

L'incertitude dans un processus de spatialisation de données climatiques

Rémi Lhotellier

Institut de Géographie Alpine

14 bis avenue Marie Reynoard – 38100 Grenoble, France

remi.lhotellier@laposte.net

MOTS - CLÉS

Incertitude
Spatialisation
Température de l'air
Montagne

RÉSUMÉ

Dans un processus de spatialisation de données climatiques à partir de mesures relevées ponctuellement, il est nécessaire de prendre en compte tous les niveaux d'imprécision et d'incertitude inhérents à chacune des étapes de la modélisation, de la mesure en elle-même jusqu'à la validité du champ obtenu. Le contexte de travail, à savoir le type de données utilisées, l'échelle d'étude ou encore le relief du terrain, doit permettre de mettre en évidence les incertitudes cumulées au sein de ce processus, mais également de pouvoir les quantifier et les cartographier. Toute carte thématique doit ainsi s'accompagner d'une carte de qualité d'estimation. Le cas présenté concerne la spatialisation de la température de l'air sur quatre départements français essentiellement alpins.

KEY WORDS

Uncertainty
Spatialization
Air temperature
Mountain

ABSTRACT

Uncertainty in the process of spatialization of climatic data

Within a spatialization process, starting from punctually measured climatic data, taking into account all levels of inaccuracy and uncertainty in each step of modelling is necessary, from the measure to the temperature field. The context of the work, namely the type of used data, the studied scale or the relief shapes, must make it possible to highlight the cumulated uncertainties within this process. It also allows us to quantify and map them. Any thematic map must thus be accompanied by a map of quality of estimate. The study deals with the spatialization of air temperature over four French alpine departments.

1. Introduction

Les paramètres climatiques, et parmi eux la température de l'air, sont généralement mesurés ponctuellement. Les besoins applicatifs s'accommodent malheureusement mal de cette dispersion des mesures. Ainsi, l'agriculteur veut connaître la température dans son champ et non à la station de mesure la plus proche, le gestionnaire de station de ski a besoin de données sur les pentes à enneiger. La spatialisation des données climatiques est donc bien aujourd'hui une exigence des utilisateurs, désormais facilitée par le recours aux systèmes d'information géographique.

Cette communication vise à présenter un aspect d'un travail de thèse sur la spatialisation des températures en zone de montagne alpine, réalisée au laboratoire SEIGAD de l'Institut de Géographie Alpine, en collaboration avec Météo-France, et dans le cadre du programme Européen *COST 719 : GIS in climatology and meteorology*. Cette action européenne, comme son nom l'indique, avait pour principal objectif d'accroître le potentiel des systèmes d'information géographique dans les domaines de la climatologie et de la météorologie, aussi bien en termes de stockage et diffusion de l'information que comme outil d'analyse et

de cartographie (Chapman et Thornes, 2003 ; Dyras *et al.*, 2005 ; Tveito *et al.*, 2005).

L'intérêt de cette recherche était de proposer une méthode pour obtenir des données spatialisées, à partir des relevés ponctuels de température de l'air des stations Météo-France, à l'échelle d'une vaste moitié nord des Alpes françaises. Nous nous attacherons ici à préciser les multiples niveaux d'incertitude inhérents à ce type de problématique, et à proposer éventuellement des évaluations et des cartographies de ces niveaux.

2. Cas d'application et méthodologie

2.1 Données et site d'étude

Du Léman, au nord, à la Durance, au sud, notre zone d'étude s'étale en latitude sur deux cent cinquante kilomètres. Dans la direction opposée, cent quatre vingt dix kilomètres séparent le Rhône, à l'ouest, des sommets alpins annonçant la frontière italienne, à l'est. Les variations rapides de relief ajoutées à l'effet de la continentalité puis à l'influence méditerranéenne engendrent sur cet espace une palette climatique singulière. Le point commun entre les quatre départements étudiés réside évidemment dans leur caractère alpin, même si le climat de l'Isère par exemple, ne peut être considéré à dominante alpine que dans sa moitié est. Autre délimitation de taille, celle généralement admise au niveau du col du Lautaret, entre Alpes du Nord et Alpes du Sud : à ce titre, le département des Hautes-Alpes s'éloigne climatiquement des trois autres par son influence méditerranéenne marquée.

La recherche a été menée à partir d'un échantillon de 168 stations météorologiques du réseau Météo-France, réparties sur quatre départements alpins, du nord au sud : Haute-Savoie, Savoie, Isère, Hautes-Alpes (soit une surface d'étude d'environ 24000 km²), durant les années 1990 à 1995. Cette période a été sélectionnée pour permettre d'utiliser les relevés effectués ponctuellement dans le contexte des Jeux Olympiques de 1992 à Albertville, en Savoie. Les températures minimales et maximales quotidiennes nous ont été fournies (soit 278897 enregistrements), en degré et dixièmes, tout comme les coordonnées géographiques des stations, au format degré minute seconde. Les notions de température maximale et de température minimale sont définies internationalement¹.

¹ Météo-France rapporte les définitions suivantes :

- « La température minimale en un lieu donné est la température atmosphérique la plus basse atteinte en ce lieu au cours d'un intervalle de temps prédéterminé. Des intervalles de temps successifs de 24 heures déterminent ainsi en chaque site de mesure la température minimale quotidienne (abréviation : TN), qui survient le plus souvent peu après le lever du jour. Afin d'assurer la synchronisation des observations à l'échelle mondiale, on convient en météorologie que cette dernière grandeur est, pour un jour J, la température la plus basse qui ait été relevée sur le site entre le jour J - 1 à 18 heures UTC et le jour J à 18 heures UTC.

