THE GRAY-MARKOV MODEL FOR FORECASTING AIR TRAFFIC FLOW BASED ON RESIDUAL CORRECTION

Shuang GUO
Master Candidate
Research Base of Air Traffic Management
Civil Aviation University of China
100 Xunhai Road, Dongli District, Tianjin. 300300 China
E-mail:guosh830819@126.com

Zhaoning ZHANG
Professor
Research Base of Air Traffic Management
Civil Aviation University of China
100 Xunhai Road, Dongli District, Tianjin. 300300 China
Fax: +86-022-24092433
E-mail:zzhaoning@263.net

Lili WANG
Associate Professor
Research Base of Air Traffic Management
Civil Aviation University of China
100 Xunhai Road, Dongli District, Tianjin. 300300 China

Yan GUO
Instructor
College of Humanities
Civil Aviation University of China
100 Xunhai Road, Dongli District, Tianjin. 300300 China

Abstract: The flight flow forecast has an important significance for the studies of strategic development of civil aviation. The Gray forecast method and the Markov forecast method are taken together, and the Gray-Markov model is established. The Gray model reveals the general trend of long-term development of the air traffic flow. The transition of states is determined by the Markov model. Through the practical application of forecasting air traffic flow, we can see that this model can obtain a high accuracy.

Key words: air traffic flow, Medium-term and long-term forecasting model, Gray-Markov model

1. FOREWORD

1.1 The necessity and significance of medium-term and long-term forecasts
Air traffic flow is the foundation of air traffic management. Along with the development of country's economies, the air traffic flow also continued a rapid growth. Medium-term and long-term scientific and accurate forecasts of air traffic flow is an effective pledge to sustained and broadened air traffic flow. It is a basis for all levels of decision-making to make development strategies and planning.

1.2 Present research condition
In the year of 2001, WANG and XIA applied the developed BP network model to the dynamic forecast of aircraft track (2001). In 2002, HE used the econometrics and tendency forecasting method to predict the freight volume of civil aviation (2002). In 2003, WEI, CHEN and YUAN presented a new flight forecasting model that is based on C-mean clustering algorithm; LIU presented a traffic flow forecasting model based on least square estimate; ZHAO used the linear regression to forecast the passengers and freights handled at Tianjin Binhai airport; CHEN used the linear regression to forecast some economic indexes of civil aviation transportation market (2003). In 2004, WEI, FENG established a model in an application of the “support vector regression method”, which is formulated on the principle of minimizing a structural risk to predicate passengers’ throughput of an airport (2004).

1.3 The methods used in this paper

Because the Gray forecast is based on the gray $GM(1,1)$ model, and its graphics is a relatively smooth exponential curve, so it is poor fit for the data which is of greater volatility, and it has a low accuracy. But the object which Markov forecast studies is a dynamic system changed randomly. It predicts the future system by the transition probability between states. It provides a new approach for the data which is of greater volatility. Gray-Markov model can accurately reflect changes of the random rule in the air traffic, the reason is not only that it can be very close to the data, but also it changes with the air traffic flow system, leading to a dynamic response.

So, based on the data of air traffic flow, firstly, the $GM(1,1)$ model was used to obtain the forecasting sequence, then, the residual correction was used to amend the gray $GM(1,1)$ model and get the forecasting sequence which has been revised, finally, the Markov model was used to get the transition probability matrix and reach a long-term forecast of air traffic flow.

2. THE GRAY $GM(1,1)$ MODEL BASED ON RESIDUAL CORRECTION

Gray theory puts emphasis on “small samples of uncertainty and poor information”. It seeks the law by sequence generated and coverage of information. Its characteristic is “less data modeling”(Liu and Dang, 1999).

