CHALMERS
UNIVERSITY OF TECHNOLOGY

Flows Generation for Synthetic Travel Demand

Master's thesis in Sustainable Energy Systems

HAO CONG

Department of Space, Earth and Environment
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023

Master's thesis 2023

Flows Generation for Synthetic Travel Demand

HAO CONG

Department of Space, Earth and Environment
Chalmers University of Technology
Gothenburg, Sweden 2023

Flows Generation for Synthetic Travel Demand

HAO CONG

Supervisor: Yuan Liao, Department of Space, Earth and Environment
Examiner: Sonia Yeh, Department of Space, Earth and Environment

Master's Thesis 2023
Department of Space, Earth and Environment
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2023

Flows Generation for Synthetic Travel Demand

HAO CONG
Department of Space, Earth and Environment
Chalmers University of Technology

Abstract

Understanding travel flows and predicting future growth and changes are essential for sustainable development and climate change mitigation. Current models predicting the travel flows, such as the Gravity model and the Radiation model, have several shortcomings, including only focusing on the spatial dependence of mobility flows without capturing the varying frequencies of repeated visits to the same location. A new model of human mobility called 'Visitation law' has recently been proposed. It considers the spatiotemporal patterns of movement flux and requires minimum input data, e.g., population density and distance. However, the new model has only been tested at the city level. Its effectiveness on a larger scale, such as at the national level, still needs exploration. This thesis applies the Visitation law-based model to Sweden and evaluates its performance. A novel hierarchical approach is implemented to reduce the computing power while achieving satisfactory accuracy. Factors that may influence the performance of the model, such as grids size, distance, and aggregation levels, have been extensively explored. This thesis also creates the projection of the future travel demand of Sweden based on five different scenarios of Shared Socioeconomic Pathways (SSPs) until 2100. The results show that the projected travel demands are only influenced by the total population growth while showing little variations of travel demand per capita and average trip distance across SSPs. In other words, the model is not sensitive to different spatial distributions of population density across SSPs or over time. To apply the approach based on a new mobility model (visitation law-based model) to a larger scale, this thesis contributes to the development of a novel hierarchical implementation to reduce the computing power of generating spatially explicit travel demand projections. More work still needs to be done, however, to explore alternative models that can better distinguish travel demand per capita given heterogenous population density, such as those in the SSPs.

Keywords: Travel demand estimation, flows generation, visitation law, human mobility model, Shared Socioeconomic Pathways.

Acknowledgements

I am so appreciative that Yuan and Sonia are my supervisor and examiner, who are the best in my mind. Thank them for giving me this great opportunity, and I am really grateful for their help, support, encouragement, and guidance during these months.

Hao Cong, Gothenburg, April 2023

Contents

List of Figures

List of Tables

1

Introduction

To achieve the temperature target of limiting global warming to $< 2°C$ relative to the pre-industrial levels, the emissions of greenhouse gas (GHG) need to be limited. At the same time, continuous transportation growth constrains the sustainable development of global cities.

The prediction of traffic and crowd flows is crucial for transportation planning and climate change mitigation, where understanding human mobility, the movements of individuals in space and time, is crucial. Among all the related studies of human mobility modelling, one fundamental problem is to generate the flows between a group of locations considering demographic and geographic characteristics without any historical information about the actual movement [1]. Besides, current transportation-related energy system models mainly rely on data that lack spatial and temporal details and are readily outdated but costly to update. Thus, it is significant to synthesise travel demand based on simple and easy-to-update data for better-informed decision-making.

At present, the commonly used flow generation models, such as the Gravity law [1], the Radiation model [2], and related approaches [3], are still inadequate as they focus on the spatial dependence of the flows without capturing the varying frequencies of repeated visits to the same location [4]. A recently proposed human mobility model considering the space-time spectrum of the migration flux [4] shows that the number of visitors to any location is inversely proportional to the square of the distance travelled and frequency of visit, which is the so-called Visitation law. The required data for this method in estimating the flows between places are only location attractiveness, e.g., population density and distance, which are easy to get. And the results of this new method are shown to be robust across multiple metropolitan regions. Despite its initial promises, this model based on visitation law has only been tested at the city level. Its performance at a larger scale, such as the country level, remains unclear. This is particularly important given that long-distance trips, contributing a large share of carbon emissions, are underrepresented when we only look at urban areas [5].

In addition to representing transport demand in its present status, long-term projections of transport demand are also crucial for future assessment of climate change. Over the past decades, projections of the possible impacts of climate change have been mainly based on scenarios presented in the Intergovernmental Panel on Climate

Change's (IPCC) Special Report on emission scenarios (SRES), which are used to represent the range of driving forces (including demographics, economic growth and technological development) and emissions in different development paths, including potential uncertainty. However, advances in scientific understanding and modelling mean that SRES scenarios are becoming obsolete. For example, earth system models now require a broader set of input data, population projections for the mid to late century are different from those used in SRES scenarios, and so on. For the resulting scenarios to have greater scientific credibility, it is up to the scientific community rather than the IPCC to lead further scenario development [6]. Therefore, a new set of global, spatially explicit population scenarios has been made publicly available, which are consistent with Shared Socioeconomic Pathways (SSPs) developed to support global change research [7]. The SSPs are Sustainability (SSP1), Middle of the road (SSP2), Regional rivalry (SSP3), Inequality (SSP4), and Fossil-fueled development (SSP5). This provides the population prediction with spatial resolution and can be used as data input for our flows generation model.

The purpose of this master thesis is to implement the mobility model based on visitation law to a country level and model the synthetic travel demand under different future scenarios, which would be beneficial for policy-making and transport planning.

1.1 Related work

Human mobility refers to the movement of people in space and time. A displacement consists of a start, an endpoint, and a specific trajectory [8]. Individual flows can be aggregated to describe the movement of people between different locations/regions. An origin-destination matrix (ODM) can be constructed from the origin and destination of all trips based on the aggregated spatiotemporal scale to quantify the travel needs of a population in an area. The models were developed to generate ODMs capture regularities in human movement trajectories and reproduce the spatiotemporal structure of travel demand. The study of human mobility has critical applications for such research as estimating migration flows, traffic forecasting, urban planning, and epidemiological modelling [3].

The rise of ubiquitous computing (e.g., mobile phones, the Internet of Things, social media platforms) provides low cost and abundant quantity to perceive human movement at different scales of time and space [9]. However, these empirical data sources of human mobility are not broadly available, and they often come in varying population and behavioural biases affected by country and culture.

Given the limitations of empirical data and their constraints in travel demand projection, we need simple, easy-to-access, and globally consistent data such as population density that can be used to synthesise travel demand estimates and projections across varying regions in the world.

1.1.1 Characterising population mobility: Origin–Destination Matrices (ODM)

Human mobility at a population level can be represented by an origin - destination matrix (ODM), F. It is the standard object of aggregate mobility research and transportation planning. It provides an estimate of the number of individuals travelling between a given region over time. An $n \times m$ OD matrix is the matrix where n is the number of different "origin" regions, m is the number of "destination" regions, and $F_{i,j}$ is the number of people travelling from zone i to zone j. Traditionally, regions are administrative or traffic analysis units, which may range in size from census and electoral units to entire cities, departments, or states. The spatial resolution of the matrix also depends on the data source [3].

A trip, containing an origin and a destination, is the essential unit of an ODM, which can be characterised by multiple indicators, including distance, start and end time, duration, purpose and the route taken (i.e., trajectory). Trip distance $d_{i,j}$ refers to the distance between the origin i and destination j. Travel distance $D_{i,j}$ refers to the actual distance/network distance. With empirical data, travel distance is usually calculated by summing up the travelling trajectory given a fine-enough sampling resolution. Travel time $TT_{i,j}$ is the time spent on traveling from one location to another location. Trip frequency $f_{i,j}$ refers to how frequently trips are established between two locations. Trip purpose $P_{i,j}$ refers to the purpose of this trip, e.g., work and leisure. Equation 1.1 shows that given the above basic indicators, the movement of individuals (p) forms a network of different positions ($\mathbf{G}_p$).

$$\mathbf{G}_p = (d, D, TT, m, f, P)_{i,j}, i, j = 1, 2, \ldots, n_p \tag{1.1}$$

Aggregating these metrics for all individuals, $p = 1, 2...N$, gives the movement flows of the total population. The basic form of origin-destination (OD) matrices without temporal dimension, has the following formulation.

$$\mathbf{G} = (d, F)_{i,j}, i, j = 1, 2, \ldots, N \tag{1.2}$$

Where $F_{i,j}$ is the total number of trips between location i and location j, and N is the total number of distinct locations for all individuals [8].

1.1.2 Modelling population mobility

The human mobility model can be divided into individual-level and population-level. Due to the free will and arbitrariness of individual actions, the travel pattern of individual mobility has a certain degree of randomness. Therefore, the concepts and methods of random walking and Brownian motion will be involved in individual mobility models. At the population level, the models aim to describe the overall mobility of many individuals through an origin - destination matrix [3]. For the purpose of generating travel demand at a large scale, this paper focuses on the population-level models.

Gravity Model

The Gravity Model was inspired by Newton's law of gravitation and assumes that the number of trips between two locations decreases with distance, which is one of the most widely spread approaches for flows generation [10].

The Gravity model has the following formulation:

$$F_{ij} = \frac{Cm_i m_j}{g\left(r_{ij}\right)} \tag{1.3}$$

Where F_{ij} is the number of trips (or the number of commuters) from location i to location j, C is a constant, m_i and m_j represent key local attributes, and $g(r_{ij})$ describes the distance dependence of population flows. The distance–decay function g typically corresponds to a power law or an exponential function [4].

Literature shows that the Gravity model performs well on scale-free networks in a variety of research backgrounds. The model has a wide range of applications in the fields of transportation [11], economy [12], geography [13], urban planning [14], public health [15], politics [16], and other spatial interactions and their intersections [1].

Although the Gravity Model has the obvious advantage of being able to interpret and require some parameters by design, it also has some disadvantages, such as not accurately capturing the structure of the actual flow and the fact that the actual flow is more variable than expected. Because the Gravity model relies on a limited set of variables, often just population and distance between locations, the resulting flow does not take into account information critical to the complexity of geographical landscapes, such as land use, diversity of points of interest (POIs), and transportation networks.

