Deep Gravity: enhancing mobility flows generation with deep neural networks and geographic information

FILIPPO SIMINI, University of Bristol, Department of Engineering Mathematics, and The Alan Turing Institute, London, UK

GIANNI BARLACCHI*, Amazon Alexa, Germany

MASSIMILIANO LUCA, Fondazione Bruno Kessler, Italy and Free University of Bolzano, Italy

LUCA PAPPALARDO, Institute of Information Science and Technologies (ISTI), National Research Council of Italy (CNR), Italy

The movements of individuals within and among cities influence key aspects of our society, such as the objective and subjective well-being, the diffusion of innovations, the spreading of epidemics, and the quality of the environment. For this reason, there is increasing interest around the challenging problem of flow generation, which consists in generating the flows between a set of geographic locations, given the characteristics of the locations and without any information about the real flows. Existing solutions to flow generation are mainly based on mechanistic approaches, such as the gravity model and the radiation model, which suffer from underfitting and overdispersion, neglect important variables such as land use and the transportation network, and cannot describe non-linear relationships between these variables. In this paper, we propose the Multi-Feature Deep Gravity (MFDG) model as an effective solution to flow generation. On the one hand, the MFDG model exploits a large number of variables (e.g., characteristics of land use and the road network; transport, food, and health facilities) extracted from voluntary geographic information data (OpenStreetMap). On the other hand, our model exploits deep neural networks to describe complex non-linear relationships between those variables. Our experiments, conducted on commuting flows in England, show that the MFDG model achieves a significant increase in the performance (up to 250% for highly populated areas) than mechanistic models that do not use deep neural networks, or that do not exploit geographic voluntary data. Our work presents a precise definition of the flow generation problem, which is a novel task for the deep learning community working with spatio-temporal data, and proposes a deep neural network model that significantly outperforms current state-of-the-art statistical models.

CCS Concepts: • **Information systems** → **Data mining**; **Information retrieval**.

Additional Key Words and Phrases: Human Mobility, Deep Learning, Migration models, Artificial Intelligence

1 INTRODUCTION

Cities are complex and dynamic ecosystems that define where people live, how they move around, who they interact with, and how they consume services [5, 12]. Most of the world's population live now in urban areas, whose evolution in structure and size influences crucial aspects of our

---

---

society such as the objective and subjective well-being [11, 15, 22, 37, 50, 52] and the diffusion of innovations [10, 14, 30]. It is therefore not surprising that the study of intra- and inter-city mobility has attracted particular interest in recent years [48], with a particular focus on the migration from rural to urban areas [41], the study and modeling of mobility patterns in urban environments [4, 7, 18, 34, 36, 49], the migration induced by natural disasters, climate change, and conflicts [3, 19, 32, 39, 42, 45], the prediction of traffic and crowd flows [16, 20, 46, 54, 56], and the forecasting of the spreading of epidemics [25, 27, 40, 44].

Among all relevant problems in the study of human mobility, the generation of synthetic flows, also known as *flow generation*, is particularly challenging. In simple words, this problem consists in generating the flows between a set of locations (e.g., how many people commute from a location to another) given the demographic and geographic characteristics of the locations (e.g., population and distance) and *without* any information about the real flows. Flow generation models have important applications, including transportation, urban planning and epidemic modeling. Traffic congestion, domestic migration, and the spread of infectious diseases are processes in which the presence of mobility flows induces a net change of the spatial distribution of some quantity of interest (vehicles, population, pathogens). The ability to accurately describe the dynamics of these processes depends on our understanding of the characteristics of the underlying spatial flows.

Existing solutions to flow generation are mainly based on *mechanistic* approaches, such as the gravity model [23], which considers both the population size of origin and destination locations and the distance between them, the radiation model [47], which additionally takes into account job opportunities in the vicinity of the origin location, and their variants [31, 33, 41, 55]. Although these approaches have the clear advantage of being simple and interpretable by design, and of requiring a few or none parameters, they suffer from several drawbacks, such as the inability to accurately capture the structure of the real flows (underfitting) and the greater variability of real flows than expected (overdispersion) [7, 28]. First of all, since mechanistic models rely on a restricted set of variables, usually just the population and the distance between locations, flows are generated without considering information that is essential to account for the complexity of the geographical landscape, such as land use, the number and diversity of points of interest (POIs), and the transportation network. Second, since mechanistic models are based on a linear combination of the variables, they cannot fully describe the non-linear interactions between the variables characterizing real-world mobility patterns. We therefore need more detailed input data and more flexible models. The former can be achieved by extracting a rich set of geographical features from data sources freely available online; the latter by using powerful non-linear models like (deep) artificial neural networks. To our knowledge, there are no models trying to combine the power of deep machine learning with the richness of geographic information to improve the realism of flow generation.

In this paper, we propose an approach to flow generation that considers a large set of variables extracted from a public voluntary geographic information system (OpenStreetMap). These variables describe important aspects of urban areas such as land use, road network features, transportation, food and health facilities, education and retail facilities. We use these geographical features to train a deep neural network, namely the Multi-Feature Deep Gravity model (MFDG), to estimate mobility flows between census areas in England (UK). We prefer neural networks over other machine learning models because they are the natural extension of the state-of-the-art model for flow generation [28], i.e., the multinomial logistic regression, which corresponds to a neural network with one softmax layer. Our approach is based on a non-linear variant of the multinomial logistic regression obtained by adding hidden layers, which introduce non-linearities and allow to build more abstract complex representations of the input geographic features.

We compare MFDG with the classic gravity model and flow generation models that use shallow neural networks or that do not exploit the geographic features, validating their realism using ground truth flow data obtained from census surveys. Specifically, for England we use information at the level of census Output Areas (OAs), the smallest unit for which census data are published in UK, which contains ~120 households.[1] We split the whole of England into equal-sized squared tiles, and we train the model on the flows between OAs within a subset of these tiles. In the test phase, we generate the flows within those tiles not been used for training, and evaluate their realism with respect to ground truth data. This procedure allows us to investigate how the realism of our model changes by the size of the tiles and the resident population within them.

Our results clearly show that MFDG outperforms all the other flow generation models, with a relative improvement in the realism with respect to the classic gravity model of around 350% in highly populated census areas, where flows are harder to predict because of the high number of relevant locations. We investigate the reasons behind this striking improvement by testing MFDG by excluding in each turn one of the geographic features. We find that, although the addition of non-linearity or the geographic features alone produces an improvement over the classic gravity model, it is the combination of deep learning and voluntary geographic information that can significantly boost the realism in the generation of mobility flows.

In summary, the main contributions of this work are:

- We propose a precise definition of the flow generation problem, which is a novel task for the deep learning community.
- We show that a characterization of the importance of geographic features, other than population and distance, can effectively improve the model performance.
- We propose a deep neural network model, called Multi-Feature Deep Gravity (MFDG), which uses a large set of variables extracted from a public voluntary geographic information system (OpenStreetMap). MFDG shows an improvement, in terms of average Common Part of Commuters (CPC), of up to 350% in highly populated areas with respect to the classic gravity model.
- We investigate the impact of population and the size of the tiles on the realism of flow generation;
- Generalization abilities are a crucial open problem in deepl-learning designed to solve human mobility tasks. We test MFDG's generalization abilities, and we show that MFDG can generate flows on a city whose flows were not used during the training phase.

2 RELATED WORK

The generation of mobility flows has attracted interest from a long time ago. Notably, in 1946 George K. Zipf proposed a model to calculate mobility flows, drawing an analogy with Newton's Law of Universal Gravitation [59]. This model, known as the *gravity model*, is based on the assumption that the number of travelers between two locations (flow) increases with the populations of the locations, while decreases with the distance between them [7].

We formally describe the gravity model and its components in Section 4. Given its ability to generate trip-flows and traffic demand between locations, the gravity model has been used in various contexts such as transport planning [17], spatial economics [24, 38, 41], and the modeling of epidemic spreading patterns [6, 13, 29, 58]. One must keep in mind that the gravity model requires the estimation of a number of free parameters from the data, making it sensitive to fluctuations or incompleteness in data [7, 28].

---

[1]A description of the characteristics of Output Areas can be found at: http://bit.ly/oa_uk

Deep learning and machine learning models exist for a different declination of the problem, namely *flow prediction*: they use the historical flows between geographic locations to forecast the future ones, but they are not able to generate flows between pairs of locations for which historical flows are not available [43]. A similar problem that is gaining interest in the literature is *crowd flow prediction*, in which the objective is to forecast the in-going (out-going) flows to (from) single locations, given historical information [16, 20, 46, 54, 56].

