



**HAL**  
open science

## Modélisation statistique multicritère de l'îlot de chaleur urbain en contexte caniculaire

Nassima Hassani, Didier Francois, Gilles Drogue

► **To cite this version:**

Nassima Hassani, Didier Francois, Gilles Drogue. Modélisation statistique multicritère de l'îlot de chaleur urbain en contexte caniculaire. 35e colloque annuel de l'Association Internationale de Climatologie, Jul 2022, Toulouse, France. hal-03655897v1

**HAL Id: hal-03655897**

**<https://hal.univ-lorraine.fr/hal-03655897v1>**

Submitted on 30 Apr 2022 (v1), last revised 10 Jul 2022 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Modélisation statistique multicritère de l'îlot de chaleur urbain en contexte caniculaire

Hassani N.<sup>1</sup>, François D.<sup>1</sup>, Drogue G.<sup>1</sup>

<sup>1</sup> Université de Lorraine, LOTERR, F-57000 Metz, France ([nassima.hassani@univ-lorraine.fr](mailto:nassima.hassani@univ-lorraine.fr))

**Résumé :** Cette étude vise à tester la robustesse d'une méthode d'interpolation spatiale multicritère de l'îlot de chaleur urbain (ICU) dans l'agglomération messine à partir d'un réseau de sondes de température sous abri. L'intensité de l'ICU est définie à partir d'un indice ICU ( $\Delta T_{u-r}$  maximum moyen sur  $n$  nuits consécutives) calculé pour chaque point de mesure par rapport à une référence péri-urbaine (station Base-vie). Trois épisodes caniculaires récents ont été testés (juin 2019, juillet 2019, août 2020) : pour chacun d'entre eux, l'indice ICU est mis en relation avec des descripteurs physiques de l'environnement afin de construire un modèle statistique. Ce dernier fait ensuite l'objet d'un test de performance et de stabilité par fractionnement de données. Les résultats montrent que le modèle est relativement stable et performant.

**Mots-Clés :** Indice ICU, régression linéaire multiple, analyse spatiale, modélisation

**Abstract:** This study aims to test the robustness of a multi-criteria interpolation method of urban heat island (UHI) in the Metz city conurbation based on a limited network of temperature probes with shelter. The intensity of the UHI is defined from an index ( $\Delta T_{u-r}$  averaged over  $n$  consecutive nights) calculated for each measurement point with respect to a peri-urban reference (Base-vie station). Three recent heatwave episodes have been tested (June 2019, July 2019, August 2020): for each of them the UHI index is related to physical descriptors of the urban environment to build a statistical model. The model is then tested for performance and stability by splitting the data. The results show that the model is relatively stable and performs well.

**Key words:** UHI index, multiple linear regression, spatial analysis, modelling.

## Introduction

Cette étude s'intéresse à l'organisation spatiale de l'ICU en période caniculaire sur la partie de l'Eurométropole de Metz instrumentée d'un réseau de sondes HOBO (Hassani et Drogue, 2020). L'objectif est de tester la performance et la robustesse d'une méthodologie permettant de connaître l'intensité de l'ICU en tout point du terrain d'étude au cours de trois épisodes de chaleur historiques qui l'ont affecté récemment lors des étés 2019 et 2020 et au cours desquels les températures ont été d'une intensité exceptionnelle au point de dépasser par endroit les 40°C. Pour cela, nous avons eu recours à une modélisation statistique multicritères. Celle-ci a déjà été employée avec succès dans de nombreux travaux en climatologie pour diverses applications (agricoles, urbaines, forestières...) et variables (précipitations, température, ensoleillement...) (Foissard, 2015; Straub et al., 2019). Cette modélisation statistique multicritère est testée ici dans une approche de géo-régression multiple couplée à du rééchantillonnage pour identifier les co-variables les plus significatives.

