

1 **FORECASTING THE PEAK-PERIOD STATION-TO-STATION**
2 **ORIGIN-DESTINATION MATRIX IN URBAN RAIL TRANSIT SYSTEM:**
3 **CASE STUDY OF CHONGQING, CHINA**

4
5
6
7 **Yan Cheng, Corresponding Author**

8 Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji
9 University

10 4800 Cao'an Road, Jiading District, Shanghai, P.R. China, Post Code 201804

11 Tel: (86)021-69583757; Email: patty_1234@126.com

12
13 **Xiafei Ye**

14 Key Laboratory of Road and Traffic Engineering of Ministry of Education, Tongji
15 University

16 4800 Cao'an Road, Jiading District, Shanghai, P.R. China, Post Code 201804

17 Tel: (86)021-69589875 Fax: (86)021-69583712; Email: yxf@tongji.edu.cn

18
19 **Lifeng Zhou**

20 Chongqing Rail Transit (Group) Co., Ltd.

21 West Jinkai Avenue, Yubei District, Chongqing, P.R. China, Post Code 401120

22 Email: 70106925@qq.com

23
24
25 Word Count: 5,030 words text + 9 tables/figures x 250 words (each) = 7,280 words

26
27 Number of References 13

28 Number of Figures 5

29 Number of Tables 4

30
31
32
33
34
35
36 Submission Date: 7/31/2017

1 **ABSTRACT**

2 The maximum one-direction section passenger flow within peak hour is an important
3 indicator for planning and design of urban rail transit. To determine it, it is necessary
4 to forecast passengers' departure time and route choice during peak period. As the
5 basis of this process, the peak-period station-to-station origin-destination (OD) matrix
6 reflects the passengers' travel needs. This paper tests traditional gravity models in
7 forecasting the peak-period station-to-station OD matrix in urban rail transit with a
8 real-world case study of Chongqing, China. To solve its over-estimation when
9 deterrence between two stations is too little, the gravity-model-based Peak Period
10 Coefficient (PPC) model is introduced. Comparing results show that with the same
11 dataset, the PPC model is superior to the gravity model. Its standard deviation is only
12 12.90 passengers, reduced by 56.02%.

13

14

15

16

17 *Keywords:* Urban rail transit, Station-to-station ridership, Gravity model, Peak period
18 coefficient, Deterrence function

1 INTRODUCTION

2 With its rapid growth, urban rail transit is now playing a more important role in
3 supporting and promoting city development. By the end of 2016, there are 133
4 operating lines of 30 cities in mainland of China, reaching the total length of 4152.8
5 km. The annual patronage reaches 16.1 billion passengers, increasing 16.6% than the
6 year of 2015. There are 5636.5 km of lines under construction in 48 cities.

7 The planning and design of urban rail transit need foresight because of its high
8 construction cost and big difficulty to modify after construction. As a key indicator in
9 planning and design phase, the maximum one-direction section passenger flow
10 (MSPF) within peak hour directly influences the determination of vehicle selection
11 and train formation. It is accumulated by the passengers who depart from different
12 stations before this section and pass through it during peak hour, whose departure time
13 should be before or during peak hour. Conventionally, this indicator is forecasted
14 through multiplying all-day passenger flow of this section by a peak hour coefficient,
15 which depends on the line attribute. However, this method hasn't taken the forming
16 mechanism of MSPF within peak hour into account, so it cannot reflect the complex
17 process.

18 In order to overcome this shortcoming, it is necessary to extend the time range
19 from peak hour to peak period, and forecast passengers' departure time and route
20 choice during this time range. For this dynamic passenger assignment process,
21 time-varying peak-period station-to-station origin-destination (OD) matrices are
22 critical inputs, because they represent the travel needs of passengers at different
23 moments. However, most of previous studies on passenger assignment model for
24 urban rail transit assumed that time-varying peak-period station-to-station OD
25 matrices were given (1-3). This becomes invalid at the planning and design stage,
26 because the actual operational data hasn't generated. To forecast time-varying
27 peak-period station-to-station OD matrices, the first step is forecasting the
28 peak-period station-to-station OD matrix, which is of fundamental importance. Errors
29 in it are transferred onto time-varying peak-period station-to-station OD matrices,
30 leading to errors in MSPF within peak hour ultimately.

31 There is little research on the forecasting of the peak-period station-to-station
32 OD matrix at present. Multiplicative models are adopted by (4) and (5). The former
33 model chose 32 factors, which can be categorized into 4 areas: built-environment,
34 travel impedance, intermodal connection and other variables. The model was
35 calibrated with the data of Seoul Metro in Korea. Meanwhile, the latter one selected
36 22 variables which can be divided into 4 areas: land use, intermodal connection,
37 station context and travel impedance, estimated with data of Metro system in Nanjing,
38 China. Although this kind of models performs well according to regression results,
39 they are totally different from traditional trip distribution models. The theoretical
40 foundation is relatively weak, which makes production and attraction constraints
41 cannot be guaranteed. In addition, it is difficult to acquire some data (such as several
42 intermodal connection variables) at the planning and design stage.

1 Traditional trip distribution models concentrate on trips between traffic
2 analyses zones (TAZs) of all modes, whose time range is a whole day. One of the
3 most important models is the gravity model (6,7), which is widely used because it is
4 simple in concept and has been well documented. Nevertheless, it has several
5 drawbacks such as it may overestimate when the distance between origin and
6 destination is too short. It still needs to be verified whether the gravity model is
7 suitable for the forecasting of the peak-period station-to-station OD matrix in urban
8 rail transit. Apart from traditional trip distribution models, disaggregate destination
9 choice models based on utility-maximization theory have been proposed by some
10 scholars in recent years to forecast OD matrix of all modes (8-11). This kind of
11 models has a good performance; however, the data requirement is too high for
12 forecasting OD matrix in urban rail transit. It cannot be clear who will choose urban
13 rail transit at the planning and design stage, let alone the investigation of their
14 information such as socio-economic characteristics and travel characteristics. So these
15 models are not taken into consideration. In summary, the forecasting of urban rail
16 transit peak-period station-to-station OD matrix still needs more in-depth study.

