

Compréhension du microclimat urbain lyonnais par l'intégration de prédicteurs complémentaires à différentes échelles dans des modèles de régression

Lucille Alonso* et Florent Renard

Université Jean Moulin Lyon 3, UMR 5600 CNRS Environnement Ville Société, Lyon, France

Résumé – Le changement climatique est un phénomène majeur actuel générant de multiples conséquences. En milieu urbain, il exacerbe celui de l'îlot de chaleur urbain. Ces deux manifestations climatiques engendrent des conséquences sur la santé des habitants et sur la sensation d'inconfort thermique ressenti en milieu urbain. Ainsi, il est nécessaire d'estimer au mieux la température de l'air en tout point d'un territoire, notamment face à la rationalisation actuelle du réseau de stations météorologiques fixes de Météo France. La connaissance spatialisée de la température de l'air est de plus en plus demandée pour alimenter des modèles quantitatifs liés à un large éventail de domaines, tels que l'hydrologie, l'écologie ou les études sur les changements climatiques. Cette étude se propose ainsi de modéliser la température de l'air, mesurée durant 4 campagnes mobiles réalisées durant les mois d'été, entre 2016 et 2019, dans Lyon par temps clair, à l'aide de modèle de régressions à partir de 33 variables explicatives issues de données traditionnellement utilisées, de données issues de la télédétection par une acquisition LiDAR (Light Detection And Ranging) ou satellitaire Landsat 8. Trois types de régression statistique ont été expérimentés, la régression partial least square, la régression linéaire multiple et enfin, une méthode de machine learning, la forêt aléatoire de classification et de régression. Par exemple, pour la journée du 30 août 2016, la régression linéaire multiple a expliqué 89 % de la variance pour les journées d'étude, avec un RMSE moyen de seulement 0,23°C. Des variables comme la température de surface, le NDVI ou encore le MNDWI impactent fortement le modèle d'estimation.

Mots-clés : température de l'air / température de surface / LiDAR / régression linéaire multiple / Landsat 8 / îlot de chaleur urbain.

Abstract – **Understanding Lyon's urban micro-climate by integrating complementary predictors at different scales in regression models.** Climate change is a major contemporary phenomenon with multiple consequences. In urban areas, it exacerbates the urban heat island phenomenon. These two climatic manifestations generate consequences on the health of the inhabitants and on the sensation of thermal discomfort felt in urban areas. Thus, it is necessary to estimate as well as possible the air temperature at any point of a territory, in particular in view of the ongoing rationalization of the network of fixed meteorological stations of Météo France. Understanding the air temperature is increasingly in demand to input quantitative models related to a wide range of fields, such as hydrology, ecology or climate change studies. This study thus proposes to model air temperature, measured during 4 mobile campaigns carried out during the summer months, between 2016 and 2019, in Lyon in clear weather, using regression models based on 33 explanatory variables from traditionally used data, data from remote sensing by LiDAR (Light Detection And Ranging) or Landsat 8 satellite acquisition. Three types of statistical regression were experimented, partial least square regression, multiple linear regression and finally, a machine learning method, the random forest regression. For example, for the day of August 30, 2016, multiple linear regression explained 89% of the variance for the study days, with a mean RMSE of only 0.23°C. Variables such as surface temperature, NDVI and MNDWI have a strong impact on the estimation model.

Keywords: Air temperature / surface temperature / LiDAR / multiple linear regression / Landsat 8 / Urban Heat Island.

* Auteure de correspondance : lucille.alonso@univ-lyon3.fr

Introduction

Le changement climatique est un phénomène majeur actuel générant de multiples conséquences environnementales, sociales et économiques (GIEC, 2019). En milieu urbain, il exacerbe le phénomène de l'îlot de chaleur urbain (ICU) (Oke, 1981 ; Weston, 1988) qui se caractérise par une différence de température entre une zone urbaine et les espaces ruraux environnants. Dans ce cas, la température des zones urbaines est plus élevée que dans les zones rurales, surtout la nuit (Oke, 1973, 1978 ; Katsoulis et Theoharatos, 1985). Ces deux manifestations climatiques engendrent des conséquences sur la santé des habitants (Alonso et Renard, 2020) et sur la sensation d'inconfort thermique ressenti en milieu urbain (Dousset et Gourmelon, 2003 ; Dousset *et al.*, 2011). En outre, l'augmentation des vagues de chaleurs est nettement démontrée en Europe et en Amérique du Nord, que ce soit lorsque l'on s'intéresse à la durée, l'intensité ou la fréquence (Meehl et Tebaldi, 2004). Les effets des canicules viennent se superposer aux caractéristiques microclimatiques des milieux urbains (Fallmann *et al.*, 2016), ainsi qu'au processus d'urbanisation croissant de la population. Cette urbanisation croissante a un impact significatif sur le changement climatique et conduit à un réchauffement des températures en ville (Kukla *et al.*, 1986 ; Wang *et al.* 2014 ; Chapman *et al.*, 2017 ; Sun *et al.*, 2019). La combinaison de ces événements a notamment pour effet une augmentation du nombre prématuré de décès aux USA liés au stress thermique (Bobb *et al.*, 2014). En ce sens, les acteurs publics locaux tentent de prévenir et de réduire les risques humains potentiellement générés par une augmentation des vagues de chaleur. Connaître et comprendre l'effet de l'îlot de chaleur urbain est une condition *sine qua non* dans la conception de ville intelligente et durable (Renard *et al.*, 2019). D'après le US Department of Energy, les Etats-Unis dépensent chaque année 10 milliards de dollars dans le domaine énergétique pour réduire l'effet de l'îlot de chaleur urbain (Rosenfeld *et al.*, 1996). De plus, atténuer la surchauffe en ville est une question financière importante puisque chaque augmentation de 1°C de la température entraîne une hausse de 2 à

4 % de la demande en électricité (Akbari *et al.*, 2001). Une étude à Chicago a démontré que l'ajout de 10 % de couverture végétale, ou la plantation d'environ trois arbres par parcelle de terrain, permet de réduire les coûts énergétiques d'environ 50 dollars à 90 dollars (environ 45 euros à 80 euros) par an et par habitation (McPherson *et al.*, 1997).

Par ailleurs, la température de l'air est une variable principale dans l'explication des conditions de l'environnement, notamment urbain. Elle est également impliquée dans de nombreux processus écologiques importants tels que l'évapotranspiration réelle et potentielle, le rayonnement net ou la distribution des espèces (Cristóbal *et al.*, 2008). Ainsi, la connaissance de la température de l'air en tout point du territoire est de plus en plus demandée pour alimenter des modèles quantitatifs liés à un large éventail de domaines, tels que l'hydrologie, l'écologie (Kustas et Norman, 1996) ou la climatologie (De Ridder *et al.*, 2017 ; Leconte *et al.*, 2017 ; Alonso et Renard, 2019). C'est pourquoi, la compréhension des modèles de température de l'air est essentielle pour de multiples applications dans les domaines de l'agriculture, de l'hydrologie, de l'aménagement du territoire ou de la santé publique. En outre, il est nécessaire d'estimer au mieux la température de l'air en tous lieux, notamment face à la rationalisation actuelle du réseau de stations météorologiques fixes de Météo France (Journal officiel du Sénat, 2012). En France, les agglomérations qui disposent de leur propre réseau de stations météorologiques fixes ne sont que peu nombreuses, comme à Rennes ou à Dijon (Foissard *et al.*, 2019 ; Richard *et al.*, 2018).

En outre, cette compréhension du microclimat urbain passe inévitablement par une meilleure connaissance de l'ICU en combinant mesures et modélisation. L'étude proposée s'inscrit dans cette thématique grâce à une modélisation de la température de l'air en zones urbaines denses sur la Métropole de Lyon (France) par l'utilisation combinée de données traditionnellement utilisées, de données issues de la télédétection par une acquisition LiDAR (Light Detection And Ranging) ou satellitaire Landsat 8 et de données produites par la mesure mobile. Ces dernières s'avèrent être d'un grand recours puisqu'il n'existe pas, jusqu'à

présent, sur Lyon, et dans la plupart des grandes agglomérations, de réseaux de stations météorologiques fixes suffisamment déployés en centre-urbain. C'est une réelle opportunité, d'autant plus que la température de l'air évolue à une échelle métrique, à moins de 100 mètres (Nichol et To, 2012 ; Tsin *et al.*, 2016). De plus, l'utilisation d'informations obtenues à partir de capteurs aéroportés ou de satellites pour observer la surface de la terre depuis le ciel ou l'espace est une méthodologie qui évalue efficacement la distribution spatiale des variables de la surface terrestre à l'échelle locale et régionale (Mira *et al.*, 2017). La connaissance précise des températures est une nécessité tant pour l'environnement que pour les politiques de santé, en particulier dans les zones urbaines, qui peuvent contribuer à l'amélioration de la planification urbaine dans le contexte de l'atténuation de l'ICU et la création d'îlots de fraîcheur urbains (IFU). Cette connaissance dépend directement de la densité du réseau de mesure. En outre, il existe une variété de techniques d'interpolation comme les splines (Boer *et al.*, 2001 ; Jarvis et Stuart, 2001) et le krigeage. Plus récemment, la modélisation par régression (Boer *et al.*, 2001) ou encore par les réseaux de neurones et les techniques d'apprentissage de machine learning (Antonić *et al.*, 2001) est apparue. De plus, de multiples études ont abordé cette question, par ces interpolations spatiales classiques déterministes (Wang *et al.*, 2017) ou stochastiques (Zhang et Du 2019) ou bien par ces régressions multiples (Cantat, 2004 ; Carrega et Rosa, 2005 ; Cristobal *et al.*, 2006 ; Hengl *et al.*, 2012 ; Zhu *et al.*, 2013 ; Chen *et al.*, 2016 ; Kastendeuch *et al.*, 2016 ; Carrega et Martin, 2017 ; Mira *et al.*, 2017 ; Richard *et al.*, 2017 ; Alonso et Renard, 2019).

Ainsi, cette étude se propose dans un premier temps de délimiter la zone d'étude, puis d'aborder les méthodes d'acquisition de la donnée ainsi que les procédés statistiques, et enfin, d'analyser les résultats. Cette dernière partie nous permet de discuter de la contribution de chaque variable prédictive à la modélisation de la température de l'air et de l'erreur de mesure. Cette recherche vise à une amélioration de la planification urbaine dans le contexte du changement climatique et de l'atténuation de l'ICU.

1. Données et méthodes

1.1. Lyon : une zone d'étude marquée par une forte diversité morphologique urbaine

La zone d'intérêt choisie dans le cadre de cette étude est l'agglomération de Lyon, et plus particulièrement le cœur urbain de la ville de Lyon ainsi qu'une partie de la ville de Villeurbanne, à la frontière avec le 6^{ème} arrondissement de Lyon (figure 1). Ce territoire présente l'intérêt de regrouper une importante diversité d'occupation du sol en milieu urbain. Il est majoritairement occupé par du tissu urbain continu (50 %) et du tissu urbain dense discontinu (12,3 %), ainsi que par des unités industrielles, commerciales, militaire ou publiques (13,8 %). Avec un peu plus de 1,4 millions d'habitants, cette agglomération de 59 communes est la deuxième plus grande de France derrière Paris. Située au sud-est de la France (45°45'35"N, 4°50'32"E), elle est caractérisée par un climat tempéré chaud et humide, de type Cfa ou Cfb selon la classification de Köppen (Köppen, 1900 ; Kottek *et al.*, 2006) aux influences méditerranéennes, avec des températures élevées au printemps et en été.

1.2. Données acquises par les appareils de mesure et journées sélectionnées

La température de l'air est la variable à estimer partout sur le territoire à partir de plusieurs indicateurs. L'échantillon d'apprentissage de cette variable est obtenu à partir de transects de mesures mobiles à l'aide d'appareils de mesure de très bonne précision, d'après les données constructeur. Le premier matériel utilisé est le EL-USB-1-RGC. Il mesure la température de l'air en continu, avec une précision de +/- 1°C (données constructeur) et un intervalle d'enregistrement minimum de 1 seconde. Le second équipement, le LOG 32, enregistre l'humidité relative et la température de l'air, avec une précision de +/- 0,5°C (données constructeur) et de +/- 3 % (de 40 à 60%) et un intervalle d'enregistrement minimum 2 secondes. De plus, l'instrument EL-USB-1-RGC, présente une réactivité aux variations de températures moins rapides que le LOG 32. Dans le premier cas, l'instrument varie de 12 à 25°C en 25 minutes, alors que pour le second, cette variation est effectuée en 22 minutes.

