des situations très fréquentes ni dans le temps, ni à l'échelle du territoire français. Pour la majorité des massifs de montagne en France, les débits minimum d'étiage sont très largement supérieurs à 10% du module.* »



Le bassin versant est soumis aux fluctuations naturelles du climat. Le débit annuel reflète cette variabilité temporelle filtrée par les processus de transformation pluie-débit. Il est évident que **calculer le module sur la base d'une moyenne sur une courte période rend sensible aux spécificités climatiques de la période d'enregistrement**. L'alternance des années sèches et humides peut biaiser l'estimation du module en particulier si elle se fonde sur quelques années de mesure prises au hasard. Si on calcule le module à partir de données prises sur une période sèche en une station, il y a un fort risque de sous-estimer la vraie valeur.

Catalogne (2010) dans le but de déterminer un jeu de données de référence pour la cartographie de variables d'étiage a réalisé des tests de sensibilité de la valeur du module à la longueur de la chronique. Cette analyse de sensibilité s'appuie sur un ré-échantillonnage au sein de séries suffisamment longues sur la période de référence 1970-2005. Une valeur de QA est calculée successivement pour des longueurs comprises entre 5 et 30 années consécutives (*i.e.* échantillonnage par blocs, sauf présence de lacunes). Un écart à la valeur de référence obtenue sur la totalité de la chronique disponible (*i.e.* jusqu'à 35 années) est ensuite calculé. Pour une durée donnée, on recense alors les stations pour lesquelles 75% des valeurs calculées s'écartent de moins de 10% de celle de référence. Le résultat est synthétisé par la figure suivante :



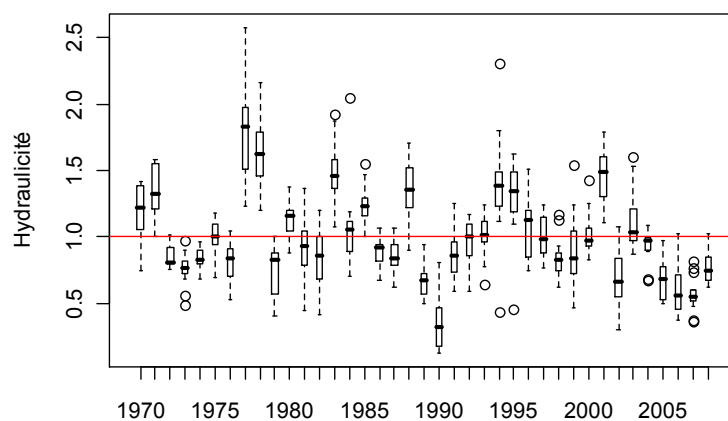
On peut y lire le fait que 50% des stations présentent 75% des valeurs calculées sur 14 années avec un écart de moins de 10%. Lorsqu'on étend à 18 ans, 75% des stations ont 75% des valeurs avec un écart de moins de 10%. Sur cette base, une durée de 18 ans semble acceptable pour obtenir une valeur robuste.

Pour exploiter les séries courtes, une procédure de réajustement doit être engagée. Une manière de procéder consiste à s'appuyer sur des débits. Il s'agit de réaliser une analyse temporelle régionale des coefficients d'hydraulicité, *i.e.* le rapport du débit annuel daté sur le module calculé sur la plus longue période de disponibilité. Il a l'avantage d'être sans dimension donc comparable d'une station à une autre. Lorsqu'il prend une valeur supérieure à 1 pour une année, cette année sera dite « humide ». Lorsqu'il est inférieur à 1, l'année en question sera qualifiée de « sèche ». La procédure suggérée consiste à faire un réajustement « climatique » des estimations des séries courtes en divisant le débit QA^* calculé sur la période d'observation par l'hydraulicité moyenne régionale $HydraulicitéRégionale^*$ (toutes stations « longues » confondues) sur cette même période :

$$QA = QA^* / HydraulicitéRégionale^*$$

Une dernière question peut être soulevée : *doit-on restreindre la période de calcul du module pour ne pas biaiser les estimations du module du fait d'un (possible) changement climatique ?* Répondre « oui » sous-entendrait : (1) qu'un effet du changement climatique est déjà présent dans les données et (2) que la réglementation impose de préserver le milieu selon des conditions hydrologiques antérieures. Le point (2) supposerait également l'existence d'une période définissant la référence climatique passée pour effectuer les calculs. Le changement global et son traitement relèvent du domaine de la réglementation et des choix en matière d'adaptation en lien avec la Directive Cadre Européenne sur l'Eau. En absence de toute démonstration scientifiquement étayée ou de tout cadre réglementaire, rien n'empêche à ce jour l'utilisation de données sur plus de 20 ans, pour réduire les incertitudes autour de l'estimation du module. **Il faut exploiter les données sur la plus grande longueur de disponibilité.** Si on se place dans une perspective de non-stationnarité et de volonté d'adaptation en conséquence, la procédure de calcul sera forcément à réviser, non pas en restreignant les données utilisables, mais plutôt en cherchant à décrire et projeter les évolutions temporelles du module.

Application : Notre exemple repose sur les longues séries disponibles du bassin de la Drôme (d'au moins 18 années d'enregistrement). La moyenne des débits annuels est calculée sur toute la période disponible. Les débits annuels ont été calculés sur les années hydrologiques entre le 1^{er} septembre et le 31 août de l'année suivante (dans notre convention, l'année hydrologique « 1971 » correspond ainsi à la période entre le 1^{er} septembre 1970 et le 31 août 1971). Pour chaque station, nous avons déterminé la chronique des coefficients d'hydraulicité. Puis, pour chaque année, nous avons rassemblé tous les coefficients et synthétisé chaque année par une boîte à moustache.

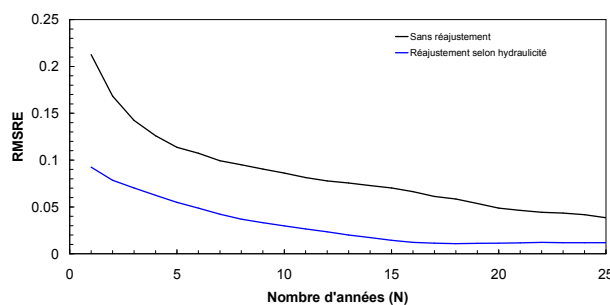


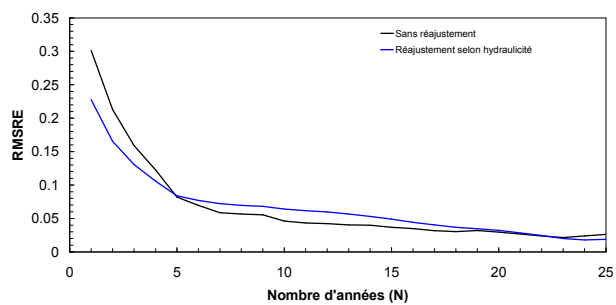
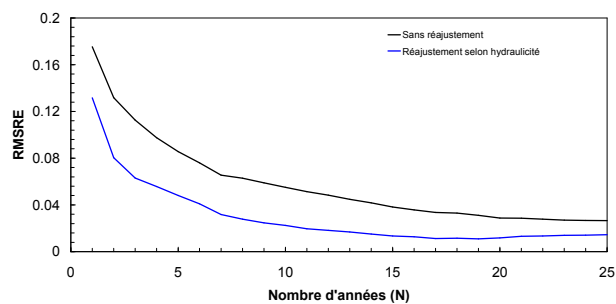
La figure ci-dessus montre les évolutions des coefficients d'hydraulicité sur la période 1970-2008. Elle met en évidence des oscillations autour de la valeur 1. Nous pouvons noter un comportement de bloc : l'hydraulicité varie de façon assez similaire sur l'ensemble des stations. Les années sèches et les années humides ont tendance à être partagées simultanément par une majorité de bassins versants. Il existe donc une forte solidarité à l'échelle du secteur. Cette solidarité est gage d'une bonne efficacité de la procédure de réajustement.