17 A long-term all-day station-to-station OD matrix with relatively high accuracy
18 can be forecasted by planners through traditional four-step procedure, which is not
19 time-dependent. So it is assumed that the all-day station-to-station OD matrix is given.
20 Based on this assumption, the paper tests traditional gravity models in forecasting the
21 peak-period station-to-station OD matrix in urban rail transit with a real-world case
22 study of Chongqing, China, and analyzes their merits and demerits. To solve the
23 drawbacks, an improved forecasting model called Peak Period Coefficient (PPC)
24 model is proposed. According to this model, the peak-period station-to-station OD
25 matrix is obtained by multiplying the given all-day station-to-station OD matrix by
26 forecasting PPC, whose expression is deduced from the gravity model. Models with
27 different configurations are estimated with the same dataset, and their performance is
28 compared with the ones of gravity models.

29 The rest of this paper is organized as follows: Section 2 describes the data of
30 Chongqing rail transit. Section 3 presents the performance of gravity models and the
31 proposal of the improved model. Section 4 compares the results of these models.
32 Section 5 provides conclusions.

33 **DATA AND SCOPE**

34 **Case Study Object**

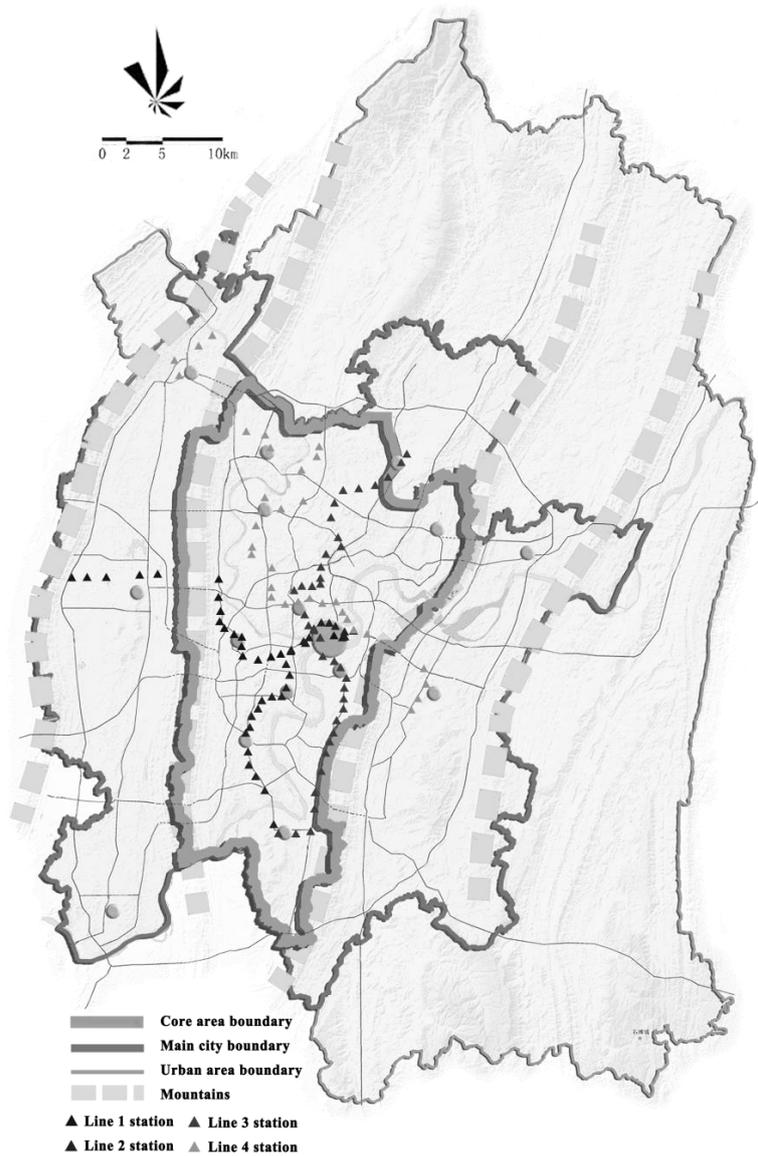
35
36 Chongqing is located in the southwest part of China, covering an area of 82400 km²,
37 650 km² of which is the built area of the main city. The registered population of
38 Chongqing is 33.752 million while the permanent resident is 29.914 million, 818.98
39 million of which belongs to the main city.
40

41 By the end of 2016, the urban rail transit network of Chongqing consists of 4
42 rail transit lines. Line 1 and Line 6 are metro systems, while Line 2 and Line 3 are

1 monorail systems. All lines are located in the main city, the total operating mileage of
 2 which reaches 202 km, with 120 stations included (Figure 1). After removing the
 3 stations with zero ridership and merging interchange stations, 110 stations are left.
 4 Due to the significant changes in land use around Guangdianyuan station and a high
 5 proportion of undeveloped land around Jiuquhe station, 108 stations are left on the list
 6 of study objects eventually.

7

8



9

10 **FIGURE 1 Chongqing rail transit stations.**

11

12 **Data Source**

13 To compare the performance of different models in forecasting the peak-period
 14 station-to-station OD matrix, the same dataset is used in this study. It mainly includes
 15 3 parts.

- 1 1. Prerequisites: The all-day station-to-station OD matrix of the study
- 2 objects;
- 3 2. Dependent variable: The peak-period station-to-station OD matrix of the
- 4 study objects;
- 5 3. Independent variables: The ridership of study objects in different periods,
- 6 the time shortest path between any two stations and its interchange times.

7 According to the automated fare collection (AFC) data of August 26, 2015
 8 provided by Chongqing Rail Transit (Group) Co., Ltd., the station-to-station OD
 9 matrices within all-day and peak period as well as the ridership of study objects are
 10 counted. It is worth noting that in order to ensure the consistency of data, one trip is
 11 classified into different periods according to its boarding time, no matter when it ends.
 12 Statistic shows that there are 8669 pairs of OD stations with ridership larger than 0
 13 during the peak period.

