# Forecast of Regional Express Volume based on Big Data

**Weijie Yuan [a], Zhenlin Wei [b, *]**

Beijing Jiaotong University, Beijing 100089, China.

[a] 17120926@bjtu.edu.cn, [b] zhlwei@bjtu.edu.cn

*Corresponding author

**Keywords:** online shopping; regional express delivery; grey system prediction model; forecast data.

**Abstract:** With the development of e-commerce, online shopping has gradually entered thousands of households, and the forecast of regional express delivery can provide a basis for the investment and development of logistics infrastructure in the region. Based on the per-capita express demand, this paper uses the grey system prediction model and the logistics growth model from the short-term and long-term perspectives respectively to obtain the forecast data for the future years, and assign the weight coefficients respectively to obtain the combined forecast values. The combined forecasting value fully considers the trend of the demand for express delivery in the long-term and short-term, and has better accuracy than the single forecasting model.

## 1. Introduction

Although e-commerce development in China started later than abroad, its development is very rapid, which has brought tremendous pressure on the express delivery industry. In 2015, the State Council issued the "Several Opinions of the State Council on Promoting the Development of the Express Industry", and pointed out that it is necessary to build a perfect service network and build express delivery. Professional logistics park, express delivery center and express delivery service platform. The construction of the service network should be based on the demand within the region, and then fully consider the environment, rent, service quality and other factors to determine the layout and scale of the network. The demand for express delivery is the basic data required for planning. Accurate measurement of the demand for express delivery can make the government and enterprises more scientific and reasonable in planning the service network and resource allocation.

## 2. Research Purpose and Meaning

In recent years, the E-commerce has experienced a spurt of development., followed by the rise of the express delivery industry, the express volume increased exponentially. In 2018, the volume of China's express service enterprises completed 50.71 billion, an increase of 26.6% [1]. It can be seen that the express delivery industry is still in a period of rapid growth. The development of express delivery industry is closely related to the degree of economic and social development. For the government, the prerequisite for rational planning of logistics facilities is to grasp the scale and development trend of logistics demand and accurately predict the demand for express delivery. It can provide a quantitative basis for the construction of the logistics infrastructure system in the region, thus contributing to the development of the entire express industry. For express delivery companies, accurately predicting the demand for express delivery can facilitate the company to systematically adjust the layout of outlets, network scale, staff configuration, facility equipment configuration, etc., reducing resource waste and improving service quality. Express delivery is an industry that provides services for various social and economic activities. It is the development of the basic industry. It provides logistical support for the development of other industries. The development of other industries provides support for the development of the express delivery industry. They support each other. Joint development and research on regional express demand are of great significance to China's economic and social development.

## 3. Regional Express Development Research

1)Long-term development

In the long run, for any industry, in the process of development and evolution, it will be restricted by various factors that inhibit its uncontrolled growth, and generally exhibit certain life cycle characteristics, which are manifested in the initial stage, growth stage, maturity stage and recession period [2]. In the initial stage, due to lack of market and experience, the industry develops at a slower speed; in the growth period, it has a certain market and the industry develops gradually; in the mature stage, the industrial development tends to be stable and the industrial pattern has formed; in the recession, due to the substitute Or the emergence of new products, the original industry market began to shrink. he expresses delivery industry has gradually gained attention with the development of the economy. Until the logistics industry was completely opened to the outside world in 2005 and the continuous development of e-commerce in recent years, the rapid growth of the express delivery industry began. In the long run, China's express delivery industry has experienced a development trajectory from formation to gradual growth, following the growth trajectory of the Logistics curve.