In recent years, the increasing availability of location data has revealed many aspects of our cities and helped researchers to better characterize cities. A notable example is Zhang et al.'s work [57], which proposes an online, semi-supervised, multimodal embedding method for geo-located information with space, time and text. The joint embeddings have been evaluated by retrieving the location for given keywords and activities. Several works have also targeted land use classification and functional area detection. For example, in [8] urban zones are represented in two ways: *(i)* as a bag-of-concepts (BOC) extracted from the Foursquare description of the area, (e.g., Arts and Entertainment, College and University, Event, Food); and *(ii)* as the same concepts above organized in a hierarchical representation, reflecting the hierarchical category structure of the activities on location-based social network platforms. Krumm et al. [26] infer the number of businesses and residences in each cell of a grid by using mobility traces. Rossi et al. [**?**] rely on the approach proposed in [8] to show how an effective urban area representation turns out to be very important for predicting the next location of a taxi ride.

Position of our work. We identify two main limitations of the current state of the art. On the one hand, existing (mechanistic) flow generation models suffer from technical issues which limit their accuracy (e.g., underfitting, overdispersion). While deep learning is used to (crowd) flow prediction, it has not been used to perform flow generation, i.e. estimating the flows between locations for which historical flows are not available. On the other hand, while a large literature exists on predicting the next-stop location in the trajectory of a single individual, little work has been done in using spatial representations and deep learning to generate mobility flows between locations without historical mobility data. Our paper aims to overcome these limitations by presenting a flow generation model that exploits, at the same time, deep learning and complex spatial representations.

3   PROBLEM DEFINITION

The problem of flow generation consists in predicting the number of trips per unit time (flow) between two locations [7]. A precise formulation of this problem requires the formal definitions of locations and flows.

A *geographical region of interest*, R, is specified as the portion of territory for which we are interested in generating the flows. Over the region of interest, a set of geographical polygons called *tessellation*, $\mathcal{T}$, is defined with the following properties: (1) the tessellation contains a finite number of polygons, l_i, called *locations*, $\mathcal{T} = \{l_i : i = 1, ..., n\}$; (2) the locations are non-overlapping, $l_i \cap l_j = \emptyset$, $\forall i \neq j$; (3) the union of all locations completely covers the region of interest, $\bigcup_{i=1}^{n} l_i = R$.

The *flow*, $y(l_i, l_j)$, between locations l_i and l_j denotes the number of trips (people moving) from location l_i to location l_j per unit time. For example, if our region of interest is the United States and our tessellation contains all US Zip codes, we can define the commuting flow $y(10001, 10005)$ as the number of people that live in location (Zip code) 10001 and go to work in location 10005 every day. The *total outflow*, O_i, from location l_i is the total number of trips per unit time originating from location l_i, i.e., $O_i = \sum_j y(l_i, l_j)$.

We formalize the problem of flow generation as follows: *given a tessellation, $\mathcal{T}$, over a region of interest R, and the total outflows from all locations in $\mathcal{T}$, estimate the flows, y, between any two locations in $\mathcal{T}$. In other words, given a location of origin and a location of destination, predict the*

flow between them. Note that this problem definition does not allow to use flows within the region of interest as input data. That is, we cannot use a subset of the flows between the locations in the region of interest to generate other flows in the same region. This means that a model tested to predict flows in region R must have been trained on a different region R', non-overlapping with R. Figure 1 shows a graphical representation of the problem formulation.
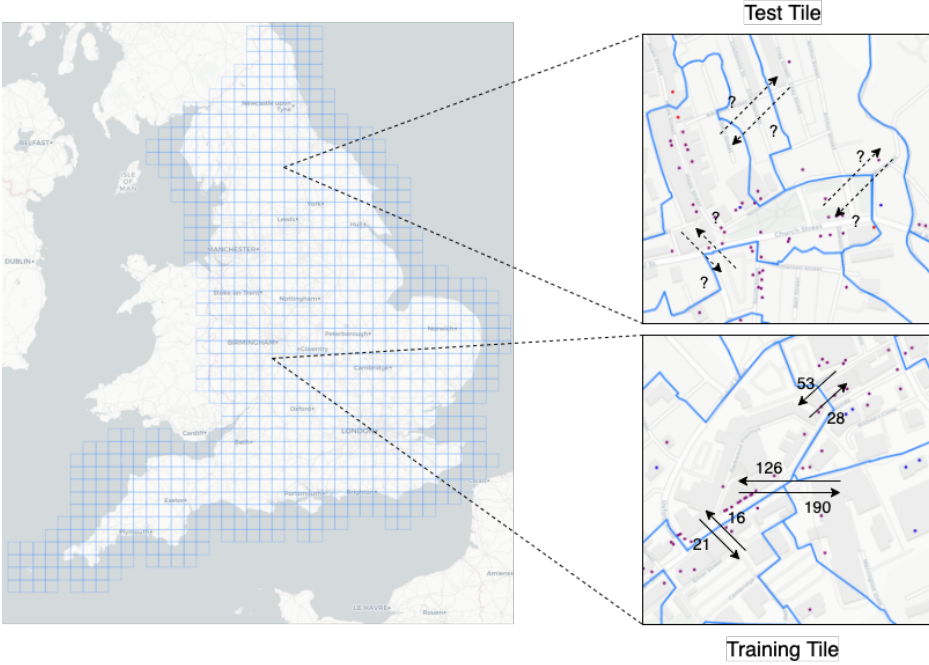

Fig. 1. **Illustration of the flow generation problem.** The geographic space (England in this example) is divided into square tiles (885 tiles in this example). Each square tile is further splitted using census output areas (OAs) to obtain a tessellation $T = \{l_i : i = 1 \ldots n\}$. During the training phase, a subset of the tiles are used for training the model and the remaining part of the tiles to test the model's performance.

4 MODEL

Our *Multi-Feature Deep Gravity model* (MFDG) originates from the observation that the state-of-the-art model of flow generation, the *Gravity model* (G) [9, 28, 59], is equivalent to a shallow linear neural network. Based on this equivalence, we naturally define the MFDG model by adding non-linearity and hidden layers to the original gravity model, as well as considering additional geographical features. The singly-constrained gravity model [7] prescribes that the expected flow, $\bar{y}$, between an origin location l_i and a destination location l_j is generated according to the following equation:

$$\bar{y}(l_i, l_j) = O_i p_{ij} = O_i \frac{m_j^{\beta_1} f(r_{ij})}{\sum_k m_k^{\beta_1} f(r_{ik})} \tag{1}$$

where m_j is the resident population of location l_j, p_{ij} is the probability to observe a trip (unit flow) from location l_i to location l_j, β_1 is a parameter and $f(r_{ij})$ is called deterrence function. Typically, the deterrence function $f(r_{ij})$ can be either an exponential, $f(r) = e^{\beta_2 r}$, or a power-law function,

$f(r) = r^{\beta_2}$, where β_2 is another parameter. In these two cases, the gravity model can be formulated as a Generalized Linear Model with a Multinomial distribution [2]. Thanks to the linearity of the model, the maximum likelihood's estimate of parameters β_1 and β_2 in Equation (1) can be found efficiently, for example using Newton's method, maximizing the model's loglikelihood:

$$LogL(\beta|y) \propto \ln\left(\prod_{i,j} p_{ij}^{y(l_i,l_j)}\right) = \sum_{i,j} y(l_i, l_j) \ln \frac{m_j^{\beta_1} f(r_{ij})}{\sum_k m_k^{\beta_1} f(r_{ik})} = \sum_{i,j} y(l_i, l_j) \ln \frac{e^{\beta \cdot x(l_i,l_j)}}{\sum_k e^{\beta \cdot x(l_i,l_k)}} \quad (2)$$

where y is the matrix of observed flows, $\beta = [\beta_1, \beta_2]$ is the vector of parameters and the input feature vector is $x(l_i, l_j) = \text{concat}[x_j, r_{ij}]$ for the exponential deterrence function ($x(l_i, l_j) = \text{concat}[x_j, \ln r_{ij}]$ for the power-law deterrence function) with $x_j = \ln m_j$. Note that the negative of loglikelihood in Equation (2) is proportional to the cross-entropy loss, $H = -\sum_i \sum_j \frac{y(l_i,l_j)}{O_i} \ln p_{i,j}$, of a shallow neural network with an input of dimension two and a single linear layer followed by a softmax layer.

