

DENSITE DES POINTS DE MESURE, TYPES ET LIMITES DES MODELES D'INTERPOLATION

Daniel JOLY ⁽¹⁾, Thierry BROSSARD ⁽¹⁾, Hervé CARDOT ⁽²⁾, Jean CAVAILHES ⁽³⁾,
Mohamed HILAL ⁽³⁾, Pierre WAVRESKI ⁽³⁾

⁽¹⁾ ThéMA, ; CNRS, Université de Franche-Comté, Besançon

Mél : {[daniel.joly.thierry.brossard](mailto:daniel.joly.thierry.brossard@univ-fcomte.fr)}@univ-fcomte.fr

⁽²⁾ Université de Bourgogne, Département de mathématique, Dijon

Mél : Herve.Cardot@u-bourgogne.fr

⁽³⁾ CESAER, INRA, Dijon

Mél : {herve.cardot . mohamed.hilal . jean.cavailhes . pierre. [wavreski](mailto:wavreski@enesad.inra.fr)}@enesad.inra.fr

Introduction

L'interpolation est un domaine clé de la climatologie puisqu'elle permet de restituer des champs continus à partir de variables mesurées ponctuellement. Parmi les nombreuses méthodes d'interpolations concurrentes, deux familles se sont dégagées au fil du temps pour la valeur de leurs résultats : le krigeage, méthode stochastique, et les régressions, méthode paramétrique. Compte tenu des contraintes statistiques qui leur sont associées, ces méthodes ne sont pas interchangeable et demandent certaines précautions d'emploi. Après avoir éclairé cette question, nous proposons de combiner krigeage et régressions dans une même procédure pour améliorer sensiblement la qualité des interpolations climatiques. Ce sera l'objectif de cet article.

Ce travail s'inscrit dans une recherche pluridisciplinaire réalisée dans le cadre d'un contrat entre la Direction Générale de l'Urbanisme, de l'Habitat et de la Construction (DGUHC) et deux institutions de recherche : le CESAER, INRA (Dijon) et ThéMA, CNRS et Université de Franche-Comté (Besançon), qui rend nécessaire une l'interpolation climatique (Mitas et Mitasova, 1999) car il convient de disposer d'une information continue sur l'ensemble du territoire national. Toutefois, au-delà du cadre bien défini de cette application de recherche, les propositions faites ici en matière d'interpolation intéressent le champ de la climatologie dans son ensemble.

Les données climatiques proviennent de Météo France ; elles concernent au premier chef les températures et les précipitations, mais aussi le nombre de jours de brouillard, la fréquence du vent, la durée d'ensoleillement, etc. Ces informations ponctuelles proviennent de stations en réseau, dont la densité diffère d'une variable à l'autre. Cette hétérogénéité spatiale initiale des données va évidemment peser sur le choix des procédures d'interpolation et leur nécessaire différenciation selon les variables traitées. Dans cette démarche en effet, la question centrale est de savoir si le même mode d'interpolation peut s'appliquer valablement à des mesures très hétérogènes comme les températures et la durée d'ensoleillement par

exemple, saisies respectivement en 1495 et 111 stations. Aussi, des tests ont permis de préciser et de comparer la qualité des estimations obtenues du krigeage et des régressions.

Ces tests porteront sur les normales de température et d'ensoleillement enregistrées sur le territoire métropolitain, Corse exclue. Les ressources d'un SIG spécifique ont été requises pour mettre en œuvre notre protocole de test.

1. DONNEES ET METHODE

1.1. Données climatiques

Les données climatiques sont collectées par Météo-France sur l'ensemble du territoire français ; elles sont ensuite filtrées (correction des données « aberrantes ») et archivées. Seront utilisées les « normales » calculées sur la base des enregistrements effectués sur 30 années consécutives (1970-2000). Les tests concerneront les valeurs mensuelles des températures exprimées en °C et de la durée de l'ensoleillement exprimée en heures.

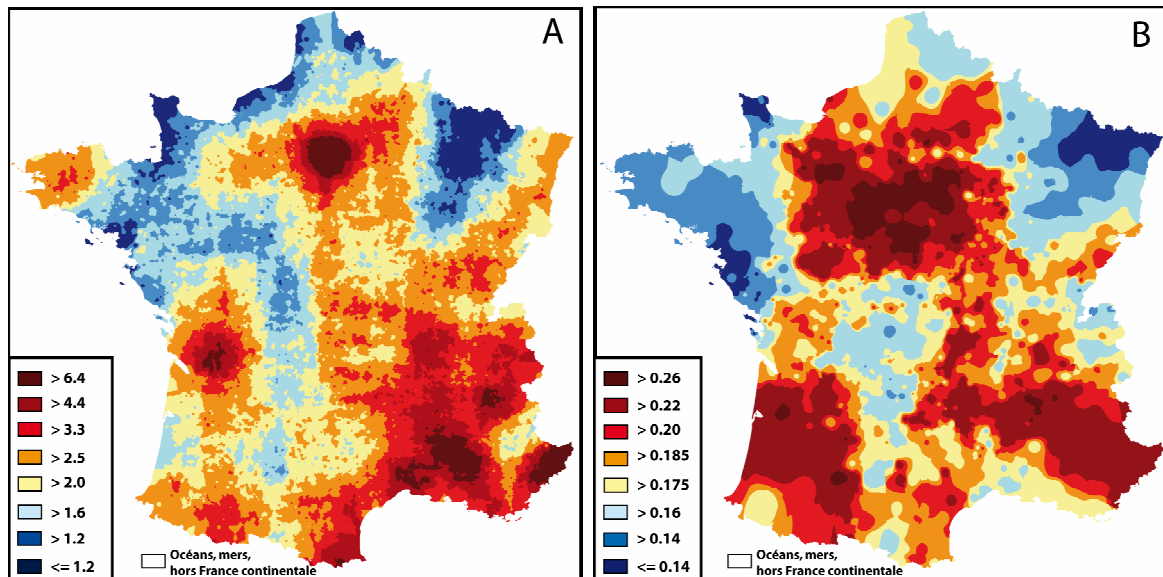


Figure 1 : Densité des réseaux (nombre de stations pour 1000 km²) enregistrant la température (A) et la durée d'insolation (B).

Au vu des cartes (figure 1), on peut être surpris de l'extrême disparité spatiale des densités de stations. Les écarts les plus forts concernent les stations thermiques : entre minimum (0,8 station/1 000 km²) et maximum (33 stations/1 000 km²) le rapport est de 1 à 40 ! Avec une densité toujours inférieure à 2 stations/1 000 km², la Lorraine, les départements littoraux allant du Nord à la Normandie, la Bretagne, Finistère excepté, la Vendée, les Pays de Loire et leur prolongement au sud vers l'Indre et l'Indre-et-Loire apparaissent comme moins bien couvertes tandis que l'Île-de-France, Poitou-Charentes, Rhône-Alpes, Auvergne, Provence-Alpes Côte d'Azur (sauf les Hautes-Alpes) et Languedoc-Roussillon atteignent 6 stations/1000 km². Pour la durée d'insolation, les écarts sont plus réduits et varient du simple au triple, de 0.1 à 0.3 station/1000 km². Les régions les moins bien dotées sont l'Alsace, la Lorraine, le Nord-Pas-de-Calais, la Bretagne, les Normandie, ainsi que l'ouest du Massif central.

La raison de ces écarts mérite considération. L'enregistrement de l'ensoleillement se fait par des stations synoptiques, généralement localisées près des aéroports des grandes villes. Les écarts de densité, faibles au demeurant, trouvent ici une part de leur explication. S'agissant des températures, l'activité agricole explique certaines différences. Les régions viticoles sont bien pourvues tandis que leurs voisines le sont peu ; la Champagne et l'Alsace s'opposent à la Lorraine, les Charentes au Limousin. Mais, d'autres différences sont plus délicates à cerner : entre les Hautes-Alpes et les deux Savoie, y a-t-il un rôle du relief, des activités touristiques où agricole ? De même, la disparité entre le Finistère, bien doté, et les Côtes d'Armor ou le Morbihan reste sans raison apparente. Quoi qu'il en soit, cette forte hétérogénéité spatiale des réseaux de mesure devra être prise en considération au moment des interpolations.