## 1. Zone d'étude et réseau de mesure fixe

L'agglomération messine se situe dans le quart nord-est de la France (figure 1). Elle s'inscrit dans le territoire de l'Eurométropole de Metz qui compte environ 220 000 habitants répartis dans 44 communes (Insee, 2016). D'une étendue de 306 km<sup>2</sup>, celle-ci est traversée par deux cours d'eau : la Moselle et la Seille. Selon la classification de Köppen-Geiger, ce territoire connaît un climat tempéré chaud de type Cfb/Csb, avec des hivers relativement froids et des étés relativement chauds (Eveno et al., 2016). Le territoire de l'agglomération messine a été équipé de 21 sondes de température de type HOBO en 2019, afin que différents types d'environnement (urbains, péri-urbains et ruraux) soient échantillonnés (Hassani et Drogue, 2020). Ce réseau a été complété en 2020 par 5 points de mesures supplémentaires dont 4 sondes HOBO et une station météorologique complète Vantage Pro 2 Plus (station 22). Ces 26 points d'observations dont les altitudes s'échelonnent entre 167 m et 232 m forment le *Réseau de Mesures Thermo-Hygrométriques de l'agglomération de Metz (REMTHAM)*. Deux

stations météorologiques gérées par Météo-France (aérodrome de Metz Frescaty) et ATMO Grand-Est sont également présentes sur le territoire étudié.

## 2. Méthodologie

L'interpolation de l'ICU est réalisée en trois étapes : le pré-traitement des données, la construction du modèle d'interpolation et la validation du modèle.

### 2.1 Prétraitement des données

Notre point de départ est le calcul d'un indice ICU durant les séquences caniculaires des étés 2019 (25-30 juin et 21-26 juillet) et 2020 (7 août - 13 août), qui ont offert la possibilité de mieux comprendre le comportement thermique du tissu urbain et la dynamique de l'ICU local lors d'extrêmes chauds. Cet indice se calcule ainsi pour chaque point de mesure :

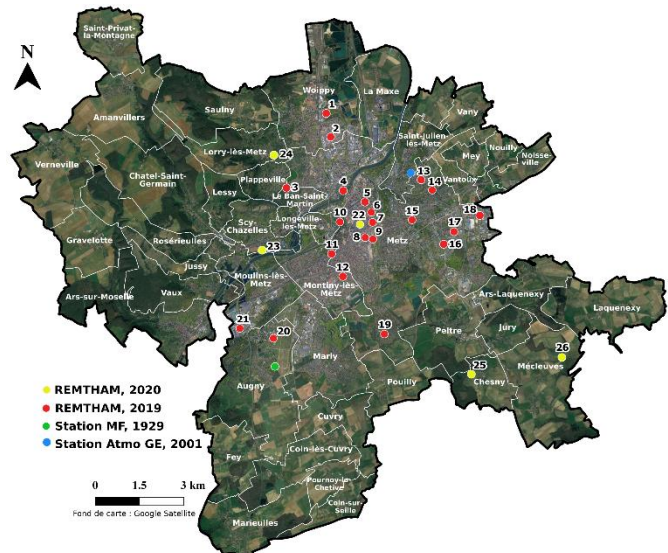
$$\Delta Tu - r_{max,i} = \frac{1}{n} \sum_{j=1}^n \Delta_{max} (T^{\circ}_{air\ sonde,i} - T^{\circ}_{Base-vie}) \quad (1)$$

où  $\Delta Tu - r_{max,i}$  représente la moyenne des écarts maximum horaires ( $\Delta_{max}$ ) entre  $T^{\circ}_{air\ sonde,i}$  et  $T^{\circ}_{air\ Base-vie}$  observés entre le coucher et le lever du soleil pour  $n$  nuits consécutives ;  $T^{\circ}_{air\ sonde,i}$  représente la température de l'air mesurée sous abri à 3 m du sol (hauteur retenue habituellement notamment pour limiter le vandalisme et le risque de vol) par une sonde  $i$  du réseau REMTHAM ;  $T^{\circ}_{air\ Base-vie}$  représente la température de l'air mesurée sous abri à 3 m du sol par la sonde péri-urbaine de Base-vie située à proximité de l'aérodrome de Metz Frescaty à environ 6 km à vol d'oiseau du cœur de ville (Hassani et Drogué, 2020).

Une base de données constituée de 801 descripteurs relatifs à la topographie (altitude, exposition, encaissement...), la couverture du sol (surfaces bâties, végétation, surfaces imperméables, NDVI...) ou la morphologie urbaine (comme le facteur de vue du ciel) a été ensuite créée à partir de la BD TOPO v.3 (2019) et du MNT RGE Alti de l'IGN. Certains descripteurs ne se distribuant pas selon une gaussienne (ex. le descripteur « surfaces imperméables » dans un rayon de 50 m), des transformations (exponentielle, logarithmique...) ont été appliquées aux données. Chaque descripteur a été déterminé sur des grilles de 5 m dans des zones tampons autour de chaque point de mesure (50 m à 100 m, 200 m, 300 m, 400 m, 500 m, 700 m, 800 m, 900 m) à l'aide du logiciel *Matlab*, dans le but d'identifier sa dimension optimale du point de vue statistique (Amorim et al., 2015 ; Foissard et al., 2019). Des indices statistiques telles que le minimum, le maximum, la somme, la moyenne et l'écart-type ont également été déterminés pour chaque rayon.