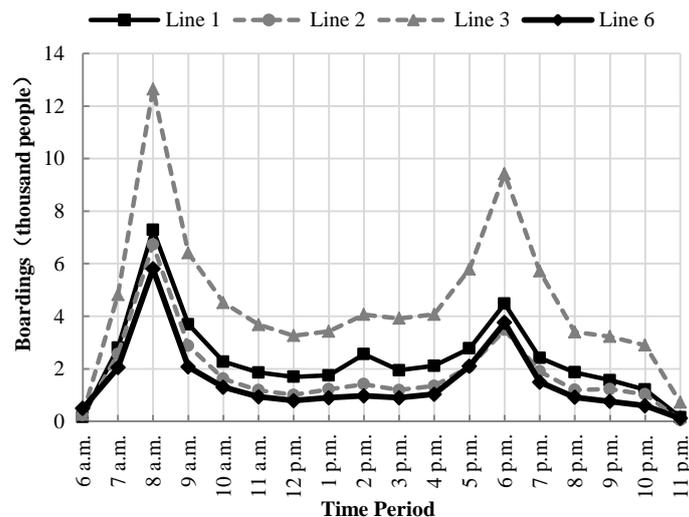
14 The time shortest path is calculated by Dijkstra algorithm, according to the
 15 network information about Chongqing Rail Transit, including the information about
 16 lines, stations, travel time between stations, stop time, interchange time and so on.

18 Temporal Scope: Peak Period

19 Urban rail transit system has the advantages of fast speed, punctuality and large
 20 capacity. Owing to these, it undertakes a lot of long- and moderate-distance trips.
 21 Their aims are mainly commuting and going to school. This type of passengers has
 22 rigid travel demands. Their specified arrival time (SAT) of destination is fixed and
 23 relatively close, so that the peak hour of urban rail transit system is usually within the
 24 morning peak period. This situation is also applicable for Chongqing Rail Transit
 25 (Figure 2). All 4 lines reach MSPF during all day between 8 a.m. and 9 a.m.

26

27



28

29 **FIGURE 2 MSPF of lines in Chongqing rail transit on one weekday.**

30

1 The same trip may be divided into different periods according to various
 2 criteria of the classification. Here, one trip is divided into different periods according
 3 to its boarding time. Because MSPF within peak hour is accumulated by the
 4 passengers departing from different stations before or during the peak hour, so the
 5 starting time of peak period should be earlier than that of peak hour. The paper
 6 provides that peak period is from 6 a.m. to 9 a.m.

8 METHODOLOGY

10 Deterrence Function

11 Deterrence function is one of the most important influence factors of station-to-station
 12 OD matrix; its mathematical form and components directly affect the final estimation
 13 results. Currently, there are 3 commonly used mathematical forms, which is power
 14 function, exponential function and the combination of both. These formulations are
 15 written as equation (1)-(3) when applied in urban rail transit system.

16 Form 1: Power function

$$17 \quad f(c_{ij}) = f(d_{ij}) = d_{ij}^{-\gamma} \quad (1)$$

18 Form 2: Exponential function

$$19 \quad f(c_{ij}) = f(d_{ij}) = \exp(-\eta \cdot d_{ij}) \quad (2)$$

20 Form 3: The combination of power function and exponential function

$$21 \quad f(c_{ij}) = f(d_{ij}) = \exp(-\eta \cdot d_{ij}) \cdot d_{ij}^{-\gamma} \quad (3)$$

22 In these formulations, c_{ij} = the generalized travel cost between station i and
 23 station j; d_{ij} = the travel time between station i and station j; and γ, η = parameters to
 24 be estimated.

25 One essential difference between urban rail transit and road traffic as well as
 26 bus transit is that its passengers can only change between various lines through
 27 interchange stations. This difference brings about the result that when measuring the
 28 generalized travel cost between two stations, interchange times is a key factor except
 29 the travel time. Since when the exponent of a power function is negative, the base
 30 mustn't be 0, and the interchange times between two stations are non-negative
 31 integers, so it is not appropriate to introduce interchange times into a deterrence
 32 function in the form of power function. This paper introduces it in the form of
 33 exponential function.

34 Form 4: Power function

$$35 \quad f(c_{ij}) = f(d_{ij}, n_{ij}) = d_{ij}^{-\gamma} \cdot \exp(-\tau \cdot n_{ij}) \quad (4)$$

36 Form5: Exponential function

$$37 \quad f(c_{ij}) = f(d_{ij}, n_{ij}) = \exp(-\eta \cdot d_{ij} - \tau \cdot n_{ij}) \quad (5)$$

1 Form 6: The combination of power function and exponential function

$$2 \quad f(c_{ij}) = f(d_{ij}, n_{ij}) = \exp(-\eta \cdot d_{ij} - \tau \cdot n_{ij}) \cdot d_{ij}^{-\gamma} \quad (6)$$

3 In these formulations, n_{ij} = the interchange times between station i and station
4 j and τ = a parameter to be estimated.

5 With the expansion of urban rail transit network, there may be several routes
6 between origin and destination stations for passengers to choose. To ensure the data
7 consistency of the study, the travel time and interchange times are the corresponding
8 values of the time shortest path between them. Models with these deterrence functions
9 will be estimated subsequently. Their performance would be compared to support the
10 analysis of the influence of deterrence function on the peak-period station-to-station
11 OD matrix.

12

13 Gravity Model

14

15 Modeling

16 The gravity model is similar to Newton's law of gravity. According to this model, in
17 urban rail transit system, the ridership between an origin station and a destination
18 station depends directly on the total boardings of the origin station, the total alightings
19 of the destination station, and depends inversely on the deterrence between two
20 stations. The peak-period ridership between two stations can be expressed as an
21 unconstrained gravity model (7), given in equation (7)

$$22 \quad t_{ij}^p = k (O_i^p)^\alpha (D_j^p)^\beta f(c_{ij}^p) \quad (7)$$

23 where t_{ij}^p = the ridership between station i and station j during peak period; O_i^p = the
24 boardings of station i during peak period; D_j^p = the sum of passengers who depart
25 from other stations during peak period and alight on station j; $f(c_{ij}^p)$ = the deterrence
26 function of station i and j; and k, α, β = parameters to be estimated.

27 On this basis, if the total boardings of origin stations are known, a production
28 -constrained gravity model can be developed, given in equation (8)

$$29 \quad t_{ij}^p = \left[O_i^p (D_j^p)^\rho f(c_{ij}^p) \right] / \left[\sum_j (D_j^p)^\rho f(c_{ij}^p) \right] \quad (8)$$

30 where ρ = a parameter to be estimated; and the meanings of other symbols are as
31 same as the ones in the unconstrained gravity model.