1.2. Constitution de la base de données géographiques

Ces données serviront à caractériser, dans leur diversité, les sites et les situations géographiques du territoire susceptibles d'expliquer la distribution spatiale des variables climatiques retenues. Pour constituer cette base, nous avons traité et géré par système d'information géographique (SIG) deux sources primaires (Fury et Joly, 1995): un modèle numérique de terrain (MNT) et une carte d'occupation du sol. L'un et l'autre sont au format raster et s'ordonnent sur un carroyage de référence dont chaque cellule fait 250 m de côté. Au total, 8 704 283 cellules sont nécessaires pour couvrir le territoire.

Le modèle numérique de terrain, produit par l'IGN, est livré avec une précision métrique en altitude. Par diverses procédures de calcul, courantes en géomatique et fondées sur l'algèbre cartographique et sur la trigonométrie, sont dérivées du MNT, outre l'altitude, de nouvelles variables telles que :

- La pente et l'orientation des versants ;
- La rugosité topographique qui est un indicateur des irrégularités de relief (elle peut être nulle en présence de plat ou sur un versant parfaitement rectiligne) ;
- L'indice d'encaissement ou de surélévation (un fond de vallée exigu prend une valeur négative tandis qu'un point dominant -crête, sommet- est doté d'une valeur positive) ;
- Le rayonnement global théorique qui est calculé pour le solstice d'été (21 juin) en tenant compte des masques topographiques jusqu'à cinq kilomètres autour de chaque point.

La carte d'occupation du sol provient de la base européenne Corine Land Cover. Plusieurs informations sont tirées de cette source :

- Un indice de végétation qui a été reconstitué à partir des types d'occupation du sol auxquels on a attribué une valeur d'indice standard (5 pour le bâti dense de centre ville, aéroport, etc. ; 250 pour la forêt compacte). Cet indice fournit une approximation de l'abondance de la biomasse au voisinage des points considérés lors des calculs (Joly, 2008) ;
- Des distances aux principaux types d'occupation du sol :
 - * distance à la forêt,
 - * distance à l'océan ou à la mer le plus proche.

Au total, la base comporte donc 11 couches, 8 dérivées du MNT et 3 de Corine Land Cover.

1.3. Méthodes d'interpolation

Nous nous contenterons de donner les caractéristiques distinctives des deux méthodes d'interpolation utilisées sans entrer dans le détail. Rappelons que les données d'observation

du climat sont stationnelles et régionalisées. Elles sont repérées dans un espace à deux dimensions (à trois si l'on tient compte de l'altitude, mais, en l'occurrence, l'altitude est considérée comme attribut des unités de carroyage et non comme dimension spécifique du référentiel spatial). Dans ses fondements théoriques, l'analyse spatiale consiste à expliquer la distribution des objets géographiques à partir de leurs coordonnées et de leur situation relative dans un contexte où d'autres objets géographiques sont présents (Arnaud et Emery, 2000). Schématiquement, l'interpolation appréhende l'espace selon deux points de vue :

- Le premier est fondé sur l'autocorrélation spatiale (Philippe et Piégay, 2001) et met en œuvre des techniques qui s'appuient uniquement sur le positionnement en coordonnées des points de mesure ;
- La seconde prend en compte les caractéristiques attributaires des points et de leur contexte, sous forme de variables « explicatives » telles que altitude, pente, etc. ; celle-ci interviendront de manière explicite dans les calculs d'interpolation.

1.3.1. Le krigeage

La localisation des points de mesure en coordonnées X et Y (longitude et latitude), autorise plusieurs types d'interpolation tels que, pour les plus connus d'entre eux, l'interpolation gravitaire, le krigeage, la spline cubique. Chaque méthode part de l'idée que la valeur (inconnue) d'un point dépend exclusivement de son positionnement dans le semis des points qui l'entourent et dont la valeur est connue (Gratton, 2002). La différence entre ces trois techniques tient à la forme de la fonction qui modélise la variation et aux facteurs de pondération qui permettent d'estimer une valeur pour tout point de l'espace (Merlier, 2001), objets de l'interpolation. Seul le krigeage a été testé ici. Il s'agit d'une méthode que l'on qualifie de « méthode optimale », d'« analyse objective » selon les domaines d'application où elle est utilisée. Son objectif est d'identifier la distance optimale à l'intérieur de laquelle les points de mesure sont autocorrélés. Cette distance, appelée portée, correspond au point d'infléchissement du variogramme et délimite la frontière entre deux échantillons :

- Celui qui, situé en deçà de la portée, regroupe les points qui présentent une organisation spatiale distinctive ;
- Celui qui, au delà de la portée, présente des variations aléatoires par rapport au précédent.

La procédure du krigeage est assez complexe et requiert une bonne connaissance des statistiques appliquées à l'espace (Burrough, 1986 ; Oliver, 1990). Le krigeage est basé sur la théorie des variables régionalisées qui stipule que la variation spatiale d'un phénomène quelconque représenté par des valeurs quantitatives, est statistiquement homogène sur toute la surface considérée : le même modèle de distribution doit régir tous les points de la zone en question. La variation spatiale est quantifiée par le semi-variogramme (qui est utilisé pour analyser la dépendance des observations spatiales, Smith *et al*, 1993) et la valeur attribuée à un point estimé est fonction des points valués qui l'entourent, elle est déterminée par le comportement spatial des données (on se sert pour cela des valeurs du variogramme) ; elle est éventuellement pondérée par la distance du point à ses voisins. Ce modèle, établi à partir du semi-variogramme, est introduit sous la forme d'un système d'équations linéaires qui fixent le poids à attribuer aux données lors de l'interpolation. Le krigeage a beaucoup été utilisé en climatologie (Courault *et Monestiez*, 1999 ; Laborde, 1995 ; Tveito *et Fjørland*, 1999).

1.3.2. Interpolations fondées sur des régressions

Cette seconde classe d'interpolation se fonde sur une démarche hypothético-déductive. Il s'agit d'identifier les facteurs susceptibles de « produire » le climat à échelle fine (Joly *et al.*, 2003). Certains facteurs physiques sont évidents :

- L'altitude qui sous-tend la relation entre pression et température (le gradient adiabatique sec est strictement dépendant de l'altitude tandis que le gradient adiabatique humide est fonction de la température et de l'humidité de la masse d'air) ;
- La pente qui contrôle les écoulements de l'air le long des versants ;
- L'orientation des versants qui détermine l'exposition aux rayons du soleil ou aux flux aérologiques, etc.

D'autres peuvent se révéler intéressants comme la distance à tel ou tel objet géographique (glacier, mer, forêt, etc.), la dimension fractale des formations arborées, etc. Ces facteurs spatialisés et archivés en tant que couche d'information dans les SIG, entrent explicitement dans les fonctions d'interpolation sous la forme de variables explicatives.

Notre démarche consiste à repérer, parmi les informations disponibles, celles qui sont significatives pour répondre au problème concret de l'interpolation. A cette fin, on estime systématiquement par régression linéaire, le coefficient de corrélation simple (de Pearson) entre la variable climatologique à interpoler et toutes les variables potentiellement explicatives qui sont archivées dans la base. Toutes les combinaisons des variables significatives au seuil de 1% sont ensuite testées de manière à identifier celle qui, présentant la meilleure valeur prédictive, sera finalement retenue.

La régression permet d'obtenir la valeur prédite de la variable climatique pour tout point de l'espace pour lequel on dispose des variables explicatives. La formalisation de ces liens donne lieu à l'élaboration d'un opérateur cartographique qui permet de transférer la connaissance acquise à partir d'une information ponctuelle à l'ensemble de l'espace. Enfin, par validation croisée (cross-validation), on compare les mesures climatiques avec les résultats obtenus de l'interpolation.