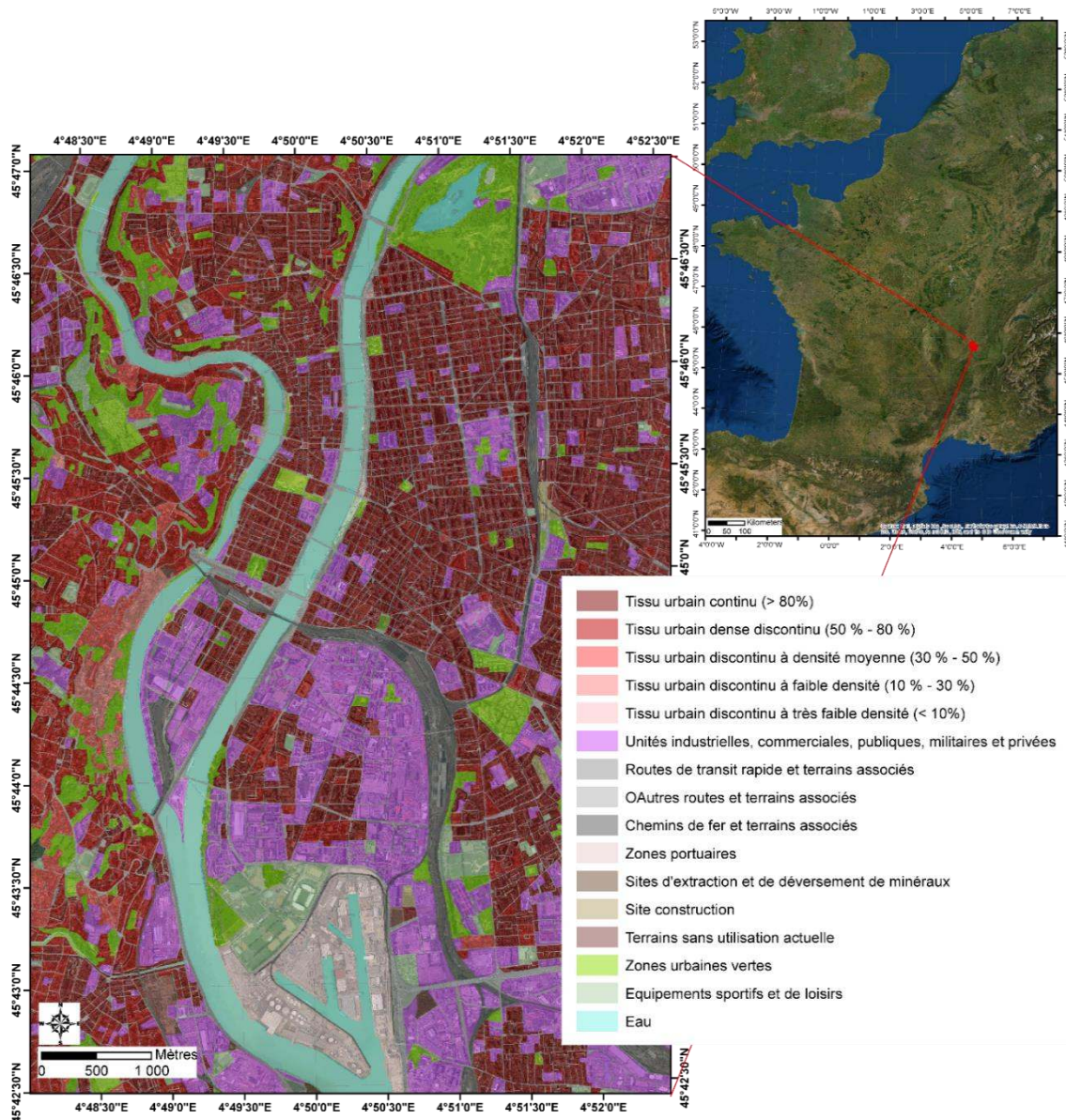


Figure 1. Localisation et occupation du sol du terrain d'étude (source : Urban Atlas 2012 et Data Grand Lyon). *Location and land use of the study area (source: Urban Atlas 2012 and Data Grand Lyon).*

L'humidité relative du LOG 32 varie de 53 à 84,1 %. Cette variation s'effectue dans le même pas de temps que pour sa température de l'air. En outre, les campagnes de mesure ont été associées à un GPS de précision (récepteur GPS/GLONASS haute sensibilité et antenne Quad Helix) afin d'enregistrer la position géographique de la mesure. La précision GPS a été validée par des tests en amont. Ainsi, lorsque la position est stabilisée, l'erreur de localisation est négligeable.

Chacune des localisations de chacun des points pour toutes les campagnes de mesures a été vérifiée

et corrigée si besoin sous système d'information géographique (SIG), par exemple, afin de s'assurer que les points de mesure ne se situent pas sur un toit de bâtiment. En effet, le milieu urbain dense peut venir créer des interférences lors de la géolocalisation de la position. Les rues en centre urbain dense peuvent être encaissées, avec une faible portion de ciel visible.

Par ailleurs, le site de Météo France de la Direction Centre-Est (DIRCE) de Lyon-Bron, situé à une latitude de 45°43'30"N, une longitude de 4°56'12"E et à une altitude de 197 mètres, a servi de

terrain d'étude à une campagne de contrôle qualité des instruments de mesures de la température de l'air et d'humidité relative. Des mesures horaires synchrones aux mesures de la station de Météo

France ont été effectués du 28 juin 2018 à 09h00 au 24 septembre 2018 à 14h00 (figure 2).

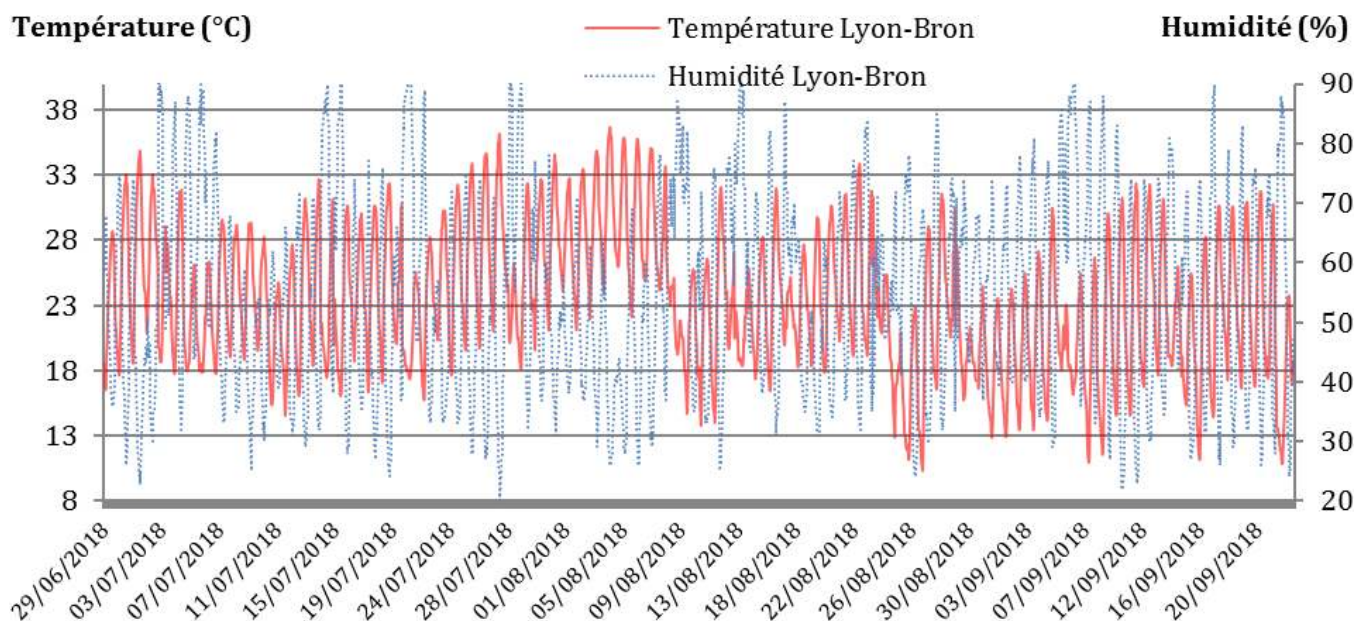


Figure 2. Mesures horaires de température (°C – trait rouge) et d’humidité (% - pointillé bleu) à la station Météo France Lyon-Bron du 28/06/18 au 24/09/18. *Hourly measurements of temperature (°C - red line) and humidity (% - blue dotted line) at the Météo France Lyon-Bron station from 06/28/18 to 09/24/18.*

Les appareils de mesures utilisés dans cette étude se sont avérés être de grande précision puisqu’après cette comparaison sur le site de Météo France sur la période estivale 2018, la corrélation de la température de l’air de ces deux différentes sources

d’acquisition présente un coefficient de corrélation au plus faible de 0,981 pour la température de l’air et de 0,977 pour l’humidité relative mesurée par le LOG 32 (tableau 1).

Tableau 1. Synthèse des coefficients de corrélation, des RMCE et MCE depuis les différents instruments de mesures utilisés par rapport à la station Lyon-Bron de Météo France. *Synthesis of correlation coefficients, RMSE and MSE from the different measurement instruments used in relation to the Météo France station Lyon-Bron.*

| Instruments / Paramètres | LOG 32 | LOG 32 | EL-USB-1-RCG | EL-USB-1-RCG |
|--------------------------|--------------|--------------|--------------|--------------|
| | n°1 | n°2 | n°1 | n°2 |
| Température | MCE : 0,892 | MCE : 0,797 | MCE : 0,516 | MCE : 0,566 |
| | RMCE : 0,944 | RMCE : 0,893 | RMCE : 0,718 | RMCE : 0,752 |
| | R : 0,983 | R : 0,981 | R : 0,989 | R : 0,987 |
| Humidité | MCE : 12,305 | MCE : 11,970 | | |
| | RMCE : 3,507 | RMCE : 3,459 | | |
| | R : 0,977 | R : 0,978 | | |

Les mesures de la température de l'air ont été effectuées lors des jours de passage du satellite Landsat 8 au-dessus de la métropole, par temps clair uniquement, c'est-à-dire une couverture nuageuse inférieure à 10 %. Ces campagnes s'échelonnent entre 2016 et 2018, exclusivement sur la période estivale : le 30 août 2016, le 1^{er} août 2017, 19 juillet 2018 et le 22 juillet 2019. Les conditions météorologiques pour chaque journée sont similaires : l'écart type de la température diurne est

de seulement 0,9°C, de 4,3 % pour l'humidité, de 2,3 m.s⁻¹ pour la vitesse du vent, de 132 degrés pour sa direction et de 3,7 hPa pour la pression. La moyenne des conditions météorologiques respectivement pour ces indicateurs est de 29,3°C, 45,3 %, 8,8 m.s⁻¹, 260,8 degrés et 1016,6 hPa (tableau 2). Respectivement, ces campagnes de mesures ont permis de récolter 573, 300, 393 et 397 points de mesure de la température de l'air et de l'humidité relative (figure 3).

Tableau 2. Paramètres météorologiques des journées d'étude à la station Lyon-Bron à 12h00 (source : Météo France).
Meteorological parameters of the study days at the Lyon-Bron station at 12:00 am (source: Météo France).

| | Température (°C) | Humidité (%) | Vitesse du vent (m/s) | Pression (hPa) | Direction du vent (degrés) |
|------------------------|------------------|--------------|-----------------------|----------------|----------------------------|
| 30 août 2016 | 27,7 | 46 | 9 | 1017,8 | 350 |
| 01 août 2017 | 29,4 | 52 | 10 | 1012,2 | 34 |
| 19 juillet 2018 | 29,8 | 42 | 5 | 1014,2 | 309 |
| 22 juillet 2019 | 30,1 | 41 | 11 | 1022 | 10 |
| Moyenne | 29,3 | 45,3 | 8,8 | 1016,6 | 260,8 |
| Ecart type | 0,9 | 4,3 | 2,3 | 3,7 | 132,0 |
| Minimum | 27,7 | 41 | 5 | 1012,2 | 34 |
| Maximum | 30,1 | 52 | 11 | 1022 | 350 |

En outre, les campagnes de mesures de température de l'air s'échelonnent parfois sur plusieurs dizaines de minutes. Il a donc été nécessaire d'apporter une correction à partir d'une équation polynomiale élaborée en fonction de l'évolution des températures de la journée enregistrées à un pas de temps de 10 min jusqu'à la fin des campagnes de mesures en milieu d'après-midi. Cette étape avant le traitement des données est primordiale et permet de ramener toutes ces mesures de température de l'air à l'heure la plus chaude la journée (Brandsma et Wolters, 2012).

