

TER - Projet ANR PICOPATT

Analyse exploratoire des données collectées pour PICOPATT

Rapport d'avancement TER du 17 décembre 2025

Réalisé par
Ayoub AKKOUH,
Anthony COMBES-AGUÉRA,
Youssef EL ALAOUI,
et Dylla Liesse IZERE

Encadré par
Sandra BRINGAY,
Luc Étienne POMMÉ-CASSIÉROU,
Vincent RAVENEAU,
et Maximillien SERVAJEAN



picopatt
patterns picoclimatiques urbains



Année universitaire 2025–2026

Table des matières

1. Introduction et contexte du TER	3
1.1. Contexte du projet PICOPATT	3
1.1.1. Problématique majeure	3
1.1.2. Présentation du projet PICOPATT	3
1.1.3. Définition de picoclimats urbains	3
1.1.4. Une station météorologique mobile	4
1.2. Les missions du TER	5
1.2.1. Compréhension et préparation des données	5
1.2.2. Construction de silhouettes picoclimatiques	5
1.2.3. Préparation des modèles d'apprentissage automatique	6
2. Présentation des données	7
2.1. Les campagnes métrologiques	7
2.1.1. Les parcours	7
2.1.2. Les mesures	8
2.1.3. Contraintes de terrain	8
2.2. Les variables mesurées	9
2.3. Organisation des fichiers	10
2.3.1. Format des fichiers	10
2.3.2. Sélection des fichiers pour l'analyse	10
3. Nettoyage des données	11
3.1. Premières approches des données	11
3.1.1. Couverture temporelle et spatiale des données	11
3.1.2. Comparaison temporelle des passages	12
3.1.3. Détection et correction des valeurs aberrantes	13
3.1.4. Validation des extrêmes réalistes et cohérence physique	14
3.1.5. Visualisation des valeurs manquantes	15
3.2. Traitement des valeurs nulles	15
3.2.1. Visualisation des valeurs nulles	15
3.2.2. Détails de l'imputation réalisée	16
3.2.3. Résultats obtenus	16
4. Analyse exploratoire des données	18
4.1. Observation de la distribution de chaque variable	18
4.1.1. Distribution globale et par parcours	18
4.1.2. Distribution par section	19
4.1.3. Distribution via boxplot et violin plot	20
4.2. Corrélations entre variables	21
4.2.1. Corrélation entre les variables par parcours	21
4.2.2. Corrélation entre les variables par <code>M_slot</code> de chaque parcours	22

4.3. Profils spatio-temporels	23
4.3.1. Agrégé par <code>M_slots</code>	23
4.3.2. A l'échelle d'une mesure	25
5. Missions restantes	28
5.1. Les silhouettes	28
5.2. Pistes d'apprentissage profond	28
6. Conclusion et perspectives	29

1. Introduction et contexte du TER

1.1. Contexte du projet PICOPATT

1.1.1. Problématique majeure

Depuis leur invention, les stations météorologiques sont la principale source de données climatiques. Ces stations sont fixes et implantées dans des sites normalisés, souvent en terrain dégagé, en périphérie des villes ou à proximité des aéroports, et à plusieurs mètres de hauteur. Si ces dispositifs permettent de caractériser efficacement le climat à grande échelle, ils sont peu adaptés à l'analyse du climat tel qu'il est réellement vécu en milieu urbain dense. En milieu urbain, la morphologie de la ville impacte fortement les conditions climatiques locales. La présence de bâtiments élevés, de rues étroites, de places ouvertes, de matériaux minéraux ou de zones végétalisées génère des contrastes thermiques importants à très petite échelle. Par exemple, une rue étroite du centre historique peut rester à l'ombre une grande partie de la journée, avec une ventilation limitée et une accumulation d'humidité, tandis qu'une place ouverte située à quelques dizaines de mètres peut être fortement exposée au rayonnement solaire et atteindre des températures nettement plus élevées. Ces contrastes sont particulièrement marqués en période estivale et jouent un rôle majeur dans le confort thermique ressenti par les usagers de l'espace urbain. Les stations météorologiques fixes ne permettent pas de capturer cette variabilité spatiale fine. Pire, leur coût limite la possibilité de multiplier les points de mesure à l'intérieur des villes, et le choix d'emplacements réellement représentatifs des situations urbaines reste complexe. Il existe ainsi un manque de données permettant de décrire finement le climat urbain à l'échelle des rues, des places et des parcours piétons. C'est dans ce contexte que s'inscrit le projet PICOPATT.

1.1.2. Présentation du projet PICOPATT

PICOPATT [1] est un projet de recherche financé par l'ANR et dédié à l'étude des picoclimats urbains. Le projet associe deux laboratoires aux compétences complémentaires. Le laboratoire AAU [2], à travers l'équipe CRENAU [3] située à Nantes, apporte son expertise en architecture, urbanisme et étude des ambiances urbaines. Le LIRMM [4], avec son équipe ADVANSE [5], contribue par ses compétences en analyse de données, modélisation et visualisation. PICOPATT vise ainsi à croiser une lecture fine des formes urbaines avec des approches quantitatives fondées sur l'analyse de données mesurées en continu afin de mieux comprendre et caractériser les picoclimats urbains.

1.1.3. Définition de picoclimats urbains

Les chercheurs travaillant sur PICOPATT ont introduit la notion de picoclimats [6], qu'ils définissent comme des variations locales du climat à l'intérieur même de la ville, observables à une échelle spatiale très fine, de l'ordre de quelques mètres. Ces picoclimats résultent des interactions entre les conditions météorologiques générales et la morphologie urbaine. Ainsi, des configurations urbaines proches géographiquement peuvent présenter des conditions climatiques très contrastées. Une rue étroite du centre historique peut rester majoritairement à l'ombre, avec une ventilation limitée et une accumulation d'humidité, tandis qu'une place ouverte située à proximité peut être fortement exposée au rayonnement solaire et atteindre des températures nettement plus élevées. Ces contrastes influencent directement le confort thermique ressenti par les habitants et constituent un enjeu majeur dans un contexte de changement climatique et d'adaptation des villes.

1.1.4. Une station météorologique mobile

Pour répondre au manque de données sur les picoclimats, les chercheurs du laboratoire AAU ont conçu une station météorologique mobile (figure 1), installée sur un chariot de golf. Ce dispositif embarque plus d'une vingtaine de capteurs permettant de mesurer simultanément des variables atmosphériques classiques, telles que la température de l'air, l'humidité relative et la vitesse du vent, ainsi que des flux radiatifs multidirectionnels et des indicateurs de confort thermique. La station est manœuvrée par des opérateurs qui parcourent à tour de rôle, des itinéraires urbains prédéfinis à une vitesse constante proche de 0,6 mètre par seconde. Les parcours sont répétés plusieurs fois par jour et à différentes dates, ce qui permet de collecter des données continues le long des trajets et de capturer la variabilité spatiale fine des conditions climatiques rencontrées dans la ville, au plus proche de l'expérience réelle des usagers.



(a) Vue de face



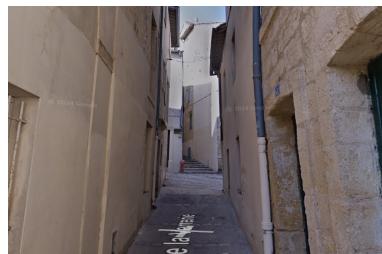
(b) Vue de profil

FIGURE 1 – Station météorologique mobile utilisée lors des campagnes PICOPATT.
Source : AAU - ENSA Nantes

À Montpellier, trois parcours ont été définis afin de représenter des environnements urbains contrastés. Le quartier Antigone est caractérisé par des espaces ouverts favorisant notamment la ventilation. Le parcours Écusson traverse le centre historique composé de ruelles étroites, où l'ombre, la faible ventilation et les matériaux utilisés impactent fortement les conditions climatiques locales. Le parcours Boulevards longe les grandes places et les principaux axes du centre-ville, caractérisés par une exposition solaire importante et une dissipation thermique plus marquée. Trois autres parcours existent également à Nantes, mais ne sont pas étudiés dans ce rapport.



(a) L'esplanade de l'Europe
Antigone



(b) La rue de la Verrerie Basse
Écusson



(c) La place de la Comédie
Boulevards

FIGURE 2 – Photographies représentatives des différents parcours de Montpellier.
Source : Google Maps

Les parcours sont découpés en une succession de sections de mesure prédéfinies. Ce découpage vise à garantir une vitesse de déplacement aussi constante que possible du début à la fin de chaque section, en évitant notamment les zones susceptibles de provoquer des arrêts imprévus, comme les passages piétons. Les sections servent à la fois de repères pour l'opérateur et de base spatiale pour l'analyse des données. Elles permettent également de limiter l'impact des rotations du chariot sur les capteurs, en particulier pour les mesures de rayonnement et de vent. Chaque section est composée d'une multitude de points mesurés toutes les secondes.

1.2. Les missions du TER

Ce TER s'inscrit dans la continuité du projet PICOPATT et porte sur l'ensemble des séries temporelles produites par la station météorologique mobile, à Nantes comme à Montpellier. Les méthodes développées ont pour but d'être génériques et applicables à tous les parcours et à toutes les campagnes. C'est pourquoi, les premières analyses sont menées sur les données montPELLIÉRaines issues des 4 premiers mois de campagne. L'objectif global du TER est d'identifier et de caractériser les picoclimats urbains à partir des profils mesurés le long des parcours. Pour cela, le travail repose sur la construction de silhouettes caractéristiques de picoclimats. Une silhouette correspond à un profil typique résumant un ensemble de séries temporelles jugées similaires dans l'espace et dans le temps. Ces silhouettes constituent un niveau intermédiaire entre les données brutes et les méthodes d'apprentissage automatique et servent de support à des tâches telles que la classification de contextes, la détection d'anomalies ou la prévision. Les missions du TER sont organisées en trois étapes, correspondant à une progression logique du traitement des données vers la modélisation. Le code a été développé sous forme de notebooks et est disponible sur le dépôt GitHub. [7]

1.2.1. Compréhension et préparation des données

Le premier semestre a été consacré en priorité à la prise en main des données PICOPATT. Cette phase a permis de comprendre en détail la structure des fichiers, la signification des variables mesurées et les spécificités des campagnes de terrain. Un travail important de nettoyage a été réalisé, incluant la gestion des valeurs manquantes, des mesures aberrantes et des fichiers partiellement exploitables. En parallèle, de nombreuses visualisations exploratoires ont été produites afin d'analyser les distributions des variables, leurs ordres de grandeur et leur cohérence physique selon les parcours et les créneaux horaires. Cette étape était indispensable pour disposer de séries temporelles fiables, alignées et interprétables et constitue une base solide pour les analyses plus avancées présentées dans la suite du rapport.

1.2.2. Construction de silhouettes picoclimatiques

La deuxième étape du TER consiste à résumer les séries temporelles nettoyées par des descripteurs statistiques. Pour chaque parcours, chaque section et chaque créneau horaire, des indicateurs tels que la moyenne, la médiane, l'écart type ou les quantiles sont calculés et combinés afin de former des vecteurs de caractéristiques. Ces vecteurs permettent de décrire la forme générale des profils climatiques et de faire émerger des silhouettes typiques de picoclimats. À l'issue du premier semestre, ce travail est en cours de construction. Les premières silhouettes ont été définies et comparées et cette phase se poursuivra au début du second semestre afin de consolider les choix méthodologiques et d'affiner les représentations.

1.2.3. Préparation des modèles d'apprentissage automatique

La troisième étape du TER correspond à la conception et à l'expérimentation de modèles d'apprentissage automatique et en particulier d'apprentissage profond, à partir des silhouettes et des séries temporelles préparées. Cette phase débutera au second semestre. L'objectif sera d'explorer des modèles capables d'apprendre automatiquement des représentations compactes des séries, comme des réseaux de neurones, des réseaux convolutionnels unidimensionnels ou des autoencodeurs. Ces modèles serviront notamment à la classification de contextes picoclimatiques, à la détection d'anomalies instrumentales ou météorologiques et à la prévision de profils le long des parcours. Ainsi, le travail réalisé au premier semestre constitue une étape structurante du projet. Il a permis de sécuriser les données, de valider les choix d'analyse exploratoire et de poser des bases méthodologiques solides pour les développements de modélisation prévus au second semestre.