1.4. Démarches

1.4.1. Recours à une seule méthode au niveau de la France entière

Dans un premier temps, krigeage et régressions feront l'objet d'une application séparée pour chacune des deux variables (température et durée d'ensoleillement), ce qui permettra de décrire avantages et inconvénients respectifs des deux méthodes. Les régressions ne tiennent pas compte de la distance qui sépare les stations ; or, en secteur homogène, ce critère a un fort pouvoir explicatif puisque, d'abord, les deux variables d'intérêt sont caractérisées par des gradients réguliers et, ensuite, les variables attributaires présentent des gradients réguliers. De ce fait, les meilleures estimations seront fournies par le krigeage qui, fondé sur l'autocorrélation, rend bien compte de ces distributions spatiales régulières. En revanche, en secteur contrasté voire chaotique, la distance explique peu les variations climatiques pour lesquelles les facteurs topographiques dont certains sont peu ou pas autocorrélés (pente, orientation, encaissement, etc.) deviennent déterminants.

1.4.2. Régression puis krigeage globaux au niveau de la France entière

L'idéal serait de recourir, dans le même processus d'analyse, à l'une et à l'autre méthode. C'est ce qui sera proposé au titre d'une seconde procédure qui enchaînera les deux méthodes primaires : calcul de régression des valeurs climatiques fournies par Météo-France sur les variables explicatives précédemment décrites (Section 1.2), puis krigeage des résidus issus de la première phase d'analyse. Le processus est, dans son principe, analogue au krigeage avec dérive externe (Goovaerts, 1997 ; Wackernagel, 2003).

Les deux méthodes, régressions et krigeage, s'appuieront globalement sur le corpus de toutes les stations disponibles (1495 pour les températures et 111 pour l'insolation). Une seule fonction d'ajustement des variables explicatives de la température et une seule fonction

d'ajustement du semi-variogramme seront calculées pour procéder à l'interpolation des 8 704 283 cellules que comporte le carroyage.

1.4.3. Régressions locales

Appliqués globalement, régressions et krigeages fournissent d'ordinaire de bons résultats lorsqu'il s'agit d'aires d'étude bien circonscrites et climatologiquement cohérentes. Toutefois, lorsque l'interpolation s'applique à une vaste zone comme la France, les résultats sont de moindre qualité. La technique fonctionne, mais la statistique est perturbée par des contraintes discordantes dont l'origine est d'ordre scalaire. En effet, l'espace en cause intègre des systèmes climatiques distincts. Or, le fonctionnement de chacun d'eux est, sinon autonome, du moins assez largement indépendant ; de sorte que les processus responsables des variations spatiales du climat ne jouent pas partout de manière équivalente et leur déterminants doivent être modulés. C'est pourquoi, les facteurs spatiaux explicatifs étant différents d'un cas à l'autre, la statistique « globale » produit un modèle général « brouillé », qui n'est satisfaisant nulle part.

C'est à la suite de ces considérations qu'il est apparu intéressant de tester une troisième démarche qui sera fondée sur des régressions locales. Le principe général est d'estimer la température en chaque point du territoire en utilisant une information locale. Plutôt que de s'appuyer sur le corpus de toutes les stations disponibles, on résoudra les équations d'estimation sur les n stations voisines de chaque point d'interpolation. Ce paramètre n est important car il détermine la surface dans laquelle l'information sera collectée pour résoudre les régressions :

- Avec un n peu élevé (par exemple $n=20$), l'aire de recrutement des stations est restreinte (pour les températures, en moyenne 7 300 km², avec, selon les écarts de densité, un minimum de 600 km² et un maximum de 24 000 km²). L'avantage est que les n stations étant proches, elles présentent une forte cohésion climatologique. Cependant, la contrepartie du petit nombre est que la statistique est peu sûre: les seuils de signification sont assez élevés et, corrélativement, le nombre des variables explicatives retenues dans l'équation de régression a de grandes chances d'être peu important.

- Avec un n élevé (100 par exemple), on conforte la statistique. Mais, l'aire de recrutement des stations devient très large (toujours pour les températures, en moyenne 37 000 km², avec, selon les écarts de densité, avec un minimum de 3 000 km² et un maximum de 120 000 km²). Aussi, la probabilité d'être situé à l'intersection de zones climatiques différentes augmente, ce qui peut poser problème pour la cohérence des modèles.

Précisons en outre, que la durée des calculs s'allonge en fonction de n . C'est pourquoi, compte tenu de ces diverses contraintes, nous avons choisi, pour les exemples développés ci-dessous, de fixer n à 30. Selon ce schéma, chaque cellule constituant le carroyage territorial se relie aux n stations voisines les plus proches. De la sorte, les cellules contiguës qui dépendent des mêmes n stations déterminent un polygone qui servira de référence spatiale pour réaliser l'interpolation.

Un test préliminaire sert à identifier tous les polygones couvrant le territoire. Avec $n=30$, le nombre de polygones, qui dépend de n et du nombre de stations constituant le réseau d'observation, est de 65 406 pour les températures et de 2 352 pour l'ensoleillement. Ensuite, la méthode d'interpolation qui a été décrite précédemment (section 1.3.2.) est appliquée à chacun des polygones : estimation du coefficient de corrélation simple, stepwise, validation croisée et résolution de l'équation par régression multiple. Finalement, les coefficients sont appliqués aux cellules composant chaque polygone avant de procéder à l'interpolation.

Outre l'interpolation, la méthode autorise d'autres développements, notamment la cartographie des constantes et des coefficients de corrélation simple attachés à chacune des variables de la base et celle du coefficient de détermination de la régression multiple. Ces

informations complémentaires sont fondamentales et ouvrent la perspective d'une nouvelle climatologie centrée sur l'étude, à échelle locale, du fonctionnement du climat. Pour des raisons qui seront explicitées ultérieurement, la méthode des régressions locales sera, dans le cadre de cette étude, appliquée aux seules températures.

2. INTERPOLATION DES TEMPERATURES

2.1. Limites spatiales de l'interpolation

La normale des températures est fournie par Météo-France pour 1 495 stations. Le nombre élevé de points de mesure permet d'appréhender correctement la variation spatiale de ce descripteur climatologique à haute résolution car la probabilité est forte que les configurations spatiales soient échantillonnées au mieux dans leurs différences. On peut le vérifier sur une variable, l'altitude, qui influence fortement la température et qui est connue à la fois pour :

- Le corpus des 8 704 283 cellules à 250 m de résolution que compte la France continentale entière,
- Le corpus des 1 495 stations où est enregistrée la température (fig. 2).

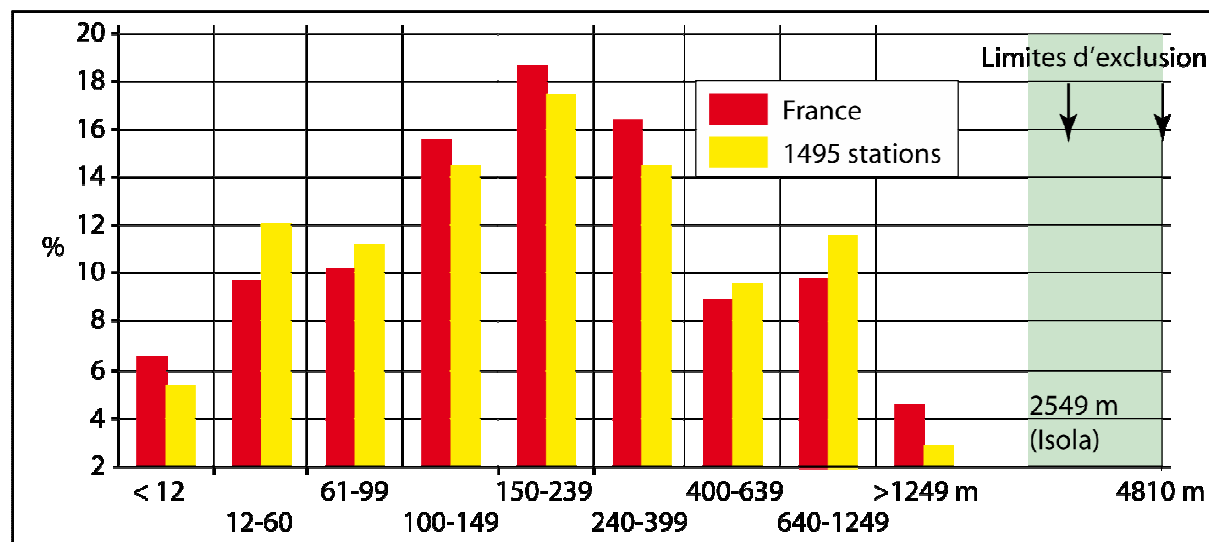


Figure 2 : Fréquences de 9 classes d'altitudes en France continentale (8 704 283 pixels) et pour le corpus des 1495 stations thermiques