Des modules de stations « courtes » ont été corrigées selon cette procédure (cf. tableau plus bas). Les modifications opérées sur les stations courtes induisent un changement de l'ordre de 15%. Il faut noter que la procédure semble fiable puisqu'on obtient plus de cohérence des modules estimés au niveau de la Drôme à Saillans. La station DREAL (V4264010), aux enregistrements étendus, annonce un module de 17.79 m³/s. La station gérée par la Compagnie Nationale du Rhône, beaucoup plus courte (V4264CNR), proposait 13.48 m³/s sans réajustement sur une période incomplète entre 1994-2008. La valeur pour la station V4264CNR intégrant le réajustement climatique, 15.6 m³/s, se rapproche de celle de la DREAL.

Code	Station	Hydraulicité moyenne sur la période d'enregistrement <i>Hydraulicité Régionale*</i>	Module sur chronique courte QA^* (m ³ /s)	Module réajusté QA (m ³ /s)	Disponibilité (année)
V4015110	Le Béal des Roses à Charpey	1.00	0.027	0.027	14.8
V4264CNR	La Drôme à Saillans [CNR]	0.87	13.5	15.6	11.3
V4275020	La Gervanne à Plan-de-Baix	1.10	1.03	0.93	7.8
V4287010	La Grenette à la Répara-Auriples	0.87	0.041	0.047	11.3
V4290CNR	La Drôme à Livron [CNR]	0.88	22.3	25.3	10.8
X1034010	Le Buech à Serres [Pont de Pierre]	1.05	14.5	13.8	9.1

Pour tester l'efficacité des procédures de réajustement par simulation, nous avons choisi trois stations aux enregistrements supérieurs à 34 ans sur le secteur d'étude. Nous allons les considérer comme peu jaugées et les traiter en conséquence, *i.e.* calculer les modules sur toutes les sous-périodes possibles de N années, puis réajuster les valeurs et les comparer au module QA calculé sur la période de disponibilité complète. Nous avons retenu finalement la racine de la moyenne des écarts relatifs quadratiques ($RMSRE$).





Les résultats graphiques montrent qu'il y a un intérêt à effectuer un réajustement en particulier pour les durées très courtes, inférieures à 5 ans (comparaison entre courbes noires « Sans réajustement » et courbes bleues). Cette opération ne semble pas introduire de biais (les courbes bleues sont en périphérie immédiate des courbes noires dans le cas où cette procédure n'est pas efficace, par exemple, sur le troisième exemple).

3 Quelles méthodes pour l'estimation en sites non jaugés ?

3.1 Les méthodes à base de formulation empirique

Eléments méthodologiques : La revue de la littérature technique montre la prééminence de ce type de méthodes. Ces relations sont le plus souvent déduites d'analyses corrélatoires sur les données brutes de débit lorsqu'on souhaite établir une formulation linéaire ou sur les logarithmes de débits lorsqu'on souhaite établir une relation de type puissance. La difficulté de ce type d'approche réside principalement :

- 1 – dans le choix des variables auxiliaires ;
- 2 – dans le choix de la forme mathématique de la relation empirique.

Concernant le point 1, la relation la plus simple pour estimer le débit de référence en l/s ou m³/s est celle qui repose uniquement sur la surface drainée S . Elle prend la forme :

$$Q = aS^k$$

Elle considère implicitement un débit spécifique ou pseudo-spécifique (Q/S^k) constant sur la région d'étude. k est généralement proche de 1². L'hypothèse faite est que le seul facteur agissant est lié au support de la donnée, les autres facteurs climatiques ou géologiques ayant une influence mineure. Ce type de relation est faussement efficace : les débits en m³/s sont toujours fortement corrélés à la surface avec des coefficients de détermination R^2 très élevés. Ces relations traduisent le fait que les volumes écoulés augmentent avec la surface de collecte. C'est une vision bien souvent trop simpliste pour accéder aux particularités aux petites échelles. Pour être efficace, d'autres variables doivent intervenir.

La sélection des variables candidates potentielles doit intégrer *les processus dominants qui déterminent les écoulements annuels. Quelles variables pourraient caractériser au mieux les interactions entre le milieu et le régime hydrologique ?* Cependant, l'insertion d'une variable dans une relation empirique, même basée sur des considérations physiques, n'est pas forcément un gage de performance. En effet, la plupart des indicateurs n'ont qu'un lien indirect avec les questions de nature hydrologique. C'est principalement le cas des données renseignant le sol ou le sous-sol. Enfin, un choix raisonné pourra s'appuyer sur les pratiques antérieures et sur la disponibilité et la pertinence statistique et physique de ces descripteurs pendant l'étude.

² Cette approche est utilisée pour les crues avec $k=0.8$; ici, la valeur de k n'a aucune raison d'être égale à 0.8.

Les variables vers lesquelles on se dirige naturellement sont d'origine climatique. Les débits sont la transformation d'entrées météorologiques en écoulement de surface. A défaut de disponibilité de ce type de données, on peut exploiter l'altitude. La variabilité spatiale des pluies et des températures à grande échelle est, en effet, expliquée par l'orographie. Sous nos latitudes, les pluies sont plus abondantes en altitude et le terme d'évapotranspiration moindre du fait de températures de l'air plus faibles qu'en plaine. C'est ce qui a motivé l'usage de cette unique variable dans les relations établies pour décrire les écoulements annuels par Sauquet (2005).

Des tests statistiques doivent être mis en œuvre pour s'assurer que les variables retenues ont un rôle significatif dans la relation établie et pour éliminer les effets de colinéarités entre variables (redondance d'information). Il peut être utile de chercher à comprendre l'origine du rejet d'une variable ou le signe de la pondération associée à une variable (ou même les coefficients de corrélation partielle, qui mesurent le degré de liaison entre la variable visée et une variable explicative, les autres étant supposées constantes, *i.e.* « les effets d'une variable toutes choses égales par ailleurs »).

Un des degrés de liberté concerne la forme mathématique de la relation empirique (point 2). *Faut-il privilégier une relation linéaire ? Une relation puissance ?* Les débits résultent d'interactions entre processus non linéaires. Ceci nous inciterait à privilégier une loi puissance. L'hydrologue pourrait être tenté de privilégier un modèle linéaire exploitant, outre la surface S , les termes du bilan hydrologique : un terme relatif aux pluies P et un autre descriptif de l'évapotranspiration ET potentielle (ETP) ou réelle (ETR), et de se rapprocher de l'équation simplifiée du bilan hydrique. Le débit de surface en m^3/s est à un facteur multiplicatif de rendement r près donné par : $Q \approx r S (P - ET)$ ou lorsque le débit est converti en lame d'eau ou débit spécifique : $Q \approx r (P - ET)$. Il n'y a pas de réponse immédiate, encore moins définitive : il faut tester les formulations, analyser les résultats des coefficients d'efficacité, examiner les pondérations associées aux variables explicatives. L'expertise doit intervenir à différents niveaux, pour éviter toute formulation « insolite » qui pourrait engendrer des extrapolations aberrantes.

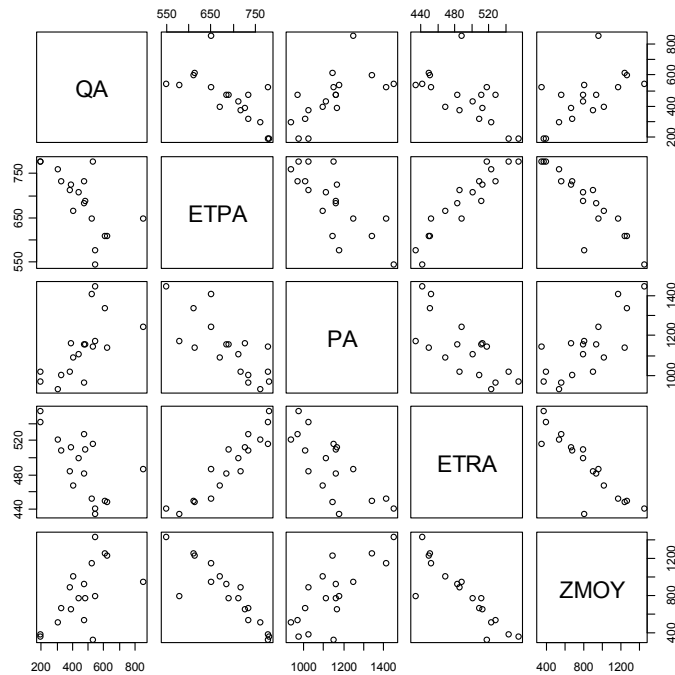
Notons que :

- la gamme de valeurs sur laquelle la relation a été établie doit être précisée. Se pose la question de la confiance accordée aux estimations pour des bassins dont les caractéristiques seraient en dehors des valeurs de calage ;
- pour être efficace, il est parfois utile de construire des relations par région, ce qui permet de rassembler des bassins soumis à des processus similaires et de mieux cerner les variables explicatives. Attention toutefois aux discontinuités entre régions.