The logistics model has obvious advantages, does not require strict assumptions, overcomes the limitations of linear equations subject to statistical assumptions, has a fixed growth rate and maximum capacity, and can easily describe the growth and evolution process of the express delivery industry [3]. The theoretical model of Logistics can be expressed as:

$$\begin{cases} \frac{dy}{dt} = yr\left(1 - \frac{y}{k}\right) \\ x(t_0) = x_0 \end{cases} \tag{1}$$

$\frac{dy}{dt}$ represents the growth rate of per capita express demand. In order to facilitate the empirical fitting analysis, the above theoretical model is transformed into an exponential growth model of per capita express demand. The specific formula is

$$\mathrm{y} = \frac{K}{1+ae^{(-rt)}} \tag{2}$$

Where y is the state variable of the per capita express demand growth, which changes with time;

R is the natural growth rate of per capita express demand growth under ideal conditions;

a is a restrictive coefficient, indicating that the growth of per capita express demand can be increased, and it has a certain inhibitory effect on its growth;

t is a time variable;

K reflects the saturation value of the growth of per capita express demand, which means the maximum amount of express delivery per capita.

2) Short-term development

In the short-term forecasting, the gray system prediction method is widely used. The gray system refers to the small sample with “partial information is known, some information is unknown”, and the uncertainty system with less information, GM (1,1) is the most commonly used. The univariate series prediction model, which uses the known series to be regarded as the discrete value of the continuous variable in its development, uses the differential equation to process the data, obtains the model calculation formula, and compares the result obtained by the above formula with the actual value. The residual is obtained, and the model is corrected by the residual, so that higher precision is obtained [4].

In this paper, the gray time series prediction is used to construct the gray prediction model with the per capita express data of each year, and the annual value is predicted. The amount of information required by the model is small, the operation is convenient, and the accuracy can be better in the case of less data [5].

3) Integrated prediction model

The factors affecting the demand for express delivery are very complicated. It is difficult to describe the change law with a single forecasting model. By using a combined forecasting model to properly combine a single model, the limitations of the single model can be compensated, and the relevant influencing factors can be fully considered. The advantages of the model make the results more convincing [6].

The focus of the combined prediction model is on the determination of the weight [7]. In this paper, the weight combination coefficient is determined by a linear combination method with the absolute error of the minimum error as the objective function.

There is a prediction method with a change with the year, there are n prediction methods, the actual T-year actual data, and assume:

$Y_t$ is the actual observation value at time point t, t= 1,2, ..., T;

$X_{it}$ is the predicted value of the i-th method at time t, i= 1,2, ..., n;

$K_{it}$ is the i-th predictive weighting coefficient at time point t, and satisfies $\sum_{i=1}^{n} k_{it} = 1$,$k_{it} \geq 0$ (i= 1, 2, …, n);

The objective function is:

$$\begin{cases} minW = \sum_{i=1}^{n} |k_{it} X_{it} - Y_t| \\ s.t. \sum_{i=1}^{n} k_{it} = 1 \quad k_{it} \geq 0 (\mathrm{t} = 1,2, \dots, \mathrm{T}) \end{cases} \tag{3}$$

For the weighting coefficients required for prediction, this paper uses the trend extrapolation method to determine the weighting coefficients that change with time, namely:

$$\begin{cases} k_{i,T+1} = \frac{1}{T} \sum_{t=1}^{T} k_{it} \\ \vdots \\ k_{i,T+j} = \frac{1}{T} \sum_{i=j}^{T+j-1} k_{it} \end{cases} \tag{4}$$

It is easy to prove that the weight coefficient obtained by this method still satisfies $\sum_{i=1}^{n} k_{it} = 1$,and $k_{it} \geq 0 (i = 1,2, \dots, n. t = T+1, \dots, T+j)$.

## 4. Empirical Analysis

Shanghai is China's first coastal open city and an international economic center. Domestic and international trade is close, and the logistics industry is developing rapidly. Domestic express delivery companies such as STO, YTO and Debon, and international express delivery headquarters such as FedEx and DHL are all established in Shanghai [8]. As the headquarters of express delivery enterprises, the development level of Shanghai's express delivery industry is at the leading level in the country. This paper selects the per capita express data of Shanghai from 2011 to 2018 for empirical analysis. The selected data source is “Shanghai Statistical Yearbook”.