2.1 Gray Model

Set the sequence of the original data as:

$$X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n)], x^{(0)}(k) \geq 0, k = 1, 2, \ldots, n$$

So $X^{(0)}$’s 1–AGO sequence is: $X^{(1)} = [x^{(1)}(l), x^{(1)}(2), \ldots, x^{(1)}(n)]$
Where \( x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \ldots, n \)

Establish the gray differential equations as \( ax^{(1)}(k) + \frac{dx^{(1)}}{dt} = u \) (3)

To remembered as: \( GM(1,1) \), parameter is: \( \hat{a}, \hat{a} = (a, b)^T \)

Parameters obtained by the least squares method is: \( \hat{a} = (B^TB)^{-1}B^TY \) (4)

\[
\begin{align*}
B &= \begin{bmatrix}
-(x^{(1)}(1) + x^{(1)}(2)) / 2 & 1 \\
-(x^{(1)}(2) + x^{(1)}(3)) / 2 & 1 \\
\vdots & \vdots \\
-(x^{(1)}(n-1) + x^{(1)}(n)) / 2 & 1
\end{bmatrix}, \\
Y &= \begin{bmatrix}
x^{(0)}(2) \\
x^{(0)}(3) \\
\vdots \\
x^{(0)}(n)
\end{bmatrix}
\end{align*}
\]

Response time sequence is: \( \tilde{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a}e^{-ak} + \frac{b}{a}; k = 1, 2, \ldots, n \) (5)

Reduction for regressive is:
\[
\tilde{x}^{(1)}(k+1) = x^{(1)}(k+1) - \tilde{x}^{(1)}(k) = (1 - e^a)(x^{(0)}(1) - \frac{b}{a})e^{-a(k-1)}, \quad k = 1, 2, \ldots, n
\] (6)

### 2.2 Residual Correction

Because the accuracy of the \( GM(1,1) \) model is limited, we use the residual correction model to correct it.

Assumed the residual sequence of \( X^{(1)} \) as: \( \epsilon^{(0)} = (\epsilon^{(0)}(1), \epsilon^{(0)}(2), \ldots, \epsilon^{(0)}(n)) \) (7)

Where \( \epsilon^{(0)}(k) = x^{(1)}(k) - \tilde{x}^{(1)}(k) \), if there is \( k_0 \), which meets:

1. \( \forall k \geq k_0 \), the symbols of \( \epsilon^{(0)}(k) \) are the same;

2. \( n - k_0 \geq 4 \);

So we can call \( \{ |\epsilon^{(0)}(k_0)|, |\epsilon^{(0)}(k_0+1)|, \ldots, |\epsilon^{(0)}(n)| \} \) as a residual tail that can be modeling. To remembered as: \( \epsilon^{(0)} = (\epsilon^{(0)}(k_0), \epsilon^{(0)}(k_0+1), \ldots, \epsilon^{(0)}(n)) \)

\( \epsilon^{(0)} \)'s \( 1 - AGO \) sequence is: \( \epsilon^{(1)} = (\epsilon^{(1)}(k_0), \epsilon^{(1)}(k_0+1), \ldots, \epsilon^{(1)}(n)) \) (8)

\( \epsilon^{(0)} \)'s \( GM(1,1) \) response time sequence is:
\[
\epsilon^{(1)}(k+1) = (\epsilon^{(0)}(k_0) - \frac{b}{a_e} \exp[-a_e(k-k_0)] + \frac{b}{a_e}, k \geq k_0
\] (9)
So \( \epsilon^{(0)} \)'s forecasting sequence is: 
\[
\tilde{\epsilon}^{(0)} = (\tilde{\epsilon}^{(0)}(k_0), \tilde{\epsilon}^{(0)}(k_0+1), \cdots, \tilde{\epsilon}^{(0)}(n))
\]  
(10)

Where 
\[
\tilde{\epsilon}^{(0)}(k_0+1) = (-a_{\epsilon}(\epsilon^{(0)}(k_0) - \frac{b_{\epsilon}}{a_{\epsilon}}) \exp[-a_{\epsilon}(k-k_0)], k \geq k_0
\]