Du point de vue technique, la régression est bien souvent la méthode retenue pour mettre en forme la relation empirique sur les variables brutes ou transformées log. Elle offre un cadre statistique objectif pour l'estimation des variables de débits en fonction de descripteurs de bassin. Laaha et Blöschl (2006) rappellent qu'il ne faut cependant pas oublier de vérifier *a minima* les propriétés

statistiques attendues des résidus (homoscédasticité, c'est-à-dire une erreur du modèle qui soit équivalente sur toute la gamme d'application, et caractère gaussien des résidus).

Application : Un exemple est donné sur le bassin de la Drôme. Sont représentés les liens entre le module QA exprimé en mm/an et d'autres variables ($ZMOY$ désigne l'altitude moyenne, PA la pluviométrie annuelle, $ETPA$ l'évapotranspiration potentielle annuelle et $ETRA$ l'évapotranspiration réelle annuelle).



Ces variables sont fortement corrélées (en particulier $ETRA$ et $ETPA$, avec $R^2 = 0.86$). La meilleure relation linéaire entre QA et une des variables météorologiques est celle qui s'appuie sur l' $ETPA$ ($R^2 = 0.44$) : $QA = -1.49 ETPA + 1484.3$

La seconde variable explicative est la pluie ($R^2 = 0.41$) : $QA = 0.6806 PA - 317.21$

Des tests ont été réalisés pour construire une relation exploitant deux variables parmi celles listées plus haut (régression multilinéaire). Les résultats montrent que l'apport d'une seconde variable explicative est faible et n'est pas significatif au sens du test de Student (à 5%) sur les coefficients de pondération.

Finalement, parmi les deux relations les plus efficaces laquelle choisir ? La première semble la plus performante au regard du seul R^2 , mais la pondération associée à l' $ETPA$ est en valeur absolue supérieure à 1, ce qui peut sembler étrange. Une raison pourrait être la sous-estimation de ce terme. La seconde relation est moins efficace mais plus cohérente physiquement. L'inconvénient majeur est pour ce cas d'application la résolution spatiale de l'information pluie (ici un point tous les 8 km), ce qui peut être insuffisant pour l'estimation en tête de bassin et en secteur très hétérogène, alors que l'ETP est

donnée à résolution de 1 km. Il n'y a pas de solution parfaite, il faut forcément expliciter les hypothèses et limites des modèles lors de leur utilisation.

L'absence de données d'ETP explique les relations nationales suggérées par la note technique du CEMAGREF (1987) :

$$QA = 0.75 PA + 0.40 ZMOY - 450 \text{ formulation réputée valide si le résultat est supérieur à 300 mm}$$

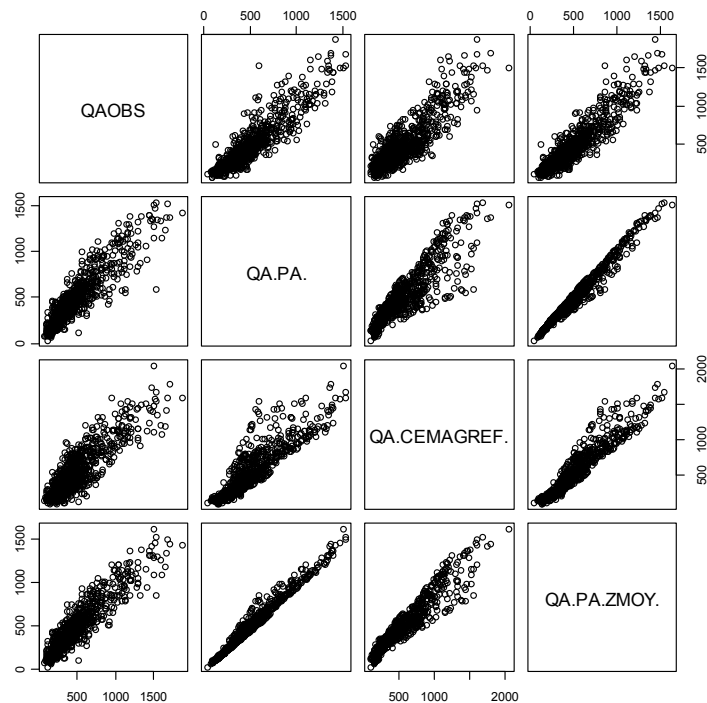
$$QA = (PA / 34)^{5/3} (ZMOY / 100)^{1/3} \text{ sinon.}$$

Les relations (QA.CEMAGREF) ont été testées sur le jeu de données de 872 stations de Sauquet *et al.* (2008), réputées peu influencées par la gestion humaine et par le karst. En parallèle nous avons ajusté des relations sur les mêmes variables. Il vient :

$$QA = 1.07 PA - 610.68 \quad (R^2 = 0.83) \quad (QA.PA)$$

$$QA = 0.11 + 0.95 PA - 545.90 \quad (R^2 = 0.85) \quad (QA.PA.ZMOY)$$

Les graphiques ci dessous montrent la bonne cohérence entre estimations et observations, exprimées en mm/an (QA.OBS). Finalement, on obtient des valeurs de *RMSRE* respectives de 0.44 pour (QA.CEMAGREF) et 0.31 pour les deux relations (QA.PA) et (QA.PA.ZMOY). Les formules recalées par nos soins sont plus efficaces à l'échelle du territoire ; ceci est logique elles s'adaptent à nos observations. Néanmoins, des erreurs relatives de 40% avec (QA.CEMAGREF) et de 30% avec les autres formules sont effectivement possibles. Les pluies et l'orographie sont des facteurs importants de variabilité du module. Ces estimations sont des formules nationales sommaires qui donnent les grandes tendances ; elles doivent être prises comme telles.

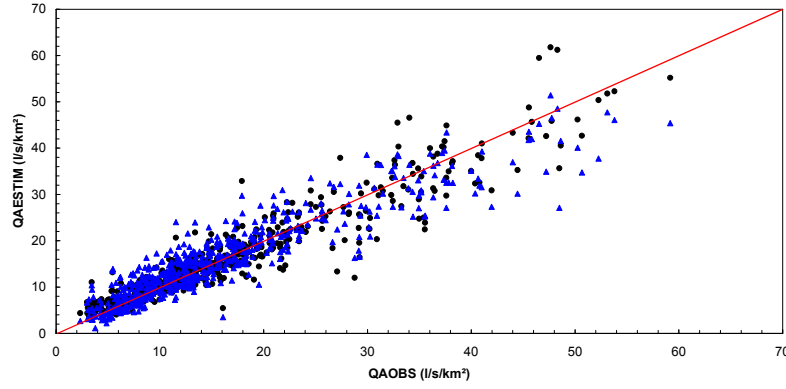


La note technique du CEMAGREF (1987) suggère, si on dispose de longues séries de pluie et de débit (PA^* , QA^*) en un site de référence drainant une surface S^* , une estimation du module en amont de cette station de référence donnée par :