Couplées à un système d'information géographique au sein duquel un modèle numérique de terrain de résolution 150 mètres constitue l'information de base, ces données de températures vont pouvoir être modélisées.

2.2. Modélisation de la température de l'air

L'intérêt premier de l'étude consiste en une modélisation puis en une spatialisation de la température de l'air à échelles spatiale et temporelle fines. La méthodologie retenue pour effectuer cette spatialisation des températures combine l'usage d'analyses statistiques et de systèmes d'information géographique : un modèle de régression multiple pas à pas est construit à partir de variables topographiques ou environnementales, grâce à un système d'information géographique (Joly *et al.*, 1994). Le système d'information géographique nous permet par exemple, à partir du modèle numérique de terrain, de dériver des informations de type valeur et orientation des pentes, caractéristiques locales du relief, radiation solaire potentielle, et ceci pour chaque station et/ou, par extension, pour chaque pixel de la zone d'étude. La première étape consiste donc en une régression linéaire multiple sur les paramètres environnementaux (Carrega, 1995) : la variable observée (température), dite variable dépendante, est exprimée en fonction de plusieurs autres variables observées, dites indépendantes. L'équation s'écrit :

$$y(s) = \beta_0 + \beta_1 x_1(s) + \dots + \beta_n x_n(s) + \varepsilon(s)$$

- où $y(s)$ est la variable dépendante, ici la température ;

- $x_1(s)$ à $x_n(s)$ sont les variables indépendantes ou explicatives, décrivant le relief au sens large (notamment altitude, valeur et orientation des pentes, radiation solaire potentielle, plus d'autres paramètres descriptifs des formes locales du relief) ;

- β sont les coefficients ;

- $\varepsilon(s)$ sont les résidus de la régression (écarts au modèle).

Les résidus de cette régression sont conservés et interpolés par krigeage, les tests effectués ayant montré que cette méthode d'interpolation semblait la plus efficace. En pratique, l'estimation réside dans l'application d'une fonction $F(s)$, où $s = (x,y)$ en un point s_p du plan à partir de valeurs connues de F en un certain nombre, m , de points environnants s_i , dont W_i représente la pondération :

- La température maximale en un lieu donné est la température atmosphérique la plus élevée atteinte en ce lieu au cours d'un intervalle de temps prédéterminé. Des intervalles de temps successifs de 24 heures déterminent ainsi en chaque site de mesure la température maximale quotidienne (abréviation : TX), qui survient le plus souvent en cours d'après-midi. Afin d'assurer la synchronisation des observations à l'échelle mondiale, on convient en météorologie que cette dernière grandeur est, pour un jour J, la température la plus élevée qui ait été relevée sur le site entre le jour J à 6 heures UTC et le jour J + 1 à 6 heures UTC » (Météo France, 2003).

$$F(s_p) = \sum_{i=1}^m W_i \cdot F(s_i)$$

La grille de résidus interpolés est ensuite additionnée à celle obtenue par régression, pour obtenir la carte finale de température, au format raster (Lhotellier, 2005). Cette démarche, appliquée ici à la température de l'air, procure également de bons résultats sur les précipitations (Prudhomme, 1999 ; Prudhomme et Reed, 1999).

L'intérêt premier de la méthode consiste donc en l'obtention, aux points de mesure incorporés dans le modèle, des valeurs exactes mesurées au lieu des valeurs estimées (la différence constitue le résidu). Le second avantage est la prise en compte par le krigeage de facteurs d'influence très locale ou au contraire plus régionale non modélisés par la régression multiple. Les deux étapes sont donc bien complémentaires.

3. Prise en compte de l'incertitude

Au sein de ce processus de spatialisation à partir de mesures ponctuelles échantillonnées, de nombreuses sources d'imprécision se juxtaposent pour former une incertitude globale dont les facteurs qui la produisent sont plus ou moins évitables ou remédiables. Dans le cas contraire, cette incertitude doit être pour le moins évaluée.

Nous prendrons ainsi appui sur les diverses étapes du processus et tenterons de présenter, corriger ou évaluer le niveau d'incertitude induit par chacune d'elle. Nous pouvons ainsi, comme Corgne (2004), établir trois grandes distinctions qui sont :

- l'incertitude et l'imprécision liées aux données ;
- l'incertitude liée aux modèles ;
- l'incertitude liée aux résultats.

L'auteur précédemment cité nous rappelle par ailleurs que « le concept d'incertitude est (...) en rapport avec le degré de confiance que l'on affecte à des faits ou un phénomène alors que l'imprécis est directement lié à l'attribut lui-même et à sa valeur » (Corgne, 2004).

Les données sont, par définition, imprécises et incertaines, dans le sens où la mesure n'est jamais qu'une représentation d'une réalité naturelle ou humaine. De plus, même si les données étaient « parfaites », l'utilisateur émet des conclusions qui détiennent forcément un degré d'incertitude, de par son raisonnement puis à travers les résultats. Il existe également en addition une incertitude propagée depuis les données jusqu'aux résultats.

L'estimation de l'incertitude peut être ainsi qualitative (prise de conscience de l'existence de l'incertitude) ou quantitative, par exemple par analyses statistiques (Goovaerts, 2002) ou calcul d'intervalles de confiances.

3.1 Incertitude et imprécision liées aux données

Les données spatiales comprennent deux composantes : une première d'ordre géométrique et une seconde qu'on peut qualifier de thématique. La composante géométrique définit la position et l'extension du phénomène pris en considération, quand la composante thématique contient l'information descriptive (Glemser et Fritsch, 1998). Ces deux termes sont individuellement à l'origine d'incertitudes. En effet, avant même le traitement statistique à proprement parler des données mesurées, apparaissent des risques d'erreur non négligeables.

