



# **Enjeux de la réduction d'échelle dans l'estimation par télédétection des déterminants climatiques**

**Mémoire**

**Hugues Yenoukoume Hangnon**

**Maîtrise en sciences géomatiques - avec mémoire**  
Maître ès sciences (M. Sc.)

Québec, Canada

# **Enjeux de la réduction d'échelle dans l'estimation par télédétection des déterminants climatiques**

**Mémoire**

**Hugues Y. Hangnon**

Sous la direction de :

Frédéric Hubert, directeur de recherche  
Alain A. Viau, codirecteur de recherche

# Résumé

Ce travail s'inscrit dans le cadre de recherche sur les maladies vectorielles de Lyme et Virus du Nil au sein de l'Agence de Santé Publique du Canada (ASPC) ayant pour finalité d'évaluer et de cartographier les risques sanitaires associés à ces maladies infectieuses liées au climat aux échelles municipales, provinciales et fédérale. Dans ce contexte, cette recherche vise à démontrer la faisabilité, la pertinence et les enjeux de recourir aux méthodes de réduction d'échelle pour obtenir à une haute résolution spatio-temporelle (100/30 m et 1 jour) avec au plus des marges d'erreur de 2 unités, des déterminants climatiques et microclimatiques (DCMC) en milieu hétérogène du Canada.

Un cadre méthodologique d'application des méthodes de réduction d'échelle, Random Forest Regression (RFR), Thermal sharpening (TsHARP), Pixel block intensity modulation (PBIM), a été proposé pour estimer la température de surface (LST) de MODIS 1000 m à 100/30 m. Des expérimentations basées sur cette approche ont été effectuées sur trois sites au Québec à différentes époques. Les résultats, spatialement représentatifs, ont été validés avec les températures de l'air et celles prises par de Landsat 08 avec des marges d'erreur autour de 2°C. L'analyse des résultats démontre la capacité effective des méthodes de réduction d'échelle à discriminer la LST dans l'espace. Toutefois, dans le contexte du projet de l'ASPC, ces résultats sont non concluants à 100/30 m en l'absence d'une plus-value significative au plan spatial de LST. Cette analyse a conduit à discuter des enjeux temporels, spatiaux, méthodologiques et de gestion de gros volumes de données en lien avec la réduction d'échelle dans le contexte du projet.

**Mots clés :** Réduction d'échelle spatiale, Température de surface, Maladies vectorielles, Québec, Canada

# Abstract

This research is part of the Public Health Agency of Canada's (PHAC) research on Lyme and West Nile Virus vector-borne diseases, which aims to assess and map the health risks associated with these climate-related infectious diseases at the municipal, provincial and federal levels. In this context, this research aims to demonstrate the feasibility, relevance and challenges of using downscaling methods to obtain high spatial and temporal resolution (100/30 m and 1 day), with margins of error of no more than 2 units, of climatic and microclimatic determinants (CMDs) in a heterogeneous Canadian environment.

A methodological framework for the application of downscaling methods, Random Forest Regression (RFR), Thermal sharpening (TsHARP), Pixel block intensity modulation (PBIM), has been proposed to estimate the surface temperature (LST) from MODIS 1000 m to 100/30 m. Experiments with our approach were carried out at three sites in Quebec at different times. The spatially representative results were validated with air and Landsat 08 temperatures with error margins around 2°C. The analysis of our results demonstrates the effective capacity of downscaling methods to discriminate LST in space. However, in the context of the ASPC project, these results are inconclusive at 100/30 m in the absence of a significant, expected increase in the spatial accuracy of LST. This analysis led to a discussion of the temporal, spatial, methodological and large data volume management issues related to downscaling in the context of the project.

**Keywords :** Spatial downscaling, Land surface temperature, Vector-borne diseases, Québec, Canada

# Table des matières

Résumé .....	ii
Abstract.....	iii
Table des matières .....	iv
Liste des figures.....	vii
Liste des tableaux.....	ix
Liste des abréviations .....	x
Remerciements.....	xi
Introduction .....	1
Contexte de l'étude.....	1
Problématique de l'étude.....	5
Question de recherche .....	6
Objectif général .....	6
Objectifs spécifiques.....	6
Hypothèse de recherche .....	7
Méthodologie de travail .....	7
Structure du document .....	11
Chapitre 1 - État de l'art sur la réduction d'échelle et des méthodes de réduction d'échelle de la température de surface (LST) .....	12
1.1. Introduction.....	12
1.2. Méthodes de réduction d'échelle et leurs limites .....	12
1.2.1. Méthodes de réduction d'échelle.....	13
1.2.2. Limites des méthodes de réduction d'échelle .....	15
1.3. Méthodes de réduction d'échelle appliquées à la LST .....	18
1.3.1. Température de surface (Land surface temperature, °C) .....	20
1.3.2. Évaluation des méthodes de réduction d'échelle de la température de surface .....	23
1.3.3. Application des trois méthodes de réduction d'échelle identifiées .....	32
1.3.4. Description des trois méthodes retenues dans cette étude.....	36
1.4. Conclusion.....	39
Chapitre 2 - Vers une approche méthodologique d'application de la réduction d'échelle de LST à 30 mètres .....	41
2.1. Introduction.....	41
2.2. Description des étapes générales de réduction d'échelle de la température de surface.....	41
2.2.1. Préparation des données .....	42
2.2.2. Définition et calcul des variables auxiliaires .....	44
2.2.3. Réalisation de la réduction d'échelle à une résolution cible .....	46
2.3. Description d'une nouvelle approche de réduction d'échelle.....	48
2.3.1. Première phase : réduction d'échelle à 100 m .....	49
2.3.2. Seconde phase : réduction d'échelle à 30 m .....	50
2.4. Conclusion.....	53

Chapitre 3 - Démarche générale d'implémentation de la nouvelle méthode de réduction d'échelle.....	54
3.1. Introduction.....	54
3.2. Identification de trois zones d'étude .....	55
3.3. Description des données.....	57
3.3.1. Création des variables de prédiction et leur sélection.....	59
3.3.2. Évaluation de la colinéarité entre les variables.....	60
3.4. Description de la mise en application de chaque méthode de réduction d'échelle.....	60
3.4.1. Méthode TsHARP .....	60
3.4.2. Méthode RFR.....	62
3.4.3. Méthode PBIM .....	63
3.5. Validation des données.....	63
3.5.1. Les indicateurs d'erreurs.....	64
3.5.2. Validation interne et externe.....	65
3.6. Conclusion.....	67
Chapitre 4 - Expérimentations .....	68
4.1. Introduction.....	68
4.2. Réduction d'échelle de MODIS LST de 1000 m à 100 m .....	68
4.2.1. Expérimentation 1 : Région de la Capitale nationale .....	68
4.2.2. Expérimentation 2 : Région d'Estrie.....	78
4.2.3. Expérimentation 3 : Région de Montérégie.....	87
4.3. Expérimentation 4 : réduction d'échelle de 100 m à 30 m dans la région de l'Estrie.....	96
4.4. Expérimentation 5 : réduction d'échelle de 1000 m à 30 m dans la région de l'Estrie.....	100
4.5. Conclusion.....	105
Chapitre 5 - Analyse des résultats et discussions.....	107
5.1. Introduction.....	107
5.2. Qualité des résultats.....	108
5.2.1. Précision spatiale des résultats de température de surface LST .....	109
5.2.2. Justesse des résultats.....	110
5.3. Utilité des résultats de LST réduites .....	112
5.4. Enjeux à la réduction d'échelle.....	113
5.4.1. Enjeux temporels .....	113
5.4.2. Enjeux spatiaux.....	115
5.4.3. Enjeux méthodologiques et du choix des méthodes .....	116
5.4.4. Enjeux de la gestion des gros volumes de données .....	119
5.4.5. Application de la méthodologie de réduction d'échelle proposée à d'autres déterminants climatiques et microclimatiques (DCMC) .....	121
5.5. Conclusion.....	123
Conclusion générale .....	124
Introduction.....	124
Acquis de l'étude sur la précision spatiale avec les méthodes de réduction d'échelle .....	125
Bilan de la validation des méthodes de réduction d'échelle de LST .....	125
Contributions.....	126

Perspectives de recherche .....	127
Amélioration de l'approche méthodologique de réduction d'échelle de la température de surface.....	128
Évaluer la contribution des drones dans la précision spatiale de LST .....	129
Intégrer la méthode de réduction d'échelle dans l'infrastructure de données massives .....	130
Bibliographie.....	132
Annexe A .....	140
Annexe B .....	158
Annexe C .....	163

# Liste des figures

Figure 0.1 : Cadre méthodologique de gestion du projet de recherche _____	8
Figure 1.1 : Schématisation du processus de réduction d'échelle _____	13
Figure 2.1 : Diagramme UML d'activités sur les étapes générales de la méthode de réduction d'échelle à 30 mètres. _____	42
Figure 2.2: Description des tâches à réaliser à la phase de préparation des données _____	43
Figure 2.3 : Détail des tâches à réaliser à la phase de définition des prédicteurs _____	45
Figure 2.4 : Description des tâches à réaliser à l'étape 3 _____	47
Figure 2.5: Processus de réalisation de la carte des émissivités de surface _____	52
Figure 3.1 : Cadre général d'implémentation de la réduction d'échelle de LST à 100 m et à 30 m _____	54
Figure 3.2 : Vue globale de la zone d'étude : région de la Capitale nationale _____	55
Figure 3.3 : Vue globale de la zone d'étude : Région de l'Estrie _____	56
Figure 3.4 : Vue globale de la zone d'étude : région de Montérégie _____	57
Figure 3.5 : Description des activités réalisées pour la mise en œuvre de l'approche proposée _____	61
Figure 3.6 : Processus d'estimation de la LST de Landsat 08/TIRS _____	66
Figure 4.1 : Région de la Capitale nationale : résultats de réduction d'échelle à 100 m suivant différentes méthodes _____	70
Figure 4.2 : Région de la Capitale nationale : zoom sur la représentativité spatiale des valeurs de LST en milieu urbain hétérogène _____	71
Figure 4.3 : Région de la Capitale nationale : zoom sur la représentativité spatiale des valeurs LST en zone homogène boisée _____	72
Figure 4.4 : Représentation spatiale des écarts de Landsat LST observée et MODIS LST prédite selon 3 méthodes à 100 m _____	75
Figure 4.5 : Région de l'Estrie : résultats de réduction d'échelle à 100 m suivant différentes méthodes _____	80
Figure 4.6: Région de l'Estrie : zoom sur la représentativité spatiale des valeurs de LST en milieu urbain hétérogène _____	81
Figure 4.7 : Région de l'Estrie : zoom sur la représentativité spatiale des valeurs de LST en zone homogène boisée _____	82
Figure 4.8 : Représentation spatiale des écarts de LST entre Landsat LST et MODIS prédite à 100 m _____	85
Figure 4.9 : Région de Montérégie : résultats de réduction d'échelle à 100 m suivant différentes méthodes _____	89
Figure 4.10: Région de Montérégie : zoom sur la représentativité spatiale des valeurs de LST en milieu hétérogène à dominance urbaine _____	90
Figure 4.11 : Région de Montérégie : zoom sur la représentativité spatiale des valeurs de LST en milieu hétérogène à dominance agricole _____	91
Figure 4.12 : Région de Montérégie : représentation des écarts de températures entre Landsat LST et MODIS LST prédite _____	95
Figure 4.13 : Occupation et utilisation des sols dans la région de l'Estrie _____	97
Figure 4.14 : Émissivités de surface de la région de l'Estrie _____	97
Figure 4.15 : Résultats de la réduction d'échelle de MODIS LST de 1000 m à 30 m _____	98
Figure 4.16 : MODIS LST prédite de 1000 m à 30 m avec RFR _____	101
Figure 4.17 : MODIS LST prédite de 1000 m à 30 m avec TsHARP _____	101
Figure 4.18 : MODIS LST observée à 1000 m _____	102
Figure 4.19 : Représentation spatiale des écarts de valeurs de LST entre Landsat LST observée et MODIS LST prédite de 1000 m à 30 m _____	104
Figure 0.1 : Région de la Capitale nationale : Spatialisation de la température de surface (°C) de MODIS à 1000 m _____	148
Figure 0.2 : Région de la Capitale nationale : résultats de réduction d'échelle de la température de surface (°C) MODIS de 1000 m à 100 m suivant la méthode RFR _____	149



Figure 0.3 : Région de la Capitale nationale : résultats de réduction d'échelle de la température de surface (°C) MODIS de 1000 m à 100 m suivant la méthode TsHARP	150
Figure 0.4 : Région de la Capitale nationale : résultats de réduction d'échelle de la température de surface (°C) MODIS de 1000 m à 100 m suivant la méthode PBIM	151
Figure 0.5 : Région de la Capitale nationale : Zoom sur la variabilité de la température de surface (°C) MODIS réduite de 1000 m à 100 m en milieu urbain suivant la méthode RFR	152
Figure 0.6 : Région de la Capitale-Nationale : distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite à 100 m	153
Figure 0.7 : Région de l'Estrie : distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite à 100 m	154
Figure 0.8 : Région de Montérégie : distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite à 100 m	155
Figure 0.9 : Région de l'Estrie: distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite directement de 1000 m à 30 m	156
Figure 0.10 : Variation de la température horaire de l'air au niveau des stations météorologiques de nos zones d'étude	157
Figure 0.1 : Exemple de cube de données matricielles à 3 dimensions (a) et à 4 dimensions (b). Les trois dimensions, la latitude, la longitude et le temps. Les bandes spectrales ajoutées donnent la dimension 4.	164
Figure 0.2 : Cadre opératoire de production de la cartographie des risques de maladies infectieuses liées au climat	166

# Liste des tableaux

Tableau 1.1: Portrait des méthodes de réduction d'échelle LST _____	25
Tableau 1.2 : Avantages et limites des méthodes de réduction d'échelle de LST _____	26
Tableau 1.3 : Évaluation comparative des modèles de réduction d'échelle de LST selon les critères du projet _____	37
Tableau 3.1 : Description des données utilisées _____	58
Tableau 3.2 : Mode de calcul des émissivités de surface basé sur les seuils de NDVI _____	63
Tableau 4.1 : Région de la Capitale nationale – analyse de corrélation linéaire univariée entre les variables _____	68
Tableau 4.2 : Indicateurs statistiques calculés entre MODIS LST observée et MODIS LST simulée à 100 m _____	74
Tableau 4.3 : indicateurs statistiques calculés entre Landsat LST observée et MODIS LST simulée à 100 m _____	74
Tableau 4.4 : Valeurs de température selon différentes sources à la localisation de la station météo _____	76
Tableau 4.5 : Écart de température entre les données brutes _____	76
Tableau 4.6 : Écart des valeurs de température entre les données simulées et données originales _____	77
Tableau 4.7 : Région de l'Estrie : analyse de corrélation linéaire univariée entre les variables auxiliaires _____	78
Tableau 4.8 : Indicateurs statistiques entre MODIS LST observée et MODIS LST simulée à 100 m _____	79
Tableau 4.9 : Indicateurs statistiques calculés entre MODIS LST simulée à 100 m et Landsat LST observée _____	84
Tableau 4.10 : Valeurs de température selon différentes sources _____	86
Tableau 4.11 : Écart de température entre les données brutes _____	86
Tableau 4.12 : Écart des valeurs de température entre les données simulées et données originales _____	87
Tableau 4.13 : Région de Montérégie : analyse de corrélation linéaire univariée entre les variables auxiliaires _____	88
Tableau 4.14 : Indicateurs statistiques calculés entre MODIS LST simulée à 100 m et MODIS LST observée _____	92
Tableau 4.15 : Indicateurs statistiques calculés entre MODIS LST simulées et Landsat LST observée _____	92
Tableau 4.16 : Valeurs de température selon différentes sources _____	92
Tableau 4.17 : Écart des valeurs de température entre les données originales _____	93
Tableau 4.18 : Écart des valeurs de température entre les données simulées et données originales _____	94
Tableau 4.19 : Indicateurs d'erreurs entre les résultats et Landsat 08 à 30 m _____	99
Tableau 4.20 : Écart des valeurs de température entre l'image simulée et les données observées _____	100
Tableau 4.21 : Indicateurs d'erreurs entre les résultats et Landsat 08 à 30 m _____	103
Tableau 4.22: Écart des valeurs de température entre les données simulées et données originales _____	103
Tableau 4.23 : Comparaison des indicateurs à 30 m selon l'approche de RE _____	103
Tableau 5.1 : Récapitulatif des indicateurs de qualité calculés sur le territoire entier des trois zones d'étude pour la LST à 100 m _____	108
Tableau 5.2 : Récapitulatif des indicateurs de qualité calculés à des stations météorologiques en zone forestière des trois zones d'étude pour la LST à 100 m _____	108
Tableau 5.3: Présentation des RMSE liées à la RE de LST selon la méthode RFR dans certaines études _____	110
Tableau 5.4 : Présentation des RMSE liées à la RE de LST selon la méthode TsHARP dans certaines études _____	111
Tableau 5.5 : Description de la température et la topographie des zones d'étude _____	117
Tableau 0.1 : Classification des méthodes de réduction d'échelle en télédétection _____	140
Tableau 0.2 : Présentation de quelques capteurs imageurs selon les résolutions temporelles et spatiales _____	140
Tableau 0.3 : Caractéristiques du capteur MODIS _____	141
Tableau 0.4 : Principales utilisations par bandes spectrales du capteur MODIS _____	141
Tableau 0.5: Analyse comparative de quelques modèles de réduction d'échelle se rapportant à la température de surface _____	143
Tableau 0.6 : Mode de calcul des différents indices _____	146
Tableau 0.7 : Déterminants environnementaux associés aux maladies transmises par les moustiques (MTM) et leurs vecteurs dont Culex pipiens-restuans, vecteur du Virus du Nil Occidental _____	147

## Liste des abréviations

ASPC	Agence de santé publique du Canada
BR	Basse résolution
BSI	Indice d'aridité ou de sol nu
DCMC	Déterminants climatiques et microclimatiques
DEC	Données environnementales et climatiques
DisTrad	Disaggregation procedure for radiometric surface temperature
Dsopt	Optimized downscaling process
ELM	Extreme Learning Machine
EMA	Erreur moyenne absolue
FVC	Fractional vegetation cover / Couverture végétale fractionnée
GPS	Genetic particle smoother
GTWR	Geographically and temporally weighted regression
HR	Haute résolution
IDE	Integrated development environment / Environnement de développement intégré
MODIS	Moderate-resolution imaging spectroradiometer
MRC	Municipalité régionale de comté
MSFAT	Multiple scale factors with adaptive thresholds
NDBI	Normalize the difference building index
NDVI	Normalized difference vegetation index / Indice de végétation différencié normalisé
NDWI	Normalized difference water index / Indice d'eau différencié normalisé
NL-DisTrad	Non linear disaggregation procedure for radiometric surface temperature
LMDS	Least median square regression downscaling
LST	Land surface temperature / Température de surface
PBIM	Pixel block intensity modulation
PRDS	Pace regression downscaling
PCV	Proportion de couverture végétale
PV	Proportion de végétation
RFR	Random Forest Regression
RMSE	Root mean square error / Erreur quadratique moyenne
SADFAT	Spatiotemporal adaptive data fusion algorithm for temperature
SAVI	Soil Adjusted Vegetation Index / Indice de végétation ajustée au sol
SPOT	Satellite pour l'observation de la terre
SVR	Support vector regression
TPS	Thin platespline
TsHARP	Thermal sharpening
VNO	Virus du Nil Occidental

# Remerciements

Je voudrais remercier ma direction de recherche, les professeurs Frédéric HUBERT. Alain A. VIAU pour m'avoir donné cette opportunité d'apprendre davantage en télédétection et surtout pour toute leur disponibilité et accompagnement durant les six sessions de formation.

Je voudrais aussi remercier Monsieur Serge-Olivier KOTCHI, point focal pour le suivi de cette étude pour le compte de l'Agence de Santé Publique du Canada (ASPC), pour ses conseils, expertises et orientations dans cette étude.

Mes remerciements vont également aux étudiants de génie en géomatique, Maxime Rioux-ROUSSEAU, Simon Gingras-GAGNON, Victor-Raul Léon MATUTE et Aidan WAGNER pour leur assistance en programmation informatique.

Je tiens à remercier le département des sciences géomatiques, le centre de recherche en données et intelligence spatiales et tout leur personnel pour nous avoir fourni un cadre adéquat de formation.

Mes sincères remerciements à mes camarades de l'Association des étudiants gradués en géomatique-AGREGÉ- et de laboratoire pour leur soutien à divers niveaux.

# Introduction

## Contexte de l'étude

Le présent sujet de recherche s'inscrit dans un projet global élaboré par l'Agence de santé publique du Canada (ASPC). Ce projet a été initié dans un contexte d'émergence des maladies vectorielles<sup>1</sup> dans diverses régions du Canada, à cause du réchauffement climatique et des modifications globales de l'environnement. Parmi elles, le virus du Nil occidental (VNO) et la maladie de Lyme dont 500<sup>2</sup> cas ont été déclarés au Québec et 1 159<sup>3</sup> cas à Ontario en 2019. À l'avenir, le nombre de cas de maladies pourrait croître avec l'augmentation de la température due aux changements climatiques (Ogden et al., 2015 ; Ogden & Gachon, 2019). L'émergence de ces maladies dues aux tiques et moustiques constitue un risque pour la santé humaine.

Face à cette émergence des cas de maladies, l'ASPC a mis en place un cadre opératoire de surveillance des mouvements et aires de répartition de ces vecteurs<sup>4</sup>. Ce cadre opératoire est basé sur l'utilisation des images d'observation de la Terre (Kotchi et al., 2019 ; Kotchi et al., 2021). Les données disponibles collectées sont utilisées pour élaborer régulièrement des cartes de risques sanitaires à l'échelle du Canada et des provinces pour la prévention des cas de maladies de Lyme et du VNO. Les cartes de risque sanitaire produites à partir d'une modélisation de données géospatiales (dont des images satellitaires) de résolution 250 mètres sont mises à jour annuellement pour les moustiques et mensuellement pour les tiques pour les prises de décision politique de gestion des risques liés à ces vecteurs. Néanmoins cette échelle de modélisation, à 250 mètres, reste faible pour une prise de décision à l'échelle des municipalités où la gestion de risques sanitaires pourrait s'avérer plus efficace. En effet, les données produites à une grande échelle (Province, Fédérale) à cause de la généralisation peuvent ne pas révéler les tendances qui existent à des échelles fines (municipalités) (Boffet, 2003 ; Simonneaux, 1996). De fait, l'ASPC vise une modélisation et une cartographie de risques sanitaires à une fine échelle de 100 mètres, voire 30 mètres avec une marge d'erreur de plus ou moins de 2 unités à partir des données géospatiales ouvertes disponibles. Pour atteindre ces échelles, il faut disposer des données aux

---

<sup>1</sup> Les maladies vectorielles encore appelées les maladies à transmission vectorielle sont des maladies infectieuses transmises par des vecteurs qui peuvent être des moustiques, tiques, ou punaises. Pour plus de détails, se réfère à <https://journals.openedition.org/confins/22854>

<sup>2</sup> De nouvelles maladies transmises par les tiques arrivent au Québec ; <https://www.protegez-vous.ca/nouvelles/sante-et-alimentation/cet-ete-attention-aux-tiques-et-aux-moustiques>

<sup>3</sup> <https://www.publichealthontario.ca/fr/diseases-and-conditions/infectious-diseases/vector-borne-zoonotic-diseases/lyme-disease>

<sup>4</sup> Les vecteurs sont des intermédiaires comprenant des agents pathogènes (virus, parasite, bactérie...) par lesquels se propagent les maladies. Dans cette étude les vecteurs désignent les tiques et moustiques.

mêmes échelles spatio-temporelles, alors que les données climatiques et environnementales d'observation disponibles et utilisées sont de sources multiples et multiéchelles diverses (données d'occupation et de l'utilisation des terres de résolution spatiale 20 à 30 mètres voire 2 à 10 mètres ; données de température de surface, l'humidité de surface et autres précipitations de résolution spatiale 100 à 1000 mètres). Ceci conduit à un problème d'inadéquation d'échelle en l'occurrence de l'échelle spatiale entre les données environnementales et climatiques (Bierkens et al., 2001 ; Wu & Li, 2009).

Les habitats des vecteurs (tiques et moustiques dans cette étude) constitués des biomes forestiers fragmentés, des zones mixtes de parcs, résidentielles, plans d'eau sont de petites tailles (micro-habitat) et peuvent être identifiés à une résolution spatiale de 30 m. Cependant, le microclimat engendré par ces habitats n'est pas déterminable à cette même résolution spatiale. La principale raison est qu'il n'existe à ce jour aucun satellite météorologique possédant une haute résolution spatiale. La seconde raison est que la mixité de plusieurs unités d'occupation du sol impacte la variabilité des données climatiques dans ces micro-habitats pour engendrer des microclimats. Or, il a été démontré que les microclimats (température et humidité de surface, précipitations), de par leur variabilité spatio-temporelle, influent sur le cycle de vie, le mouvement et l'abondance des tiques et moustiques (Burtis et al., 2016 ; McCabe & Bunnell, 2004 ; Ogden et al., 2004). Il va en résulter une variabilité spatiale dans la distribution des vecteurs.

Par voie de conséquence, dans une même journée, les vecteurs peuvent se mouvoir d'une aire géographique à une autre en fonction des fluctuations des microclimats. De fait, le suivi quotidien des microclimats dans les micro-habitats paraît nécessaire pour la modélisation des cartes de risques, ce suivi va demander de disposer à la fois des données climatiques, de hautes résolutions temporelle et spatiale. Ce que n'offre aucun capteur d'image satellitaire à ce jour. La plupart des capteurs d'images satellitaires météorologiques ont une résolution temporelle d'un jour au minimum. C'est l'exemple des capteurs AVHRR<sup>5</sup> ou MODIS<sup>6</sup>. Si les images satellitaires météorologiques ont une haute résolution temporelle, leurs faibles résolutions spatiales ne permettent cependant pas d'avoir des informations détaillées de la température de surface, de l'humidité de surface ou des précipitations à une fine échelle de 100 m ou 30 m. Précisons que les capteurs d'images satellitaires thermiques (température de surface) de Landsat 08 TIRS font exception à ces caractéristiques sus-citées. Ils fournissent des images satellitaires de résolution spatiale de 100 m et de résolution temporelle de 16 jours. La résolution temporelle de 16 jours ne répond pas non plus à l'exigence de haute résolution temporelle désirée. C'est aussi cette haute résolution temporelle (d'un jour au minimum) qui fait également défaut chez des capteurs d'images satellitaires dédiés à la surveillance de l'occupation du sol et à l'environnement. Il se pose également la question de non-concordance d'échelle, en l'occurrence de l'échelle temporelle des sources des données.

---

<sup>5</sup> AVHRR pour Advanced Very High Resolution Radiometer

<sup>6</sup> Moderate Resolution Imaging Spectroradiometer

C'est en réponse à cette question d'échelle à laquelle sont toujours confrontées les études environnementales (Bierkens et al., 2001) que le projet de l'ASPC a intégré un volet géomatique pour le suivi des maladies vectorielles émergentes transmises par les tiques et les moustiques. L'objectif général est d'optimiser le processus de caractérisation des milieux de vie des tiques et moustiques à partir des images satellitaires en utilisant la mise à l'échelle ou le changement d'échelle (Bierkens et al., 2001 ; Wu & Li, 2009) pour une uniformisation des données environnementales, climatiques et microclimatiques. À travers le changement d'échelle, on cherche à avoir des données environnementales et climatiques (DEC) à une fine résolution à partir de résolutions grossières des images satellitaires pour mieux connaître et caractériser les habitats des vecteurs. Ce besoin vise également à cartographier la progression des vecteurs de maladie à l'échelle des municipalités, des provinces et au niveau fédéral. L'enjeu scientifique de ce sujet de recherche est donc la réduction d'échelle (« downscaling », en anglais) pour l'estimation des DEC appliqués au cas des maladies vectorielles. Il s'agit donc de ramener l'information contenue dans les images satellitaires (données environnementales et climatiques) à une résolution fine, proche de la résolution spatiale à laquelle les habitats sont caractérisables au sol et ainsi de tester la faisabilité de la réduction d'échelle avant de l'appliquer à tout le territoire canadien. Une résolution spatiale de 100 m ou 30 m pour l'expérimentation a été fixée par le projet principal de l'ASPC. La réduction d'échelle a été utilisée pour des applications dans les domaines de l'hydrologie, climatologie, météorologie, agrométéorologie, de la gestion des ressources en eau, etc. (Agam et al., 2007 ; Bartkowiak et al., 2019 ; Bindhu et al., 2013 ; Jeganathan et al., 2011 ; Peng et al., 2017) pour pallier les lacunes d'échelle spatiale et temporelle des capteurs d'images satellitaires. La température de surface, l'humidité de surface, l'évapotranspiration ou les précipitations sont les paramètres souvent estimés par réduction d'échelle à l'échelle des parcelles agricoles, du bassin versant. Plusieurs méthodes de réduction d'échelle sont disponibles en fonction des milieux, quel que soit le paramètre à estimer : la méthode Thermal Shapering (TSHARP) (Agam et al., 2007 ; Hutengs & Vohland, 2016 ; Peng et al., 2019) qui s'adapte mieux aux zones disposant d'une bonne couverture végétale, la méthode lissage Laplacien (TPS) appliquée en zones de faible couverture végétale (Singh Rawat et al., 2019), la méthode de Random Forest Regression (Bartkowiak et al., 2019 ; Ebrahimi & Azadbakht, 2019 ; Hutengs & Vohland, 2016) utilisée en zones hétérogènes, etc. Malgré la multiplicité de méthodes de réduction d'échelle, conséquence de l'absence de robustesse d'une seule méthode reproductible en tout milieu et conditions saisonnières (Malbêteau, 2017), le mode opératoire est quasi identique. L'ensemble des méthodes existantes établissent une relation de liaison statistique ou un modèle physique entre la variable d'intérêt à une échelle grossière et les variables auxiliaires à une échelle fine (Bierkens et al., 2001). Les variables auxiliaires sont des variables explicatives qui ont (ou peuvent avoir) une influence sur la variable à expliquer. En exemple, la relation entre la température de l'air et l'altitude dans laquelle la variabilité de température de l'air est influencée par la topographie du milieu observé.

Les résultats obtenus dans plusieurs études en fin de processus de réduction d'échelle présentent une précision moyenne en valeur absolue de 1,5 à 5 d'unités en termes d'erreur quadratique moyenne (RMSE) en fonction du facteur d'échelle cible et de la qualité des données sources. Par contre, dans cette étude, les résultats sont attendus avec une précision de plus ou moins 2 unités. Rappelons que la méthode de réduction d'échelle a déjà été expérimentée au sein de l'ASPC pour modéliser la température de surface de 1 km à 500/250 m selon nos échanges avec le point focal de l'ASPC en charge du volet géomatique.

Le sujet de recherche s'intéresse uniquement aux données climatiques et microclimatiques (DCMC), à cause de leur prééminence sur les données environnementales dans l'émergence des maladies de Lyme et du virus de Nil (Gasmi et al., 2019). De manière précise, ce sujet se focalise sur la réduction d'échelle de la température de surface parmi les DCMC à cause de son rôle prépondérant dans la distribution des vecteurs en dépit de l'influence des actions combinées des DCMC sur la biologie et prolifération des vecteurs. En effet, les précipitations, la température de surface et l'humidité de surface, dénommées ici « déterminants » sont les principales variables climatiques qui impactent plus les moustiques et les tiques (Adde et al., 2016 ; Bouchard et al., 2011, 2019). Au sein des trois déterminants, la température de surface a été admise comme la plus significative devant les précipitations et l'humidité de surface dans l'expansion et la croissance des tiques et moustiques (Bouchard et al., 2019 ; Githeko et al., 2001). C'est ce qui justifie que ce travail se focalise sur la réduction d'échelle de la température de surface, laquelle est influencée par plusieurs facteurs dont les états de surface désignés ici « données ». À cause de cette relation d'interdépendance, l'estimation de la température de surface requiert des données d'occupation des territoires (Chokmani et al., 2001 ; Hachem, 2008). Or, les données d'occupation des territoires proviennent d'autres capteurs imageurs et sont disponibles à une résolution spatiale de 2,5 mètres pour le satellite pour l'observation de la terre (SPOT) 5 et 0,8 mètres pour les images satellite IKONOS. Ces différences dans les caractéristiques des imageries satellitaires amènent à procéder à la réduction d'échelle par l'association des images de diverses sources pour un suivi dynamique à une échelle plus fine.

Toutefois, selon les résultats des études exploratoires de Kochi et al. (2014, 2019), les images satellitaires MODIS offrent l'avantage de réduire les sources de données à utiliser. En effet, ce capteur fournit des données thématiques variées (environnement, atmosphère, océanographie, météorologie...) à une bonne résolution temporelle d'un jour (voire cinq minutes pour le Canada), mais avec une résolution spatiale faible (varie de 250 à 1000 mètres). Mais l'utilisation d'une seule source de données peut aussi altérer la qualité des informations à obtenir (Peng et al., 2017) due notamment à la faible résolution spatiale de MODIS. L'enjeu de la réduction d'échelle est plus spatial que temporel. Par conséquent, en visant une résolution spatiale de 100/30 m, il subsiste toujours la difficulté de la source de données en considérant l'imagerie MODIS comme unique source.



Les travaux de la présente étude s'inscrivent dans la continuité de l'expérimentation débutée à l'ASPC. Le but de cette étude vise à estimer par réduction d'échelle les déterminants climatiques et microclimatiques à partir des données multispectrales et d'évaluer l'effet de réduction d'échelle sur les résultats aux échelles de 100 m et 30 m.

## **Problématique de l'étude**

Le besoin d'information à l'échelle des territoires pour mieux caractériser les habitats associés aux vecteurs des maladies vectorielles comme celle de Lyme et du virus du Nil Occidental, émergentes au Canada, amène à désagréger des données géospatiales grâce à la réduction d'échelle. La réduction d'échelle utilise les données multispectrales et de sources multiples pour parvenir à une résolution fine. La qualité des résultats obtenus en fin du processus de réduction d'échelle est inhérente à chaque méthode. Elle est dépendante de la qualité des données géospatiales utilisées, du facteur d'échelle (le rapport entre la résolution spatiale de l'image à réduire et celle de la résolution spatiale de l'image réduite), de la morphologie du territoire et du nombre de variables auxiliaires utilisées. Ainsi, elle engendre une perte d'informations due à des pixels mixtes et plus le facteur d'échelle augmente, grande est la marge d'erreur. Or dans le cadre de cette étude, l'ASPC souhaite atteindre une échelle réduite de 100 m et de 30 m avec une marge d'erreur d'au plus de 2 unités partant d'une image de 1000 m de résolution spatiale, soit un facteur d'échelle de 10/33. Ce qui pose le problème de la qualité des résultats attendus.

De même, la réduction d'échelle soulève les enjeux liés au territoire (morphologie du territoire) et au temps (saison/mois), car il existe une multitude de méthodes de réduction d'échelle appliquées selon les états de surface du territoire. La précision des résultats obtenus est également fonction des saisons de production des images sources exploitées. Pour une même méthode de réduction par exemple, la qualité du résultat est meilleure en été qu'en automne pour la température de surface. Ce qui est une limite dans le contexte du Canada, où le territoire à couvrir est composé de 7 régions physiographiques présentant différentes morphologies dans l'occupation des territoires où règnent une multitude de microclimats. En effet, les terres arctiques et subarctiques, les Rocheuses canadiennes, les plaines intérieures, les basses-terres de la baie d'Hudson, le terrain forestier du Bouclier canadien, les basses-terres du Saint-Laurent et les Appalaches sont les 7 régions physiographiques et présentent différentes spécificités biophysiques et climatiques.

Bien qu'il y ait eu des expérimentations de réduction d'échelle à 250 m en lien avec des DCMC associées aux maladies vectorielles à l'ASPC, il n'y a pas, en l'état actuel des connaissances, encore d'études qui aient fait cas de la réduction d'échelle à 100 m et à 30 m en lien avec des DCMC associés aux maladies vectorielles. De

tout ce qui précède, il résulte qu'en dépit des limites des méthodes la réduction d'échelle, elles restent et demeurent à ce jour la réponse face au changement d'échelle. Elles peuvent s'adapter et s'appliquer à cette étude dans l'estimation des DCMC identifiés pour caractériser les habitats des tiques et moustiques.

Les insuffisances identifiées justifient cette étude. Cette recherche s'attardera sur un cas d'étude par une approche comparative dans les régions de l'Estrie, de Montérégie et de Québec en investiguant les différentes stratégies de réduction d'échelle à mettre en œuvre pour la température de surface. Il s'agira aussi d'examiner si l'approche de réduction d'échelle proposée peut s'appliquer à tout DCMC.

## **Question de recherche**

Dans le contexte du projet de l'ASPC, il y a un besoin des données climatiques et microclimatiques à haute résolution spatio-temporelle pour la modélisation de risques liés aux maladies du Virus du Nil Occidental (VNO) et de Lyme à l'échelle locale, le changement d'échelle revêt toute son importance, car il permet de produire des informations détaillées à l'échelle municipale afin de favoriser la mise en place des politiques publiques sur le plan sanitaire à cette même échelle. **Dans un tel contexte, est-ce que les méthodes de réduction d'échelle permettent de déterminer significativement la température de surface (LST en °C) à 100 m et 30 m en milieu hétérogène du Canada en utilisant les images satellitaires MODIS (1000 m) ?**

## **Objectif général**

L'objectif général de ce projet de recherche est de démontrer la faisabilité et la pertinence de réduire l'échelle de données MODIS de 1000 m à 100 m et 30 m pour déduire la température de surface avec une marge d'erreur de 2°Celsius. Ceci revient à une application concrète des méthodes de réduction d'échelle pour montrer leur capacité à fournir des résultats. Ensuite, il faut évaluer la performance des méthodes appliquées en analysant les résultats obtenus à 100 m et 30 m afin de conclure en l'intérêt de leur application. Les résultats de la température de surface réduite aux échelles fines vont servir de base pour l'estimation des DCMC associés aux maladies vectorielles de Lyme et du virus du Nil occidental pour le territoire canadien tout en considérant les grands volumes de données géospatiales multisources.

## **Objectifs spécifiques**

Pour répondre à cet objectif général, il a été décomposé en 4 sous-objectifs :

- Sous-objectif 1 : Identifier les méthodes de réduction d'échelle applicables à ce projet de recherche selon une revue de littérature la plus « exhaustive » possible.

- Sous-objectif 2 : Qualifier les méthodes de réduction d'échelle existantes recensées pour estimer la température de surface MODIS de 1000 m à 100/30 m avec une marge d'erreur d'au plus de 2° Celsius (°C). Dans cet objectif, il sera procédé à la classification des méthodes de réduction d'échelle pour en retenir trois pour être appliquées dans cette étude.
- Sous-objectif 3 : Valider la précision des résultats de LST obtenus à 100 m et à 30 m afin de mieux comprendre les améliorations obtenues. À la suite de l'expérimentation, il s'agit de questionner si avec les images LST de 1000 m réduites à 100/30 m, des précisions suffisantes proches de la réalité ou non ont été obtenues. La comparaison de l'incidence liée à la précision des résultats permettra de déduire la pertinence de viser les deux résolutions spatiales 100/30 m afin de questionner les enjeux de la réduction d'échelle de MODIS LST.
- Sous-objectif 4 : Identifier et formuler les enjeux liés à l'application de la méthode optimale de réduction d'échelle pour l'identification des habitats des tiques et moustiques sur des sites d'étude et sur le territoire du Canada.

## Hypothèse de recherche

Pour atteindre les objectifs de cette recherche, des hypothèses de recherches ont été formulées. L'hypothèse générale (HG) postule qu'une approche méthodologique de réduction d'échelle basée sur les émissivités de surface améliorerait les prédictions de la température de surface à la résolution spatiale de 30 m. En effet, les valeurs d'émissivité basée sur l'occupation et l'utilisation du sol minimisent l'impact des effets atmosphériques qui détériorent les informations contenues dans les images satellitaires. Cette hypothèse générale est déclinée en des sous-hypothèses :

Sous-hypothèse 1: Les méthodes de réduction d'échelle appliquées au niveau des zones d'études hétérogènes permettent d'obtenir une image de MODIS LST avec une marge d'erreur en deçà de 2° Celsius.

Sous-hypothèse 2 : Les modèles de réduction d'échelle permettent d'obtenir une précision spatiale du MODIS LST à 30 m de qualité équivalente à celle obtenue à 100 m.

## Méthodologie de travail

La démarche méthodologique adoptée pour cette étude s'articule en deux étapes complémentaires décrites sous forme de diagramme d'activités illustré par la Figure 0.1: (1) le cadre théorique de la réduction d'échelle dans le contexte des maladies vectorielles et (2) l'application des méthodes de réduction d'échelle retenues dans un contexte spatio-temporel varié.

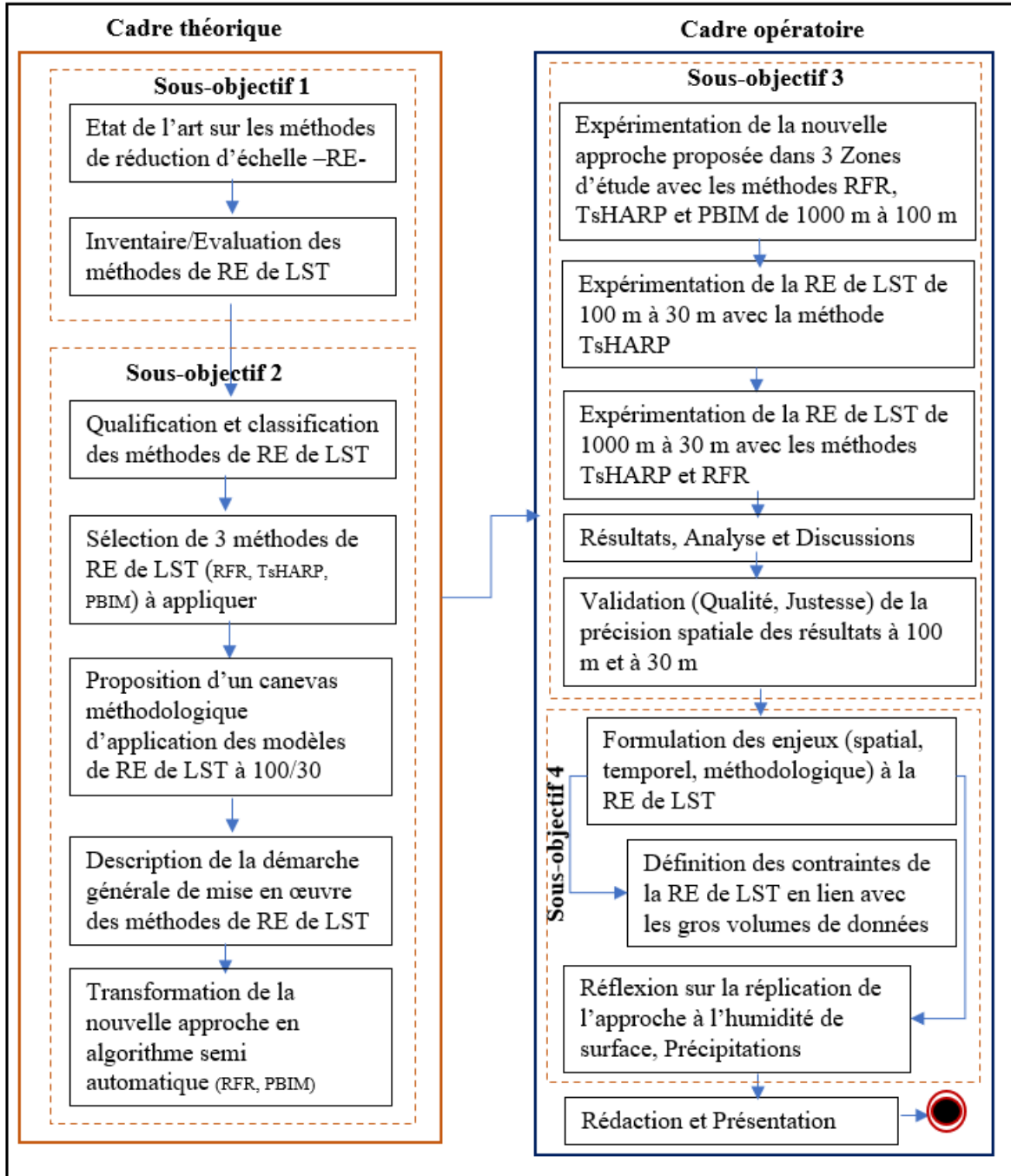


Figure 0.1 : Cadre méthodologique de gestion du projet de recherche

### Étape 1 : Cadre théorique

Le cadre théorique vise à répondre aux sous-objectifs 1 et 2 de cette étude. Il va de l'état de l'art jusqu'à la description des méthodes de réduction d'échelle de LST à appliquer dans un canevas méthodologique de réduction d'échelle à 100/30 m. Les activités du sous-objectif 1 sont l'état de l'art sur les méthodes de réduction

d'échelle et la classification de celles appliquées à la température de surface. La revue de littérature, transversale dans toute l'étude, a permis de circonscrire ce projet d'étude après l'état des connaissances sur la réduction d'échelle, les domaines et contexte d'application puis, de dégager les faiblesses et atouts des méthodes générales de réduction d'échelle. Ensuite ce même exercice d'évaluation (avantages/limites) a été fait pour les méthodes de réduction d'échelle de LST recensées.

S'agissant du sous-objectif 2, il se subdivise en 5 activités sur les méthodes de réduction d'échelle de la température de surface. La première est la qualification et la catégorisation des méthodes de réduction d'échelle de LST selon les modèles mathématiques ou statistiques et l'algorithmique. Cette classification a débouché sur l'analyse de performance comparative entre les méthodes de réduction d'échelle de LST en se basant sur les limites et avantages identifiés dans l'état de l'art. Ensuite, la deuxième activité a porté sur la sélection de trois méthodes de réduction d'échelle de LST – TsHARP (Thermal sharpening), PBIM (Pixel Block Intensity Modulation), RFR (Random Forest Regression) – suivant des critères généraux et spécifiques au projet. De manière spécifique, le tri entre plusieurs méthodes de réduction d'échelle de LST a été fait sur des critères généraux comme la typologie des modèles, l'algorithmique, les données géospatiales en entrée et l'unité des LST (Celsius ou Kelvin). Les critères du projet sont par exemple, la marge d'erreur, la saison d'étude, le type d'image satellitaire de température de surface ou la répliquabilité de la méthode en tout lieu et saison. La troisième activité a été la proposition d'un canevas de réduction d'échelle en deux étapes pour réduire la MODIS LST de 1000 m à 100 m et de 100 m à 30 m. La nouvelle approche de réduction d'échelle résulte de la connaissance descriptive du fonctionnement des méthodes de réduction d'échelle, des données utilisées et des méthodes de validation des résultats, le cheminement général des méthodes de réduction d'échelle de LST. De fait, en se basant sur l'appropriation de ce processus général et des limites des méthodes de réduction d'échelle de LST, un nouveau canevas pour atteindre une résolution fine de 30 m a été défini. La quatrième activité a été la description de la démarche générale de mise en œuvre du nouveau canevas de réduction d'échelle de LST proposé. Ainsi, la réalisation de cette activité a nécessité des tâches comme le choix des zones d'étude, la description des données, les outils de traitements et d'analyse des données et la procédure de validation des résultats. La dernière activité a consisté à la semi-automatisation des chaînes de traitements de réduction d'échelle de LST pour les méthodes RFR et PBIM. À cette fin, une collaboration avec des étudiants en génie géomatique a permis de traduire le canevas méthodologique proposé en algorithme semi-automatique pour permettre la célérité des traitements des données et faciliter les expérimentations. Ce cadre théorique défini à l'étape 1 a été opérationnalisé dans la seconde étape.

## **Étape 2 : Application des méthodes de réduction d'échelle retenues**

L'étape 2 est consacrée à l'expérimentation des méthodes de réduction d'échelle de LST, les résultats obtenus, l'analyse de ces résultats et leur discussion pour répondre aux sous-objectifs 3 et 4 de l'étude. Le cadre conceptuel défini à l'étape 1 sert de référence pour des expérimentations à faire avec des données de zones spatio-temporelles différentes. Le sous-objectif 3 comprend 3 activités : une portant sur les expérimentations, une autre consistant à la présentation des résultats, leur analyse suivit des discussions et la dernière activité est la validation des résultats selon le contexte de l'étude et le seuil de 2°C. Trois grandes expérimentations ont été réalisées: la première est la prédiction de MODIS LST de 1000 m à 100 m, puis de 100 m à 30 m à la deuxième expérimentation. La troisième expérimentation est la prédiction de MODIS LST de 1000 m à 100 m, et a été réalisée pour vérifier les résultats obtenus dans les deux premières expérimentations. De fait, nous avons identifié trois zones d'étude dans la province de Québec en nous basant sur des rapports documentés par le laboratoire national de santé publique sur les profils épidémiologiques des maladies du Virus du Nil Occidental (VNO) et de Lyme à l'échelle du Canada et de chaque province. Les données géospatiales relatives aux trois zones d'étude ont été téléchargées dans la banque de données de Earthexplorer pour faire les expériences. Les résultats obtenus ont été analysés et discutés. L'analyse des résultats a permis d'évaluer le comportement et la performance des méthodes en zone géographique variée à des dates différentes. Il s'agit notamment d'apprécier dans cette analyse la précision spatiale des résultats par rapport au seuil de 2°C avec des données de référence à 100 m et à 30 m. Cette analyse donne également lieu à des discussions sur les gains réels obtenus en appliquant les méthodes de réduction d'échelle à 100/30 m en se basant sur la revue de littérature et de questionner les résultats en prenant en compte le contexte du projet de l'ASPC.

Le sous-objectif 4 décomposé en 3 activités que sont la formulation des enjeux de réduction d'échelle de LST, la définition des contraintes liées à l'application des méthodes de réduction d'échelle avec les gros volumes de données et une analyse sur l'adaptabilité du nouveau canevas proposé à d'autres déterminants climatiques (humidité de surface, précipitations). Les enjeux de la réduction d'échelle de LST sont en lien avec la morphologie des zones d'étude et aux temporalités des données. Les résultats contextualisés permettent également de réfléchir sur les stratégies de mise en œuvre des méthodes de réduction d'échelle tenant compte des objectifs de l'étude et des contextes environnementaux du Canada. Par ailleurs, le temps de traitement des données et la qualité des résultats ont permis de discuter sur les modalités de traitement, la préparation des données, les goulots d'étranglement dans la gestion des gros volumes de données. Pour finir, on s'est posé la question de savoir si l'approche méthodologique de réduction d'échelle de LST à 30 m et les méthodes appliquées peuvent s'appliquer à l'humidité de surface et les précipitations. Ces deux étapes ont permis de tester les hypothèses de recherche afin de répondre à la question de recherche et aux objectifs de l'étude.

## **Structure du document**

Le mémoire est organisé en cinq (5) chapitres hormis la conclusion et l'introduction. Le chapitre 1 est consacré au cadre théorique de la réduction d'échelle, les méthodes de réduction d'échelle de la température de surface et la sélection des méthodes de réduction d'échelle de LST à appliquer dans cette étude. Le chapitre 2 est dédié à la description du cadre méthodologique d'application des méthodes de réduction d'échelle pour atteindre une résolution spatiale de 100/30 m. On y aborde aussi la description des étapes générales de la méthode de réduction d'échelle puis l'application opérationnelle de ces méthodes pour atteindre 100 m et 30 m. Quant au chapitre 3, il présente la méthodologie à suivre pour l'implémentation des méthodes de réduction d'échelle dans le contexte environnemental de Québec. Les activités à réaliser pour chacune des trois méthodes sont décrites en plus de la présentation des zones d'étude et les stratégies de validation des résultats. Les résultats des tests réalisés sont présentés au chapitre 4. Les analyses des résultats et les discussions sont présentées au chapitre 5. Au terme des discussions des résultats, il y a la conclusion suivie des perspectives qui font la synthèse de cette recherche et les potentielles orientations futures.

# **Chapitre 1 - État de l'art sur la réduction d'échelle et des méthodes de réduction d'échelle de la température de surface (LST)**

## **1.1. Introduction**

Ce thème de recherche met en exergue la réduction d'échelle à partir des images satellitaires pour prédire les déterminants climatiques et microclimatiques, plus précisément la température de surface. L'état des connaissances est axé sur les méthodes de réduction d'échelle et leurs limites puis celles appliquées uniquement à la température de surface. Ainsi, ce chapitre présente premièrement les grandes familles de méthodes de réduction d'échelle et le contexte lié à leur utilisation. En second lieu, les méthodes de réduction d'échelle existantes sur les températures de surface, leurs avantages et limites sont décrits.

## **1.2. Méthodes de réduction d'échelle et leurs limites**

La réduction d'échelle est assimilée au passage d'une échelle plus petite à une plus grande échelle pour passer des informations généralisées à des informations détaillées. En télédétection, c'est le fait de passer d'une résolution spatiale grossière de 1km par exemple à une résolution spatiale fine de 30 m. La réduction d'échelle permet de faire des projections spatiales à des échelles fines partant des hypothèses ou connaissances établies sur la relation entre les résolutions spatiales (Atkinson, 2004). La relation entre les résolutions spatiales est une information auxiliaire à la variable étudiée permettant l'opérationnalisation du processus. C'est le cas par exemple avec l'estimation de la température de surface en se basant sur l'indice de végétation normalisée (NDVI), l'humidité de sol (Ebrahimi & Azadbakht, 2019) ou la topographie et le NDVI (Bartkowiak et al., 2019).

L'usage de la réduction d'échelle est dû aux besoins d'échelles fines pour des prédictions en hydrologie, climatologie, agrométéorologie, pédologie afin de prévenir les risques de sécheresses, d'inondation ou de canicule. Dans la littérature, la réduction d'échelle est plus utilisée dans l'estimation des paramètres climatiques que les paramètres environnementaux. Les paramètres estimés sont ainsi principalement la température de surface de sol (Bartkowiak et al., 2019 ; Ebrahimi & Azadbakht, 2019 ; Singh Rawat et al., 2019), l'humidité de surface de sol (Peng et al., 2017), l'évapotranspiration (Baldocchi et al., 2019 ; Zhang et al., 2019). La réduction d'échelle utilise différentes informations biophysiques (indices spectraux, par exemple), topographiques (altitude, pente ou orientation) en lien avec la variable étudiée pour passer à une résolution spatiale plus fine que celle de l'image d'entrée au moyen de différentes méthodes. Ces données biophysiques et topographiques sont des variables auxiliaires appelées des prédicteurs et sont à haute ou très haute résolution spatiale.



Au plan opérationnel, la réduction d'échelle s'opère comme la décomposition d'une valeur moyenne d'un pixel en plusieurs pixels de valeurs uniques sur la même superficie (cf. Figure 1.1). La figure ci-dessous montre le fractionnement avec 3 types de méthode (i.e. basée sur des images satellitaires, la géométrie ou la modélisation précédemment décrite) chacune des cases (résolution spatiale 25 km\* 25 km) du grand carré en de petites cases (résolution spatiale de 1 km \* 1 km). À la fin de ce processus, on constate une dispersion et un mélange des couleurs dans l'image obtenue. Ainsi si la case remplie de couleur blanche du grand carré apparaît comme une unité de conifères à 25 km, sur l'image réduite on retrouve 25 petites cases de 1km dont certaines de couleur blanche et d'autres de couleur grise. La couleur grise, inexistante avant dans la case de 25 km, apparaît dans les petites cases de 1 km et peut être assimilée à une unité de feuillus. C'est ce qui est désigné par le mélange de pixels dans les unités d'occupation du sol. D'un point de vue de déterminant climatique et microclimatique, si l'unité de conifère a une température de surface moyenne de 18°Celsius dans la grande case par exemple, la réduction d'échelle doit permettre une décomposition de cette valeur moyenne selon l'abondance de conifères ou de feuillus dans chacune des 25 petites cases.

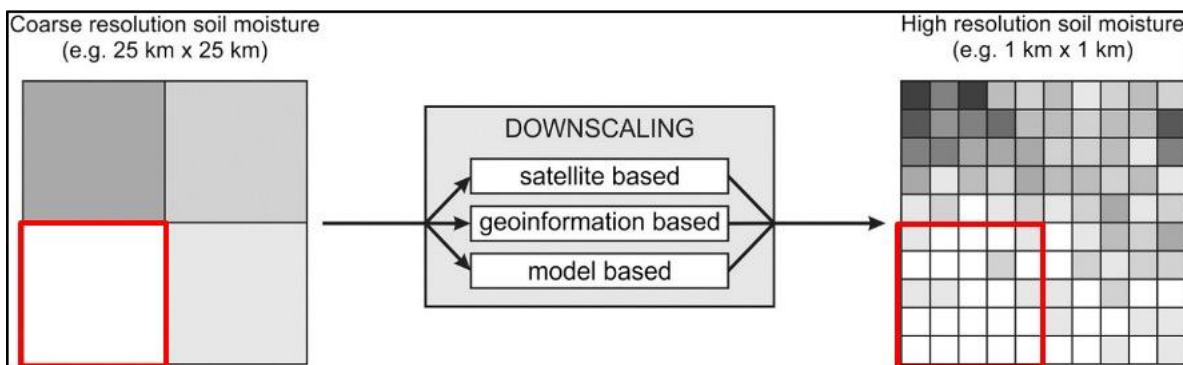


Figure 1.1 : Schématisation du processus de réduction d'échelle  
Source : Peng et al, 2017

L'objectif de la réduction d'échelle est de prédire l'emplacement de chaque classe de couverture terrestre à l'intérieur de chaque pixel de l'image de sortie. Autrement, il y a une incertitude sur la valeur moyenne de chaque classe de couverture dans chaque pixel de sortie.

Cette section présentera en deux sous-sections la typologie des méthodes de réduction d'échelle et leurs limites.

### 1.2.1. Méthodes de réduction d'échelle

La réduction d'échelle utilise différentes méthodes et techniques. Ces méthodes diffèrent en fonction d'une part, de la nature de la variable à prédire, ou des objectifs poursuivis (Atkinson, 2013) et d'autre part, selon le type de données d'entrée (i.e. données radar, données optiques/thermiques) et des caractéristiques du modèle d'échelle (i.e. physique et statistique) (Peng et al., 2017). La variable d'intérêt est de nature discrète (i.e.

l'humidité du sol ou la température) ou continue (i.e. la biomasse, l'indice de végétation normalisé...). Les méthodes de réduction d'échelle sont des modèles de régression, de géostatistique, statistique ou de télédétection faisant appel à des algorithmes pour l'extraction d'informations. L'ensemble des méthodes existantes établissent une relation de liaison statistique ou un modèle physique entre la variable d'intérêt à une échelle grossière et les variables auxiliaires à une échelle fine (Bierkens et al., 2001).

De manière générale, les différentes méthodes de réduction d'échelle en télédétection peuvent être catégorisées en trois groupes (Peng et al., 2017): (1) méthodes par satellite, (2) méthodes utilisant des données géospatiales et (3) méthodes basées sur des modèles. Chacune des catégories est subdivisée en différentes approches et chaque approche en des méthodes (cf. Tableau 0.1 en annexe A).

1. Les **méthodes par satellite** utilisent les images satellitaires. Elles consistent à fusionner une image à faible résolution spatiale avec une image de haute résolution spatiale du même capteur ou d'un capteur différent afin d'obtenir une image à résolution spatiale et spectrale fine. C'est le cas du capteur MODIS où deux images A et B de résolution spatiale respective de 1000 m et 250 m ont été fusionnées pour obtenir en résultat des informations contenues dans l'image A à 250 m. Dans ce groupe de méthodes, on peut citer par exemple la méthode de détection de changement, celle de fusion bayésienne ou la modélisation bayésienne. Les méthodes par satellites ont l'avantage d'être simples, faciles à appliquer et ont un grand potentiel d'amélioration de la résolution spatiale.
2. Les **méthodes utilisant les données géospatiales** sont des méthodes mixtes, combinant les méthodes par satellites et les modèles. Ces méthodes se basent sur des observations au sol, les attributs topographiques à résolution fine (Werbylo & Niemann, 2014) collectés sur le terrain ou extraits des images satellitaires utilisées pour réduire l'humidité de surface. Dans ces méthodes, il y a les méthodes de la fonction orthogonale empirique (EOF) et le modèle Equilibrium Moisture from Topography (EMT).
3. Les méthodes basées sur des modèles utilisent des modèles statistiques ou de surface qui peuvent être soit déterministes ou stochastiques (Bierkens et al., 2001). Ces méthodes arrivent à établir des liens directs entre les images grossières et celles prédites à fine échelle à partir des fonctions de distribution de probabilité ou de densité de probabilité. Comme exemple de méthodes basées sur des modèles, il y a la méthode Thermal Shapering (TsHARP) ou la méthode de lissage Laplacien (TPS) (Singh Rawat et al., 2019), la méthode de régression forestière aléatoire, la méthode de CoKrigage (Atikens, 2014), etc.

## 1.2.2. Limites des méthodes de réduction d'échelle

Plusieurs limites associées aux méthodes de réduction d'échelle ont été identifiées. Ces limites sont relatives à la qualité des résultats, aux incertitudes associées aux différentes méthodes et au mode de validation des résultats.

### 1.2.2.1. *Qualité des résultats*

Il ressort de plusieurs études que l'exactitude des résultats obtenus par réduction d'échelle est fonction du processus utilisé, du facteur de réduction d'échelle, de l'incertitude des données géospatiales (Peng et al., 2017 ; Wu & Li, 2009 ; Zawadzka et al., 2020). En effet, plus le facteur de réduction d'échelle, qui est le rapport entre la résolution spatiale observée et la résolution spatiale réduite (Zhou et al., 2016), est élevé, plus grande est la marge d'erreur de l'image réduite. Par exemple une image satellitaire de 1000 m de résolution spatiale réduite à 100 m est moins altérée que la même image réduite à 20 m, car le facteur d'échelle respectivement, de 10 et 50, est plus petit dans le premier cas que le second. Les principales raisons de cette dégradation sont la variabilité spatiale des informations à différentes échelles et l'hypothèse de dépendance ou d'indépendance entre les variables auxiliaires et la variable à réduire (Zhou et al., 2016). En outre le facteur d'échelle, les incertitudes liées aux données d'imagerie satellitaire alimentant les modèles peuvent accroître les marges d'erreur en raison des erreurs systématiques liées aux algorithmes de récupération des informations à partir de chaque capteur ou au système d'enregistrement de l'image (Agili et al., 2014 ; Lunetta, 1991 ; Toutin, 2004 ; Yang & Weng, 2011). Ces erreurs proviennent également des parasites et bruits tels les nuages, gaz et aérosols présents dans l'atmosphère et leur élimination nécessite des méthodes de correction atmosphérique (Agili et al., 2014 ; Toutin, 2004).

La marge d'erreur peut aussi être engendrée par le processus de réduction d'échelle. D'une part, l'opérationnalisation de la réduction d'échelle se base sur des données auxiliaires, comme les indices spectraux (le NDVI, la couverture fractionnée de végétation) et la topographie dérivés des images satellitaires optiques à haute ou très haute résolution. Les incertitudes associées à la détermination de ces indices peuvent se propager à toutes les étapes du processus de réduction d'échelle et augmenter les marges d'erreur. De plus, l'éventuel rééchantillonnage à appliquer sur ces données auxiliaires peut donner lieu à des pertes d'informations, si les résolutions auxquelles elles existent ne sont pas celles à atteindre.

D'autre part, la considération ou non des interrelations existantes entre les déterminants climatiques et microclimatiques (DCMC) dans la procédure d'estimation peut donner lieu à des biais dans les résultats. Pour certains modèles par exemple, l'estimation de l'humidité de surface nécessite au préalable la détermination de

la température de surface dépendante à son tour des indices spectraux et de l'émissivité à calculer. Ainsi, les sources d'erreur, augmentant, conduisent à une erreur cumulative et soulèvent le problème de la fiabilité des résultats. Si les indicateurs d'erreur, l'erreur moyenne absolue ou l'erreur quadratique moyenne servent à mesurer la qualité des résultats dans différentes études, aucune de ces études n'a mentionné la valeur seuil à partir de laquelle on peut conclure à la bonne qualité des résultats. La plupart de ces études considèrent souvent la plus petite valeur de RMSE pour déclarer bon un résultat obtenu.

#### *1.2.2.2. Incertitudes associées aux méthodes de réduction d'échelle*

La synthèse de littérature révèle l'existence d'une variété de méthodes de réduction d'échelle développées en fonction des thématiques d'étude. Par exemple pour la température de surface, il y a plusieurs méthodes telles : Disaggregation procedure for radiometric surface temperature (DisTrad) (Kustas et al., 2003), Thermal Shapening (Agam et al., 2007; Kustas et al., 2003), Forêt aléatoire de régression (Random Forest Regression -RFR), Machine d'apprentissage extrême (Extreme Learning Machine -ELM) (Ebrahimi & Azadbakht, 2019), etc. Pour l'humidité de surface, il y a la méthode de détection des changements, la méthode de fusion bayésienne, la méthode DISaggregation based on Physical And Theoretical scale Change (DISPATCH), l'algorithme de fusion active passive (Malbêteau, 2017 ; Peng et al., 2017). Les modèles probabilistes tels que le modèle linéaire généralisé (GLM), la forme vectorielle des modèles linéaires généralisés (VGLM), la forme vectorielle des additifs généralisés (VGAM), les densités des réseaux de neurones conditionnels, les distributions Pareto généralisées (GP) ou la régression des moindres écarts absolus (LAD Least absolute deviation) sont souvent utilisés pour prédire les précipitations à une fine échelle. Ces méthodes pour la plupart basées sur modèles statistique ou physique se différencient cependant par les conditions liées à leur fonctionnalité.

Certaines méthodes font appel à des équations paramétriques avec l'hypothèse de l'existence d'une relation linéaire et invariable entre les variables d'étude. En revanche, d'autres méthodes sont axées sur des modèles non paramétriques avec l'hypothèse de l'absence de relation linéaire et invariable entre les variables. Ces hypothèses influencent la prise en compte de tous les processus ou facteurs liés au phénomène étudié dans la structure des modèles élaborés. La diversité des méthodes est liée aux insuffisances constatées au niveau de certaines d'entre elles ou la recherche d'une grande exactitude dans les résultats. En effet, l'exactitude des résultats dépend de l'hétérogénéité ou l'homogénéité des états de surface de la zone d'étude. Cette diversité de méthodes révèle donc l'influence des états de surface sur les modèles associés aux différentes méthodes et aussi l'absence de consensus sur et autour d'une méthode de réduction d'échelle. Il en résulte la non-réplication

des méthodes dans toute zone d'étude en l'absence de toute similitude à cause de la variabilité dans l'espace des paramètres étudiés (Peng et al., 2017).

D'un point de vue fonctionnel, toutes les méthodes de réduction d'échelle sont sensibles à certaines unités d'occupation du sol ou à des valeurs extrêmes ou encore à la saison (Agam et al., 2007 ; Bindhu et al., 2013 ; Peng et al., 2017 ; Zawadzka et al., 2020). Cette sensibilité peut engendrer des erreurs. Des études menées dans des zones hétérogènes ou homogènes ont montré par exemple la faible capacité de n'importe quel algorithme à simuler parfaitement la température de surface dans les zones humides. C'est aussi le cas en hiver où la présence de la glace et la neige à la surface du sol influe sur la justesse des températures de surface réduites pour tout modèle (Ebrahimi & Azadbakht, 2019 ; Mechri et al., 2016 ; Stathopoulou & Cartalis, 2009 ; Yang, Cao, et al., 2017). Ce qui montre l'inadaptation des méthodes selon les saisons. D'autres études ont montré l'insuffisance des modèles à simuler adéquatement les valeurs de température de surface les plus basses et les plus chaudes à partir des images satellitaires (Bindhu et al., 2013 ; Zawadzka et al., 2020).