   This equivalence suggests to interpret the flow generation problem as a classification problem, where each observation (trip or unit flow from an origin location) should be assigned to the correct class (the actual location of destination) chosen among all possible classes (all locations in $\mathcal{T}$). According to this interpretation, the gravity model is a linear classifier based on two explanatory variables (features x), i.e. population and distance. The interpretation of the flow generation problem as a classification problem allows us to naturally extend the gravity model's shallow neural network introducing hidden layers and non-linearities.

4.1 Architecture of the Multi-Feature Deep Gravity model (MFDG)

To generate the flows from a given origin location (e.g. l_i), the MFDG model uses a number of input features to compute the probability $p_{i,j}$ that any of the n locations in the region of interest (e.g. l_j) is the destination of a trip from l_i. Specifically, the model output is a n-dimensional vector of probabilities $p_{i,j}$ for $j = 1, ..., n$. These probabilities are computed in three steps (see Figure 2).

   First, the input vectors $x(l_i, l_j) = \text{concat}[x_i, x_j, r_{i,j}]$ for $j = 1, ..., n$ are obtained performing a concatenation of the following input features: *(i)* x_i, the feature vector of the origin location l_i; *(ii)* x_j the feature vector of the destination location l_j; *(iii)* and the distance between origin and destination $r_{i,j}$. For each origin location (e.g. l_i), n input vectors $x(l_i, l_j)$ with $j = 1, ..., n$ are created, one for each location in the region of interest that could be a potential destination.

   Second, the input vectors $x(l_i, l_j)$ are fed in parallel to the same feed-forward neural network. The network has 15 hidden layers of dimensions 256 (the bottom six layers) and 128 (the other layers) with LeakyReLu activation function, a. Specifically, the output of hidden layer h is given by the vector $z^{(0)}(l_i, l_j) = a(W^{(0)} \cdot x(l_i, l_j))$ for the first layer ($h = 0$) and $z^{(h)}(l_i, l_j) = a(W^{(h)} \cdot z^{(h-1)}(l_i, l_j))$ for $h > 0$, where W are matrices whose entries are parameters learned during training.

   The output of the last layer is a scalar $s(l_i, l_j) \in [-\infty, +\infty]$ called score: the higher the score for a pair of locations (l_i, l_j), the higher the probability to observe a trip from l_i to l_j according to the model. Finally, scores are transformed into probabilities using a softmax function, $p_{i,j} = e^{s(l_i,l_j)}/\sum_k e^{s(l_i,l_k)}$, which transforms all scores into positive numbers that sum up to one.

4.2 Feature Vectors

The location feature vector x_i provides a spatial representation of an area, and it contains features describing some properties of location l_i, e.g., the total length of residential roads or the number of restaurants therein. Its dimension, d is equal to the total number of features considered. The location features we use include geographical features extracted from OpenStreetMap (OSM) [1] belonging to the following categories:

- *Land use areas (5 features)*: total area (in km^2) for each possible land use class, i.e., residential, commercial, industrial, retail and natural;
- *Road network (3 features)*: total length (in km) for each different types of roads, i.e., residential, main and other;
- *Transport facilities (2 features)*: total count of Points Of Interest (POIs) and buildings related to each possible transport facility, e.g., bus/train station, bus stop, car parking;
- *Food facilities (2 features)*: total count of POIs and buildings related to food facilities, e.g., bar, cafe, restaurant;
- *Health facilities (2 features)*: total count of POIs and buildings related to health facilities, e.g., clinic, hospital, pharmacy;
- *Education facilities (2 features)*: total count of POIs and buildings related to education facilities, e.g., school, college, kindergarten;
- *Retail facilities (2 features)*: total count of POIs and buildings related to retail facilities, e.g., supermarket, department store, mall.

In addition, we included as feature of MFDG the *distance*, $r_{i,j}$, between two locations l_i, l_j, which is defined as the geographical distance (the distance measured along the surface of the earth) between the centroids of the two polygons representing the locations. All values of features for a given location (excluding distance) are normalized dividing them by the location's area.
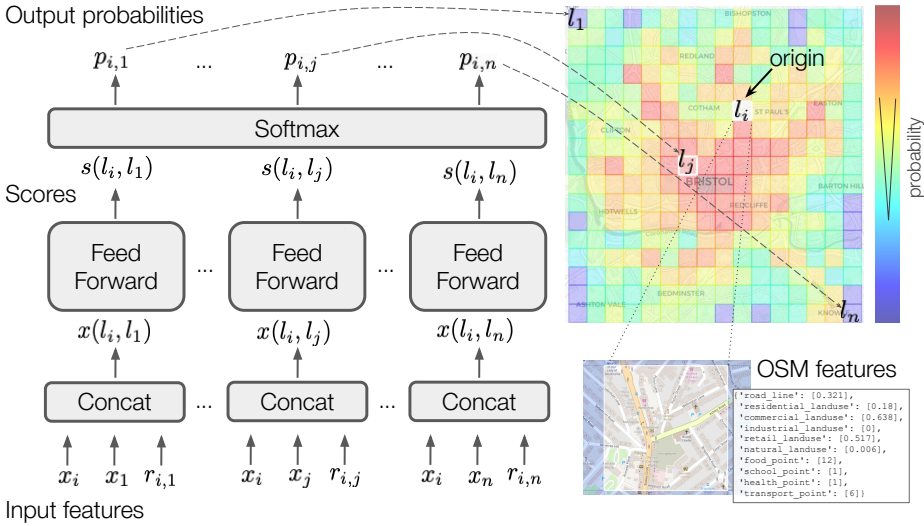

Fig. 2. **Architecture of the Multi-Feature Deep Gravity model (MFDG).** The input features x_i (feature vector of the origin location l_i), x_j (feature vector of the destination location l_j), and $r_{i,j}$ (distance between origin and destination) are concatenated to obtain the input vectors $x(l_i, l_j)$. These vectors are fed, in parallel, to the same feed forward neural network with 15 hidden layers with LeakyReLu activation functions. The output of the last hidden layer is a score $s(l_i, l_j) \in [-\infty, +\infty]$. The higher this score for a pair of locations (l_i, l_j), the higher the probability to observe a trip from l_i to l_j. Finally, a softmax function is used to transform the probabilities $p_{i,j}$, which are positive numbers that sum up to one.

4.3 Training

The loss function of MFDG is the cross-entropy loss:

$$H = -\sum_i \sum_j \frac{y(l_i, l_j)}{O_i} \ln p_{i,j}, \tag{3}$$

where $y(l_i, l_j)/O_i$ is the fraction of observed flows from l_i that go to l_j and $p_{i,j}$ is the model's probability of a unit flow from l_i to l_j. Note that the sum over i of the cross-entropies of different origin locations follows from the assumption that flows from different locations are independent events, which allows us to apply the additive property of the cross-entropy for independent random variables. The network is trained for 150 epochs with the RMSprop [51] optimizer using batches of size 100 origin locations. To reduce the training time, we use negative sampling and consider up to 700 randomly selected locations of destination for each origin location.

5 EXPERIMENTS

In this section, we present the experimental setup and the data used for the experiments (Section 5.1), the results obtained and a characterization of the importance of the introduced geographic features (Section 5.2).

5.1 Experimental Setup

Urban data provide information about a physical object (e.g., POIs) in the urban environment. Additionally, official statistics offer information about commuting flows between areas in a country. In this work, we combine the two data sources to estimate mobility flows in different regions of interest in England.

Regions of interest. First, we define a squared tessellation over the original polygonal shape of England. Formally, let C be the polygon composed by q vertices $v_1, ..., v_q$ that define the polygon boundary. We define the grid G as the square tessellation covering C with $L_x \times L_y$ regions of interest: $G = \{R_{ij}\}_{i=1,..,L_x; j=1,...,L_y}$, where R_{ij} is the square cell (i, j) defined by two vertices representing the top-left and bottom-right coordinates. Depending on the nature of the problem, such tessellation can be formed with any triangle or quadrilateral tile or, as in the case of Voronoi tessellation, with tile defined as the set of points closest to one of the points in a discrete set of defining points. In this paper, we build a square grid (see Figure 3) using the tessellation builder from the python library *scikit-mobility* [35], defining 885 regions of interest of 25 by 25 km², which cover the whole of England. Half of these regions are used to train the models and the other half are used for testing: the regions of interest have been randomly allocated to the train and test sets in a stratified fashion based on the regions' populations, so that the two sets have the same number of regions belonging to the various population deciles.