Therefore, more detailed input data and a more flexible model are needed to generate a more realistic flow [1]. For example, Liao et al. (2022) [9] combine the Gravity model with geolocations of Twitter data and develop a novel way of using geotagged tweets as an attraction generator to better reflect travel demand. In addition, the basic Gravity model generally performs well on coarse-grained data but the performance decays for the finer-scale prediction of flows generation [17]. Thus, the basic Gravity model is not a good choice for simulating urban mobility networks on finer spatial and temporal scales.

Radiation Model

In recent years, a radiation model derived from diffusion dynamics has been widely used because of its simple form and non-parametric properties [18], and its introduction provides new insights into the long history of modelling human mobility. The framework of the Radiation model was first introduced by Simini et al. [2] based on first principles, but it can overcome the drawbacks of the Gravity model to some extent.

The Radiation model is defined as the following formulation:

$$F_{ij} = \frac{F_i m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})} \tag{1.4}$$

Where F_{ij} is the average commuting flux from location i to j, F_i is the total number of commuters per time starting their journey from location i, which is proportional to the population of location i, m_i and m_j are the number of opportunities at the origin and at the destination, respectively, and s_{ij} is the number of opportunities within a circle of radius r_{ij} centred in i [2].

Although many studies of the Radiation model are promising and they overall give competitive results, the dilemma of "simplicity versus diversity" has also been a major challenge to the Radiation model [18]. Certain significant elements, such as spatial scale and heterogeneity, are ignored in the model, and the thermodynamic limit assumption of the original Radiation model obviously underestimates the commuter flow in large cities [19].

Therefore, there are studies that improve the Radiation model. For example, the PWO model, which is a population-weighted opportunities model without any adjustable parameters, is developed as an alternative to the Radiation model to reproduce and predict the movement behaviour of cities of different sizes, economic levels and cultural backgrounds [20].

Visitation law-based approach

A scaling law that captures the space-time spectrum of population movements has recently been presented [4]. According to this law, which is derived from data on large-scale population movements in different cities around the world including Greater Boston, Portugal, Senegal, Ivory Coast and Singapore, the number of tourists in any location decreases inversely as the product of the frequency of visits multiplied by the distance travelled. Based on this visitation law, a model of human mobility for flows generation is constructed, and the study allows the prediction of recurrent flows, providing the application basis for urban planning, traffic engineering and epidemic mitigation [4]. More details about the Visitation law are described in 'Methods'.

1.1.3 Global population scenarios consistent with the Shared Socioeconomic Pathways

The predicted future population size and spatial distribution are widely used to predict energy consumption and carbon emissions, which are important factors in the study of climate and global environmental change. A new set of global-scale, spatially explicit population projection scenarios is now presented, and these scenarios are consistent with new shared socioeconomic pathways (SSPs) developed to facilitate climate global change research [7].

These SSPs describe five alternative outcomes that incorporate trends in social factors such as demographics, economy, technological development, lifestyle, and governance [6]. The SSPs include not only qualitative statements of future development [21], but also quantitative projections of key factors [22], such as population growth and education composition, urbanization, and economic growth at the national level. A short description of each situation is shown in Figure 1.1 [6].

One of the most important SSPs outcomes is a set of predictions of spatial distribution of the population in the future, which is consistent with the five SSPs. The latest population projection scenarios show that the five SSPs result in significantly different spatial demographic outcomes at continental, national, and sub-national scales. Overall, the ranking of the three factors that most influenced the results were country-level population change, urbanization rates, and assumptions about spatial styles of development. However, the relative importance of these factors depends on the magnitude of projected changes in total population and urbanization in each country and across SSPs [7].

The data required for the Visitation law is population density and distance between different grids, so the population distributions with a spatial resolution of five SSPs can be taken as the input for travel demand projections. The prediction results would show the patterns of future travel demand, and the difference between five development pathways, which contribute to better guidance of GHG emissions mitigation.

1.2 Thesis objectives

To achieve emissions reductions, we need a model that works well at the national level for better policy planning, because long-distance travel between countries accounts for a large share of total carbon emissions. However, the visitation model has only been tested at the city level at present [4]. Whether it is suitable for flows generation on a larger scale, such as the national level, remains to be explored.

The primary goal of this thesis is to extend the latest approach of flows generation from urban-level application to country-level. We perform an experiment to explore the impacts on mobility given future scenarios of population growth and spatial relocation (e.g. urbanization). The specific goals of this thesis are the following:

- **Country-level travel demand estimates**

  Test the validity of a state-of-the-art mobility model, which has been tested at the city level [4] to the country level, using Sweden as a case study. Then, validate the synthesized travel demand against the other well-established data sources (travel surveys) to quantify its performance.

- **Experiment: Revealing future scenarios**

  Apply the fine-grained global population data, which have a spatial resolution of 0.4 km^2 in flows generation, to model the synthetic travel demand in Sweden over the years (till 2100) across different future scenarios (SSP1–5) and summarize the patterns.

Scenarios	Description
Sustainability (SSP 1)	It assumes the world achieves its development goals relatively well. Its salient features are more efficient global cooperation driven by international organizations and institutions, and significantly less resource intensity and dependence on fossil fuels.
Middle of the Road (SSP 2)	It depicts a world where typical trends of the last two decades would continue, with some progress towards achieving development goals, resource and energy intensity declining at historic rates, and fossil fuel dependence slowly declining.
Fragmentation (SSP 3)	It is a situation in which the world is divided into regional blocks, and there is almost no coordination between them. Not only have these countries failed to meet global development goals, but they have also made little progress in reducing resource intensity and fossil fuel dependence.
Inequality (SSP 4)	It describes a world in which there is high inequality within and between countries, and hence frequent social conflict and unrest. Most of the emissions shall be the responsibility of the relatively small wealthy global elite, and they will be able to reduce emissions at low cost, yet the greater poverty group contribution to emissions has almost no.
Conventional development (SSP5)	It is one in which the world focuses on self-interested market-driven development with economic growth-oriented solutions to social and economic problems. It is characterized by the achievement of human development goals, strong economic growth, rapid urbanization, highly engineered infrastructure, as well as highly managed ecosystems. At the same time, because people would have confidence in their ability to manage societies and ecosystems, they would not have to make concrete positive efforts to avoid potential global environmental impacts. The preference for rapid conventional development will make the energy system dominated by fossil fuels continually, leading to high emissions. SSP5 can also be called the scenario of fossil-fueled development.

Figure 1.1: Description of SSP (1-5).

2

Methods

2.1 Visitation law-based flow generation

2.1.1 Model description

The universal visitation law of human mobility (Equation 2.1) determines for each location i the set of unique users who visited the corresponding cell and grouped them according to the distance r of their home location and according to their visitation frequency f (number of days over a period T during which they visited for a minimum duration τ). To factor out the effects of area size, the resulting visitor counts are normalized, $N_i(r, f)$, by the area of their origin, giving $\rho_i(r, f) = N_i(r, f)/A(r)$, with $A(r) \approx 2\pi r \delta r$. The proportionality constant μ_i determines the magnitude of the flows and thus reflects the location-specific attractiveness. f is the visitation frequency, r is the travel distance, and scaling exponent $\eta \approx 2$. The form of visitation law is shown below.

$$\rho_i(r, f) = \frac{\mu_i}{(rf)^\eta} \tag{2.1}$$

This visitation law shows that $\rho_i(r, f)$ does not depend on r and f separately but on the single rescaled variable rf. The regularity of the visitation law is largely robust to site-specific conditions, including big changes in the surrounding population density or level of economic or infrastructure development. Due to the robustness of this method, we can use it to get relatively reliable predictions for future travel demand.

An application of the visitation law is flows generation. The parameter μ_j, which can reflect location attractiveness, is determined by population density based on the aforementioned law. Then μ_j can later be used to estimate the flows of the population. $\rho_{pop}(j)$ is the population density at location j, and r_j is the distance to the boundary of the location. Then the value of μ_j can be calculated from Equation 2.2.

$$\rho_{\mathrm{pop}}(j) \approx \int_{f_{\mathrm{home}}}^{\infty} \rho_j\left(r_j, f\right) \mathrm{d} f = \frac{\mu_j}{r_j^2} \frac{1}{f_{\mathrm{home}}} \tag{2.2}$$

Equation 2.2 is based on three assumptions: (1) individuals return back to their home location on a daily basis, which generates a local flow with a minimum frequency of $f_{home} \approx 1d^{-1}$. (2) The population density $\rho_{pop}(j)$ within the location's area of radius r_j is equal to the population density at its boundary. (3)$f \geq f_{min} \approx f_{home}$. Thus, the magnitude of the flows can be approximated as $\mu_j \approx \rho_{pop}(j) r_j^2 f_{home}$.

Then, the average daily number of trips V_{ij} of those individuals who live at an origin i and visit destination j can be calculated using Equation 2.3. Equation 2.4 can be used to calculate the total number of trips V_{ij}^{tot} which also included individuals living at j and returning home. Because our ground-truth data include the flows of returning home (more details about ground-truth data can be seen in Section Model experiment and evaluation), so during the following calculation work, we choose Equation 2.4.

$$V_{ij} \approx \mathcal{A}_i \int_{f_{\min}}^{f_{\max}} f \rho_j \, \mathrm{d}f = \mu_j \mathcal{A}_i / r_{ij}^2 \ln \left(f_{\max} / f_{\min} \right) \tag{2.3}$$

$$V_{ij}^{tot} = V_{ji}^{tot} = \left(\mu_j \mathcal{A}_i + \mu_i \mathcal{A}_j \right) / r_{ij}^2 \ln \left(f_{\max} / f_{\min} \right) \tag{2.4}$$

With A_i, A_j being the area of origin location i and the destination j, r_{ij} being the distance between the origin and destination. $f_{min} = 1/T$, where $T \gg 1d$ is the observation period and $f_{max} = 1d^{-1}$ (for $r \gg r_j$).

Given the above formula, we generate the flows of THE population between places following two steps. First, we define spatial zones. Because r_j is the distance to the boundary of location j, we should separate the area we studied into the regular grids with equally sized square cells, and in this way, the average radius would be more precise than that of irregular grids. We use $r = \sqrt{\frac{A}{\pi}}$ to calculate r of each grid. The smaller the grid size, the more spatially-accurate flows generation we can obtain.

Then, we evaluate μ_j representing zonal attractiveness so that the total daily number of trips V_{ij}^{tot} from any origin location to any destination can be calculated.