32

33 Estimation

34 For an unconstrained gravity model, since it is non-linear, the authors take logarithms
35 of equation (7) as follows

$$36 \quad \ln t_{ij}^p = \ln k + \alpha \ln O_i^p + \beta \ln D_j^p + \ln f(c_{ij}^p) \quad (9)$$

37 where $\ln f(c_{ij}^p)$ varies with the deterrence function, and it is always a linear
38 combination of $\ln d_{ij}^p, d_{ij}^p$ and n_{ij}^p . According to equation (9), the parameters can be

1 estimated by Ordinary Least Squares (OLS).

2 For a production-constrained gravity model, (12) proposed the estimation
3 method. The linear equations to be estimated are summarized in equation (10)-(13)

$$4 \quad \ln\left(r_{ij}^p / \bar{r}_i^p\right) = \rho \ln\left(D_j^p / \bar{D}^p\right) + \ln\left(f(c_{ij}^p) / \overline{f(c_{ij}^p)}\right) \quad (10)$$

$$5 \quad \bar{r}_i^p = \left(\prod_{j=1}^J r_{ij}^p\right)^{1/J} \quad (11)$$

$$6 \quad \bar{D}^p = \left(\prod_{j=1}^J D_j^p\right)^{1/J} \quad (12)$$

$$7 \quad \overline{f(c_{ij}^p)} = \left(\prod_{j=1}^J f(c_{ij}^p)\right)^{1/J} \quad (13)$$

8 where $\ln\left(f(c_{ij}^p) / \overline{f(c_{ij}^p)}\right)$ varies with the deterrence function, and it is always a linear
9 combination of $\ln\left(d_{ij}^p / \bar{d}_i^p\right)$, $d_{ij}^p - \frac{1}{J} \sum_{j=1}^J d_{ij}^p$ and $n_{ij}^p - \frac{1}{J} \sum_{j=1}^J n_{ij}^p$. \bar{d}_i^p is calculated by
10 equation(14).

$$11 \quad \bar{d}_i^p = \left(\prod_{j=1}^J d_{ij}^p\right)^{1/J} \quad (14)$$

12

13 **TABLE 1 Bivariate Correlation Analysis of Gravity Models**

14

Unconstrained Gravity Models						
Variables	$\ln r_{ij}^p$	$\ln O_i^p$	$\ln D_j^p$	$\ln d_{ij}^p$	d_{ij}^p	n_{ij}^p
Asymptotic significance of K-S test	0.000	0.000	0.000	0.000	0.000	0.000
Correlation coefficient	1.000	0.390**	0.495**	-0.497**	-0.497**	-0.404**
Production-constrained gravity models						
Variables	$\ln\left(r_{ij}^p / \bar{r}_i^p\right)$	$\ln\left(D_j^p / \bar{D}^p\right)$	$\ln\left(d_{ij}^p / \bar{d}_i^p\right)$	$d_{ij}^p - \frac{1}{J} \sum_{j=1}^J d_{ij}^p$	$n_{ij}^p - \frac{1}{J} \sum_{j=1}^J n_{ij}^p$	
Asymptotic significance of K-S test	0.000	0.000	0.000	0.000	0.000	0.000
Correlation coefficient	1.000	0.610**	-0.571**	-0.579**	-0.480**	

15 Note: ** Correlation is significant at the 0.01 level (2-tailed).

16

17 The result of Kolmogorov–Smirnov test shows the asymptotic significances of
18 variables' distributions mentioned above are all less than 0.050, which means none of
19 them follows a normal distribution. Bivariate correlations among variables in different
20 kinds of gravity models are calculated through Spearman correlation analysis. It has
21 proven that no matter in which kind of gravity models, the independent variables are

1 significantly correlated with the dependent variable, and the value is relatively high.
 2 The signs of correlation coefficients are in accordance with the inference of models.
 3 The values of correlation coefficients among variables in the production-constrained
 4 gravity model are comparatively higher (Table 1).

5 There are 12 unconstrained gravity models and 6 production-constrained
 6 gravity models to be estimated. The main differences among unconstrained gravity
 7 models are whether k is equal to one and deterrence functions. For
 8 production-constrained gravity models, because there is no constant to be estimated,
 9 so the main difference is their deterrence functions.

10 The forecast peak-period station-to-station OD matrices are iterated by Fratar
 11 Method (13) to ensure the satisfaction of production and attraction constraints, which
 12 is represented by equation (15)-(16).

13
$$\sum_i t_{ij}^p = O_i^p \tag{15}$$

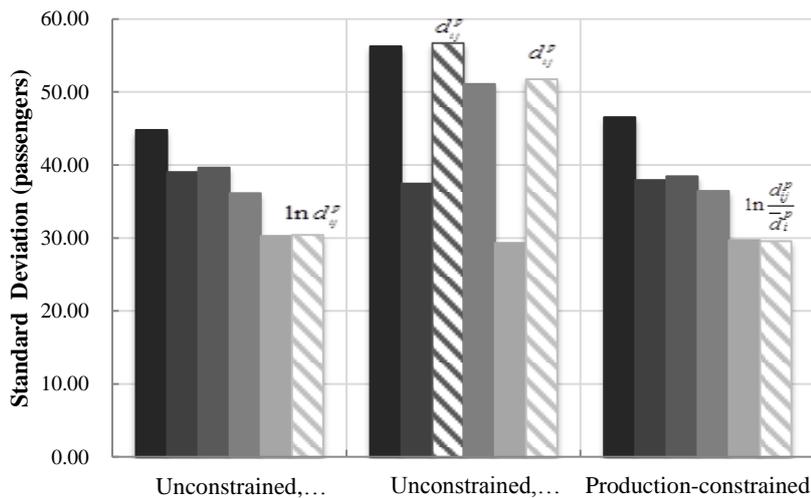
14
$$\sum_j t_{ij}^p = D_j^p \tag{16}$$

15 As a measurement of model performance, the standard deviation, which is also
 16 called root-mean-square deviation (RMSD), is taken as the comparison standard. In
 17 this study, it is calculated by equation (17)

18
$$\sigma = [(\varepsilon_1^2 + \varepsilon_2^2 + \dots + \varepsilon_l^2 + \dots + \varepsilon_n^2) / n]^{1/2} \tag{17}$$

19 where ε_l = the difference between predicted ridership and observed ridership between
 20 the l th pair of OD stations; and n = the number of station pairs whose peak-period
 21 ridership is larger than 0.