2. Présentation des données

2.1. Les campagnes météorologiques

Les données utilisées dans ce premier rapport proviennent des campagnes de météorologie mobile menées à Montpellier. Pour rappel, chaque mesure consiste à parcourir un parcours prédéfini avec la station météorologique mobile, en enregistrant en continu les variables climatiques et la position du chariot. La personne chargée de la mesure doit pousser le chariot à une vitesse proche de 0,6 mètre par seconde. Cela permet de limiter l'influence de la vitesse de marche sur les capteurs.

2.1.1. Les parcours

Les parcours sont mesurés en rotation. Un jour est réservé à un seul parcours et le même trajet est répété quatre fois dans la journée. Les journées dédiées à un même parcours sont espacées de neuf à dix jours. Cela permet d'échantillonner des situations météorologiques suffisamment rapprochées pour être comparables, tout en laissant le temps à la météo de varier. Les parcours ont été tracés par les chercheurs architectes du laboratoire AAU. La durée d'un parcours a été pensée pour être au alentour de 75 minutes mais selon les conditions de circulation cela peut aller jusqu'à 115 minutes.

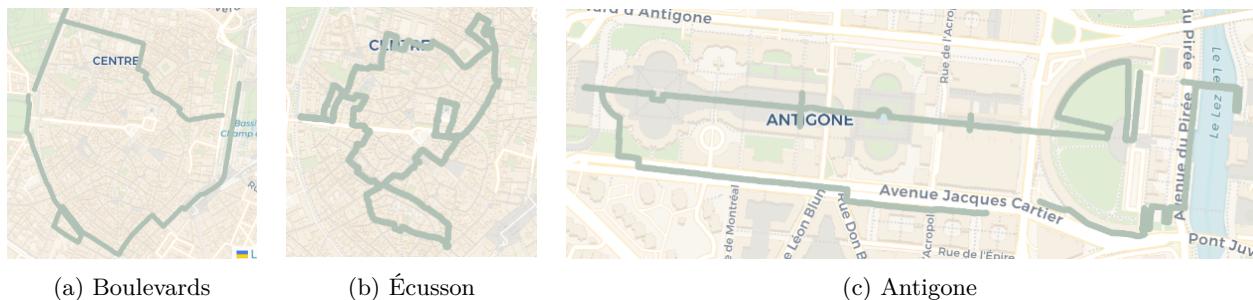


FIGURE 3 – Captures d'écran des tracés des différents parcours.
Source : Picopatt ExploreR (ADVANSE - LIRMM)

Les trois parcours montPELLIÉRAINS se trouvent dans le centre-ville mais possèdent de fortes différences comme décrit dans l'introduction à la section 1.1.4. Le tableau 1 en propose une synthèse.

Nom du parcours	Nombre de sections	Nombre de points	Caractéristiques climatiques attendues
Boulevards	51	115 962	Grands axes urbains fortement exposés au rayonnement solaire et à des hautes températures.
Écusson	75	103 862	Ruelles étroites avec variation d'ombre et de soleil et une humidité plus marquée.
Antigone	53	121 457	Quartier récent caractérisé par une meilleure ventilation et une proximité avec le Lez.

TABLE 1 – Synthèse des caractéristiques climatiques attendues selon les parcours.
Source : AAU – ENSA Nantes

2.1.2. Les mesures

Les campagnes sont organisées deux jours par semaine. Lors de chacune de ces journées, quatre mesures sont réalisées sur un même parcours, notées de M1 à M4. Les départs de ces quatre créneaux sont espacés de trois heures. Lors des changements d'heure, ces créneaux sont ajustés afin de rester alignés sur la même période solaire et d'éviter un décalage entre les mesures d'hiver et d'été. Chaque mesure M1 à M4 correspond à une situation climatique différente pour une même journée et un même parcours. M1 décrit les conditions du matin, souvent plus fraîches. M2 et M3 couvrent le début et la fin d'après-midi, qui sont les périodes les plus ensoleillées. M4 correspond à la fin de journée, lorsque le rayonnement solaire diminue et que la température commence à baisser. La table 2 offre un descriptif de chaque mesure.

Mesure	Heure d'hiver	Heure d'été	Situation climatique typique
M1	8h30 – 11h00	9h30 – 12h00	Conditions fraîches et faiblement rayonnées
M2	11h30 – 14h00	12h30 – 15h00	Augmentation du rayonnement et de la température
M3	14h30 – 17h00	15h30 – 18h00	Rayonnement maximal et températures élevées
M4	17h30 – 20h00	18h30 – 21h00	Coucher du soleil et baisse de la température

TABLE 2 – Descriptif des mesures M1 à M4.

Source : AAU - ENSA Nantes

Lors des mesures, le pousseur doit à la fois respecter une vitesse de déplacement proche de 0,6 mètre par seconde et suivre précisément le tracé du parcours. Cette rigueur est nécessaire car le chariot est géolocalisé en continu. L'objectif est de pouvoir analyser les mesures en fonction de leur position dans l'espace urbain. Par exemple, certaines observations sont réalisées au centre de la rue, tandis que d'autres sont effectuées plus près des murs. Cette précision permet d'étudier l'influence de la position dans la rue sur les variables mesurées et d'identifier d'éventuelles différences liées à la proximité des bâtiments ou aux zones d'ombre.

2.1.3. Contraintes de terrain

Les mesures étant réalisées en milieu urbain, elles peuvent être perturbées par différents facteurs extérieurs. Les conditions météorologiques constituent une première contrainte. En cas de pluie durant une mesure, le protocole prévoit une mise en pause temporaire. Si la pluie cesse en moins de dix minutes, la mesure peut reprendre. Au-delà de ce délai, la mesure est annulée, situation déjà rencontrée lors de forts épisodes orageux. Les contraintes liées à la vie urbaine affectent également certaines campagnes. A Montpellier, les parcours Boulevards et Écusson, situés en centre-ville dense, sont régulièrement sujets à des blocages ponctuels de rues, notamment en raison des livraisons pour les commerces et les restaurants. Le parcours Antigone est lui aussi impacté, lors de la tenue d'une brocante hebdomadaire située au début du trajet les jeudis, jour de mesure. Enfin, des événements urbains exceptionnels, tels que le passage du Tour de France, ont pu perturber certaines campagnes en modifiant les conditions de circulation ou en entraînant l'annulation de mesures. L'ensemble de ces contraintes peut générer des interruptions, des retards ou des discontinuités dans les séries temporelles, qui doivent être prises en compte lors des étapes de nettoyage et d'analyse des données.

2.2. Les variables mesurées

La station météorologique mobile enregistre, à chaque instant, un ensemble de variables permettant de décrire finement l'état du picoclimat urbain traversé. Ces variables couvrent à la fois les conditions atmosphériques classiques, les flux radiatifs multidirectionnels, la géométrie solaire ainsi que des informations spatiales et temporelles nécessaires à l'analyse des profils. Les mesures sont enregistrées simultanément à chaque point du parcours et constituent des séries temporelles denses. Les principales familles de variables sont résumées dans le tableau 3.

Famille de variables	Description
Localisation et métadonnées	Coordonnées GPS (RTK et projetées sur le tracé du parcours), identifiants de parcours, de sections et de points, permettant l'analyse spatiale et le découpage des profils.
Température de l'air (T_{air})	Températures mesurées par différents capteurs (thermo-hygromètre, thermocouples, anémomètre), permettant de caractériser l'air ambiant et de comparer les mesures selon les dispositifs.
Humidité relative (rh)	Humidité de l'air mesurée sous abri, utilisée pour l'analyse du confort thermique et l'identification de zones plus humides ou rafraîchies.
Vent	Vitesse (ws) et direction ($wdir$) du vent mesurées par l'anémomètre, permettant d'identifier des zones ventilées, abritées ou des couloirs de vent urbains.
Rayonnement solaire (sw)	Flux radiatifs de courtes longueurs d'onde mesurés dans six directions autour de la station (haut, bas, avant, arrière, gauche, droite), afin de caractériser l'exposition solaire et les effets d'ombre.
Rayonnement thermique (lw)	Flux radiatifs de grandes longueurs d'onde mesurés dans les mêmes six directions, reflétant l'émission thermique des surfaces urbaines proches.
Indices de confort	Température moyenne radiante ($tmrt$) et Physiological Equivalent Temperature (pet), calculées à partir des variables mesurées, pour évaluer le confort thermique ressenti.
Géométrie solaire	Azimut et hauteur du soleil, calculés pour chaque timestamp, utilisés pour interpréter les variations de rayonnement et les alternances d'ombre et de soleil.

TABLE 3 – Synthèse des principales variables mesurées lors des campagnes PICOPATT.
Source : AAU - ENSA Nantes

L'ensemble de ces variables permet d'analyser les variations spatio-temporelles du picoclimat urbain et constitue la base des analyses exploratoires présentées dans la suite du rapport.

2.3. Organisation des fichiers

A la fin de chaque mesure, les données sont sauvegardées sous forme de fichiers au format CSV, chaque fichier correspondant à une mesure réalisée sur un parcours donné et à un créneau horaire précis. Une structuration claire et cohérente des fichiers est essentielle pour le traitement des données à grande échelle. Elle permet d'automatiser les étapes de lecture, de filtrage et d'agrégation des séries temporelles, tout en réduisant les risques d'erreurs lors des analyses. Dans le cadre du TER, cette organisation est centrale, car elle permet d'associer explicitement chaque série temporelle à son contexte spatial (ville, parcours, section) et temporel (date et créneau horaire), condition indispensable à la comparaison des mesures et à la construction des silhouettes picoclimatiques.

2.3.1. Format des fichiers

Les fichiers sont nommés selon une convention stricte, ce qui permet d'identifier immédiatement le contexte de chaque mesure. Le nom des fichiers suit le format suivant :

`picopatt_<ville>_<parcours>_YYYYMMDD_HHMM.csv`

où `<ville>` correspond à la ville de la mesure (Nantes ou Montpellier), `<parcours>` au parcours mesuré (Antigone, Écusson ou Boulevards), `YYYYMMDD` à la date de la campagne et `HHMM` à l'heure de début de la mesure. Cette convention facilite l'automatisation du chargement des données et le tri des fichiers par parcours, saison ou créneau horaire.

Tous les fichiers partagent une structure commune et contiennent un ensemble de champs généraux présents pour chaque point de mesure. On retrouve notamment :

- un horodatage précis de la mesure (`timestamp`),
- des identifiants spatiaux permettant le repérage le long du parcours (`track_id`, `section_id`, `point_id`),
- les coordonnées géographiques du chariot ainsi que l'ensemble des variables climatiques et radiatives décrites dans la section précédente.

Cette organisation homogène garantit la comparabilité des séries temporelles entre les différentes campagnes, quels que soient le parcours, la saison ou le créneau de mesure.

2.3.2. Sélection des fichiers pour l'analyse

Pour la suite du rapport, nous nous sommes concentrés sur les fichiers de données issus des campagnes menées à Montpellier entre fin octobre 2024 et mi-janvier 2025. Au total, 73 fichiers étaient disponibles pour cette période. Parmi eux, trois fichiers ne correspondaient pas à des mesures réalisées lors de campagnes météorologiques et ont donc été écartés de l'analyse. L'ensemble des analyses présentées dans ce rapport repose ainsi sur 70 mesures exploitables, chacune correspondant à un fichier et à un créneau horaire donné.

3. Nettoyage des données

3.1. Premières approches des données

Avant toute opération de nettoyage approfondie, une première exploration des fichiers a été réalisée afin de comprendre leur structure, la fréquence d'échantillonnage et les ordres de grandeur des variables mesurées. Cette étape préliminaire avait également pour objectif d'évaluer la répartition des observations dans le temps et dans l'espace.

3.1.1. Couverture temporelle et spatiale des données

Nous avons vérifié la couverture des campagnes afin de savoir si les trois parcours et les quatre créneaux M1 à M4 sont représentés de façon comparable. Cette étape est importante car la construction de silhouettes et les comparaisons inter-parcours sont sensibles aux déséquilibres d'échantillonnage. Dans les fichiers bruts, le créneau horaire n'était pas fourni explicitement. Il a donc été construit à partir de l'horodatage, puis ajouté dans les données sous la forme d'une variable `M_slot`. La figure 4 résume la règle utilisée pour affecter un créneau `M_slot` à partir de l'heure de début.

