A Novel Excess Commuting Framework Considering the Commuting Efficiency and Equity

Abstract: Excess commuting, which concerns the differences between the current commute and the optimal commute, is of long-standing interest to transport policy-makers and planners. Regarding optimal commute calculation, the existing excess commuting framework has focused on measuring commuting efficiency but overlooking commuting (in)equity, which is defined as the variation in commuting cost of all workers in a given commuting pattern. Furthermore, existing excess commuting only considers the spatial arrangement between home and workplaces over space. The variation in commuting frequency of different workers over time has yet been paid much attention, which also affects the overall commuting cost. In this paper, we propose a novel excess commuting framework using a Greedy-Initialisation-based Genetic Algorithm (GI-GA), in which the optimal commute is calculated by considering commuting efficiency, equity and the variation in commuting frequency simultaneously. We illustrate and calibrate the framework using a one-month metro smart card data in Shanghai. Comparing with three other existing models, the GI-GA based framework is able to generate a commuting pattern that balances commuting efficiency and equity for better excess commuting study.

Keywords: excess commuting, commuting frequency, equity, genetic algorithm, smart card data

1 Introduction

Commuting, the recurrent trip between one's residence and workplace, accounts for a significant share of daily trips among urban residents. For a given set of job and housing locations, referred to as a fixed urban form, there are numerous commuting patterns. For non-optimal commuting patterns (i.e. excess commuting), the majority of commuters may live far away from their workplaces, i.e., only few enjoy a short commuting distance. This has caused some of the most challenging urban problems of our times:

traffic congestion, air pollution and even social inequality (Zhou and Long, 2016). Therefore, the study of optimal commute for excess commuting analysis remains an important concern for planners and policy-makers who care about environmental sustainability and social equity (Niedzielski et al., 2013).

The excess commuting framework has been used to study many urban issues for more than thirty years, such as job-housing imbalance problem (Hamilton, 1982; Horner, 2002; Yang and Ferreira, 2008; J Zhou et al., 2018; X Zhou et al., 2018). The concept of excess commuting is concerned with the empirical differences between the actual commuting pattern with the optimum commuting pattern to measure the commuting efficiency (Zhou & Long, 2016). Regarding excess commuting, the key step is to calculate the optimal commuting pattern, which, in other words, captures the potential for a city to reduce its overall commuting given a fixed urban form, assuming that jobs and workplaces are exchangeable among workers (Hu & Wang, 2015). There are two major issues of existing excess framework, summarised in two aspects.

First, in the existing excess commuting framework, the optimal commute is derived by minimising the system-wide daily overall commuting cost of workers in a city. Such cost can be measured by commuting distance or time, which reflect the spatial arrangement of homes and workplaces of workers (spatial domain). However, the framework overlooks the differences in commuting frequency over time (temporal domain). Due to various benefits, there are many different and flexible work schedules besides the traditional five-day work arrangement (Zhang & Cheng, 2019). In countries like the United States, the United Kingdom, Australia and Canada, many employees work four 10-hour days (called "compressed workweeks") to exchange for a weekday off every fortnight and even to work part time (e.g., 2 work days each week) continuously (Arbon et al., 2012). The variation in workdays among employees results in different commuting frequencies per week. This has significant impacts on the optimal commuting pattern.

Second, in the existing commuting framework, the optimal commute focuses on improving commuting efficiency (minimising the overall commuting cost) but overlooking commuting equity, which is understood as the fairness in commuting cost across among different commuters. If only the commuting efficiency is considered, we may find a more efficient but less equal commuting pattern than before the optimisation. To avoid this unwanted outcome, it is important for us to improve the level of efficiency and equity simultaneously as our optimisation objectives.

In this paper, we propose a novel excess commuting framework to fill the gaps mentioned above. Specifically, we propose a Greedy-Initialisation-Based Genetic Algorithm (GI-GA) to compute the optimum commute by considering efficiency and equity simultaneously. The GI-GA is also able to minimise the overall commuting cost that accounts for variations in commuting frequency and in the daily commuting distance simultaneously. To illustrate, we have conducted an empirical study that applies the GI-GA, with one-month metro smart card data (SCD) of Shanghai, China.

2 Literature review

The excess commuting is the most frequently used framework to study commuting efficiency. It was first introduced by Hamilton and Röell (1982), defined as the differences between the actual commute T_{act} and optimal commute T_{min}. Hamilton and Röell (1982) calculated the T_{min} using a monocentric model which was based on the assumption that workers chose the residences to maximise utility by trading off commuting and housing costs. White (1988) relaxed the assumption of monocentricity in Hamilton (1982) and employed the "Transportation Problem" algorithm, first proposed by Hitchcock (1941), to find the theoretical minimum commute T_{min}. In this approach, the commuters' locations of jobs and houses were exchanged in a way that minimised the system-wide travel cost (distance or time) given that the total number of commuters and jobs in each zone remained the same.

Built on the foundational work done by White (1988), many scholars have attempted to advance the excess commuting framework. Horner (2002) introduced the theoretical maximum commute T_{max}. He also termed the range between the T_{min} and T_{max} as the commute potential or carrying capacity. Charron (2007) suggested a 'commuting possibilities' framework to calculate the commuting upper bound T_{rand}. About the same time, Yang and Ferreira (2008) provided an alternative method to calculate the maximum commute, called proportionally matched commuting, which can be demonstrated to be equal with T_{rand}. Murphy and Killen (2011) expanded the T_{rand} by proposing the normalised commuting economy index to measure the commuting potential utilised. Jun et al. (2018) and Ha et al. (2018)

investigated the relationship between urban form (or spatial structure) and excess commuting using a comparative approach. Xu et al. (2019) evaluated the transport policies in Xiamen City, China, based on the excess commuting framework.

However, in the existing framework, the main approach to deriving the optimal commute is the Transportation Problem of Linear Programming (TPLP), which was first introduced by White (1988) into the study of excess commuting. Later, it was adopted by many other researchers (e.g., (Horner, 2002; Hu and Wang, 2015; Taaffe, 1996; Zhou and Long, 2016)). In TPLP, all commuters are considered as homogeneous and can be enticed to any jobs and/or residences without losing any utilities (Zhou & Long, 2016). Such homogeneity assumption, of course, can be challenged and some scholars have done so. O'Kelly and Lee (2005) and Zhou et al. (2018), for instance, disaggregated the commuters by job types when calculating various excess commuting indicators. Horner and Schleith (2012) considered the incomes of commuters to highlight the differences in people's accessibility to jobs. Schleith et al. (2016) examined the jobs-housing balance of different categories of workers across 26 metropolitan regions in the United States. However, in all above-mentioned excess framework, the optimal commute based on the TPLP is to minimise the daily overall commuting distances or time given an urban form, which is only about the 'spatial' relationships between workers' residential and job locations. To the best of our knowledge, no authors have considered the different commuting frequency of commuters, which reflects the variation of commuting behaviours in the 'temporal' domain. For example, the weekly commuting frequency implies 'when' commuting takes place during a week and it could vary considerably among different workers. Hence, a worker's weekly commuting distances do not only depend on the spatial distance between home and workplace, but also the commuting frequency. If we overlooked such variation when optimising the commute, we would introduce biases and even errors.

