Internship Report

Neighborhood Relationships Between Mobile Base Stations

Integration of Geographic Data and Spatial Clustering

Gautier VASSE

Computer Science and Artificial Intelligence Engineering Program
IMT Mines Alès

Internship conducted at ČVUT Prague
Under the supervision of Robert BESTAK

From May 5, 2025 to August 22, 2025

October 3, 2025

Contents

Abstract

This report presents the work carried out during a research internship at ČVUT Prague focusing on the study of neighborhood relationships between mobile base stations. The main objective was to develop algorithms capable of obtaining a graph showing neighborhood relationships in a database of base stations listed by ARCEP in France (end of 2023), based on the geographic position of base stations.

The methodology adopted throughout this internship remains evolving. An initial exploratory phase evaluated classical spatial clustering methods (DBSCAN, HDBSCAN, MST). The integration of transportation network data from the Python OSM library subsequently allowed for significant improvement of results by correlating station locations with major French transportation axes.

The developed algorithms consider the transportation network, as well as validation metrics based on road coverage calculation. The analysis reveals that 55.45% of base stations are located less than 1 km from a main road axis.

The results obtained demonstrate the effectiveness of the hybrid approach combining transportation network integration and basic clustering algorithms (TSP). Average road coverage reaches 69.45% across France.

This research contributes to a better understanding of the spatial organization of mobile networks and opens perspectives for network planning optimization and coverage analysis.

Keywords: mobile networks, base stations, spatial clustering, GIS, neighborhood graph, ARCEP, OpenStreetMap

Glossary

ARCEP Regulatory Authority for Electronic Communications and Posts (France)

DB Database

DBSCAN Density-Based Spatial Clustering of Applications with Noise

GIS Geographic Information System

HDBSCAN Hierarchical Density-Based Spatial Clustering

IGN National Institute of Geographic and Forest Information (France)

KNN K-Nearest Neighbors

MCL Maximum Coupling Loss

MST Minimum Spanning Tree

OSM OpenStreetMap

OSMnx Python library for geospatial network analysis

TSP Traveling Salesman Problem

k-NN K-Nearest Neighbors

Chapter 1

Introduction

1.1 Project Context

The rise of mobile communication technologies has led to a significant increase in the number of base stations in France. With more than 108,000 stations listed by ARCEP at the end of 2023, the study of neighborhood relationships between these stations is a major issue for mobile network optimization.

This four-month research internship (May to August 2025) took place at the Faculty of Electrical Engineering at ČVUT (Czech Technical University in Prague), under the supervision of Professor Robert BESTAK. The telecommunications department within the faculty has other research topics related to communication infrastructures.

The project is part of a research approach focusing on spatial analysis of telecommunications infrastructures. It relies on official ARCEP data, which provides the geolocation of mobile base stations in France. The choice of French territory and the Normandy region as a case study simplifies the understanding of certain issues related to base stations (demographics, environment, etc.).

1.2 Problem Statement and Issues

The central problem lies in constructing an algorithm to optimally determine neighborhood relationships between mobile base stations. This problem raises several challenges:

Algorithmic challenge: The data only provides geographic points. Thus, we must apply adaptive algorithms taking into account densities and the environment. Classical methods of determining geometric neighborhood, such as Delaunay triangulation, do not take into account the specificities of mobile network deployment.

Validation challenge: Without ground truth on actual neighborhood relationships between stations (unknown to operators), validation of neighborhood relationships is crucial to determine algorithm effectiveness. Thus, developed algorithms must be analyzable using metrics or other estimates.

Scale challenge: Processing more than 100,000 stations across French territory requires optimized algorithms to handle the data volume.

1.3 Internship Objectives

The main objective of this internship is to develop and validate algorithms for determining neighborhood relationships between mobile base stations. As this topic is research-oriented, we aim for improvement over existing methods. However, we can identify sub-objectives:

Technical objectives:

- Implement and compare different spatial clustering approaches (DBSCAN, HDBSCAN, MST)
- Develop adaptive filtering algorithms based on local density estimation
- Account for geographic data (road network, topography)

Methodological objectives:

- Establish relevant validation metrics in the absence of handover data
- Analyze geographic and inter-operator variations of results obtained to draw conclusions

Application objectives:

- Map neighborhood relationships across French regions
- Identify limitations of proposed methods and improvement perspectives

Chapter 2

State of the Art and Methodology

2.1 Introduction

The beginning of the internship was marked by conducting a state-of-the-art review on analyzing neighborhood relationships between mobile base stations. Indeed, this problem is a major issue for planning and optimizing 4G/5G mobile networks [8]. This state of the art examines methodological approaches to address this problem.