To summarize, the model takes the input of population density and distance between origin i and destination j, and generates an ODM quantifying the flows of the population between pre-defined spatial zones in an average day.

2.1.2 Hierarchical implementation

We first define spatial zones by dividing the study area into small grids of different resolutions, e.g., 1 km^2. The number of trips between 1 km^2 size grids from the model is spatially-accurate at the country level. However, this leads to calculating $\sim 10^{12}$ OD pairs (for Sweden), which is computationally intensive. Therefore, this thesis proposes a hierarchical implementation of flows generation.

Instead of estimating the flows from every grid to every other grid, we create big grids on top of small grids, e.g., 10×10 km^2. The overall flows estimation for the study area is done hierarchically. We first estimate flows between small grids within each big grid, and then calculate the flows between big grids. This hierarchical implementation reduces the computation load drastically (from $\sim 10^{12}$ to $\sim 10^7$), but at the cost of the accuracy of flows estimation. However, the number of population trips largely depends on the distance between the origins and the destinations. Most of the trips are short-distance trips. Therefore, we hypothesise that the accuracy drop is not salient, given the careful choice of grid sizes for both small and big grids.

Equation 2.4 calculates the total number of trips between different grids ($i \neq j$). When $i = j$, the r_{ij} would be 0. In other words, the equation does not calculate the number of trips inside each grid. Therefore the diagonal of ODM ($i = j$) would be 0. Using small grids based on Equation 2.4 to calculate the number of trips inside 10×10 km^2 big grids, so that the flows inside each big grid can be obtained.

$$V_{ij}^{10\times10}(i = j) = \sum v_{ab}^{1\times1} \tag{2.5}$$

Inside big 10×10 km^2 grids, $v_{ab}^{1\times1}$ is the total number of trips between the origin small 1 km^2 grids a to destination small 1 km^2 grids b. $v_{ab}^{1\times1}$ can be calculated using Equation 2.4, so that the number of trips inside each 10×10 km^2 grid, $V_{ij}^{10\times10}$ ($i = j$), is obtained. The total number of trips between different 10×10 km^2 grids $V_{ij}^{10\times10}$ ($i \neq j$) can be calculated using Equation 2.4.

The proposed implementation of flows generation can not get the number of trips between two 1 km^2 grids if they belong to different 10×10 km^2 grids. However, the proposed approach aims to strike a good balance between computation cost and accuracy. For example, if a person wants to go to Stockholm from Göteborg, the exact grids where he/she starts, and ends can be well approximated by the two city centres. Therefore, this hierarchical way of estimating flows is capable of depicting the travel demand of both short and long distances with correspondingly reasonable spatial resolutions.

2.2 Model experiment and evaluation

This thesis has a framework shown in Figure 2.1. It first calculates the distances between grids and then takes the population density to model ODMs of the entire Sweden. Then Equation 2.4 based on the Visitation law is used to calculate the total number of trips between different big grids ($i \neq j$) and the the number of trips between different small grids inside each big grid ($i = j$). After aggregation and normalisation, the obtained model output ODMs are compared with the ground-truth data. The similarity metric (SSI) is applied to evaluate the similarity between model output and the ground-truth data.

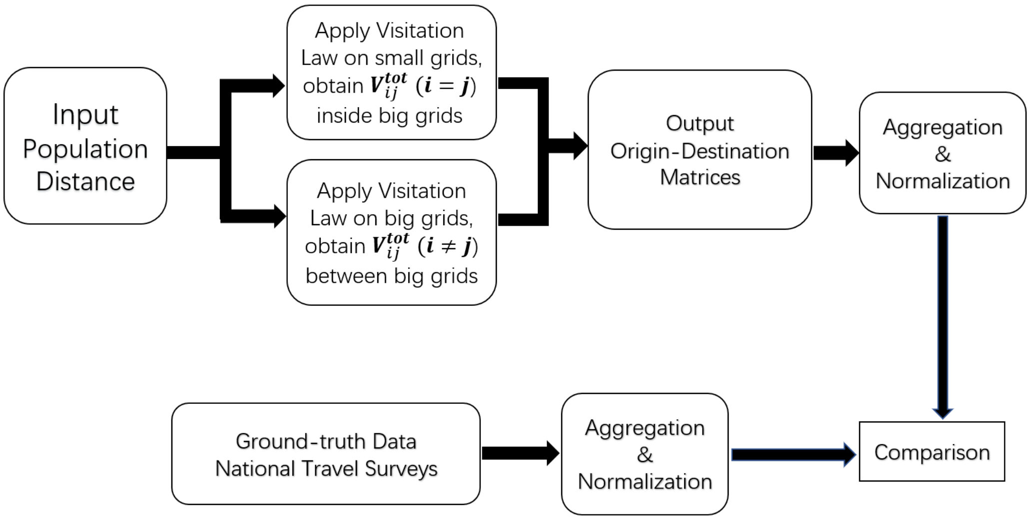


Figure 2.1: Flow chart describing the experimental design of the model and its evaluation.

2.2.1 Data Description

To evaluate the model, we use the travel survey data covering detailed trip information from Statistics Sweden (SCB), including the origin, destination and distance for individual trips, as ground-truth data [23].

A one-day travel diary from 2011 to 2016 was collected by the Swedish National Travel Survey and Collection (Swedish Official Statistics, 2016). 171,553 trips for 38,258 participants and 2,189 recorded days are included in this survey (submitted anonymously, 2021). This data set contains the origin and destination of the trip and the distance to travel. The spatial resolution is the DeSO region (Demographic Statistics Areas), defined by Statistics Sweden as 5,984 demographic regions. Thus, the travel survey data can provide the ODMs of the actual movements as the ground-truth data. The corresponding total number of trips for each origin DeSO zone and destination DeSO zone is shown in the ground-truth data.

Each DeSO zone is represented by a nine-bit code. The first four positions show which county and city the area belongs to, as it is made up of county and city codes. The fifth position is A letter: A, B or C, which divides DeSO into three different categories. Because some DeSO zones are very small, which are even smaller than the size of the grids, and some are located inside others, we need to aggregate DeSO zones into larger DeSO zones based on the first 4 or 5 digital numbers when we compare the ground-truth data with our model output.

Figure 2.2 shows the maps of Sweden and Västra Götalandsregionen (VG). Sweden, whose area is 528,447 km^2, has 25 provinces. VG, whose area is 25,247 km^2, is one of those provinces, and there are 992 DeSO zones in VG.

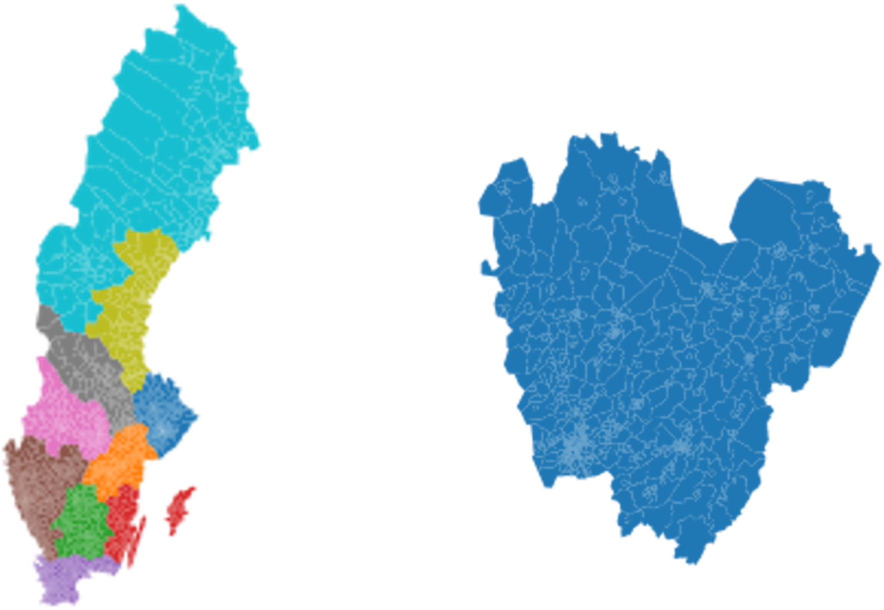


Figure 2.2: Map of Sweden and Västra Götalandsregionen (VG).

The thesis starts with VG as a study area to test the selection of grids size on the performance of model output. DeSO zones are aggregated based on the first five digital numbers. Figure 2.3 shows the spatial DeSO zones and aggregated DeSO zones of VG. There are 127 Aggregated DeSO zones in VG. For data input, we choose different grids size, which is 1 km^2,2×2 km^2,3×3 km^2,4×4 km^2,5×5 km^2 and 10×10 km^2, respectively.

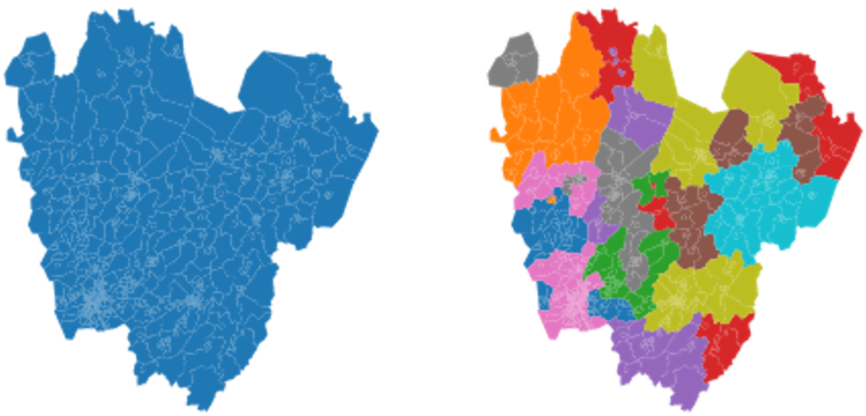


Figure 2.3: DeSO zones and aggregation DeSO zones of VG, Sweden.

For Sweden, we choose the big grids size of 10×10 km^2 and small grids sizes as

1×1 km^2. There are 5984 DeSO zones in Sweden, and we aggregate DeSO zones based on the first four digital numbers. Then we get 290 aggregated DeSO zones.

2.2.2 Aggregating flows for comparison

We aggregate the total number of trips between grids to the level of DeSO zones to compare with the ground-truth data so that the model output and ground-truth data have the same spatial system.