Additionally, existing optimal commute means only to optimise the commuting efficiency. However, commuting equity has been overlooked. Here, we propose that commuting equity is measured by the variation in the commuting cost among all workers. A commuting pattern with a smaller variation of commuting costs for all workers is more equitable than other patterns. Without consciously considering commuting equity, an 'optimised' commuting pattern simply based on TPLP compromises commuting equity. In this regard, Zhou and Long (2016) proposed a concept of ‘losers’, those who suffer from longer commuting distances in the optimal commute. However, what they were interested in was to examine the severity of the loss, the loss’ spatial patterns and its influencing factors. They didn’t consider commuting equity as we have defined above.

In light of the above, we contend that when ‘optimising’ commute for excess commuting study, one need to account for both the mean of, and variation in the commuting cost of all workers simultaneously, as well as the differences of workers’ commuting frequency. We formulate this optimisation problem as a dual response problem (DRP) as described in (Costa, 2010). DRP is a special case of multi-objective optimisation problem. Compared to TPLP (a linear programming problem), DRP is a non-linear programming problem. In most cases, the two objectives, minimising the mean while minimising the variation, are conflicting. Therefore, we need to find a ‘sweet spot’ between two objectives, that is, an acceptable solution that tolerates certain trade-off between the two objectives. To solve multi-objective optimisation problems and incorporate the heterogeneous commuting frequencies for optimisation, the evolutionary algorithms are often adopted, in particular, Genetic Algorithm (GA) (Konak et al., 2006). In this research, we also adopt and adapt GA to find the acceptable solution.

3 Methodology

3.1 Problem illustration

Commuting pattern optimisation at an aggregate zone-level is usually reformulated into a Transportation Problem of Linear Programming (Charron, 2007b; O’Kelly and Lee, 2005). However, the traditionally TPLP can only minimise the total commuting cost of all commuters given spatial arrangement of jobs and homes, overlooking the heterogeneous commuting frequencies of individuals in temporal scale and the equity concerns. In this study, we aim to simultaneously optimise the mean (efficiency) and variation (equity) of commuting cost considering the different commuting frequencies. We further illustrate the practical meanings using the examples in Figure 1.

First, the differences of commuters’ commuting frequencies have been overlooked in existing research. The overall commuting cost depends not only on the commuting distance over space but also commuting frequency over time. As shown in Figure 1(a), suppose there are two commuters P_1 and P_2

in the residential area H_1 with different commuting frequencies and there are two available workplaces W_1 and W_2.The minimum commute then achieves when letting P_1 work at W_1 and P_2 work at W_2, because pattern 2 has lower overall commuting cost than pattern 1 considering commuting frequency. Note that there is a risk to suppress the Additionally, Figure 1(b) further illustrates the importance of considering commuting equity. The two commuting patterns in Figure 1(b) have the same (or similar) level of efficiency, but we think the pattern 2 is better since its variation of commuting cost is smaller than that of the pattern 1, i.e., the pattern 2 is more equal. However, in practice, mean and variation can hardly achieve the minimum values simultaneously. We need to find the trade-off between them, which means we would like to sacrifice the mean cost, to a certain extent, in order to reduce the variation to ensure the equity.

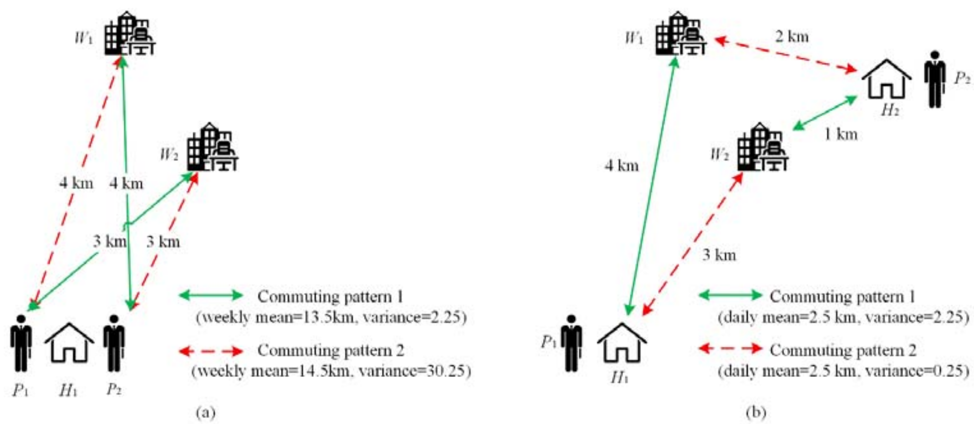

Figure 1. Optimal commuting pattern considering (a) commuting frequency; (b) commuting equity

3.2 Greedy-Initialisation-based Genetic Algorithm

In this paper, optimising the aggregated commuting pattern relies on the assumption that commuters can freely swap their workplaces in a city so as to minimise the overall commuting cost in the city, which is a widely adopted assumption in the existing relevant literature (Ma and Banister, 2006; Zhou and Long, 2016). In the following description, we assume that commuters could freely change their workplaces but live at fixed home locations. As stated previously, we aim to decrease both the mean and variation of commuting cost simultaneously considering the heterogeneity of commuting frequencies to optimise the zone-level commuting pattern. To solve this problem, a Greedy-Initialisation based Genetic Algorithm (GI-GA) model is proposed. The scope of application of the algorithm is not limited to this special problem, it also adapts to more generalised zone-level commuting optimisation with heterogeneous commuting patterns (e.g., different job types).

Figure 2 provides the general framework of GA. In operations research, GA is an evolutionary algorithm inspired by the process of natural selection following the principles of "survival of the fittest". An implementation of GA starts with a population of random chromosomes. Each individual among the population represents a point in a search space and a candidate solution (Step 1). Then the fitness of individuals can be evaluated according to an objective function (Step 2). Those individuals of higher fitness scores in "competition" will have higher possibility to be selected to produce the offspring than those performing poorly (Step 3). Mutation randomly occurs with a certain possibility (Step 4). During this process, the successive generation will become more suited to the objective function. However, in different circumstances, the data structure of solutions and initialisation, objective function, crossover, and mutation operation should be designed for specific problems. In the following subsections, we will introduce the steps in order.