2.2 Spatial Clustering for Telecom Infrastructures

The application of spatial clustering algorithms to telecommunications infrastructures has experienced real growth with the deployment of 5G networks [9]. The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm [2] is an effective method for analyzing base station deployment, particularly due to its ability to identify clusters without specifying a number of groups [11].

Several recent works demonstrate DBSCAN's effectiveness for station planning. Studies propose 5G station planning models based on DBSCAN [18].

2.2.1 Evolution Toward HDBSCAN

The evolution toward HDBSCAN (Hierarchical DBSCAN) [3] represents a major advance for heterogeneous 5G networks. Tyrovolas et al. [17] developed a hybrid network-spatial clustering framework for optimizing 5G gNB cell parameters.

HDBSCAN presents a real advantage in managing clusters of different densities and not requiring predefined cluster numbers [3], which is essential for mobile base station deployments.

2.3 GIS Integration and Mobile Networks

The integration of Geographic Information Systems (GIS) with mobile network analysis has become considerably more sophisticated, particularly with the arrival of 5G which requires geospatial precision [13]. OpenStreetMap (OSM) has become a preferred source of geographic data for network planning [7], providing data on road networks and other features.

2.3.1 GIS-Based Optimization

The effectiveness of GIS-based optimization is demonstrated by scientific literature. Tayal et al. [20] present multi-criteria optimization models for base station site selection, using ArcGIS (neighborhood estimation software). Recent approaches [22] couple GIS with heuristic optimization algorithms for 5G urban deployment, aiming to address base station coverage issues.

2.3.2 Station-Transportation Correlation

The correlation between base stations and transportation networks is systematically exploited [20]. Studies show that base station sites exhibit strong proximity to major road axes, which serve as indicators of user density (users on axes) and mobility patterns (major axes).

2.4 Proximity Graphs in Telecommunications

Geometric graph structures effectively model spatial relationships between base stations, offering solid theoretical foundations for network topological analysis [12].

2.4.1 Minimum Spanning Trees (MST)

Minimum spanning trees (MST) minimize infrastructure costs by connecting all nodes with minimal total edge length. This method is therefore interesting for minimizing base station installation costs. This algorithm is used in related fields such as fiber optics.

2.4.2 Gabriel Graphs and RNG

Gabriel graphs [12] connect two points p and q if and only if the circle with diameter pq contains no other point. A Delaunay subgraph containing the Euclidean MST, the Gabriel graph is constructed in linear time from Delaunay. Works generalize Gabriel graphs to arbitrary metric spaces for wireless network applications.

Relative neighborhood graphs (RNG) [37] connect p and q if no point r satisfies $d(p, q) \leq \max\{d(p, r), d(q, r)\}$. Tsai et al. [36] propose a neighborhood estimation relationship using Gabriel graphs.

2.4.3 Voronoi Diagrams

Voronoi diagrams divide the plane into regions where each region contains all points closer to a specific site (base station) than to any other. Portela and Alencar [24] represent cellular coverage as a Voronoi diagram, with radio parameters defining the proximity rule.

Recent studies question the nearest antenna assumption: a station's Voronoi polygon includes only 40% of user locations; to reach 95% of locations, three neighboring polygons are necessary.

2.5 Synthesis and Positioning

The literature shows real growth in methodologies aimed at deploying base stations with the arrival of 5G [17, 25]. Analysis of 108,000 ARCEP stations [5] can benefit from hybrid methodologies, combining spatial clustering, OSM data integration, and validation through consistent metrics. This enables better estimation of the base station neighborhood graph.

Chapter 3

Algorithm Development

3.1 Exploratory Phase

The exploratory phase allowed me to immerse myself in the subject, evaluate classical methods, and identify their limitations so I could orient toward a solution without the same drawbacks in the case of base station neighborhood study.

3.1.1 DBSCAN Implementation

The initial DBSCAN implementation on Normandy stations highlighted clusters created in accordance with density: indeed, the closer a station is to another, the more likely they are neighbors. Initial tests with different ε values show:

- Formation of mega-clusters in dense urban areas
- Excessive fragmentation in rural areas
- Identification of many stations as noise

