
Analyses de la sensibilité du territoire aux fortes chaleurs des espaces naturels

Etude de cas du Parc Naturel Régional des Vosges du Nord

CARMIER Morgane¹, FRIESS Clément¹, LEDERMANN Quentin¹, THILL Ella¹

1. Master 2 Observation de la Terre et géomatique – 3 rue de l'Argonne, 67000 Strasbourg – adresses mails

CETTE ÉTUDE ANALYSE LA SENSIBILITÉ AUX FORTES CHALEURS DU PARC NATUREL RÉGIONAL DES VOSGES DU NORD EN ADAPTANT LA CLASSIFICATION DES ZONES CLIMATIQUES LOCALES (LCZ) AUX MILIEUX RURAUX. EN CONFRONTANT L'OUTIL STANDARDISÉ GEOCLIMATE À UNE MÉTHODE SUPERVISÉE PAR RANDOM FOREST ET À L'EXPLOITATION DE L'OCs GRAND EST, LES TRAVAUX METTENT EN ÉVIDENCE LES LIMITES DES APPROCHES AUTOMATISÉES EN MILIEU PEU DENSE. L'INTÉGRATION DE DONNÉES LiDAR VIA LE MACHINE LEARNING PERMET TOUTEFOIS D'AFFINER LA DÉTECTION DU BÂTI AVEC UNE PRÉCISION DE 72,4 %. CES RÉSULTATS SOULIGNENT L'IMPÉRATIF D'ADAPTER LES PROTOCOLES DE MODÉLISATION URBAINE AUX SPÉCIFICITÉS DES TERRITOIRES NATURELS POUR ORIENTER EFFICACEMENT LES POLITIQUES D'AMÉNAGEMENT DURABLE.

MOTS-CLÉS : GeoClimate, Machine Learning, Random Forest, Low Climate Zone

1. Introduction

Dans un contexte de changement climatique avec des conditions climatiques plus intenses, l'organisation urbanistique des villes et des villages est un aspect qui gagne en importance depuis le début des années 2000. Avec des canicules estivales plus fréquentes, des pluies plus intenses ou même des périodes de sécheresse extrêmes, ce sont les aménagements des villes et villages qui permettent aux habitants de garder un cadre de vie le plus agréable possible.

Néanmoins, ce ne sont pas uniquement les espaces urbains qui sont touchés par les températures croissantes, mais également les espaces naturels comme les Parcs Naturels Régionaux en France.

Même si ces territoires, majoritairement rural, sont probablement moins sensibles face au changement climatique que les territoires métropolitains, leur exposition aux vagues de chaleur entraîne également une augmentation de la température de l'air, des cours d'eau et des lacs ; les conséquences néfastes pour les espaces naturels ne doivent pas être négligées.

Une option pour étudier les différentes zones d'occupations du sol face aux fortes chaleurs est la représentation des *Local Climate Zones (LCZ)* conçues et présentées dans un article par Steward et Oke au début des années 2010. Les LCZ représentent un système de classification universel de régions de couverture de sol, de matériaux et d'activités humaines uniformes à une échelle locale. Ces zones sont normalement définies de quelques centaines de mètres à plusieurs kilomètres et chaque zone est conçue pour standardiser les études des îlots de chaleur urbain. Cette étude s'inscrit dans le cadre du remplacement de la distinction binaire « urbaine – rurale » par 17 catégories standards, dont 10 sont des types de bâtis et 7 sont des types de couverture terrestres. Les zones se différencient par des propriétés physiques spécifiques, telles que la géométrie du bâti, la perméabilité du sol et la chaleur anthropique qui génèrent un régime thermique distinct à hauteur d'abri. (I.D. Steward et T.R. Oke, 2012)

C'est en 2021, qu'un article présentant un outil de calcul de ces LCZ à partir de données open-source est publié. L'outil de *GeoClimate* conçu par un projet franco-suisse mené par le chercheur CNRS Erwan Bocher du laboratoire LAB-STICC de Bretagne-Sud présente une solution libre de droit, de calcul des LCZ. Il s'agit d'un outil de facile prise en main, qui permet la création automatisée des indicateurs

climatiques sur l'ensemble d'un territoire d'étude à partir de données locales issues de la BD Topo ou encore à partir des données globales d'OpenStreetMap. (E. Bocher et al., 2021)

GeoClimate a un fonctionnement qui s'effectue par deux systèmes d'unités spatiales : la Référence Spatiale Unitaire (RSU) et le quadrillage régulier. De ces deux unités de calcul possibles, celle du RSU est la plus précise comme elle prend en compte les caractéristiques géographiques et topographiques spécifiques du territoire. L'outil de *GeoClimate* a pour objectif final de sortir divers indicateurs, les LCZ, de la zone d'étude, permettant par la suite de d'estimer des îlots de chaleur urbain (ICU). (E. Bocher et al., 2021)

Les îlots de chaleur désignent des effets de création d'une sorte de dôme d'air chaud qui couvre une ville. En milieu urbain, les températures de l'air dépendent principalement des températures ambiantes de la région. Néanmoins, elles sont augmentées par la forme, l'architecture, les matériaux de construction et encore la quantité de végétation que présente ces milieux urbains. Plus précisément, les immeubles bloquent plus ou moins la circulation de l'air et les matériaux captent les chaleurs, les températures en ville sont plus élevées et créent ainsi cet effet d'îlot de chaleur urbain. (L. Kouadria, 2025)

Face aux limites inhérentes aux outils automatisés dont la structure interne reste parfois opaque, cette étude mobilise également une approche de classification supervisée basée sur l'algorithme *Random Forest*. Introduite par L. Breiman (2001), cette méthode d'apprentissage automatique (*Machine Learning*) est largement reconnue en télédétection pour sa robustesse et sa capacité à traiter des jeux de données complexes (G. Biau & E. Scornet, 2016). Contrairement à l'approche standardisée de *GeoClimate*, le *Random Forest* offre une maîtrise complète des variables d'entrée. Il permet notamment de dépasser la simple utilisation de la BD Topo en intégrant des données tridimensionnelles fines issues du LiDAR HD ainsi que des signatures spectrales extraites d'orthophotographies. Cette flexibilité méthodologique vise à affiner la distinction des classes LCZ, en prenant en compte la hauteur réelle de la végétation et la rugosité du bâti, des paramètres cruciaux pour la caractérisation microclimatique d'un territoire de parc naturel.

L'étude qui fait l'objet de cet article se focalise sur le cas du Parc Naturel Régional des Vosges du Nord. Comme les PNR en France sont des espaces de protection et de conservation d'espaces naturels fragiles face à cette augmentation des températures (J.B. Bouron, 2024).

La question de recherche suivante se pose ainsi : Comment peut-on détecter les espaces sensibles aux fortes chaleurs face aux enjeux du réchauffement climatique, dans le cas d'un Parc Naturel Régional ?

L'étude vise à faire des comparaisons intra et inter années des classifications en Zones Climatiques Locales de la région d'étude.

L'étude vise ainsi à faire une analyse à plusieurs échelles des Local Climate Zones dans le Parc Naturel Régional des Vosges du Nord entre 2010 et 2025.

Il s'agit de faire une comparaison entre les résultats obtenus par l'outil *GeoClimate* et le calcul supervisée *Random Forest* pour l'année 2025.

L'étude vise à donner une solution simple et fiable à la détection des Zones climatiques locales dans un contexte rural.

2. Présentation du site d'étude et des données

2.1. Le site d'étude

Cette étude de la sensibilité du territoire aux fortes chaleurs se focalise sur le cas du Parc Naturel Régional des Vosges du Nord. En France, les Parc Naturels Régionaux sont des espaces gérés par un syndicat mixte de collectivités locales, dans un but de protection et de conservation d'espaces naturels fragiles, tout en promouvant un développement économique rural. (J.B. Bouron, 2024)

Le Parc Naturel Régional des Vosges du Nord a été créé le 30 décembre 1975 dans l'actuelle région du Grand Est, à cheval sur la Moselle (57) et le Bas-Rhin (67). A cette époque, le parc se constituait de 98 communes, mais depuis 2014, il se compose de 111 communes, dont 37 se trouvent en Moselle et 74 dans le Bas-Rhin. (PNR Vosges du Nord, 2025)

Le PNR des Vosges du Nord, occupe ainsi 127 666 hectares, dont 83 535 hectares sont des forêts (66%) et 15 955 hectares (13%) sont des espaces protégés. La Figure 1 permet de situer le Parc Naturel Régional des Vosges du Nord dans son contexte plus global. (PNR Vosges du Nord, 2026)



Figure 1. Carte de localisation du PNR des Vosges du Nord (Q. Ledermann, 2026)

2.2. Les données disponibles

La réalisation de cette étude se compose en deux grandes parties, celle des calculs des Zones Climatiques Locales par l'outil *GeoClimate*, ainsi que celle des calculs des *LCZ* par la méthodologie de *Random Forest*.