### *1.2.2.3. Inadéquation des méthodes de validation des résultats obtenus par des méthodes de réduction d'échelle*

Selon Merlin et al. (2016), l'évaluation d'une méthode de réduction d'échelle reste un exercice délicat en raison de toutes les sources d'erreur propagées à toutes les étapes du processus, en l'absence de mesures de terrain spatialisées et l'éventuelle non-représentativité des mesures faites sur le terrain. Idéalement, pour une meilleure comparaison, les mesures in situ devraient être densifiées et se faire au même moment que le satellite passe dans la zone d'étude. La réalisation de ce protocole exigerait de gros investissements en temps, équipements et moyens humains. Pour pallier cette difficulté, l'alternative trouvée est le recours à l'utilisation des images satellitaires de Landsat et Aster (à 100 m de résolution spatiale) pour la LST dont le passage est quasi le même que le capteur imageur MODIS. Ce qui n'est cependant pas le cas avec l'humidité de surface. En effet, bien que les données radar aient un potentiel de suivi à haute résolution, il n'existe pas à ce jour d'algorithme opérationnel d'inversion de l'humidité de surface à partir des données fournies par radar (Malbêteau, 2017 ; Rodenas, 2017). Et même si cela est possible, le processus d'obtention par inversion reste limité sur une période de 1 mois et sur des superficies de moins de 1000 km<sup>2</sup>.

Dans ce contexte, une nouvelle stratégie de validation des données spatiales réduites a été proposée à l'aide de l'indicateur  $G_{down}$ . (Merlin et al., 2010). La plus-value de cet indicateur est d'évaluer le gain entre l'image réduite et l'image brute observée. Cette nouvelle approche s'est révélée performante pour l'humidité de surface plus en été qu'en hiver et dans les zones semi-arides que tempérées. En résumé, la validation des images réduites reste complexe et limitée dans le temps et l'espace.

La plupart de ces méthodes présentent encore des déficiences dans leur stratégie de validation. Les validations sont séquentielles et se font avec des données mesurées au sol dont les valeurs diffèrent selon le capteur utilisé. Ceci limite la robustesse des modèles et la qualité des résultats obtenus bien que ceux-ci présentent des marges d'erreur relativement faibles. La mise en place d'un référentiel unique de validation permettra une comparaison objective entre les méthodes.

En conclusion, il n'y a pas de consensus autour d'une méthode unique de réduction d'échelle encore moins pour une méthode de réduction d'échelle selon la thématique étudiée. Par exemple, Peng et al. (2017) ont montré dans leur article sur l'humidité de surface que 13 méthodes ont été utilisées. Nous avons également recensé plus de 10 méthodes appliquées à la réduction d'échelle de température de surface. Cette pluralité de méthodes est liée à la morphologie du territoire, aux saisons, aux intrants dans les modèles et le besoin d'une grande exactitude des résultats. Certaines méthodes de réduction d'échelle se basent sur les hypothèses de linéarité et stationnarité entre les variables étudiées ; ce qui est une simplification de la réalité et dont l'impact pourrait être l'imprécision dans les résultats. Celles qui postulent à l'absence de toute relation linéaire et invariable entre les variables ne prennent pas en compte tous les facteurs ou conditions dans la schématisation des modèles décrivant le phénomène étudié. Ainsi, la simplification des modèles peut engendrer des biais avec la réalité. Aucune des études n'a véritablement évalué la performance liée au temps de calcul de chacune des méthodes. Le choix d'une méthode dépend de la thématique étudiée, des objectifs poursuivis et des données à utiliser ou disponibles.

Après avoir décrit la typologie des méthodes de réduction d'échelle, la section suivante présente les méthodes de réduction d'échelle de la température de surface recensées dans la revue de littérature.

### **1.3. Méthodes de réduction d'échelle appliquées à la LST**

Les méthodes de réduction d'échelle de LST se répartissent en deux catégories : celle des méthodes basées sur la fusion et la catégorie des méthodes sur les noyaux (Bai et al., 2015 ; Feng Gao et al., 2006 ; Mao et al., 2021 ; Tang et al., 2021 ; Weng et al., 2014 ; Zhu et al., 2010). Nous nous focalisons plus sur les méthodes axées sur les noyaux que celles basées sur la fusion, car il ressort de la revue de littérature menée que les premières sont les plus utilisées à cause de leur facilité d'utilisation et performance (Ebrahimi & Azadbakht, 2019 ; Yangbao et al., 2017). Les méthodes basées sur les noyaux se subdivisent, au plan algorithmique, en approches statistiques et physiques (Weng et al., 2014).

Les **approches physiques** exploitent les relations physiques entre la LST et les facteurs liés à l'état de surface pour procéder à la réduction d'échelle. C'est le cas des méthodes de modulation d'intensité par bloc de pixels,

Pixel Block Intensity Modulation -PBIM- (Feng & Li, 2020 ; Guo & Moore, 1998) ou de modulation de l'émissivité (Chybicki & Łubniewski, 2017 ; Nichol, 2009 ; Stathopoulou & Cartalis, 2009). Les **approches statistiques** établissent une relation statistique entre la LST et les variables auxiliaires à haute résolution spatiale issues des bandes d'ondes plus courtes pour réduire la LST. Dans cette rubrique les modèles de réduction d'échelle de LST sont subdivisés en deux classes : la classe des modèles linéaires et la classe des modèles non linéaires.

- Les modèles linéaires sont fondés sur le postulat d'une relation linéaire et stationnaire entre la LST et un indice spectral dominant en faisant fi de la variabilité spatio-temporelle. Nous pouvons citer les modèles de régression statistique linéaire DisTrad, TsHARP, LM<sub>SD</sub>, PR<sub>DS</sub>, etc. (basé sur les indices de végétation comme le NDVI, la couverture végétale fractionnée -CVF-). Les modèles linéaires ont donné des résultats satisfaisants pour les milieux homogènes à dominance de végétation et où le sol est homogène avec peu de variations d'air (Gao et al., 2012). Par contre, pour des milieux hétérogènes, les résultats obtenus présentent des incertitudes, quoiqu'acceptables (Crosson et al., 2020 ; Inamdar & French, 2009 ; Jeganathan et al., 2011 ; Mukherjee et al., 2014 ; Peng et al., 2019). Ces incertitudes révèlent les limites de l'hypothèse de linéarité et d'invariabilité de la LST et des indices de végétation dans les milieux hétérogènes. Les limites des modèles linéaires ont été prises en compte dans les modèles non linéaires.
- Les modèles non linéaires sont des modèles de régression multiple non linéaire, basés sur la relation entre la LST et les paramètres biophysiques, les facteurs de terrain (topographie, aspect, ombre) et l'occupation du sol. Ainsi, outre les indices de végétation comme prédicteurs, ces modèles font appel aux différentes classes d'occupation du sol, la topographie, l'albédo de surface, l'émissivité de surface, l'état de l'humidité du sol, l'angle d'incidence solaire et l'angle de vue du ciel. Ils sont basés sur l'hypothèse que la relation existante entre la LST et les variables auxiliaires à une basse échelle peut aussi s'appliquer à des échelles fines (Yang, Li, et al., 2017). En exemple de modèles de régression non linéaire appliqués à la réduction d'échelle de LST, il y a Non Linear DisTrad (Bindhu et al, 2013), Geographically and Temporally Weighted Regression – GTWR- (Peng et al, 2019), Random Forest Regression -RFR- (Bartkowiak et al., 2019 ; Bisquert et al., 2016 ; Ebrahimi & Azadbakht, 2019 ; Hutengs & Vohland, 2016), etc. Il ressort de plusieurs études que les modèles de régression non linéaire ont donné de bons résultats avec moins d'incertitudes en milieux hétérogènes par rapport aux modèles de régression linéaire (Bartkowiak et al., 2019 ; Bindhu et al., 2013 ; Ebrahimi & Azadbakht, 2019 ; Gao et al., 2012 ; Guijun Yang et al., 2010 ; Hutengs & Vohland, 2016 ; Mukherjee et al., 2014 ; Peng et al., 2019a ; Wu & Li, 2019). Contrairement aux modèles linéaires, plus simples et faciles d'utilisation, les modèles non linéaires sont difficiles à mettre en œuvre (Ebrahimi & Azadbakht, 2019 ; Yangbao et al., 2017 ; Hutengs et Vohland, 2014) à cause de leur complexité structurelle à intégrer une multitude de variables de prédiction.

Cette section, subdivisée en 4 sous-sections, décrit en premier lieu la température de surface et sa variabilité spatio-temporelle, suivie de l'évaluation des méthodes de réduction d'échelle de LST (°C), puis de l'application des trois méthodes retenues pour cette étude. La dernière sous-section est consacrée à la description des méthodes sélectionnées pour les expérimentations.

### 1.3.1. Température de surface (Land surface temperature, °C)

Cette sous-section traite de la variabilité spatiale temporelle de la température de surface et sa mesure à l'aide des images satellitaires.

#### 1.3.1.1. *Température de surface et les facteurs affectant sa variabilité spatio-temporelle*

La température de surface (LST, °C) est un paramètre clé utilisé dans l'estimation de l'évapotranspiration, de la teneur en humidité des sols, des effets de l'îlot de chaleur urbain, etc. La LST est comme une interface entre la terre et l'atmosphère et permet de saisir l'interrelation entre elles par son état d'équilibre thermodynamique (Kotchi, 2015 ; WenFeng et al., 2013). Selon Gillespie (2014), la LST est la température moyenne d'un élément de la surface exacte de la terre calculée à partir de la radiance mesurée (Hulley et al., 2019 ; Njoku, 2014). La température de surface est la somme relative de la température du couvert végétal (canopée) dominant et la température du sol, toutes deux définies en fonction de la couverture végétale fractionnée (Kotchi, 2015). Ainsi, la température de surface mesurée par le capteur infrarouge thermique intègre tous les objets circonscrits dans sa résolution spatiale.

La température de surface (°C), résultat du bilan énergétique de la surface terrestre (Cao et al., 2008), connaît une grande variabilité spatio-temporelle à toute échelle. À l'échelle locale, les différents facteurs prédominants dans cette variabilité sont l'albédo de surface, l'humidité du sol, les conditions d'émissivité, les unités d'occupation et d'utilisation des sols, le bilan radiatif, la topographie, l'activité photosynthétique de la végétation et les facteurs météorologiques (Crum & Jenerette, 2017; Fridley, 2009; Ibrahim et al., 2016; Inamdar & French, 2009; Mirchooli et al., 2020; Sandholt et al., 2002; Small, 2006; Zhang, Zhong, Wang, et al., 2009a; Zhang, Zhong, Feng, et al., 2009b).

Fridley (2009) a montré dans ses recherches que la variation sur de fines échelles spatiales (moins de 1000 m d'étendue) des régimes de température à proximité du sol dans les paysages complexes est due au processus topographique (altitude, pente, aspect). D'autres études (Cao et al., 2008 ; Mirchooli et al., 2020 ; Zhang et al., 2013) ont constaté que le développement urbain entraînant des modifications dans les états de surface a pour conséquence une augmentation de la température de surface. En effet, selon la synthèse bibliographique de ces études, la dégradation de la couverture végétale dans une zone modifie les processus thermiques de conduction, de convection et de rayonnement qui dépendent des propriétés biophysiques de la végétation et des unités d'occupation du sol. Zhang et al, (2009a) étudiant les relations entre la LST et la végétation en milieu urbain, ont bien identifié l'influence significative de la couverture de végétation fractionnée sur l'environnement thermique. En plus, l'étude souligne que l'abondance de la végétation et les surfaces imperméables sont les



principaux facteurs de couverture terrestre qui contribuent plus aux variations de LST (Zhang, Zhong, Feng, et al., 2009b). Dans cette même lignée, une autre étude quantifie la relation LST et la végétation et mentionne que la structure, la composante et la distribution spatiale de la végétation déterminent également la variabilité spatiale et temporelle de la LST (Zhang, Zhong, Wang, et al., 2009a). Small (2006) à travers une analyse comparative de la réflectance urbaine et la LST a quantifié la dépendance de la LST agrégée vis-à-vis de chaque unité d'occupation et d'utilisation des sols dans 24 divers milieux urbains. Comme résultat, une forte relation de dépendance a été trouvée entre les variations globales des champs thermiques urbains et les effets de l'abondance de la végétation, de l'exposition du sol, de l'albédo et de l'ombrage (Small, 2006). S'agissant des variations temporelles (horaire, jour, nuit, saison), Filon (2015) a identifié les éclaircissements et la température de l'air dans les paramètres climatiques qui influent sur la LST. Ces variations temporelles peuvent être aussi instantanées à cause de l'état de l'atmosphère, les propriétés de la surface, le vent et l'humidité du sol (Coops et al., 2007). La température de surface est dérivée des mesures de radiance réalisées avec les capteurs imageurs.

#### *1.3.1.2. Images satellitaires, mesures de LST (°C) et changement d'échelle*

En raison de la grande variabilité spatio-temporelle de la température de surface, même l'interpolation des températures de l'air mesurées par des stations météorologiques, qu'importe leur densité et bonne répartition spatiale, ne peut refléter la réalité ou permettre de saisir avec précision la distribution spatiale de la température de surface (Beltrami et al., 2003 ; Coops et al., 2007 ; Xu et al., 2014). La télédétection satellitaire permet dans ce contexte de combler cette insuffisance, car elle permet d'avoir des informations continues sur de grandes échelles. La plupart des capteurs imageurs fournissent des informations de LST dans l'infrarouge thermique. Les images satellitaires de type Landsat, MODIS (Moderate Resolution Imaging Spectroradiometer), ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer) et AVHRR (Advanced Very High Resolution Radiometer) fournissent des LST à travers leurs bandes infrarouges thermiques (Essa et al., 2012; Kustas et al., 2003). Les caractéristiques radiométriques, les résolutions spectrale, spatiale et temporelle spécifiques de ces capteurs déterminent celles des images satellitaires de LST mises à disposition des utilisateurs. Les capteurs Landsat et ASTER ont une résolution spatiale variant entre 60 m à 100 m pour les bandes thermiques avec une résolution temporelle de 16 jours. Par contre, les bandes TIR des capteurs imageurs MODIS et AVHRR ont une résolution spatiale principale respective de 1000 m, 1100 m et une résolution temporelle d'un jour (cf. Tableau 0.2 en annexe A). De tous ces capteurs météorologiques, le capteur MODIS offre plus de possibilités pour les études de la température à long terme à diverses échelles spatiales (Zhang et al., 2013). De plus dans le contexte de cette étude où des variabilités instantanées de la LST ont un fort impact sur les mouvements des vecteurs, le capteur MODIS offre également la possibilité d'un suivi horaire

de tous les autres déterminants environnementaux, climatiques et microclimatiques excepté les précipitations (Kotchi et al., 2019).

Au cours des vingt dernières années, la télédétection satellitaire a réalisé d'énormes progrès dans l'amélioration des résolutions temporelle et spatiale des capteurs imageurs multispectrales depuis leur avènement en 1972. La très haute résolution spatiale est de l'ordre métrique ou centimétrique avec les images Ikonos<sup>7</sup>, Quickbird<sup>8</sup>, Orbview<sup>9</sup>, Sentinel<sup>10</sup> pour fournir plus de détails et de précision dans les analyses environnementales. La résolution temporelle varie de minutes à l'heure ou au jour avec des satellites SPOT<sup>11</sup>, MODIS<sup>12</sup>, NOAA-AVHRR<sup>13</sup> et le radiomètre imageur SEVIRI<sup>14</sup> est en suivi continu dans le temps (cf. Tableau 0.2 en annexe). En dépit des performances des satellites, des limites liées aux caractéristiques des capteurs existent toujours en termes de couverture globale de la terre et de l'observation de l'environnement à des échelles locales fines (Chybicki & Łubniewski, 2017).

Le capteur imageur MODIS embarqué sur deux plateformes distinctes (sur le satellite TERRA en 1999 puis sur le satellite AQUA en 2002) est classé dans la catégorie des satellites à moyenne résolution spatiale. Il fournit les informations sur l'environnement (évolution et changement du couvert végétal) pour un suivi régulier, la température de surface des terres et des océans, l'atmosphère (neige, nuage, vapeur d'eau), les sols, les aérosols, les feux de végétation grâce à ses résolutions radiométrique, spectrale, spatiale et temporelle (cf. Tableau 0.3 en annexe A). Les 36 bandes spectrales de MODIS (cf. Tableau 0.4 annexe A) couvrant plusieurs thématiques cadrent parfaitement avec cette étude pour l'estimation des DCMC. Mais la différence d'échelle entre ses résolutions spatiales (RS varie de 250 m à 1000 m selon ses bandes spectrales) constitue une limite dans la discrimination des données à une fine échelle de 30 m. Sa résolution spatiale la plus haute (250 m) vaut plus de 8 fois la résolution de 30 m souhaitée. Or, plusieurs autres capteurs imageurs tels SPOT 5 permettent d'avoir une précision spatiale du couvert végétal entre 2,5 m et 5 m en panchromatique et 10 m en multispectrale. De même, le satellite Landsat 8 OLI TIRS avec une résolution spatiale de dix fois moins que celle de MODIS fournirait plus de détails sur les DCMC, mais a une résolution temporelle de 16 jours. Ce qui ne permet pas le suivi horaire ou quotidien des vecteurs. En conséquence, pour avoir une température de surface à une fine résolution spatiale et temporelle, il faut procéder à un changement d'échelle.

---

<sup>7</sup> <https://www.satimagingcorp.com/satellite-sensors/ikonos/>

<sup>8</sup> <https://www.satimagingcorp.com/satellite-sensors/quickbird/>

<sup>9</sup> <https://directory.eoportal.org/web/eoportal/satellite-missions/o/orbview-3>

<sup>10</sup> <https://sentinel.esa.int/web/sentinel/missions>

<sup>11</sup> <https://e-cours.univ-paris1.fr/modules/uved/envcal/html/vegetation/role-teledection/basse-resolution.html>

<sup>12</sup> <https://modis.gsfc.nasa.gov/about/specifications.php>

<sup>13</sup> <https://earth.esa.int/web/guest/missions/3rd-party-missions/current-missions/noaa-avhrr>

<sup>14</sup> <https://e-cours.univ-paris1.fr/modules/uved/envcal/html/msg/2-performancesmsg/seviri.html>

En effet, il est rare de trouver un même capteur possédant à la fois une bonne résolution spatiale et temporelle. Une difficulté conduisant à associer plusieurs images satellites de caractéristiques différentes et de diverses sources. Ces images doivent être mises à la même échelle par un processus de changement d'échelle avant toute exploitation. Bien que certains capteurs imageurs, comme MODIS, fournissent de plus en plus des données de diverses thématiques pour réduire le recours à de multiples sources, le changement d'échelle s'impose lors des besoins d'informations à des échelles fines. Le changement d'échelle consiste à modifier la résolution spatiale de l'ensemble de données matricielles et à établir des règles d'agrégation ou de désagrégation des valeurs pour la nouvelle taille des pixels. Lorsque le processus de changement fait passer d'une échelle plus petite à une échelle plus grande, le résultat est nommé la réduction d'échelle (« downscaling », en anglais).

### 1.3.2. Évaluation des méthodes de réduction d'échelle de la température de surface

Cette sous-section se base sur la revue de littérature pour analyser les méthodes de réduction d'échelle de LST afin de montrer ses avantages/limites, la relation existante entre la résolution spatiale cible et les différentes images satellitaires utilisées dans la réduction d'échelle. Elle met également l'accent sur les difficultés des méthodes de réduction d'échelle de LST par rapport aux pixels mixtes et enfin sur la qualité des résultats en fonction des états de surface.

#### 1.3.2.1. *Avantages et limites des modèles de réduction d'échelle de LST*

Afin de mieux comprendre les spécificités, les forces et faiblesses des méthodes de réduction d'échelle pour la température de surface, nous avons produit deux tableaux (cf. Tableau 1.1 et

Tableau 1.2) reprenant une partie de la revue de littérature (cf. Tableau 0.5 en annexe A) faite pour décrire les articles scientifiques exploités selon les critères sur la typologie de couverture du sol, la superficie des zones d'étude, les images satellitaires utilisées, les résolutions cibles atteintes lors des prédictions de LST, la typologie du modèle, de l'algorithmique, la saison et la marge d'erreur obtenue. Plus spécifiquement, nous nous sommes intéressés dans le Tableau 1.1 aux images satellitaires en entrée des modèles et la marge d'erreur (RMSE) des résultats obtenus par différentes méthodes de réduction d'échelle. Le Tableau 1.1 permet de faire une lecture globale sur les méthodes recensées. Le

Tableau 1.2 permet une fine analyse sur le fonctionnement de ces méthodes par rapport aux critères : typologie de couverture du sol, typologie du modèle, algorithmique, saison d'étude, seuil de 2°C pour la RMSE. La

comparaison de ces critères a servi à évaluer la performance des méthodes. Les commentaires qui suivent sont basés sur la lecture concomitante des deux tableaux.

L'ensemble des méthodes présentées dans le Tableau 1.1 et le

Tableau 1.2, expérimentées sur des superficies allant de 300 km<sup>2</sup> à 130 000 km<sup>2</sup>, ont montré leur capacité à réduire la température de surface sur différentes catégories d'utilisation et d'occupation des sols. (zone hétérogène urbaine, zone hétérogène boisée, zone hétérogène agricole). Les précisions statistiques varient en fonction des paramètres d'entrée, de la typologie du modèle, de l'algorithmique et la saison.

Tableau 1.1: Portrait des méthodes de réduction d'échelle LST

Méthodes	Typ Couv		Sup Km²	Res Spatiale des images sat utilisées	Résultat	RMSE	Sources
	Het Veg	Het Urb					
<b>Méthodes ayant exploitée une seule Variable auxiliaire</b>							
TsHARP	1	0	NA	Landsat ETM+ 60 m; MODIS11A1	60 m	NA	Agam et al 2007
TsHARP	1	0	5 665	MOD09GA; MOD11A1; Landsat 7 ETM + 60 m	60 m	1,52 K	Bindhu et al ; 2013
NL- DisTrad	1	0	5 665	MOD09GA; MOD11A1; Landsat 7 ETM + 60 m	60m	0,84 K	Bindhu et al ; 2013
TsHARP	0	1	6 587	Landsat 8OLI 30 m; Landsat 8 TIRS 100m;	90 m	1.14°C	Yang et al., 2017)
DisTrad	0	1	6 587	Landsat 8OLI 30 m; Landsat 8 TIRS 100m;	90 m	1.01°C	Yang et al., 2017)
TsHARP		1	3115 ; 1536	Landsat 8 OLI ; Landsat 8 TIRS, MOD11A1	100 m	2,85°C	Peng et al, 2019
Dsopt	0	1	300 ; 14 256	AVHRR 1,1 km; Landsat 8 OLI 100 m	100 m	2,2°C	Chybicki et Lubniewski, 2017
PBIM optimisé	0	1	300 ; 14 256	AVHRR 1,1 km; Landsat 8 OLI 100 m	100 m	2,8°C	Chybicki et Lubniewski, 2017
PBIM modifié en 4 sous modèles	0	1	NA	AVHRR 1,1 Km; Land cover Land use Landsat TM ( 5, 6, 7)	120 m	4,9-5,3°C	Stathopoulou et Cartalis, 2009
TSHARP	1	0	5 170	Landsat 8 (OLI, TRS) 100 m ; MODIS11A1 1 km, MODIS09GA	240 m	3,02°C	Ebrahimi & Azadbackht, 2019
TSHARP + TPS	1	0	NA	MODIS 11 A1 1 km , Landsat 7 ; MODIS09 GQ	250 m	0,307°C	(Singh Rawat et al., 2019)
DisTrad	1	0	34 225	MOD09GA; MOD11A1	500m/250 m	1,52°C/1,65 °C	(Mukherjee et al., 2014)
TsHARP avec variante Locale	1	0	34 225	MOD09GA; MOD11A1	250 m	1,57° C	(Mukherjee et al., 2014)
TsHARP	1	0	34 225	MOD09GA; MOD11A1	250 m	1,59°C	(Mukherjee et al., 2014)
LMS <sub>DS</sub>	1	0	34 225	MOD09GA; MOD11A1	500m/250 m	1,35°C/1.43 C	(Mukherjee et al., 2014)
PR <sub>DS</sub>	1	0	34 225	MOD09GA; MOD11A1	250 m	1.62 °C	(Mukherjee et al., 2014)
DisTrad	1	0	NA	MOD11L2, 1km,MOD09GA	500 m	3,82°C	Yang et al., 2017)
<b>Méthodes ayant exploitées plusieurs Variables auxiliaires</b>							
TsHARP multivariable		1		Landsat OLI 30 m et Landsat TIRS 100 m	30m	1,5 °C	(Govila et al., 2019)
R D A P R	0	1	7 400	Landsat 5 TM (1-5 et 7) 30 m et Landsat 5TM (6) 120 m; levés avec drone 1 m	30 m	3°C	Bonafoni et al, 2017
MSFAT	0	1	6 587	Landsat 8OLI 30 m; Landsat 8 TIRS 100m;	90 m	1.13°C	Yang et al., 2017)
GTWR	0	1	3115 ; 1536	Landsat 8 OLI ; Landsat 8 TIRS, MOD11A1	100m	1.62 °C	Peng et al, 2019
SADFAT	0	1	NA	MOD11A1; MOD09GA; Landsat TM	120m	Entre 1.3k et 2K	Weng et al., 2014)
RFR		1	32 000	MOD11A1; MOD11B1, MOD09GA; MCD12Q1, Landsat 7 ETM, STRM	240 m	2,24K	Hutengs & Vohland, 2014
Random Forest regression	1	0	5170	Landsat 8 (OLI, TRS) 100 m ; MODIS 11 A1 1 km, MODIS 09	240 m	2,25°C	Ebrahimi & Azadbackht, 2019
ELM	1	0	5170	Landsat 8 (OLI, TRS) 100 m ; MODIS 11 A1 1 km, MODIS 09	240 m	2,53°C	Ebrahimi & Azadbackht, 2019
SVR	1	0	5170	Landsat 8 (OLI, TRS) 100 m ; MODIS 11 A1 1 km, MODIS 09	240 m	2,56°C	Ebrahimi & Azadbackht, 2019
RFR (BM; EM 1 et EM2)	1	0	7400	MODIS11A1; 1 km; MODIS NDVI; Landast 5 et 7 ASTER GDEM	250 m	1,81-2,51 K	Barthowiak et al, 2019
GPS Downscaling	1	0	4539	Météosat (SEVERI) 1 km , Landsat 7, 60 m	1 km, 60 m	2,7 K	Mechri et al, 2016
RFR (MIRF)		1	NA	MOD11L2, 1km,MOD09GA	500m	2,42°C	Yang et al., 2017)
RFR (base)	1		NA	MOD11L2, 1km,MOD09GA, Landcover	500 m	2,60°C	Yang et al., 2017)

Tableau 1.2 : Avantages et limites des méthodes de réduction d'échelle de LST

Méthodes	Caractéristiques environnementales			Typologie modèle		Algorithm		Saison				Performance		
	Zone hétér. Agric	Zone hétér urb	Zone hétér bois	Lin	Non Lin	Stat	Phy	Prin	Eté	Aut	Hiv	Cons. Caract Therm	RMSE <=2°C/K	Sur/sous/estimation
Thermal sharpening (TsHARP)	Oui	Limitée	Oui	Oui		Oui		Oui	Oui	Oui	Oui	Non	Non	Oui
Random Forest Regression (RFR)		Oui			Oui	Oui		Oui	Oui	Oui	Non	Non	Non	Oui
Extreme Learning Machine (ELM)		Oui			Oui	Oui		Oui	Oui	Oui		Non	Non	Oui
Support Vector Regression (SVR)		Oui			Oui	Oui		Oui		Oui		Non	Non	Oui
Optimized downscaling process Dsopt		Oui		Oui			Oui		Oui			Oui	Non	
Pixel Block Intensity Modulation (PBIM)	Oui	Limitée	Oui	Oui			Oui		Oui			Non	Non	
Data assimilation: Genetic particle smoother (GPS) Downscaling				Oui		Oui			Oui			Non	Non	Oui
PBIM: Pixel Block Intensity Modulation modifié en 4 sous modèles	Oui	Limitée	Oui	Oui			Oui		Oui			Oui	Non	
Random Forest Algorithm décomposé en 3 sous modèle (BM, EM 1 et EM2)		Oui			Oui	Oui		Oui	Oui			Non	Non	
Combinaison de Thin plateSpline TPS+ TsHARP	Oui			Oui		Oui				Oui		Non	Oui	
Regressive disaggregation approach: polyomial regression: R D A P R	Oui		Oui		Oui	Oui			Oui			Non	Non	
Non Linear DisTrad			Oui		Oui	Oui			Oui			Non	Oui	Oui
DisTrad: Disaggregation procedure for radiometric surface temperature	Oui	Non	Non	Oui		Oui				Oui		Non	Oui	Oui
TsHARP avec variante locale	Oui			Oui		Oui			Oui			Non	Oui	
LMS <sub>DS</sub>	Oui		Oui	Oui		Oui			Oui			Non	Oui	
PR <sub>DS</sub> <i>Pace regression downscaling</i>	Oui	Non	Non	Oui		Oui			Oui			Non	Oui	Oui
MSFAT		Oui			Oui	Oui			Oui			Non	Oui	
SADFAT		Oui			Oui	Oui			Oui			Oui	Oui	
GTWR Geographically and Temporally Weighted Regression		Oui			Oui	Oui			Oui			Non	Oui	

La lecture du Tableau 1.1 montre que la plus faible RMSE est de 0,30°C (Rawat et al., 2019) et la plus élevée est de 5,3°C (Stathopoulou & Cartalis, 2009) en zone hétérogène urbaine. La moyenne générale des marges d'erreur (2,42°C en considérant uniquement celles en °C) oscille autour de 2°C. Les conclusions de ces articles scientifiques révèlent que l'analyse qualitative comparative des résultats de réduction d'échelle avec les images de référence des études citées dans le Tableau 1.1 montre une grande justesse dans la répartition des valeurs de LST et plus fondamentalement dans les sous-ensembles uniformes exempts d'humidité de sol. De plus, dans les images réduites (à haute résolution spatio-temporelle) apparaissent les structures et formes des différentes unités d'occupation du sol caractérisant chaque milieu d'étude. L'avantage fondamental de ces méthodes, pour la plupart statistique (cf.

Tableau 1.2), est de fournir une image à haute résolution spatiale de LST pour des applications environnementales spécifiques pour lesquelles la télédétection reste limitée à ce jour.

Il ressort de manière générale du

Tableau 1.2 que les méthodes linéaires ou non linéaires appliquées majoritairement en été (soit 84% des cas) ne conservent pas les caractéristiques thermiques des images d'origine. Seules les méthodes Optimized downscaling process (Dsopt) et Pixel Block Intensity Modulation (PBIM) modifié en quatre sous-modèles y font exception ; et ces deux méthodes au plan algorithmique sont des modèles physiques contrairement aux autres basées sur des modèles statistiques. Sur le critère seuil de marge d'erreur de 2°C, 53% des résultats de LST prédites sont au-dessus de cette marge contre 47% en deçà ou égale de ce seuil pour toutes les méthodes. On remarque pour les marges d'erreurs en dessous ou égale à 2°C que d'une part, ces résultats ont été obtenus avec les modèles linéaires dans des zones hétérogènes agricoles, et d'autre part, avec les modèles non linéaires en zones urbaines hétérogènes. En revanche, avec les méthodes d'apprentissage automatique (RFR, SVR, ELM ou la variante de RFR), expérimentées en toute saison dans les zones urbaines hétérogènes, la marge d'erreur est au-dessus du seuil de 2°C. Sur le critère de sensibilité aux valeurs extrêmes, la quasi-totalité des méthodes sous-estime ou surestime ces valeurs.

Nous retenons de cette sous-section que la saison d'été reste celle propice pour les études de réduction d'échelle de LST. Au plan algorithmique, les modèles statistiques sont largement utilisés bien qu'ils ne permettent pas la conservation des caractéristiques thermiques et ce peu importe les caractéristiques environnementales de la zone d'étude. Malgré les avantages des différentes méthodes, elles montrent des limites et incertitudes qui affectent la qualité des résultats. C'est ce qui est mis en lumière dans les sections suivantes à travers des analyses détaillées.

### 1.3.2.2. *Résolution spatiale cible et typologie des images satellitaires*

Cette analyse se base sur le Tableau 1.1. La valeur de résolution spatiale obtenue par réduction d'échelle varie de 30 m à 500 m selon les cas et dépendrait des résolutions des images satellitaires en entrée (cf. Tableau 1.1). Les résolutions de 30/90 m par exemple ont été obtenues à partir d'une image Landsat LST de 100 ou 120 m. Par contre, les résolutions de 60 à 500 m ont été acquises à partir de l'image MODIS LST de 1000 m ou AVHRR de 1100 m. Ces deux constats supposent que pour atteindre 30 m de résolution spatiale, il faudrait une image LST de résolution spatiale moyenne de 120 m au plus. Parmi les méthodes ayant exploité une seule variable auxiliaire, notamment un indice de végétation ou l'émissivité effective, la plus fine résolution cible atteinte est de 60 m. En revanche, deux méthodes (TsHARP multivariable et RDAPR) ont été appliquées pour une réduction d'échelle à 30 m parmi les méthodes qui ont exploité plusieurs variables auxiliaires. Nous remarquons pour cette cible de 30 m, les images satellitaires exploitées (thermique et optique) proviennent du même capteur Landsat, contrairement aux autres cas où il y a au moins deux sources différentes. Or, selon Rawat et al. (2019), il est préférable d'avoir une seule source pour l'ensemble des données pour minimiser ou éviter les erreurs.

Les différences entre les capteurs (nombre de bandes spectrales, temps d'acquisition, date/heure de passage ; angle de vue, couverture effective des pixels, largeur de bande, paramètre d'orbite, erreurs de géolocalisation, fonction de réponse spectrale...) affectent la qualité des données et par conséquent, le résultat qui en découle (Hutengs & Vohland, 2014 ; Mukherjee et al., 2014 ; Weng et al., 2014). Pour mieux illustrer la différence entre la qualité des images de Landsat et MODIS, des méthodes de réduction d'échelle (TsHARP, LM<sub>DS</sub>) ont été appliquées séparément sur ces deux types d'images pour obtenir des résolutions de 500/250 m (Mukherjee et al., 2014) ; 240 m, 120 m et 60 m (Agam et al., 2007). Les résultats révèlent moins d'erreurs pour Landsat LST que MODIS LST. Mais les erreurs étaient encore plus élevées avec l'application des méthodes sur les deux capteurs après la normalisation et la coregistration entre les bandes spectrales (Bisquert et al., 2016).

Par ailleurs, une analyse quantitative basée sur plusieurs indicateurs d'erreur dont l'erreur quadratique moyenne montre de manière globale que plus l'échelle spatiale de réduction diminue, plus l'imprécision augmente, quel que soit le modèle (Agam et al., 2007 ; Govil et al., 2019 ; Mukherjee et al., 2014) à cause de la variabilité élevée des sous-pixels et du facteur d'échelle. Par exemple, Agam et al. (2007) ont utilisé TsHARP pour réduire la MODIS LST à 240 m, 120 m et 60 m. L'erreur moyenne absolue associée respectivement à ses résolutions est 0,68°C, 0,82°C et 0,91°C. La même expérience a été reprise uniquement sur les surfaces agricoles avec les mêmes résolutions, l'erreur moyenne absolue donne dans le même ordre 0,80°C ; 0,94°C et 1,06°C. Ce même



constat a été fait dans l'étude de Mukherjee et al. (2014) sur une zone hétérogène à dominance agricole avec cinq méthodes linéaires (TsHARP, TsHARP avec variante locale, LM<sub>DS</sub>, DisTrad et PR<sub>DS</sub>) pour réduire la MODIS LST de 1000 m à 500 m et 250 m. Avec la méthode LM<sub>DS</sub>, la RMSE obtenue est de 1,35°C et 1,43°C (cf. Tableau 1.1). Avec DisTrad, la RMSE est de 1,52°C et 1,65°C (cf. Tableau 1.1). Ces conclusions sont en concordance avec les résultats de l'étude de Govil et al. (2019) réalisée dans un paysage urbain hétérogène (la ville de Raipur en Inde) de la région tropicale humide en utilisant la technique TsHARP pour réduire la Landsat LST de 960 m à 30, 120, 240 et 480 mètres. Partant de ces observations, il se pose la question de la résolution spatiale optimale qui engendrerait le moins d'incertitude dans cette étude. En d'autres termes, réduire l'image MODIS LST de 1000 m à 100 m et à 30 m engendrerait quel degré de justesse ou d'exactitude ? Les marges d'erreur seront-elles en dessous ou égale à 2°C ?

La comparaison qualitative et quantitative des résultats révèle des différences de l'exactitude, la justesse et la sensibilité des méthodes de réduction d'échelle selon les états de surface. Les auteurs Ebrahimi & Azadbakht et al., (2019) soulignent l'incapacité de tous les modèles à reproduire intégralement les variations spatiales et la distribution complète des LST de référence. Ceci se traduit par la présence de biais. Ces biais, mis en évidence par l'apparition des effets de flou dans les images réduites, sont associés en partie à la procédure de correction résiduelle intervenant dans le processus de réduction d'échelle (Ebrahimi & Azadbakht, 2019 ; Hutengs & Vohland, 2014). Ainsi, on observe des pixels mixtes non ou peu simulés, une faible conservation des caractéristiques thermiques de LST. Ceci impacte donc la précision spatiale des résultats obtenus avec ces méthodes.

### 1.3.2.3. *Sensibilité des modèles aux pixels hétérogènes et aux valeurs extrêmes de LST*

L'analyse sur la capacité des modèles à simuler les pixels hétérogènes et sur/sous-estimer les valeurs extrêmes de LST est basée sur l'indicateur de performance dans le

Tableau 1.2. L'indicateur de performance est composé de trois critères qui sont : la conservation des caractéristiques thermiques, la RMSE et la sensibilité aux valeurs extrêmes. Pour une meilleure compréhension de cette sous-section, nous citons des valeurs de RMSE associées aux résolutions cibles que l'on ne retrouve pas forcément autant dans le Tableau 1.1 que le

Tableau 1.2.

Les valeurs élevées de RMSE dépassant le seuil de 2°C semblent être attribuées à la présence des pixels mixtes non décomposés qui contribuent à l'augmentation des marges d'erreur. Le plus souvent, ces

imprécisions dans les résultats de LST (°C) sont associées à la faiblesse des modèles linéaires appliqués en zone hétérogène urbaine. Ainsi, selon Bindhu et al., (2013), dans ce contexte d'hétérogénéité urbaine, la principale raison de l'imprécision spatiale est la difficulté des modèles utilisant un indice de végétation à saisir la variation de LST dans les pixels mixtes. Il s'ensuit une RMSE élevée, attribuable à une sensibilité du modèle aux valeurs aberrantes ou extrêmes (Agam et al., 2007). Or, il se remarque dans le Tableau 1.1 des valeurs élevées de RMSE même avec les modèles non linéaires pour ces mêmes zones hétérogènes urbaines ou hétérogènes agricoles/boisées. C'est le constat fait par Bartkowiak et al., (2019), dans sa zone d'étude hétérogène et boisée, dominée par des montagnes avec la présence de la vallée de l'Adige/Etsch. Cette zone est couverte de forêts, de prairies, de pâturages et de zones agricoles. Les variables auxiliaires utilisées sont les seuils de végétation et la topographie. Les images de MODIS LST 1000 m réduites à 250 m à partir d'une méthode de RFR ont révélé après comparaison avec une image Landsat LST de référence des écarts allant de -4,1 à 19 Kelvins. Ainsi, les auteurs en concluent que la méthode RFR n'a pu simuler les variations thermiques au niveau des sous-pixels dans la zone d'étude. Les recherches de Bindhu et al. (2013) menées sur une zone hétérogène à dominance boisée quasi identique que la précédente ont abouti à cette même conclusion. Ces auteurs ont appliqué TsHARP et NL Distrad (exploitant un indice de végétation comme variable auxiliaire) à une zone hétérogène dans la partie sud du Tamil Nadu en Inde, pour réduire la MODIS LST 960 m à 60 m. Les unités d'occupation dans cette zone sont des terres irriguées, des zones agricoles, des forêts, des eaux de surface, des prairies et des sols nus. Bien que l'évaluation des résultats basée sur l'indice de Nash Sutcliffe Efficiency -NSE-indique une nette précision pour les deux modèles (NSE inférieur à 1 ; cf. Tableau 1.1), le problème des pixels mixtes subsiste. La méthode NL-DisTrad n'a pu modéliser que partiellement les valeurs de LST les plus élevées. Ces valeurs élevées de LST correspondaient aux pixels mixtes dans MODIS LST à 960 m. Les mêmes ou d'autres méthodes appliquées sur des zones hétérogènes urbaines, boisées ou agricoles, ont révélé la tendance à surestimer ou sous-estimer les valeurs élevées ou faibles de LST. L'étude de Mukherjee et al., (2014) a expérimenté d'autres modèles en zone hétérogène à dominance agricole dans l'Uttar Pradesh, en Inde. Les modèles utilisés sont TsHARP, DisTrad, TsHARP avec variante locale, LMS<sub>DS</sub> et PR<sub>DS</sub> pour prédire la MODIS LST 960 m à des résolutions de 500 m et 250 m. La variable auxiliaire utilisée est un indice de végétation. Les modèles LMS<sub>DS</sub>, TsHARP avec variante locale et TsHARP ont donné des meilleurs résultats avec des RMSE respectives de 1,43 °C, 1,57 °C et 1,59 °C (cf. Tableau 1.1). Malgré la précision des résultats, le modèle LM<sub>DS</sub> a montré sa sensibilité aux valeurs extrêmes.

D'autres recherches (Yang, Cao, et al., 2017 ; Yang, Li, et al., 2017) ont supposé que l'introduction de plusieurs variables auxiliaires dans les modèles pour réduire la LST en zone hétérogène pouvait améliorer la sensibilité aux valeurs extrêmes et prendre en compte les pixels mixtes. Mais ces insuffisances demeurent toujours. Par exemple le modèle MFSAT a intégré les seuils optimaux de SAVI, NMDI, MNDWI et NDBI pour réduire la

Landsat 08 LST 100 m à 90 m, mais les résultats ont révélé la sur/sous-estimation des valeurs de LST dans les zones bâties hautes/denses et des rives des rivières. Il en est de même pour les algorithmes de RFR qui ont intégré jusqu'à 18 variables auxiliaires (Ebrahimi & Azadbakht, 2019 ; Hutengs & Vohland, 2014, 2016).

De ce qui précède, nous retenons que la faible capacité des méthodes de réduction d'échelle à simuler les valeurs extrêmes de LST est indépendante des zones d'étude, des modèles et du nombre de variables auxiliaires. Toutefois, le recours à plusieurs variables auxiliaires permet de réduire les incertitudes bien que le problème reste non résolu. La problématique des valeurs extrêmes varie selon les unités d'occupation et d'utilisation des sols. C'est ce qui est mis en lumière dans cette dernière sous-section.

#### *1.3.2.4. Variabilité spatio-temporelle des méthodes de réduction de LST selon les états de surface*

Dans le but d'identifier la variabilité des erreurs, des analyses approfondies de réduction d'échelle ont été faites sur différentes unités de surface de la zone d'étude ou parfois en différentes saisons. Les conclusions de ces travaux montrent une variabilité saisonnière dans les résultats de réduction d'échelle (cf.

Tableau 1.2). L'étude de Ebrahimi & Azadbakht, (2019), menée dans la province de Téhéran en Iran couvrait une zone hétérogène variée : 50 % de pâturages, 25,5 % de zones agricoles, 15,5 % de zones résidentielles, 8,5 % de terres nues et 0,5 % de plans d'eau. Les méthodes de TsHARP et les modèles de Machine Learning (RFR, ELM, SVR) ont été appliqués pour réduire MODIS LST 1000 m à 240 m. Une douzaine de variables auxiliaires a été exploitée pour la RFR, ELM et SVR contre une (la CVF) pour le TsHARP. Les images MODIS LST ont été prises au printemps (20 juin), en été (7 août) et à l'automne (24 novembre) 2016. Les marges d'erreur sont quasi identiques pour toutes les saisons et pour chacun des modèles. La marge d'erreur donne respectivement, 3,02°C, 2,25°C, 2,53°C et 2,56°C pour le TsHARP, le RFR, ELM et SVR. Une analyse comparative selon les unités d'occupation en fonction de la saison montre sur les images réduites qu'en automne, printemps et été, on obtient une meilleure précision dans les zones agricoles et de pâturages. Tandis que cette précision est observée en zones résidentielles uniquement qu'en printemps et automne et enfin, dans les zones nues en été. Les tests de Yang et al., (2017) sur six sites (zone humide, maïs, verger, Gobi, zone désertique et zone sauvage) avec les méthodes de RFR (multivariable MIRF), RFR de base, DisTrad, ont donné de meilleurs résultats en automne plus qu'en été, printemps et hiver. Pour la saison d'automne, la plus grande précision a été obtenue sur les sites de vergers et de maïs, avec respectivement des valeurs de biais de 0,06 °C et -0,11 °C. La réduction d'échelle des LST dans les zones humides et les sites désertiques était moins satisfaisante, avec des biais respectifs de 2,45 °C et -2,64 °C. Dans une autre étude, l'été par contre s'est révélé être la saison où l'algorithme MSFAT (Multi-scale-factor downscaling approach based on adaptive threshold MSFAT) a donné de bons résultats contrairement aux trois autres saisons dans la zone hétérogène urbaine de

Nanjing, capitale de la province de Jiangsu (Chine). En ce qui concerne les états de surface, les surfaces de végétation ont une RMSE de 0,40°C contre 2,79°C, 2,50°C et 3,34°C respectivement pour les surfaces en eau, imperméables et nues (Yang et al, 2017). Les travaux de Govil et al (2019) avec TsHARP dans un paysage urbain hétérogène (la ville de Raipur en Inde) de la région tropicale humide désignent les saisons post-mousson et hivernale comme celles présentant les moins d'erreurs pour toutes les quatre unités d'occupation du sol (eau, végétation, surfaces nues, surfaces imperméables). Les travaux de Agam et al. (2007) sur les hautes plaines de Texas où domine l'agriculture et de Mukerdje (2014) sur une zone hétérogène à dominante agricole en Inde, basés sur les indices de végétation ont révélé une bonne précision dans les zones agricoles plus que dans les zones humides, de végétation, nues et urbaines. Toutefois, les LST des champs irrigués dans les zones agricoles n'ont pu être bien réduites (Agam et al., 2007).

Il ressort des études mentionnées que les zones humides ou les zones présentant une variabilité d'humidité du sol dans leur unité d'occupation ont des précisions plus faibles pour tous les modèles. L'imprécision de LST réduite en zones humides serait liée aux faibles valeurs dans ces zones. Selon Yang et al., (2019), plus la valeur de LST observée est élevée, meilleure est la qualité de LST réduite. Inversement, une trop faible valeur de LST observée est sous-estimée dans la LST réduite. Pour Mukerdjee (2014), les performances des modèles de réduction d'échelle sont sensibles à la variation du niveau d'humidité du sol dans les unités d'occupation ou d'utilisation du sol. De fait, en hiver, la glace et la neige à la surface du sol influencent la justesse des LST réduites. Il faut souligner que ces imprécisions varient selon les résolutions cibles à atteindre, des types d'images satellitaires utilisées et leur résolution initiale, du nombre de variables auxiliaires utilisées dans les modèles.

Par ailleurs, dans les articles scientifiques exploités, la plupart des méthodes ont été appliquées sur des sites dont la superficie la plus grande est 34 225 km<sup>2</sup>. Aucun des articles n'a rapporté la performance et les précisions de leur application sur de très grandes superficies telle celle de Québec (1,668 000 000 km<sup>2</sup>) par exemple. À travers les informations sur les points forts et faibles de chaque modèle, un cadre de référence a été établi dans la section ci-dessous (cf. section 1.3.3) pour identifier les méthodes prometteuses et cadrant avec le contexte particulier du Canada et selon les besoins du projet.

### 1.3.3. Application des trois méthodes de réduction d'échelle identifiées

La sélection d'une méthode à appliquer dans cette étude nécessite une analyse comparative entre elles en tenant compte de leurs spécificités dans le paramétrage, au plan algorithmique et les incertitudes qui leur sont associées. De même l'exactitude des résultats liée à chaque méthode est singulière et dépend fortement des caractéristiques environnementales des zones d'étude et de la qualité des données utilisées. Par conséquent,

leur évaluation doit se baser sur les mêmes normes. À travers les informations sur les points forts et faibles de chaque modèle (cf. Tableau 0.5 en annexe), un cadre de référence a été établi pour identifier trois méthodes prometteuses et cadrant avec le contexte particulier du Canada et selon les besoins du projet.

#### 1.3.2.1. *Cadre théorique sur l'évaluation des performances des méthodes de réduction d'échelle de LST*

L'évaluation de la performance de différents modèles pour aboutir à une classification et en choisir le meilleur se base sur des critères précis. Massalabi (1996) dans son étude comparative des algorithmes de classification d'image satellitaire a identifié trois grandes catégories de critère : ceux se rapportant au modèle (*mesure d'homogénéité, fonctions discriminantes, risques d'erreur*) puis ceux axés sur les résultats (*précision, robustesse du modèle, comportement par rapport à la géométrie, des objets, comportement par rapport aux paramètres statistiques de l'objet, le comportement aux frontières des objets*) et enfin ceux liés au fonctionnement et l'utilisation du modèle (*sensibilité, stabilité des résultats, rapidité d'exécution des calculs, convivialité de l'algorithme, rentabilité de la méthode et la capacité de traitement selon le type, la qualité et les volumes des données*). Dans d'autres études, l'analyse comparative se base sur une évaluation quantitative ou qualitative en recourant à certains des critères sus-cités : la précision, l'ergonomie, l'homogénéité, l'exactitude, l'incertitude, la robustesse auxquels s'ajoutent la reproductibilité, la sensibilité. En télédétection, pour la fusion d'images, l'évaluation qualitative est une analyse visuelle entre le résultat et l'image native sur la restitution fidèle des couleurs de chaque objet contenu dans l'image ou le rendu visuel des hautes fréquences spatiales. Cette analyse visuelle est globale (sur l'ensemble de l'image) ou locale (sur chaque classe d'objets identifiée dans l'image). En revanche, l'évaluation quantitative porte sur la distorsion radiométrique ou géométrique des objets de l'image et fait appel aux paramètres mathématiques (Thomas, 2006). Selon ce même auteur, les critères se rapportant aux résultats ne peuvent être des comparateurs fiables (Thomas, 2006) à cause des paramètres (*sensibilité à la variabilité des thèmes présents dans la scène acquise par le capteur satellite, à la taille des objets de la scène, dont la variabilité est liée à une bande de fréquence spatiale, la différence de résolution spatiale des images multispectrales à fusionner par exemple*) impactant la qualité des résultats issus d'une fusion d'images satellitaires. Partant des travaux de Kristjánsson, (1991), Raffy, (1993), Wald et al., (2009) et Ning Kong & Vidal-Madjar, (1988) qui ont démontré que la qualité est une fonction complexe et aléatoire de la résolution, les auteurs ont soulevé des réserves sur l'utilisation des propriétés statistiques de l'image fusionnée et les images de référence comme critères de comparaison.

Dans cette étude, tous ces critères précédemment cités ne sont pas pris en compte, car toutes les informations en lien avec ces critères ne sont pas toujours précisées dans les articles recensés. De plus, tous les modèles

n'ont pas été appliqués sur des données d'un même site d'étude. Les résultats n'ont pas été non plus validés avec les mêmes données de référence. Bien que les critères statistiques soient déconseillés par plusieurs auteurs (cf. paragraphe précédent), ils ont été pris en compte, puisque dans la présente recherche, ils permettent la prise de décision sur la mise en pratique des méthodes à notre contexte d'étude. Ainsi, les critères de précision, notamment l'erreur quadratique moyenne RMSE ont été retenus dans l'évaluation comparative. De plus, la discrimination des modèles a également été faite à partir d'une analyse qualitative et selon les objectifs du projet de recherche.

#### 1.3.2.2. *Analyse comparative des modèles selon les critères généraux et objectifs du projet*

Les critères généraux ont permis de discriminer pour présélectionner graduellement un lot de méthodes parmi les 22 recensées, dont certaines étaient citées plusieurs fois dans différents articles. Les critères spécifiques au projet ont été appliqués sur cette présélection. Les critères généraux retenus se rapportent à : (1) la catégorie d'appartenance des méthodes (Fusion ou Noyaux), (2) l'algorithmique (statistique ou physique) ; (3) l'unité des RMSE (°Celsius ou °Kelvin) et (4) la typologie des données à l'entrée des méthodes. Nous basant sur la littérature, nous avons éliminé les méthodes de fusion, puis de type physique, les méthodes ou enregistrements dont les résultats sont exprimés en degré Kelvin et enfin, les enregistrements mentionnant que les images satellitaires utilisées sont uniquement du capteur Landsat. En effet, il ressort des études de l'efficacité avérée des méthodes de type statistique relevant de la catégorie des méthodes noyaux (Yang, Cao, et al., 2017) par rapport à celles de fusion. L'unité des RMSE en degré Kelvin n'a pas été retenue afin de faciliter la comparaison des modèles dont les RMSE sont majoritairement exprimées en degré Celsius.

Le résultat de ce premier tri a permis de retenir 10 méthodes classées selon huit critères. Le projet vise l'estimation des biotopes des vecteurs responsables de propagation des maladies du Virus du Nil Occidental (VNO) et Lyme à l'échelle du Canada à une échelle spatiale de 30 mètres avec les données MODIS. Partant de cette information, les critères définis sont : la résolution spatiale de 30 m, l'utilisation des données MODIS, la superficie des zones d'étude, la couverture terrestre, la saison d'étude, le temps de calcul, la marge d'erreur souhaitée, la reproductibilité et la typologie des modèles.

- **Résolution spatiale de 30 m** : le modèle qui a expérimenté la désagrégation de LST à 30 m reçoit le chiffre 1 et en cas contraire le chiffre 0. Les résolutions spatiales recensées dans la littérature varient de 30 à 500 m et sont classées en 3 : celle inférieure ou égale à 30 m, entre 30 m et 100 m et au-delà de 100 m.
- **Utilisation des données MODIS** : on a affecté le chiffre 1 au modèle ayant utilisé les données MODIS.
- **Superficie de la zone d'étude** : l'échelle spatiale du projet est celle du Canada, soit neuf millions neuf cent quatre-vingt-cinq mille kilomètres carrés (9 985 000 km<sup>2</sup>). Dans les articles parcourus, aucune

zone d'étude n'a cette superficie. Néanmoins, pour toute zone d'étude de superficie supérieure ou égale à 10 000 km<sup>2</sup>, le chiffre 1 a été attribué.

- **Couverture terrestre** : elle est définie en fonction des conditions environnementales de développement des vecteurs responsables de la propagation des maladies du Virus du Nil Occidental (VNO) et de Lyme. Les habitats spécifiques aux moustiques, responsables du VNO sont les milieux urbains, milieux humides naturels en zones rurales tandis que les tiques transmetteuses de la maladie de Lyme se reproduisent dans les biomes forestiers à couvert denses ou fragmentés (Kotchi et al., 2019 ; Ruiz et al., 2010). Ainsi, nous avons défini trois caractéristiques d'occupation : zone hétérogène à dominante agricole (Het. Agr), zone hétérogène à dominante forestière (Het. For) et zone hétérogène à dominante urbaine avec plus de zones résidentielles ou construites (Het. Urb). Une zone d'étude classée dans l'une des 3 zones est codée 1. Donc si les zones sont cochées alors la somme donne 3.
- **Saison d'étude** : Selon différentes études (Gasmi et al., 2019 ; Ogden et al., 2015 ; Ogden et al., 2004 ; Ruiz et al., 2010), la saison influence l'augmentation et la prolifération des moustiques et tiques. De fait, la prise en compte de la saison est déterminante dans les prédictions. Deux modalités ont été définies : modalité PEA pour le printemps, l'été et l'automne ; la modalité Hiv pour l'hiver. Le modèle de réduction d'échelle LST ayant expérimenté le processus pour chacune des modalités est coté 1. Une expérience faite dans les deux cas donne donc 2.
- **Marge d'erreur** : elle dépend des données en entrée, du nombre de variables prédictives utilisées, du type de variables utilisées et de la résolution cible à atteindre. Elle relève de l'exactitude.
- **Exactitude** (assimilée à la précision ou justesse) : elle se rapporte aux données et mesure l'écart par exemple entre les LST réduites et les LST de référence. L'exactitude définie dans le vocabulaire international de métrologie -VIM<sup>15</sup>- comme l'« *étroitesse de l'accord entre une valeur mesurée et une valeur vraie d'un mesurande*, renvoie à la conjonction de la justesse et la fidélité des mesures selon la norme ISO 1994. Son appréciation nécessite la racine carrée de l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R<sup>2</sup>), le biais relatif et l'efficacité relative de Nash-Sutcliffe efficiency (NSE). La marge d'erreur cible à atteindre dans cette étude est de 2°C. De fait, nous avons défini trois modalités : les valeurs de RMSE inférieures ou égales à 2°C, celles supérieures à 2°C et inférieures ou égales à 3°C et enfin les valeurs de RMSE supérieures à 3°C. Ces modalités sont codées respectivement 1, 0.5 et 0.
- **Reproductibilité** est la capacité du modèle à donner des résultats avec la même justesse en des conditions différentes (zone d'étude différente, saison différente, données différentes en entrée...). Étant donné la diversité morphologique du territoire canadien, les modèles doivent être adaptables aux spécificités de chaque territoire. Ainsi, les modèles qui ont été testés dans différents contextes sont codés 1.
- **Typologie des modèles** : les modèles se basent sur des hypothèses d'existence d'une relation linéaire ou non linéaire entre LST et ses variables explicatives. Partants de la synthèse de littérature qui révèlent l'efficacité des modèles non linéaires, nous leur avons attribué la note 1 et 0.5 aux modèles linéaires.

La somme des valeurs attribuées à chaque critère a permis de classer les 10 méthodes. Le résultat de cette classification est présenté dans le Tableau 1.3 ci-dessous.

---

<sup>15</sup> [JCGM 200:2012 International vocabulary of metrology - Basic and general concepts and associated terms \(VIM\) \(bipm.org\)](https://www.bipm.org/en/activities/vim2012/)

L'évaluation donne le modèle TsHARP comme le plus performant, suivi du modèle RFR, puis des modèles DisTrad et GTWR. Nous retenons alors dans cette étude les méthodes **TsHARP et la RFR** expérimentées sur différents types d'occupation et d'utilisation du sol. Ce sont des modèles flexibles répliquables. Le modèle TsHARP est un modèle de base à partir duquel des variantes ont été mises au point ou d'autres modèles ont été conçus. Il a servi de référence de comparaison à toute nouvelle méthode. La troisième méthode qui devrait être DisTrad a été remplacée par la méthode PBIM, une recommandation de l'ASPC.

#### 1.3.4. Description des trois méthodes retenues dans cette étude

De l'évaluation des 10 méthodes de réduction d'échelle présélectionnées, les méthodes Thermal Sharpering (TsHARP) et Random Forest Regression (RFR) ayant obtenu les meilleurs scores ont été retenues. Pour chacune des méthodes, une description théorique a été faite. Une troisième méthode, Pixel Block Intensity Modulation (PBIM), recommandée par l'ASPC a été aussi décrite, car elle a également fait l'objet d'expérimentation à titre comparatif avec les 2 autres méthodes.

##### 1.3.4.1. Méthode Thermal Sharpering (TsHARP)

Le modèle TsHARP est la version améliorée du modèle DisTrad (disaggregation temperature radiometric) élaboré par Kustas et al. (2003). DisTrad utilise la relation inverse entre la température de surface et le NDVI. Afin de corriger les biais de ce modèle et augmenter la précision des données réduites, le NDVI a été substitué par la couverture végétale fractionnée -CVF- (Agam et al., 2007) avec l'hypothèse d'une relation linéaire entre la température radiative et la CVF. En différenciant l'impact de la photosynthèse active ou non dans la variabilité spatiale de la LST, Merlin et al. (2010) ont modifié les paramètres du modèle pour réduire les biais et erreurs. Cette version de TsHARP a apporté beaucoup d'amélioration dans la réduction d'échelle de LST, mais moins satisfaisants entre les LST d'origine en zone humide, zone de couverture hétérogène ou de zones nues et les LST désagrégées. L'approfondissement des recherches a alors montré une variation spatiale de LST en fonction de l'humidité du sol de surface, la sénescence de la végétation, de la topographie et des paramètres météorologiques (Agam et al., 2007 ; Merlin et al., 2010), facteurs non intégrés dans l'algorithme de TsHARP. Par la suite, le perfectionnement du modèle par Merlin et al., (2012) a consisté à différencier et intégrer dans le modèle les indices des plans d'eau libres et l'état hydrique des sols nus. Avec ces variables, en milieu hétérogène, il y a moins de lacunes dans les résultats de LST fournis par TsHARP.

Par ailleurs, il a été démontré que les faibles variations (amplitude) de LST d'une zone demeurent non significatives pour être simulées par TsHARP (Agam et al., 2007). De fait, il est recommandé d'appliquer le modèle sur des scènes prises en milieu de matinée ou d'après-midi. Cette méthode empirique demeure une



des plus utilisées pour désagréger la LST à des échelles locales (parcelle agricole), régionales (ville, zone forestière) malgré ses limites.

Tableau 1.3 : Évaluation comparative des modèles de réduction d'échelle de LST selon les critères du projet

Modèles	Résolution spatiale (m)			Données MODIS		Car Zone couverte			Saison		RMSE (°C)			Typologie Modèle		Reproductibilité	Superficie >=10 000	Somme
	<=30	31-100	>100	Oui	Non	Het Agr	Het For	Het Urb	P E A	H	<= 2	>2 et <=3	> 3	Lin	Non Lin			
TsHARP		1	0	1	0	1	1	1	1	1			0	0,5		1	1	9,5
TsHARP avec variante Locale			0	1	0	1	0	0	1	0	1			0,5		1	1	6,5
RFR			0	1	0	1	1	1	1	0		0,5		1		1	1	8,5
ELM			0	1	0	0	1	0	1	0		0,5		1		0	0	4,5
SVR			0	1	0	0	1	0	1	0		0,5		1		0	0	4,5
RFA (BM ; EM 1 et EM2)			0	1	0	0	1	0	1	1		0,5		1		0	0	5,5
DisTrad		0	0	1	0	1	0	0	1	0	1			0,5		1	1	6,5
LMS <sub>DS</sub>		0	0	1	0	1	0	0	1	0	1			0,5		0	1	5,5
PR <sub>DS</sub>		0	0	1	0	1	0	0	1	0	1			0,5		0	1	5,5
GTWR		1	0	1	0	0	0	1	1	0	1			1		0	0	6

#### 1.3.4.2. Méthode de Random Forest Regression (régression Forestière aléatoire)

Les forêts aléatoires (Random Forest RF) font partie des méthodes d'apprentissage par machine (Machine Learning) basée sur un ensemble. Ce sont des familles de méthodes statistiques non paramétriques développées par Breiman en 2001 (Breiman, 2001). Les forêts aléatoires présentent l'avantage d'une grande précision avec une tolérance à la multicolinéarité des prédicteurs (Breiman, 2001 ; Yang, Cao, et al., 2017). Elles ont la capacité de modéliser les relations complexes entre les variables prédictives et les variables de réponse sans aucune relation non fonctionnelle prédéfinie entre les variables (Hutengs & Vohland, 2016). Les variables peuvent être des données catégorielles (les unités d'occupation et d'utilisation des sols) et continues (les indices de végétation -NDVI, EVI2, SAVI...- ou indices d'urbanisation -NDBI, UI,...-). L'insensibilité à la multicolinéarité admet l'ajout de multiples variables dans les forêts aléatoires.

La régression forestière aléatoire (RFR) est une approche basée sur un tirage aléatoire avec remplacement (Bootstrap sampling) au sein de l'ensemble des données (N) d'apprentissage en utilisant une partition récursive pour diviser les données en des sous-ensembles homogènes, désignés, arbres de régression et pour chaque arbre de régression, des prévisions forestières aléatoires sont calculées. Puis la moyenne arithmétique des résultats de tous les arbres de régression est faite pour avoir le prédicteur RF-RI. Cette agrégation résultant du vote majoritaire en classification permet d'améliorer la précision prédictive et de contrôler le surajustement. En effet, le principe consiste en une sélection d'un sous-ensemble de variables parmi plusieurs sous-échantillons générés aléatoirement qui décrivent les échantillons d'entraînement et, les sous-ensembles de variables non sélectionnés, Out-of-Bag (OOB) servent à évaluer les performances du modèle (Breiman, 2001).

La RFR offre plusieurs avantages dans la prévision des LST, car elle peut s'adapter à la disponibilité des données, intégrer plusieurs types de variables de prédiction : la réflectance de surface, les indices spectraux, les classes d'occupation des sols, les facteurs topographiques, etc. Le calcul de score d'importance entre les variables d'entrée permet d'établir des relations de proximité entre elles et la variable dépendante, LST. La RFR peut combiner les données multisources et multiéchelles de résolutions grossière ou fine, et minimise les erreurs du résultat final grâce au mode de classification (Bisquert et al., 2016 ; Hutengs & Vohland, 2016 ; Wu & Li, 2019). Comme autres avantages, il y a le risque minimisé de surajustement et le nombre relativement faible de paramètres du modèle à spécifier par l'utilisateur. Cependant, l'ajout de multiples types de données à la RFR peut être source de complexité de surcharge lorsque les arbres de décision sont cultivés en profondeur (Wu & Li, 2019). Il n'est pas aussi avéré que plus le nombre de variables augmente, plus il y a d'informations significatives dans le résultat final. De fait, il est suggéré d'accorder de l'importance à la rationalité des variables d'entrée plutôt qu'à la quantité en se basant sur les scores d'importance des corrélations relatives. La RFR a également l'avantage d'être un algorithme simple dans sa conception et le temps de traitement de la RF semble également beaucoup plus court que les réseaux de neurones artificiels (*Artificial Neural Network ANN*) et les

machines à vecteurs (*Support Vector Machine -SVM-*). Les résultats de réduction d'échelle de LST avec la RFR ne présentent pas des effets de lissage (Li et al., 2019).

#### 1.3.4.3. *Algorithme Pixel Block Intensity Modulation (PBIM)*

La méthode de modulation de l'intensité des blocs de pixels (PBIM) est une méthode physique linéaire basée sur les émissivités de surface. Cette méthode est recommandée par l'ASPC à des fins de comparaison avec celles retenues. Elle avait été proposée en 1998 par Guo et Moore (Guo & Moore, 1998) pour réduire les pixels thermiques de 120 m à 30 m en se basant sur les bandes spectrales optiques de Landsat TM (1, 5 et 7) et les rasters de topographie. Elle a été également utilisée comme méthode de fusion d'images multispectrales et panchromatiques. Les résultats ont montré une fidélité à l'image thermique originale. La méthode PBIM a été récemment adaptée et utilisée dans deux études pour réduire l'AVHRR LST de 1100 m à 100 m (Chybicki & Łubniewski, 2017) ou à 120 m (Stathopoulou & Cartalis, 2009) dans des zones hétérogènes urbaines. Les émissivités de surface ont été utilisées comme variables prédictives contrairement aux bandes spectrales utilisées par Guo & Moore. Il ressort de ces études que la méthode s'adapte mieux aux zones hétérogènes urbaines présentant peu de diversités en unités d'occupation du sol et avec des effets topographiques accentués. Les résultats obtenus avec cette méthode sont relativement plus précis dans des zones de petites superficies (300-14 256 km<sup>2</sup>). Dans la littérature recensée, cette méthode empirique a été très peu utilisée pour réduire la LST.

La méthode PBIM est une méthode directe qui suppose à priori la dépendance entre l'émissivité et la température de la surface terrestre.

La description du cadre opératoire des 3 méthodes avec les formules est faite en annexe B.