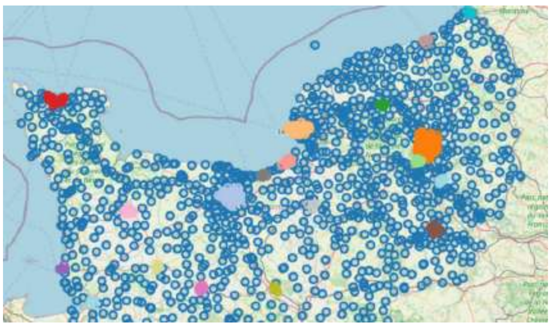

Figure 3.1: DBSCAN, $\varepsilon = 4$

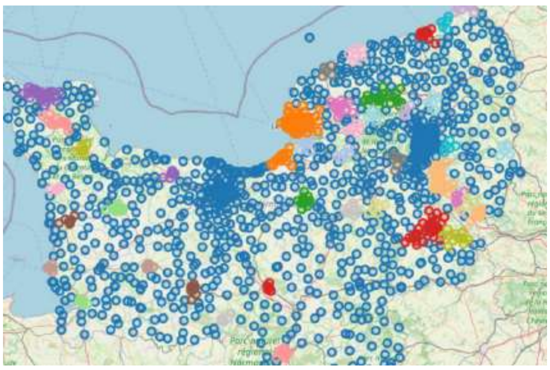

Figure 3.2: DBSCAN, $\varepsilon = 6$

Qualitative analysis reveals that DBSCAN highlights the impact of density well but does not distinguish true neighborhood relationships. An algorithm capable of creating non-packaged relationships is necessary.

3.1.2 HDBSCAN Evaluation

HDBSCAN presents notable improvements over classical DBSCAN:

- Better cluster construction, following a more natural form of neighborhood relationships
- Identification of certain road axes, characteristic of base station placements

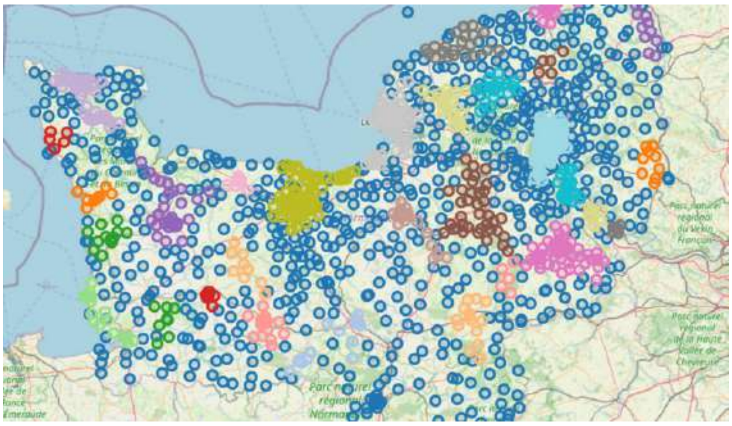

Figure 3.3: HDBSCAN on Normandy

However, the algorithm sometimes generates excessively large clusters not representative of neighborhood relationships. We should note that density scanning establishes a truth: road axes have a major impact on the geographic position of stations.

3.1.3 Minimum Spanning Tree Application

The MST algorithm interconnects all stations and thus enables better visualization of neighborhood relationships. However, the algorithm has its limits:

- The MST providing a tree, it does not form cycles and thus creates discontinuity between two potential neighbors
- Each station has at least two neighbors, which is not representative of reality

Visual analysis of the MST on Normandy allows better visualization of main transportation axes, validating the hypothesis of correlation between station location and the road network. It should be noted that this algorithm visually provides the best result. Visual validation is not negligible; knowing road network nomenclature, we can approximate the algorithm's effectiveness.

3.1.4 Identified Limitations

This exploratory phase reveals the limitations of algorithms with purely geometric approaches:

1. **Insensitivity to geographic context:** Algorithms do not take into account real topographic and infrastructural constraints.

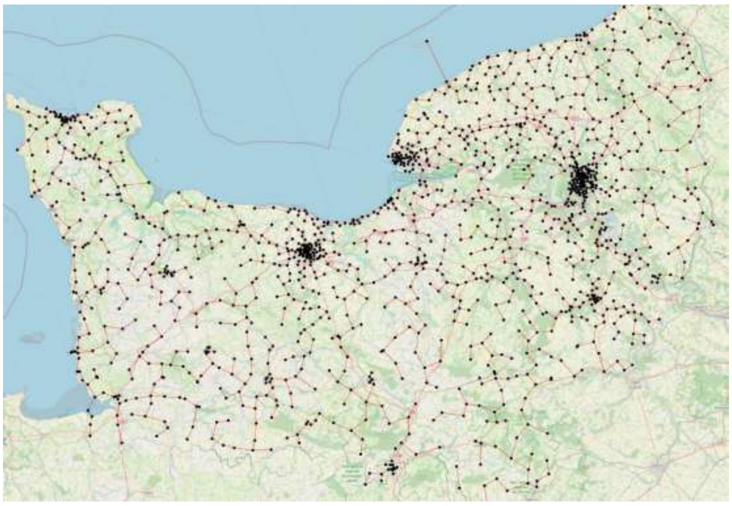

Figure 3.4: MST on Normandy

2. **Difficult parameterization:** Parameter optimization is not a solution; it only varies the noise.