In Figure 2.4, we have A, B, and C aggregated DeSO zones and grids 1, 2, and 3. Which aggregated DeSO zone the grid is located in is decided by its centroid. If grids are in the same aggregated DeSO zone, they have no contribution to V_{ij}^{tot} $(i \neq j)$, but have a contribution to V_{ij}^{tot} $(i = j)$. For example, we should add the value v_{12}^{tot} and v_{21}^{tot} to V_{AA}^{tot}, but v_{12}^{tot} adds 0 to V_{AC}^{tot}. If grids are in the different aggregated DeSO zones, they have accumulated effect to V_{ij}^{tot} $(i \neq j)$. For example, the value of v_{23}^{tot} should be added to V_{AB}^{tot}.

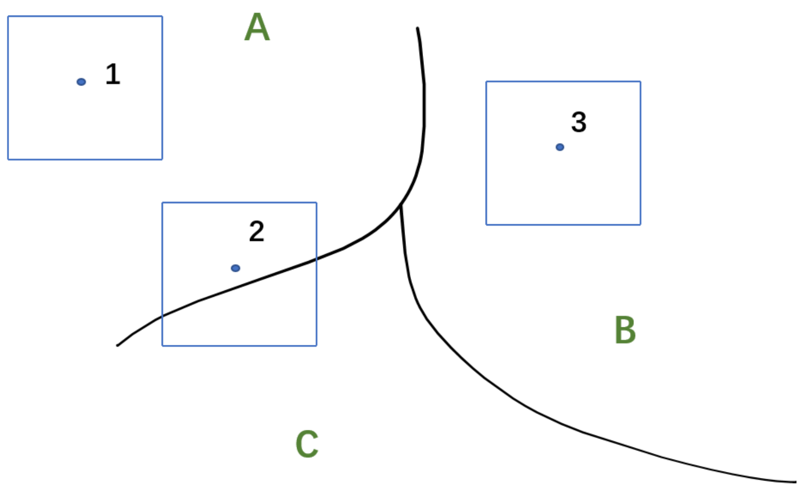


Figure 2.4: Model output aggregation: total number of trips (V_{ij}^{tot}) between aggregated DeSO zones.

We do a similar handling method for ground-truth data. In Figure 2.5, if DeSO zones are in the same aggregated DeSO zone, they have no contribution to V_{ij}^{tot} $(i \neq j)$, but have a contribution to V_{ij}^{tot} $(i = j)$. If DeSO zones are in the different aggregated DeSO zones, they have accumulated effect to V_{ij}^{tot} $(i \neq j)$. For example, $V_{EE}^{tot} = v_{AB}^{tot} + v_{BA}^{tot} + v_{AA}^{tot} + v_{BB}^{tot}$, $V_{EF}^{tot} = v_{AC}^{tot} + v_{AD}^{tot} + v_{BC}^{tot} + v_{BD}^{tot}$

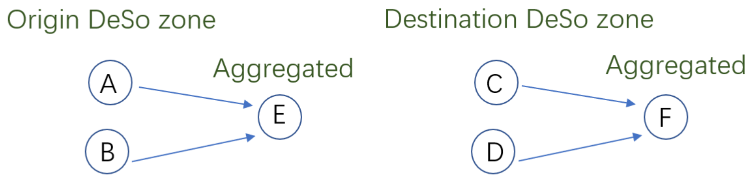


Figure 2.5: Ground-truth data aggregation: total number of trips (V_{ij}^{tot}) between aggregated DeSO zones.

2.2.3 Model performance evaluation: Similarity metrics

The total number of trips between aggregated DeSO zones for model output is compared with ground-truth data, where Sørensen-Dice Similarity Index (SSI) in ecology is applied to quantify the similarity between model output and the ground-truth data [24].

$$SSI = 2\sum_{i,j} min(V_{ij}^{model}, V_{ij}^{gt})/(\sum_{i,j} V_{ij}^{model} + \sum_{i,j} V_{ij}^{gt}) \tag{2.6}$$

Because the travel survey does not capture the total number of trips inside Sweden, we normalised the ground-truth data and our model output to compare them. SSI is a similarity index ranging from 0 to 1; 0 means the predicted results and the benchmark do not match, and 1 means they are identical.

Long-distance trips and short-distance trips contribute differently to travel demand, as well as greenhouse gas emissions and transportation policy making. Therefore, we also quantify the similarity between the visitation-law-based output and the ground-truth data regarding different distance groups. Specifically, we multiply the number of trips by its corresponding distance to get a distance-weighted SSI.

There are two types of distance: straight-line distance and travel distance, i.e., network distance which is closer to reality. The actual travel distance depends on many factors in reality and varies case by case. In this thesis, we roughly estimate it by deriving the distance ratio (network distance/straight-line distance) first using either simulation or empirical data.

Here we choose two ways to get the distance ratio, which are travel survey-based distance ratio, and simulation-based distance ratio. The former comes from travel surveys of Sweden and the Netherlands which contain participant-reported travel distance, and the trip distance is calculated as the straight-line distance between centroids of statistics zones, e.g., DeSO zones for Sweden. The latter uses simulation for ten selected urban regions whose area varies between 272.3 – 4878.7 in km^2. The simulation conducted creates a multiplier function of straight-line distance that determines the ratio of network distance over straight-line distance.

After having distance ratios, we can calculate straight-line distance, network distance, and corresponding distance-weighted SSIs. SSI_d stands for the similarity index weighted by straight-line distance. SSI_D_sy and SSI_D_sim represent the multiplying the number of trips by network distance based on travel survey and simulation, respectively.

Taking straight-line distance, for example, Figure 2.6 shows the way of multiplying the number of trips by distance, which is represented by T_{AB}. A and B stand for aggregated DeSO zones. i and j stand for 10 km^2 grids inside aggregated DeSO zones. a and b represent 1 km^2 grid inside each 10 km^2 grid.

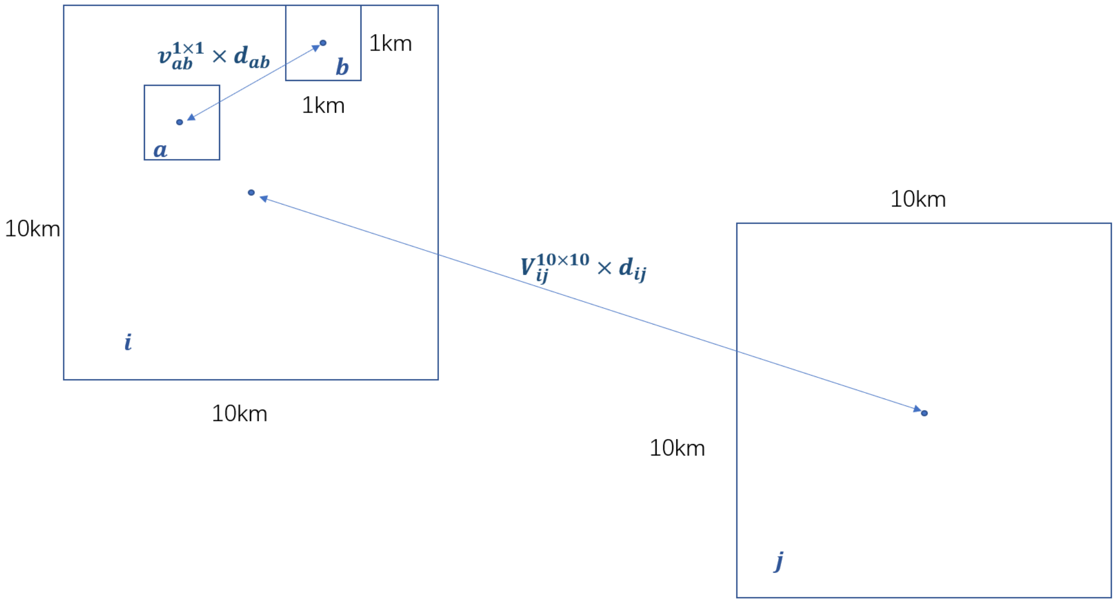


Figure 2.6: Multiplying V_{ij}^{tot} by straight-line distance.

$$T_{AB} = \sum V_{ij}^{10\times10} \times d_{ij}(i \neq j) + \sum V_{ij}^{10\times10} \times d_{ij}(i = j) \tag{2.7}$$

$$V_{ij}^{10\times10} \times d_{ij}(i = j) = \sum V_{ab}^{1\times1} \times d_{ab} \tag{2.8}$$

For ground-truth data, the way of multiplying the number of trips by distance is the same, and we also do the normalization for model output and ground-truth data. In total, we have four similarity metrics, SSI, SSI_d, SSI_D_sy, and SSI_D_sim, evaluating how similar the model-based ODMs are compared to the ground-truth data.

3

Results

In this chapter, Sections 3.1.1 and 3.1.2 discuss the model performance regarding grid size, and a proposed hierarchical implementation of the model in the Västra Götalands region (VG), respectively. Section 3.2 also shows the model performance for Sweden and explores the impacts of aggregation level and trip distance on the model performance. Finally, Section 3.3 presents the results of travel demand projection in VG for SSP (1-5) until the end of this century.

3.1 Impacts of model design

This section shows the results of the Västra Götalands region (VG) given the choice of grid sizes and hierarchical implementation using four different similarity metrics to evaluate their performance.

3.1.1 Impact of grid size on model performance

Figure 3.1 shows the results of the similarity metric in response to different grid sizes. It is clear that the larger the grids' size, the lower similarity. This is consistent with the hypothesis because the travel demand inside each grid would be ignored in the model.

In addition to the similarity metric SSI, shown in Figure 3.1, we notice that the values of distance-weighted SSIs are lower than the ones of SSIs. For distance-weighted SSIs, the longer distances the trips have, the higher their weight. But long-distance trips happen less frequently than short-distance trips in real life, and the number of long-distance trips captured in the ground-truth data is low due to the nature of a one-day travel diary. This may explain why distance-weighted SSIs are lower than that SSIs.

When comparing the two network distance-weighted SSIs with straight-line-distance-weighted SSIs, the network-distance ones outperform straight-line SSIs. This is because realistic travel distances are more capable of explaining the number of trips.