Deux classes seulement présentent des divergences sensibles d'un corpus à l'autre. En premier lieu, les stations climatiques sont surreprésentées dans la classe qui regroupe les altitudes comprises entre 12 et 50 m (12,1 %, contre 9,7 % pour le total). En revanche, les altitudes élevées (>1249 m.) sont mal échantillonnées avec seulement 2,9 % contre 4 % en distribution réelle. Isola (Alpes Maritimes) étant la station la plus haute (2 549 m), aucune des altitudes supérieures à cette valeur n'est prise en compte : cette tranche d'altitude compose une *terra incognita* du point de vue climatologique. Autant la surreprésentation des altitudes les plus faibles ne gêne pas les analyses statistiques, autant l'absence de donnée sur 2 300 m d'amplitude verticale constitue un obstacle à l'interpolation, même si, par ailleurs, la technique l'autorise. C'est pourquoi nous donnerons une limite au développement des calculs en gardant une marge de 10 % au-delà de la limite maximale renseignée. Ainsi, la plage altitudinale d'interpolation ira de 0 à 2 750 m.

Lorsque l'interpolation est fondée sur des régressions multiples, c'est l'ensemble des prédicteurs qui devront être analysés, ainsi qu'il vient d'être fait pour l'altitude, afin d'établir les limites de validité de l'interpolation. En conséquence, l'espace de validité des modèles se réduit à mesure que le nombre des prédicteurs introduits augmente.

2.2. Qualité des analyses

La qualité des interpolations consiste à apprécier la valeur des résidus obtenus de la validation croisée : valeurs extrêmes, fréquence des valeurs élevées, écart type calculée pour chaque moyenne mensuelle de température et pour chacun des quatre types d'analyse. L'ensemble de ces indices et de ces mesures de dispersion révèle des écarts importants d'une méthode à l'autre.

2.2.1. Valeur des résidus de juillet

- **La méthode des régressions.** Ponctuellement, certaines stations présentent des erreurs d'estimation très élevées. Parmi les résidus négatifs, le maximum, en valeur absolue, est atteint à Sizun (29) en août avec $-4,7^{\circ}\text{C}$. En juillet, les résidus négatifs les plus forts sont atteints dans certaines stations des Alpes du nord, d'Alsace et des côtes de Bretagne et de Normandie (fig. 4A). Le résidu positif extrême, $4,9^{\circ}\text{C}$, affecte Roquesteron (06) en août. Les résidus fortement positifs se trouvent dans les Alpes du Sud et dans les Pyrénées. Enfin, précisons que 1,9 % des résidus dépassent 3°C en valeur absolue, tandis que les résidus faibles inférieurs à 1°C , représentent 60 % du corpus.

- **Le krigeage** fournit de meilleurs résultats même si, ponctuellement (fig. 3B), des écarts importants apparaissent ici aussi : les résidus maximums s'établissent respectivement à $-6,9^{\circ}\text{C}$ à Isola en août et à $4,9^{\circ}\text{C}$ à Mouthiers en avril. La fréquence des résidus supérieurs à 3°C en valeur absolue est de 1,4 % ; la fréquence des résidus faibles est de 83 %, confirmant le caractère plus resserré de la dispersion.

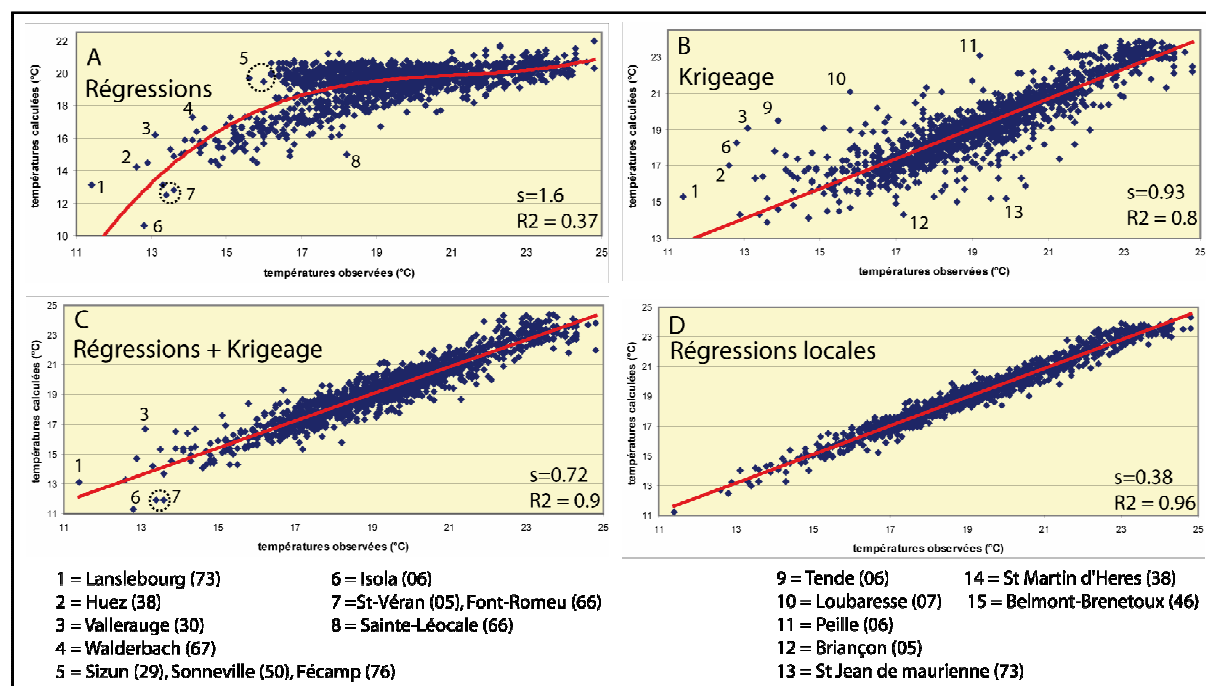


Figure 3 : Comparaison entre les températures observées et les températures prédites par les quatre démarches pour les 1495 stations thermiques (valeurs de juillet).

La procédure qui enchaîne **régressions et krigeage** améliore sensiblement les résultats (fig. 3C). En effet, le coefficient de Moran tiré des résidus issus de la phase de régression, vaut 0.88 et révèle une forte autocorrélation : un krigeage s'imposait. Les résidus ainsi obtenus prennent des valeurs extrêmes de $-3,7^{\circ}\text{C}$ et de $3,5^{\circ}\text{C}$. Enfin, la fréquence des résidus forts ($> 3^{\circ}\text{C}$ en valeur absolue) devient insignifiante (0,1%) alors que celle des résidus faibles dépasse 90 %. Cela étant, la méthode des **régressions locales** est encore plus efficace puisque les résidus ne sont jamais inférieurs à -2°C ou supérieurs à $1,9^{\circ}\text{C}$ et que la fréquence des valeurs comprises entre -1°C et 1°C s'établit désormais à 98%. Quant au R^2 , il atteint 0,95 (fig. 3D).

2.2.2. Variation spatiale des résidus de juillet

Les 1495 valeurs ponctuelles des résidus de juillet ont été interpolées par krigeage afin d'obtenir un champ continu des erreurs d'estimation produites par les quatre modèles (fig. 4). De la carte 5A à la carte 5D, la restriction graduelle des plages magenta foncé (forts résidus positifs) et bleu nuit (forts résidus négatifs) souligne la diminution des écarts type et donc la performance accrue des modèles respectifs.