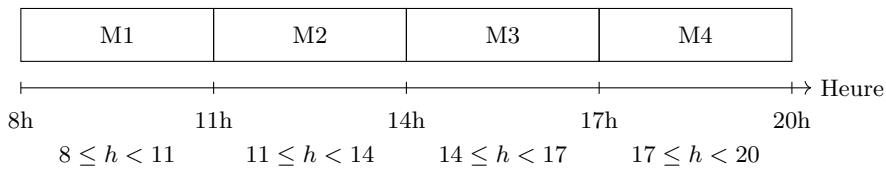


FIGURE 4 – Schéma de la règle d'affectation du créneau `M_slot` à partir de l'heure de début (heure d'hiver).

Nous définissons un passage distinct comme une combinaison unique (`date, M_slot`) pour un parcours donné. La figure 5 présente le nombre de passages distincts par parcours et par créneau horaire. Les résultats montrent une couverture hétérogène. Le parcours Antigone est le plus complet et le plus équilibré avec 7 passages pour chacun des quatre créneaux (M1 à M4), soit 28 passages au total. Le parcours Boulevards présente une couverture légèrement moins homogène avec 22 passages au total (5 en M1, 6 en M2, 5 en M3, 6 en M4). Enfin, le parcours Écusson est le plus incomplet avec 20 passages, et une sous-représentation marquée du créneau M4 (6 en M1, 5 en M2, 5 en M3, 4 en M4). Au-delà du nombre de passages, la quantité de données effectivement exploitables peut varier selon la durée réelle des mesures, d'éventuelles interruptions et la qualité des enregistrements.

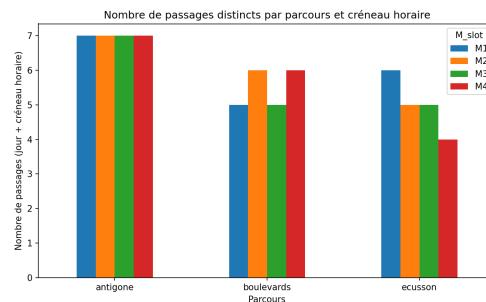


FIGURE 5 – Diagramme en barres groupées du nombre de passages distincts par parcours et par créneau.

La figure 6 complète donc l'analyse en montrant, pour chaque couple parcours-créneau, le nombre total d'échantillons enregistrés. Antigone reste stable et dense sur l'ensemble des créneaux, Boulevards présente une variabilité plus marquée avec un maximum sur M2, et Écusson confirme une densité nettement plus faible sur M4. Cette information est utile pour interpréter les statistiques agrégées, car un créneau peu couvert contribue moins aux moyennes et augmente l'incertitude sur les comparaisons.

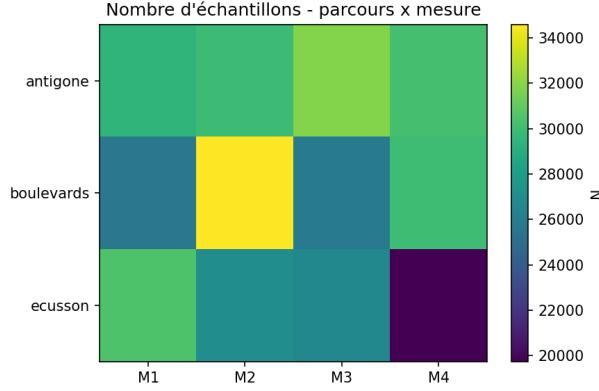


FIGURE 6 – Carte de chaleur du nombre total d'échantillons enregistrés pour chaque couple parcours-créneau.

3.1.2. Comparaison temporelle des passages

En complément de la couverture globale, nous analysons la régularité temporelle des mesures, c'est-à-dire la présence ou l'absence d'un passage pour chaque date et chaque créneau. La figure 6 représente cette complétude sous forme de cartes de chaleur, où une case indique si un passage existe pour une date donnée et un créneau `M_slot`.

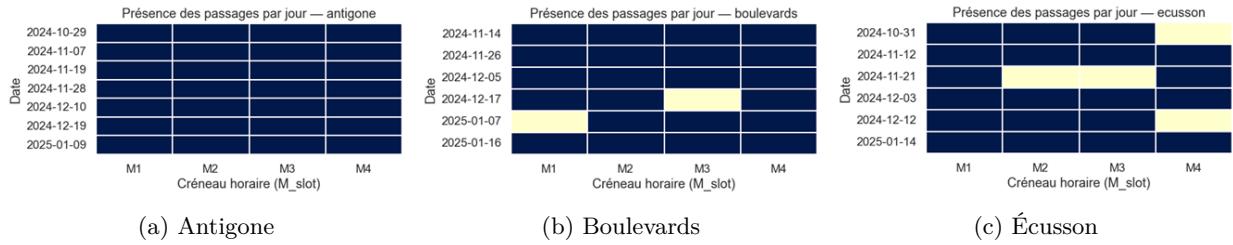


FIGURE 7 – Cartes de chaleur de la présence des passages par date et créneau horaire pour chaque parcours.

Pour le parcours Antigone, la couverture est complète sur toute la période étudiée. Chaque date comporte les quatre créneaux, ce qui garantit une comparabilité temporelle robuste et fournit un parcours de référence pour les analyses ultérieures. Pour le parcours Boulevards, la couverture reste globalement satisfaisante mais légèrement moins homogène. On observe quelques absences ponctuelles, notamment sur le créneau M1 du 7 janvier 2025 et sur le créneau M3 du 17 décembre 2024. Ces manques restent limités, mais ils réduisent localement la comparabilité des statistiques par créneau. Le parcours Écusson est le plus irrégulier. Plusieurs absences apparaissent, en particulier sur le créneau M4 à certaines dates. Ce comportement est cohérent avec les contraintes du centre historique et les conditions de fin de journée en hiver. Pour la suite, ces déséquilibres devront être pris en compte lors des agrégations par créneau, afin d'éviter d'attribuer à un effet de parcours ce qui relève d'un effet de couverture.

3.1.3. Détection et correction des valeurs aberrantes

Après la vérification temporelle, une analyse de qualité a été menée afin d'identifier et de traiter les valeurs aberrantes susceptibles de biaiser les statistiques descriptives et les comparaisons inter-parcours. Dans un protocole de mesure mobile, des défauts instrumentaux ponctuels peuvent produire des valeurs extrêmes sans signification physique. L'analyse se concentre sur les variables météorologiques principales, en particulier les températures `tair_thermohygro`, `tair_tc1` et `tair_tc2`. Pour réduire le bruit au point seconde, les données ont été agrégées par passage, défini comme une combinaison unique (`track_id`, `date`, `M_slot`). Pour chaque passage, les statistiques de base ont été calculées, dont la moyenne, l'écart-type, le minimum et le maximum. Une première détection visuelle a été réalisée à l'aide de boîtes à moustaches représentant la distribution des températures moyennes par parcours figure 7.

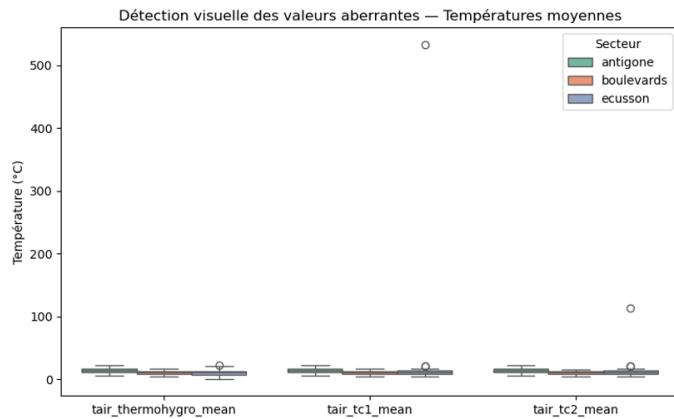


FIGURE 8 – Boîtes à moustaches des températures moyennes par capteur et par parcours.

Ce graphique met en évidence deux valeurs extrêmes sur le parcours Écusson, lors du passage M1 du 14 janvier 2025, `tair_tc1_mean` avec 533.5 °C et `tair_tc2_mean` avec 112.8 °C. Ces valeurs sont physiquement impossibles et indiquent un défaut ponctuel des capteurs thermocouples. Au même passage, `tair_thermohygro_mean` reste cohérente pour un matin d'hiver, ce qui confirme une anomalie localisée aux capteurs `tc1` et `tc2`. Afin d'éviter que ce passage ne biaise les analyses, les mesures associées à `tair_tc1` et `tair_tc2` sur ce fichier ont été neutralisées en valeurs manquantes, puis l'agrégation a été recalculée.

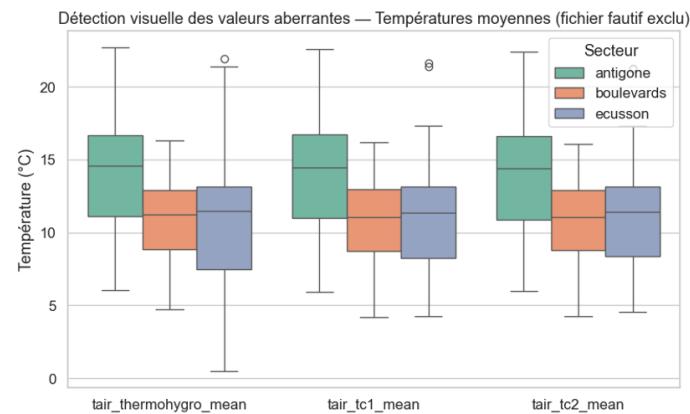


FIGURE 9 – Boîtes à moustaches des températures moyennes après exclusion des valeurs aberrantes.

La figure 8 montre la distribution après correction. Les valeurs redeviennent cohérentes, et les trois capteurs présentent des plages comparables, compatibles avec la période automne-hiver. Pour confirmer le caractère isolé de l'événement, la figure 9 présente l'évolution temporelle de `tair_tc1_mean` sur le créneau M1. Le pic apparaît comme une rupture unique sur Écusson au 14 janvier 2025, alors que les autres dates et parcours restent stables. Cela confirme que c'est bien une erreur de capteur ponctuel plutôt qu'un phénomène météorologique.

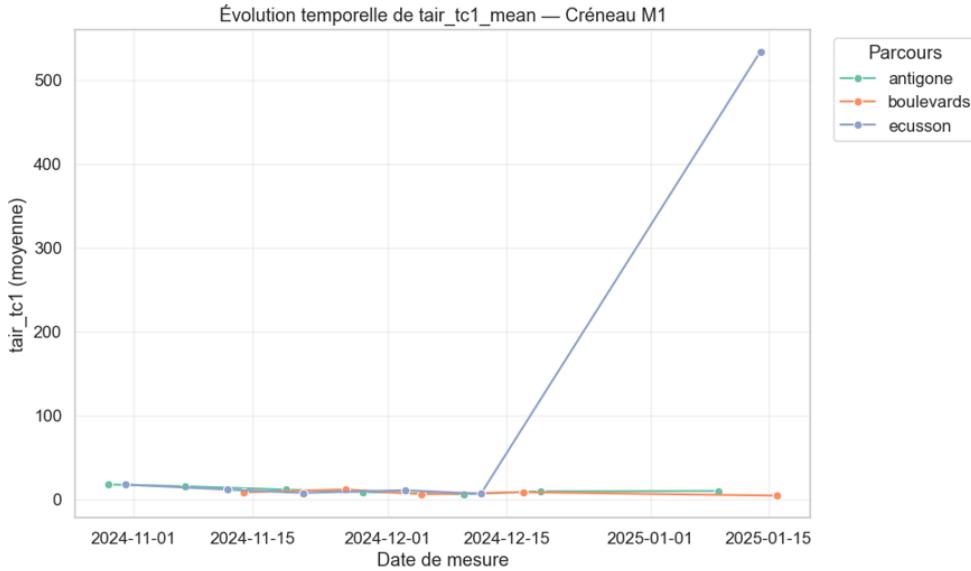


FIGURE 10 – Profil de la variable `tair_tc1_mean` sur le créneau M1.

3.1.4. Validation des extrêmes réalistes et cohérence physique

Lors de l'analyse des indices thermiques, une valeur de PET négative ($\text{PET} = -0.41^\circ\text{C}$) a été identifiée. Cette valeur est atypique mais peut être réaliste en conditions hivernales. Pour distinguer une anomalie d'une situation plausible, nous avons vérifié la cohérence entre le PET et la température radiante moyenne `tmrt`. La figure 10 montre la relation entre `tmrt` et le PET sur l'ensemble des passages. Le point correspondant au PET négatif suit la structure générale observée et ne présente pas d'isolement extrême par rapport aux autres valeurs. Cette observation suggère que cette valeur reflète une situation de froid marquée plutôt qu'un défaut de calcul ou de capteur.