## 1.4. Conclusion

En conclusion de ce chapitre porté sur l'état de l'art sur les méthodes de réduction d'échelle en général, et particulièrement celles dédiées à la LST, il a été montré qu'il existe différentes catégories de réduction d'échelle présentant toutes des incertitudes sur la qualité des résultats et leurs processus de validation. S'agissant spécifiquement de la LST, l'ensemble des méthodes de réduction d'échelle ont montré leur capacité à prédire la LST avec des marges d'erreurs autour de 2°C. Il faut préciser que la qualité des résultats est subordonnée à la qualité des données utilisées en entrée, au facteur d'échelle, l'incertitude liée aux algorithmes de réduction d'échelle eux-mêmes et la morphologie de la zone d'étude.

Il ressort également de ce chapitre que parmi la panoplie de méthodes de réduction d'échelle de LST existantes, celles relevant au plan algorithmique des statistiques sont largement les plus utilisées à cause de leur facilité de mise en œuvre et de la qualité de leur résultat. Dans cette même classe, les modèles linéaires utilisant une seule variable explicative ont donné de bons résultats en zone homogène boisée et des résultats limités avec des marges d'erreur plus grandes en zones hétérogènes urbaines. Les modèles non linéaires ont apporté plus de précisions dans les zones hétérogènes urbaines. Pour tout type de méthode, la revue bibliographique a montré leur sensibilité à la saison et dans différentes unités d'occupation et d'utilisation du sol. De plus, toutes les méthodes n'arrivent pas à bien simuler les pixels possédant des valeurs extrêmes de température de surface. Partant des critères spécifiques au projet (précision, type d'image satellitaire, résolution spatiale, réplicabilité, saison...), nous avons effectué une étude comparative de ces méthodes. Ceci a permis de discriminer les méthodes. Au final, les méthodes RFR (modèle non linéaire) et TsHARP (modèle linéaire) outre la méthode PBIM (qui nous a été recommandée par l'ASPC) ont été sélectionnées pour être expérimentées dans une nouvelle approche de réduction d'échelle proposée pour atteindre 30 m de résolution spatiale. La description générale de l'approche méthodologique de réduction d'échelle de LST à 100/30 m est consacrée au Chapitre 2

- .

# **Chapitre 2 - Vers une approche méthodologique d'application de la réduction d'échelle de LST à 30 mètres**

## **2.1. Introduction**

Le chapitre précédent a fait la synthèse sur les familles des méthodes de réduction d'échelle puis s'est intéressé à celles appliquées à la température de surface avec leurs avantages et limites. Ceci a contribué à la sélection de trois méthodes pour faire les expérimentations. Dans ce chapitre, nous présentons, en deux sections, la description des étapes générales de la réduction d'échelle de LST puis l'approche méthodologique permettant d'atteindre 30 m de résolution spatiale.

## **2.2. Description des étapes générales de réduction d'échelle de la température de surface**

Dans cette section, nous décrivons les étapes générales de l'approche méthodologique de réduction d'échelle de la température de surface. L'ensemble des activités réalisées dans ce processus se regroupent en trois grandes étapes telles que définies sur le diagramme de classes UML (Unified Modeling Language) (cf. Figure 2.1). Ces trois grandes étapes sont partie intégrante des 5 étapes correspondant au canevas proposé pour atteindre une résolution de 30 m (cf. Figure 2.1). Cette sous-section présente uniquement ces trois étapes qui sont communes aux méthodes statistiques et physiques (PBIM). La première étape porte sur la préparation des données suivie de l'étape de calcul des variables auxiliaires définies et, enfin, de l'application d'un modèle de réduction de réduction d'échelle sur les données pour obtenir une image de LST à haute résolution spatio-temporelle. Chacune de ces étapes est décrite dans les sous-sections ci-dessous.

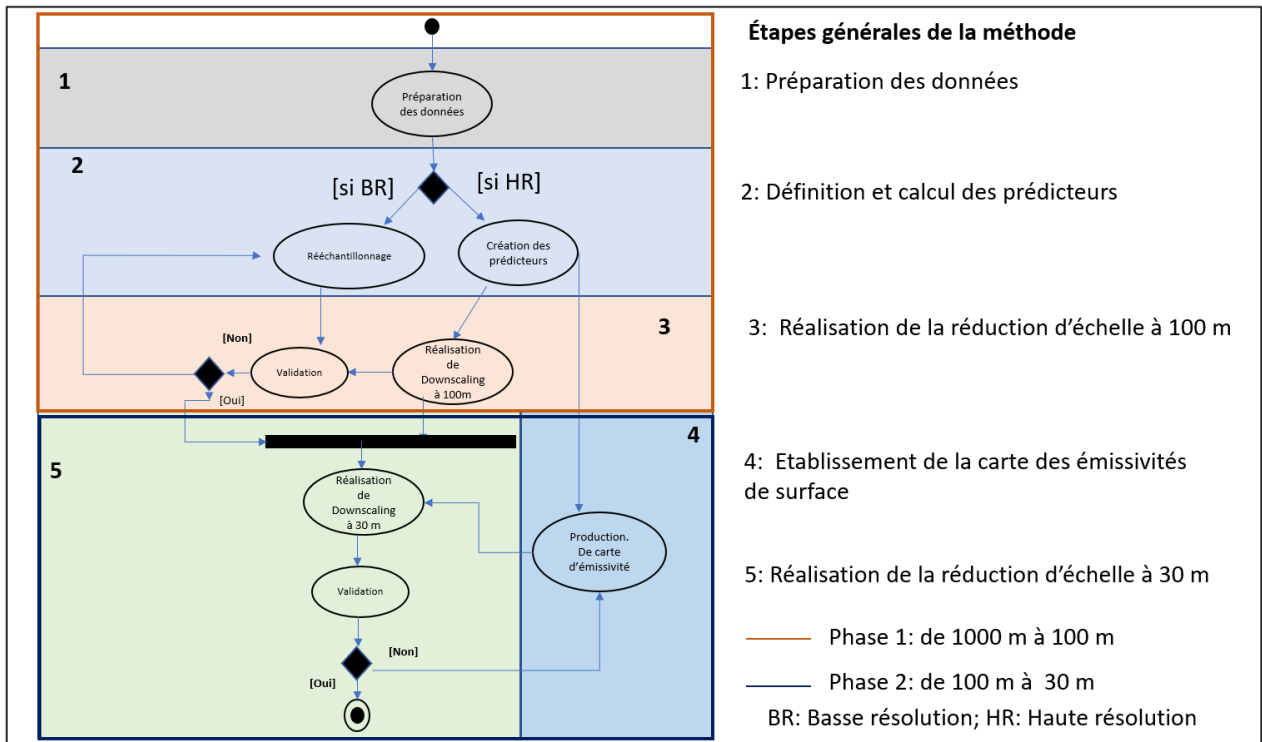


Figure 2.1 : Diagramme UML d'activités sur les étapes générales de la méthode de réduction d'échelle à 30 mètres.

## 2.2.1. Préparation des données

Cette étape consiste en des prétraitements et traitements des données géospatiales (images satellitaires) ; il s'agit de les évaluer qualitativement avant leur utilisation. Comme le montre la Figure 2.2, les prétraitements à faire sont : la reprojection, la correction géométrique, l'agrégation, le découpage et l'alignement des données géospatiales. Les traitements concernent, le masquage des données pour éliminer les bruits et les parasites et la transformation des images thermiques en donnée de température.

### 2.2.1.1. Correction géométrique et agrégation des données

L'ensemble des données géospatiales de diverses résolutions spatio-temporelles sont toutes reprojétées dans un même et unique système de projection (cartésienne de préférence). Les images téléchargées sont notamment les images thermiques de MODIS 1000 m et de Landsat 08 à 100 m, les images optiques de Landsat 08 à 30 m ou de MODIS à 250 m, et les images topographiques GDEM à 30 m de résolution.

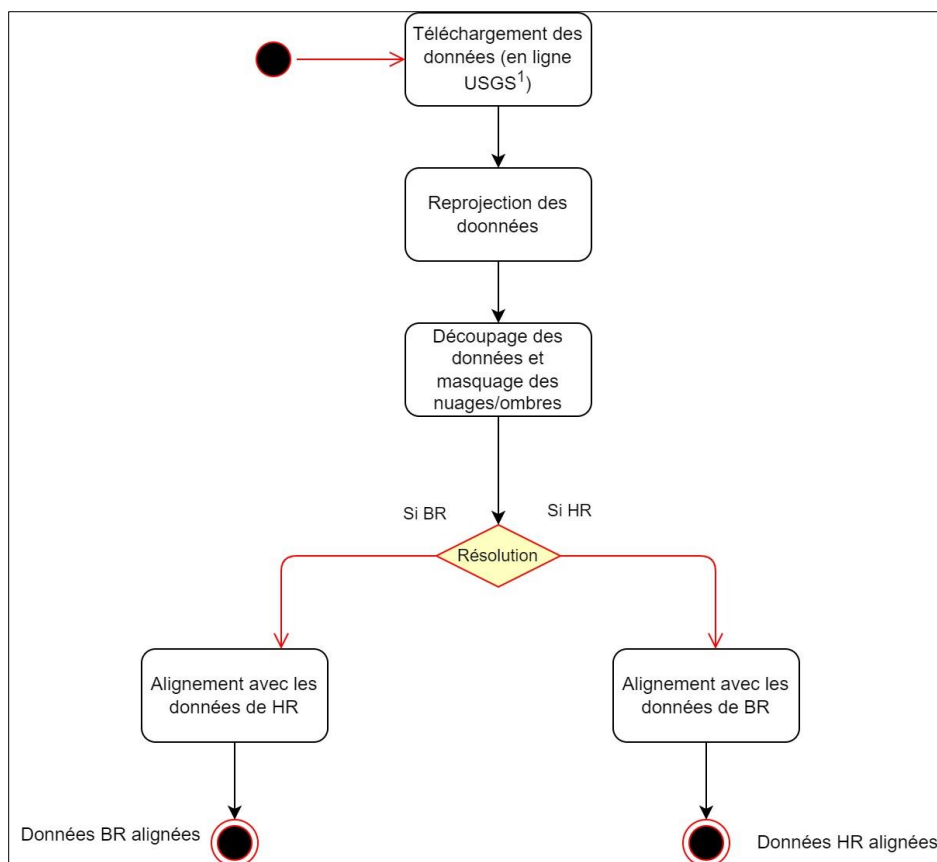


Figure 2.2: Description des tâches à réaliser à la phase de préparation des données

1 : <https://lpdaacsvc.cr.usgs.gov/appeears/>

Les résolutions spatiales (30 m, 100 m, 250 m et 1000 m par exemple) des différentes images ne sont pas équivalentes entre elles ou ne sont pas divisibles entre elles. Autrement dit, la taille des pixels d'une image A équivaut la taille de pixel de l'image B si la taille de pixel A est un multiple de la taille du pixel B ou inversement. En l'absence de ce facteur multiple, il est important de procéder à l'agrégation des données par une méthode de rééchantillonnage spatial des valeurs spectrales afin d'avoir une conformité entre les découpages spatiaux.

La méthode d'agrégation par la moyenne a été souvent appliquée, car, selon la littérature (Bian & Butler, 1999 ; Roy, 2008), elle est plus précise et conserve les propriétés originales au cours de l'opération. Ainsi pour avoir des tailles de pixels divisibles par 100 ou multiple de 100, les images satellitaires de 250 m sont rééchantillonnées à une résolution de 200 m ; et celles à 30 m de résolution sont rééchantillonnées de 30 m à 100 m tandis que les images satellitaires à 1000 m de résolution spatiale maintiennent leur résolution native respective de 1000 m.



### 2.2.1.2. Découpage, alignement et prétraitement des données

Dans l'objectif de disposer des images satellitaires de même étendue spatiale et ayant aussi le même nombre de pixels lors du passage de haute résolution à la basse résolution ou inversement, il faut découper et aligner les images. De fait, une image satellitaire sert de référence pour le découpage et l'alignement du reste. La fonction d'alignement dans le logiciel QGIS permet, par exemple, de faire en même temps ces deux opérations. Dans la mise en œuvre de cette opération, les images de basse résolution (BR) vont premièrement être rééchantillonnées à la haute résolution (HR) selon la méthode du plus proche voisin, tandis que les images de HR sont agrégées à la BR par la méthode de moyenne spatiale. En pratique, cela consiste à ramener les images thermiques de 1000 m à 100 m. En revanche, les images de couverture terrestre Landsat 08 OLI sont agrégées de 100 m à 1000 m. Les données topographiques dérivées d'ASTER GDEM doivent être aussi agrégées de 100 m à 1000 m par moyennage spatial. À l'issue de ces opérations, les données géospatiales BR sont appariées avec les données géospatiales HR pour un alignement et vice versa.

À la suite de l'alignement des données, on procède à la conversion des radiances en valeurs de température pour l'image thermique. Ainsi l'image thermique de MODIS est transformée en image de température en appliquant le facteur d'échelle (\*0,02) aux pixels ayant des valeurs supérieures à 273 Kelvin. La formule s'écrit : « Image MODIS TH » \*0,02 \*(« MOD11A1 »\*0,02>273). Les valeurs de température en Kelvin sont converties en degré Celsius.

Pour les images optiques, si elles ne sont pas préalablement traitées, elles sont transformées en données de réflectance après leur correction radiométrique. Le reste des prétraitements va consister à masquer les nuages et supprimer les valeurs aberrantes dans toutes les images. Le masque de nuage de l'image thermique est également appliqué aux images optiques et celui de ces dernières est également appliqué à l'image thermique. Ainsi, les pixels de toutes les images s'apparient les uns par rapport aux autres.

### 2.2.2. Définition et calcul des variables auxiliaires

Comme il a été démontré que divers facteurs biophysiques, topographiques, météorologiques, etc. influencent la variabilité spatio-temporelle de la température de surface, pour les caractériser, on recourt aux indices spectraux, données de réflectance de surface et paramètres topographiques.

La définition des variables auxiliaires est fonction des caractéristiques de la zone d'étude et de la méthode de réduction d'échelle à appliquer. Les variables auxiliaires encore appelées des prédicteurs servent à la simulation des températures de surface. Sur la Figure 2.3, nous présentons une vue détaillée des prédicteurs à calculer selon la typologie des données HR alignées. De manière générale, quatre grandes familles de données

auxiliaires sont définies à partir des données spectrales et topographiques: 1) les données de réflectance en HR, 2) les indices de végétation et d'urbanisation, 3) les paramètres topographiques et 4) les classes d'occupation et d'utilisation du sol.

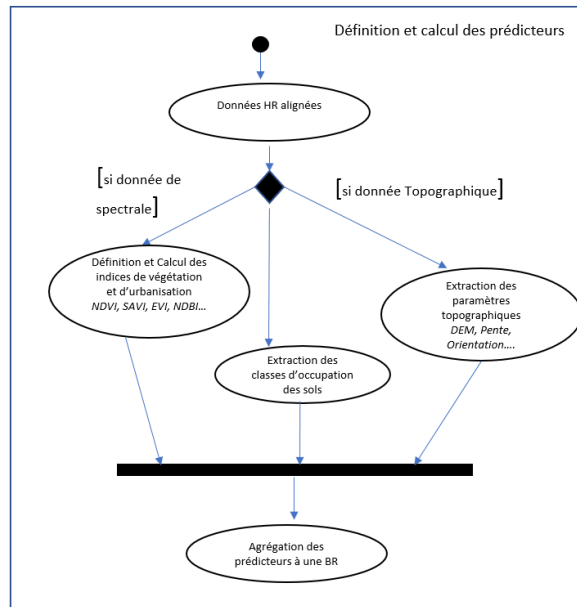


Figure 2.3 : Détail des tâches à réaliser à la phase de définition des prédicteurs

### 2.2.2.1. Données de réflectance à haute résolution HR et indices spectraux

Les données de réflectance sont caractérisables par les bandes spectrales 2, 3, 4 (bandes Bleu, Vert, Rouge) du domaine visible et les bandes infrarouges à ondes courtes (bandes 5, 6 et 7) du capteur Landsat 08/OLI sont souvent utilisées comme variables auxiliaires intégrées dans les modèles non linéaires basés sur plusieurs variables. Les bandes 6 et 7 permettent de différencier la terre humide de la terre sèche, les roches des sols. La bande 5 apporte plus d'informations sur l'écologie, notamment la biomasse, la densité du couvert végétal et la santé des plantes tandis que les bandes 2, 3 et 4 permettent la description de la couverture terrestre et l'utilisation du sol.

Les indices spectraux, relatifs aux indices de végétation et d'urbanisation, permettent de caractériser également l'occupation et l'utilisation du sol. Le calcul de ces indices se base sur les différentes bandes spectrales du capteur imageur à haute résolution spatiale retenu. Il ressort de la littérature que les indices de végétation retenus varient en fonction des objectifs d'étude et des caractéristiques de la zone d'étude. Comme exemple d'indices spectraux, il y a : les indices de végétation différenciés normalisés (NDVI), végétation améliorée (EVI), végétation ajustée au sol (SAVI), couverture végétale fractionnée (FVC) considérés comme des indicateurs renseignant sur la densité de la végétation et des biomasses, la variation de la structure de la canopée, la densité foliaire et la santé de la végétation. Les indices de construction par différence normalisée (NDBI), Indice

d'eau différencié normalisé (NDWI), Indice d'eau différencié normalisé modifié (MNDWI), Indice urbain (IU), Indice de construction basé sur un indice (IBI), Indice d'aridité ou de sol nu (BSI), renseignent spécifiquement sur les bâtiments, les surfaces humides, l'urbanisation et les sols nus. La définition et le mode de calcul de certains indices spectraux avec les bandes spectrales optiques de Landsat 08 OLI sont présentés dans la section 3.3.1. Le Tableau 0.6 en annexe présente aussi le mode de calcul des indices recensés dans la littérature en lien avec la réduction de LST.

#### *2.2.2.2. Facteurs topographiques et carte d'occupation/utilisation des sols*

Les paramètres topographiques décrivent le relief de la zone d'étude. Les plus utilisés comme variables de prédiction de la LST sont la pente, l'altitude, l'exposition et le modèle numérique de terrain. Ces paramètres sont dérivés des images ASTER GDEM ou ASTER STRM. L'exposition donne l'orientation de la pente par rapport au nord dans le sens horaire. Elle varie de 0 à 360°. La valeur 0 indique le Nord.

La nomenclature adoptée pour réaliser la carte d'occupation et d'utilisation des sols dépend de la morphologie de chaque zone d'étude et du niveau de détails à atteindre dans la description de la couverture terrestre. Les unités d'occupation et d'utilisation du sol sont extraites le plus souvent d'une classification d'images satellitaires optiques ou de bases de données existantes. Les grandes classes d'occupation rencontrées sont : la classe de végétation, les zones bâties/imperméables, la classe des zones humides, la classe des zones agricoles, etc.

#### **2.2.3. Réalisation de la réduction d'échelle à une résolution cible**

Sur la Figure 2.4, nous présentons une vue détaillée des tâches à réaliser pour réduire une image de LST (°C) de basse résolution (BR) à haute résolution (HR). L'ensemble des tâches à réaliser sont regroupées en trois sous-étapes dont : (1) la normalisation des données de prédiction BR et HR en raison des caractéristiques différentes au niveau des résolutions spectrales des capteurs d'acquisition des images satellitaires, (2) l'application d'une méthode de réduction d'échelle, (3) la correction des résidus pour finaliser la dérivation de LST à 100 m ou 30 m (par exemple).

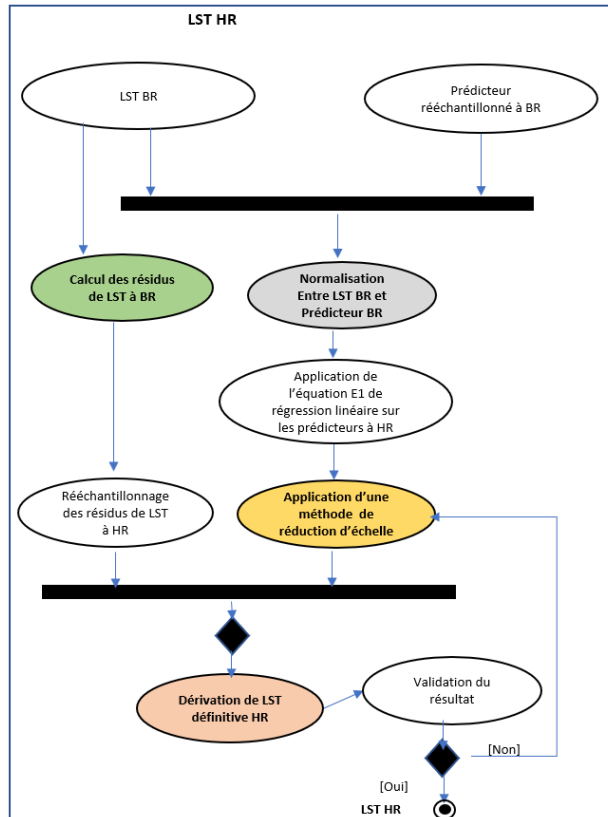


Figure 2.4 : Description des tâches à réaliser à l'étape 3

### 2.2.3.1. Normalisation et application d'une méthode de réduction d'échelle

Dans le processus de réduction d'échelle, la normalisation est indispensable surtout si les images satellitaires proviennent de différents capteurs afin de minimiser les différences entre les images thermiques et les images optiques. La normalisation permet d'avoir des données spectrales équivalentes entre les deux capteurs. En effet, il existe des différences structurelles entre les capteurs (Bindhu et al., 2013). Ces différences sont liées à la résolution spatiale, la correction atmosphérique, l'angle de vue, à l'empreinte des pixels ou encore à la différence de temps d'acquisition. La normalisation est faite à partir des coefficients de conversion intercapteurs obtenus par régression linéaire. Cette étape de normalisation doit être appliquée à toute paire d'images HR et BR à utiliser. Dans le processus de normalisation, il faut rééchantillonner les données de prédiction (HR) à la même résolution que l'image de LST (BR), puis y appliquer une régression linéaire afin d'identifier les paramètres « a » et « b » pour la définition de l'équation de régression linéaire. L'équation est appliquée aux données prédictives de HR pour la normalisation. Après la normalisation, on applique une méthode de réduction d'échelle sur les données prétraitées pour obtenir une image réduite primaire de LST à haute HR.

### *2.2.3.2. Correction de LST Haute résolution primaire avec des résidus calculés*

Le calcul des résidus est nécessaire parce que l'ensemble des prédicteurs utilisés n'expliquent pas les variabilités de la LST. Ainsi, le résidu est l'effet local de LST non expliqué par les variables auxiliaires retenues. Son calcul permet d'intégrer les autres facteurs locaux explicatifs dans le processus de réduction d'échelle.

Le calcul des résidus se fait avec les données LST BR observées et les LST BR calculées, puis le résultat est rééchantillonné à une HR. Le résultat rééchantillonné à HR est ajouté à la première image réduite précédemment obtenue pour obtenir définitivement une LST HR. Ce résultat de LST HR sera validé qualitativement et quantitativement par comparaison avec une image LST de référence (Landsat 08 TIRS LST par exemple). Si les marges d'erreur sont au-dessus du seuil de précision défini alors, le processus est repris à partir de l'application de la méthode de réduction.

À travers cette section, nous avons montré la procédure globale de réduction d'échelle de la LST. Les étapes préalables indispensables sont le prétraitement des données et la définition des variables auxiliaires pour la prédiction. Le choix d'une ou plusieurs variables de prédiction est fonction du modèle (linéaire, non linéaire), de l'algorithmique (statistique, physique) de la méthode de réduction d'échelle à appliquer ou encore des caractéristiques morphologiques (forte hétérogénéité ou grande homogénéité) de la zone d'étude. Enfin, après l'application de la méthode de réduction d'échelle, le calcul des résidus permet de prendre en compte la variabilité spatio-temporelle de la LST non expliquée par les prédicteurs utilisés.

La mise en application de ces procédures dans cette recherche s'insère dans un cadre opératoire décrit dans la section suivante.

## **2.3. Description d'une nouvelle approche de réduction d'échelle**

L'approche proposée pour aboutir à une réduction d'échelle à 30 m est un canevas en deux phases : la première consiste à réduire la MODIS LST de 1000 m à 100 m, puis, d'affiner ce résultat à 30 m à la seconde phase. À chacune des phases, les méthodes TsHARP et RFR ont été appliquées. La différence fondamentale est le choix des variables auxiliaires à chaque phase. Nous ne proposons pas une nouvelle méthode de réduction d'échelle, mais il s'agit d'utiliser celles existantes dans ce canevas pour atteindre la résolution finale de 30 m. Pour la méthode PBIM, cette approche en deux phases n'a pas été appliquée vu que la variable auxiliaire (les émissivités de surface à utiliser à la seconde phase) reste la même pour les deux phases définies. En conséquence, le mode opératoire décrit concerne uniquement les modèles statistiques et ne s'applique pas aux méthodes physiques (PBIM).

L'approche proposée se veut simple et est basée sur le postulat que les émissivités de surface permettent de prédire la LST à 30 m avec une plus grande justesse. Le nouveau cadre méthodologique de réduction d'échelle se décline en cinq étapes générales représentées en diagramme (cf. Figure 2.1). La première phase (dans l'encadré en orange) comprend la réalisation des étapes 1, 2 et 3 précédemment décrites dans la section 2.2 et la seconde (dans l'encadré en bleu foncé) porte sur les étapes 4 et 5 que sont l'établissement de la carte des émissivités de surface et la réalisation de la réduction d'échelle à 30 m. Ces deux phases sont présentées en deux sous-sections dans les lignes suivantes.

### 2.3.1. Première phase : réduction d'échelle à 100 m

La première phase étant décrite à la sous-section 2.2.3 de la section 2.2 alors à cette étape, nous apportons de précisions sur l'adaptation de la méthode TsHARP à cette étude et la précision sur les variables de prédiction. La méthode TsHARP basée généralement sur une seule variable a été adaptée à cette étude avec plusieurs variables descriptives de la zone d'étude.

La méthode TsHARP a été testée dans différentes zones aux caractéristiques environnementales variées (zone urbaine hétérogène, zone hétérogène à dominante agricole ou boisée) et en toutes saisons (Hiver, printemps, automne et été). Différents auteurs ont pris la méthode TsHARP comme référence et ont comparé leur nouvelle méthode à TsHARP. La plupart de ces études n'ont utilisé qu'une seule variable avec TsHARP, la couverture de végétation fractionnée. Or, la couverture végétale fractionnée ou l'indice de végétation n'est parfois pas adapté (e) aux conditions environnementales de la zone d'étude. En milieu urbain hétérogène, où il y a peu de végétation, l'indice spectral caractérisant le mieux ce type de milieu est l'indice urbain du bâti ou l'indice urbain. En effet, plusieurs études ont conclu à l'inadaptation du NDVI pour réduire la LST dans les zones urbaines hétérogènes (Dominguez et al., 2011 ; Essa et al., 2012 ; Govil et al., 2019 ; Stathopoulou & Cartalis, 2009 ; Zakšek & Oštir, 2012). En remplacement de l'indice de végétation, l'indice de pixel pur ou le pourcentage de surface imperméable a été proposé à cause de la forte relation linéaire entre la LST et ces indices (Essa et al., 2012 ; Guijun Yang et al., 2010). Comme le milieu urbain hétérogène est caractérisé par différentes unités d'occupation et d'utilisation du sol, en intégrant plus de variables auxiliaires adaptées et représentatives des états de surface, le résultat de réduction d'échelle présenterait certainement moins de biais. C'est ce modèle avec plusieurs variables auxiliaires que nous mettons en œuvre avec l'algorithme TsHARP dans cette étude. Govil et al. (2019) ont déjà utilisé le TsHARP avec quatre indices différents NDBI, NDVI, NDWI, NMDI pour réduire la LST de 100m à 30 m pour un milieu urbain hétérogène. Dans leur approche, l'objectif était d'identifier lequel des indices permettait d'avoir une bonne précision de LST.

À la différence de leur approche, nous proposons une régression linéaire multiple. Les variables basées sur les indices spectraux et la topographie ont été intégrées au modèle TsHARP. La LST HR sera calculée pour chaque variable de prédiction définie et la moyenne de l'ensemble donnera la LST HR de la zone d'étude.

Le choix des variables prédictives est lié à la morphologie de la zone d'étude et les objectifs de cette recherche. Outre les spécificités de la zone d'étude (milieu hétérogène et mixte), ce sont également les biotopes des différents vecteurs responsables des maladies du Virus du Nil Occidental (VNO) et de Lyme qui ont orienté dans le choix des indices de végétation et biophysiques. Du fait que les tiques transmetteurs de maladie de Lyme se reproduisent dans les biomes forestiers à couvert denses ou fragmentés (Kotchi et al., 2019 ; Ruiz et al., 2010), nous avons retenu les indices de végétation différencié normalisé (NDVI), végétation améliorée (EVI), végétation ajustée au sol (SAVI), couverture végétale fractionnée (FVC) considérés comme des indicateurs renseignant sur la densité de la végétation et des biomasses, la variation de la structure de la canopée, la densité foliaire et la santé de la végétation. Les indices de construction par différence normalisée (NDBI), Indice d'eau différencié normalisé (NDWI), Indice d'eau différencié normalisé modifié (MNDWI), Indice urbain (IU), Indice de construction basé sur indice (IBI), Indice d'aridité ou de sol nu (BSI) ont été retenus, car les biotopes spécifiques aux moustiques, responsables du VNO sont les milieux urbains, milieux humides naturels en zones rurales. L'altitude, la pente et l'exposition au soleil complètent ces variables.

Une fois les variables de prédiction sélectionnées, on applique les modèles mathématiques de TsHARP ou de RFR en suivant la procédure de réduction d'échelle de LST décrite au Chapitre 2 - .

### 2.3.2. Seconde phase : réduction d'échelle à 30 m

À ce stade, une seule variable auxiliaire, en l'occurrence les émissivités de surface, a été utilisée pour réduire la température de surface à 30 m. Le recours aux valeurs d'émissivité de surface est justifié par la relation existante entre les émissivités de surface et la température de surface. La connaissance des émissivités de surface est souvent nécessaire dans l'estimation précise de la température de surface à partir des mesures de radiance (Sobrino et al., 2008). Ainsi, plusieurs techniques ou modèles de dérivation de la température de surface y recourent dans la modélisation du transfert radiatif.

Les émissivités de surface avaient déjà été utilisées dans les modèles de type physique (PBIM) pour réduire la LST des zones hétérogènes urbaines (Chybicki & Łubniewski, 2017 ; Nichol, 2009 ; Stathopoulou & Cartalis, 2009). Elles ont également été utilisées pour ajuster la LST réduite avec un modèle statistique non linéaire comme RFR (Bonafoni et al., 2016). Bonafoni et ses collègues ont réduit la Landsat LST 160m de la ville de Florence à 30 m à partir de 18 variables auxiliaires. Le résultat a été ensuite ajusté grâce aux émissivités de surface basées sur la couverture terrestre. Cet ajustement selon les auteurs a permis de réduire les écarts des

données entre l'image LST réduite, l'ASTER LST de référence. Ainsi, nous nous inspirons de cette méthode pour l'appliquer dans la nouvelle approche méthodologique. Ici, la carte des émissivités de surface va plutôt servir de variable auxiliaire. Nous présentons dans la sous-section suivante le cadre théorique du calcul des émissivités.

### *2.3.2.1. Cadre théorique de calcul des émissivités*

L'émissivité de surface terrestre ( $\varepsilon$ ) intervient dans la caractérisation des surfaces et la correction des données thermiques, notamment la température réelle de surface. L'émissivité est assimilée à l'énergie émise par un objet ou un corps sous forme de radiations électromagnétiques. Autrement, c'est le rapport entre les énergies émises par le corps observé et celles rayonnées par le corps noir dans les mêmes conditions. Les valeurs d'émissivité varient de 0 à 1. Les corrections des températures de surface à l'aide des valeurs d'émissivité peuvent éviter des marges d'erreur de 0,2 à 1,4 pour  $\varepsilon = 0.98$  et une hauteur par rapport au sol de 0 km (Dash et al., 2002). Plus la valeur estimée de l'émissivité est précise, soit un intervalle de confiance de 99,5%, la température de surface est plus précise (Sobrino et al., 1991).

L'estimation de l'émissivité de surface se base sur deux approches : l'une requiert l'estimation du NDVI et la fraction de végétation et l'autre consiste à attribuer des valeurs de l'émissivité aux différentes classes d'utilisation des terres. Ces valeurs sont des mesures fixes (banque d'émissivité spectrale) ou répertoriées dans la littérature existante (Benmecheta, 2017 ; Zhang et al., 2014). L'approche basée sur le NDVI se décompose en trois variantes : la variante axée sur l'image NDVI, celle sur le rapport végétation/sol nu et enfin la méthode sur le seuil du NDVI (Benmecheta, 2017). L'approche basée sur la classification de l'image paraît plus simple et a été utilisée dans certaines études de réduction d'échelle de la température de surface ou simplement dans le calcul de température de surface. Pour cette étude, nous avons opté pour le calcul des émissivités basé sur la classification de l'image.

Les principales actions à réaliser pour le calcul des émissivités de surface sont : la classification des unités d'occupation du sol suivant une nomenclature précise, l'identification d'une banque de données des émissivités pour le satellite Landsat et le calcul des émissivités effectives (cf. Figure 2.5). Ces actions sont décrites dans les sous-sections qui suivent.



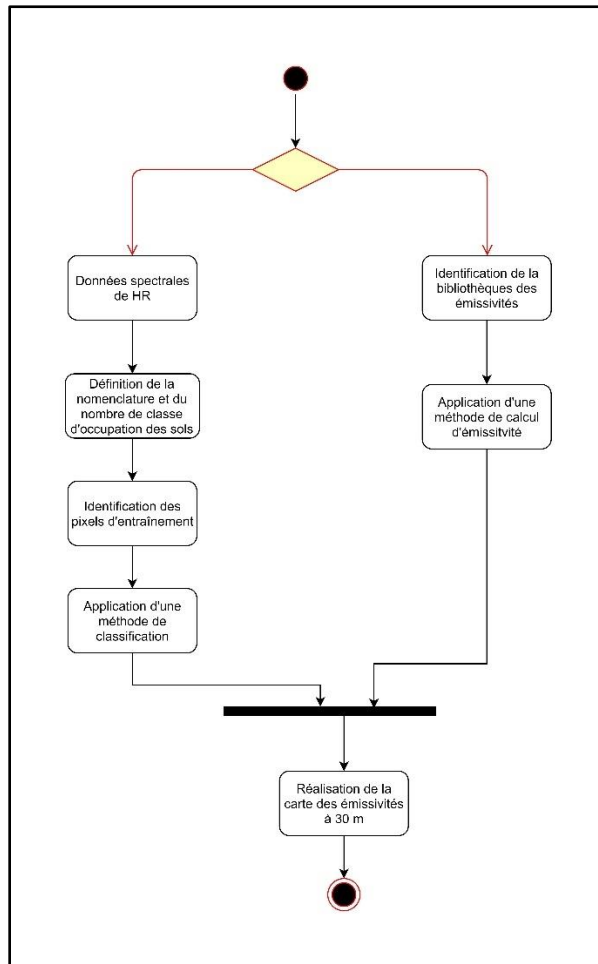


Figure 2.5: Processus de réalisation de la carte des émissivités de surface

### 2.3.2.2. Classification des unités d'occupation et d'utilisation du sol

La nomenclature proposée dans le document projet de l'ASPC a servi à la classification des déterminants environnementaux (travaux réalisés par les étudiants en génie géomatique ; cf. Tableau 0.7 en annexe A), mais ne convient pas pour l'identification des déterminants climatiques ou microclimatiques. En effet, c'est la fonction jouée par l'environnement en lien avec la température dans la prolifération des vecteurs étudiés qui est mise en avant. De fait, ce sont les couvertures du sol ayant des émissivités proches qui ont guidé le choix et le nombre de classes. Au total, 6 classes (surface imperméable/Bâti, sol nu, Zone humide, Eau de surface, Végétation et Forêt dense) ont été définies. Les sous-classes des tableaux en annexe ont été affectées à l'une ou l'autre des 6 classes retenues. Ainsi, la classe de végétation renferme les herbacés, les prairies, les arbustes et les zones agricoles.

Afin d'identifier ces classes, les bandes 2, 3, 4, 5, 6 et 7 optiques de Landsat 8 OLI ont été assemblées. L'approche utilisée est la méthode de classification semi-supervisée pour les zones d'apprentissage puis l'algorithme par maximum de vraisemblance a été appliqué à l'ensemble de la zone d'étude pour la classification.

### *2.3.2.3. Estimation de l'émissivité effective*

Selon la littérature, pour chaque unité d'occupation du sol, la valeur de l'émissivité varie de 0,90 à 0,99 (Kerr et al., 1992 ; Snyder et al., 1998). Les valeurs d'émissivité mesurées sont des valeurs théoriques dont les corrections donnent les valeurs d'émissivité effective. Dans l'infrarouge thermique, l'émissivité est proche de 1 pour l'eau, elle est comprise entre 0,97 et 1 pour la végétation (Van De Griend & Owe, 1993) et, de façon générale, elle varie selon les terrains de 0,90 à 0,98 (Kerr et al., 1992 ; Snyder et al., 1998). La base de données des émissivités de divers matériaux nommée Ecosystem Spaceborne Thermal Radiometer Experiment on Space Station (ECOSTRESS) a été téléchargée sur le site <https://speclib.jpl.nasa.gov/>. Les valeurs d'émissivité affectées aux unités de surfaces sont : zones artificialisées  $\epsilon = 0,964$ , eau de surface  $\epsilon = 0,988$ , Végétation  $\epsilon = 0,985$ , forêt dense  $\epsilon = 0,967$  et sol nu  $\epsilon = 0,967$ , zones humides  $\epsilon = 0,983$  (Kotchi, 2015 ; Li & Meng, 2018).

À ce stade, pour la réduction d'échelle à 30 m, on applique également la procédure générale de réduction d'échelle de LST (°C).

## **2.4. Conclusion**

Ce chapitre a résumé les étapes descriptives de la réduction d'échelle de LST et le canevas méthodologique pour atteindre une résolution spatiale de 30 m. Ce canevas méthodologique consiste à réduire une image LST MODIS 1000 m à 100 m puis de 100 m à 30 m. Les méthodes à appliquer dans ce cadre méthodologique sont le TsHARP, le RFR et le PBIM. La réduction d'échelle de LST avec le TsHARP et le RFR est basée sur divers prédicteurs tels les indices de végétation, d'urbanisation et topographiques contre un seul prédicteur, les émissivités de surface pour la méthode PBIM. Les étapes de mise en application de ce canevas méthodologique sont mises en exergue dans le Chapitre 3 - .

# Chapitre 3 - Démarche générale d'implémentation de la nouvelle méthode de réduction d'échelle

## 3.1. Introduction

Cette section présente le cadre général de mise en œuvre de la nouvelle approche de réduction d'échelle proposée et présentée dans le précédent chapitre. Le diagramme d'activités illustre la conception de ce cadre général à la Figure 3.1 avec les activités suivantes : l'identification de trois zones d'étude, la description des données, l'analyse de colinéarité entre les prédicteurs définis, la mise en application des méthodes de TsHARP, RFR et PBIM à 100 m puis à 30 m et les méthodes de validation des résultats.

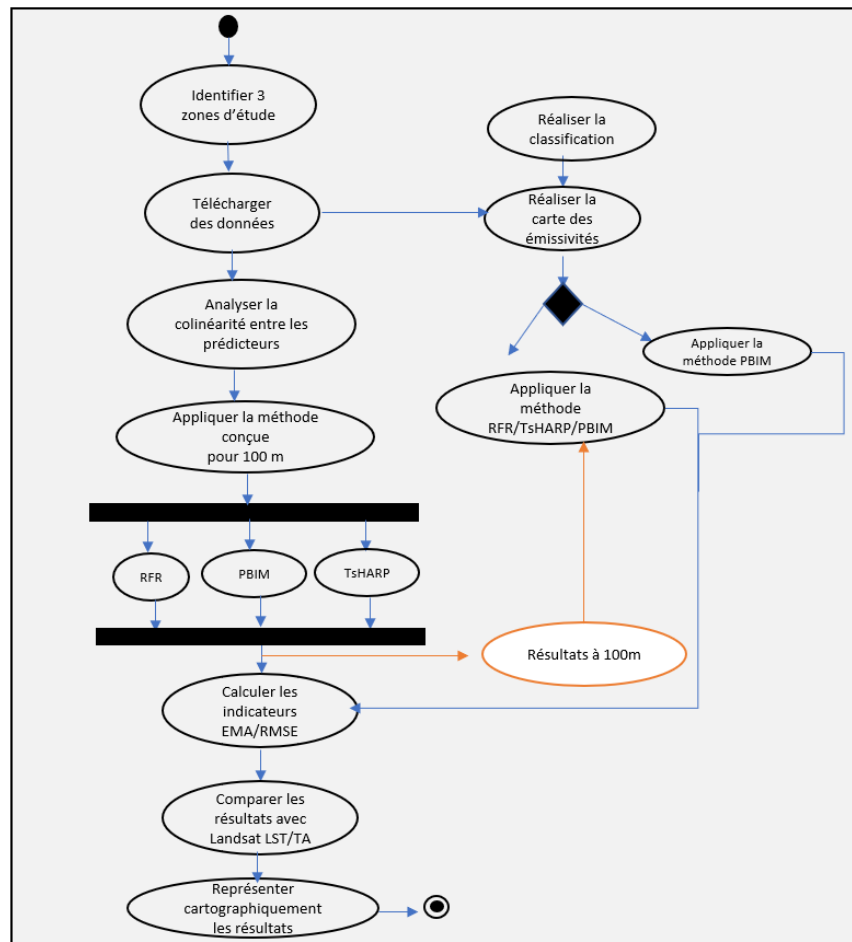


Figure 3.1 : Cadre général d'implémentation de la réduction d'échelle de LST à 100 m et à 30 m

### 3.2. Identification de trois zones d'étude

Au Canada, la province de Québec est la plus touchée en 2018 par le Virus du Nil Occidental (VNO) selon l'ASPC. C'est pourquoi nous avons choisi trois zones d'étude dans cette province pour faire nos expérimentations. Au Québec, les régions sociosanitaires d'Estrie, de Montérégie, et de Mauricie/Centre de Québec ont un taux d'incidence annuel respectif de 35,2 ; 4,5 et 1,9 pour 100 000 personnes selon le rapport de surveillance de la Maladie de Lyme en 2017. En outre de ces données, le critère de la couverture terrestre a également été pris en compte dans le choix des zones d'étude. Ainsi, une zone hétérogène à dominance urbaine, une zone hétérogène à dominance forestière et une zone hétérogène à dominance agricole correspondent respectivement aux régions de la Capitale-Nationale, de l'Estrie et de Montérégie.

**La région de la Capitale nationale** couvre une superficie d'environ 20 000 km<sup>2</sup> et regroupe au plan administratif six municipalités régionales de comté (MRC) outre l'agglomération de Québec et de la ville de Lévis. Selon le plan d'affectation du territoire public, c'est une zone à dominante forestière, car son territoire est couvert à 87 % par la forêt contre 7% pour les zones urbaines et agricoles. La zone d'échantillonnage (Figure 3.2) couvre en grande partie la communauté métropolitaine de Québec qui est un milieu de grande hétérogénéité en occupation et utilisation du sol.

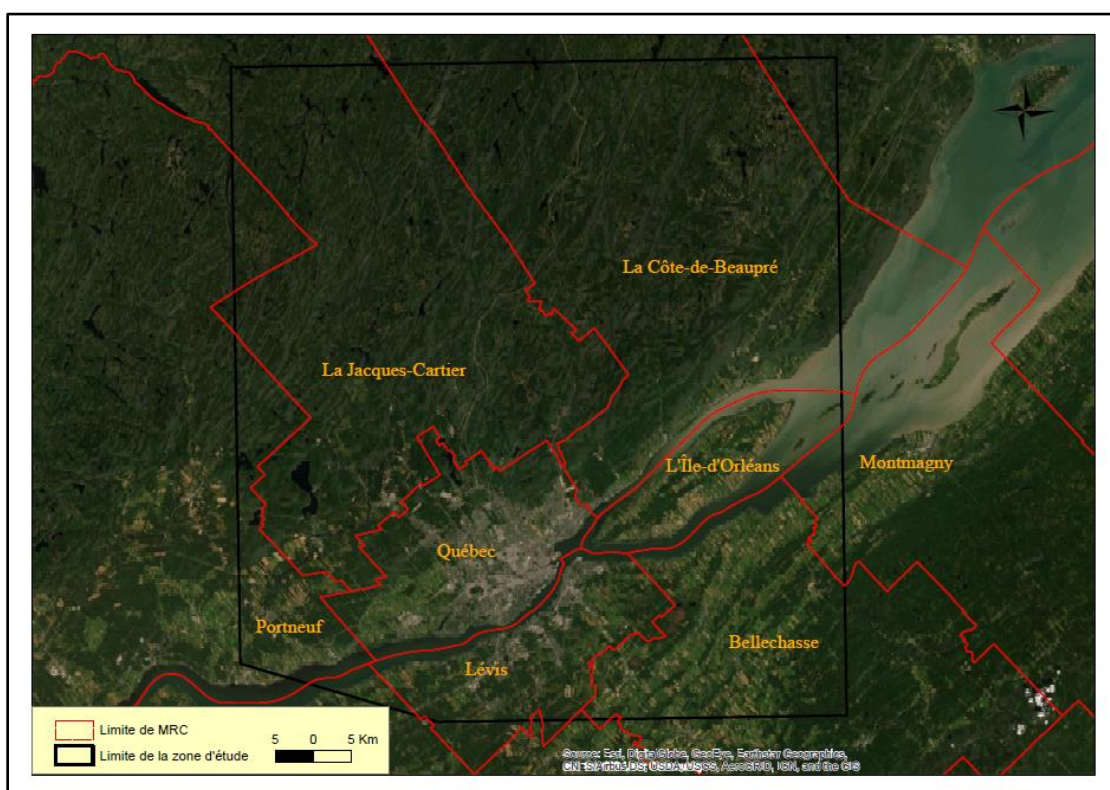


Figure 3.2 : Vue globale de la zone d'étude : région de la Capitale nationale

**La région de l’Estrie** compte administrativement sept municipalités régionales de comté. Son territoire est également dominé par le couvert forestier à 76%. La zone d’étude couvre entièrement les municipalités régionales de comté du Haut-Saint-François, Val-Saint-François, ville de Sherbrooke, et les Sources, et partiellement le reste de MRC. La zone d’étude (cf. Figure 3.3) s’étend partiellement sur deux MRC (La Haute Yamaska et Acton) de la région d’Estrie et deux autres MRC (Drummond et Arthabaska) de la région du Centre du Québec. Les zones urbaines dominantes de la zone d’étude sont les MRC de Sherbrooke et Drummond.

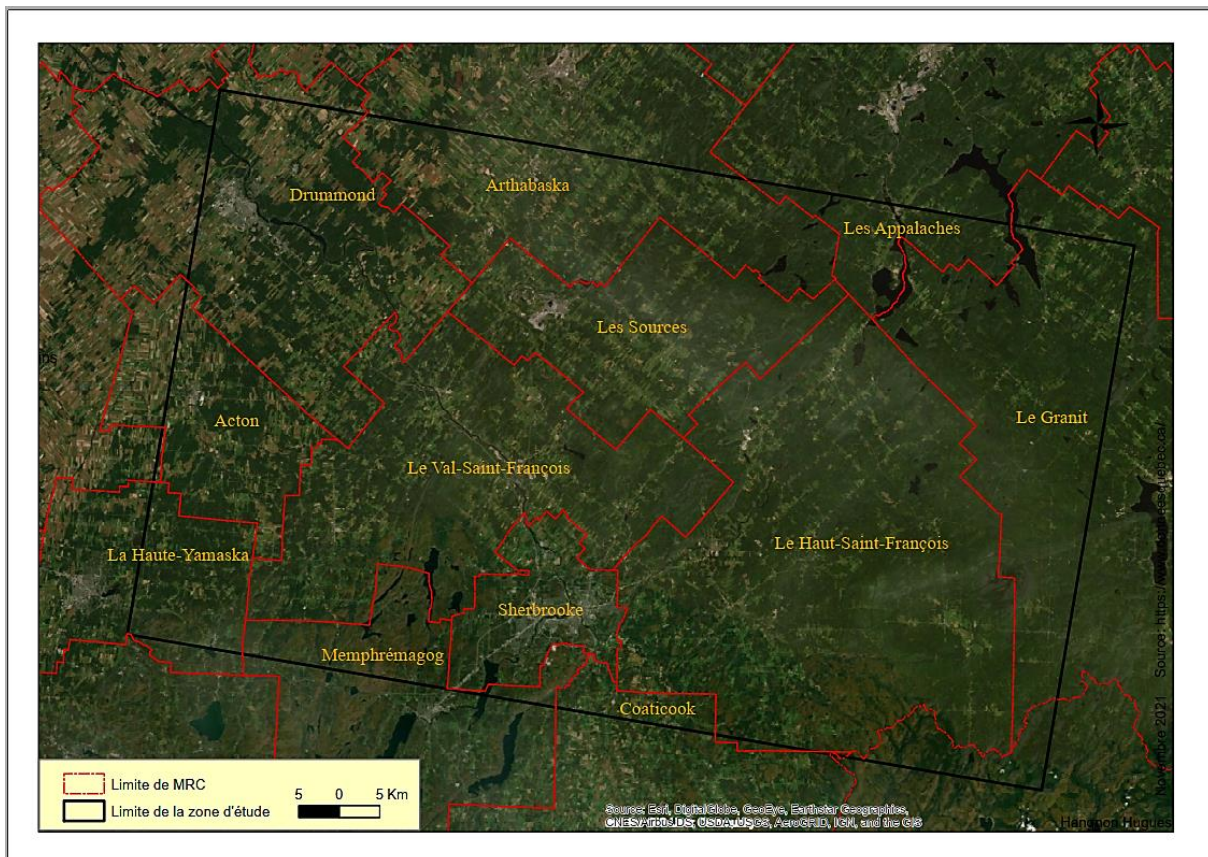


Figure 3.3 : Vue globale de la zone d’étude : Région de l’Estrie

**La région administrative de Montérégie**, au sud-ouest du Québec, est composée de 12 municipalités régionales de comté (MRC) et l’agglomération de Longueuil. Selon le rapport GéoMont (2018), l’occupation et l’utilisation du sol sont à dominance agricole. Le territoire régional de la Montérégie est couvert à 81 % par de zones agricoles en dépit du récent développement urbain ou industriel observé. La zone d’expérimentation couvre entièrement les MRC de Beauharnois-Salaberry, Roussillon, Les Jardins-de-Napierville, Le Haut-Richelieu et partiellement, Le Haut-Saint-Laurent, Vaudreuil-Soulanges, Brome-Missisquoi et Rouville (Figure 3.4).

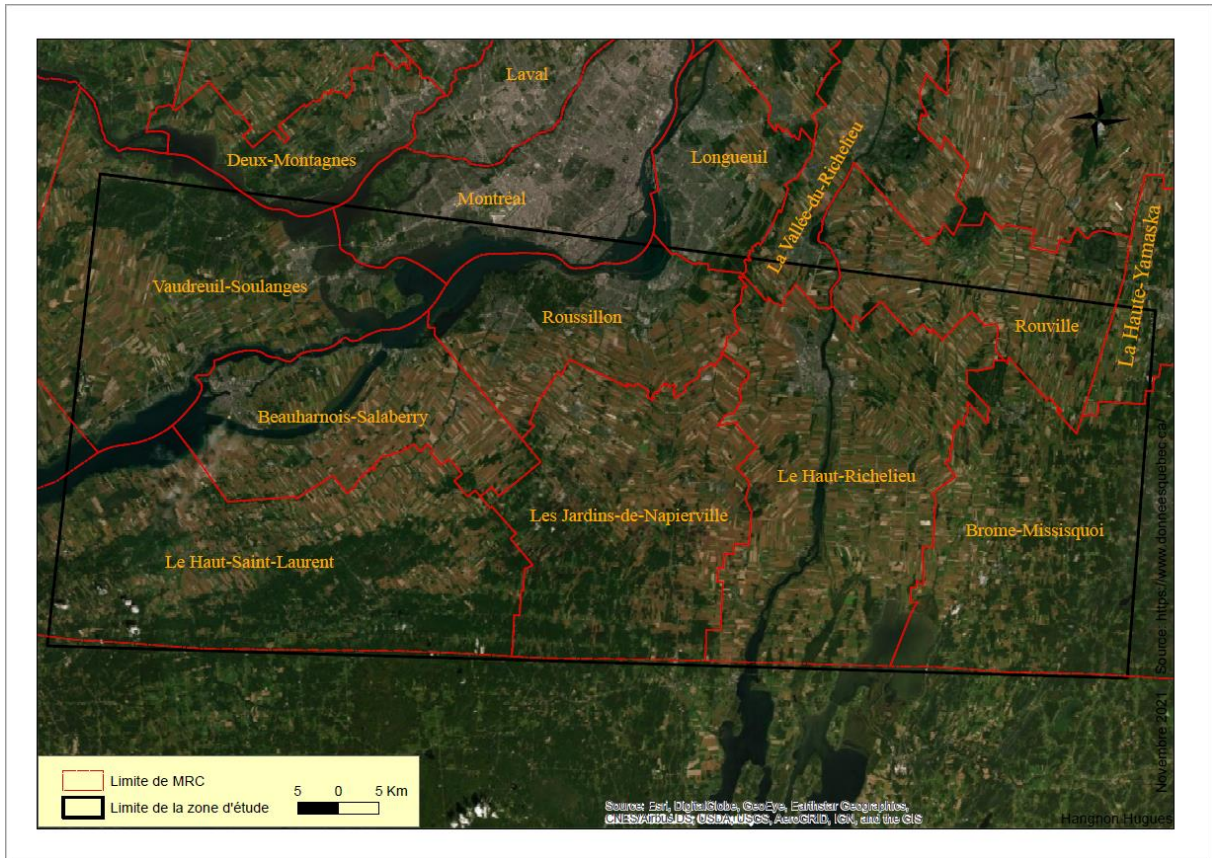


Figure 3.4 : Vue globale de la zone d'étude : région de Montérégie

### 3.3. Description des données

Les données thermiques, de réflectance et topographiques sont les types de données acquises et utilisées dans cette étude (cf. Tableau 3.1). La donnée thermique de MODIS est enregistrée au moins quatre fois par jour pour l'ensemble du territoire canadien par les deux satellites Aqua/Terra qui embarquent le capteur MODIS. Les données de réflectance proviennent de l'imagerie Landsat 08/OLI dont le capteur a une revisite tous les 16 jours. Enfin, les données topographiques sont fournies par le capteur ASTER<sup>16</sup> à bord du satellite Terra.

Les données thermiques MODIS LST (MOD11A1 version 6) de résolution spatiale de 1 km sont des températures moyennes quotidiennes de surface terrestre. Le facteur d'échelle (0,02) a été appliqué pour obtenir premièrement la température de surface en unité de Kelvin ; puis de ce résultat, nous avons soustrait 273,15 pour convertir la température de surface en degré Celsius. Les données de réflectance de surface de Landsat Analysis Ready Data (ARD) CONUS sont des données prises à un instant « T » de la journée tous les 16 jours et sont d'une résolution spatiale de 30 m. Le facteur d'échelle (0,0001) a été appliqué sur les pixels de chacune des sept bandes pour avoir des images avec des réflectances de surface. Les données ASTER Global

<sup>16</sup> Advanced Spaceborne Thermal Emission and Reflection Radiometer

Digital Elevation Model (GDEM) Version 3, statiques dans le temps, représentent le modèle numérique d'élévation des zones terrestres à une résolution spatiale de 1 seconde d'arc (environ 30 mètres d'affichage horizontal à l'équateur). Les détails sur ces données sont présentés dans le Tableau 3.1 ci-dessous.

Tableau 3.1 : Description des données utilisées

Zones d'étude	Types	Résolution spatiale (m)	Capteur	Sources	Date d'acquisition
Région de la Capitale-Nationale	Thermique MOD 11 A1	1000	MODIS	<a href="https://lpdaacsvc.cr.usgs.gov/appeears/task/area">https://lpdaacsvc.cr.usgs.gov/appeears/task/area</a>	12/07/2013
	Réflectance de surface	30	Landsat 08 OLI		
	Topographique GDEM2	30	ASTER		
Région de la Montérégie	Thermique MOD 11 A1	1000	MODIS	<a href="https://lpdaacsvc.cr.usgs.gov/appeears/task/area">https://lpdaacsvc.cr.usgs.gov/appeears/task/area</a>	05/08/2019
	Réflectance de surface	30	Landsat 08 OLI		
	Topographique GDEM2	30	ASTER		
Région de l'Estrie	Thermique MOD 11 A1	1000	MODIS	<a href="https://lpdaacsvc.cr.usgs.gov/appeears/task/area">https://lpdaacsvc.cr.usgs.gov/appeears/task/area</a>	23/07/2017
	Réflectance de surface	30	Landsat 08 OLI		
	Topographique GDEM2	30	ASTER		

Toutes les données acquises ont été acquises pendant la saison estivale et plus précisément en juillet pour les régions de la Capitale-Nationale et de l'Estrie, et en août pour la région de Montérégie (cf. Tableau 3.1 ci-dessus). L'ensemble des données ont été acquises sur une même plateforme : Application for Extracting and Exploring Analysis Ready Samples « AppEEARS »<sup>17</sup>. L'intérêt de cette URL via LP DAAC's pour le téléchargement des données satellitaires est la particularité d'avoir des données mosaïquées selon la zone d'intérêt (shapefile en polygone) définie avec la projection native ou de sélectionner une autre projection prédéfinie. De plus, ces données sont prétraitées et peuvent être directement utilisées. Même si ces données présentent peu de parasites, car les nuages, bruits, aérosols, vapeurs d'eau et ombres ont été supprimés, il reste toujours quelques pixels de nuage et leur ombre. Nous avons supprimé ces nuages résiduels par application des masques créés avec les données de qualité de MODIS et de Landsat 08 OLI.

<sup>17</sup> URL: <https://lpdaacsvc.cr.usgs.gov/appeears/>

Les limites des données offertes sur cette plateforme sont que les données n'existent pas toujours selon les critères prédéfinis ou si elles existent, elles sont parfois aberrantes. En effet, nous avons téléchargé des données couvrant l'ensemble des municipalités régionales de comté de la Montérégie, mais les données de réflectance de surface Landsat 08 ne couvraient qu'une partie de la zone définie. Par contre, les données thermiques MODIS LST de jour (MOD11A1) couvraient la zone d'étude et avaient plus des 2/3 des pixels supprimés. Cet exercice répété plusieurs fois (cinq dates différentes) donnait les mêmes résultats pour la même zone de Montérégie. Nous avons également essayé de télécharger les données d'une zone dont les limites correspondent aux dimensions d'une tuile complète (Path 014, Row 028) précédemment acquise sur la plateforme Earthexplorer<sup>18</sup>. Le constat est le même. Partant de ces constats, il est souhaitable d'acquérir des données brutes et de les prétraiter soi-même avant utilisation.

### 3.3.1. Création des variables de prédiction et leur sélection

Dans cette sous-section, nous parlons de la sélection des variables prédictives, notamment celles basées sur les indices spectraux et les facteurs topographiques pour prédire la température de surface à une fine résolution. Les liens entre la température de surface et les facteurs environnementaux ou topographiques ont été mentionnés dans la littérature scientifique et rapportés dans l'état de l'art (Malbêteau, 2017 ; Merlin et al., 2010 ; Song et al., 2014). Nous rappelons que la température de surface est la somme relative de la température du couvert végétal (canopée) dominant et la température du sol, toutes deux définies en fonction de la couverture végétale fractionnée (Kotchi, 2004). Ainsi, la température de surface mesurée par le capteur infrarouge thermique intègre tous les objets circonscrits dans sa résolution spatiale. Les facteurs environnementaux de surface, décrits par les indices spectraux et facteurs topographiques, servent donc de variables de prédiction de la température de surface.

Les variables de prédiction sélectionnées dépendent, comme mentionné à la sous-section 2.3.1 du Chapitre 2 - , de la couverture de surface de la zone d'étude ainsi que des biotopes des tiques et moustiques. Ce sont des indices calculés avec les bandes de réflectance de surface de Landsat 08/OLI. Ils caractérisent la végétation (NDVI, SAVI, EVI, EVI2, FVC), les surfaces en eau (NDWI), les zones artificialisées (NDBI), les facteurs topographiques (altitude, Pente et Orientation). Seules les variables prédictives utilisées dans cette étude sont décrites ci-dessous.

---

<sup>18</sup> <https://earthexplorer.usgs.gov/>



S'agissant des paramètres topographiques, l'altitude, la pente et l'exposition ont été dérivées du raster d'ASTER GDEM dans l'environnement d'analyse spatiale du SIG ArcMap 10.7. L'exposition donne l'orientation de la pente par rapport au nord dans le sens horaire. Elle varie de 0 à 360°. La valeur 0 indique le Nord.

### 3.3.2. Évaluation de la colinéarité entre les variables

Cette sous-section montre que plusieurs variables prédictives expliquent la LST, mais toutes ne peuvent pas être introduites à la fois dans les modèles de réduction d'échelle pour éviter des biais dans les résultats de LST prédite.

Les méthodes de réduction d'échelle TsHARP et RFR sont des méthodes statistiques basées sur plusieurs variables explicatives de la variable dépendante LST. Toutes les variables explicatives jouent un même rôle malgré leur diversité. Toutefois, des liaisons fortes entre deux ou plusieurs variables explicatives introduites simultanément dans la méthode peuvent créer des biais et exercer une influence importante sur l'efficacité de la méthode. De fait, afin d'éviter des biais dans les résultats, il est indispensable de vérifier la colinéarité entre les variables explicatives et ne retenir que les variables indépendantes pour alimenter l'équation de la méthode.

La mesure de colinéarité entre les variables explicatives a été réalisée avec l'outil « statistiques de groupe de canaux » du sous-module « Multivariés » de l'extension « Analyse spatiale » du SIG ArcMap version 10.7. Le résultat est un tableau de contingence nommé matrice de corrélation de Pearson Bravais. La valeur seuil utilisée dans cette étude est en valeur absolue 0,89. Toutes les variables ayant un coefficient de corrélation « r » supérieure ou égale à |0,89| sont considérées colinéaires entre elles et ne peuvent être utilisées conjointement dans un même modèle. Cette analyse préalable a été faite sur l'ensemble des variables des trois zones d'étude. Les résultats de cette analyse sont présentés dans le Chapitre 4 - .

## **3.4. Description de la mise en application de chaque méthode de réduction d'échelle**

Cette section se décline en 3 sous-sections portant chacune sur les méthodes de réduction d'échelle à appliquer dans cette recherche.

### 3.4.1. Méthode TsHARP

L'implémentation de la méthode de réduction d'échelle TsHARP a été faite sous le logiciel SIG QGIS « QGIS 3.12.2 » intégrant GRASS 7.8.2. avec les fonctionnalités d'analyse spatiale sur les rasters. Les opérations réalisées successivement sont décrites en trois étapes dans la Figure 3.5 ci-dessous. Les activités de l'étape 1

(en gris à gauche dans le diagramme) concernent la préparation des données réalisée avec l'outil d'algèbre spatial « calculatrice raster » de QGIS. Le rééchantillonnage de 100 m à 1000 m a été fait par moyennage spatiale avec la fonction « r.resamp.stats » de la commande GRASS.

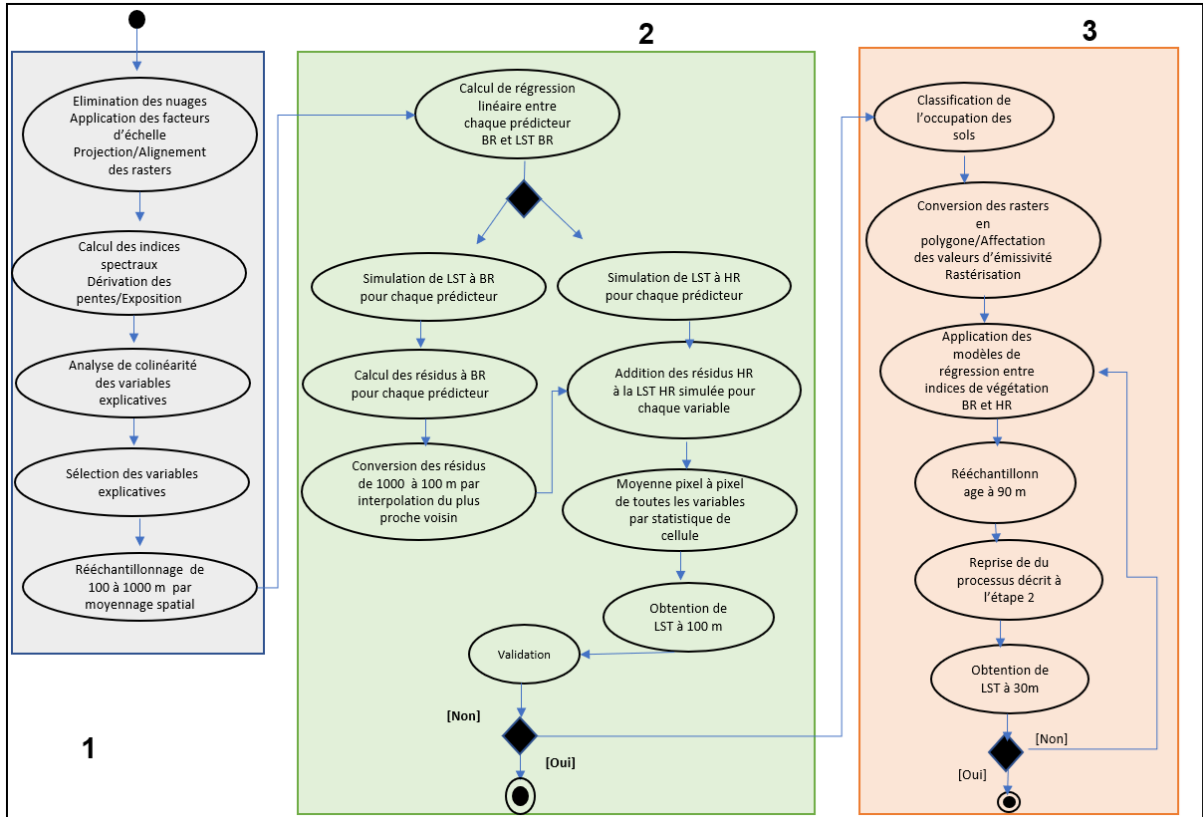


Figure 3.5 : Description des activités réalisées pour la mise en œuvre de l'approche proposée

L'étape 2 (en vert, centre du diagramme) décrit les traitements réalisés jusqu'à l'obtention de la LST à 100 m. Il s'agit notamment de la construction de la régression linéaire entre la MODIS LST à 1000 m et chacune des variables explicatives rééchantillonnées à 1000 m. Les paramètres (a et b) de régression ont servi à la simulation de LST à 1000 m avec chaque variable dans la calculatrice raster suivant la fonction de régression  $y = a + b \cdot x$  ; x étant la variable explicative (NDVI ou NDBI par exemple) et la LST à réduire. Avec ces mêmes coefficients, la LST à 100 m a été simulée. Nous avons par la suite calculé les résidus entre la LST simulée à 1000 m et la LST native à 1000 m par une opération de soustraction dans la calculatrice raster. Les résidus obtenus à 1000 m ont été rééchantillonnés à 100 m par la méthode d'interpolation du plus proche voisin avec la fonction « r.resamp.interp » de la commande GRASS. Enfin les résidus rééchantillonnés à 100 m ont été additionnés aux LST précédemment simulées à 100 m pour une LST à 100 m pour chaque variable. La LST finale unique résulte de la moyenne pixel à pixel de toutes variables explicatives utilisées.

