

## SYSTEMES D'INFORMATION GEOGRAPHIQUE ET TRAITEMENT DE DONNEES CLIMATIQUES : L'EXEMPLE DE L'INTERPOLATION DES TEMPERATURES

Rémi Lhotellier<sup>1</sup>

**Mots clefs:** SIG, krigage des résidus, température, prédicteurs, validation

### RESUMÉ

Le potentiel des systèmes d'information géographique (SIG) est désormais célèbre dans de nombreux domaines de recherche. En climatologie, il demeure souvent encore sous valorisé. En effet, il est généralement difficile de recourir à une information climatique car celle-ci est ponctuelle et doit être validée dans le temps (les normales climatiques sont ainsi calculées sur des durées de 30 ans minimum). D'autre part, le problème de validité des mesures (postes manuels relevés par des bénévoles, lacunes dans la base, dérives) est constant et doit bien sûr être abordé avant toute application plus poussée. Sans information fiable à la base, aucun traitement scientifique ne peut être réalisé. Cet article présente une démarche ayant recours à des traitements statistiques sur des bases de données climatiques combinés à des outils de type SIG. Nous évoquerons ici des questions liées à l'interpolation des températures, en précisant la démarche suivie pour estimer spatialement ces données, la nécessaire étape de validation des résultats, et deux brefs exemples de traitement de variables climatiques dans un logiciel de systèmes d'information à référence spatiale. La région d'étude se situe dans la partie ouest de la chaîne alpine, en France.

**1. Méthodologie retenue pour l'interpolation des températures** Interpoler, c'est intercaler des valeurs calculées dans une série de valeurs existantes et connues. Par extension, interpoler des températures consiste à réaliser, à partir de données mesurées ponctuellement dans les stations de mesure, des champs de températures, de manière à connaître une température estimée dans chaque point de l'espace étudié.

Les variations de la température, en relief contrasté, sont complexes. La corrélation des températures relevées avec l'altitude reste la principale caractéristique guidant leur répartition. Néanmoins, cette relation n'est pas constante dans le temps comme dans l'espace et elle ne suffit donc pas pour effectuer une cartographie valide des champs de température.

La méthode retenue consiste à dériver d'un modèle numérique de terrain à échelle fine (pixel de 150 mètres de côté dans notre exemple) des paramètres topographiques expliquant partiellement les variations des températures relevées. Les méthodes d'interpolation purement spatiales (krigeage ordinaire par exemple) ne proposent pas de résultats satisfaisants. La définition des facteurs de variabilité spatiale des données climatiques est donc une étape essentielle, ceux-ci doivent être définis sur des échelles emboîtées dans l'espace comme dans le temps : ainsi, les paramètres locaux n'auront pas la même influence sur des valeurs extrêmes quotidiennes ou sur des moyennes annuelles. Les résultats statistiques sont pour ces raisons nettement meilleurs en utilisant moins de variables explicatives sur des échelles de temps longues (moyenne annuelle notamment).

---

<sup>1</sup> SEIGAD, Université Joseph Fourier, 2061 rue de la piscine, BP53X, 38041 Grenoble CEDEX 9, France  
fax: (33)(0)4 76 51 45 96, e-mail : remi.lhotellier@laposte.net

L'utilisation des méthodes géostatistiques n'est donc pas valable en relief contrasté, ou plutôt elle doit être précédée par la création d'un modèle statistique opératoire. La prise en compte de paramètres environnementaux, dans des échelles d'applications spatiales comme temporelles demeure donc essentielle : des variables représentatives des processus physiques de terrain permettent la mise en place d'un modèle de régression multiple, éventuellement corrigé par un paramètre à évolution spatiale et temporelle d'une toute autre échelle : les types de temps [3].

La première étape consiste donc en une régression linéaire multiple sur les paramètres environnementaux : la variable observée (température), dite variable dépendante, est exprimée en fonction de plusieurs autres variables observées, dites indépendantes. L'équation s'écrit :

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

avec  $y(s)$  : variable dépendante  
 $x_1(s)$  à  $x_n(s)$  : variables indépendantes ou explicatives  
 $\beta$  : coefficients  
 $\varepsilon(s)$  : résidus de la régression (écarts au modèle).