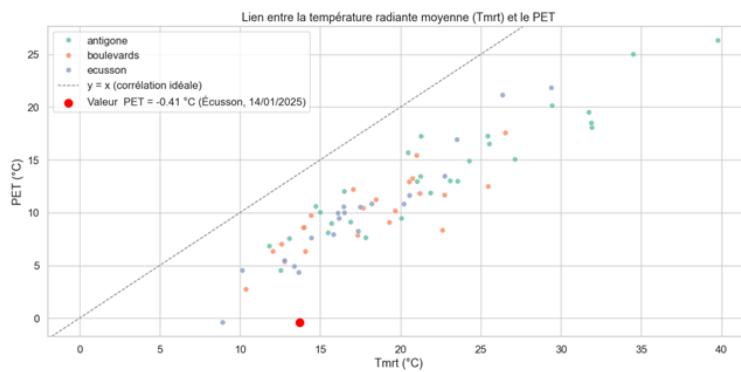


FIGURE 11 – Nuage de points reliant la température radiante moyenne `tmrt` et le PET.

3.1.5. Visualisation des valeurs manquantes

En parallèle, nous avons vérifié la présence de valeurs manquantes codées en NaN. La figure 12 présente le taux de valeurs manquantes par variable pour les variables météorologiques. Sur la période analysée, aucune variable ne présente de taux de NaN.

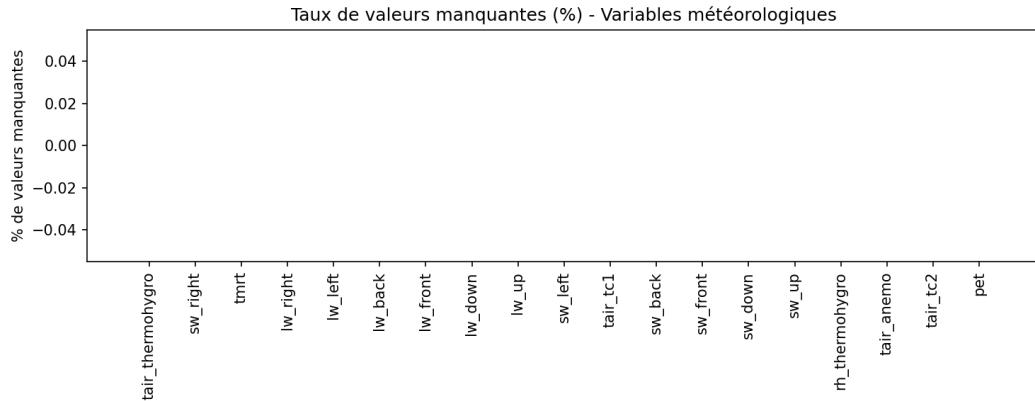


FIGURE 12 – Diagramme en barres du taux de valeurs manquantes par variable.

3.2. Traitement des valeurs nulles

3.2.1. Visualisation des valeurs nulles

La visualisation des valeurs nulles est essentielle, car l'exploration initiale a montré que certains défauts capteurs étaient codés par 0. Dans ce contexte, les zéros sont interprétés comme des valeurs invalides lorsque leur fréquence est très élevée ou lorsqu'ils apparaissent sous forme de séquences répétées, ce qui est typique d'une coupure de capteur ou d'une défaillance d'acquisition. L'analyse globale du pourcentage de zéros sur les variables numériques a mis en évidence deux variables critiques dont le taux est trop élevé pour envisager une correction robuste. Comme le montre la table 4, les variables **sw_down** et **sw_left** présentent respectivement 76.25% et 49.74% de zéros au global. Ces proportions sont compatibles avec un capteur défaillant ou inactif sur de longues périodes. Pour éviter une imputation spéculative, ces deux variables ont été exclues de la suite des analyses.

Antigone		Boulevards		Écusson		Global	
Variable	%	Variable	%	Variable	%	Variable	%
sw_down	65.08	sw_down	76	sw_down	89.7	sw_down	76.25
sw_left	34.18	sw_left	51.47	sw_left	66.13	sw_left	49.74
sw_up	26.73	sw_up	27.7	sw_up	25.08	sw_up	26.56
ws	2.8	ws	5.15	ws	4.72	ws	4.18

TABLE 4 – Classement des pourcentages de valeurs nulles par variable, au global et par parcours

Les autres variables numériques présentaient un taux de zéros plus faible et par conséquent plus gérable, ne justifiant pas une exclusion définitive. À titre d'exemple, la vitesse de l'air (**ws**) enregistrait seulement 4.18%

de zéros à l'échelle globale, et la température de l'air (`tair_tc2`) uniquement 0.59%. Ces défauts ne sont pas interprétés comme des pannes majeures, mais plutôt comme des erreurs de transmission intermittentes ou des arrêts de capteurs de très courte durée. Les variables ciblées pour la correction par imputation étaient ainsi les flux radiatifs restants (`sw_up`, `sw_back`, `sw_right`), la température (`tair_tc2`) et la vitesse de l'air (`ws`). Ces variables nécessitaient un traitement spécifique pour garantir la continuité des séries temporelles sans en altérer la dynamique fondamentale, condition essentielle à la construction des silhouettes picoclimatiques.

3.2.2. Détails de l'imputation réalisée

Pour préserver la continuité des séries nécessaires à la construction de silhouettes, une imputation locale a été appliquée sur un ensemble restreint de variables (`sw_up`, `sw_back`, `sw_right`, `ws`, `tair_tc2`). Le traitement est réalisé fichier par fichier, afin de détecter correctement les séquences consécutives de zéros. La règle d'imputation suit la logique suivante : une séquence de zéros consécutifs est corrigée uniquement si sa longueur est inférieure ou égale à un seuil `MAX_RUN` égale à 30 échantillons et si elle est encadrée par deux valeurs numériques finies. Dans ce cas, chaque zéro du run est remplacé par la moyenne des deux valeurs d'encadrement :

$$x_i = \frac{x_{i_0-1} + x_{i_1+1}}{2}.$$

Si la séquence est plus longue que `MAX_RUN`, si elle touche le début ou la fin du fichier, ou si l'une des valeurs d'encadrement est manquante, alors l'ensemble du run est converti en valeur manquante. La figure 13 schématisse la méthode utilisée.



- (a) Les zéros sont remplacés par la moyenne des deux valeurs encadrantes.
(b) Les points du run sont convertis en valeur manquante.

FIGURE 13 – Schéma de la règle d'imputation locale appliquée aux séquences de zéros.

Cette séparation entre l'imputation et la conversion en NaN est volontaire. Elle permet de corriger uniquement des coupures courtes, qui ressemblent à des défauts d'acquisition ponctuels, tout en évitant de reconstruire artificiellement des portions entières de série lorsque le capteur semble inactif sur une durée trop longue. Le choix d'un seuil à 30 échantillons correspond ici à une coupure de l'ordre de quelques dizaines de secondes, ce qui reste compatible avec une continuité picoclimatique le long du parcours.

3.2.3. Résultats obtenus

Pour chaque variable, nous avons comptabilisé le nombre total de zéros avant correction, puis distingué deux issues : les points effectivement imputés lorsque la séquence respecte le seuil `MAX_RUN` et les points convertis en NaN lorsque la séquence est trop longue ou non encadrée. Ce bilan permet d'évaluer l'ampleur réelle de l'imputation et d'identifier les variables pour lesquelles les défauts sont principalement ponctuels ou au contraire structurels. On observe que certaines variables, comme `ws`, présentent de nombreux zéros mais

la majorité correspond à des runs courts, ce qui conduit à un grand nombre de points imputés. À l'inverse, `sw_up` concentre une part importante de runs non imputables, transformés en NaN. Ce résultat suggère un défaut plus structurel sur cette mesure, cohérent avec les taux élevés de zéros observés dans l'analyse précédente. Enfin, l'absence de points imputés pour `tair_tc2` indique que les zéros y apparaissent surtout sous forme de séquences longues ou mal encadrées, ce qui justifie de ne pas les corriger automatiquement. Au-delà des volumes, cette étape a un impact direct sur les analyses suivantes. D'une part, l'imputation limite les ruptures artificielles dans les profils et facilite le calcul de statistiques par section ou par passage. D'autre part, la conversion en NaN permet de conserver la trace des segments réellement non mesurés, ce qui évite d'introduire un faux signal et permet de contrôler la complétude lors des agrégations et des comparaisons inter-parcours. La table 5 résume le bilan final de ce traitement.

Variable	Zéros avant	Après traitement	
		Imputés	Convertis en NaN
<code>sw_back</code>	4 000	2 672	1 328
<code>sw_right</code>	1	1	0
<code>sw_up</code>	88 756	2 345	86 411
<code>tair_tc2</code>	2 011	0	2 011
<code>ws</code>	14 018	13 724	294

TABLE 5 – Synthèse de l'imputation des zéros par variable

Nous avons ensuite mesuré le taux de valeurs manquantes après application du traitement d'imputation. On observe donc qu'une partie des défauts initialement codés par des zéros est convertie en NaN. La figure 14 présente le taux de valeurs manquantes par variable sur les variables météorologiques après correction. On observe que les manques concernent principalement `sw_up`, qui est donc retiré de nos analyses.

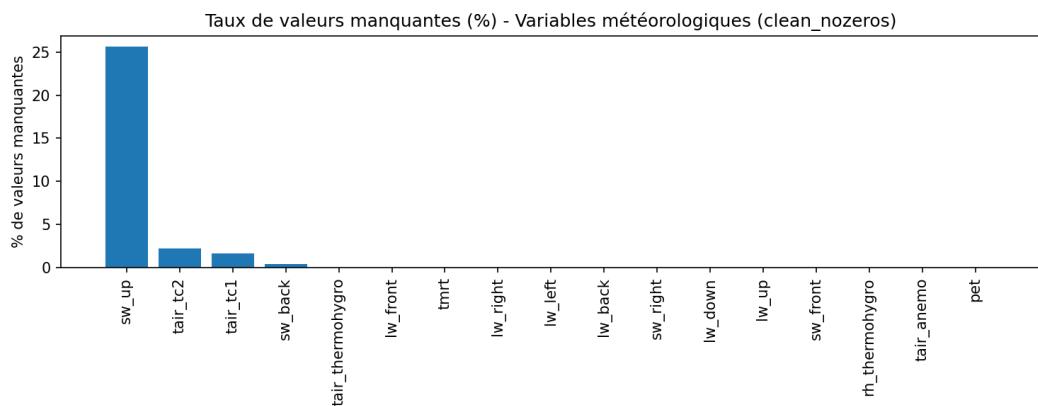


FIGURE 14 – Diagramme en barres du taux de valeurs manquantes par variable après imputation.

4. Analyse exploratoire des données

4.1. Observation de la distribution de chaque variable

4.1.1. Distribution globale et par parcours

Les distributions globales mettent en évidence des comportements très différents selon la nature physique des variables. Les flux d'ondes longues (`lw_*`) présentent des distributions unimodales et relativement concentrées, ce qui suggère une acquisition stable et cohérente entre capteurs. À l'inverse, les flux d'ondes courtes (`sw_*`) sont fortement asymétriques, avec une forte concentration de valeurs faibles et une traîne vers des valeurs élevées. Ce comportement est attendu en contexte urbain (ombrage, variations d'exposition, période hivernale), mais implique que les statistiques basées sur la moyenne peuvent être sensibles à quelques valeurs fortes. La vitesse du vent (`ws`) suit également une distribution très asymétrique, dominée par de faibles vitesses, ce qui est cohérent avec un écoulement perturbé par la rugosité urbaine. Enfin, les indices thermiques (`PET` et `tmrt`) présentent des distributions étalées, avec une variabilité marquée liée aux conditions radiatives et aux contrastes d'exposition. La présence de valeurs de `PET` proches de 0°C reste compatible avec des situations hivernales et a été contrôlée par cohérence avec `tmrt`.

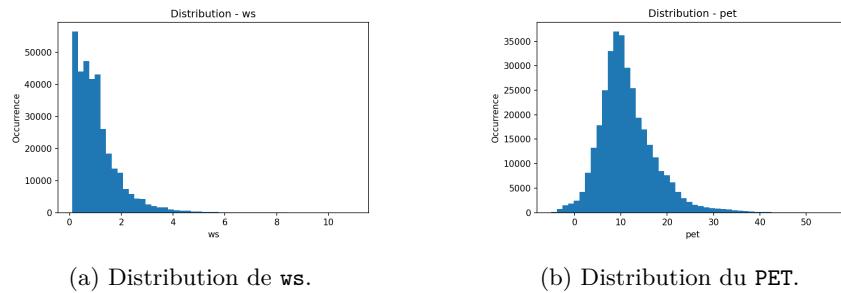


FIGURE 15 – Distributions globales de deux variables.