L'étape 3 (en orange, droite du diagramme) conduit à l'obtention de la LST à 30 m en utilisant une seule variable explicative : les émissivités de surface. Les émissivités de surface sont basées sur les unités d'occupation et d'utilisation des sols obtenues par la méthode de classification supervisée selon le maximum de vraisemblance (cf. la sous-section 2.3.2.) La classification des images a été faite sous le SIG ArcMap 10.7. La nomenclature définie compte 6 grandes classes thématiques : les eaux de surface, les zones humides, les zones artificialisées, les zones de végétation, les forêts et les sols nus. Cette nomenclature est liée, telle que précisée dans la description de la méthode plus haut, à la similitude des valeurs d'émissivité des différentes unités de surface. En appliquant les opérations de l'étape 2 et 3 (cf. Figure 2.1) nous obtenons la LST à 30 m. La cartographie des données a été faite sous ArcMap.

### 3.4.2. Méthode RFR

La méthode RFR a été mise en œuvre dans le langage de programmation Python sous l'environnement de développement intégré (IDE) Pycharm édition 2021 grâce à la collaboration avec les étudiants en génie géomatique dans le cadre du projet des génies. En effet, cette collaboration a été nécessaire pour deux raisons. La première raison est pour atteindre l'objectif 4 de cette étude qui vise à la mise en place de l'automatisation de la méthode de réduction d'échelle proposée ; la seconde raison est liée au temps et à nos propres compétences en programmation qui sont en cours d'amélioration. Nous nous sommes initiés à la programmation puis à la programmation orientée objet au cours de la session d'hiver 2020. Malgré les habiletés développées, la mise en place d'un tel projet va demander plus de temps, c'est ce qui justifie la mise à contribution des étudiants de génie géomatique pour ce projet. Le programme développé a été utilisé pour faire nos expérimentations.

La librairie principale utilisée, outre celles de Gdal, Numpy, Pandas, Seaborn, Matplotlib qui permettent la manipulation des rasters, leur visualisation, ou traitements statistiques, est Scikit-Learn. Il s'agit d'une librairie d'apprentissage machine dont fait partie la méthode de forêt aléatoire, « Random Forest ». Le programme développé est semi-automatisé. Par exemple, les données à l'entrée, les variables explicatives à prendre en compte et la résolution spatiale à atteindre doivent être spécifiées par l'utilisateur après avoir mesuré la colinéarité entre les variables explicatives. Dans son fonctionnement, les opérations de prétraitements, de rééchantillonnage et d'entraînement sont toutes automatisées. L'importance des prédicteurs, les indicateurs de validation interne tels l'erreur moyenne absolue EMA, ou l'erreur quadratique moyenne RMSE sont calculés à mi-étape avant l'obtention du résultat final. Le résultat final a été mis en page après la composition cartographique sous l'environnement ArcMap du logiciel SIG ArcGIS 10.7.

### 3.4.3. Méthode PBIM

Le PBIM est une méthode physique qui utilise une seule variable explicative qu'est l'émissivité des unités de surface. L'émissivité de surface est basée sur l'occupation et l'utilisation des sols réalisée précédemment lors de traitement des données avec TsHARP. La mise en œuvre de l'approche PBIM a été testée dans un premier temps sous l'environnement d'ArcMap du logiciel SIG ArcGIS 10.7 avec la méthode de « statistique zonale » du sous-module « Zonaux » de l'extension d'analyse spatiale. Auparavant les rasters ont été rééchantillonnés selon la méthode d'interpolation bilinéaire avec la fonction « Ré-échantillonner » du sous-module « Traitement raster » dans le module « Raster » de l'extension d'analyse spatiale.

Pour la suite des travaux d'expérimentation, nous avons utilisé la méthode automatisée réalisée par les étudiants de génie géomatique. Le programme automatisé développé dans le langage de programmation Python version 3.7 sous l'IDE Pycharm ne prend pas en compte le mode de calcul des émissivités basé sur l'occupation et l'utilisation des sols. Dans ce programme automatisé, le mode de calcul des émissivités des unités de surface est lié aux valeurs seuils de NDVI (cf. Tableau 3.2) développées par Sobrino et ses pairs (Sobrino et al., 2008) et adaptées pour les deux bandes thermiques (B10 et B11) de Landsat 08 TIRS.

Tableau 3.2 : Mode de calcul des émissivités de surface basé sur les seuils de NDVI

Landsat 08 TIRS	Émissivité de surface	NDVI
B10	$0,973 - 0,047B_4$	NDVI < 0,2
	$0,9863P_v + 0,9668(1 - P_v) + C_i$	$0,2 \leq \text{NDVI} \leq 0,5$
	$0,9863 + C_i$	NDVI > 0,5
B11	$0,984 - 0,026B_4$	NDVI < 0,2
	$0,9896P_v + 0,9747(1 - P_v) + C_i$	$0,2 \leq \text{NDVI} \leq 0,5$
	$0,9896 + C_i$	NDVI > 0,5

Source : (Yu et al., 2014)

Où  $B_4$  est la bande rouge de Landsat 08 ;  $P_v$  est la proportion de végétation,  $C_i$  est le coefficient de rugosité de la surface pour la bande  $i$  ( $C_i = 0$  pour les surfaces planes). De manière générale, la valeur affectée à  $C$  est :  $C = 0,005$ . Les équations liées à la Bande 11 sont seulement valides pour les zones hétérogènes. L'équation (Équation 0.8) de calcul de la proportion de végétation ( $P_v$ ) est présentée en annexe B.

## 3.5. Validation des données

L'objectif de cette section est de décrire les indicateurs d'erreurs et les modes de validation des résultats prédictifs de la LST (°C).

### 3.5.1. Les indicateurs d'erreurs

La validation des résultats a été faite à deux niveaux : une validation interne et une validation externe. Pour chacun des niveaux, une évaluation qualitative et quantitative a été faite. L'évaluation qualitative est une analyse visuelle entre deux images, observée et simulée, pour comparer la fidélité dans la distribution spatiale des différentes plages de températures de surface. Tandis que l'évaluation quantitative se base sur les indicateurs statistiques tels l'erreur moyenne absolue -EMA- (Équation 0.4) et l'erreur quadratique moyenne -RMSE- (Équation 0.5), présentées en annexe B. L'erreur quadratique moyenne, EMA, est une mesure de l'erreur quadratique moyenne des prédictions du modèle. Elle calcule la moyenne de la différence carrée entre la LST prédite ( $LST_s$ ) et la LST observée.

L'erreur quadratique moyenne (Root Mean Square Error - RMSE) est une mesure de l'erreur moyenne, pondérée par le carré de l'erreur et indique la magnitude de l'erreur de la prévision<sup>19</sup>. Une valeur faible de cet estimateur signifie une bonne précision des résultats et que la séquence d'erreurs est caractérisée, à la fois, par une faible variance et une erreur moyenne réduite.

Nous rappelons, selon l'état de l'art, qu'il existe d'énormes difficultés en lien avec la validation des résultats de réduction d'échelle. Les difficultés sont liées d'une part, aux effets des incertitudes associées aux données géospatiales multisources avec des résolutions spatiales/temporelles variées et, d'autre part à la non-représentativité spatiale des mesures de terrain servant de référence à la validation des LST (Yalbetau ; Merlin, 2010). En dépit de ces insuffisances, les pratiques et usages n'ont pas changé. Les indicateurs d'erreur moyenne absolue, l'erreur quadratique moyenne ou le coefficient de corrélation sont régulièrement utilisés pour quantifier la qualité des LST réduites. Ainsi, l'écart entre la donnée simulée et une donnée de référence est calculé.

La validation des résultats de réduction d'échelle revient à évaluer leur exactitude par rapport aux données de référence. L'exactitude définie dans le vocabulaire international de métrologie -VIM- comme l'« *étroitesse de l'accord entre une valeur mesurée et une valeur vraie d'un mesurande* », renvoie à la conjonction de la justesse et la fidélité des mesures selon la norme ISO 1994. Selon le VIM, la fidélité est l'« *étroitesse de l'accord entre les indications ou les valeurs mesurées obtenues par des mesurages répétés du même objet ou des objets similaires dans des conditions spécifiées* » que sont par exemple des conditions de répétabilité ou de reproductibilité. La fidélité est caractérisée par l'écart type, la variance ou le coefficient de variation. Quant à la

---

<sup>19</sup> [https://help.salesforce.com/articleView?id=bi\\_edd\\_model\\_metrics\\_cross\\_validation.htm&type=5](https://help.salesforce.com/articleView?id=bi_edd_model_metrics_cross_validation.htm&type=5) consulté le 12 avril 2022

justesse, elle exprime l'« étroitesse de l'accord entre la moyenne d'un nombre infini de valeurs mesurées répétées et une valeur de référence » (VIM, 2012).

Contrairement à la fidélité qui est une grandeur pouvant être exprimée numériquement, l'exactitude et la justesse ne sont pas des grandeurs à caractériser numériquement. Dans cette étude, les concepts d'exactitude, de justesse et de fidélité ont été utilisés pour la validation des résultats. La justesse a été employée pour l'évaluation qualitative des résultats. La fidélité a été utilisée pour vérifier la performance des résultats et donc des méthodes après les trois expérimentations. En cas de faibles erreurs ( $\pm 2$  °C) entre les résultats et les données de référence, on peut qualifier d'exact le résultat.

### 3.5.2. Validation interne et externe

La validation interne a consisté à utiliser le même jeu de données pour estimer la LST pour valider les résultats prédits. Nous avons aléatoirement sélectionné près des points communs aux images originales de MODIS et prédites. Les valeurs de LST respectives ont servi à calculer les indicateurs d'erreurs RMSE et EMA. Pour la validation externe, les résultats obtenus par réduction d'échelle ont été validés avec la température de surface de Landsat 08 OLI/TIRS et la température de l'air près de surface relevée au niveau de stations météorologiques localisées sur le site d'étude. Les images satellitaires thermiques exploitées sont des données horaires fournies tous les 16 jours. La Landsat 08 TIRS LST a été dérivée des bandes spectrales thermiques 10 et 11 selon l'approche décrite dans le guide d'utilisation de Landsat 08 OLI/TIRS (USGS, 2019)<sup>20</sup>. Le processus d'estimation de Landsat 08 LST est résumé dans le diagramme ci-dessous (cf. Figure 3.6).

---

<sup>20</sup> [https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574\\_L8\\_Data\\_Users\\_Handbook-v5.0.pdf](https://prd-wret.s3.us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574_L8_Data_Users_Handbook-v5.0.pdf) consulté le 12 avril 2022

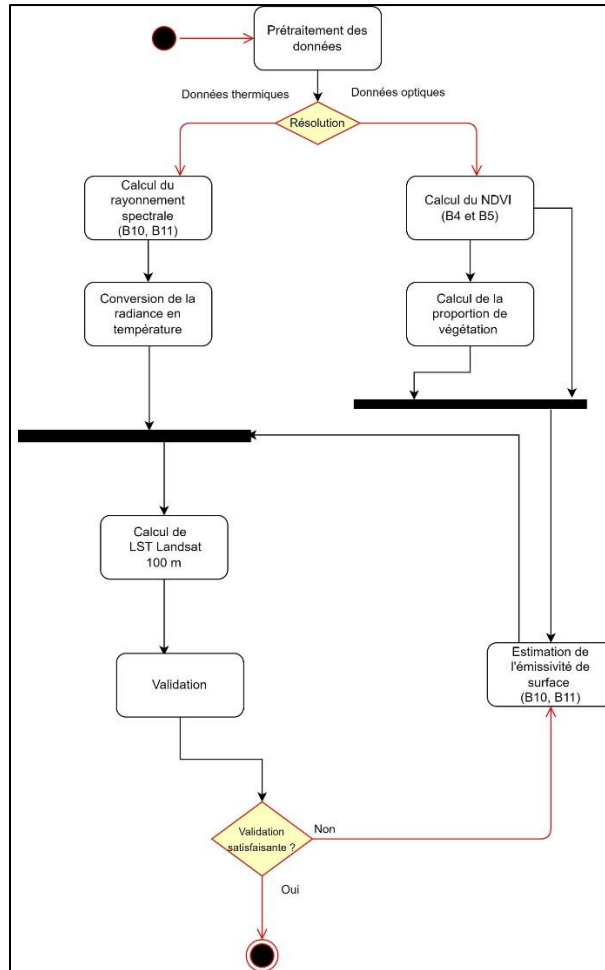


Figure 3.6 : Processus d'estimation de la LST de Landsat 08/TIRS

### Conversion des données Landsat 08 TIRS en LST

Le processus de dérivation de la température de surface se fait en différentes étapes (cf. Figure 3.6) : le calcul du rayonnement spectral ou radiance (Top of Atmospheric TOA) avec l'Équation 0.6, la conversion de la température TOA en luminosité (Équation 0.7), le calcul du NDVI (Équation 0.10), le calcul de la proportion de végétation (PV) (Équation 0.8), le calcul de l'émissivité (Tableau 3.2) et enfin le calcul de la LST selon l'approche standard (Équation 0.9). Les équations sont présentées en annexe B.

La différence entre la Landsat 08 TIRS LST et chaque LST prédite a été calculée et représentée spatialement sur une carte pour identifier les zones et les unités de surface ayant des valeurs au-dessus du seuil de +/-2°C tolérable dans cette étude. Les valeurs de température ont été réparties en 3 classes : une classe des valeurs de LST au-dessus de 2°C, une classe des valeurs de LST comprises entre -2°C et 2°C et enfin, la classe des valeurs de LST en dessous de 2°C.

Les résultats de température de surface LST ont été aussi validés avec la température maximale de l'air près de la surface acquise sur le site de l'environnement Canada<sup>21</sup>. En effet, plusieurs études ont montré la corrélation existante entre ces deux variables (Kawashima et al., 2000 ; Morin et al., 2018a), soit en estimant la température de l'air à partir des températures de surface (Kloog et al., 2014; Morin et al., 2018a; Nieto et al., 2011; Xu et al., 2014) ou en vérifiant la qualité des LST des images satellitaires par comparaison aux températures de l'air près de la surface (Coll et al., 2010; Duan et al., 2019; Ermida et al., 2014; Kotchi et al., 2021; Srivastava et al., 2009). Il résulte de ces études que les valeurs de ces deux variables présentent un faible écart en conditions atmosphériques normales sans stress surtout dans les zones boisées avec une canopée complète. La précision entre ces deux types de températures est de +/- 3 à 5% en moyenne.

Dans cette étude, pour identifier les stations météorologiques implantées en zone de végétation dense, nous avons isolé les valeurs de NDVI se situant au-dessus du seuil 0,75, car la classe de 0,6 à 1 de NDVI représente la végétation dense. Ainsi nous avons retenu dans le réseau de stations météorologiques celles qui sont localisées dans les zones présentant un NDVI élevé. Après cette étape, nous avons utilisé l'outil « Point Sampling Tool » dans l'extension du logiciel SIG QGIS pour extraire les valeurs de température à ces points d'observation de manière simultanée pour toutes les images satellitaires de LST exploitées (MODIS, Landsat et MODIS LST prédites).

### **3.6. Conclusion**

À travers ce chapitre, les étapes nécessaires pour l'application des trois méthodes sont décrites sous forme de diagrammes UML composés des actions structurées. Par la suite, le profil environnemental des trois sites d'étude identifiés dans la province de Québec et les données géospatiales, thermiques, topographiques et de réflectance, à exploiter sont présentés. Les outils comme le QGIS, ArcGIS, Python avec la bibliothèque libre scikit-learn sont retenus pour réaliser les tests. Enfin, le mode de validation des résultats et les indicateurs de qualité des résultats ont également été exposés. Après cette étape vient celle de l'opérationnalisation présentée au Chapitre 4 -

---

<sup>21</sup> [https://climat.meteo.gc.ca/historical\\_data/search\\_historic\\_data\\_f.html](https://climat.meteo.gc.ca/historical_data/search_historic_data_f.html)



# Chapitre 4 - Expérimentations

## 4.1. Introduction

Dans cette section, des expérimentations ont mis en application les méthodes TsHARP, RFR et PBIM pour les 3 zones d'études identifiées (région de la Capitale nationale, région de l'Estrie et région de Montérégie). Premièrement, les résultats obtenus à 100 m sont présentés selon les trois méthodes et pour chacune des zones d'étude. Deuxièmement, les résultats des expérimentations de la réduction d'échelle de 100 m à 30 m selon la méthode TsHARP et enfin, les résultats obtenus de 1000 m à 30 m selon les méthodes de TsHARP et RFR seront décrits et illustrés. Ces différentes expérimentations permettent de valider la nouvelle approche en deux étapes pour réduire la MODIS LST de 1000 m à 30 m en passant par une étape intermédiaire de 100 m.

## 4.2. Réduction d'échelle de MODIS LST de 1000 m à 100 m

Cette section subdivisée en trois sous-sections présente les résultats des expérimentations et leurs validations réalisés respectivement en régions de la Capitale nationale, de l'Estrie et de Montérégie.

### 4.2.1. Expérimentation 1 : Région de la Capitale nationale

#### 4.2.1.1. Résultats

Les résultats des traitements sont présentés sous forme de cartes ci-dessous. Leur interprétation est basée sur des évaluations qualitative et quantitative. Il apparaît important de montrer en premier les résultats de l'analyse de colinéarité, car c'est à partir des résultats de ce tableau que les variables prédictives alimentant les modèles sont sélectionnées. Cette analyse de colinéarité entre les variables a été faite pour les trois expérimentations. Le Tableau 4.1 ci-dessous est la matrice de corrélation linéaire univariée entre la MODIS LST et les variables explicatives de la région de la Capitale nationale.

Tableau 4.1 : Région de la Capitale nationale – analyse de corrélation linéaire univariée entre les variables

Variables	LST	Orientation	Pente	PV	NDVI	EVI2	SAVI	NDBI	NDWI
LST	1								
Orientation	0,06231	1							
Pente	-0,27903	0,1586	1						
PV	-0,29682	0,03318	0,06891	1					
NDVI	-0,30263	0,0398	0,06068	0,99446	1				
EVI2	0,11069	0,32834	0,28939	0,13499	0,1283	1			
SAVI	0,03915	0,367	0,33162	0,16744	0,16298	0,97809	1		
NDBI	0,3269	-0,01524	-0,10093	-0,85518	-0,85147	-0,03636	-0,08635	1	
NDWI	0,19418	-0,01633	-0,0379	-0,86456	-0,82603	-0,19152	-0,19169	0,64649	1

Le Tableau 4.1 montre une colinéarité entre les indices de végétation PV et le NDVI d'une part et d'autre part entre le SAVI et l'EVI 2, car, les valeurs de leur coefficient de corrélation « r » sont supérieures à la valeur seuil de |0,89| (cf. valeur en rouge dans le Tableau 4.1). Cette colinéarité présume l'intégration d'une seule des variables colinéaires dans le modèle de régression pour éviter des biais. Ainsi, l'indice NDVI a été préféré à l'indice PV dont le calcul se base sur les valeurs de NDVI. Par contre, aucun des indices SAVI et EVI2 n'a été retenu. Il s'ensuit que les variables retenues pour cette expérimentation sont des variables indépendantes entre elles et permettent de caractériser les grandes unités d'occupation du sol: NDVI, NDBI, NDWI, Pente. Le NDVI renvoie à la zone de végétation, le NDBI permet de caractériser les zones urbanisées, les zones humides par le NDWI et enfin, la pente décrit la topographie. Bien qu'il n'existe pas de colinéarité entre les variables décrivant la topographie, nous avons préféré la pente à l'orientation parce que la pente est l'indicateur fréquemment utilisé dans les modèles.

L'image 1 de la Figure 4.1 correspond à l'image brute la température de surface de MODIS LST (°C) à 1000 m tandis que les images 2, 3 et 4 de la Figure 4.1 représentent les images de MODIS LST prédites à 100 m respectivement avec les méthodes TsHARP, RFR et PBIM. L'ensemble des images sont représentées spatialement en différentes classes d'amplitude de 3. Sur l'image réduite avec TsHARP (image 2 de la Figure 4.1), les valeurs de LST varient de 10,76°C à 37,35°C, et de 14,98°C à 40,20°C avec celle obtenue avec RFR (Figure 4.1. Image 3). Les valeurs de LST pour la méthode PBIM vont de 0,12°C à 40°C. Les températures les plus élevées sont observées dans les milieux urbains et les faibles températures dans les zones boisées où abonde le couvert végétal.

Une comparaison visuelle des images 2, 3, et 4 de la Figure 4.1 représentant les LST (°C) obtenues par réduction d'échelle à 100 m respectivement selon les méthodes TsHARP, RFR, PBIM, montre que les distributions spatiales générales des LST correspondant aux températures MODIS à 1000 m (image 1 de Figure 4.1) sont similaires. Sur les images 2, 3 et 4 de la Figure 4.1, nous observons des formes, textures ou patrons des objets d'occupation plus prononcés et rugueux par rapport à ce qui est observé sur l'image 1 de la même figure. De même, sur ces images, certaines classes de température telle la classe [35-40,9], ont une étendue spatiale plus large que ce qui est observé sur l'image 1 de la Figure 4.1, (Cf. les ellipses en couleur rouge sur les images. Cette observation est plus explicite sur les images 1,2 et 3 de la Figure 4.2 et aussi sur la Figure 0.1 à la Figure 0.5 en annexe A). Ceci signifie que le nombre de pixels dans cette plage de température a augmenté. Cette augmentation est encore plus importante sur les images 3 et 4 par rapport à l'image 2 (Figure 4.1 et Figure 4.2).

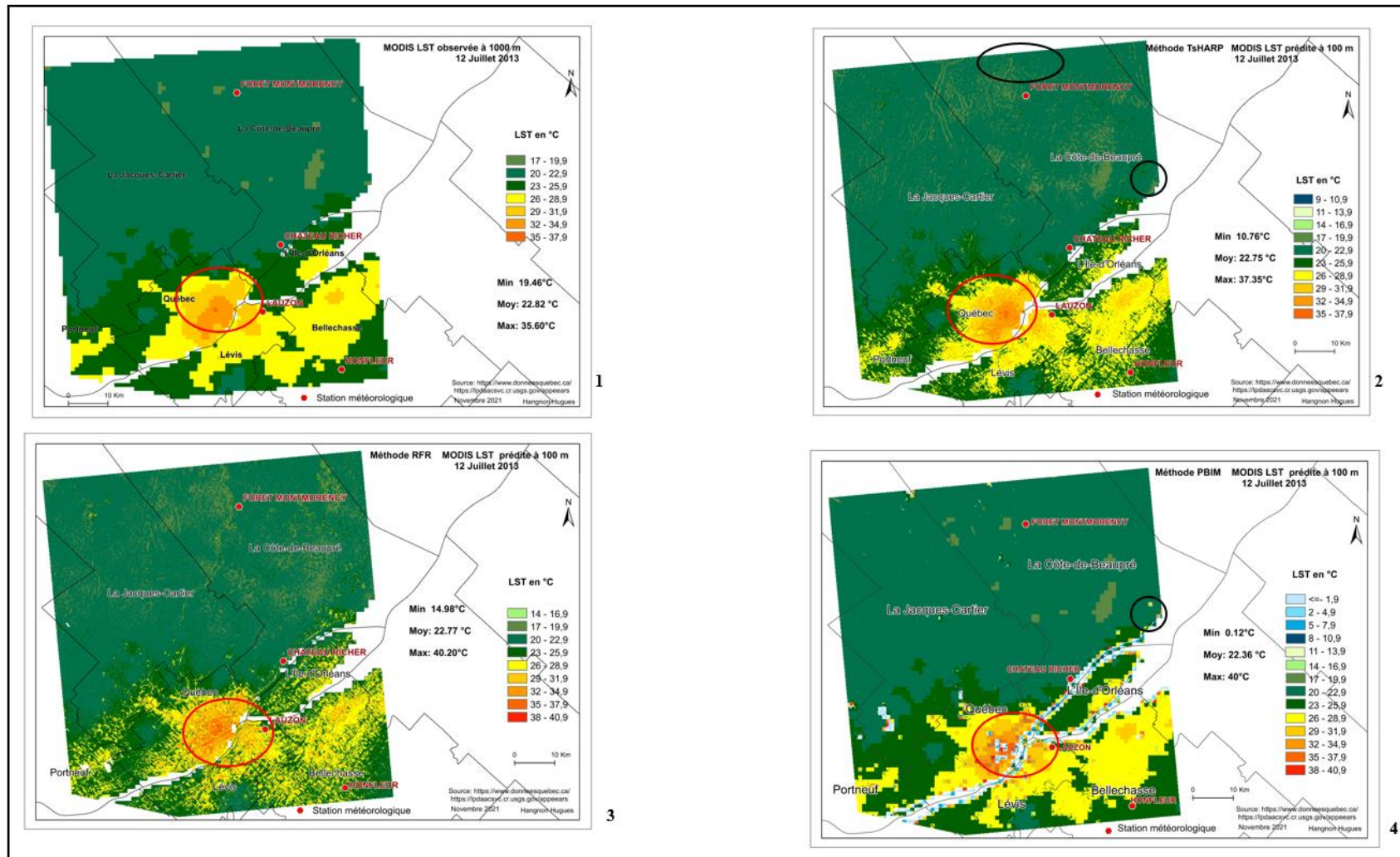


Figure 4.1 : Région de la Capitale nationale : résultats de réduction d'échelle à 100 m suivant différentes méthodes

1 : MODIS LST observée à 1000 m ; 2 : MODIS LST réduite à une résolution de 100 m par la méthode TsHARP ; 3 : MODIS LST réduite à une résolution de 100 m par la méthode RFR ; 4 : MODIS LST réduite à une résolution de 100 m avec PBIM

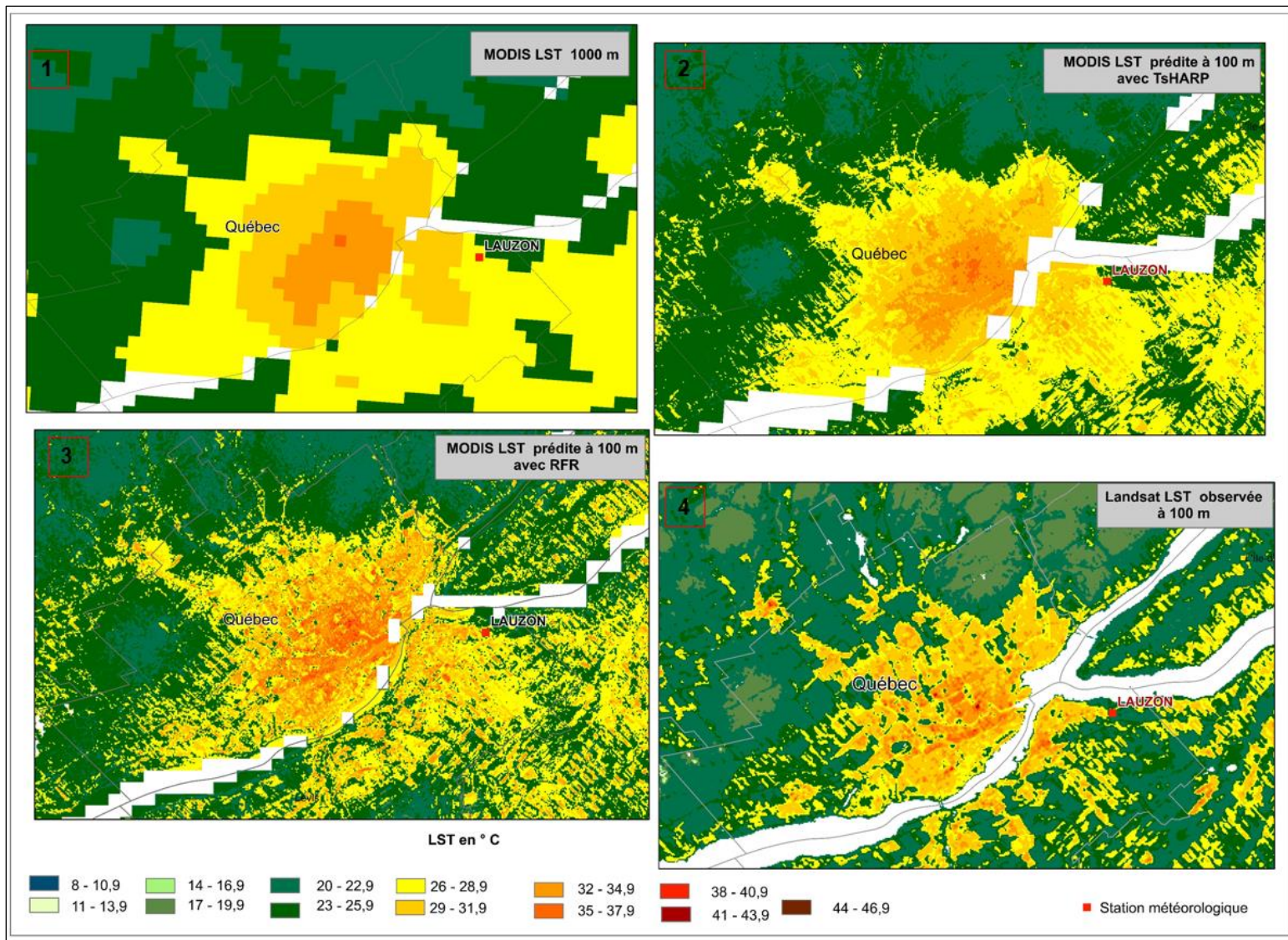


Figure 4.2 : Région de la Capitale nationale : zoom sur la représentativité spatiale des valeurs de LST en milieu urbain hétérogène

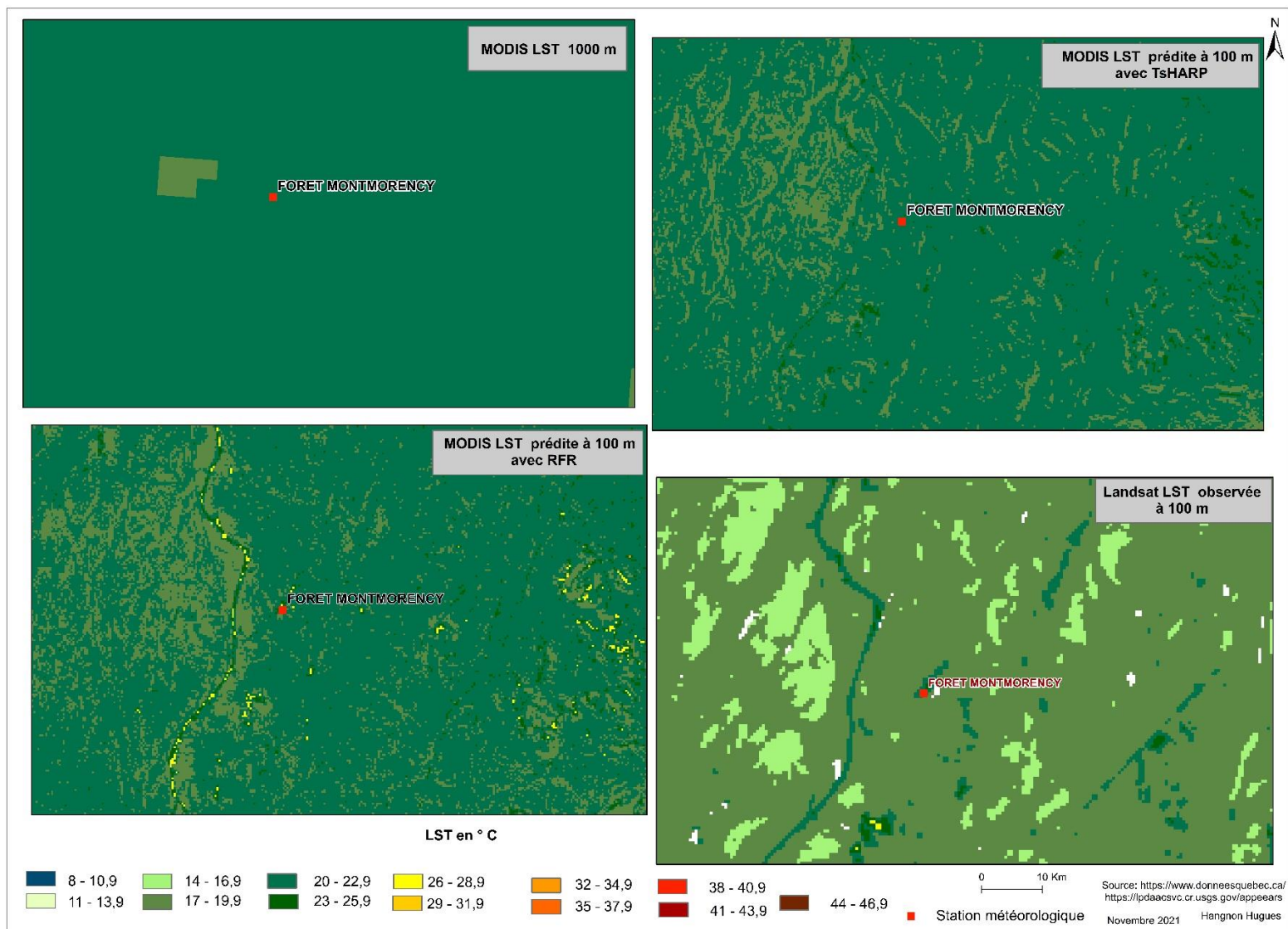


Figure 4.3 : Région de la Capitale nationale : zoom sur la représentativité spatiale des valeurs LST en zone homogène boisée

L'analyse visuelle montre également l'apparition de nouvelles classes de température sur les images 2, 3 et 4 par rapport à l'image 1 (cf. Figure 4.1). En effet, pour l'image 2, les valeurs des nouvelles classes vont de 9 à 16,9 °C. Ces classes de température sont quasi inexistantes sur l'image 1 à cause du nombre insignifiant (0.002% sur le total des pixels) de pixels qui s'y retrouve. Un zoom plus grand montre que ces pixels se retrouvent à proximité du plan d'eau (ellipse de couleur noire dans le MRC « La Côte-de-Beaupré ») ou dans la zone proche de la forêt Montmorency. Les classes [9-10,9] et [11-13,9] comptent chacune 1 pixel. Cependant, sur les images 3 et 4, les pixels des classes de températures de 38 à 40,9°C en faible proportion (0.011%), sont apparus dans la zone artificialisée et urbanisée de la ville de Québec. Afin de constater les effets de la réduction d'échelle sur la qualité des résultats, nous avons fait un zoom dans certaines unités d'occupation du sol.

Un zoom des résultats a été fait sur la zone urbaine de la ville de Québec et la spatialisation des valeurs de LST (°C) selon les différentes méthodes a été présentée sur Figure 4.2. Sur cette Figure 4.2, nous avons de gauche vers la droite en haut, l'image de MODIS LST à 1000 m suivies de l'image de MODIS LST à 100 m prédite avec TsHARP ; et en bas, l'image de MODIS LST à 100 m avec RFR. La 4ième image est celle de l'image de Landsat LST à 100 m, qui sert de référence pour la validation des images prédites. À partir de la légende commune à ces 4 images, on constate sur la Figure 4.2 les mêmes classes de températures entre la MODIS LST brute à 100 m et la MODIS LST prédite avec TsHARP à 100 m, ainsi que sur la carte de LST prédite avec RFR, où apparait la classe [38-40,9]. En revanche, sur la carte de référence Landsat LST, les classes de température atteignent 46,9 °C avec plus de détails spatiaux. Malgré ces différences sur les différentes images, plus perceptibles en zone urbaine avec les points chauds sur la dernière image, il y a une grande similitude sur les résultats prédits et l'image de référence Landsat.

La même opération a été faite en zone boisée (Figure 4.3). Sur le panorama de cartes de cette Figure 4.3, on observe l'apparition de trois tons de couleurs éparses sur toute la zone des cartes de LST à 100 m obtenues par TsHARP et RFR contre deux tons de couleurs sur la carte de MODIS LST à 1000 m. Ainsi, nous constatons sur la carte de MODIS LST à 1000 m uniquement que les classes de température varient de 17 à 22,9 °C alors que sur la carte de LST à 100 m obtenue par TsHARP, les classes de température s'étendent de 17°C jusqu'à 25,9°C. En revanche sur la carte de LST 100 m obtenue avec RFR, les classes de températures vont de 17°C à 28,9°C.

L'analyse statistique donne une EMA de 0,55 °C et une RMSE de 0,78 °C pour la LST prédite avec TsHARP, contre respectivement 0,94 et 1,44 pour la LST réduite avec le RFR (cf. Tableau 4.2). Les valeurs de ces indicateurs statistiques étant en dessous de 2°C pour TsHARP et RFR, le seuil de qualité fixé dans cette étude, amènent à conclure sur la justesse des résultats obtenus.

Tableau 4.2 : Indicateurs statistiques calculés entre MODIS LST observée et MODIS LST simulée à 100 m

Québec	TsHARP	RFR	PBIM
RMSE (°C)	0,78	1,44	3.30
EMA (°C)	0,55	0,94	0.76

#### 4.2.1.2. Validation des résultats

Pour la validation des résultats, nous les avons comparés avec la LST estimée à partir de l'imagerie thermique Landsat 08 TIRS (cf. Figure 4.2 et Figure 4.3, en bas à droite). Les écarts entre la Landsat LST et les MODIS LST réduites à 100 m sont représentés spatialement (cf. images 1, 2 et 3 de la Figure 4.4) afin de visualiser les unités de surface ou les zones dans lesquelles apparaissent ces écarts.

L'imagerie Landsat 08 produite le 12 juillet 2013 à 15h 33 a été acquise sur la plateforme <https://earthexplorer.usgs.gov>. L'erreur quadratique moyenne -RMSE- et l'erreur moyenne absolue -EMA- entre le MODIS LST à échelle réduite et la Landsat LST ont été calculées pour évaluer la qualité des résultats (cf. Tableau 4.3). Les images 1, 2 et 3 (cf. Figure 4.4) représentent respectivement la spatialisation des écarts de LST entre Landsat et MODIS prédite à 100 m avec RFR, TsHARP et PBIM.

Les RMSE sont inférieures à 3°C (Tableau 4.3), soit environ 1°C de plus que le seuil de 2°C souhaité par le projet. Ceci peut s'expliquer par la prise en compte de divers facteurs, dont la qualité des données sources, les erreurs liées aux modèles et l'intercalibration des données de sources diverses.

Par ailleurs, les valeurs de température prédites ont été également comparées à celles de température de l'air mesurée à des stations météorologiques localisées dans des zones de couverture végétale totale ou dominante. À cet effet, trois stations météorologiques ont été identifiées : Lauzon, Forêt Montmorency et Château-Richer. La valeur maximale de la température de l'air disponible, au pas de temps journalier pour ces stations, a été retenue comme valeur de référence, car la MODIS LST réduite dans cette étude est une image de jour. Pour ces mêmes stations, nous présentons les valeurs de MODIS LST native (MOD\_LST\_1000), MODIS LST simulée par TsHARP (MOD\_P\_TSHP) et par RFR (MOD\_P\_RFR) et Landsat 08 LST(LST\_LAND) dans le Tableau 4.4 afin de percevoir les écarts de température.

Tableau 4.3 : indicateurs statistiques calculés entre Landsat LST observée et MODIS LST simulée à 100 m

	TsHARP	RFR	PBIM
RMSE (°C)	2,57	2,82	3,39
EMA (°C)	2,3	2,46	2,66

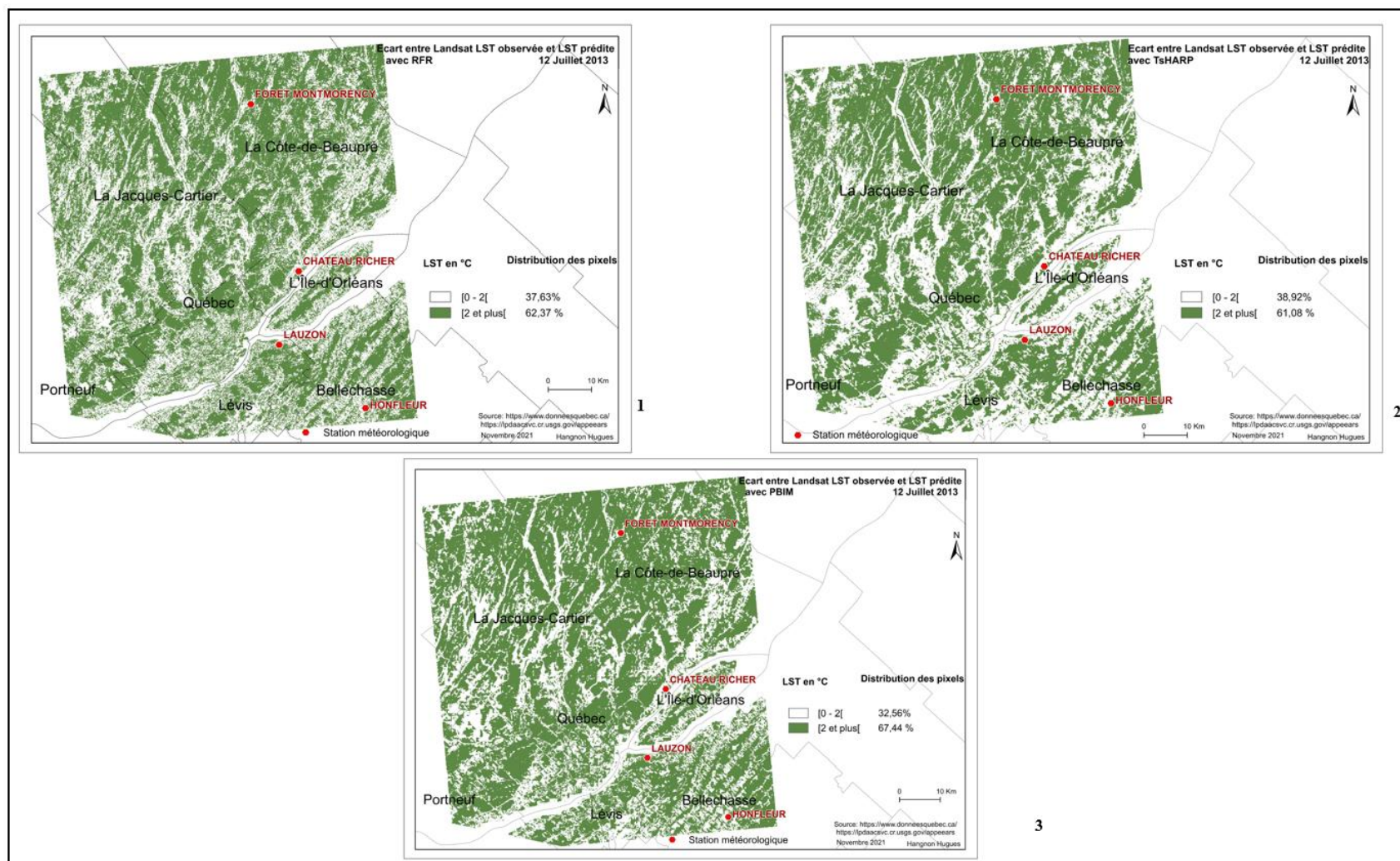


Figure 4.4 : Représentation spatiale des écarts de Landsat LST observée et MODIS LST prédite selon 3 méthodes à 100 m

1 : Écart de LST entre Landsat LST et MODIS LST prédite avec RFR

2 : Écart de LST Landsat et MODIS prédite avec TsHARP

3 : Écart de LST Landsat et MODIS prédite avec PBIM



Tableau 4.4 : Valeurs de température selon différentes sources à la localisation de la station météo

Nom de station Météo	MOD_P_RFR	MOD_P_TSHP	MOD_P_PBIM	MOD_LST_1000	LST_LAND	TA
Foret Montmorency	20,12	21,63	20,54	20,35	20,74	23,9
Lauzon	24,15	25,74	25,01	27,87	21,51	26
Château-Richer	20,46	21,33	22,09	22,09	22,25	25,5

P : Prédite ; LAND : Landsat ; TA : température de l'air ; RFR : Random Forest Regression, TSHP : TsHARP, LST : Land surface temperature, MOD : MODIS, PBIM : Pixel block intensity modulation

La lecture du Tableau 4.4 montre une variation des valeurs de température dans le même ordre de grandeur entre les MODIS LST (°C) prédites avec Landsat LST (°C) hormis, pour les stations de Lauzon où les écarts sont au-dessus de 2 °C. En revanche, avec la température de l'air, les écarts sont plus élevés surtout dans les stations de Forêt de Montmorency et Château-Richer. Dans une première étape, nous présentons dans le Tableau 4.5 les écarts de température entre les LST de référence (Landsat 08 LST et la température maximale de l'air) et MODIS LST 1000 pour apprécier les marges d'erreur entre elles. Puis, dans une seconde étape, les écarts entre les résultats obtenus et les données de référence sont présentés dans le Tableau 4.6.

Il ressort du Tableau 4.5 qu'en valeur absolue, les écarts les plus élevés sont entre Landsat 08 LST et la température de l'air. Les faibles écarts sont entre MODIS LST et Landsat 08 LST. Cette perception diffère si l'on considère les écarts par station. La station de Lauzon présente des écarts dont deux en valeur absolue sont au-dessus de la forte moyenne.

Tableau 4.5 : Écart de température entre les données brutes

Nom de station Météo	E-MOD-LAND	E-MOD-TA	E-LAND-TA
Foret Montmorency	-0,39	-3,55	-3,16
Lauzon	6,36	1,87	-4,49
Château-Richer	-0,16	-3,41	-3,25
Moyenne	1,94	-1,70	-3,63

E : écart ; MOD : MODIS ; LAND : Landsat ; TA : température de l'air

Ces écarts peuvent être liés aux caractéristiques intrinsèques des capteurs (calibration), ou des données elles-mêmes et les conditions d'acquisition des températures. En effet, la MODIS LST native observée à partir de laquelle des prédictions sont faites est une valeur moyenne de jour. C'est le même cas pour les données de stations météorologiques. Or, la Landsat 08 LST observée à 15h 33' est une valeur horaire. Ces détails peuvent éventuellement expliquer les différences de valeurs entre les LST.

Dans le Tableau 4.6, les écarts au-dessus de 3 °C en valeur absolue sont mis en évidence (couleur rouge). Les écarts, hormis ceux au-dessus de 3°C, sont faibles et tiennent dans des marges acceptables de 2°C si nous

considérons que les données alimentant les méthodes présentaient des erreurs. Nous remarquons également qu'il y a plus de marge d'erreur avec les températures de l'air que Landsat 08 LST. Cet écart entre la température de l'air et la température de surface peut s'expliquer par une grande hétérogénéité dans les zones de végétation. En conséquence, les données de référence Landsat 08 LST valident mieux les résultats obtenus dans la RCN que les données de référence au sol pour cette expérimentation. La visualisation de ces écarts sur des cartes va nous permettre d'identifier les zones et les unités d'occupation qui en sont à l'origine.

Tableau 4.6 : Écart des valeurs de température entre les données simulées et données originales

Nom de station Météo	E- RFR-LAND	E-RFR-TA	E-TSHP-LAND	E- TSHP-TA	E-PBIM-LAND	E-PBIM-TA
Foret Montmorency	-0,62	<b>-3,78</b>	0,89	-2,27	-0,2	<b>-3,36</b>
Lauzon	2,64	-1,85	<b>4,23</b>	-0,26	<b>3,5</b>	-0,99
Château-Richer	-1,79	<b>-5,04</b>	-0,92	<b>-4,17</b>	-0,16	<b>-3,41</b>
Moyenne	0,08	<b>-3,56</b>	1,40	-2,23	1,05	-2,59

E : écart ; LAND : Landsat ; TA : température de l'air ; RFR : Random Forest Regression, TSHP : TsHARP,

De manière générale, la lecture des images 1, 2 et 3 de la Figure 4.4 montre une variation des valeurs des écarts de température de surface en deux classes de valeurs absolues ( $[0 - 2[$  et  $[2$  et plus  $]$ ), selon les méthodes RFR, TsHARP et PBIM. Les écarts de température de surface de la classe  $[0 - 2[$  représentent des pixels, dont les valeurs de température de surface de MODIS 1000 m ont été intégralement simulées. Ces pixels dont les proportions avoisinent 38%, 39% et 33% respectivement pour les méthodes RFR, TsHARP et PBIM (cf. Figure 4.4 et Figure 0.6 en annexe A) sont spatialement répartis, mais sont concentrés au sud et bien plus en zone urbaine qu'en zone de végétation. En revanche les pixels de la classe  $[2$  et plus $]$  sont les valeurs de températures de surface de MODIS 1000 m partiellement simulées. Ces pixels (en vert sur les figures) apparaissent principalement sur plus de la moitié de la zone et plus, vers le nord où dominent les zones de végétation. Les pixels de classes  $[2$  et plus $]$  dans une proportion de 62%, 61% et 67% respectivement pour RFR, TsHARP et PBIM (cf. Figure 4.4 et les diagrammes de distribution des pixels de la Figure 0.6 en annexe A) indiquent bien qu'ils appartiennent à des zones boisées et normalement homogènes, il y a dans l'image de MODIS LST native des pixels hétérogènes dans lesquels dominent des unités artificialisées (route, bâti,...) dans cette même zone. Dans ce contexte, on devrait avoir une fragmentation des valeurs de température associée à chaque unité d'occupation. En conséquence, la variabilité spatiale de température au niveau subpixels n'a pas été prise en compte par les deux méthodes de réduction d'échelle.

## 4.2.2. Expérimentation 2 : Région d'Estrie

Les résultats du deuxième test réalisé dans la région de l'Estrie et leur validation sont présentés dans cette sous-section.

### 4.2.2.1. Résultats

À l'issue de l'analyse de colinéarité entre les variables explicatives (cf. Tableau 4.7 où les fortes liaisons sont mentionnées en rouge) on constate une multicolinéarité entre les couples d'indices (NDVI, NDWI), (NDVI, PV), (NDWI, PV) et (SAVI, EVI2). De ces indices, ceux de NDVI et SAVI ont été retenus. Le choix de NDVI, indice le plus utilisé dans les études de calcul de LST, amène à délaissier à la fois les indices NDWI et PV. Aux indices NDVI et SAVI, l'altitude, le NDBI et l'Orientation ont été ajoutés comme variables explicatives de simulation pour les deux méthodes. Les résultats obtenus sont présentés en classes de température et mis en page (cf. les images 1, 2, 3 et 4 de la Figure 4.5) sous le logiciel de SIG ArcGIS. Les amplitudes des classes sont égales à 3. Le même ton de couleur est associé à chacune des classes identiques sur les cartes. Des agrandissements sur deux sous-ensembles de zones ont été faits pour mieux apprécier les résultats. Il s'agit d'une zone urbaine et une zone boisée.

Tableau 4.7 : Région de l'Estrie : analyse de corrélation linéaire univariée entre les variables auxiliaires

Variables	NDVI	NDBI	NDWI	Altitude	Orientation	Pente	PV	SAVI	EVI2
NDVI	1								
NDBI	-0,57021	1							
NDWI	<b>-0,96982</b>	0,42119	1						
Altitude	0,22756	-0,24254	-0,21185	1					
Orientation	0,11455	0,02491	-0,1372	0,01908	1				
Pente	0,15022	-0,01429	-0,16781	0,13008	0,0558	1			
PV	<b>0,99455</b>	-0,62041	<b>-0,9557</b>	0,243	0,10674	0,15019	1		
SAVI	0,8652	-0,43914	-0,86282	0,10758	0,08263	0,15904	0,87253	1	
EVI2	0,83035	-0,44638	-0,82282	0,09999	0,06962	0,15608	0,84248	<b>0,996</b>	1

Les images 1, 2, 3 et 4 de la Figure 4.5 représentent distinctement dans le même ordre, les températures de surface de MODIS de résolution 1000 m observées, les températures de surface réduites à 100 m selon les méthodes de TsHARP, RFR et PBIM. L'évaluation statistique avec l'erreur quadratique moyenne et l'erreur moyenne absolue (cf. Tableau 4.8) donne des valeurs en dessous de 1°C pour les résultats selon les trois méthodes. Ces valeurs pour TsHARP et RFR sont en deçà du seuil de 2°C souhaité. Ceci peut signifier que les résultats sont satisfaisants sur l'ensemble des pixels des deux images obtenues avec TSHARP et RFR.

Tableau 4.8 : Indicateurs statistiques entre MODIS LST observée et MODIS LST simulée à 100 m

ESTRIE	TsHARP	RFR	PBIM
RMSE (°C)	0.46	0.88	2.33
EMA (°C)	0.31	0.57	0.51

Ainsi, cette précision dans les résultats de LST s'observe de manière générale sur les images 2, 3 et 4 (Figure 4.5) où les distributions des classes de température sont fidèles à celles observées sur l'image 1 de la même figure **Erreur ! Source du renvoi introuvable.** La tendance générale observée est une évolution graduelle croissante des températures des zones d'abondante végétation vers les zones peu pourvues en végétation ou artificialisées. Les textures, formes et patrons plus prononcés sur les images réduites (images 2, 3 et 4 de la Figure 4.5) permettent de différencier et d'identifier plus de détails dans les unités d'occupation de surface de la zone d'étude.

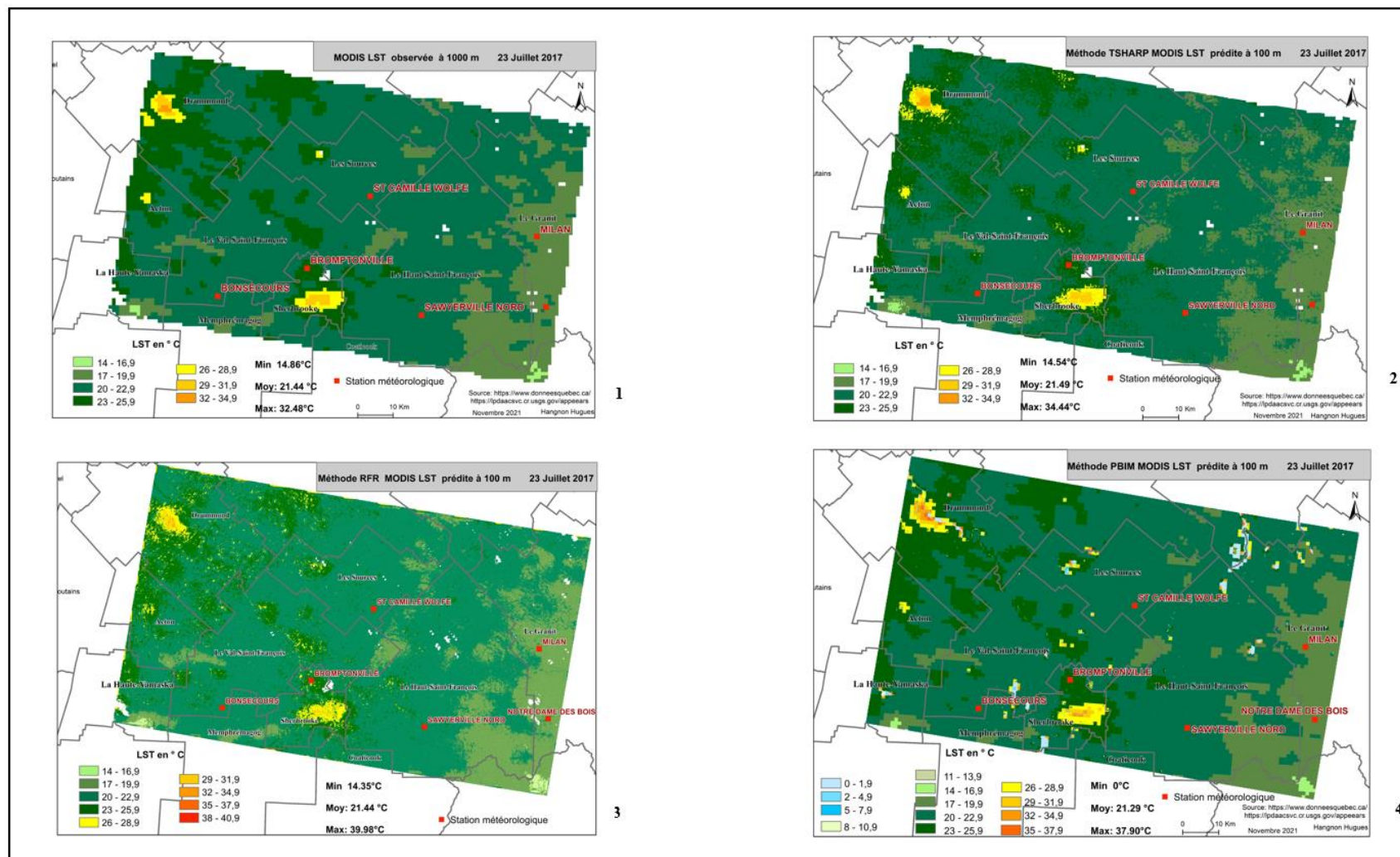


Figure 4.5 : Région de l'Estrie : résultats de réduction d'échelle à 100 m suivant différentes méthodes

1 : MODIS LST observée à 1000 m ; 2 : MODIS LST réduite à 100 m par la méthode TsHARP ; 3 : MODIS LST réduite à 100 m par la méthode RFR ; 4 : MODIS LST réduite à 100 m par la méthode PBIM

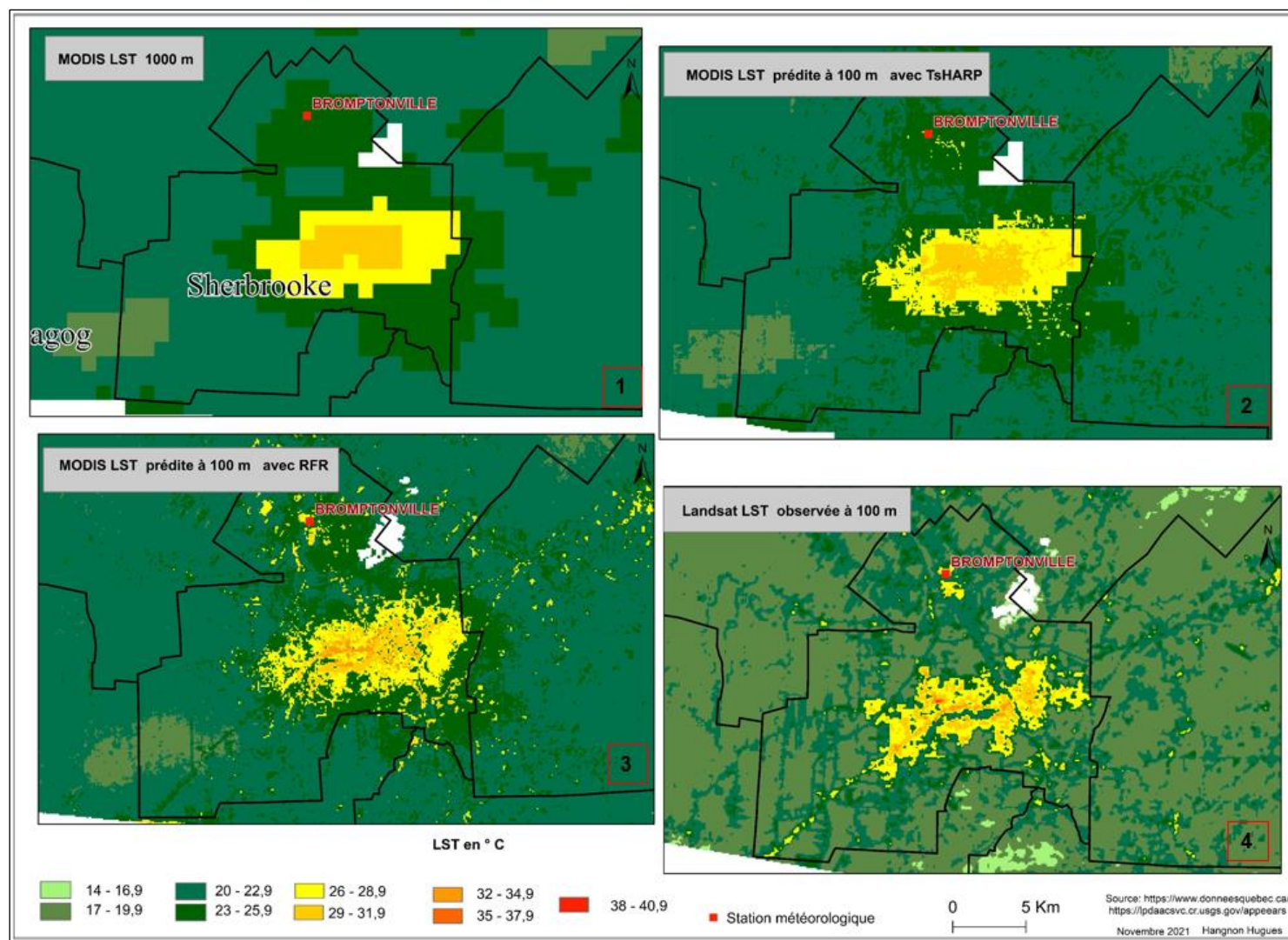


Figure 4.6: Région de l'Estrie : zoom sur la représentativité spatiale des valeurs de LST en milieu urbain hétérogène

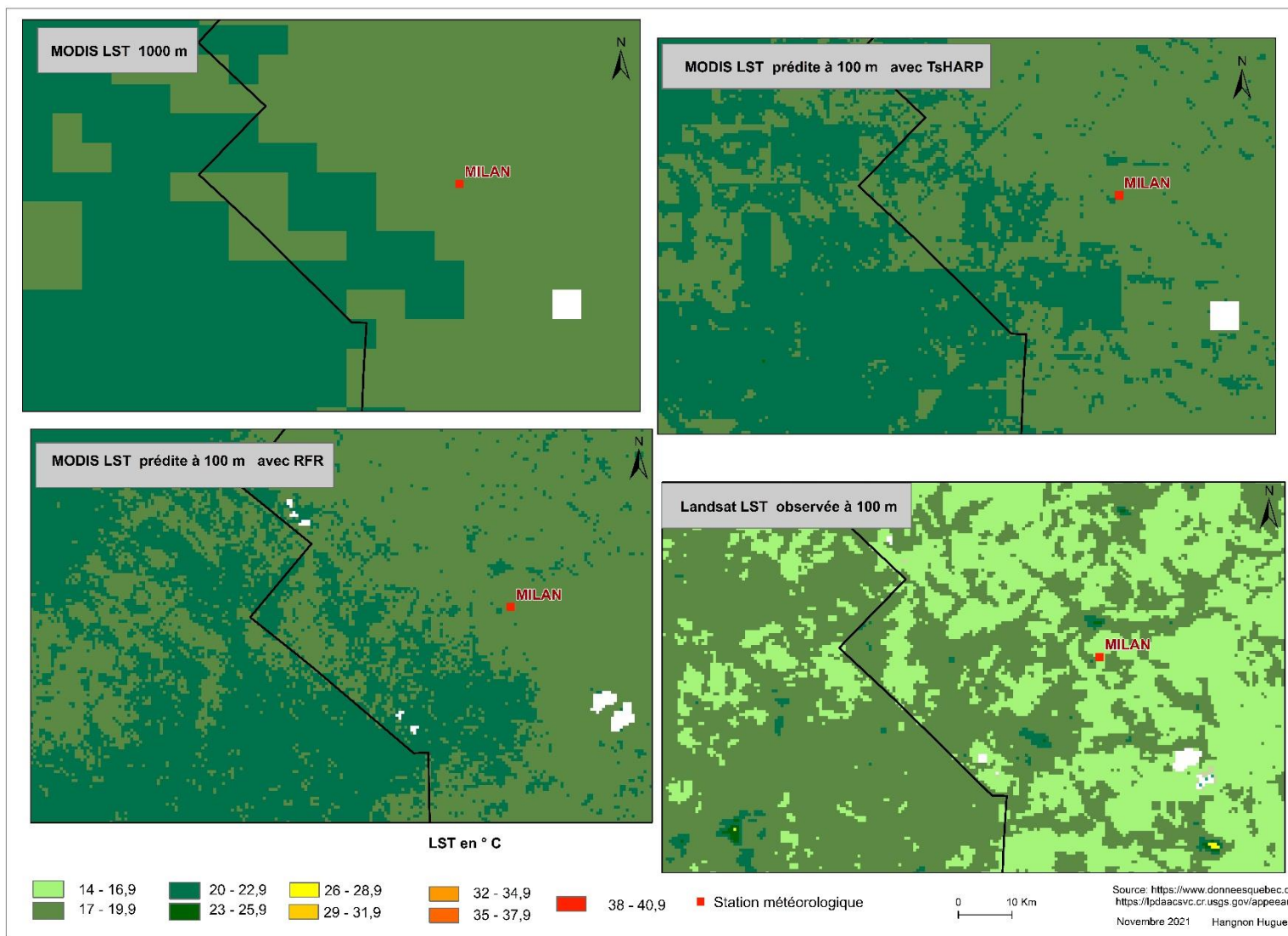


Figure 4.7 : Région de l'Estrie : zoom sur la représentativité spatiale des valeurs de LST en zone homogène boisée

Les classes des valeurs de température de [14 -34,9] sont identiques sur les images 1, 2, et 3 (Figure 4.5). Cependant, les classes de valeurs de température de [35 – 40,9], inexistantes sur les 2 premières images apparaissent à l'image 3 de la Figure 4.5. Cette classe bien sur l'image 3 de la Figure 4.6. Toutefois, le nombre de pixels appartenant à cette classe de température est très minime, voire inexistant (soit 0.0001% par rapport à la totalité des pixels). Cette nouvelle gamme de température observée en zone artificialisée est certainement liée au niveau de détails d'observation et de résolution spatiale plus fine. En fait, en résolution spatiale de 1000 m, les informations sur les objets dans des zones de superficie en dessous de 1000 m étaient généralisées et n'apparaissent pas dans les pixels de 1000 m. Aussi, dans ce contexte où un pixel de 1000 m est composé de plusieurs unités de surface, seule la valeur moyenne de température de surface de l'ensemble des unités de surface est affectée au pixel de 1000 m. Sur l'image 4 de la Figure 4.5, les valeurs de températures varient de 0°C à 37,9°C. Les pixels appartenant à l'intervalle de 0°C à 13,9°C apparaissant aux abords des plans d'eau sont en petit nombre et peuvent être des pixels de bruit non entièrement éliminés. Hormis ces pixels, les classes de valeurs de température sont similaires à l'image 3 de la même figure. Une analyse comparée sur la zone hétérogène urbaine de Sherbrooke montre une harmonie au sein des mêmes gammes de température allant de [14 – 34,9] sur les trois premières images de la figure 4.6. Tandis que, sur l'image 4 (Figure 4.6) de Landsat LST observée, on y voit les classes de température au-dessus de 34,9 °C pour le même milieu à cause des détails spatiaux qu'apportent la résolution spatiale à 100 m. Sur cette même image 4 de Landsat LST observée, la classe de température de [14-16,9] occupe plus d'étendue et englobe même les zones où ces valeurs de température ont été remarquées sur les 3 premières cartes.

Ce même constat est fait lorsqu'on examine les classes de température dans la zone boisée (Figure 4.7), où prédominent les LST allant de 14 à 22,9 °C regroupées en trois classes. Tandis que sur l'image de Landsat LST, on perçoit sept classes de température variant de 8 à 28,9 °C.

Les résultats obtenus montrent effectivement la capacité des méthodes appliquées à réduire les températures de surface de 1000 m à 100 m en associant différents indices de végétation, d'urbanisation et facteurs topographiques dans des zones hétérogènes qu'homogènes. La réduction d'échelle à 100 m a permis d'avoir plus de détails sur les unités de surface avec parfois de nouvelles classes de température par rapport à la MODIS LST brute à 1000 m. Les valeurs de RMSE inférieures à 1°C entre les paires d'images réduites et originales peuvent attester de la justesse des résultats et du peu de biais sur l'ensemble de la zone d'étude. Il reste à valider ces résultats avec des données externes.