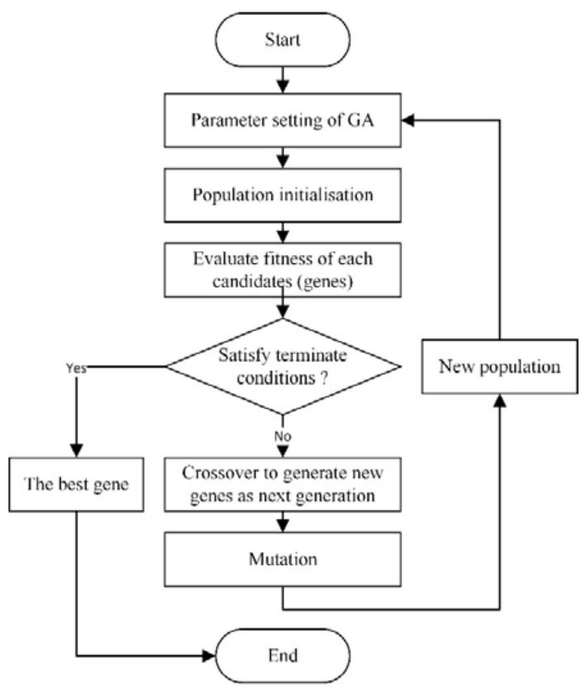

Figure 2. An overview of the process of Genetic Algorithm

Considering the commuting pattern optimisation, suppose there are N zones for residence and M zones for working, and the total population of commuters is P. The number of home locations and workplaces in each residential and working zone is $\boldsymbol{H} = \{H_1, H_2, \cdots, H_N\}$ and $\boldsymbol{W} = \{W_1, W_2, \cdots, W_M\}$, respectively. And it satisfies the following condition:

$$\sum_{i=1}^{N} H_i = \sum_{i=1}^{M} W_i = P \tag{1}$$

The origin-destination (OD) commuting cost matrix is denoted as $T = (t_{ij})(1 \le i \le N, 1 \le j \le M)$, where $t_{ij} \ge 0$ is the commuting cost (e.g., time or distance) from residential zone *i* to working zone *j*.

In GA, the structure of a solution is a matrix $V = (v_{ij})(1 \le i \le N, 1 \le j \le M)$, where v_{ij} indicates the number of commuters who commute from residential zone *i* to working zone *j*, and it should satisfy the following:

(1) $v_{ij} \ge 0$ for all $1 \le i \le N$ and $1 \le j \le M$

(2) $\sum_{j=1}^{M} v_{ij} = H_i$ for $1 \le i \le N$

(3) $\sum_{i=1}^{N} v_{ij} = W_j$ for $1 \le j \le M$

Similarly, the commuting frequency matrix for the *N* residential zones is denoted as $C = (c_{ik})(1 \le i \le N, 1 \le k \le 7)$, where c_{ik} means the number of commuters who live in the *i*-th residential zone and commute *k* times per week and it satisfies the following conditions:

(1) $c_{ik} \ge 0$ for all $1 \le i \le N$ and $1 \le k \le 7$

(2) $\sum_{k=1}^{7} c_{ik} = H_i$ for $1 \le i \le N$

3.2.1 Greedy initialisation

Traditionally, an initial solution to a "Transportation Problem" is a two-dimension matrix $V = (v_{ij})$ which should satisfy the above-mentioned constraints. However, GA with randomly generated solutions runs quite slowly to achieve an acceptable solution. To accelerate the optimisation process, we propose to use the greedy algorithm (Edmonds, 1971) to initialise the population of GA. A greedy algorithm is an algorithmic paradigm that makes the locally optimal choice at each stage with the objective to find a global optimum (Cormen et al., 2009). The greedy strategy for commuting optimisation is the following heuristic: At each stage, randomly select a residential zone, commuters living in that zone should be reassigned to working zones with available workplaces and nearest to the current residential zone.

3.2.2 Objective function and Fitness evaluation

The objective is to simultaneously optimise the mean (efficiency) and variation (equity) of the daily commuting cost of commuters, which is commonly referred to the Dual Response Problem (Costa, 2010). To achieve this goal, a popular approach is to aggregate the mean and variation into a single function (Messac et al., 2000). Thus, we propose the following objective function for GA:

$$objective = \max \frac{U_\mu - \hat{\mu}}{\hat{\sigma}^{2k}}, \text{ subject to } \hat{\mu} < U_\mu, \tag{2}$$

where $\hat{\mu}$ and $\hat{\sigma}^2$ is the mean and the variation of commuting cost, respectively. k is a weight coefficient to trade off the priority of $\hat{\mu}$ and $\hat{\sigma}^2$. Larger positive k indicates that the change of has more influence on the fitness. In addition, the U_μ is the upper bound of $\hat{\mu}$, which keeps the mean cost at an acceptable low value. The value of U_μ can be determined by the average commuting cost of the initial solutions generated by the greedy algorithm. This objective function allows assigning different weights to mean and variation to find trade-off solutions between them. Besides, via tuning the parameter k and U_μ, the priority of mean and variation can be adjusted during optimisation.

The value of the objective function is called "fitness". When considering different commuting frequencies, the objective can be calculated as follows. For the *i*-th residential zone, we expand the *i*-th row in the commuting frequency matrix *C* as an extended commuting frequency vector:

$$\begin{aligned} \boldsymbol{f}_i &= \left[f_{i1}, f_{i2}, \cdots, f_{iH_i} \right] \\ &= [\underbrace{1, \cdots, 1}_{c_{i1}}, \underbrace{2, \cdots, 2}_{c_{i2}}, \cdots, \underbrace{7, \cdots, 7}_{c_{i7}}] \end{aligned} \tag{3}$$

where H_i is the number of commuters in the *i*-th residential zone, f_{ij} $\left(1 \le i \le N, 1 \le j \le H_i\right)$ indicates the commuting frequency of the *j*-th commuter living in the *i*-th residential zone. Similarly, the extended commuting cost vector of the *i*-th row in the commuting cost matrix is:

$$\begin{aligned} \boldsymbol{d}_i &= \left[d_{i1}, d_{i2}, \cdots, d_{iH_i}, \right] \\ &= [\underbrace{t_{i1}, \cdots, t_{i1}}_{v_{i1}}, \underbrace{t_{i2}, \cdots, t_{i2}}_{v_{i2}}, \cdots, \underbrace{t_{iN}, \cdots, t_{iN}}_{v_{i7}}] \end{aligned} \tag{4}$$

where d_{ij} $\left(1 \le i \le N, 1 \le j \le H_i\right)$ is the commuting cost of the *j*-th commuter living in the *i*-th residential zone. To achieve the minimum commute, the commuter with lower frequency would be assigned to a farther workplace. We do so by sorting the extended commuting cost vector from largest to smallest, denoted as:

$$\begin{aligned} \boldsymbol{d}_i' &= \text{sorted}\left(\boldsymbol{d}_i\right) \text{ from largest to smallest} \\ &= \left[d_{i1}', d_{i2}', \cdots, d_{iH_i}' \right] \end{aligned} \tag{5}$$

After sorting, the total weekly commuting cost of commuters in *i*-th residential zone is easy to calculate as $\boldsymbol{f}_i \cdot \boldsymbol{d}_i'^T$. For the *j*-th commuter in residential zone *i*, the weekly total commuting cost is $d_{ij}' f_{ij} \left(1 \le i \le N, 1 \le j \le H_i\right)$. Then we use $d_{ij}' f_{ij} / 5$ (default five-day workweek) as an individual's commuting cost to calculate fitness value according to Eq. (2). When the fitness achieves convergence within the maximum number of GA iterations, it stops.