Les relevés Météo-France sont vérifiés en interne, mais ils n'en comportent pas moins quelques erreurs potentielles : différents types de stations de mesure engendrant des précisions différentes dans les relevés, erreurs de coordonnées de station (figure 1), déménagements de station sans modification des coordonnées dans la base, erreurs de relevés, lacunes, dérives...

Il nous est donc possible de résumer les erreurs et imprécisions de mesure de la sorte :

- les instruments de mesure sont techniquement limités dans le niveau de précision ;
- il existe des erreurs de mesure par la machine ou de relevé par l'homme ;
- il existe des problèmes aléatoires dans le fonctionnement des instruments ;
- parfois, la non-correspondance de données collectées par différentes sources est avérée ;
- les coordonnées des stations sont parfois erronées (en x , y ou z). Notons sur ce dernier point, que les bases de données très récentes sont nettement plus correctes que les bases historiques.

3.2 Incertitude liée aux modèles

Les modèles utilisés pour la spatialisation sont statistiques. L'échantillon est le mode de collecte de toute donnée climatique : cet échantillon n'est pas choisi, il nous est imposé. Il est donc nécessaire, d'une part, que les modèles s'appuient sur un échantillon dont on connaît la représentativité, et d'autre part, que l'utilisateur puisse quantifier l'incertitude produite par eux.

Les données climatiques se caractérisent par une forte variabilité dans le temps et l'espace. De plus, en zone de montagne, les questions d'échelle et de densité des réseaux météorologiques doivent être, plus que dans n'importe quel autre milieu, mis en exergue (Choisnel et Seguin, 1987). La représentativité spatiale d'une station unique et d'un ensemble de postes de mesures nécessite en effet, plus qu'en plaine, d'être évaluée : si un département a pour altitude la plus fréquente 1000 mètres, il sera évidemment mieux décrit par des relevés

effectués à cette altitude plutôt que par une station plus basse ou plus élevée.

L'importance de la représentativité nécessaire de l'échantillon doit donc être examinée avant toute tentative de modélisation ou d'interpolation (Dumolard, 2006). Différentes méthodes permettent d'estimer cette incertitude liée à l'échantillonnage. L'une d'entre elle repose sur le rapport des fréquences simples des altitudes des stations de mesure sur celles du modèle numérique de terrain : pour un intervalle d'altitude donné, un rapport $F_{éch}/F_{MNT}$ égal à 1 sera synonyme d'une bonne représentativité, quand une valeur proche de zéro traduira un échantillonnage très peu représentatif (sous-représentation). On obtient un indice de validité altitudinale du réseau qu'il est possible de cartographier.

3.3 Incertitude liée aux résultats

Diverses méthodes existent pour évaluer l'incertitude des résultats. La validation croisée est considérée comme une méthode « objective » pour mesurer la qualité d'une interpolation. Elle possède l'avantage d'être automatique dans la plupart des logiciels. Le second avantage est l'exécution de la validation sur le nombre maximal de valeurs disponibles.

Le processus est relativement simple : au sein d'un échantillon comportant n individus, on retire à tour de rôle chaque individu pour obtenir en ses coordonnées une estimation réalisée à partir des $n-1$ individus restant. Ainsi, le logiciel retire la première observation O_1 , estime la valeur E_1 au point P_1 avec les $(n-1)$ valeurs suivantes et effectue la différence entre mesure et estimation, générant l'erreur d'estimation en ce point, qui s'écrit :

$$\varepsilon_1 = E_1 - O_1$$

L'opération se répète n fois (le logiciel retire la seconde observation O_2 et estime la valeur E_2 en ce point P_2 grâce aux $(n-1)$ autres valeurs). On obtient ainsi n estimations et n erreurs aux n points de mesure.

Parmi les autres méthodes de validation, il en existe une à la fois plus simple et plus contraignante : elle repose sur une séparation manuelle de l'échantillon de mesure en deux sous-échantillons. Le premier est alors utilisé pour la construction du modèle, et le second, non intégré dans la modélisation, pour la validation. La méthode nécessite un nombre de valeurs important qui permette d'en évincer certaines du modèle : c'est une contrainte majeure de ce type de validation, par ailleurs sévère car n'ayant pas recours à des données incluses dans la construction du modèle. En pratique, on spatiale les températures d'un échantillon n_1 puis on compare aux coordonnées des points d'un échantillon n_2 les températures estimées avec les valeurs réellement mesurées, cet échantillon n_2 n'ayant pas servi pour la mise en place du modèle.

4. Résultats

4.1 Incertitude et imprécision liées aux données

La figure 1 nous propose ainsi la comparaison des altitudes des stations fournies par Météo-France avec celles extraites du modèle numérique de terrain, à résolution 150 mètres. Le nuage de point suit une forme très nettement linéaire, mais on constate quelques individus très éloignés de la tendance générale : quelques stations connaissent une différence de plusieurs centaines de mètres entre les deux altitudes comparées. Cette différence peut provenir de deux éléments :

- soit une (ou plusieurs) des coordonnées de la base Météo-France est (sont) fausse(s), ce qui peut être éventuellement corrigé, par exemple en relevant par GPS les coordonnées exactes ;
- soit la station de mesure est située dans une zone à variation de relief très marquée, et dans ce cas il n'est pas impossible que l'altitude donnée par le pixel de la couche raster soit éloignée de celle donnée par les coordonnées précises de la station.