#### 4.2.2.2. Validation des résultats

Pour la validation par référence à l'image thermique de Landsat 08, une donnée horaire acquise le 23 juillet 2017 à 15h 31'59", l'écart entre Landsat 08 LST observée et MODIS LST prédite à 100 m avec RFR, TsHARP et PBIM a été calculé. Le résultat de cet écart est représenté spatialement pour chacune des méthodes : RFR, TsHARP et PBIM (cf. les images 1, 2, et 3 de la Figure 4.8). Une évaluation visuelle des trois figures montre que les valeurs des écarts de température supérieures à 2°C sont observées dans leur grande majorité sur toute l'étendue de la zone d'étude et surtout en des zones où abonde la végétation. Les pixels de la classe de température [2°C et plus] représentent au moins 70% de l'ensemble des pixels sur chacune des trois images (cf. Figure 4.8 et la Figure 0.7 de la distribution des valeurs de pixels en annexe A), contre une moyenne autour de 28% pour les pixels de la classe [0°C – 2°C]. Il s'observe également sur ces cartes que la plupart des pixels des zones bâties ont été bien réduits. En revanche, les écarts de température de surface au-dessus de 2°C constatés en zones urbaines (MRC de Sherbrooke) et le long des plans ou cours d'eau révèlent la présence des pixels mixtes dont les informations renfermées n'ont pu être entièrement fractionnées en des valeurs de LST associées aux différentes unités de surface composant ces pixels. Ces pixels mixtes sont des indicateurs de milieux hétérogènes.

Ces écarts de température constatés entre l'image de référence et les images simulées ont servi au calcul de l'erreur quadratique moyenne et l'erreur moyenne absolue associées à chacune des méthodes et présentées dans le Tableau 4.9.

Tableau 4.9 : Indicateurs statistiques calculés entre MODIS LST simulée à 100 m et Landsat LST observée

	TsHARP	RFR	PBIM
RMSE (°C)	2,75	2,82	3,26
EMA (°C)	2,49	2,54	2,73

Les indicateurs d'erreur pour les méthodes TsHARP et RFR varient dans le même ordre de grandeur. Les RMSE sont au-dessus de 2,5 °C et en deçà de 3°C. Ces valeurs de RMSE au-dessus du seuil souhaité peuvent être dues aux valeurs extrêmes et aberrantes de température observées dans la Landsat 08 LST. Des facteurs, comme expliqué plus haut, liés à la qualité des données initiales elles-mêmes (résidus de nuages ou leur ombre non supprimés) et à la propagation des erreurs générées à chaque étape dans la mise en œuvre peuvent expliquer les valeurs élevées des erreurs du résultat final. Pour la méthode PBIM, la valeur de RMSE est au-dessus de 3°C et celle d'EMA s'en rapproche.

Soulignons que la donnée de référence Landsat 08 LST est une donnée horaire acquise le 23 juillet 2017 à 15h 31'59" alors que la donnée simulée est une moyenne de la journée. Ceci peut engendrer des écarts au niveau des données brutes qu'on peut constater à la lecture du Tableau 4.10.



Figure 4.8 : Représentation spatiale des écarts de LST entre Landsat LST et MODIS prédite à 100 m

1 : Écart de LST entre Landsat et MODIS prédite avec RFR à 100m ; 2 : Écart de LST entre Landsat et MODIS prédite avec TsHARP à 100 m ; 3 : Écart de LST entre Landsat et MODIS prédite avec PBIM à 100 m

Tableau 4.10 : Valeurs de température selon différentes sources

Nom de station Météo	MOD_P_RFR	MOD_P_TsHARP	MOD_P_PBIM	MOD_LST_1000	LST_Landsat	TA
Bonsecours	22,06	21,73	21,60	21,59	20,38	22,60
Sawyer ville Nord	20,74	20,57	21,23	20,89	17,55	23,30
St Camille Wolfe	20,98	21,07	21,73	21,23	17,66	22,20

Le Tableau 4.10 comprend les valeurs d'observation de MODIS LST à 1000 m, Landsat LST à 100 m et la température maximale de l'air (TA) et les résultats de prédiction par RFR (MOD\_P\_RFR) et TsHARP (MOD\_P\_TsHARP) au niveau d'un même point d'observation (station météorologique). Les données de température de l'air sont des moyennes journalières. Pour les stations disposant des données, nous avons calculé les écarts entre MODIS LST et Landsat 08 LST, MODIS LST et TA et enfin entre Landsat 08 LST et TA. Nous rappelons que ces trois stations ne sont pas représentatives à l'échelle de la zone d'étude, mais permettent de saisir les différences dans les données brutes et par conséquent, dans les résultats finaux.

Tableau 4.11 : Écart de température entre les données brutes

Nom de station Météo	E-MOD-LAND	E-MOD-TA	E-LAND-TA
Bonsecours	1,21	-1,01	-2,22
Sawyer ville Nord	3,34	-2,41	-5,75
St Camille Wolfe	3,57	-0,97	-4,54
<b>Moyenne</b>	<b>2,70</b>	<b>-1,46</b>	<b>-4,17</b>

E : écart ; MOD : MODIS ; LAND : Landsat ; TA : température de l'air

Il se déduit de ce Tableau 4.11 que les écarts existants entre les données brutes dépassent au moins 1°C en valeur absolue. Les écarts les plus élevés sont entre Landsat LST et la température de l'air ; les plus faibles sont entre MODIS LST et la température de l'air. Or théoriquement, en l'absence de stress, de mélange de toute couche d'inversion de surface et en présence de ciel dégagé, la température de l'air proche de la surface est sensiblement égale à la LST dérivée des images satellitaires dans les zones boisées avec une canopée complète.

Les écarts entre MODIS LST 1000 m et Landsat LST ; MODIS LST 1000 m et T° AIR sont similaires à ceux obtenus avec les MODIS LST prédites à 100 m avec les mêmes données de référence (Landsat LST et TA). Que ce soit pour MODIS LST obtenue par TsHARP, RFR ou PBIM, l'écart dépasse les 3°C en comparaison avec Landsat 08 LST au niveau de deux stations météorologiques (cf. Tableau 4.12). Tandis que les écarts avec les températures de l'air sont en dessous de 2 °C au niveau des stations Bonsecours et St Camille Wolfe, et au-dessus de 2.5 °C à la station de Sawyer ville (-2,56 °C et -2,73 °C respectivement pour RFR et TsHARP).

Tableau 4.12 : Écart des valeurs de température entre les données simulées et données originales

Nom de station Météo	E- RFR-LAND	E-RFR-TA	E-TSHP-LAND	E-TSHP-TA	E-PBIM-LAND	E-PBIM-TA
Bonsecours	1,68	-0,54	1,35	-0,87	1,22	-1
Sawyer ville Nord	3,19	-2,56	3,02	-2,73	3,68	-2,07
St Camille Wolfe	3,32	-1,22	3,41	-1,13	4,07	-0,47
<b>Moyenne</b>	<b>2,73</b>	<b>-1,44</b>	<b>2,59</b>	<b>-1,58</b>	<b>2,99</b>	<b>-1,18</b>

E : écart ; MOD : MODIS ; LAND : Landsat ; TA : température de l'air

L'analyse des écarts selon la méthode de réduction d'échelle montre qu'en dehors de la station St Camille Wolfe où l'écart de TsHARP (3,41°C) est au-dessus de celui de RFR (3,32 °C), les écarts sont faibles au niveau des deux autres stations en considérant les écarts avec Landsat 08 LST. En revanche, c'est l'inverse en prenant les écarts avec la température de l'air, excepté toujours St Camille Wolfe où l'écart de RFR est plus faible que celui de TsHARP et PBIM.

Dans cette deuxième expérimentation, les résultats sont plus proches des températures de l'air que la LST (°C) dérivée de Landsat 08. Aussi, la méthode RFR donne un meilleur résultat au plan quantitatif par rapport à TsHARP.

#### 4.2.3. Expérimentation 3 : Région de Montérégie

Dans cette troisième expérimentation, nous présentons d'abord les résultats puis de leur validation.

##### 4.2.3.1. Résultats

L'analyse de colinéarité entre les neuf variables explicatives calculées à travers la matrice de corrélation de Bravais Pearson (Tableau 4.13) révèle de liens forts entre tous les indices de végétation, deux à deux, et entre chacun des indices de végétation et l'indice différencié de l'eau (Cf. les valeurs de liaison mises en rouge, Tableau 4.13). Il en résulte que les choix des variables explicatives pour cette simulation pour les méthodes de TsHARP et de RFR sont le NDVI, NDBI, Orientation et l'altitude. Les résultats obtenus sont représentés spatialement sous forme de cartes sur les images 1, 2, 3 et 4 de la Figure 4.9. Les images 2, 3 et 4 sont les résultats de LST (°C) prédites à 100 m obtenus respectivement avec les méthodes TsHARP, RFR et PBIM tandis que l'image 1 est celle de LST (°C) brute de MODIS à 1000 m. Les valeurs de LST ont été catégorisées en différentes classes d'amplitude égale à 3. La même légende a été adoptée pour toutes les cartes.

Tableau 4.13 : Région de Montérégie : analyse de corrélation linéaire univariée entre les variables auxiliaires

Variables	SAVI	EVI2	NDVI	PV	NDBI	NDWI	Orientation	Pente	Altitude
SAVI	1								
EVI2	0,9864	1							
NDVI	1	0,98637	1						
PV	0,99254	0,99885	0,99252	1					
NDBI	-0,07989	-0,21167	-0,0798	-0,17586	1				
NDWI	-0,97738	-0,94112	-0,97739	-0,95354	-0,09508	1			
Orientation	0,35155	0,32074	0,35156	0,33005	0,16511	-0,38576	1		
Pente	0,31282	0,29468	0,31283	0,30029	0,10779	-0,3316	0,16765	1	
Altitude	0,29662	0,30952	0,29662	0,30462	-0,02106	-0,27613	0,10929	0,16837	1

L'analyse visuelle des images de MODIS LST réduites à 100 m (cf. Figure 4.9) par la méthode TsHARP (image 2), la méthode RFR (image 3) et la méthode PBIM (image 3) par rapport à celle de MODIS LST originale de résolution spatiale 1000 m (image 1) montre une forte concordance entre elles avec plus d'affinement sur les images 3 et 4 que pour l'image 2. Ainsi, certaines classes de LST couvrent plus de superficies sur les images simulées que l'espace occupé sur l'image originale. Tel est l'exemple de la classe de LST (°C) de [20-22,9] repérée par une ellipse sur les trois figures. Les structures, formes et patrons des unités de surface sont assez perceptibles sur les images réduites. L'allure générale des unités de surface s'affinant davantage donne des indications dans l'identification de ces unités. Ces descriptions sont significatives en analysant les images de la Figure 4.10 et la Figure 4.11. La Figure 4.10 est une vue partielle d'une zone hétérogène à dominance urbaine et la Figure 4.11 montre un milieu hétérogène à dominance agricole avec des parcelles agricoles, des plans d'eau et de bâtis.

Sur la Figure 4.10, nous remarquons, par exemple, des zones artificialisées disparates dans les MRC de Haut Richelieu (comme Iberville), Roussillon ou des Jardins de Napierville et un cours d'eau passant par Iberville. Nous remarquons également le même nombre (5) de classes de température sur les images 1 et 2 contre 11 sur les images 3 et 4 (cf. Figure 4.9) si l'on considère les pixels de la classe [-1, 16,9] comme contaminés par les bruits. Les valeurs de LST oscillent entre 17 et 49 °C avec une moyenne générale autour de 26 °C dans la zone d'étude.

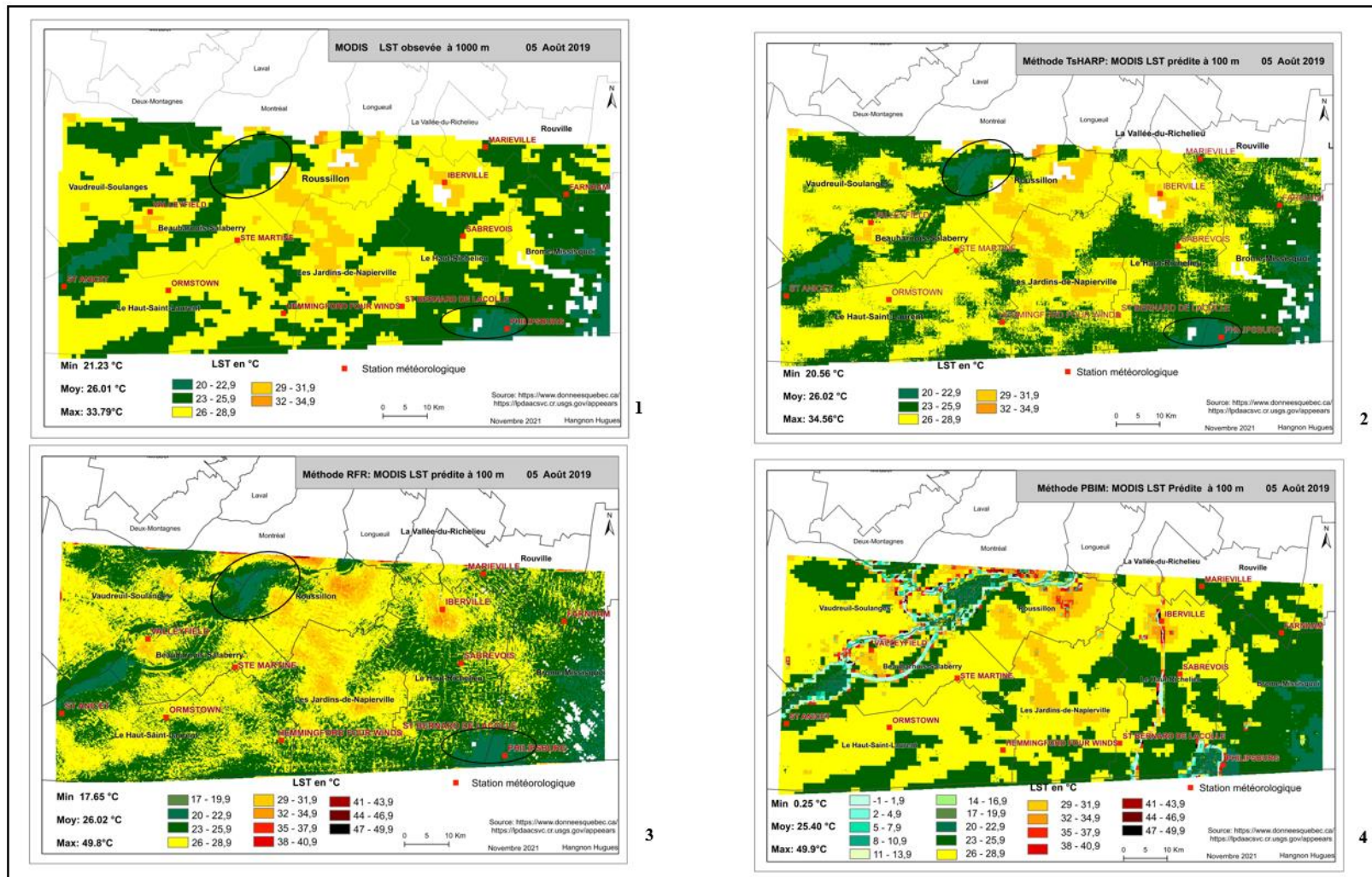


Figure 4.9 : Région de Montérégie : résultats de réduction d'échelle à 100 m suivant différentes méthodes

1 : MODIS LST observée à 1000m ; 2 : MODIS LST prédite à 100 m par la méthode TsHARP ; 3 : MODIS LST prédite à 100 m par la méthode RFR ; 4 : MODIS LST prédite à 100 m par la méthode PBIM

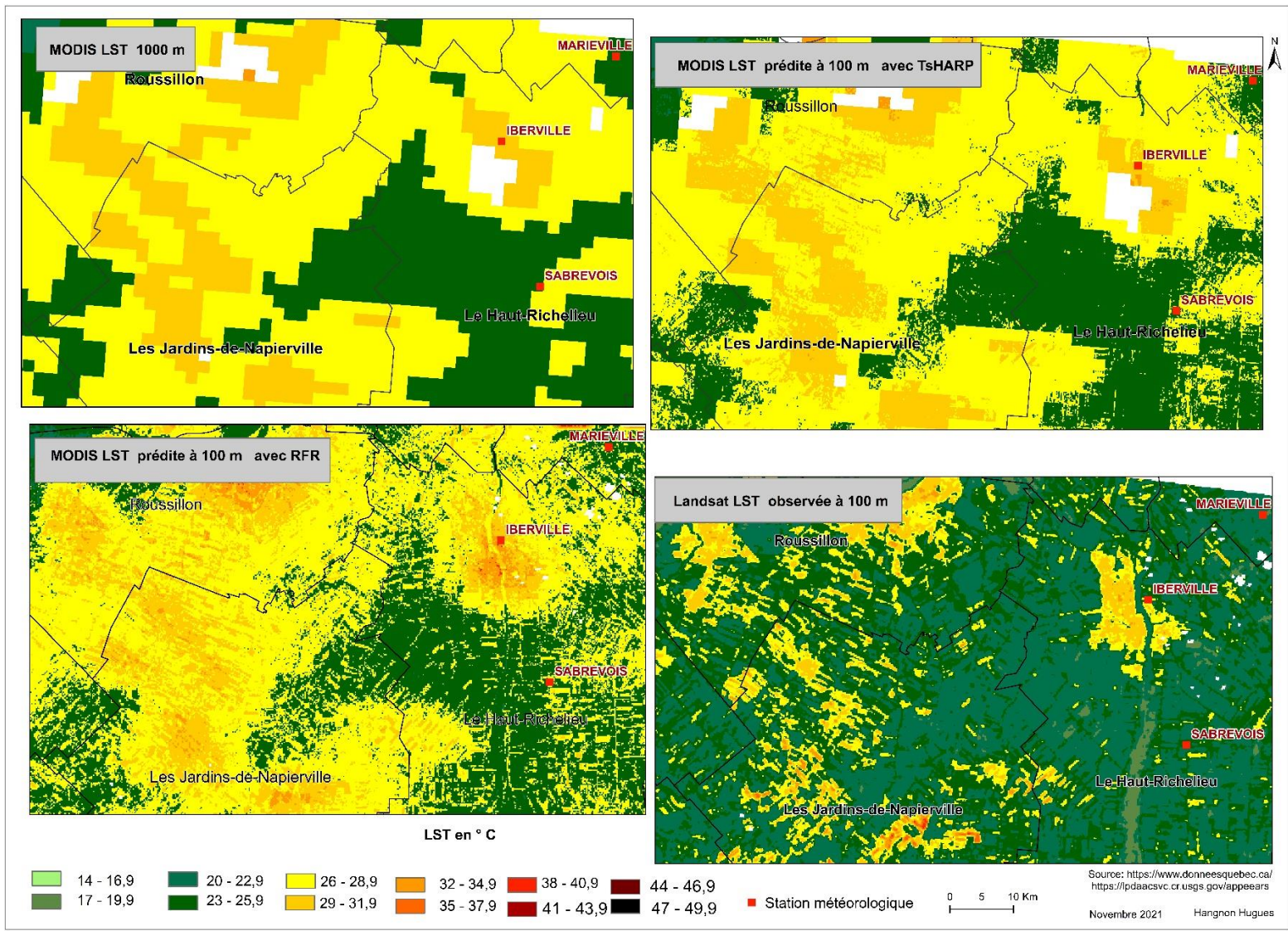


Figure 4.10: Région de Montérégie : zoom sur la représentativité spatiale des valeurs de LST en milieu hétérogène à dominance urbaine

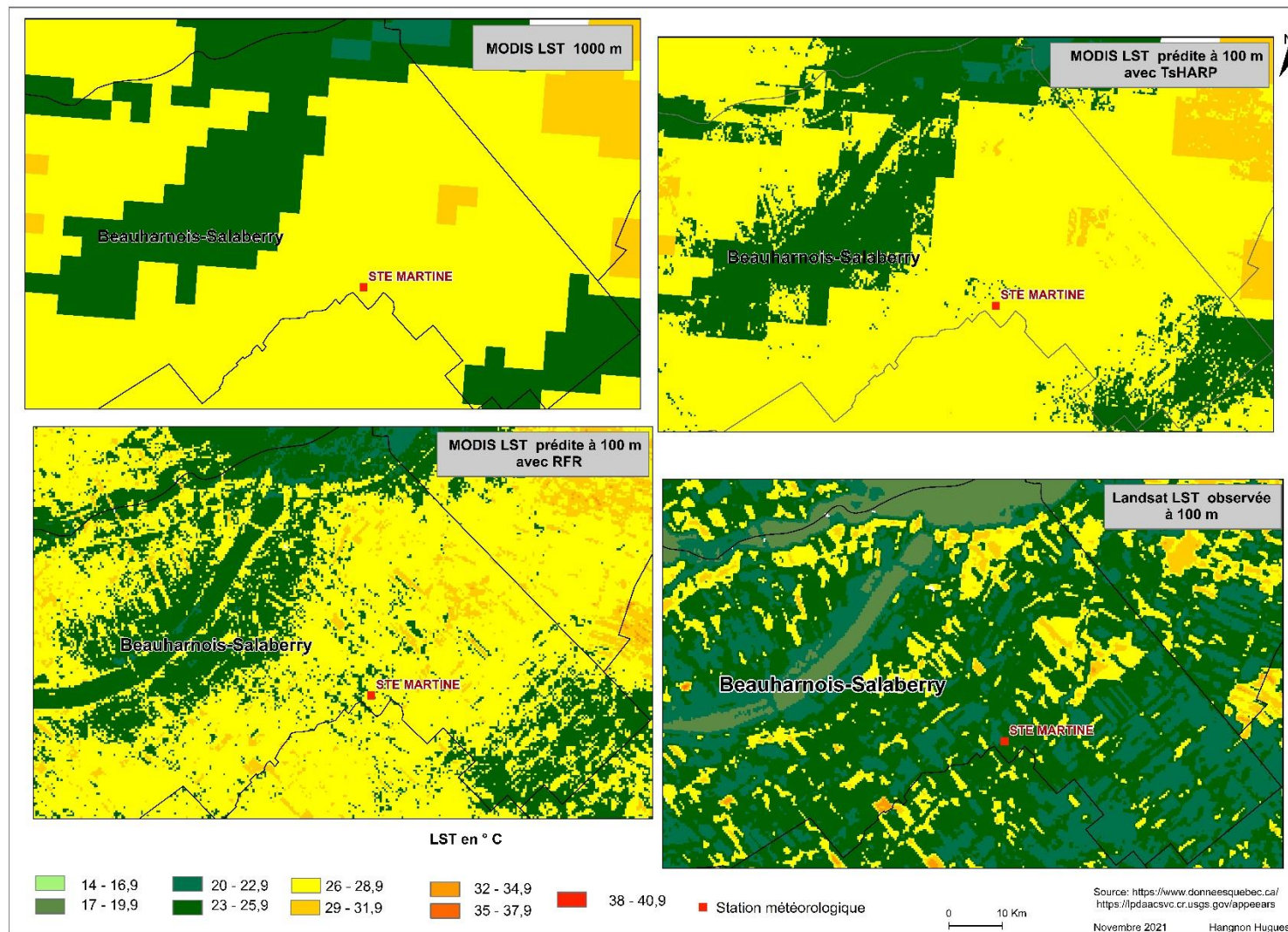


Figure 4.11 : Région de Montérégie : zoom sur la représentativité spatiale des valeurs de LST en milieu hétérogène à dominance agricole



Les indicateurs d'erreurs associés aux méthodes de réduction d'échelle présentés dans le Tableau 4.14 sont en deçà de 1°C pour TsHARP contre 1,5 °C pour le RFR. Ces valeurs en deçà du seuil de 2°C, retenu comme marge d'erreur tolérée, peuvent attester que les résultats sont satisfaisants. Par contre, une RMSE de 4,72 °C pour la méthode PBIM indique la présence des valeurs extrêmes ou aberrantes à éliminer.

Tableau 4.14 : Indicateurs statistiques calculés entre MODIS LST simulée à 100 m et MODIS LST observée

Montérégie	TsHARP	RFR	PBIM
RMSE (°C)	0,44	1,33	4,72
EMA (°C)	0,29	0,97	1,40

Ces résultats vont être comparés à d'autres données sources, Landsat 08 TIRS LST et la température de l'air pour vérifier leur justesse dans la sous-section sur la validation des résultats à venir.

#### 4.2.3.2. Validation des résultats

La comparaison de résultats obtenus avec les données de température de l'air et la LST de l'imagerie Landsat 08 TIRS, prise le 05/08/2019 à 15h38'06'', permet de vérifier si les résultats sont proches de la réalité. Les erreurs quadratiques moyennes et erreurs absolues moyennes calculées (Cf. Tableau 4.16) entre les images réduites à 100 m et la Landsat LST 08 sont toutes au-dessus de 3 °C. Ces marges d'erreur, au-dessus du seuil de 2°C, pourraient indiquer que nos résultats ne sont pas représentatifs de la réalité.

Tableau 4.15 : Indicateurs statistiques calculés entre MODIS LST simulées et Landsat LST observée

	TsHARP	RFR	PBIM
RMSE	3,76	3,71	5,55
EMA	3,37	3,31	4.18

Tableau 4.16 : Valeurs de température selon différentes sources

Nom de station Météo	MOD_P_RFR	MOD_P_TSHP	MOD_LST_1000	MOD_P_PBIM	LST_LAND	TA
Hemmingford Four Winds	25,53	25,95	25,79	26,41	25,13	28,5
Marieville	25,12	25,43	<b>25,51</b>	25,47	<b>21,38</b>	26,9
Ormstown	26,07	26,26	<b>26,72</b>	26,44	<b>22,25</b>	27,5
St Anicet	24,42	23,63	23,93	24,26	22,55	28,5

Cependant, ces marges d'erreur élevées peuvent aussi être imputables aux données sources elles-mêmes à la lecture du Tableau 4.16. Le Tableau 4.16 présente des valeurs de température de sources différentes relevées pour un même point de station météorologique. Bien que ces 4 points ne soient pas représentatifs à l'échelle de la zone d'étude, ils permettent de comprendre les probables sources d'erreurs. Il comprend les valeurs de température prédites par RFR (MOD\_P\_RFR) ou par TsHARP (MOD\_P\_TsHARP), les températures de MODIS

LST à 1000 m et la température maximale de l'air (TA). Les valeurs dans les trois premières colonnes sont quasiment similaires. Par contre, quand elles sont comparées à celles des deux dernières colonnes, des écarts sont grands. Les données brutes MODIS LST et Landsat LST (LST LAND) malgré les prétraitements pour éliminer les bruits, ils existent toujours et peuvent toujours engendrer des erreurs tout au long du processus de réduction. Ainsi, en plus de l'erreur existante au niveau des données brutes, il y a celles liées aux modèles eux-mêmes. C'est ce qui expliquerait cette augmentation des erreurs au niveau du Tableau 4.15.

S'agissant de la qualité des données, les données utilisées pour alimenter les modèles proviennent de trois sources différentes (Capteur MODIS TERRA, ASTER, Landsat 08/TIRS) présentant chacune ses spécificités. Leur utilisation dans les études comme celle-ci nécessite de procéder à l'intercalibration pour corriger les erreurs intrinsèques. Ce processus peut engendrer des erreurs qui vont se propager à chaque étape de la chaîne de traitement.

Pour une meilleure lecture, nous avons calculé, dans le Tableau 4.17, les écarts de température entre la donnée source MODIS LST à 1000 m et les données de référence (LST Landsat et TA), ainsi que les écarts entre les données réduites et les données de référence dans le Tableau 4.18.

Tableau 4.17 : Écart des valeurs de température entre les données originales

Nom de station Météo	E-MOD-LAND	E-MOD-TA	E-LAND-TA
Hemmingford Four Winds	0,66	-2,71	-3,37
Marieville	4,13	-1,39	-5,52
Ormstown	4,47	-0,78	-5,25
St-Anicet	1,38	-4,57	-5,95

E : écart ; MOD : MODIS ; LAND : Landsat ; TA : température de l'air

Il ressort du Tableau 4.17 qu'entre MODIS LST brute et les données de référence, les écarts varient en valeur absolue de 0,66 °C à 4,57°C pour la température de l'air (E-MOD-TA) et de 5,95°C entre Landsat et température de l'air (E-LAND-TA). Les écarts les plus élevés sont observés à Marieville et Ormstown dans la première colonne, à St-Anicet dans la deuxième colonne et enfin à toutes les stations météorologiques pour la troisième colonne, qui compare les données LANDSAT, prise le 05/08/2019 à 15h38'06", à la température de l'air. Bien que ces données aient été acquises à différents moments, ces valeurs renseignent sur la variabilité spatiale de la température de surface à différentes résolutions et altitude. En effet, peu de facteurs influencent la variabilité de la LST à l'échelle d'une station météorologique. Alors que ces facteurs peuvent augmenter à l'échelle spatiale de Landsat 08 et, plus encore à celle de MODIS (Bois, 2012). En plus de ces facteurs, la différence dans les périodes d'acquisition des données pourrait engendrer des erreurs entre elles. Ainsi, ces écarts peuvent se répercuter sur les résultats finaux après la réduction d'échelle.

Les résultats finaux obtenus par RFR, TsHARP ou PBIM comparés aux données de référence révèlent des écarts au-dessus de 3°C (valeurs en couleur rouge dans le Tableau 4.18) au niveau des stations Marieville, Ormstown et St Anicet. Nous remarquons également que parmi les valeurs au-dessus de 3°C, quatre d'entre elles proviennent du calcul impliquant la Landsat 08 LST.

Tableau 4.18 : Écart des valeurs de température entre les données simulées et données originales

Nom de station Météo	E- RFR-LAND	E- RFR-TA	E-TSHP-LAND	E-TSHP-TA	E-PBIM-LAND	E-PBIM-TA
Hemmingford Four Winds	0,41	-2,97	0,83	-2,55	1,28	-2,09
Marieville	3,74	-1,78	4,05	-1,47	4,09	-1,43
Ormstown	3,81	-1,43	4,01	-1,24	4,19	-1,06
St Anicet	1,87	-4,08	1,08	-4,87	1,71	-4,24
<b>Moyenne</b>	<b>2,46</b>	<b>-2,57</b>	<b>2,49</b>	<b>-2,53</b>	<b>2,82</b>	<b>-2,21</b>

E : écart ; RFR : Random Forest Regression ; LAND : Landsat ; TA : température de l'air, TSHP : TsHARP; MOD : MODIS, LST : Land surface temperature, PBIM : Pixel block intensity modulation

En faisant la comparaison par méthode, nous constatons, qu'excepté à la station St Anicet, avec la donnée de référence Landsat 08 LST, les écarts pour le RFR sont plus faibles que ceux de TsHARP et de PBIM. Tandis que, c'est le cas contraire avec la donnée de référence température de l'air, hormis toujours la station St-Anicet où l'erreur avec TsHARP ou PBIM dépasse celle de RFR.

Partant des constats des Tableau 4.17 et 4.18, nous affirmons que les données de température de l'air valident mieux les résultats de réduction d'échelle obtenus en Montérégie. De plus, la méthode TsHARP semble fournir un bon résultat par rapport à la méthode de RFR si l'on tient compte des indicateurs RMSE/EMA de la validation avec la Landsat LST.

Les écarts quantitatifs avec la Landsat 08 LST sont spatialisés pour chacune des méthodes et présentés sur la Figure 4.12 avec les images 2 (RFR), 3 (TsHARP) et 4 (PBIM). Une analyse visuelle de ces illustrations montre une similitude entre les distributions spatiales des écarts de température en valeur absolue répartis en deux classes [0 °C – 2°C] et [2°C et plus]. La majorité des pixels, répartis sur l'ensemble de la zone d'étude, appartient à la classe de [2°C et plus] en des proportions de 77%, 78% et 79% respectivement pour les méthodes de RFR, TsHARP et PBIM. (cf. les images 2, 3 et 4 de la Figure 4.12 et les diagrammes de distribution des écarts de la Figure 0.8 en annexe A). Ces pixels de valeur de température de surface au-dessus de la valeur seuil (2°C) sont localisés dans les milieux hétérogènes à dominance agricole et les zones humides et faiblement dans les zones construites. Les pixels de la classe [0 °C, 2 °C] représentant 23 % (RFR), 22% (TsHARP) et 21% (PBIM) de tous les pixels sont ceux des deux images (observée et prédite) qui s'ajustent le mieux entre eux. On retrouve ces pixels notamment dans les zones bâties et imperméables, les zones nues et de manière éparse dans les zones agricoles.

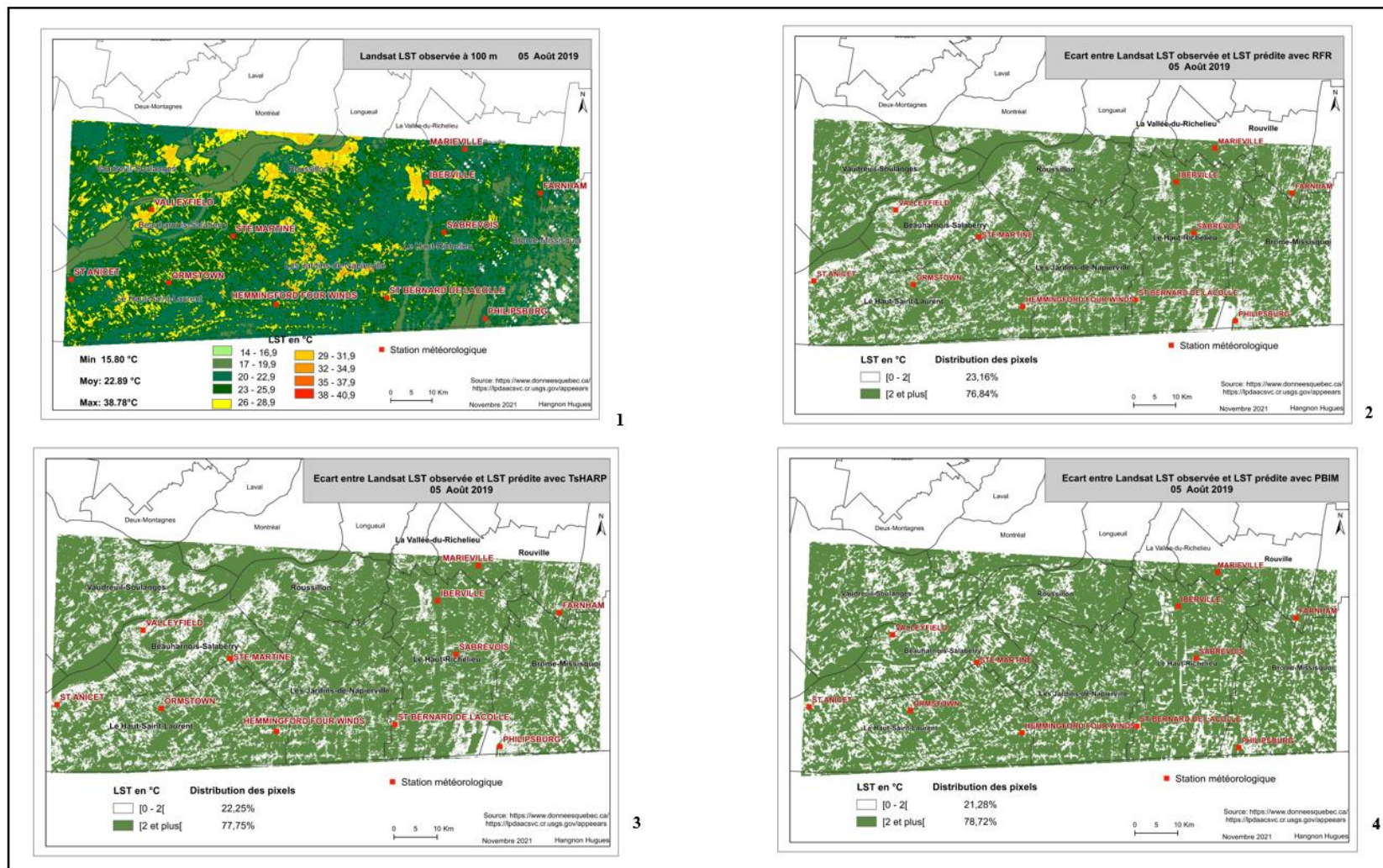


Figure 4.12 : Région de Montérégie : représentation des écarts de températures entre Landsat LST et MODIS LST prédite

1 : Landsat LST observée à 100 m ; 2 : Écart de LST entre Landsat et MODIS prédite avec RFR à 100 m ; 3 : Écart de LST entre Landsat et MODIS prédite avec TsHARP à 100 m ; 4 : Écart de LST entre Landsat et MODIS prédite avec PBIM à 100 m

Les écarts négatifs constatés pourraient signifier que les valeurs de pixels de MODIS LST à 1000 m paraissent faibles pour être réduites à cause de la présence de l'humidité du sol. Cela peut signifier aussi la présence de pixels mixtes. Pour les écarts positifs, nous les attribuons aux valeurs élevées de pixels des zones urbaines mesurées par le satellite Landsat 08 TIRS. Les valeurs de températures LST des zones bâties mesurées par Landsat 08 TIRS paraissent plus précises avec plus de détails que celles mesurées par le capteur imageur MODIS. Même avec la réduction d'échelle de MODIS LST à 100 m, où les valeurs de pixels ont été décomposées en fonction des unités d'occupation du sol, l'effet de généralisation subsiste toujours sur les images réduites. Dans cette expérimentation, il semble que les pixels mixtes qui n'ont pu être affinés sont dans les zones agricoles et les zones humides.

### **4.3. Expérimentation 4 : réduction d'échelle de 100 m à 30 m dans la région de l'Estrée**

Dans cette quatrième expérimentation, l'objectif est d'atteindre une résolution de 30 m en partant des résultats précédemment obtenus à 100 m tel que décrit dans la nouvelle approche méthodologique (cf. la sous-section 2.3.1.) Ainsi, la variable auxiliaire à utiliser est l'image des émissivités de surface à 30 m. L'image des émissivités de surface a été utilisée comme variable explicative. La méthode de réduction d'échelle appliquée est TsHARP étant donné que des trois méthodes, c'est elle qui a donné les faibles marges d'erreur.

Pour disposer des images des émissivités, nous avons procédé à la classification des images suivant une nomenclature à 4 niveaux pour la région de l'Estrée : zone artificialisée (urbanisée), zone humide, zone de forêt et la zone de végétation. La Figure 4.13 montre les résultats de l'occupation et l'utilisation des sols. Elle montre un milieu où la couverture boisée (végétation et forêt) est dominante suivie de zone urbanisée éparse, mais plus concentrée au sud. On y trouve aussi des zones humides.

Comme décrit dans la section 2.3.2 portant sur la réduction d'échelle à 30 m, le calcul des émissivités de surface basé sur l'occupation des sols utilise des valeurs extraites de la base de données ECOTRESS et de la thèse de Kotchi (2015). Par exemple, nous avons affecté aux zones humides 0,983, 0,964 aux surfaces artificialisées et 0,985 à la végétation. Pour la forêt dense, cette valeur est 0,99 et 0,967 pour le sol nu, etc. Ceci a permis de produire la carte de la Figure 4.14.

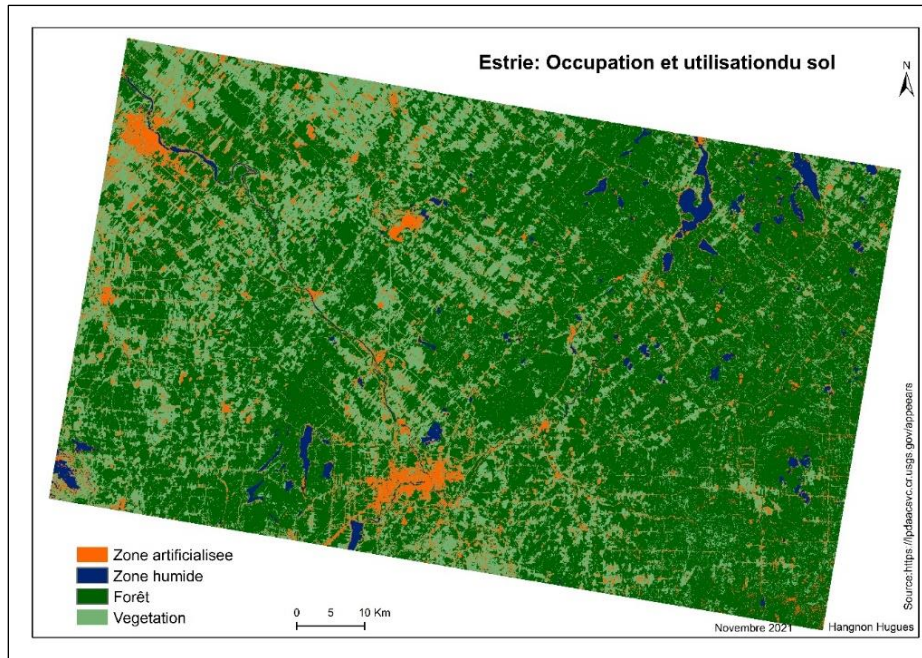


Figure 4.13 : Occupation et utilisation des sols dans la région de l'Estrie

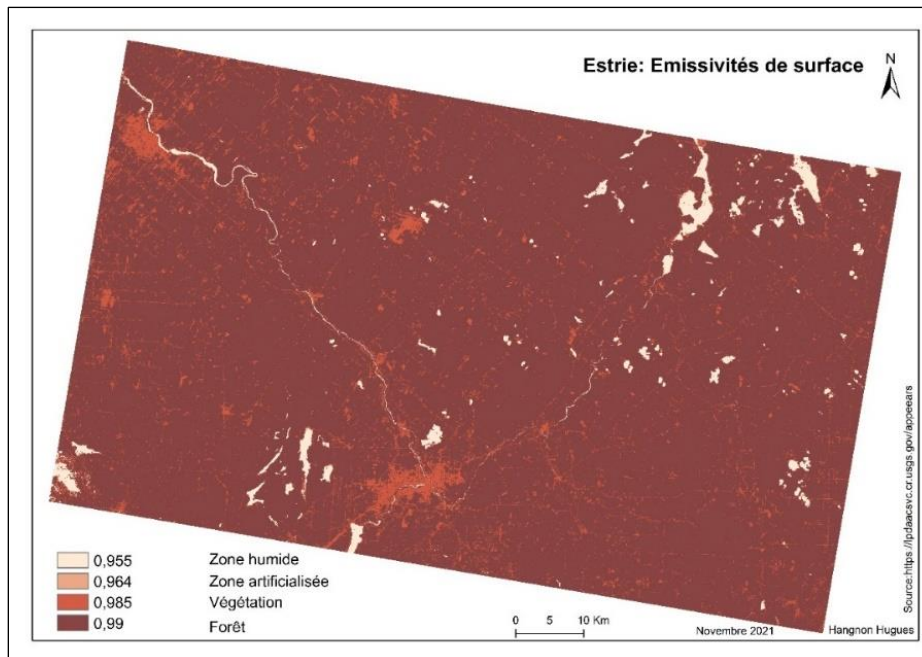


Figure 4.14 : Émissivités de surface de la région de l'Estrie

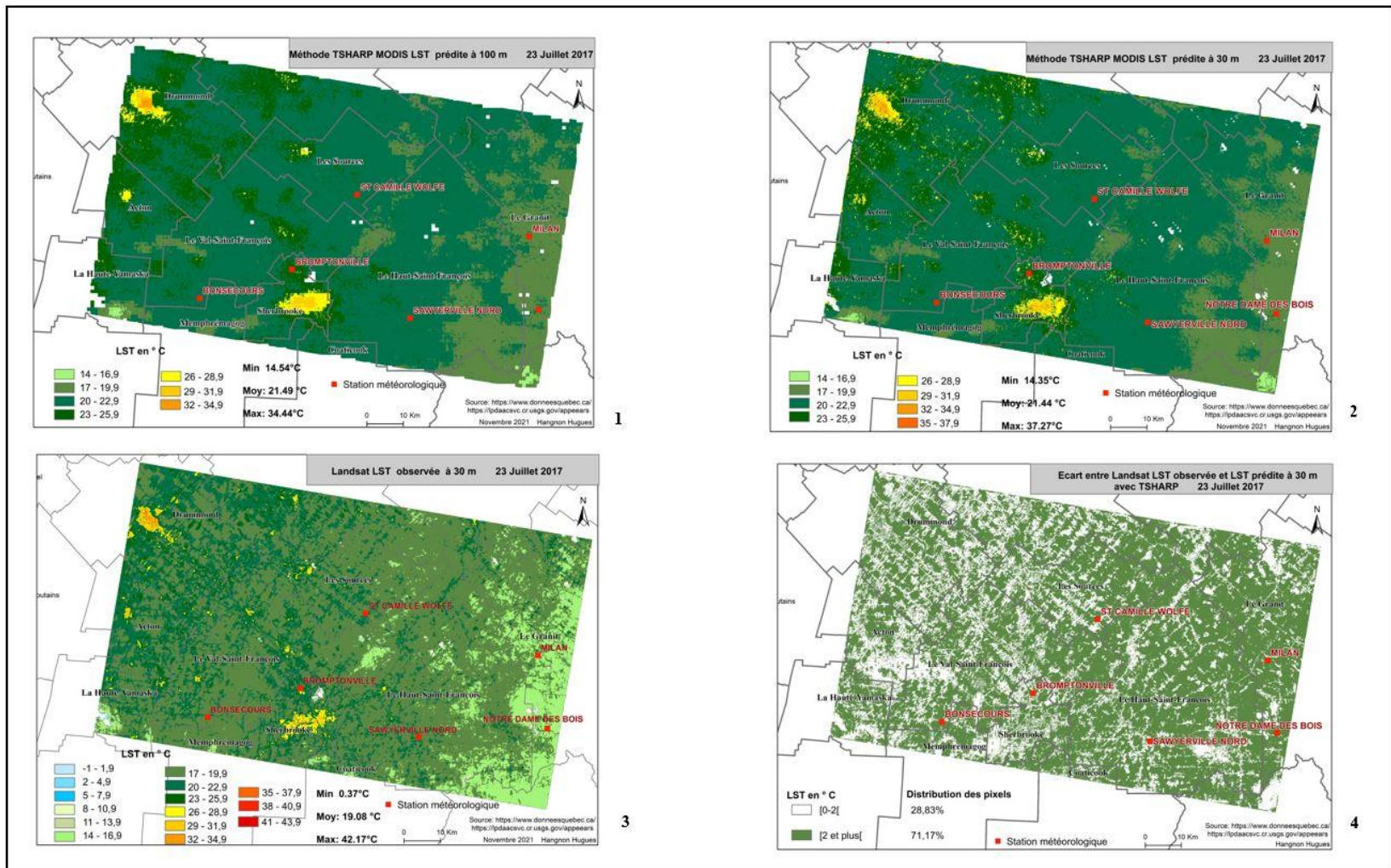


Figure 4.15 : Résultats de la réduction d'échelle de MODIS LST de 100 m à 30 m

1 : MODIS LST prédite de 1000 m à 100 m avec TsHARP ;  
 3 : Landsat 08 TIRS LST observée à 30 m

2 : MODIS LST prédite de 100 m à 30 m avec TsHARP (émissivité) ;  
 4 : Écart entre Landsat 08 TIRS LST et MODIS LST prédite à 30 m

#### 4.3.1.1. Présentation des résultats

L'image 2 de la Figure 4.15 est le résultat de la simulation de 100 m à 30 m avec la méthode TsHARP. En la comparant visuellement avec l'image 1 de la Figure 4.15 représentant MODIS LST à 100 m, on observe une cohérence dans la distribution spatiale des différents tons de couleur. Une nouvelle classe de température [35°C-37,9°C] est apparue sur l'image réduite à 30 m. Ceci démontre que les pixels de LST ont été bien désagrégés et que les émissivités de surface ont amélioré la prédiction. Pour évaluer la précision de cette réduction, nous l'avons comparée avec l'image LST de Landsat 08 TIRS observée à 30 m (cf. image 3 de la Figure 4.15). La différence entre cette image observée et l'image réduite a donné des écarts spatialisés sur l'image 4 de la Figure 4.15. Comme il a été montré précédemment, nous constatons également que les écarts faibles (<-2°C) sont les plus représentatifs et correspondent aux zones hétérogènes d'abondance de végétation. Par contre, les pixels de valeur supérieure à 2°C sont des poches localisées notamment dans les zones construites.

#### 4.3.1.2. Validation des résultats

La comparaison des résultats avec la Landsat LST 08 a donné une marge d'erreur RMSE sensiblement égale à 3°C et une erreur moyenne absolue EMA de 2,55°C (cf. Tableau 4.19). Ces indicateurs sont plus élevés que ceux observés (RMSE/EMA= 2,75°C/2,49°C) à 100 m avec la méthode TsHARP. Soit un écart respectif de 0,13 °C et 0,06 °C. Ce qui paraît logique à cause de la propagation des erreurs obtenues auparavant à 100 m. Ce constat corrobore les conclusions de l'état de l'art rapportant que, plus le facteur d'échelle est grand, plus grande est la marge d'erreur.

Tableau 4.19 : Indicateurs d'erreurs entre les résultats et Landsat 08 à 30 m

	TsHARP (émissivité)
RMSE (°C)	2,88
EMA (°C)	2,55

Par ailleurs, la comparaison des valeurs de température de Landsat 08 TIRS et de l'air au niveau de trois stations météorologiques (Tableau 4.20) avec l'image LST prédite à 30 m montre de faibles écarts avec la température de l'air. Ainsi, la LST prédite à 30 m se rapproche plus de la température de l'air.



Tableau 4.20 : Écart des valeurs de température entre l'image simulée et les données observées

Nom de station Météo	E-TSHP-LAND	E-TSHP-TA
Bonsecours	1,89	-0,57
Sawyer ville Nord	3,12	-2,57
St Camille Wolfe	3,32	-1,22
<b>Moyenne</b>	<b>2,78</b>	<b>-1,45</b>

E : écart ; LAND : Landsat ; TA : température de l'air, TSHP : TsHARP

Dans la section suivante, nous comparons ces indicateurs d'erreur avec ceux résultant des images réduites directement de 1000 m à 30 m afin d'évaluer la pertinence de l'approche en deux étapes.

#### 4.4. Expérimentation 5 : réduction d'échelle de 1000 m à 30 m dans la région de l'Estrie

Bien que nous ayons constaté une amélioration de la réduction d'échelle de 100 m à 30 m en combinant les émissivités de surface, il faudrait valider cette hypothèse en réduisant directement de 1000 m à 30 m sans utiliser les émissivités de surface comme variables prédictives. C'est ce qui est fait dans cette section. L'objectif de cette expérimentation de réduction d'échelle de MODIS LST de 1000 m à 30 m est double : il s'agit d'une part de comparer la précision des résultats à 30 m (celui obtenu en deux étapes, soit de 1000 à 100 m puis de 100 m à 30 m, et celui à prédire de 1000 m à 30 m en une seule étape) et d'autre part de valider la nouvelle approche méthodologique. De fait, nous avons réalisé l'expérimentation sur la région de l'Estrie. Les mêmes variables auxiliaires (NDVI, Altitude, NDBI, SAVI et Orientation) utilisées pour la réduction de 1000 m à 100 m ont été reprises ici pour les méthodes de TsHARP et RFR pour estimer la LST de 1000 m à 30 m. Les résultats obtenus sont présentés dans la Figure 4.16 pour le RFR et la Figure 4.17 pour le TsHARP.

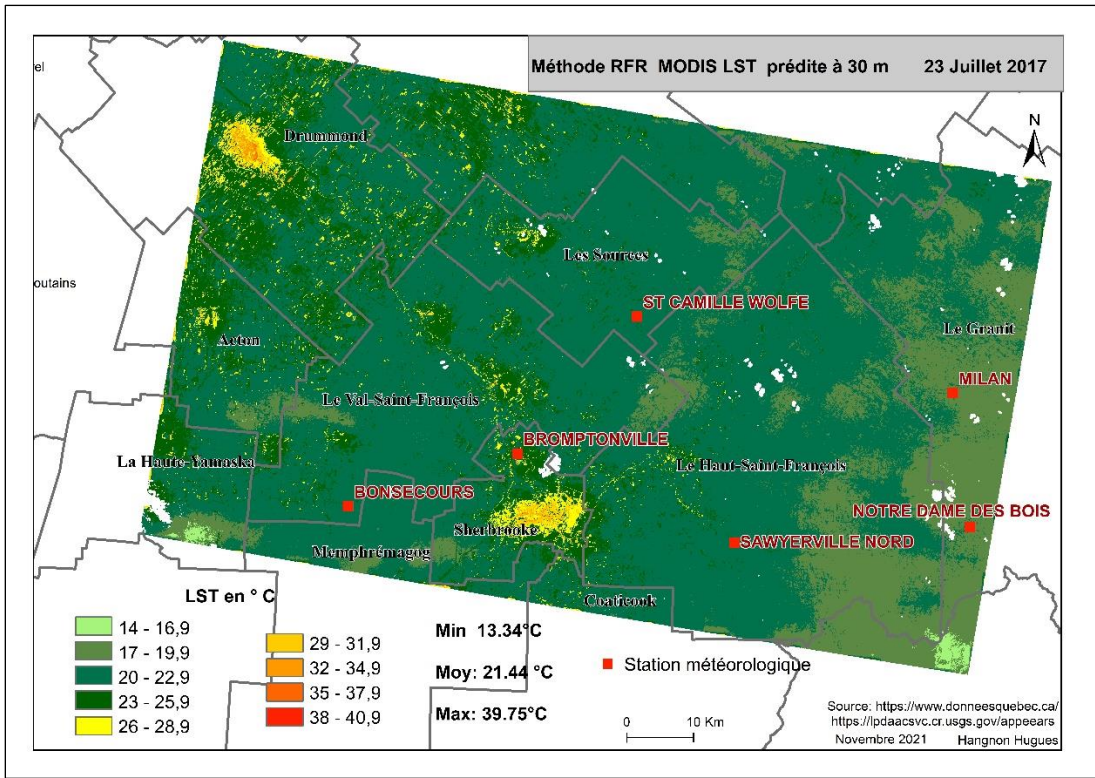


Figure 4.16 : MODIS LST prédite de 1000 m à 30 m avec RFR

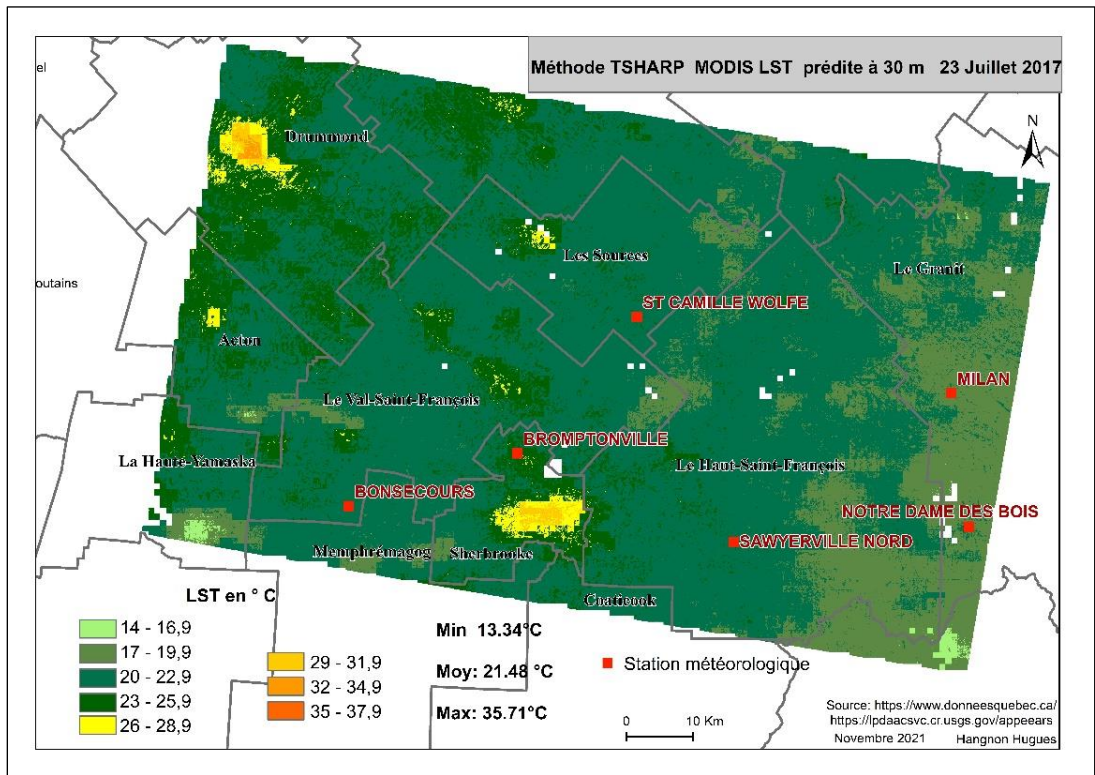


Figure 4.17 : MODIS LST prédite de 1000 m à 30 m avec TshARP

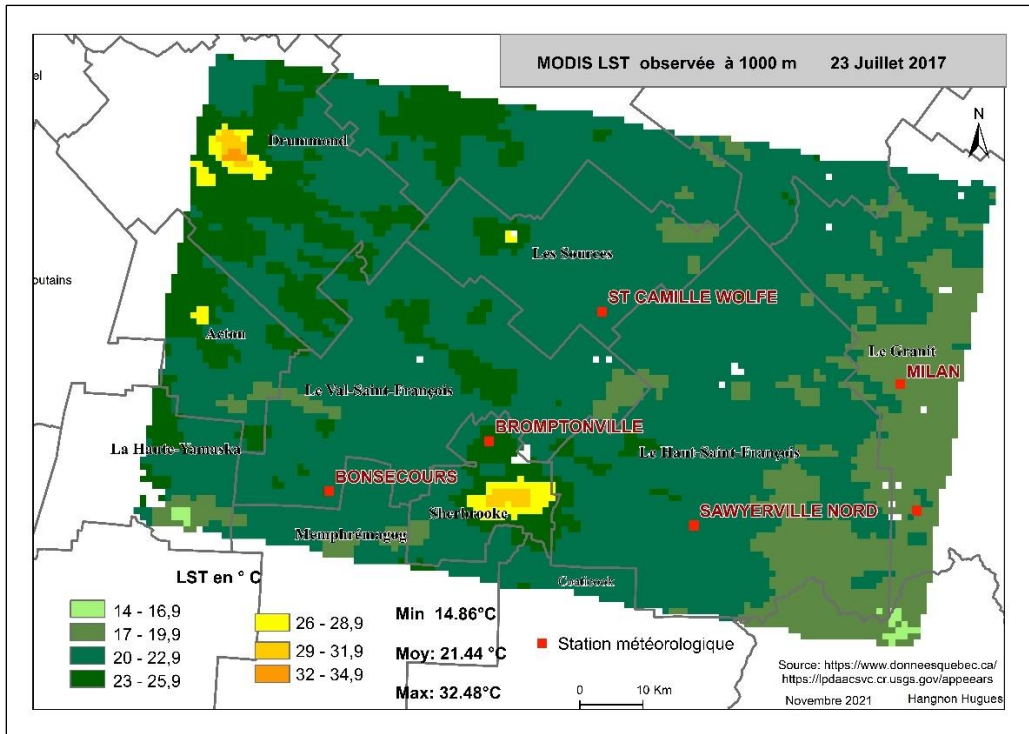


Figure 4.18 : MODIS LST observée à 1000 m

### Présentation des résultats et leur validation

Nous constatons sur la Figure 4.16 et la Figure 4.17 une similitude dans la distribution spatiale des valeurs de LST à 30 m avec celles de MODIS LST à 1000 m (Figure 4.18). Les valeurs minimales (13,44°C) des images réduites sont égales contrairement aux valeurs maximales qui sont de 39,75°C pour la méthode de RFR et 35,71°C pour le TSHARP. Le maximum est de 32,48°C pour la MODIS LST 1000 m. Ainsi, par la méthode TSHARP, la réduction d'échelle a permis d'ajouter une classe de plus ([35-37,9°C]) dans les températures chaudes contre deux ([35-37,9°C] et [38-40,9°C]) pour la méthode RFR. Ces détails de température apparaissent dans les zones urbaines. Les valeurs de RMSE et EMA (Tableau 4.21) entre les images réduites et l'image Landsat 08 TIRS LST à 30 m amènent à conclure à une justesse des données obtenues avec RFR par rapport à la méthode TSHARP.

La lecture du Tableau 4.22 montrant les écarts avec la Landsat LST et la température de l'air selon chaque méthode révèle de faibles écarts avec la température de l'air. Ainsi, la température de l'air valide mieux ces résultats à 30 m. Les images 1 et 2 de la Figure 4.19 représentent la spatialisation respective des écarts selon RFR et TSHARP. Aucun changement notable n'a été constaté sur les mêmes types de cartes interprétées en sus.

Tableau 4.21 : Indicateurs d'erreurs entre les résultats et Landsat 08 à 30 m

	TsHARP	RFR
RMSE (°C)	2,77	2,50
EMA (°C)	2,51	2,77

Tableau 4.22: Écart des valeurs de température entre les données simulées et données originales

Nom de station Météo	E-RFR-LAND	E-RFR-TA	E-TSHP-LAND	E-TSHP-TA
Bonsecours	2,4	-0,06	1,15	-1,31
Sawyer ville Nord	3,14	-2,55	3,42	-2,27
St Camille Wolfe	3,33	-1,21	3,66	-0,88
Moyenne	2,95	-1,27	2,74	-1,48

E : écart ; RFR : Random Forest Regression ; LAND : Landsat ; TA : température de l'air, TSHP : TsHARP

Nous avons comparé les valeurs de RMSE/EMA de la réduction d'échelle de 1000 m à 30 m obtenue en une seule étape et celle réalisée en deux étapes avec l'émissivité (soit de 1000 m à 100 m et de 100 m à 30 m) (Tableau 4.23). À la lecture de ce Tableau 4.23, on constate que l'approche en deux étapes présente des marges d'erreur plus grandes selon la méthode TsHARP pour les deux indicateurs. Par contre, selon la méthode RFR, la RMSE reste inférieure à celle de TsHARP et l'inverse pour l'EMA.

Tableau 4.23 : Comparaison des indicateurs à 30 m selon l'approche de RE

	TsHARP	RFR	TsHARP (Émissivité)
RMSE (°C)	2,77	2,50	2.88
EMA (°C)	2,51	2,77	2.55

Ces remarques sur les indicateurs d'erreur (Tableau 4.23) s'expliquent en partie par les niveaux de détails de la couverture du sol contenus dans les variables auxiliaires utilisées et le processus de réduction d'échelle lui-même. Selon l'approche directe (de 1000 m à 30 m), les variables de prédiction ont conservé leur résolution spatiale initiale de 30 m et donc, elles présentent plus de détails spatiaux des différentes unités d'utilisation et d'occupation du sol. Par contre, selon l'approche à deux étapes, les résolutions des variables de prédiction ont été rééchantillonnées de 30 m à 100 m pour une réduction d'échelle à 100 m à la première étape.

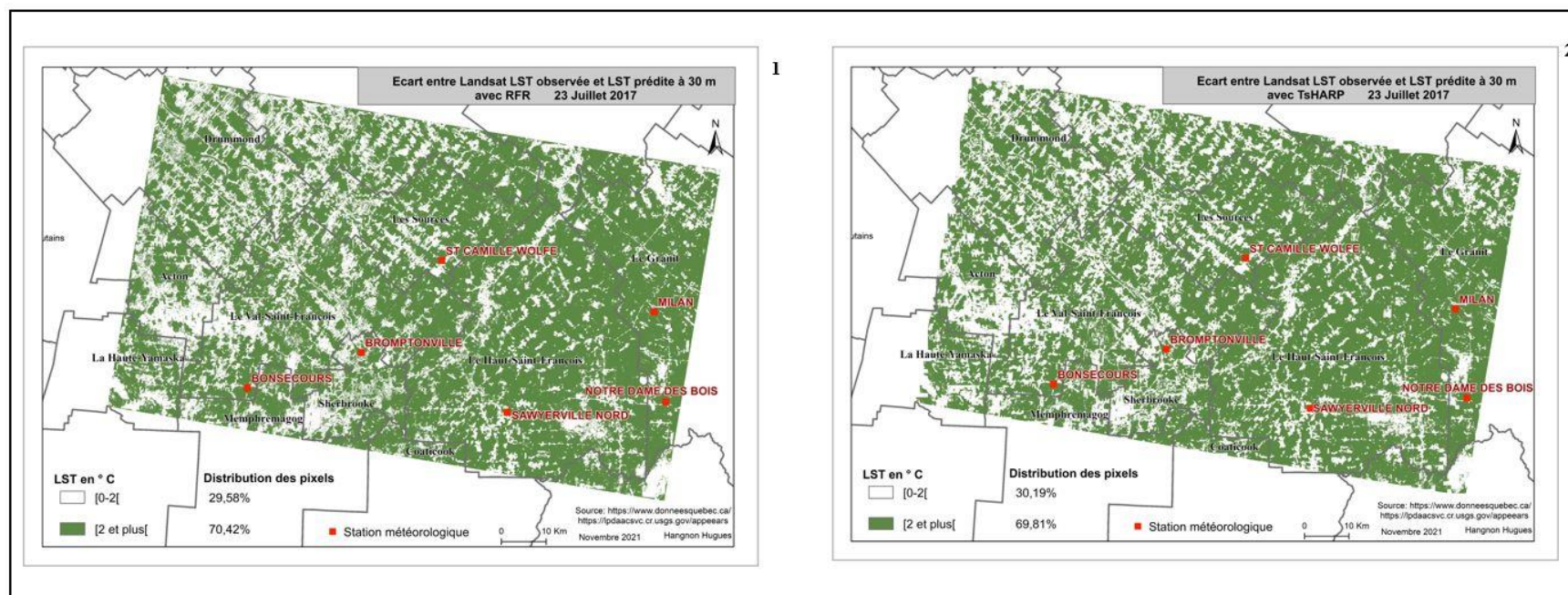


Figure 4.19 : Représentation spatiale des écarts de valeurs de LST entre Landsat LST observée et MODIS LST prédite de 1000 m à 30 m

1 : Écart entre Landsat 08 LST et MODIS LST simulée à 30 m avec RFR ; 2 : Écart entre Landsat 08 LST et MODIS LST simulée à 30 m avec TsHARP

Ce rééchantillonnage a donné lieu à une généralisation des détails de la couverture du sol et donc à une perte d'informations. À la seconde étape du processus, ce sont les valeurs d'émissivité de surface qui ont été utilisées comme variables de prédiction. La production de la carte des émissivités de surface, basées sur la carte de l'occupation et de l'utilisation du sol, peut engendrer aussi des pertes d'informations liées à l'interprétation lors de la classification.

Quant au processus de réduction d'échelle, celui décomposé en deux étapes donne lieu à une double propagation des erreurs contre une pour la méthode directe, outre la qualité de chaque algorithme et des données d'entrée.

Concernant la validation du postulat selon lequel, les émissivités de surface apportent plus de précision à l'image réduite à 30 m et testé dans une approche en deux étapes, nous avons constaté plus de détails spatiaux sur l'image réduite à 30 m que sur l'image originale à 100 m. En effet, sur l'image réduite à 30 m le minimum, la moyenne et le maximum sont respectivement de 14,35°C, 21,44°C et 37,27°C alors que sur l'image à 100 m, les valeurs sont dans ce même ordre avec 14,54°C, 21,49°C et 34,44°C. De fait, les émissivités de surface ont effectivement amélioré les résultats à 30 m, mais en engendrant plus de marges d'erreur qu'à 100 m à cause du facteur d'échelle final pondéré à 13 (facteur d'échelle =  $1000/100 = 10$  plus  $100/30 = 3,33$ ) contre 10 à 100 m.

## 4.5. Conclusion

En conclusion des expérimentations faites, nous avons obtenu des résultats satisfaisants à l'issue de la réduction d'échelle de 1000 à 100/30 m selon les méthodes TsHARP, RFR et PBIM. Les résultats ont été validés avec les références de Landsat LST à 100 m/30 m et la température moyenne maximale journalière de l'air en se basant sur la RMSE ou l'écart moyen absolu EMA. Nous avons constaté que la méthode RFR apporte plus de détails spatiaux que TsHARP et PBIM, autant à 100 m comme à 30 m, par comparaison visuelle des images réduites entre elles-mêmes, et entre elles et les images d'origine à réduire. Malgré cette amélioration des données au plan qualitatif, la méthode RFR a engendré des marges d'erreur plus grandes que TsHARP pour les expériences dans les régions de la Capitale nationale et de l'Estrie à 100 m. Cependant, pour l'expérience dans la Montérégie, sa marge d'erreur reste plus faible que TsHARP. La méthode PBIM a engendré des marges d'erreur plus grandes que les autres méthodes dans les trois régions étudiées.