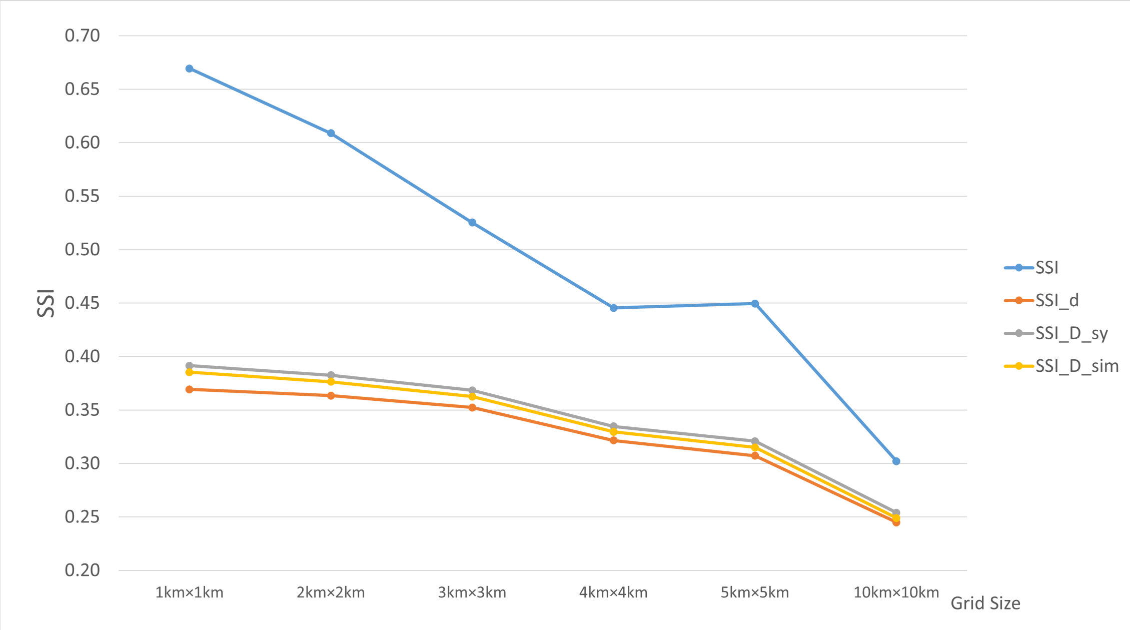


Figure 3.1: The similarity between the model-based ODMs and the ground-truth data in VG, Sweden.

3.1.2 Feasibility of hierarchical implementation

The hierarchical implementation is designed to improve computational efficiency, even if it means sacrificing some accuracy. Nevertheless, it is believed that the reduction in accuracy is minimal and acceptable. The test results are shown in Figure 3.2, where we see the number of trips for 1 km^2 and 5×5 km^2 with 1 km^2 in comparison with the ground-truth data. The SSI results are similar, suggesting the feasibility of the proposed hierarchical implementation. In other words, we avoid dividing the entire study area into small grids, which leads to high computation costs, but still, we can achieve relatively accurate model output.

Grid Size	SSI	SSI_d	SSI_D_sy	SSI_D_sim
1km×1km	0.6694	0.3693	0.3915	0.3853
2km×2km	0.6088	0.3635	0.3825	0.3764
3km×3km	0.5253	0.3523	0.3683	0.3625
4km×4km	0.4455	0.3215	0.3347	0.3296
5km×5km	0.4496	0.3072	0.3208	0.3149
10km×10km	0.3022	0.2449	0.2540	0.2490
5km×5km with 1km×1km	0.6659	0.3350	0.3564	0.3501
10km×10km with 1km×1km	0.6271	0.2787	0.2994	0.2932

Figure 3.3: Model performance of different grid sizes vs hierarchical implementation.

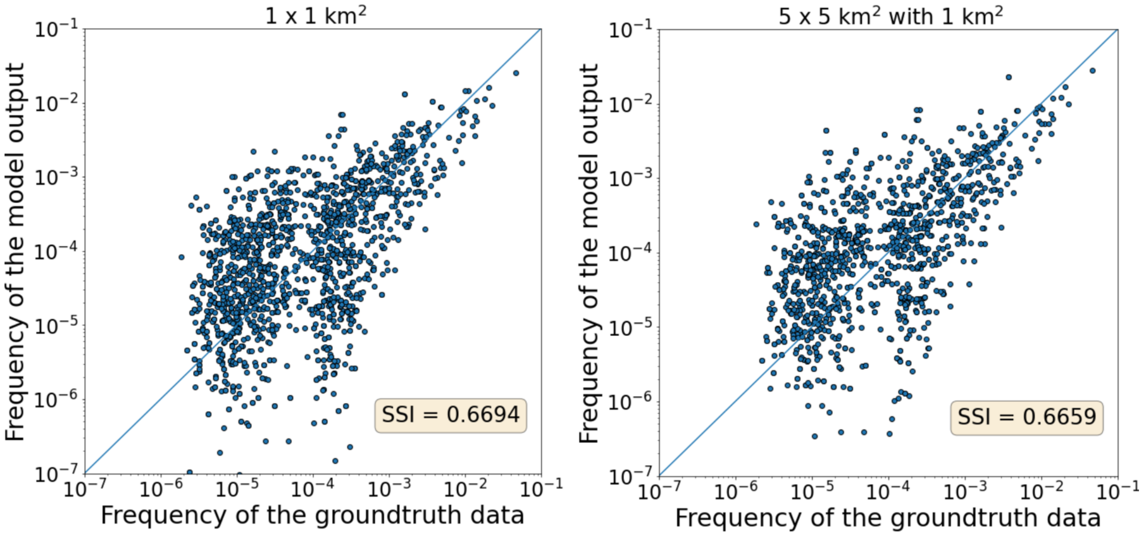


Figure 3.2: Performance of 1 km^2 vs a hierarchical implementation with 5×5 km^2 plus 1 km^2.

Figure 3.3 summarises the model performance quantified by SSI and distance-weighted SSIs, given different selections of grid sizes. The bottom two lines indicate just a slight decrease in model performance when we use the hierarchical implementation of flows generation, compared with using 1 km^2. However, the selection of a greater grid size of big grids in the proposed method leads to a greater drop in performance, especially for distance-weighted SSIs. The error introduced for the grids size of 10×10 km^2 may not be small for the VG region. We should choose a reasonable grid size concerning the size of the study area.

Balancing the cost of performance and computation, the proposed hierarchical implementation can obtain almost the same accuracy level as 1 km^2. Therefore we apply it to the entire of Sweden.

3.2 Model Performance in Sweden

For Sweden, comparing the model output with the ground-truth data under the spatial system of 10×10 km^2 plus 1 km^2, SSI is 0.7558, which is comparable with the model performance of selected cities being 0.70 in the original study proposing the visitation law [4]. Thus, the visitation law and the hierarchical implementation proposed by this thesis at the country level are feasible. However, it is worth noting that the ground-truth data in this thesis has limited spatial resolution compared with the original study, and the aggregation level affects the evaluation results.

3.2.1 Impact of aggregation levels

Different aggregation levels lead to varying results of SSI and distance-weighted SSI. Because some grids are located on the boundary of aggregated DeSO zones, some error is introduced if we allocate the grids to which aggregated DeSO zones according to their centroid locations.

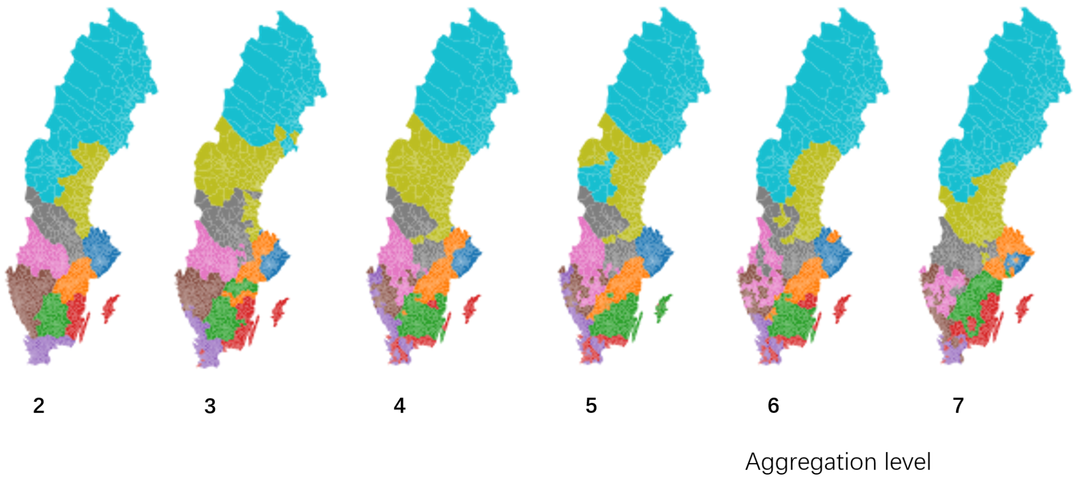


Figure 3.4: Visualization of different aggregation levels. An aggregation level X means the same number of the first X digits in the DeSO zone code being considered as one region.

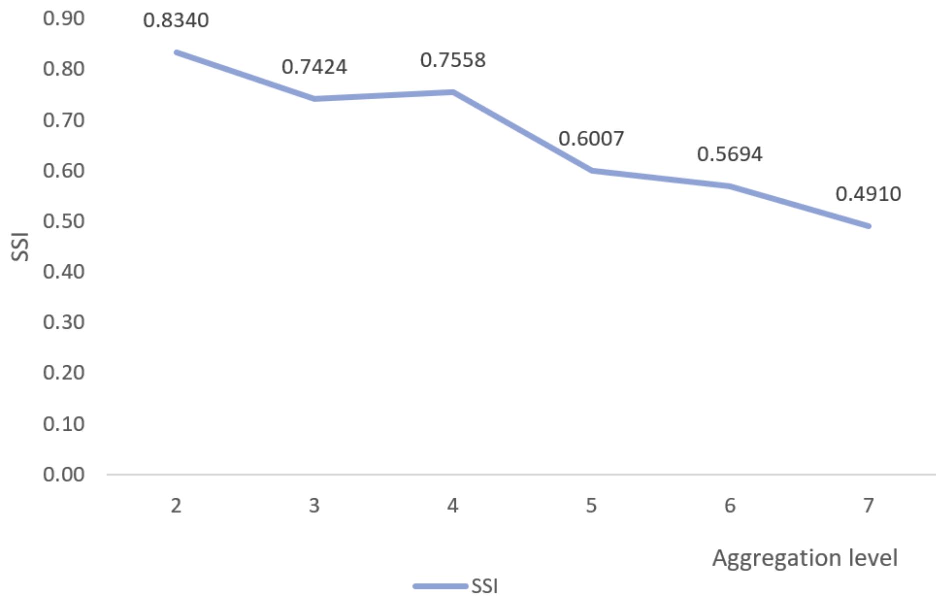


Figure 3.5: SSI of different aggregation levels.