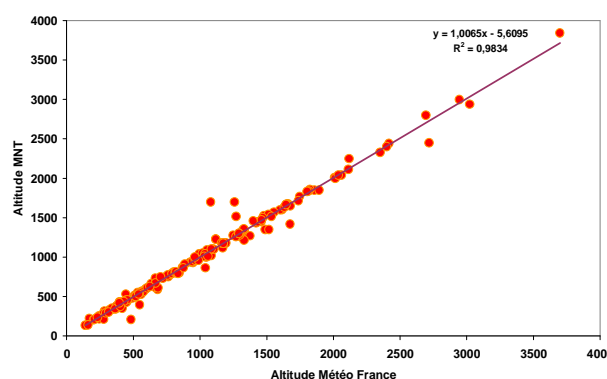


Figure 1. Comparaison des altitudes des stations fournies par Météo-France avec celles extraites du modèle numérique de terrain.

Les coordonnées fournies (au format degré minute seconde) sont transformées pour être compatibles avec le modèle numérique de terrain utilisé (Lambert 2 étendu) au sein d'un système d'information géographique, et les tests effectués ont montré que les logiciels utilisaient des algorithmes de conversion différents (figure 2). On constate ainsi des différences de plusieurs dizaines de mètres pour un même jeu de coordonnées suivant qu'on utilise ArcInfo® (Esri), Circe® (Institut Géographique National) ou Geodus® (France Telecom, CNET). Ainsi, à titre d'exemple, on mesure entre les calculs réalisés sous ArcInfo et Circe, en moyenne, 40 mètres de décalage en longitude et 109 mètres en latitude. Il est alors difficile de considérer qu'un résultat est meilleur que l'autre. Il faut s'obliger à utiliser toujours le même algorithme pour ne pas

multiplier des erreurs supplémentaires, et pour que toutes les coordonnées soient extraites sur une base de référence commune.

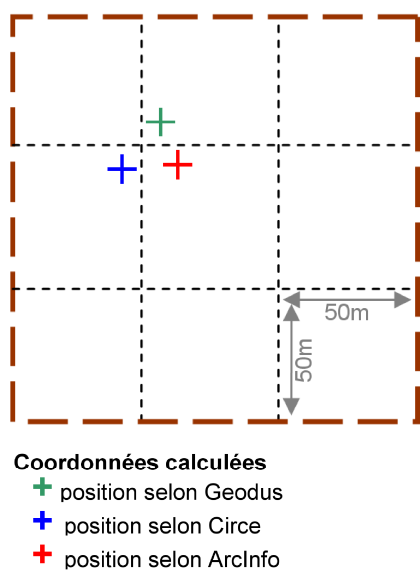


Figure 2. Coordonnées obtenues pour une station d'après trois logiciels de conversion de coordonnées.

Une incertitude spatiale naît par ailleurs de l'affectation d'une station, donnée ponctuelle par définition (car définie par deux coordonnées dans un plan), à un pixel, donnée surfacique, variant de 2500 m² à 22500 m² selon la résolution spatiale du modèle numérique utilisé, puisque le travail est effectué en mode image. Si l'on ajoute les incertitudes dues à la conversion des coordonnées, une même station peut finalement être décrite par trois pixels différents. Il faut en être conscient lorsqu'on travaille à échelle fine, plus particulièrement en relief contrasté : un même pixel peut en effet recouvrir, résumer, voire faire disparaître des réalités ponctuelles minoritaires. Une même station peut ainsi être décrite selon la conversion réalisée par deux pixels différents... sur deux versants opposés.

La figure 3 propose, pour la variable altitude uniquement, la représentativité de l'échantillon par rapport à l'intégralité des altitudes de nos quatre départements d'étude. Le graphique présente les fréquences cumulées des altitudes des stations de mesure, comparées à celles du modèle numérique de terrain, qui constitue notre référence (même si, bien sûr, ce dernier est aussi à l'origine d'incertitudes). Le constat est sans appel : la représentativité de l'échantillon est excellente jusqu'à 500 mètres environ, puis elle se dégrade constamment jusqu'à 2000 mètres, qui est approximativement la limite de validité des modèles fondés sur des mesures en domaine alpin, cette

limite étant celle des mesures denses. Au-delà, quelques rares stations fournissent des données, malheureusement souvent lacunaires car difficiles d'accès (conditions climatiques difficiles, réparation non immédiate), donc non intégrables au modèle. Les stations sont majoritairement installées en vallée pour des raisons évidentes de commodité, d'accessibilité, de coût de collecte et de maintenance.

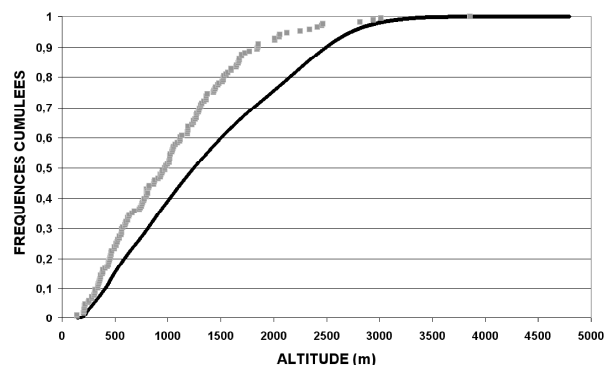


Figure 3. Comparaison des fréquences cumulées des altitudes des stations de mesure ($F_{éch}$ en gris) avec celles du modèle numérique de terrain (F_{MNT} en noir).

Cette représentativité de l'échantillon varie d'ailleurs avec la variable considérée et l'influence du relief sur elle. Lors de la modélisation par régression, il est indispensable que l'échantillon traité (à savoir les coordonnées des stations de mesure) soit représentatif du terrain d'étude : cette représentativité n'est pas toujours évidente, le graphique le montre, notamment aux hautes altitudes. En effet, la fréquence des postes de mesure diminue avec l'altitude, mais dans les vallées, la densité spatiale des stations se révèle semblable à celle des régions de plaine. Ainsi, l'indice $F_{éch}/F_{MNT}$ est élevé (bonne représentativité) dans les plaines et vallées, notamment vers 400/500 mètres, altitude la plus fréquente observée sur le réseau. Il est très faiblement inférieur dans le fond des basses vallées (200 mètres et moins), mais diminue surtout lorsque l'altitude s'élève.