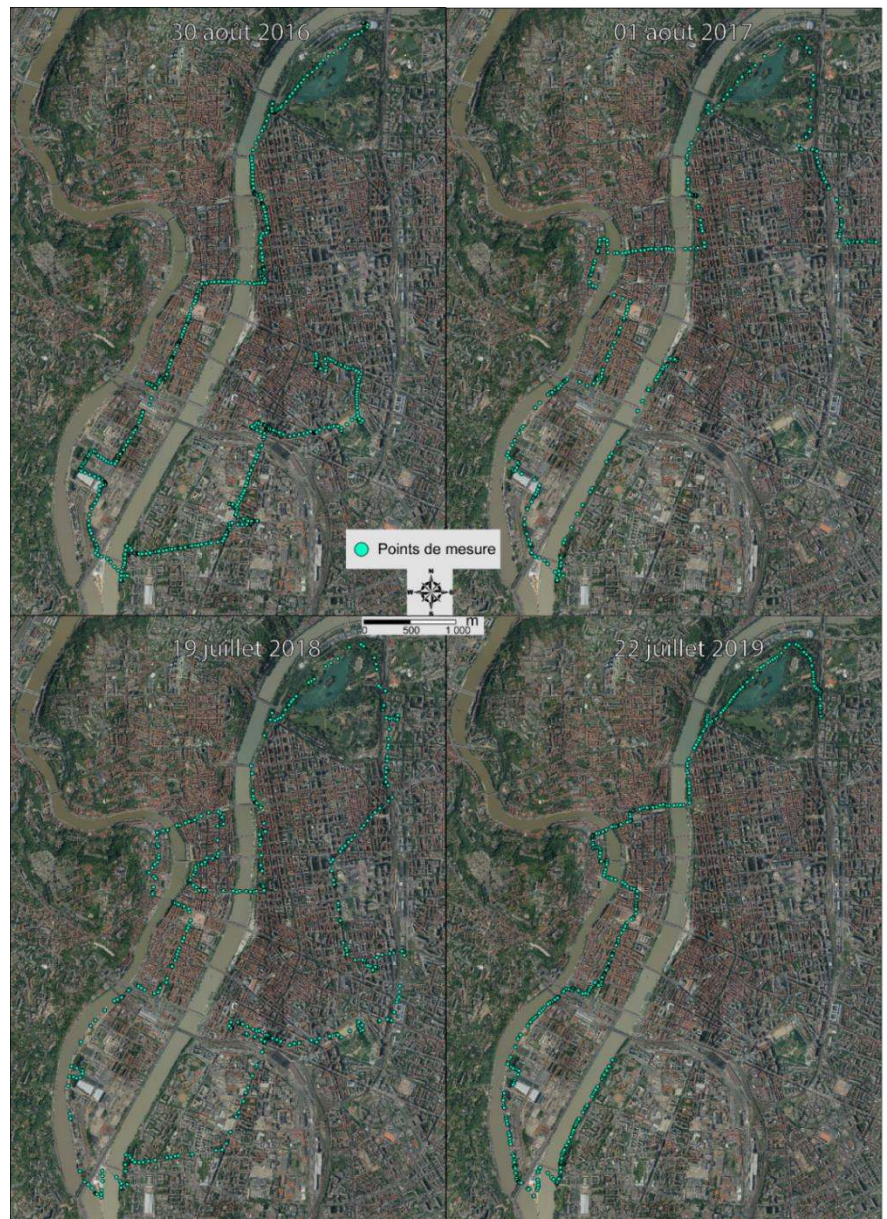
1.3. Les descripteurs morphologiques intervenant dans l'estimation de la température de l'air

Les modifications des modes d'occupation du sol liées à la fabrique urbaine contribuent à structurer spatialement le paysage urbain, qui, par ailleurs, ont une influence sur la transmission et la balance énergétique (Xie *et al.*, 2011 ; Liu *et al.*, 2018). Ces changements sont considérés comme

étant une cause directe de la formation des ICU (Chen *et al.*, 2012 ; Yang *et al.*, 2017). Ainsi, la mise en relation des variations de la température de l'air avec l'utilisation des terres et l'occupation des sols est pertinente.

Dans cette étude, trente-huit variables explicatives contribuent à estimer la température de l'air sur la zone d'étude, comme utilisées dans plusieurs autres études sur différents territoires (Sobrino *et al.*, 2004 ; Jin et Sader, 2005 ; Chen *et al.*, 2006 ; Liu et Zhang, 2011 ; Nguyen *et al.*, 2016). Elles appartiennent à diverses catégories telles que des données climatiques issues de la télédétection, des variables topographiques, des indices de végétation, de la présence d'eau, d'humidité, de sol nu, de bâti, de rayonnement, de morphologie urbaine ou encore de proximité à diverses occupations du sol (tableau 3). Les sources d'acquisition sont multiples et proviennent du satellite Landsat 8, de points LiDAR et d'autres produits cartographiques téléchargés depuis la plateforme d'open data du Grand Lyon.

Figure 3. Points de mesure de la température de l'air en après-midi pour le 30 août 2016 (haut-gauche), 01 août 2017 (haut-droite), 19 juillet 2018 (bas-gauche) et le 22 juillet 2019 (bas-droite – source : Data Grand Lyon). *Afternoon air temperature measurement points for August 30th, 2016 (top-left), August 01st, 2017 (top-right), July 19th, 2018 (bottom-left) and July 22nd, 2019 (bottom-right - source: Data Grand Lyon).*



Ces descripteurs morphologiques sont acquis à une précision spatiale pouvant descendre à l'échelle du centimètre. De ce fait, les informations relevées sont denses et nous permettent d'acquérir un état du milieu urbain précis dans le but de modéliser la température de l'air.

1.4. La procédure statistique suivie

1.4.1. Une zone tampon explicative qui varie en fonction de l'indicateur

Cette étude a pour but de modéliser la température de l'air à l'aide de régressions linéaires,

multiples et *partial least square*, et de régression non linéaire par la forêt d'arbres aléatoires, à partir des prédicteurs sélectionnés. Dans un premier temps, l'échelle présentant une meilleure corrélation entre température de l'air et variables explicatives est retenue pour chaque indicateur à partir d'une analyse par tampons de proximité (de 5 à 1000 m; figure 4). Ainsi, la zone tampon sélectionnée varie pour les indicateurs de présence de végétation, d'eau, d'humidité, de sol nu et de bâti, les indices de rayonnement, de proximité aux occupations du sol, de morphologie urbaine et enfin les données climatiques (tableau 4).

Tableau 3. Liste des descripteurs morphologiques intervenant dans l'estimation de la température de l'air. *List of morphological descriptors used to estimate air temperature.*

| | Variables | Source d'acquisition | | Variables | Source d'acquisition | |
|--|--|---|-------------------------|--|----------------------|-----------|
| Données climatiques issues de la télédétection | Températures de surface (°C) | Landsat 8 | Indices de bâti | NDBI <i>Normalized Difference Built-Up Index</i> | Landsat 8 | |
| | UTFVI | Landsat 8 | | UI <i>Urban Index</i> | Landsat 8 | |
| | Durée d'ensoleillement de la journée d'étude (h) | Données LiDAR et modélisation par ESRI ARCGIS | | IBI <i>Index-based Built-Up Index</i> | Landsat 8 | |
| | Radiation reçue pour la journée d'étude (WH/m²) | Données LiDAR et modélisation par ESRI ARCGIS | | Densité de bâti | LiDAR | |
| Indices de végétation | NDVI <i>Normalized Difference Vegetation Index</i> | Landsat 8 | Topographie | Modèle numérique d'élévation | LiDAR | |
| | SAVI | Landsat 8 | | Pente (°) | Data Grand Lyon | |
| | EVI | Landsat 8 | | Exposition | LiDAR | |
| | Tasseled cap greenness (GVI) | Landsat 8 | | Rugosité | Data Grand Lyon | |
| | Densité de végétation basse | LiDAR | | Sky View Factor | LiDAR | |
| | Densité de végétation moyenne | LiDAR | | Variation de la hauteur des bâtiments (écart-type : Building Height) | Data Grand Lyon | |
| | Densité de végétation haute | LiDAR | | Distance aux voies ferrées | Data Grand Lyon | |
| Indices de présence d'eau | MNDWI <i>Modified Normalized Difference Water Index</i> | Landsat 8 | Morphologie urbaine | Distance aux points d'intérêt touristique | Data Grand Lyon | |
| | NDWI | Landsat 8 | | Distance aux bouches de métros | Data Grand Lyon | |
| Indices d'humidité | Tasseled cap Wetness | Landsat 8 | Occupations du sol | Distance aux fontaines | Data Grand Lyon | |
| | NDMI <i>Normalized Difference Moisture Index</i> | Landsat 8 | | Zones en eau | Data Grand Lyon | |
| Indices de sol nu | NDBaI <i>Normalized Difference Bareness Index</i> | Landsat 8 | | Occupations du sol | Densité de sol nu | LiDAR |
| | BI <i>Bare Soil Index</i> | Landsat 8 | | | Radiance spectrale | Landsat 8 |
| | EBBI <i>Enhanced Built-Up and Bareness Index</i> | Landsat 8 | Emissivité | | Landsat 8 | |
| | Densité de sol nu | LiDAR | Tasseled Cap Brightness | | Landsat 8 | |
| Indices de rayonnement | Radiance spectrale | Landsat 8 | | | | |
| | Emissivité | Landsat 8 | | | | |
| | Tasseled Cap Brightness | Landsat 8 | | | | |

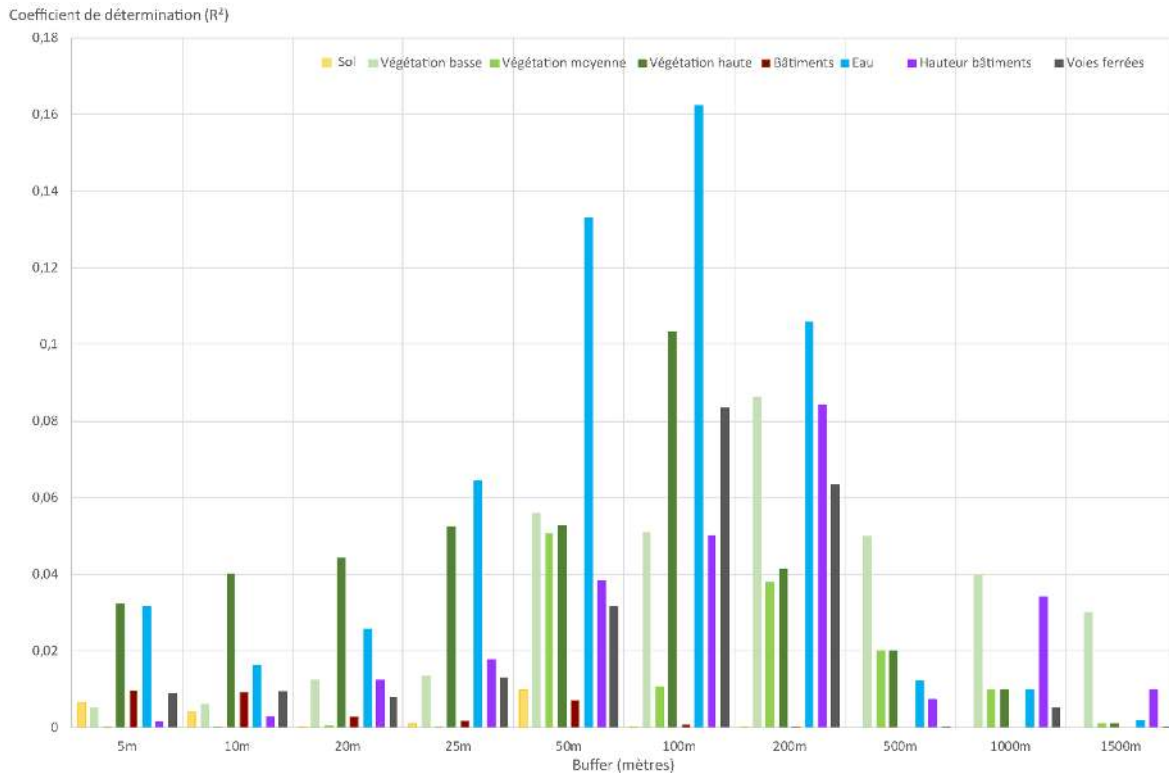


Figure 4. Exemple de variation de la corrélation entre prédicteur et température de l'air en fonction de l'échelle d'étude. *Example of the variation of the correlation between predictor and air temperature according to the study scale.*

1.4.2. Trois méthodes de régression complémentaire dans la modélisation

Trois méthodes de régression de la modélisation de la température de l'air sont comparées dans cette étude. Il s'agit de deux régressions, linéaires multiples (Chen *et al.*, 2016 ; Mira *et al.*, 2017 ; Alonso et Renard, 2019) et par les moindres carrés partiels (Tenenhaus *et al.*, 2005 ; Guo *et al.*, 2020), et d'une non linéaire, la forêt d'arbres aléatoire (Breiman, 1996 ; Di Paola *et al.*, 2018 ; Sekulić *et al.*, 2020). L'objectif est de choisir la meilleure régression pour cette modélisation. Cette évaluation s'effectue essentiellement en comparant les coefficients de détermination et les racines de la moyenne des carrés des erreurs (RMCE) obtenues pour les échantillons. Les conditions d'utilisation pour chacune des régressions ont aussi été vérifiées.