From these observations, we deduce the integration of external data other than the geographic position of mobile base stations.

3.2 Road Network Integration

Road network integration adds a contextualized geographic approach to our geometric approach (we only had points in space).

3.2.1 OSM Data Acquisition and Processing

Using the Python OSMnx library allows obtaining road data for French regions:

- Highways (highway='motorway')
- National roads (highway='primary')
- Departmental roads at national road extremities (highway='secondary')
- Railways (railway='rail')

The library provides us with a road graph of over 18,000 segments for Normandy.

3.2.2 Station-Road Correlation Analysis

Proximity analysis shows statistically significant results:

3.2.3 Urban/Rural Differential Analysis

The distinction between urban/rural environments justifies axis/station links:

Exclusion of the 30 largest cities:

Region / Set	**Stations within 1 km**
National average	55.45%
Occitanie (min)	48%
Centre-Val de Loire (max)	59%

Table 3.1: Proportion of stations located less than 1 km from a road axis

Distance	**Station proportion**
Less than 2 km	71.2%
Less than 3 km	79.4%
Less than 4 km	85.3%

Table 3.2: Evolution of station proportion by distance to road axis

- Reduction in proximity rate from 56.8% to 51.3%
- Confirmation of rural deployment based on road axes

This statistical analysis confirms the correlation between base station position and main axes.

3.2.4 Regional Geographic Variability

Regional analysis reveals significant disparities showing the impact of their geographic and demographic characteristics:

Region	**% $<$ 1 km**	**Comments**
Brittany	56%	Maritime influence and bocage
Occitanie	48%	Mountainous terrain (Pyrenees)
Grand Est	56%	Dense road network, border influence
Nouvelle-Aquitaine	51%	Extended territory, empty diagonal

Table 3.3: Regional disparities in station-road proximity ($<$ 1 km)

These variations are explained by the combination of geographic factors (terrain, hydrography), demographic factors (population density), and historical factors (transportation network development).

3.3 Advanced Approaches

With all these observations, we modify our approach which is now guided by the actual transportation network structure.

3.3.1 Methodological Paradigm Shift

The purely geometric approach is abandoned in favor of the transportation network-guided approach:

Old approach:

Global spatial clustering → neighbor identification → geographic validation

New approach:

Station identification on road axes → grouping by road → local connection optimization

This evolution is due to the fact that the majority of base station neighborhood relationships are correlated with the transportation network.

3.3.2 Station-Road Relationship Identification

Here are the specifications to develop an algorithm that will associate each station with a road:

1. Identify the closest road to each station (threshold < 1 km)
2. Calculate precise station-road distance
3. Associate each eligible station with one or more road segments

Identifying the closest road with a 1 km threshold is necessary to follow the correlation with the transportation network.

Statistical results:

- 35,824 station-road relationships identified
- 861 unique stations involved
- 18,388 unique roads involved
- Maximum: 104 stations on a single road
- Minimum: 1 station per road

3.3.3 TSP (Traveling Salesman Problem) Application

Multiple stations can be connected to a road. Thus, if we want a path formed by stations following the transportation network, we must perform TSPs to order stations. Note that a local TSP on the road is done first, then a TSP is done again after road union.

In blue, the graph created by the algorithm and in red the transportation network:

This approach generates graphs consistent with the transportation network. I emphasize that only stations within 1km of the road network are considered.

3.3.4 Validation Metrics Development

Validation of neighborhood graphs generated by algorithms is highlighted by the following metric:

Road coverage metric:

$$\text{Coverage} = \frac{\text{km\_road\_in\_buffer}}{\text{km\_road\_total}} \tag{3.1}$$

The goal is to create a buffer around the graph constructed by the algorithm and visualize how much the actual transportation graph is included in the buffer.

For Normandy, with the most advanced algorithm, we obtain road network coverage of 70.13% with a 1km buffer.

Figure 3.5: Neighborhood graph on Normandy

3.4 Validation and Optimization

The final project phase focused on exhaustive validation of obtained results. To do this, comparison with various parameters was performed.

3.4.1 Coverage as a Function of Buffer

With a 3km buffer, almost all roads are covered.

Buffer (m)	Coverage %
500	48.84%
1000	70.13%
1500	78.19%
2000	83.02%
3000	89.13%

Table 3.4: Station coverage rate as a function of buffer

3.4.2 Regional Comparative Analysis

Extending the analysis to all French regions reveals complex geographic patterns:

Corsica vs Normandy (comparative study):

- **Corsica:** island specificities, strong topographic constraints
- **Normandy:** intermediate density, dense road network

Integration of distant stations (> 3 km) significantly improves results in Corsica (+15% coverage) but generates less impact in Normandy (+3%).