3.2.3 Crossover

Crossover is to vary the candidate solutions from one generation to the next. In each generation, parent solutions are randomly selected by using Roulette wheel selection method. Therefore, parent solutions with higher fitness can have higher possibility to produce the offspring than those perform poorly. Assume that two potential solutions $V^1 = \left(v_{ij}^1\right)$ and $V^2 = \left(v_{ij}^2\right)$ are selected as parents for the crossover operation. We adopt the following method proposed by Vignaux and Michalewicz (1991) to generate a pair of offsprings. First, create a divisibility matrix $DIV = \left(div_{ij}\right)$ and a remainder matrix $REM = \left(rem_{ij}\right)$ according to V^1 and V^2 as follows:

$$div_{ij} = \left\lfloor \left(v_{ij}^1 + v_{ij}^2\right) / 2 \right\rfloor \tag{6}$$

$$rem_{ij} = \left(v_{ij}^1 + v_{ij}^2\right) \bmod 2 \tag{7}$$

Obviously, the matrix *DIV* records the average value of the two parents and the matrix *REM* can tack the odevity of each value in $V^1 + V^2$. Then, *REM* is split into two matrices REM^1 and REM^2 satisfying the following conditions:

$$REM = REM^1 + REM^2 \tag{8}$$

$$sCol_i^1 = sCol_i^2 \text{ for all } 1 \le i \le N \tag{9}$$

$$sRow_j^1 = sRow_j^2 \text{ for all } 1 \le j \le M \tag{10}$$

where $sRow_j^k = \sum_{i=1}^{N} rem_{ij}^k$ and $sCol_i^k = \sum_{j=1}^{M} rem_{ij}^k$ are the marginal sum of rows of columns of REM^k $(k=1,\ 2)$ respectively. Then, the pair of children is generated as follows:

$$V^3 = DIV + REM^1 \tag{11}$$

$$V^4 = DIV + REM^2 \tag{12}$$

3.2.4 Mutation

The mutation is used to maintain genetic diversity from one generation to the next. Its operation on a $(N \times M)$ parent $V = (v_{ij})$ can be implemented as follows:

1) Randomly select *p* rows and *q* columns from *V*, and construct a $(p \times q)$ submatrix using the intersected cells of these selected rows and columns, denoted as $V' = (v'_{ij})$.

2) Similarly, extract the corresponding $(p \times q)$ submatrix from the original OD commuting cost matrix $T = (t_{ij})(1 \le i,\ j \le N)$, denoted as $T' = (t'_{ij})$.

3) Use the Greedy Initialisation method to reassign new values to the submatrix. Here, the three inputs should be $\sum_{j=1}^{q} v'_{ij}$, $\sum_{i=1}^{p} v'_{ij}$, and $T' = (t'_{ij})$, respectively.

3.3 Performance metrics

Here we propose several performance metrics to quantitatively evaluate models' performance from the perspective of efficiency and equity, ignoring/considering distinct commuting frequency, respectively.

3.3.1 Ignoring commuting frequency

(1) Efficiency metrics

- **EC:** excess commuting at the zone level. It is formulated by solving a TPLP model as the following:

$$EC = \left(\frac{T_{ac} - T_{min}}{T_{ac}} \right) \times 100 \tag{13}$$

where T_{ac} is the average actual commuting cost and T_{min} is the minimum commuting cost found by TPLP.

- **NPCT**: the number of commuters with positive commuting cost after commuting pattern optimisation.

- **Avg**: the average commuting cost for all commuters.

- **Avg_PCT**: the average commuting time for commuters with positive commuting cost (referred as "survival" commuters) after optimisation.

(2) Equity metrics

- **Var**: the variation of commuting cost for all commuters.

- **Gini**: Gini coefficient, first proposed by (Gini, 1936), is a well-known measure of inequality among values of a frequency distribution. Theoretically, it can range from 0 to 1, where a higher Gini coefficient indicates a more unequal (heterogeneous) distribution. According to its definition, if there are *n* commuters with commuting cost $t_1 \le t_2 \le \cdots \le t_n$, the Gini coefficient is:

$$G = \frac{1}{n}\left\{ n + 1 - 2\left[\frac{\sum_{i=1}^{n}(n+1-i)t_i}{\sum_{i=1}^{n} t_i} \right] \right\} = \frac{\sum_{i=1}^{n}(2i-n-1)t_i}{n\sum_{i=1}^{n} t_i} \tag{14}$$

- **Losers**: the number of losers after commuting optimisation. Losers are those who have to commute longer than the status quo (Zhou & Long, 2016).

- **AICTL**: average increment of commuting cost of losers.

- **MICTL**: maximum increment of commuting cost of losers.

3.3.2 Considering different commuting frequency

In this circumstance, if a commuter commutes *t* distances/minutes for *f* times per week, the average daily commuting cost is defined as *tf*/5 (default a five-day work mode). Then, the second group of metrics is proposed to evaluate the performance of models considering commuters' distinct commuting frequencies.

(1) Efficiency measurements

- **ECF**: excess commuting considering weekly commuting frequency.

- **AvgF**: the the average daily commuting cost of all commuters considering distinct weekly commuting frequency.
- **AvgF_PCT**: the average daily commuting cost of commuters with positive commuting cost after commuting optimisation considering distinct weekly commuting frequency.

(2) Equity metrics

- **VarF**: the variation the average daily commuting cost of all commuters considering distinct weekly commuting frequency.
- **VarF_PCT**: the variation commuting cost of commuters with positive commuting cost after commuting optimisation considering distinct weekly commuting frequency.
- **GiniF:** Gini coefficient, considering distinct weekly commuting frequency.

4 Case study

4.1 Study area and data

Shanghai, with a 24.26 million population, is one of the most populous cities in the world. In the case study, we focus on the metro commuters as metro is the most common public transit for daily commuting in Shanghai. Shanghai Metro System has 13 lines and 288 stations.