Pour la réduction d'échelle directe de 1000 m à 30 m, la méthode RFR a produit des marges d'erreur plus faibles que TsHARP. Nous avons également constaté que dans les zones hétérogènes à dominance boisée ou agricole, tous les pixels à basse résolution n'ont pu être réduits. Ainsi des écarts de valeurs de LST de la classe

[2°C et plus] ont été observés dans les zones de végétation ou agricole. Ces écarts de LST se retrouvent également plus dans les zones urbaines où leurs réelles valeurs LST n'avaient pas été mesurées par MODIS, mais révélées par la Landsat 08 TIRS à cause de sa résolution spatiale. En revanche, la majorité des pixels des milieux urbains ont été bien décomposés dans les images réduites. En Montérégie et en Estrie, les valeurs de LST réduites se rapprochent mieux de la température de l'air. À Québec, ce sont les LST de Landsat 08 TIRS qui valident mieux les résultats.

Les résultats présentés suscitent des questions sur leur qualité intrinsèque, l'intérêt de l'application de ces méthodes de réduction d'échelle, le bien-fondé de viser une résolution spatiale de 30 m. Ces résultats donnent également lieu à s'interroger sur le processus méthodologique proposé, de l'utilité de recourir aux méthodes de réduction d'échelle dans le contexte du projet de l'ASPC et enfin les implications auxquelles on doit s'attendre avec la généralisation du processus sur tout le Canada. La réponse aux points cités donne lieu à des analyses et discussions dans le chapitre suivant.

# Chapitre 5 - Analyse des résultats et discussions

## 5.1. Introduction

L'objectif général de cette étude visait à démontrer la faisabilité et la pertinence de réduire l'échelle de données MODIS de 1000 m à 100 m et 30 m pour déduire la température de surface avec une marge d'erreur de 2°C. De manière spécifique, il s'agit d'évaluer à partir de la qualité des résultats obtenus, la pertinence de faire une réduction d'échelle de MODIS LST à 30 m au lieu de 100 m, d'explorer les conditions et les contraintes liées à l'application de la méthode de réduction d'échelle sur de grands volumes de données. Afin d'atteindre ces objectifs, l'hypothèse émise est qu'une approche méthodologique de réduction d'échelle basée sur les émissivités de surface améliorerait les prédictions de la température de surface à la résolution spatiale de 30 m. Pour tester cette hypothèse, des expérimentations ont été faites dans trois zones de morphologies différentes avec des données géospatiales thermiques de 1000 m, des indices spectraux et données topographiques à haute résolution spatiale de 30 m. Les données utilisées sont des mois de juillet 2013, 2017 et août 2019 de la saison d'été.

Les résultats de réduction d'échelle obtenus à 100 m montrent une marge d'erreur au-dessus de 2°C sur l'ensemble du territoire de chaque zone d'étude, quelle que soit la méthode de réduction d'échelle (Tableau 5.1). C'est le cas aussi avec la réduction à 30 m où les marges d'erreur sont au-dessus de 2°C (Tableau 4.23). En revanche, à des points d'observations de stations météorologiques en zone forestière, les écarts de température sont autour de 2°C (Tableau 5.2). Les sections qui suivent viennent discuter de ces résultats selon différents points de vue :

- (Q1) Quelle plus-value avons-nous avec ces résultats obtenus à partir des méthodes de réduction d'échelle ?
- (Q2) Est-ce que les résultats obtenus peuvent répondre aux besoins de précision des outils de modélisation pour la cartographie du risque lié aux maladies vectorielles associées aux tiques et moustiques ?
- (Q3) Auxquels enjeux serons-nous confrontés si la réduction d'échelle de LST de 1000 m à 100/30 m est réalisée sur l'ensemble du territoire canadien ?

Ces 3 questions sont répondues dans les 3 sous-sections suivantes.



Tableau 5.1 : Récapitulatif des indicateurs de qualité calculés sur le territoire entier des trois zones d'étude pour la LST à 100 m

Z E	RMSE °C			EMA °C		
	TsHARP	RFR	PBIM	TsHARP	RFR	PBIM
RCN (juillet 2013)	2,57	2,8	3,39	2,3	2,46	2,66
RE (Juillet 2017)	2,75	2,82	3,26	2,49	2,54	2,73
RMT (Août 2019)	3,76	3,71	5,55	3,37	3,31	4,18
<b>MOY</b>	<b>3,03</b>	<b>3,11</b>	<b>4,07</b>	<b>2,72</b>	<b>2,77</b>	<b>3,19</b>

ZE : zone d'étude ; RCN : région de la Capitale nationale, RE : région de l'Estrie, RMT : région de Montérégie

Tableau 5.2 : Récapitulatif des indicateurs de qualité calculés à des stations météorologiques en zone forestière des trois zones d'étude pour la LST à 100 m

Z E	ET °C (MODIS LST-Landsat 100 m)			ET °C (MODIS LST 100 m-TA)		
	TsHARP	RFR	PBIM	TsHARP	RFR	PBIM
RCN (juillet 2013)	1,40	0,08	1,05	-2,23	-3,56	-2,59
RE (Juillet 2017)	2,59	2,73	2,99	-1,58	-1,44	-1,18
RMT (Août 2019)	2,49	2,46	2,81	-2,53	-2,57	-2,21
<b>MOY</b>	<b>2,16</b>	<b>1,75</b>	<b>2,28</b>	<b>-2,11</b>	<b>-2,52</b>	<b>-1,99</b>

ZE : zone d'étude ; RCN : région de la Capitale nationale, RE : région de l'Estrie, RMT : région de Montérégie

## 5.2. Qualité des résultats

Les différences observées dans les valeurs de RMSE/EMA peuvent s'expliquer d'une part par les effets d'échelle survenant généralement dans les processus de changement d'échelle. En effet, les changements d'échelle peuvent engendrer des effets d'échelle d'un point de vue de l'analyse, du traitement et de la modélisation, car une même cible à différentes échelles présentera des caractéristiques différentes (Wu & Li, 2009). De plus, selon Wu et Li (2009), la combinaison des caractéristiques de linéarité ou de non-linéarité des modèles et de l'hétérogénéité des surfaces engendre des effets d'échelle. De fait, une canopée homogène par exemple peut toujours entraîner une hétérogénéité de la température de surface en raison de l'ombre, tel que le montre le Tableau 5.2 avec les marges d'erreur. D'autre part, le facteur de réduction d'échelle qu'est le rapport entre la résolution spatiale des images natives et prédites peut aussi occasionner des erreurs lors du processus de réduction d'échelle (Agam et al., 2007). Le facteur d'échelle (1/10 si l'on passe de 1000 m à 100 m, 3/10 de 100 m à 30 m et de 3/100 de 1000 m à 30 m) dans cette étude peut expliquer la variabilité et représentativité spatiale des valeurs de LST des images réduites. Les valeurs élevées des incertitudes calculées à Montérégie

où, dominant les unités de surface agricole peut s'expliquer par la présence d'humidité de surface (Agam et al., 2007).

Nous répondons dans cette section à la question Q1 en deux sous-sections portant sur la précision spatiale et la justesse des résultats. Il s'agit de comparer nos résultats à 100 m et à 30 m et de les confronter avec ceux d'autres études relevées dans la revue de littérature.

### 5.2.1. Précision spatiale des résultats de température de surface LST

De manière générale, nous remarquons que les marges d'erreur de MODIS LST réduite à 100 m ou 30 m est quasi identique et se situent dans la marge d'erreur de 2°C (Tableau 5.2 et pour les RCN et de RE avec les méthodes TsHARP et RFR) constatée dans la revue de littérature (2,42°C en moyenne) pour tout type de méthode ou pour la LST mesurée dans l'espace. Cette précision ne peut être meilleure que celle de MODIS originale. En effet, pour la MODIS LST (daytime) du jour utilisée dans cette étude, la RMSE associée à sa précision est de 1,3 Kelvin sur des sites homogènes (Duan et al., 2019). Dans une autre étude, une analyse sur les températures moyennes journalières a donné une RMSE de 2,46°C et 4,90 °C pour les températures maximales journalières (Morin et al., 2018b). Nos marges d'erreur tiennent dans ces proportions.

De plus, la comparaison des résultats obtenus avec les données de référence à 100 m ou 30 m, Landsat 08 LST et la température de l'air, révèle que l'écart moyen dans la marge d'erreur est faible ; elle est autour de 1°C pour la température de l'air et de +/- 2°C pour Landsat 08 LST. De fait, le niveau de précision spatiale de LST entre MODIS LST et Landsat LST est similaire. Ce qui nous amène à affirmer que hormis, la haute résolution temporelle de MODIS (un jour au plus), la Landsat 08 LST peut valablement lui être substituée d'autant plus que les images Landsat 08 LST couvrent une superficie précise avec plus de détails spatiaux.

S'agissant de la réduction d'échelle de 100 m à 30 m, la marge d'erreur constatée à 100 m a été conservée à 30 m et se tient à l'intérieur de la marge d'erreur des 2°C (2,50 °C, 2,77°C, 2,88°C ; Tableau 5.2). Cependant, la meilleure précision spatiale de LST attendue à 30 m n'a pas été constatée. Partant de là, il n'y a pas un gain substantiel entre les images réduites à 100 m et à 30 m et donc, le recours à la réduction d'échelle pour prédire la LST à 30 m ne semble pas concluant.

Au travers de ces résultats, les méthodes de réduction d'échelle ont démontré leur capacité à discriminer la LST dans l'espace et n'ont cependant pas apporter plus de précisions spatiales que les images originales brutes. Pour évaluer la performance des méthodes appliquées notamment celles de RFR et de TsHARP, nous faisons une comparaison des RMSE avec les études existantes ayant utilisé les mêmes méthodes.

### 5.2.2. Justesse des résultats

L'analyse ici consiste en la comparaison de nos résultats avec ceux de différentes études pour en ressortir les gains substantiels. Le Tableau 5.3 et le Tableau 5.4 présentent nos résultats et ceux obtenus dans des études antérieures ayant appliqué respectivement la même méthode RFR et TsHARP. Ces études ont été réalisées dans les milieux hétérogènes sous différents climats et conditions environnementales différentes. Les informations sur la méthode adoptée, le type et la résolution observée, la résolution spatiale de l'image réduite, la RMSE, la source de l'étude, la superficie de la zone d'étude et le nombre de variables y sont décrites.

Tableau 5.3: Présentation des RMSE liées à la RE de LST selon la méthode RFR dans certaines études

Méthodes	Image/Résolution	Résultat (m)	RMSE (°C)	Sources	Superficie km <sup>2</sup>	Nbr. variables
<b>RFR</b>	<b>MODIS 1000 m</b>	<b>100</b>	<b>2,80</b>	<b>Notre étude</b>	<b>6 780</b>	<b>4</b>
<b>RFR</b>	<b>MODIS 1000 m</b>	<b>100/30</b>	<b>2,82/2,77</b>	<b>Notre étude</b>	<b>9 422</b>	<b>5</b>
<b>RFR</b>	<b>MODIS 1000 m</b>	<b>100</b>	<b>3,71</b>	<b>Notre étude</b>	<b>6 047</b>	<b>4</b>
RFR	MODIS 1000 m	240	2,25°C	Ebrahimi & Azadbackht, 2019	5 170	11
RFR	MODIS 1000 m	240	3,13°C	Ebrahimi & Azadbackht, 2021	6 400	6
RFR	MODIS 1000m	500	2,42°C	Yang et al, 2017	NA	5
RFR	MODIS 1000m	500	2.60 °C	Yang et al, 2017b	NA	7

Le Tableau 5.3 révèle que peu d'études ont essayé d'aboutir à une résolution cible de 100/30 m avec l'image MODIS LST avec la méthode RFR. Les résolutions spatiales visées vont de 240 m à 500 m, avec une valeur moyenne supérieure de 2,4°C par rapport au seuil de référence de 2°C. Bien que le facteur d'échelle de la présente étude (1/10 si l'on passe de 1000 m à 100 m, 3/10 de 100 m à 30 m et de 3/100 de 1000 m à 30 m) soit deux fois supérieur à celui de ces études (1/5 en passant de 1000 m à 500 m et de 6/25 pour 1000 m à 240 m), les marges d'erreur sont sensiblement égales. Ainsi, en référence aux valeurs de RMSE des autres études, nous affirmons que nos résultats sont en concordance avec ceux des études antérieures.

De plus, le nombre de variables prédictives utilisées varie de 5 à 11 dans les études antérieures contre 4 à 5 dans la présente étude. Outre les indices spectraux et l'altitude, les autres variables introduites dans leur modèle sont l'occupation et l'utilisation du sol, les bandes spectrales de Landsat 08/OLI et le facteur solaire. Il est fort probable d'avoir une faible marge d'erreur dans les images réduites en introduisant ces prédicteurs. Toutefois, précisons que le nombre de prédicteurs ne garantit pas forcément plus de précision bien que, selon Zhan et al., (2012) et Zhou et al., (2016), les marges d'erreur sont sensibles au nombre de variables explicatives et de la résolution cible. De fait, il est suggéré d'accorder de l'importance à la rationalité des variables d'entrée plutôt qu'à la quantité en se basant sur les scores d'importance des corrélations relatives (Wu & Li, 2019). En ce qui concerne la résolution cible, en se basant sur le Tableau 5.3, la marge d'erreur est quasi similaire et est autour de 2°C.

Tableau 5.4 : Présentation des RMSE liées à la RE de LST selon la méthode TsHARP dans certaines études

Méthodes	Image/Résolution	Résultat (m)	RMSE (°C)	Superficie (km <sup>2</sup> )	Sources
<b>TsHARP</b>	<b>MODIS 1000</b>	<b>100</b>	<b>2,57</b>	<b>6 780</b>	<b>Notre étude</b>
<b>TsHARP</b>	<b>MODIS 1000</b>	<b>100/30/30</b>	<b>2,75/2,88/2,77</b>	<b>9 422</b>	<b>Notre étude</b>
<b>TsHARP</b>	<b>MODIS 1000</b>	<b>100</b>	<b>3,76</b>	<b>6 047</b>	<b>Notre étude</b>
TSHARP	MODIS 1000	240	3,02	5 170	Ebrahimi & Azadbackht, 2019
TsHARP	MODIS 1000	60	8,3	NA	Agam et al 2007
TsHARP	MODIS 1000	500/ 250	1,47/ 1,59	34 225	Mukherjee et al., 2014
TsHARP	Landsat 8 100	90	1.14	6 587	Yang et al., 2017
TsHARP avec variante Locale	MODIS 1000	500 /250	1.45/ 1,57	34 225	(Mukherjee et al., 2014)
TsHARP : 4 Var.	Landsat 08 100	30	1,5	NA	Govil et al, 2019

Il ressort du Tableau 5.4 que la réduction d'échelle selon la méthode TsHARP a donné des résultats avec des RMSE en deçà de 2°C, excepté l'étude de Agam (2007) dans laquelle l'écart est de 8,3 °C. Cet écart élevé serait lié au facteur d'échelle de 17 (passant de 1000 m à 60 m) de la réduction d'échelle et à l'image MODIS LST 1000 m utilisée. Les autres études qui ont exploité MODIS LST 1000 m ont visé 500 m ou 250 m en résolution spatiale. En revanche, les études de Yan et al. (2017), Govil et al., (2019) ont utilisé les images thermiques de Landsat 08 (100 m de résolution spatiale) pour atteindre respectivement 90 m et 30 m avec une RMSE respective de 1,14°C et 1,5°C. Hormis l'étude de Govil, qui a utilisé 4 prédicteurs, les autres études ont été basées sur une variable, le NDVI ou la CVF. La comparaison de ces études avec cette recherche révèle que nos résultats malgré une valeur élevée de RMSE sont satisfaisants et corroborent ceux des autres études.

En définitive, la revue de littérature réalisée présente les modèles de réduction d'échelle comme un moyen de prédiction des LST à de fines résolutions et ont fait leurs preuves dans diverses études appliqués surtout en agroéconomie, hydrologie, etc. Cependant, dans le cas de cette étude où le besoin est de discriminer spatialement la LST pour enfin définir les déterminants climatiques et microclimatiques, les résultats obtenus par réduction d'échelle à 100 m/30 m pour les 3 cas d'étude sont intéressants et tiennent dans la marge d'erreur de 2°C ; ce qui démontre de la capacité des méthodes à discriminer la LST dans l'espace. La précision ne saurait être plus fine que celle-là à moins d'utiliser d'autres moyens de mesure en surface comme les drones. Mais, du point de vue de résolution spatiale, les résultats de réduction d'échelle sont non concluants, car que ce soit à 100 m ou à 30 m la précision est quasi identique (c'est le cas aussi avec la Landsat LST à 100 m ou à 30 m). Cette conclusion peut être justifiée par une agrégation d'échelle simple de 100 m à 30 m sans appliquer aucune méthode de réduction d'échelle. La réduction d'échelle de 100 m à 30 m n'a pas apporté une plus-value conséquente à la précision spatiale et par conséquent, il n'est pas pertinent de faire le passage de 100 m à 30 m.

### 5.3. Utilité des résultats de LST réduites

Dans cette section, nous répondons à la question Q2 posée dans la sous-section 5.1 portant sur la précision des outils de modélisation pour la cartographie du risque liée aux maladies vectorielles. Dans cette analyse, le contexte d'étude a été pris en compte pour apprécier les résultats obtenus.

Pour arriver à la cartographie de risque du Virus du Nil Occidental et la maladie de Lyme, il faudrait identifier les zones potentiellement à risque dans les micro-habitats où règnent des conditions microclimatiques en se basant sur des critères de classification de l'occupation du sol. Les zones de prédilection des moustiques et des tiques sont les biomes forestiers, les zones humides, les zones agricoles ou suburbaines à dominance forestières. Par contre, les zones urbaines et suburbaines ne sont pas les zones potentielles de rétention des tiques et moustiques, bien qu'on puisse les y retrouver du fait par la prolifération avec les hôtes (oiseaux, rongeurs, animaux...). Dès lors, ces zones urbaines et suburbaines ne sont pas représentatives dans l'identification des habitats des vecteurs. Or, selon les résultats de réduction d'échelle de LST à 100 m ou à 30 m dans les 3 zones d'expérimentation, les zones d'imprécision spatiale de LST (les marges d'erreur RMSE sont supérieures/inférieures à  $[-2 ; 2 \text{ °C}]$ ) (en valeur absolue  $[0-2\text{°C}]$ ) sont surtout ces zones forestières, les zones agricoles hétérogènes. De fait, les résultats dans leur état actuel limitent la possibilité de leur application. En conséquence l'objectif n'est pas totalement atteint en utilisant la méthode de réduction d'échelle ou encore, il faut améliorer les résultats. Par exemple, pour améliorer ces résultats pour être plus proche de  $2\text{°C}$ , on peut tester les méthodes de réduction d'échelle en utilisant d'autres variables prédictives comme les unités d'occupation du sol, les ombres, l'albédo de surface, l'angle d'incidence du rayonnement solaire ou encore en faisant d'autres combinaisons de variables prédictives afin de mesurer leur sensibilité sur les résultats.

Cependant, si les résultats ne s'amélioreraient (précision spatiale de  $2\text{°C}$ ) pas en dépit du recours à d'autres variables prédictives et en l'absence d'une plus-value de LST ( $\text{°C}$ ), il est nécessaire de trouver d'autres solutions pour disposer des données à haute résolution spatiale et temporelle. Une solution possible est de trouver d'autres critères ou informations supplémentaires discriminants pour l'identification graduelle des zones d'intérêt. Une autre solution possible serait de se concentrer sur des zones d'imprécision spatiale révélées par réduction d'échelle et utiliser en complément les drones par exemple pour avoir plus de détails et précisions au centimètre près ou utiliser uniquement les drones au détriment de la méthode de réduction d'échelle. En effet, depuis la conception du projet par l'Agence de santé publique de Canada en 2016 à ce jour, il y a des évolutions technologiques comme le drone qui peuvent être explorées. Il y a aussi des améliorations spatiales au niveau

des capteurs imageurs des données météorologiques tel le Visible Infrared Imaging Radiometer Suite (VIIRS)<sup>22</sup> qui fournit des images LST de 750 mètres de résolution spatiale et d'au plus un jour de résolution temporelle. En résumé, les résultats de LST obtenus par réduction d'échelle n'apportent pas plus d'informations dans la discrimination spatiale des milieux forestiers, agricoles, ruraux à dominance forestière, susceptibles de renfermer les tiques et moustiques. Dès lors, la détection des déterminants climatiques et microclimatiques (DCMC) ne saurait être effective sans recourir à d'autres mesures de LST avec outils plus performants (très haute résolution spatiale comme les drones).

## 5.4. Enjeux à la réduction d'échelle

Relativement à la question Q3 de la section 5.1, nous abordons ici les enjeux temporels, spatiaux et des enjeux méthodologiques.

### 5.4.1. Enjeux temporels

La différence de marge d'erreur (RMSE) constatée selon le mois d'étude peut s'expliquer par la variabilité des valeurs de température de chaque zone et au cours des mois d'été. En effet, dans la province de Québec, le mois de juillet est habituellement celui le plus chaud alors que le mois d'août l'est moins (cf. les images de la Figure 0.10 en annexe A). Ce qui se vérifie sur les images de MODIS LST 1000 m originales, où les étendues entre les valeurs de LST sont de 16,2°C, 14,02°C et 12,46°C respectivement pour la Région de Capitale nationale, l'Estrie et la Montérégie. Ainsi, toutes les méthodes ont été sensibles à la faible variation de température du mois d'août où la LST oscille entre 21,33°C et 33,79°C. Les résultats de cette étude contrastent avec ceux de la littérature rapportant des sensibilités de réduction d'échelle selon les saisons, mais pas spécifiquement au sein d'une même saison. Par exemple, l'étude de Ebrahimi & Azadbakht, (2019), menée dans une zone hétérogène variée dans la province de Téhéran en Iran a donné des marges d'erreur quasi identiques pour toutes les 3 saisons en 2016 (printemps -20 juin-, été -7 août- et automne -24 novembre-) et pour chacun des modèles (TsHARP, le RFR, ELM et SVR). Par contre, les tests de Yang et al., (2017) sur six sites du bassin de Heihe dans la ville de Zhangye (Chine) avec les méthodes de RFR (multivariable MIRF), RFR de base, DisTrad, ont donné de meilleurs résultats en automne plus qu'en été, le printemps et l'hiver. Dans une autre étude menée dans la zone hétérogène urbaine de Nanjing (en Chine) avec l'algorithme MSFAT, l'été (13 août) a donné des résultats plus précis contrairement aux 3 autres saisons (printemps -28 mars-, Automne

---

<sup>22</sup> <https://earthdata.nasa.gov/earth-observation-data/near-real-time/download-nrt-data/viirs-nrt>

– 14 octobre- et hiver -20 décembre). **L'enjeu temporel** de cette variabilité de LST est d'identifier et retenir un mois unique ou plusieurs mois comme étant celui/ceux dont les données sont susceptibles d'aboutir à un meilleur résultat de réduction d'échelle pour tout type de couverture au sol.

**L'enjeu temporel** concerne aussi l'ensemble des données utilisées dans la réduction d'échelle de LST. Par exemple, à l'échelle du Canada dans une même journée, nous pouvons avoir les images MODIS LST, ce qui n'est pas le cas avec les images Landsat 08 OLI/TIRS dont les prises de vue se font à différentes dates. Par exemple pour le mois de juillet 2015, le même capteur Landsat qui prend une image de la région de la Capitale nationale le 2 juillet 2015 à 15h 31, poursuit son chemin pour photographier la région de Montérégie le 09 juillet 2015 à 15h 37 puis la région de Montréal à le 16 juillet 2015 à 15h 44. Soit un décalage de 7 jours entre ces images. Or, les données de Landsat 08 OLI servent aux calculs des prédicteurs tandis que celles de Landsat 08 TIRS permettent de valider les résultats. Si le décalage de date impacte peu l'établissement de la relation LST et les prédicteurs à cause de l'invariabilité des états de surface sur une courte durée, il affecte néanmoins l'évaluation des résultats de MODIS LST réduite à cause de la grande variabilité spatio-temporelle de LST au cours d'une même journée. Si nous considérons le besoin en données de LST au pas de temps degré jour et hebdomadaire pour un suivi respectif des tiques et moustiques, alors ce décalage temporel soulève la préoccupation de validation des résultats de LST avec la Landsat 08 LST. À défaut de disposer des images MODIS LST et Landsat LST de la même date pour une comparaison, il faut mettre en place des stratégies de sélection par zone jusqu'à couvrir tout le territoire canadien. Ce choix par zone pourrait être une contrainte et non bénéfique en cas de situation d'urgence ou de catastrophe d'épidémie nécessitant le suivi en temps réel.

Une alternative à ce décalage temporel, c'est de recourir aux **données de température de l'air disponibles au pas de temps horaire ou journalier** pour valider les résultats de LST réduites. Bien qu'au plan spatial, il y aura une perte de l'information sur la variabilité spatiale de la température de surface, au moins les stations d'observation de la température de l'air au sol serviront à valider les résultats en divers points sur l'ensemble du Canada.

Si nous considérons, la période de mise à jour des cartes finales de risques liés aux maladies (annuelle pour les moustiques et mensuelle pour les tiques) et vu qu'en un mois ou une année, les états des surfaces varient peu au niveau des affectations (une zone de forêt transformée en zone urbaine), alors les variables de prédiction biophysique (NDVI, MNDWI, NDBI, émissivité de surface...) peuvent être établies pour la saison estivale ou à une date « D » et exploitées pour l'année.

#### 5.4.2. Enjeux spatiaux

Nous avons constaté que les valeurs de LST des zones nues, bâties et imperméables dans les trois zones d'étude ont été bien réduites contrairement aux zones de végétation, de cultures et de milieux humides où l'ensemble des pixels n'a pu être simulé. Ceci pourrait s'expliquer par la présence d'humidité au sol dans ces unités. Dans la zone d'étude de région de l'Estrie, selon la classification réalisée, le territoire est occupé par 7% de zone artificialisée, 2% de zone humide, 63% de forêt et 28% de végétation. Parmi ces unités de couverture, près de 90% de pixels de végétation se retrouvent dans la classe [0-; 2°C] (cf. image 4 de la Figure 4.15), contre 80 % pour les zones humides, 70% des zones artificialisées et 50% de zone forestière. La majorité des pixels hors de la classe [0- 2°C] sont localisés dans des zones de faible pente (moins de 10°) et forte pente (plus de 35°). Dans la zone d'étude de la région de la Capitale nationale, où on retrouve 6% d'eau de surface, 1% de zone humide, 68% de forêt, 18% de végétation et 7% de zone artificialisée, les zones forestières sont celles qui présentent le taux de précision le plus élevé, car seulement 40% de ses pixels sont de la classe [0 - 2°C] (cf. image 3 de la Figure 4.4). En revanche, sur le site étudié dans la région de Montérégie, l'occupation du sol constituée de 55% de végétation (dont 44% de zones agricoles), 24% de forêt, 9% de milieu urbain/bâti ; 8% d'eau de surface, 3% de terres humides et 1% de sols nus, les marges d'erreur les plus grandes ont été engendrées dans les zones agricoles (cf. image 2 de la Figure 4.12).

Nos résultats se rapprochent de ceux d'autres études réalisées sur des sites similaires. Les études de Yang et al. (2017) dans la zone hétérogène urbaine de Nanjing, les zones de végétation ont donné de faibles incertitudes comme celles observées dans les régions de l'Estrie et de la Capitale-Nationale. Les recherches de Bisquert et al. (2016) dans la ferme expérimentale « Las Tiasas » dans le Barrax en Espagne ont montré que les erreurs les plus importantes observées dans les terres irriguées et les champs de culture augmentent significativement dans les images d'été (Bisquert et al., 2016). Bien que la réduction d'échelle dans les zones à dominance agricole sur les hautes plaines du Texas ait donné de bons résultats dans les surfaces agricoles par rapport aux zones de végétation et d'eaux libres, les erreurs sont plus importantes pour la classe agricole dans les images du septembre (Agam et al., 2007). Les grandes erreurs relevées dans les zones forestières pour les images de juillet et les zones agricoles des images d'août peuvent s'expliquer par la présence de l'humidité du sol ou de l'humidité des canopées par phénomène de transpiration dans les végétations saines. De fait, les terres cultivées sont plus humides à cause de l'irrigation (Bisquert et al., 2016 ; Guijun Yang et al., 2010) et les pixels de végétation naturelle sont plus humides que le sol nu. L'autre explication serait liée à la sensibilité des modèles aux valeurs extrêmes ou aux faibles variations de valeurs de LST comme dans la région de Montérégie où elles fluctuent de 21,33°C et 33,79°C.



**Sans considérer les marges d'erreur dans chaque unité d'occupation du sol**, soulignons que les zones dans lesquelles il y a plus d'imprécisions dans des résultats de LST réduite (zones de végétation, zones rurales, de cultures et humides) sont celles susceptibles d'abriter les tiques et les moustiques. Il s'ensuit :

- qu'on ne peut avoir plus précisions spatiales de LST dans ces zones et dans lesquelles il faudrait utiliser d'autres moyens (ajouter des critères de discrimination) ou des outils comme les drones pour acquérir des valeurs de LST plus précises.
- à l'échelle du territoire canadien avec les gros volumes de données, au lieu d'appliquer les méthodes de réduction d'échelle sur l'ensemble des unités d'occupation du sol, on peut les appliquer uniquement sur ces unités d'occupation d'intérêt et représentatives. Ce qui pourrait apporter une plus-value **en temps de calcul, ressources et coût**.

L'enjeu spatial à l'échelle du Canada est également relatif aux prédicteurs. En effet, le choix des prédicteurs à utiliser pour une réduction d'échelle de LST varie en fonction des caractéristiques environnementales de chaque zone. Ce choix tient aussi compte du degré de corrélation entre la LST et les prédicteurs. Avec la généralisation des traitements, il se posera la difficulté d'assigner à chaque zone ses prédicteurs intrinsèques. Face à cette probable difficulté, des stratégies de traitement doivent être définies. On peut opter pour obtenir:

- des traitements par région physiographique ou par région administrative si l'on considère le Canada. Une des solutions serait l'utilisation des méthodes d'intelligence artificielle comme apprentissage profond (Deep Learning) ou la régression de forêt aléatoire (RFR) pour procéder à une analyse préalable des spécificités de chaque zone et y affecter les prédicteurs adaptés ;
- des traitements unité par unité d'occupation (uniquement sur les zones forestières, ou hétérogènes agricoles, zones humides.) à l'échelle du Canada ;
- des traitements focalisés uniquement sur les zones représentatives et d'intérêt préalablement identifiées comme celles susceptibles de contenir les tiques et moustiques. Ce mode de traitement paraîtrait efficace si des modélisations sont faites au préalable pour identifier et connaître les caractéristiques des habitats des vecteurs. Ainsi par des modèles d'apprentissage machine, ces caractéristiques seront identifiées dans l'ensemble des unités d'occupation et d'utilisation des sols.

#### 5.4.3. Enjeux méthodologiques et du choix des méthodes

Les résultats obtenus ne permettent pas de dégager une méthode entre TsHARP et RFR comme étant la meilleure à adopter pour être appliquée sur toute l'étendue du territoire canadien. Il est nécessaire de les expérimenter sur d'autres types de couverture du sol à différentes dates avant de tirer une conclusion définitive. Néanmoins nous suggérons leur application selon différentes caractéristiques environnementales et climatiques de la zone d'étude. Il est également question de la validation de l'hypothèse de recherche et du cadre méthodologique proposé.

#### 5.4.3.1. Stratégies de sélection d'une méthode de réduction d'échelle

À travers le Tableau 5.1 sur les RMSE/EMA, le Tableau 5.5 sur la variabilité/topographie de la LST et la Figure 0.6, Figure 0.7, Figure 0.8 et la Figure 0.9 sur les statistiques descriptives des écarts d'erreurs en annexe A, des sites étudiés, nous évaluons la performance des méthodes en fonction des spécificités morphologiques. Nous constatons que pour la réduction d'échelle à 100 m, la méthode TsHARP a donné des résultats ayant les plus faibles RMSE/EMA dans les zones hétérogènes à dominance forestière comme en Estrie et à la Capitale nationale. En revanche, en Montérégie où la couverture du sol est à dominance agricole, c'est la méthode RFR qui donne des résultats plus précis. Ces résultats peuvent être également liés à la topographie et la variation de la température. Nous constatons que la zone de Montérégie est un milieu de plaine où la pente est faible (une moyenne 8,76° degré) avec une altitude moyenne de 57 m et de surface uniforme. Par contre, dans la région de l'Estrie, la pente moyenne est de 10° degré avec une altitude moyenne élevée de 288 m et marquée par de fortes dénivelées. À l'image de la région d'Estrie, le relief de la région de la Capitale-Nationale est caractérisé par une pente forte dont la moyenne est de 15° avec une altitude moyenne d'environ 400 m.

Tableau 5.5 : Description de la température et la topographie des zones d'étude

Z E	MODIS LST 1000m			Relief	
	Min °C	Moy °C	Max °C	Altitude (m)	Pente(°)
RCN (juillet 2013)	19.40	22.02	35.60	400	15
RE (Juillet 2017)	14.86	21.44	32.48	288	10.07
RMT (Août 2019)	21.33	26.01	33.79	57	8.76

ZE : zone d'étude ; RCN : région de la Capitale nationale, RE : région de l'Estrie, RMT : région de Montérégie

S'agissant de la variation de la LST, les valeurs de MODIS LST 1000 m connaissent peu de fluctuation dans la région de Montérégie avec une étendue de 12,46°C contrairement à Estrie et dans la Capitale-Nationale où l'amplitude respective est de 14,02°C et 16,2°C.

Ces caractéristiques par zone d'étude peuvent constituer un référentiel pour orienter dans le choix d'une méthode de réduction d'échelle entre TsHARP et RFR. Par exemple, une préférence peut être donnée à la méthode TsHARP pour toute zone présentant un profil environnemental similaire à celui de la région d'Estrie ou celui de la région de la Capitale nationale. En revanche, la méthode RFR aura la primauté si la zone présentait les mêmes similitudes que la région de Montérégie. Nous précisons que les deux méthodes TsHARP et RFR ou d'autres peuvent être employées selon des profils environnementaux différents.

Pour la réduction d'échelle à 30 m, les expérimentations n'ont été faites que dans la région de l'Estrie et ne permettent d'avoir un avis tranché sur la performance des méthodes en fonction des spécificités spatio-temporelles. Les résultats obtenus en deux étapes (de 1000 m à 100 m puis de 100 m à 30 m) ou directement de 1000 m à 30 m pour la région de l'Estrie montrent que la méthode de RFR présente la meilleure précision (RMSE/EMA : 2,50°C/2,51°C) pour le processus en une seule étape (cf. Tableau 4.23). Toutefois, les paramètres statistiques de dispersion et valeurs centrales décrits dans le tableau de la Figure 0.6 en annexe A révèlent qu'il n'y a pas une grande différence entre les méthodes et l'approche méthodologique à 2 étapes.

#### *5.4.3.2. Validation de l'hypothèse de recherche et du cadre méthodologique*

L'hypothèse émise est qu'une approche méthodologique de réduction d'échelle basée sur des émissivités de surface améliorerait les prédictions de la température de surface à 30 m. Les méthodes testées dans cette étude ont effectivement permis de prédire les MODIS LST à 100 m et 30 m avec des marges d'erreur semblables à celles rapportées dans d'autres études pour des résolutions de 240 m ou 500 m en utilisant toujours MODIS LST 1000 m. Les faibles incertitudes observées avec TsHARP par rapport à RFR (Tableau 5.1) sont à l'opposé de ce qui est rapporté dans la littérature. En effet, pour la plupart des études (Ebrahimi & Azadbakht, 2019 ; Peng et al., 2019 ; Yang, Li, et al., 2017), les modèles statistiques non linéaires (RFR, ELM, SVR, GTWR ou MSFAT) comparés au modèle statistique linéaire (TsHARP) apportaient plus de précisions. Ainsi, l'étude de Ebrahimi & Azadbakht, (2019), menée dans la province de Téhéran en Iran dans une zone hétérogène à dominance forestière avec les méthodes de TsHARP, RFR, ELM, SVR ont donné comme RMSE respective, 3,02°C, 2,25°C, 2,53°C et 2,56°C. Quant à Peng et al., (2019), ils ont testé les méthodes GTWR et TsHARP pour réduire MODIS LST 1000 m à 100 m d'un milieu urbain hétérogène et ont eu comme RMSE respective 1,62°C et 2,85°C. Ces résultats contradictoires avec les nôtres s'expliquent par les variables explicatives utilisées. Dans cette étude, le même nombre de variables explicatives (indices spectraux et topographiques) caractérisant le mieux la zone d'étude a été utilisé. Tandis que les études citées n'ont pas adapté la méthode TsHARP et ont utilisé le NDVI ou la CVF, comme seule variable explicative. Cette différence constatée est une plus-value pour les méthodes statistiques linéaires de réduction d'échelle.

La seconde étape de la réduction d'échelle de 100 m à 30 m, si elle a permis d'avoir plus de détails spatiaux dans les zones construites avec une classe de LST de plus que la LST prédite à 100 m, sa marge d'erreur (RMSE/EMA: 2,88°C/2,55°C) est plus grande que celles obtenues à 100 m (RMSE/EMA : 2,75°C/2,49°C). Toutefois, en comparant les marges d'erreur à 30 m à celles obtenues à 240 m dans le Tableau 5.3 avec les méthodes de RFR, nous pouvons affirmer que les émissivités de surface ont amélioré les LST à l'échelle réduite.

En revanche, si cette comparaison est faite avec les résultats en une seule étape de 1000 m à 30 m (sans les émissivités de surface parmi les variables explicatives) pour laquelle la RMSE/EMA donne 2,77°C/2,51°C (TsHARP) ou 2,50°C/2,77°C (RFR), nous pouvons dire que les émissivités n'ont pas véritablement d'effets sur le résultat final à 30 m. Cette conclusion serait définitive si des expériences de 1000 m à 30 m en une étape intégraient les émissivités de surface dans les variables explicatives et que les marges d'erreur se révèlent toujours plus faibles.

#### 5.4.4. Enjeux de la gestion des gros volumes de données

L'application de la méthode de réduction d'échelle à tout le territoire canadien signifie la manipulation et le traitement d'un grand nombre d'images satellitaires multispectrales variées, volumineuses avec des périodes de mise à jour courtes et régulières. En conséquence, la recherche de l'efficacité et l'efficience font émerger l'enjeu de la gestion de grande quantité de données géospatiales. La manipulation de gros volumes de données suscite des interrogations.

##### 5.4.4.1. Définition des gros volumes des données

La réduction d'échelle utilise des données géospatiales constituées en majorité des images satellitaires volumineuses. Par exemple, la procédure de réduction d'échelle appliquée sur une zone de la région de la Capitale nationale basée sur des variables prédictives a nécessité un jeu de données de 13 images satellitaires (7 bandes spectrales pour les données de réflectance et 1 (une) bande de qualité ; deux images satellitaires de MODIS LST dont la bande de qualité et 3 bandes pour les données topographiques) volumineux de près de 1 Giga-octet. Pour la validation des résultats, deux bandes thermiques de Landsat 08 et la bande de qualité, puis les bandes optiques 5 et 4 ont été utilisées. La capacité totale des jeux de données utilisées dépasse 1 Go. De fait, les caractéristiques des images de télédétection (multisources, multiéchelles, multidimensionnelles, multi-résolutions, dynamiques et non linéaires) font d'elles de gros volume de données spatiales (données massives spatiales -DMS- ou spatial Big Data -SBD-), distinguables par une croissance significative du volume, la vitesse d'acquisition, de traitement ou de mise à jour, la complexité et la variété des domaines couverts (Liu, 2015 ; Ma et al., 2015 ; Rathore et al., 2015). Ainsi, la transposition de la méthode de réduction d'échelle à l'ensemble du territoire canadien (9 984 670 km<sup>2</sup>)<sup>23</sup> engendrera de gros volumes de données à gérer, sources d'impacts dans la chaîne de traitements et la mise à jour régulière des données. En plus de l'aspect superficielle à couvrir, il y a également l'hétérogénéité du territoire (états de surface variés) et la dynamique temporelle des données surtout,

---

<sup>23</sup> <https://www150.statcan.gc.ca/n1/pub/11-402-x/2011000/chap/geo/geo-fra.htm>

pour la LST qui varie constamment. La variabilité sur un laps de temps horaire par exemple exige de disposer de plus de données satellitaires MODIS LST pour un suivi horaire.

La manipulation de telle quantité de données satellitaires demande de disposer de méthodes adéquates de traitement et d'analyse pour gagner en efficacité. Soulignons que dans la procédure de réduction d'échelle, la préparation et le traitement des données nécessitent des actions répétitives, fastidieuses et énergivores en temps. Partant de là, il se pose la question de gestion des flux de données massives<sup>24</sup>, leur cycle de vie et leur automatisation. Les défis liés aux données massives sont autant sur le plan technique/technologique, structurel et organisationnel. Dans le traitement et l'analyse de ces gros volumes de données, les enjeux spatio-temporels sont à prendre en compte. Une des solutions est de développer un algorithme automatisé pour faciliter le traitement des données en vue de les reproduire sur d'autres cas d'étude et à grande échelle.

#### *5.4.4.2. Automatisation des chaînes de traitements de MODIS LST de 1000 m à 100 m*

Le recours à l'automatisation n'est pas nouveau en géomatique. L'automatisation est utilisée en télédétection dans la classification des images satellitaires basée sur des pixels ou des objets (Grippa et al., 2017), en cartographie dans la production des cartes (Touya, 2017) ou encore dans la mise à jour des données (Latour, 2017). Le recours à l'automatisation en géomatique se justifie également par l'actuelle époque qu'est l'ère des données multiples (images satellitaires multisources et multirésolutions), où la production des informations exploitables doit être robuste et fiable (Stratoulas et al., 2017). L'automatisation accroît la qualité des résultats par réduction des erreurs liées à la répétitivité de manipulations (Lamy & Dubreuil, 2014). Elle réduit les tâches exécutées manuellement et les coûts de production pour gagner en efficacité par la rapidité d'exécution (Bonin, 1981 ; Thomas, 2017). Cette optimisation paraît importante dans cette étude si on se retrouve avec une quantité considérable de données, à mettre à jour mensuellement pour les moustiques et chaque semaine pour les tiques.

Il existe plusieurs solutions pour l'automatisation : celle de type Extract, Transform and Load -ETL- (exemple du logiciel Feature manipulation engine FME), de modèles graphiques de type « Modèle Builder », celles basées exclusivement sur des langages de programmation (Python, Java, C++) ou celles en ligne comme Google Earth Engine. Nous avons opté pour la solution basée sur des langages de programmation, car avec des scripts, l'automatisation est plus complète et peut se réaliser sur plusieurs territoires à la fois à partir d'un même et

---

<sup>24</sup> Les flux de données massives renvoient aux quantités de données géospatiales (surtout satellitaires) et à leur fréquence temporelle de production. Ainsi pour les données MODIS LST, la fréquence est quasi continue à toutes les 15 minutes et de 16 jours pour des données Landsat 08 OLI. En fonction de l'option retenue, les traitements de ces données peuvent être synchronisés ou non avec la fréquence de production.

unique script (Latour, 2017). Ce choix a donné lieu comme mentionné dans le chapitre 6 au développement d'un algorithme de réduction d'échelle de LST de 1000 m à 100 m pour les méthodes RFR et PBIM.

La fonctionnalité de ces algorithmes a été testée. Ainsi, le traitement des données avec 4 prédicteurs jusqu'à l'obtention du résultat final à 100 m avec l'algorithme automatisé de RFR a duré 35 minutes. Cette durée augmente chaque fois qu'on ajoute un prédicteur supplémentaire. Tandis qu'avec l'algorithme de PBIM, basé sur un seul prédicteur, le traitement a duré 10 h pour la même zone d'étude. Nous avons constaté avec les deux autres expérimentations que le coût des calculs en temps et ressources a baissé en raison de la taille des fichiers, moins volumineuse par rapport à celle de région de la Capitale nationale. Cette description donne un aperçu sur les coûts en traitement des données auxquels on peut s'attendre avec un gros flux de données géospatiales. Il reste à tester ces algorithmes automatisés sur de grandes zones d'études pour évaluer son efficacité et adaptabilité. C'est seulement à travers des tests que nous pouvons savoir s'ils répondent bien ou s'il faut modifier leur conception. L'automatisation permet de faire face à la pénibilité dans le traitement des données.

#### 5.4.5. Application de la méthodologie de réduction d'échelle proposée à d'autres déterminants climatiques et microclimatiques (DCMC)

Dans cette section, nous démontrons la capacité des méthodes TsHARP et RFR ou de l'approche proposée à prédire les précipitations et l'humidité de surface. Les résultats obtenus avec LST démontrent qu'outre les méthodes, ce sont les variables de prédiction qui sont importantes.

L'humidité de surface et les précipitations ont été prédites par réduction d'échelle en se basant sur le même principe, l'utilisation des variables explicatives à haute résolution spatiale. Ainsi, entre autres méthodes, les méthodes statistiques de régression, dont la RFR, SVR, l'algorithme Cubist ou autres, algorithmes d'apprentissage machine (Chen et al., 2021 ; Guevara & Vargas, 2019 ; Liu et al., 2018 ; López López et al., 2018 ; Peng et al., 2017 ; Sheng et al., 2019 ; Zhao, 2021) ont été souvent appliquées dans différentes études. Par analogie à la prédiction de la LST, les variables explicatives utilisées sont les données biophysiques décrites par les indices spectraux (indice de surface foliaire, albédo, indice d'eau, indice de végétation...), topographiques (altitude, aspect, pente, MNT), de couverture terrestre (occupation et utilisation du sol), les données météorologiques (LST,...) et autres variables explicatives comme la latitude, longitude. Comme on peut le constater, la prédiction de l'humidité de surface ou des précipitations recourt aux mêmes variables prédictives que la LST hormis que la LST à cause de l'interrelation existante entre ces trois déterminants climatiques. De plus, la LST elle-même devient une variable explicative dans l'estimation des deux autres.

L'utilisation des mêmes types de variables prédictives est un avantage pour une adaptation rapide de l'algorithme semi-automatisé de RFR réalisé.

Partant du fait que le mode opératoire de la réduction d'échelle reste inchangé, alors les méthodes de TsHARP avec plusieurs variables et de RFR peuvent aussi être utilisées dans la réduction d'échelle de l'humidité de surface et des précipitations. Par exemple, Immerzeel et ses pairs ont appliqué la méthode TsHARP pour simuler les précipitations à 1km de résolution. Dans cette étude, le NDVI a été la variable prédictive (Immerzeel et al., 2009). Li et al. (2018) ont intégré comme variables explicatives l'altitude, la pente, l'aspect, le NDVI, la longitude et la latitude à l'équation de la méthode TsHARP pour réduire les précipitations à 1km de résolution spatiale.

Bien que les méthodes TsHARP et RFR peuvent être appliquées aux précipitations et à l'humidité de surface, est-ce que l'approche méthodologique en deux étapes convient dans ce cas précis ? C'est la deuxième étape de cette approche basée sur les émissivités de surface qu'il fait questionner. Selon Zhou et al. (2018), il existe une relation entre l'émissivité de surface et l'humidité de surface. Toutefois, dans les articles exploités, aucune étude n'a utilisé les émissivités de surface pour prédire l'humidité de sol ou les précipitations, ce qui constitue une opportunité à expérimenter avant de tirer les conclusions.

En conclusion sur la possibilité d'appliquer la méthode de réduction d'échelle à d'autres déterminants climatiques et microclimatiques (humidité de surface et les précipitations), nous retenons que les résultats obtenus pour la LST et confrontés avec les études similaires ont permis de les valider. Ces résultats démontrent également la capacité des méthodes appliquées à prédire la MODIS LST jusqu'à 30 m de résolution spatiale malgré les incertitudes s'élevant à plus de 2°C. Les émissivités de surface, comme variable prédictive, ont apporté une plus-value aux résultats obtenus. Cependant, les marges d'erreur similaires à celles issues de la réduction d'échelle de 1000 m à 30 m en une seule étape amènent à se questionner sur sa nécessité.

La bonne performance (faible marge d'erreur) de la méthode TsHARP dans deux zones hétérogènes devant la RFR, qui elle, a montré plus de détails spatiaux dans les images réduites limite la sélection d'une seule méthode comme étant meilleure. De fait, la qualité des résultats obtenus selon les spécificités environnementales et topographiques de chaque zone oriente dans le choix d'une méthode adaptée.

Étendre la mise en œuvre de la réduction d'échelle sur de grandes superficies est synonyme de gestion des données massives de télédétection. Leur manipulation nécessite la mise en place d'un cadre organisé pour atténuer les éventuelles contraintes liées aux données et au mode de traitement.

Nous avons également démontré théoriquement que les méthodes TsHARP et RFR peuvent s'adapter en fonction des données en entrée pour réaliser la réduction d'échelle des autres déterminants climatiques et microclimatiques dans le cadre du projet de cartographie du risque associé aux maladies vectorielles émergentes au Canada.

## **5.5. Conclusion**

En synthèse de ce chapitre consacré à l'analyse et aux discussions des résultats, nous avons montré à travers les analyses des résultats qu'il y a une concordance des résultats de LST à 100 m ou 30 m obtenus par réduction d'échelle avec ceux des différentes études recensées dans la revue de littérature. La précision obtenue à 100 m et à 30 m est quasi identique et les marges d'erreur sont autour de 2°C. Cependant malgré la capacité des méthodes testées à prédire aux fines échelles la LST, la précision spatiale souhaitée au niveau de LST (moins de 2°C pour la RMSE à 30 m) dans le contexte de cette étude pour permettre le calcul d'autres DCMC et la caractérisation des habitats des tiques et moustiques n'est pas atteinte. Dès lors, en l'absence d'une plus grande incidence dans la précision spatiale la pertinence de prédire la LST à 30 m ne paraît pas nécessaire. Ce qui, par voie de conséquence, amène à questionner la nécessité de recourir aux méthodes de réduction d'échelle dans le cadre du projet de l'ASPC. Alors, en vue d'améliorer les résultats pour avoir des données proches de la réalité terrain, l'utilisation de la télédétection par drone peut être une alternative.

Partant des expérimentations faites et des constats sur la sensibilité des résultats selon la morphologie de la zone d'étude, des unités d'occupation du sol et des dates d'acquisition des données, nous avons discuté des préoccupations spatio-temporelles et des méthodes liées à l'application des modèles de réduction d'échelle. Un des enjeux spatiaux est de l'intérêt de se focaliser uniquement sur des zones forestières, humides et rurales susceptibles d'abriter les tiques et moustiques que d'appliquer les méthodes de réduction d'échelle sur toutes les unités d'occupation de la zone d'étude. L'enjeu temporel souligne la nécessité d'identifier un mois de la saison d'été comme référence pour l'acquisition des données à traiter. Ces préoccupations sont à prendre en compte pour éviter des contraintes dans le traitement des gros volumes de données si la réduction d'échelle s'appliquait à l'ensemble du territoire canadien.



# Conclusion générale

## Introduction

Dans cette étude, l'objectif général visait à démontrer la faisabilité et la pertinence de réduire l'échelle de données MODIS de 1000 m à 100 m et 30 m pour déduire la température de surface avec une marge d'erreur de 2°C. De manière spécifique, il s'agit aussi de questionner si avec les images MODIS LST de 1000 m réduites à 100/30 m, nous avons des précisions suffisantes proches de la réalité ou non. La LST réduite à l'échelle de 100/30 m contribuera au calcul des déterminants climatiques et microclimatiques comme l'humidité de surface, les précipitations basés sur la LST. En effet, l'identification des méthodes de réduction d'échelle appliquées dans cette étude a été traitée dans le Chapitre 1 - portant sur l'état de l'art et une analyse comparative de celles dédiées à la LST. Cette analyse comparative basée sur des critères généraux et spécifiques du projet de l'APSC a permis de retenir 3 méthodes de réduction d'échelle (TsHARP, RFR, PBIM).

L'hypothèse qui sous-tend cet objectif général est comme les émissivités de surface améliorent les valeurs de LST, alors en les utilisant pour calculer une LST réduite à 30 m, la précision serait plus grande et significative. Pour tester cette hypothèse, nous avons fait des expérimentations dans trois zones distinctes : les régions de la Capitale nationale, de l'Estrie et de la Montérégie, composées de différentes unités de surface (zone humide, plan d'eau, zone artificialisée, zone agricole, zone forestière, zone nue). Les expérimentations ont été faites suivant un processus opérationnel en deux étapes : la première étape consiste à simuler MODIS LST 1000 m pour atteindre 100 m en se basant sur les indices de végétation, d'urbanisation et facteurs topographiques à haute résolution spatiale. La seconde étape consiste à réaliser la réduction d'échelle à 30 m en utilisant une seule variable, les émissivités de surface. Pour les deux étapes, ce sont les méthodes statistiques de TsHARP et de RFR qui ont été utilisées. La troisième méthode PBIM expérimentée en parallèle est la méthode physique basée uniquement sur les émissivités de surface.

Au terme de ce projet de recherche, nous présentons en deux sections les acquis de cette recherche et les pistes de recherches prospectives.

## Acquis de l'étude sur la précision spatiale avec les méthodes de réduction d'échelle

Cette section s'articule autour deux points : le résumé de cette recherche et les contributions faites pour répondre aux objectifs de l'étude.

### Bilan de la validation des méthodes de réduction d'échelle de LST

Les résultats obtenus dans cette étude permettent de tirer **des enseignements sur trois plans** :

- Si nous considérons **les méthodes**, les résultats avec les marges d'erreur aux alentours de 2 °C démontrent la capacité des trois méthodes de réduction d'échelle appliquées à déduire la LST dans l'espace. Les marges d'erreur sont semblables à celles variant autour de 2 °C identifiées dans la revue de littérature. Les analyses qualitatives révèlent également pour les 3 expérimentations que la méthode RFR apporte plus de détails spatiaux que TsHARP et PBIM, autant à 100 m qu'à 30 m. En revanche, les analyses quantitatives montrent que la méthode TsHARP a de faibles marges d'erreur. Bien sûr, d'autres expérimentations sont nécessaires pour conforter ces conclusions. Nos résultats comparés avec ceux des études ayant appliqué les mêmes méthodes donnent des marges d'erreur sensiblement égales (autour de 2°C) en dépit des différences entre les résolutions spatiales cibles. En conséquence, nos résultats sont concordants avec ceux des études antérieures et donc satisfaisants. Cependant, vu que la marge d'erreur n'est pas inférieure au seuil de 2°C, alors, l'hypothèse affirmant que les méthodes de réduction d'échelle appliquées au niveau des zones d'études hétérogènes permettent d'obtenir une image MODIS LST avec une marge d'erreur en deçà de 2°C est infirmée.
- En considérant **les résultats** à 100 m et à 30 m pour les méthodes TsHARP et RFR, l'écart le plus élevé entre les RMSE est d'au plus 0.31°C. Cet écart est marginal et révèle que le niveau de précision spatial de LST est quasi identique en passant de 100 m à 30 m même en se basant sur les émissivités de surface. Dès lors, nos résultats sont non concluants en l'absence de grands écarts dans les marges d'erreur à 100 m comme à 30 m et aussi en passant de 100 m à 30 m. L'hypothèse de recherche postulant que les modèles de réduction d'échelle permettent d'obtenir une précision spatiale du MODIS LST à 30 m de qualité équivalente à celle obtenue à 100 m est confirmée et amène à conclure qu'avec les images MODIS LST réduite à 100 m ou 30 m, la précision ne peut être meilleure et être en deçà de 2°C. Or, en appliquant une méthode de réduction d'échelle surtout de 100 m à 30 m, on s'attendait à avoir une plus grande précision de la LST, proche de la réalité dans l'espace. Il s'ensuit qu'une réduction d'échelle jusqu'à 30 m n'est pas nécessaire et pertinente.
- En considérant le **contexte environnemental**, le projet de l'ASPC vise la caractérisation de l'habitat des tiques et moustiques en se basant sur les déterminants climatiques et microclimatiques à calculer grâce à la LST déduite et des critères d'occupation du sol pour un suivi vectoriel efficace. Or, les résultats de LST obtenus par réduction d'échelle n'apportent pas plus d'informations dans la discrimination spatiale des milieux forestiers, agricoles, ruraux à dominance forestière, où l'on retrouve les tiques et moustiques. Dès lors, la détection des DCMC ne saurait être effective sans recourir à d'autres mesures de LST avec des outils plus performants comme les drones. En effet, les drones ont été déjà utilisés pour la caractérisation et la compréhension des températures de surface aux échelles locales dans différentes études (Duffy et al., 2021 ; Naughton & McDonald, 2019). Les faibles marges d'erreur (moins de 1°C) obtenues surtout en zones hétérogènes offrent d'intéressantes opportunités pour améliorer les imprécisions des méthodes de réduction d'échelle ou l'incapacité des

capteurs imageurs à fournir des images satellitaires à une très haute résolution spatio-temporelle adéquate.

Si l'applicabilité des méthodes de réduction d'échelle a été validée par les expérimentations réalisées, les résultats obtenus montrent que pour atteindre une grande précision spatiale à 100/30 m, il faudrait envisager utiliser autres données que MODIS LST ou appliquer d'autres méthodes que celles de réduction d'échelle dans le contexte du projet de l'ASPC.

## Contributions

La principale contribution dans cette étude a été la **définition d'un canevas de réduction d'échelle de MODIS LST à 30 m dans le contexte des maladies vectorielles**. Les études existantes n'ont pu atteindre 30 m avec l'image MODIS LST 1000 m. Pour atteindre les 30 m de résolution spatiale, nous avons utilisé les émissivités de surface. Les résultats de LST obtenus à 30 m sont satisfaisants, mais n'apportent pas plus de détails spatiaux et une plus grande précision spatiale que les résultats à 100 m. Ce constat démontre la faible efficacité des méthodes de réduction d'échelle à 30 m. Par ailleurs, il n'existe pas non plus de différence importante entre les résultats de LST réduites à 100 m et les images LST de Landsat 08 TIRS à 100 m. En résumé, la plus-value des méthodes de réduction d'échelle de MODIS LST de 1000 m à 100/30 m est minime dans le contexte des maladies vectorielles. Si nous jugeons du temps investi dans la préparation des données, du calcul des variables de prédiction et de l'application de ces modèles pour aboutir à des résultats peu exploitables, alors l'efficacité matérielle des méthodes de réduction d'échelle est faible. Toutefois, il se peut qu'en se focalisant uniquement sur des zones ciblées, l'application des méthodes de réduction d'échelle se révèle être utile. Sur ce point, seules les expérimentations peuvent le confirmer. Cette contribution a permis de répondre aux trois premiers objectifs fixés pour cette étude (Objectifs spécifiques). **Ces sous-objectifs font référence à l'identification de méthodes de réduction d'échelle, leur application et l'analyse des résultats pour le cas de LST pour juger de leur intérêt pour cette recherche.**

Nous avons par la suite défini un cadre méthodologique au Chapitre 2 - pour l'application de ces 3 méthodes pour atteindre les résolutions spatiales de 100/30 m. Il s'ensuit au Chapitre 3 - la proposition d'un cadre de mise en œuvre de la méthodologie définie. L'opérationnalisation de la démarche proposée a été faite au Chapitre 4 - avec des expérimentations dont les résultats ont permis de discuter sur la précision ou la représentation spatiale de LST à 100 m/30 m avec une marge d'erreur en deçà de 2°C. Les analyses et discussions faites ont abouti aux enseignements présentés dans la précédente sous-section (cf. la page 125).

Une deuxième contribution a été également de mettre en évidence à travers les discussions **les enjeux méthodologiques et spatio-temporels liés aux méthodes de réduction d'échelle de LST sur l'ensemble du territoire**. Par exemple, nous avons montré du fait de la sensibilité des méthodes de réduction d'échelles à

la morphologie du territoire que le choix d'une méthode entre RFR ou TsHARP avant son application peut se baser sur des caractéristiques environnementales (topographie, diversité dans l'occupation du sol...) et climatiques (variation des températures de surface, l'amplitude thermique, les valeurs extrêmes...) de chaque zone d'étude.

En ce qui concerne l'enjeu temporel, il s'agira d'identifier le mois idéal de la saison d'été pour l'acquisition des données de température de surface à réduire pour espérer avoir de bons résultats pour tout type de couverture du sol. De plus, le choix de la date (jour et heure) de la température de surface MODIS à réduire et celles des données de validation Landsat 08 TIRS ou température de l'air) doivent idéalement correspondre ou à défaut avoir un décalage horaire d'au plus 6 heures pour le même jour.

S'agissant de l'enjeu spatial, il est en lien avec les habitats des tiques et des moustiques. Les résultats de LST prédite sur l'ensemble de la zone d'étude s'étant avérés peu significatifs dans l'identification précise de ces habitats, il apparaît utile de se focaliser uniquement sur des unités d'occupation et y appliquer la réduction d'échelle. Une alternative qui peut aboutir à des résultats plus significatifs et s'avérer efficace, voire efficiente, en temps de calcul, ressources et coût.

**À travers cette contribution, le sous-objectif 4 de cette étude a été atteint.** Ce sous-objectif visait spécifiquement à identifier et décrire les opportunités, contraintes et problématiques à prendre en compte dans la gestion de gros volumes de données avec les méthodes de réduction d'échelle. Cette problématique des gros volumes de données et leur gestion ont été théoriquement discutées à la sous-section 5.4.4 du Chapitre 5 -

## **Perspectives de recherche**

De l'analyse et discussions des résultats présentés dans cette étude, des approfondissements sur la réduction d'échelle dans le contexte des maladies vectorielles sont nécessaires pour apporter plus d'informations pour orienter l'ASPC dans ses choix sur les méthodes, la gestion de données massives et sur les stratégies. Cette section vient également souligner ce qui peut être amélioré dans cette étude et des pistes de réflexion pour des études le futur.

## Amélioration de l'approche méthodologique de réduction d'échelle de la température de surface

**À court terme**, prioritairement, l'ensemble du processus de réduction d'échelle de 1000 m à 30 m intégrant le téléchargement des données pourrait être automatisé pour des méthodes de RFR et TsHARP. Ceci facilitera plusieurs expérimentations suivant différentes temporalités et morphologies du territoire pour identifier par exemple la période idéale des images LST à réduire. Des tests de réduction d'échelle sur de grandes superficies et sur de longues séries chronologiques (par exemple des MODIS LST de jour de 2014 à 2020 pour le mois de juillet) peuvent aussi se faire dans un contexte de gestion des gros volumes de données, afin d'évaluer la performance et la robustesse des méthodes sur de grandes étendues.

Afin d'améliorer les résultats de LST, d'autres variables prédictives telles les occupations du sol et les émissivités, la longitude, la latitude, l'albédo, l'angle d'incidence solaire, etc. ou l'association d'autres variables prédictives pour espérer avoir des précisions spatiales des déterminants climatiques et microclimatiques de 1000 m à 100 m/30 m. La question qui se pose est : « **est-ce que l'ajout d'autres variables ou de leur association permettra d'avoir une précision spatiale significative en deçà de celle obtenue dans cette étude qui est aux alentours de 2°C ?** ».

Bien que les résultats de cette recherche amènent à se questionner sur la nécessité de recourir aux méthodes de réduction d'échelle pour avoir des données géospatiales spatio-temporelles à haute résolution, elles restent pour le moment l'outil de substitution le plus utilisé dans la production des données de haute résolution spatio-temporelle. Cependant des incertitudes sur ces méthodes ont été relevées dans l'état de l'art. L'une d'elles est la sensibilité des méthodes de réduction d'échelle à l'occupation et l'utilisation du sol. De fait, l'étude des méthodes de réduction d'échelle peut être menée **à moyen terme** afin d'identifier celles qui s'appliquent le mieux à des contextes particuliers des territoires du Canada. Dans cette étude, nous avons testé 3 méthodes et n'avons pas identifié celle qui peut correspondre à tout type du territoire (ce qui n'est pas un des objectifs de cette étude). Or, l'idéal serait de dégager une seule méthode à déployer surtout si l'on prend en compte le cadre opératoire de modélisation des cartes de risques lié aux maladies vectorielles (Figure 0.2 ) pour permettre d'atteindre une efficacité et efficacité du système dans sa globalité. Dans le cas contraire, **de nouvelles recherches peuvent proposer une nouvelle méthode intégrant les limites ou insuffisances identifiées des méthodes existantes pour permettre de disposer de données plus précises**. À défaut de mettre en place une nouvelle méthode de réduction d'échelle, on peut envisager une stratégie visant à disposer de plusieurs méthodes de réduction d'échelle pour les appliquer selon les contextes/états des données entrantes par exemple. D'autre part, dans le contexte des maladies vectorielles, des recherches peuvent être menées sur **des méthodes alternatives à celles de réduction d'échelle pour évaluer l'opportunité, coût-bénéfice lié**

**les unes par rapport aux autres.** Ceci est fondamental surtout avec les évolutions des nouvelles technologies de l'information et la mise en orbite de nouveaux capteurs imageurs tels le Landsat 09 OLI/TIRS depuis le 27 septembre 2021.

### Évaluer la contribution des drones dans la précision spatiale de LST

L'évaluation des résultats de LST estimée par réduction dans cette étude a montré que la précision spatiale de l'image de LST prédite ne peut être meilleure que celle de l'image originale en dépit d'une amélioration dans la représentation spatiale de LST. Étant donné que nos résultats de LST réduite à 30 m n'apportent pas un gain significatif en précision spatiale pour favoriser l'identification des conditions microclimatiques des zones de rétention des vecteurs, alors on pourrait recourir à des données acquises par drones en complément aux méthodes de réduction d'échelle pour viser une précision spatiale en deçà de 2°C.

**L'émergence des drones dans la télédétection** vient combler les limites de la télédétection satellitaire à fournir des images satellitaires à haute/très haute résolution spatiale et temporelle notamment en environnement. De plus, son utilisation vient également combler les lacunes en représentativité spatio-temporelle des capteurs ou enregistreurs implantés au sol. À la seule différence qu'elle ne peut couvrir une grande superficie, la télédétection par drone peut être déployée en toute zone et en tout temps d'un territoire. Les caméras, comme les caméras thermiques radiométriques pour la LST par exemple, embarquées sur les drones, offrent l'avantage d'avoir des données géospatiales de résolution spatiale d'ordre centimétrique. Ainsi, on peut aujourd'hui acquérir des données géospatiales aux échelles spatiales fines des environnements plus complexes comme les milieux hétérogènes forestiers à canopées denses ou les milieux hétérogènes urbains ou encore agricoles. **La plus-value avec les drones est l'acquisition des données de grandes précisions avec plus de détails spatiaux et reflétant la réalité terrain** (Duffy et al., 2021 ; Naughton & McDonald, 2019). Les drones ont déjà été utilisés dans divers contextes où les microclimats jouent un rôle prépondérant. Il s'agit par exemple l'étude des microclimats dans des habitats forestiers (Webster et al., 2018), de l'influence de l'artificialisation des états de surface dans les paysages agricoles (Tucci et al., 2019), etc. L'utilisation de la télédétection par drone pour mesurer la température de surface nécessiterait une caméra thermique infrarouge embarquée sur un drone. Selon le type de drone utilisé, la durée moyenne maximale de vol est de 3h. L'altitude maximale autorisée par le Canada est de 122 mètres<sup>25</sup> pour les drones. Il existe également différents outils et logiciels permettant à la fois la gestion de vol, le paramétrage et l'exploitation des données obtenues par le capteur imageur.