La régression linéaire multiple (RLM) est une méthode de modélisation de la donnée qui nécessite plusieurs étapes statistiques avant son application (Dempster, 1969). Dans un premier temps, il est nécessaire de vérifier la distribution normale des

séries du jeu de données à l'aide du test de Shapiro-Wilk (s'applique aux échantillons de moins de 5000 observations ; Shapiro et Wilk, 1965). Ce test a été invalidé, c'est pourquoi la matrice de corrélation de Spearman a été utilisée. Elle permet de ne pas inclure les variables redondantes dans le modèle de régression. Un des deux indicateurs pour lequel la paire présente un $|r| > 0,7$ dans la matrice de corrélation de Spearman ainsi qu'un VIF (*Variable Importance for the Projection*) > 5 a été supprimé (OCDE, 2008 ; Dormann *et al.*, 2013). De plus, une validation croisée *holdout* a été réalisée en raison de sa capacité à détecter les *over-fitting* de la régression multiple (80 % de données d'apprentissage, 20 % de données de validation) (Shandas *et al.*, 2019).

D'une manière complémentaire à la RLM, la régression *partial least square* est une méthode s'appliquant lorsqu'un grand nombre de variables explicatives est présent et que celles-ci sont susceptibles de présenter de fortes colinéarités entre elles (Dempster, 1969).

Tableau 4. Zones tampons sélectionnées pour chaque indicateur explicatif. *Buffer zones selected for each explanatory indicator.*

| | Variables | Zone tampon (m) | | Variables | Zone tampon (m) |
|--|-------------------------------|-----------------------------|----------------------------------|-------------------------|-------------------------|
| Données climatiques issues de la télédétection | Températures de surface (°C) | 500 | Indices de sol nu | NDBal | 1000 |
| | UTFVI | 500 | | BI | 50 |
| Indices de végétation | NDVI | 1000 | | Indices de bâti | EBBI |
| | SAVI | 1000 | Densité de sol nu | | 50 |
| | EVI | 50 | Morphologie urbaine | NDBI | 1000 |
| | Tasseled cap greenness | 1000 | | UI | 1000 |
| | Densité de végétation basse | 200 | | IBI | 500 |
| | Densité de végétation moyenne | 50 | | Densité de bâti | 5 |
| | | Densité de végétation haute | 100 | | STD Building Height |
| Indices de présence d'eau | MNDWI | 500 | Indices de rayonnement | Radiance spectrale | 1000 |
| | NDWI | 500 | | Emissivité | 500 |
| Indices d'humidité | Tasseled cap Wetness | 50 | | | Tasseled Cap Brightness |
| | NDMI | 1000 | Proximité aux occupations du sol | Densité de zones en eau | 100 |

Ainsi, cette méthode nous permet de modéliser et prédire des valeurs de la température de l'air en fonction d'une combinaison linéaire de plusieurs variables explicatives quantitatives (ou qualitatives) en s'affranchissant des contraintes de la régression linéaire pour la distribution et le nombre de variables incluses. Il n'est donc pas nécessaire de supprimer les variables colinéaires. Le modèle donne une valeur pour chaque *VIP*. Une variable explicative est considérée comme importante lorsque le *VIP* est supérieur à 0,8 (Wold *et al.*, 2000). Un coefficient normalisé est alors généré pour chacune d'elles (Tenenhaus *et al.*, 2005).

Le troisième type de régression testé est la forêt aléatoire de régression. Il s'agit d'un modèle prédictif qui met en œuvre des arbres de décision binaire (Breiman *et al.*, 1984). A partir d'un échantillon d'observation, la méthode *bagging* va générer plusieurs possibilités avant de n'en sélectionner qu'une seule. Cette technique d'apprentissage automatique (Hastie *et al.*, 2009) est articulée autour des arbres *Classification And Regression Trees* (CART). Ces derniers sont construits à partir de différents échantillons *bootstrap*, choisis au hasard avec une remise aléatoire, afin d'obtenir, après agrégation, un

ensemble de prédicteurs robuste et performant de la température de l'air (Breiman, 1996). L'importance de chaque variable est calculée par l'accroissement moyen de l'erreur d'un arbre dans la forêt, c'est-à-dire lorsque les valeurs de chacune des variables est échangée au hasard dans les échantillons *out-of-bag* (OOB). Les variables utilisées dans la modélisation de la température de l'air par la régression non linéaire de la forêt d'arbres aléatoires sont issues de la sélection des régressions linéaires multiples pour chacune des journées. La forêt aléatoire de classification et de régression présente l'avantage de réduire le bruit blanc et donc potentiellement d'améliorer les coefficients de corrélation et les RMCE déjà obtenus par régression linéaire multiple.

En outre, la régression Lasso n'est pas utilisable dans cette étude. Elle n'est d'usage que lorsque le nombre de variables prédictives est supérieur au nombre d'observations (Tibshirani, 1996 ; Reid *et al.*, 2016). Or, ici, le nombre d'observations est beaucoup plus élevé que le nombre de prédicteurs.

1.4.3. Contrôle de qualité sur les modélisations par l'identification spatiale des clusters des erreurs

L'autocorrélation spatiale de la différence entre la température de l'air modélisée et la température de l'air mesurée par les mesures mobiles est analysée, d'une part, à l'aide de l'indicateur d'association spatiale locale *Anselin Local Moran I* LISA (Anselin, 1995), et, d'autre part, grâce au degré de regroupement des valeurs d'intensité élevée et faible par le Getis Ord General G (G_i^* ; Getis et Ord, 1992 et 1996).

Le LISA permet de regrouper, pour des résultats statistiquement significatifs ($p < 0,05$), la similitude d'une unité spatiale avec ses voisins. Il permet d'identifier des agrégats spatiaux d'entités avec des valeurs hautes ou basses ainsi que des points spatiaux aberrants. Une représentation cartographique qui montre un type de cluster pour chaque entité statistiquement significative est ainsi obtenue. Sous système d'information géographique (SIG), il est distingué à la fois, un groupe de valeurs élevées (HH) statistiquement significatif, un groupe de valeurs faibles (LL), une valeur aberrante dans

laquelle une valeur élevée est entourée principalement de valeurs faibles (HL) et une valeur aberrante dans laquelle une valeur faible est entourée principalement de valeurs élevées (LH).

L'application locale de la statistique G générale est une méthode particulière de celle de Getis Ord G_i^* . Elle est utilisée pour identifier des grappes spatiales statistiquement significatives ($p < 0,05$) de haute et basse intensité (ou analyse de point chaud). Ainsi, pour les scores Z positifs, plus ils sont élevés, plus le groupe de valeurs de forte intensité est fort (erreur surestimant la température de l'air). En revanche, plus les scores Z négatifs sont faibles, plus le groupe de valeurs de faible intensité est élevé (erreur sous-estimant la température de l'air).

2. Résultats

2.1. La modélisation par régression linéaire multiple

Après avoir supprimé les variables colinéaires, pour chaque journée, les prédicteurs intervenant dans la modélisation de la température de l'air permettent d'obtenir des coefficients de détermination significatifs. Ces derniers varient de 0,60 pour le 22 juillet 2019 à 0,89 pour le 30 août 2016 avec respectivement des RMCE de seulement 0,96°C et 0,23°C. Par ailleurs, chaque variable retenue dans le modèle est caractérisée par un coefficient normalisé qui correspond au poids de cette variable explicative. Ce poids varie en fonction des journées d'étude (figure 5).

Par exemple, pour le 19 juillet 2018, les variables contribuant à avoir un impact positif sur le modèle sont la proximité aux bouches de métros, le BI, la longitude, le TCT Wetness, la densité de végétation basse et le NDBaI. Les variables intervenant négativement dans le modèle sont, quant à elles, le Sky View Factor, la densité de végétation haute, la proximité aux points d'intérêt touristiques et la densité de sol anthropisé¹. A partir de l'équation obtenue, il est ainsi possible de modéliser la température de l'air en tout point (figure 6).

¹ Nuage de points LIDAR 2015 de la Métropole de Lyon : <https://www.data.gouv.fr/fr/datasets/nuage-de-points-lidar-2015-de-la-metropole-de-lyon/>

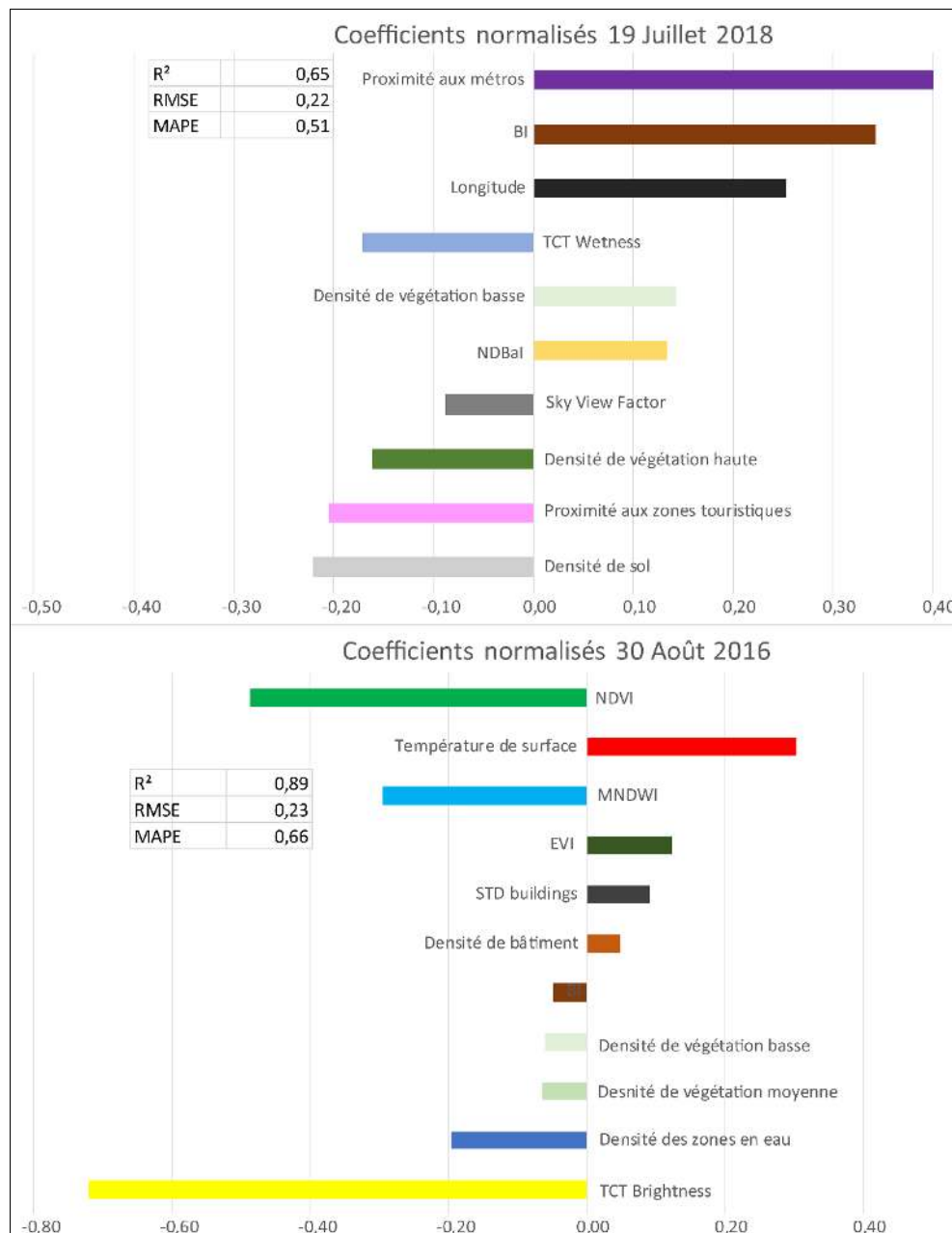


Figure 5. Exemple de poids des variables retenues le 30 août 2016 et le 19 juillet 2018. *Example of weights for variables selected on August 30th, 2016 and July 19th, 2018*

2.2. La modélisation par régression *partial least square*

La modélisation par régression *partial least square* ne révèle pas une grande cohérence dans la prédiction de la température de l'air puisque le coefficient de détermination moyen pour l'ensemble des quatre journées d'étude n'est que de 0,62 avec un maximum de 0,79 pour le 30 août 2016 et un minimum de 0,53 pour le 22 juillet 2019 (tableau 5). En outre, un grand nombre de variables

explicatives sont retenues, au maximum 26 pour la journée du 22 juillet 2019. Certaines variables influencent à la fois positivement et négativement le modèle en fonction de la journée. Par exemple, le MNDWI a un impact positif important pour le 30 août 2016 et négatif pour le 1^{er} août 2017, le 19 juillet 2018 et le 22 juillet 2019. De ce fait, les résultats de modélisation de la température de l'air sont bien moins pertinents que par la régression linéaire multiple.