Premièrement, pour le calcul des *LCZ* par l'outil *GeoClimate*, les seules données nécessaires sont celles de la BD Topo de l'Institut National de l'Information Géographique et Forestière (*IGN*). L'outil fonctionne également avec des données *OpenStreetMap*, cependant, comme les données sont plus précises en provenance de l'*IGN* et comme la zone d'étude se trouve uniquement sur le territoire français, la réalisation du travail est donc plus pertinente sur des produits français. Plus précisément, ce sont la version 2, pour les années 2008 à 2018, et la version 3, pour les années depuis 2019, de la BD Topo qui ont été utilisées pour la réalisation de ce projet. En ce qui concerne l'utilisation de la version 2 de la BD Topo, treize indicateurs ont été gardés pour l'insertion dans l'outil, tandis que pour la version 3 ce sont quatorze indicateurs qui ont été retenus. Le récapitulatif des indicateurs selon la version peut être consulté sur le Tableau 1.

Tableau 1. Indicateurs utilisés pour le calcul des LCZ à l'aide de GeoClimate (E. Thill, 2026)

VERSION BD TOPO	SOURCE & DATE	INDICATEURS
BD TOPO VERSION 2	IGN 2008 - 2018	BATI_INDIFFERENCIE
		BATI_INDUSTRIEL
		BATI_REMARQUABLE
		COMMUNE
		CONSTRUCTION_SURFACIQUE
		ROUTE
		PISTE_AERODROME
		SURFACE_ACTIVITE
		SURFACE_EAU
		SURFACE_ROUTE
		TERRAIN_DE_SPORT
		TRONCON_DE_VOIE_FERREE
		ZONE_DE_VEGETATION
		BD TOPO VERSION 3
BATIMENT		
CIMETIERE		
COMMUNE		
CONSTRUCTION_SURFACIQUE		
EQUIPEMENT_DE_TRANSPORT		
PISTE_AERODROME		
RESERVOIR		
SURFACE_HYDROGRAPHIQUE		
TERRAIN_DE_SPORT		
TRONCON_DE_ROUTE		
TRONCON_DE_VOIE_FERREE		
ZONE_DE_VEGETATION		
ZONE_D_ACTIVITE_OU_D_INTERER		

Deuxièmement, la méthodologie de calcul des LCZ à l'aide de la méthode de *Random Forest* nécessite trois données différentes. Pour la réalisation de cette partie, il y a les données du LiDAR HD, les Orthophotos IRC (infra-rouge couleur) et les données de l'occupation du sol à grande échelle (OCS GE) qui ont été utilisées. Les trois types de données sont également de provenance de l'IGN et datent de 2025 et ont dû être soumises à des prétraitements, pour les calculs des LCZ.

3. Méthodologie

3.1. GeoClimate

3.1.1. Fonctionnement de GeoClimate

GeoClimate est un logiciel développé en Java, téléchargeable sous la forme d'un exécutable compatible avec tous les types de systèmes d'exploitation. Sa diffusion est ainsi facilitée par un téléchargement rapide depuis la page GitHub du projet.

L'algorithme suit une chaîne de traitement en quatre étapes. Il commence par une division de la zone géographique en deux types d'unités spatiales : un maillage régulier (ex. carré de 100m), ou des *RSU* (Unité Spatiale de Référence). Ces unités spatiales divisent l'espace sur la base de contraintes physiques : axe de routes, voie ferrée, limite de végétation, de surface d'eau, de mer, ou encore les limites administratives et les zones urbaines.

GeoClimate nettoie les données pour chaque unité avec une fusion des géométries adjacentes sur un plan en 2D et gère les priorités (ex. un bâtiment à un poids plus important que la pelouse sur laquelle il est situé, pour éviter de comptabiliser deux fois la même surface on conserve l'emprise la plus importante). Certaines géométries sont seuillées comme la végétation et les surfaces hydrographique pour ne pas prendre d'élément non influent et d'autres sont simplifiées comme les fonction urbaine.

L'algorithme va pouvoir à partir de ce système de référence commencer le calcul des indicateurs morphologiques en extrayant les caractéristiques physiques de chaque unité. Pour cela trois indicateurs sont calculés : la densité du bâti (*BSF*) qui représente le pourcentage de surface au sol occupé par des bâtiments, la hauteur moyenne, qui donne un indicateur sur la dimension verticale des

bâtiments et le *Sky View Factor (SVF)* qui représente le ratio de ciel visible depuis le sol calculé selon l'analyse géométrique des obstacles environnants.

Ces premiers indicateurs vont poser la base d'un processus de classification suivant une logique de boucle de conditions pour chaque unité spatiale en testant d'abord la proportion de surface végétalisée sur la zone, si cette dernière est supérieure à 80% il compare alors les ratios pour définir le type de zone naturelle, il effectue ainsi plusieurs tests de positionnement successif, pour déterminer l'ensemble des classes. (Annexe 1)

Enfin le logiciel va filtrer la classification pour éviter d'obtenir des dispositions aléatoires des *LCZ* et obtenir une cohérence territoriale. Il analyse le voisinage de chaque maille afin d'éviter des changements d'occupation du sol trop abruptes, forêt au milieu d'une ville dense, et de lisser les zones pour rendre la carte réaliste et exploitable.

3.1.2. Prétraitement des données

La première étape est la mise en forme la donnée. Pour cette étude, quatre années distinctes (2010, 2015, 2020 et 2025) sont prises en compte et donc un regroupement des couches est nécessaire au traitement d'un seul et unique dossier. Selon la version de BD Top utilisée, V2 ou V3, des mises à jour sur l'organisation des couches ont été menées par l'IGN, pour faire des comparaisons de différentes dates différents fichiers de configuration sont créés menant vers les couches nécessaires de la BD Topo.

Le fichier de configuration est la base du traitement de *GeoClimate*, il est le seul par lequel les paramètres de traitement sont directement modifiables en donnant des dossiers spécifiques pour les couches d'entrée et de sortie, le *SCR* des données utilisées, ainsi que les différents indicateurs à calculer. Pour calculer les *LCZ* de l'entièreté du PNR, les traitements ont été spatialement divisés entre le Bas-Rhin et la Moselle, chaque fichier contenant les communes respectives du PNR sur le département.

3.1.3. Utilisation de *GeoClimate*

Le logiciel, outre les nombreuses entrées modifiables du fichier de configuration, est facile à comprendre. Avec des paramètres basiques trouvable sur le read-me de la page on peut facilement générer des *LCZ* sur une large zone, en seulement deux lignes de commande dans le terminal. Une première étape et de définir le dossier de travail qui doit contenir le .jar ainsi que le fichier de configuration, puis dans le terminal de commande (PowerShell sous Windows) une première ligne permettant l'activation du logiciel : « *java -jar geoclimate-1.0.0.jar -h* » et la seconde permettant le lancement de l'algorithme en appelant le fichier de configuration :

```
« java -jar Geoclimate-1.0.0.jar -f Config_geoclimate.json -w BDTOPO_V3 »
```

L'outil va calculer commune par commune les indicateurs et les exporte sous forme de GeoJSON dans un dossier nommé au format : « *bdtopo_n°version_CodeINSEE* ». Automatiquement des fichiers GeoJSON contenant les indicateurs clé sont créés avec « *rsu_indicators.geojson* », l'emprise de la zone d'étude, un fichier *LCZ* « *rsu_lcz.geojson* ». Voir Figure 2 l'onglet « données de sorties ».