SCD, collected by automatic fare collection systems, can provide detailed onboard/outboard transactions of massive transit riders. The SCD used here covers a one-month period in April 2015, provided by Shanghai Open Data Apps (SODA) contest[*]. The dataset contains over 0.24 billion records made by about 11 million passengers. Trip records are extracted from the original dataset referring to (Wang, Correia, de Romph, & Timmermans, 2017). After pre-processing, about 122.6 million trip records of 10.5 million passengers can be identified. Each trip record contains the anonymous user ID, start/end stations, start/end time, and trip cost.

4.2 Metro commuters in Shanghai

Commuters are detected using the method proposed by Long and Thill (2015). About 1.82 million commuters are identified from the SCD. In further analysis, we use the commuting time as commuting

[*] http://soda.datashanghai.gov.cn

cost. The population of commuter flow between Origin-Destination (OD) and average commuting time, as well as the commuting frequency of individuals, can be computed using SCD.

4.2.1 Job-housing mismatch

Figure 3(a) shows the spatial distribution of job-housing ratio and Figure 3(b) displays the number of home locations versus the number of workplaces near each metro station. Both subfigures give an intuitive visualisation of job-housing spatial mismatch and imbalance.

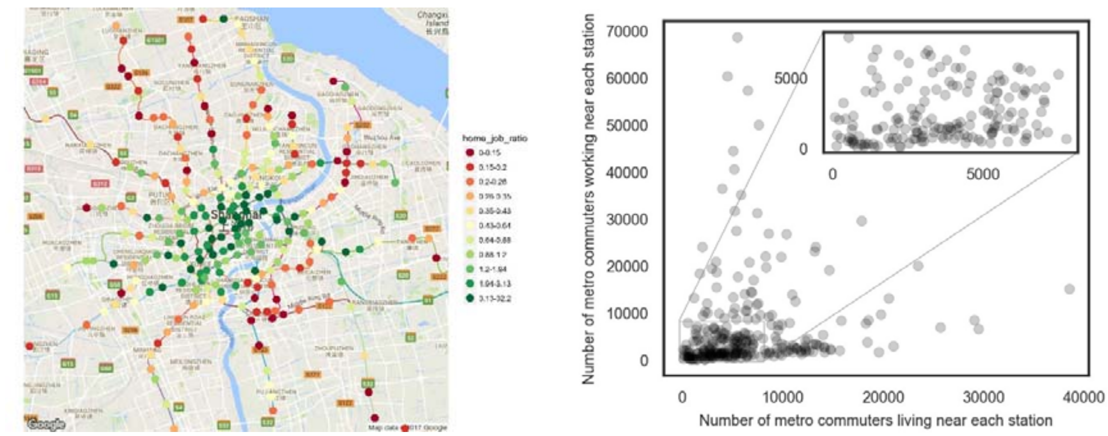

Figure 3. Job-housing ratio distribution (a) and the scatter plot (b) of metro commuters living/working near each station

4.2.2 Commuting time and frequency

The average and median commuting time of detected metro commuters is about 34.60 and 32.18 minutes, respectively. A day is regarded as a commuting day if at least one commuting between home and work can be detected from SCD to eliminate interference of the before-work or after-work activities. For example, if parents need to send their children to school before going to work, the home-to-work trip cannot be found. After that, the maximum weekly commuting days during the study period are taken as a commuter's weekly commuting frequency. This is because commuters may occasionally change the commuting mode (e.g., taxi) or sometimes the before- and after-work activities happened on the same day. Finally, the proportion of commuters with different weekly commuting frequency is 0.18% (once), 4.67% (twice), 12.39% (three times), 23.33% (four times), 44.54% (five times), 11.40% (six times), 3.49% (seven times), respectively. Notably, the five-day workweek is the most common work mode, accounting for about 45% of the whole commuters. However, other commuting patterns cannot be overlooked.

4.3 Optimisation results and comparison

4.3.1 Parameter settings for GI-GA

For GI-GA model, we set 20 as the size of the initial population, in which each gene is a 288×288 matrix. We let crossover rate and mutation rate to be 0.9 and 0.8, respectively, the size of submatrix in mutation step to be 50×50, and the maximum iteration to be 1000 times. As aforementioned, the parameter settings of the objective function have impacts on the trade-off between mean and variation of commuting time. Thus, we fix U_{μ} to be 17.5 and compare three different values of parameters *k* of the objective function, denoted as Efficiency-First (EFF, *k*=0.5), Equity-First (EQF, *k*=5) and Efficiency-Equity-Simultaneous (EES, *k*=2), respectively. Figure 4 presents the change of the mean and the variation of commuting time during iteration. A decreasing trend of the mean but a rising trend of the variation of EEF can be observed over interation. The opposite tendency appeared under EQF settings. For EES model, it is worth noting that the mean and variation declined alternatively, and both decreasing amplitudes were larger than EEF and EQF. It means this parameter settings speed up the optimisation process. In the following experiments, we let *k* be 2.

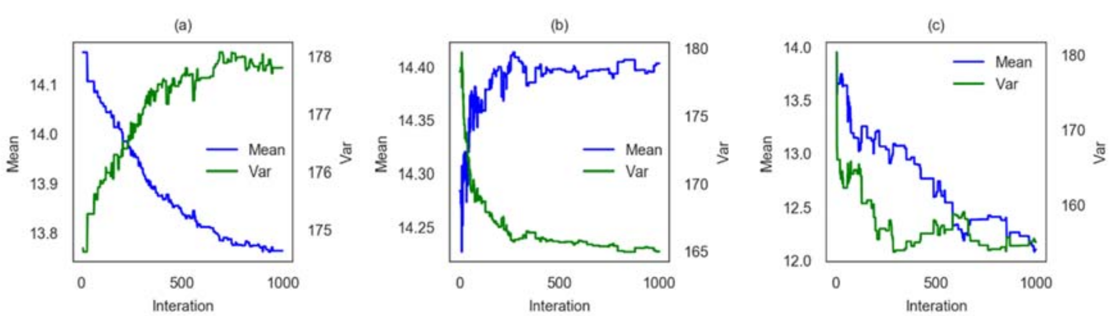

Figure 4. The change of mean and variation of commuting time during optimisation by GI-GA with different parameter settings (a) EFF; (b) EQF; (c) EES.

4.3.2 Comparison results

Here, we compare the optimised commuting pattern generated by our model with another three models, as well as the original commuting patterns in Shanghai. The description of each model is given below:

- **Status quo**: current commute in Shanghai.
- **TPLP**: Optimised commuting pattern using TPLP by minimising mean commuting time as the only goal we care about.
- **GI-GA-V**: Optimised commute using GI-GA algorithm, and the objective function is just to minimise the variation of commuting time, ignoring commuting frequency.
- **GI-GA-MV**: Optimised commute using GI-GA algorithm, and the objective function is to iminimise the mean and variation of commuting time simultaneously, but ignoring commuting frequency.
- **GI-GA-MVF**: Optimised commute using GI-GA algorithm, and the objective function is to minimise the mean and the variation of commuting time simultaneously considering distinct weekly commuting frequency.