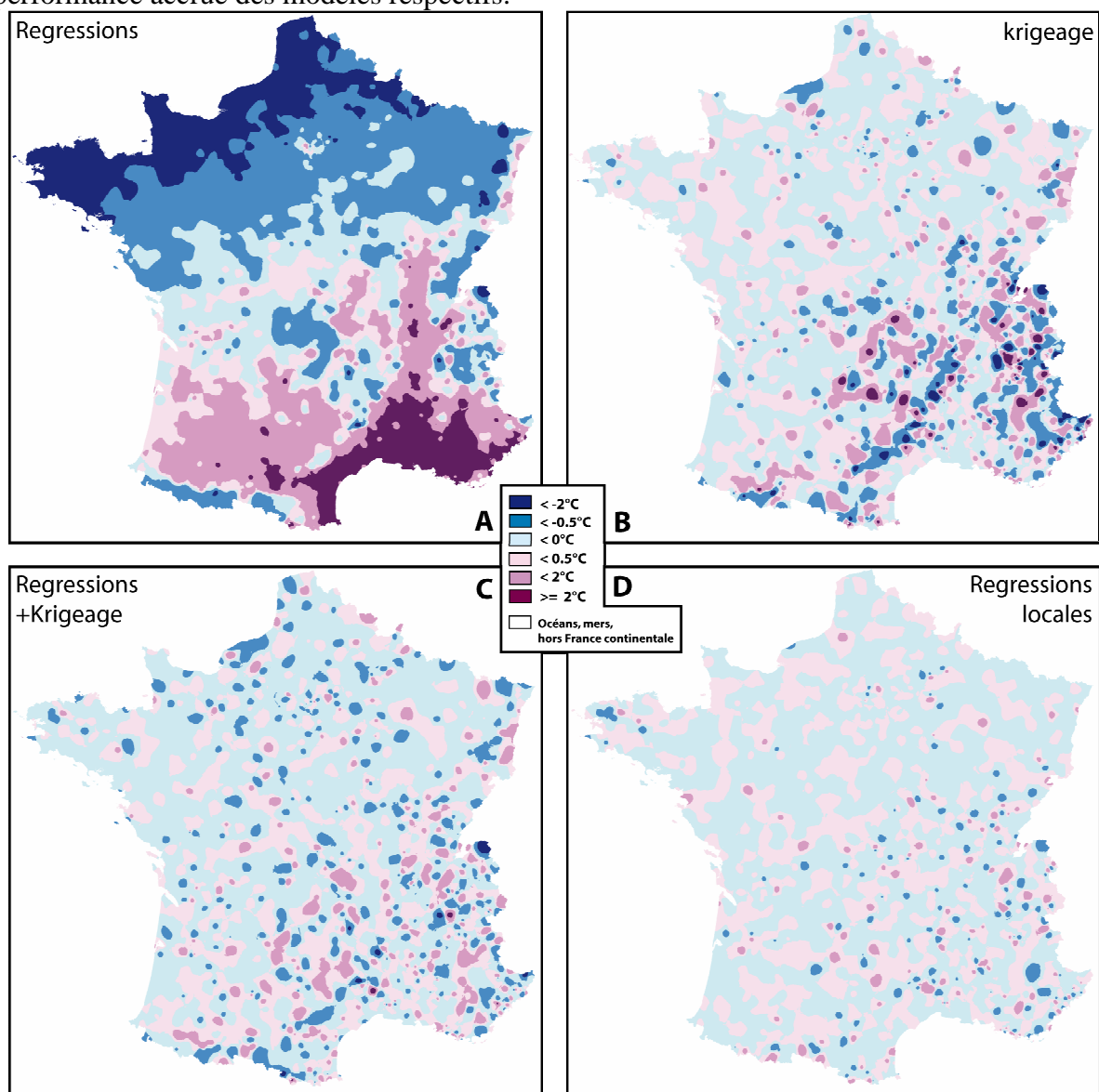


Figure 4 : Résidus issus des quatre démarches (valeurs de juillet)

En effet, la méthode des régressions est mal adaptée à l'interpolation des températures de juillet. De vastes surfaces en Bretagne, en Normandie et le long du littoral de la mer du Nord présentent des résidus inférieurs à 2°C tandis que le pourtour méditerranéen, la Provence et la vallée du Rhône se caractérisent par des valeurs supérieures à +2°C. Les trois autres cartes (fig. 4B, 4C et 4D) sont marquées par la domination des résidus faibles, compris entre -0,5 et +0,5°C. Les résidus élevés n'ont que peu d'impact. Les résidus moyens amènent toutefois des nuances ; sensibles dans la figure 4B (krigeage), ils diminuent ensuite pour devenir négligeable dans la figure 4D (régressions locales).

Notons toutefois que, dans chacune de ces trois cartes, les résidus les plus élevés se concentrent surtout dans les massifs montagneux où la juxtaposition, à faible distance, de valeurs négatives et positives traduit un effet de pépité qui révèle des situations locales en fort contraste difficiles à modéliser.

2.2.3. Variation annuelle des écarts type des résidus

La hiérarchie des quatre méthodes selon leur aptitude à fournir une bonne estimation de la température est confirmée par la variation des écarts-types mois par mois. Avec la méthode des régressions, les résidus sont hétérogènes tout au long de l'année. La figure 5 indique que la valeur d'écart type est minimale en hiver (1°C) et culmine à plus de 1,6°C en été. Les trois autres méthodes d'interpolation présentent de moindres variations mensuelles avec de meilleures performances pour le krigeage simple par rapport aux *régressions + krigeage* et surtout aux régressions locales. On assiste à une amélioration graduelle de l'interpolation avec une diminution de 0,2°C d'une méthode à l'autre. Du krigeage aux régressions locales, on passe ainsi d'une valeur de 1°C à 0,44°C en août et de 0,79°C à 0,48°C en novembre.

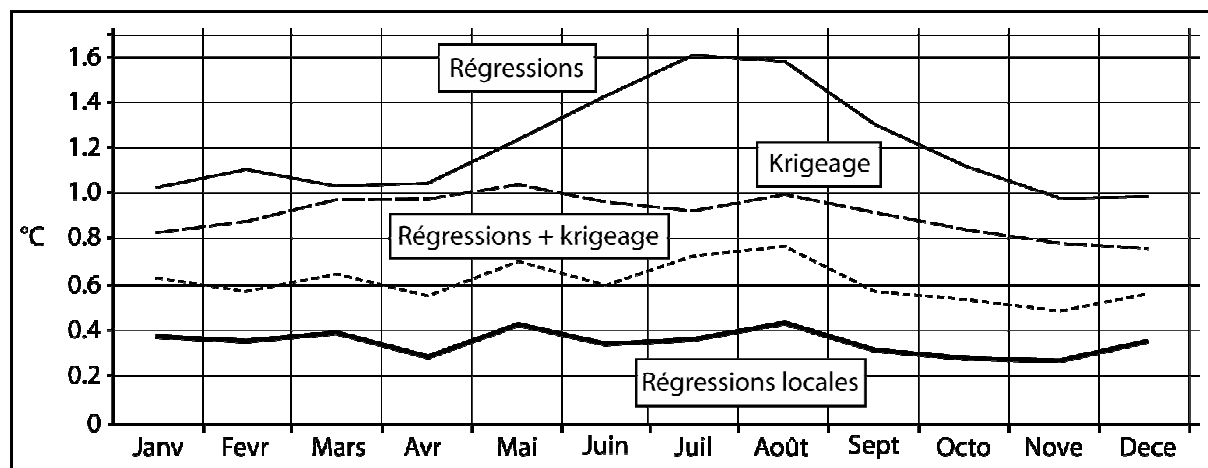


Figure 5 : Ecart type des résidus obtenus des 4 types d'analyse.

2.3. Cartes des températures de juillet

Les résultats de chaque procédure sont cartographiés (fig. 6). Sur la carte issue du **modèle de régression**, on note surtout l'influence de l'altitude puisque les massifs montagneux, froids, s'opposent au reste du territoire où les versants apparaissent avec une température légèrement supérieure aux zones planes. Ce constat est dû à la nature des trois estimateurs sélectionnés par le stepwise puis intégrés dans la régression multiple puisqu'il s'agit de l'altitude ($r=-0,56$), de la rugosité topographique ($r=-0,24$) et de l'orientation des versants ($r=0,21$). On notera les plages plus ou moins extensives de gris qui correspondent aux zones non interpolées en raison de leur altitude supérieure à 2 750 m.