（1）Long-term forecast

The per capita express delivery volume of Shanghai in 2009-2018 is shown in Table 1. The trend of change is shown in Figure 1.

Table 1. Shanghai per capita express delivery in 2009-2018

| **Year** | 2011 | 2012 | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
|---|---|---|---|---|---|---|---|---|
| **Per Capita Express Demand (unit: piece / person)** | 17.46 | 25.20 | 39.33 | 52.76 | 70.79 | 107.4 | 129.0 | 143.9 |

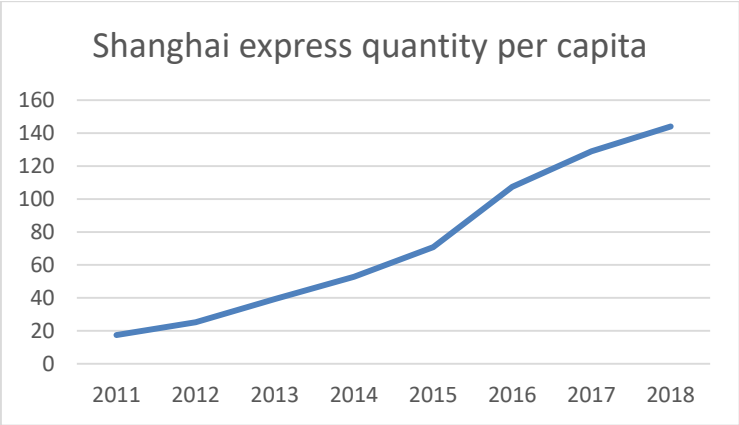


Figure 1. Shanghai express quantity per capita

As can be seen from the figure, the per capita express volume in 2018 is in a state of rapid growth, but in theory, it is impossible to grow indefinitely. Generally speaking, when the per capita express volume is small, the growth is faster and faster. As the population grows to a certain number, due to the saturation of the population and the limitation of natural resources, the growth rate will gradually become full, that is, the growth rate will become smaller until it reaches 0. At this time, the per capita express volume will tend to in a constant, this constant is the maximum value that the environment can accommodate.

Bringing the 2011 data into the logistics model, the change rule of per capita express volume is:

$$y = \frac{K}{1 + (\frac{K}{17.47} - 1)e^{(2011-t)r}} \tag{5}$$

There are two unknown parameters in the above formula. Set the initial value to (K, r) = (100,1). Use the "lsqnonlin" command in MATLAB software to fit the above data. The fitting effect is shown in Figure 2. The fitting effect is good, and changing the initial value repeatedly, it is found that the initial value of the growth rate parameter r has a certain influence on the fitting effect of the function. When the initial value of r is selected to be greater than 20, the fitting effect of the model is not good. When the initial value of r is [0, 20], the results obtained by the fitting method are relatively stable. After repeatedly changing the initial value and combining the fitting effect, the best result is (K, r) = (210.2068, 0.4649), and the change function is:

$$y = \frac{210.2068}{1+11.03e^{-0.4649t}} \tag{6}$$

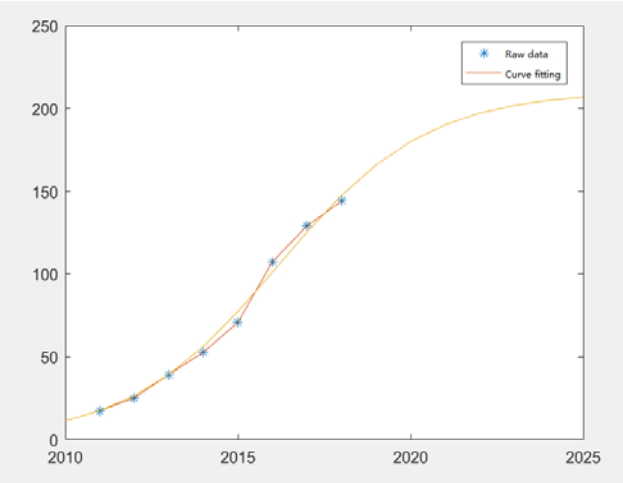


Figure 2. Logistics model fitting curve

In order to predict the per-capita express volume for the next four years, each year is substituted into equation (6), and the predicted values are shown in Table 2.