La comparaison par parcours confirme la forme globale de `ws` (forte concentration près de 0 et traîne à droite), ce qui indique que le régime de vent est majoritairement faible dans les trois contextes urbains. Les différences entre parcours portent surtout sur la dispersion et l'importance de la traîne. Un parcours plus ouvert et exposé peut présenter davantage de vitesses modérées, tandis qu'un parcours plus encaissé ou contraint favorise des vitesses très faibles et plus fréquentes. Ces écarts restent cohérents avec les caractéristiques urbaines attendues et justifient d'analyser `ws` à une échelle plus fine (par créneau ou par section) pour relier la variabilité aux configurations locales.

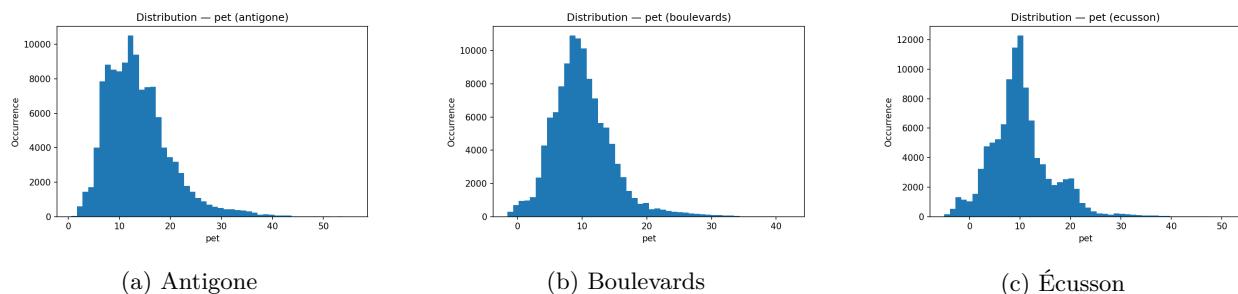


FIGURE 16 – Distribution de `PET` par parcours.

La comparaison par parcours met en évidence une structure globale similaire du PET, une distribution asymétrique avec très peu de valeurs élevées, ce qui indique que la majorité des situations correspondent à un confort thermique modéré, avec quelques épisodes plus contraignants. Boulevards présente une distribution plus resserrée, centrée autour de valeurs intermédiaires. Antigone apparaît légèrement décalé vers des PET plus élevés, ce qui suggère des conditions un peu plus souvent favorables au réchauffement ressenti (exposition, rayonnement ou moindre effet d'ombrage selon les sections). Enfin, Écusson est le plus hétérogène, avec une dispersion plus marquée et un renflement secondaire vers des PET plus élevés, ce qui est compatible avec une alternance forte entre zones ombragées et zones très exposées dans le centre historique. Ces écarts justifient une analyse plus fine par créneau M_slot et par section afin de relier ces variations à la morphologie urbaine et aux conditions radiatives.

4.1.2. Distribution par section

Afin de descendre à une échelle plus fine, nous avons étudié la distribution du PET section par section au sein d'un même parcours. Cette approche met en évidence la variabilité spatiale du confort thermique. Il faut toutefois interpréter ces histogrammes avec précaution, car les sections ne contiennent pas toutes le même nombre de points (durée réelle, interruptions, vitesse variable). Comparer directement la hauteur des barres (occurrences) peut donc être trompeur. Pour comparer correctement les formes, il est préférable d'utiliser une échelle normalisée (densité) ou, à défaut, de raisonner principalement sur la position du mode, la dispersion et la présence d'une traîne à droite. À titre d'illustration, certaines sections présentent une distribution relativement concentrée autour de valeurs faibles à modérées, tandis que d'autres sections montrent une dispersion plus forte et une traîne plus marquée, signe de passages ponctuels dans des zones plus exposées ou plus minérales.

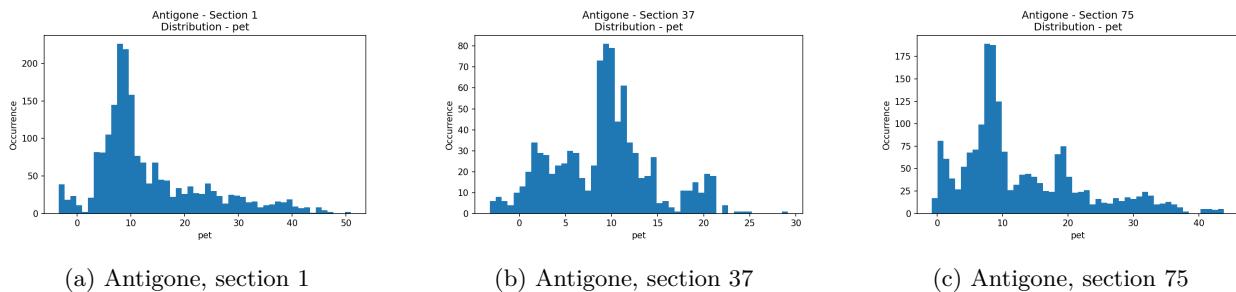


FIGURE 17 – Distributions du PET à l'échelle d'une section.

A noter que les sections n'ayant pas le même nombre de points, l'échelle en ordonnée (occurrences) n'est pas directement comparable.

Pour résumer la variabilité spatiale, nous avons construit une représentation synthétique des distributions pour chaque variable et pour chaque parcours. Chaque vignette correspond à une section et montre l'histogramme de la distribution locale. Des repères verticaux indiquent la moyenne et la médiane du parcours, ainsi que la moyenne de la section, ce qui permet d'identifier rapidement les sections systématiquement plus chaudes ou plus fraîches que la tendance globale. La couleur de fond traduit l'écart entre la moyenne de la section et celle du parcours, ce qui facilite la lecture des contrastes. Cette visualisation aide à repérer des sections durablement au-dessus ou en dessous de la moyenne, des sections très dispersées, et des sections plus homogènes où les conditions restent stables.

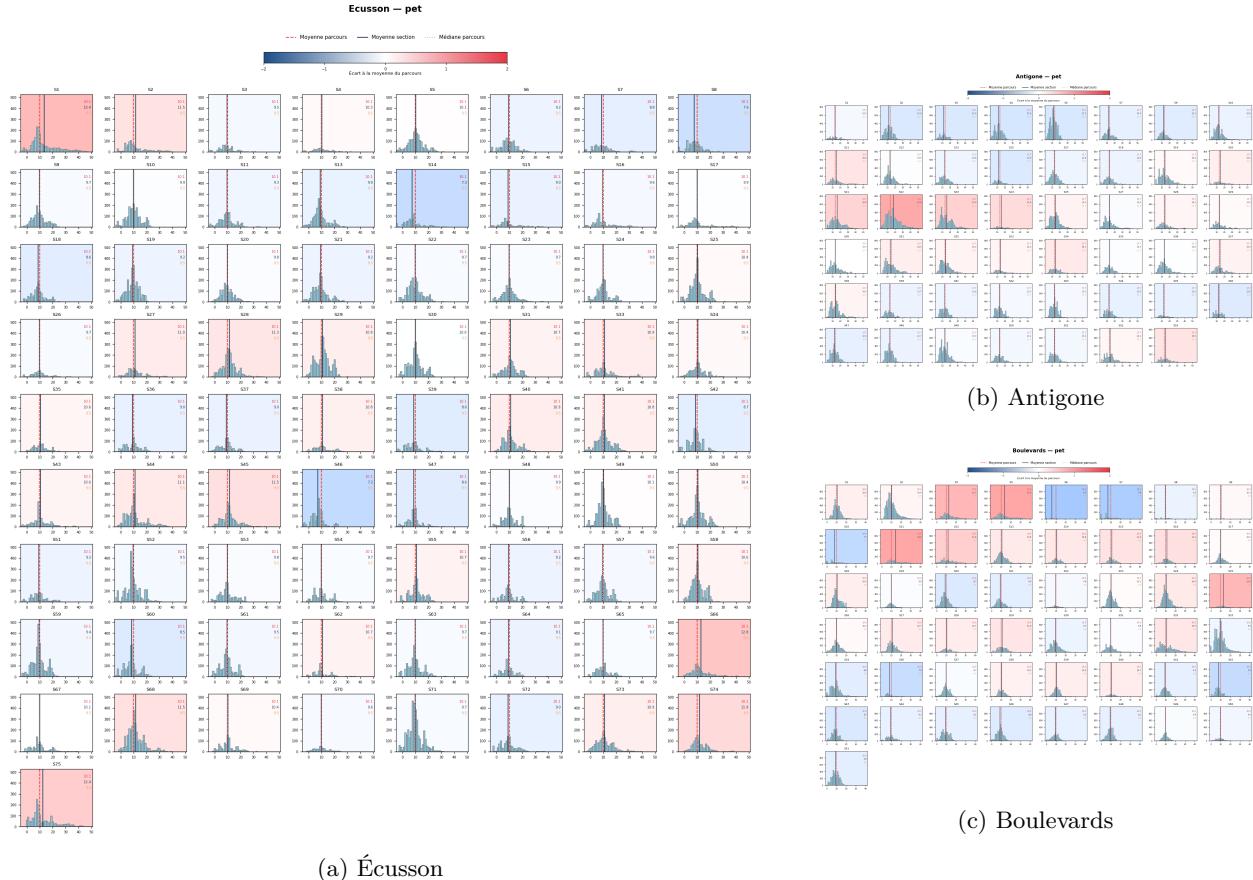


FIGURE 18 – Ensemble des distributions du PET par section pour chaque parcours.

4.1.3. Distribution via boxplot et violin plot

Pour comparer les variables entre parcours (Antigone, Boulevards, Écusson), nous avons représenté leurs distributions avec deux graphiques complémentaires. Le boxplot fournit un résumé robuste et rapide, centré sur la médiane et l'étendue interquartile. Il permet de comparer les niveaux typiques et la dispersion entre parcours, tout en rendant visibles les valeurs extrêmes. Le violin plot complète cette lecture en affichant la forme de la distribution via une densité estimée. Il permet de voir immédiatement si la distribution est très asymétrique, si elle comporte plusieurs régimes (multimodalité), ou si les valeurs extrêmes correspondent à une traîne rare ou à une structure plus régulière. Cette double lecture est utile ici car les mesures agrègent des situations urbaines très contrastées (ombre vs exposition, canyon urbain vs espaces ouverts), ce qui peut produire des distributions non gaussiennes. Deux exemples illustrent l'intérêt de cette approche. D'abord, l'humidité relative `rh_thermohygro` montre des différences nettes entre parcours. Antigone présente un niveau médian plus élevé et une dispersion plus large que Boulevards et Écusson. Surtout, le violonplot suggère plusieurs régimes d'humidité, ce qui indique que la variabilité ne se résume pas à un simple "bruit" autour d'une moyenne. Cela motive des analyses plus fines en séparant par créneau horaire (`M_slot`) ou par sections, afin de relier ces régimes à des contextes spatio-temporels différents.

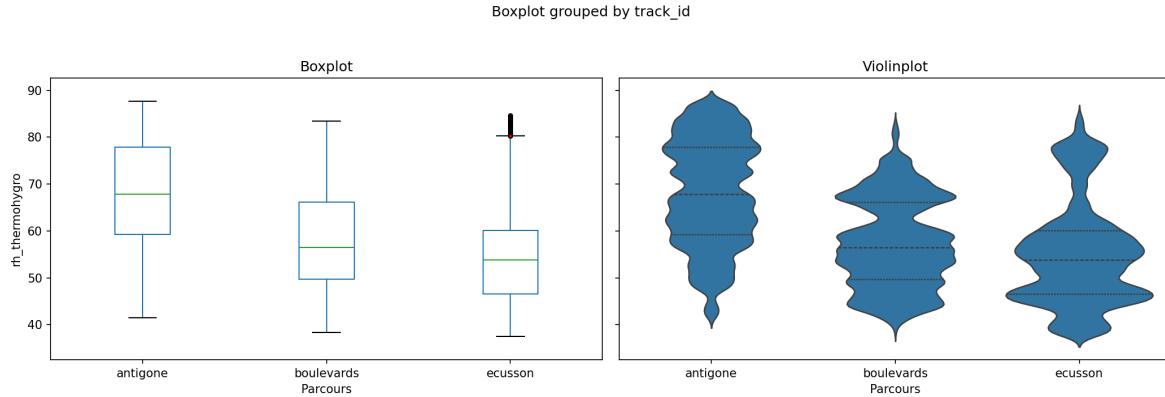


FIGURE 19 – Distribution de rh_thermohygro par parcours (boxplot et violin plot).