Locations. The area covered by each region of interest is further divided into locations using a tessellation $\mathcal{T}$ provided by the UK Census in 2011. The UK Census defines 232,296 non-overlapping polygons called census Output Areas (OAs, see Figure 3), which cover the whole of England. By construction, OAs should all contain a similar number of households (125), hence in cities and urban areas where population density is higher, there is a larger number of Output Areas and they have a smaller size than average. For a given region of interest R_{ij}, its locations are defined as all OAs whose centroids are contained in R_{ij}.

Location Features. We collect information about the geographic features of each location from OpenStreetMap[2] (OSM) [1], an online collaborative project aimed to create an open source map of the world with geographic information collected from volunteers. The OSM data contain three types of geographical objects: nodes, lines and polygons. Nodes are geographic points, stored as latitude and longitude pairs, which represent points of interests (e.g., restaurants, schools). Lines are ordered lists of nodes, representing linear features such as streets or railways. Polygons are lines that form a closed loop enclosing an area and may represent, for example, landuse or buildings. We use OpenStreetMap (OSM) data to compute the 23 features listed in Section 4.2. Regarding the population, we include the number of inhabitants for each location as an input feature and we use the number of residents in each OA provided by the UK Census for the year 2011. To evaluate the performance of the model, we compare the generated flows with empirical flows using the Sørensen-Dice index, also called Common Part of Commuters (CPC) [28], which is a well-established measure in human mobility to compute the similarity between real flows, y^r, and generated flows, y^g [7, 28]:

$$CPC = \frac{2 \sum_{i,j} min(y^g(l_i, l_j), y^r(l_i, l_j))}{\sum_{i,j} y^g(l_i, l_j) + \sum_{i,j} y^r(l_i, l_j)} \tag{4}$$

CPC is an always positive number contained in the closed interval (0,1) with 1 indicating a perfect match between the generated flows and the ground truth and 0 highlighting bad performance with no overlap.
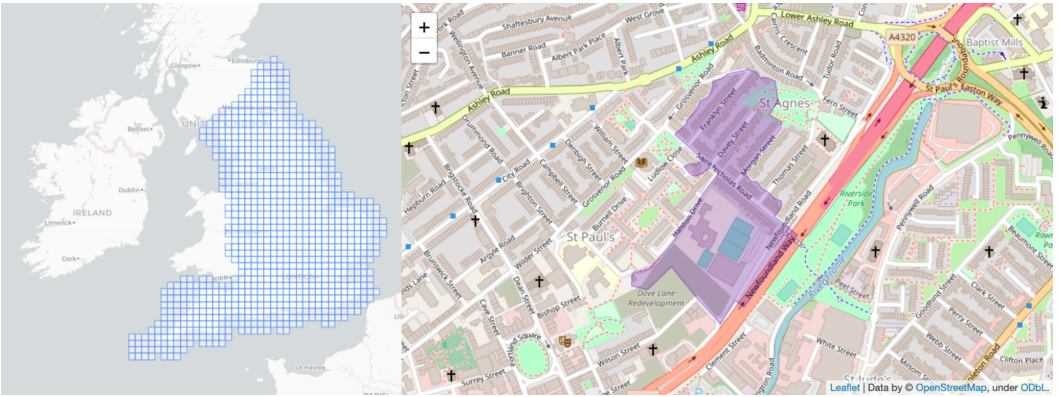

Fig. 3. **Example of tiles and Output Area in England.** (Left) Visualization of the regions of interest of 25 by 25 km^2 in England: half of the regions are used for training and half for testing. (Right) An example of census Output Area. The illustration are made using the scikit-mobility library [35].

Mobility Flows. The UK Census collects information about commuting flows between Output Areas. We use the commuting flows collected by the UK Census in 2011 and consider flows that have origin location and destination location in the same region of interest only. The UK Census covers 30,008,634 commuters with an average flow from an OA to another of 1.78 and a standard deviation of 3.21.
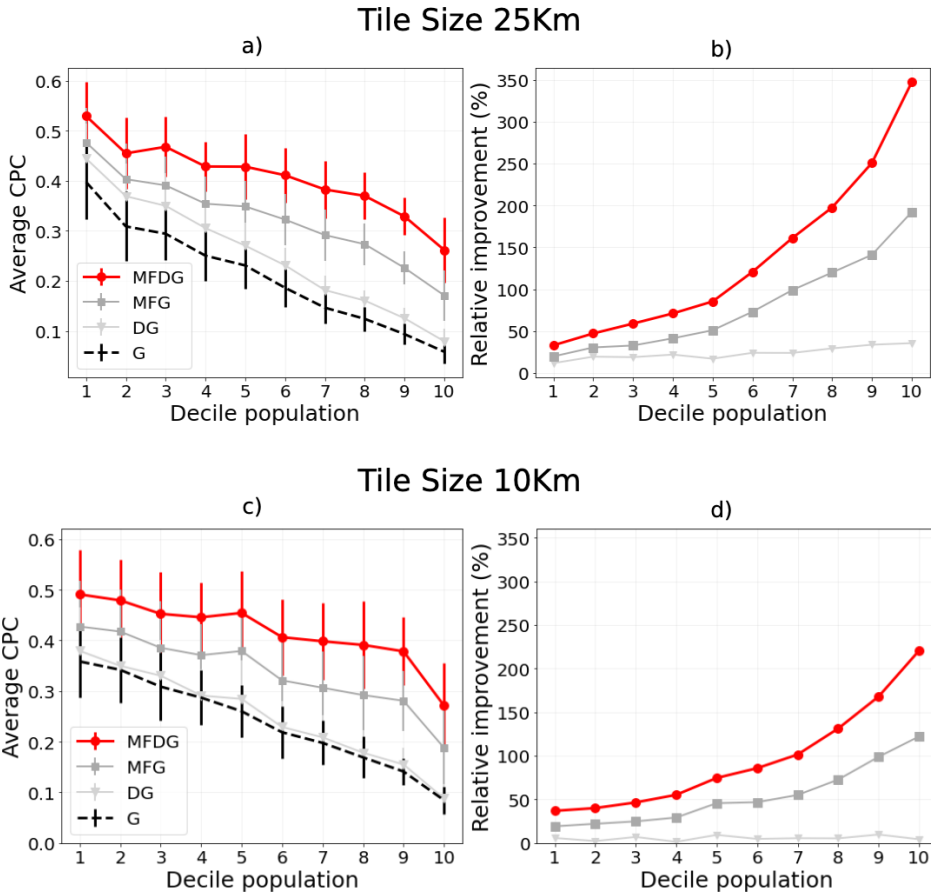
Fig. 4. **Models comparison.** (Left) Comparison of the performance (CPC) of Gravity (G), Deep Gravity (DG), Multi-Feature Gravity (MFG), and Multi-Feature Deep Gravity (MFDG), varying the decile of the population and for tile sizes of 25km (a), 10km (c) and 5km (e). MFDG is by far the approach with the best average CPC, regardless of the decile of the population. (Right) Relative improvement with respect to G of DG, MFG, and MFDG, varying the decile of the population and for tile sizes of 25km (b), 10km (d) and 5km (f). The higher the population the higher is the improvement of MFDG and MFG with respect to G, while DG shows no improvement. For example, for tile size 25km and the last decile of the population, MFDG has an improvement of around 350% with respect to the traditional gravity model G.

5.2 Results

Our experiments aim at demonstrating the effectiveness of the models in generating mobility flows within the region of interest belonging to the test set.