4.3.2.1 Comparison results ignoring commuting frequency

Table 1 shows the comparison results by using the first group of metrics. Summarily, since TPLP can generate the optimal solution when aiming to minimise the mean commuting time alone, TPLP perform best on most of the efficiency metrics (NPCT, Avg) without considering commuting frequency. The excess commuting decreases from 63.73% to around 13% by using three GI-GA based model. However, after optimisation, numerous commuters can work near home without commuting by metro.Additionally, TPLP performs the worst in terms of Avg_PCT, which indicates that using TPLP, the ‘survival’ commuters need to travel longer than using GI-GA models.

Table 1. Model comparison results without considering commuting frequency

Metrics		Status quo	TPLP	GI-GA (V)	GI-GA (MV)	GI-GA (MVF)
Efficiency	**EC**	63.73%	0	12.42%	8.67%	9.06%
	NPCT	1829622	941259	1509281	1421291	1362413
	Avg	34.60	12.55	14.33	13.74	13.80
	Avg_PCT	34.60	24.39	18.58	18.93	18.87
Equity	**Var**	292.24	206.86	177.26	173.07	195.8
	Gini	0.28	0.61	0.47	0.49	0.50
	Losers	-	39711	147346	155286	3266
	AICTL	-	5.67	10.75	10.20	6.38
	MICTL	-	37.68	61.68	62.28	27.78

The last five indicators in Table 1 show the comparison results in terms of commuting equity. The variation of commuting time can be reduced from 292.24 to about 200 by TPLP or GI-GA-MVF, and to around 175 by GI-GA-V or GI-GA-MV. Contrarily, the Gini coefficient rises up after optimisation, which is because the majority of commuters in original commuting patterns need to commute about 20~40 minutes, but after optimisation, considerable commuters do not need to take metro anymore. The most exciting result is about 'losers'. From Table 1, we can see the number of losers after GI-GA-MVF optimisation is just about 3000 out of total 1.82 million commuters, far less than other models. To further illustrate, Figure 5 visualises 1166 commuters living near Wuzhou Avenue Station in Shanghai before and after TPLP and GI-GA-MVF optimisation. After re-distributing commuters' workplaces by TPLP, 359 commuters who currently work near those stations within the yellow circles in Figure 5(b) become the losers. Meanwhile, no losers were found after GI-GA-MVF optimisation. Additionally, AICTL and MICTL are used to examine the severity of the loss among all losers. The loss is defined as the increments of commuting time, as shown in Figure 6. AICTLs of TPLP and GI-GA-MVF are smaller than the others, and GI-GA-MVF performs best on MICTL. This indicates even if there are losers after optimisation, our model effectively limit the loss of losers.

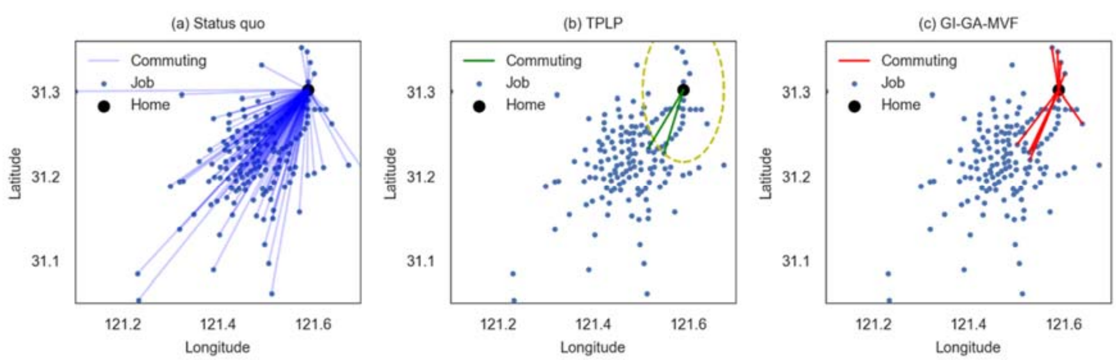

Figure 5. 1,166 commuters living near Wuzhou Avenue Station (a) before optimisation; (b) after TPLP optimisation (359 losers); (c) GI-GA-MVF optimisation (no losers)

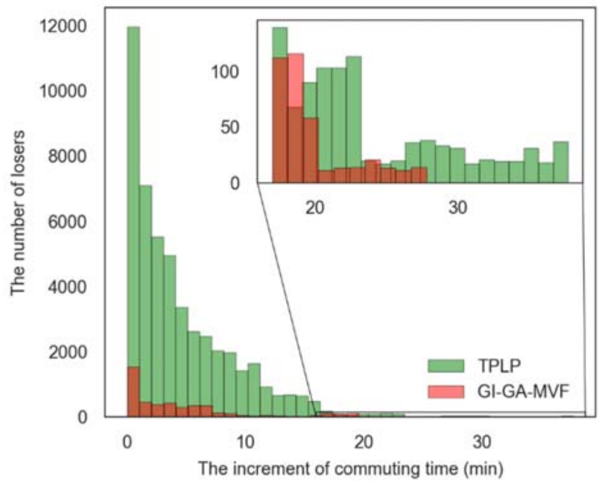

Figure 6. The histogram of loss of all losers in TPLP and GI-GA-MVF model

4.3.2.2 Comparison results considering commuting frequency

Table 2 shows the strength of our proposed model in tems of efficiency and equity optimisation considering different commuting frequency against others. First, comparing with TPLP, the average commuting time of GI-GA-MVF is slightly longer, but AvgF_PCT of GI-GA-MVF is considerably better than that of TPLP. As for equity metrics, TPLP is notably inferior to GI-GA-MVF. Then, clearly, GI-GA-MVF thoroughly precedes the other two GI-GA based models on all performance metrics. Overall, the proposed framework can be used to derive a better metirc for excess commuting by considering commuting frequency and commuting equity. Additionally, the excess commuting increases in terms of ECF, which implies that overlooking the commuting frequency may underestimate excess commuting. This result also shows that the commuting frequency has great impacts on commuting costs. Thus, it shall not be overlooked in the excess commuting framework.

Table 2. Model comparison results considering commuting frequency

Metrics		**Status quo**	**TPLP**	**GI-GA-V**	**GI-GA-M**	**GI-GA-MVF**
Efficiency	**ECF**	66.90 %	-	23.61%	20.51%	8.09%
	AvgF	31.34	10.34	13.58	13.05	11.25
	AvgF_PCT	31.34	20.16	16.46	16.80	15.84
Equity	**VarF**	298.07	150.93	158.14	157.75	112.40
	VarF_PCT	298.07	95.93	144.25	140.08	86.81
	GiniF	0.30	0.62	0.49	0.51	0.49

Finally, Figure 7 provides a visual comparison among the city-wide commuting patterns without optimization, optimised by TPLP or GI-GA-MVF. Comparing with the original commute, TPLP and our model can obviously decompose the excess commuting. Comparing Figure 7(b) and (c), an interesting finding is that the commuters who live extremely far away from the central area are prone to becoming ‘losers’ after optimisation (e.g. areas within red-dashed circles). However, our model can let most of those commuters work near their home locations. This once again confirms the advantage of our model.