Ensuite, la température radiante moyenne tmrt met en évidence une variabilité radiative fortement dépendante de l'exposition. Les distributions présentent une asymétrie marquée, avec une majorité de valeurs modérées et une traîne vers des valeurs plus élevées correspondant à des passages en forte contrainte radiative. Le boxplot permet de comparer rapidement le niveau central et la dispersion entre parcours, tandis que le violin plot renseigne sur la structure de densité, en distinguant une masse principale de mesures et des épisodes plus extrêmes. Pour la suite, ce type de variable justifie l'usage de statistiques robustes (médiane, quantiles) et potentiellement une transformation (par exemple $\log(1+x)$) ou une normalisation robuste avant apprentissage, afin de limiter l'influence des valeurs très élevées.

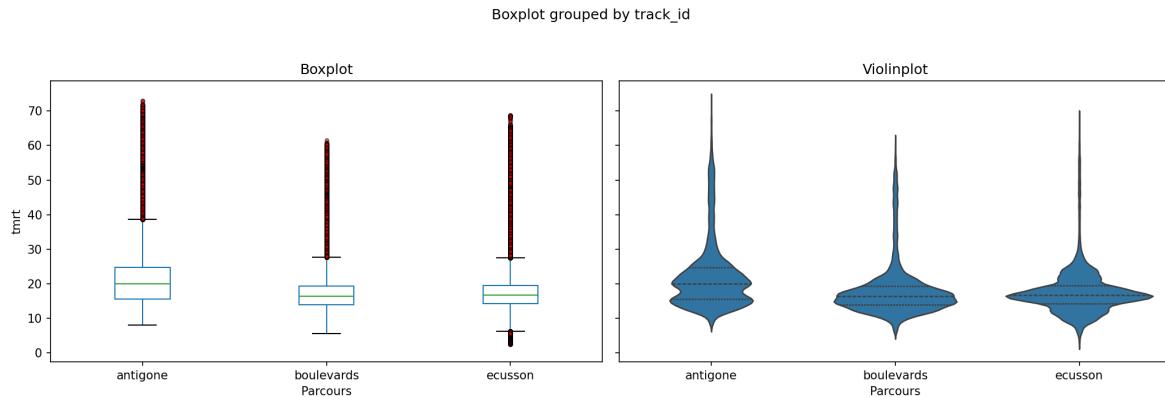


FIGURE 20 – Distribution de tmrt par parcours (boxplot et violonplot).

4.2. Corrélations entre variables

4.2.1. Corrélation entre les variables par parcours

Afin de caractériser les liens entre les variables météorologiques et radiatives mesurées, nous avons calculé des matrices de corrélation pour chacun des trois parcours montpelliérais (Boulevards, Antigone et Écussion). Cette analyse permet de comparer les sites entre eux, d'identifier des structures communes, et de repérer des comportements spécifiques liés au contexte urbain local. Les figures suivantes présentent les matrices obtenues pour chaque parcours.

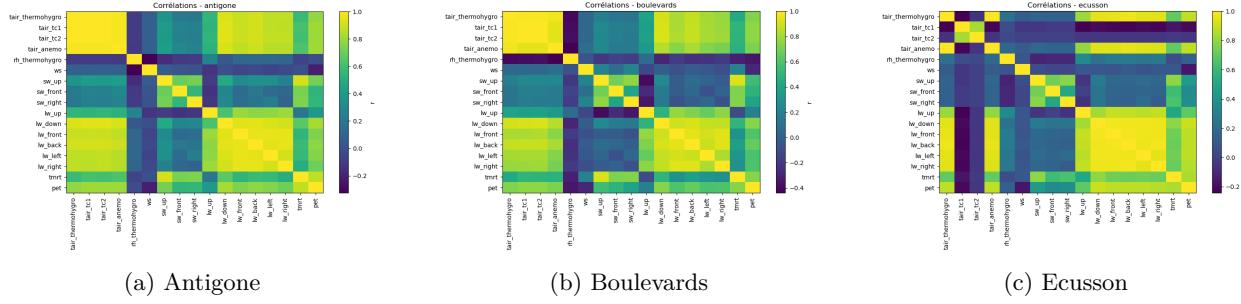


FIGURE 21 – Cartes de corrélation par parcours.

Sur Boulevards et Antigone, les mesures de température de l'air issues des différents capteurs sont fortement corrélées, ce qui indique une évolution cohérente des capteurs au cours du temps. À l'inverse, l'Écusson présente des corrélations plus faibles entre certaines températures, suggérant des variations plus hétérogènes et un micro-environnement plus contrasté (rues étroites, zones d'ombre, ventilation réduite). Concernant le rayonnement, les composantes du rayonnement solaire (**sw_**) montrent généralement des corrélations positives entre elles sur Boulevards et Antigone, traduisant une réponse similaire aux conditions d'ensoleillement. Pour l'ensemble des parcours, les variables radiatives de courte longueur d'onde sont aussi positivement corrélées aux indices de confort (**tmrt** et **PET**), souvent plus nettement que les ondes longues, ce qui reflète l'influence directe du rayonnement solaire sur le bilan radiatif reçu par l'individu. Les corrélations entre rayonnement solaire et infrarouge (**1w_**) sont en revanche plus modestes, cohérentes avec une relation plus indirecte (stockage et réémission par les surfaces). Enfin, les différentes composantes du rayonnement infrarouge évoluent de manière très proche sur les trois sites, avec de fortes corrélations internes au groupe **1w_**. Globalement, Boulevards et Antigone présentent des structures de corrélation assez similaires, tandis que l'Écusson se distingue par des relations moins stables et plus variables.

4.2.2. Corrélation entre les variables par M_slot de chaque parcours

Pour compléter l'analyse, les corrélations ont aussi été étudiées en distinguant les créneaux de mesure (M1 à M4). Cette approche permet d'évaluer si les relations entre variables restent constantes au cours de la journée ou si elles varient selon les conditions (rayonnement, évolution thermique, fin de journée).

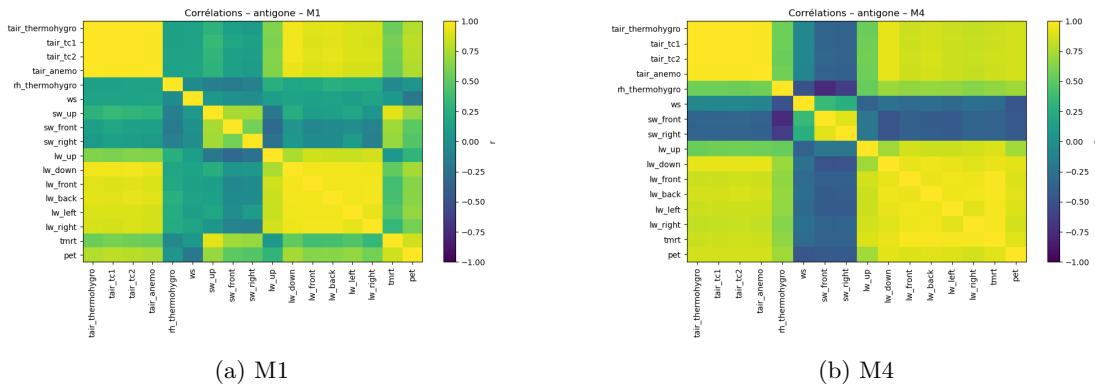


FIGURE 22 – Matrices de corrélation pour Antigone selon le créneau de mesure.

À Antigone, les matrices M1, M2 et M3 présentent des structures très proches, ce qui suggère des rela-

tions globalement stables entre variables sur la majeure partie de la journée. En M4, certaines corrélations deviennent plus faibles ou se réorganisent, ce qui peut refléter un changement des conditions environnementales en fin de journée (baisse de l'ensoleillement, modification du bilan radiatif, évolution de l'humidité). Sur Boulevards, les matrices sont très similaires d'un créneau à l'autre (M1 à M4). Les relations entre variables varient peu, ce qui traduit un comportement plus homogène et une dynamique relativement stable du site.

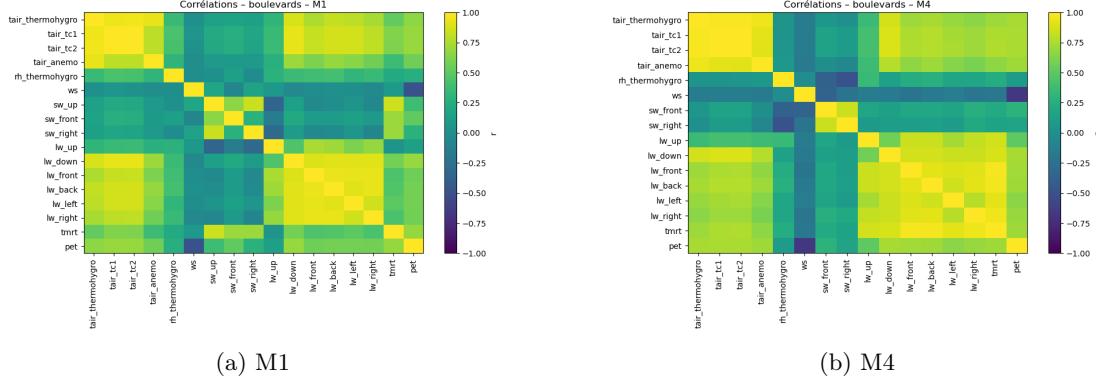


FIGURE 23 – Matrices de corrélation pour Boulevards selon le créneau de mesure.

Pour le parcours de l'Écusson, les matrices de corrélation présentent des différences plus marquées entre les périodes M1 à M4. Les relations entre les variables évoluent davantage au cours de la journée, même si les périodes M2 et M3 montrent des structures de corrélation relativement proches, traduisant un fonctionnement intermédiaire plus stable.

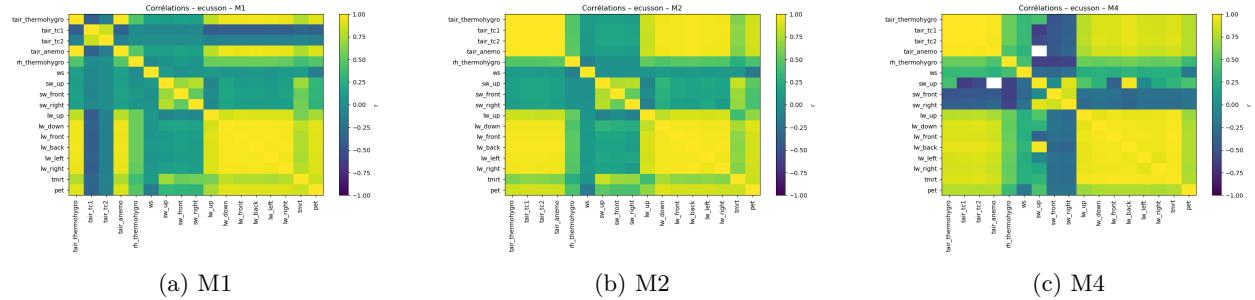


FIGURE 24 – Matrices de corrélation pour Écusson sur plusieurs créneaux.

4.3. Profils spatio-temporels

4.3.1. Agrégé par M_slots

Afin de compléter l'analyse des corrélations présentée en figure 21, nous avons construit des profils moyens par créneau horaire (`M_slot`) et par parcours, pour des variables représentatives du bilan radiatif et des conditions d'ambiance. Nous retenons ici une composante de rayonnement solaire directionnel `sw_right`, un flux d'ondes longues atmosphériques `lw_down`, l'indice radiatif intégré `tmrt` et la vitesse du vent `ws`. Ces variables et leurs capteurs associés sont décrits dans le tableau 3.