■ Deterrence function in Form 1 ■ Deterrence function in Form 2 ■ Deterrence function in Form 3
 ■ Deterrence function in Form 4 ■ Deterrence function in Form 5 ■ Deterrence function in Form 6



22

Note: The slash-filled model fails t-test.

23

24

25

FIGURE 3 The standard deviation of gravity models.

1 Figure 3 indicates that:

2 1. Some models' variables fail t-test, because there are two variables related
3 to travel time included in deterrence functions. When the influence of travel time is
4 mainly reflected by one of them, the other one cannot pass the significant test;

5 2. The models with deterrence functions including interchange times have
6 better performance. The obvious decrease of standard deviation proves that it is
7 necessary to consider the influence of interchange times on deterrence in urban rail
8 transit system;

9 3. The optimal model in the framework is the unconstrained gravity model
10 with a deterrence function in Form 5 and $k=1$, the standard deviation of which is
11 29.33 passengers.

12

13 *Analysis*

14 Although gravity model is widely used, the following drawbacks still exist: (1) it
15 hasn't taken human behavior into consideration basically; (2) it may overestimate
16 when the distance between origin and destination is too short; (3) the trips within one
17 TAZ is hard to be forecasted; (4) trip distance in one TAZ is not a fixed value; (5)
18 deterrence between TAZs varies significantly according to traffic modes.

19 In urban rail transit system, passengers move between stations instead of TAZs.
20 Thanks to this change, drawback (3)-(5) of gravity model can be avoided. For (3),
21 there is no passenger travelling from a station to its own in one trip; for (4), when
22 origin and destination stations are determined, trip distance is fixed because of the
23 infrastructure characteristics of urban rail transit; and for (5), the deterrence between
24 stations is not influenced by traffic modes because there is only one mode. But
25 drawback (2) is further aggravated with travel modes reducing from all modes to
26 urban mass transit alone. When the deterrence is too little, the passengers travelling
27 between two stations are more likely to choose bus transit and non-motorized
28 transportation due to travel cost. This phenomenon makes the ridership between OD
29 stations decline. But traditional gravity models are strictly in accordance with the
30 trend that with the decrease of deterrence, the ridership between two stations increases,
31 which doesn't conform to the fact.

32

33 **Peak Period Coefficient (PPC) Model**

34

35 *Modeling*

36 The Peak Period Coefficient (PPC) model is proposed to solve the over-estimation
37 caused by the gravity model. According to this model, the peak-period
38 station-to-station OD matrix is obtained by multiplying the given all-day
39 station-to-station OD matrix by forecasting PPC.

40 The PPC of station-to-station ridership is the proportion of station-to-station
41 ridership during peak-period to the one within all day. The trip purposes of passengers
42 who travel during various periods are quite different. For passengers travelling during

1 the morning peak period, their main purpose is commuting and going to school, so the
 2 ridership between residential-land-based origin stations and destination stations whose
 3 surrounding land is mainly related to jobs and schools is tend to be high. When the
 4 time range is extended to all day, the composition of passengers is more complex,
 5 with three parts included. The first one is the morning-peak-period passengers as
 6 mentioned above. The second part is the off-peak-period passengers living around the
 7 station with flexible travel demands. And the last part is the evening-peak-period
 8 passengers working around the station, whose trip purpose is going home. Because of
 9 the difference and the diversification of land use around urban rail transit stations, it is
 10 not recommended to multiply the all-day station-to-station OD matrix by an identical
 11 PPC to obtain the peak-period station-to-station OD matrix. This simple conversion
 12 hasn't considered the properties of origin and destination stations.

13 Based on the hypothesis that the station-to-station ridership follows an
 14 unconstrained gravity model, the PPC is calculated as follows

$$15 \quad P_{ij} = t_{ij}^p / t_{ij}^d = \left[k_1 \cdot (O_i^p)^{\alpha_1} \cdot (D_j^p)^{\beta_1} \cdot f(c_{ij}^p) \right] / \left[k_2 \cdot (O_i^d)^{\alpha_2} \cdot (D_j^d)^{\beta_2} \cdot f(c_{ij}^d) \right] \quad (18)$$

16 where P_{ij} = the PPC of ridership between station i and station j ; t_{ij}^p, t_{ij}^d = the ridership
 17 between station i and station j during peak period and all day, respectively; O_i^p, O_i^d =
 18 the boardings of station i during peak period and all day, respectively; D_j^p = the sum
 19 of passengers who depart from other stations during peak period and alight on station
 20 j , D_j^d = the alightings of station j during all day; and $f(c_{ij}^p), f(c_{ij}^d)$ = the deterrence
 21 function of station i and station j during peak period and all day, respectively.

22 Equation (19) can be obtained by dividing the numerator and the dominator of
 23 P_{ij} by $(O_i^d)^{\alpha_1} \cdot (D_j^d)^{\beta_1}$ simultaneously, which is represented by

$$24 \quad P_{ij} = (k_1 / k_2) \cdot \left[f(c_{ij}^p) / f(c_{ij}^d) \right] \cdot \left[(O_i^p / O_i^d)^{\alpha_1} \cdot (D_j^p / D_j^d)^{\beta_1} \right] / \left[(O_i^d)^{\alpha_2 - \alpha_1} \cdot (D_j^d)^{\beta_2 - \beta_1} \right] \quad (19)$$

25 After substituting $P_{O_i} = O_i^p / O_i^d$ and $P_{D_j} = D_j^p / D_j^d$ into equation (20), it can be
 26 written as follows

$$27 \quad P_{ij} = (k_1 / k_2) \cdot \left[f(c_{ij}^p) / f(c_{ij}^d) \right] \cdot \left[(P_{O_i})^{\alpha_1} \cdot (P_{D_j})^{\beta_1} \right] / \left[(O_i^d)^{\alpha_2 - \alpha_1} \cdot (D_j^d)^{\beta_2 - \beta_1} \right] \quad (20)$$

$$28 \quad P_{ij} = k' \cdot \left[f(c_{ij}^p) / f(c_{ij}^d) \right] \cdot \left[(P_{O_i})^{\alpha_1} \cdot (P_{D_j})^{\beta_1} \right] / \left[(O_i^d)^{\alpha'} \cdot (D_j^d)^{\beta'} \right] \quad (21)$$

29 where P_{O_i} = the proportion of peak-period boardings of station i ; P_{D_j} = the proportion
 30 of peak-period alightings of station j ; and $k', \alpha_1, \beta_1, \alpha', \beta'$ = parameters to be
 31 estimated, where $k' = k_1 / k_2$; $\alpha' = \alpha_2 - \alpha_1$ and $\beta' = \beta_2 - \beta_1$.