### 2.2 Construction du modèle statistique

En premier lieu, un préfiltrage des descripteurs a été effectué grâce à une analyse de multicollinéarité pour ne conserver que les descripteurs indépendants. Ensuite, nous avons automatisé la méthode de construction de la régression linéaire multiple (RLM) en utilisant l'outil *Regsubsets* du logiciel *R* qui explore toutes les combinaisons possibles de descripteurs, les compare pour ensuite les classer selon des critères prédéfinis. Le nombre de co-variables retenues dans la régression est alors paramétré (trois au maximum par souci de parcimonie), de même que le résultat graphique du classement des modèles possibles. Ensuite, nous avons réalisé un tri entre les co-variables à l'aide du *Bayesian Information Criterion (BIC)* pour sélectionner le meilleur modèle de régression (au sens de la minimisation des erreurs).



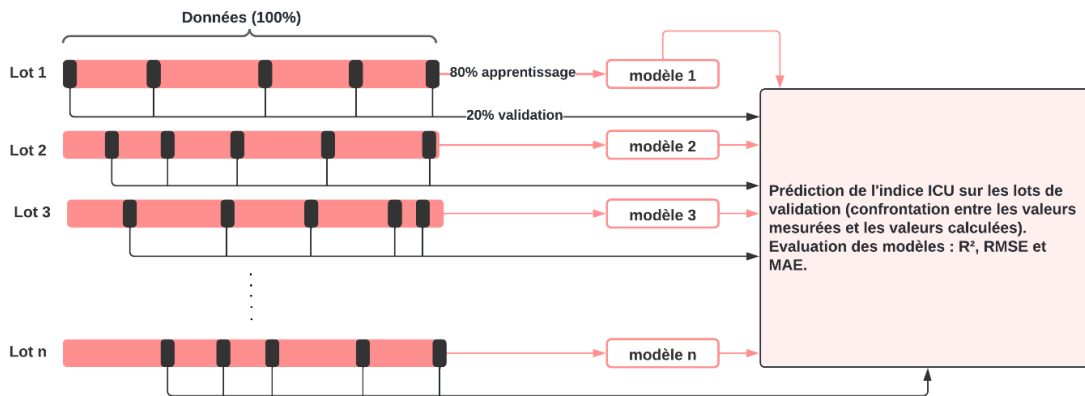
**Figure 1** : Carte de répartition des sondes du réseau REMTHAM sur le territoire de l'Eurométropole de Metz.

La qualité du modèle de régression a été vérifiée sur la base de métriques d'erreur classiques ( $R^2$  et  $R^2_{aj} > 0,80$  et erreur quadratique moyenne RMSE  $< 0,6^\circ\text{C}$ ). En outre, les résidus du modèle ont été analysés afin de vérifier leur normalité, leur homoscédasticité et leur éventuelle structuration dans l'espace. Les modèles de régression finaux ont été implémentés à l'aide du module d'analyse spatiale d'*ArcGIS* afin de produire les cartes de l'indice ICU présentées dans la section 3.

### 2.3. Test de stabilité et de performance par fractionnements de données

Pour tester la stabilité de la paramétrisation du modèle statistique et son niveau de performance, une technique de rééchantillonnage a été appliquée aux données collectées lors de chaque épisode caniculaire. La variante testée est celle des échantillons fractionnés ou *split-sample test*. Comme son nom l'indique, cette technique consiste à évaluer les performances d'un modèle en divisant l'ensemble du jeu de données en deux sous-ensembles : l'un sert à l'apprentissage, au calage du modèle ; l'autre à sa validation.

Le partitionnement utilisé est du type 80/20 : 80% des données sont utilisés pour la phase d'apprentissage, 20% pour la phase de validation. À chaque itération, les données d'apprentissage et les données de validation sont tirées au hasard (Figure 2). Grâce à cette technique, un grand nombre de modèles a été ajusté (440 pour les canicules de juin et juillet 2019 et 1000 pour celle d'août 2020). Les modèles sont évalués en calculant trois métriques :  $R^2$  (variance expliquée), RMSE (erreur quadratique) et MAE (biais de l'erreur).