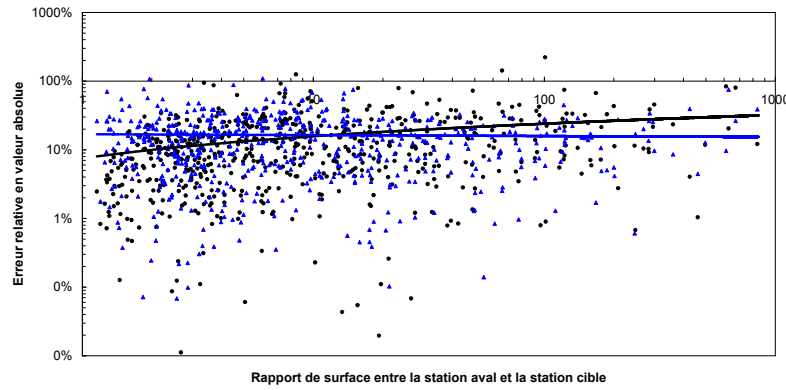
$$QA = QA^* (S / S^*) (PA / PA^*)^2 \quad (QA.S.PA^2)$$

où QA est le module recherché en m^3/s , S la surface drainée et PA la pluie moyenne annuelle au point recherché. Ceci signifie que la quantité $QA/S/PA^2$ est un invariant, point à vérifier dans la région où l'on souhaite appliquer cette relation.

Nous avons appliqué cette approche considérant $QA/S/PA^2$ comme une constante le long du réseau hydrographique, donnée par la première station rencontrée en aval de la station cible. 571 stations ont une station jaugée parmi les 872 du jeu de données de Sauquet *et al.* (2008). Ces 571 stations sont, tour à tour, considérées comme stations cibles pour lesquelles une estimation de QA est souhaitée. Le résultat est donné sur la figure suivante ($\bullet$ = une station). A titre de comparaison, nous avons calculé pour ces 571 stations les valeurs issues de l'application de la formule nationale (QA.PA.ZMOY), exprimées en $l/s/km^2$ ($\blacktriangle$ = une station). La première bissectrice est reportée en rouge.



Les incertitudes induites par $(QA.S.PA^2)$ sont légèrement supérieures à celles induites par $(QA.PA.ZMOY)$ ($RMSRE = 24\%$ contre 22%). Les performances de deux approches sont sensiblement équivalentes. Pour aller plus loin dans l'analyse des incertitudes, nous avons tracé la valeur absolue de l'erreur relative selon les deux méthodes d'estimation en fonction du rapport S^*/S qui indique la proximité au bassin aval S^* .



L'erreur du modèle d'interpolation $ErrRel$ est faiblement corrélée avec S^*/S :

$$|ErrRel| = 0.0358 \ln(S^*/S) + 0.0748 \quad (R^2 = 0.07)$$

(courbe tracée en noir) et est cohérente avec ce qui est attendu : plus la station cible est éloignée de la station d'appui, moins l'estimation est pertinente ; cependant, le facteur d'éloignement n'est pas la seule origine des écarts. La corrélation entre S^*/S et l'erreur relative de $(QA.PA.ZMOY)$ est en revanche

quasiment nulle ($R^2 = 0.004$) (courbe tracée en bleue). Les deux courbes se croisent vers $S^*/S \approx 10$, pour $S \leq S^* < 10 S$, la courbe noire est au-dessus de la courbe bleue : il y a intérêt à utiliser l'estimation par extrapolation de $QA/S/PA^2$, pour $10 S < S^*$ c'est l'inverse : il y a intérêt à utiliser la formulation nationale (QA.PA.ZMOY). Comme pour la formule du CEMAGREF de 1987, des plages d'application à privilégier, ici selon la surface drainée, devraient être distinguées.

3.2 Les méthodes à base d'interpolation

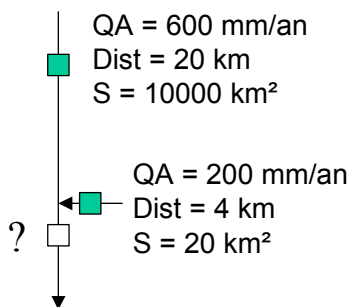
Eléments méthodologiques : Certaines techniques fournissent une estimation d'une variable (débit ou coefficients associés) z en un site non jaugé $\mathbf{x}_0$ par pondération des valeurs de cette variable observées sur des stations voisines $z(\mathbf{x}_i)$ sur le domaine étudié :

$$z^*(\mathbf{x}_0) = \sum_{i=1}^N \lambda_i z(\mathbf{x}_i)$$

où λ_i désignent les N coefficients de pondération associés aux N observations $z(\mathbf{x}_i)$.

Une des difficultés réside dans la définition du voisinage. Le plus simple est de le définir géographiquement sur la base d'une distance. Que penser de la définition d'une distance euclidienne entre stations hydrométriques ? Les débits résultent de transformation pluie-débit en amont de la station. Considérer le point de mesure comme point représentatif du bassin est donc incohérent. En effet, intéressons-nous à deux points en amont immédiat d'une confluence, par exemple, la Saône et le Rhône à Lyon avant qu'ils ne se rejoignent. Ces deux points sont très proche géographiquement, mais éloignés vis-à-vis de l'hydrologie : l'un mesure la Saône au régime pluvial et l'autre le Rhône au régime influencé par la neige.

Quelles seraient les conséquences d'une pondération définie par l'inverse de la distance dans la configuration suivante :



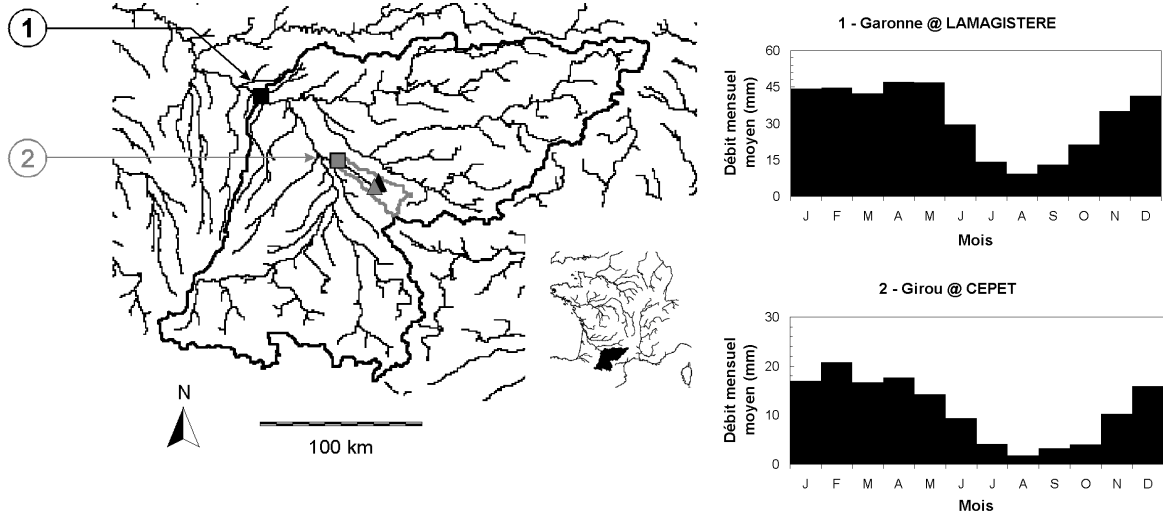
Une estimation est souhaitée sur le point identifié par le carré blanc et les sites jaugés sont repérés par un carré vert. Les distances au site non jaugé sont de 20 km et 4 km. Au site non jaugé,

$$QA = \frac{1}{1/20 + 1/4} (600/20 + 200/4) = 267 \text{ mm}$$

Le chiffre obtenu est aberrant (cela impliquerait une perte non négligeable d'eau entre la station amont drainant 10000 km² et la station aval). Ce type d'interpolation du fait de la distance conduit bien souvent à des valeurs erronées. Une étude récente de la DREAL Franche Comté l'a bien signalé

(Vallaud, 2010). Nous considérons que **le recours à l'interpolation avec la distance euclidienne entre exutoires est à exclure.**

Une autre solution est de mesurer la proximité entre bassins versant par la distance entre les centres de gravité. Elle est plus adaptée mais reste imparfaite dans certaines configurations du fait de recouvrement des surfaces de bassin. Considérons deux bassins jaugés : la Garonne à Lamagistère (32 350 km²) et un de ses affluents, le Girou à Cepet (522 km²).