GéoClimate

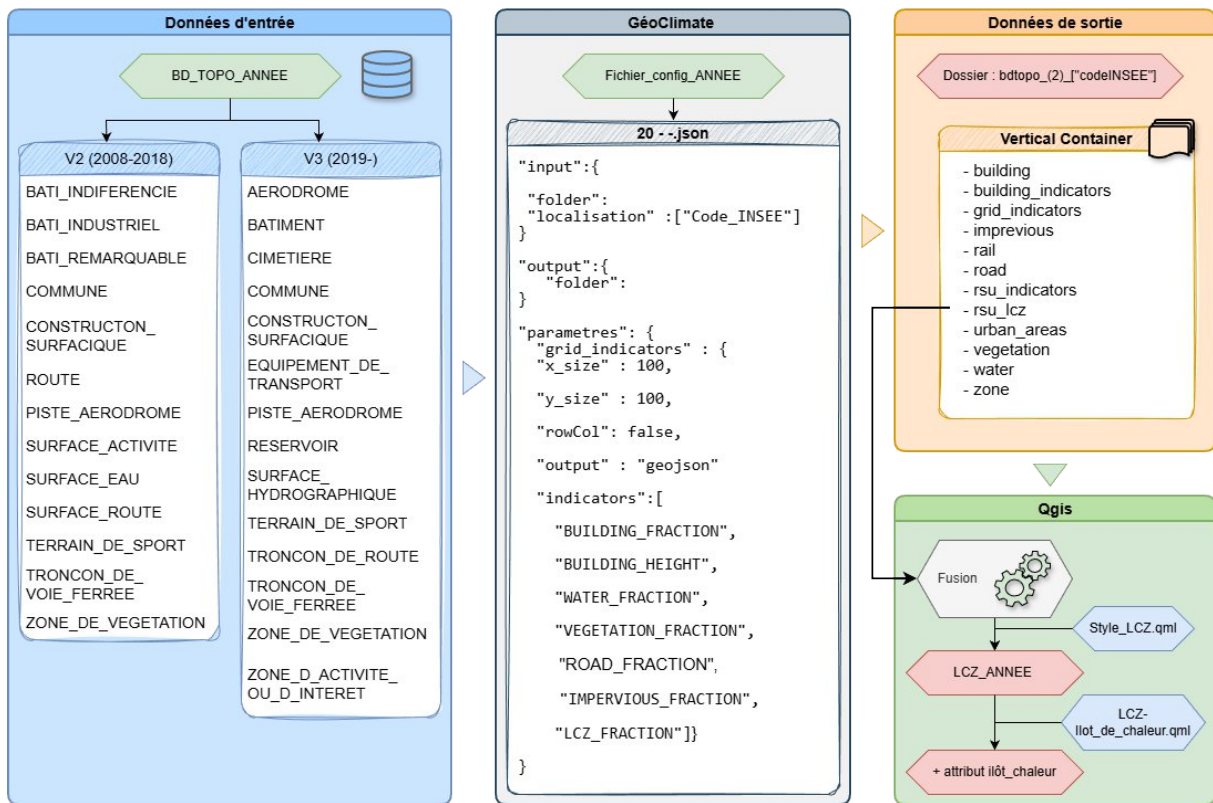


Figure 2. Organigramme de traitement pour GeoClimate (M. Carmier, 2026)

Le fichier « rsu_lcz » généré par GeoClimate ainsi que le « grid_indicators » permettent d'obtenir directement les LCZ sur l'emprise de la géographie d'entrée et en appliquant un fichier de style prévu par la documentation, de visualiser directement la classification.

Pour permettre un export rapide des données vers la géovisualisation nous avons supprimé les données vides des fichiers grid en 100 mètres sur 100 mètres et fusionné les GeoJSON de LCZ communale vers une entité unique.

3.2. Random Forest

La caractérisation des Zones Climatiques Locales nécessite de dépasser la simple observation visuelle pour quantifier la structure physique du territoire. À partir des données brutes (LiDAR et orthophotographie), une chaîne de traitement géomatique a été déployée pour constituer une base de données vectorielle. Cette étape a permis de calculer, pour chaque maille de la grille de référence, un ensemble d'attributs morphologiques et spectraux destinés à alimenter le modèle de classification.

3.2.1. Constitution de l'espace vectoriel

La discrimination des classes LCZ repose sur la constitution d'un vecteur de quinze variables (Tableau 2) prédictives calculées pour chaque maille de 100 m. Ce jeu de données a été construit en croisant trois sources d'information complémentaires extraites des données spatiales :

1. La Morphologie (3D) : Les modèles numériques de surface et de terrain (MNS/MNT) ont permis de dériver des indicateurs de rugosité urbaine, tels que la hauteur moyenne du bâti (H_{AVG}) et le Sky View Factor (SVF). Ce dernier est déterminant pour distinguer les tissus urbains compacts (LCZ 3) des zones ouvertes (LCZ 6).

2. L'Occupation du Sol : Des ratios de densité, tels que la fraction de surface bâtie (BSF) et le taux d'imperméabilisation (ISF), ont été calculés pour quantifier l'artificialisation des sols.
3. La Signature Spectrale : Les statistiques zonales (moyenne, écart-type) issues des bandes Infrarouge et Rouge de l'orthophotographie apportent une information texturale indispensable pour séparer les surfaces végétales complexes (forêts) des cultures.

Tableau 2. Les variables Random Forest (Q. Ledermann, 2026)

Variable	Couche d'entrée	Calcul
OCCUPATION DU SOL		
ndvi_std	NDVI	STD
rsf	Route	Mean
tree	Arbre	Mean
isf	Imperméabilité	Mean
GÉOMÉTRIE 3D		
bsf	Bâti	Mean
slope	Pente	Mean
svf	SVF	Mean
h_max	MNH	Max
b_num	Bati centroïde	Count
h_avg	MNH	Mean
SPECTRAL ET TEXTURE		
h_std	MNH	STD
red_mean	Ortho IRC	Mean
red_std	Ortho IRC	STD
PIR_mean	Ortho IRC	MEAN
PIR_std	Ortho IRC	STD

3.2.2. Modélisation par Random Forest

La classification a été opérée via l'algorithme Random Forest (Breiman, 2001), implémenté sous Orfeo ToolBox. Cette méthode d'apprentissage supervisé a été retenue pour sa capacité à hiérarchiser des variables hétérogènes (métriques géométriques et radiométriques) sans normalisation préalable. L'entraînement du modèle s'est fondé sur une vérité terrain constituée par photo-interprétation (50 polygones de référence par classe). Afin d'optimiser la performance du classifieur sur ce territoire complexe, les hyperparamètres ont été ajustés spécifiquement : le nombre d'arbres a été fixé à 200 pour la stabilité statistique, et la profondeur maximale (Max depth) portée à 25. Cette profondeur élevée permet au modèle de capturer des règles de décision fines, nécessaires pour discriminer les classes aux frontières sémantiques ténues (ex. transition verger/forêt). La Figure 3 ci-dessous, récapitule la chaîne de traitement mise en place.

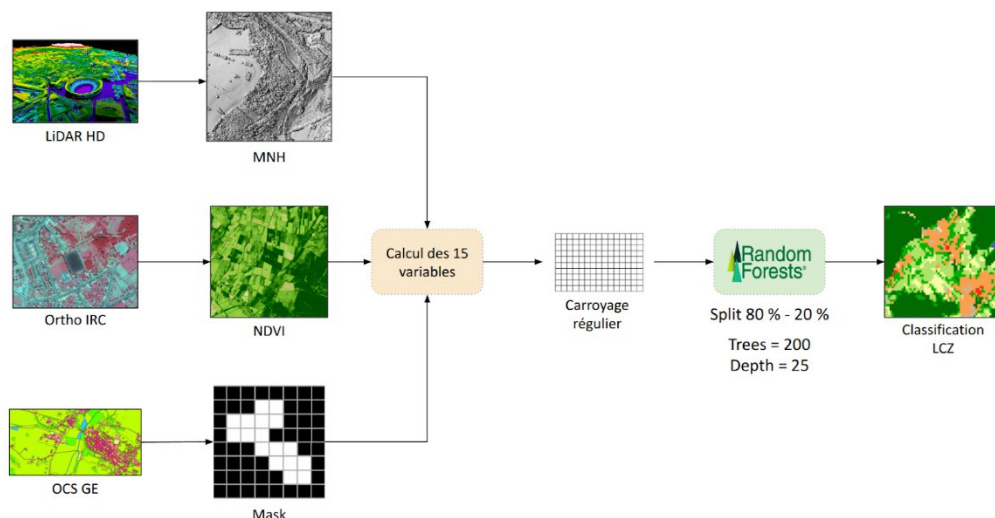


Figure 3. Organigramme de traitement Random Forest (Q. Ledermann, 2026)

3.2.3. Post-traitements

L'analyse des résultats bruts a mis en évidence certaines confusions thématiques, notamment une sur-détection de la classe industrielle (LCZ 8) au sein des tissus villageois. Pour garantir la cohérence géographique de la carte finale, un post-traitement déterministe a été appliqué. Des règles de décision logiques ont été implémentées pour contraindre le modèle par la réalité physique :

- Correction morphologique : Les pixels classés comme urbains mais dépourvus d'emprise bâtie réelle ($B_NUM = 0$) ont été réaffectés à la classe naturelle dominante.
- Correction contextuelle : Les zones classées en LCZ F (Sol nu). Cette classe a été utilisée pour éviter les confusions entre la végétation basse et les champs nus, ont été reclassées en végétation basse (LCZ D), corrigeant ainsi la symbologie et la rendant cohérente avec les méthodologies comme *GeoClimate* et *WUDAPT*.

4. Résultats

4.1. GeoClimate

Une classification LCZ a donc pu être réalisée pour 4 années distinctes (2010, 2015, 2020 et 2025) sur l'ensemble du PNR grâce à l'outil *GeoClimate*. En voici un exemple en Figure 4 pour l'année 2025.