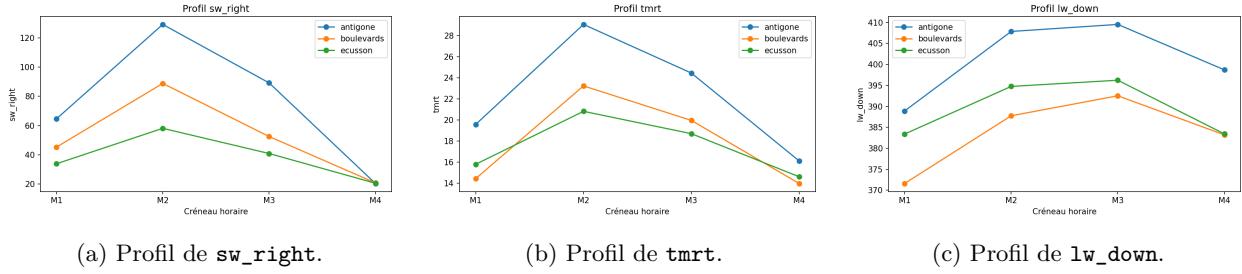


FIGURE 25 – Profils moyens aggrégés par créneau M_{slot} pour chaque parcours.

Les profils de `sw_right` et de `tmrt` présentent des évolutions très proches. Ce résultat est cohérent avec les matrices de corrélation, où les composantes `sw_*` sont parmi les variables les plus liées aux indices de confort, et en particulier à `tmrt` (voir figure 21). Dans les trois parcours, `sw_right` augmente fortement entre M1 et M2, puis diminue en M3 et devient faible en M4 (figure 25a). Cette dynamique correspond au cycle journalier du rayonnement, avec un maximum autour du milieu de journée, puis une chute nette en fin d'après-midi. Le signal reste systématiquement plus élevé sur Antigone, puis Boulevards, puis Écusson, ce qui est cohérent avec les morphologies attendues : Antigone et Boulevards comportent davantage d'espaces ouverts et d'exposition directe, alors que l'Écusson est plus encaissé, avec davantage d'ombre et d'obstructions. La variable `tmrt` suit le même schéma, avec un pic marqué en M2, puis une décroissance jusqu'à M4 (figure 25b). L'écart entre parcours est net, avec Antigone au-dessus des deux autres. Comme `tmrt` intègre l'ensemble des flux radiatifs, ce résultat indique que la contribution radiative, en particulier solaire, pilote fortement le confort thermique radiatif dans ces situations hivernales. Le flux `lw_down` augmente également de M1 à M3 puis diminue en M4 (figure 25c). Cette évolution est plus lisse que celle de `sw_right`. C'est attendu pour l'infrarouge atmosphérique, plus lié à l'état thermique de l'atmosphère et à la couverture nuageuse qu'à l'exposition directionnelle instantanée. On observe toutefois un niveau plus élevé à Antigone sur l'ensemble des créneaux, ce qui peut traduire des conditions atmosphériques légèrement distinctes lors des passages, ou un contexte local influençant le signal infrarouge reçu.

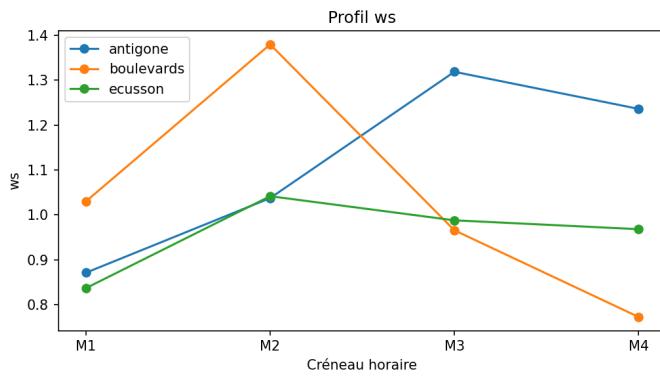


FIGURE 26 – Profil moyen de la vitesse du vent `ws` agrégé par créneau M_{slot} .

La vitesse du vent `ws` se distingue des variables radiatives par une dynamique moins régulière au cours de la journée (figure 26). Les corrélations entre `ws` et les variables radiatives ou `tmrt` sont globalement plus faibles dans la figure 21, ce qui confirme que le vent ne suit pas mécaniquement le cycle solaire. Boulevards présente un maximum en M2 puis une baisse progressive jusqu'à M4, ce qui est compatible avec une ventilation

plus marquée en milieu de journée sur de grands axes. Antigone augmente de M1 à M3 puis reste élevé en M4, ce qui est cohérent avec un quartier plus ouvert où la ventilation peut s'exprimer plus régulièrement. Enfin, l'Écusson reste bas et relativement stable, ce qui correspond à un contexte de canyon urbain où la ventilation est plus contrainte. Au total, ces profils par `M_slot` confirment une cohérence physique forte entre le rayonnement solaire directionnel (`sw_right`) et l'indice radiatif intégrateur (`tmrt`), tandis que `lw_down` suit une évolution journalière plus amortie. La variable `ws` apporte une information complémentaire, davantage contrôlée par la dynamique locale de l'écoulement et la géométrie urbaine que par le cycle radiatif.

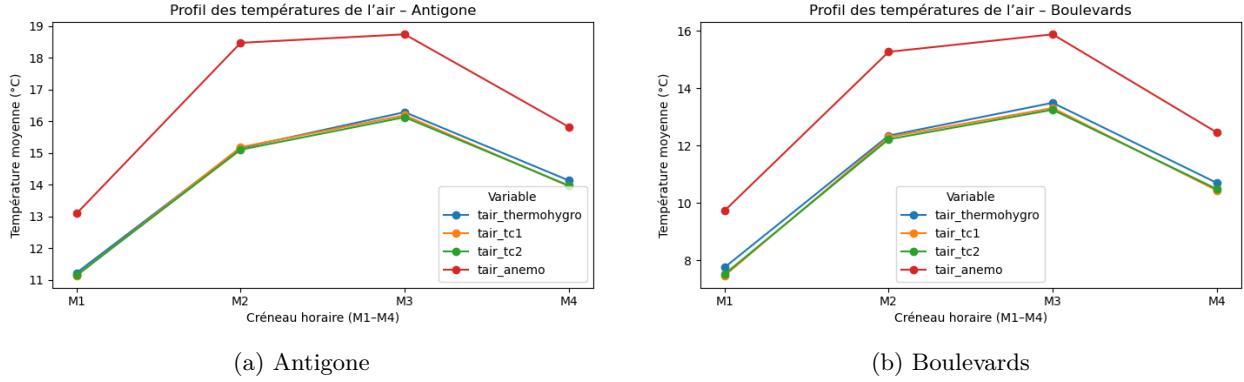


FIGURE 27 – Profils moyens des températures de l'air par capteur, agrégés par créneau `M_slot`.

Les profils de température de l'air confirment la cohérence globale des capteurs au cours de la journée. Sur Antigone et Boulevards, les trois mesures `tair_thermohydro`, `tair_tc1` et `tair_tc2` suivent une silhouette très similaire, avec une hausse marquée entre M1 et M2, un maximum en M3, puis une baisse en M4 (figure 27). Cette structure est cohérente avec l'évolution journalière attendue en période automne hiver, et avec les matrices de corrélation où les températures issues de capteurs différents sont fortement corrélées sur ces deux parcours (figure 21). On observe toutefois un décalage systématique de `tair_anemo`, qui reste plus élevée que les autres capteurs sur l'ensemble des créneaux, avec un écart particulièrement net en M2 et M3. Ce comportement suggère un biais de mesure ou un effet d'exposition spécifique au capteur associé à l'anémomètre, par exemple une moindre protection, une influence radiative plus directe, ou un positionnement différent sur le chariot. Ce point est important pour la suite car il indique que, même si la dynamique temporelle est cohérente, toutes les mesures de température ne sont pas strictement interchangeables. Dans la construction de silhouettes, il est donc préférable de privilégier `tair_thermohydro` comme référence de température de l'air, et d'utiliser `tair_tc1` et `tair_tc2` comme capteurs de comparaison ou de contrôle qualité plutôt que comme source principale.

4.3.2. A l'échelle d'une mesure

Après les profils moyens par créneau (figure 25), nous analysons ici les séries à l'échelle d'un fichier, c'est-à-dire d'un passage unique défini par (`track_id`, `date`, `M_slot`). L'objectif est de visualiser la variabilité spatiale le long du parcours, puis de comparer les créneaux d'une même journée, et un même créneau entre plusieurs dates. Les profils sont tracés en fonction de `point_id`, qui correspond à l'index séquentiel des points mesurés le long du trajet. Les exemples ci-dessous utilisent la variable `lw_down`, décrite dans le tableau 3.

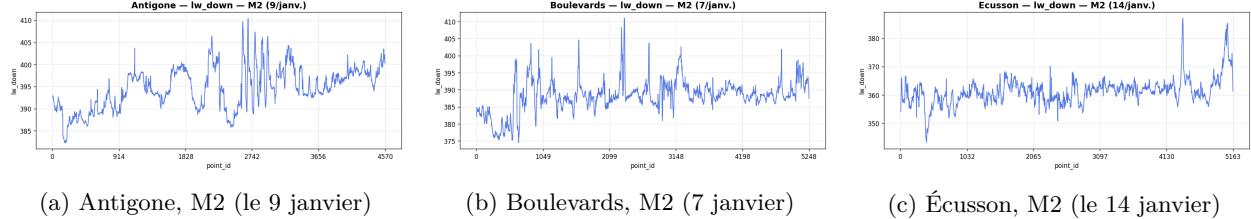


FIGURE 28 – Profils pour la variable `lw_down` en fonction de `point_id` pour un passage unique (un fichier).

On observe des variations locales et des pics marqués, qui traduisent des changements rapides du champ radiatif reçu par la station au cours du trajet. À Antigone, plusieurs pics très abrupts apparaissent au milieu du parcours (figure 28a), suggérant des transitions franches d'environnement, par exemple entre zones plus ouvertes et zones plus contraintes. Boulevards montre une série plus hachée avec des pics isolés (figure 28b), ce qui est compatible avec un parcours alternant places et axes, où la géométrie du canyon urbain change fréquemment. À l'Écusson, le profil est plus régulier sur de longues portions, avec quelques événements ponctuels plus nets en fin de trajet (figure 28c).

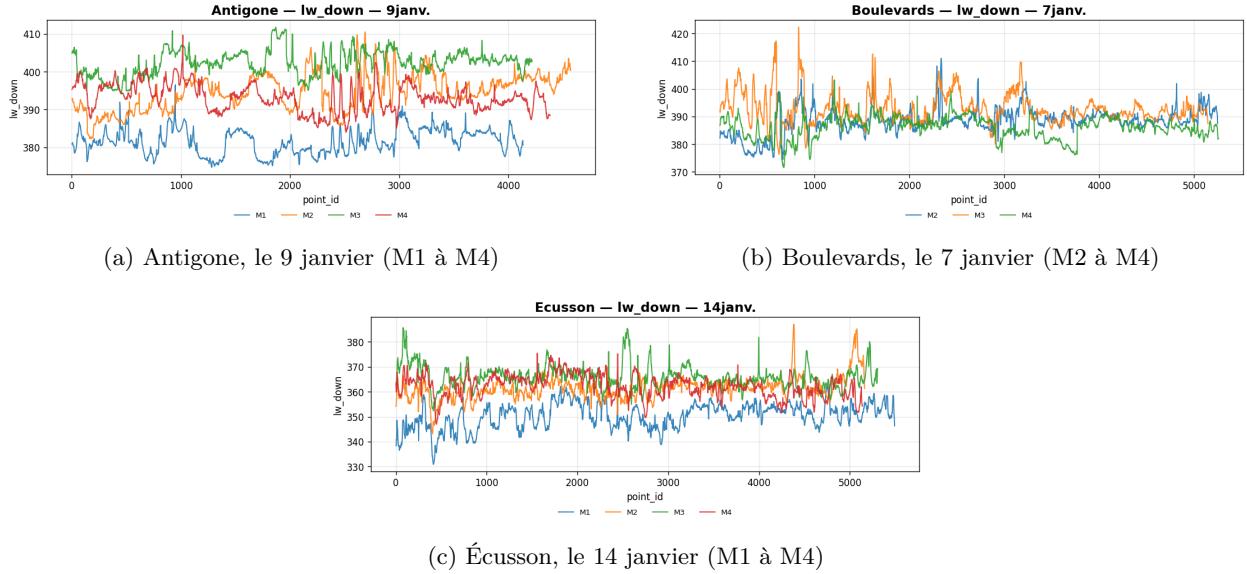


FIGURE 29 – Comparaison des profils `lw_down` sur une même journée.

A noter que la mesure M1 du parcours Boulevards n'a pas été réalisée le 7 janvier 2025.