**Figure 2 :** Chaîne de traitement décrivant la méthodologie suivie pour le calage et la validation croisée du modèle d'interpolation spatiale.

## 3. Résultats

### 3.1. Géorégression et sous-échantillonnage

Les résultats des régressions multiples appliquées aux trois périodes caniculaires sont présentés dans le tableau 1. Dans les trois cas, le pourcentage de variance expliquée excède 80%, ce qui est conforme à l'état de l'art (voir Foissard, 2015 ; Straub et al., 2019 par exemple).

**Tableau 1 :** Synthèse des statistiques des modèles de régression linéaire multiple.

Type d'épisode	Modèles de régression		
	$R^2$	$R^2$ ajusté	RMSE ( $^\circ\text{C}$ )
Canicule juin 2019	0,82	0,80	0,42
	$\Delta Tu - r_{max,i} = -1,60 + 0,08 * \text{Bat}_{900} + 0,80 * \log(\text{Sim}_{50})$		
Canicule juillet 2019	0,85	0,82	0,54
	$\Delta Tu - r_{max,i} = -13,11 + 0,14 * \text{Bat}_{900} + 0,81 * \text{Altitude} + 1,16 * \log(\text{Sim}_{50})$		
Canicule août 2020	0,89	0,88	0,45
	$\Delta Tu - r_{max,i} = -6,43 + 0,14 * \text{Bat}_{900} + 0,08 * \text{Eau}_{800} + 0,03 * \text{Altitude}$		

Sans surprise, plusieurs co-variables se rapportant à l'environnement bâti (Bat\_900 : fraction bâtie dans un rayon de 900 m et Sim\_50 : surface imperméable dans un rayon de 50 m), la présence d'eau (Eau\_800 : surfaces en eau dans un rayon de 800 m) et la topographie (Altitude) sont identifiées. La fraction bâtie dans une zone tampon de 900 m est commune aux trois modèles d'interpolation et traduit l'effet réchauffant (par rugosité et piégeage infrarouge) des parties de l'agglomération les plus densément bâties. La co-variable Sim\_50 traduit un effet réchauffant plus localisé lié au stockage d'énergie dans les surfaces artificialisées (routes, parkings et places). Le signe de la corrélation est cohérent pour ces deux co-variables. En revanche pour les deux autres co-variables (Eau\_800 et Altitude), le signe de la corrélation est contre-intuitif même si concernant l'eau, on sait que son effet rafraîchissant est faible en raison de son inertie thermique. Notons que la végétation n'est pas identifiée comme co-variable sans doute en raison d'une sous-représentation dans les environnements échantillonnés par nos sondes.

Le tableau 2 représente les co-variables les plus fréquentes dans les modèles ajustés sur les sous-ensembles d'apprentissage. Les co-variables dominantes sont identiques à celles du modèle « global » ajusté sur l'ensemble des données disponibles. Cela démontre une certaine stabilité des paramétrisations. L'augmentation du nombre de capteurs entre 2019 et 2020 a visiblement permis de gagner en stabilité puisque les co-variables Bat\_900 et Altitude sont ultradominantes.

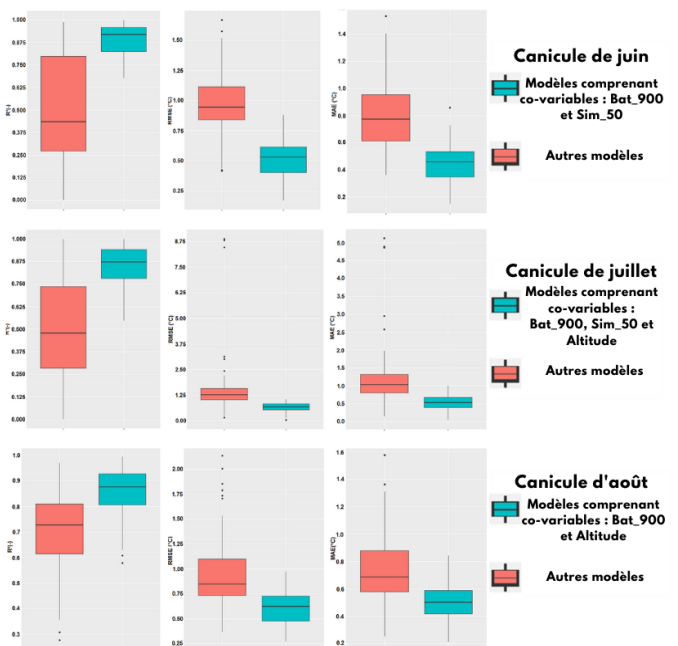
La figure 3 synthétise les statistiques issues de l'application des modèles en validation. Les modèles intégrant les co-variables récurrentes du tableau 2 sont représentés par les boîtes à moustaches bleues, tandis que les autres modèles sont représentés par les boîtes à moustaches rouges. Les statistiques prouvent que les modèles intégrant les co-variables récurrentes sont bien les plus compétitifs (gain en  $R^2$  moyen de 0,15 à 0,5 entre août 2020 et juin 2019 ; chute des RMSE et MAE moyennes de quelques dixièmes pour les trois épisodes testés). Il existe cependant des sous-ensembles qui génèrent de mauvaises simulations (score en  $R^2 < 0,5$  ; RMSE et MAE  $> 1^\circ\text{C}$ ).

