URBAN CLIMATE AND MOBILITY:

A Graph Neural Network-based Approach to Predict the Effect of Urban Climate on Personal Mobility Choices in Seoul, South Korea

Michele Giampaolo

October 2025

TU Delft

URBAN CLIMATE AND MOBILITY:

A Graph Neural Network-based Approach to Predict the Effect of Urban Climate on Personal Mobility Choices in Seoul, South Korea

Michele Giampaolo

October 2025

A thesis submitted to the Delft University of Technology
in partial fulfillment of the requirements for

The degree of Master of Science in Geomatics for the Built Environment

1st Supervisor:	Azarakhsh Rafiee
2nd Supervisor:	Martín Mosteiro Romero
Co-reader:	Saeed Rahmani
Delegate:	Herman de Wolff

Cover: Photo by Jieun Kim

Source code for reproducibility of this thesis is available at:
github.com/michelegiampaolo/mobility-seoul

Acknowledgements

I would like to thank the following people, each of whom immeasurably helped me finish this research and this master program.

My supervisors, Azarakhsh and Martín, for their continuous patience and guidance, from the initial inception of the research topic to the valuable feedback during our last weekly meetings.

My co-reader, Saeed, whose expertise and rigor greatly influenced the completion of the project.

My partner, Josh, for all the love and support, without which I could not have done this.

My friends and family across the Netherlands, Italy, and Australia for all the help, laughs and meals spent together that helped keep me going throughout this whole process.

Thank you!

Abstract

Urban climate affects how people move through cities, but its influence is difficult to capture with models based on generalized comfort indices that ignore individual experiences of climate. This thesis instead explores a bottom-up approach that uses daily Global Navigation Satellite System (GNSS) traces of people traversing an urban environment, which inherently contain each individual's personal influences on their mobility. A machine learning model was developed and trained using this dataset with the purpose of predicting future mobility values, while assessing the role that climate played in such predictions. The model employed a Spatio-Temporal Graph Neural Network (STGNN) architecture to capture both potential spatial dependencies between visited locations and temporal patterns in their activity.

The work draws on the Seoul Cozie dataset, which recorded six weeks of GNSS location data from wearable devices of 22 participants in Seoul during autumn 2023. Positions were aggregated into a graph structure with road intersections as nodes and transitions between them as edges. Climate features (temperature, humidity and PM10) were interpolated from over 1,000 weather stations using a Triangulated Irregular Network method and added as dynamic node features. STGNN variants were trained and compared based on whether they included climate node features.

Results show forecasts of node visits with low Mean Square Error of around 0.12 . However, precision and recall values for visited/unvisited node detection are low, peaking at 56.51%, reflecting strong class imbalance in the input. Adding climate attributes produced only minor and inconclusive improvements, in part due to the dataset's short time span. The thesis proposes a reproducible framework linking climate and mobility, while underlining the need for richer datasets and for more flexible model architectures, capable of addressing class imbalances and representing personal mobility datasets.

Contents

Figure List

Table List

Acronyms

ARMA AutoRegressive Moving Average.

CAU Chung-Ang University.

CDR Call Detail Records.

CNN Convolutional Neural Network.

GD Green Distance.

GNN Graph Neural Network.

GNSS Global Navigation Satellite System.

GrCR Green Coverage Ratio.

GRU Gated Recurrent Unit.

GSI Ground Space Index.

HAS Humans as Sensors.

HDBSCAN Hierarchical Density-Based Spatial Clustering of Applications with Noise.

IDW Inverse Distance Weighting.

LSTM Long Short-Term Memory.

MSE Mean Squared Error.

NC Node Count.

NN Nearest Neighbor.

NP Node Presence.

OD Origin–Destination.

OSM Open Street Map.

PET Physiological Equivalent Temperature.

PM10 Particulate Matter with a diameter of 10 micrometers or less.

PM2.5 Particulate Matter with a diameter of 2.5 micrometers or less.

PMV Predicted Mean Vote.

PoI Point of Interest.

RNN Recurrent Neural Network.

S-DoT Seoul Data of Things.

STGNN Spatio-Temporal Graph Neural Network.

STGCN Spatio-Temporal Graph Convolutional Network

SVR Support Vector Regression.

TIN Triangulated Irregular Network.

UTCI Universal Thermal Climate Index.

VAR Vector AutoRegression.

WMBH area-Weighted Mean Building Height.

XGBoost Extreme Gradient Boosting.

1. Introduction

1.1 Context and Motivation

Urban climate and its effects on people's comfort and movement have become central pillars in contemporary city design (Mauree et al., 2019; Ye and Niyogi, 2022). Studies have shown that not considering urban climate can have detrimental effects ranging from safety-related, like the creation of dangerously high heat island effects (Jeon et al., 2023; Gupta et al., 2025), to effectiveness-related, like the underutilization of new developments due to unexpected meteorological conditions (Fallmann and Emeis, 2020), and to cost-related, like expensive retrofitting and long-term operational costs (Erell, 2008; Fallmann and Emeis, 2020). It has therefore become apparent to governments and designers around the world that climate context-aware city planning is pivotal to creating cities that people want to and can safely live in (Hebbert, 2014).

However, this has proven to be a challenging goal to achieve due to the complexities of modern cities and the inherently personal notion of comfort. The former problem particularly impacts approaches aimed at simulating the urban ecosystem. The complexity stems from the sheer amount of high quality and up-to-date datasets required (Alva et al., 2023; Liu et al., 2024; Ignatius et al., 2024), which may not always be available or may be too computationally expensive to analyse concurrently. Additionally, these digital twins must cover a variety of interconnected but distinct systems that make up the overall urban experience, like 3d building morphology, urban material placement, seasonal tree shading, water management systems, traffic affluence, emissions levels, indoor heating usage, meteorological data and more, which removes the possibility of analysing phenomena potentially impacted by excluded systems (Shahat et al., 2021; Xia et al., 2022; Jeddoub et al., 2023; Mazzetto, 2024).

The second problem instead impacts the use of deterministic models in assessing the impact of various climate factors on human behaviour. Models such as the Predicted Mean Vote (PMV) allow for the categorization and calculation of predicted comfort-based levels on a variety of variables (van Hoof, 2008). This has proven effective in creating a baseline approach to understanding human comfort. However, since its results are meant to be generalized values it is unable to consider the individual responses to climate that differ from person to person. Other indices, like the Physiological Equivalent Temperature (PET) and the Universal Thermal Climate Index (UTCI), extend these approaches to outdoor urban conditions by integrating heat balance models and environmental parameters (Honjo, 2009; Bröde et al., 2013; Zhang et al., 2023). While such indices capture meteorological influences more comprehensively, they still operate on assumptions of an "average"

person and do not reflect individual variation. In fact, climate comfort is impacted by many high-level parameters, such as temperature, humidity and wind speed, but is also driven by personal, non-standardizable, low-level ones, such as activity level, body composition, metabolism, eating habits and individual preferences (Jayathissa et al., 2020; Upasani et al., 2024; Zafarmandi and Matzarakis, 2025). These cannot be considered by generalized approaches like the ones mentioned.

Other methods have been built around the idea of the individuality of perception by trying to employ a 'bottom up' approach in opposition to the 'top down' one of the methods presented. What this entails is understanding that personal reactions to climate are challenging to predict based solely on climate values themselves. Instead, people's own behaviour can be recorded and used to model how these changes actually cause people to act and move. This way, conclusions can come directly from human data, including all the participants' individual motivations, preferences and real-world influences, without having to simulate or generalize them (Jayathissa et al., 2020; Ignatius et al., 2024; Gottkehaskamp, 2024; Liu et al., 2024).

This human-centred, 'human-as-sensors' (HAS) method is explored in this research to assess the effect of climate on people's movements at an urban scale. The approach is based on the Soul Cozie dataset, which recorded six weeks of Global Navigation Satellite Systems (GNSS) traces from 22 participants in Seoul during autumn 2023 using wearable devices. It recorded the individual movements of individuals across the city, intrinsically including the personal influences that led to their mobility choices. The use of this dataset aims to directly address the lack of personal influences by the climate comfort models presented before. In order to be able to use the Seoul Cozie dataset to analyse the effects of climate on personal mobility, a machine learning model was developed and trained with the purpose of predicting future mobility values and assessing the role that climate played in such predictions. This, paired with the HAS input would, in theory, help identify the mobility choices most likely affected by climate-related factors. The model was developed using a Spatio-Temporal Graph Neural Network (STGNN)-based method, chosen due to their proven implementations in urban traffic and mobility prediction tasks and their unique ability to detect spatial dpendencies of the input, which are hypothesized to be present in the Seoul Cozie dataset.

1.2 Relevance

The application of STGNNs and Graph Neural Networks (GNN) to urban mobility is extensive but has rarely touched upon personal mobility datasets, which record mobility by following individual people rather than using static frames of reference (Rico et al., 2021; Jiang and Luo, 2022). Many of the currently proposed models are based on data from traffic sensors placed along major transport lines. These models

excel in making congestion and traffic predictions and do not to capture individual mobility, due to the nature of their static networks. Alternatively, Point of Interest (PoI) GNNs use identified important locations, like restaurants, schools and parks, to assess the relevance between them, usually for social media and online navigation platforms. While the granularity of what a node represents is closer to that of personal mobility, the eventual application is far from the context of the urban space.

In comparison, this research aims to create a model that is trained and based on a personal mobility dataset on a single-person scale and includes a variety of visited location types and transport modes taken. Additionally, the use of mobility to assess urban climate comfort in this way is also less explored, but promising. The HAS method has been applied to interior spaces before but has not been thoroughly explored on a larger urban scale with a dataset of this type.

1.3 Thesis Outline

Chapter 2. will define the goals and research questions that drove the thesis.
Chapter 3. will explore papers related to research and current state-of-the-art developments in urban mobility STGNNs and climate comfort.
Chapter 4. will introduce the methodology used in the research, including a detailed description of the Seoul Cozie dataset which defined the project.
Chapter 5. will present the results of the most recent STGNN implementations.
Chapter 6. will discuss these results and their meaning relating to the initial research questions.

2. Research Goals

2.1 Main Research Question

How *do urban climate factors impact people's mobility choices in the urban landscape of Seoul and how can this be captured using a STGNN model?*

The research aims to analyse the relationship between climate and mobility in the urban context of Seoul. To do this, it proposes a STGNN-based method for predicting affluence at several main node locations across the city by leveraging historical position data, enriched with additional climate and urban morphology attributes. The following secondary research questions help guide the research.

2.2 Secondary Research Questions

+ *To what extent can the developed STGNN model be used to predict urban mobility?*
+ *How is the STGNN graph structure constructed and how does this impact the insights the model is able to give on mobility?*
+ *How can a personal mobility dataset, like the Seoul Cozie dataset, be integrated in a STGNN framework and what are its limitations?*

Through these secondary aims of the research, particular attention is put on assessing the plausibility of the proposed STGNN method. They investigate whether the proposed STGNN method produces promising results but also question whether its theoretical set up in this research might have impacted its performance.

2.3 Scope

The primary objective of the research is to develop an initial STGNN model for predicting mobility patterns and evaluating the inclusion of climate parameters on its performance. The implementation and assessment of the model are therefore what most of the research focuses on to answer the main research question.

This research is not meant to be an exhaustive look into this approach, but only an initial exploration of the STGNN case study. Additional model configurations or frameworks that could be implemented to achieve

similar analyses are not experimented or compared with, although these are considered as future work to assess the overall feasibility of the presented method. Additionally, the research is defined and limited by the Seoul Cozie dataset. Evaluation of the model framework on other datasets and contexts is also not in the scope.

3. Related Work

3.1 STGNNs in Mobility Applications

3.1.1 GNN and STGNN methods

The use of GNN models for forecasting applications in the urban mobility context has been widely studied. In recent years, this topic has seen particular interest and development of new approaches. Compared to earlier forecasting methods, like AutoRegressive Moving Average (ARMA) and Vector AutoRegression (VAR) and non-graph-based machine learning methods, like Support Vector Regression (SVR) and Extreme Gradient Boosting (XGBoost), GNNs can handle non-linear relationships in their input data and integrate crucial information from close locations in the learning process (Rico et al., 2021; Jiang and Luo, 2022). Success has also been found with other deep learning methods, such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU), especially regarding sequential forecasting. However, they still lack the integration of spatial relations that a graph structure can provide.

Among GNNs, Spatio-Temporal GNNs (STGNNs) have seen particular use due to the time forecasting goals of most urban mobility-related

PAPER	MODEL	DATASETS	DATASET TYPE	OUTPUT	DF	DS
Yu et al. (2018)	STGCN	BJER4, PeMSD7	Fixed sensors	Future speeds	✓	×
Wu et al. (2019b)	Graph WaveNet	METR-LA, PEMS-BAY	Fixed sensors	Future speeds	✓	~
Shleifer et al. (2019)	Improved GWN	METR-LA, PEMS-BAY	Fixed sensors	Future speeds	✓	~
Kong et al. (2020)	STGAT	METR-LA, PEMS-BAY	Fixed sensors	Future speeds	✓	~
Fang et al. (2021)	STWave	PeMSD3/4/7/8	Fixed sensors	Future flows	✓	~
Ma et al. (2022)	Multi-Modal	TaxiNYC, BikeNYC	OD flows	Future demand	✓	~
Roy et al. (2022)	SST-GNN	PeMSD4/7/8	Fixed sensors	Future speeds	✓	×
Sharma et al. (2023)	STGGAN	PeMSD4/8	Fixed sensors	Future speeds	✓	×
Rossi et al. (2020)	TGN	Wikipedia, Reddit, Twitter	Social media	Future edges	✓	✓
Cini et al. (2023)	GDL	METR-LA, PEMS, CER-E	Multiple types	Multi-step forecasts	✓	✓
Liu et al. (2023)	GLSP	Foursquare NYC, Tokyo	POI check-ins	Next-POI	✓	~
S.K.B et al. (2024)	GT-LSTM	NYC-style multimodal	Mobile traces	Travel times	✓	✓
Wang et al. (2024)	Mode-Aware GNN	Wuhan CDRs	Mobile traces	OD flows	✓	×
Zhang et al. (2025)	MoSTGTN	Tianjin GPS	Mobile traces	Congestion levels	✓	✓

Table 1. Overview of STGNN models and datasets used in related work. DF: Dynamic Features, DS: Dynamic Structure.

models. Their main advantages over regular GNNs are that they can capture temporal relationships along with the spatial ones in the graph by introducing some level of dynamism in the definition of the model structure (Wu et al., 2019b; Kong et al., 2020). Their application has seen an improvement in forecasting accuracy, especially when comparing to RNN-based models (Fang et al., 2021). STGNNs share many of the main drawbacks common to GNNs, where the definition of the graph structure, composed of nodes and edges, greatly impacts the quality of the results. Additionally, the use of a spatial and temporal graph structure can sometimes lead to over-smoothing and low interpretability of the learned dependencies (Jiang and Luo, 2022; Roy et al., 2021; Rico et al., 2021). Despite this, STGNN-based methods have seen great success and have become standard in many mobility forecasting applications (Jiang and Luo, 2022; Jin et al., 2023).

3.1.2 Dataset Diversity

STGNNs are employed in a wide range of methods, both regarding the definition of the graph structure and the model components. Table 1 shows the differences between the models employed by the papers discussed in these sections.

Defining aspects of these models that vary among applications are the input dataset and the predicted output. One of the most common setups is the use of traffic sensor data to create a static graph structure with dynamic node traffic features. Nodes usually represent the sensor locations and links represent distance values or the infrastructure network. Important benchmark datasets, such as METR-LA (traffic speed dataset from detectors on the LA County road network) and PeMS (traffic flow and speed dataset from the CalTrans Performance Measurement System), have become crucial and widely used in STGNN research due to their wide availability and ease in application to a graph structure (Yu et al., 2018; Wu et al., 2019a; Kong et al., 2020; Fang et al., 2021).