Nous pouvons donc confirmer que le PNR est bien une zone à prédominance de classe de végétation en particulier par la LCZ A (Espace densément arboré) ou par de la LCZ D (Végétation basse).

Cette classification nous permet de détecter facilement les zones urbaines majeures présentes dans le PNR, avec Wissembourg au Nord-Est, Bitche au Nord ou encore Niederbronn-Les-Bains et Reichshoffen au Sud. Ces communes sont remarquables par l'amplitude des classes de LCZ urbaine qui sont les plus larges avec au niveau du cœur des communes certaines classes dites « compact ». Cette classification nous permet également de voir l'ensemble des zones bâties de la centaine de communes moins grandes souvent représenté par la classe LCZ 9 : Implantation diffuse et espacée de maisons.

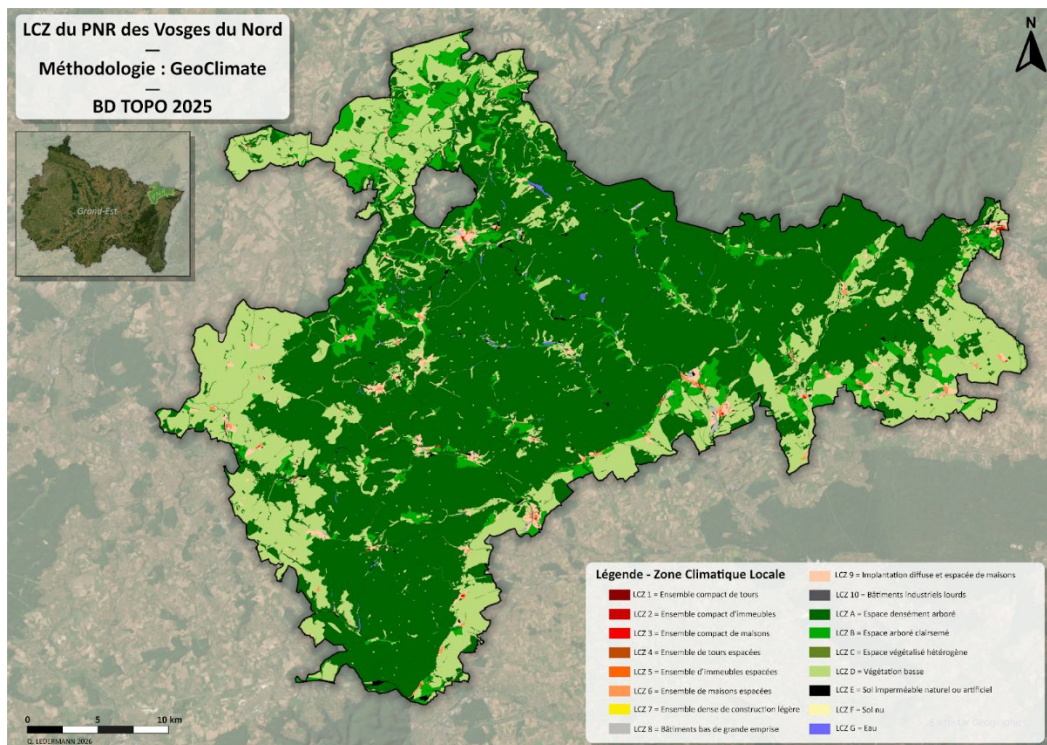


Figure 4. Carte des LCZ dans le PNR des Vosges du Nord pour 2025 (C. Friess et Q. Ledermann, 2026)

Naturellement le PNR est majoritairement constitué de classe de végétation donc observer des changements de classe bâti à cette échelle n'est pas pertinent. C'est cette raison que les graphiques représentant l'évolution ne représentent uniquement les classes de LCZ selon leur type (Bâti / naturel) sur nos 4 années étudiées.

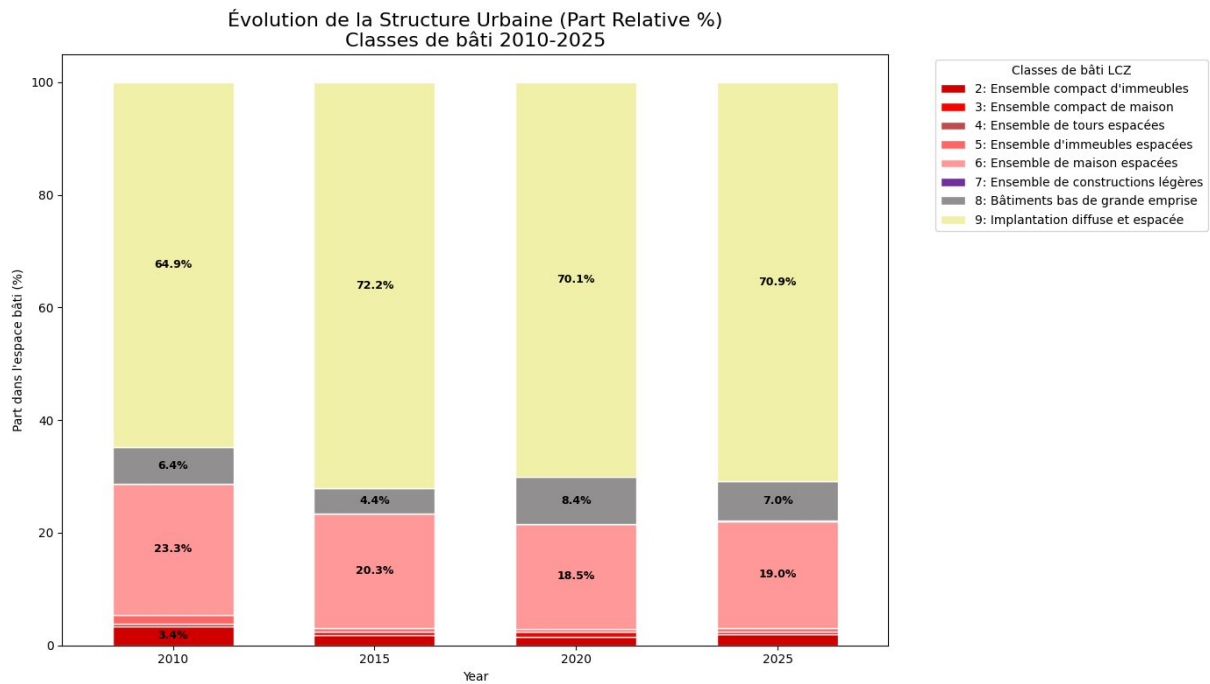


Figure 5. Graphique d'évolution des classes de bâti selon la part totale des surfaces bâties (C. Friess, 2026)

Avec la Figure 5 que la structure urbaine entre 2010 et 2025 est caractérisée par une hégémonie de l'implantation diffuse et espacée (LCZ 9), qui se stabilise à environ 71 % de la surface bâtie totale.

Les ensembles de maisons espacées (LCZ 6) constituent la deuxième catégorie la plus importante, malgré une légère érosion de leur part relative, passant de 23,3 % à 19,0 %. Parallèlement, les bâtiments bas de grande emprise (LCZ 8) affichent une dynamique fluctuante : après une croissance ponctuelle en 2020 (8,4 %), leur proportion s'établit à 7,0 % en fin de période.

Les zones urbaines denses et compactes (LCZ 2 à 5) occupent une place marginale et constante, représentant moins de 4 % de la répartition globale sur l'ensemble de la décennie.

On observe que les zones créées par la méthode de base de *GeoClimate* sont susceptibles de comporter des incohérences : notamment avec de la LCZ 2 dans des villes moyennes ainsi que certains quartiers orientés vers les extérieurs des agglomérations qui ne sont pas détectés.