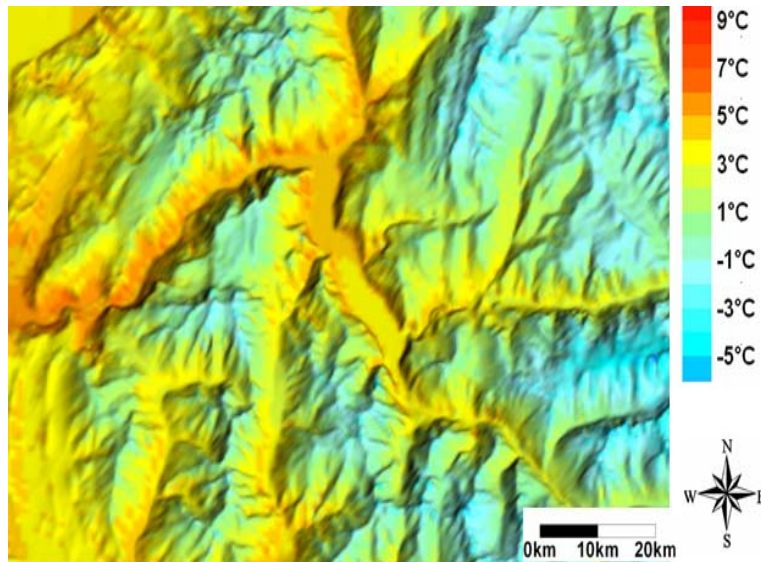
Dans notre modèle, nous avons sélectionné 13 variables explicatives (altitude, latitude, longitude, pente, orientation selon deux composantes, concavité / convexité du relief, convexité horizontale, convexité transversale, radiation potentielle calculée aux 21/12, 21/01, 21/02 et 21/03). Bien sûr, tous les prédicteurs n'entrent pas quotidiennement dans le modèle. Plus l'échelle temporelle s'allonge, plus le nombre de variables explicatives diminue en moyenne. Le premier prédicteur en terme de fréquence d'occurrence et de coefficient de corrélation est évidemment l'altitude. Les autres variables explicatives oscillent selon la saison, le type de temps, ou le type d'extrême (minimum ou maximum) étudié.

L'équation de régression ainsi définie est appliquée dans un logiciel de type SIG sur les matrices représentant les variables de terrain, permettant une cartographie des températures estimées par le modèle (estimation notée  $y'$ ).

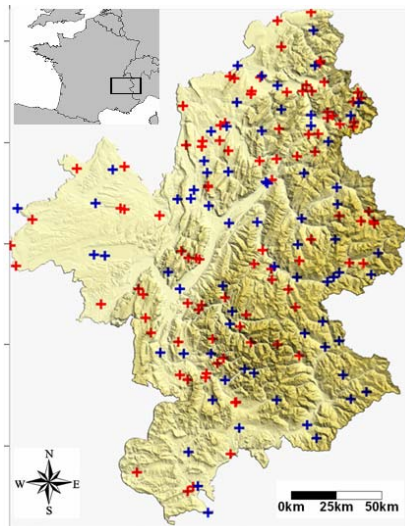
Les résidus  $\varepsilon(s)$  du modèle de régression multiple calculés en chaque site, alors considérés comme variables aléatoires suivant le modèle intrinsèque, sont ensuite interpolés par krigeage de manière à obtenir un champ de résidus. La carte qui en est déduite est additionnée à celle précédemment calculée à partir des  $y'$ . Ainsi, une troisième cartographie, qui contient à la fois le modèle de régression et les résidus interpolés, est construite. Celle-ci prend donc en compte à la fois les facteurs environnementaux, locaux, liés au terrain et ceux non intrinsèques au modèle, qui sont spatialisés par krigeage. Ces derniers, même s'ils ne sont pas expliqués par le modèle de régression, apparaissent donc dans la carte finale.

La prise en compte du type de temps permet d'améliorer significativement les résultats. Pour exemple, la figure 1 propose un résumé schématique de la méthode d'interpolation et la carte 1 présente l'interpolation des températures minimales sur une portion des Alpes Française à la date du 26 décembre 1995. La température minimale est plus faible dans le fond de vallée (jaune, orange pâle) que sur certains versants (orange plus prononcé). Au dessus d'une certaine altitude, l'inversion thermique disparaît et la température recommence à décroître avec l'altitude.

Le modèle est réalisé à partir des données acquises dans les stations climatiques disposant de données complètes sur la période d'étude 1990-1995 (la carte 2 propose leur représentation cartographique). Sur cette carte 2, l'altitude varie entre 140 mètres, à l'ouest, et près de 5000 mètres à l'est (4810 mètres au Mont Blanc).



**Carte 1** : Agrandissement d'une zone propice aux inversions thermiques : les versants (orange) apparaissent plus chauds que les fonds de vallée sur la carte des températures minimales du



+ en bleu, stations météorologiques disposant de l'ensemble des données quotidiennes pour la période 1990-1995, utilisées pour la mise au point du modèle

+ en rouge, stations météorologiques ne disposant pas de l'ensemble des données quotidiennes pour la période 1990-1995, utilisées pour la validation à l'échelle quotidienne

**Carte 2** : Stations de mesure superposées à la zone d'étude, qui s'étend du Lac Léman au Nord à la Durance au Sud, et de la vallée du Rhône à l'Ouest à la frontière italienne à l'Est.