Other dataset types often found in the literature are Origin-Destination (OD) flows. They describe specific movements across an urban environment, like bicycle or taxi rides, by indicating the location and time of the start and end instances. The graph structure of OD flows STGNNs usually defines the nodes as locations and the edges as the recorded trips between two specific locations (Ma et al., 2022; Yeghikyan et al., 2020). In a similar way, Point of Interest (PoI) datasets are utilized to represent trips between specified locations. While the application of OD flows is generally more related to the study of transportation patterns, PoI research is closer to spatial social networks, where these datasets usually originate, like in the case of Foursquare NYC and TKY (Liu et al., 2023). In contrast with the traffic sensor-based models mentioned before, the OD flow and PoI ones analyse mobility from a more personal perspective, as they are tied to the movements patterns of specific people.

A less prevalent but still relevant dataset type is that of mobile trace data gathered through GNSS. These rely on mobile devices like phones, smartwatches or GNSS receivers to transmit a continuous record of positions in a certain time period. They can vary from large signalling datasets from mobile providers (Wang et al., 2024) to GNSS traces from a group of vehicles (Zhang et al., 2025; Yeghikyan et al., 2020). Compared to the datasets introduced so far, they are distinguished by the much more irregular patterns of locations that are not predefined and that usually require some sort of semantic grouping. This means that they are rich in information regarding individual and trip-level data and represent human movements more closely. This, however, comes at the cost of graph construction being less straightforward than for fixed sensor or location networks (Rico et al., 2021; Jiang and Luo, 2022).

3.1.3 Architecture Diversity

STGNNs applications vary widely in their architecture definitions and new approaches are quickly evolving. Earlier convolutional approaches, like the Spatio-Temporal Graph Convolutional Network (STGCN) introduced by Yu et al. (2018), replaced more traditional RNNs with temporal Convolutional Neurlal Networks (CNN), combining graph convolutions with gated temporal ones. This saw improvements in the model's efficiency and the creation of a fully convolutional STGNN. Models capable of adaptively calculating the graph adjacency have also seen success, in particular with the introduction of Graph WaveNet by Wu et al. (2019a) and its refinement by Shleifer et al. (2019). Attention mechanisms have also been introduced within STGNNs to more meaningfully weigh neighbours Kong et al. (2020). Additionally, more complex and hybrid architectures have been proposed, which include using new temporal modules, like wavelet decomposition and separation of current and historical patterns Fang et al. (2021); Roy et al. (2021) and combining some of the aforementioned features with other components, like edge features and transformers Sharma et al. (2023); Zhang et al. (2025).

The graph structure of STGNN models is also one of its defining factors. They can be broadly divided into static and dynamic, based on whether its nodes and links can change throughout the model or if they are constant. Most research employs static graphs, as these are best suited for sensor networks, grids and fixed location datasets (Yu et al., 2018; Roy et al., 2021; Sharma et al., 2023). Dynamic graphs aim to capture certain structural changes in a dataset and, so, require a different graph definition from non-dynamic ones. This is partly the case for STGNNs that use an adaptive adjacency, but these are also usually based on a static structure (Wu et al., 2019a; Kong et al., 2020). Fully dynamic graphs remain relatively rare, although progress has been made to allow calculating changes based on memory or similarity components (Rossi et al., 2020; Zhang et al., 2025).

3.2 Human-Centric Modelling

Conventional approaches to climate comfort tend to rely on deterministic indices such as PMV or UTCI, which assume that a set of physical variables can adequately represent the experience of individuals. While such models are useful for establishing baselines, they can reduce a large variety of personal responses into an average value. In practice, comfort is not only shaped by temperature, humidity or wind speed but also by factors that are much harder to standardize. This gap left by traditional approaches has led to studies that place the human subject at the centre of measurement.

Jayathissa et al. (2020) propose this concept as “humans-as-sensors” in the context of buildings. Their work collected repeated subjective comfort feedback from occupants and linked it to concurrent environmental conditions, showing that the frequency of responses allowed the construction of more precise indoor comfort models than standard methods. Upasani et al. (2024) instead emphasized that comfort cannot be separated from the characteristics of the individual and the specific building they inhabit. By explicitly modelling personal and contextual attributes they showed that comfort prediction can improve significantly.

This more individualized logic has also been applied to the outdoor space. Ignatius et al. (2024) embedded wearable data into a digital twin that also included weather records and street-view imagery, linking physiological signals directly to the surrounding built environment and its walkability implications. Liu et al. (2024) developed a human-centric digital twin for Singapore that integrates morphology, meteorology and remote sensing data to estimate outdoor comfort distributions across the city. Alva et al. (2023) extended this perspective by presenting a bottom-up digital twin platform that combines city-scale datasets on energy, mobility and emissions, demonstrating how integrated urban data can support multiple use cases for planning and management.

Walkability and urban mobility have also been studied using this human-scale approach. Jonietz (2016) argued that pedestrian’s different priorities and trajectories cannot be summarized into single general walkability score. Jonietz and Bucher (2017) later introduce a more holistic framework for movement trajectory analysis that considers the spatiotemporal context in addition to just their geometry. Some work on mobility prediction through graph-based methods (Terroso-Sáenz and Muñoz, 2021) and data fusion for synthetic populations (Vo et al., 2025) shows that bottom-up datasets can, in fact, support predictive modelling at multiple scales.

3.3 Climate Comfort and Urban Design

Urban climate research has long been linked to city design. Hebbert (2014) reviewed the evolution of this field and showed how climatology entered planning in the mid-twentieth century, with tools such as the Klimaatlas designed to translate atmospheric analysis into planning guidance. Erell (2008) described the same gap from another angle: although impacts of form, material and density on microclimate are well established, their systematic use in practice has been limited due to technical, organizational and economic constraints.

Several reviews highlight the complexity of connecting climate processes with planning. Mauree et al. (2019) evaluated methods for outdoor comfort, building energy demand and energy systems and noted that these domains are usually studied separately despite strong interdependence. Ye and Niyogi (2022) argued for a convergence of urban climate science and planning practice to address increasing risks from heat, flooding and extreme rainfall. Gupta et al. (2025) focused on thermal hazards and urban heat islands, pointing to “blue-green” infrastructure, optimized morphology and governance as central strategies while also stressing financing and equity barriers.

Case studies show how these issues appear at different scales. Jeon et al. (2023) analysed seasonal land surface temperatures in Seoul and found that vegetation and water reduce heat compared to dense fabric. Peng et al. (2022) used climate walks along a waterfront and showed that shading and water proximity directly shape comfort perception. Liu et al. (2024) and Ignatius et al. (2024) applied digital twins, one combining weather and morphology at city scale and the other fusing wearables, weather and imagery for pedestrian comfort. Fallmann and Emeis (2020) reviewed measures such as greening, reflective materials and biophilic design from a meteorological perspective and stressed that translation into planning requires interdisciplinary dialogue. Despite different methods and scales, these studies converge in showing that thermal comfort depends not only on climate variables but also on vegetation, shading, building form, water and planning decisions.

3.4 Open Challenges and Future Directions

Several surveys and reviews point to open challenges in the use of STGNNs. Among them, the use of dynamic graphs is an area that can see particular growth. Models such as those proposed by Rossi et al. (2020) and Longa et al. (2023) show the difference between snapshot-based and event-based dynamic graphs, noting that most applications still rely on static or semi-dynamic adjacency. Another open problem regards the integration of heterogeneous data sources. Reviews like Rico et al. (2021), Jiang and Luo (2022) and Jin et al. (2023) emphasize that traffic forecasting problems often require additional factors such as weather or calendar effects.

Several recent works have started to address some of these issues. S.K.B et al. (2024) integrates mobility traces with public transport schedules, land use information and weather data in a multi-modal framework. Wang et al. (2024) includes transport mode choice explicitly by building separate car and transit-based adjacency matrices and Zhang et al. (2025) combine taxi and bike GNSS data while also calculating dynamic adjacency at each time step. Despite this progress, studies seem to not consider climate as a key driver of mobility in a spatiotemporal graph model. Climate data is rarely integrated, even though identified as an important external factor (Rico et al., 2021; Jiang and Luo, 2022).

4. Methodology

4.1 Overview

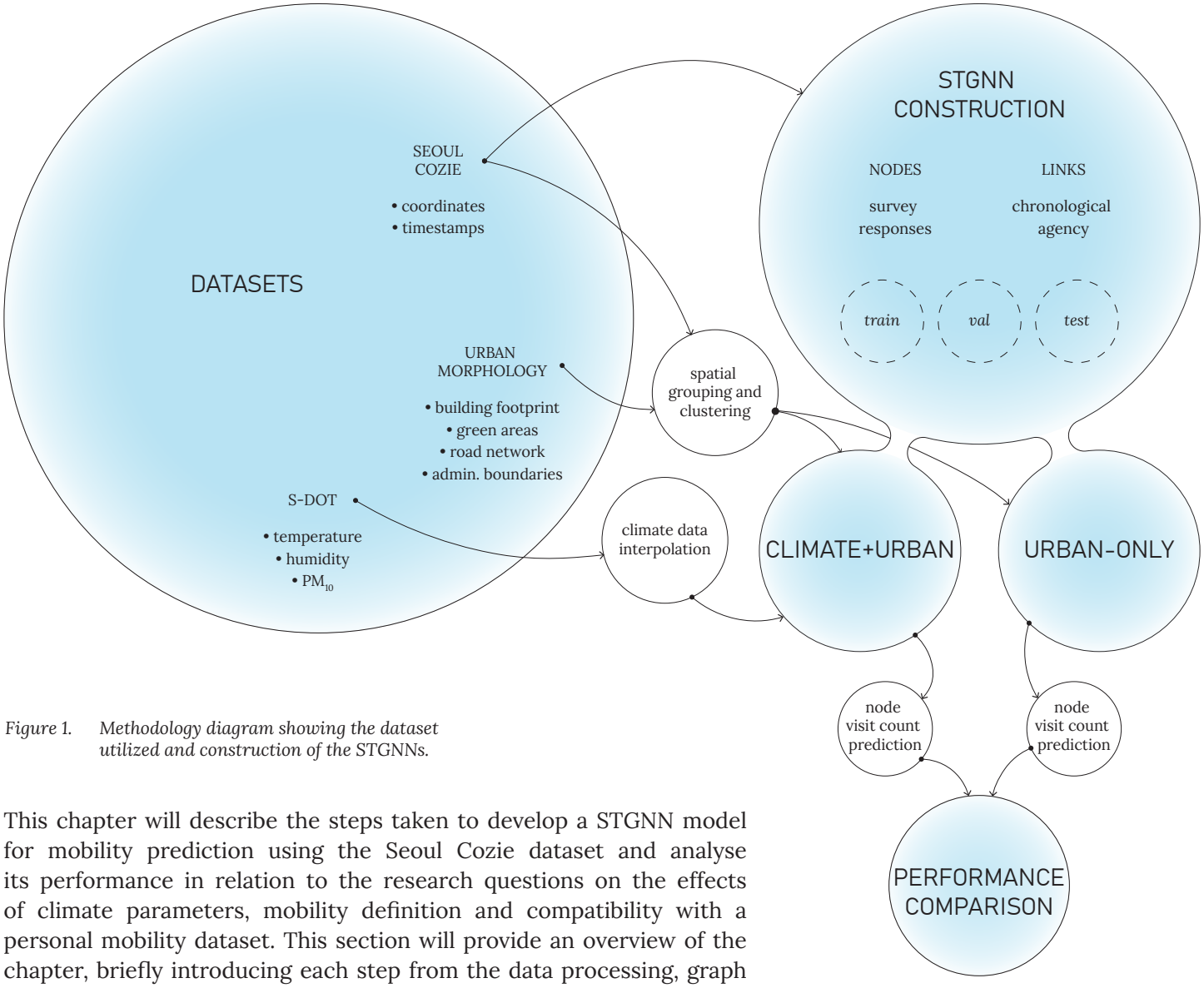


Figure 1. Methodology diagram showing the dataset utilized and construction of the STGNNs.

This chapter will describe the steps taken to develop a STGNN model for mobility prediction using the Seoul Cozie dataset and analyse its performance in relation to the research questions on the effects of climate parameters, mobility definition and compatibility with a personal mobility dataset. This section will provide an overview of the chapter, briefly introducing each step from the data processing, graph construction, architecture definition, evaluation and experimentation, as presented in figure 1.

The first step saw the acquisition, cleaning and processing of the used datasets to prepare them as input to the graph construction (Section 4.2). The Seoul Cozie dataset is cleaned to account for errors in the GNSS traces and timestamp inaccuracies. Readings from the Seoul Data of Things (S-DoT) weather station network are used to interpolate climate parameters across the city during the same time frame of the Seoul Cozie dataset. Urban morphology data in the form as building footprints and green areas are also gathered to enrich the input.

After processing, the datasets are used to create the graph structure (Section 4.3). Different methods of meaningfully group the Seoul Cozie coordinates are tested to eventually define the graph nodes and edges. The graph elements are then enriched with features from the climate and urban datasets and divided into training, validation, and testing time bin.

The architecture of the STGNN model is then defined, as it takes the past snapshots of the graph to predict its next-bin state (Section 4.4). The model is trained to output the node visit counts. Once trained, the model is evaluated and experimented on to answer the initial research questions (Section 4.5). Two almost identical models, Climate+Urban and Urban-only, were trained with the same architecture and graph structure above that differed on the inclusion of climate parameters to assess their influence on model performance. The evaluation step saw a quantitative analysis of performance metrics and a qualitative look at implementation of the models in real case scenarios. From these insights the results and conclusions of the research are drawn.

4.2 Datasets

4.2.1 Seoul Cozie Dataset

The dataset utilized to develop and evaluate the proposed STGNN model is the Seoul Cozie dataset gathered by Mosteiro-Romero et al. (2024) during their experiment on the emergent role of district-scale and occupant-scale data in urban environments. Conducted at Chung-Ang University (CAU) in Seoul, South Korea, the experiment aimed to collect subjective from individuals in a real-world setting. The study saw 22 university students, aged 20 to 31, agree to have their location and physiological data recorded for a period of 6 weeks (October 4 to November 13, 2023). This was done using Apple Watches that were loaned or previously owned by the participants. Through the iOS application Cozie, GNSS coordinates and body measurements, such as resting heart rate and wrist temperature, were collected (Tartarini et al., 2023). Additionally, micro surveys would be sent out on an hourly basis to ask participants about their past activities, current location and temperature comfort. Each submission of these surveys would also record their GNSS coordinates.

The final gathered dataset was composed of 54,322 records with individual and valid recorded coordinates. Due to the number of participants and their similar profiles, the recorded positions are mostly concentrated in the vicinity of CAU buildings (fig. 2) but also extend to other popular areas of the city and even beyond its boundaries, capturing trips to Cheongju, Gangneung, Daegu, Jeju Island and more (fig. 3).