### 3.2. Cartes de l'indice ICU

Les cartes produites (figure 4) fournissent les principales tendances de l'indice ICU. La carte de la canicule de juin 2019 se démarque des deux autres : l'indice ICU en cœur de ville est inférieur à  $3^\circ\text{C}$  et l'altitude n'est pas un déterminant de l'indice, contrairement aux deux autres canicules. Sur les cartes de juillet 2019 et d'août 2020, l'indice augmente au niveau des zones d'activités (situées au nord, à l'est et au sud-ouest de la carte) et sur les coteaux de la Moselle, ce qui se traduit par l'apparition de petits îlots de chaleur. L'influence de l'altitude est prégnante surtout sur la carte d'août 2020 : les écarts de température « ville-campagne » sont plus élevés lorsque l'altitude augmente et l'on voit apparaître des noyaux de chaleur en milieu forestier sur des points hauts (Mont Saint-Quentin). De

**Tableau 2 :** Fréquence des co-variables dans les modèles de régression ajustés sur les sous-ensembles d'apprentissage.

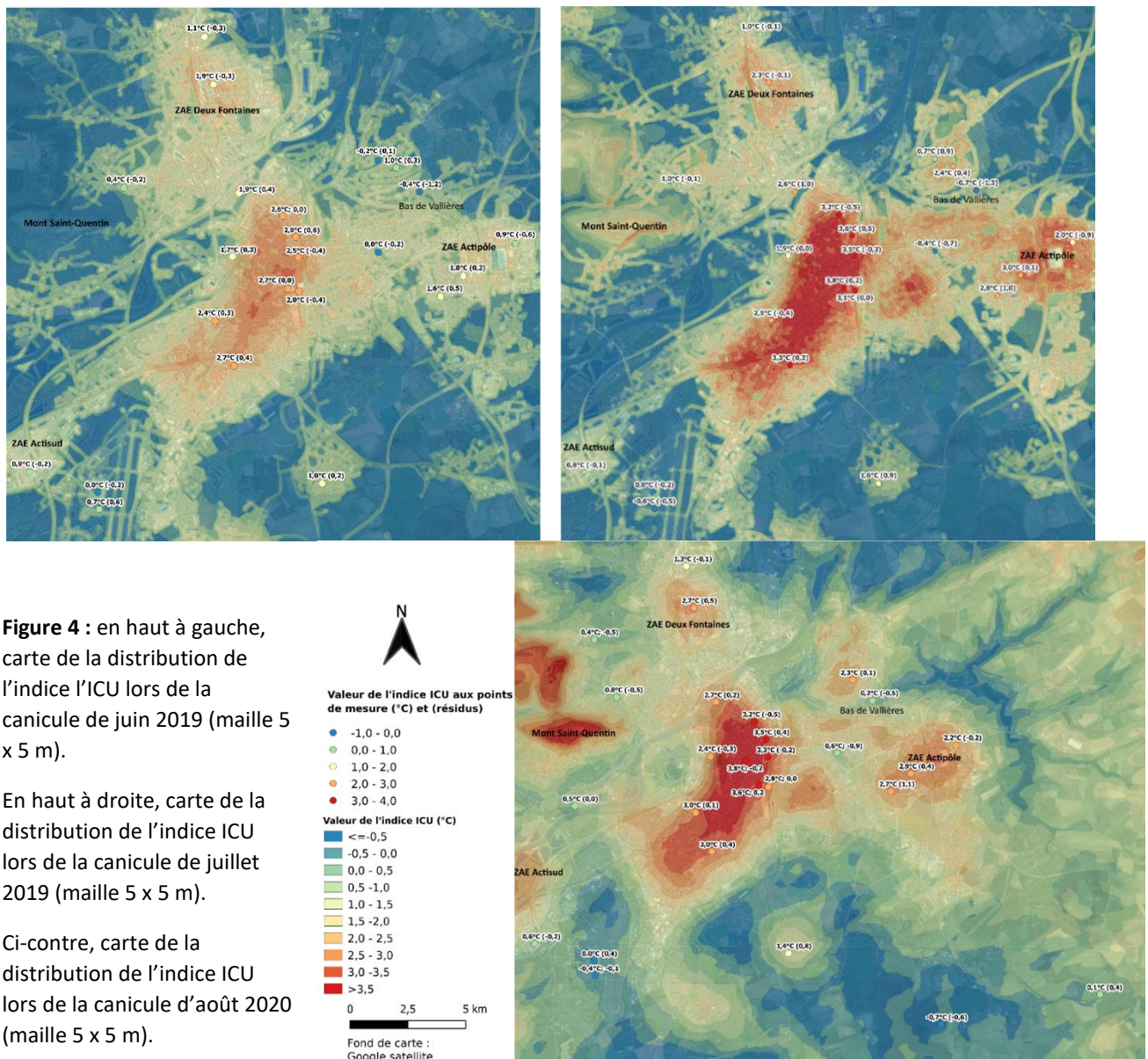
Co-variables	Fréquence (%)		
	juin 2019	juillet 2019	août 2020
<b>Bat_900</b>	58,4	47,9	88,50
<b>Sim_50</b>	53,18	55,2	12,70
<b>Altitude</b>	/	55,7	87,48
<b>Eau_800</b>	/	/	2,5