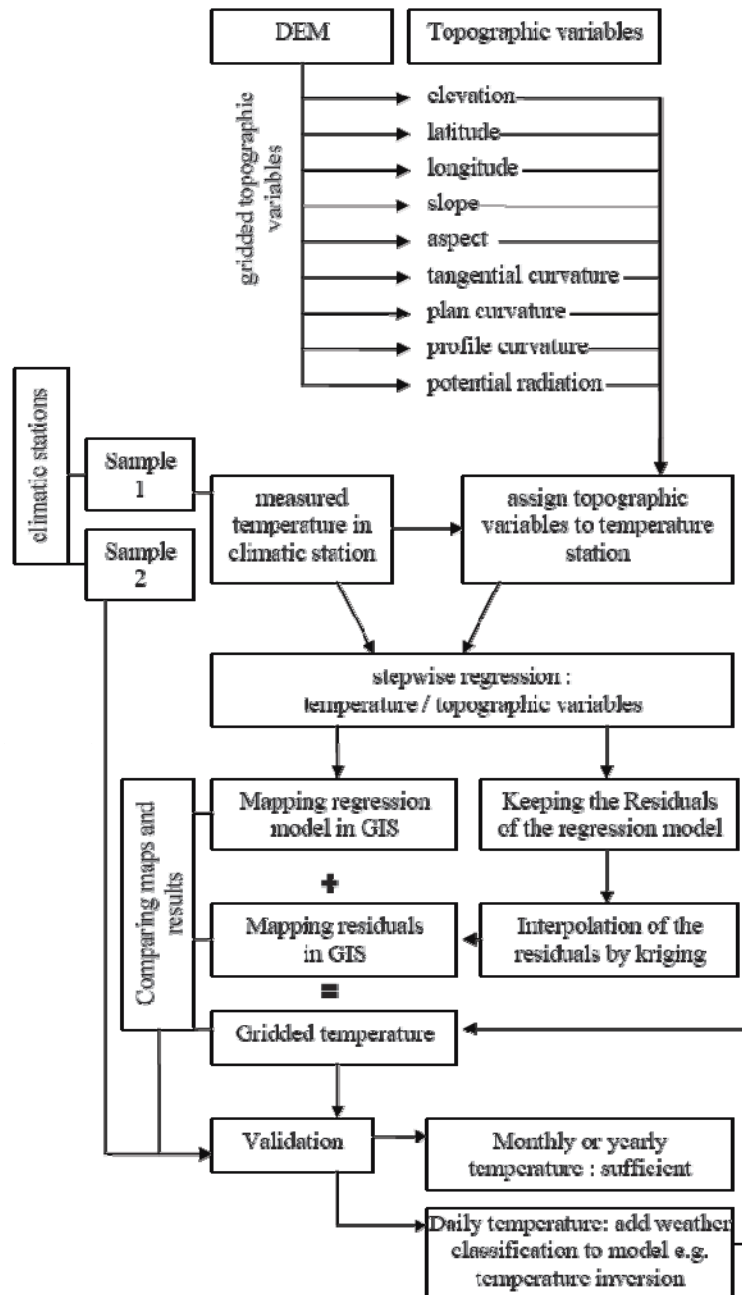
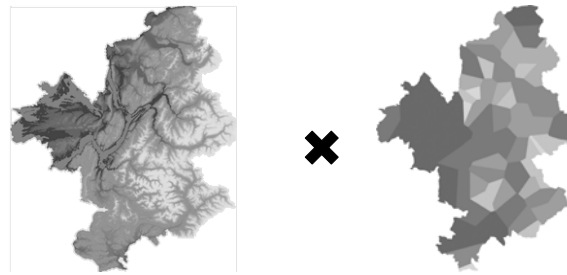


Figure 1 : Démarche retenue pour l'interpolation des températures.