La représentativité planimétrique (en longitude et latitude, figure 4) peut elle être illustrée par le recours aux polygones de Thiessen autour des stations du réseau, en assignant des valeurs de meilleure qualité aux surfaces les plus réduites, puisqu'à priori plus on s'éloigne d'une station, plus la modélisation devient aléatoire. Pour obtenir des valeurs continues (figure 5), il est également possible de se fonder sur l'écart type du krigeage (conservé lors de l'interpolation des résidus), ou encore faire appel à une modélisation 3D (Dumolard, 2006).

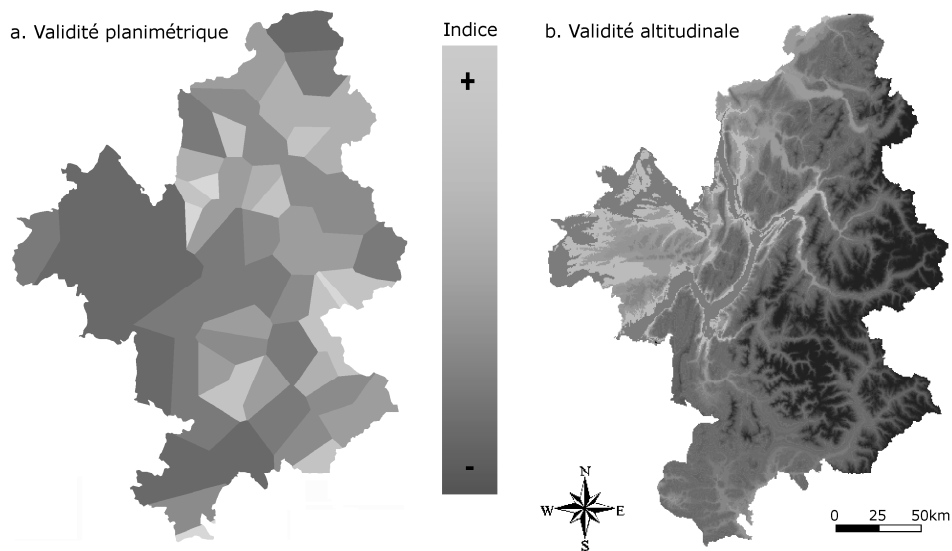


Figure 4. Indices de validité du champ de température. Les valeurs varient du gris clair (indice fort, validité bonne) au gris foncé (indice faible, validité incertaine)

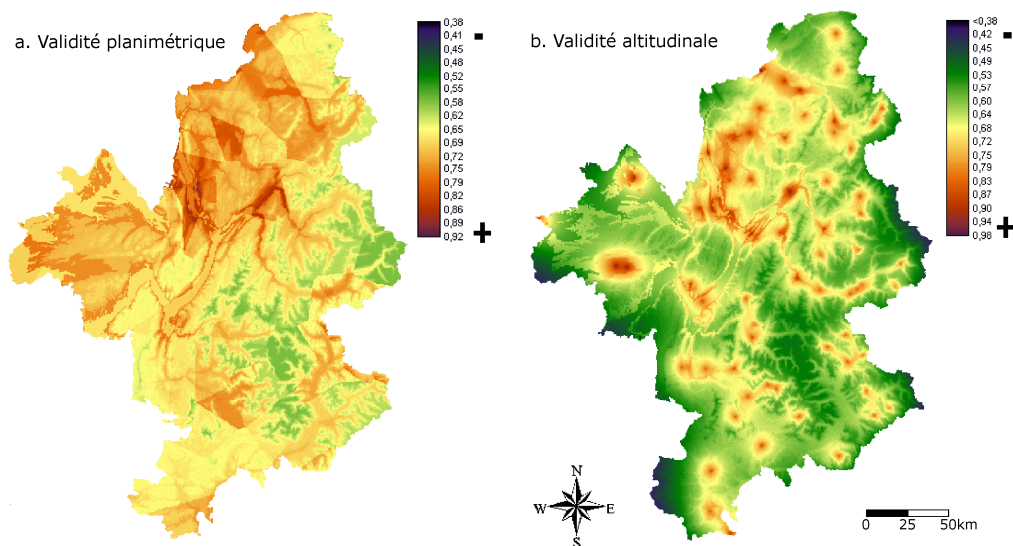


Figure 5. Validité des températures spatialisées. Construction par polygones de Thiessen et rapport des altitudes réseau/terrain (à gauche). Construction par écart-type du krigeage et rapport des altitudes réseau/terrain (à droite).

Ce type de carte est dépendant du nombre de stations incluses dans le modèle d'estimation. Ici, seules les stations à données complètes sur la période étant intégrées, l'indice de validité prend des valeurs plus faibles qu'avec l'inclusion du réseau de mesure complet. Les deux cartes de la figure 5 proposent des similitudes :

- les vallées se distinguent nettement par la conjonction d'une altitude relativement basse et d'un réseau dense ;
- les massifs montagneux sont peu dotés en station donc mal notés ;
- la région de l'Avant-Pays Savoyard fait figure de privilégiée.

Cependant, la première carte présente un plus grand nombre de valeurs moyennes quand la seconde a tendance à focaliser sur les extrêmes.

L'intérêt d'une telle carte, donnant un indice de validité de la température spatialisée, réside dans l'évaluation du niveau de confiance à accorder à la température calculée en différents points de la carte. Cette confiance sera d'autant plus forte que le point étudié se situe à proximité d'un point de mesure ou/et dans une tranche d'altitude fortement représentée par le réseau. La conjonction des deux facteurs est bien sûr préférable pour plus de sûreté dans l'affirmation.