La station de jaugeage est identifiée par un carré et le centre de gravité du bassin versant par un triangle. Les centres de gravité des deux bassins sont distants de 3.6 km. Selon ce critère, les deux bassins versants sont très voisins. Cependant, les régimes hydrologiques observés sont très différents. La Garonne reçoit une contribution amont non négligeable des cours d'eau pyrénéens. Le maximum des écoulements est observé en avril ; son régime est qualifié de pluvio-nival selon Pardé (1955). Les variations des débits mensuels moyens du Girou trahissent un régime hydrologique océanique pluvial. Les écoulements annuels sont plus abondants pour la Garonne à Lamagistère que pour le Girou à Cepet.

La proximité entre centres de gravité de deux bassins ne garantit pas la similitude des régimes hydrologiques. En l'absence de données sur la Garonne, un algorithme s'appuyant sur cette distance affectera un poids disproportionné au Girou dans l'évaluation de débits caractéristiques à Lamagistère. Gottschalk (1993) a proposé une alternative : la distance entre deux entités géographiques S_1 et S_2 de superficie respective A_1 et A_2 est la moyenne arithmétique des distances entre bipoints appartenant à ces deux entités :

$$d = \frac{1}{A_1 A_2} \int_{u_1 \in S_1} \int_{u_2 \in S_2} |u_1 - u_2| du_1 du_2$$

Pour des bassins relativement éloignés, les valeurs calculées ainsi seront proches de celles de la distance entre centres de gravité. D'un point de vue pratique, ce type de distance est calculé à l'aide d'un système d'information géographique, à partir des coordonnées des pixels qui forment les deux secteurs.

Le choix d'une distance appropriée n'est pas anodin. On peut appliquer une méthode d'interpolation complexe ; cette dernière fournira des résultats moins convaincants si la distance est mal choisie.

Une autre forme d'interpolation est proposée par la note technique du CEMAGREF (1987) dans une configuration bien précise. Le voisinage est défini par deux stations, une en amont du site cible et la seconde en aval ($N=2$). La distance entre deux bassins est mesurée par la surface drainée le long d'un même tronçon hydraulique. L'hypothèse est que l'écoulement augmente au prorata de la surface drainée :

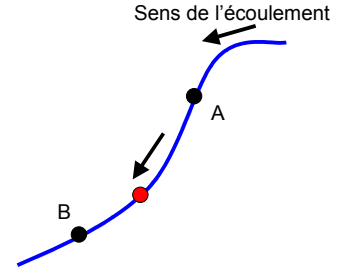
$$QA = (S - S_A)(QA_B - QA_A)/(S_B - S_A) + QA_A$$

avec QA en m^3/s . Cette estimation est assortie d'une incertitude donnée par :

$$(1 + (S_B - S_A)/(5S)) ((S - S_A)(\Delta QA_B - \Delta QA_A)/(S_B - S_A) + \Delta QA_A)$$

où ΔQA_B et ΔQA_A sont les incertitudes pesant sur les modules aux deux stations d'appui. S'il n'y a pas d'incertitudes sur ces termes, selon cette formule, il n'y en a pas concernant le site cible.

Il est clair que la formule suggérée n'inclut pas les erreurs dues à la procédure d'interpolation elle-même.



Parmi les autres méthodes qui retiennent le principe d'interpolation spatiale, citons :

- les interpolations inversement proportionnelles à la distance qui sépare $\mathbf{x}_0$ des N points d'observation $\mathbf{x}_i$: les poids sont donnés par :

$$\lambda_i = \frac{w_i}{\sum_{j=1}^N w_j} \text{ avec } w_i = 1/[dist(\mathbf{x}_0 ; \mathbf{x}_i)]^k$$

Il n'y a pas de raison objective pour guider le choix de k ;

- les moyennes régionales simples sur un jeu de N points d'observation formant une région homogène (la distance peut être utilisée comme définissant les frontières de ces régions) ;
- la famille des méthodes stochastiques qui se fondent sur un cadre théorique solide et dont les variantes se différencient principalement par les hypothèses explicites faites sur la variable d'intérêt. Elles présentent l'avantage de laisser les variables s'exprimer librement sans autre intervention. Nous retiendrons la technique du krigeage (Matheron, 1965) qui repose sur l'ajustement d'une fonction décrivant les liaisons spatiales, le *variogramme*.

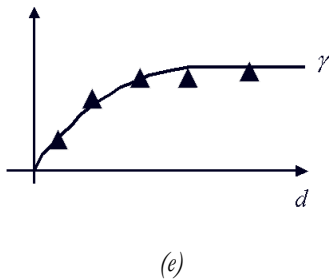
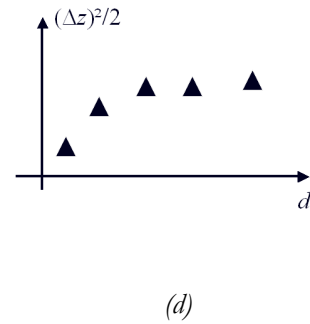
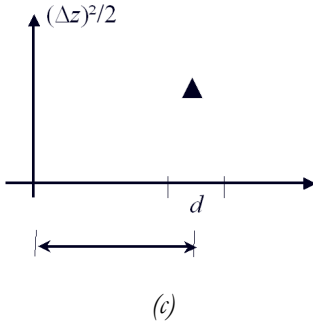
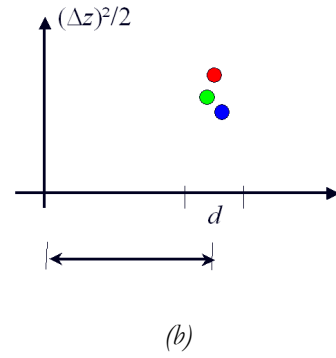
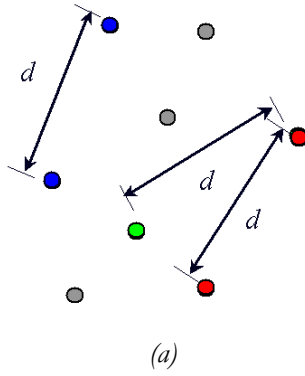
Pour obtenir le variogramme, on étudie le carré des écarts entre valeurs en fonction de l'éloignement. Le variogramme est déterminé expérimentalement à partir des observations selon :

$$\hat{\gamma}(d) = \frac{1}{2N(d)} \sum_{i,j}^{N(d)} (\hat{z}(\mathbf{x}_i) - \hat{z}(\mathbf{x}_j))^2 = \frac{1}{N(d)} \sum_{i,j}^{N(d)} \Delta \hat{z}(i,j)^2 / 2$$

où $N(d)$ est le nombre de couples de points distants d'environ d . En pratique (cf. figure ci dessous), on identifie les couples de points distants d'environ d (a), on calcule pour chacun de ces couples les écarts quadratiques divisés par 2 (b) et la moyenne (c). On renouvelle ces opérations autant de fois que nécessaire (d).