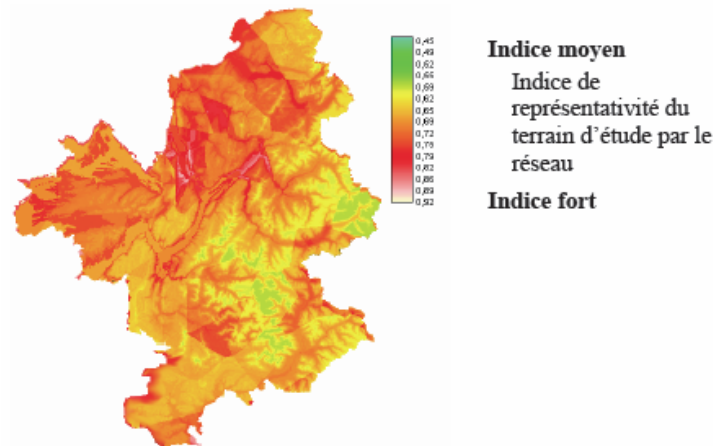
## 2. Validation des résultats

**2.1. La question de la représentativité du réseau** La validation des résultats obtenus est bien évidemment une étape essentielle post-interpolation. Cependant, il nous paraît également tout aussi important de tester la représentativité de l'échantillon de stations de mesures utilisé par rapport au terrain d'étude. Pour se faire, deux approches ont été combinées : une représentativité altitudinale des stations ainsi qu'une représentativité spatiale ont été définies. Pour la première, le rapport entre fréquences cumulées des altitudes des stations et celles du modèle numérique de terrain est calculé puis cartographié. La densité relative des stations en altitude fait ainsi apparaître un évident déséquilibre entre plaines et massifs montagneux, au bénéfice des premières (carte 5).

**Error!**



**Cartes 5 & 6** : Carte de représentativité altitudinale du réseau de mesure (à gauche) et carte de représentativité spatiale du réseau de mesure (à droite).



**Carte 7** : Combinaison des cartes de représentativité du réseau de postes climatiques sur les trois dimensions spatiales x, y et z. Plus l'indice est élevé, plus la qualité de représentation spatiale du terrain est bonne et donc plus l'interpolation doit être juste.

Pour des raisons économiques et humaines, les stations de mesure des paramètres climatiques sont en effet principalement situées dans des zones accessibles, donc en plaine et vallées. Cependant, on peut noter que les altitudes les mieux pourvues en stations ne sont pas les plus basses (l'altitude minimale est de 140 mètres, quand le maximum de densité des stations est atteint vers 400/500 mètres). Là encore, il s'agit

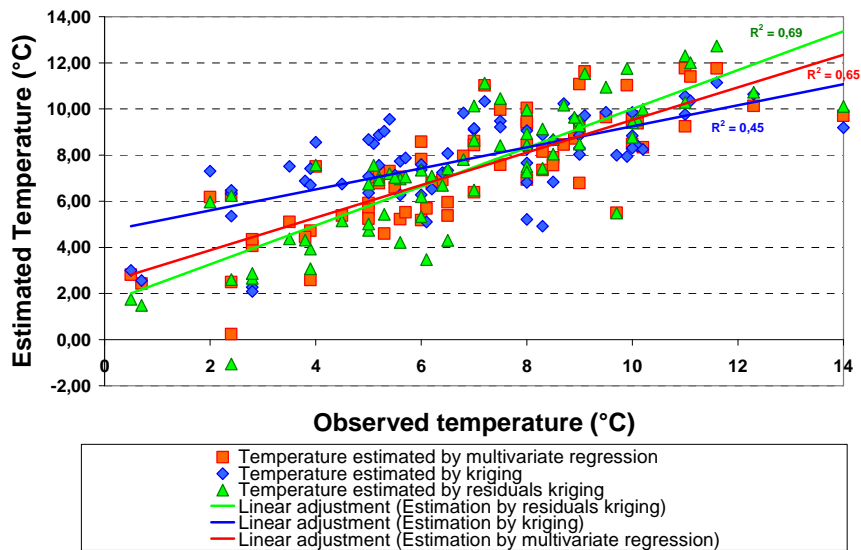
d'une corrélation avec les bassins de développement de la région. En terme de densité spatiale, la réalisation de polygones de Voronoï centrés sur les stations climatiques a permis la création de surfaces inversement proportionnelles à la densité des points de mesure (carte 6). On peut en effet supposer que l'influence d'une mesure diminue avec l'éloignement de celle-ci.

Un coefficient a ensuite été assigné aux valeurs sur les deux cartes de manière à pouvoir les multiplier et prendre en compte les valeurs relatives en x, y et z (carte 7).

**2.2. Validation statistique** Il est indispensable de valider les résultats obtenus : la technique la plus simple (et probablement aussi la plus efficace) repose sur la comparaison de valeurs mesurées dans un échantillon de stations de mesure qui n'a pas servi à la construction du modèle avec les valeurs estimées par celui-ci.

Une journée pour laquelle les résultats obtenus par le modèle ne sont pas très bons a été sélectionnée (figure 2). Le temps est fortement anticyclonique sur la France (1030 hPa), cette situation entraîne la formation d'inversions thermiques dans les vallées alpines.