La comparaison au sein d'une même journée met en évidence un point important pour la notion de silhouette. À parcours fixé, les profils des différents `M_slot` restent structurés par des variations qui apparaissent souvent aux mêmes zones de `point_id`, même si le niveau moyen évolue au cours de la journée (figure 29). Autrement dit, une partie de la forme du signal semble liée à l'espace traversé, tandis que le décalage vertical global reflète l'évolution temporelle des conditions atmosphériques (température de l'air, humidité, nébulosité) et du contexte radiatif. On observe par exemple un ordre global plus élevé en milieu de journée sur Antigone, alors que M1 reste plus bas, ce qui est cohérent avec les créneaux décrits dans le tableau 2. Sur Boulevards, tous les créneaux ne sont pas toujours disponibles sur une date donnée, ce qui est cohérent avec les absences ponctuelles discutées lors de l'analyse de complétude (figure 7).

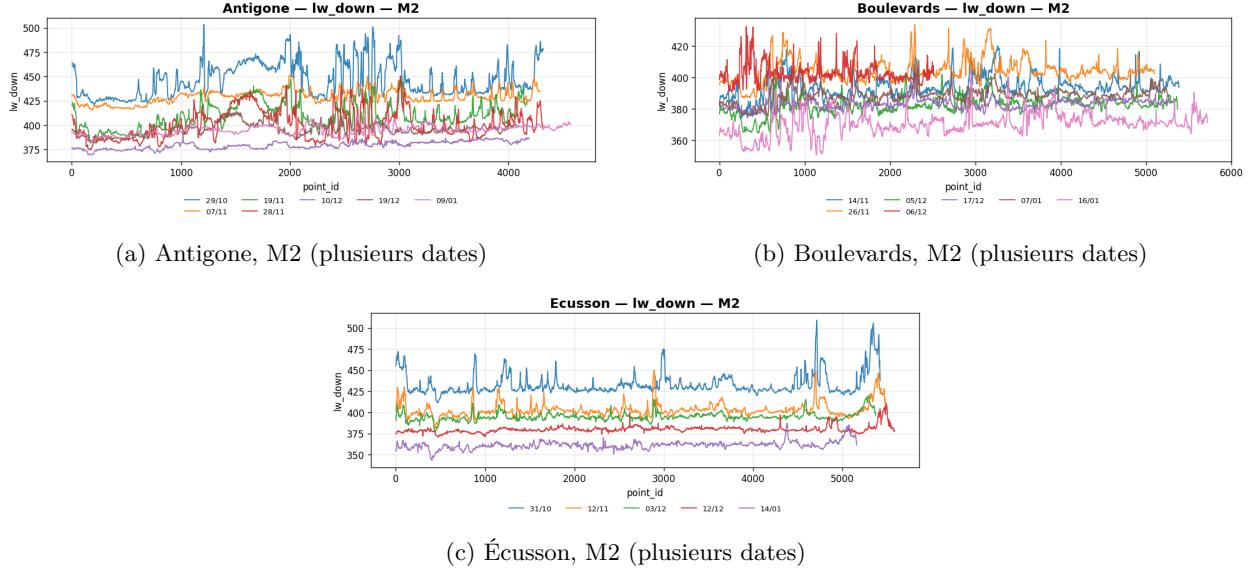


FIGURE 30 – Comparaison de profils `lw_down` pour un même créneau `M_slot` (M2), sur plusieurs dates.

Enfin, la comparaison d'un même `M_slot` entre plusieurs dates permet de séparer plus clairement les effets temporels et les effets spatiaux (figure 30). On observe des niveaux moyens très différents selon les dates, ce qui est attendu car `lw_down` dépend fortement de l'état thermique de l'atmosphère et de la nébulosité. Malgré ces décalages, certaines zones du parcours présentent des motifs récurrents et des ruptures localisées qui réapparaissent sur plusieurs dates, ce qui suggère une signature spatiale relativement stable.

5. Missions restantes

5.1. Les silhouettes

Dans notre approche, une silhouette est un résumé numérique d'un passage (un fichier, donc une combinaison `date` et `M_slot` pour un parcours). Elle est construite à partir de statistiques descriptives calculées sur les séries temporelles mesurées le long du trajet. L'objectif est de transformer un profil bruité point par point en un vecteur de caractéristiques stable et comparable entre passages. Actuellement, la silhouette est composée, pour chaque variable retenue, d'un ensemble d'indicateurs robustes qui décrivent le niveau moyen et la dispersion : moyenne, médiane, écart-type, minimum, maximum, ainsi que des quantiles (par exemple q_{25} et q_{75}) qui capturent la variabilité sans être trop sensibles aux valeurs extrêmes. Cette composition pourra être ajustée ensuite, mais elle constitue déjà une base cohérente pour comparer les passages entre eux. La stratégie prévue est la suivante. Nous construisons d'abord une silhouette pour chaque fichier afin d'obtenir une base homogène sur l'ensemble des passages. Ensuite, nous affinerons en construisant des silhouettes plus locales en ciblant les zones les plus informatives du parcours, par exemple certaines sections ou portions où les contrastes sont les plus nets. Ce travail sur les silhouettes n'est pas encore terminé. Il sera finalisé peu de temps après le rendu de ce rapport.

5.2. Pistes d'apprentissage profond

Après une première étape basée sur des descripteurs statistiques, une suite naturelle est d'apprendre automatiquement des représentations compactes des séries temporelles. L'idée est de produire, pour chaque passage, un vecteur latent qui résume la dynamique du profil et qui joue le rôle de silhouette. Ce vecteur permet ensuite de comparer des passages entre eux, de regrouper des situations similaires et d'identifier des passages atypiques, sans dépendre uniquement de statistiques construites à la main. Plusieurs familles de modèles sont adaptées à ce contexte. Les CNN 1D sont une option simple et robuste pour capter des motifs locaux dans les séries (ruptures, variations rapides, plateaux) et fonctionnent bien sur des fenêtres ou des sections. Les réseaux récurrents (LSTM, GRU) sont pertinents si l'ordre temporel détaillé porte une information importante, mais ils sont plus lourds sur de longues séquences. Les autoencodeurs apprennent un code latent en reconstruisant la série, ce qui fournit à la fois une silhouette apprise et un signal direct pour la détection d'anomalies via l'erreur de reconstruction. Enfin, les Transformers temporels peuvent capturer des dépendances plus longues et des interactions entre variables, mais ils demandent plus de ressources et ne sont utiles que si les gains sont réels au regard du surcoût. Un point important est que les étiquettes de picoclimats sont rarement disponibles au départ. Il est donc pertinent de privilégier des approches auto-supervisées. La reconstruction (autoencodeur) est la plus directe. On peut aussi apprendre des représentations par contraste en créant deux vues proches d'une même série (bruit léger, masquage de points, sous-échantillonnage) et en forçant le modèle à les rapprocher dans l'espace latent. L'objectif est d'obtenir des silhouettes stables malgré le bruit, les coupures courtes et les défauts capteurs. Ces représentations peuvent ensuite être exploitées pour trois usages principaux. D'abord, regrouper et comparer les passages afin de faire émerger des silhouettes typiques. Ensuite, détecter des anomalies instrumentales ou des passages perturbés, via une forte erreur de reconstruction ou un isolement dans l'espace latent. Enfin, tester la capacité du modèle à prédire l'évolution d'une variable le long du trajet, ce qui sert aussi à valider que la représentation capture des régularités exploitables. L'évaluation reposera sur des vérifications simples et interprétables, comme la cohérence des regroupements entre parcours et créneaux, et des visualisations de l'espace après réduction de dimension.

6. Conclusion et perspectives

Ce rapport d'avancement présente les premières étapes d'analyse des données issues des campagnes PICOPATT, avec un focus sur les mesures réalisées à Montpellier entre fin octobre 2024 et mi-janvier 2025. L'objectif général du TER est de progresser vers une caractérisation quantitative des picoclimats urbains, à partir de profils mesurés en continu le long de parcours contrastés, et de préparer la construction de silhouettes capables de résumer ces profils de façon stable et comparable. Un premier résultat important concerne la mise en place d'une base de données exploitable. La compréhension de la structure des fichiers, des variables mesurées et du protocole de terrain a permis d'identifier les points sensibles liés aux campagnes mobiles. Les analyses de couverture ont montré une complétude globalement satisfaisante sur Antigone, plus hétérogène sur Boulevards, et plus irrégulière sur Écusson, ce qui correspond aux contraintes de terrain et devra être intégré dans les comparaisons. Le nettoyage a ensuite permis de traiter des erreurs clairement non physiques, comme des températures extrêmes dues à des défauts ponctuels de capteurs, et de distinguer des valeurs rares mais plausibles, comme certains minima de PET. Enfin, la gestion des valeurs nulles a mis en évidence des variables trop dégradées pour être utilisées telles quelles, et a conduit à une stratégie prudente d'imputation limitée aux coupures courtes, tout en conservant une traçabilité via des valeurs manquantes lorsque la reconstruction serait trop artificielle. Au terme de cette étape, les données sont suffisamment cohérentes pour supporter des analyses descriptives robustes et la construction de représentations synthétiques.

L'analyse exploratoire a ensuite permis de mettre en évidence des structures cohérentes avec la physique des phénomènes observés. Les distributions globales distinguent clairement les comportements des variables radiatives, des variables atmosphériques et des indices de confort thermique. Les matrices de corrélation confirment le rôle central du rayonnement de courte longueur d'onde dans la variabilité de tmrt et, par extension, du PET, tandis que les ondes longues apparaissent plus stables et moins directement liées aux alternances d'ombre et de soleil. Les profils moyens par créneau (**M_slot**) montrent une dynamique journalière attendue pour les variables radiatives, avec un maximum en milieu de journée et une baisse en fin d'après-midi. À l'inverse, la vitesse du vent présente des évolutions moins régulières et davantage dépendantes de la morphologie urbaine, ce qui renforce son intérêt comme variable discriminante entre contextes. Enfin, l'observation des profils à l'échelle d'un passage met en évidence un point clé pour la suite du travail : au-delà des niveaux moyens qui varient d'une date à l'autre, certaines structures du signal réapparaissent à des zones similaires du trajet, ce qui suggère une signature spatiale récurrente compatible avec l'idée de picoclimats locaux.

À ce stade, la démarche a donc permis de sécuriser les données et de produire des constats solides sur la variabilité spatio-temporelle des mesures. La suite du TER vise maintenant à transformer ces observations en objets de comparaison systématique. Le premier travail à finaliser est la construction des silhouettes par passage, puis par zones plus informatives du parcours, en s'appuyant sur des statistiques robustes et en vérifiant la stabilité des résultats selon les parcours et les créneaux. Une fois ces silhouettes stabilisées, elles serviront de base à des regroupements, à la mise en évidence de passages atypiques et à la comparaison inter-parcours. En parallèle, des pistes d'apprentissage profond pourront être explorées de manière progressive, en privilégiant des approches simples et auto-supervisées pour apprendre des représentations compactes des séries. L'enjeu sera de vérifier que ces représentations améliorent réellement la séparation de situations, la robustesse au bruit et la capacité à généraliser entre dates, tout en restant interprétables au regard des mécanismes urbains et radiatifs étudiés dans PICOPATT.

En résumé, ce premier rapport consolide une chaîne de traitement fiable, établit des premiers résultats cohérents avec les caractéristiques urbaines attendues, et prépare la transition vers l'étape centrale du projet : la définition de silhouettes pertinentes et exploitables pour caractériser, comparer et, à terme, identifier des picoclimats urbains le long des parcours.

Références

- [1] *Projet ANR PICOPATT.* <https://anr.fr/Projet-ANR-23-CE22-0005>.
- [2] *Laboratoire AAU Ambiances Architectures Urbanités.* <https://aau.archi.fr/>.
- [3] *Équipe CRENAU du laboratoire AAU.* <https://aau.archi.fr/equipes/crenau/>.
- [4] *LIRMM - Laboratoire d'Informatique, de Robotique et de Microélectronique de Montpellier.* <https://www.lirmm.fr/>.
- [5] *Équipe ADVANSE - Advanced Analytics for Data Science.* <https://www.lirmm.fr/advanse/>.
- [6] *PICOPATT Patterns picoclimatiques urbains.* <https://aau.archi.fr/contrat-de-recherche/picopatt-patterns-picoclimatiques-urbains/>.
- [7] *Dépôt GitHub du projet TER.* <https://github.com/bouyaka34/TER-Picopatt>.