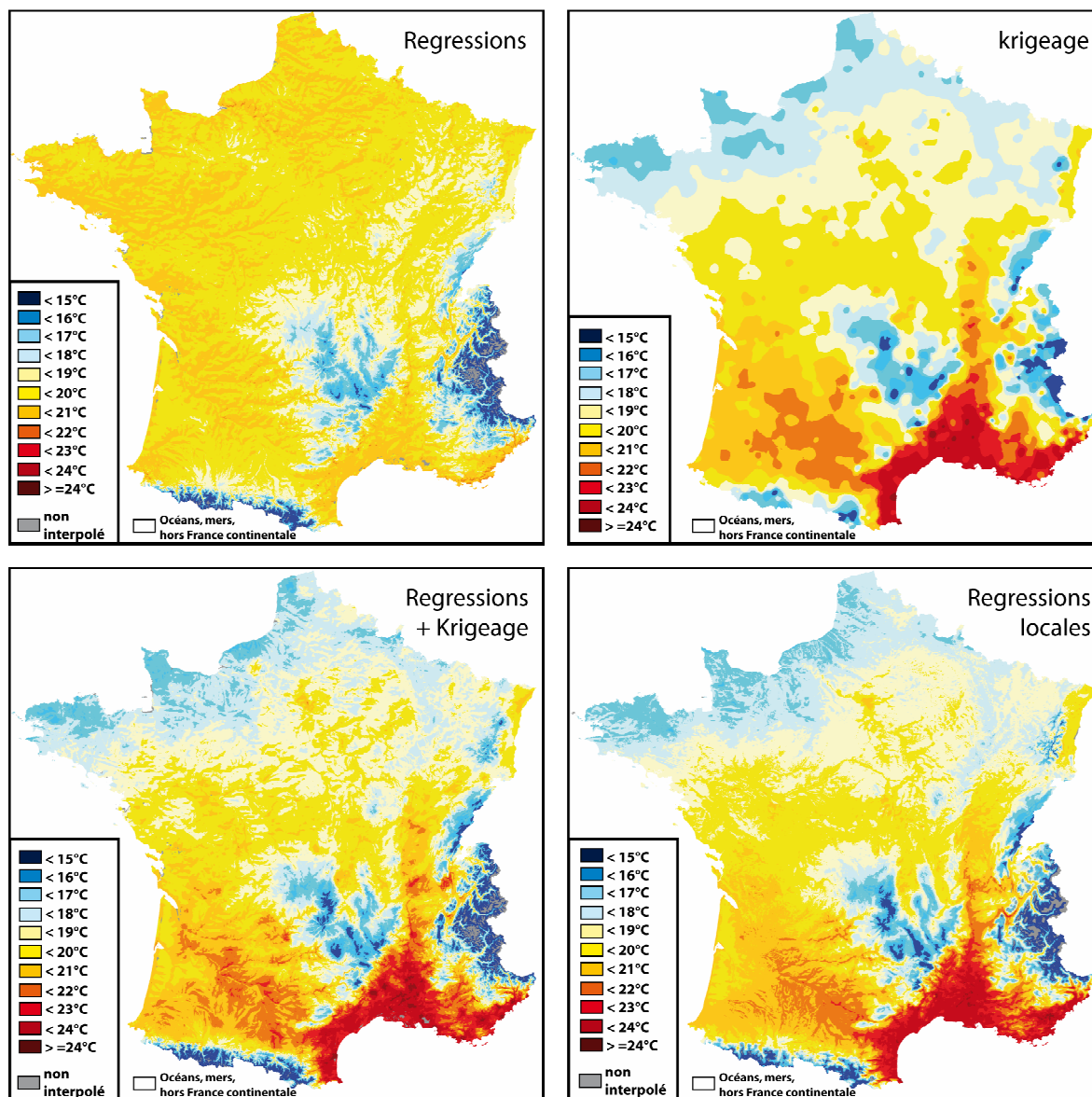


Figure 6 : Températures des quatre démarches ; mois de juillet.

Le **krigeage** délivre une carte plus conforme aux attentes. On y voit ressortir la montagne et ses plus hauts ensembles avec une température inférieure à 16°C. En opposition, la bordure méditerranéenne et son prolongement vers le nord, le long de la vallée du Rhône prennent des valeurs qui dépassent 22°C. Entre ces deux extrêmes, notons les deux pôles secondaires que sont d'abord le Bassin aquitain (21-22°C), et ensuite, le littoral de la Manche et de la Mer du Nord (17-18°C). La carte issue d'un couplage entre **régressions et krigeage des résidus** montre bien le rôle distinct de chaque modèle : le krigeage tend à lisser les distributions tandis que les régressions révèlent le rôle joué par la topographie dans la finesse de ses contrastes où s'exprime la variabilité locale.

Enfin, la carte des **régressions locales** est voisine dans ses grands traits de la précédente. A l'issue de cette phase, la question se pose de savoir s'il existe une autocorrélation qui pourrait convenir au krigeage. C'est effectivement le cas puisque le coefficient de Moran

appliqué aux résidus issus des régressions locales pour les valeurs de juillet, vaut 0,55. La figure 7 montre toutefois que l'autocorrélation n'est pas homogène sur le territoire : faible dans les Alpes, les Pyrénées, le Roussillon et les Cévennes, les Pays de la Loire et la Normandie, le Nord, elle est élevée dans les Ardennes, la Lorraine, la Bretagne, la Bourgogne, le Lyonnais, le Massif central et le Bassin aquitain. Une telle situation est gênante car la théorie du krigeage stipule qu'il ne peut être bien utilisé que si le même modèle de distribution spatiale est observé en tout lieu ; or ce n'est pas le cas.

De fait, le krigeage, calculé globalement à partir d'une situation aussi hétérogène, améliore très peu les performances des régressions locales et entraîne même, parfois, une augmentation de la valeur des résidus : en juillet, le krigeage dégrade les résultats obtenus à l'issue de la phase de régression avec un écart type qui passe de 0,35°C à 0,41°C. Seul un krigeage local, opéré sur le même mode que les régressions, pourrait théoriquement améliorer les performances de cette quatrième procédure. C'est l'un des développements sur lequel nous travaillons actuellement.

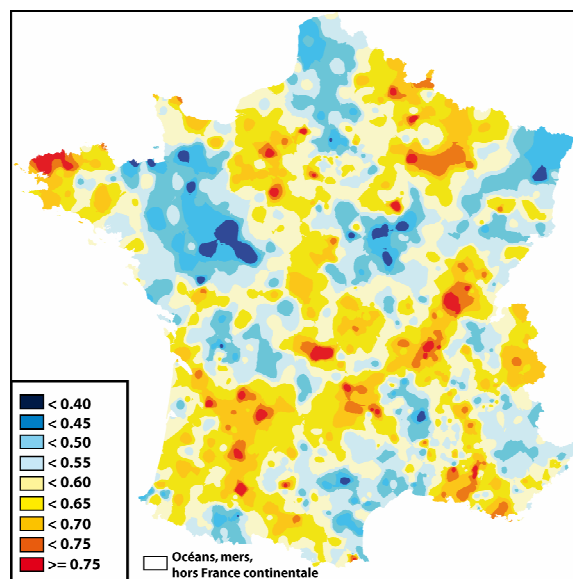


Figure 7: Coefficient de Moran ; mois de juillet

3. INTERPOLATION DE L'ENSOLEILLEMENT

3.1. Limites spatiales induites par les régressions

La régression multiple qui ajuste au mieux la durée de l'ensoleillement combine trois estimateurs : l'altitude ($r = -0,24$), la pente ($r = 0,23$) et l'indice de végétation ($r = -0,18$). La durée de l'ensoleillement est enregistrée en 111 stations. Ce faible nombre de points de mesure entraîne un mauvais échantillonnage de l'espace français pour les trois estimateurs. Les altitudes faibles, inférieures à 240 m (60 % de l'espace français) sont sur-échantillonnées, avec 80 % des stations (fig. 8). Corrélativement, les altitudes moyennes (400-1 000 m) sont sous-échantillonnées. De surcroît, Brenoux (48), la station la plus élevée du réseau n'est située qu'à 1 018 m, de sorte que la plage d'exclusion de l'interpolation s'étend sur une verticale de 3 600 m. Ainsi, l'interpolation n'est pas valide, sur le critère altitudinal, pour un bon nombre de pixels situés dans les Alpes et les Pyrénées (fig. 10).