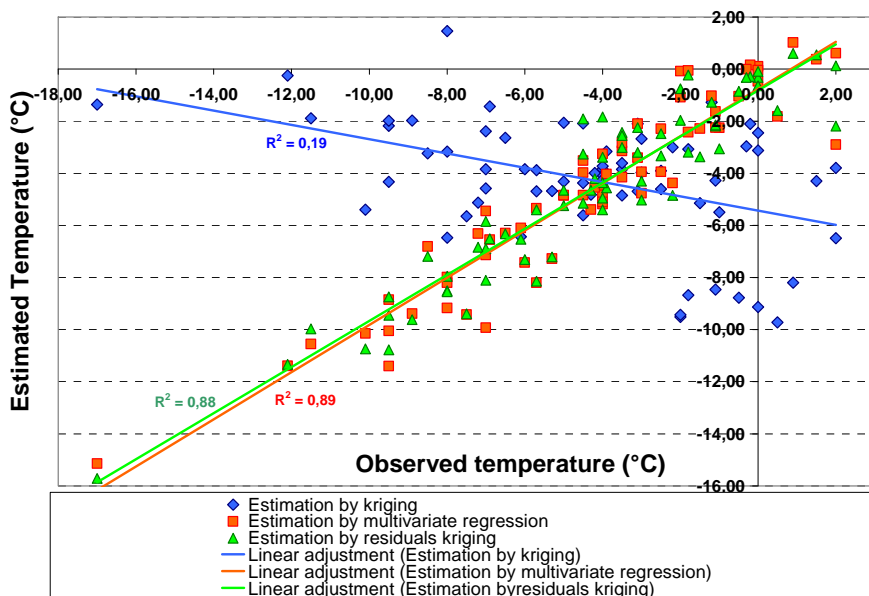
Le coefficient de détermination entre valeurs estimées et mesurées varie de 44% en utilisant le krigeage ordinaire (directement sur les valeurs de températures) à 69% en combinant régression multiple et résidus krigés, ce qui demeure un résultat moyen (la régression linéaire multiple seule atteint 65%). La journée choisie ne procure pas de résultats exceptionnels mais souligne la stabilité de la méthode. Rappelons que des coefficients de plus de 90% sont atteints certains jours. Cependant, dans le type de cas présent (journée du 3 février 1995), une meilleure prise en compte du relief générateur d'inversion thermique serait nécessaire.



**Figure 2 :** Comparaison des températures estimées par différentes méthodes d'interpolation avec les températures mesurées (échantillon de validation non inclus dans la construction du modèle), 03/02/1995, type de temps à caractère anticyclonique avec champ de hautes pressions centré sur la France.

Une seconde journée, le 25 décembre 1993 (figure 3), a été choisie pour représenter un modèle de régression multiple plus fonctionnel. La situation météorologique est tout autre puisque la France subit un flux de Nord-Ouest perturbé (phénomène qui peut être à l'origine de chutes de neige conséquentes sur le relief). En pratique, les résultats sur des journées à type de temps perturbé sont toujours meilleurs car la régression multiple linéaire, fondée principalement sur les effets engendrés par le relief sur les températures, profite de l'instabilité atmosphérique. Le krigeage des résidus n'apporte presque aucune information nouvelle (le coefficient de détermination entre températures observées et températures modélisées est de 89% pour la technique de régression multiple plus résidus krigés contre 88% pour la régression multiple seule). Néanmoins, l'interpolation par krigeage ordinaire des valeurs de températures ne fonctionne pas du tout ce jour (coefficient de détermination de 19% seulement). La méthode purement spatiale du krigeage démontre ainsi ses limites.

Quand les paramètres explicatifs utilisés par la régression multiple sont fortement corrélés avec les températures (le type de temps perturbé est idéal à ce titre), le krigeage des résidus n'apporte qu'une faible part explicative supplémentaire. Par contre, son usage est intéressant dans la prise en compte des effets régionaux lorsque la situation météorologique rend les variables indépendantes moins efficaces dans la modélisation.