4.2 Incertitude liée aux modèles et leurs résultats

Sur les modèles statistiques en eux-mêmes, la méthode choisie pour la spatialisation est celle qui laisse le moins de place à l'incertitude quant aux résultats. Ainsi, si on compare trois méthodes de spatialisation pour une date donnée (à échelle quotidienne), celle qui permet d'expliquer le plus grand pourcentage de variance des températures est dans une très grande majorité des cas la technique par régression multiple pas à pas puis interpolation par krigeage des résidus. En relief très contrasté, les techniques d'interpolation spatiales se révèlent en effet inefficaces. Par contre, l'ajout des résidus interpolés au modèle régressif permet souvent d'améliorer quelque peu les estimations. À titre d'exemple, des tests sur la température du 25 décembre 1993, réalisés sur deux échantillons tests, ont donné les résultats suivants :

- l'interpolation par krigeage directement sur les valeurs de température procure un coefficient de détermination entre observation et mesure de seulement 0,19 ;
- la régression multiple pas à pas offre un r^2 de 0,88 ;
- la régression multiple pas à pas complétée par l'interpolation des résidus (méthode connue sous le

nom de « *residual kriging* ») permet un r^2 de 0,89, soit le meilleur résultat.

Le choix de la méthode conditionne donc fortement l'incertitude liée aux résultats.

L'incertitude des résultats est également nettement corrélée à l'échelle temporelle d'étude (figure 6). Les valeurs des coefficients de détermination obtenus sur des moyennes annuelles de températures minimales et maximales sont tout à fait acceptables (figure 6a). On observe toujours plus de 75% de variance expliquée sur les minima : 76% pour 1990 avec seulement deux variables explicatives retenues, l'altitude et la pente mais 86% pour 1995 avec deux paramètres supplémentaires : la longitude et la radiation potentielle. Quant aux maxima, ils permettent d'atteindre sur les six années étudiées des valeurs toujours supérieures à 90%, avec une très grande régularité dans les résultats. À l'échelle décadaire (figure 6b), les coefficients de détermination sont la plupart du temps supérieurs à 70% sur les minima, à 80% sur les maxima, mais les valeurs s'effondrent en hiver. Le type de circulation atmosphérique et le type de temps qui lui est associé sont à l'origine des variations soudaines de pourcentage de variance expliquée par le modèle de régression multiple.

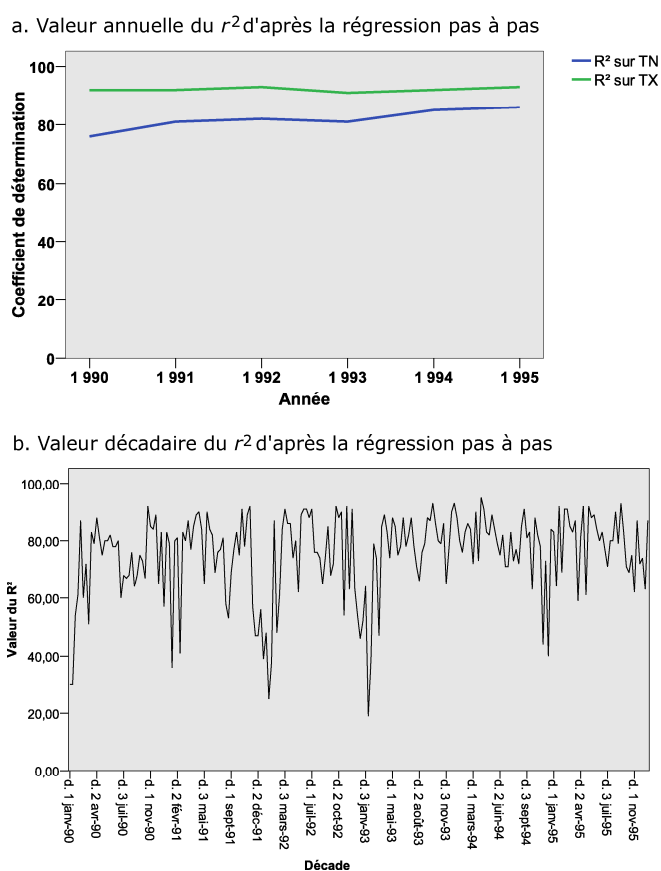


Figure 6. Comparaison de la qualité des estimations selon l'échelle de temps étudiée : température minimale et maximale moyenne annuelle entre 1990 et 1995 (a), température minimale moyenne par décennie entre 1990 et 1995 (b)

Outre l'incertitude fondée sur l'échantillonnage spatial, il existe donc une forte diversité de qualité des résultats causée par le temps, à la fois fonction de l'échelle temporelle (température extrême quotidienne, température moyenne mensuelle, température moyenne annuelle) et de la succession de situations météorologiques diverses. Par exemple, à l'échelle quotidienne, les températures hivernales sont souvent, pour ne pas dire toujours, les plus difficiles à cartographier, principalement à cause des surfaces persistantes de hautes pressions sur les Alpes et des inversions thermiques qu'elles engendrent. Il faut donc dans certains cas améliorer les modèles, en ayant recours à des variables topographiques plus fines, en modélisant par exemple les dépressions dans lesquelles l'air froid pourra stagner.

Il est aussi possible d'effectuer les calculs non plus sur les valeurs des pixels exacts de positionnement des stations, mais sur des variables moyennées à l'intérieur de filtres de tailles diverses. Ainsi, la comparaison de construction d'un modèle de régression pas à pas sur les valeurs des variables décrites aux coordonnées des

stations et sur des grilles filtrées a été menée aux échelles mensuelle et quotidienne, en prenant toujours soin de séparer les maxima des minima. Certains jours d'hiver, le coefficient de détermination a pu être multiplié par deux.