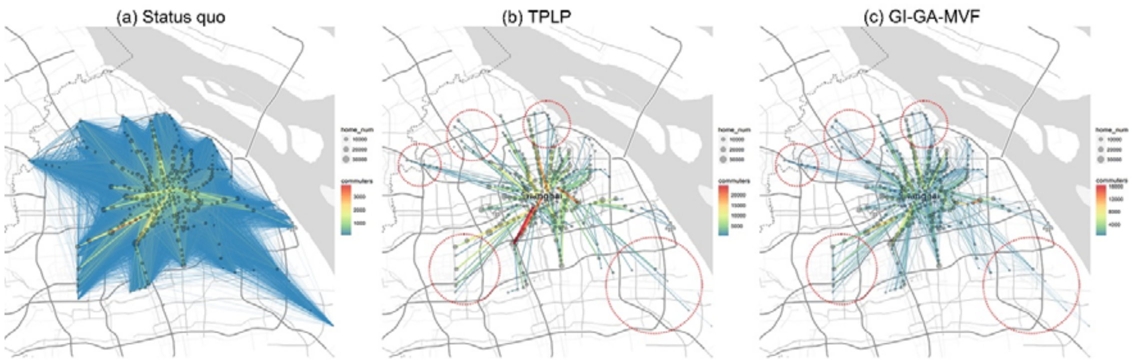

Figure 7. The city-wide commuting networks: (a) status quo, (b) optimized by TPLP, and (c) optimized by GI-GA-MVF (line width is proportional to commuter flow volume, and bubble size displays the number of commuters living near each metro station)

5 Conclusion and future directions

Commuting, defined as the trip between a worker’s residence and workplace, accounts for a significant share of daily trips. Commuting pattern optimisation in cities remains an important concern for transport planners and policy-makers for excess commuting measurement. In this paper, we propose a novel GI-GA model to determine the optimal commute. The GI-GA is able to minimise the mean and variation of commuting cost both considering the spatial arrangement of job and works and the heterogeneous commuting frequency in temporal domain. The optimised commute also balances the efficiency and equity of commuting, providing a better benchmark for excess commuting analysis. The case study of Shanghai shows that the proposed method can effectively improve the fairness during the commuting pattern optimisation process.

Despite of the above methodological advancements and empirical findings, we still feel that there is room for improvements in the future. First, at the current stage of our study, we only focus on optimising commuting patterns of metro commuters. To better understanding the general commuting patterns in cities, it is also important to include commuters using different travel modes (such as bus, bicycle, and walk). Second, in this study, each commuter is assumed to have only one job and we are unable to consider multiple-job commuters. In the future, more complicated scenarios should be studied, such as commuters with multiple jobs, different occupations or levels of income. And existing commuting framework only considers the access to current jobs, but not the access to future job opportunities. Future research should the latter to better reflect worker's residential choice with respect to both current and future job accessibility. One possible way to materialise such goals is to combine SCD and traditional survey data to access more socioeconomic information about travellers. Additionally, the commuters' characteristics and their variations can be reflected not only at a zone level but also at an individual level: each commuter is rightfully regarded as having specific commuting frequency and socioeconomic attributes (Hu and Wang, 2015). And finally, we can conduct comparative analyses across different cities so as to identify generic and localised factors conductive to commuting efficiency and/or equity.

References

Arbon CA, Facer RL and Wadsworth LL (2012) Compressed Workweeks - Strategies for Successful Implementation. *Public Personnel Management*, SAGE PublicationsSage CA: Los Angeles, CA 41(3): 389–405. Available from: http://journals.sagepub.com/doi/10.1177/009102601204100301 (accessed 6 August 2018).

Charron M (2007a) From excess commuting to commuting possibilities: More extension to the concept of excess commuting. *Environment and Planning A* 39(5): 1238–1254. Available from: http://journals.sagepub.com/doi/abs/10.1068/a3897 (accessed 10 June 2017).

Charron M (2007b) From excess commuting to commuting possibilities: More extension to the concept of excess commuting. *Environment and Planning A*, Elsevier Ltd 39(5): 1238–1254. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2015.03.002.

Cormen TH, Leiserson CE, Rivest RL, et al. (2009) *Introduction to Algorithms*. MIT press. Available from: https://books.google.co.uk/books?hl=zh-CN&lr=&id=aefUBQAAQBAJ&oi=fnd&pg=PR5&dq=+Introduction+to+algorithms.+&ots=dN5qTwXJd0&sig=OGKeJt8w6hz49U_KB_qJxyT98Mc&redir_esc=y#v=onepage&q=Introduction to algorithms.&f=false (accessed 6 August 2018).

Costa NRP (2010) Simultaneous Optimization of Mean and Standard Deviation. *Quality Engineering*, Taylor & Francis Group 22(3): 140–149. Available from: http://www.tandfonline.com/doi/abs/10.1080/08982110903394205 (accessed 6 August 2018).

Edmonds J (1971) Matroids and the greedy algorithm. *Mathematical Programming*, Springer-Verlag 1(1): 127–136. Available from: http://link.springer.com/10.1007/BF01584082 (accessed 6 August 2018).

Gini C (1936) On the Measure of Concentration with Special Reference to Income and Statistics. *Colorado College Publication: General Series* 208: 73–79. Available from: https://www.jstor.org/stable/2964918?origin=crossref (accessed 6 August 2018).

Ha J, Lee S and Kwon SM (2018) Revisiting the Relationship between Urban Form and Excess Commuting in US Metropolitan Areas. *Journal of Planning Education and Research*: 1–18. Available from: http://journals.sagepub.com/doi/10.1177/0739456X18787886 (accessed 10 April 2019).

Hamilton BW (1982) Wasteful Commuting. *Journal of Political Economy* 90(5): 1035–1053. Available from: http://www.journals.uchicago.edu/doi/10.1086/261107 (accessed 1 November 2017).

Hitchcock FL (1941) The Distribution of a Product from Several Sources to Numerous Localities. *Journal of Mathematics and Physics* 20(1–4): 224–230. Available from: http://doi.wiley.com/10.1002/sapm1941201224 (accessed 1 November 2017).

Horner MW (2002) Extensions to the Concept of Excess Commuting. *Environment and Planning A* 34(3): 543–566. Available from: http://journals.sagepub.com/doi/10.1068/a34126 (accessed 1 November 2017).