32 Equation (20) shows that $P_{O_i}, P_{D_j}, O_i^d, D_j^d$ and deterrence between two stations
 33 are the main influence factors of PPC. The former two factors need to be forecasted
 34 according to land use characteristics around stations, while the other four factors can
 35 be calculated from given data.

36 Based on equation (20), the PPC model is written in equation (21)

$$t_{ij}^p = P_{ij} \cdot t_{ij}^d = k' \cdot [f(c_{ij}^p) / f(c_{ij}^d)] \cdot [(P_{O_i})^{\alpha_1} \cdot (P_{D_j})^{\beta_1} / (O_i^d)^{\alpha'} \cdot (D_j^d)^{\beta'}] \cdot t_{ij}^d \quad (21)$$

2 The model can be further simplified with the following assumptions.

3 **Assumption (1):** The station-to-station OD matrices during different time
4 periods are only affected by the boardings of origin station, the alightings of
5 destination station and deterrence between two stations.

6 In an unconstrained gravity model, the parameter k is regarded as a
7 comprehensive correction coefficient caused by other influence factors except the
8 three influence factors mentioned in Assumption (1). If Assumption (1) holds then
9 $k_1 = k_2$, that is to say, $k' = 1$. Equation (21) can be simplified as shown in equation
10 (22).

$$t_{ij}^p = [f(c_{ij}^p) / f(c_{ij}^d)] \cdot [(P_{O_i})^{\alpha_1} \cdot (P_{D_j})^{\beta_1} / (O_i^d)^{\alpha'} \cdot (D_j^d)^{\beta'}] \cdot t_{ij}^d \quad (22)$$

12 **Assumption (2):** The deterrence between origin and destination stations
13 during different time periods keep consistent.

14 Usually, the deterrence function is a function of the travel time between origin
15 and destination stations as well as interchange times. The characteristic of urban rail
16 transit system leads to the result that once the lines and stations construction is
17 complete, these two independent variables have little changes during different time
18 periods. If Assumption (2) holds then $f(c_{ij}^p) = f(c_{ij}^d)$, simplifying the model to
19 equation(23).

$$t_{ij}^p = k' \cdot [(P_{O_i})^{\alpha_1} \cdot (P_{D_j})^{\beta_1} / (O_i^d)^{\alpha'} \cdot (D_j^d)^{\beta'}] \cdot t_{ij}^d \quad (23)$$

21 Whether these two assumptions are valid or not in the model still needs more
22 analysis.

24 *Estimation*

25 Since the all-day station-to-station OD matrix is given, the forecasting part is the PPC.
26 The logarithms of PPC model is shown in equation (24)

$$27 \ln P_{ij} = \ln k' + \alpha_1 \ln P_{O_i} + \beta_1 \ln P_{D_j} - \alpha' \ln O_i^d - \beta' \ln D_j^d + \ln f(c_{ij}^p) - \ln f(c_{ij}^d) \quad (24)$$

28 The parameters can be estimated by Ordinary Least Squares (OLS) since the
29 equation is linear. It should be noted that to prevent the inaccuracy of forecast P_{O_i}
30 and P_{D_j} influencing the performance of the PPC model, here the actual values of
31 both variables are used for estimation.

32 Table 2 shows that the signs of correlation coefficients are consistent with the
33 inference of PPC model. All variables are significantly correlated. The coefficients of
34 $\ln P_{O_i}$ and $\ln P_{D_j}$ are higher than the ones of $\ln O_i^d$ and $\ln D_j^d$, while the ones of d_{ij}^p , $\ln d_{ij}^p$
35 and n_{ij}^p are a little lower. Compared with the traditional gravity model, the
36 correlation coefficients of d_{ij}^p , $\ln d_{ij}^p$ and n_{ij}^p in the PPC model decline markedly,

1 proving that the influence of these variables on the PPC model gets weaken.

2

3 **TABLE 2 Bivariate Correlation Analysis of PPC Models**

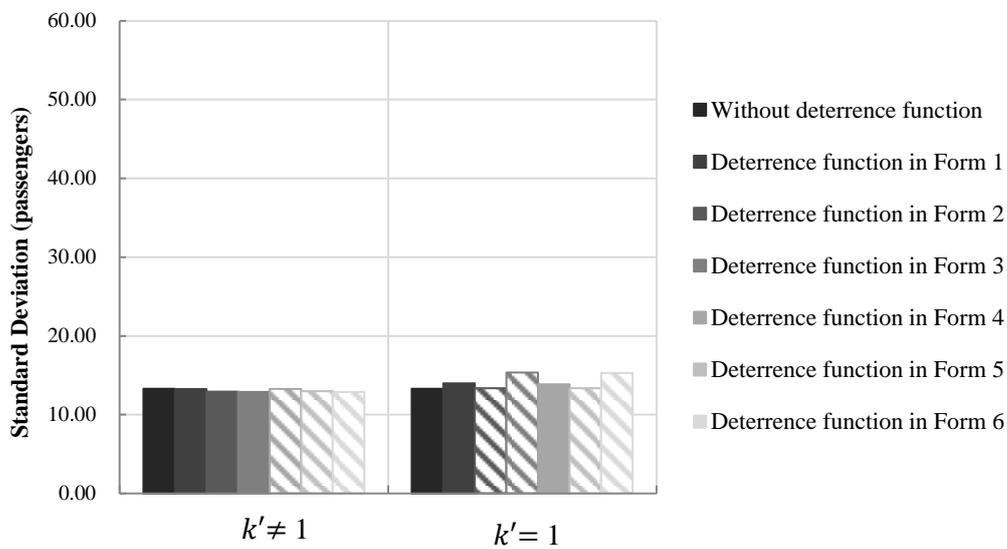
4

Variables	$\ln P_{ij}$	$\ln P_{O_i}$	$\ln P_{D_j}$	$\ln O_i^d$	$\ln D_j^d$	$\ln d_{ij}^p$	d_{ij}^p	n_{ij}^p
Asymptotic significance of K-S test	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Correlation coefficient	1.000	0.498**	0.516**	-0.241**	-0.162**	-0.044*	-0.044*	0.022*

5 Note: ** Correlation is significant at the 0.01 level (2-tailed);* Correlation is significant at the 0.05
 6 level (2-tailed); $d_{ij}^p = d_{ij}^d, n_{ij}^p = n_{ij}^d$.