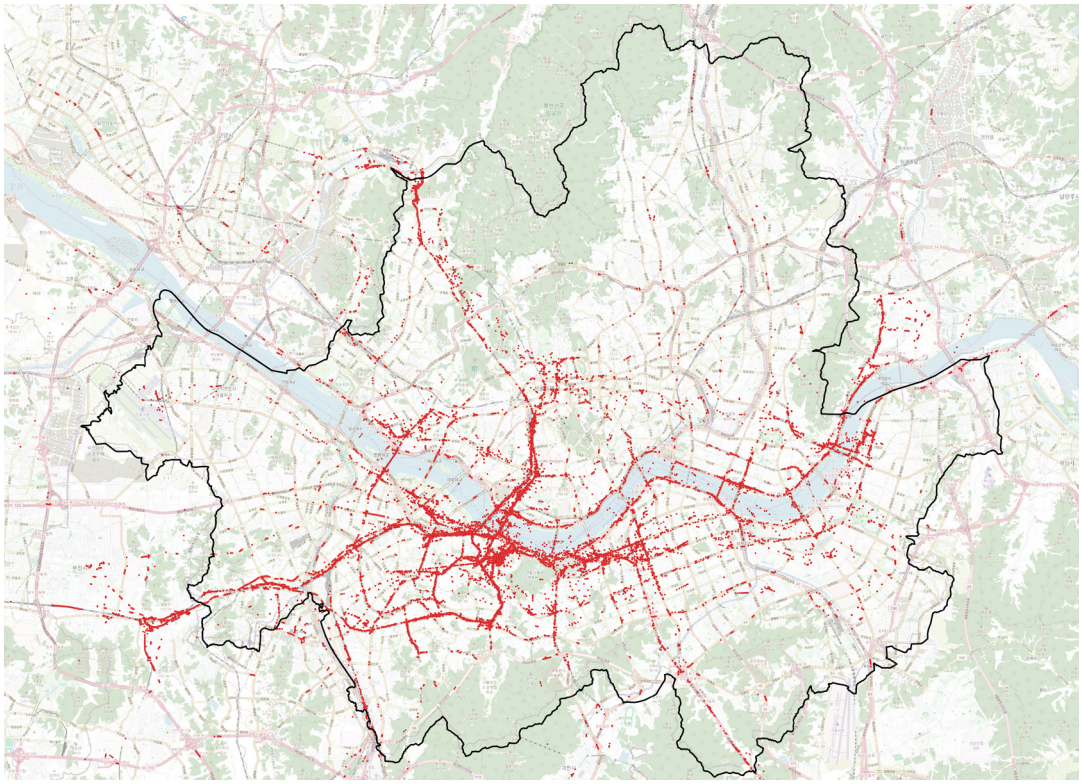


Figure 2. Recorded positions of the Cozie dataset in Seoul.

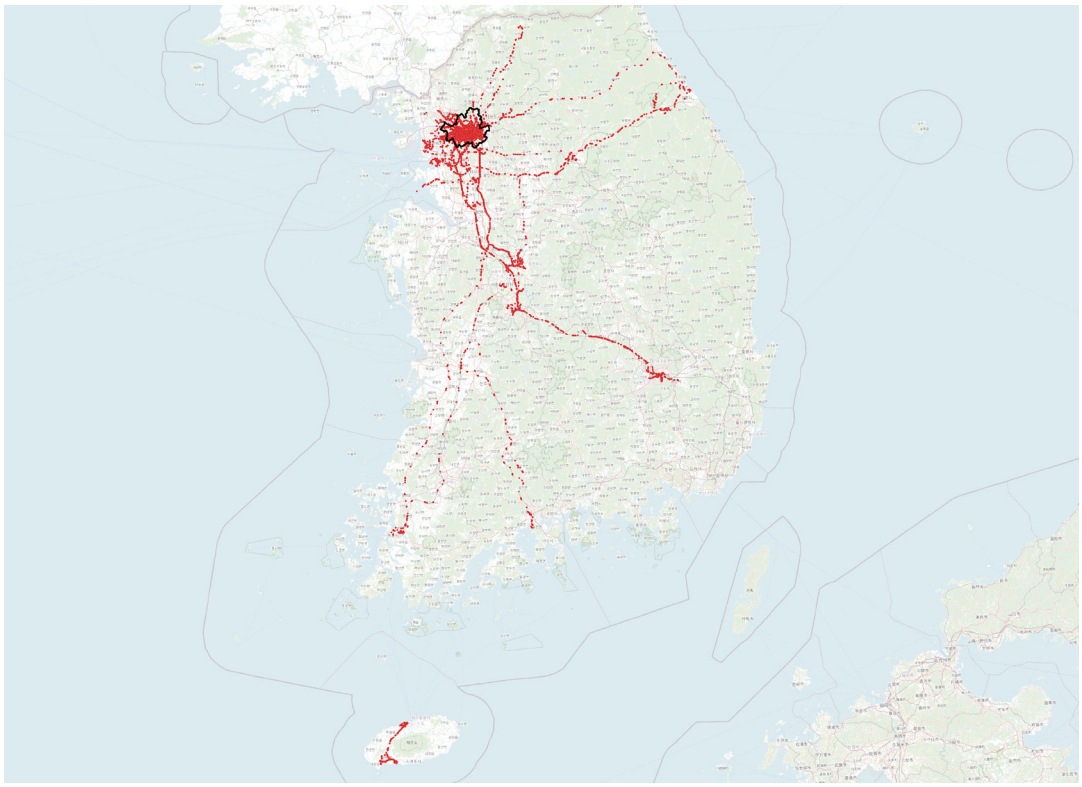


Figure 3. Recorded positions of the Cozie dataset across South Korea.

Due to the dataset not being explicitly gathered with the purpose of STGNN development, certain preprocessing steps had to be performed. Firstly, several features and records that were not relevant to the research were omitted, including all physiological readings and records that were not geolocated or had invalid (0.0, 0.0) coordinates. Records that were outside of the administrative boundaries of Seoul were also discarded to focus on the urban scale.

Additional cleaning was done to tackle irregularities found in the coordinate records. Occasional big “jumps” would appear caused by a distant coordinate interrupting a single participant’s path (fig. 4). To highlight these and other suspicious nodes, a speed value was calculated based on the distance between consecutive coordinate records.

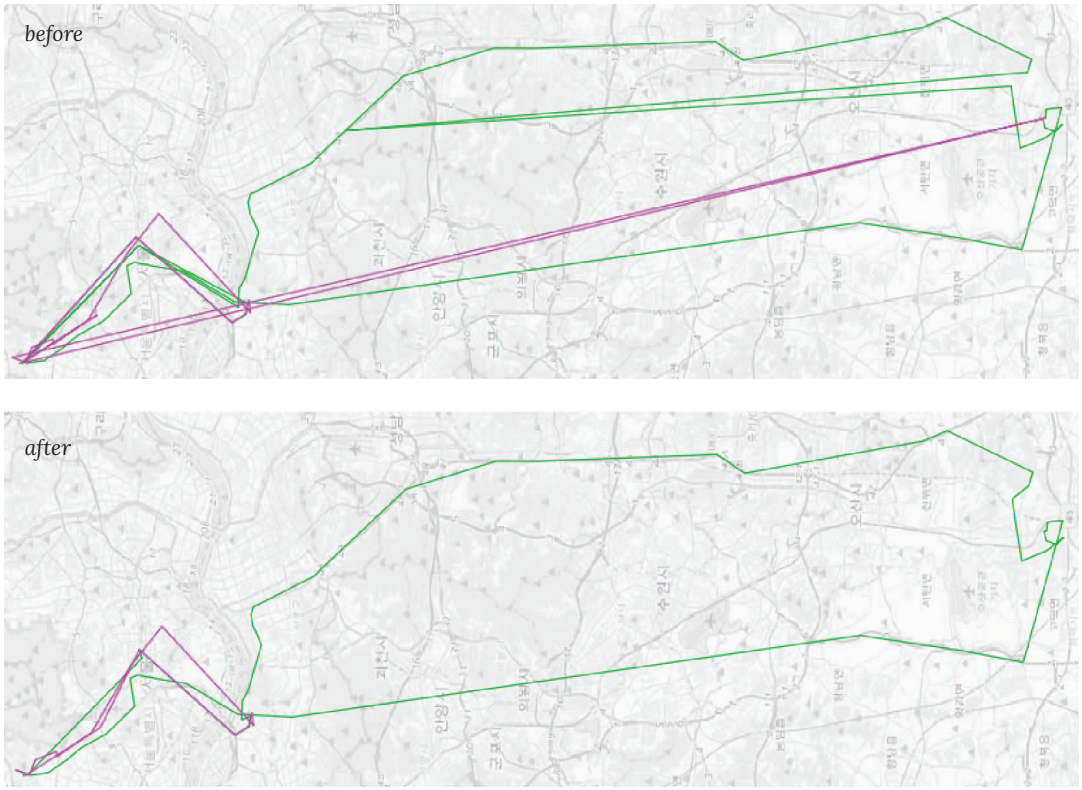


Figure 4. ”Jumps” error before and after cleaning.

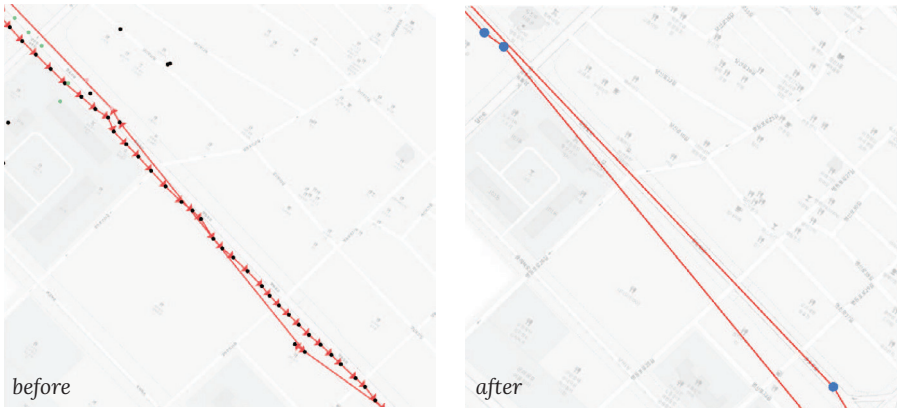


Figure 5. Redundant coordinate error before and after cleaning.

Coordinates with unfeasibly high-speed values (above 400 km/h) were selected. This allowed to identify different groups of errors and address them individually.
A large portion of errors came from a mismatch between the Cozie timestamps and coordinates, due to them both differentiating between readings triggered by completing a survey and ones triggered by opening the Cozie app or by background tasks of the device. This was solved by creating unified coordinate and timestamp fields that took into account each type of record.

Another group was identified as redundant coordinates that recorded the participant's position multiple times over less than one second. This extremely short time distance between records, supposedly caused by a device or application error, caused the speed value to balloon. While not necessarily incorrect data points, their temporal accuracy was widely out of scale with the rest of the dataset, which instead saw intervals between coordinates in the 20 minutes to 1 hour range. Therefore, coordinates that were recorded less than five minutes after the previous one were dropped (fig. 5).

The data cleaning also addressed problems caused by overlapping consecutive positions, which were usually followed by a third position with an impossibly high speed value. The identical values of the coordinates implied some error in the data writing process, again probably due to a temporary malfunction with the participant's device or application which caused a previous coordinate to overwrite the current one. Since the dataset stored coordinates as latitude and longitude with up to 15 decimal places, it was highly unlikely to produce the same value for even the same location, making these errors easy to identify. The inaccurate larger distance between records caused by this paired with their accurate timestamps, led to the high speed values. This was resolved by omitting records that had identical coordinate values to the previous record.

The dataset cleaning therefore tackled most time/coordinate inconsistencies, "jumps", redundant positions and overlapping coordinates to obtain a final dataset most closely resembling the participants' everyday trajectories. Due to the dynamic and inconsistent nature of the dataset, it is possible that additional processing could have been applied. This, however, was deemed an acceptable level for the purposes and scope of the research.

Finally, the experiment was performed on weekdays (Monday to Friday) from 09:00 to 18:00, and participants were not required to respond to surveys or wear the devices outside these hours. For this reason, the number of records on weekends and between 24:00 and 08:00 see a drastic fall in the number of records. Interestingly, the ranges of 18:00 to 24:00 and 08:00 to 09:00 do not see this same fall, despite being outside the official experiment duration. Therefore, the records were also filtered to only include ones on weekdays and from 08:00 to 24:00.

4.2.2 S-DoT Dataset

FEATURE	UNIT	DESCRIPTION	U?	C?	NR?	G?
Temperature	C	Air temperature near the station	✓	✓	✓	✓
Humidity	%	Relative humidity of the air	✓	✓	✓	✓
PM_{10}	$\mu g/m^3$	Concentration of particles ≤ 10 μm in air	✓	✓		✓
$PM_{2.5}$	$\mu g/m^3$	Concentration of particles ≤ 2.5 μm in air		✓		✓
Wind speed	m/s	Air movement past the sensor			✓	✓
Illuminance	lux	Level of visible light at ground		✓	✓	
UV	UV	Intensity of UV sun radiation		✓	✓	
Noise	dB	Ambient sound level at the site		✓	✓	
Vibration	mm/s	Vibration velocity along 3 axes		✓	✓	
NO_2	ppm	Nitrogen dioxide concentration			✓	
SO_2	ppm	Sulfur dioxide concentration			✓	
NH_3	ppm	Ammonia concentration			✓	
H_2S	ppm	Hydrogen sulfide concentration			✓	
O_3	ppm	Ozone concentration			✓	

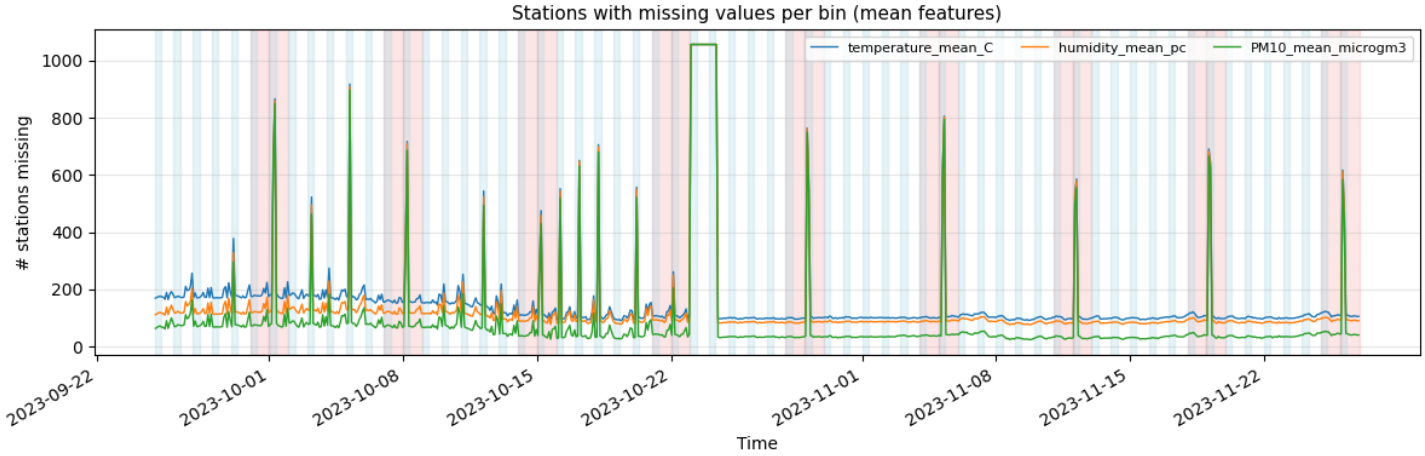
Table 2. Features captured by S-DoT sensors and the qualities that influenced them being included as node features. U: used as node features, C: is the data (mostly) complete?, NR: is the feature not directly related with another?, G: can it be generalized across space?

To enrich the nodes of the STGNN that would take into account climate, additional features were integrated from the S-DoT network. S-DoT comprises over 1000 weather stations distributed across Seoul. They record temperatures, humidity, air quality, wind speed, illumination, UV radiation, vibration and noise levels at two-minute intervals which are then averaged into hourly values (Seoul Metropolitan Government, 2021; Song et al., 2023). The stations are installed at 3-4m height on poles and building walls predominantly in dense urban areas (87%) as well as near rivers (9%) and in mountain areas and parks (4%), giving a spatial resolution of around 0.75km (Song et al., 2023). The data can be on occasion subject to missing records due to overheated sensors, power failures or telecom failures (Kim et al., 2023).