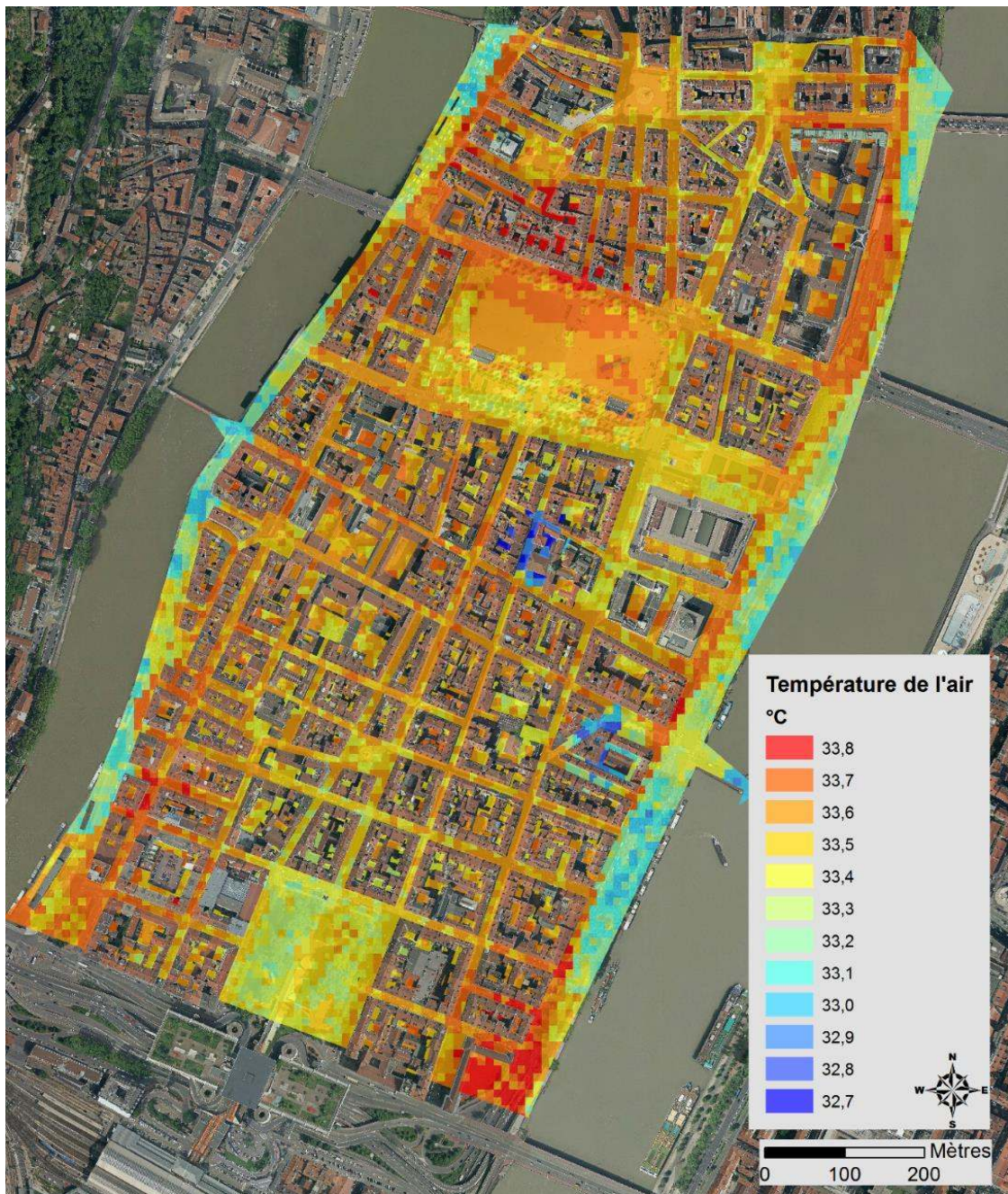


Figure 6. Modélisation de la température de l'air dans le centre urbain dense de Lyon le 19 juillet 2018 (source : Data Grand Lyon). *Air temperature modelling in the dense urban center of Lyon on July 19th, 2018 (source: Data Grand Lyon)*

2.3. La modélisation par la forêt aléatoire de classification et de régression

Pour les quatre journées d'étude, les coefficients de détermination obtenus sont forts : 0,98 pour le 30 août 2016, 0,96 pour le 1^{er} août 2017, 0,95 pour le 19 juillet 2018 et 0,92 pour le 22 juillet 2019 (tableau 6). Ainsi, en moyenne, un coefficient de détermination de 0,95 est obtenu, avec un RMCE de

seulement de 0,17°C et une erreur Out Of Bag (OOB) de 0,05.

En outre, la mesure d'importance pour chacune des variables est mesurée par l'accroissement moyen de l'erreur d'un arbre dans la forêt lorsque les valeurs observées de cette variable sont permutées au hasard dans les échantillons *Out Of Bag* (OOB ; figure 7).

Tableau 5. Paramètres statistiques des trois variables explicatives retenues dans la modélisation de la température de l'air par régression partial least square. *Statistical parameters of the three explanatory variables retained in the modelling of air temperature by partial least square regression.*

| Date | R ² | MCE | RMCE | Variables | Paramètre du modèle en valeur absolue | Impact sur le modèle |
|------------|----------------|------|------|------------|---------------------------------------|----------------------|
| 30/08/2016 | 0,79 | 0,11 | 0,33 | LST | 0,0675 | Négatif |
| | | | | NDVI | 1,71 | Positif |
| | | | | MNDWI | 4,53 | Positif |
| 01/08/2017 | 0,77 | 0,03 | 0,18 | BI | 0,58 | Positif |
| | | | | NDMI | 0,51 | Négatif |
| | | | | NDBI | 0,51 | Positif |
| 19/07/2018 | 0,37 | 0,09 | 0,07 | Emissivité | 2,1128 | Négatif |
| | | | | Longitude | 1,3906 | Positif |
| | | | | NDBal | 1,2262 | Positif |
| 22/07/2019 | 0,53 | 1,13 | 1,06 | Emissivité | 7,4782 | Positif |
| | | | | BI | 3,0472 | Positif |
| | | | | NDBal | 2,5931 | Positif |
| 22/07/2019 | 0,53 | 1,13 | 1,06 | NDVI | 0,4544 | Négatif |
| | | | | UTFVI | 0,2812 | Positif |
| | | | | NDBI | 0,2637 | Positif |
| Moyenne | 0,62 | 0,34 | 0,41 | | | |

Tableau 6. Récapitulatif des coefficients de détermination, de l'erreur Out Of Bag et de racine de la moyenne des carrés des erreurs de la modélisation par forêt aléatoire de classification et de régression. *Summary of the coefficients of determination, Out Of Bag error and root mean square error of random forest regression modelling.*

| Date | R ² | Out Of Bag | RMCE |
|------------|----------------|------------|------|
| 30/08/2016 | 0,98 | 0,0071 | 0,08 |
| 01/08/2017 | 0,96 | 0,0045 | 0,07 |
| 19/07/2018 | 0,95 | 0,0071 | 0,08 |
| 22/07/2019 | 0,92 | 0,19 | 0,44 |
| Moyenne | 0,95 | 0,05 | 0,17 |

3. Discussion

3.1. Spatialisation de l'erreur

L'erreur de modélisation constatée est minime pour les modélisations par régression linéaire multiple et par la forêt d'arbres aléatoires. Pour l'ensemble des journées d'étude, la médiane par régression linéaire multiple est de 0,02°C et par

forêt aléatoire de classification et de régression, de 0,002°C (σ respectivement de 0,44 et 0,17 ; tableau 7). Contrairement à la proximité des valeurs de la médiane et de la moyenne par ces deux méthodes de modélisation, l'entendue est plus forte pour la régression linéaire multiple que pour la forêt aléatoire de classification et de régression (figure 8).

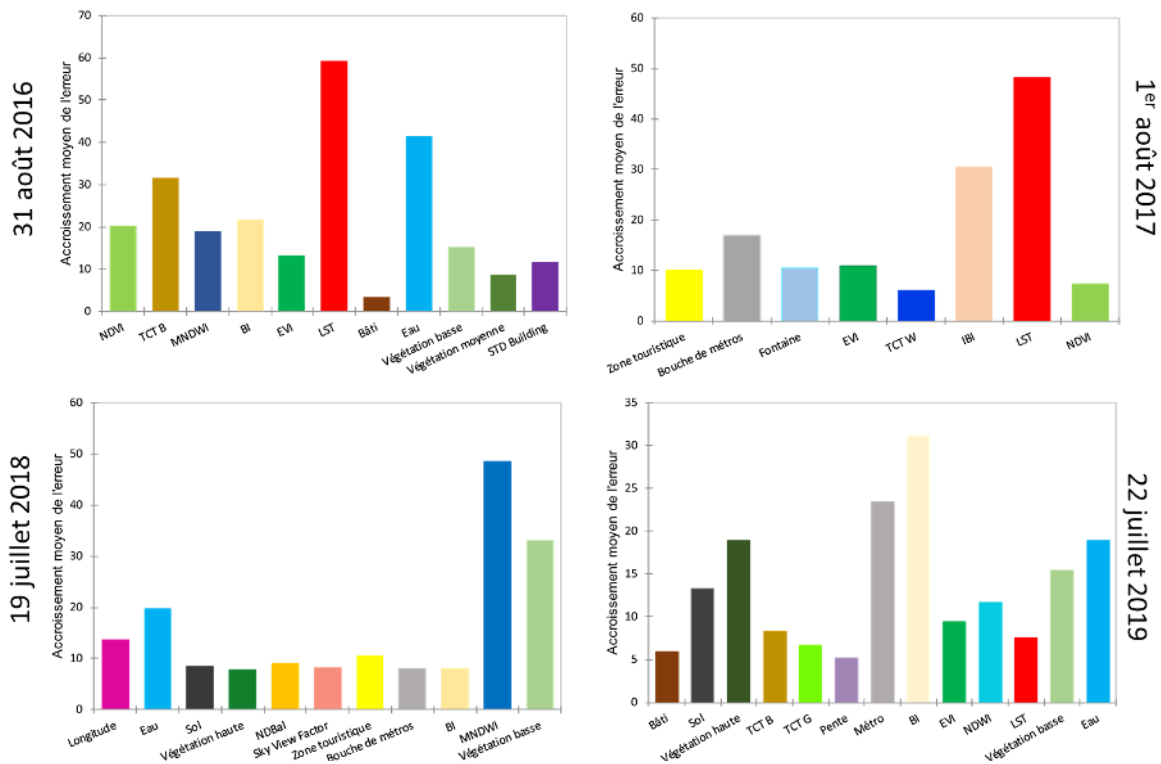
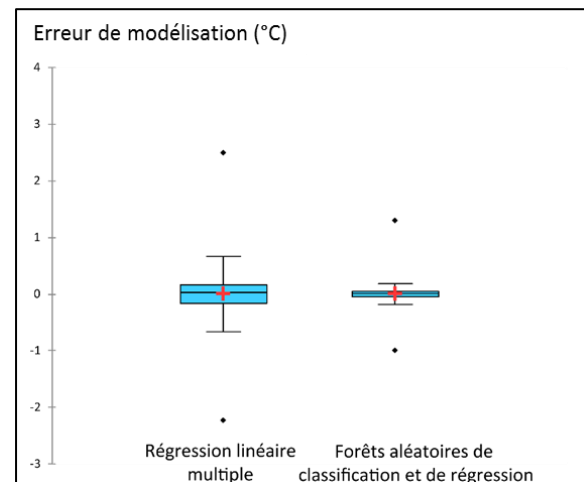


Figure 7. Evolution de l'importance des variables retenues dans la modélisation par forêt aléatoire de classification et de régression. *Evolution of the importance of the variables retained in the random forest regression modeling.*

Tableau 7. Statistiques descriptives de l'erreur de modélisation par régression linéaire multiple (RLM) et par forêt aléatoire de classification et de régression (RDF). *Descriptive statistics of the modeling error by multiple linear regression (MLR) and by random forest regression (RDF).*

| | RLM | RDF |
|----------------------|-------|-------|
| Minimum | -2,23 | -0,99 |
| Maximum | 2,50 | 1,29 |
| 1er Quartile | -0,17 | -0,05 |
| Médiane | 0,02 | 0,002 |
| 3ème Quartile | 0,17 | 0,05 |
| Moyenne | 0,01 | 0,003 |
| Variance | 0,19 | 0,03 |
| Ecart-type | 0,44 | 0,17 |

Figure 8. Représentation par boîte à moustache de l'erreur de modélisation par régression linéaire multiple et par forêt aléatoire de classification et de régression. *Box plot representation of the modelling error by multiple linear regression and by random forest regression.*



Lorsque l'on s'intéresse à la localisation des erreurs de la modélisation de la température de l'air entre la méthode par régression linéaire multiple et la forêt d'arbres aléatoire, on observe des similarités entre les deux. Les modèles surestiment les températures de l'air vers les zones en eaux sur

Confluence (sud de la presqu'île) et à proximité de la gare Perrache. Ils sous-estiment la température de l'air dans les rues encaissées, à proximité des espaces verts et au sud de la rive gauche du Rhône (figure 9).