7

8



9 Note: The slash-filled model fails t-test.

10

11 **FIGURE 4 The standard deviation of PPC models.**

12

13 The combination of two assumptions and different deterrence functions forms
 14 14 PPC models to be estimated. Similarly, the forecast peak-period station-to-station
 15 OD matrices are iterated by Fratar Method (13) to ensure the satisfaction of
 16 production and attraction constraints.

17 Excluding the models with variables failing t-test (Figure 4), the optimal
 18 model in the PPC model framework is the one with a deterrence function in Form 3,
 19 whose $k \neq 1$. Its standard deviation is 12.90 passengers. In the model, there is a
 20 difference between the deterrence during peak period and all day, which means the
 21 perceived travel time during peak period is longer than the one during other periods
 22 even though the actual travel time is totally the same. This is caused by the crowding
 23 discomfort in trains during the peak period.

24 For Assumption (1), the estimation results show that the standard deviation
 25 differences between models whose $k=1$ and $k \neq 1$ are less than 1 passenger, except the

1 models with deterrence function in the form of a combination of power function and
 2 exponential function. So when the difference between these two kinds of models is
 3 acceptable, it is assumed to be valid. For Assumption (2), whether it is valid is
 4 supported by the proofs from two aspects. On one hand, n_{ij}^p and d_{ij}^p fail t-tests in
 5 some models (Table 3). The failure of n_{ij}^p is owing to that it doesn't change through
 6 all day, so its impact on the PPC is negligible. Meanwhile, both d_{ij}^p and $\ln d_{ij}^p$ are
 7 proxies for travel time. When only one of them is used to represent the deterrence, the
 8 modulus of regression coefficient is little or the variable fails the t-test; when both of
 9 them are in the function, the signs of their regression coefficients are opposite. So the
 10 influence of travel time is balanced. On the other hand, it is easy to find from Figure 4
 11 that there is no big difference among models with various deterrence functions and the
 12 one without a deterrence function. The standard deviation of the model without a
 13 deterrence function is 13.30 passengers, which is 0.40 higher than the one of the
 14 optimal PPC model, but still lower than most of other models. So Assumption (2) is
 15 regarded to be valid when the difference of standard deviation is tolerable or the
 16 network is not very crowded during peak period.

17

18 **TABLE 3 The Parameters of PPC Models**

19

PPC Models	$\ln k'$	$\ln P_{O_i}$	$\ln P_{D_j}$	$\ln O_i^d$	$\ln D_j^d$	$\ln d_{ij}^p$	d_{ij}^p	n_{ij}^p
1	2.952	0.818	0.853	-0.124	-0.089	-	-	-
2	3.088	0.819	0.849	-0.126	-0.091	-0.031	-	-
3	3.140	0.820	0.841	-0.130	-0.096	-	-0.003	-
4	2.812	0.817	0.836	-0.132	-0.098	-0.009	0.170	-
5	3.086	0.819	0.850	-0.126	-0.091	-0.030	-	<u>-0.002</u>
6	3.143	0.818	0.840	-0.131	-0.096	-	-0.003	<u>0.011</u>
7	2.803	0.818	0.836	-0.132	-0.097	-0.009	0.173	<u>-0.008</u>
8	-	0.790	0.778	0.044	0.049	-	-	-
9	-	0.789	0.797	0.029	0.039	0.078	-	-
10	-	0.789	0.780	0.044	0.049	-	<u>0.000</u>	-
11	-	0.790	0.782	<u>-0.004</u>	0.010	-0.014	0.391	-
12	-	0.795	0.800	0.030	0.039	0.090	-	-0.038
13	-	0.789	0.780	0.044	0.049	-	<u>0.000</u>	<u>0.000</u>
14	-	0.796	0.786	<u>-0.004</u>	<u>0.009</u>	-0.014	0.406	-0.042

20 Note: The variable with the coefficient in underlined and bold style fails t-test.

21

22 **RESULT AND DISCUSSION**

23 The optimal models in two frameworks are given in equation (25)-(26)

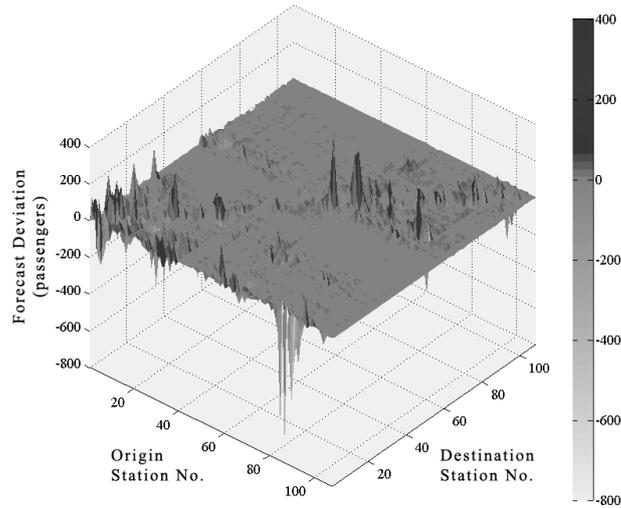
24 The gravity model

25
$$t_{ij}^p = (O_i^p)^{0.210} (D_j^p)^{0.314} \exp(-0.033d_{ij}^p - 0.730n_{ij}^p) \quad (25)$$

26 The PPC model

27
$$t_{ij}^p = 16.64(P_{O_i})^{0.817} (P_{D_j})^{0.836} / (O_i^d)^{0.132} (D_j^d)^{0.098} \left[(d_{ij}^p)^{0.170} \exp(-0.009d_{ij}^p) \right] \cdot t_{ij}^d \quad (26)$$