The residual correction \( GM(1,1) \)'s response time sequence is:

\[
\tilde{x}^{(0)}(k+1) = \begin{cases} 
(1-e^a)x^{(0)}(1) - \frac{b}{a} e^{-a(k-1)}, & k<k_0 \\
(1-e^a)x^{(0)}(1) - \frac{b}{a} e^{-a(k-1)} \pm a_{\epsilon}(\epsilon^{(0)}(k_0) - \frac{b_{\epsilon}}{a_{\epsilon}}) e^{-a_{\epsilon}(k-k_0)}, & k \geq k_0 
\end{cases}
\]  
(11)

3. THE GRAY-MARKOV FORECASTING MODEL

Figure of the \( GM(1,1) \) model is a smooth curve, so the precision of the model is low if the sequence of data fluctuates tempestuously and it can not reflect variation rules of stochastic data. The Markov process forecasts the trend of system’s development based on the transition probability matrix (Liu and Dang, 2005). The model reflects influence of random factors and transition rules of states, so it can be used to the forecasting of random data sequence.

Therefore the advantages of \( GM(1,1) \) model and Markov process combined to establish the Gray-Markov model, the \( GM(1,1) \) model is used to reflect development trend of data and the Markov process is used to reflect transition rules of states. The Gray-Markov model not only makes sufficient use of the information of data, but also improves the precision of forecasting sequence.

3.1 Data Preprocess

Based on the data requirements of Markov model, we choose a percentage of the year incremental raw of air traffic flow as the original data.

The forecasting sequence of air traffic flow get by the gray \( GM(1,1) \) model based on residual correction is: 
\[
\tilde{X}^{(0)} = [\tilde{x}^{(0)}(1), \tilde{x}^{(0)}(2), \cdots, \tilde{x}^{(0)}(n)]
\]

The sequence of the percentage of the year increment is:

\[
D = (d(1), d(2), \cdots, d(n-1))
\]  
(12)
Where \( d(k) = \frac{\tilde{x}^{(0)}(k+1) - \tilde{x}^{(0)}(k)}{\tilde{x}^{(0)}(k)} \times 100, k = 1, 2, \cdots, n-1 \)

### 3.2 State Divided
The state division is based on the \( D \)-Curve, it divided to some strip regions which are intersect with the \( D \)-Curve, and every regional is a state.

It assumes that there are \( n \) types of states. every state \( \varTheta_i \) can be indicated as \( \varTheta_i = [\theta_{si}, \theta_{ti}] \), where \( \theta_{si} = A_i, \quad \theta_{ti} = B_i \quad (i = 1, 2, \cdots, n) \)

### 3.3 State transition probability matrix’s calculation
\[
P_{ij} = \frac{M_{ij}(m)}{M_i} \quad (i,j = 1, 2, \cdots, n) \tag{13}
\]

\( M_{ij}(m) \) is the number of original data samples which is the state \( \varTheta_j \) that is transferred from \( \varTheta_j \) by \( m \) steps; \( M_i \) is the number of original data samples in state \( \varTheta_i \).

Transition probability matrix is:
\[
R(m) = \begin{bmatrix}
P_{11} & P_{12}(m) & \cdots & P_{1m}(m) \\
P_{21} & P_{22}(m) & \cdots & P_{2m}(m) \\
\vdots & \vdots & \ddots & \vdots \\
P_{n1} & P_{n2}(m) & \cdots & P_{nm}(m)
\end{bmatrix}
\]

\( R(m) \) reflects the rules of transition between the states. \( P_{ij}(m) \) reflects the probability of the state \( \varTheta_j \) that can be transferred from \( \varTheta_j \) by \( m \) steps.