En théorie, et dans la version pour l'interpolation des variables ponctuelles (pluie ou température par exemple), pour que les calculs des pondérations puissent être menés à leur terme, il est nécessaire de faire usage d'un modèle ajusté au variogramme empirique choisi au sein d'une famille assez limitée de fonctions mathématiques. Les modèles autorisés les plus communément rencontrés sont :

- le modèle sphérique :
$$\gamma(d) = \begin{cases} \gamma_0 - (\gamma_\infty - \gamma_0) \left(\frac{3}{2} \frac{d}{b} - \frac{1}{2} \left(\frac{d}{b} \right)^3 \right), & d \leq b \\ \gamma(d) = \gamma_\infty, & d > b \end{cases}$$
- le modèle exponentiel :
$$\gamma(d) = \gamma_0 + (\gamma_\infty - \gamma_0) \left(1 - \exp\left(-\frac{d}{b}\right) \right)$$



Les paramètres descriptifs sont : la portée, la valeur du palier γ_∞ et la pépité γ_0 . Ils sont ajustés graphiquement ou plus rarement numériquement pour obtenir la meilleure adéquation au variogramme empirique (e). La portée est la distance au-delà de laquelle la corrélation est supposée nulle entre site. Le lien entre paramètre d'échelle et portée dépend du modèle (ainsi, la portée est égale à b pour le modèle sphérique et est atteinte asymptotiquement pour le modèle exponentiel). La pépité doit être

toujours positive. Une pépite non nulle indique des discontinuités dues à des effets d'échelle locaux autour de la mesure.

Pour le cas d'une variable à support de mesure non ponctuel (par exemple, les débits dont les supports de données sont des surfaces drainées non-nulles), la procédure de calage est plus complexe. Au final, nous obtenons également un modèle théorique de variogramme qui est exploité dans autant de systèmes d'équations linéaires que de points pour lesquels une estimation est recherchée et la déclinaison est connue dans la littérature sous le nom de « krigeage de/par bloc » (cf. Skøien *et al.* (2006) pour les éléments théoriques).

La pratique du krigeage s'est démocratisée (elle devient accessible dans certaines « tool box » de SIG). Cependant, il faut prendre du recul par rapport à ces outils livrés clefs en main pour une application aux débits. Les options proposées par défaut ne sont pas adaptées (la variante « krigeage de/par bloc » n'est pas proposée, ni celle plus critiquable se fondant sur les centres de gravité) : bien souvent, seul le krigeage avec des valeurs rapportées à la station hydrométrique, *i.e.* distance sur la base des distances entre stations, n'est possible. Des outils sont en cours de développement au Cemagref de Lyon (logiciel HYDRODEM, développé Leblois (2010) pour faciliter l'usage du krigeage de/par bloc. L'approche consiste à considérer que le module en rivière est la somme sur le bassin versant d'une production d'écoulement ponctuel (assimilable à $P-ETR$, qui sont des termes à représentativité ponctuelle). La résolution suit d'évidence mais requiert un pré-traitement de la zone d'étude (dont le plan de drainage).

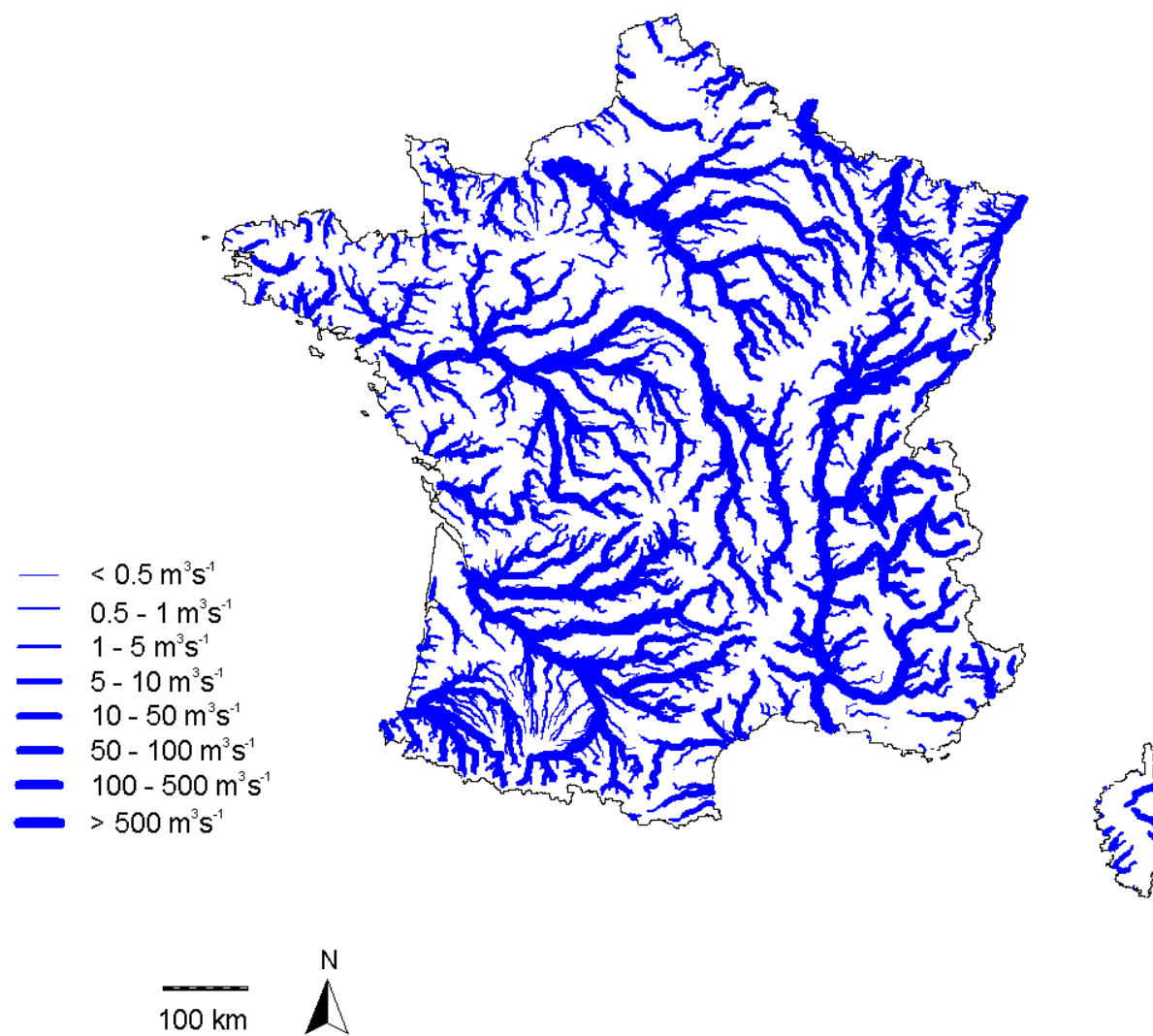
Applications : Sauquet (2005) propose une déclinaison du krigeage garantissant la continuité le long du réseau hydrographique et le traitement du karst. Des relations empiriques QA^* entre QA et un ensemble de variables dont une description spatiale sur le domaine étudié est connue sont ajustées sur un découpage en grands secteurs hydrographiques :

$$QA^* = g(X_1, X_2, \dots, X_K)$$

ε est le résidu de la relation empirique ajustée, estimé sur les sites jaugés par $\varepsilon = QA - QA^*$, et est considéré comme un terme correcteur structuré dans l'espace et donc interpolable. La valeur du résidu en sites non jaugés est obtenue par krigeage. Les incertitudes obtenues par validation croisée sont données en mm par gamme de surface. Nous retrouvons une erreur médiane de 10% de la lame d'eau à l'échelle du territoire toute gamme de surface considérée.

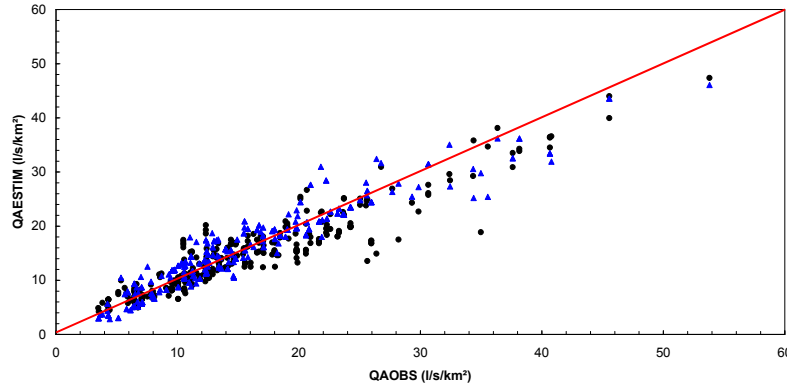
Classe de surface de bassins	Premier quartile	Médiane	Dernier quartile
$S > 0 \text{ km}^2$	14	31	70
$50 \text{ km}^2 < S < 500 \text{ km}^2$	19	43	86
$S > 500 \text{ km}^2$	9	19	40

Signalons que Riffard (2010) a choisi une approche similaire sans toutefois aller jusqu'au krigeage. La méthode consiste à regrouper un certain nombre de bassins situés dans le voisinage immédiat de la station cible, d'établir une relation empirique et d'introduire non pas un terme correctif additif mais multiplicatif : $QA = QA^* \times \varepsilon$ avec $QA^* = g(X_1, X_2, \dots, X_K)$.