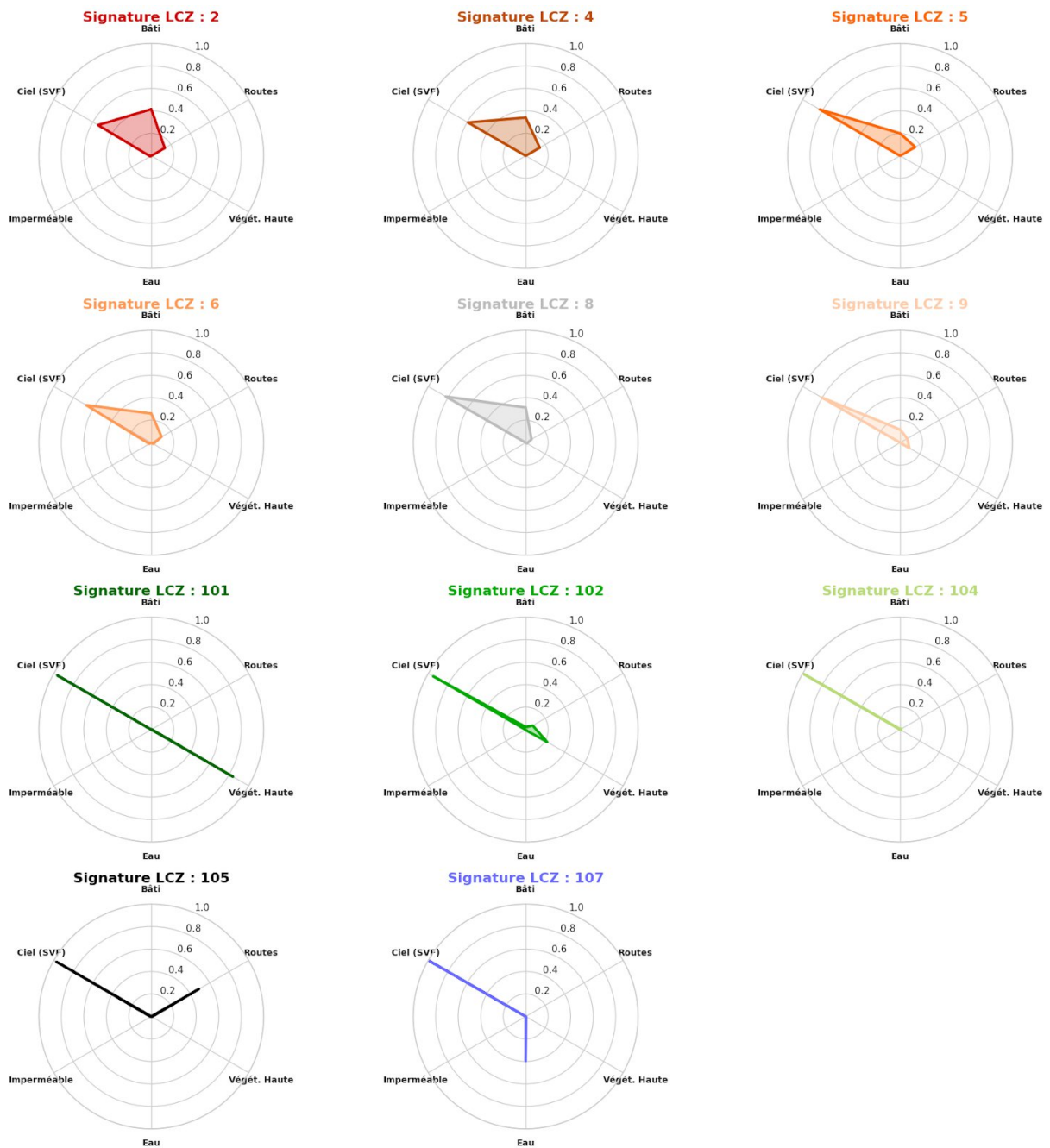


Figure 6. Graphiques des variables explicatives des LCZ (M. Carmier, 2026)

La Figure 6 démontre que le Sky View Factor constitue la variable la plus explicative pour la classification des LCZ. Toutefois, la distinction des zones urbaines repose exclusivement sur la densité du bâti, ce qui génère des biais d'identification. L'absence de prise en compte de la morphologie des toitures par GeoClimate provoque notamment une confusion entre les immeubles et les maisons à toiture haute.

4.2. Random Forest

4.2.1. Analyse spatiale des résultats

Les traitements réalisés via l'algorithme *Random Forest* sur l'ensemble du Parc Naturel Régional des Vosges du Nord permettent de dresser un état des lieux précis de l'occupation du sol et des régimes climatiques potentiels pour l'année 2025.

La cartographie issue de la classification (Figure 7) met en évidence la structure dichotomique du territoire. Le PNR se caractérise par une matrice paysagère largement dominée par la couverture forestière dense (LCZ A), qui occupe la quasi-totalité des reliefs vallonnés des Vosges du Nord. Cette continuité forestière constitue un réservoir de fraîcheur majeur à l'échelle régionale.

L'anthropisation du territoire apparaît strictement contrainte par la topographie. Les zones bâties s'organisent de manière linéaire le long du réseau hydrographique et des axes routiers de fond de vallée. On observe une typologie d'urbanisation en « chapelet », où les tissus villageois (LCZ 6) alternent avec des espaces agricoles ouverts (LCZ D), créant des corridors de ventilation naturelle. Les cœurs urbains denses (LCZ 3) sont rares et ponctuels, limités aux centres historiques des bourgs principaux (ex. Niederbronn-les-Bains, Wissembourg), tandis que les zones d'activités (LCZ 8) apparaissent rejetées en périphérie immédiate des agglomérations.

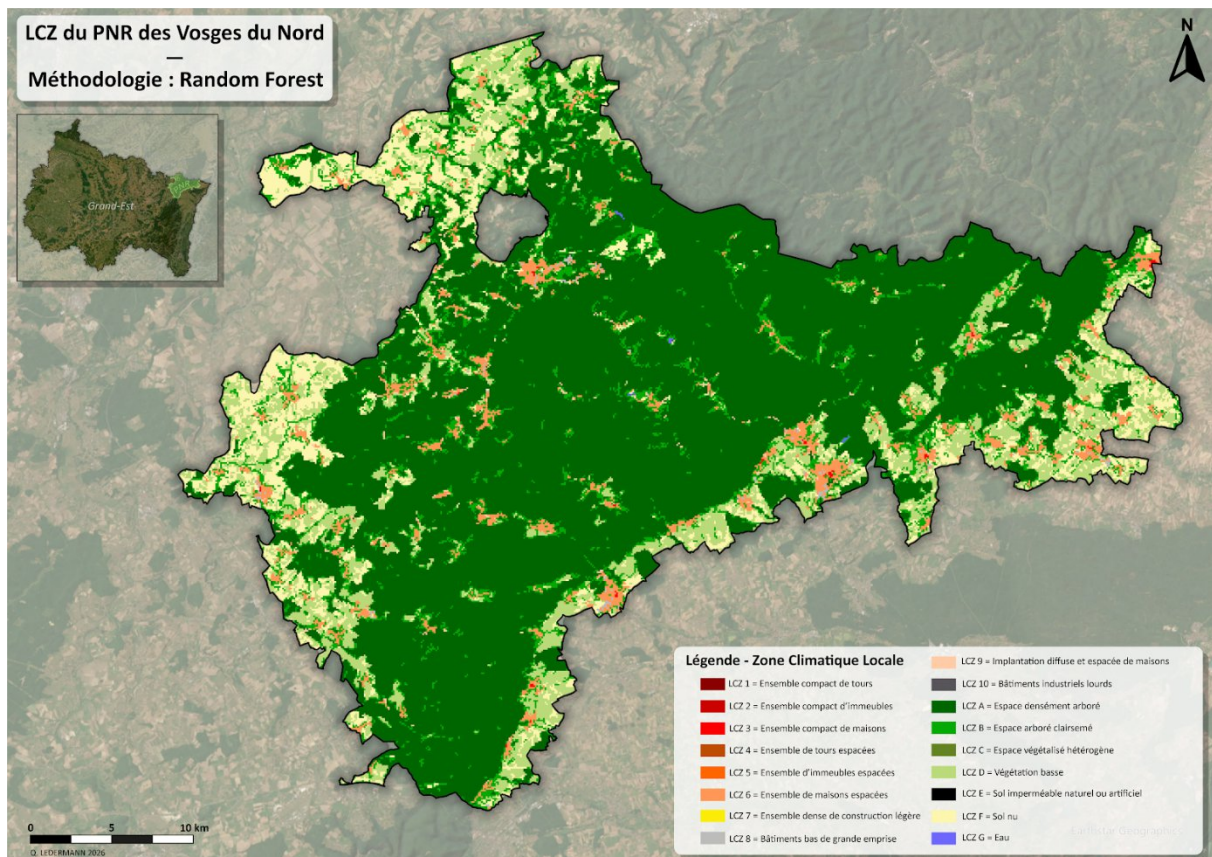


Figure 7. LCZ du PNR des Vosges du Nord par Random Forest (Q. Ledermann, 2026)

4.2.2. Bilan statistique de l'occupation du sol

L'analyse quantitative des surfaces (Tableau 3) confirme le caractère éminemment rural et naturel du parc. Les classes naturelles cumulent plus de 95 % de la surface totale du territoire étudié. La LCZ A (Arbres denses) est hégémonique, couvrant 83 055 hectares, soit 63,0 % du parc. Elle est suivie par la LCZ D (Plantes basses/Agriculture) avec 21,3 % et la LCZ B (Arbres dispersés) avec 10,8 %.