---

<sup>25</sup> <https://tc.canada.ca/fr/aviation/securite-drones/apprenez-regles-avant-piloter-votre-drone/utiliser-votre-drone-facon-securitaire-legale>

**Le recours aux images de drone dans cette étude vise à saisir la précision spatiale de la variabilité de LST dans les zones d'imprécision** après l'application des méthodes de réduction d'échelle. Il s'agit notamment des zones forestières, agricoles hétérogènes, les zones humides, les zones rurales. Les mesures in situ répétées dans ces zones sous différentes conditions météorologiques peuvent servir de données de référence pour la modélisation de LST dans des ces zones. Ainsi, ces données de référence peuvent aider en second lieu à la prédiction de LST en se basant sur des outils de l'intelligence artificielle (apprentissage machine ou profond). **Il apparait alors dans cette perspective d'évaluer les apports des images de LST acquises par drone dans la précision spatiale des images de MODIS LST estimées par réduction d'échelle à 100/30 m. De manière approfondie, cela vient aussi questionner la précision en allant à une meilleure résolution que 30 m. Pourrait-on atteindre jusqu'à 1 mètre de résolution spatiale pour la LST ?**

Intégrer la méthode de réduction d'échelle dans l'infrastructure de données massives

Comme on peut le constater sur la Figure 0.2 qui présente les étapes sur la cartographie des risques de maladies infectieuses liées au climat, la réduction d'échelle intervient dans la partie de l'étape 1 dédiée à la production des données pour alimenter les modèles pour la prédiction des risques. La réalisation de cette cartographie de risque à l'échelle du Canada va générer d'importantes quantités de données volumineuses à traiter. Ceci pose des défis et contraintes structurels, technologiques, informatiques pour l'archivage, la manipulation, l'analyse et la diffusion des données massives. En prenant en compte l'enjeu de gestion des données massives, des pistes de recherches peuvent viser l'automatisation du système composé des 4 parties décrites sur cette Figure 0.2 en annexe C. La mise en place d'un tel système doit tenir compte du processus de réduction d'échelle dans les infrastructures de données massives en tenant compte de l'existant à l'ASPC.

Selon l'ASPC, la gestion des grands volumes de données dépasse la capacité de leur département de technologie informationnelle<sup>26</sup>. En outre, dans un objectif de célérité dans la modélisation/production des cartes de risques pour éclairer les prises de décision notamment en situation d'urgence, il est souhaité de disposer d'une solution efficace qui prend en compte toutes les étapes liées à cette production cartographique. Cette solution sous la forme d'un système automatisé doit comprendre les outils ou modèles assurant des fonctions dont:

- un accès amélioré à des données géospatiales ouvertes, multisources, prétraitées et prêtes à l'utilisation ;
- le téléchargement de données géospatiales ouvertes sur internet ;

---

<sup>26</sup> <https://www.ic.gc.ca/eic/site/101.nsf/fra/00029.html>

- l'archivage, l'analyse, l'affichage et la diffusion des données et résultats.

Ces fonctions correspondent aux 4 composantes clés de l'infrastructure des données massives : collecter, stocker, analyser et visualiser les données traitées. Pour mettre en place cette infrastructure, l'ASPC a mandaté en 2018 des firmes informatiques. Comme mentionné plus haut, nous ne disposons pas plus d'informations sur ce qui se fait exactement. Toutefois, selon la communication du chef de l'unité géomatique de l'ASPC en novembre 2021 au Centre de recherche en données et intelligence spatiales (CRDIG)<sup>27</sup>, les travaux en cours évoluent et sont à la phase 2 après leur interruption due à la crise sanitaire de COVID 19. Cette phase 2 est dédiée au développement du prototype de l'infrastructure des données massives dans les nuages (le Cloud avec les cubes de données) et en parallèle à la mise en place de solutions de téléchargement, prétraitement des données satellitaires prêtes à utiliser et de calcul des indicateurs de base (NDVI, SAVI, NDBI...). Cette phase 2 intervient après la phase 1 consacrée au développement conceptuel de la solution envisagée.

Ce besoin renvoie à l'automatisation de l'ensemble des chaînes de traitement pour une mise à jour à des fréquences régulières des estimations des données et des cartes des risques sanitaires liés aux maladies de Lyme et du virus du Nil. Ainsi, chaque sous-partie de chacune des 4 étapes de la Figure 0.2 (en annexe C) doit être autonome dans son fonctionnement tout en restant en interrelation avec les autres sous-parties de manière que l'ensemble du système aboutisse au résultat final. **Comment arriver à implanter la sous-partie relative à la réduction d'échelle dans cet ensemble tout en garantissant le bon fonctionnement du système ?** C'est ce sur quoi des recherches futures doivent se pencher. Ces recherches doivent prendre en compte les 4 composantes essentielles d'une infrastructure de données massives. Des pistes de propositions (en lien avec la réduction d'échelle) à améliorer ou approfondir sont présentées en 3 sous parties en annexe C.

Cette recherche sur les déterminants climatiques et microclimatiques représente une sous-composante du cadre opératoire du projet de l'ASPC présenté sur la Figure 0.2 en annexe C. La production de la cartographie de risque est basée sur des données géospatiales dont notamment les images MODIS pour la température de surface (LST). Bien que nos résultats à 30 m n'apportent pas assez de précision spatiale, ce projet peut toujours continuer en se limitant à une résolution spatiale de 100 m pour la LST. Il faudrait toutefois avoir à l'esprit le biais que peut introduire la tolérance aux marges d'erreur dans la représentation spatiale des résultats des DCMC. Cette recherche vient éclairer et orienter ce qui se faisait déjà au niveau de ce projet de l'ASPC, à savoir la réduction d'échelle des DCMC à des résolutions moins fines de 250 m.

---

<sup>27</sup> <https://crdig.ulaval.ca/la-geographie-medecale-et-la-tele-epidemiologie-comme-outils-dinformation-sur-les-maladies-infectieuses/>



# Bibliographie

- Adde, A., Dusfour, I., Roux, E., Girod, R., & Briolant, S. (2016). Anopheles fauna of coastal Cayenne, French Guiana : Modelling and mapping of species presence using remotely sensed land cover data. *Memórias Do Instituto Oswaldo Cruz*, 111(12), 750-756. <https://doi.org/10.1590/0074-02760160272>
- Agam, N., Kustas, W. P., Anderson, M. C., Li, F., & Colaizzi, P. D. (2007). Utility of thermal sharpening over Texas high plains irrigated agricultural fields. *Journal of Geophysical Research*, 112(D19), D19110. <https://doi.org/10.1029/2007JD008407>
- Agli, H., Daniel, S., & Chokmani, K. (2014). *Revue Des Méthodes De Prétraitement Des Données D'imagerie Hyperspectrale Acquisées Depuis Un Drone*. 14.
- Atkinson, P. M. (2004). Chapter 4 Resolution Manipulation and Sub-Pixel Mapping. 20.
- Atkinson, P. M. (2013). Downscaling in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 22, 106-114. <https://doi.org/10.1016/j.jag.2012.04.012>
- Augustin, H., Sudmanns, M., Tiede, D., Lang, S., & Baraldi, A. (2019). Semantic Earth Observation Data Cubes. *Data*, 4(3), 102. <https://doi.org/10.3390/data4030102>
- Bai, Y., Wong, M., Shi, W.-Z., Wu, L.-X., & Qin, K. (2015). Advancing of Land Surface Temperature Retrieval Using Extreme Learning Machine and Spatio-Temporal Adaptive Data Fusion Algorithm. *Remote Sensing*, 7(4), 4424-4441. <https://doi.org/10.3390/rs70404424>
- Baldocchi, D., Dralle, D., Jiang, C., Ryu, Y., (2019). How Much Water Is Evaporated Across California? A Multiyear Assessment Using a Biophysical Model Forced With Satellite Remote Sensing Data. *Water Resources Research*, 55(4), 2722-2741. <http://dx.doi.org.acces.bibl.ulaval.ca/10.1029/2018WR023884>
- Bartkowiak, P., Castelli, M., & Notarnicola, C. (2019). Downscaling Land Surface Temperature from MODIS Dataset with Random Forest Approach over Alpine Vegetated Areas. *Remote Sensing*, 11(11), 1319. <https://doi.org/10.3390/rs11111319>
- Baumann, P., Misev, D., Merticariu, V., Huu, B. P., & Bell, B. (2018). Datacubes : A Technology Survey. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 430-433. <https://doi.org/10.1109/IGARSS.2018.8518920>
- Baumann, P., Rossi, A. P., Bell, B., Clements, O., Evans, B., Hoenig, H., Hogan, P., Kakaletis, G., Koltsida, P., Mantovani, S., Marco Figuera, R., Merticariu, V., Misev, D., Pham, H. B., Siemen, S., & Wagemann, J. (2018). Fostering Cross-Disciplinary Earth Science Through Datacube Analytics. In P.-P. Mathieu & C. Aubrecht (Eds.), *Earth Observation Open Science and Innovation* (p. 91-119). Springer International Publishing. [https://doi.org/10.1007/978-3-319-65633-5\\_5](https://doi.org/10.1007/978-3-319-65633-5_5)
- Beltrami, H., Gosselin, C., & Mareschal, J. C. (2003). Ground surface temperatures in Canada : Spatial and temporal variability: spatial variation of surface temperature. *Geophysical Research Letters*, 30(10), n/a-n/a. <https://doi.org/10.1029/2003GL017144>
- Benmecheta, A. (2017). Estimation de la température de surface a partir de l'imagerie satellitale ; validation sur une zone côtière d'Algérie. 118.
- Bian, L., & Butler, R. (1999). Comparing Effects of Aggregation Methods on Statistical and Spatial Properties of Simulated Spatial Data. *Photogrammetric Engineering and Remote Sensing*, 65, 73-84.
- Bierkens, M. F. P., Finke, P., & Willigen, P. (2001). Upscaling and Downscaling Methods for Environmental Research. Dordrecht, etc., Kluwer, 2000. *Dev. Plant Soil Sci.* 88, 190 pp, 88.
- Bindhu, V. M., Narasimhan, B., & Sudheer, K. P. (2013). Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration. *Remote Sensing of Environment*, 135, 118-129. <https://doi.org/10.1016/j.rse.2013.03.023>
- Bisquert, M., Sanchez, J. M., & Caselles, V. (2016). Evaluation of Disaggregation Methods for Downscaling MODIS Land Surface Temperature to Landsat Spatial Resolution in Barrax Test Site. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(4), 1430-1438. <https://doi.org/10.1109/JSTARS.2016.2519099>
- Boffet, A. (2003). Généralisation cartographique et description des espaces urbains (Cartographic generalisation and description of urban areas). *Bulletin de l'Association de géographes français*, 80(2), 207-219. <https://doi.org/10.3406/bagf.2003.2328>
- Bonafoni, S., Anniballe, R., Gioli, B., & Toscano, P. (2016). Downscaling Landsat Land Surface Temperature over the urban area of Florence. *European Journal of Remote Sensing*, 49(1), 553-569. <https://doi.org/10.5721/EuJRS20164929>

- Bonin, S. (1981). Perspectives nouvelles pour l'enseignement de la cartographie. *Norois*, 109(1), 31-44. <https://doi.org/10.3406/noroi.1981.3942>
- Bouchard, C., Beauchamp, G., Nguon, S., Trudel, L., Milord, F., Lindsay, L. R., Bélanger, D., & Ogden, N. H. (2011). Associations between Ixodes scapularis ticks and small mammal hosts in a newly endemic zone in southeastern Canada : Implications for Borrelia burgdorferi transmission. *Ticks and Tick-Borne Diseases*, 2(4), 183-190. <https://doi.org/10.1016/j.ttbdis.2011.03.005>
- Bouchard, C., Dibernardo, A., Koffi, J., Wood, H., Leighton, P., & Lindsay, L. (2019). Increased risk of tick-borne diseases with climate and environmental changes. *Canada Communicable Disease Report*, 45(4), 83-89. <https://doi.org/10.14745/ccdr.v45i04a02>
- Breiman, L. (2001). Machine Learning, Volume 45, Number 1—SpringerLink. *Machine Learning*, 45, 5-32. <https://doi.org/10.1023/A:1010933404324>
- Burtis, J. C., Sullivan, P., Levi, T., Oggenfuss, K., Fahey, T. J., & Ostfeld, R. S. (2016). The impact of temperature and precipitation on blacklegged tick activity and Lyme disease incidence in endemic and emerging regions. *Parasites & Vectors*, 9(1), 606. <https://doi.org/10.1186/s13071-016-1894-6>
- Cao, X., Bao, A., Chen, X., & Xia, Y. (2008). Land surface temperature in response to land use/cover change based on remote sensing data in Sangong River (W. Gao & H. Wang, Éd.s.; p. 70830K). <https://doi.org/10.1117/12.791576>
- Chen, C., Hu, B., & Li, Y. (2021). Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy. *Hydrology and Earth System Sciences*, 25(11), 5667-5682. <https://doi.org/10.5194/hess-25-5667-2021>
- Chokmani, K., Viau, A. A., & Bourgeois, G. (2001). Analyse de l'incertitude de quatre modèles de phytoprotection relative à l'erreur des mesures des variables agrométéorologiques d'entrée. *Agronomie*, 21(2), 147-167. <https://doi.org/10.1051/agro:2001100>
- Chybicki, A., & Łubniewski, Z. (2017). Optimized AVHRR land surface temperature downscaling method for local scale observations : Case study for the coastal area of the Gulf of Gdańsk. *Open Geosciences*, 9(1). <https://doi.org/10.1515/geo-2017-0032>
- Coll, C., Galve, J. M., Sanchez, J. M., & Caselles, V. (2010). Validation of Landsat-7/ETM+ Thermal-Band Calibration and Atmospheric Correction With Ground-Based Measurements. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1), 547-555. <https://doi.org/10.1109/TGRS.2009.2024934>
- Coops, N. C., Duro, D. C., Wulder, M. A., & Han, T. (2007). Estimating afternoon MODIS land surface temperatures (LST) based on morning MODIS overpass, location and elevation information. *International Journal of Remote Sensing*, 28(10), 2391-2396. <https://doi.org/10.1080/01431160701294653>
- Crosson, W. L., Al-Hamdan, M. Z., & Insaf, T. Z. (2020). Downscaling NLDAS-2 daily maximum air temperatures using MODIS land surface temperatures. *PLOS ONE*, 15(1), e0227480. <https://doi.org/10.1371/journal.pone.0227480>
- Crum, S. M., & Jenerette, G. D. (2017). Microclimate Variation among Urban Land Covers : The Importance of Vertical and Horizontal Structure in Air and Land Surface Temperature Relationships. *Journal of Applied Meteorology and Climatology*, 56(9), 2531-2543. <https://doi.org/10.1175/JAMC-D-17-0054.1>
- Dash, P., Götsche, F.-M., Olesen, F.-S., & Fischer, H. (2002). Land surface temperature and emissivity estimation from passive sensor data : Theory and practice-current trends. *International Journal of Remote Sensing*, 23(13), 2563-2594. <https://doi.org/10.1080/01431160110115041>
- Dhu, T., Giuliani, G., Juárez, J., Kavvada, A., Killough, B., Merodio, P., Minchin, S., & Ramage, S. (2019). National Open Data Cubes and Their Contribution to Country-Level Development Policies and Practices. *Data*, 4(4), 144. <https://doi.org/10.3390/data4040144>
- Dominguez, A., Kleissl, J., Luvall, J. C., & Rickman, D. L. (2011). High-resolution urban thermal sharpener (HUTS). *Remote Sensing of Environment*, 115(7), 1772-1780. <https://doi.org/10.1016/j.rse.2011.03.008>
- Duan, S.-B., Li, Z.-L., Li, H., Götsche, F.-M., Wu, H., Zhao, W., Leng, P., Zhang, X., & Coll, C. (2019). Validation of Collection 6 MODIS land surface temperature product using in situ measurements. *Remote Sensing of Environment*, 225, 16-29. <https://doi.org/10.1016/j.rse.2019.02.020>
- Duffy, J. P., Anderson, K., Fawcett, D., Curtis, R. J., & Maclean, I. M. D. (2021). Drones provide spatial and volumetric data to deliver new insights into microclimate modelling. *Landscape Ecology*, 36(3), 685-702. <https://doi.org/10.1007/s10980-020-01180-9>
- Ebrahimi, H., & Azadbakht, M. (2019). Downscaling MODIS land surface temperature over a heterogeneous area : An investigation of machine learning techniques, feature selection, and impacts of mixed pixels. *Computers & Geosciences*, 124, 93-102. <https://doi.org/10.1016/j.cageo.2019.01.004>
- Ermida, S. L., Trigo, I. F., DaCamara, C. C., Götsche, F. M., Olesen, F. S., & Hulley, G. (2014). Validation of remotely sensed surface temperature over an oak woodland landscape—The problem of viewing and illumination geometries. *Remote Sensing of Environment*, 148, 16-27. <https://doi.org/10.1016/j.rse.2014.03.016>

- Essa, W., Verbeiren, B., van der Kwast, J., Van de Voorde, T., & Batelaan, O. (2012). Evaluation of the DisTrad thermal sharpening methodology for urban areas. *International Journal of Applied Earth Observation and Geoinformation*, 19, 163-172. <https://doi.org/10.1016/j.jag.2012.05.010>
- Feng Gao, Masek, J., Schwaller, M., & Hall, F. (2006). On the blending of the Landsat and MODIS surface reflectance : Predicting daily Landsat surface reflectance. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2207-2218. <https://doi.org/10.1109/TGRS.2006.872081>
- Feng, X., & Li, J. (2020). Evaluation of the Spatial Pattern of the Resolution-Enhanced Thermal Data for Urban Area. *Journal of Sensors*, 2020, 1-15. <https://doi.org/10.1155/2020/3427321>
- Fox, E. W., Ver Hoef, J. M., & Olsen, A. R. (2020). Comparing spatial regression to random forests for large environmental data sets. *PLOS ONE*, 15(3), e0229509. <https://doi.org/10.1371/journal.pone.0229509>
- Fridley, J. D. (2009). Downscaling Climate over Complex Terrain : High Finescale (&lt;1000 m) Spatial Variation of Near-Ground Temperatures in a Montane Forested Landscape (Great Smoky Mountains) \*. *Journal of Applied Meteorology and Climatology*, 48(5), 1033-1049. <https://doi.org/10.1175/2008JAMC2084.1>
- Gao, F., Kustas, W., & Anderson, M. (2012). A Data Mining Approach for Sharpening Thermal Satellite Imagery over Land. *Remote Sensing*, 4(11), 3287-3319. <https://doi.org/10.3390/rs4113287>
- Gasmi, S., Ogden, N. H., Ripoche, M., Leighton, P. A., Lindsay, R. L., Nelder, M. P., Rees, E., Bouchard, C., Vrbova, L., Rusk, R., Russell, C., Pelcat, Y., Mechai, S., Kotchi, S.-O., & Koffi, J. K. (2019). Detection of municipalities at-risk of Lyme disease using passive surveillance of Ixodes scapularis as an early signal : A province-specific indicator in Canada. *PLOS ONE*, 14(2), e0212637. <https://doi.org/10.1371/journal.pone.0212637>
- Githeko, A. K., Lindsay, S. W., & Confalonieri, U. E. (2001). Changement climatique et maladies a` transmission vectorielle : Une analyse re`gionale. 4, 11.
- Gomes, V., Queiroz, G., & Ferreira, K. (2020). An Overview of Platforms for Big Earth Observation Data Management and Analysis. *Remote Sensing*, 12(8), 1253. <https://doi.org/10.3390/rs12081253>
- Goodfellow, I. (2017). NIPS 2016 Tutorial: Generative Adversarial Networks. ArXiv:1701.00160 [Cs]. <http://arxiv.org/abs/1701.00160>
- Govil, H., Guha, S., Dey, A., & Gill, N. (2019). Seasonal evaluation of downscaled land surface temperature : A case study in a humid tropical city. *Heliyon*, 5(6), e01923. <https://doi.org/10.1016/j.heliyon.2019.e01923>
- Grippa, T., Lennert, M., Beaumont, B., Vanhuyse, S., Stephenne, N., & Wolff, E. (2017). An Open-Source Semi-Automated Processing Chain for Urban Object-Based Classification. *Remote Sensing*, 9(4), 358. <https://doi.org/10.3390/rs9040358>
- Guevara, M., & Vargas, R. (2019). Downscaling satellite soil moisture using geomorphometry and machine learning. *PLOS ONE*, 14(9), e0219639. <https://doi.org/10.1371/journal.pone.0219639>
- Guijun Yang, Ruiliang Pu, Wenjiang Huang, Jihua Wang, & Chunjiang Zhao. (2010a). A Novel Method to Estimate Subpixel Temperature by Fusing Solar-Reflective and Thermal-Infrared Remote-Sensing Data With an Artificial Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4), 2170-2178. <https://doi.org/10.1109/TGRS.2009.2033180>
- Guijun Yang, Ruiliang Pu, Wenjiang Huang, Jihua Wang, & Chunjiang Zhao. (2010b). A Novel Method to Estimate Subpixel Temperature by Fusing Solar-Reflective and Thermal-Infrared Remote-Sensing Data With an Artificial Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, 48(4), 2170-2178. <https://doi.org/10.1109/TGRS.2009.2033180>
- Guo, L. J., & Moore, J. McM. (1998). Pixel block intensity modulation : Adding spatial detail to TM band 6 thermal imagery. *International Journal of Remote Sensing*, 19(13), 2477-2491. <https://doi.org/10.1080/014311698214578>
- Hachem, S. (2008). Cartographie Des Temperatures De Surface, Des Indices De Gel Et De Degel Et De La Repartition Spatiale Du Pergelisol A L'aide Du Moderate Resolution Imaging Spectroradiometer (Modis). 176.
- Hulley, G. C., Ghent, D., Göttsche, F. M., Guillevic, P. C., Mildrexler, D. J., & Coll, C. (2019). Land Surface Temperature. In *Taking the Temperature of the Earth* (p. 57-127). Elsevier. <https://doi.org/10.1016/B978-0-12-814458-9.00003-4>
- Hutengs, C., & Vohland, M. (2014). Downscaling land surface temperatures from MODIS data to mesoscale resolution with Random Forest regression. 11.
- Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178, 127-141. <https://doi.org/10.1016/j.rse.2016.03.006>
- Ibrahim, I., Abu Samah, A., Fauzi, R., & Noor, N. M. (2016). THE LAND SURFACE TEMPERATURE IMPACT TO LAND COVER TYPES. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B3, 871-876. <https://doi.org/10.5194/isprsarchives-XLI-B3-871-2016>

- Immerzeel, W. W., Rutten, M. M., & Droogers, P. (2009). Spatial downscaling of TRMM precipitation using vegetative response on the Iberian Peninsula. *Remote Sensing of Environment*, 113(2), 362-370. <https://doi.org/10.1016/j.rse.2008.10.004>
- Inamdar, A. K., & French, A. (2009). Disaggregation of GOES land surface temperatures using surface emissivity : DISAGGREGATION OF GOES LST. *Geophysical Research Letters*, 36(2), n/a-n/a. <https://doi.org/10.1029/2008GL036544>
- Jeganathan, C., Hamm, N. A. S., Mukherjee, S., Atkinson, P. M., Raju, P. L. N., & Dadhwal, V. K. (2011). Evaluating a thermal image sharpening model over a mixed agricultural landscape in India. *International Journal of Applied Earth Observation and Geoinformation*, 13(2), 178-191. <https://doi.org/10.1016/j.jag.2010.11.001>
- Kasprzyk, J.-P., & Devillet, G. (2021). A Data Cube Metamodel for Geographic Analysis Involving Heterogeneous Dimensions. *ISPRS International Journal of Geo-Information*, 10(2), 87. <https://doi.org/10.3390/ijgi10020087>
- Kawashima, S., Ishida, T., Minomura, M., & Miwa, T. (2000). Relations between Surface Temperature and Air Temperature on a Local Scale during Winter Nights. *JOURNAL OF APPLIED METEOROLOGY*, 39, 10.
- Kerr, Y. H., Lagouarde, J. P., & Imbernon, J. (1992). Accurate land surface temperature retrieval from AVHRR data with use of an improved split window algorithm. *Remote Sensing of Environment*, 41(2-3), 197-209. [https://doi.org/10.1016/0034-4257\(92\)90078-X](https://doi.org/10.1016/0034-4257(92)90078-X)
- Killough, B. (2018). Overview of the Open Data Cube Initiative. *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 8629-8632. <https://doi.org/10.1109/IGARSS.2018.8517694>
- Kloog, I., Nordio, F., Coull, B. A., & Schwartz, J. (2014). Predicting spatiotemporal mean air temperature using MODIS satellite surface temperature measurements across the Northeastern USA. *Remote Sensing of Environment*, 150, 132-139. <https://doi.org/10.1016/j.rse.2014.04.024>
- Kolajo, T., Daramola, O., & Adebiji, A. (2019). Big data stream analysis : A systematic literature review. *Journal of Big Data*, 6(1), 47. <https://doi.org/10.1186/s40537-019-0210-7>
- Kotchi, S., Bouchard, C., Ludwig, A., Rees, E., & Brazeau, S. (2019). Utilisation des images d'observation de la terre pour améliorer la cartographie des risques de maladies associées au changement climatiques. *Relevé des maladies transmissibles au Canada*, 45(5), 148-158. <https://doi.org/10.14745/ccdr.v45i05a04f>
- Kotchi, S. O. (2015). Estimation et évaluation d'incertitude d'indicateurs agrométéorologiques par télédétection en vue de supporter la lutte phytosanitaire. 476.
- Kotchi, S. O., Bouchard, C., Brazeau, S., & Ogden, N. H. (2021). Earth Observation-Informed Risk Maps of the Lyme Disease Vector *Ixodes scapularis* in Central and Eastern Canada. *Remote Sensing*, 13(3), 524. <https://doi.org/10.3390/rs13030524>
- Kristjánsson, J. E. (1991). Cloud parametrization at different horizontal resolutions. *Quarterly Journal of the Royal Meteorological Society*, 117(502), 1255-1280. <https://doi.org/10.1002/qj.49711750207>
- Kustas, W. P., Norman, J. M., Anderson, M. C., & French, A. N. (2003). Estimating subpixel surface temperatures and energy fluxes from the vegetation index–radiometric temperature relationship. *Remote Sensing of Environment*, 85(4), 429-440. [https://doi.org/10.1016/S0034-4257\(03\)00036-1](https://doi.org/10.1016/S0034-4257(03)00036-1)
- Lamy, C., & Dubreuil, V. (2014). Impact du changement climatique sur les sécheresses en Bretagne. *Automatisation d'un bilan hydrique avec ArcGis et Python. Revue internationale de géomatique*, 24(3), 355-375. <https://doi.org/10.3166/riq.24.355-375>
- Latour, F. (2017). Master 2 « Géographies Numériques ». 78.
- Lewis, A., Oliver, S., Lymburner, L., Evans, B., Wyborn, L., Mueller, N., Raevksi, G., Hooke, J., Woodcock, R., Sixsmith, J., Wu, W., Tan, P., Li, F., Killough, B., Minchin, S., Roberts, D., Ayers, D., Bala, B., Dwyer, J., ... Wang, L.-W. (2017). The Australian Geoscience Data Cube—Foundations and lessons learned. *Remote Sensing of Environment*, 202, 276-292. <https://doi.org/10.1016/j.rse.2017.03.015>
- Li, T., & Meng, Q. (2018). A mixture emissivity analysis method for urban land surface temperature retrieval from Landsat 8 data. *Landscape and Urban Planning*, 179, 63-71. <https://doi.org/10.1016/j.landurbplan.2018.07.010>
- Li, W., Ni, L., Li, Z.-L., Duan, S.-B., & Wu, H. (2019). Evaluation of Machine Learning Algorithms in Spatial Downscaling of MODIS Land Surface Temperature. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(7), 2299-2307. <https://doi.org/10.1109/JSTARS.2019.2896923>
- Liu, J., Zhang, W., & Nie, N. (2018). Spatial Downscaling of TRMM Precipitation Data Using an Optimal Subset Regression Model with NDVI and Terrain Factors in the Yarlung Zangbo River Basin, China. *Advances in Meteorology*, 2018, 1-13. <https://doi.org/10.1155/2018/3491960>
- Liu, P. (2015). A survey of remote-sensing big data. *Frontiers in Environmental Science*, 3. <https://doi.org/10.3389/fenvs.2015.00045>

- López López, P., Immerzeel, W. W., Rodríguez Sandoval, E. A., Sterk, G., & Schellekens, J. (2018). Spatial Downscaling of Satellite-Based Precipitation and Its Impact on Discharge Simulations in the Magdalena River Basin in Colombia. *Frontiers in Earth Science*, 6, 68. <https://doi.org/10.3389/feart.2018.00068>
- Lunetta, R. S. (1991). Remote Sensing and Geographic Information System Data Integration: Error Sources and Research Issues. *PHOTOGRAMMETRIC ENGINEERING*, 11.
- Ma, Y., Wu, H., Wang, L., Huang, B., Ranjan, R., Zomaya, A., & Jie, W. (2015). Remote sensing big data computing : Challenges and opportunities. *Future Generation Computer Systems*, 51, 47-60. <https://doi.org/10.1016/j.future.2014.10.029>
- Malbêteau, Y. (2017). Suivi des ressources en eau par une approche combinant la télédétection multi-capteur et la modélisation phénoménologique. 194.
- Mao, Q., Peng, J., & Wang, Y. (2021). Resolution Enhancement of Remotely Sensed Land Surface Temperature : Current Status and Perspectives. *Remote Sensing*, 13(7), 1306. <https://doi.org/10.3390/rs13071306>
- Massalabi Amani. (1996). Étude comparée de quelques algorithmes de classification spectro-spatiale d'images de télédétection appliqués à l'agriculture. <https://savoirs.usherbrooke.ca/handle/11143/11448>
- Mazzetti, P., Roncella, R., Mihon, D., Bacu, V., Lacroix, P., Guigoz, Y., Ray, N., Giuliani, G., Gorgan, D., & Nativi, S. (2016). Integration of data and computing infrastructures for earth science : An image mosaicking use-case. *Earth Science Informatics*, 9(3), 325-342. <https://doi.org/10.1007/s12145-016-0255-5>
- McCabe, G. J., & Bunnell, J. E. (2004). Precipitation and the Occurrence of Lyme Disease in the Northeastern United States. *Vector-Borne and Zoonotic Diseases*, 4(2), 143-148. <https://doi.org/10.1089/1530366041210765>
- Mechri, R., Ottlé, C., Pannekoucke, O., Kallel, A., Maignan, F., Courault, D., & Trigo, I. (2016). Downscaling Meteosat Land Surface Temperature over a Heterogeneous Landscape Using a Data Assimilation Approach. *Remote Sensing*, 8(7), 586. <https://doi.org/10.3390/rs8070586>
- Merlin, O., Duchemin, B., Hagolle, O., Jacob, F., Coudert, B., Chehbouni, G., Dedieu, G., Garatuza, J., & Kerr, Y. (2010). Disaggregation of MODIS surface temperature over an agricultural area using a time series of Formosat-2 images. *Remote Sensing of Environment*, 114(11), 2500-2512. <https://doi.org/10.1016/j.rse.2010.05.025>
- Minichiello, F. (2017). Machines, données et apprentissage : Relations et enjeux. *Revue internationale d'éducation de Sèvres*, 74, 12-15. <https://doi.org/10.4000/ries.5755>
- Mircholi, F., Sadeghi, S. H., & Khaledi Darvishan, A. (2020). Analyzing spatial variations of relationships between Land Surface Temperature and some remotely sensed indices in different land uses. *Remote Sensing Applications: Society and Environment*, 19, 100359. <https://doi.org/10.1016/j.rsase.2020.100359>
- Morin, G., Le Roux, R., Sturman, A., & Quénel, H. (2018a). Évaluation de la relation entre températures de l'air et températures de surface issues du satellite modis : Application aux vignobles de la vallée de Waipara (Nouvelle-Zélande). *Climatologie*, 15, 62-83. <https://doi.org/10.4267/climatologie.1345>
- Morin, G., Le Roux, R., Sturman, A., & Quénel, H. (2018b). Évaluation de la relation entre températures de l'air et températures de surface issues du satellite modis : Application aux vignobles de la vallée de Waipara (Nouvelle-Zélande). *Climatologie*, 15, 62-83. <https://doi.org/10.4267/climatologie.1345>
- Mukherjee, S., Joshi, P. K., & Garg, R. D. (2014). A comparison of different regression models for downscaling Landsat and MODIS land surface temperature images over heterogeneous landscape. *Advances in Space Research*, 54(4), 655-669. <https://doi.org/10.1016/j.asr.2014.04.013>
- Naughton, J., & McDonald, W. (2019). Evaluating the Variability of Urban Land Surface Temperatures Using Drone Observations. *Remote Sensing*, 11(14), 1722. <https://doi.org/10.3390/rs11141722>
- Nichol, J. (2009). An Emissivity Modulation Method for Spatial Enhancement of Thermal Satellite Images in Urban Heat Island Analysis. *Photogrammetric Engineering & Remote Sensing*, 75(5), 547-556. <https://doi.org/10.14358/PERS.75.5.547>
- Nieto, H., Sandholt, I., Aguado, I., Chuvieco, E., & Stisen, S. (2011). Air temperature estimation with MSG-SEVIRI data : Calibration and validation of the TVX algorithm for the Iberian Peninsula. *Remote Sensing of Environment*, 115(1), 107-116. <https://doi.org/10.1016/j.rse.2010.08.010>
- Ning Kong, X., & Vidal-Madjar, D. (1988). Effet de la résolution spatiale sur des propriétés statistiques des images satellites : Une étude de cas. *International Journal of Remote Sensing*, 9(8), 1315-1328. <https://doi.org/10.1080/01431168808954939>
- Njoku, E. G. (Éd.). (2014). *Encyclopedia of Remote Sensing*. Springer New York. <https://doi.org/10.1007/978-0-387-36699-9>
- Ogden, N., & Gachon, P. (2019). Climate change and infectious diseases : What can we expect? *Canada Communicable Disease Report*, 45(4), 76-80. <https://doi.org/10.14745/ccdr.v45i04a01>
- Ogden, N. H., Lindsay, L. R., Beauchamp, G., Charron, D., Maarouf, A., O'Callaghan, C. J., Waltner-Toews, D., & Barker, I. K. (2004). Investigation of Relationships Between Temperature and Developmental Rates of Tick Ixodes

- scapularis (Acari: Ixodidae) in the Laboratory and Field. *Journal of Medical Entomology*, 41(4), 622-633. <https://doi.org/10.1603/0022-2585-41.4.622>
- Ogden, N., Koffi, J., Lindsay, L., Fleming, S., Mombourquette, D., Sanford, C., Badcock, J., Gad, R., Jain-Sheehan, N., Moore, S., Russell, C., Hobbs, L., Baydack, R., Graham-Derham, S., Lachance, L., Simmonds, K., & Scott, A. (2015). Surveillance de la maladie de Lyme au Canada, de 2009 à 2012. Relevé des maladies transmissibles au Canada, 41(6), 152-168. <https://doi.org/10.14745/ccdr.v41i06a03f>
- Peng, J., Loew, A., Merlin, O., & Verhoest, N. E. C. (2017). A review of spatial downscaling of satellite remotely sensed soil moisture: Downscale Satellite-Based Soil Moisture. *Reviews of Geophysics*, 55(2), 341-366. <https://doi.org/10.1002/2016RG000543>
- Peng, Y., Li, W., Luo, X., & Li, H. (2019). A Geographically and Temporally Weighted Regression Model for Spatial Downscaling of MODIS Land Surface Temperatures Over Urban Heterogeneous Regions. *IEEE Transactions on Geoscience and Remote Sensing*, 57(7), 5012-5027. <https://doi.org/10.1109/TGRS.2019.2895351>
- Raffy, M. (1993). Remotely-sensed quantification of covered areas and spatial resolution. *International Journal of Remote Sensing*, 14(1), 135-159. <https://doi.org/10.1080/01431169308904325>
- Rapiński, Bednarczyk, & Zinkiewicz. (2019). JupyTEP IDE as an Online Tool for Earth Observation Data Processing. *Remote Sensing*, 11(17), 1973. <https://doi.org/10.3390/rs11171973>
- Rathore, M. M. U., Paul, A., Ahmad, A., Chen, B.-W., Huang, B., & Ji, W. (2015). Real-Time Big Data Analytical Architecture for Remote Sensing Application. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(10), 4610-4621. <https://doi.org/10.1109/JSTARS.2015.2424683>
- Rodenas, B. M. M. (2017). Différence d'échelle spatiale entre les mesures satellitaires et in situ d'humidité du sol : Analyse par des approches spatio-temporelles. 219.
- ROY Charles -André. (2008). Cartographie Végétale Du Col De La Vallée De Sverdrup, Île D'ellesmere, Avec Le Ndvi Modis Et Analyses Des Changements D'échelles. UQTR. <http://depot-e.uqtr.ca/id/eprint/1238/1/030096709.pdf>
- Ruiz, M. O., Chaves, L. F., Hamer, G. L., Sun, T., Brown, W. M., Walker, E. D., Haramis, L., Goldberg, T. L., & Kitron, U. D. (2010). Local impact of temperature and precipitation on West Nile virus infection in Culex species mosquitoes in northeast Illinois, USA. *Parasites & Vectors*, 3(1), 19. <https://doi.org/10.1186/1756-3305-3-19>
- Sandholt, I., Rasmussen, K., & Andersen, J. (2002). A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sensing of Environment*, 79(2-3), 213-224. [https://doi.org/10.1016/S0034-4257\(01\)00274-7](https://doi.org/10.1016/S0034-4257(01)00274-7)
- Sheng, J., Rao, P., & Ma, H. (2019). Spatial downscaling of the FY3B soil moisture using random forest regression. 8th International Conference on Agro-Geoinformatics, Agro-Geoinformatics 2019, July 16, 2019 - July 19, 2019, CSISS Foundation Inc.; USDA NIFA. <https://doi.org/10.1109/Agro-Geoinformatics.2019.8820253>
- Simonneaux, V. (1996). Généralisation d'images satellites classées pour leur intégration à un SIG.
- Singh Rawat, K., Sehgal, V. K., & Ray, S. S. (2019). Downscaling of MODIS thermal imagery. *The Egyptian Journal of Remote Sensing and Space Science*, 22(1), 49-58. <https://doi.org/10.1016/j.ejrs.2018.01.001>
- Small, C. (2006). Comparative analysis of urban reflectance and surface temperature. *Remote Sensing of Environment*, 104(2), 168-189. <https://doi.org/10.1016/j.rse.2005.10.029>
- Snyder, W. C., Wan, Z., Zhang, Y., & Feng, Y.-Z. (1998). Classification-based emissivity for land surface temperature measurement from space. *International Journal of Remote Sensing*, 19(14), 2753-2774. <https://doi.org/10.1080/014311698214497>
- Sobrino, J. A., Jimenez-Munoz, J. C., Soria, G., Romaguera, M., Guanter, L., Moreno, J., Plaza, A., & Martinez, P. (2008). Land Surface Emissivity Retrieval From Different VNIR and TIR Sensors. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2), 316-327. <https://doi.org/10.1109/TGRS.2007.904834>
- Sobrino, J., Coll, C., & Caselles, V. (1991). Atmospheric correction for land surface temperature using NOAA-11 AVHRR channels 4 and 5. *Remote Sensing of Environment*, 38(1), 19-34. [https://doi.org/10.1016/0034-4257\(91\)90069-I](https://doi.org/10.1016/0034-4257(91)90069-I)
- Song, J., Du, S., Feng, X., & Guo, L. (2014). The relationships between landscape compositions and land surface temperature: Quantifying their resolution sensitivity with spatial regression models. *Landscape and Urban Planning*, 123, 145-157. <https://doi.org/10.1016/j.landurbplan.2013.11.014>
- Srivastava, P. K., Majumdar, T. J., & Bhattacharya, A. K. (2009). Surface temperature estimation in Singhbhum Shear Zone of India using Landsat-7 ETM+ thermal infrared data. *Advances in Space Research*, 43(10), 1563-1574. <https://doi.org/10.1016/j.asr.2009.01.023>
- Stathopoulou, M., & Cartalis, C. (2009). Downscaling AVHRR land surface temperatures for improved surface urban heat island intensity estimation. *Remote Sensing of Environment*, 113(12), 2592-2605. <https://doi.org/10.1016/j.rse.2009.07.017>

- Stratoulas, D., Tolpekin, V., de By, R., Zurita-Milla, R., Retsios, V., Bijker, W., Hasan, M., & Vermote, E. (2017). A Workflow for Automated Satellite Image Processing : From Raw VHSR Data to Object-Based Spectral Information for Smallholder Agriculture. *Remote Sensing*, 9(10), 1048. <https://doi.org/10.3390/rs9101048>
- Sudmanns, M., Tiede, D., Lang, S., Bergstedt, H., Trost, G., Augustin, H., Baraldi, A., & Blaschke, T. (2020). Big Earth data : Disruptive changes in Earth observation data management and analysis? *International Journal of Digital Earth*, 13(7), 832-850. <https://doi.org/10.1080/17538947.2019.1585976>
- Talebi, H., Peeters, L. J. M., Otto, A., & Tolosana-Delgado, R. (2022). A Truly Spatial Random Forests Algorithm for Geoscience Data Analysis and Modelling. *Mathematical Geosciences*, 54(1), 1-22. <https://doi.org/10.1007/s11004-021-09946-w>
- Tang, K., Zhu, H., & Ni, P. (2021). Spatial Downscaling of Land Surface Temperature over Heterogeneous Regions Using Random Forest Regression Considering Spatial Features. *Remote Sensing*, 13(18), 3645. <https://doi.org/10.3390/rs13183645>
- Thomas, C. (2006). Fusion d'images de résolutions spatiales différentes. 197.
- Thomas, G. (2017). Automatisation d'une méthode d'interpolation pour la modélisation de l'îlot de chaleur urbain sous QGIS. 105.
- Toutin, T. (2004). Review article : Geometric processing of remote sensing images: models, algorithms and methods. *International Journal of Remote Sensing*, 25(10), 1893-1924. <https://doi.org/10.1080/0143116031000101611>
- Touya, G. (2017). Vers l'automatisation de la production de cartes. 96.
- Tucci, G., Parisi, E., Castelli, G., Errico, A., Corongiu, M., Sona, G., Viviani, E., Bresci, E., & Preti, F. (2019). Multi-Sensor UAV Application for Thermal Analysis on a Dry-Stone Terraced Vineyard in Rural Tuscany Landscape. *ISPRS International Journal of Geo-Information*, 8(2), 87. <https://doi.org/10.3390/ijgi8020087>
- Van De Griend, A. A., & Owe, M. (1993). On the relationship between thermal emissivity and the normalized difference vegetation index for natural surfaces. *International Journal of Remote Sensing*, 14(6), 1119-1131. <https://doi.org/10.1080/01431169308904400>
- VoPham, T., Hart, J. E., Laden, F., & Chiang, Y.-Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI) : Potential applications for environmental epidemiology. *Environmental Health*, 17(1), 40. <https://doi.org/10.1186/s12940-018-0386-x>
- Wald, L., Ranchin, T., & Mangolini, M. (2009). Fusion Of Satellite Images Of Different Spatial Resolutions: Assessing The Quality Of Resulting Images. 26.
- Webster, C., Westoby, M., Rutter, N., & Jonas, T. (2018). Three-dimensional thermal characterization of forest canopies using UAV photogrammetry. *Remote Sensing of Environment*, 209, 835-847. <https://doi.org/10.1016/j.rse.2017.09.033>
- WenFeng, Z., YunHao, C., Ji, Z., JinFei, W., WenYu, L., Voogt, J., Zhu, X. L., JinLing, Q., & Jing, L. (2013). Disaggregation of remotely sensed land surface temperature : Literature survey, taxonomy, issues, and caveats. *Remote Sensing of Environment*, 119-139. <https://doi.org/10.1016/j.rse.2012.12.014>
- Weng, Q., Fu, P., & Gao, F. (2014). Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data. *Remote Sensing of Environment*, 145, 55-67. <https://doi.org/10.1016/j.rse.2014.02.003>
- Werbylo, K. L., & Niemann, J. D. (2014). Evaluation of sampling techniques to characterize topographically-dependent variability for soil moisture downscaling. *Journal of Hydrology*, 516, 304-316. <https://doi.org/10.1016/j.jhydrol.2014.01.030>
- Wu, H., & Li, W. (2019). Downscaling Land Surface Temperatures Using a Random Forest Regression Model With Multitype Predictor Variables. *IEEE Access*, 7, 21904-21916. <https://doi.org/10.1109/ACCESS.2019.2896241>
- Wu, H., & Li, Z.-L. (2009). Scale Issues in Remote Sensing : A Review on Analysis, Processing and Modeling. *Sensors*, 9(3), 1768-1793. <https://doi.org/10.3390/s90301768>
- Xu, Y., Knudby, A., & Ho, H. C. (2014). Estimating daily maximum air temperature from MODIS in British Columbia, Canada. *International Journal of Remote Sensing*, 35(24), 8108-8121. <https://doi.org/10.1080/01431161.2014.978957>
- Yang, G., Pu, R., Zhao, C., Huang, W., & Wang, J. (2011). Estimation of subpixel land surface temperature using an endmember index based technique : A case examination on ASTER and MODIS temperature products over a heterogeneous area. *Remote Sensing of Environment*, 115(5), 1202-1219. <https://doi.org/10.1016/j.rse.2011.01.004>
- Yang, H., & Weng, F. (2011). Error Sources in Remote Sensing of Microwave Land Surface Emissivity. *IEEE Transactions on Geoscience and Remote Sensing*, 49(9), 3437-3442. <https://doi.org/10.1109/TGRS.2011.2125794>
- Yang, Y., Cao, C., Pan, X., Li, X., & Zhu, X. (2017). Downscaling Land Surface Temperature in an Arid Area by Using Multiple Remote Sensing Indices with Random Forest Regression. *Remote Sensing*, 9(8), 789. <https://doi.org/10.3390/rs9080789>

- Yang, Y., Li, X., Pan, X., Zhang, Y., & Cao, C. (2017). Downscaling Land Surface Temperature in Complex Regions by Using Multiple Scale Factors with Adaptive Thresholds. *Sensors*, 17(4), 744. <https://doi.org/10.3390/s17040744>
- Yu, X., Guo, X., & Wu, Z. (2014). Land Surface Temperature Retrieval from Landsat 8 TIRS—Comparison between Radiative Transfer Equation-Based Method, Split Window Algorithm and Single Channel Method. *Remote Sensing*, 6(10), 9829-9852. <https://doi.org/10.3390/rs6109829>
- Zakšek, K., & Oštir, K. (2012). Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sensing of Environment*, 117, 114-124. <https://doi.org/10.1016/j.rse.2011.05.027>
- Zawadzka, J., Corstanje, R., Harris, J., & Truckell, I. (2020). Downscaling Landsat-8 land surface temperature maps in diverse urban landscapes using multivariate adaptive regression splines and very high resolution auxiliary data. *International Journal of Digital Earth*, 13(8), 899-914. <https://doi.org/10.1080/17538947.2019.1593527>
- Zhan, W., Chen, Y., Wang, J., Zhou, J., Quan, J., Liu, W., & Li, J. (2012). Downscaling land surface temperatures with multi-spectral and multi-resolution images. *International Journal of Applied Earth Observation and Geoinformation*, 18, 23-36. <https://doi.org/10.1016/j.jag.2012.01.003>
- Zhang, K., externe, L. vers un site, fenêtre, celui-ci s'ouvrira dans une nouvelle, Zhu, G., Ma, J., Yang, Y., Shang, S., & Gu, C. (2019). Parameter Analysis and Estimates for the MODIS Evapotranspiration Algorithm and Multiscale Verification. *Water Resources Research*, 55(3), 2211-2231. <http://dx.doi.org.acces.bibl.ulaval.ca/10.1029/2018WR023485>
- Zhang, X., Ding, F., Peng, X., Wu, W., & Fan, P. (2014). Fast retrieval of land surface emissivity from landsat data through IDL programming. 2014 Third International Workshop on Earth Observation and Remote Sensing Applications (EORSA), 76-80. <https://doi.org/10.1109/EORSA.2014.6927853>
- Zhang, X., Zhong, T., Feng, X., & Wang, K. (2009). Estimation of the relationship between vegetation patches and urban land surface temperature with remote sensing. *International Journal of Remote Sensing*, 30(8), 2105-2118. <https://doi.org/10.1080/01431160802549252>
- Zhang, X., Zhong, T., Wang, K., & Cheng, Z. (2009). Scaling of impervious surface area and vegetation as indicators to urban land surface temperature using satellite data. *International Journal of Remote Sensing*, 30(4), 841-859. <https://doi.org/10.1080/01431160802395219>
- Zhang, Y., Odeh, I. O. A., & Ramadan, E. (2013). Assessment of land surface temperature in relation to landscape metrics and fractional vegetation cover in an urban/peri-urban region using Landsat data. *International Journal of Remote Sensing*, 34(1), 168-189. <https://doi.org/10.1080/01431161.2012.712227>
- Zhao, L. (2016). Geographical information system parallelization for spatial big data processing : A review. *Cluster Comput*, 14.
- Zhao, N. (2021). An Efficient Downscaling Scheme for High-Resolution Precipitation Estimates over a High Mountainous Watershed. *Remote Sensing*, 13(2), 234. <https://doi.org/10.3390/rs13020234>
- Zhou, D. K., Larar, A. M., & Liu, X. (2018). On the relationship between land surface infrared emissivity and soil moisture. *Journal of Applied Remote Sensing*, 12(01), 1. <https://doi.org/10.1117/1.JRS.12.016030>
- Zhou, J., Liu, S., Li, M., Zhan, W., Xu, Z., & Xu, T. (2016). Quantification of the Scale Effect in Downscaling Remotely Sensed Land Surface Temperature. *Remote Sensing*, 8(12), 975. <https://doi.org/10.3390/rs8120975>
- Zhu, X., Chen, J., Gao, F., Chen, X., & Masek, J. G. (2010). An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions. *Remote Sensing of Environment*, 114(11), 2610-2623. <https://doi.org/10.1016/j.rse.2010.05.032>



# Annexe A

Tableau 0.1 : Classification des méthodes de réduction d'échelle en télédétection

Catégories	Approches	Méthodes
Méthodes par satellites	Méthodes actives et passives de fusion de données par micro-ondes	Méthode de détection de changement
		Méthode de fusion bayésienne
		Modélisation bayésienne
	Fusion optique/thermique et hyperfréquence	Réduction d'échelle par ajustement polynomial et améliorée
		Désagrégation basée sur le changement d'échelle physique et théorique (DISPATCH, basé sur l'évaporation)
		Méthode de réduction d'échelle (UCLA) et optimisée
		Techniques d'IA (réseaux neuronaux artificiels, machines à vecteurs pertinents)
Régression des vecteurs de soutien		
Méthodes utilisant des données géospatiales		Relation linéaire entre la SM, la topographie, la végétation et les propriétés du sol
Méthodes basées sur des modèles	Modèle statistique	Description géostatistique de la distribution spatiale
		Technique de réduction d'échelle par ondelettes multi résolution
		Interpolation fractale
		CoKigeage
		krigeage avec dérive externe (KED)
		Régression Forestière aléatoire
	Modèle impliquant les états de surface	krigeage par régression de zone à point (ATPRK)
		Machine d'apprentissage extrême (ELM)
		Deterministic downscaling: approche d'algorithme génétique
		Thin plate spline (TPS) : lissage Laplacien
		Statistical downscaling: regression linéaire
Data assimilation: filtre de Kalman		

Tableau 0.2 : Présentation de quelques capteurs imageurs selon les résolutions temporelles et spatiales

Images satellitaires	R S	RT	Usages majeurs
ASTER	15 m, 20 m, 90m (TIR)	16 j	Environnement/Météorologique
Sentinel 3	300 m	2 j	Environnement/Océanographie
IKONOS	0,82 - 3,2 m	3j	Environnement
Quickbird	61 cm	1 - 3,5 j	Environnement
Orbview	1- 4 m	3 j	Environnement
SPOT 5	2,5 à 5 m	1 j	Météorologie/Environnement
MODIS	250 m, 500m, 1 km	1 j	Atmosphère/Environnement/Océanographie
NOAA-AVHRR	1 - 1,1 km	1 j	Météorologique/Environnement
SEVERI	1 - 4,8 km	15 min	Météorologie
Meteosat	2,5 - 5 km	30 min	Météorologie
Landsat 8 TIRS	100 m (TIRS)	16 j	Température

Source : voir les notes de bas de page RS : résolution spatiale ; RT : résolution temporelle

Tableau 0.3 : Caractéristiques du capteur MODIS

Caractéristiques spécifiques	Description
Orbite	705 Km, 10h30 nœud descendant (Terra) ou 13h30 nœud ascendant (Aqua) ; héliosynchrone, quasi polaire, circulaire
Dimension de la fauchée	2330 km (perpendiculairement à la trace) sur 10 km (le long de la trace au nadir)
Poids	228,7 kg
Taille	1,0 x 1,6 x 1,0 m
Codage des données	12 bits
Délai de ré observation	1 à 2 jours
Bandes spectrales	Bandes 1-2 : 0.62 - 0.87 $\mu\text{m}$ ; Bandes 3-7 : 0.54 - 2.15 $\mu\text{m}$ Bandes 8-36 : 0.405 – 14.38 $\mu\text{m}$
Résolution spatiale	250 m (bandes 1-2) ; 500 m (bandes 3-7) ; 1000 m (bandes 8-36)

Source : <https://modis.gsfc.nasa.gov/about/specifications.php>

Tableau 0.4 : Principales utilisations par bandes spectrales du capteur MODIS

Utilisation principale	Bande	Largeur de la bande		Utilisation principale	Bande	Largeur de la bande
	**				**	
Terre/Nuages/Limites des aérosols	1	620 - 670		Surface/Température des nuages	20	3,660 - 3,840
	2	841 - 876			21	3,929 - 3,989
Terre/Nuages/Propriétés des aérosols	3	459 - 479			22	3,929 - 3,989
	4	545 - 565		23	4,020 - 4,080	
	5	1230 - 1250		Température Atmosphérique	24	4,433 - 4,498
	6	1628 - 1652			25	4,482 - 4,549
	7	2105 - 2155		Nuage Cirus/Vapeur d'eau	26	1,360 - 1,390
Couleur des océans/Phytoplancton/ Biogéochimique	8	405 - 420			27	6,535 - 6,895
	9	438 - 448			28	7,175 - 7,475
Couleur des océans/Phytoplancton/ Biogéochimique	10	483 - 493		Propriétés des nuages	29	8,400 - 8,700
	11	526 - 536		Ozone	30	9,580 - 9,880
	12	546 - 556		Surface/Température des nuages	31	10,780 - 11,280
		13	662 - 672		32	11,770 - 12,270
	14	673 - 683		Altitude maximale des nuages	33	13,185 - 13,485
	15	743 - 753			34	13,485 - 13,785
	16	862 - 877			35	13,785 - 14,085
		17	890 - 920			36
Atmosphère/Vapeur d'eau	18	931 - 941				
	19	915 - 965				

Source : <https://modis.gsfc.nasa.gov/about/specifications.php>

\*\* Bandes 1 à 19 sont en nm ; Bandes 20 à 36 sont en  $\mu$

Tableau 0.5: Analyse comparative de quelques modèles de réduction d'échelle se rapportant à la température de surface

Méthodes	Variante étudiée	Avantages	Limites	Sources
Thermal sharpening (TsHARP)	LST	Bonne performance du modèle dans des zones de végétation assez homogène Amélioration et similarité de la LST réduite par rapport et entre LST grossière en zone urbaine mixte Bons résultats en zone agricole homogène; Modèle simple à mettre en œuvre Modèle adapté aux zones d'abondance végétation TsHARP tient compte de l'humidité du sol à l'échelle la plus grossière Forte capacité de l'algorithme TsHARP à reconstruire la variabilité sub-pixel de la LST à partir uniquement des données NDVI Distribution spatiale cohérente de LST réduite en zone urbaine avec LST MODIS, LST Landsat 8 d'origine Forte similarité de LST réduite avec LST originale en zone hétérogène à dominance agricole Peut s'appliquer aux régions de couverture terrestre mixte mais avec une précision acceptée ou limitée	Incapacité du modèle de retrouver les variations spatiales et la distribution complète des LST de référence Tendance à sous-estimer ou surestimer les faibles valeurs de LST dans les zones montagneuses et les zones humides Peu performant dans les milieux de grandes hétérogénéités Non adapté dans les zones urbaines ou les zones à faible couverture végétale ; Faible capacité à réduire la LST en zones humides Peu performant en zone hétérogène urbaine Ne parvient pas à détecter la variabilité sub-pixel à des résolutions plus fines ; Incapacité du modèle à saisir la variation de la LST avec le NDVI dans les pixels non uniformes Ne peut simuler correctement la LST dans les milieux humides (lac, rivières, ...) Résolution spatiale de la LST réduite possible jusqu'à 250 m pour une image MODIS réelle sur un paysage hétérogène	(Ebrahimi & Azadbakht, 2019) ;  Agam et al., 2007 (Stathopoulou & Cartalis, 2009) Yang et al., (2017)  (Singh Rawat et al., 2019)  (Mukherjee et al., 2014) (Y. Peng et al., 2019b)
Random Forest Regression (RFR)	LST	Similarité dans la répartition spatiale entre l'image de LST réduite et la LST des images MODIS et Landsat 8 Originales Meilleure précision dans les classes d'agriculture et de pâturages sur les images d'été, printemps et d'automne Forte corrélation entre LST et NDVI à une résolution spatiale plus fine Meilleure précision globale et meilleure efficacité de calcul	Incapacité du modèle de retrouver les variations spatiales et la distribution complète des LST de référence Tendance à sous-estimer ou surestimer les faibles valeurs de LST dans les zones montagneuses et les zones humides Grande complexité structurelle et dans la mise en œuvre	(Ebrahimi & Azadbakht, 2019)
Extreme Learning Machine (ELM)	LST	Similarité dans la répartition spatiale entre l'image de LST réduite et la LST des images MODIS et Landsat 8 originales Niveau de précision élevée dans les zones dénudées sur l'image d'été Faible temps de calcul ; Facile à mettre en œuvre Forte capacité à modéliser les phénomènes environnementaux complexes non linéaires Forte corrélation entre LST et NDVI à une résolution spatiale plus fine Meilleure précision globale	Incapacité du modèle de retrouver les variations spatiales et la distribution complète des LST de référence Tendance à sous-estimer ou surestimer les faibles valeurs de LST dans les zones montagneuses et les zones humides	(Ebrahimi & Azadbakht, 2019)
Support Vector Regression (SVR)	LST	Similarité dans la répartition spatiale entre l'image de LST réduite et la LST des images MODIS et Landsat 8 Originales Niveau de précision élevée dans les zones résidentielles sur les images de printemps et d'automne Forte corrélation entre LST et NDVI à une résolution spatiale plus fine	Incapacité du modèle de retrouver les variations spatiales et la distribution complète des LST de référence Tendance à sous-estimer ou surestimer les faibles valeurs de LST dans les zones montagneuses et les zones humides Grande complexité structurelle et dans la mise en œuvre	(Ebrahimi & Azadbakht, 2019)

Source : recherche documentaires articles ; BM : basic modelling; EM : enhanced modelling; MSFAT :Multi-scale-factor downscaling approach based on adaptive threshold; SADFAT : Spatio-temporal Adaptive Data Fusion Algorithm for Temperature mapping; LMSDS : Least median square regression downscaling;

Méthodes	Variable étudiée	Avantages	Limites	Sources
Optimized downscaling process Dsopt	LST	Conservation des caractéristiques radiométriques des images en entrée Plus performant à l'échelle locale ; sur des superficies relativement petites (d'ordre de 300 Km <sup>2</sup> ) S'adapte aux cas où les conditions physiques locales influencent significativement les caractéristiques finales de la relation LST VI (vegetation Index) Modèle auto adaptative	Moins efficace avec des images sous influence des aérosols e aérosols	(Chybicki & Lubniewski, 2017)
Pixel Block Intensity Modulation (PBIM)		Conservation des caractéristiques radiométriques des images en entrée S'adapte aux milieux où les variations de température sont principalement causées par la topographie	Ne s'adapte pas pour les zones urbaines où la topographie est essentiellement plate Précision limitée dans les zones présentant une grande diversité et spécifiquement en zones humides	(Chybicki & Lubniewski, 2017)
Data assimilation: Genetic particle smoother (GPS) Downscaling	LST	Approche dynamique; efficace sur les pixels mixtes; Repose sur la cartographie plus fine de couverture du sol; Résultat instantané au moment de l'observation et même en couverture nuageuse Potentialité de la méthode de réduction d'échelle jusqu'à l'échelle hectométrique	Basé sur le postulat que la LST est stationnaire pour chaque d'unité d'occupation du sol; limité au pas du temps du modèle (chaque 30 min) Surestimation des biais sur l'image de la LST réduite par rapport à ceux de l'image LST originale pour les sols nus Modèle à affiner davantage pour ajuster les paramètres	(Mechri et al., 2016b)
PBIM: Pixel Block Intensity Modulation modifié en 4 sous modèles	LST	Conservation des caractéristiques radiométriques des images en entrée; Fusion des images multi capteurs et multi résolutions différentes; s'applique à des bandes individuelles	Difficultés avec les images AVHRR à distinguer les variations de LST en milieu urbain et environs	(Stathopoulou & Cartalis, 2009)
Random Forest Algorithm décomposé en 3 sous modèle (BM; EM 1 et EM2)	LST	Prédire avec précision similaire le LST sous pixel quel que soit la méthode de sélection des pixels; Conservation des caractéristiques des données thermiques originales de 1 km en entrée Meilleurs résultats pour le modèle BM avec MODIS LST réduite Modélisation ne tient pas compte des relations non linéaires entre les variables	Ne peut simuler les variations thermiques au niveau des sous pixels Qualité finale de LST réduite altérée à cause des nuages et de l'hétérogénéité des pixels de 1 km	(Bartkowiak et al., 2019)
Combinaison de Thin plateSpline TPS+ TsHARP	LST	Le couple (TPS + TsHARP) est performant dans les zones hétérogènes car le TPS est efficace en zone de faible végétation Résultat de LST réduite amélioré et concordant en comparaison aux données originales		(Singh Rawat et al., 2019)
Regressive disaggregation approach:polyomial regression	LST	S'adapte mieux aux milieux urbains avec peu de classes d'occupation de sol	Faible capacité à faire la différence en détails thermique spécifique en milieu urbain	(Bonafoni et al., 2016)

Source : recherche documentaire d'articles ; BM : basic modelling; EM : enhanced modelling; MSFAT :Multi-scale-factor downscaling approach based on adaptive threshold; SADFAT : Spatio-temporal Adaptive Data Fusion Algorithm for Temperature mapping; LMSDS : Least median square regression downscaling;

Méthodes	Variable	Avantages	Limites	Sources
Non Linear DisTrad	LST	Capacités à préserver les propriétés statistiques des images observées Peut prédire la LST à une résolution plus fine même en milieu hétérogène Permet également l'estimation de l'évapotranspiration	Modèle basé sur le postulat que la relation entre le NDVI et la LST des capteurs dérivés est invariante à l'échelle. Ne modélise que partiellement la LST aux extrêmes les plus élevés	(Bindhu et al., 2013)
DisTrad <i>Disaggregation procedure for radiometric surface temperature</i>	LST	Bons résultats en zone agricole homogène Modèle simple Plus performant en zone homogène et à dominance de végétation Amélioration et similarité de la LST réduite par rapport et entre LST grossière en zone urbaine mixte	Peu performant en zone hétérogène, complexe et urbaine Incapacité du modèle à simuler efficacement la LST en paysage hétérogène (à dominance agricole mixte et végétalisé) Sensibilité du modèle aux valeurs aberrantes Faible capacité du modèle à réduire la LST en zones humides Résolution spatiale de la LST réduite possible jusqu'à 250 m pour une image MODIS réelle sur un paysage hétérogène	Agam et al, 2007; (Mukherjee et al., 2014) Yang et al., 2017) Bartkowiak et al., 2019
TsHARP avec variante locale	LST	Distribution spatiale cohérente de LST réduite en zone urbaine avec LST MODIS, LST Landsat 8 d'origine Offre une meilleure précision dans la réduction d'échelle Modèle adapté aux paysages agricoles et végétalisés Peut s'appliquer aux régions de couverture terrestre mixte mais avec une précision acceptée ou limitée	Faible capacité du modèle à simuler la LST dans les classes d'eau, surfaces urbaines et nues. Non recommandé pour les plans d'eau et les lits de rivière sablonneux/sols arides où il n'existe pas de relation inverse entre la LST et le NDVI Résolution spatiale de la LST réduite possible jusqu'à 250 m pour une image MODIS réelle sur un paysage hétérogène	(Mukherjee et al., 2014)
LMSDs	LST	Modèle s'appliquant aux zones de paysages hétérogènes (à dominance agricole mixte et végétalisé) Forte similarité de LST réduite avec LST originale	Modèle non adapté aux zones humides et zones arides/nues Résolution spatiale de la LST réduite possible jusqu'à 250 m pour une image MODIS réelle sur un paysage hétérogène	(Mukherjee et al., 2014)
PRDs <i>Pace regression downscaling</i>	LST	Modèle s'appliquant aux zones de paysages hétérogènes (à dominance agricole mixte) Modèle performant avec Landsat 8	Modèle peu efficace avec MODIS LST en zones de paysages hétérogènes (à dominance agricole mixte) Résolution spatiale de la LST réduite possible jusqu'à 250 m pour une image MODIS réelle sur un paysage hétérogène	(Mukherjee et al., 2014)
MSFAT	LST	Modèle efficace dans les zones urbaines mixtes Conserve les caractéristiques thermiques et de distribution spatiale de l'image LST originale Méthode applicable en toute saison et avec de meilleurs résultats en été Peut estimer la LST dans les milieux humides Adapté aux régions à topographie inégale mais plate	Modèle moins performant en hiver et dans les zones humides, hautement construites et dense en milieu urbain Incapacité du modèle à s'adapter en région montagneuse Modèle non appliqué à des images de haute résolution temporelle (MODIS LST ; AVHRR LST ; ASTER LST)	(Y. Yang et al., 2017)
SADFAT	LST	S'adapte aux milieux hétérogènes et surtout dans les zones urbaines (la zone d'étude à une topographie élevée -montagnes) Prise en compte des variations temporelles (entre deux dates) de la LST par le modèle ; Permet d'approximer le changement annuel de la LST	Incapacité du modèle à prédire les changements de LST non reflétés dans les pixels des images satellitaires originales	(Weng et al., 2014)
GTWR <i>Geographically and Temporally Weighted Regression</i>	LST	S'adapte aux milieux urbains hétérogènes Méthode basée sur des relations de non-stationnarité spatiale et temporelle entre les facteurs environnementaux et LST Modèle apportant plus de détails spatiaux sur les images de LST réduite Meilleure concordance entre la LST réduite et la LST de Landsat 8 à 100 m Bonne cohérence entre la LST réduite en zone humide (lac, rivières) avec la LST de Landsat 8	Modèle ne tient pas compte des saisons, de la topographie et l'aspect de zone d'étude	(Y. Peng et al., 2019b)

Source : recherche documentaire d'articles ; BM : basic modelling; EM : enhanced modelling; MSFAT : Multi-scale-factor downscaling approach based on adaptive threshold; SADFAT : Spatio-temporal Adaptive Data Fusion Algorithm for Temperature mapping; LMSDs : Least median square regression downscaling