**Figure 3 :** Boîtes à moustaches représentant les métriques de validation ( $R^2$ , RMSE et MAE) des modèles de régression intégrant (bleues) ou non (rouges) les co-variables les plus fréquentes.

petites poches de fraîcheur se dessinent dans les parcs et points d'eau, ainsi que dans les creux topographiques.

Notons que sur la carte de juin 2019, le résidu le plus important est celui de la sonde « Bas de Vallières » avec une valeur de  $-1,2^{\circ}\text{C}$  (figure 4). Cette sonde est située au fond d'un vallon où coule le ruisseau de Vallières et se démarque des autres puisque son écart n'est que de  $-0,4^{\circ}\text{C}$  avec la sonde de référence, soit  $2^{\circ}\text{C}$  à  $2,5^{\circ}\text{C}$  de moins que les sondes urbaines. Elle restitue donc une ambiance typiquement « rurale » en milieu suburbain. Nous supposons que durant la nuit, sur les versants du vallon, l'air au contact du sol se refroidit par conduction, devient plus dense et s'écoule gravitairement vers les points « bas » ou creux topographiques alimentant une légère brise de versant. Le modèle a du mal à détecter ces axes de fraîcheur provoqués par la topographie en creux sans doute en raison d'une sous-représentation de ce phénomène dans les mesures. Cela se vérifie également pour juillet 2019 malgré l'intégration de la co-variable altitude. En 2020 ce résidu « rentre dans la norme » grâce à un *pattern* qui est davantage structuré par la topographie.



**Figure 4** : en haut à gauche, carte de la distribution de l'indice ICU lors de la canicule de juin 2019 (maille 5 x 5 m).

En haut à droite, carte de la distribution de l'indice ICU lors de la canicule de juillet 2019 (maille 5 x 5 m).

Ci-contre, carte de la distribution de l'indice ICU lors de la canicule d'août 2020 (maille 5 x 5 m).