Concernant les surfaces artificialisées, elles restent minoritaires mais significatives en termes d'impact microclimatique local :

- Le tissu pavillonnaire ouvert (LCZ 6) représente la typologie urbaine dominante avec 4 739 hectares (3,6 %).
- L'habitat dispersé (LCZ 9), caractéristique du mitage périurbain, couvre 1,0 % du territoire.
- Les zones à fort potentiel de réchauffement, telles que les zones industrielles et commerciales (LCZ 8), restent contenues à 0,3 % (340 ha), mais constituent des points chauds localisés.
- Le bâti compact (LCZ 3) est anecdotique à l'échelle du parc (0,1 %), soulignant l'absence de grande agglomération dense intra-muros.

Tableau 3. Statistiques d'occupation du sol des LCZ (Q. Ledermann, 2026)

Code LCZ	Description (Type)	Surface estimée (ha)	Proportion (%)
Types Naturels		124 850,00	95,0 %
LCZ A (101)	Arbres denses (Forêt)	83 055,00	63,0 %
LCZ D (104)	Plantes basses (Agriculture/Prairies)	27 960,00	21,3 %
LCZ B (102)	Arbres épars (Vergers/Transition)	13 400,00	10,2 %
LCZ G (107)	Eau	435,00	0,3 %
Types Bâti		6 550,00	5,0 %
LCZ 6	Ouvert de faible hauteur (Pavillonnaire)	4 740,00	3,6 %
LCZ 9	Bâti dispersé (Habitat isolé)	1 315,00	1,0 %
LCZ 8	Grand bâti de faible hauteur (Indus/Co)	395,00	0,3 %
LCZ 3	Compact de faible hauteur (Centre-bourg)	100,00	0,1 %
TOTAL	Ensemble du Parc	131 400,00	100,0 %

4.2.3. Validation du modèle

La validation quantitative a été réalisée sur un échantillonnage de 655 points de contrôle indépendants. La matrice de confusion (Tableau 4) révèle une Précision Globale (*Overall Accuracy*) de 72,4 % et un coefficient Kappa de 0,68. Ces métriques témoignent d'une performance satisfaisante du modèle compte tenu de l'hétérogénéité du milieu périurbain, bien que des disparités existent selon les classes.

Les classes « stables » comme la forêt dense (LCZ 101) et l'eau (LCZ 107) obtiennent d'excellents scores de rappel (respectivement 94,7 % et 96,7 %), validant la robustesse des indicateurs spectraux. En revanche, les principales confusions se concentrent sur les zones de transition. La classe LCZ 102 (Arbres épars) présente un rappel plus faible (54,3 %), étant fréquemment confondue avec la forêt dense ou les zones pavillonnaires très végétalisées (LCZ 6). De même, le bâti dispersé (LCZ 9) peine à être distingué des surfaces agricoles (LCZ 104) en raison de la prédominance de la végétation au sein de la maille de 100 mètres (effet de *mixels*). Le post-traitement contextuel a néanmoins permis de réduire significativement la surestimation initiale des zones industrielles (LCZ 8).

Tableau 4. Matrice de confusion issue de Random Forest (Q. Ledermann, 2026)

Classe Réelle \ Prédite	LCZ 3(Compact)	LCZ 6(Open)	LCZ 8(Large)	LCZ 9(Sparse)	LCZ 101(Forest)	LCZ 102(Sparse Forest)	LCZ 104(Low Vege.)	LCZ 106(Bare soil)	LCZ 107(Water)	Total	Rappel
LCZ 3	41	9	4	0	0	0	0	1	0	55	74.5 %
LCZ 6	8	62	3	7	2	14	4	0	0	100	62.0 %
LCZ 8	3	4	31	0	0	0	2	5	0	45	68.9 %
LCZ 9	0	8	1	29	2	5	15	0	0	60	48.3 %
LCZ 101	0	1	0	1	142	6	0	0	0	150	94.7 %
LCZ 102	0	12	0	4	11	38	5	0	0	70	54.3 %
LCZ 104	0	3	2	11	1	6	72	5	0	100	72.0 %
LCZ 106	0	0	6	0	0	0	9	30	0	45	66.7 %
LCZ 107	0	0	0	0	1	0	0	0	29	30	96.7 %
Total (Pred)	52	99	47	52	159	69	107	41	29	655	
Précision	78.8 %	62.6 %	66.0 %	55.8 %	89.3 %	55.1 %	67.3 %	73.2 %	100 %		OA= 72.4 %

5. Discussion

5.1. Validation de la méthode de GeoClimate

5.1.1. Biais de la méthode

L'analyse des résultats produits par la méthode *GeoClimate* révèle à l'échelle du PNR des résultats très instables, si l'outil est performant en milieu urbain dense, son application à des tissus plus lâches révèle des limites structurelles. En effet, la classification *LCZ* via *GeoClimate* s'avère inadaptée aux territoires ruraux en raison d'un déséquilibre structurel entre les surfaces naturelles et bâties, où la dominance massive des classes de végétation, notamment la *LCZ A*, neutralise la sensibilité des calculs microclimatiques.

Cette instabilité engendre des incohérences spatiales tel que la présence de *LCZ 2*, correspondant à des zones de forte densité de bâtiment moyen (3 à 9 étages), comparable à des espaces urbains comme le centre de Paris, qui se retrouve au sein de territoires ruraux. Certaines petites communes voient leur densité largement surestimée et leur étalement urbain à l'inverse, sous-estimé, avec la disparition de nombreux quartiers résidentiels épars au profit de champs. Les limites de forêt dense sont elles aussi simplifiées, en s'arrêtant brusquement supprimant les espaces de transition moins dense précédant des espaces plus ouverts.

Les traitements introduisent aussi de nombreuses modifications non naturelles de *LCZ* entre les années, celle-ci sont dues aux versions de la BD Topo. En effet, même si le projet PAENDORA a permis l'intégration de la BD Topo V3 dans l'algorithme de traitement de *GeoClimate*, des biais sont introduits par la modification de la structure et des attributs des couches entre chaque année. Ce phénomène est particulièrement observable dans les résultats de la BD V2 de 2010, qui engendre une baisse significative de la qualité des résultats, dû à la précision de celle-ci qui n'est pas à la hauteur de ses versions ultérieures.

5.1.2. Origine des erreurs

En prenant en compte les résultats issus de la méthode *GeoClimate*, certaines incohérences mises en évidence se corrélaient directement avec la structure de la BD Topo. En effet, le réseau routier, en bordure de la tache urbaine, est souvent composé d'impasses qui ne forment pas de boucles fermées. Cette configuration contraint le logiciel à créer des *RSU* (unités de référence) disproportionnées, qui englobent à la fois des bâtiments et de grands espaces naturels.

Dans ces conditions, la végétation devient majoritaire dans le calcul de l'unité, et le logiciel classe alors ces secteurs comme des zones naturelles (type forêt ou champs). C'est une erreur de classification majeure : le bâti disparaît des statistiques au profit de la végétation, ce qui fausse la réalité du terrain.

Malheureusement, il n'y a pas vraiment de solution à proposer pour remédier à ces erreurs d'interprétation, les données d'entrée étant la BD Topo la correction de l'ensemble des attributs pour chacune des communes du PNR représenterait des traitements manuels trop longs.

5.1.3. Intégration du contexte rurale dans la méthode

Plusieurs pistes ont été explorées pour affiner la classification *LCZ* dans les communes rurales. Un premier test a consisté à appliquer un masque sur la végétation haute, avec pour objectif de réduire l'influence des strates arborées et de favoriser la détection du bâti. Les résultats n'ont toutefois pas été concluants : les classes de forêt restent prédominantes et continuent de saturer la classification dans les zones urbaines.

Une seconde approche a porté sur la modification des données de la BD Topo, en augmentant artificiellement la largeur des tronçons de route. Cette méthode visait à multiplier le découpage des *RSU* pour obtenir une maille plus fine et limiter les confusions spatiales. Cependant, cette modification

s'est avérée insuffisante. La structure du réseau routier rural, marqué par de nombreuses impasses, ne permet pas une fragmentation efficace des unités. En conséquence, les zones forestières conservent une emprise trop importante dans les calculs, empêchant une caractérisation correcte des secteurs bâtis

Une solution viable a finalement pu être explorée pouvant améliorer la classification en LCZ exclusivement dans les zones bâties. Plutôt que de prendre les limites communales comme emprise de base pour le calcul des LCZ, la solution s'est portée sur la création d'un polygone représentant exclusivement la tâche bâtie, puis relancer la classification pour une seule commune (Niederbronn-les-Bains).

Bilan, le résultat est bien plus pertinent voir Figure 8, des confusions existent toujours, mais chacune des zones contenant des bâtiments précédemment classés en zone de végétation sont maintenant classifiées comme implantation diffuse de bâtiment. Les résultats de cette solution restent naturellement discutables mais constitue toutefois une certaine amélioration.