Table 2. Statistical model prediction results

| **Year** | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|
| **Results (unit: piece / person)** | 165.8445 | 179.9666 | 190.1371 | 197.1355 |

(2) Short-term forecast

There are many factors affecting regional express delivery, such as economic factors, population, per capita income, regional policies, etc. [9] These factors are affected by many factors and have uncertainties, we can use the gray system prediction model to Better predict the change in per capita express demand in the short term.

1)The raw data is the above-mentioned per capita express demand of the city from 2009 to 2017:

$$x^{(0)}(\mathrm{t}) = \quad [17.47\ 25.21\ 39.34\ 52.77\ 70.80\ 107.45\ 129.01\ 143.99]$$

2)Accumulate the original data $x^{(0)}(\mathrm{t})$:

$$x^{(1)}(\mathrm{t}) = [17.47\ 42.68\ 82.0\ 134.79\ 205.59\ 313.04\ 442.05\ 586.04]$$

3) Get the data matrix B and the data vector y

$$\mathrm{B}=\begin{bmatrix} -\frac{1}{2}[x^{(1)}(2)+x^{(1)}(1)] & 1 \\ -\frac{1}{2}[x^{(1)}(3)+x^{(1)}(2)] & 1 \\ \cdots & 1 \\ -\frac{1}{2}[x^{(1)}(n)+x^{(1)}(n-1)] & 1 \end{bmatrix}=\begin{bmatrix} -30.0750 & 1 \\ -62.3500 & 1 \\ -108.4050 & 1 \\ -170.1900 & 1 \\ -259.3150 & 1 \\ -377.5450 & 1 \\ -514.0450 & 1 \end{bmatrix}$$

$$\mathrm{y}= [25.21\ 39.34\ 52.77\ 70.80\ 107.45\ 129.01\ 143.99]^{\mathrm{T}}$$

4) Find the parametric model and derive the time corresponding equation:

$$x^{(1)}(\mathrm{k}+1) = x^{(0)}(1)-\frac{b}{a})e^{-ak}+\frac{b}{a} = 119.588e^{0.254194*t}-102.118 \qquad (7)$$

The historical year is brought into the application and subtracted to obtain the predicted value and test result of the short-term forecast as shown in Table 3.

The prediction accuracy is one level, indicating that the model is effective for the per capita express quantity prediction problem. The above results are used to predict the data from 2019 to 2023, and the data is shown in Table 4.

(3) Integrated prediction

According to the objective function given in the third section, the historical annual weight coefficient is obtained by using lingo, and the relative error value of each prediction method is obtained, as shown in Table5.

Table 3. Predicted values and test results of short-term forecasts

| Year | Actual Value | Predictive Value | Residual Value |
|---|---|---|---|
| 2011 | 17.46569 | 17.4657 | 0 |
| 2012 | 25.20553 | 34.6102 | 9.40 |
| 2013 | 39.33503 | 44.6277 | 5.29 |
| 2014 | 52.76871 | 57.5447 | 4.78 |
| 2015 | 70.79954 | 74.2004 | 3.40 |
| 2016 | 107.4513 | 95.6769 | -11.77 |
| 2017 | 129.0147 | 135.3695 | -5.65 |
| 2018 | 143.99 | 159.0775 | 15.09 |
| **Test Result** | Posterior variance ratioC:0.0305<br>Small error probabilityP:100%<br>Prediction accuracy level: First level, good | | |

Table 4. Short-term forecast data for the next 4 years

| **Year** | 2019 | 2020 | 2021 | 2022 |
|---|---|---|---|---|
| **Results (unit: piece / person)** | 205.1102 | 264.4736 | 341.0182 | 439.7165 |