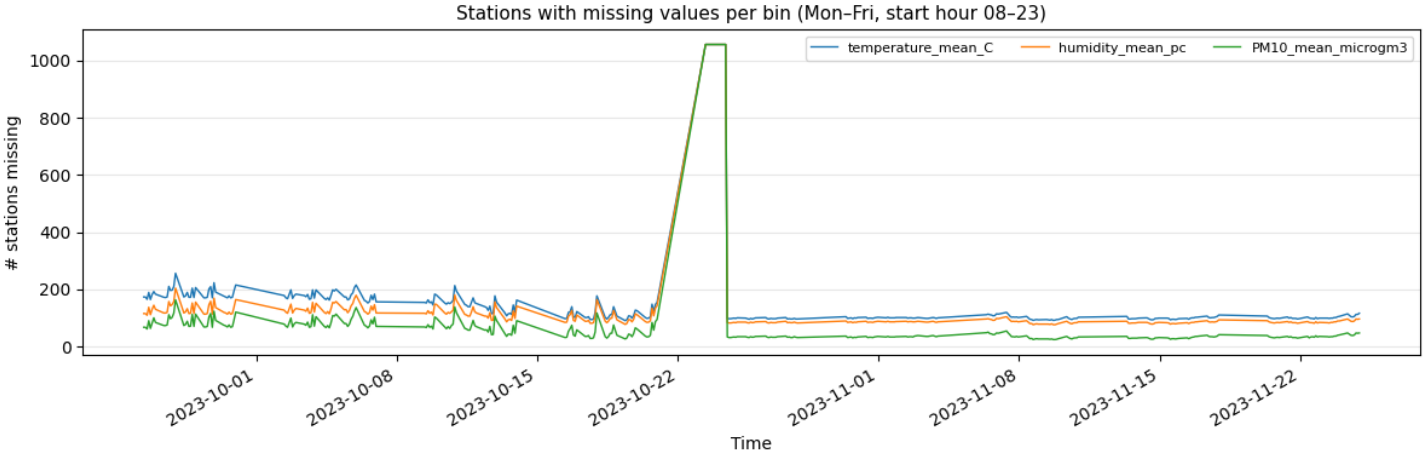
S-DoT has been used to clearly detect urban heat island patterns, consistently showing one to three degrees of difference between the mountain and downtown areas. They also function as a public data source for the Smart Seoul Map to guide heat stress prevention and air quality information (Seoul Metropolitan Government, 2021).

Out of the 16 measured parameters, 3 were chosen to be implemented in the STGNN: temperature, humidity and PM10 (coarse particulate matter with a diameter of 10 micrometers or less). Other parameters

were omitted due to having too many missing values, being too reliant on the positioning of the station and not being generalized to the area around it or being too correlated with another field (specifically for PM2.5). Table 2 gives a full overview of the captured S-DoT parameters and the reasoning for their omission.



a. Invalid values per feature. Weekends in red; 24:00–08:00 in blue.



b. After excluding weekends and 24:00–08:00.

Figure 6. S-DoT station invalid-value counts before and after filtering.

To assign to each node in the STGNN graph structure a temperature, humidity and PM10 value at each time bin, some cleaning needed to be done on the S-DoT dataset that would account for occasional null values. Then, a triangular irregular network (TIN) method was applied to interpolate climate features at each node's coordinate from the S-DoT locations.

Although the chosen S-DoT features were chosen in part due to their high data availability, they still encountered some data loss that needed to be dealt with before interpolation. Many stations

saw a spike in null values during late hours, specifically between 24:00 and 08:00 (fig. 6a). Another noticeable rise in null values was apparent from 24:00 of October 23 to 08:00 of October 24, 2023. During this period, none of the S-DoT stations recorded any data (fig. 6b). Both these patterns affected a large number of stations and might have been caused by some known issues cited by Kim et al. (2023), such as power or telecom outages. The 24:00 to 08:00 time frame was already filtered out from the Seoul Cozie dataset, due to its own low data points in that period, so no additional cleaning was required. However, the 32-hour time period between October 23 and 24 needed to be addressed, and so it had to be filtered out to prevent multiple time-bins having no climate data. The omission of these records was also applied to the STGNN that did not use climate features to make sure it would not affect their comparison.

After addressing these most noticeable and largest missing data records, some station-specific missing values still existed throughout the dataset, possibly caused by local issues of overheating or temporary breakages. A single approach was utilized for them: if the station had a valid reading from at most 90 minutes before or after the missing value timestamp, a new value would be generated by performing a simple linear interpolation between the real values. Otherwise, the null value would remain and be dealt with during the interpolation phase.

The chosen interpolation method was a TIN-based one. TIN interpolation has been proven to be time efficient while being able to account for irregularly placed data points, in contrast to other methods that usually excel only in one of the two (Ledoux et al., 2024). This was ideal for the S-DoT dataset due to the unevenly distributed locations of the weather stations, which often avoided parks, mountains and the Han River crossing the city.

The interpolation worked by firstly creating a Delaunay triangulation of the S-DoT network using the stations as the vertices and creating triangles that were as equilateral as possible. Then, at each time bin, a node's location within one of these triangles and its vertices' weather features was used to calculate the node's interpolated temperature, humidity and PM10 values (fig. 7). This was done by transforming the node's coordinates into barycentric coordinates, which represent its position in relation to the three stations. These barycentric coordinates then acted as weights to the weather values of the stations to calculate the interpolated values, meaning that a node would have more similar values to the closest stations in its triangle (fig. 8).

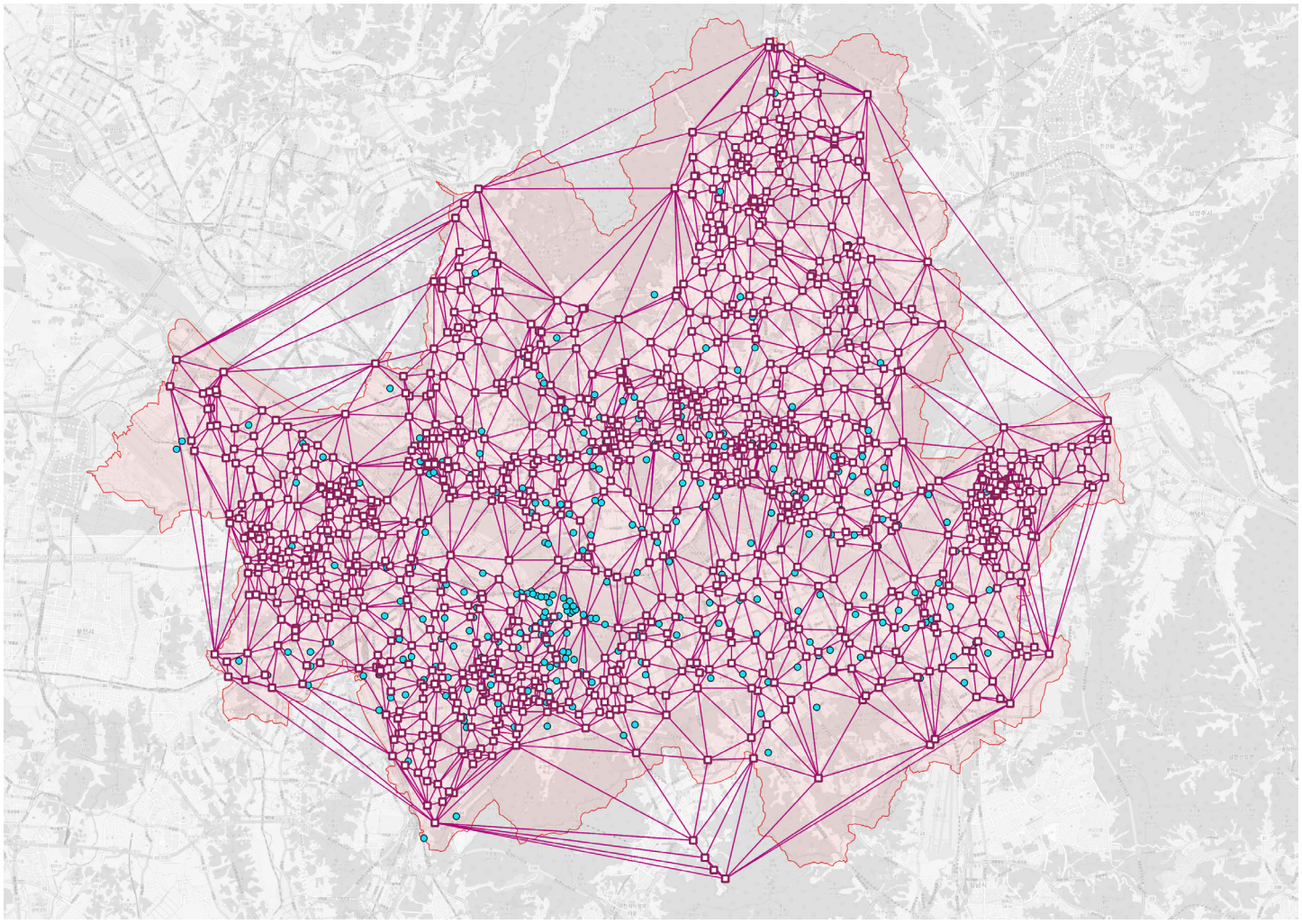


Figure 7. *Triangulation of the weather stations (purple) and the node positions (cyan) overlaid on the Seoul extents.*

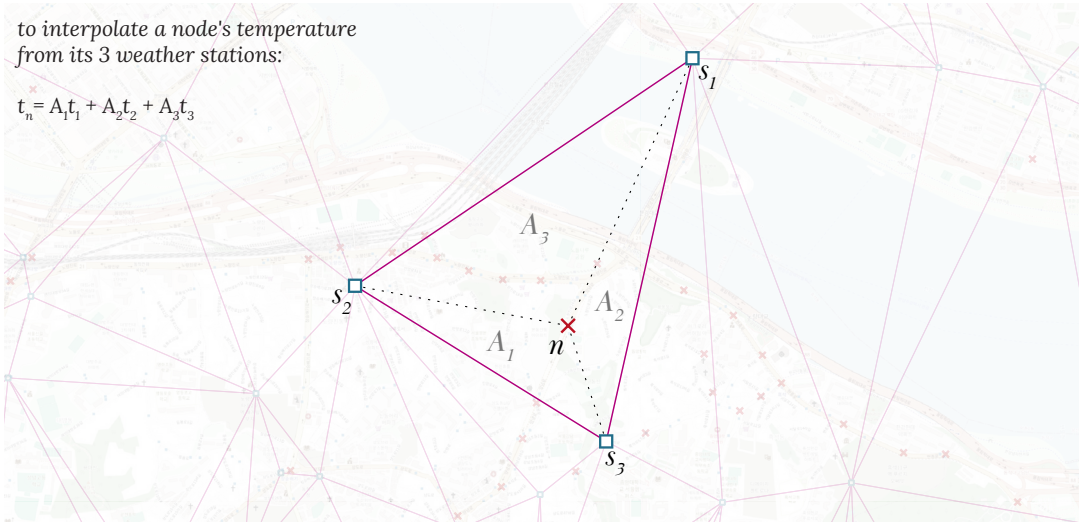


Figure 8. *Barycentric coordinates of a triangle. The weight wi of a point pi can be thought of as the area Ai of its corresponding triangle created by the point x and the other two points pj and pk (Ledoux et al., 2024).*

This general approach was applied to most nodes across time-bins, but some exceptional cases needed to be dealt with individually. Firstly, nodes that were outside the convex hull of the triangulation, would not have been inside any triangle and so could not have barycentric coordinates and interpolated values according to the TIN method. The simple nearest neighbour (NN) method was used that would take the value of the closest weather station. This same method was also applied in the cases of triangles with at least one station having a remaining null value from the initial S-DoT cleaning. Since the null value would not allow for valid barycentric coordinates, this was a simple fix to still make sure that each node would have valid weather features, without having to rebuild a triangulation for every time bin with invalid station readings.

Table 3 shows other methods that were considered and their reason for ultimately not being chosen: nearest neighbour (NN) could be too simplistic and give the same value to differently placed nodes, inverse distance weighting (IDW) would have been better at estimating different values for the nodes but would have struggled in areas with less concentrated stations, C1 or C2 smoothed TIN methods would have been unnecessary and overly complicated for the simple weather dataset and costed in computational efficiency, Kriging could have provided more realistic values but would have required additional contextual datasets and development time which went outside the scope of the research.

METHOD	PROS	CONS
TIN	Time efficient; handles irregular station placement well; straightforward to implement; flexible fallback to NN for nodes outside convex hull or triangles with null station values	Cannot interpolate outside convex hull without fallback; barycentric interpolation only uses three nearest stations (may miss broader patterns)
NN	Very simple; fast to compute; always produces a value	Too simplistic, assigns identical values to many nodes; loses spatial granularity
IDW	Produces smoother variation than NN; relatively easy to implement	Sensitive to irregular or sparse station placement (e.g. gaps along rivers or linear distributions); may be unreliable in poorly covered areas
C1/C2 TIN	Generates smoother surfaces; better continuity across triangles	Computationally more expensive; unnecessary complexity for relatively simple weather datasets
Kriging	Statistically rigorous; can incorporate spatial autocorrelation; potentially provides most realistic estimates	Requires more contextual data; high computational cost; timeconsuming and outside scope of this research

Table 3. Comparison of interpolation methods considered for assigning S- DoT climate features to STGNN nodes.

4.2.3 Urban Morphology Datasets

Datasets on the morphology and features of Seoul's urban environment were also used for the purpose of defining the graph structure and enriching the model features. Building footprint and heights were acquired from the Municipality of Seoul and required cleaning to ensure valid and non-overlapping geometries. Green areas, including city parks, riversides, and forests, were acquired from Open Street Map (OSM). Cleaning again had to account for overlapping geometries, both within the dataset and with the building footprints. When an overlap between the two datasets occurred, the building geometry took precedence, and the shared area was excluded from the respective green area. OSM Road network lines were also used and cleaned to merge double lines representing double-way lanes. Finally, the administrative boundaries of Seoul, its municipalities, and its sub-municipalities were also acquired from the Municipality of Seoul.

4.3 Graph Structure

4.3.1 Mobility as a Graph

The STGNNs were built using the same graph structure. Nodes represented locations across Seoul gathered from semantically grouping the Seoul Cozie coordinates, as explained in section 4.3.2. Edges connected pairs of node locations that were visited in succession by at least one participant. Constructing this graph required translating the raw Cozie GNSS traces into a form that could capture meaningful urban positions since the coordinates collected by the devices were irregular and not directly tied to distinct places in the city. A single GNSS point cannot by itself represent a location in the urban sense, as it just represents a coordinate point with no semantic value. To address this, a method was needed to group multiple records into nodes that could stand for semantically comparable locations. This was the central step in defining the graph: deciding at what spatial scale positions would be merged and how edges would then connect these nodes based on observed participant movements.

4.3.2 Node Grouping

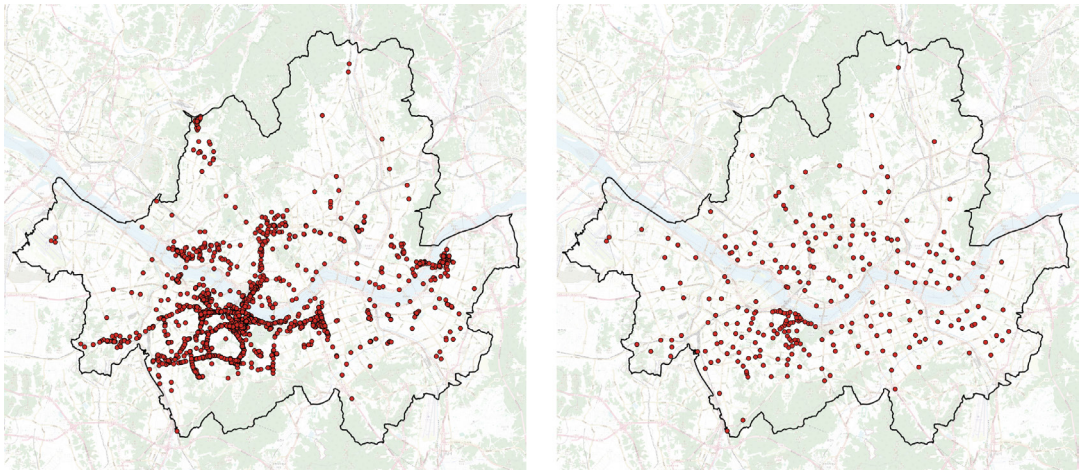


Figure 9. Nodes from morphological grouping (left) and from intersections grouping (right)

A first grouping strategy explored was based on urban morphology. The aim was to preserve as much detail as possible, letting nodes represent the actual places participants visited. Several datasets were combined for this: the Cozie survey responses, the building footprint dataset from the Seoul Metropolitan Government and the OSM green areas layer. Indoor responses were matched to the nearest building, outdoor responses inside polygons of green areas were grouped as park nodes and the remaining outdoors points were clustered using the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) algorithm with a minimum cluster size of 5 and a minimum sample size of 3. (fig. 9). In principle, this would allow nodes to correspond to categories like specific buildings, parks, or clustered street locations. In practice, however, most recorded positions did not have survey responses attached indicating whether the participant was inside a building, which meant many indoor coordinates were incorrectly treated as outdoors. The varying horizontal accuracy of GNSS coordinates contributed to this as locations close to buildings could not be reliably assigned as indoors. Furthermore, the method produced very fine-grained results near the university, where data was dense, but quickly broke down in the rest of the city. Locations that were visited once or by a single participant created nodes with extremely low activity, resulting in strong class imbalance between frequently and rarely visited nodes. This approach was tested but discarded after it produced poor model performance.