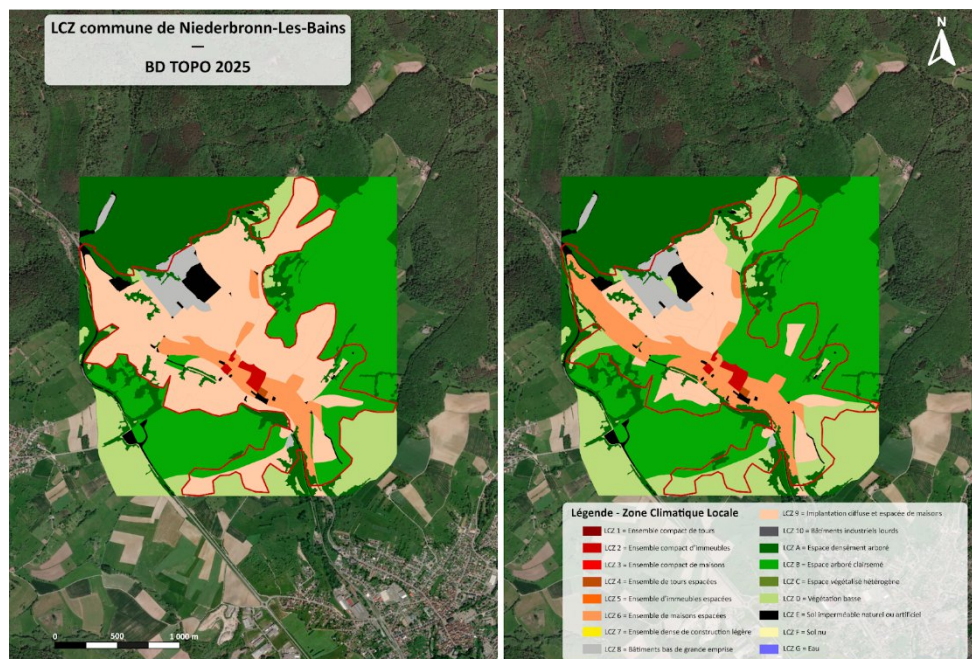


Figure 8. Représentation des LCZ de la commune de Niederbronn-les-Bains avec l'hypothèse d'amélioration de GeoClimate (C. Friess et Q. Ledermann, 2026)

Le contour rouge représente l'emprise de la tâche bâtie de la commune, sur la gauche de la figure on observe que les bâtiments sont ainsi mieux pris en compte lors du calcul des LCZ avec la nouvelle méthode, contrairement à la méthode originale qui sous estime l'importance des bâtiments en bordure de la tâche bâtie.

5.2. Les apports et les limites de Random Forest

5.2.1. Apport de la 3D et du post-traitement

La plus-value majeure de cette méthodologie réside dans l'intégration des données LiDAR (3D). Contrairement à une analyse purement optique, l'ajout de la hauteur du bâti et du Sky View Factor a permis de discriminer finement les typologies urbaines, notamment la distinction critique entre les centres villageois compacts (LCZ 3) et les zones pavillonnaires (LCZ 6). Par ailleurs, le post-traitement par règles expertes a prouvé son efficacité en corrigeant les erreurs sémantiques grossières, telle que la confusion entre les grands corps de ferme et les zones industrielles (LCZ 8).

5.2.2. Limites de la résolution

La contrainte principale demeure la résolution kilométrique (100 m) imposée par le standard *WUDAPT*, pour faire une comparaison post-traitement. Elle engendre un effet de mélange spectral (*mixel*) dans les zones hétérogènes. C'est ce qui explique la faible performance du modèle sur l'habitat dispersé (LCZ 9) et les interfaces ville-forêt, où une seule maille agglomère de la végétation, de la voirie et du bâti, brouillant ainsi la signature spectrale. Malgré ces artefacts locaux, la précision globale de 72,4 % reste suffisante pour valider les grandes tendances climatiques du territoire. Il est bien sûr possible de descendre la résolution jusqu'à un maillage de 0,5 m (résolution minimum du LiDAR HD), mais on peut questionner la pertinence d'une échelle aussi fine.

5.2.3. Avantage comparatif

Enfin, cette approche supervisée se distingue de la chaîne de traitement standardisée *GeoClimate* par sa capacité d'adaptation au contexte local. Là où les outils automatisés appliquent des règles de décision génériques parfois inadaptées aux subtilités rurales, le modèle *Random Forest* offre une maîtrise totale des variables prédictives. L'intégration de données « sur-mesure » (LiDAR HD, indices spectraux spécifiques) et la constitution d'un échantillonnage manuel permettent de dépasser les limites des bases de données contributives (type OpenStreetMap) souvent lacunaires dans les territoires peu denses, garantissant une classification plus fidèle à la réalité du terrain.

5.3. Propositions de méthodes meilleures

Au-delà des approches vectorielles ou supervisées testées, le protocole *WUDAPT* constitue une alternative de référence pour la production de cartographies LCZ standardisées à l'échelle mondiale. Contrairement à *GeoClimate*, qui dépend de la précision structurelle de bases de données comme la BD Topo, *WUDAPT* repose sur une grille régulière (100 m) et une méthodologie universelle favorisant l'inter-comparaison climatique inter-villes. Si cette approche par maillage peut induire des effets de « mixels » dans les zones rurales hétérogènes, elle offre une robustesse accrue face aux disparités des données géospatiales locales, garantissant une cohérence thématique indispensable aux modèles climatiques globaux.

Une deuxième solution de réalisation de LCZ qui permet de donner des résultats plus précis que ce n'est le cas pour les deux méthodes présentées au cours de cette étude est celle de la reclassification des données du projet OCS Grand Est de DataGrandEst. L'OCS Grand Est est une base de données régionale qui présente les données d'occupation du sol de la région Grand Est à grande échelle. Elle inclut des données multi-dates pour les années 2010, 2019 et 2021 et peut être affichée à plusieurs niveaux. Le Niveau 1 présente la classification en classes principales comme les territoires agricoles, les espaces forestiers, les surfaces en eaux, les territoires artificialisés ou autres. Les niveaux 2 à 5 présentent des niveaux plus détaillés. Ainsi, pour ces couches vectorielles il y a des spécifications techniques plus précises, notamment sur la perméabilité des sols. Cette base de données a été conçue à l'aide d'orthophotographies aériennes de résolution de 20 cm pour les millésimes 2021/2022 et des données exogènes comme la BD TOPO de l'IGN, le Registre Parcellaire Graphique Agricole (RPGA) et les fichiers fonciers. (Région Grand Est, 2024)

De cette base de données, l'OCS Grand Est de 2021 au Niveau 3 a été choisie pour faire une reclassification en LCZ. Cette subdivision en 34 classes d'occupation du sol a donc servi de base pour la reclassification dans les classes de LCZ. Le Tableau 5, permet de montrer comment les classes de OCS GE ont été reclassées en LCZ.

Tableau 5. Conversion de classes OCS GE en LCZ (E. Thill, 2026)

Catégorie LCZ	Classes d'origine regroupées
LCZ 3	Bâti continu
LCZ 6	Bâti discontinu
LCZ 8	Emprises d'activités, militaires, Équipements collectifs, Réseaux (eau, énergie, déchets, TIC)
LCZ 9	Espaces en transition, Espaces libres urbains, Espaces verts urbains, Bâti isolé
LCZ A	Forêts (conifères, feuillus, mixtes), Peupleraies et sapinières
LCZ B	Arboriculture, Formations pré-forestières
LCZ C	Exploitations agricoles, Vignes, Prairies, friches et délaissés agricoles
LCZ D	Bosquets et haies, Coupes à blanc, Cultures (annuelles, pluri-annuelles, spécifiques), Ripisylves, Surfaces enherbées
LCZ E	Réseaux routiers et ferroviaires, Roches nues
LCZ F	Zones d'extraction, Zones de sinistre, Plages et sables
LCZ G	Milieus liés à l'eau, Bassins artificiels, Cours d'eau, Plans d'eau

Finalement, cette nouvelle méthodologie, plus simple, a permis de rendre un résultat clair des LCZ théoriques à partir des OCS GE. C'est ainsi la Figure 9 qui permet de montrer ce résultat. Il s'agit d'une carte avec une précision meilleure que ce n'est le cas pour le calcul à l'aide de l'outil GeoClimate ou encore à l'aide de *Random Forest*. C'est ainsi, que cette méthode, qui pourrait encore être précisée avec l'utilisation d'un niveau de OCS GE avec plus de cartes, présente une solution simple et efficace pour le calcul des LCZ dans le Parc Naturel Régional des Vosges du Nord.