Les méthodes statistiques appliquées sur les variables estimatives diffèrent donc selon l'échelle temporelle visée, et l'ajout d'informations supplémentaires (ACP, filtrage des grilles, influence du type de temps...) permet, quand la régression n'est pas suffisante, d'apporter des améliorations significatives dans les résultats.

Les méthodes de validation procurent enfin une illustration supplémentaire de l'incertitude de la spatialisation. La figure 7 présente ainsi les erreurs d'estimation obtenues par validation croisée lors de l'étape du krigeage des résidus de la régression sur la température minimale moyenne de décembre 1995. La qualité de l'interpolation peut ainsi être appréciée ; il n'est pas contre pas toujours aisé de commenter géographiquement les différences observées.

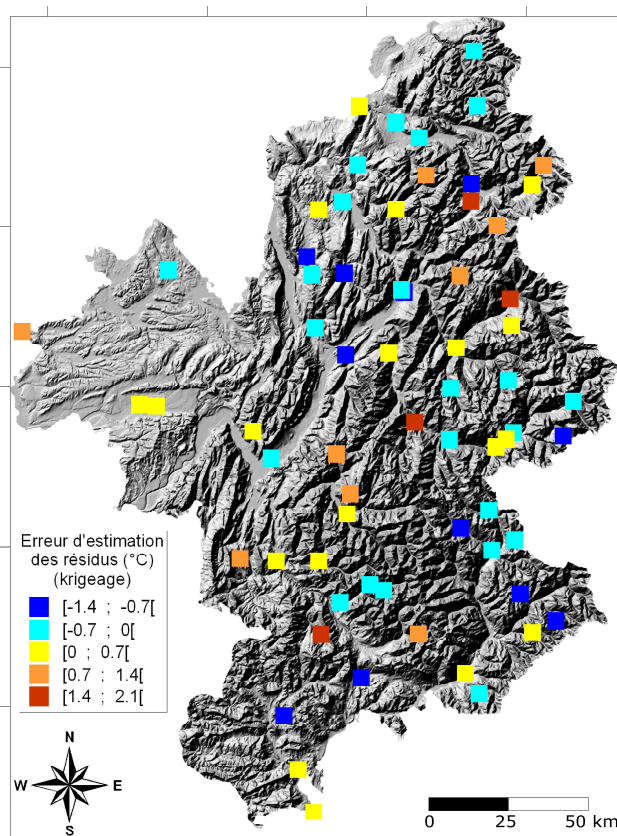


Figure 7. Exemple de cartographie d'erreur d'estimation : erreur d'estimation calculée par validation croisée sur les résidus de la régression multiple concernant la température minimale de décembre 1995.

La figure 8 propose quant à elle une illustration de validation par la deuxième méthode, dite des échantillons-tests, avec calcul de l'erreur d'estimation en un certain nombre de points dont les données n'ont pas été incluses dans la construction du modèle. Pour cette

journée du 6 décembre 1995, les erreurs d'estimation de la température minimale comprises entre -1°C et $+1^{\circ}\text{C}$ sont majoritaires. Les faibles erreurs d'estimation sont donc prépondérantes, mais des erreurs de 5 degrés sont ponctuellement atteintes.

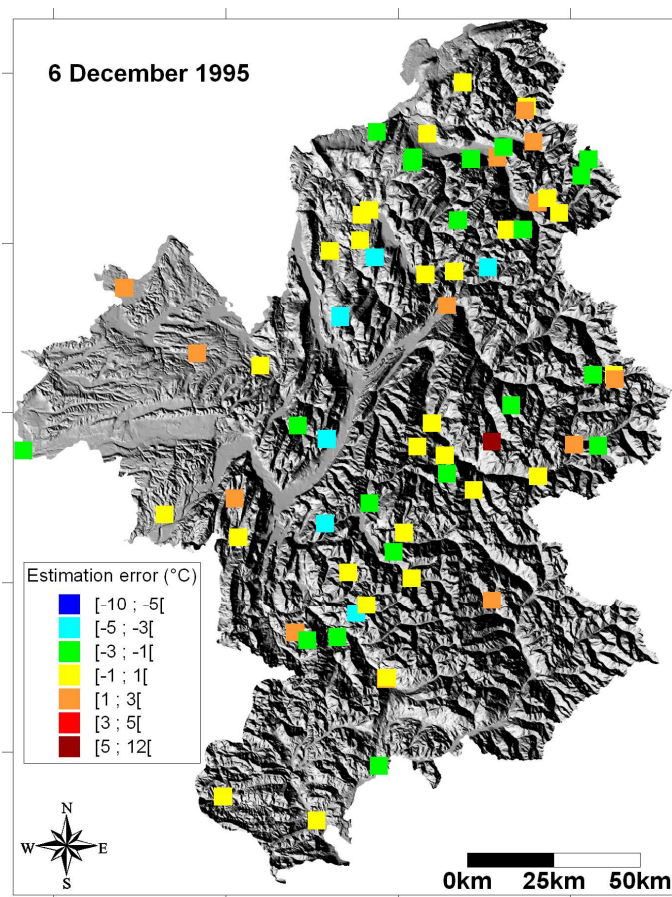


Figure 8. Exemple de cartographie d'erreur d'estimation : erreur d'estimation sur la température minimale du 6 décembre 1995.