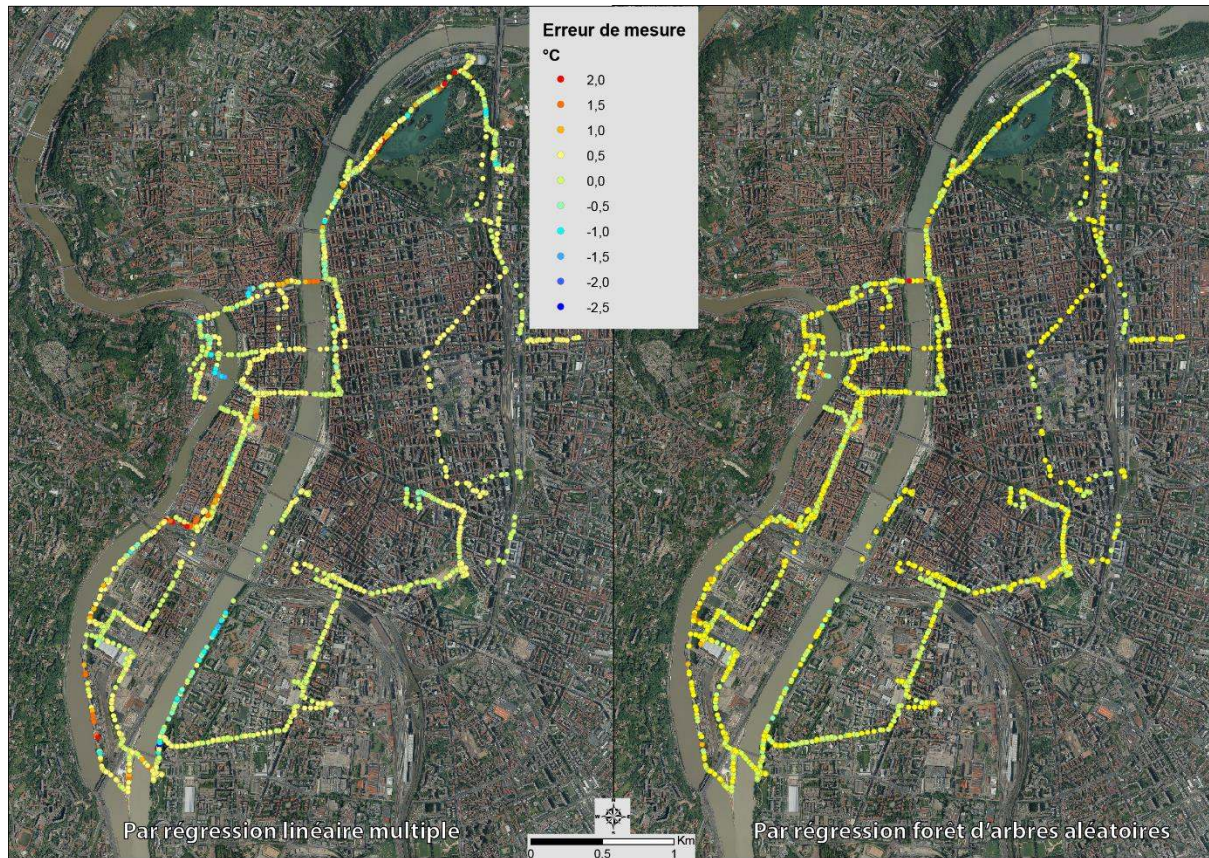


Figure 9. Localisation de l'erreur de mesure modélisée de la température de l'air par régression linéaire multiple (gauche) et par forêt d'arbres aléatoires (droite) pour l'ensemble des journées d'étude (source : Data Grand Lyon). *Localization of the modelled measurement error of the air temperature by multiple linear regression (left) and by random forest regression (right) for all the study days (source: Data Grand Lyon).*

Si on analyse la localisation de ces erreurs journée par journée, on remarque que pour le 30 août 2016, le modèle par régression linéaire multiple surestime les températures de l'air à proximité des zones en eau, sur les ponts et sud de la rive gauche du Rhône. A l'inverse, il sous-estime cette grandeur physique sur les espaces ouverts comme la place Bellecour. En ce qui concerne les journées du 1^{er} août 2017, 19 juillet 2018 et 22 juillet 2019, le modèle surestime la température de l'air également à proximité des cours d'eau et sur

les ponts mais également sur les espaces ouverts. Les rues encaissées semblent, en revanche, plus fraîches selon le modèle qu'avec les mesures mobiles *in situ* (figure 10). Le même constat peut être visible sur la modélisation par la forêt d'arbre aléatoire (figure 11). En outre, on observe une surestimation de la température de l'air à proximité des espaces verts du Parc de la Tête d'Or pour les journées du 30 août 2016, 19 juillet 2018 et 22 juillet 2019.

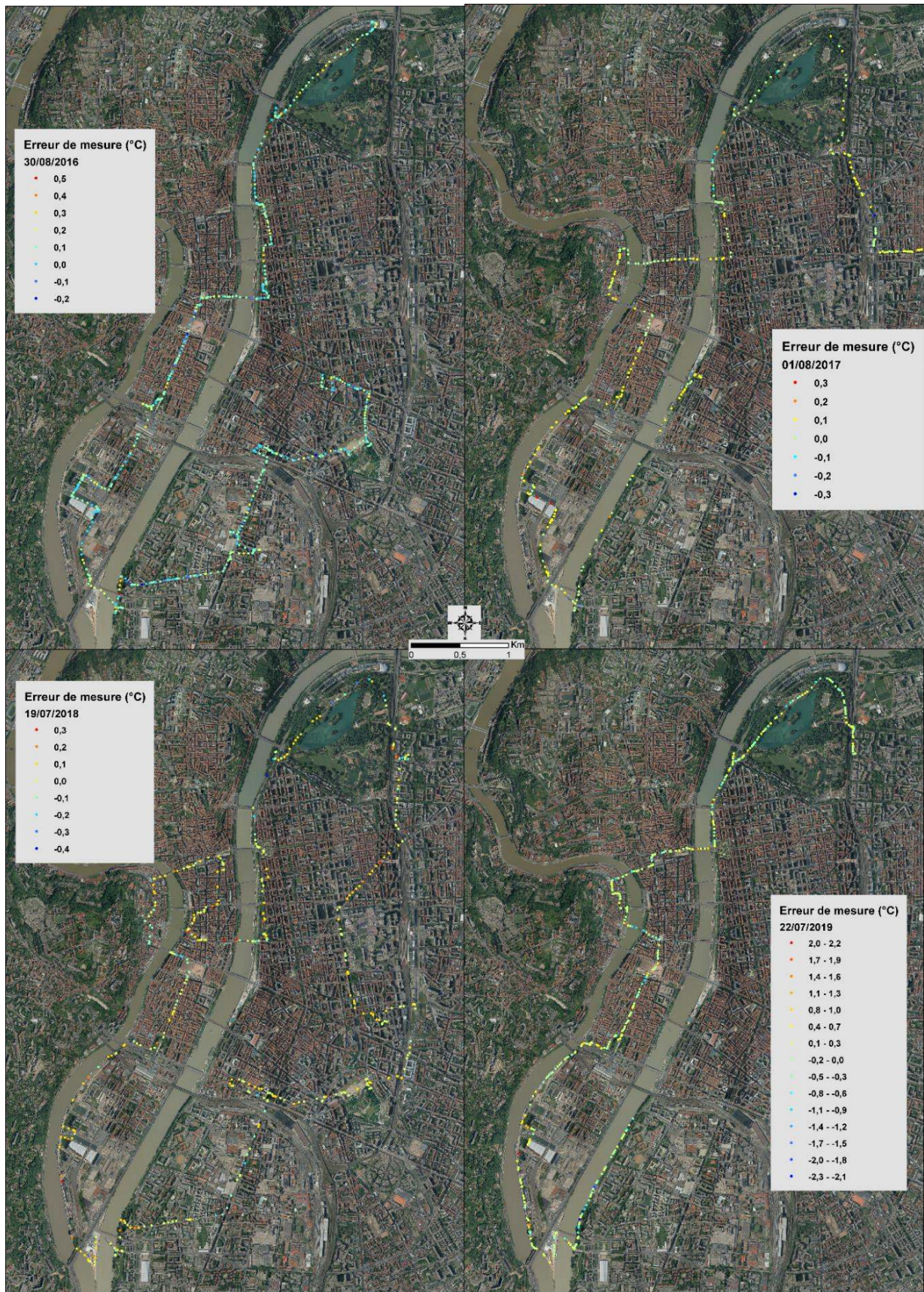


Figure 10. Localisation de l'erreur de mesure de la modélisation de la température de l'air par régression linéaire multiple. *Localization of the measurement error of air temperature modelling by multiple linear regression.*

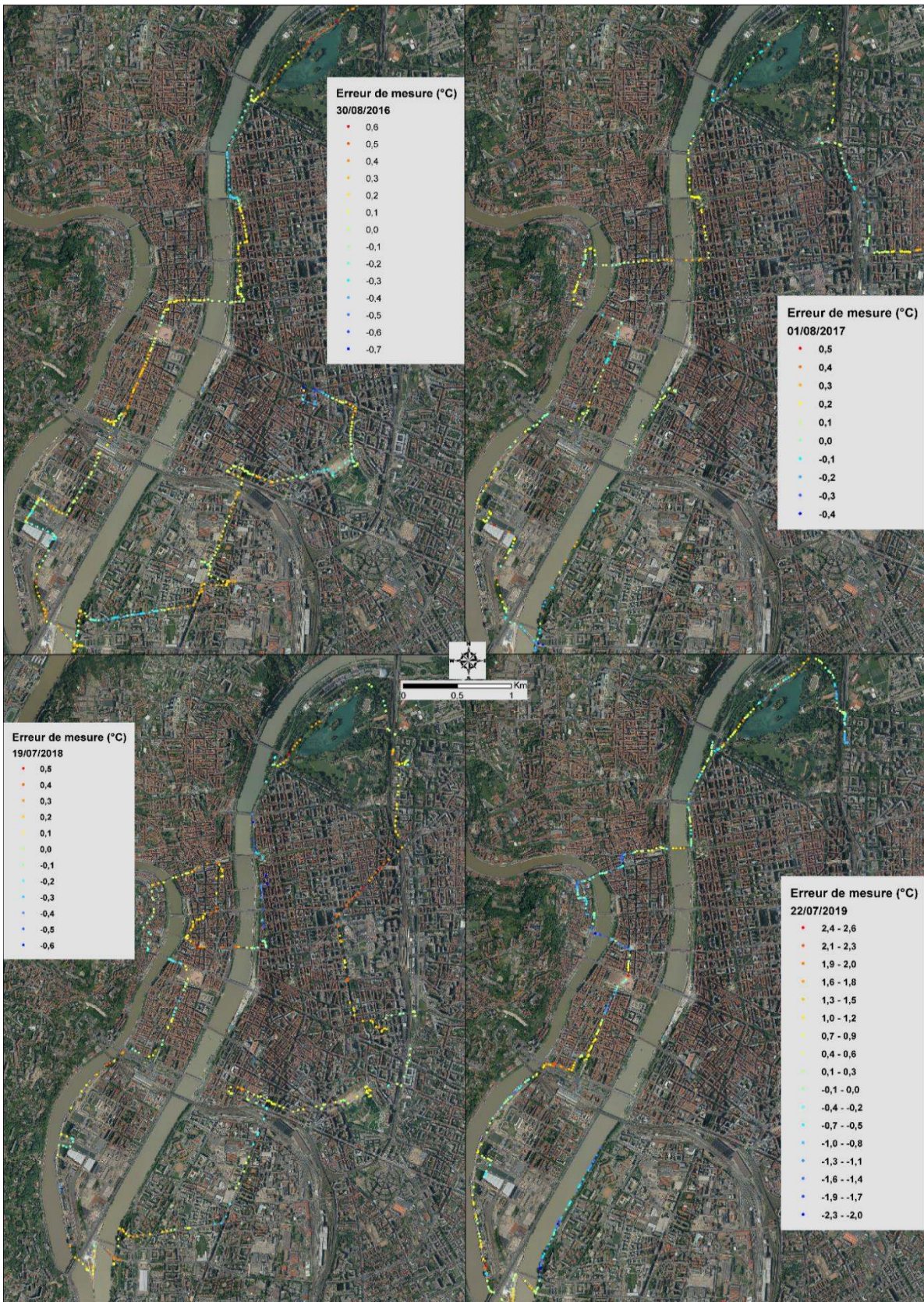


Figure 11. Localisation de l'erreur de mesure de la modélisation de la température de l'air par la forêt aléatoire de classification et de régression. *Localization of the measurement error of the air temperature modelling by the random forest regression.*

3.2. Regroupement des erreurs similaires

Les regroupements spatiaux des valeurs statistiquement proches des différences entre la température de l'air modélisée et la température de l'air mesurée sont évalués grâce au LISA (figure 12) et au Gi (figure 13). Entre les deux méthodes de régression (linéaire multiple et forêt d'arbres aléatoires), des similitudes dans la localisation des types de regroupement de l'erreur par le LISA et le Gi* sont observées.

Dans un premier temps, par la méthode LISA, des regroupements d'erreurs faibles (LH), de sous-estimation du modèle par rapport aux valeurs mesurées, sont identifiables sur la rive gauche du Rhône, dans les rues encaissées de la presqu'île et du Vieux Lyon et sur les ponts. Les zones qui présentent une valeur forte (HL), c'est-à-dire une

surestimation du modèle, sont observables à proximité de la gare Perrache, dans la zone de Confluence et à proximité des espaces verts du parc de la Tête d'Or (figure 12).