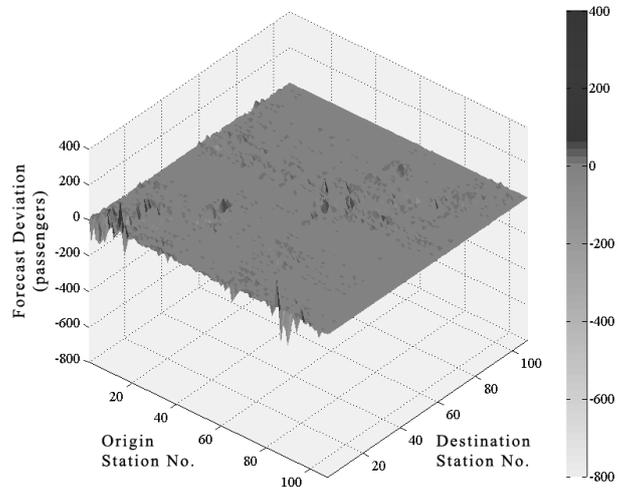
1



2

(a) The gravity model

3



4

(b) The PPC model

5

FIGURE 5 The forecast deviation of the optimal model.

6

7

TABLE 4 Statistics of Model Deviations

8

Statistics		Gravity model	PPC model
Median		1.00	0.00
Standard deviation		29.33	12.90
Variance		860.26	166.51
Minimum		-684.00	-237.00
Maximum		384.00	199.00
Percentile	25	-2.00	-1.00
	50	1.00	0.00
	75	5.00	3.00
Average deviation of little-deterrence OD stations		29.95	-0.54

10

1 Figure 5 shows the deviation between actual value and the results forecasted
2 by the optimal gravity model and PPC model and iterated by Fratar method. In Figure
3 5 (a), the deviation near the 45 ° diagonal is relatively high, because the deterrence of
4 these OD stations is little. But this problem doesn't exist in Figure5 (b). The
5 origin-destination stations with travel time between them less than 5 min and no
6 interchange are defined as little-deterrence OD stations. Table 4 shows that the
7 average deviation of little-deterrence OD stations of Model I is as much as 29.95
8 passengers, while the one of Model II is only -0.54 passengers. It is proved that the
9 over-estimation is weakened effectively in the PPC model.

10 From the figure and table, it also can be seen that the PPC model is obviously
11 superior to the gravity model in terms of overall performance or case result. The
12 standard deviation of the PPC model is 16.43 passengers less than the one of the
13 gravity model, reduced by 56.02%.

14 15 **CONCLUSIONS**

16 With the case study of Chongqing, China, this study proposes a gravity-model-based
17 Peak Period Coefficient (PPC) model. According to this model, the peak-period
18 station-to-station OD matrix in urban rail transit is forecasted by multiplying the given
19 all-day station-to-station OD matrix by the forecasting PPC. Results show that:

- 20 1. When adopting traditional gravity models in forecasting the peak-period
21 station-to-station OD matrix in urban rail transit, several drawbacks of these
22 models can be avoided. But the drawback of over-estimation is further
23 aggravated with travel modes reducing from all modes to urban mass transit
24 alone.
- 25 2. The PPC model can effectively weaken the over-estimation when the
26 deterrence between OD stations is too little. Its performance is much better
27 than the traditional gravity model, with standard deviation reduced by 56.02%.
- 28 3. In the PPC model, interchange times isn't the main influence factor any
29 longer. And there is no big difference in the performance among models with
30 various deterrence functions and the one without a deterrence function. This
31 result means there may be some difference among the deterrence during
32 various periods, but the influence of the difference on the PPC is not obvious.

33 When applying the PPC model in planning and design phase, the proportion of
34 peak-period boardings of the origin station and the proportion of peak-period
35 alightings of the destination station can be forecasted based on the socio-economic
36 and land use characteristics around stations.

37

REFERENCES

1. Poon, M. H., S. C. Wong and C. O. Tong. A dynamic schedule-based model for congested transit networks. *Transportation Research Part B: Methodological*, Vol.38, No.4, 2004, pp. 343-368.
2. Liu, X. Research on the Dynamic Flow Assignment Model Based on Train Schedule for Urban Subway Network, Chang'an University. Xi'an, 2013.
3. Yang, D. Research on Schedule-based Rail Transit Passenger Flow Assignment, Beijing Jiaotong University, 2013.
4. Choi, J., Y. J. Lee, T. Kim and K. Sohn. An analysis of Metro ridership at the station-to-station level in Seoul. *Transportation*, Vol.39, No.3, 2012, pp. 705-722.
5. Zhao, J., W. Deng, Y. Song and Y. Zhu. Analysis of Metro ridership at station level and station-to-station level in Nanjing: an approach based on direct demand models. *Transportation*, Vol.41, No.1, 2014, pp. 133-155.
6. Wilson, A. G. A family of spatial interaction models, and associated developments. *Environment and Planning A*, Vol.3, No.1, 1971, pp. 1-32.
7. Shao, C. *Traffic Planning*, China Railway Publishing House. Beijing, China, 2004.
8. Southworth, F. A Disaggregated Trip Distribution Model with Elastic Frequencies and Expenditures. *Journal of Transport Economics and Policy*, Vol.13, No.2, 1979, pp. 209-224.
9. Ben-Akiba, M., H. F. Gunn and L. A. Silman. Disaggregate trip distribution models. *Proceeding of Japan Society of Civil Engineers*, No.347, 1984, pp. 1-17.
10. Lam, W. H., Z. X. Wu and K. S. Chan. Estimation of transit origin - destination matrices from passenger counts using a frequency-based approach. *Journal of Mathematical Modelling and Algorithms*, Vol.2, No.4, 2003, pp. 329-348.
11. Bekhor, S. and J. N. Prashker. GEV-based destination choice models that account for unobserved similarities among alternatives. *Transportation Research Part B: Methodological*, Vol.42, No.3, 2008, pp. 243-262.
12. Nakanishi, M. and L. G. Cooper. Parameter estimation for a multiplicative competitive interaction model: least squares approach. *Journal of Marketing Research*, 1974, pp. 303-311.
13. Fratar, T. Vehicular trip distribution by successive approximations. *Traffic Quarterly*, Vol.8, No.1, 1954, pp. 53-56.