En ajoutant le critère de pente, la zone retenue se réduit encore car elle exclut les valeurs supérieures à 5 degrés (fig. 9A). De fait, 92 % des sites de mesure offrent une pente inférieure ou égale à 2 degrés.

La forêt, dernier critère pris en compte pour le calcul de l'ensoleillement (fig. 9B), est absente de l'immédiate proximité des stations. L'indice de végétation maximum atteint est de 170, ce qui correspond à une couverture de buissons épars entrecoupés d'espaces ouverts et minéraux (routes, bâti). En revanche, les indices faibles (< 107) liés au bâti et aux formations ouvertes, sont sur-représentés autour des stations.

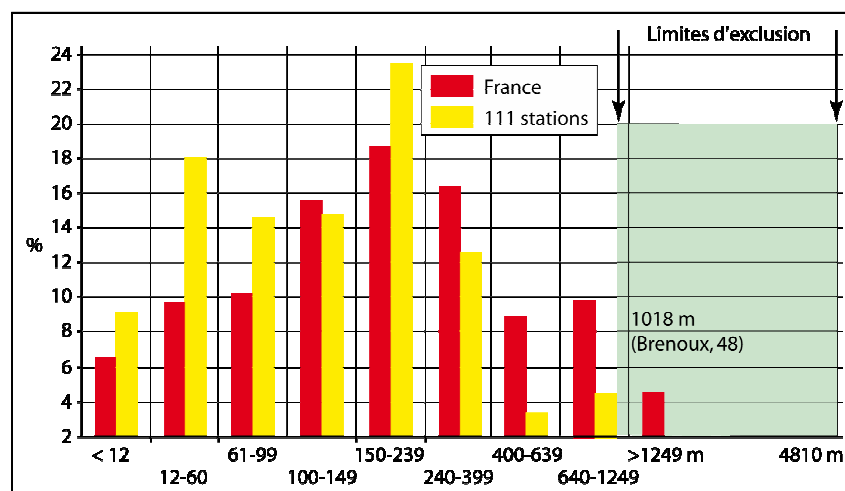


Figure 8 : Fréquences de 9 classes d'altitudes en France (8 704 283 pixels) et pour le corpus des 111 stations enregistrant la durée d'ensoleillement

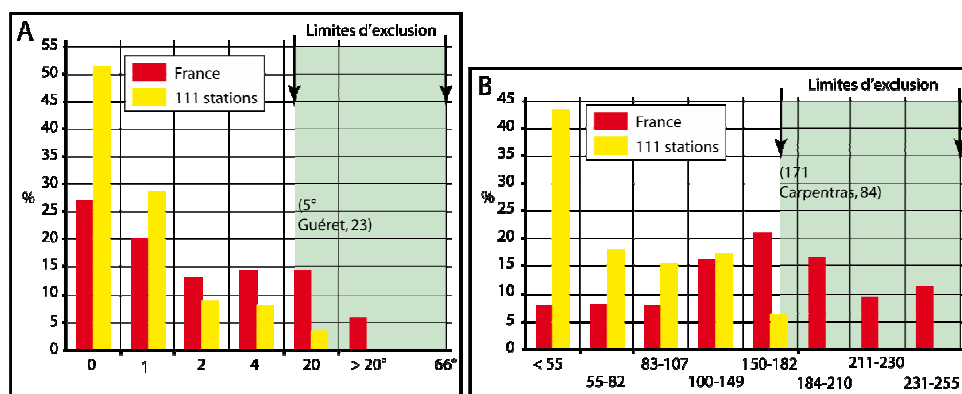


Figure 9 : Fréquences de 6 classes de pentes (A) et de 8 classes d'indice de végétations (B) en France (8 704 283 pixels) et pour le corpus des 111 stations enregistrant l'ensoleillement

Alors que le critère altitudinal, en éliminant de l'interpolation la tranche supérieure à 1 020 m, limite la zone d'exclusion aux massifs montagneux, le critère de pente ajoute à celle-ci le piémont Pyrénéen, un liseré autour des Alpes, la bordure orientale du Massif central et certains versants du Jura, des Vosges, des Maures et de l'Estérel (fig. 10). Mais l'indice de végétation est encore plus exclusif puisque ce critère laisse tout juste 15 % du territoire comme apte à l'interpolation.

Même si les estimateurs intégrés dans la régression multiple sont significatifs au seuil de 1 %, leur validité au sens spatial du terme est rejetée. Faute d'une représentativité suffisante, l'espace de non interpolation est beaucoup trop vaste pour que les régressions, seules ou combinées avec un krigeage, puissent servir de support au calcul. Les interpolations locales sont encore plus restrictives car, plutôt que de s'appuyer sur 111 stations, échantillonnage, déjà insuffisant compte tenu de l'uniformité des sites sur lesquels sont installées les postes météorologiques, elles reposent sur seulement 30 stations. Le risque qu'elles accentuent encore cette homogénéité est important, ce qui réduit d'autant leur intérêt. Le krigeage est la seule méthode possible d'interpolation de la durée de l'ensoleillement.

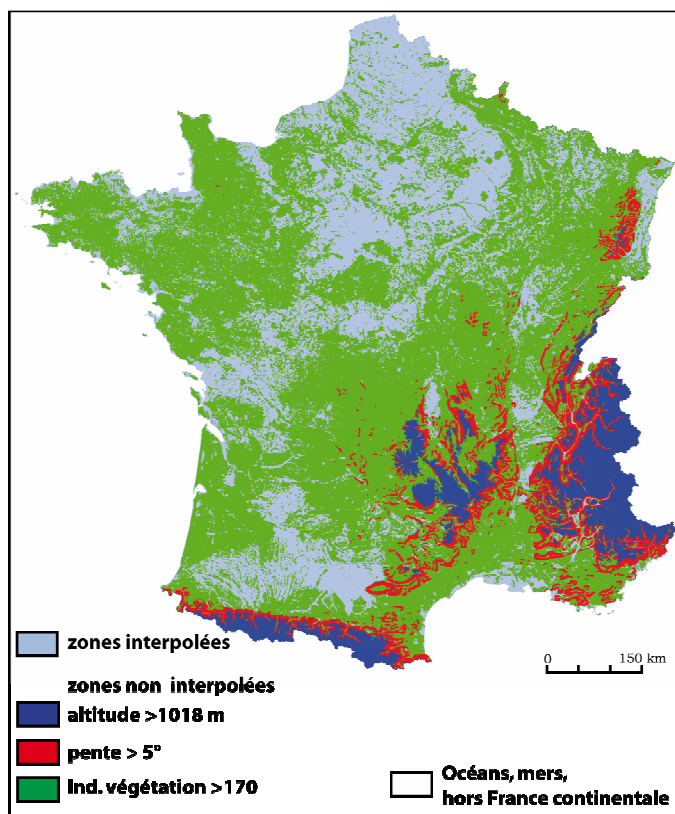


Figure 10 : Zones où l'interpolation de la durée d'insolation est/n'est pas possible sur les critères des altitudes et de la pente

3.2. Qualité du krigeage

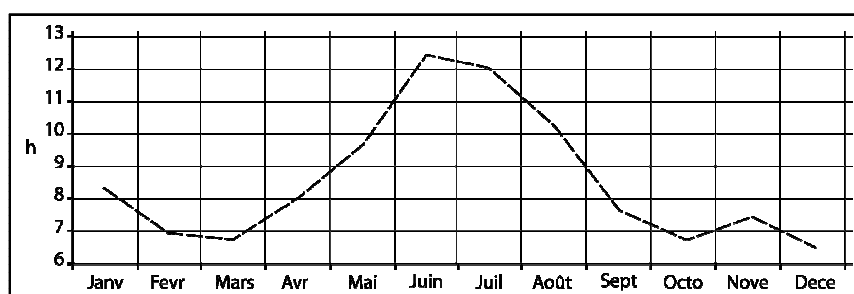


Figure 11 : Ecart-type des résidus obtenus du krigeage de l'ensoleillement

La qualité du krigeage dépend d'abord de la saison. L'écart-type des résidus, pour l'ensemble de l'année, s'établit à 8 h 40, mais passe de 6 h 30 en hiver à plus de 12 h en été (fig. 11). L'ajustement, même pour les mois d'été, reste très satisfaisant (fig. 12). Le nuage de points s'aligne bien sur la droite de régression et peu de stations s'écartent de cette dernière. Ponctuellement, les résidus extrêmes sont de l'ordre de - 30 h et + 30 h.