The compared models are listed on the first column of Table 1. Given the formal similarity between the MFDG and the gravity model, we use the Gravity model (G) as a baseline to assess MFDG's improved predictive performance. Additionally, we define two *hybrid* models to understand the performance gain obtained by adding either multiple non-linear hidden layers or complex geographical features to the Gravity model:

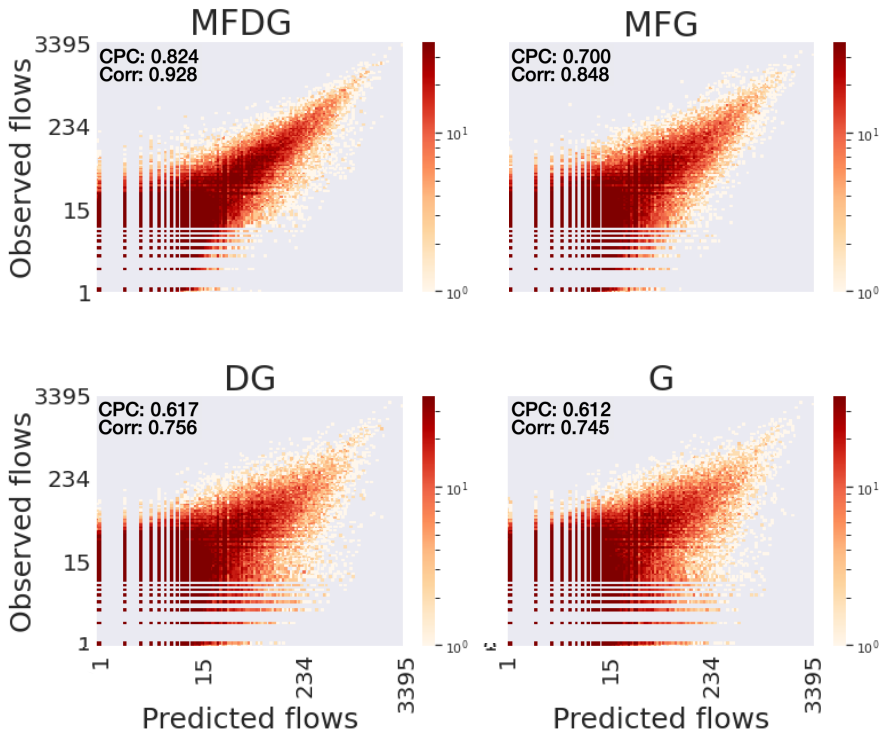
[2].https://www.openstreetmap.org/

Fig. 5. **Observed vs predicted flows.** Observed flows versus predicted flows for MFDG, MFG, DG and G. The color, in a gradient from yellow to red, indicates the density of points (flows). For each plot we also specify the CPC and the Pearson correlation between the observed and the predicted flows.

- the *Deep Gravity model* (DG) uses a feed-forward neural network with the same structure of the MFDG model, but, similarly to G, DG's input features are only population and distance.
- the *Multi-Feature Gravity model* (MFG) has the same multiple input features of MFDG, including various geographical variables extracted from OSM but, similarly to G, these features are processed by a shallow single-layer linear neural network.

We split the regions of interest into ten equal-sized groups, i.e., deciles, based on their population, where decile 1 includes the regions of interest with the smaller population and decile 10 includes the regions of interest with the larger population, and we analyze the performance of the four models in each decile (Figure 4 and Table 1). Figure 8 shows an example of the flows observed and the flows generated by the models on 101 OAs. We find two main results.

First, all models degrade (i.e., CPC decreases) as the decile of the population increases, denoting that they are more accurate in sparsely populated output areas. Nevertheless, the relative improvement of MFG and MFDG with respect to G increases as the population increases. In other words, the performance of these degrades less as population increases. This is a remarkable outcome because in highly populated regions of interest there are many relevant locations, and hence predicting the correct destinations of trips is harder. MFDG improves especially where current models are unrealistic.

Second, we find that the introduction of the geographic features (MFG) and of non-linearity and hidden layers (MFDG) leads to a significant improvement of the overall performance. As Figure 4
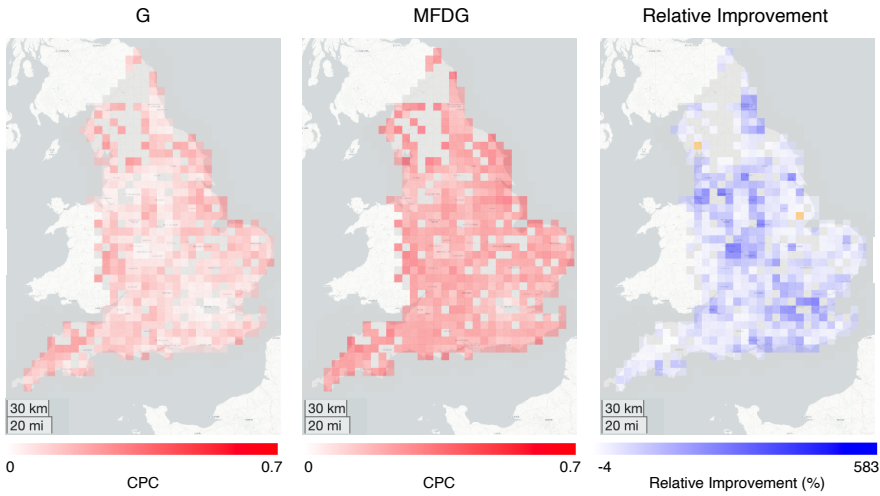
Fig. 6. **CPC by tile in England.** (Left) CPC in each tile in England according to G. (Center) CPC in each tile according to MFDG. The CPC in each tile is the average CPC over all the runs in which that tile has been selected in at least one test set. Grey tiles have never been selected in the test set. (Right) Relative improvement of MFDG with respect to G for each tile. The two orange tiles are the ones for which G works slightly better than MFDG (G has an improvement of about 4% and 1% with respect to MFDG).
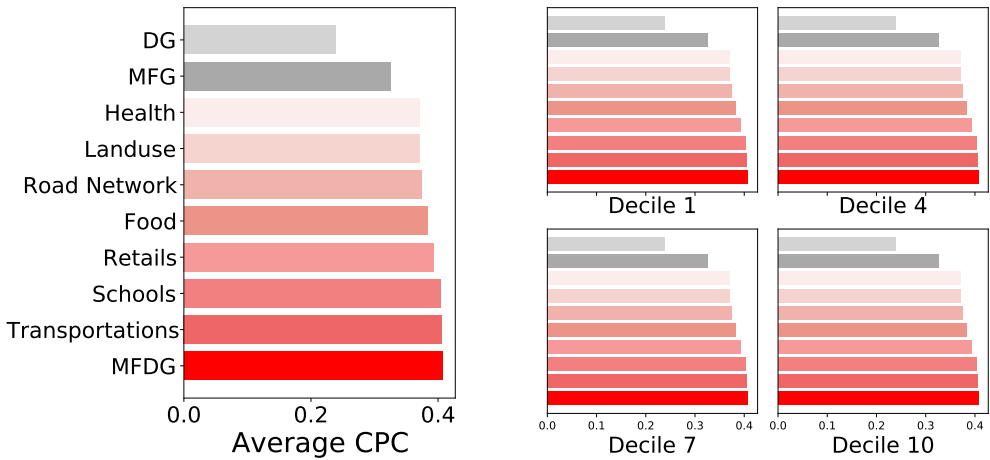

Fig. 7. **Variable importance.** (Left) Average Common Part of Commuters (CPC) of DG, MFG, MFDG, and MFDG on test cases in which each geographic feature is set in turn to zero. (Right) Average CPC for the OAs in the first, fourth, seventh and last decile of the population only.

shows, the relative improvement of MFG and MFDG with respect to G is significant, with values of about 190% and 350%, respectively, in the last decile of population. Even in the first decile of the population, we find a relative improvement of MFDG with respect to G of 33%. Note that, as depicted in Figure 6, MFDG's improvement on G is a common characteristic. There are only two poorly populated regions of interest out of the 885 ones in which G performs better than MFDG. Moreover, we find that also the Pearson correlation between the predicted and observed CPC
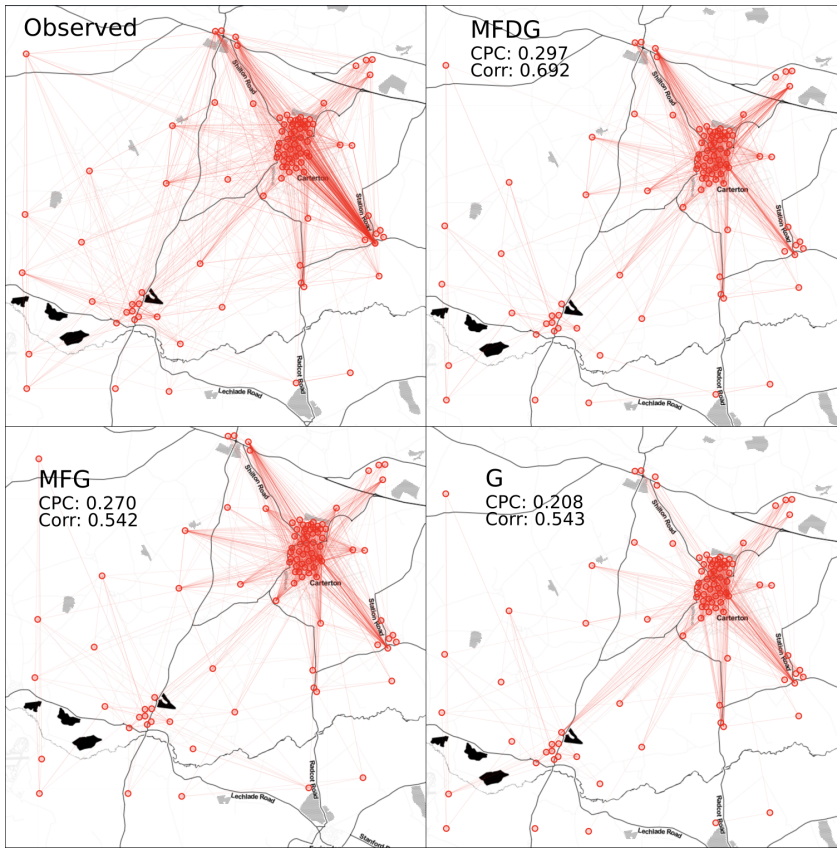
Fig. 8. **Generated flows sample.** Visualization of a sample of the observed flows and the flows generated by MFDG, DG, and G between 101 scarcely populated OAs. For each model, we highlight the CPC and the Pearson correlation coefficient for the specific area depicted. Note that the CPC is relatively low because 1) we use just a sample of the available OAs, and 2) the OAs selected are scarcely populated and, as shown in Figure 5, flows in these OAs are the most difficult to capture.