The DeSO zones are aggregated based on their first few digits. The fewer digits we take, the larger size of aggregated DeSO zones and the fewer grids on the boundary we should take care of. Figure 3.4 shows the varying distribution of aggregated DeSO zones of different aggregation levels. In Figure 3.5, X-axis shows the level of aggregation; for example, number 2 means the DeSO zones are aggregated based on their first two digits. The similarity index decreases when the aggregated DeSO zones become smaller. We continue with the first four digital numbers as our aggregation level for the entire Sweden to compare with the ground-truth data.

3.2.2 Impact of distance

Long-distance trips and short-distance trips have different contributions to transportation emissions. Thus, this section explores the model performance on trips of varying distances. First, we rank the number of trips of all OD pairs based on their distances and separate them into ten distance groups based on the ten percentile groups ranging between 10% and 100%. Then we calculate SSI for the ten distance groups respectively.

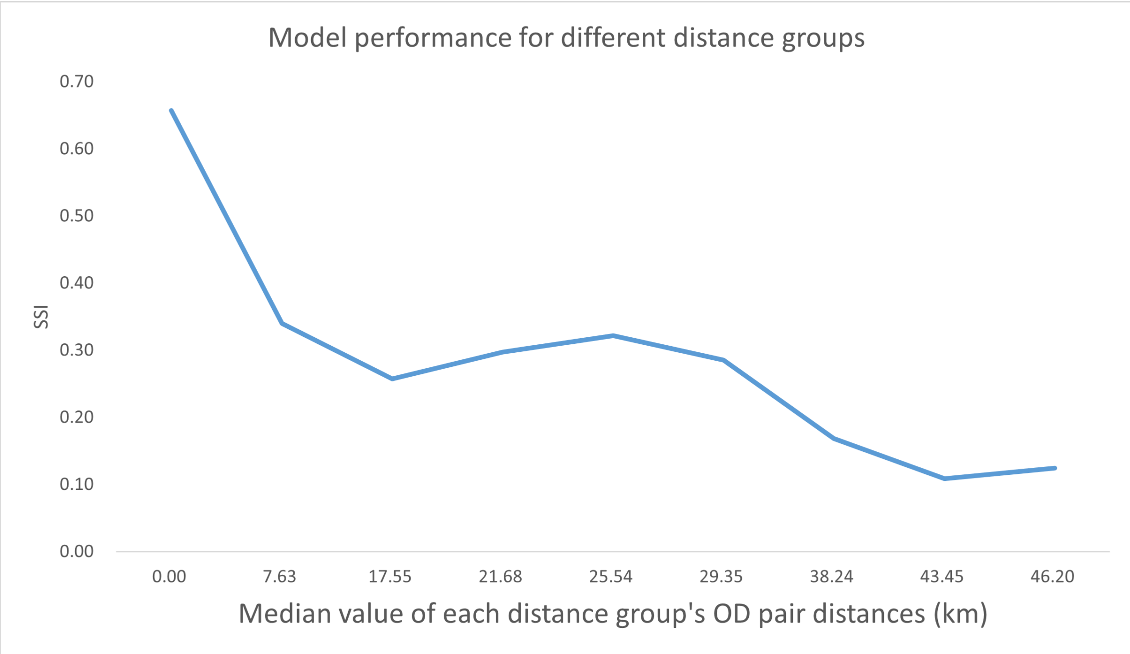


Figure 3.6: SSI and distance-weighted SSIs as a function of distance.

Figure 3.6 shows that the visitation law-based model performs much better for short-distance trips than long-distance trips. The general decreasing trend of SSI with increasing distance is consistent with the hypothesis mentioned above that long-distance trips are rare and can not be captured by ground-truth data adequately.

3.3 Model experiment on travel demand projection in Västra Götalandsregionen

After verifying the feasibility of the visitation law-based model, we use it to predict future travel demand in Västra Götalandsregionen. The fine-grained global prediction data of population, which are consistent with Shared Socioeconomic Pathways

(SSPs), are used to synthesise travel demand until 2100. The SSPs are Sustainability (SSP1), Middle of the road (SSP2), Regional rivalry (SSP3), Inequality (SSP4), and Fossil-fueled development (SSP5).

3.3.1 Total population forecast

Figure 3.7 shows the population development in VG. While the population of SSP5 continues to increase by the end of this century, the others gradually level off or even decrease, consistent with the characteristics of the corresponding scenarios.

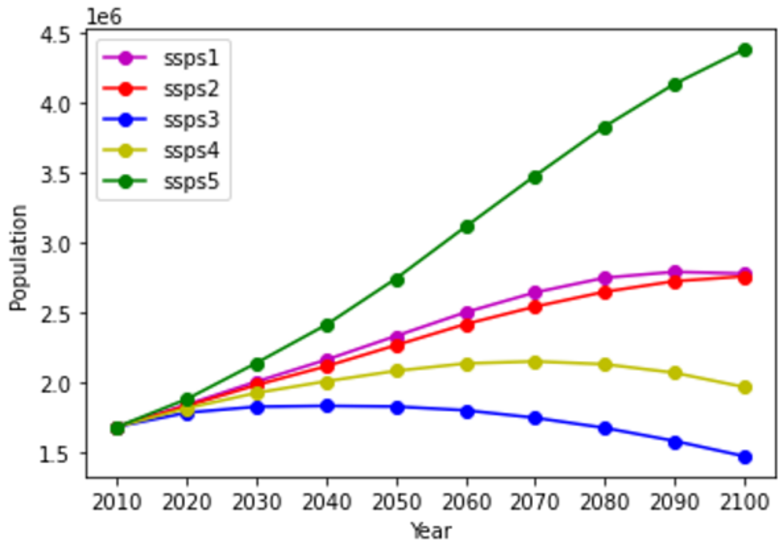


Figure 3.7: Population of SSP(1-5) in VG, Sweden.

SSP1 (Sustainability) and SSP5 (Fossil-fueled development) both desire to promote a relatively rapid demographic transition through higher investment in education, health, and relatively high-income growth. In countries with low fertility rates, such as Sweden [25], where VG is located, optimism about the economic outlook has kept fertility rates at medium (SSP1) or even high (SSP5). And the numbers of immigrants in both scenarios are huge [7]. As a result, the population of VG continues to increase in SSP5, and also shows an overall rising trend in SSP1. Compared to other scenarios, SSP2 represents a middle-of-the-road outcome with moderate population density and variation [7], which is also consistent with the population of VG shown in Figure 3.7. Both SSP3 (Regional rivalry) and SSP4 (Inequality), especially SSP3, will lead to relatively low fertility and low population growth (or even decline) in currently low-fertility countries, as well as relatively low numbers of migrants [7]. This is consistent with the results shown in Figure 3.7. Sweden, where VG is located, already has a low fertility rate especially compared to developing countries all around the world [25], so the population shows a downward trend in both SSP3 and SSP4 scenarios, and SSP3 has the largest decline in population among all the scenarios.

3.3.2 Travel demand projection

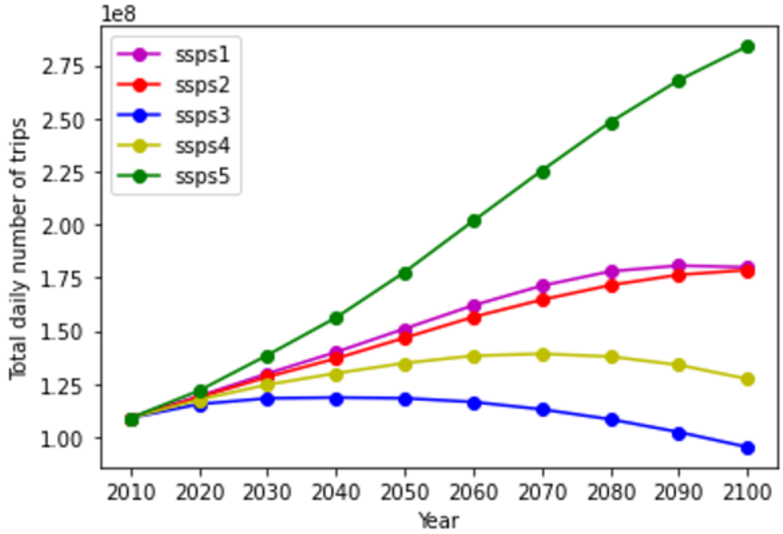
The total daily number of trips in VG for five scenarios suggests that the trends are almost identical to that of population because the visitation law is mainly based on population density (Figure 3.8 (a)). Except for SSP3 and SSP4, the total daily number of trips increase 75 - 175% between 2010 and 2100, suggesting demand increase and therefore, intelligent and rational transportation planning and infrastructure are necessary.

Figure 3.8 (b) and Figure 3.8 (c) represent the daily number of trips per capita and average trip distance of five scenarios in VG. The corresponding value for different years and SSP has no significant change from the figures. The value of the daily number of trips per capita shown in Figure 3.8 (b) doesn't make sense for real life, which is obviously much larger. In addition to the numbers of returning home included in the model output should be subtracted, the parameters set by assumptions in the Visitation law model may need to be adjusted, like f_{min} and f_{max}.

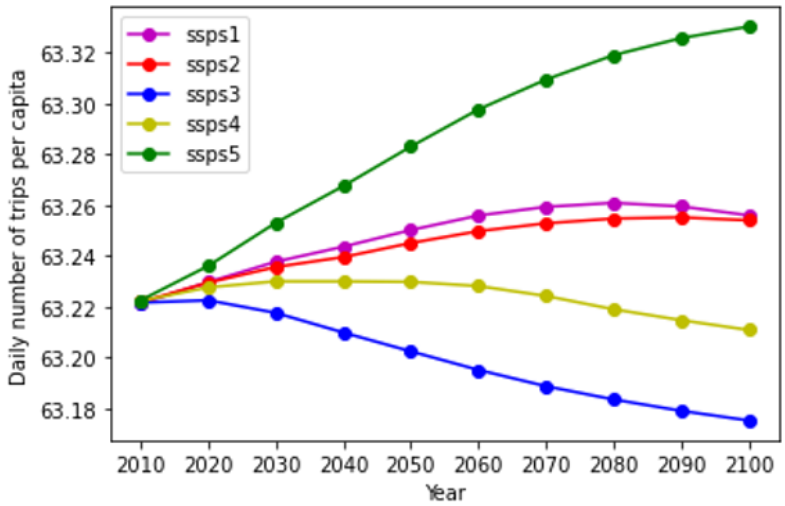
It seems that the proposed model fails to capture the variation of travel patterns corresponding to the distinct population scenarios over time. Specifically, the computed average trip distance and trip number per capita are not sensitive to different spatial distributions of the population of SSPs or over time. This thesis had this null hypothesis that the proposed model, taking the grid population of different SSPs as input, captures the impact of population distribution on travel patterns. However, the results reject the hypothesis.