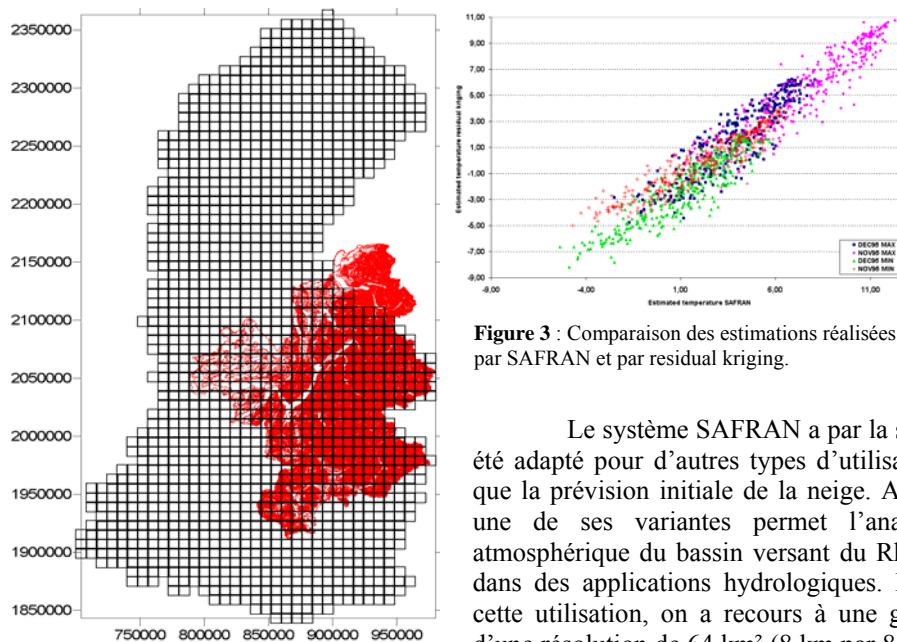


**Figure 3:** Comparaison des températures estimées par différentes méthodes d'interpolation avec les températures mesurées (échantillon de validation non inclus dans la construction du modèle), 25/12/1993, type de temps à caractère dépressionnaire : temps perturbé, flux de nord ouest sur les Alpes.

Des évolutions saisonnières et quotidiennes sont observables dans les résultats : ceux-ci sont les meilleurs au printemps et durant les journées perturbées [4]. Ils sont parfois plus médiocres durant la saison hivernale et quand la circulation atmosphérique est de type anticyclonique. La persistance hivernale des zones de hautes pressions

atmosphériques est l'un des cas particuliers de difficultés à résoudre : la modélisation est plus efficace dans ce cas à l'intérieur des vallées étroites, parfois plus délicate dans les larges vallées.

**2.3. Comparaison des résultats à un modèle existant : SAFRAN** Le système SAFRAN (Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige, Durand et al, 1993) a été développé par le Centre d'Etude de la Neige (Météo France) pour fournir, à l'échelle de chaque massif des Alpes des paramètres météorologiques : température de l'air au voisinage de la surface, vent de surface, humidité de l'air, nébulosité totale intégrée, quantité et phase des précipitations, rayonnement atmosphérique et rayonnement incident. Les massifs ainsi délimités ne constituent pas de maillage régulier : leur surface varie entre 400 et 1000 km<sup>2</sup>.



**Figure 3 :** Comparaison des estimations réalisées par SAFRAN et par residual kriging.

**Carte 8 :** Grille SAFRAN superposée à notre zone d'étude

Le système SAFRAN a par la suite été adapté pour d'autres types d'utilisation que la prévision initiale de la neige. Ainsi, une de ses variantes permet l'analyse atmosphérique du bassin versant du Rhône dans des applications hydrologiques. Pour cette utilisation, on a recours à une grille d'une résolution de 64 km<sup>2</sup> (8 km par 8 km) interpolée de la façon suivante : chaque maille se caractérise par la position de son centre et par son altitude. La transformation du zonage SAFRAN en maillage régulier se fait par affectation de chaque maille à la zone qui contient son centre [2]. Sur cette région discrétisée en maille de 8 kilomètres de côté, SAFRAN permet d'interpoler sur la grille les paramètres météorologiques nécessaires aux autres modules climatiques. L'application climatologique de cette base consiste en une simulation avec un réseau de station complet (sont ajoutées ensuite les stations climatiques non automatiques, non incluses d'office dans le système initial, qui contient uniquement les mesures des stations synoptiques et automatiques). Ainsi, SAFRAN peut nous fournir à échelle quotidienne, entre autres paramètres de surface, les températures minimales et maximales. Il nous est alors possible de comparer nos résultats avec ceux fournis par le système SAFRAN (il faut tout de même conserver à l'esprit que cette comparaison est limitée, de par la différence de résolution de travail des deux modèles : la confrontation



du modèle de régression multiple et résidus krigés dégradé à 8 kilomètres et la grille SAFRAN n'aura qu'une valeur indicative et illustrative, cf. carte 8 et figure 3).

La figure 3 illustre la comparaison des résultats estimés sur près de 400 points de grille par SAFRAN et par régression multiple suivie du krigeage des résidus (residual kriging). Le test présenté a été réalisé pour les mois de novembre et décembre 1995, concernant les températures minimales et maximales. Par rapport à SAFRAN, notre méthode semble sous estimer les basses valeurs de température (notamment en décembre 1995). Il faut admettre que la comparaison entre les deux modèles s'avère difficile par construction ; de nombreux biais sont introduits pour obtenir une estimation de notre modèle à 8 kilomètres de résolution : dégradation de notre modèle de 150 mètres de côté à 8000 mètres et coordonnées extrêmes différentes entraînent une non correspondance des pixels de 64 km<sup>2</sup> obtenus. Notre travail se base sur les effets locaux quand le modèle physique résonne sur des processus atmosphériques à l'échelle de l'ensemble du pays. Par ailleurs, les valeurs de SAFRAN sont celles du centre de la grille, quand notre modèle effectue la moyenne du pixel. Les estimations, par essence, ne peuvent donc correspondre intégralement et des écarts significatifs sont observés.