A second method was then developed to address these issues by lowering the spatial granularity of the nodes. Instead of trying to represent nodes as exact places, each coordinate was assigned to its closest road intersection derived from the OSM street network. In this way, one node represented one intersection, which positioned the graph more at the scale of an urban transport network, while still utilizing the personal mobility records. The model no longer tried to capture

whether a participant was inside a specific building or in a certain park, but rather which intersection they were closest to. The idea was similar in spirit to what Ma (2022) described, although in this case the grouping logic itself was changed rather than merging nodes afterwards. At the edges of the city, where visits were rare and often only by one or two participants, even this scale was still too fine and resulted in many commonly unvisited nodes. These cases were merged further into single nodes representing whole municipalities or sub-municipalities, depending on their number of records and unique visitors (fig. 9). This method reduced the number of nodes, enriched the data available per node and lowered the class imbalance. The imbalance did not disappear entirely, as central intersections remained much more visited than peripheral ones, but it was less severe than in the morphology-based approach. This second grouping definition was therefore adopted as the graph structure used in the rest of the project.

4.3.3 Node and Edge Features

Once the Cozie coordinates were grouped into nodes and the edges were constructed by connecting these nodes based on the participant's movements, additional information was added to both graph components.

Nodes were enriched with both static features, that remain the same throughout the time bins and dynamic ones, which would change based on the timebin. These were the node coordinates (static), urban morphology parameters (static) and a visit count of the number of times participants passed through the node during the time bin duration (dynamic). Dynamic climate features were also added to the nodes of the climate STGNN as described in section 4.2.2. Each node's urban features were calculated using a 100-meter buffer around its coordinates and incorporating building footprint and green area datasets (fig. 10). The following metrics were chosen due to their ability to generally describe and differentiate different urban areas (Maiullari 2023): Ground Space Index (GSI), area-Weighted Mean, Building Height (WMHB), Green Coverage Ratio (GrCR), GD (Green Distance). Table 4 details how each of them are calculated.

Edges were enriched with only dynamic features, as the edges themselves were also dynamic. These were the number of movements between the two nodes connected by the edges during the time bin duration and the average transport mode derived from the average speed of such movements. The transport mode was categorized based on the average collective speed of movements on the edge. A 7.5 m/s threshold on this value was used to distinguish between slower personal mobility, like walking running, and biking, from faster motorised mobility, like cars, metro, and trains.

FEATURE		DESCRIPTION
GSI	Ground Space Index	*Footprint area / buffer area*
WMBH	area-Weighted Mean Building Height	*∑ building height • footprint / ∑ footprint*
GrCD	Green Coverage Ratio	*green area / buffer area*
GD	Green Distance	*distance to closest green area*

Table 4. Description of node urban features.

Figure 10. Examples of nodes, their buffers and urban features calculated based on building and green morphology.

4.3.4 Binning

The graph structure described was used to create multiple graphs bins that would represent different time instances across the Seoul Cozie dataset duration. Bins were defined as two-hour periods, that

would be able to consistently represent different parts of the day. Each bin's graph was composed of all the nodes, with their static features remaining unchanged and their dynamic ones being aggregated to the two-hour period. Node count visits were summed, and climate features were averaged. Edges between nodes were included in a time bin only when at least one person travelled between them. Edge counts were averaged, and transport was calculated using the speed values of only the movements occurring in the bin duration.

Following this, the bins were divided into training, testing and validation groups. Entire days, and all two-hour bins inside them, were assigned to each group following a four (training) - two (validation) - three (training) - one (testing) recurring pattern, which was chosen to have an even distribution of weekdays in all groups across the entire dataset. This was done to limit the effects of patterns that might emerge only during specific weekdays or at different time periods during the Seoul Cozie dataset time span and to keep a train/validation/test split of 70%/20%/10% (fig. 11).

It was also enforced that bins would be able to act as prediction targets only for their own binning group to avoid data leakage between them. Moreover, any bins used as input for a prediction could be reused as a target only for inputs in its same group. This meant that when chronologically switching between bin groups a gap would remain where the bins were only used are inputs for the next bins of the same type, but not as targets for the previous bin group.

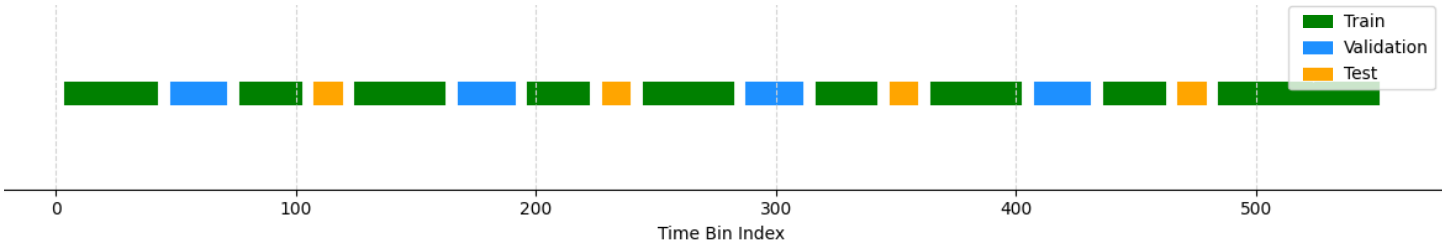


Figure 11. Distribution of binning groups across the total time frame.

4.4 STGNN Architecture and Training

4.4.1 Architecture Overview

The STGNN model was designed to take as input graph instances from five previous time bins, representing a ten-hour period, and outputting predictions for visit counts for all nodes at the next time bin. The utilized architecture is presented in this section, further detailed in the subsections below, and shown in figure 12.

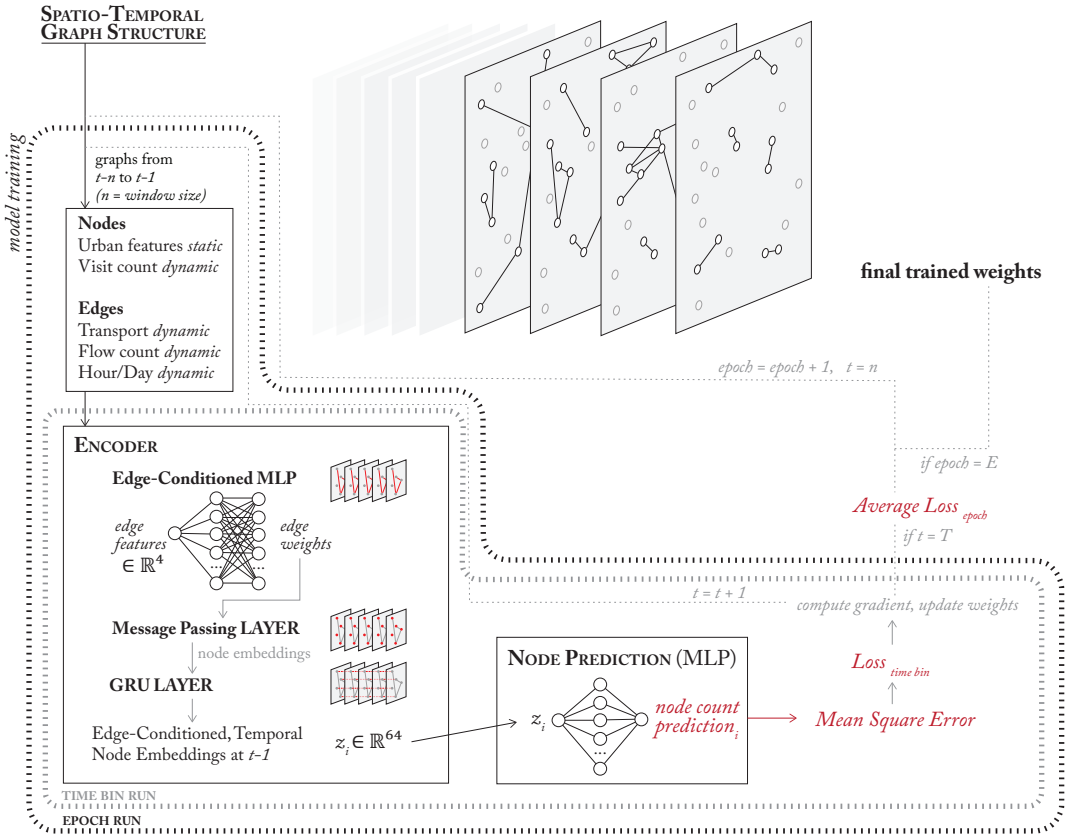


Figure 12. Diagram explaining model architecture and components.

The input graph bins pass first through a decoder. Its first two layers use the node features (visits, climate, morphology) and the dynamic edges (count, average transport mode) at each past time bin. An edge-conditioned MLP encodes the edge features into weights for each one. These are then used in a message passing layer along with node features to exchange information between neighbours and calculate embeddings for each node. Then, a Gated Recurrent Unit (GRU) layer is applied at each node to retain important information from its embedding in past bins and updates the embeddings once again.

The decoder takes the final learned node embeddings from the encoder and maps each one to a single predicted node visit count value. The model is iteratively trained using Mean Square Error (MSE) loss over all train bins and over 60 epochs.

4.4.2 Input

The input consisted of a sliding window of five consecutive time bins, $(t - W, ..., t - 1)$ with $W = 5$, each represented as a graph with nodes, edges and their attributes. Predictions were then made for the next unseen bin t. This setting allowed the network to condition its forecasts not only on the spatial structure of the graph but also on the short-term temporal dynamics of mobility.

Each node contained both static and dynamic features. The static part included the normalized coordinates of the node and, when available, the surrounding morphology features. These attributes did not vary across time. The dynamic part was recalculated for each bin and included the number of visits recorded at the node during that interval.

Each edge carried attributes describing the temporal context (hour and day, normalized), the number of movements recorded and the average speed of these movements. The speed values were normalized across the dataset, with a threshold of 7.5 m/s separating slower personal mobility from faster motorised travel.

Formally, a graph at time k can be written as

$$G_k = \left(V,\ E_k, X_k, E_k^f\right)$$

where V is the set of nodes, E_k the set of edges observed in bin k, X_k the node features and E^f_k the edge features.

4.4.3 Encoder

The encoder combined edge-conditioned spatial message passing with a re- current temporal update. For each bin in the input window, node embed- dings were first updated with an edge conditioned convolution (NNConv). The weight matrices used in the message passing were produced by a small neural network applied to the edge attributes, making the aggregation dependent on the type of connection.

For a node i, the update at step k can be expressed as

$$m_i^{(k)} = \sum_{j \epsilon N(i)} \phi\left(e_{ij}^{(k)}\right) x_j^{(k)}$$

where $e^{(k)}_{ij}$ are the features of edge (i, j) at time k, ϕ is the edge network mapping them to a weight matrix and $x^{(k)}_j$ is the feature vector of neighbour j.

The outputs of the convolution were then passed through a GRU which maintained a hidden state across the five time steps. This allowed the encoder to integrate the temporal evolution of each node. The recurrence can be written as

$$h_i^{(k)} = GRU\left(m_i^{(k)}, h_i^{(k1)}\right)$$

with $h^{(0)}_i$ initialized to zero. After processing the full sequence, the final embedding $h^{(W)}_i$ represented the node at the end of the input window.

4.4.4 Decoder

From the node embeddings, the decoder produced predictions for the next bin. It was conditioned on the temporal encoding of the target bin, given by the normalized hour and day values Tt. To predict the number of visits per i node, for node i, the embedding $h^{(W)}_i$ was concatenated with Tt and passed through a regression network to give the predicted visit count $\widehat{y}_i$:

$$\widehat{y}_i = f_{node}\left(\left[h_i^{(w)}, \tau t\right]\right)$$

The dataset was split along entire days to prevent overlap between training and evaluation windows. Optimization used an initial learning rate of 0.01, which reduced whenever the validation loss plateaued. The objective used mean squared loss to train the node visits.

4.5 Experiments and Evaluation

4.5.1 Model Configurations

To isolate the effect of contextual information on predictive performance, a total of four model configurations were trained under the same experimental method. All variants share the same architecture detailed in section 4.4 and they use the same graph structure, optimizer, learning rate and train-test-validation split. The difference between them only concerns the node features utilised.

The Climate+Urban model is the main configuration that incorporates both climate features interpolated from the S-DoT dataset (temperature, humidity and PM10) and urban features from the morphological datasets (GSI, WMBH, GrCR, and GD) aligned to the 2-hour time bins. The Urban-only variant removes the climate features, while keeping the urban morphology ones. These two configurations were the ones in the initial scope of the research, as the latter was meant to count as the baseline for the former. However, a Climate-only and a Baseline variant were also introduced to isolate the results from the urban features, if necessary. The former omits the urban features but retains the climate ones, while the latter removes both.

	CLIMATE+URBAN	CLIMATE-ONLY	URBAN-ONLY	BASELINE
Climate node features	✓	✓	×	×
Urban node features	✓	×	✓	×

Table 5. Presence of climate and urban features across model configurations.

4.5.2 Metrics

The evaluation of the models focuses on the next-bin node count regression trast and a derived node presence value. For counts, MSE between predicted and observed counts is reported (table 6). Presence is instead a value derived from the predicted counts that informs on whether a node is visited by at least one person. It is defined from the ground truth as visited when the node count is at least one, while for the model outputs a threshold of 0.30 is used. We use this binary inactive/active classification to describe Presence through the metrics of Accuracy, Precision, Recall and F1-score.

METRIC	DESCRIPTION	RANGE	NOTES
NC MSE	Mean squared error between predicted and observed node visit counts per time bin.	≥ 0	Sensitive to scale; dominated by many zeros.
NP Accuracy	Share of correctly classified nodes (active/inactive) after thresholding predicted counts.	[0, 1]	Can be inflated by class imbalance. Threshold = 0.3.
NP Precision	TP/(TP + FP) for the *active* class; proportion of predicted-active nodes that were truly active.	[0, 1]	Reflects false positive control. Threshold = 0.3.
NP Recall	TP/(TP + FN) for the active class; proportion of truly active nodes detected.	[0, 1]	Reflects sensitivity. Threshold = 0.3.
NP F1 Score	Harmonic mean of precision and recall (2 PR/(P + R)) for the active class.	[0, 1]	Balances precision/recall. Threshold = 0.3.
Pearson Corr.	Pearson correlation coefficient between predicted and observed node counts.	[−1, 1]	Can appear high when both series are mostly zeros; interpret with care.

Table 6. Evaluation metrics used in this study on node count (NC) and node presence (NP). NP metrics are computed by thresholding predicted counts; a node is active if its predicted count > 0.3.