of MFDG is higher than the one obtained with G, DG and MFG (Figure 5). We do not find any significant improvement of DG with respect to G, despite the use of the non-linearity introduced by the deep neural network.

Neural network models applied to spatial data tend to suffer from low generalization capabilities when applied to different geographical regions than the ones used for training. To mitigate such weakness, MFDG is trained using a geographic representation of an area rather than geographical coordinates. To further test such generalization abilities of MFDG, we designed an additional experiment. In the previous experiment, it may happen that given a major city - covered by multiple regions of interest - some of its tiles are used during the training phase. Even if MFDG achieves outstanding performances, a more challenging problem consists in testing its ability to generate flows to areas never seen before. Moreover, by analyzing the characteristic of the model, we can discover whether we can generate flows for a city for which we do not have *any* information, a peculiarity that we cannot fully investigate if the model partially see a city (e.g., use some of the tiles of a city in the training phase). In this sense, differently from the experiment previously described,
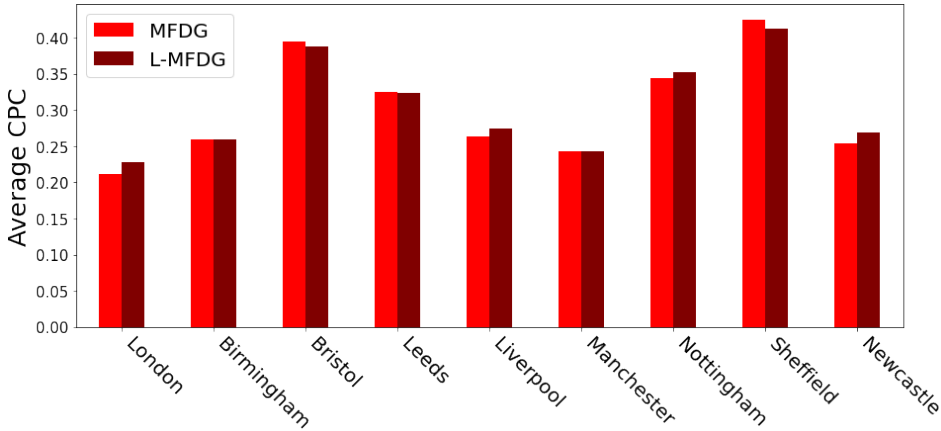
Fig. 9. Average CPC of T-MFDG on the Core Cities and London according to the leave-one-city-out validation.

we design specific training and testing datasets so that a city is never seen during the training phase. Given the nine England major cities, i.e., the so-called Core Cities [3] and London, the training dataset contains the tiles and the information of eight cities and the test set contains information on the city excluded from the training. In particular, we selected 15 tiles corresponding to London, eight to Leeds, seven to Sheffield, five to Birmingham, four to Bristol, Liverpool, Manchester and Newcastle, and three to Nottingham. In this way, we can test whether MFDG is able to generalize by analyzing its performances according to a *leave-one-city-out* validation mechanism, i.e., generate flows on a city whose tiles never appear in the training set. We denote this implementation with *Leave-one-city-out-MFDG* (L-MFDG). In Figure 9, we show that L-MFDG produces average CPCs that are remarkably close to the MFDG's ones. For instance, by testing L-MFDG on London's tiles using the tiles of the other cities as training, the average CPC slightly improves. Similar results are obtained by testing the model on Newcastle, Liverpool and Nottingham. The average CPC slightly decreases when tested on Bristol and Sheffield, while it does not change significantly with respect to MFDG on Leeds, Birmingham and Manchester. In summary, MFDG can generate flows also for regions of interest for which no knowledge is available about the mobility flow patterns.

Role of tile size. We investigate the impact of the size of the regions of interest on the model performance by replicating the experiments using tiles of size 10km. We find that the performance of all models does not change significantly as the tile size decreases. However, the relative improvement of MFDG over G is smaller with a tile size of 10km (Figure 4b, d). For instance, for a tile size of 10km, the improvement of the MFDG over G on the last decile is about 220%, i.e., 130% less than the improvement on the same decile for a tile size of 25km. Hence, MFDG has a higher relative improvement over G in larger regions of interest (25 km), where there are more locations and predicting the correct destination of trips is harder.

Importance of geographic variables. To understand the contribution of the various geographic variables in the striking performance of MFDG, we perform a series of tests setting to zero in turn all the features of one category (Health, Landuse, Road Network, Food, Retails, Schools, Transportations). For example, for the "Health" test, we set to zero the two health-related features: the total count of POIs and the total number of buildings related to health facilities, such as clinics,

---

[3]https://www.corecities.com/

		Decile of Population									
		1	2	3	4	5	6	7	8	9	10
		Tile Size 25km									
G	Mean CPC	0.39	0.30	0.29	0.25	0.23	0.18	0.14	0.12	0.09	0.05
	std CPC	0.07	0.06	0.05	0.05	0.04	0.03	0.03	0.02	0.02	0.02
	Mean CPC	0.44	0.36	0.34	0.30	0.27	0.23	0.18	0.16	0.12	0.07
DG	std CPC	0.06	0.06	0.06	0.05	0.04	0.03	0.02	0.02	0.02	0.02
	Rel. Imp.	11.78	19.50	18.94	22.00	17.10	24.12	23.98	29.37	33.79	35.67
	Mean CPC	0.47	0.40	0.39	0.35	0.34	0.32	0.29	0.27	0.22	0.17
MFG	std CPC	0.06	0.07	0.05	0.05	0.05	0.05	0.05	0.04	0.03	0.04
	Rel. Imp.	19.82	30.53	32.83	41.42	51.06	73.14	98.98	119.94	141.08	192.18
	Mean CPC	**0.52**	**0.45**	**0.46**	**0.42**	**0.42**	**0.41**	**0.38**	**0.36**	**0.32**	**0.26**
MFDG	std CPC	0.06	0.07	0.05	0.04	0.06	0.05	0.05	0.04	0.03	0.06
	Rel. Imp.	**33.09**	**47.22**	**58.99**	**71.10**	**85.43**	**120.72**	**161.07**	**197.59**	**250.75**	**347.15**
		Tile Size 10km									
G	Mean CPC	0.35	0.34	0.30	0.28	0.26	0.21	0.19	0.16	0.14	0.08
	std CPC	0.07	0.06	0.06	0.05	0.05	0.05	0.04	0.04	0.02	0.02
	Mean CPC	0.37	0.34	0.33	0.29	0.28	0.22	0.20	0.17	0.15	0.08
DG	std CPC	0.08	0.05	0.07	0.05	0.07	0.05	0.04	0.04	0.03	0.02
	Rel. Imp.	5.82	2.29	7.01	1.42	9.37	4.76	5.58	5.38	9.79	4.35
	Mean CPC	0.42	0.41	0.38	0.37	0.37	0.32	0.30	0.29	0.28	0.18
MFG	std CPC	0.08	0.08	0.09	0.06	0.07	0.08	0.07	0.07	0.05	0.07
	Rel. Imp.	19.23	22.08	24.84	29.23	45.71	46.83	55.07	72.68	98.77	122.13
	Mean CPC	**0.49**	**0.47**	**0.45**	**0.44**	**0.45**	**0.40**	**0.39**	**0.39**	**0.37**	**0.27**
MFDG	std CPC	0.08	0.08	0.08	0.06	0.08	0.07	0.07	0.08	0.06	0.08
	Rel. Imp.	**36.92**	**40.08**	**46.57**	**55.28**	**74.64**	**85.82**	**101.45**	**131.19**	**167.86**	**220.46**