Dans un second temps, les regroupements des erreurs de sous-estimation et de surestimation de la modélisation de la température de l'air par rapport à celle mesurée par la méthode du Gi* sont localisés dans des zones similaires au LISA. Ces espaces récurrents pour les valeurs de score z négatifs statistiquement faibles sont les rues encaissées de la presqu'île et du Vieux Lyon et le sud de la rive gauche du Rhône. Les valeurs de score z positifs statistiquement élevées sont le quartier de la gare Perrache et Confluence, la proximité aux espaces verts du parc de la Tête d'Or mais aussi le pont Morand (figure 13).



Figure 12. LISA des différences entre les températures de l'air modélisées et mesurées pour l'ensemble des journées d'étude. *LISA of the differences between modelled and measured air temperatures for all study days.*

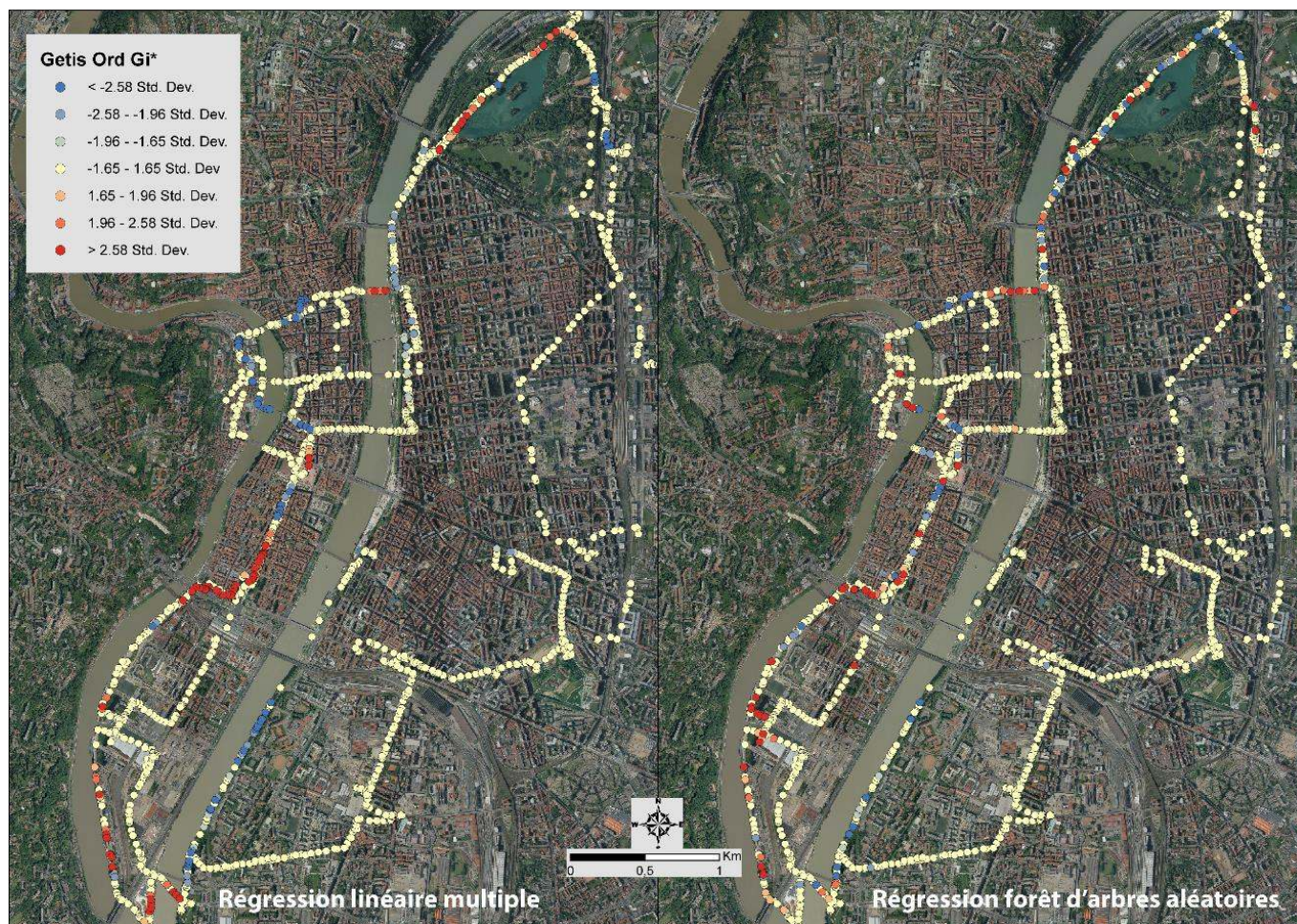


Figure 13. Gi^* des différences entre les températures de l'air modélisées et mesurées pour l'ensemble des journées d'étude. *Gi^* of the differences between modelled and measured air temperatures for all study days.*

Conclusion

L'objectif de cette étude était d'identifier la méthode de régression la plus appropriée et qui présente les meilleurs résultats pour la modélisation de la température de l'air en milieu urbain à partir d'un grand nombre de variables explicatives de nature diverse. L'intégration de ces prédicteurs dans des régressions linéaires multiples puis dans des régressions non-linéaire de la forêt d'arbres aléatoires affiche des résultats très satisfaisants.

La part de la variance expliquée par régression linéaire multiple dans la modélisation des températures de l'air pour chaque journée d'étude est dans l'ensemble forte avec des coefficients de détermination allant de 0,60 à 0,89. Cette cohérence de la modélisation en tous points du territoire est confirmée par le coefficient de détermination par la forêt aléatoire de classification et de régression

indiqué. En effet, pour cette dernière, le coefficient de détermination moyen est de 0,95 pour une RMCE de seulement 0,17°C et un OOB de 0,05. En revanche, la régression linéaire par *partial least square* ne fournit pas de coefficient de détermination assez élevé.

Ainsi, chaque équation qui contribue à la modélisation sur l'ensemble de la zone d'étude n'est valable que pour une journée d'étude. En revanche, des variables prépondérantes sont récurrentes comme le NDVI, la température de surface ou encore la densité de zones en eau. Ainsi, les changements de morphologie urbaine influencent les réponses thermiques de la ville. L'intégration de prédicteurs satellitaires est un avantage certain dans la modélisation du microclimat urbain par modèle de régression linéaire, et à partir de mesures mobiles de la température de l'air.

Lorsque l'on s'intéresse à la localisation des erreurs dans la modélisation de la température de l'air entre la méthode de régression linéaire multiple et la forêt aléatoire, des similitudes entre les deux sont observées. Les modèles surestiment les températures de l'air vers les plans d'eau de Confluence (sud de la presqu'île) et à proximité de la gare de Perrache. Ils sous-estiment la température de l'air dans les rues étroites, à proximité des espaces verts et au sud de la rive gauche du Rhône. Ainsi, les résultats de cette étude ont confirmé le rôle refroidissant joué par les espaces verts et les plans d'eau, ainsi que les problèmes liés à la densité des bâtiments sans végétation dans leurs actions sur la surchauffe urbaine. De plus, la végétation basse présentait un faible pouvoir de refroidissement, principalement en raison d'une absence d'ombre par rapport à la végétation haute et à la végétation de faible densité assurant peu d'évapotranspiration. Cela met en évidence le réel besoin d'utiliser des solutions d'espaces verts et bleus afin de limiter l'ICU et d'améliorer le confort thermique.

Références

- Akbari H., Pomerantz M., Taha H., 2001. Cool Surfaces and Shade Trees to Reduce Energy Use and Improve Air Quality in Urban Areas. *Solar Energy*, 70(3), 295-310. doi: 10.1016/S0038-092X(00)00089-X.
- Alonso L., Renard F., 2019. Integrating Satellite-Derived Data as Spatial Predictors in Multiple Regression Models to Enhance the Knowledge of Air Temperature Patterns. *Urban Science*, 3(4), 101. doi: 10.3390/urbansci3040101.
- Alonso L., Renard F., 2020. A Comparative Study of the Physiological and Socio-Economic Vulnerabilities to Heat Waves of the Population of the Metropolis of Lyon (France) in a Climate Change Context. *International Journal of Environmental Research and Public Health*, 17(3), 1004. doi: 10.3390/ijerph17031004.
- Anselin L., 1995. Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 27(2), 93-115. doi: 10.1111/j.1538-4632.1995.tb00338.x.
- Oleg A., Križan J., Marki A., Bukovec D., 2001. Spatio-Temporal Interpolation of Climatic Variables over Large Region of Complex Terrain Using Neural Networks. *Ecological Modelling*, 138(1), 255-63. doi: 10.1016/S0304-3800(00)00406-3.
- Benas N., Chrysoulakis N., Cartalis C., 2017. Trends of Urban Surface Temperature and Heat Island Characteristics in the Mediterranean. *Theoretical and Applied Climatology*, 130(3), 807-816. doi: 10.1007/s00704-016-1905-8.
- Bobb J., Roger D., Peng M., Dominici F., 2014. Heat-Related Mortality and Adaptation to Heat in the United States. *Environmental Health Perspectives*, 122(8), 811-816. doi: 10.1289/ehp.1307392.
- Boer, E., de Beurs K., Hartkamp A., 2001. Kriging and Thin Plate Splines for Mapping Climate Variables. *International Journal of Applied Earth Observation and Geoinformation*, 3(2), 146-154. doi: 10.1016/S0303-2434(01)85006-6.
- Brandsma T., Wolters D., 2012. Measurement and Statistical Modeling of the Urban Heat Island of the City of Utrecht (the Netherlands). *Journal of Applied Meteorology and Climatology*, 51(6), 1046-1060.
- Breiman, L. Friedman J., Olshen R., Stone C., 1984. *Classification and Regression Trees*. CRC Press.
- Breiman L., 1996. Bagging Predictors. *Machine Learning*, 4(2), 123-140. doi: 10.1007/BF00058655.
- Cantat O. 2004. L'îlot de chaleur urbain parisien selon les types de temps. *Norois - Environnement, aménagement, société*, 191, 75-102. doi: 10.4000/norois.1373.
- Carrega P., Martin N., 2017. *Le climat urbain de Nice, dans un contexte méditerranéen*. Actes de l'Association Internationale de Climatologie, Sfax (Tunisie), 6 pages.
- Carrega P., Rosa F., 2005. Aspects Du Climat Urbain de Nice : Îlot de Chaleur, Îlot de Fraîcheur, Effets de Paroi. *Publications de l'Association Internationale de Climatologie*, 18. Gênes, 8.
- Chapman S., Watson J., Salazar A., Thatcher M., McAlpine C., 2017. The Impact of Urbanization and Climate Change on Urban Temperatures: A Systematic Review. *Landscape Ecology*, 32(10), 1921-1935. doi: 10.1007/s10980-017-0561-4.
- Chen A., Sun R., Chen D., 2012. Studies on Urban Heat Island from a Landscape Pattern View: A Review. *Acta Ecologica Sinica*, 32, 4553-4565. doi: 10.5846/stxb201106280965.
- Chen X., Zhao H., Li P., Yin J., 2006. Remote Sensing Image-Based Analysis of the Relationship between Urban Heat Island and Land Use/Cover Changes. *Remote Sensing of Environment*, 104, 133-146. doi: 10.1016/j.rse.2005.11.016.
- Chen Y., Quan J., Zhan W., Guo Z., 2016. Enhanced Statistical Estimation of Air Temperature Incorporating Nighttime Light Data. *Remote Sensing*, 8(8), 656. doi: 10.3390/rs8080656.
- Cristóbal J., Ninyerola M., Pons X., 2008. Modeling Air Temperature through a Combination of Remote Sensing and GIS Data. *Journal of Geophysical Research: Atmospheres*, 113 (D13). doi: 10.1029/2007JD009318.
- Cristobal J., Ninyerola M., Pons X., Pla M., 2006. Improving Air Temperature Modelization by Means of Remote Sensing Variables. *IEEE International Symposium on Geoscience and Remote Sensing*, 2251-2254
- De Ridder K., Maiheu B., Lauwaet D., Daglis I., Keramitsoglou I., Kourtidis K., Manunta P., Paganini M., 2017. Urban Heat Island Intensification during Hot Spells - The Case of Paris during the Summer of 2003. *Urban Science*, 1(1), 3. doi: 10.3390/urbansci1010003.
- Dempster P., 1969. *Elements of Continuous Multivariate*

Analysis. Reading, Mass.: Addison-Wesley Pub. Co.

Dempster, P., 1969. Upper and Lower Probability Inferences for Families of Hypotheses with Monotone Density Ratios. *The Annals of Mathematical Statistics*, 40(3), 953-969. doi: 10.1214/aoms/1177697600.

Di Paola F., Ricciardelli E., Cimini D., Cersosimo A., Di Paola A., Gallucci D., Gentile S., Gherardi E., Larosa S., Nilo S., Ripepi E., Romano F., Sanò P., Viggiano M., 2018. MiRTaW: An Algorithm for Atmospheric Temperature and Water Vapor Profile Estimation from ATMS Measurements Using a Random Forests Technique. *Remote Sensing*, 10(9), 1398. doi: 10.3390/rs10091398.