La validation par comparaison au modèle SAFRAN représente donc une difficulté, cependant l'inclusion d'un indice de représentativité des stations et la validation sur des températures mesurées non intégrées dans le modèle nous permettent d'affirmer la robustesse du modèle en conditions perturbées. Une amélioration de la prise en compte des régimes anticycloniques hivernaux est en cours.

**3. SIG et données climatiques : deux illustrations** Nous avons choisi de présenter deux illustrations de phénomènes climatiques modélisés sous SIG : la radiation potentielle sur une journée hivernale (21 février) et l'amplitude diurne annuelle pour l'année 1995 sur un département alpin. Ces deux indices climatiques révèlent le potentiel des systèmes d'information géographique dans ce domaine.

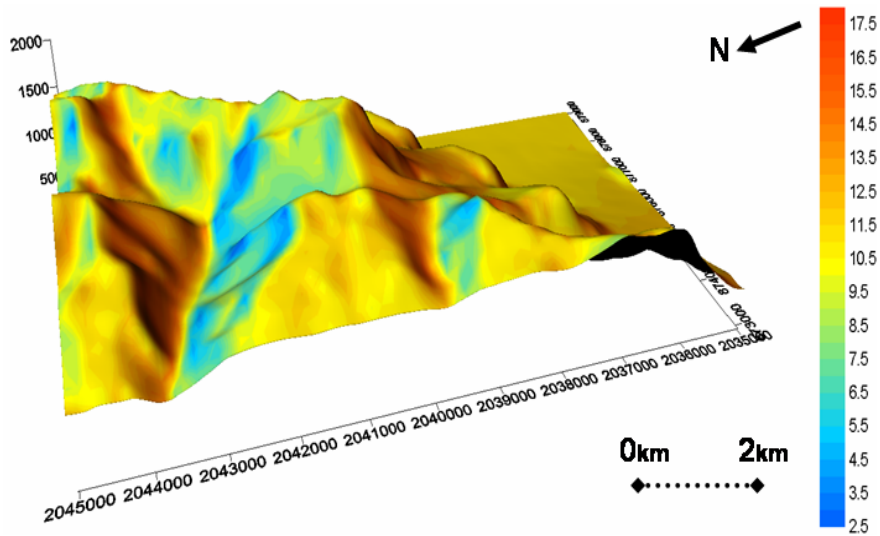
La première carte (carte 9) présente l'une des bases de données incluses dans le modèle de régression utilisé pour la spatialisation des températures. En effet, de nombreux processus d'ordre climatique sont sous l'influence directe de l'énergie solaire reçue à la surface de la Terre. L'ombrage par le relief environnant et l'exposition des versants (adret/ubac) sont, visuellement, les éléments prépondérants à la répartition de la radiation. Le relief constaté est facteur de fortes fluctuations des valeurs calculées sur des surfaces minimes.

Les secteurs à l'ombre d'un relief peuvent, durant l'hiver, rester gelés pendant plusieurs semaines, quand le versant adret profite pendant ce temps des moindres rayons pour voir sa température s'élever. D'un autre point de vue, le sol de certains versants montagneux est parfois plus chaud que celui des plaines environnantes. L'orientation des pentes doit donc se voir d'abord comme un facteur d'exposition à la radiation solaire, donc au réchauffement dans le cas d'une exposition sud (mais aussi une illumination matinale dans le cas des versants Est, tardive pour les versants Ouest). Les évolutions saisonnières, même si elles n'apparaissent pas sur l'illustration, sont aussi très marquées.