5. Conclusion

L'incertitude provient donc essentiellement de trois éléments : les données à disposition, les modèles utilisés, les résultats obtenus. À cela vient s'ajouter la propagation de l'incertitude initiale : par exemple, lorsqu'on utilise les données de température dont on sait qu'elles sont nécessairement imprécises, on propage, sur les résultats, cette imprécision. Il apparaît donc nécessaire de, sinon quantifier, au minimum qualifier les grands types d'imprécision. L'échelle géographique d'étude paraît être un moyen simple d'effectuer une classification. Ainsi, pour un travail à échelle fine sur un secteur d'étude limité (une vallée), l'incertitude et les imprécisions inhérentes aux données auront tendance à supplanter celles liées aux modèles et aux résultats. À une échelle plus vaste, celle du massif alpin, les erreurs dans les données initiales vont plutôt s'effacer devant les incertitudes relevant de la modélisation statistique.

Des modèles peu complexes de régression multiple offrent d'excellents résultats quand ils sont construits sur des valeurs moyennes mensuelles ou annuelles (Lhotellier, 2004), ou encore sur des valeurs minimales ou maximales quotidiennes sous conditions perturbées. Dans les autres cas de figure, le cas extrême étant une

journée d'hiver en situation anticyclonique, la nécessité d'utiliser d'autres étapes (fenêtres mobiles par exemple) s'impose et permet d'améliorer significativement les résultats obtenus (Joly *et al.*, 1995).

Un modèle n'étant, en géographie comme dans toute autre science, jamais parfait, il reste des résidus dont il convient d'exploiter les variations spatiales, sous peine d'omettre une partie de l'information climatique disponible. Leur interpolation par krigeage semble constituer l'une des meilleures alternatives, pour peu que la densité du réseau le permette et que les résidus possèdent une structure spatiale et statistique convenable.

Tout au long de ce processus permettant d'aboutir à une carte de champs de températures, l'utilisateur ne doit omettre de quantifier les incertitudes, pour éventuellement les cartographier. En effet, la validité statistique de la carte doit autoriser sa réalisation : il est tout à fait possible de cartographier la température par un modèle qui explique 5% des variations mesurées (coefficient de détermination de 0,05), mais quel est alors l'intérêt d'une telle carte ?

Le niveau d'incertitude, mesuré à partir de l'échantillon spatial, d'un coefficient de détermination, d'une validation croisée ou encore plus simplement en comparant estimation et mesure, est donc l'élément

essentiel qui doit permettre d'accepter de proposer une cartographie d'un phénomène ou non à un utilisateur éventuel. Il devrait accompagner toute carte thématique, notamment dans le cadre de l'aide à la décision.

Ainsi, une intervention de lutte contre le gel (en agriculture ou dans le domaine routier) pourra être en priorité portée sur les espaces dépassant un seuil thermique et où l'incertitude est la plus faible.

6. Références bibliographiques

- Carrega P., 1995, A method for the reconstruction of mountain air temperatures with automatic cartographic applications, *Theoretical and Applied Climatology*, 52, 69-84.
- Chapman L., Thornes J. E., 2003, The use of Geographical Information Systems in climatology and meteorology, *Progress in Physical Geography*, 27(3), 313-330.
- Choisnel E., Seguin B., 1987, Problème d'échelle et réseaux météorologiques en zone de moyenne montagne, *Agrométéorologie des régions de moyenne montagne*, Colloque de l'INRA, 39, 37-49.
- Corgne S., 2004, Modélisation prédictive de l'occupation des sols en contexte agricole intensif : application à la couverture hivernale des sols en Bretagne, Thèse de doctorat, Université Haute-Bretagne, Rennes.
- Dumolard P., 2006, Uncertainty from spatial sampling: a case study in the French Alps, in Dobesch H., Dumolard P., Dyras I. (Ed.), *Spatial interpolation for climate data: the use of GIS in climatology and meteorology*, ISTE, Hermes, London.
- Dyras I. et al., 2005, The use of Geographical Information Systems in climatology and meteorology: COST719, *Meteorological Applications*, 12(1), 1-6.
- Glemser M., Fritsch D., 1998, Data uncertainty in a hybrid GIS, *Symposium on GIS – Between visions and applications*, IAPRS, 32-4., ISPRS Commission IV, Stuttgart.
- Goovaerts P., 2002, Geostatistical modelling of spatial uncertainty using p-field simulation with conditional probability fields, *International Journal Geographical Information Science*, 16(2), 167-178.
- Joly D., Bert H., Fury R., Vermot-Desroches B., Jaquinot J.-P., 1994, Interpolation des températures à grande échelle, *Revue Internationale de Géomatique*, 4(1), 55-86.
- 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*, 23 (3), 1637-1654.
- Lhotellier R., 2004, Spatial interpolation of daily minimum and maximum temperatures in the French Alps, *Proceedings of the Conference on the Spatial Interpolation Techniques in Climatology and Meteorology*, HMS, Budapest, 209-216.
- Lhotellier R., 2005, Spatialisation des températures en zone de montagne alpine, Thèse de doctorat, Université Joseph Fourier, Grenoble.
- Météo-France, 2003, *Glossaire Météo-France*, disponible gratuitement en ligne à l'adresse suivante: <http://www.meteofrance.com/FR/glossaire/index.jsp>.
- Prudhomme C., 1999, Mapping a statistic of extreme rainfall in a mountainous region, *Physics and Chemistry of the Earth, Part B: Hydrology, Oceans and Atmosphere*, 24(1-2), 79-84.
- Prudhomme C., Reed D. W., 1999, Mapping extreme rainfall in a mountainous region using geostatistical techniques: a case study in Scotland, *International Journal of Climatology*, 19(12), 1337-1356.
- Tveito O. E., Bjordal I., Skjelvag O., Aune B., 2005, A GIS-based agro-ecological decision system based on gridded climatology, *Meteorological Applications*, 12(1), 57-68.