Carte des modules sur la période 1981-2000 (extrait de Sauquet (2005))

Nous avons souhaité vérifier la performance de la méthode par interpolation selon la surface de la note de 1987. Un sous-échantillon des 872 stations de Sauquet *et al.* (2008) formé des stations positionnées entre deux stations hydrométriques (365 sur les 872) a été constitué. La procédure d'interpolation a été appliquée. Nous avons repris les mêmes représentations que pour l'analyse des résultats de (QA.S.PA²) : représentation des estimations en fonction des observations pour la méthode d'interpolation (● = une station) et pour la formulation (QA.PA.ZMOY) (▲ = une station). En rouge est tracée la première bissectrice.

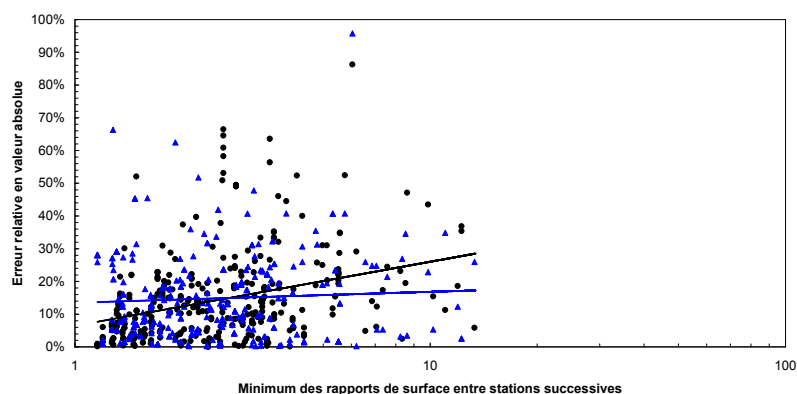


Les deux approches présentent des incertitudes équivalentes ($RMSRE = 0.19\%$). Comme pour la méthode d'extrapolation en amont (QA.S.PA²), nous avons examiné si l'erreur de reconstitution est dépendante de la proximité entre station cible et une des stations aux extrémités du tronçon (A ou B). Cette proximité est quantifiée par le recouvrement de surface entre la station cible et une des deux stations aux extrémités A ou B , *i.e.* le minimum des rapports S/S_A et S_B/S , toujours supérieur à 1 ($Rmin$). Nous avons donc tracé la valeur absolue de l'erreur relative selon les deux méthodes d'estimation en fonction $Rmin$. L'erreur du modèle d'interpolation est faiblement corrélée $Rmin$:

$$|ErrRel| = 0.0851 \ln(Rmin) + 0.0646 \quad (R^2 = 0.11)$$

(courbe tracée en noir). Le recouvrement est un facteur explicatif de l'erreur commise, mais n'est pas prédominant. Ici encore : plus la station est éloignée des extrémités, moins l'interpolation est pertinente.

Il n'y a pas de corrélation entre $Rmin$ et l'erreur relative de (QA.PA.ZMOY) ($R^2 = 0.004$) (courbe tracée en bleue). Les deux droites de régression se croisent vers 2.7, pour $Rmin < 2.7$ la courbe noire est au-dessus de la courbe bleue : l'estimation par interpolation est plus performante, pour $Rmin \geq 2.7$ l'inverse est observé : la formulation nationale (QA.PA.ZMOY) doit être privilégiée.



3.3 Les méthodes à base de modélisation pluie-débit

Une dernière possibilité est offerte par la modélisation pluie-débit. Plutôt que d'estimer une caractéristique du régime hydrologique, l'objectif est de reconstituer des chroniques de débit à l'aide de modélisations distribuées conceptuelles ou physiques en tout point du territoire.

L'approche à base de modèles conceptuels requiert d'abord une phase de calage sur des observations réparties sur le territoire puis une opération de régionalisation qui concerne non plus directement les écoulements mais les données d'entrée et les paramètres internes de la transformation pluie-débit. Une approche à base de relations empiriques est mise en œuvre malgré tout (Servat et Dezetter, 1995 ; Fernandez *et al.*, 2000, Merz et Blöschl, 2005).

Pour les modèles à base physique, aucune donnée de calage n'est nécessaire. Ils restent difficiles à mettre en œuvre : l'obtention des paramètres physiques requiert une connaissance détaillée du terrain au travers de campagnes ou d'instrumentations ou d'une base de données environnementales conséquentes et de nombreuses données d'entrée décrivant les forçages météorologiques sont nécessaires.

L'inconvénient est que la méthode n'est pas optimisée pour estimer le débit caractéristique, c'est bien la chronique qui est la cible. On peut comprendre ainsi que Engeland *et al.* (2006) jugent la méthode à base statistique plus performante que celle exploitant le modèle conceptuel HBV.

En France, la démarche à base conceptuelle est représentée par le logiciel LOIEAU (Folton et Lavabre, 2006, 2007). LOIEAU propose de calculer les chroniques de débits mensuels datés sur les stations hydrométriques grâce à un modèle à deux paramètres basé sur GR2M (Michel et Makhlouf, 1994). Il permet d'accéder de façon dynamique au module et au débit Q_{MN45} , à partir d'une analyse statistique des chroniques. Ceci est possible pour l'ensemble des cours d'eau d'une zone géographique, pour laquelle les deux paramètres du modèle ont été préalablement régionalisés. Les zones géographiques pour lesquelles LOIEAU est actuellement opérationnel : le bassin Adour Garonne, la quasi-totalité du bassin Rhône-Méditerranée-Corse, l'amont du bassin Loire Bretagne et le bassin de la Seine-Normandie. Pour obtenir des estimations en sites non jaugés, des cartes des paramètres ont été

élaborées en s'appuyant sur différents descripteurs (occupation du sol, couverture sol et altitude) et sur la base de l'interprétation physique des paramètres du modèle. Une opération de régionalisation est nécessaire mais elle est masquée et concerne des termes indirectement liés aux débits.

4 Conclusion et perspectives

Ce document n'a pas la prétention d'être exhaustif. Il présente un certain nombre d'approches possibles pour l'estimation des modules, en « complétant » des séries courtes ou par interpolation à partir d'un jeu d'observations. Différentes applications sont proposées à l'échelle nationale, elles donnent une idée sur les performances des approches. Les conclusions ne sont pas définitives sur leur rang. Si on souhaite une application dans un secteur plus réduit, il faut de nouveau optimiser les méthodes et les comparer sur des débits affranchis des effets d'échelle spatiale (c'est-à-dire en l/s/km² ou mm) ou sur la base des erreurs relatives, sinon le risque est de privilégier l'estimation/l'adéquation des forts débits au détriment des petits débits. Aucune méthode ne peut prétendre être universelle. Les contextes climatiques et hydrologiques sont divers et les hétérogénéités peuvent compromettre l'efficacité d'une méthode plutôt qu'une autre, si la première est de fait moins souple d'application.

Le seul moyen objectif de valider est de se placer artificiellement dans une configuration non jaugée. La validation croisée permet de le faire. Elle consiste à exclure tour à tour une ou un ensemble de station(s) parmi les N de l'échantillon initial, puis de proposer une estimation sur le(s) bassin(s) exclus à partir du jeu de stations restantes. Ainsi, on mesure la capacité à estimer en extrapolation spatiale. L'analyse de performance ne doit pas porter uniquement sur le seul coefficient de corrélation, mais sur des notions d'erreurs relatives (plus concrètes). Dans une perspective de comparaison objective, il faut si possible, tester les méthodes dans les mêmes conditions.