Tableau 0.6 : Mode de calcul des différents indices

Symbole	Indice	Formule	Sources
NDVI	Indice de végétation différencié normalisé	$NDVI = \frac{(B5 - B4)}{(B5 + 4)}$	Rouse et al. (1974)
EVI	Indice de végétation améliorée	$EVI = \frac{2.5(B5 - B4)}{1 + B5 + 6B4 - 7.5B2}$	
SAVI	Indice de végétation ajustée au sol	$SAVI = \frac{(B5 - B4) * 1.5}{(B5 + B4) + 0.5}$	Huete (1998)
CVF	Couverture végétale fractionnée	$FC = 1 - \left( \frac{NDVI_{max} - NDVI}{NDVI - NDVI_{min}} \right)^{0.625}$	Huete, (1988)
VC	Couverture végétale	$VC = -4.3 - 3.7 * NDVI + 161.9 * NDVI^2$	
NDWI	Indice d'eau différencié normalisé	$NDWI = \frac{(B3 - B5)}{(B3 + B5)}$	Gao, (1996)
MNDWI	Indice d'eau différencié normalisé modifié	$MNDWI = \frac{(B3 - B6)}{(B3 + B6)}$	Xu (2006)
NDBI	Indice de construction différencié normalisé	$NDBI = \frac{(B6 - B5)}{(B6 + B5)}$	Zha et al., (2003)
UI	Indice urbain	$UI = \frac{(B7 - B5)}{(B7 + B5)}$	Kawamura et al., (1996)
IBI	Indice de construction basé sur indice	$IBI = \frac{NDBI - (SAVI + MNDWI) / 2}{NDBI + (SAVI + MNDWI) / 2}$	H. Xu, (2008)
Albédo	Albédo de surface	$0.356 * B2 + 0.130 * B4 + 0.373 * B5 + 0.085 * B6 + 0.072 * B7 - 0.0018$	
BSI	Indice de sol nu	$BSI = \frac{(B3 + B5)}{(B3 - B5)}$	

Tableau 0.7 : Déterminants environnementaux associés aux maladies transmises par les moustiques (MTM) et leurs vecteurs dont *Culex pipiens-restuans*, vecteur du Virus du Nil Occidental

Classe	Sous-classes
Végétation	Herbacé
	Prairie
Forêt	Feuille
	Conifère
	Mixte
	Arbuste
Urbain	Résidentiel
	Parc
	Commercial
	Zone récréative
	Végétalisé
	Boisé
Suburbain	Résidentiel
	Végétalisé
	Boisé
Agriculture	Arbre et arbuste
	Herbacé
	Prairie
	Grandes cultures, céréales, maraîcher
	Terre humide/irriguée
Parcs et aires récréatives	Parcs et aires récréatives
Zones humides	Zones régulièrement inondées
	Zones temporairement inondées
	Bassins de rétention/détention (zone de stockage d'eau permanente ou temporaire)
	Canaux/fossés
	Terres humides
	Marais
	Marécage boisé
	Marécage
	Tourbière

	Tourbière boisée
	Étang
	Site d'excavation (ex. carrières)
	Site de traitements des eaux usées
Eau	Eau stagnante
	Eau riche en matières organiques
	Eau en zones urbaine, suburbaine, agricole, parcs (ex. étang d'irrigation, eau de surface non permanente créée par la pluie, etc.)
	Eau dans les contenants artificiels (ex. tas de pneus)



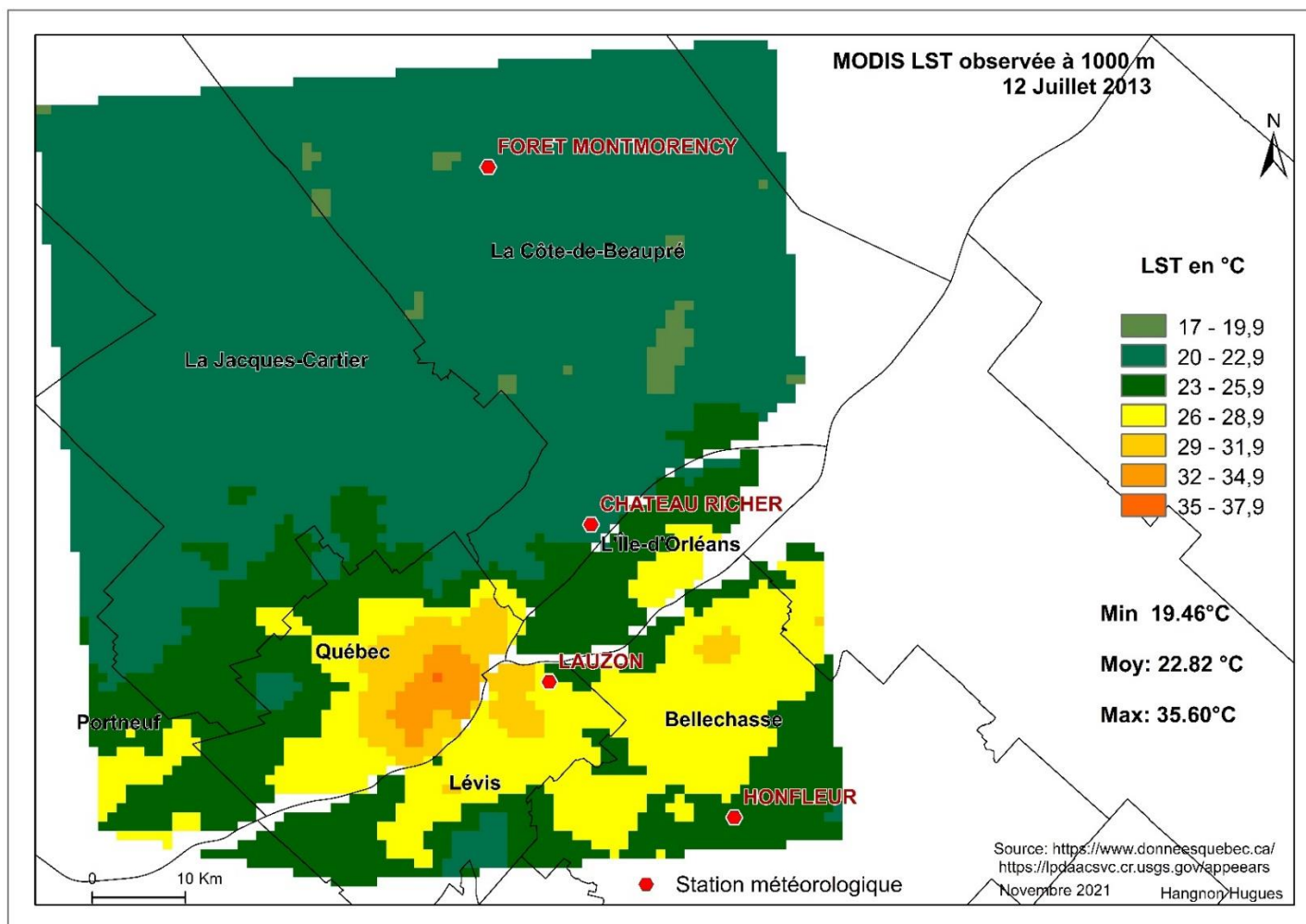


Figure 0.1 : Région de la Capitale nationale : Spatialisation de la température de surface (°C) de MODIS à 1000 m

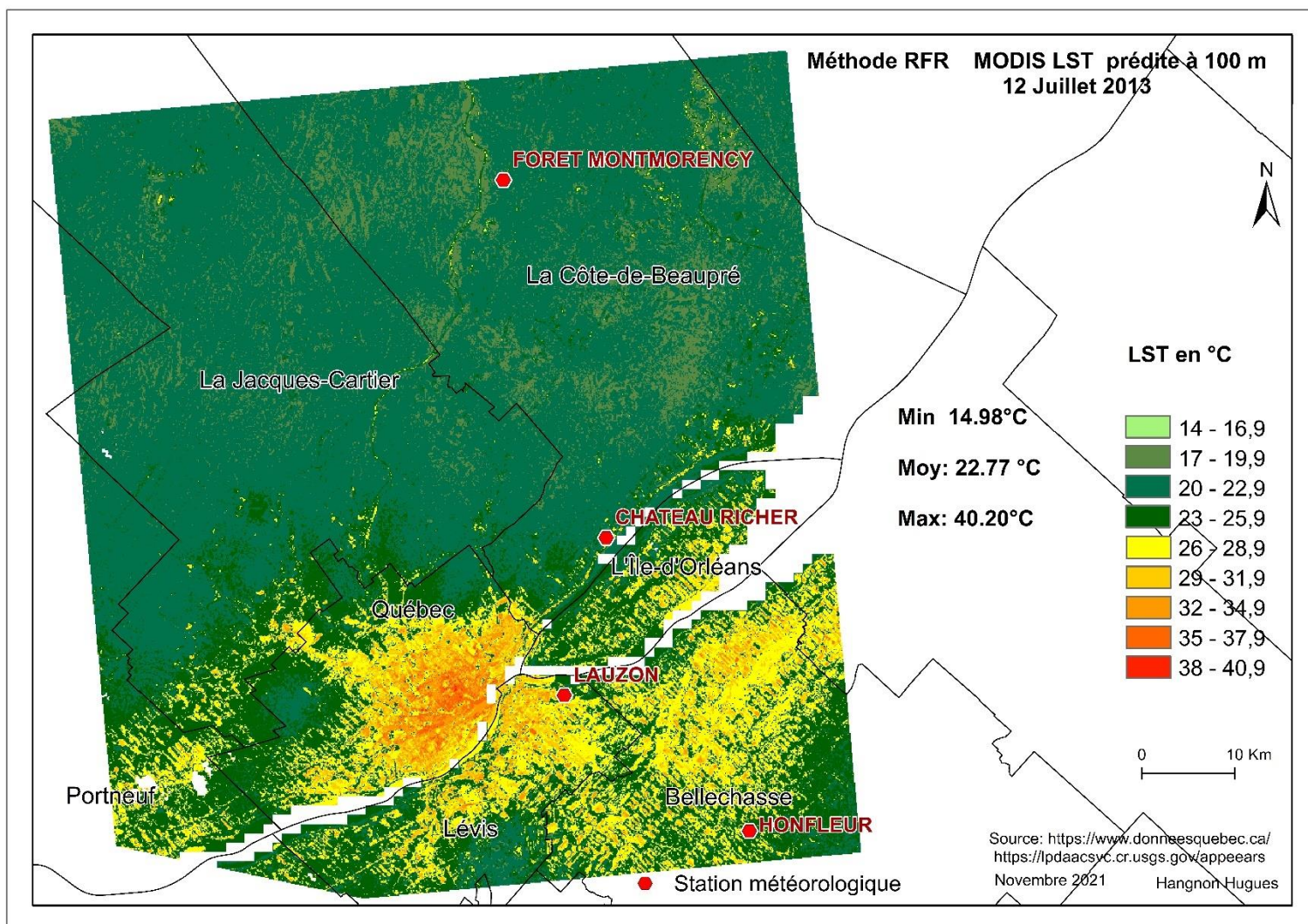


Figure 0.2 : Région de la Capitale nationale : résultats de réduction d'échelle de la température de surface (°C) MODIS de 1000 m à 100 m suivant la méthode RFR

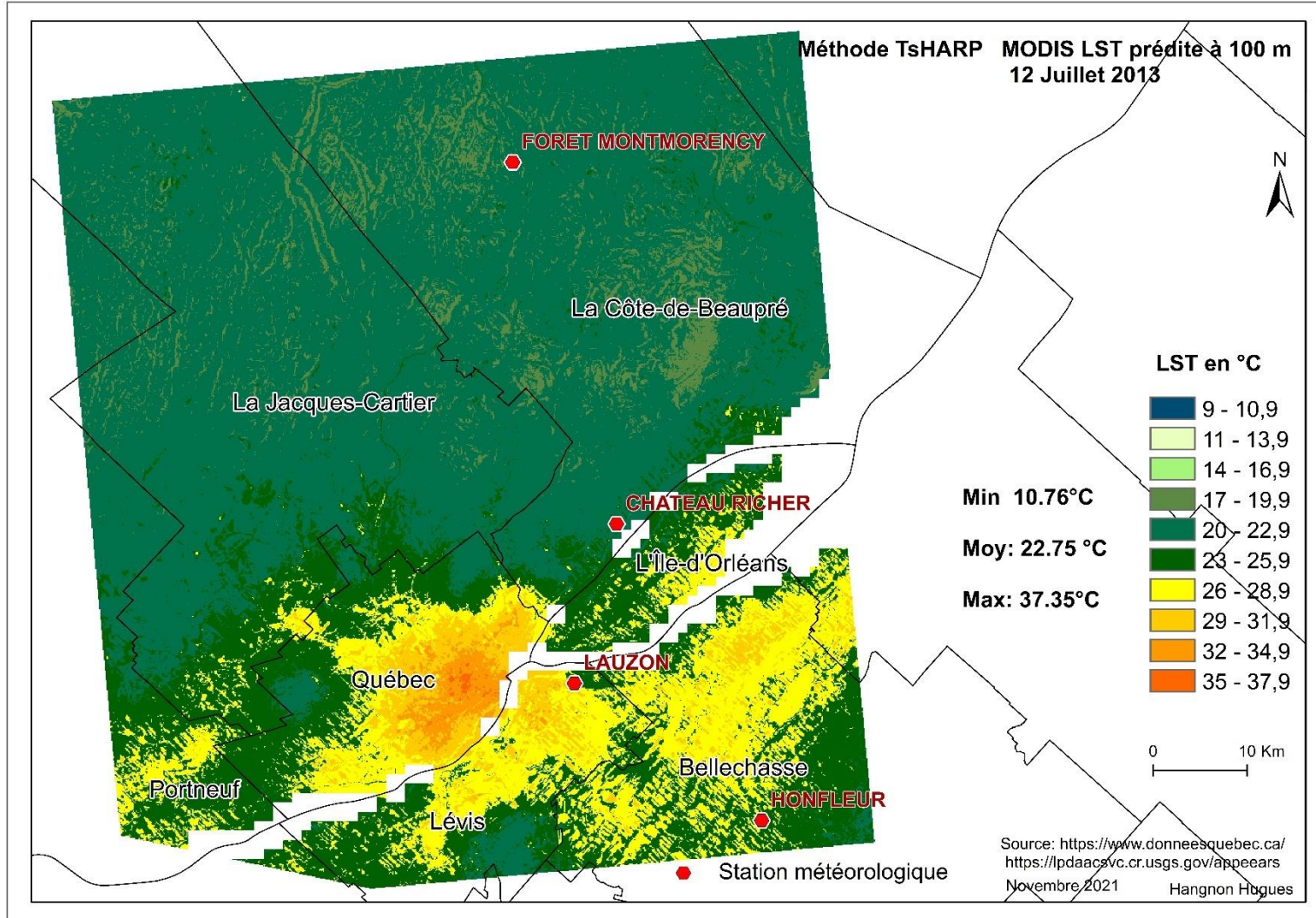


Figure 0.3 : Région de la Capitale nationale : résultats de réduction d'échelle de la température de surface (°C) MODIS de 1000 m à 100 m suivant la méthode TsHARP

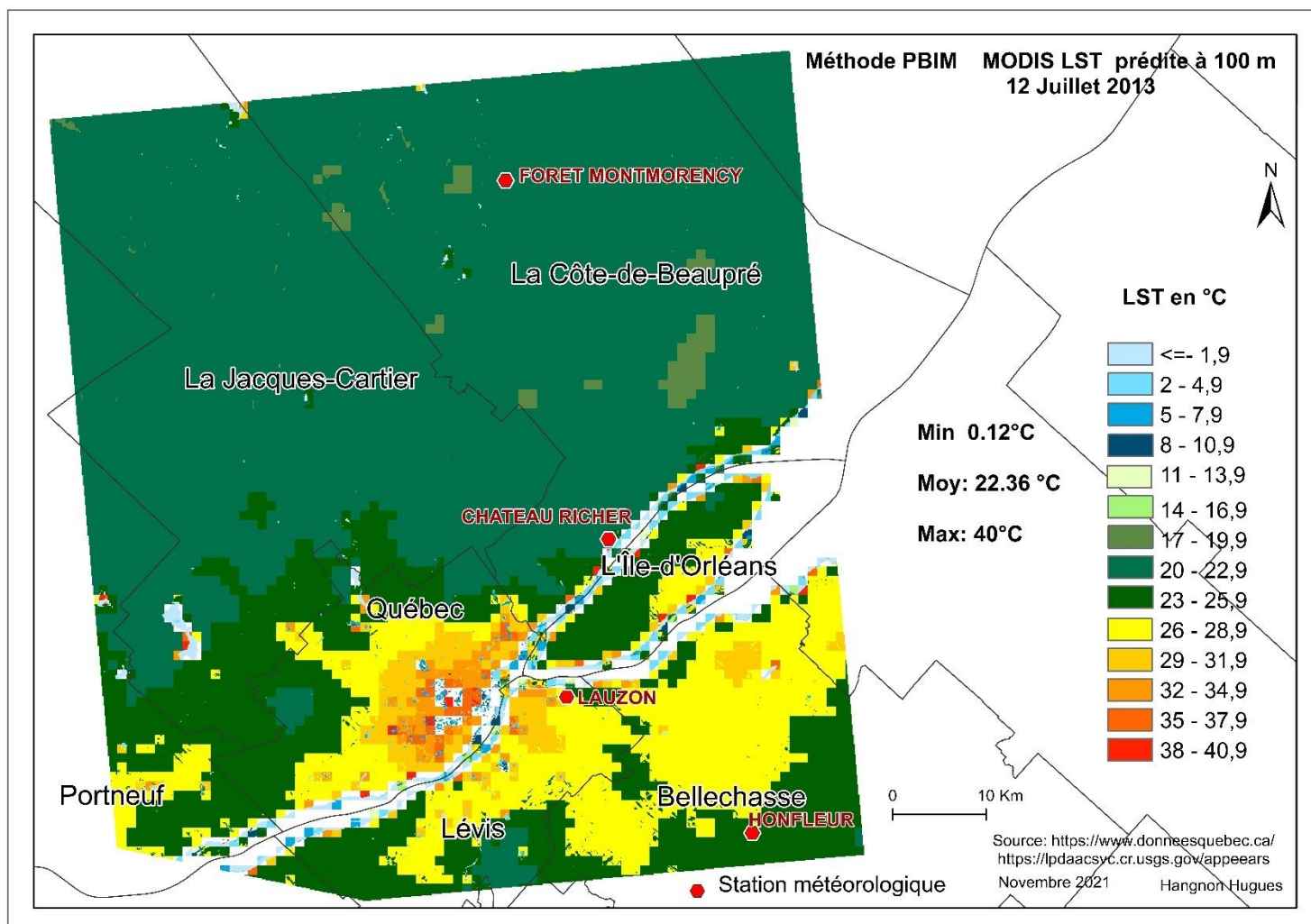


Figure 0.4 : Région de la Capitale nationale : résultats de réduction d'échelle de la température de surface (°C) MODIS de 1000 m à 100 m suivant la méthode PBIM

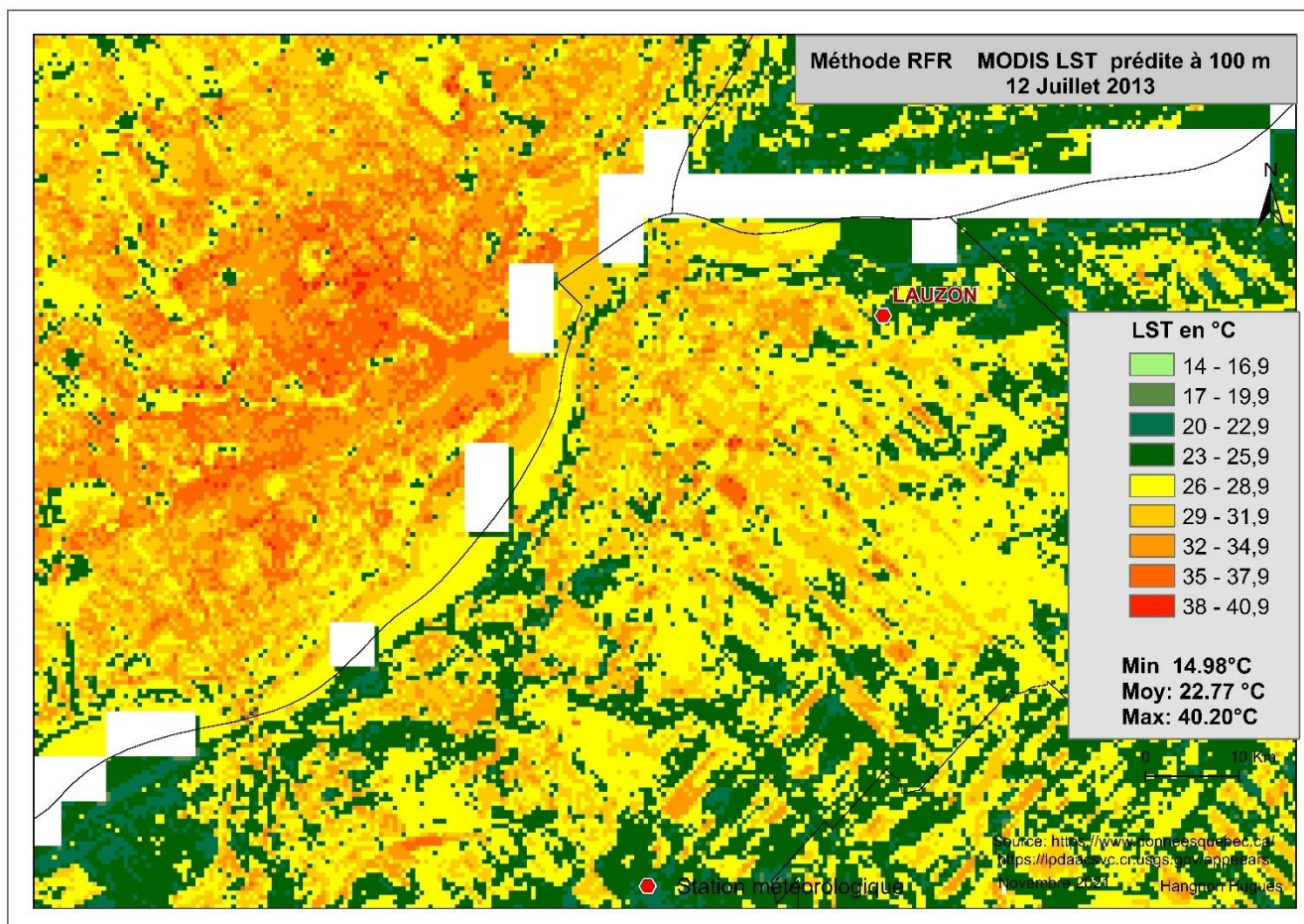
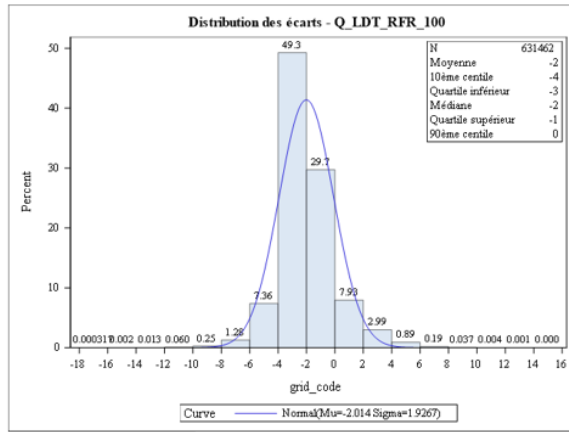
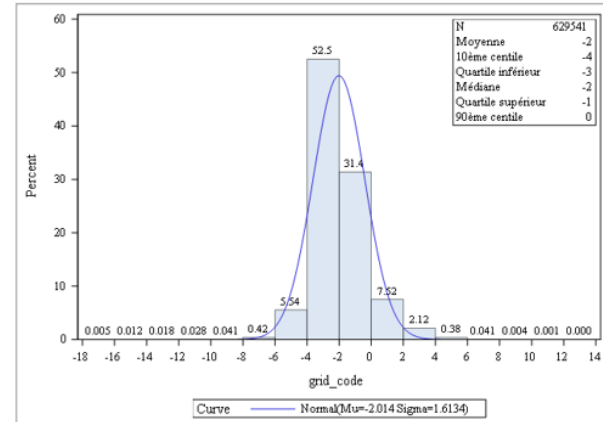


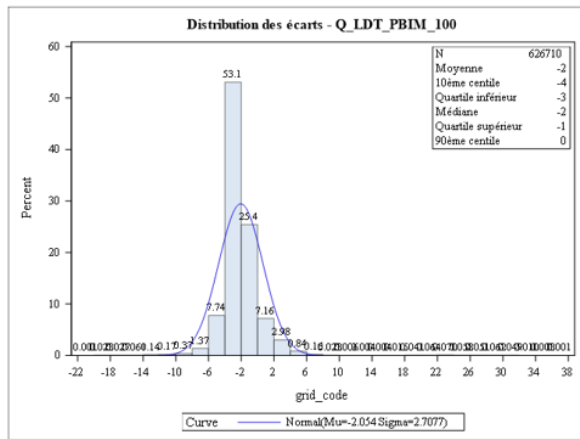
Figure 0.5 : Région de la Capitale nationale : Zoom sur la variabilité de la température de surface (°C) MODIS réduite de 1000 m à 100 m en milieu urbain suivant la méthode RFR



A: Distribution des écarts de LST RFR 100 m



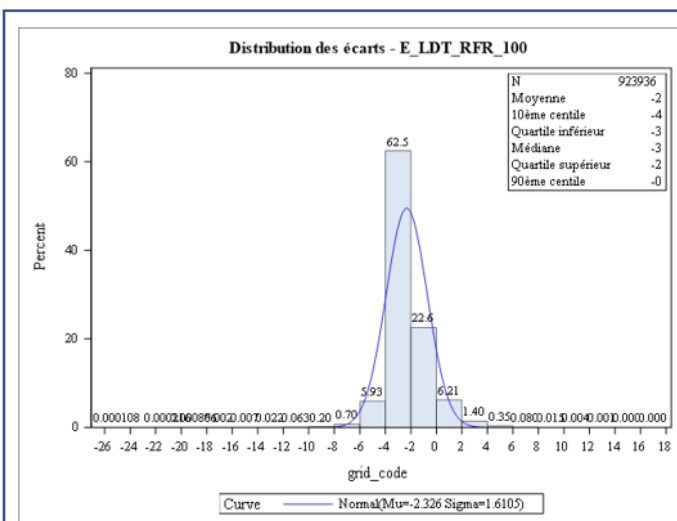
B: Distribution des écarts de LST TsHARP 100 m



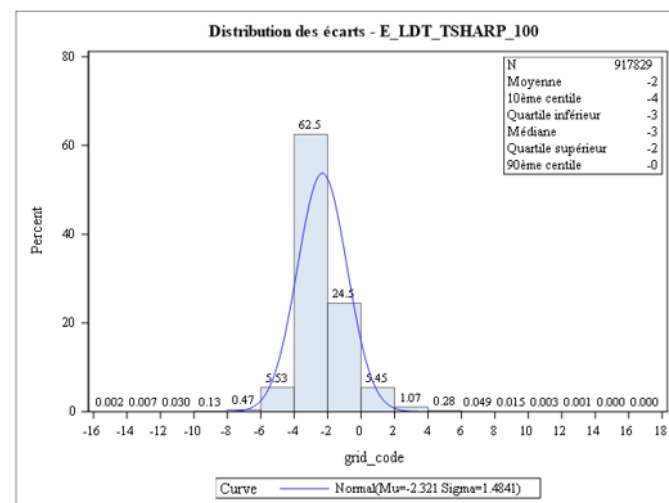
C: Distribution des écarts de LST PBIM 100 m

Région de Québec	RFR	<u>TsHARP</u>	PBIM
Moyenne °C	-2,01	-2,01	-2,05
Médiane °C	2,25	2,24	-2,36
Ecart -Type °C	1,92	1,61	2,7
Variance °C	3,71	2,6	7,33

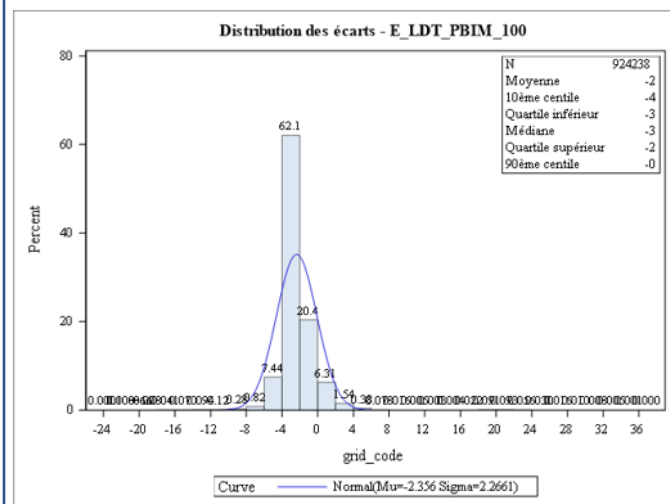
Figure 0.6 : Région de la Capitale-Nationale : distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite à 100 m



A: Distribution des écarts de LST RFR 100 m



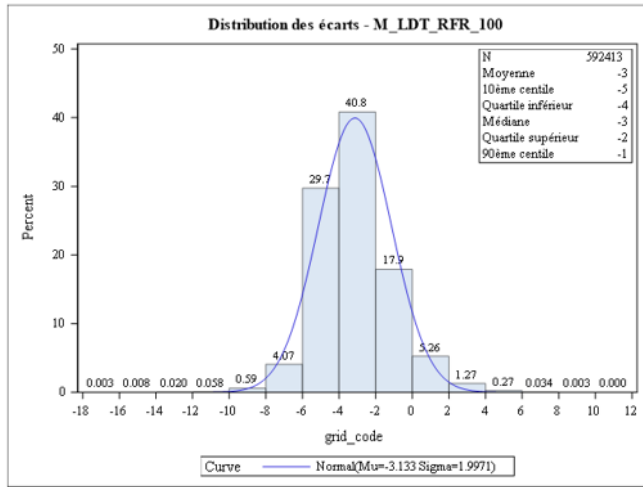
B: Distribution des écarts de LST TsHARP 100 m



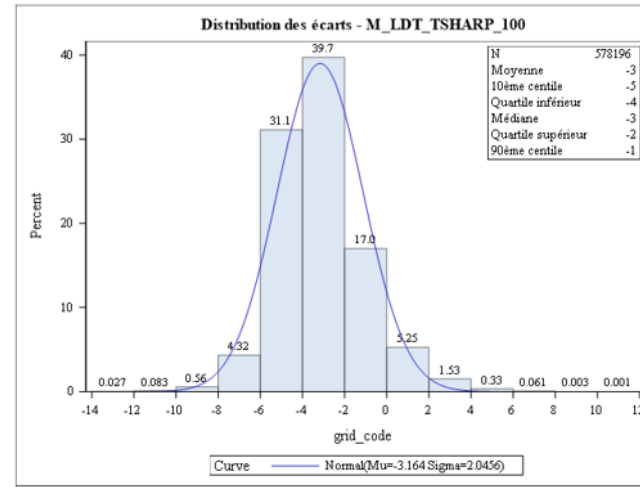
C: Distribution des écarts de LST PBIM 100 m

Région de	RFR	TsHARP	PBIM
<b>ESTRIE</b>			
Moyenne °C	-2,32	-2,32	-2,35
Médiane °C	-2,61	-2,58	-2,72
Ecart -Type °C	1,61	1,48	2,26
Variance °C	2,59	2,2	5,13

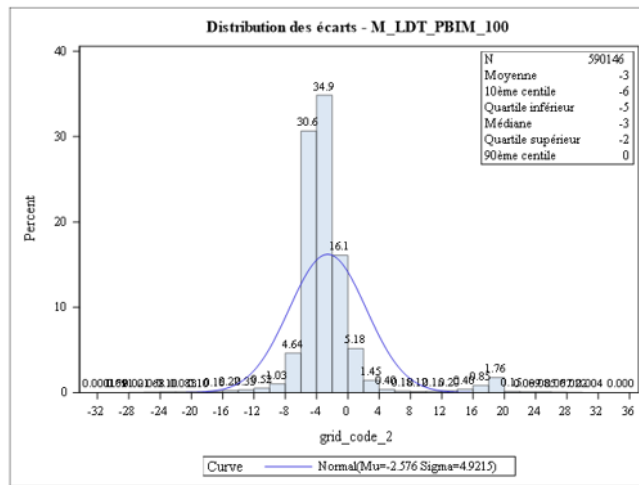
Figure 0.7 : Région de l'Estrie : distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite à 100 m



A: Distribution des écarts de LST RFR 100 m



B: Distribution des écarts de LST TsHARP 100 m

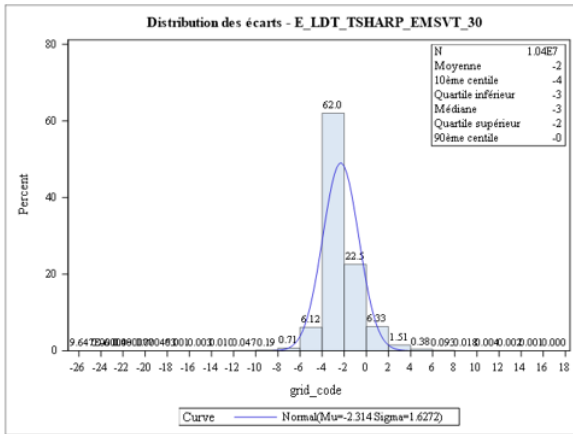


C: Distribution des écarts de LST PBIM 100 m

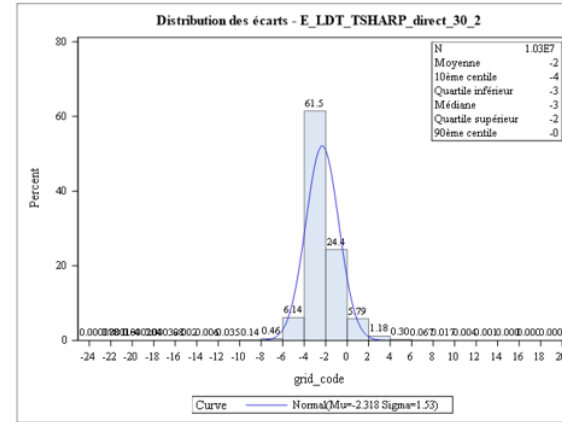
Région de Montréal	RFR	TsHARP	PBIM
Moyenne °C	-3,13	-3,16	-2,57
Médiane °C	-3,34	-3,4	-3,41
Ecart -Type °C	1,99	2,04	4,92
Variance	3,98	4,18	24,22

Figure 0.8 : Région de Montréal : distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite à 100 m

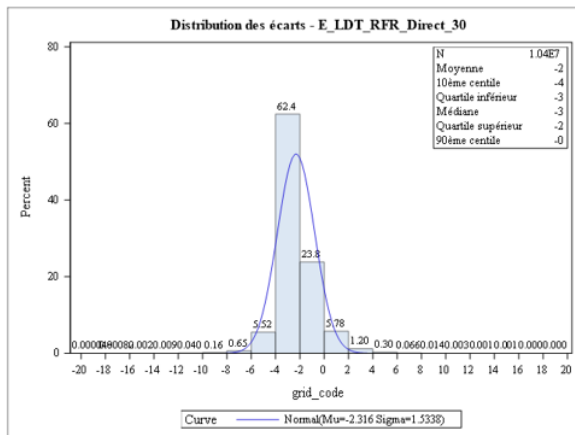




A: Distribution des écarts de LST TsHARP de 100 m à 30 m avec des émissivités de surface



B: Distribution des écarts de LST TsHARP de 1000m à 30 m

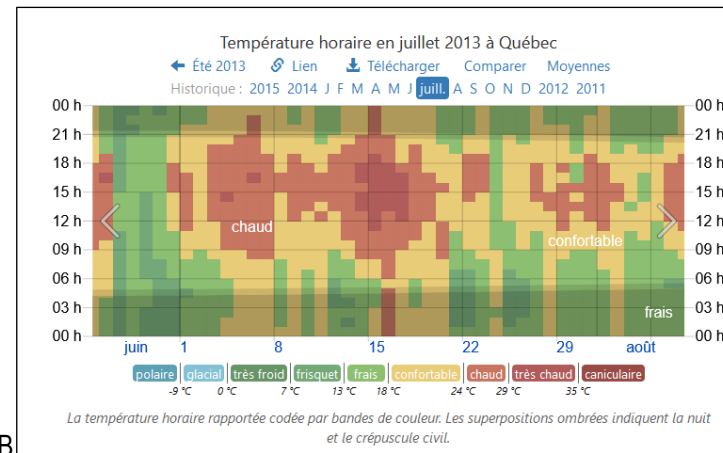
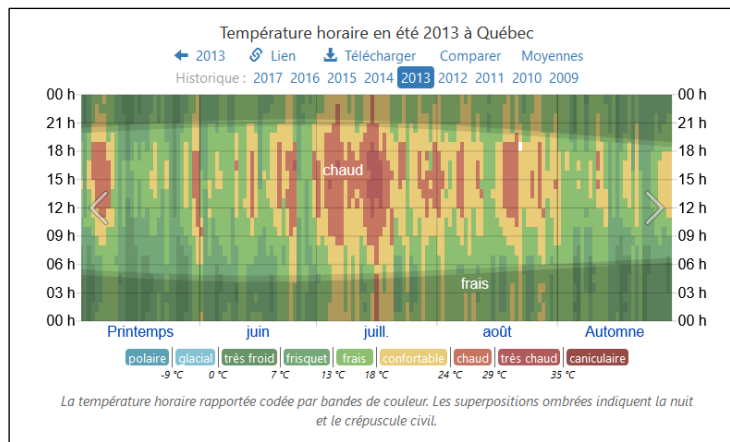


C: Distribution des écarts de LST RFR de 1000 m à 30 m

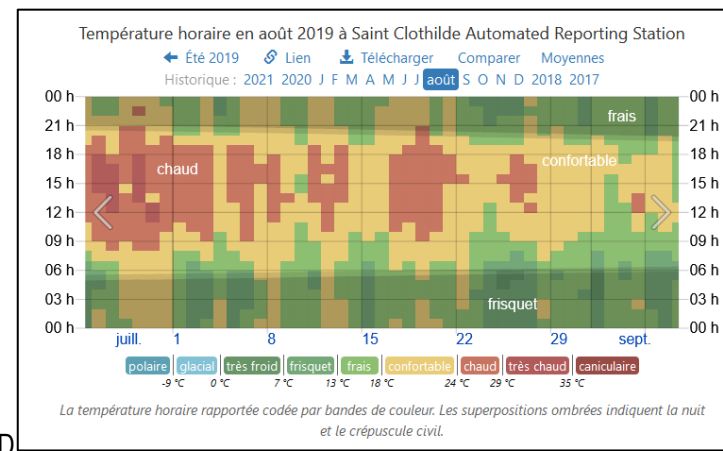
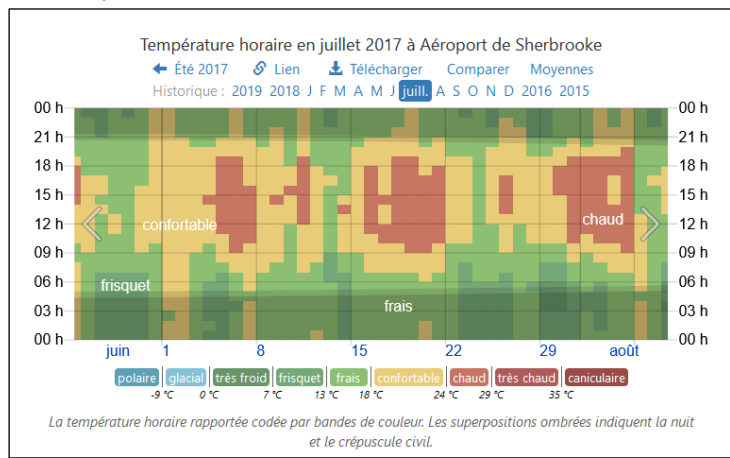
Région de ESTRIE	TsHARP Emissivité	TsHARP D 30	RFR D 30
Moyenne °C	-2,32	-2,31	-2,31
Médiane °C	-2,58	-2,59	-2,58
Ecart -Type °C	1,48	1,53	1,53
Variance °C	2,2	2,34	2,35

D\_30: réduction d'échelle directe de 1000 m à 30 m

Figure 0.9 : Région de l'Estrie: distribution des écarts de LST Landsat 08 TIRS et MODIS LST réduite directement de 1000 m à 30 m



Source <https://fr.weatherspark.com/h/s/26469/2013/1/M%C3%A9t%C3%A9o-historique-en-%C3%A9t%C3%A9-2013-%C3%A0-Qu%C3%A9bec-Canada#Figures-ColorTemperature>



Source : <https://fr.weatherspark.com/h/m/147280/2017/7/M%C3%A9t%C3%A9o-historique-en-juillet-2017-%C3%A0-A%C3%A9roport-de-Sherbrooke-Canada#Figures-ColorTemperature>

Source : <https://fr.weatherspark.com/h/m/147205/2019/8/M%C3%A9t%C3%A9o-historique-en-ao%C3%BBt-2019-%C3%A0-Saint-Clothilde-Automated-Reporting-Station-Canada>

Figure 0.10 : Variation de la température horaire de l'air au niveau des stations météorologiques de nos zones d'étude

## Annexe B

Présentation des équations associées aux différentes méthodes de réduction d'échelle ou au calcul de LST ou à certains indices de végétation

Équation 0.1 : Formule générale de la méthode RFR

$$F(x) = \frac{\sum_{j=1}^N T_j(x)}{N}$$

Où N indique le nombre d'arbres, T<sub>j</sub> représente chaque arbre et F est une prédiction à un nouveau point x comme une prédiction moyenne basée sur les arbres créés.

Équation 0.2 : Calcul de la température de surface avec la méthode PBIM

$$LST_{HR} = \left( \frac{MODIS\ LST_{1000} * \epsilon_{HR}}{M \epsilon_{HR}} \right)$$

où LST<sub>HR</sub> est la température de surface cible à haute résolution, MODIS LST<sub>1000</sub> est la température de surface terrestre du pixel en résolution grossière.

ε<sub>HR</sub> est l'émissivité de surface à HR et Mε<sub>HR</sub> est la valeur moyennée de l'émissivité de surface. Le calcul de l'émissivité de surface est souvent basé sur les unités d'occupation et d'utilisation ou sur le seuil de NDVI.

Équation 0.3 : Calcul de l'indice proportion de végétation (PV)

$$p_v = \left( \frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \right)^2$$

Où P<sub>v</sub> est la proportion de végétation, NDVI<sub>min</sub> est la valeur minimale de NDVI calculé et NDVI<sub>max</sub> est la valeur maximale de NDVI calculé.

### Équations associées aux indicateurs des marges d'erreur

Équation 0.4 : Calcul de l'erreur moyenne absolue pour la LST

$$EMA = 1/N \sum_{i=1}^N (LST_s - LST_{ref})$$

Équation 0.5 : Calcul de l'erreur quadratique moyenne RMSE

$$RMSE = \sqrt{1/N \sum_{i=1}^N (LST_s - LST_{ref})^2}$$

$LST_s$  = LST simulée (réduite) et  $LST_{ref}$  = LST de référence

Équations associées à la conversion des données Landsat 08 TIRS en température de surface

Le processus de dérivation de la température de surface se fait en différentes étapes (cf. Figure 3.6): le calcul du rayonnement spectral ou radiance (Top of Atmospheric TOA) avec l'Équation 0.6, la conversion de la température TOA en luminosité (Équation 0.7), le calcul du NDVI (Équation 0.10), le calcul de la proportion de végétation (Pv) (Équation 0.8), le calcul de l'émissivité (Tableau 3.2) et enfin le calcul de la LST selon l'approche standard (Équation 0.9).

Équation 0.6 : Calcul de la radiance spectrale du capteur (sensor spectral radiance)

$$L_\lambda = M_L Q_{cal} + A_L$$

Où  $L_\lambda$  est la radiance spectrale du capteur (sensor spectral radiance);  $M_L$ = facteur de rééchantillonnage spécifique à la bande fournie dans les métadonnées,  $Q_{cal}$ = Produit standard quantifié et calibré DN du pixel, et  $A_L$ = facteur de redimensionnement additif spécifique à la bande décrite en métadonnées.

Équation 0.7 : Calcul de la température de luminosité (Brightness Temperature BT)

$$BT = \frac{k_2}{\ln\left[\left(\frac{k_1}{L_\lambda}\right)+1\right]} - 273,15$$

Où BT Brightness Temperature (température de luminosité);  $K_1$ ;  $K_2$  est le coefficient de conversion thermique spécifique respectivement aux bandes 10, 11 fourni avec les fichiers de métadonnées d'imagerie Landsat 8, et  $L_\lambda$  est la radiance spectrale du capteur (sensor spectral radiance).

Équation 0.8 : Calcul de la proportion de végétation

$$p_v = \left(\frac{NDVI - NDVI_{min}}{NDVI_{max} - NDVI_{min}}\right)^2$$

Où  $P_v$  est la proportion de végétation,  $NDVI_{min}$  est la valeur minimale de NDVI calculé et  $NDVI_{max}$  est la valeur maximale de NDVI calculé.

Équation 0.9 : Calcul de la température de surface

$$LST = \frac{BT}{\left(1 + \left(\frac{\lambda BT}{\rho}\right) \ln \epsilon \lambda\right)}$$

Où BT est la température de luminosité en Celsius ( $C^\circ$ ),  $\lambda$  est la longueur d'onde moyenne de bande 10 ou 11,  $\epsilon \lambda$  est l'émissivité calculée à partir de l'équation 5 et  $\rho$  est ( $h \times (c / \sigma)$ ) qui est =  $1,438 \times 10^{-2}$  mk dans laquelle,  $\sigma$  est la constante de Boltzmann =  $1,38 \times 10^{-23}$  J/K,  $h$  est la constante de Plank =  $6,626 \times 10^{-34}$  Js et  $C$  est la vitesse de lumière =  $2.998 \times 10^8$  m/s.

Cette partie de l'annexe B est consacrée à la description opérationnelle des méthodes TsHARP et RFR.

## Description opérationnelle de la méthode de réduction d'échelle à l'aide de TsHARP

Le processus de réduction d'échelle décrit ici est celui basé sur l'algorithme développé par Agam et al. (2007) avec l'hypothèse d'invariabilité de la relation entre la LST et le NDVI à toute résolution spatiale.

**Étape 1** : L'algorithme de TsHARP est un modèle de régression linéaire entre la LST à une résolution grossière et le couvert végétal décomposé suivant les valeurs de NDVI à une résolution fine dans une image satellitaire (Équation 0.10). Le NDVI est utilisé comme covariable dans la technique TsHARP pour dériver la LST correspondante à une résolution spatiale fine.

**Étape 2** : À cette étape, des résidus de LST générés à l'échelle grossière sont rééchantillonnés et intégrés à la LST modélisée à une fine résolution.

**Étape 3** : L'erreur résiduelle du pixel à une résolution grossière est ajoutée à la température réduite pour augmenter la précision.

Équation 0.10 : Calcul de l'indice NDVI

$$NDVI = \frac{(NIR - RED)}{(NIR + RED)}$$

NDVI est : Normalized Difference Vegetation Index

R et PIR représentent respectivement la réflectance dans la bande spectrale rouge (0.62-0.78µm) et la bande proche infrarouge (0,78-3µm)

NDVI min et NDVI max représentent respectivement la valeur minimale et la valeur maximale estimée.

Équation 0.11 : Fraction de couverture végétale

$$f_{cs} = 1 - \left( \frac{NDVI_{max} - NDVI}{NDVI_{max} - NDVI_{min}} \right)^{0.325}$$

$f_{cs}$  désigne la fraction de couverture végétale

Équation 0.12 : Mode de calcul

$$f_{cs} = 1 - (1 - NDVI)^{0.625}$$

Une régression des moindres carrés est effectuée entre TR et fCs, après que le NDVI eut été agrégé à la résolution thermique plus grossière (NDVI low) :

Équation 0.13 : Calcul de la température de surface (LST) à basse résolution avec l'indice NDVI

$$\hat{TR}(NDVI\ low) = a_0 + a_1 f_{cs\ low} = a'_0 - a'_1 (1 - NDVI\ low)^{0.625}$$

Avec  $a$  et  $a'$  comme le paramètre de régression et le  $\hat{TR}$  indique une valeur de température prédite à l'aide de l'équation de régression de l'indice de végétation (VI). Cette relation de régression est ensuite appliquée aux données NDVI à leur résolution native plus fine (NDVI high) :

Équation 0.14 : Calcul de LST à haute résolution avec l'indice NDVI

$$\dot{TR}(NDVI\ high) = a'0 - a'1(1 - NDVI\ high)^{0.625}$$

Les erreurs dans la régression des moindres carrés, dues à des forces entraînant des températures de surface autres que la quantité de couverture végétale (par exemple, l'humidité du sol), peuvent être évaluées à l'échelle grossière avec l'Équation 0.15:

Équation 0.15 : calcul des résidus de LST à basse résolution

$$\Delta \dot{TR}_{low} = TR_{low} - \dot{TR}(NDVI\ low)$$

Le champ résiduel à l'échelle grossière est réintégré dans la carte affinée, de sorte que le champ de température d'origine est récupéré par réagrégation. Les températures des sous-pixels affinés dans chaque pixel grossier sont donc calculées par l'intermédiaire de l'Équation 0.16:

Équation 0.16 : calcul de la LST définitive à haute résolution

$$TR\ high = TR(NDVI\ high) + \Delta TR_{low}$$

où le premier terme du côté droit est évalué à l'aide de l'Équation 0.14, et le second terme résiduel de l'**Erreur ! Source du renvoi introuvable.** est constant sur la zone du pixel grossier.

### Description opérationnelle de la méthode de réduction d'échelle à l'aide de Random Forest Regression

La mise en œuvre de la réduction d'échelle de la température de surface (LST) avec le RFR se réalise en trois étapes : i) la simulation de LST à échelle grossière, ii) la correction des résidus de LST et iii) l'estimation de LST à haute résolution.

- i) Simulation de LST à échelle grossière

Le modèle de régression RF est traduit mathématiquement par l'Équation 0.17:

Équation 0.17 : Modèle de RFR

$$y = F(x)$$

Où y est la variable dépendante, la LST et x la variable explicative, les prédicteurs et F () la fonction non linéaire entre les deux variables. L'équation peut intégrer autant de variables explicatives selon leur disponibilité. Généralement les types de variables de prédiction utilisées sont : la réflectance de surface ( $\rho$ ), les indices spectraux (S), les facteurs terrain (TF), l'occupation des sols. Sur la base des types de variables prédictives cités, cette équation prend la forme suivante :

Équation 0.18 : Modèle de RFR appliqué à la LST avec les différentes variables

$$LST_c = F((\rho_i)_c, (s_i)_c, (TF_i)_c, (LC)_c)$$

Où l'indice c représente la résolution grossière et l'indice i la nième variable

Les variables prédictives d'entrée (à résolution fine) sont agrégées à la résolution grossière de LST MODIS en utilisant une méthode de moyenne spatiale. À ce stade, la relation statistique entre les variables explicatives et la LST peut être établie à un niveau grossier.

ii) Correction des résidus (Erreur d'estimation du modèle)

Le modèle de régression RF n'explique pas toute la variation spatio-temporelle de la distribution de LST, alors une correction résiduelle est effectuée pour améliorer la LST prédite à haute résolution par ajout à l'équation (E<sub>1</sub>) de LST simulée. La température résiduelle se calcule à partir de la différence entre LST originale -LST<sub>0</sub>- (résolution grossière) et la LST simulée LST<sub>c</sub> (Équation 0.19). La correction résiduelle est appelée l'erreur d'estimation du modèle.

$$\Delta LST_c = LST_0 - LST_c$$

Équation 0.19 : Calcul des résidus

Par conséquent, à partir de la LST à résolution grossière, l'équation (E<sub>1</sub>) de la LST simulée à résolution grossière (LST<sub>c</sub>) devient:

Équation 0.20

$$LST_c = F((\rho_i)_c, (s_i)_c, (TF_i)_c, (LC)_c) + \Delta LST_c$$

iii) Dérivation LST à haute résolution

La LST finale à haute résolution est estimée par l'Équation 0.21 après un rééchantillonnage de l'erreur d'estimation du modèle à une haute résolution. Étant donné l'invariance d'échelle entre les relations entre la LST et les variables prédictives, l'erreur d'estimation du modèle rééchantillonnée à haute résolution est attribuée à chaque pixel.

$$LST_f = F((\rho_i)_f, (s_i)_f, (TF_i)_f, (LC)_f) + \Delta LST_c$$

Équation 0.21

Où f représente la haute résolution.

## Annexe C

### *I. Analyse d'impacts des données massives de télédétection (DMS) sur la méthode de réduction d'échelle*

L'optimisation des méthodes de travail reste un objectif primordial dans tout travail. C'est cette véritable question qui se pose avec les gros volumes de données de télédétection à traiter tout en prenant en compte leur variété spécifique tant au contenu qu'à la source de provenance. L'algorithme automatisé de réduction d'échelle permet d'aboutir à un résultat final à partir de son exécution. Toutefois, pour savoir si cette solution est optimale pour les DMS afin de conclure sur l'adaptabilité de cette solution aux DMS, il faudrait faire des tests dans ce sens. Ainsi, des expérimentations supplémentaires sont nécessaires pour tester sa performance et faire de potentielles améliorations.

Par ailleurs, dans la chaîne de traitement de la réduction d'échelle de LST, l'exécution des tâches n'est pas linéaire et successive. Bien que certaines activités soient préalables et dépendantes d'autres, certains traitements sont indépendants et peuvent se réaliser parallèlement. Par exemple, la classification de l'occupation et l'utilisation des sols sont indépendantes du calcul des variables prédictives. De même, la création des pentes et de l'orientation est aussi indépendante du reste des activités. La réalisation du raster des émissivités de surface est consécutive à la classification de carte des occupations du sol. La définition de l'ordre des opérations est primordiale pour permettre leur fluidité et de décharger les entrées et sorties.

Nous avons identifié des contraintes ou obstacles liés aux DMS pour lesquelles des approches de solution sont répertoriées. Les contraintes courantes sont relatives : i) aux données, ii) aux méthodes de traitement et d'analyse des données.

- i. Les obstacles liés aux DMS elles-mêmes sont leur disponibilité, leur structuration, qualité et sécurité. En effet, les données existent et sont accessibles sur des plateformes de fournisseurs. Face à leur grande quantité, de plus en plus des données sont prétraitées selon des normes prédéfinies, converties en cube de données et stockées dans des bases de données adaptées aux DMS telles Rasdaman, PostGis Raster,... Ces données converties en des formats web selon les normes de l'Open Geospatial Consortium OGC sont accessibles en ligne sur des plateformes web. Cependant, la qualité des données mises à disposition et prêtes pour l'utilisation reste sujette à débat en absence de consensus sur les normes de qualité (Baumann, Rossi, et al., 2018). La tendance générale est le stockage des données en ligne ; ce qui pose le problème de la sécurité et la souveraineté des données et des résultats obtenus à partir d'elles, cette problématique découle aussi des éventuels cas de panne électrique ou le refus du fournisseur de permettre l'accès aux données. Dans le cadre de la méthode de réduction d'échelle, lequel des accès aux données par zone d'intérêt ou par tuile permet l'optimisation de l'algorithme ? Au regard de notre expérience dans les différentes expérimentations, l'accès par tuile pourrait être facilité par la disponibilité des données selon les dates d'acquisition.
- ii. Dans le domaine de traitement et d'analyse des DMS, les obstacles sont surtout la faible capacité des outils et méthodes à intégrer et assimiler les DMS pour l'exécution des opérations à moindre coût et obtenir de bons résultats. Ces limites ont donné lieu à l'avènement de nouvelles méthodes de traitement/analyse et le développement de nouvelles infrastructures informatiques. Comme nouvelles méthodes nous avons l'apprentissage profond (Deep Learning) ou l'apprentissage automatique (Machine Learning) relevant de l'intelligence artificielle.

L'intelligence artificielle désigne des machines et techniques qui donnent un sens réel au monde, en automatisant des processus qui créent des aperçus évolutifs à partir de grandes données (Goodfellow, 2017 ; Minichiello, 2017). Son implémentation dans le domaine géospatiale (geoAI) procure d'importants avantages dans la modélisation, notamment à cause de leur capacité à intégrer de grandes quantités de données spatiales



et temporelles sous différents formats (VoPham et al., 2018). Le geoAI utilise les algorithmes d'apprentissage machine, d'apprentissage automatique et profond, ainsi que l'exploration des données pour extraire des informations significatives à partir des données volumineuses. Les avancées de cette science dans la classification des images satellitaires et de modélisation prédictive du trafic ont été présentées au cours d'un colloque aux USA en 2017. Les recherches en cours dans le geoAI portent sur le développement des méthodes de réduction d'échelle des données spatio-temporelles massives (VoPham et al., 2018). L'algorithme de RFR utilisé dans cette étude qui est un algorithme d'apprentissage automatique (Machine Learning) fait aussi partie de geoAI. Cet algorithme a déjà été utilisé pour les DMS (Fox et al., 2020 ; Talebi et al., 2022)

L'importance de développement de nouvelles méthodes de gestion est le gain de temps dans le traitement des données massives. Ainsi, de nouvelles approches ou composantes informatiques comme le recours aux supercalculateurs (calculateur de haute performance HPC), le calcul distribué ou encore les techniques d'encapsulation, l'équilibrage des charges (Load Balancing) ont été développées pour paralléliser les traitements. Avant la présentation des méthodes de gestion de DMS, nous présentons d'abord la structuration des DMS.

## II. Structuration des données massives spatiales en cube de données

Dans le domaine de la télédétection, les données massives d'observation de la terre sont organisées **en cube de données spatiales** (Spatial data cube) pour valoriser leurs dimensions spatio-temporelles (latitude, longitude et temps). La Figure 0.1 ci-dessous montre deux cas de cube de données matricielles en trois et quatre dimensions. Cette structuration se fait en utilisant une indexation ou une ingestion (Augustin et al., 2019). Ainsi, un cube de données est un tableau multidimensionnel (temps, lieu, typologie) structuré dans lequel les variables sont indexées pour faciliter le stockage, l'organisation, l'accès, la gestion et l'analyse des données. L'application de la technologie des cubes de données est une solution dans la gestion des DMS notamment celles des multicapteurs d'observation de la terre (Augustin et al., 2019 ; Baumann, Misev, et al., 2018 ; Kasprzyk & Devillet, 2021). Le stockage des données est fait suivant le critère d'accès au sein d'une même structure de données. EarthServer, Digital Earth Australia (jadis Australian Geoscience Data Cube -AGDC-), Earth system Data Cube de l'Agence spatiale européenne ou les services d'accès aux données et à l'information -DIAS- de Copernicus sont quelques exemples de cube de données géospatiales (Dhu et al., 2019 ; Gomes et al., 2020 ; Killough, 2018 ; Lewis et al., 2017).

Dans la perspective d'optimisation de gestion des DMS, Augustin et al. (2019) propose un cube de donnée sémantique d'observation de la terre axée sur la disponibilité des informations catégorielles (nuage, végétation, eau, autres) à des fins de requêtes dans la même instance. De fait, un utilisateur peut sélectionner dans un cube de données des mosaïques sans nuages, ou une mosaïque ne contenant que l'eau. La plus-value est de disposer des blocs de bases interoperables d'informations thématiques extraites dans un cube de données.

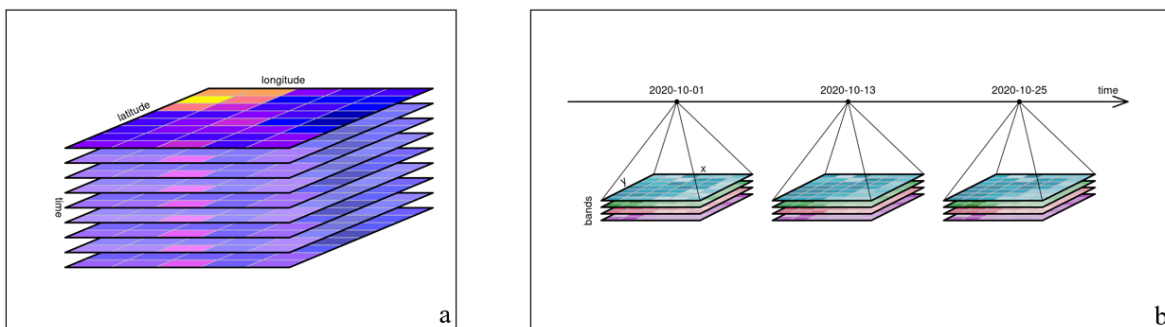


Figure 0.1 : Exemple de cube de données matricielles à 3 dimensions (a) et à 4 dimensions (b). Les trois dimensions, la latitude, la longitude et le temps. Les bandes spectrales ajoutées donnent la dimension 4.

Source : <https://keen-swartz-3146c4.netlify.app/datacube.html>

La tendance actuelle des cubes de données est de rapprocher les données satellitaires des utilisateurs en fournissant des données prétraitées (mosaïquées calibrées et prêtes à l'utilisation) comme celles de la collection 2, niveau 2 de Landsat ; accessibles en ligne sur les plateformes en « nuage » (Cloud).

En conséquence, l'utilisateur n'a plus besoin de télécharger des images satellitaires et de les stocker sur son ordinateur avant de les exploiter. De plus, d'autres plateformes comme Google Earth Engine proposent une plateforme de calcul au détriment des données (Sudmanns et al., 2020). Dès lors, les cubes de données d'observation de la terre se présentent comme une solution alternative pour suppléer les contraintes techniques liées au stockage, l'organisation, d'archivage et d'analyse des données rasters.

**Dans le cas de la réduction d'échelle**, les cubes de données d'observation de la terre réalisées avec les données prêtes à l'utilisation **revêtent un important intérêt dans le contexte de gros volumes de données**. En effet, peu de temps serait consacré à la préparation des données qui occupe les  $\frac{3}{4}$  du processus de la réduction d'échelle. Ainsi, l'accès à des données prétraitées ou des indices dérivés des images optiques comme le NDVI, la pente, l'altitude, l'occupation du sol, etc. impacterait positivement la performance de l'algorithme de réduction d'échelle. De manière précise, il y aurait **une plus-value en temps d'exécution, l'allocation des ressources et la consommation d'énergie**. De plus la capacité d'extraire des données chronologiques mosaïquées sur de grandes zones au-delà des limites des tuiles des images satellitaires serait un autre impact positif. En situation d'urgence en cas de catastrophes sanitaires liées aux maladies de Lyme et du VNO, l'utilisation des cubes de données prétraitées est une opportunité pour la production des données multiterritoires en continu et en temps réel.

### *III. Méthodes de gestion des données massives spatiales (rasters) DMS*

S'agissant des méthodes de traitement des données, étant donné que dans la méthodologie de réduction d'échelle certaines tâches peuvent être faites séquentiellement, d'autres de manière autonome ou d'autres encore consécutivement à la suite des résultats de précédents traitements, alors il faudrait des techniques évolutives, tolérantes et permettant d'équilibrer les charges (Kolajo et al., 2019 ; Mazzetti et al., 2016 ; Rapiński et al., 2019). En effet, des techniques évolutives sont celles capables d'intégrer de nouvelles données et la tolérance se rapporte aux pannes dont leur survenance ne peuvent handicaper le fonctionnement des opérations. L'équilibrage de la charge permet aux techniques d'être adaptatives et d'éviter le délestage (Kolajo et al., 2019 ; L. Zhao, 2016). Ainsi, des techniques d'encapsulation, la conteneurisation et l'équilibrage des charges peuvent être une solution à explorer. Ces techniques permettent en environnement distribué des traitements parallèles, l'équilibrage des charges pour rendre autonome et performante le système opérationnel de gestion des données massives. L'équilibrage des charges ou répartition de charge assure l'optimisation de la capacité du système par le partage de charges de travail sur plusieurs serveurs afin de limiter d'éventuelles surcharges ou pannes du système. Tandis que, la conteneurisation (conteneurs Docker)<sup>28</sup> est une plateforme qui permet un regroupement d'algorithmes et ses dépendances en un seul lieu. Son avantage est d'éviter tout blocage de tout le système en cas de panne d'un des conteneurs. Les conteneurs de par leur fonctionnement augmentent l'optimisation des ressources et infrastructures dans le traitement des DMS. Cette technologie permet de réduire les coûts de calculs. Il existe d'autres techniques en ligne comme Google Earth Engine ou Earthserver. Le cadre opérationnel de cartographie de risque (Figure 0.2 ci-dessous) peut permettre des traitements parallèles dont les résultats seront incorporés dans le système global.

Par ailleurs **la manipulation de gros volume de données** et la mise à jour fréquente de la cartographie de risque orientent vers la mise en place des traitements automatisés et parallèles sur différents serveurs. À ce niveau pour l'automatisation, on peut recourir à l'intelligence artificielle (apprentissage machine, apprentissage profond).

---

<sup>28</sup> <https://www.docker.com/resources/what-container/>

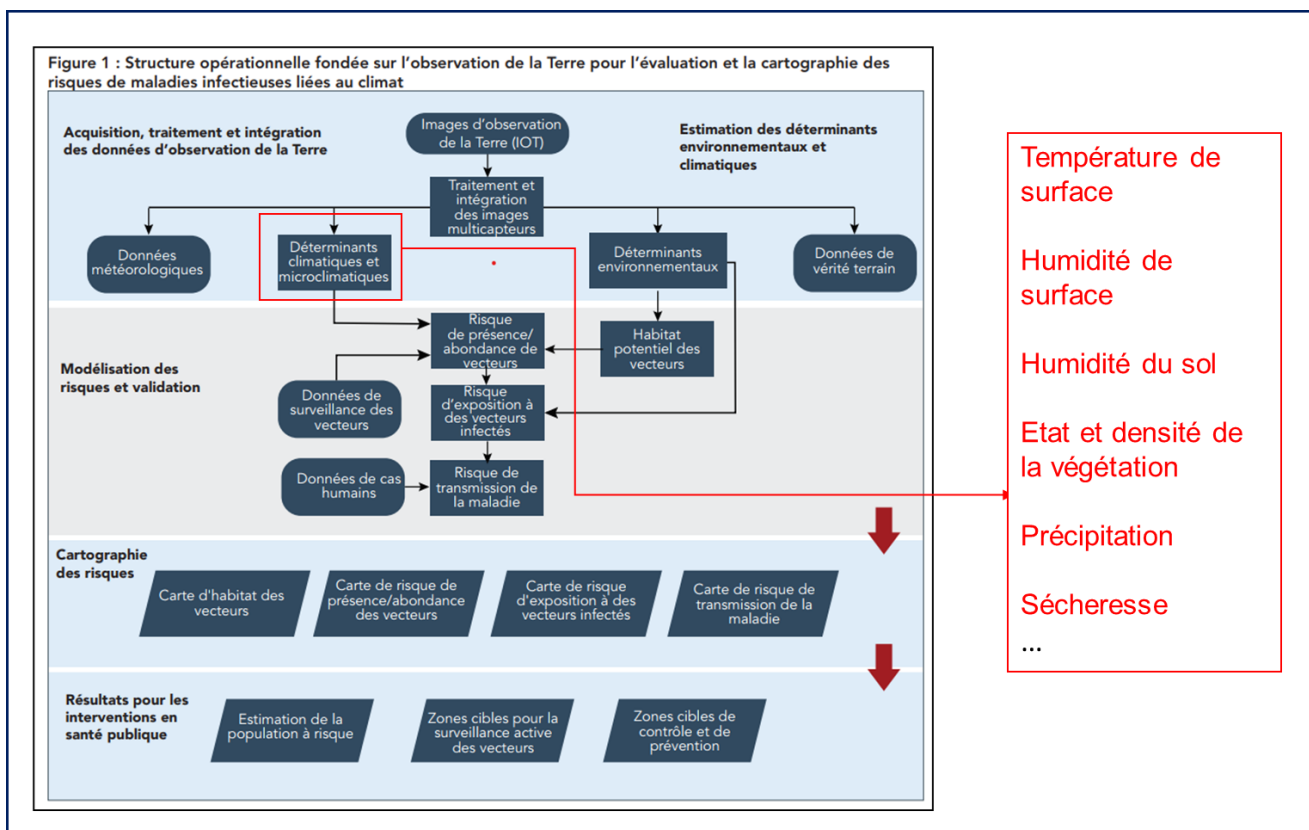


Figure 0.2 : Cadre opératoire de production de la cartographie des risques de maladies infectieuses liées au climat

Source : Kotchi et al., 2019