Dormann F., Elith J., Bacher S., Buchmann C., Carl C., Carré G., García Marquéz J., Gruber B., Lafourcade B., Leitão P., Münkemüller T., McClean C., Osborne P., Reineking B., Schröder B., Skidmore A., Zurell D., Lautenbach S., 2013. Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance. *Ecography*, 36(1), 27-46. doi: 10.1111/j.1600-0587.2012.07348.x.

Dousset B., Gourmelon F., 2003. Satellite Multi-Sensor Data Analysis of Urban Surface Temperatures and Landcover. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58, 43-54. doi: 10.1016/S0924-2716(03)00016-9.

Dousset B., Gourmelon F., Laaidi K., Zeghnoun A., Giraudet E., Bretin P., Mauri E., Vandentorren S., 2011. Satellite Monitoring of Summer Heat Waves in the Paris Metropolitan Area. *International Journal of Climatology*, 31(2), 313-23. doi: 10.1002/joc.2222.

Fallmann J., Forkel R., Emeis S., 2016. Secondary Effects of Urban Heat Island Mitigation Measures on Air Quality. *Atmospheric Environment*, 125, 199-211. doi: 10.1016/j.atmosenv.2015.10.094.

Foissard X., Dubreuil V., Quénot 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*, 29, 100490. doi: 10.1016/j.uclim.2019.100490.

Getis A., Ord J., 1996. *A Research Agenda for Geographic Information Science. Spatial analysis and modeling in a GIS environment*. Robert B. McMaster, E. Lynn Usery.

Getis A., Ord J., 1992. The Analysis of Spatial Association by Use of Distance Statistics - Getis - 1992 - Geographical Analysis - Wiley Online Library. *Geographical Analysis*, 24 (3).

GIEC, 2019. *Rapport spécial du GIEC sur le réchauffement planétaire de 1,5°C*. Organisation météorologique mondiale. Retrieved January 6, 2020.

Guo Y., Han J., Zhao X., Dai X., and Zhang H., 2020. Understanding the Role of Optimized Land Use/Land Cover Components in Mitigating Summertime Intra-Surface Urban Heat Island Effect: A Study on Downtown Shanghai, China. *Energies*, 13(7), 1678. doi: 10.3390/en13071678.

Hastie T., Tibshirani R., Friedman J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd ed. New York: Springer-Verlag.

Hengl T., Heuvelink G., Perčec Tadić M., Pebesma E., 2012.

Spatio-Temporal Prediction of Daily Temperatures Using Time-Series of MODIS LST Images. *Theoretical and Applied Climatology*, 107(1), 265-277. doi: 10.1007/s00704-011-0464-2.

Jarvis C., Stuart N., 2001. A Comparison among Strategies for Interpolating Maximum and Minimum Daily Air Temperatures. Part II: The Interaction between Number of Guiding Variables and the Type of Interpolation Method. *Journal of Applied Meteorology*, 40(6), 1075-1084.

Jin S., Sader S., 2005. Comparison of Time Series Tasseled Cap Wetness and the Normalized Difference Moisture Index in Detecting Forest Disturbances. *Remote Sensing of Environment*, 94, 364-372. doi: 10.1016/j.rse.2004.10.012.

Journal officiel du Sénat, 2012. *Fermetures de Stations Météo-France et Avenir Du Service Public Météorologique Français*. senat.fr/questions/base/2011/qSEQ110317685.html

Kastendeuch P., Philips N., Nerry F., Roupioz L., Colin J., Luhhe R., Najjar G., 2016. *Mesures pour l'étude des ambiances climatiques à Strasbourg lors de la canicule de juillet 2015*. Actes du colloque de l'Association Internationale de Climatologie, Besançon.

Katsoulis D., Theoharatos G., 1985. Indications of the Urban Heat Island in Athens, Greece. *Journal of Climate and Applied Meteorology*, 24(12), 1296-1302. doi: 10.1175/1520-0450(1985)024<1296:IOTUHI>2.0.CO;2.

Köppen W., 1900. Versuch Einer Klassifikation Der Klimate, Vorzugsweise Nach Ihren Beziehungen Zur Pflanzenwelt. (Schluss). *Geographische Zeitschrift*, 6(12), 657-679.

Kottek M., Grieser J., Beck C., Rudolf B., Rubel F., 2006. World Map of the Köppen-Geiger Climate Classification Updated. *Meteorologische Zeitschrift*, 259-263. doi: 10.1127/0941-2948/2006/0130.

Kukla G., Gavin J., Karl T., 1986. Urban Warming. *Journal of Climate and Applied Meteorology*, 25(9), 1265-1270. doi: 10.1175/1520-0450(1986)025<1265:UW>2.0.CO;2.

Kustas P., Norman M., 1996. Use of Remote Sensing for Evapotranspiration Monitoring over Land Surfaces. *Hydrological Sciences Journal*, 41(4), 495-516. doi: 10.1080/02626669609491522.

Leconte F., Bouyer J., Claverie R., Pétrissans M., 2017. Analysis of Nocturnal Air Temperature in Districts Using Mobile Measurements and a Cooling Indicator. *Theoretical and Applied Climatology*, 130(1), 365-376. doi: 10.1007/s00704-016-1886-7.

Lin L., Zhang Y., 2011. Urban Heat Island Analysis Using the Landsat TM Data and ASTER Data: A Case Study in Hong Kong. *Remote Sensing*, 3(7), 1535-1552. doi: 10.3390/rs3071535.

Yanxu L., Peng J., Wang Y., 2018. Efficiency of Landscape Metrics Characterizing Urban Land Surface Temperature. *Landscape and Urban Planning*, 180, 36-53.

McPherson G., Nowak D., Heisler G., Grimmond S., Souch C., Grant R., Rowntree R., 1997. Quantifying Urban Forest Structure, Function, and Value: The Chicago Urban Forest Climate Project. *Urban Ecosystems*, 1(1), 49-61. doi:

- 10.1023/A:1014350822458.
- Meehl A., Tebaldi C., 2004. More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science*, 305(5686), 994-997. doi: 10.1126/science.1098704.
- Mira M., Ninyerola M., Batalla M., Pesquer L., Pons X., 2017. Improving Mean Minimum and Maximum Month-to-Month Air Temperature Surfaces Using Satellite-Derived Land Surface Temperature. *Remote Sensing*, 9(12), 1313. doi: 10.3390/rs9121313.
- Nguyen K., Liou Y., Li M., and Anh Tran T., 2016. Zoning Eco-Environmental Vulnerability for Environmental management and Protection. *Ecological Indicators*, 69(2016). doi: 10.1016/j.ecolind.2016.03.026.
- Nichol E., Hang To P., 2012. Temporal Characteristics of Thermal Satellite Images for Urban Heat Stress and Heat Island Mapping. *ISPRS Journal of Photogrammetry and Remote Sensing*, 74, 153-162. doi: 10.1016/j.isprsjprs.2012.09.007.
- OCDE, 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OCDE.
- Oke T. 1973. City Size and the Urban Heat Island. *Atmospheric Environment (1967)*, 7, 769-779. doi: 10.1016/0004-6981(73)90140-6.
- Oke T., 1978. Boundary Layer Climates, Londres. *Géographie physique et Quaternaire*, 32(3), 290-291. doi: <https://doi.org/10.7202/1000310ar>.
- Oke T., 1981. Canyon Geometry and the Nocturnal Urban Heat Island: Comparison of Scale Model and Field Observations. *Journal of Climatology*, 1(3), 237-254. doi: 10.1002/joc.3370010304.
- Reid S., Tibshirani R., Friedman J., 2016. A study of error variance estimation in lasso regression. *Statistica Sinica*, 26(1), 35-67.
- Renard F., Alonso L., Fitts Y., Hadjiosif A., Comby J., 2019. Evaluation of the Effect of Urban Redevelopment on Surface Urban Heat Islands. *Remote Sensing*, 11(3), 299. doi: 10.3390/rs11030299.
- Richard Y., Emery J., Dudek J., Pergaud J., Chateau-Smith C., Zito S., Rega M., Vairet T., Castel T., Thévenin T., Pohl B., 2018. How relevant are local climate zones and urban climate zones for urban climate research? Dijon (France) as a case study. *Urban Climate*, 26, 258-274. <https://doi.org/10.1016/j.uclim.2018.10.002>
- Richard Y., Pohl B., Pergaud J., Zito S., Rega M., Thevenin T., Emery J., Vairet T., Dudek J., Thévenin D., Codet-Hache C., Dodet M.-F., 2017. *Évaluation Des Effets Thermiques Des Leviers d'action Visant à Réduire l'Îlot de Chaleur Urbain : L'exemple de Dijon*. Actes du colloque de l'Association Internationale de Climatologie, Sfax (Tunisie).
- Rosenfeld A., Romm J., Akbari H., Pomerantz M., and Taha H., 1996. Policies to Reduce Heat Islands: Magnitudes of Benefits and Incentives to Achieve Them. *ACEEE Summer Study on Energy Efficiency in Buildings*, 9, 177-186.
- Sekulić A., Kilibarda M., Heuvelink G., Nikolić M., Bajat B., 2020. Random Forest Spatial Interpolation. *Remote Sensing*, 12(10), 1687. doi: 10.3390/rs12101687.
- Shandas V., Voelkel J., Williams J., Hoffman J., 2019. Integrating Satellite and Ground Measurements for Predicting Locations of Extreme Urban Heat. *Climate*, 7(1), 5. doi: 10.3390/cli7010005.
- Shapiro S., Wilk M., 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika*, 52(3-4), 591-611. doi: 10.1093/biomet/52.3-4.591.
- Sobrino J., Jimenez-Munoz J.-C., Paolini L., 2004. Land Surface Temperature Retrieval from LANDSAT TM 5. *Remote Sensing of Environment*, 90, 434-440. doi: 10.1016/j.rse.2004.02.003.
- Sun Y., Gao C., Li J., Wang R., Liu J., 2019. Quantifying the Effects of Urban Form on Land Surface Temperature in Subtropical High-Density Urban Areas Using Machine Learning. *Remote Sensing*, 11(8), 959. doi: 10.3390/rs11080959.
- Tenenhaus M., Pagès J., Ambroisine L., Guinot C., 2005. PLS Methodology to Study Relationships between Hedonic Judgements and Product Characteristics. *Food Quality and Preference*, 16(4), 315-325. doi: 10.1016/j.foodqual.2004.05.013.
- Tibshirani R., 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, 58(1), 267-288.
- Tsin K., Knudby A., Krayenhoff E., Chak Ho H., Brauer M., Henderson S., 2016. Microscale Mobile Monitoring of Urban Air Temperature. *Urban Climate*, 18, 58-72. doi: 10.1016/j.uclim.2016.10.001.
- Wang M., He G., Zhang Z., Wang G., Zhang Z., Cao X., Wu Z., Liu X., 2017. Comparison of Spatial Interpolation and Regression Analysis Models for an Estimation of Monthly Near Surface Air Temperature in China. *Remote Sensing*, 9(12), 1278. doi: 10.3390/rs9121278.
- Wang Y., Ji W., Yu X., Xu X., Jiang D., Wang Z., Zhuang D., 2014. The Impact of Urbanization on the Annual Average Temperature of the Past 60 Years in Beijing. *Advances in Meteorology*, e374987.
- Weston J., 1988. Boundary Layer Climates. *Quarterly Journal of the Royal Meteorological Society*, 114(484), 1568-1568. doi: 10.1002/qj.49711448412.
- Wold S., Sjöström M., Andersson P., Linusson A., Edman M., Lundstedt T., Nordén B., Sandberg M., Uppgård L.-L., 2000. *Multivariate Design and Modelling in QSAR, Combinatorial Chemistry, and Bioinformatics*. Molecular Modeling and Prediction of Bioactivity, edited by K. Gundertofte and F. S. Jørgensen. Boston, MA: Springer US.
- Xie M., Wang Y., Fu M., 2011. An Overview and Perspective about Causative Factors of Surface Urban Heat Island Effects. *Progress in geography*, 30(1), 35-41. doi: 10.11820/dlxxjz.2011.01.004.
- Yang Q., Huang X., Li J., 2017. Assessing the Relationship between Surface Urban Heat Islands and Landscape Patterns across Climatic Zones in China. *Scientific Reports*, 7. doi:

10.1038/s41598-017-09628-w.

Zhang Z., Du Q., 2019. A Bayesian Kriging Regression Method to Estimate Air Temperature Using Remote Sensing Data. *Remote Sensing*, 11(7), 767. doi: 10.3390/rs11070767.

Zhu W., Lú A., Jia S.. 2013. Estimation of Daily Maximum and Minimum Air Temperature Using MODIS Land Surface Temperature Products. *Remote Sensing of Environment*, 130, 62-73. doi: 10.1016/j.rse.2012.10.034.

Citation de l'article : Alonso L. et Renard F., 2020. Compréhension du microclimat urbain lyonnais par l'intégration de prédicteurs complémentaires à différentes échelles dans des modèles de régression. *Climatologie*, 17, 2.