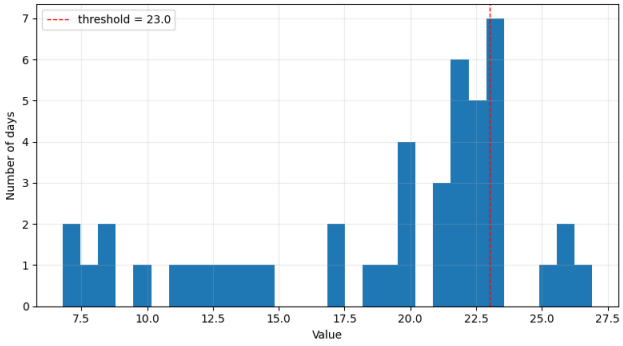
4.5.3 Scenarios

To further investigate the performance and applicability of the different model configurations, performance was also analysed in specific contexts of the Seould Cozie dataset. Two types of scenarios are considered: time-of-day and weather. In both cases scenarios are constructed from bins that meet the requirements of the scenario, and the five preceding bins are used as input.

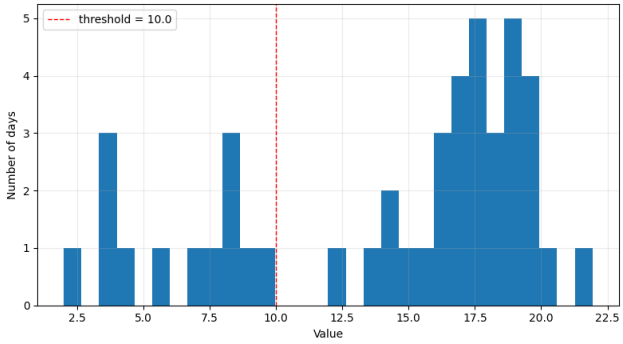
The chosen time-of-day scenarios were four two-hour time bins that cover typical daily patterns: 8:00-10:00 for morning rush, 12:00-14:00 for mid-day/lunch period, 16:00-18:00 for afternoon rush and 18:00-20:00 for early evening. These scenarios are built to reveal whether the models capture differences in activity pattern across daily cycles.

For the weather scenarios, city-level daily weather statistics were calculated using the S-DoT dataset by taking, for each day, the median

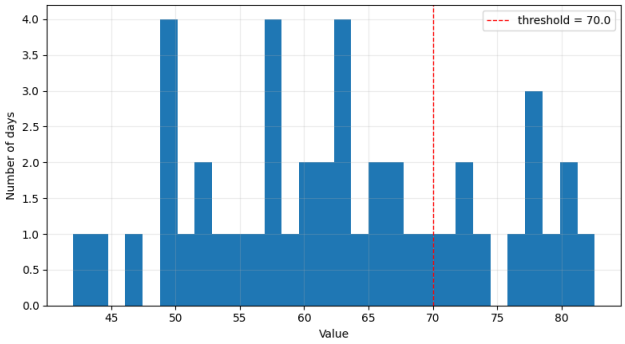
across stations of each station's daily mean temperature, humidity and PM10 values. Days with insufficient coverage of less than 30% of the stations reporting are discarded. Three scenarios were defined: hot days with a mean temperature above 23°C (fig. 13a), cold days with a mean temperature below 10°C (fig. 13b), and humid days with mean humidity above 70% (fig. 13c). For each, the days that meet the threshold are selected and evaluate within the same four time-of-day windows. The resulting slices allow us to examine whether climate-sensitive model configurations have improved performance relative to those without climate features.



a. Hot days (threshold: ≥23°C)



b. Cold days (threshold: ≤10°C)



c. Humid days (threshold: ≥70%)

Figure 13. Histograms of days with mean S-DoT values used for the scenarios sampling.

5. Results

5.1 Main Models Results

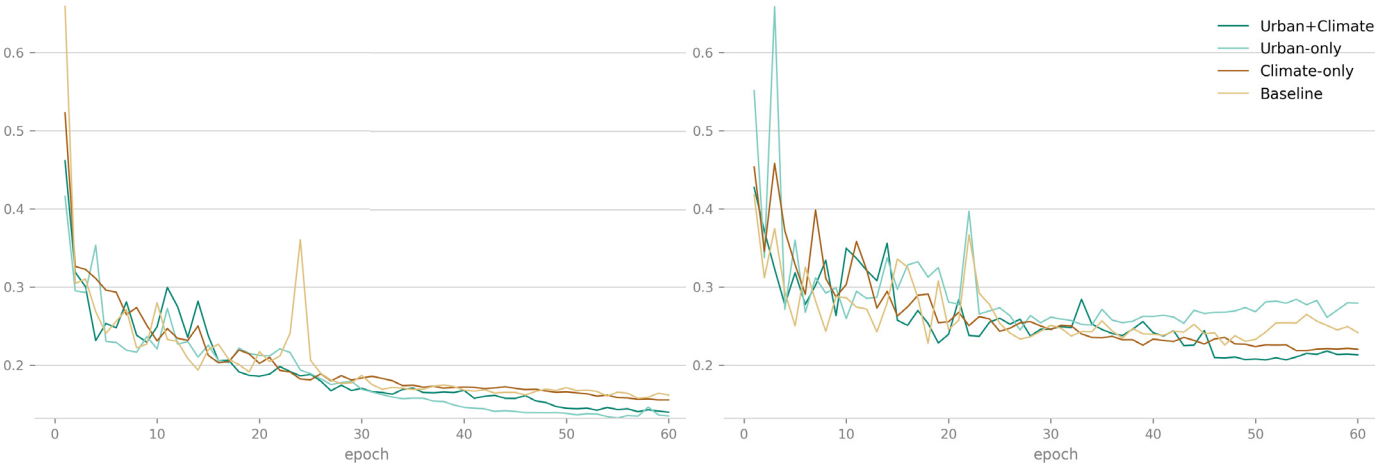
The Climate+Urban and Urban-only STGNN models performed similarly in their node count prediction task. They both achieved low overall MSE values of around 0.13 and high accuracies above 98% (table 7). When looking into their node presence metrics, they again perform similarly with Precision at around 76% and Recall at a lower 56%. Although similar, Climate+Urban performs slightly better in all metrics, especially in node presence Precision and Recall, where it surpassed Urban-only by a whole percentage point in both.

CONFIGURATION	NODE COUNT ACCURACY	NODE PRESENCE ACCURACY (%)	NODE PRESENCE PRECISION (%)	NODE PRECISION RECALL (%)	NODE PRESENCE F1 SCORE (%)
Climate+Urban	0.1250	98.09	76.41	56.51	0.6497
Urban-only	0.1314	98.04	75.33	55.53	0.6393
Cliamate-only	0.1240	97.77	67.16	56.27	0.6123
Baseline	0.1174	97.99	74.33	54.79	0.6308

Table 7. Evaluation results across model configurations.

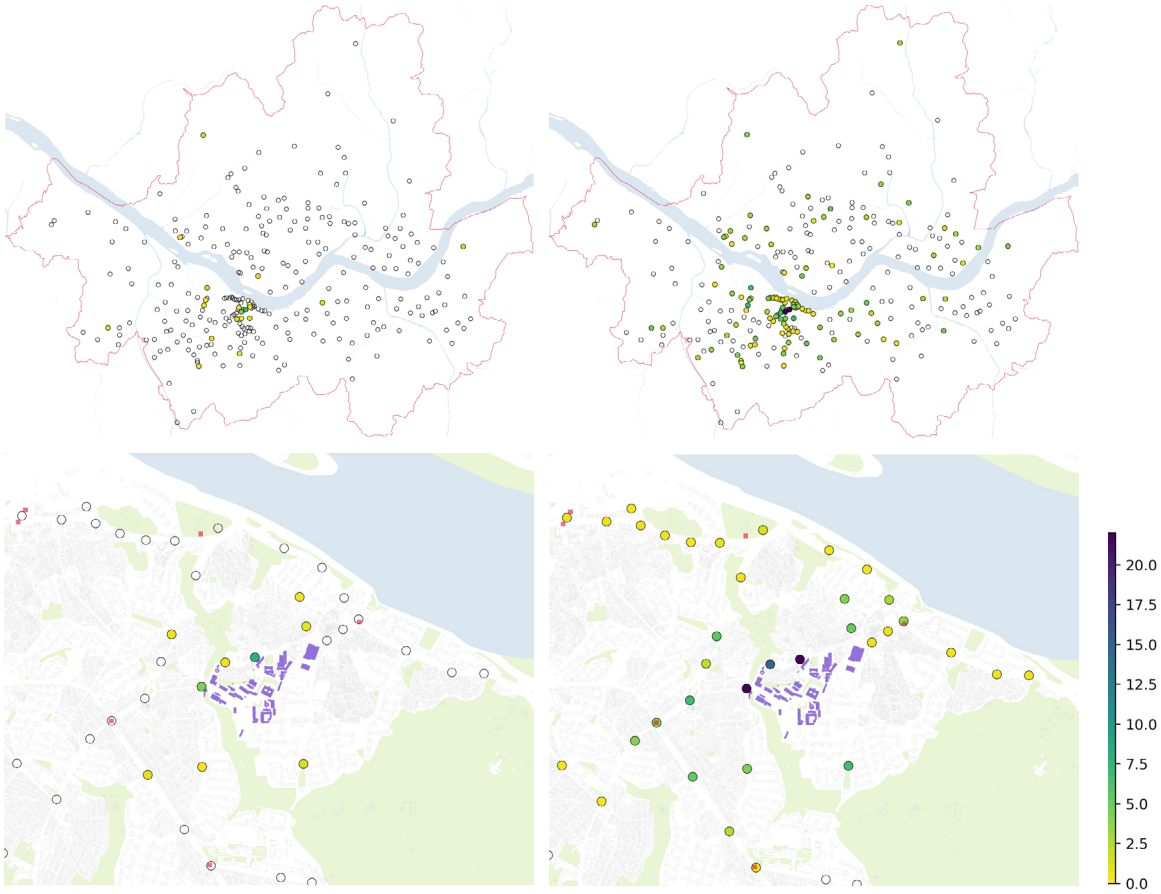
Figures 14a and 14b show the training and validation losses of all models. The training loss curves of Climate+Urban shows a big drop in the first 20 epochs and then continues to go down more slowly until leveling off at epoch 50 until 60. Its validation loss curve similarly shows big drops in the first 20 epoch but then continues to have a few small spikes until dropping again in the last 15 epochs. The Urban-only training loss also sees initial large drops with gradual decreases in the last 30 epochs. Its validation loss curve, in contrast to the Climate+Urban one, sees most of its drop in the first five epochs and then sees a few medium spikes until plateauing and even seeing a slight increase towards the last epochs. Although the exact trends vary between curves, it is noticeable that all losses see a constant decline indicating that the models are learning and optimizing to some level.

Plotting the results from the Climate+Urban model shows the distribution of the node prediction in the urban space of the best performing model (fig. 15). It is notable how the nodes around CAU buildings see the highest level of both overall mean values as well as peak values from the entire time frame. Other nodes that see particular activity are ones in proximity of the university ones, especially following in the North-South direction. Other points with medium to low activity seem scattered throughout the city and its periphery. Lastly, the rest of the nodes, which make up most the majority, appear to have overall low and null values indicating that they are frequently predicted as unvisited.



a. Training losses *b. Validation losses*

Figure 14. Node count MSE loss curves across model configurations



a. Mean values *b. Peak values*

Figure 15. Climate-Urban node count predictions at urban and neighborhood scale. CAU buildings are highlighted in purple and railway stations in red.

5.2 Urban Features Ablation

The Climate-only and Baseline STGNN models were evaluated in the same way as the main models, with the difference being their omission of urban features to control for their influence on the results. Again, the models do not perform too differently from each other when looking at their evaluation metrics (table 7), with close MSE values of 0.124 and 0.1174 and Accuracies of 97.77% and 97.99% respectively. The node presence metrics instead show some differences. Climate-only performs poorer in Precision at 67.17% compared to Baseline's 74.33%, while better in Recall with 56.27% compared to 54.79%. These values show that the Climate-only model was able to more often identify real active nodes, at the cost of more false active node predictions.

The loss curves also perform similarly to the other models, with the training losses gradually decreasing until the last epoch (fig. 14a). A notable spike is seen in the Baseline model at epoch 24, but it quickly stabilises. The losses of Climate-only and Baseline reach around 0.15, compared to the lower 0.10 achieved by the previous models. The validation losses also act similarly (fig. 14b), with more spikes throughout the curve that still overall sees the values decrease to around 0.22 and 0.25. Interestingly, these losses at the last epoch are slightly lower than the Urban-only model's but are still higher than the Climate+Urban model's.

When comparing these models' performance with the Climate+Urban and Urban-only models, we can see how they are outperformed by them in each metric. Climate+Urban remains the best performing model of the experiment, with the lowest MSE and highest Precision, Recall, and F1 score, indicating that it was the best model at both predicting node counts and when nodes will be active. In a similar way, Urban-only outperforms Baseline to be the best model to not include climate features. This improvement in metrics in the models containing urban features, however, is not drastic and sees only slight improvements, apart from Precision between Climate+Urban and Climate-only. This seems to point to the urban features adding some valuable additional information to the model but not enough to make a stark difference in its prediction task.

5.3 Scenario Results

The scenarios defined in section 4.5.3 were used to apply the trained models to specific subgroups of the total dataset timeframe. The day-of-time scenario was performed using the best performing model, Climate+Urban, and grouped by morning (08:00-10:00), mid-day (12:00-14:00), afternoon (16:00-18:00) and evening (18:00 to 20:00). For each, Figures 16 and 17 show the mean and peak node count values predicted by the model. Looking at the mean values, the nodes closest to the CAU

buildings see the most predicted activity at all times, with the highest mean visits occurring in the mid-day window with around 12 to 15 visits. Other nodes outside the university neighbourhood also see mean predicted activity, but this is more sporadic with values around 1 to 2 visits. The peak values show the same trend of centralized activity at CAU during mid-day and afternoon but also help show trends of less consistently visited nodes. In the morning, peaks are higher in nodes surrounding the university neighbourhood reaching about 7 peak visits. During the mid-day and afternoon, some more spread-out activity is

Figure 16. Mean predicted node count values of the time-of-day scenarios at urban and neighborhood scale.