Hu Y and Wang F (2015) Decomposing excess commuting: A Monte Carlo simulation approach. *Journal of Transport Geography* 44: 43–52. Available from:

http://www.sciencedirect.com/science/article/pii/S0966692315000423 (accessed 10 June 2017).

Jun M-J, Choi S, Wen F, et al. (2018) Effects of urban spatial structure on level of excess commutes: A comparison between Seoul and Los Angeles. *Urban Studies* 55(1): 195–211. Available from: http://journals.sagepub.com/doi/10.1177/0042098016640692 (accessed 10 April 2019).

Konak A, Coit DW and Smith AE (2006) Multi-objective optimization using genetic algorithms: A tutorial. *Reliability Engineering & System Safety*, Elsevier 91(9): 992–1007. Available from: https://www.sciencedirect.com/science/article/pii/S0951832005002012 (accessed 6 August 2018).

Long Y and Thill JC (2015) Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing. *Computers, Environment and Urban Systems*, Elsevier Ltd 53: 19–35. Available from: http://dx.doi.org/10.1016/j.compenvurbsys.2015.02.005.

Ma K-R and Banister D (2006) Extended Excess Commuting: A Measure of the Jobs-Housing Imbalance in Seoul. *Urban Studies* 43(11): 2099–2113. Available from: http://journals.sagepub.com/doi/10.1080/00420980600945245 (accessed 31 October 2017).

Messac A, Sundararaj GJ, Tappeta R V., et al. (2000) Ability of Objective Functions to Generate Points on Nonconvex Pareto Frontiers. *AIAA Journal* 38(6): 1084–1091. Available from: http://arc.aiaa.org/doi/10.2514/2.1071 (accessed 6 August 2018).

Murphy E and Killen JE (2011) Commuting Economy: An alternative Approach for Assessing Regional Commuting Efficiency . *Urban Studies* 48(6): 1255–1272. Available from: http://usj.sagepub.com/content/48/6/1255.abstract.

Niedzielski MA, Horner MW and Xiao N (2013) Analyzing scale independence in jobs-housing and commute efficiency metrics. *Transportation Research Part A: Policy and Practice*, Elsevier Ltd 58: 129–143. Available from: http://dx.doi.org/10.1016/j.tra.2013.10.018.

O'Kelly ME and Lee W (2005) Disaggregate Journey-to-Work Data: Implications for Excess Commuting and Jobs–Housing Balance. *Environment and Planning A* 37(12): 2233–2252. Available from: http://journals.sagepub.com/doi/10.1068/a37312 (accessed 1 November 2017).

Schleith D, Widener M and Kim C (2016) An examination of the jobs-housing balance of different categories of workers across 26 metropolitan regions. *Journal of Transport Geography*, Elsevier

Ltd 57: 145–160. Available from: http://dx.doi.org/10.1016/j.jtrangeo.2016.10.008.

Taaffe E (1996) *Geography of Transportation*. Morton O'kelly. Available from: https://books.google.com/books?hl=zh-CN&lr=&id=N60qf7WynaEC&oi=fnd&pg=PR1&dq=Geography+of+transportation&ots=TPhR0KBHNo&sig=1ONwyW4Dp2IDKas9wtg8FJRgFaA (accessed 14 February 2019).

Vignaux GA and Michalewicz Z (1991) A genetic algorithm for the linear transportation problem. *IEEE Transactions on Systems, Man, and Cybernetics* 21(2): 445–452. Available from: http://ieeexplore.ieee.org/document/87092/ (accessed 6 August 2018).

White MJ (1988) Urban Commuting Journeys Are Not "Wasteful" *Journal of Political Economy* 96(5): 1097–1110. Available from: http://www.journals.uchicago.edu/doi/10.1086/261579 (accessed 1 November 2017).

Xu W, Yang L and Zhang W (2019) Evaluation of transport policy packages in the excess commuting framework: The case of Xiamen, China. *Cities*, Pergamon 87: 39–47. Available from: https://www.sciencedirect.com/science/article/pii/S0264275118310813 (accessed 10 April 2019).

Yang J and Ferreira J (2008) Choices versus choice sets: A commuting spectrum method for representing job - Housing possibilities. *Environment and Planning B: Planning and Design* 35(2): 364–378.

Zhou J and Long Y (2016) Losers and Pareto optimality in optimising commuting patterns. *Urban Studies* 53(12): 2511–2529. Available from: http://journals.sagepub.com/doi/10.1177/0042098015594072 (accessed 31 October 2017).

Zhou J, Murphy E and Corcoran J (2018) Integrating road carrying capacity and traffic congestion into the excess commuting framework: The case of Los Angeles. *Environment and Planning B: Urban Analytics and City Science*, SAGE PublicationsSage UK: London, England: 239980831877376. Available from: http://journals.sagepub.com/doi/10.1177/2399808318773762 (accessed 6 August 2018).

Zhou X, Yeh AG, Li W, et al. (2018) A commuting spectrum analysis of the jobs–housing balance and self-containment of employment with mobile phone location big data. *Environment and Planning B: Urban Analytics and City Science* 45(3): 434–451. Available from:

http://journals.sagepub.com/doi/10.1177/2399808317707967 (accessed 31 October 2017).

Costa, N. R. P. (2010). Simultaneous optimization of mean and standard deviation. *Quality Engineering*, *22*(3), 140-149.

Hamilton, B. W., & Röell, A. (1982). Wasteful commuting. *Journal of Political Economy, 90*(5), 1035-1053.

Horner, M. W. (2002). Extensions to the Concept of Excess Commuting. *Environment and Planning A*, *34*(3), 543-566. doi:10.1068/a34126

Hu, Y., & Wang, F. (2015). Decomposing excess commuting: a Monte Carlo simulation approach. *Journal of Transport Geography, 44*, 43-52. doi:https://doi.org/10.1016/j.jtrangeo.2015.03.002

O'Kelly, M. E., & Lee, W. (2005). Disaggregate journey-to-work data: implications for excess commuting and jobs–housing balance. *Environment and planning A*, *37*(12), 2233-2252.

Wang, Y., Correia, G. H. d. A., de Romph, E., & Timmermans, H. J. P. (2017). Using metro smart card data to model location choice of after-work activities: An application to Shanghai. *Journal of Transport Geography*, *63*(Supplement C), 40-47. doi:https://doi.org/10.1016/j.jtrangeo.2017.06.010

White, M. J. (1988). Urban commuting journeys are not" wasteful". *Journal of Political Economy*, *96*(5), 1097-1110.

Zhang, Y., & Cheng, T. (2019). A Deep Learning Approach to Infer Employment Status of Passengers by Using Smart Card Data. *IEEE Transactions on Intelligent Transportation Systems*. doi:10.1109/TITS.2019.2896460

Zhou, J., & Long, Y. (2016). Losers and Pareto optimality in optimising commuting patterns. *Urban Studies*, *53*(12), 2511-2529.