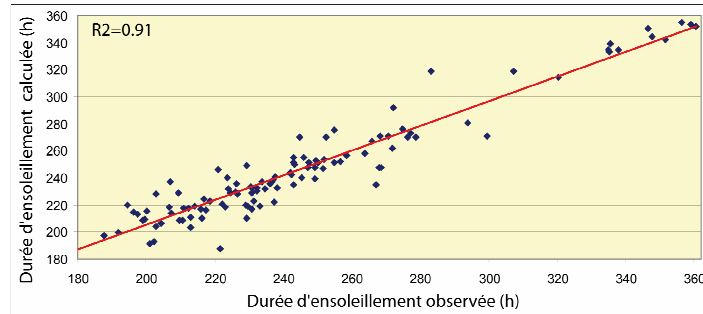


Figure 12 : Comparaison des durées d'ensoleillement observées et estimées par krigeage

La carte résultante met surtout en évidence un effet nord-sud (fig. 13). Une bande large de 150 km entre Bretagne et Nord-Pas-de-Calais, avec moins de 200 h d'ensoleillement, s'oppose à la Provence, et notamment à la basse vallée du Rhône marquée par des pointes à plus de 350 h. Entre ces deux extrêmes, notons une plage de valeurs moyennes en Charente qui ressort au milieu de valeurs plus faibles, et une zone sur les Pyrénées centrale et occidentale où la durée d'ensoleillement se réduit.

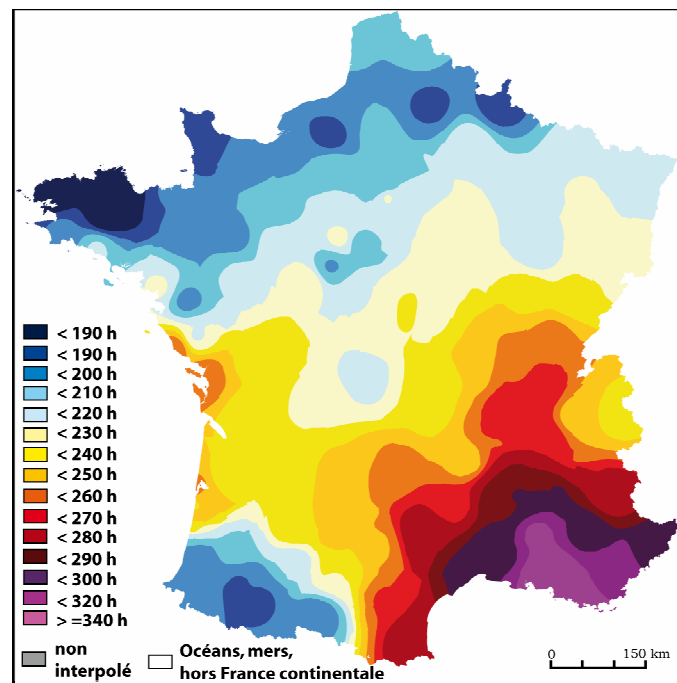


Figure 13 : Ensoleillements calculée par krigeage ; mois de juillet

Discussion, conclusions

Les deux variables proposées à titre d'exemple, température et durée de l'ensoleillement, ont été choisies en raison des différences qui les oppose en terme de densité de mesures. Ce qui conduit à élaborer des procédures d'interpolation adaptées à chaque cas. Ce n'est pas tant le nombre de postes qui est en cause : les 111 stations mesurant l'ensoleillement constituent un nombre tout à fait raisonnable du strict point de vue statistique. C'est surtout leur localisation en des sites aux caractéristiques mal ajustées à la diversité des situations, qui pose problème que ce soit pour l'altitude, les pentes ou l'occupation du sol. Les sites échantillonnés ne représentent bien que 15 % du territoire, les 85 % restants sont mal pris en compte en raison de leur altitude trop forte (supérieure à 1 020 m), de leur pente trop vigoureuse (supérieure à 6°), ou de leur végétation trop prégnante. Inversement, avec 1 495 stations enregistrant la température, la probabilité que toutes les configurations présentes sur l'espace d'étude soient échantillonnées est grande. Et de fait, l'échantillonnage est bon en terme de représentativité spatiale, sauf pour les altitudes très élevées. Ainsi, alors que les températures ont pu être interpolées avec toute la palette des méthodes envisagées, seul le krigeage a pu être retenu pour la durée d'ensoleillement.

La qualité des résultats fournis par les quatre procédures a été testée et comparée pour les températures. Un classement a été établi grâce à l'analyse des écarts type des résidus. Les méthodes de régression globale et de krigeage ordinaire donnent des résultats de moindre qualité. En revanche, lorsque ces deux méthodes sont associées au sein d'une démarche qui enchaîne le krigeage des résidus après régression, les résultats sont bien meilleurs. Mais c'est la dernière procédure, fondée sur des régressions locales, qui délivre les résidus les plus faibles et donc les meilleurs résultats.

Ainsi la qualité des résultats d'une interpolation dépend tout autant des propriétés de l'information traitée que des méthodes utilisées.

Bibliographie

- Arnaud M., Emery X., 2000. *Estimation et interpolation spatiale : méthodes déterministes et méthodes géostatistiques*. Hermès, Paris, 221 p.
- Courault D., Monestiez P. 1999. Spatial interpolation of air temperature according to atmospheric circulation patterns in southeast France. *International Journal of Climatology*, **19**: 365-378.
- Cressie N., 1993. *Statistics for Spatial Data*. Revised edition, Wiley ed., NY, 900 p.
- Fury R., Joly D., 1995. Interpolation spatiale à maille fine des températures journalières. *La Météorologie*, 8^{ième} Série, n°11, pp. 36-43.
- Goovaerts P., 1997. Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of hydrology*, **228** : 113-129.
- Gratton Y., 2002. Le krigeage: la méthode optimale d'interpolation spatiale. (disponible à www.iag.asso.fr).
- Joly D., 2008. L'information géographique au service de la climatologie. Chapitre 1. Traité IGAT (Information Géographique et Aménagement du Territoire), « *Information géographique et climatologie* », Hermes ed., pp.23-72.
- Joly D., Nilsen L., Fury R., Elvebakk A., Brossard T., 2003. Temperature interpolation at a large scale; test on a small area in Svalbard. *International Journal of Climatology*, vol. 23, pp. 1637-1654.

- Laborde J.P., 1995. Les différentes étapes d'une cartographie automatique : exemple de la carte pluviométrique de l'Algérie du Nord, *Publications de l'Association Internationale de Climatologie*, vol. 8, pp. 37-46.
- Merlier C., 2001. *Interpolation des données spatiales en climatologie et conception optimale des réseaux climatologiques*. Annexe du rapport de Météo-France concernant ses activités en rapport avec la commission for climatology (CCI) de l'OMM., 21 p.
- Mitas L., Mitasova H, 1999. Spatial interpolation. In *Geographical information systems : principles and technical issues* (vol. 1), P.A. Longley, M.F. Goodchild, D.J. Maguire, D.W. Rhind, John Wiley & sons, New-York, pp. 481-492.
- Oliver M.A., 1990. Kriging: 'A Method of Interpolation for Geographical Information Systems'. *International Journal of Geographic Information Systems*, vol. 4, n°4, 313-332.
- Philippe A., Piégay H., 2001. Pratique de l'analyse de l'autocorrélation spatiale en géomorphologie : définitions opératoires et tests. *Géographie physique et quaternaire*, vol. 55, n°2, pp. 111-129.
- Smith JF, Halvorson JJ, Robert IP, 1993. Using multiple variable indicator kriging for evaluating soil quality. *Soil Science Soc. Am.*, vol 57, pp 743-749.
- Tveito O., Førland E.J., 1999. Mapping temperatures in Norway applying terrain information, geostatistics and GIS. *Norsk geografisk tidsskrift*, vol. 53, pp. 202-212.
- Wackernagel H., 2003. *Multivariate Geostatistics : an introduction with applications*. Springer-Verlag ed., Berlin, Third completely revised edition.