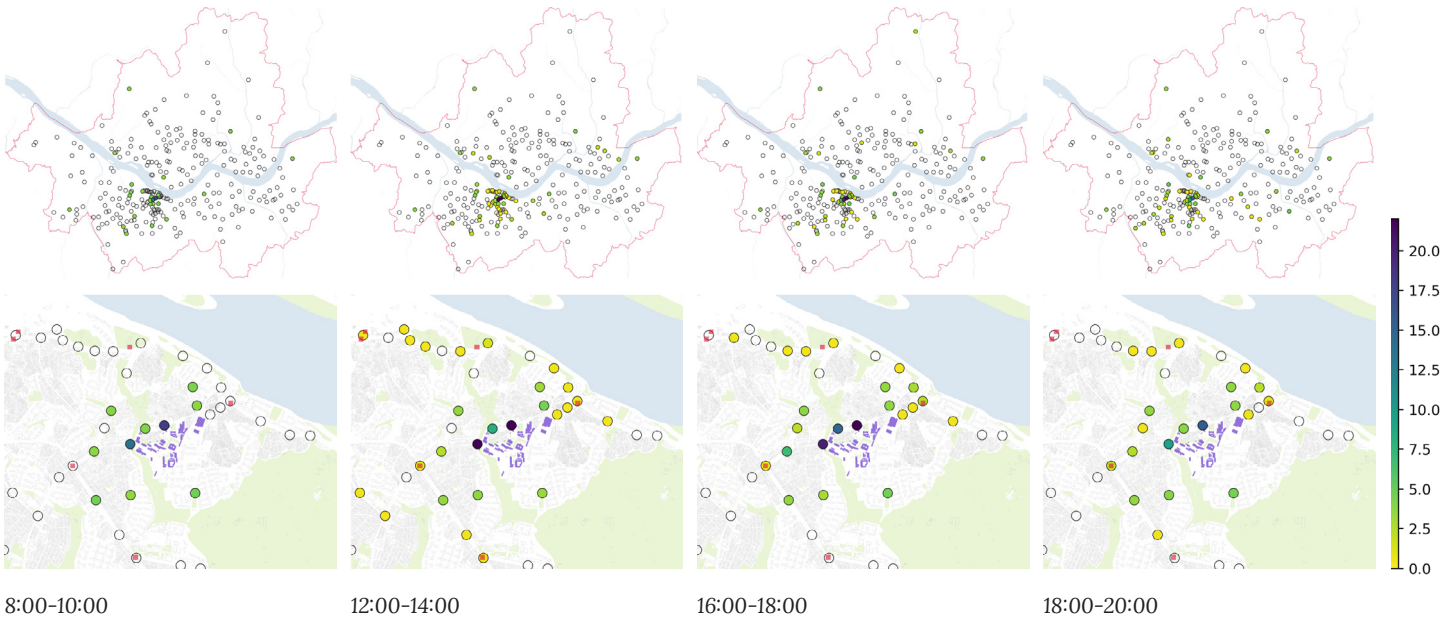
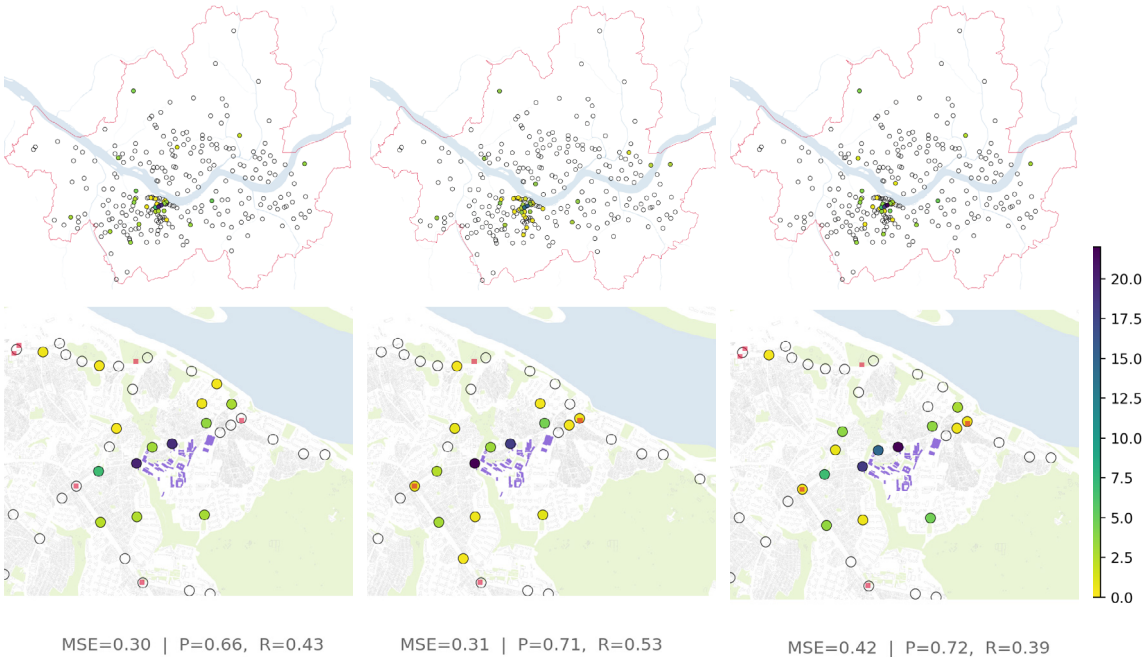


Figure 17. Peak predicted node count values of the time-of-day scenarios at urban and neighborhood scale.

predicted south of CAU, on nodes close to train stations and near the river. The evening continues this trend and the nodes closest to CAU buildings decrease in activity.

The weather scenarios that looked at hot, cold and humid days were performed on both the Climate+Urban and Urban-only models to compare the addition of climate features in more extreme weather conditions. To this end, node count MSE, node presence Precision and Recall were also calculated. Only peak values are shown in figure 18, as they seemed more expressive in capturing predictive patterns compared to mean values. As in the time-of-day scenarios, peaks in all weather scenarios are highest in the nodes closest to CAU buildings at mid-day and in the afternoon. However, some specific behaviour for the weather scenarios can still be seen: hot days seem to have slightly higher peaks in nodes closer to the river during the afternoon (fig. 18a), cold days seem to have more spread out node counts during midday (fig. 18b), and humid days seem to have more highly concentrated peak values in the nodes closest to the CAU buildings (fig. 18c).

The calculated metrics between Climate+Urban and Urban-only models in these scenarios show only minor differences. The former usually has slightly higher MSE and Recall than the latter, while the opposite occurs for Precision. This could mean that climate features make the model overall less cautious, leading to a slight increase of correctly predicted active nodes at the cost of precision. Interestingly, all the evaluation metrics calculated on these scenarios perform lower than the overall metrics shown in section 5.1, reaching lows of 0.48 for MSE, 53% for Precision and 0.26 for Recall, although this might be due to a lower number of considered time bins, sometimes reaching just 8 bins, which might make the values easily skewed.



a. Hot days (≥23°C) 16:00-18:00 *b. Cold days (≤10°C) 12:00-14:00* *c. Humid days (≥70%) 16:00-18:00*

Figure 18. Peak predicted node count values of the weather scenarios using the Climate+Urban model at urban and neighborhood scale.

6. Conclusion

6.1 Discussion

The results across the tested node feature model variations did not overall stray too far from each other. The models all achieved very low levels of MSE, between 0.1174 and 0.1314, and high node presence Accuracies, between 97.77% and 98.09%. While initially these may appear as good performance indicators, node presence Precision and Recall paint a more realistic picture. Precision values between 0.6716 and 0.7641 show that when the model predicts a node to be visited at least once, it is mostly correct. Recall values between 0.5479 and 0.5651 instead show that the model can identify only slightly more than half of visited nodes and often predicts them as inactive. These values reveal that the models are highly susceptible to the class imbalance of visited and unvisited nodes even after having reduced the spatial granularity of the graph through the intersection node grouping. This causes them to skew towards predicting zero visit counts in order to achieve lower MSE. For this reason, the node presence Presence and Recall metrics became more meaningful metrics in comparing the different model configurations than the stable MSE and Accuracy levels.

Looking at the role of climate and urban features in the model variations' performances, some small differences can be found despite their close results. Models that included either urban or climate features overall performed better than their counterparts, especially in Recall, and the best performing one in all metrics was Climate+Urban, which included both. The improvement is, however, not too drastic and saw on average a 1-2% increase in Presence and Recall, with the exception of the Climate model which underperformed in Precision. Looking qualitatively at the scenario experiments shows that the models are able to capture some general trends as they relate to time of day and weather conditions, but nothing substantial that drastically differentiates the model variations. Generally, it is possible that the climate features provided a positive effect on the model performance, as highlighted by the small increase in node presence metrics, but, along with the urban features, they had less of an impact that was expected. The short time span and low changes in weather of the Seoul Cozie dataset might have played a role. With only six weeks in an autumn that saw late a seasonal switch, the differences in temperature and humidity were small, which limited the variability available for the model to learn from.

The node visit count prediction task of the STGNN meant that Mobility within the model was defined as location based. The road intersection grouping of the Seoul Cozie positions created a network that somewhat resembled a sensor network used in vehicle traffic forecasting models, but personalized to the locations visited by the participants across all transportation types. Therefore, the output of the model when looked at a single instance or averaged over long periods of times can appear static. Activity peaks can be observed at the highly visited nodes, like

those close to the university neighbourhood, and lower values portray which locations across the city are less popular. This model output by itself can show some relations between activity levels between neighbouring nodes but does not give too much insight of the actual movements between them present in the original dataset. However, using multiple outputs of tailored scenarios to compare activity across different times and weather conditions can indirectly portray patterns in movements between the locations and across the urban environment. This allows do draw some qualitative conclusions on the movement of people as time passes.

6.2 Limitations

The research encountered various limitations in its process to reach the research goals, which encompassed the input datasets, the chosen model architecture, and the integration of node features.

The Seoul Cozie dataset was the basis for much of the research and for the construction of the STGNN model. This allowed to reach the results presented in the previous sections but was also the cause of some setbacks. Its size was relatively small for the use that it served. With 22 participants and spanning less than two months, the datapoints were often not dense enough to train a model at an urban scale and was one of the causes of the large number of unvisited nodes. At the same time, the spatial and temporal resolution of the GNSS traces also were not adequate for the neighbourhood-scale. Only the areas surrounding the CAU buildings could be analysed more in-depth due to their higher visits by all participants, but even then, precise positions could often not be determined, which led to the use of intersection-based nodes. Additionally, the dataset required extensive cleaning relating to GNSS irregularities, jumps, repeating values, and mismatched timestamps. Many of these were addressed, but it is possible that errors remained and were then fed into the developed models, introducing noise and overall incorrect inputs.

When defining the graph structure, the node grouping method simplified and flattened the dynamic nature of the Seoul Cozie dataset. Although necessary based on the dataset's own limitations, the aggregation to street intersection and municipality nodes decreased the level of detail. A node, therefore, did not represent an exact place, building, or park, but rather entire blocks or, in some cases, entire municipalities. Other than limiting the intput and output detail of the model, this also affected the precision and meaningfulness of a node's climate and morphology features. These were caluclated using the node's coordinates to represent positions inside its buffer, which might not have always provided accurate and rich feature information.

The developed STGNN architecture also could also have introduced limitations to the findings of the research. Using a simple decoder,

composed of an edge-conditioned message passing layer and a GRU layer, might not have given the model enough chances to learn the spatial and temporal pattern present in the input structure. Additionally, the window used as input was of five time bins encompassing a total of ten hours. While enough to potentially predict short-term patterns, the model would not have been able to identify longer-term influences. Additionally, the choice of a GNN-based architecture in the first place could have also hindered the research, as it was sensitive to the way its structure was defined and highly dependent on spatial links that would inform the message passing. These factors, together with the large class imbalance of unvisited nodes, might have caused the graph to be too sparse, directly impacting its performance more drastically than other non-graph-based methods.

The factors mentioned above also contributed to limitations in how the concept of mobility could be approached in relation to climate. The model predictions of node count values across time bins, while allowing some qualitative recognition of patterns in the scenario experiments, depict mobility as location-based and static. This was in large part due to the dataset resolution and graph structure, which would not allow for more granular detection, like the choice of specific paths, shaded areas, or indoor buildings. Additionally, mobility was largely affected by regular university commuting and free time, like weekends and late evenings, could not be trained on due to an inconsistent number of records at those times. Therefore, the hours where participants were most likely to enact mobility choices based on their own choices and, possibly in reaction to the climate, could not be captured in the research. In a similar way and again due to the low dataset resolution and strict graph structure, usage of transport methods was not implemented in the prediction stage, which might have given insight on transport preferences in differing weather conditions.

Finally, the integration of node climate features was also constrained. Out of the variables collected by the S-DoT stations, only three were used: temperature, humidity and PM10. These were chosen due to their low missing data and relatively stability across stations. Other potentially relevant parameters, such as wind speed, solar radiation, or shading effects, did not meet these requirements and were excluded. This means that only some of the climate conditions experienced by participants was accounted for in the model. Additionally, when interpolating the features, values were efficiently assigned to the nodes but naturally required a level of smoothening that could not reproduce potentially relevant micro-scale variations. For example, temperature differences between shaded and sunlit areas or airflow along narrow streets would have been too detailed to capture.

6.3 Conclusion and Takeaways

This research set out to explore whether urban climate factors leave a measurable trace in people's mobility choices and whether a STGNN

could capture such effects. This was approached by running versions of the same model that differed only in the inclusion of dynamic node climate features. Their performance in a node visit count prediction task was analysed. Overall, the models performed similarly and were able to produce stable forecasts of node visits from the Seoul Cozie dataset. However, the outcomes showed that much of high accuracy values (>98%) stemmed from the dominance of inactive nodes in the graph structure. This led to an over-prediction of unvisited nodes, which greatly hindered the capabilities of the model. Nonetheless, the Recall values of predicted active nodes remained above 50%, indicating that the model at least identified them at a better rate than random chance. Adding climate features lead to minor improvements. Performance showed a small increase in metrics that point to climate influencing the mobility dataset. However, this was less than initially expected and as such relations or patterns between urban climate and mobility could not be definitively identified. The efficacy of this STGNN approach in mobility prediction therefore proved to be limited during evaluation. The task that the model was trained for, predicting next-step node visits, allowed to clearly identify peaks in activity in different areas of the city but proved more difficult in directly portraying mobility. A possible real use of such a model was experimented through the use of specific scenarios. By applying the model on these specifically chosen time bin subsets, the node count predictions could be used as indicators of shifting activity based on the chosen factors, such as time of day, temperature and humidity.

Many of the conclusions above stemmed from the way the construction of a graph structure from the Seoul Cozie dataset was approached. The grouping of GNSS traces into road intersection nodes allowed for a working node set that aimed at balancing their spatial expressiveness and data richness. This, while allowing to have a working implementation of the model, still could not counteract the effect of the class imbalance between visited locations inherent to the dataset. Additionally, the way mobility could be interpreted largely relied on this decision. Whether by looking at overall peaks or comparing scenarios, the model would only be able to give readings of people's activity in relation to their closest road intersection, a scale in between the urban and neighbourhood level.

Overall, the work made several contributions. It demonstrated that raw GNSS traces, like the Seoul Cozie dataset, can be successfully used as input into a dynamic graph and applied in a spatiotemporal framework. It showed that both morphology and climate attributes can be integrated in this structure, and it established a foundation that can be extended with richer datasets and alternative graph structures. Although no strong claims about the effect of climate on mobility could be definitively reached, an initial STGNN framework using a personal mobility dataset was proposed to be built upon.

6.4 Future Work

Building on this study, several directions could extend and strengthen the results. Starting from the input used, new datasets could be gathered more tailored to the goal of STGNN construction. They could be based off individual peoples' mobile traces, much like the Seoul Cozie dataset, but improve in more consistent positional information that would allow to create much more precise node profiles. The time variations between recorded coordinates across participants could be more standardised. Input from the user on their location information, indicating whether they are inside, outside, or taking transportations, could also be more frequent, or at least enough to infer the rest from the existing ones. To further aid the creation of meaningful nodes, a locationing method, like Wi-Fi fingerprinting, could be used to identify when participants are visiting specific public locations, like a university building or railway station. A larger study size, comprising of more than 22 participants from different backgrounds and spanning a longer time period, would also help in enriching the graph structure with movement information and in capturing more meaningful patterns with changing weather conditions.

The construction of the graph could also be improved by future research. A dynamic, event-based, temporal structure could be explored in comparison to the binning one used in the researched, which could allow for finer and less static temporal sensitivity. Different ways of grouping the Seoul Cozie dataset into nodes could be investigated, like grouping based on city blocks or with a uniform grid system. Additionally, distance-aware edges between nodes could be introduced to improve the model's understanding of relationships between close nodes. The features considered could also be expanded to include more relevant information to urban factors, such as terrain height, sky view factor, shading, tree coverage, and more.

The chosen STGNN architecture could be focused on to improve its learning capabilities. Additional components commonly used in GNNs for mobility and transport forecasting tasks might be beneficial. Explicit long/short-term split modules would give the model a better understanding of daily patterns and more localized events. An attention component applied to the edge or nodes might also prove to be effective in more consistently helping the model identify important relationships between nodes. More broadly, the analysis of mobility and urban climate within the Seoul Cozie dataset could be approached with alternatives to the STGNN used. Edge-centric GNNs would model flows directly, representing movements as edges and predicting origin-destination counts rather than node ones. Sequence models, like RNNs and LSTMs can learn regularities in node visits using per-person sequences, with spatial context added as exogenous features. Each choice could come as a trade-off between spatial expressiveness and simplicity, and future work could quantify when each is preferable.