Table 1. Comparison of the performance, in terms of Common Part of Commuters (CPC), of Gravity (G), Deep Gravity (DG), Multi-Feature Gravity (MFG), and Multi-Feature Deep Gravity (MFDG), varying the decile of the population of the output areas. For each model, and for each decile of the distribution of population within the output areas, we show the average CPC and the standard deviation of the CPC obtained over three runs of the model. For DG, MFG, and MFDG we also show the relative improvement in terms of CPC with respect to model G. We put in bold the values over the deciles that correspond to the best mean CPC and relative improvement.

hospitals and pharmacies. As Figure 7 shows, the usage of the geographic features allows us to achieve significantly better performance than DG. Although some feature categories bring a slightly better contribution than the others (e.g., Health and Landuse) on the other hand, removing some categories (e.g., Transportations and Schools) do not affect the performances. In general, there is no single category which is significantly more important than the others. Note that the addition of geographic features alone is not sufficient to explain the improvement of MFDG over the other models: the non-linearity plays a crucial role as well. Indeed, although MFG is slightly better than DG, it is still significantly worse than MFDG. In summary, our results clearly show that, although the addition of non-linearity or the geographic features alone produces an improvement over the classic gravity model, it is the combination of deep learning and voluntary geographic information that can significantly boost the realism in the generation of mobility flows.

6 CONCLUSION

In this work, we presented the Multi-Feature Deep Gravity model (MFDG), an approach to mobility flow generation based on deep neural networks and voluntary geographic information. The

comparison of the performance of MFDG with models that do not use non-linearity or do not include the geographic information reveals two key results. First, all models are more accurate on scarcely populated regions, where most destinations are concentrated into few relevant locations and are thus easier to predict. More importantly, in highly populated regions where there are many relevant locations and hence predicting the correct destinations of trips is harder, the improvement of MFDG with respect to its competitors becomes much higher, suggesting that our model improves especially where current models fail. We observed that MFDG still outperforms all its competitors even when using a smaller tile size.

To gain more insights on this striking improvement, we investigated the role of the geographic features by zeroing in turn all geographic variables of one category. We discover that the addition of the features from OpenStreetMap are crucial to boost the realism of our approach. However, our analysis clearly shows that it is the combination of deep learning and voluntary geographic information that significantly boosts the realism in the generation of mobility flows, paving the road to a new breed of data-driven flow generation models.

Third, we showed that MFDG is a geographic agnostic model able to generate flows on urban agglomerations never seen before. This opens the doors to a set of intriguing questions such as the investigation on the limits regarding the geographical transferability of such models.

In this sense, a future improvement of the model may consist in analyzing whether we can apply geographic transferability on other scales. Indeed, it would be interesting to observe the impact of the type of area on the tranferrability to answer questions such as: can we use rural areas flows to generate flows in cities? On the other hand, can we use cities' flows to generate flows in rural areas?

Other possible improvements are related to the tessellation. In this experiment we used a squared tessellation but we may also consider different types of tessellations. In fact, using administrative areas may lead to improved performances as theoretically, administrative areas should identify parts of similar cities. For example, while an administrative area should clearly divide suburbs from the center of a city, a squared tessellation may partially cover both the city center and the outskirts.

Finally, in the proposed model, we count the number of POIs and relative building per category. However, there are deep-learning based approaches to solve other mobility tasks that highlight the importance of efficiently encode geographic features to boost the performances of the models. Examples of such embedding mechanisms are [21, 53]. Use such approaches combined with MFDG may lead to an improvement of the performances.

ACKNOWLEDGMENTS

REFERENCES

[1] 2018. *OpenStreetMap*. https://www.openstreetmap.org/
[2] Alan Agresti. 2015. *Foundations of linear and generalized linear models.* John Wiley & Sons.
[3] Mohammed N Ahmed, Gianni Barlacchi, Stefano Braghin, Francesco Calabrese, Michele Ferretti, Vincent PA Lonij, Rahul Nair, Rana Novack, Jurij Paraszczak, and Andeep S Toor. 2016. A Multi-Scale Approach to Data-Driven Mass Migration Analysis.. In *SoGood@ ECML-PKDD*.
[4] Laura Alessandretti, Piotr Sapiezynski, Vedran Sekara, Sune Lehmann, and Andrea Baronchelli. 2018. Evidence for a conserved quantity in human mobility. *Nature Human Behaviour* 2, 7 (2018), 485–491. https://doi.org/10.1038/s41562-018-0364-x

[5] Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Guido Caldarelli, Paolo Cintia, Stefano Cresci, Angelo Facchini, Fosca Giannotti, Aristides Gionis, Riccardo Guidotti, et al. 2020. (So) Big Data and the transformation of the city. *International Journal of Data Science and Analytics* (2020). https://doi.org/10.1007/s41060-020-00207-3

[6] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. 2010. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *Journal of computational science* 1, 3 (2010), 132–145.

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

[8] Gianni Barlacchi, Alberto Rossi, Bruno Lepri, and Alessandro Moschitti. 2017. Structural Semantic Models for Automatic Analysis of Urban Areas. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 279–291.

[9] Marc Barthélemy. 2011. Spatial networks. *Physics Reports* 499, 1-3 (2011), 1–101.

[10] Luís MA Bettencourt, José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B West. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the national academy of sciences* 104, 17 (2007), 7301–7306.

[11] Luís MA Bettencourt, José Lobo, Deborah Strumsky, and Geoffrey B West. 2010. Urban scaling and its deviations: Revealing the structure of wealth, innovation and crime across cities. *PloS one* 5, 11 (2010), e13541.

[12] D Byrne. 1998. Class and Ethnicity in Complex Cities: The Cases of Leicester and Bradford. *Environment and Planning A: Economy and Space* 30, 4 (1998), 703–720. https://doi.org/10.1068/a300703 arXiv:https://doi.org/10.1068/a300703

[13] Serhan Cevik. 2020. Going Viral: A Gravity Model of Infectious Diseases and Tourism Flows. (2020).

[14] Deqiu Chen, Huasheng Gao, Jiang Luo, and Yujing Ma. 2020. The effects of rural–urban migration on corporate innovation: Evidence from a natural experiment in China. *Financial Management* 49, 2 (2020), 521–545.

[15] Marco De Nadai, Yanyan Xu, Letouzé Emmanuel, Marta C González, and Bruno Lepri. 2020. Socio-economic, built environment, and mobility conditions associated with crime: a study of multiple cities. *Scientific Reports (Nature Publisher Group)* 10, 1 (2020).

[16] Zeinab Ebrahimpour, Wanggen Wan, Ofelia Cervantes, Tianhang Luo, and Hidayat Ullah. 2019. Comparison of main approaches for extracting behavior features from crowd flow analysis. *ISPRS International Journal of Geo-Information* 8, 10 (2019), 440.

[17] Sven Erlander and Neil F Stewart. 1990. *The gravity model in transportation analysis: theory and extensions*. Vol. 3. Vsp.

[18] Marta C. González, César A. Hidalgo, and Albert-László Barabási. 2008. Understanding individual human mobility patterns. *Nature* 453 (05 06 2008), 779 EP –. https://doi.org/10.1038/nature06958

[19] Clark L Gray and Valerie Mueller. 2012. Natural disasters and population mobility in Bangladesh. *Proceedings of the National Academy of Sciences* 109, 16 (2012), 6000–6005.

[20] Kasthuri Jayarajah, Andrew Tan, and Archan Misra. 2018. Understanding the Interdependency of Land Use and Mobility for Urban Planning. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (Singapore, Singapore) *(UbiComp '18)*. Association for Computing Machinery, New York, NY, USA, 1079–1087. https://doi.org/10.1145/3267305.3274163

[21] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. 2019. Tile2vec: Unsupervised representation learning for spatially distributed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3967–3974.

[22] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C. Gonzalez. 2016. The TimeGeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences* 113 (08 2016), 201524261. https://doi.org/10.1073/pnas.1524261113

[23] Woo-Sung Jung, Fengzhong Wang, and H Eugene Stanley. 2008. Gravity model in the Korean highway. *EPL (Europhysics Letters)* 81, 4 (2008), 48005.