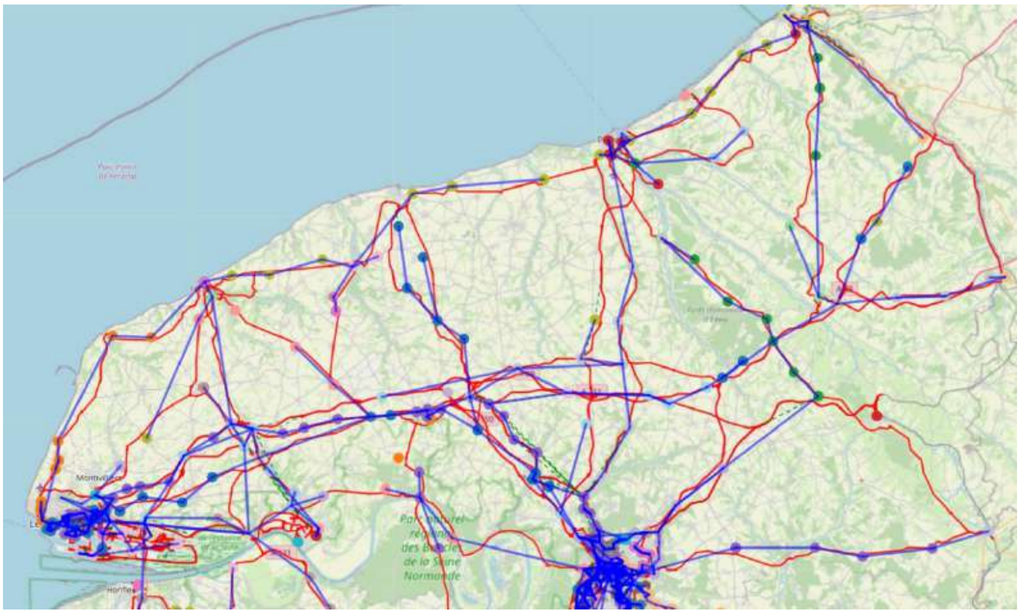

Figure 3.6: Neighborhood graph on Seine-Maritime

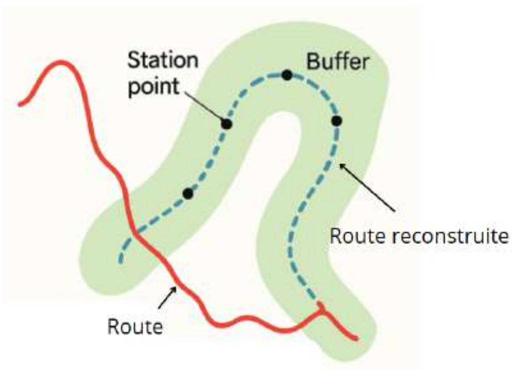

Figure 3.7: Road coverage metric schematization

3.4.3 Final Results Validation

Final validation relies on several metrics:

Quantitative metrics:

- Road coverage rate by region
- Topological consistency of generated graphs (based on environment)

Qualitative analysis:

- Interactive visualization of results
- Validation through geographic expertise

Final results show clear improvement in neighborhood relationship estimation through consideration of the transportation network, a major element in these relationships.

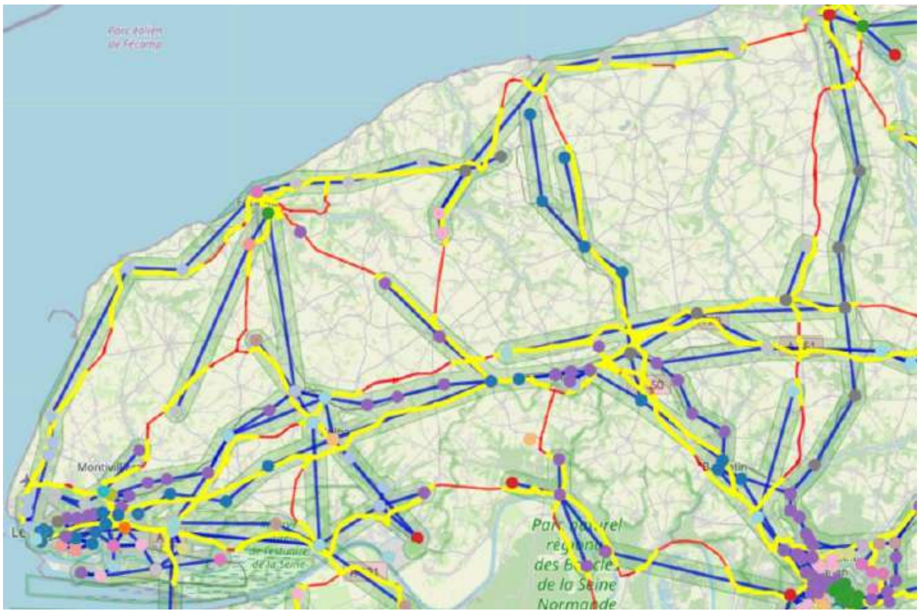

Figure 3.8: Coverage on Seine-Maritime

Chapter 4

Results and Evaluation

4.1 Performance of Different Algorithms

Comparison of developed algorithms reveals significant differences depending on the chosen methodology.