Table 5. Historical year weight coefficient

| Year | Actual Value | Long-term Value | Long-term Relative Error | Short-term Value | Short-term Relative Error | Combined weight coefficient | | Combined Forecast | Combined Relative Error |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Long-term Forecast | Short-term Forecast | | |
| 2011 | 17.47 | 17.47 | 0.00 | 17.47 | 0.00 | 0 | 1 | 17.47 | 0.00 |
| 2012 | 25.21 | 26.51 | 0.05 | 34.61 | 0.37 | 1 | 0 | 26.51 | 0.05 |
| 2013 | 39.34 | 39.27 | 0.00 | 44.63 | 0.13 | 0.98 | 0.02 | 39.37 | 0.00 |
| 2014 | 52.77 | 56.28 | 0.07 | 57.54 | 0.09 | 1 | 0 | 56.28 | 0.07 |
| 2015 | 70.80 | 77.34 | 0.09 | 74.20 | 0.05 | 0 | 1 | 74.20 | 0.05 |
| 2016 | 107.45 | 101.10 | 0.06 | 95.68 | 0.11 | 1 | 0 | 101.10 | 0.06 |
| 2017 | 129.01 | 125.28 | 0.03 | 135.37 | 0.05 | 0.63 | 0.37 | 129.01 | 0.00 |
| 2018 | 143.99 | 147.43 | 0.02 | 159.08 | 0.10 | 1 | 0 | 147.43 | 0.02 |

Analysis of the data in the above table, the following conclusions are obtained:

1) It can be seen from the data in the table that the long-term forecasting model can well predict the per-capita express demand. In the short-term forecasting model, the relative error value of individual years is large, and the prediction accuracy fluctuates, but it still reflects well. The trend of demand development indicates the effectiveness of the single prediction model selected in this paper.

2) Comparing the relative error value of the single prediction with the relative error value of the combined prediction, the combined prediction relative error value is smaller than the two single predictions, indicating that the combined prediction can improve the prediction accuracy to a certain extent.

## 5. Conclusion

The regional express volume forecast can provide a basis for the strategic development of the enterprise and the formulation of government policies. This paper aims at the forecasting of the demand volume of express delivery, combining the long-term and short-term forecasting models, applying the trend extrapolation method to obtain the weight coefficient, the accuracy of the combined forecasting model is better than that of the single forecasting model through the historical data of Shanghai historical year. It proves the validity of the combined forecasting model and provides ideas for forecasting the express demand in other regions.


## Acknowledgments

The research is supported by grants from National Key R&D Program of China (2018YFB1601600).


## References


[1] Xiaojun Li, Construction and Empirical Study of Forecast Model for Express Demand Combination, Southwest Jiaotong University,2017.

[2] Guotao Ji, Logistic Curve Analysis of Logistics Industry Growth in Liaoning Province, Productivity research, 2016(09):67-69.

[3] Lei Xiao and Donglin Xu, Mathematical modeling experiment, first ed., Beijing,2012.

[4] Heping Zhang and Xiaolong Xie, Research on Combination Optimization Model Based on Grey Correlation Degree 2019(18):17-18.

[5] Yuting Jiang, forecast of winter jujube price based on grey GM (1,1) model, Scientific and technological innovation 2019(18):17-18.

[6] Yanzhi Li, Research on Regional Express Traffic Forecast and Location Selection of Connection Point, Zhejiang Gongshang University,2015.

[7] Guanghui An, Long-term power load combination forecasting method and case study, North China Electric Power University,2012.

[8] Miao Song, Xin Guo and Chuan Chen, Discussion on the Economic Development Model of Shanghai Express Industry Headquarters, Logistics Engineering and Management, 2015, 37 (07):37-39.

[9] Shuai Li, Research and Implementation of Time Series Analysis Based on Large-scale Express Data, Beijing University of Posts and Telecommunications,2018.