#### 4. Discussion

Malgré un échantillonnage de données assez limité, ce travail a permis d'établir des relations statistiquement robustes entre un indice ICU calculé à partir de données de température et des

descripteurs physiques de l'environnement à l'échelle de l'agglomération messine pendant des épisodes caniculaires. Une approche multicritère a été utilisée pour identifier les deux ou trois co-variables les plus significatives. Celles-ci sont principalement liées aux caractéristiques de l'environnement bâti à deux échelles : la densité du bâti à l'échelle locale et la présence de surfaces imperméables à l'échelle ultra-locale. Cette paramétrisation est cohérente pour une agglomération de taille moyenne. Dans les trois cas testés (juin et juillet 2019, août 2020), la régression sur des sous-ensembles de sondes permet d'expliquer plus de 80 % de la variabilité spatiale de l'indice ICU et les niveaux d'erreur en validation fractionnée sont conformes à l'état de l'art. La paramétrisation des modèles de régression est assez stable. Ceci prouve que les sites de mesure ont été bien choisis et que l'information apportée par les sondes reste pertinente même lorsque le réseau de mesure est dégradé. Si les relations entre l'indice ICU et les co-variables se rapportant au milieu urbain paraissent réalistes (*i.e.* physiquement interprétables) ; en revanche, l'influence univoque de l'altitude détectée pour deux canicules sur trois est contre-intuitive. Elle permet de restituer l'effet d'une inversion thermique dans les vallons mais tend probablement à surestimer l'indice ICU sur certains points hauts de la topographie (dépassant 250 m) *a fortiori* lorsque ceux-ci sont boisés, et à sous-estimer l'indice ICU dans le chenal de la Moselle par un rafraîchissement excessif. L'absence de point de mesure dans ces sites ne nous permet pas à ce stade de formuler un avis définitif. Les cartes de l'indice ICU produites grâce à l'automatisation des calculs montrent des structures spatiales communes aux trois cartes (figure 4) : le cœur de ville où le phénomène d'ICU est le plus sensible, les zones d'activités qui restent chaudes la nuit et l'armature urbaine (petites villes périphériques). Les cartes se différencient principalement par le lissage qu'opère la topographie sur la carte d'août 2020 tandis que les deux autres cartes sont davantage structurées par la densité du bâti et les surfaces imperméables incluant les grandes voies de circulation (ce qui paraît singulier). Ces différences mettent en évidence le fait que la cartographie de l'indice ICU n'est pas figée dans le temps (influence des conditions atmosphériques et des écoulements d'air locaux) et dépend de la stratégie d'échantillonnage.

## Conclusion

Les limites de l'échantillonnage des données (seulement 28 sondes dans le meilleur des cas) incitent à la prudence vis-à-vis de cette première modélisation statistique qui rappelons-le, n'a pas la prétention d'estimer avec précision la valeur exacte de l'indice ICU en tout point du territoire mais plutôt d'esquisser les grands traits du phénomène. Nos travaux futurs s'orientent vers la mobilisation des sources de données complémentaires (images ASTER, données de campagne mobiles) pour documenter les températures urbaines dans les espaces dépourvus de mesure et juger du réalisme des modèles statistiques et des cartes produites.

**Remerciements** : les auteurs remercient l'Eurométropole de Metz pour son soutien et son accompagnement dans la pose des sondes ainsi que l'ADEME Grand Est pour son aide financière.

## Bibliographie

- Amorim M., Dubreuil V., Cardoso R., 2015 : *Modelagem espacial da ilha de calor urbana em presidente prudente (SP) – Brasil*. Rev. Bras. Climatol., 16, pp. 29-45.
- Eveno M., Planchon O., Oszwald J., Dubreuil V., Quénoel H., 2016 : *Variabilité et changement climatique en France de 1951 à 2010 : Analyse au moyen de la classification de Köppen et des types de climats annuels*. Climatologie, 13, 47–70.
- Foissard X., 2015 : *L'îlot de chaleur urbain et le changement climatique : application à l'agglomération rennaise*. Thèse de doctorat, Université de Rennes 2, Rennes, France, 248 p.
- Foissard X., Dubreuil V., Quénoel H., 2019 : *Defining scales of the land use effect to map the urban heat island in a mid-size European city: Rennes (France)*. Urban Climate, Volume 29.
- Hassani N. et Drogue G., 2020 : *Mesure et spatialisation de l'îlot de chaleur urbain dans l'aire urbaine de Metz Métropole : premiers résultats de la campagne de mesure 2019*. Climatologie, 17, 8.
- Straub A., Berger K., Breitner S., Cyrus J., Gerschkat U., Jacobeit J., Kühnbach B., Kusch T., Philipp A., Schneider A., Umminger R., Wolf K., Beck C., 2019 : *Statistical modelling of spatial patterns of the urban heat island intensity in the urban environment of Augsburg, Germany*. Urban Climate, Volume 29.