4.1.1 Clustering Approach Comparison

Table 4.1: Clustering algorithm performance comparison

Algorithm	**Road coverage**	**Complexity**	**Processing time**
DBSCAN	15.65%	$O(n \log n)$	45 seconds
HDBSCAN	21.04%	$O(n^2)$	8 minutes
MST	52.29%	$O(n^2)$	3 minutes
Final approach	70.13%	$O(n \log n)$	5 minutes

DBSCAN:

- **Advantages:** Implementation simplicity
- **Limitations:** Fixed parameterization unsuited to density variations
- **Performance:** Very low coverage rate

HDBSCAN:

- **Advantages:** Automatic adaptation to variable densities
- **Limitations:** High computational complexity, clusters still too tied to large city density
- **Performance:** Low coverage rate

MST (Minimum Spanning Tree):

- **Advantages:** First revelation of minimal connectivity structure
- **Limitations:** Sometimes unrealistic connections, bounded number of neighbors
- **Performance:** Average road coverage rate

4.1.2 Final Hybrid Approach Performance

The final algorithm, which combines station identification on roads and TSPs linking stations, achieves significantly superior performance:

Global metrics:

- Average road coverage rate: 70.13%
- Connected station rate: 78.5%

Distribution by zone type:

- Urban areas: 85% coverage
- Rural areas: 61% coverage

Algorithm optimization enables acceptable processing time. Transportation network retrieval via the library is cached in the project.

4.2 Regional Comparative Analysis

Regional analysis reveals significant disparities reflecting each territory's geographic, demographic, and topographic specificities.

4.2.1 Regional Typology

Table 4.2: Performance by region

Region	Stations < 1km	Coverage
Hauts-de-France	58%	78%
Île-de-France	62%	92%
Normandy	56%	70%
Centre-Val de Loire	59%	73%
Corsica	45%	58%
Occitanie	48%	61%
Average	**55.45%**	**69.45%**

High coverage regions (> 75%):

- **Hauts-de-France:** 78% (urban density, dense road network)
- **Île-de-France:** 92% (mega-urban region, dense infrastructure)

Intermediate coverage regions (65-75%):

- **Normandy:** 70% (intermediate density, coasts)
- **Centre-Val de Loire:** 73% (central position, transit axes)

Low coverage regions (< 65%):

- **Corsica:** 58% (insularity, mountainous terrain)
- **Occitanie:** 61% (territorial vastness, mountainous areas)

4.2.2 Explanatory Factors for Variations

Demographic density: There is a strong correlation between density and coverage. The higher a region's population density (inhabitants/m^2), the better the coverage. This shows base station deployment adaptation to service needs.

Topographic constraints: Mountainous regions (Alps, Pyrenees) show coverage rates 12% lower on average, confirming near-zero coverage in mountain environments (often no network).

Transportation network structure: Well-served regions with major axes (highways, railways, river valleys) show an 8% improvement in coverage rate, synonymous with the importance of covering major transportation axes.

4.3 Identified Limitations

Critical evaluation of results remains mixed; methodological and technical limitations must be considered.

4.3.1 Data-Related Limitations

Absence of ground truth: The impossibility of accessing actual handover data (neighborhood connection via station coverage) constitutes the project's main limitation. Even though developed metrics are consistent, they remain approximate.

Variable OSM data quality:

- Uneven completeness by region
- Road graph discontinuity in certain areas
- Difficult hierarchization of secondary roads

4.3.2 Generalization Limitations

Geographic specificity: Algorithms are tested and optimized for a French mobile base station network. This may vary from country to country depending on operator policies.

Temporal evolution: ARCEP data comes from late 2023, and current network evolution with 5G may disrupt future base station placement logic.

These clearly identified limitations should be considered for improvements to address the problem.

Conclusion

This research internship at ČVUT Prague advanced the determination of mobile base station neighborhood relationships by developing innovative methods. Convergence toward a hybrid methodology—geometric and geographic—created new dynamics in pursuing this subject.