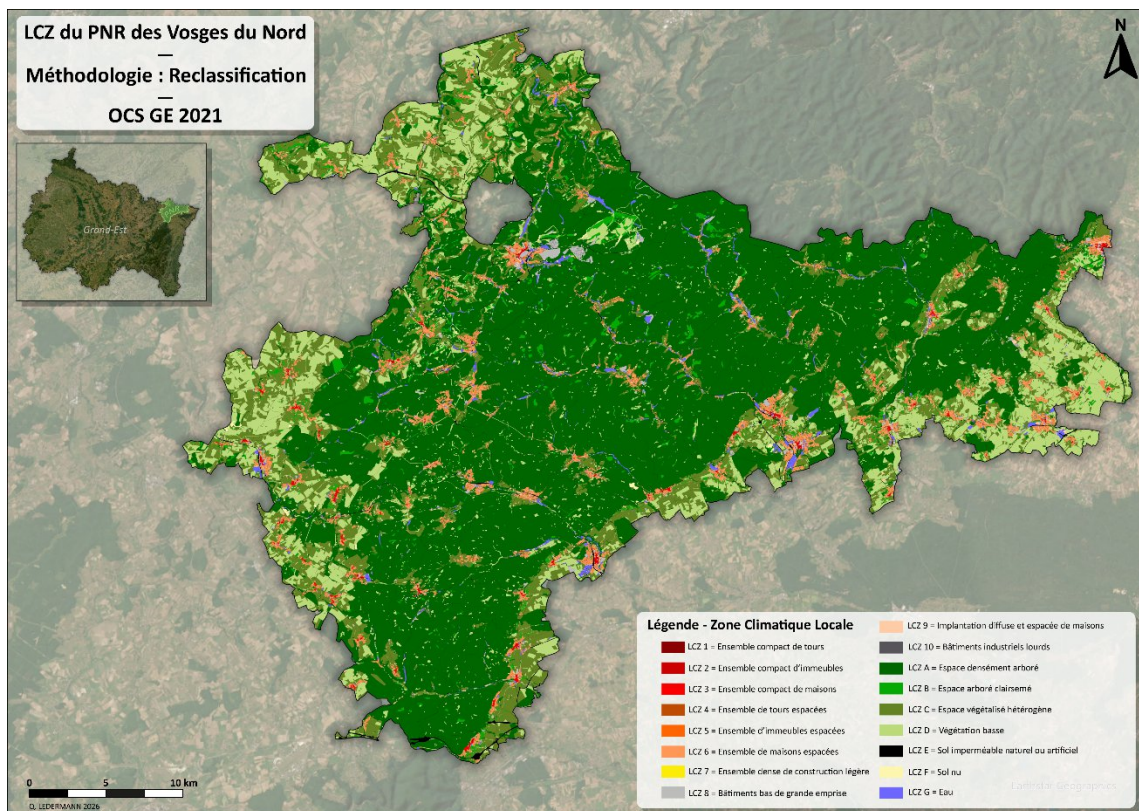


Figure 9. Cartographie des LCZ par reclassification de la base de données OCS GE (E. Thill et Q. Ledermann, 2026)

6. Conclusion et perspectives

Cette étude a permis de caractériser la structure climatique du Parc Naturel Régional des Vosges du Nord en confrontant plusieurs méthodologies de classification en Zones Climatiques Locales (LCZ). Il apparaît que l'outil standardisé GeoClimate, conçu pour les milieux urbains denses, présente des biais structurels majeurs en contexte rural, tendant à sous-estimer l'habitat diffus au profit de la matrice forestière. À l'inverse, l'approche supervisée par l'algorithme Random Forest, enrichie par les données LiDAR, offre une discrimination plus fine des typologies bâties, bien que la résolution hectométrique du standard WUDAPT induise inévitablement des effets de mélange spectral. Finalement, la reclassification directe des données d'occupation du sol régionales (OCS GE) émerge comme une alternative pragmatique et performante. Ces résultats soulignent la nécessité d'adapter les protocoles de modélisation urbaine aux spécificités des territoires de faible densité pour accompagner efficacement les politiques d'aménagement durable face au changement climatique.

Bibliographie

- G. Biau and E. Scornet, « A random forest guided tour, » *Test*, vol. 25, pp. 197–227, Apr. 2016. [Online]. Available : <https://link.springer.com/article/10.1007/s11749-016-0481-7>
- E. Bocher, G. Petit, J. Bernard, and S. Palominos, « A geoprocessing framework to compute urban indicators: The MAPUCE tools chain, » *Urban Climate*, vol. 24, pp. 153–174, 2018. [Online]. Available : <https://doi.org/10.1016/j.uclim.2018.01.008>
- E. Bocher et al., « GeoClimate : a Geospatial processing toolbox for environmental and climate studies, » *Journal of Open-Source Software*, vol. 6, no. 65, 15 pages, Sep. 2021. [Online]. Available : <https://shs.hal.science/halshs-03359757v1>
- J.-B. Bouron, « Parc national en France, parc naturel régional (PNR), » *Géoconfluences*, 2024. [Online]. Available : <https://geoconfluences.ens-lyon.fr/glossaire/parcs-nationaux-et-parcs-naturels-regionaux-pnr>. [Accessed Jan. 2026].
- L. Breiman, « Random forests, » *Machine Learning*, vol. 45, pp. 5–32, 2001. [Online]. Available : <https://doi.org/10.1023/A:1010933404324>
- M. Demuzere, J. Kittner, and B. Bechtel, « LCZ Generator : A Web Application to Create Local Climate Zone Maps, » *Frontiers in Environmental Science*, vol. 9, 2021. [Online]. Available : <https://doi.org/10.3389/fenvs.2021.637455>
- J. Hidalgo, M. Aranda, J. Bernard, and E. Bocher, « Paendora 2 : Pour la gestion du confort estival : Données, Outils et recherche-action, Partie1 : Suivi du projet, » La librairie ADEME, 20 pages, Dec. 2024.
- J. Hidalgo, M. Aranda, J. Bernard, and E. Bocher, « Paendora 2 : Pour la gestion du confort estival : Données, Outils et recherche-action, Partie2 : Rapport final, » La librairie ADEME, 48 pages, Dec. 2024.
- L. Kouadria, « Îlot de chaleur urbain, » *Géoconfluences*, 2025. [Online]. Available : <https://geoconfluences.ens-lyon.fr/glossaire/ilot-de-chaleur-urbain>. [Accessed Jan. 2026].
- « Modélisation géoclimatique : Outils, limites et perspectives, » *OpenScience*. [Online]. Available : <https://www.openscience.fr/Modelisation-geoclimatique-outils-limites-et-perspectives>. [Accessed Jan. 12, 2026].
- Parc Naturel Régional des Vosges du Nord, « Présentation du Parc - Chiffres Clés, » 2026. [Online]. Available : <https://www.parc-vosges-nord.fr/le-parc/presentation>. [Accessed Jan. 2026].
- Région Grand Est, « Production 2021/2022 de la base de données 'Occupation des Sols' à grande échelle en Région Grand Est (OCS Grand Est), » *DataGrandEst*, version 1, 90 pages, Apr. 2024. [Accessed Jan. 2026].
- I. D. Stewart and T. R. Oke, « Local Climate Zones for urban temperature studies, » *Bulletin of the American Meteorological Society*, vol. 93, no. 12, 22 pages, Dec. 2012. [Online]. Available : <https://journals.ametsoc.org/view/journals/bams/93/12/bams-d-11-00019.1.xml>

Annexes

Classe	Nom de la zone	Hauteur Moyenne (H)	Densité Bâtie (BSF)	Sky View Factor (SVF)
LCZ 1	Compact High-rise	> 25 m	0,4 - 0,6	0,2 - 0,5
LCZ 2	Compact Mid-rise	10 - 25 m	0,4 - 0,7	0,3 - 0,6
LCZ 3	Compact Low-rise	3 - 10 m	0,4 - 0,7	0,2 - 0,6
LCZ 4	Open High-rise	> 25 m	0,2 - 0,4	0,5 - 0,7
LCZ 5	Open Mid-rise	10 - 25 m	0,2 - 0,4	0,5 - 0,8
LCZ 6	Open Low-rise	3 - 10 m	0,2 - 0,4	0,6 - 0,9
LCZ 7	Lightweight Low-rise	2 - 4 m	0,6 - 0,9	0,2 - 0,5
LCZ 8	Large Low-rise	3 - 10 m	0,3 - 0,5	> 0,7
LCZ 9	Sparsely Built	3 - 10 m	0,1 - 0,2	> 0,8
LCZ 10	Heavy Industry	5 - 15 m	0,2 - 0,3	0,6 - 0,9

Annexe 1. Seuils de classification utilisé par GeoClimate pour les zones urbaines (LCZ 1 à 10) (C. Friess, 2026)