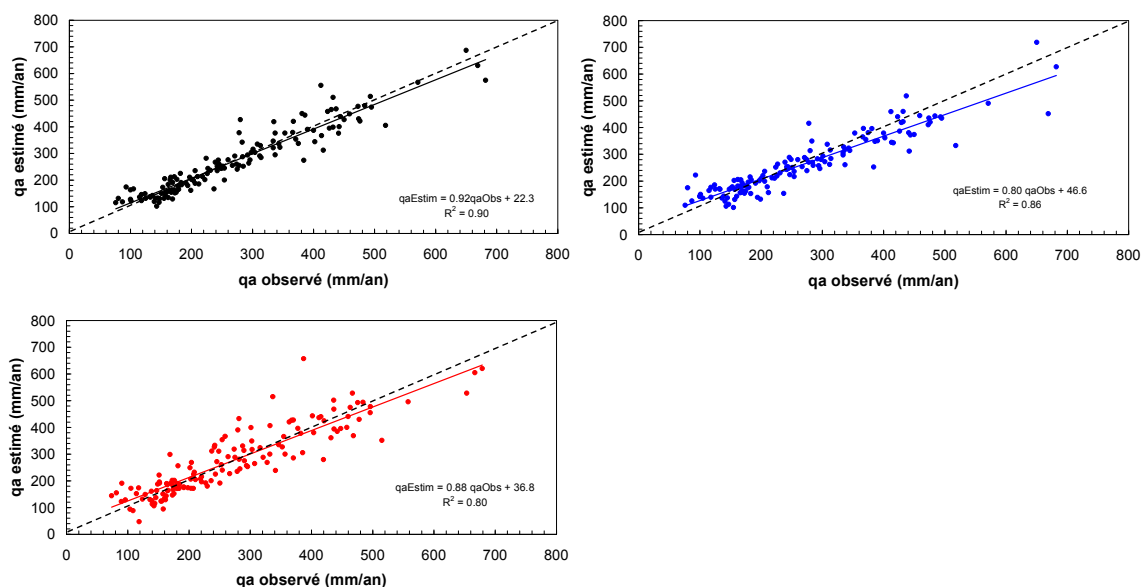
Outre l'analyse de performance, un regard sur la continuité des écoulements peut être révélateur d'incohérences hydrologiques. Le module exprimé en m³/s est additif le long du réseau hydrographique : les débits en amont d'une confluence se cumulent pour donner les débits en aval. Cette propriété valide en temps courant se transmet sur les débits moyens annuels et mensuels. Elle doit se retrouver dans les cartes produites.

Il faut veiller à travailler sur des données réajustées climatiquement si les séries sont courtes (durée inférieure à 18 ans). En travaillant sur des périodes d'observation différentes - parfois même disjointes - il y a un risque de représenter une variabilité temporelle plutôt qu'un effet spatial. Par conséquent, il est exclu de manipuler des variables hydrologiques extraites de chroniques non concomitantes. Il faut donc homogénéiser la période d'étude en visant la plus longue.

Il ne faut pas hésiter à incorporer des stations hydrométriques en périphérie immédiate dans les différentes procédures, cela permet de mieux « border » les estimations à la frontière des bassins versants.

Cette note ne propose pas une méthode à appliquer. Des techniques de régionalisation sont encore en développement ou testées, notamment au Cemagref dans le cadre de convention de recherche avec l'ONEMA et l'Agence Seine Normandie. Pour ce dernier secteur, trois équipes du Cemagref ont appliqué des méthodes d'estimation différentes (deux d'origine statistique et la troisième s'appuyant sur LOIEAU). L'objectif annoncé est de comparer la performance des techniques et de consolider les estimations en construisant une carte de consensus (par « pondération » des valeurs estimées par chaque méthode selon les incertitudes pesant sur les estimations). L'idée est de valoriser les avantages de chaque méthode là où elle s'exprime.

A titre d'illustration, les incertitudes mesurées par validation croisée sur les trois méthodes sont données dans la figure suivante (chaque point est un bassin jaugé, l'estimation est faite sur la base de l'information extraite des autres stations, la droite en pointillé est la première bissectrice, la droite de corrélation est tracée en couleur en trait plein) :



Bibliographie

Baran, P., Courret, D., Larinier, M., 2008. Les méthodes d'aide à la détermination des valeurs de débit réservé au niveau des aménagements hydroélectriques, Note technique Cemagref-Onema-ENSEEIH, novembre 2008.

Catalogne, C., Sauquet, E., 2010. Interpolation des courbes des débits classés Rapport d'avancement. Rapport d'étude Cemagref-ONEMA, 84 pages.

Catalogne, C., 2010. Amélioration des méthodes de prédétermination des étiages en valorisant les données disponibles en site peu ou pas jaugé. Rapport d'avancement de première année de thèse.

CEMAGREF, 1987. Guide méthodologique en vue de l'estimation du module du cours d'eau. Note technique, 10 pages.

Engeland, K., Hisdal, H., Beldring, S., 2006. Predicting low flows in ungauged catchments. In: Climate Variability and Change—Hydrological Impacts (Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006). IAHS Publ. 308. IAHS Press, Wallingford, UK, 63-168

Fernandez, W., Vogel, R.M., Sankarasubramanian, A., 2000. Regional calibration of a watershed model. Hydrological Sciences Journal, 45(5): 689-707.

Folton N., Lavabre J., 2006. Regionalisation of a monthly rainfall-runoff model for the southern half of France based on a sample of 880 gauged catchments. In: Climate Variability and Change—Hydrological Impacts (Proceedings of the Fifth FRIEND World Conference held at Havana, Cuba, November 2006). IAHS Publ. 308. IAHS Press, Wallingford, UK, 264-277.

Folton N., Lavabre J., 2007. Approche par modélisation pluie-débit pour la connaissance régionale de la ressource en eau: application à la moitié du territoire français. La Houille-Blanche, 3, 64-70.

Gottschalk, L., 1993. Interpolation of runoff applying objective methods. Stochastic Hydrology and Hydraulics 7, 269-281.

Laaha, G., Blöschl, G., 2006. Seasonality indices for regionalizing low flows. Hydrological Processes, 20: 3851–3878.

Leblois, E., 2010. Interpolation des débits au sein d'HYDRODEM. Note technique, 18 pages.

Matheron, G., 1965. Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux sciences de la nature. Ed. Masson, Paris (in French).

Merz, R., Blöschl, G., 2005. Flood frequency regionalisation – spatial proximity vs. catchment attributes. Journal of Hydrology 302, 283–306.

Michel, C., Makhlof, Z., 1994. A two parameters monthly water balance model for the French watersheds, Journal of Hydrology, 162, 299-318.

Pardé, M., 1933. Fleuves et rivières. Collection Armand Colin, Paris, 224 pages.

Riffard, M., 2010. Estimation des débits de référence sur le bassin de la Seine. Rapport d'avancement de première année de thèse.

Sauquet E., Gottschalk L., Krasovskaia I., 2008. Estimating mean monthly runoff at ungauged locations : an application to France. *Hydrology Research*, vol. 39, n° 5-6, p. 403-423.

Servat, E., Dezetter, A., 1993. Rainfall-runoff modelling and water resources assessment in northwestern Ivory Coast. tentative extension to ungauged catchments. *Journal of Hydrology*, 148(1-4): 231-248.

Skøien, J. O., Merz, R., Blöschl, G., 2006. Top-kriging – geostatistics on stream networks. *Hydrol. Earth Syst. Sci.*, 10, 277–287, 2006.

Vallaud, R., 2010. Méthodologie d'estimation des débits réservés en région karstique. Rapport de TFE, ENTPE Vaux en Velin, juin 2010, 91 pages + annexes.



Direction générale
Parc de Tourvoie
BP 44 - 92163 Antony cedex
Tél. 01 40 96 61 21 - Fax 01 40 96 62 25
www.cemagref.fr