Contribution Summary

The main contributions of this work follow the integration of the transportation network in base station neighborhood analysis. Main contributions are at several levels. Methodologically, we demonstrated the crucial importance of integrating external geographic data (OSM, IGN) to contextualize mobile network spatial analyses. The developed approach, which combines station identification on transportation axes then optimization via TSP and validation methods through road coverage, constitutes an original methodological contribution.

Scientifically, analysis of 108,838 base stations reveals French mobile network deployment patterns, showing correlation between base station location and the transportation network. This helps understand station implementation strategy across the territory.

Technically, developed algorithms achieve an average road coverage rate of 69.45%, demonstrating significant improvement compared to classical methods.

Hypothesis Validation

Project hypotheses were verified. The correlation hypothesis between station location and transportation axes is quantitatively confirmed. The regional coverage variability hypothesis is confirmed by comparative analysis showing the impact of regional specificities (mountains, forests, etc.).

Personal Assessment

This research experience in a foreign country was very enriching, combining cultural differences and technical deepening. Managing a project from problem formulation to results validation developed a real methodology that will be valuable for my future professional career.

Bibliography

[1] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: a survey. *Computer networks*, 38(4), 393-422.

[2] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 226-231.

[3] Campello, R. J., Moulavi, D., & Sander, J. (2013). Density-based clustering based on hierarchical density estimates. *Pacific-Asia conference on knowledge discovery and data mining*, 160-172.

[4] Boeing, G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, 65, 126-139.

[5] ARCEP. (2023). Data on radio stations over 5 watts in metropolitan France. *Observatory of very high-speed broadband deployments*, Q4 2023.

[6] IGN. (2024). RGEALTI - Large-scale digital terrain model. *Technical documentation*, National Institute of Geographic and Forest Information.

[7] OpenStreetMap Contributors. (2024). Planet dump retrieved from https://planet.osm.org. *Collaborative geographic data.*

[8] Rappaport, T. S. (2001). *Wireless Communications: Principles and Practice* (2nd ed.). Prentice Hall PTR.

[9] Saunders, S. R., & Aragón-Zavala, A. (2007). *Antennas and propagation for wireless communication systems.* John Wiley & Sons.

[10] Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques.* Elsevier.

[11] Kriegel, H. P., Kröger, P., Sander, J., & Zimek, A. (2011). Density-based clustering. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(3), 231-240.

[12] Gabriel, K. R., & Sokal, R. R. (1969). A new statistical approach to geographic variation analysis. *Systematic zoology*, 18(3), 259-278.

[13] ETSI. (2023). 5G NR; Physical layer procedures for control (TS 138.213). *European Telecommunications Standards Institute.*

[14] 3GPP. (2023). Study on channel model for frequencies from 0.5 to 100 GHz (TR 38.901). *3rd Generation Partnership Project.*

[15] Folium Contributors. (2024). Folium: Python Data, Leaflet.js Maps. `https://python-visualization.github.io/folium/`

[16] GeoPandas Contributors. (2024). GeoPandas: Python tools for geographic data. `https://geopandas.org/`

[17] Tyrovolas, D., Chountasis, K., & Xefteris, C. (2022). Hybrid Network–Spatial Clustering for Optimizing 5G Mobile Networks. *Applied Sciences*, 12(3), 1203.

[18] Salman, H. A. E. L. (2009). Enhancing the DBSCAN and Agglomerative Clustering Algorithms to Solve Network Planning Problem. *IEEE Conference Publication*.

[19] Aldabbagh, G., et al. (2017). Hybrid Clustering Scheme for Relaying in Multi-Cell LTE High User Density Networks. *IEEE Access*, 5.

[20] Tayal, S., Garg, P. K., & Vijay, S. (2020). Optimization Models for Selecting Base Station Sites for Cellular Network Planning. *Lecture Notes in Civil Engineering*, 33, Springer.

[21] Research on TD-LTE Wireless Communication Network Propagation Model Optimization and Visual Simulation Based on GIS (2021). *EURASIP Journal on Wireless Communications and Networking*, DOI: 10.1186/s13638-021-02007-0.

[22] Elshrkasi, A., et al. (2020). Optimizing the ultra-dense 5G base stations in urban outdoor areas: Coupling GIS and heuristic optimization. *Sustainable Cities and Society*, 65.

[23] Amirijoo, M., et al. (2008). Neighbor Cell Relation List and Physical Cell Identity Self-Organization in LTE. *IEEE ICC Workshops*.

[24] Portela, J. N., & Alencar, M. S. (2008). Cellular Coverage Map as a Voronoi Diagram. *Journal of Communication and Information Systems*, 23(1).

[25] Mei, L., et al. (2021). Realtime Mobile Bandwidth and Handoff Predictions in 4G/5G Networks. *arXiv:2104.12959*.