[24] David Karemera, Victor Iwuagwu Oguledo, and Bobby Davis. 2000. A gravity model analysis of international migration to North America. *Applied economics* 32, 13 (2000), 1745–1755.

[25] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Louis Du Plessis, Nuno R Faria, Ruoran Li, William P Hanage, et al. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 6490 (2020), 493–497.

[26] John Krumm and Kora Krumm. 2019. Land Use Inference from Mobility Traces. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*. ACM, 1–4.

[27] Shengjie Lai, Andrea Farnham, Nick W Ruktanonchai, and Andrew J Tatem. 2019. Measuring mobility, disease connectivity and individual risk: a review of using mobile phone data and Health for travel medicine. *Journal of travel medicine* 26, 3 (2019), taz019.

[28] Maxime Lenormand, Aleix Bassolas, and José J Ramasco. 2016. Systematic comparison of trip distribution laws and models. *Journal of Transport Geography* 51 (2016), 158–169.

[29] Xinhai Li, Huidong Tian, Dejian Lai, and Zhibin Zhang. 2011. Validation of the gravity model in predicting the global spread of influenza. *International journal of environmental research and public health* 8, 8 (2011), 3134–3143.

[30] Francesco Lissoni. 2018. International migration and innovation diffusion: an eclectic survey. *Regional Studies* 52, 5 (2018), 702–714.

[31] Er-Jian Liu and Xiao-Yong Yan. 2020. A universal opportunity model for human mobility. *Scientific reports* 10, 1 (2020), 1–10.

[32] Candice A Myers, Tim Slack, and Joachim Singelmann. 2008. Social vulnerability and migration in the wake of disaster: the case of Hurricanes Katrina and Rita. *Population and Environment* 29, 6 (2008), 271–291.

[33] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. 2012. A Tale of Many Cities: Universal Patterns in Human Urban Mobility. *PLOS ONE* 7, 5 (05 2012), 1–10. https://doi.org/10.1371/journal.pone.0037027

[34] Luca Pappalardo and Filippo Simini. 2018. Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery* 32, 3 (2018), 787–829. https://doi.org/10.1007/s10618-017-0548-4

[35] Luca Pappalardo, F Simini, G Barlacchi, and R Pellungrini. 2019. scikit-mobility: A Python library for the analysis, generation and risk assessment of mobility data. *arXiv preprint arXiv:1907.07062* (2019).

[36] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. 2015. Returners and explorers dichotomy in human mobility. *Nature Communications* 6, 1 (2015), 8166. https://doi.org/10.1038/ncomms9166

[37] Luca Pappalardo, Maarten Vanhoof, Lorenzo Gabrielli, Zbigniew Smoreda, Dino Pedreschi, and Fosca Giannotti. 2016. An analytical framework to nowcast well-being using mobile phone data. *International Journal of Data Science and Analytics* 2, 1-2 (2016), 75–92.

[38] Roberto Patuelli, Aura Reggiani, Sean P Gorman, Peter Nijkamp, and Franz-Josef Bade. 2007. Network analysis of commuting flows: A comparative static approach to German data. *Networks and Spatial Economics* 7, 4 (2007), 315–331.

[39] Bimal Kanti Paul. 2005. Evidence against disaster-induced migration: the 2004 tornado in north-central Bangladesh. *Disasters* 29, 4 (2005), 370–385.

[40] Emanuele Pepe, Paolo Bajardi, Laetitia Gauvin, Filippo Privitera, Brennan Lake, Ciro Cattuto, and Michele Tizzoni. 2020. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. *Scientific data* 7, 1 (2020), 1–7.

[41] Rafael Prieto Curiel, Luca Pappalardo, Lorenzo Gabrielli, and Steven Richard Bishop. 2018. Gravity and scaling laws of city to city migration. *PLOS ONE* 13, 7 (07 2018), 1–19. https://doi.org/10.1371/journal.pone.0199892

[42] Rafael Reuveny. 2007. Climate change-induced migration and violent conflict. *Political geography* 26, 6 (2007), 656–673.

[43] Caleb Robinson and Bistra Dilkina. 2018. A Machine Learning Approach to Modeling Human Migration. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies* (Menlo Park and San Jose, CA, USA) *(COMPASS '18)*. Association for Computing Machinery, New York, NY, USA, Article 30, 8 pages. https://doi.org/10.1145/3209811.3209868

[44] 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. 2020. Assessing the impact of coordinated COVID-19 exit strategies across Europe. *Science* 369, 6510 (2020), 1465–1470. https://doi.org/10.1126/science.abc5096 arXiv:https://science.sciencemag.org/content/369/6510/1465.full.pdf

[45] Albert Ali Salah, Alex Pentland, Bruno Lepri, Emmanuel Letouzé, Patrick Vinck, Yves-Alexandre de Montjoye, Xiaowen Dong, and Ozge Dagdelen. 2018. Data for refugees: the D4R challenge on mobility of Syrian refugees in Turkey. *arXiv preprint arXiv:1807.00523* (2018).

[46] Yan Shi, Haoran Feng, Xiongfei Geng, Xingui Tang, and Yongcai Wang. 2019. A Survey of Hybrid Deep Learning Methods for Traffic Flow Prediction. In *Proceedings of the 2019 3rd International Conference on Advances in Image Processing* (Chengdu, China) *(ICAIP 2019)*. Association for Computing Machinery, New York, NY, USA, 133–138. https://doi.org/10.1145/3373419.3373429

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

[48] Alina Sirbu, Gennady Andrienko, Natalia Andrienko, Chiara Boldrini, Marco Conti, Fosca Giannotti, Riccardo Guidotti, Simone Bertoli, Jisu Kim, Cristina Ioana Muntean, et al. 2020. Human migration: the big data perspective. *International Journal of Data Science and Analytics* (2020), 1–20.

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

[50] Victor Soto, Vanessa Frias-Martinez, Jesus Virseda, and Enrique Frias-Martinez. 2011. Prediction of socioeconomic levels using cell phone records. In *International Conference on User Modeling, Adaptation, and Personalization*. Springer, 377–388.

[51] T. Tieleman and G. Hinton. 2012. Lecture 6.5—RmsProp: Divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks for Machine Learning.

[52] Vasiliki Voukelatou, Lorenzo Gabrielli, Ioanna Miliou, Stefano Cresci, Rajesh Sharma, Maurizio Tesconi, and Luca Pappalardo. 2020. Measuring objective and subjective well-being: dimensions and data sources. *International Journal of Data Science and Analytics* (2020). https://doi.org/10.1007/s41060-020-00224-2

[53] Zhecheng Wang, Haoyuan Li, and Ram Rajagopal. 2020. Urban2Vec: Incorporating Street View Imagery and POIs for Multi-Modal Urban Neighborhood Embedding. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 1013–1020.

[54] Peng Xie, Tianrui Li, Jia Liu, Shengdong Du, Xin Yang, and Junbo Zhang. 2020. Urban flow prediction from spatiotemporal data using machine learning: A survey. *Information Fusion* 59 (2020), 1 – 12. https://doi.org/10.1016/j.inffus.2020.01.002 arXiv:https://royalsocietypublishing.org/doi/pdf/10.1098/rsif.2014.1128

[55] Xiao-Yong Yan, Wen-Xu Wang, Zi-You Gao, and Ying-Cheng Lai. 2017. Universal model of individual and population mobility on diverse spatial scales. *Nature Communications* 8, 1 (2017), 1639. https://doi.org/10.1038/s41467-017-01892-8

[56] Xueyan Yin, Genze Wu, Jinze Wei, Yanming Shen, Heng Qi, and Baocai Yin. 2020. A Comprehensive Survey on Traffic Prediction. *arXiv preprint arXiv:2004.08555* (2020).

[57] Chao Zhang, Keyang Zhang, Quan Yuan, Fangbo Tao, Luming Zhang, Tim Hanratty, and Jiawei Han. 2017. ReAct: Online Multimodal Embedding for Recency-Aware Spatiotemporal Activity Modeling. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 245–254.

[58] Yahua Zhang, Anming Zhang, and Jiaoe Wang. 2020. Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. *Transport Policy* 94 (2020), 34 – 42. https://doi.org/10.1016/j.tranpol.2020.05.012

[59] George Kingsley Zipf. 1946. The P 1 P 2/D hypothesis: on the intercity movement of persons. *American sociological review* 11, 6 (1946), 677–686.