One speculation is that the grid population of SSPs does not have a high enough resolution to generate a 1 km grid population. In other words, there was a disaggregating process, which resulted in low data quality. As a consequence, the subtle reallocation of the whole population is not able to be captured by the visitation law-based model. Another speculation is that the visitation law-based model is static and abstract, representing the overall strength of the connection between zones dominated by population size, among many other factors affecting zones' attractiveness. The sole use of population density for long-term travel demand projection is not sufficient to get reasonable results.

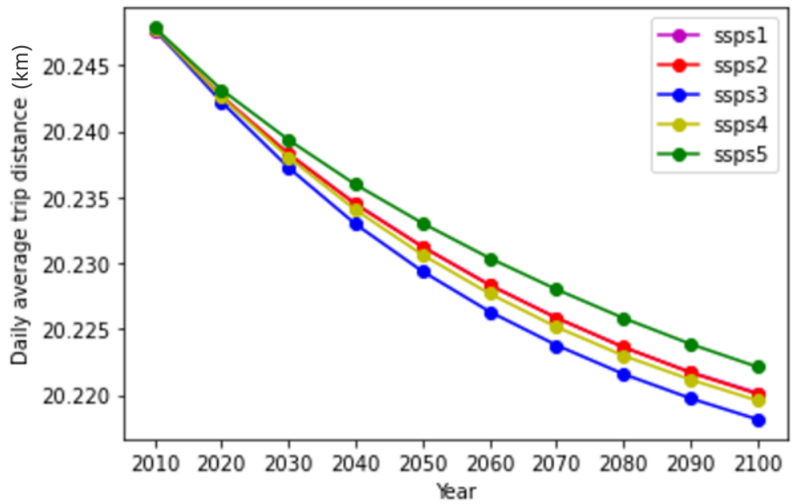
The Visitation law provides the objective rule of the relationship between the number of trips, distance, and frequency, which is applicable to different regions and times to a certain extent. Therefore, this approach can be applied to flow generation in a certain period of time in a region, from the city level to the country level, and the model output can be compared well with the ground-truth data. However, for the travel demand forecast of different conditions in the same region, in addition to the high-resolution population distribution as the data input, more influential factors about the surrounding cases should be added into the model as adjustment parameters, such as economic situations or infrastructure conditions. Population density alone cannot capture enough detailed data. Moreover, the low resolution used in this paper is equivalent to the regional aggregation of the model output, which will make the slight difference in travel demand between each small grid more difficult to detect, and more elaborate data will be lost. Higher spatial resolution or larger study areas maybe improve the current status.



(a)



(b)



(c)

Figure 3.8: Travel demand projection of SSP(1-5) in VG, Sweden. Notice the extremely small range in the Y-axis for (b) and (c).

4

Discussion

This thesis aims to extend the latest approach of flow generation from urban-level application to country-level and experiment on long-term travel demand projection given future scenarios of population growth and spatial relocation (e.g. urbanization). This thesis first tests the performance of the visitation law-based model in VG and shows the impact of grid size on SSI and distance-weighted SSIs, quantifying the performance of the model in comparison with the ground-truth data. Then we develop a hierarchical implementation of the model, which achieves a good balance between computation cost and accuracy in VG. Based on the model design, we continue to apply this method to the entire of Sweden and explore the impact of aggregation levels and distance on the model performance. Finally, we perform a model experiment in Västra Götalandsregionen to estimate travel demand for different future scenarios (SSP1-SSP5) from 2010 to 2100.

4.1 Implementing flows generation based on visitation law

This thesis proposed a hierarchical implementation of flow generation, accounting for the trade-offs between computation cost and accuracy. From Figure 3.1 and Figure 3.3 in Section Results, it is clear that the smaller the size of the grid we choose, the higher the accuracy of the model results would be. The grid size of 1×1 km^2 makes the similarity SSI as high as 0.6694, which is a desirable outcome. Larger grids size can not capture enough travel demand because more trips inside each grid would be ignored, which makes the precision decrease. In fact, if we choose the grid size of 500×500 m^2, the results of SSI would be higher. However, small grids would lead to a geometric increase in computer computation load, which results in a high requirement on the hardware, and consumes a lot of time. Therefore, we develop a hierarchical implementation with 5×5 km^2 plus 1 km^2, which can calculate the number of trips inside 5×5 km^2 grids. In this way, the similarity SSI has been increased from 0.4496 to 0.6659 compared with the original method of using 5×5 km^2 grids only. Though it's still a little less than the results of using a 1×1 km^2 grid size, we consider this an acceptable sacrifice of accuracy because the computation cost slashes. There must be a trade-off between computation load and accuracy here.

The solution is not always satisfactory, so we should choose the bigger grid size and smaller grid size inside carefully based on the size of the area we research and the outcome we want to achieve. For example, from Figure 3.3, the hierarchical implementation with 10×10 km^2 plus 1 km^2 is obviously not suitable for VG, because the area of VG is 25247 km^2, and there are only about 252 grids of 10×10 km^2 in VG, which would make the calculation too rough and the travel demand between different grids deviates greatly from reality. In contrast, when we choose a grid size of 10×10 km^2 plus 1 km^2 for the entire Sweden, the SSI can reach 0.7558, which means this choice of grid size is quite appropriate for Sweden. Therefore, the area we study should be large enough with respect to the bigger grid size while adopting the hierarchical implementation.

In this thesis, the smaller grid size we apply in the hierarchical implementation is always 1 km^2, but it's sure that if we choose 500×500 m^2 as the smaller grid size inside the bigger grid size, the similarity SSI would be higher. But it would increase the computation load much more, and existing results of 1 km^2 have already achieved a satisfactory degree. There is no need to expend that much computing load in pursuit of a very limited SSI improvement. Certainly, if the area we study is very small, choosing the smaller grid size of 500×500 m^2 is necessary. Even more, elaborate calculations would also be needed for acceptable accuracy.

4.2 Potentials and limitations for visitation law-based model

The similarity metric SSI of this approach based on the Visitation law achieves 0.7558 at the country level if DeSO zones in Sweden are aggregated by the first four digits, which means the results are magnitude-wise reasonable, and the hierarchical implementation is feasible. However, the similarity index decreases with a more elaborate degree of aggregation, which shows this approach can not capture detailed data with sufficient accuracy in ODMs when the aggregated DeSO zones become smaller. In addition to this being a possible limit of the approach, limited spatial resolution in this thesis may also contribute to this result to some extent. Higher resolution would improve the possibility of capturing more detailed generation flows.

The result of the predicted total daily number of trips reveals a significant difference between different future scenarios, and the trends are consistent with their respective population's growth. This phenomenon is reasonable for the formula of visitation law because the core parameters of visitation law are population density and the distance from i to j. For different scenarios, the distance between each zono keeps the same, so the only core distinction is the population density. The results in Sections 3.1, 3.2 and 3.3.2 show that this approach based on the visitation law is capable of making the model output of the total daily number of trips between grids in an area magnitude-wise reasonable after a certain degree of aggregation.

However, the results of the daily number of trips per capita and average trip distance can't reflect an obvious distinction between different future scenarios of population growth and spatial relocation, which shows that this Visitation law-based model is not precise enough to capture the variation of travel demand patterns corresponding to the different population distributions of SSPs or over time.

There are two possible reasons for this situation. One is that the spatial resolution of the grid population for SSP (1-5) is not sufficient to generate a population distribution of 1 km grids. Low data quality prevents the Visitation law-based model from capturing the subtle redistribution of the entire population. If the existing resolution is used, but a larger study area is selected, such as the national level, the relative area of the study region and the grids will be larger. Then the situation may be improved, because changes in the reallocation of the population may be captured. The other is caused by some ignored elements like heterogeneity and spatial scale in the model of visitation law. There would always be a dilemma that simplicity and diversity in human mobility models couldn't be achieved at the same time. There are still many other factors that affect the attractiveness of a region. Thus it is unreasonable to use population density as a single variable to forecast long-term travel demand.

The visitation law obviously overestimates the flow generation per capita in SSPs. The parameters in the formula, like f_{min} and f_{max} should be amended according to the different assumptions set for SSP (1-5), respectively. In short, the model experiment on travel demand projection shows the limitations of the current Visitation law-based model, which is not suitable for projecting the travel demand related to per capita and the average for different scenarios of the same region.

4.3 Future work

Given the limitations of the Visitation-law model, methods to improve the application of this model could be further explored in the future. One way is to add adjusted parameters to reproduce and predict the flow generation of cities or countries with different sizes, economic levels, and cultural backgrounds. These aspects can also be used to improve the representation of place attractiveness beyond the simple population density. Here are a few factors that may have impacts on the results:

GDP. There is the "iron law" of coupling between GDP per capita and travel demand, which has been a constant in recent economic history [26]. The travel demand would increase with GDP, so GDP is a key factor that should be considered in the model.

Speed & Convenience in transport systems. The supply of transport and the technology capabilities can affect the travel speed, and the policy setting in the transport system involves the supply of transport, influencing the convenience of travel. These will all lead to people's willingness to travel.

Transportation expense & Income. The smaller the cost of transport relative to people's income, the more willing people are to go out. Also, people's expectation of social progress also affects their willingness to travel. If they are optimistic about economic development, they are more willing to travel. In contrast, if they have a pessimistic attitude, they will choose to reduce travel demand in order to save money.

Seasonality. There are also periodic changes in travel demand. In the warm summer, especially during the festivals and holidays, the number of trips would be higher than in cold winters, because of more outdoor activities and increased willingness to travel.