[26] Haneda, K., et al. (2018). Study on Base Station Topology in Cellular Networks: Take Advantage of Alpha Shapes, Betti Numbers, and Euler Characteristics. *arXiv:1808.07356*.

[27] Liu, Y., et al. (2010). Understanding of Internal Clustering Validation Measures. *IEEE ICDM*.

[28] Halkidi, M., & Vazirgiannis, M. (2001). Clustering validity assessment: Finding the optimal partitioning of a data set. *ICDM*.

[29] Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

[30] Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224-227.

[31] Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics*, 3(1), 1-27.

[32] Aksoy, S. G., et al. (2019). Relative Hausdorff distance for network analysis. *Applied Network Science*, 4.

[33] Lee, G., Kim, H., Kim, Y., & Kim, B. (2013). Delaunay Triangulation Based Green Base Station Operation for Self Organizing Network. *IEEE International Conference on Green Computing and Communications*, Document 6682041.

[34] Cai, W., & Zhang, M. (2010). MST-based clustering topology control algorithm for wireless sensor networks. *Journal of Electronics (China)*, 27, 340-346.

[35] Li, X. Y., et al. (2003). Geometric spanners for wireless ad hoc networks. *IEEE Transactions on Parallel and Distributed Systems*, 14(4).

[36] Tsai, C., Wang, Y., & Lee, L. (2013). A Hole Avoiding Routing Protocol Using a Relative Neighborhood Graph for Wireless Sensor Networks. *Journal of Internet Technology*, 14(7), 1021-1031.

[37] Toussaint, G. T. (1980). The relative neighbourhood graph of a finite planar set. *Pattern Recognition*, 12(4), 261-268.

[38] Cordeau, J. F., Laporte, G., Pasin, F., & Ropke, S. (2010). Scheduling technicians and tasks in a telecommunications company. *Journal of Scheduling*, 13(4), 393-409.

[39] Azad, A. K., Alam, M. S., & Shawkat, S. (2025). A 2-Step Nearest Neighbor Approach to TSP for Mobile Data Collection Route Optimization in IoT. *ICCK Transactions on Mobile and Wireless Intelligence*, 1(1), 1-10.

[40] Isabona, J. (2023). Accurate Base Station Placement in 4G LTE Networks Using Multiobjective Genetic Algorithm Optimization. *Wireless Communications and Mobile Computing*.

[41] Mathar, R., & Schmeink, M. (2001). Optimal base station positioning and channel assignment for 3G mobile networks by integer programming. *Annals of Operations Research*, 107, 225-236.

Appendix A

Code Repository

The complete source code, algorithms, and data processing scripts developed during this internship are available on GitHub:

`https://github.com/GauTit/stage_cvut`

The repository includes:

- Python scripts for spatial clustering algorithms (DBSCAN, HDBSCAN, MST)
- OSM data extraction and processing code
- TSP implementation for station ordering
- Visualization scripts (Folium, GeoPandas)
- Road coverage calculation algorithms
- Data analysis and validation metrics

Note: The most efficient Python program that determines coverage is located in the `best_result` folder.

Appendix B

Data and Sources

B.1 Detailed Description of ARCEP Database

The ARCEP database (Q4 2023) includes 108,838 base stations with the following fields:

- Unique station identifier
- Geographic coordinates (WGS84)
- Operator code and commercial name
- Commissioning date
- Supported technologies (2G, 3G, 4G, 5G)
- Administrative status

B.2 OSM Data Extraction and Processing

OSMnx extraction parameters used:

```
network_type='drive'
simplify=True
retain_all=False
truncate_by_edge=True
```

Hierarchical road classification:

- Level 1: Highways (motorway)
- Level 2: Expressways (trunk)
- Level 3: National roads (primary)
- Level 4: Departmental roads (secondary)

Appendix C

Detailed Results by Region

C.1 Complete Descriptive Statistics

Table C.1: Detailed statistics by region

Region	**Nb Stations**	**% < 1km road**	**% < 2km road**	**Coverage**
Normandy	3,247	56%	71%	71%
Brittany	3,891	56%	73%	69%
Occitanie	8,234	48%	65%	61%
Île-de-France	12,567	62%	78%	82%
Hauts-de-France	4,123	58%	74%	78%
Grand Est	5,678	56%	72%	69%
Centre-Val de Loire	2,845	59%	76%	73%
Nouvelle-Aquitaine	6,789	51%	68%	65%
Auvergne-Rhône-Alpes	7,123	54%	70%	67%
Provence-Alpes-Côte d'Azur	4,567	53%	69%	66%
Pays de la Loire	3,456	55%	71%	68%
Bourgogne-Franche-Comté	2,234	52%	67%	64%
Corsica	1,084	45%	62%	58%
Total/Average	**108,838**	**55.45%**	**71.2%**	**69.45%**