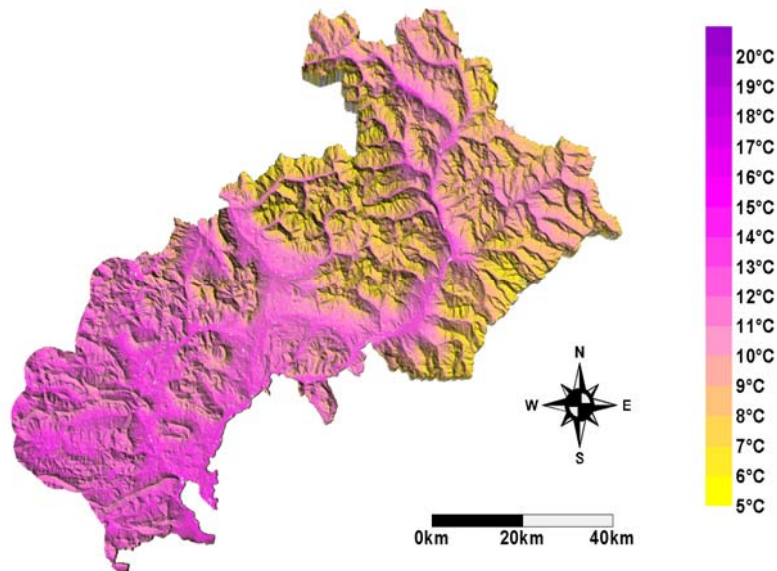
La seconde carte (carte 10) est, au contraire, calculée à partir des modèles de températures précédemment décrits. Elle illustre notamment le caractère déterminant de l'altitude dans les variations d'amplitude thermique : il est reconnu que les fonds de vallées subissent de plus grands écarts de températures que les sommets. L'amplitude varie également en fonction de la région, du relief environnant, du type de temps ou encore de la saison. La température minimale est généralement plus élevée lorsque la

nébulosité est forte, et la maximale au contraire plus élevée par temps clair. Les nuages ont en effet un rôle de régulateur thermique et ont donc tendance à atténuer à la fois le refroidissement nocturne et le réchauffement diurne. L'amplitude est de ce fait normalement plus forte en été qu'en hiver.

Il existe beaucoup d'autres utilisations possibles dans le domaine : on pourrait citer par exemple l'évaluation de la durée de l'hiver thermique ou celle de la durée de la période végétative, etc.



**Carte 9** : Exemple de données climatiques traitées par système d'information géographique : carte de radiation potentielle reçue le 21 février, en MJ.m<sup>-2</sup>.



**Carte 10** : Exemple de carte dérivée : carte d'amplitude thermique diurne moyenne en 1995 sur le département des Hautes-Alpes (France).

**4. Conclusion** Les systèmes d'information géographique, couplés à une analyse statistique, sont des outils privilégiés pour le traitement de données climatiques. L'exemple de spatialisation des températures en zone de montagne par régression linéaire multiple sur paramètres de terrain à l'emplacement de chaque station de mesure puis l'interpolation des résidus de la régression par krigeage souligne la qualité et la complexité des modèles réalisables. Les SIG permettent également de stocker et dériver de l'information, dans notre exemple pour l'obtention de variables explicatives (radiation potentielle notamment) ou la création d'autres indices climatiques. Les résultats estimés se doivent d'être validés par confrontation à des données mesurées non incluses dans le modèle ou par comparaison avec d'autres systèmes.

De manière à améliorer la prise en compte du terrain dans nos modèles d'interpolation (pour les journées avec inversion thermique, généralement en hiver), nous envisageons d'enrichir le système par calcul de coefficients de corrélation entre les paramètres de terrain et les températures sur des fenêtres glissantes centrées sur les stations de mesure et de largeur variable (3\*3 pixels, 5\*5 pixels...). Les premiers résultats traduisent une augmentation du coefficient de corrélation en hiver avec la croissance de la fenêtre, situation qui n'apparaît pas en été : ils confortent donc ce choix pour un affinement de la prise en compte du relief local et régional dans nos modèles.

## BIBLIOGRAPHIE

1. Benichou, P., Le Breton, O. (1987) – *Prise en compte de la topographie pour la cartographie des champs pluviométriques statistiques*, La Météorologie 7<sup>ème</sup> série, 1987, 14 p.
2. Etchevers, P., Durand, Y. – *Analyse SAFRAN des paramètres météorologiques de surface sur le bassin versant du Rhône de 1981 à 1994*, Note de travail du centre d'études de la neige, n°13, Météo-France.
3. Lhotellier, R. (2003) – *Influence of geographic factors on minimum and maximum temperatures in the French Alps*, Geographical Information Systems and Remote Sensing: Environmental Applications, Proceedings of the International Symposium, Volos, Greece, 7-9 November 2003, 10 p.
4. Lhotellier, R. (2004) – *Spatial interpolation of daily temperatures in the French Alps*, *Proceedings of the conference on spatial interpolation in climatology and meteorology*, Budapest, Hongrie, 25-28 Octobre 2004, 9 p.
5. Tveito, O. E., Schoner, W. (2002) – *Applications of spatial interpolation of climatological and meteorological elements by the use of geographical information systems (GIS)*, met.no report 28, COST 719, 44 p.
6. Van Dam, O. (2000) – *Modelling incoming Potential Radiation on a land surface with PCRaster*. POTRAD5.MOD manual, Utrecht Centre for Environment and Landscape dynamics, Utrecht University, the Netherlands.