References

P. Alva, M. Mosteiro-Romero, W. Pei, A. Bartolini, C. Yuan, and R. Stouffs. Bottom-up approach for creating an urban digital twin platform and use cases. In *Proceedings of the 28th Conference on Computer Aided Architectural Design Research in Asia (CAADRIA) [Volume 1]*, CAADRIA 2023. CAADRIA, 2023. doi: 10.52842/conf.caadria.2023.1.605.

F. Benita, G. Bansal, and B. Tunçer. Public spaces and happiness: Evidence from a large-scale field experiment. *Health amp; Place*, 56:9–18, Mar. 2019. ISSN 1353-8292. doi: 10.1016/j.healthplace.2019.01.014.

P. Bröde, K. Blazejczyk, D. Fiala, G. Havenith, I. Holmér, G. Jendritzky, K. Kuklane, and B. Kampmann. The universal thermal climate index utci compared to ergonomics standards for assessing the thermal environment. *Industrial Health*, 51(1):16–24, 2013. ISSN 1880-8026. doi: 10.2486/indhealth.2012-0098.

T. Cheung, S. Schiavon, T. Parkinson, P. Li, and G. Brager. Analysis of the accuracy on pmv – ppd model using the ashrae global thermal comfort database ii. *Building and Environment*, 153:205–217, Apr. 2019. ISSN 0360-1323. doi: 10.1016/j.buildenv.2019.01.055.

A. Cini, I. Marisca, D. Zambon, and C. Alippi. Graph deep learning for time series forecasting. Oct. 2023. doi: 10.48550/ARXIV.2310.15978.

E. Erell. The application of urban climate research in the design of cities. *Advances in Building Energy Research*, 2(1):95–121, Jan. 2008. ISSN 1756-2201. doi: 10.3763/aber.2008.0204.

J. Fallmann and S. Emeis. How to bring urban and global climate studies together with urban planning and architecture? *Developments in the Built Environment*, 4:100023, Nov. 2020. ISSN 2666-1659. doi: 10.1016/j.dibe.2020.100023.

Y. Fang, Y. Qin, H. Luo, F. Zhao, B. Xu, C. Wang, and L. Zeng. Spatio-temporal meets wavelet: Disentangled traffic flow forecasting via efficient spectral graph attention network, 2021.

J. Gehl. *Cities for People*. Island Press, Washington, DC, 2010. ISBN 9781597269841. Description based on publisher supplied metadata and other sources.

A. Gupta, B. De, S. Das, and M. Mukherjee. Thermal hazards in urban spaces: A review of climate-resilient planning and design to reduce the heat stress. *Urban Climate*, 59:102296, Feb. 2025. ISSN 2212-0955. doi: 10.1016/j.uclim.2025.102296.

M. Hebbert. Climatology for city planning in historical perspective. *Urban Climate*, 10:204–215, Dec. 2014. ISSN 2212-0955. doi: 10.1016/j.uclim.2014.07.001.

T. Honjo. Thermal comfort in outdoor environment. *Global Environmental Research*, 13:43–47, 2009.

M. A. Humphreys and J. Fergus Nicol. The validity of iso-pmv for predicting comfort votes in every-day thermal environments. *Energy and Buildings*, 34(6):667–684, July 2002. ISSN 0378-7788. doi: 10.1016/s0378-7788(02) 00018-x.

M. Ignatius, J. Lim, B. Gottkehaskamp, K. Fujiwara, C. Miller, and F. Biljecki. Digital twin and wearables unveiling pedestrian comfort dynamics and walkability in cities. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W5-2024:195–202, June 2024. ISSN 2194-9050. doi: 10.5194/isprs-annals-x-4-w5-2024-195-2024.

P. Jayathissa, M. Quintana, M. Abdelrahman, and C. Miller. Humans-as-a-sensor for buildings—intensive longitudinal indoor comfort models. *Buildings*, 10(10):174, Oct. 2020. ISSN 2075-5309. doi: 10.3390/buildings10100174.

I. Jeddoub, G.-A. Nys, R. Hajji, and R. Billen. Digital twins for cities: Analyzing the gap between concepts and current implementations with a specific focus on data integration. *International Journal of Applied Earth Observation and Geoinformation*, 122:103440, Aug. 2023. ISSN 1569-8432. doi: 10.1016/j.jag.2023.103440.

G. Jeon, Y. Park, and J.-M. Guldmann. Impacts of urban morphology on seasonal land surface temperatures: Comparing grid- and block-based approaches. *ISPRS International Journal of Geo-Information*, 12(12):482, Nov. 2023. ISSN 2220-9964. doi: 10.3390/ijgi12120482.

W. Jiang and J. Luo. Graph neural network for traffic forecasting: A survey. *Expert Systems with Applications*, 207:117921, Nov. 2022. ISSN 0957-4174. doi: 10.1016/j.eswa.2022.117921.

G. Jin, Y. Liang, Y. Fang, Z. Shao, J. Huang, J. Zhang, and Y. Zheng. Spatio-temporal graph neural networks for predictive learning in urban computing: A survey, 2023.

D. Jonietz. *Personalizing Walkability: A Concept for Pedestrian Needs Profiling Based on Movement Trajectories*, pages 279–295. Springer International Publishing, 2016. ISBN 9783319337838. doi: 10.1007/978-3-319-33783-816.

D. Jonietz and D. Bucher. Towards an analytical framework for enriching movement trajectories with spatio-temporal context data. In A. Bregt, T. Sarjakoski, R. Van Lammeren, and F. Rip, editors, *Societal Geo-Innovation: Short Papers, Posters and Poster Abstracts of the 20th AGILE Conference on Geographic Information Science*, number 133, Wageningen, May 2017. Association of Geographic Information Laboratories for Europe.

D. Kim, W. Nam, and K. C. Park. Application of urban computing to explore living environment characteristics in seoul : Integration of s-dot sensor and urban data. *Journal of Internet Computing and Services*, 24(4):65–76, Aug. 2023. URL https://koreascience.or.kr/article/JAKO202326445689018.page.

X. Kong, W. Xing, X. Wei, P. Bao, J. Zhang, and W. Lu. Stgat: Spatial-temporal graph attention networks for traffic flow forecasting. *IEEE Access*, 8:134363–134372, 2020. ISSN 2169-3536. doi: 10.1109/access.2020.3011186.

H. Ledoux, K. Arroyo Ohori, R. Peters, and M. Pronk. *Computation modelling of terrains*. 2024.

J. Liu, Y. Chen, X. Huang, J. Li, and G. Min. Gnn-based long and short term preference modeling for next-location prediction. *Information Sciences*, 629: 1–14, June 2023. ISSN 0020-0255. doi: 10.1016/j.ins.2023.01.131.

X. Liu, Z. Gou, and C. Yuan. Application of human-centric digital twins: Predicting outdoor thermal comfort distribution in singapore using multi-source data and machine learning. *Urban Climate*, 58:102210, Nov. 2024. ISSN 2212-0955. doi: 10.1016/j.uclim.2024.102210.

A. Longa, V. Lachi, G. Santin, M. Bianchini, B. Lepri, P. Lio, F. Scarselli, and A. Passerini. Graph neural networks for temporal graphs: State of the art, open challenges and opportunities. *Transactions on Machine Learning Research*, Aug. 2023. ISSN 2835-8856. URL https://openreview.net/forum? id=pHCdMat0gI.

H. Ma, M. Zhou, X. Ouyang, D. Yin, R. Jiang, and X. Song. Forecasting regional multimodal transportation demand with graph neural networks: An open dataset. In *2022 IEEE 25th International Conference on Intelligent Transportation Systems* (ITSC), pages 3263–3268. IEEE, Oct. 2022. doi: 10.1109/itsc55140.2022.9922512.

D. Maiullari. *Urban form influence on microclimate and building cooling demand: An analytical framework and its application in the Rotterdam case*. phd thesis, Delft University of Technology, 2023.

D. Mauree, E. Naboni, S. Coccolo, A. Perera, V. M. Nik, and J.-L. Scartezzini. A review of assessment methods for the urban environment and its energy sustainability to guarantee climate adaptation of future cities. *Renewable and Sustainable Energy Reviews*, 112:733–746, Sept. 2019. ISSN 1364-0321. doi: 10.1016/j.rser.2019.06.005.

S. Mazzetto. A review of urban digital twins integration, challenges, and future directions in smart city development. *Sustainability*, 16(19):8337, Sept. 2024. ISSN 2071-1050. doi: 10.3390/su16198337.

B. Monnot, E. Wilhelm, G. Piliouras, Y. Zhou, D. Dahlmeier, H. Y. Lu, and W. Jin. *Inferring Activities and Optimal Trips: Lessons From Singapore's National Science Experiment*, pages 247–264. Springer International Publishing, 2016. ISBN 9783319296432. doi: 10.1007/978-3-319-29643-219.

M. Mosteiro-Romero, Y. Park, and C. Miller. Converging smartwatch and urban datasets for sustainable city planning: A case study in Seoul, South Korea. *E3S Web of Conferences*, 562:03004, 2024. ISSN 2267-1242. doi: 10. 1051/e3sconf/202456203004.

M.-S. Park and K. Baek. Quality management system for an iot meteorological sensor network—application to smart Seoul data of things (s-dot). *Sensors*, 23(5):2384, Feb. 2023. ISSN 1424-8220. doi: 10.3390/s23052384.

Z. Peng, R. Bardhan, C. Ellard, and K. Steemers. Urban climate walk: A stop-and-go assessment of the dynamic thermal sensation and perception in two waterfront districts in Rome, Italy. *Building and Environment*, 221: 109267, Aug. 2022. ISSN 0360-1323. doi: 10.1016/j.buildenv.2022.109267.

J. Rico, J. Barateiro, and A. Oliveira. Graph neural networks for traffic forecasting. 2021. doi: 10.48550/ARXIV.2104.13096.

E. Rossi, B. Chamberlain, F. Frasca, D. Eynard, F. Monti, and M. Bronstein. Temporal graph networks for deep learning on dynamic graphs, 2020.

A. Roy, K. K. Roy, A. A. Ali, M. A. Amin, and A. K. M. M. Rahman. Sst-gnn: Simplified spatio-temporal traffic forecasting model using graph neural network, 2021.

Seoul Metropolitan Government. Analysis of city data "s-dot" collected by 1,100 sensors in Seoul is released, Mar. 2021. URL https://english.seoul.go.kr/analysis-of-city-data-s-dot-collected-by-1100-sensors-in-seoul-is-released/.

E. Shahat, C. T. Hyun, and C. Yeom. City digital twin potentials: A review and research agenda. *Sustainability*, 13(6):3386, Mar. 2021. ISSN 2071-1050. doi: 10.3390/su13063386.

A. Sharma, A. Sharma, P. Nikashina, V. Gavrilenko, A. Tselykh, A. Bozhenyuk, M. Masud, and H. Meshref. A graph neural network (gnn)-based approach for real-time estimation of traffic speed in sustainable smart cities. *Sustainability*, 15(15):11893, Aug. 2023. ISSN 2071-1050. doi: 10.3390/su151511893.

S. Shleifer, C. McCreery, and V. Chitters. Incrementally improving graph wavenet performance on traffic prediction, 2019.

S. S.K.B, S. K. Mathivanan, H. Rajadurai, J. Cho, and S. V. Easwaramoorthy. A multi-modal geospatial–temporal lstm based deep learning framework for predictive modeling of urban mobility patterns. *Scientific Reports*, 14(1), Dec. 2024. ISSN 2045-2322. doi: 10.1038/s41598-024-74237-3.

F. Tartarini, M. Frei, S. Schiavon, Y. X. Chua, and C. Miller. Cozie apple: An ios mobile and smartwatch application for environmental quality satisfaction and physiological data collection. *Journal of Physics: Conference Series*, 2600 (14):142003, Nov. 2023. ISSN 1742-6596. doi: 10.1088/1742-6596/2600/14/142003.

F. Terroso-Sáenz and A. Muñoz. Nation-wide human mobility prediction based on graph neural networks. *Applied Intelligence*, 52(4):4144–4160, July 2021. ISSN 1573-7497. doi: 10.1007/s10489-021-02645-3.

N. Upasani, O. Guerra-Santin, and M. Mohammadi. Developing building-specific, occupant-centric thermal comfort models: A methodological ap- proach. *Journal of Building Engineering*, 95:110281, Oct. 2024. ISSN 2352- 7102. doi: 10.1016/j.jobe.2024.110281.

J. van Hoof. Forty years of fanger's model of thermal comfort: comfort for all? *Indoor Air*, 18(3):182–201, June 2008. ISSN 1600-0668. doi: 10.1111/j. 1600-0668.2007.00516.x.

K. D. Vo, E.-J. Kim, and P. Bansal. A novel data fusion method to leverage passively-collected mobility data in generating spatially-heterogeneous synthetic population. *Transportation Research Part B: Methodological*, 191: 103128, Jan. 2025. ISSN 0191-2615. doi: 10.1016/j. trb.2024.103128.

K. Wang, Z. Peng, M. Cai, H. Wu, L. Liu, and Z. Sun. Coupling graph neural networks and travel mode choice for human mobility prediction. *Physica A: Statistical Mechanics and its Applications*, 646:129872, July 2024. ISSN 0378- 4371. doi: 10.1016/j.physa.2024.129872.

Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. 2019a. doi: 10.48550/ARXIV.1901.00596.

Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang. Graph wavenet for deep spatial-temporal graph modeling, 2019b.

H. Xia, Z. Liu, M. Efremochkina, X. Liu, and C. Lin. Study on city digital twin technologies for sustainable smart city design: A review and bibliometric analysis of geographic information system and building information modeling integration. *Sustainable Cities and Society*, 84:104009, Sept. 2022. ISSN 2210-6707. doi: 10.1016/j.scs.2022.104009.

X. Ye and D. Niyogi. Resilience of human settlements to climate change needs the convergence of urban planning and urban climate science. *Computational Urban Science*, 2(1), Feb. 2022. ISSN 2730-6852. doi: 10.1007/ s43762-022-00035-0.

G. Yeghikyan, F. L. Opolka, M. Nanni, B. Lepri, and P. Lio'. Learning mobility flows from urban features with spatial interaction models and neural networks, 2020.

B. Yu, H. Yin, and Z. Zhu. Spatio-temporal graph convolutional networks: A deep learning framework for traffic forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, IJCAI- 2018, pages 3634–3640. International Joint Conferences on Artificial Intelligence Organization, July 2018. doi: 10.24963/ijcai.2018/505.

S. Zafarmandi and A. Matzarakis. Investigating thermal comfort indices in relation to clothing insulation value: A survey of an outdoor space in Tehran, Iran. *Sustainable Cities and Society*, 118:105993, Jan. 2025. ISSN 2210- 6707. doi: 10.1016/j.scs.2024.105993.

J. Zhang, Y. Chen, T. Wang, C.-Z. T. Xie, and Y. Tian. Mixture of spatial–temporal graph transformer networks for urban congestion prediction using multimodal transportation data. *Expert Systems with Applications*, 268:126108, Apr. 2025. ISSN 0957-4174. doi: 10.1016/j. eswa.2024.126108.

S. Zhang, X. Zhang, D. Niu, Z. Fang, H. Chang, and Z. Lin. Physiological equivalent temperature-based and universal thermal climate index-based adaptive-rational outdoor thermal comfort models. *Building and Environment*, 228:109900, Jan. 2023. ISSN 0360-1323. doi: 10.1016/j. buildenv.2022.109900.

Y. Zhu, Y. Zhang, and F. Biljecki. Understanding the user perspective on urban public spaces: A systematic review and opportunities for machine learning. *Cities*, 156:105535, Jan. 2025. ISSN 0264-2751. doi: 10.1016/j.cities.2024.105535.