In addition to appending adjusted parameters to improve the model, more work on the choice of grids size should be continued. α is the ratio of the bigger grid size to the smaller grid size in the method of hierarchical implementation, and β is the ratio of the area we study to the bigger grid size. The higher α and β, then the better accuracy is. We need to further apply the model to more areas so that the selection of these parameters can strike a good balance between computational efficiency and performance, which suits different research purposes.

Besides, the model output of the other models, like the Gravity model and the Radiation model, should be calculated in the future and be compared with the ground-truth data of Sweden, which we applied in this thesis. Then the corresponding similarity metrics SSI should be compared with the SSI of the Visitation law-based model to more thoroughly reveal the capability of this newly proposed approach.

5

Conclusion

Human mobility is very important for the prediction of traffic, which contributes a crucial role in the reduction of carbon emissions and sustainable development. However, current mobility models, such as the Gravity model and the Radiation model, mostly have some drawbacks, like only focusing on the spatial dependence of mobility flows but not capturing the varying frequencies of repeated visits to the same location. A recently proposed model of human mobility, "Visitation law", which takes into account the space-time spectrum of migratory fluxes, finds that the number of visitors to any given location is inversely proportional to the square of the distance travelled and the frequency of visits. The input required in this method, which is population density and distance, is relatively easy to obtain, and the results of this new method have been shown to be feasible in several metropolitan areas. But for larger areas, such as at the national level, the model's effectiveness has not yet been tested. This is necessary because long-distance trips lead to a large share of GHG emissions. Besides, long-term projections of travel demand are also important for transport planning and future assessment of climate change.

In this thesis, we apply the visitation law in Västra Götalandsregionen to validate the visitation law, and find that the accuracy increases as the size of the grids decreases. Smaller grids size could achieve more accurate results, but that would lead to a geometric growth of computation cost, constraining its feasibility to scale up to larger areas. Based on the characteristics of the visitation law, we developed a hierarchical implementation, i.e., combining small grids and big parental grids, which proved to have a good balance between the cost of model performance and computational cost.

After the model design, we apply the hierarchical implementation of the model to the entire Sweden. We find that the selection of small and big grids size and aggregation levels should be treated carefully concerning the relative size of the study area. The similarity metric SSI of this approach would decrease with smaller aggregated DeSO zones. In addition, this model performs differently for short and long-distance trips. Since long-distance trips are rarer and not easily captured by ground-truth data, SSI shows a tendency to decrease with the increase in distance.

Finally, we perform the model experiment to predict how the travel demand will evolve in Sweden until 2100. A new set of global, spatially explicit population scenarios is taken as the input, which is consistent with the Shared Socio-economic

Pathways (SSPs) developed to support research on global change. Combining the Visitation law and the fine-grained global population distribution data can model the synthetic travel demand over the years (till 2100) across different future scenarios (SSP1–5). The prediction results of total trip numbers show an obvious difference across SSPs and over time, which is influenced by the total population variation. But for the computed average trip distance and trip number per capita, there is no obvious difference, which means this current approach has no sensitivity to different spatial distributions of population density.

From the performance and experiments of the Visitation law-based model, in addition to showing the magnitude-wise feasibility of this approach, some shortcomings have also been exposed, which is unable to capture the variation of travel demand patterns for different scenarios over time. And a high spatial resolution of population distribution is required for travel demand projection. Hence, more work, like introducing adjusted parameters, should be done in the future to modulate the method. The performance of this approach should also be compared with that of other models, such as the Gravity model and the Radiation model, to reveal the capabilities of this newly proposed approach more comprehensively.

Bibliography

[1] Filippo Simini, Gianni Barlacchi, Massimilano Luca, and Luca Pappalardo. A deep gravity model for mobility flows generation. *Nature communications*, 12(1):1–13, 2021.

[2] Filippo Simini, Marta C. González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, Apr 2012.

[3] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.

[4] Markus Schläpfer, Lei Dong, Kevin O'Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B West. The universal visitation law of human mobility. *Nature*, 593(7860):522–527, 2021.

[5] Yuan Liao. Ride-sourcing compared to its public-transit alternative using big trip data. *Journal of Transport Geography*, 95:103135, 2021.

[6] Kristie L. Ebi, Tom Kram, Detlef P. van Vuuren, Brian C. O'Neill, and Elmar Kriegler. A new toolkit for developing scenarios for climate change research and policy analysis. *Environment: Science and Policy for Sustainable Development*, 56(2):6–16, 2014.

[7] Bryan Jones and Brian C O'Neill. Spatially explicit global population scenarios consistent with the shared socioeconomic pathways. *Environmental Research Letters*, 11(8):084003, 2016.

[8] Yuan Liao. *Understanding Mobility and Transport Modal Disparities Using Emerging Data Sources: Modelling Potentials and Limitations.* Chalmers Tekniska Hogskola (Sweden), 2021.

[9] Yuan Liao, Sonia Yeh, and Jorge Gil. Feasibility of estimating travel demand using geolocations of social media data. *Transportation*, 49(1):137–161, Feb 2022.

[10] Fan Yang, Peter J. Jin, Yang Cheng, Jian Zhang, and Bin Ran. Origin-destination estimation for non-commuting trips using location-based social networking data. *International Journal of Sustainable Transportation*, 9(8):551–564, 2015.

[11] Yan Shi, Haoran Feng, Xiongfei Geng, Xingui Tang, and Yongcai Wang. A survey of hybrid deep learning methods for traffic flow prediction. ICAIP 2019, page 133–138, New York, NY, USA, 2020. Association for Computing Machinery.

[12] Rafael Prieto Curiel, Luca Pappalardo, Lorenzo Gabrielli, and Steven Richard Bishop. Gravity and scaling laws of city to city migration. *PLOS ONE*, 13(7):1–19, 07 2018.

[13] Alina Sîrbu, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Marco Conti, Fosca Giannotti, Riccardo Guidotti, Simone Bertoli, Jisu Kim, Cristina Ioana Muntean, Luca Pappalardo, Andrea Passarella, Dino Pedreschi, Laura Pollacci, Francesca Pratesi, and Rajesh Sharma. Human migration: the big data perspective. *International Journal of Data Science and Analytics*, 11(4):341–360, May 2021.

[14] Chaoming Song, Zehui Qu, Nicholas Blumm, and Albert-László Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[15] N. W. Ruktanonchai, J. R. Floyd, S. Lai, C. W. Ruktanonchai, A. Sadilek, P. Rente-Lourenco, X. Ben, A. Carioli, J. Gwinn, J. E. Steele, O. Prosper, A. Schneider, A. Oplinger, P. Eastham, and A. J. Tatem. Assessing the impact of coordinated covid-19 exit strategies across europe. *Science*, 369(6510):1465–1470, 2020.

[16] Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Guido Caldarelli, Paolo Cintia, Stefano Cresci, Angelo Facchini, Fosca Giannotti, Aristides Gionis, Riccardo Guidotti, Michael Mathioudakis, Cristina Ioana Muntean, Luca Pappalardo, Dino Pedreschi, Evangelos Pournaras, Francesca Pratesi, Maurizio Tesconi, and Roberto Trasarti. (so) big data and the transformation of the city. *International Journal of Data Science and Analytics*, 11(4):311–340, May 2021.

[17] Chiawei Hsu, Chao Fan, and Ali Mostafavi. Limitations of gravity models in predicting fine-scale spatial-temporal urban mobility networks. *arXiv preprint arXiv:2109.03873*, 2021.

[18] Chaogui Kang, Yu Liu, Diansheng Guo, and Kun Qin. A generalized radiation model for human mobility: spatial scale, searching direction and trip constraint. *PloS one*, 10(11):e0143500, 2015.

[19] A Paolo Masucci, Joan Serras, Anders Johansson, and Michael Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Physical Review E*, 88(2):022812, 2013.

[20] Xiao-Yong Yan, Chen Zhao, Ying Fan, Zengru Di, and Wen-Xu Wang. Universal predictability of mobility patterns in cities. *Journal of The Royal Society Interface*, 11(100):20140834, 2014.

[21] Brian O'Neill, Elmar Kriegler, Kristie Ebi, Eric Kemp-Benedict, Keywan Riahi, Dale Rothman, Bas van Ruijven, Detlef Vuuren, Joern Birkmann, Kasper Kok, Marc Levy, and William Solecki. The roads ahead: Narratives for shared socioeconomic pathways describing world futures in the 21st century. *Global Environmental Change*, 42, 02 2015.

[22] Samir KC and Wolfgang Lutz. The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100. *Global Environmental Change*, 42:181–192, 2017.

[23] Yuan Liao, Kristoffer Ek, Eric Wennerberg, Sonia Yeh, and Jorge Gil. A mobility model for synthetic travel demand from sparse traces. *IEEE Open Journal of Intelligent Transportation Systems*, 3:665–678, 2022.

[24] A. Paolo Masucci, Joan Serras, Anders Johansson, and Michael Batty. Gravity versus radiation models: On the importance of scale and heterogeneity in commuting flows. *Phys. Rev. E*, 88:022812, Aug 2013.

[25] Max Roser. Fertility rate. *Our World in Data*, 2014. https://ourworldindata.org/fertility-rate.

[26] Yves Crozet et al. *The prospects for inter-urban travel demand*, volume 14. OECD Publishing, 2009.

A

Notations

Notation	Definition
p	Individual index
i	Index of origin location
j	Index of distinction location
r	Travel distance
f	Visitation frequency
$N_i(r, f)$	Visitor counts
$\rho_i(r, f)$	Normalized visitor counts
$A(r)$	Area size
μ_i	The magnitude of the flows, reflecting the location-specific attractiveness
η	Scaling exponent
r_j	Distance to the boundary of the location j
$\rho_{pop}(j)$	Population density at location j
f_{home}	Frequency of returning home
f_{min}	Minimum frequency
f_{max}	Maximum frequency
A_i	Area of origin location i
A_j	Area of destination location j
T	Observation period
V_{ij}	Average daily number of trips of individuals who live at origin i and visit destination j
V_{ij}^{tot}	Average total daily number of trips of individuals who live at an origin i and visit destination j, and individuals living at j and returning home
T_{AB}	Distance weighted number of trips, which means the multiplying the number of trips by distance
α	The ratio of the bigger grid size to the smaller grid size in the method of hierarchical implementation
β	The ratio of the study area to the bigger grid size

Table A.1: Lookup table with the main symbols and relevant notations used in this thesis.