### 3.4 Determine the fluctuation interval and the forecast
Through investigation of the one step transition probability matrix, we decide on the future transfer state, so the \( \theta_{si}, \theta_{ti} \) are decided, and the fluctuation interval is \( [\theta_{si}, \theta_{ti}] \). The most likely forecast \( Y(k) \) can be calculated by the following formula:
\[
Y(k) = (\theta_{si} + \theta_{ti}) / 2
\]

So:
\[
Y(k) = \tilde{X}^{(0)} + (A_i + B_i) / 2 \tag{14}
\]

### 3.5 Data Post process
Reduction of data that is the value of air traffic flow forecast is:
\[ \tilde{Q} = (\tilde{q}(1), \tilde{q}(2), \cdots, \tilde{q}(n)) \]

Where \( \tilde{q}(k) = q(k-1) \cdot [1 + Y(k)] / 100, \ k = 1, 2, \cdots, n \)

4. THE GRAY-MARKOV MODEL BASED ON RESIDUAL CORRECTION

The steps of the gray-Markov model based on residual correction can be drawn from the above description:

1> Put the sequence of the original data \( X^{(0)} = [x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)] \) to

\[ X^{(1)} = [x^{(1)}(1), x^{(1)}(2), \cdots, x^{(1)}(n)], \] where \( x^{(1)}(k) = \sum_{i=1}^{k} x^{(0)}(i), k = 1, 2, \cdots, n. \)

2> The gray GM(1,1)'s response time sequence is:

\[ \tilde{x}^{(1)}(k+1) = (x^{(0)}(1) - \frac{b}{a})e^{-ak} + \frac{b}{a}; k = 1, 2, \cdots, n \]

where \( \hat{a} = (a, b)^{T} = (B^{T}B)^{-1}B^{T}Y, \ B = \begin{bmatrix} -\frac{1}{2} & 1 \\ -\frac{1}{2} & 1 \\ \vdots & \vdots \\ -\frac{1}{2} & 1 \end{bmatrix}, \ Y = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix} \)

Reduction for regressive is:

\[ \tilde{x}^{(0)}(k+1) = \tilde{x}^{(1)}(k+1) - \tilde{x}^{(1)}(k) = (1 - e^{a})(x^{(0)}(1) - \frac{b}{a})e^{-a(k-1)}, \ k = 1, 2, \cdots, n \)

3> Assumed the residual sequence of \( X^{(1)} \) as: \( \varepsilon^{(0)} = (\varepsilon^{(0)}(1), \varepsilon^{(0)}(2), \cdots, \varepsilon^{(0)}(n)) \),

Where \( \varepsilon^{(0)}(k) = x^{(1)}(k) - \tilde{x}^{(1)}(k) \),

\( \varepsilon^{(0)} \)'s 1-AGO sequence is: \( \varepsilon^{(1)} = (\varepsilon^{(1)}(k_{0}), \varepsilon^{(1)}(k_{0}+1), \cdots, \varepsilon^{(1)}(n)) \)

4> Establish the residual correction GM(1,1) model, its response time sequence is:

\[ \tilde{\varepsilon}^{(0)}(k+1) = \begin{cases} 
(1 - e^{a})(x^{(0)}(1) - \frac{b}{a})e^{-a(k-1)}, & k < k_{0} \\
(1 - e^{a})(x^{(0)}(1) - \frac{b}{a})e^{-a(k-1)} \pm a_{\varepsilon}(\varepsilon^{(0)}(k_{0}) - \frac{b}{a_{\varepsilon}})e^{-a_{\varepsilon}(k-k_{0})}, & k \geq k_{0}
\end{cases} \]

Where \( \hat{a}_{\varepsilon} = (a_{\varepsilon}, b_{\varepsilon})^{T} = (B_{\varepsilon}^{T}B_{\varepsilon})^{-1}B_{\varepsilon}^{T}Y_{\varepsilon} \)

\[
B_ε = \begin{bmatrix}
-ε^{(1)}(1) + ε^{(1)}(2) / 2 & 1 \\
-ε^{(1)}(2) + ε^{(1)}(3) / 2 & 1 \\
... & ...
\end{bmatrix}, \quad Y = \begin{bmatrix}
ε^{(0)}(2) \\
ε^{(0)}(3) \\
... \\
ε^{(0)}(n)
\end{bmatrix}
\]

So the air traffic flow forecasting sequence is:

\[
\tilde{X}^{(0)} = [\tilde{x}^{(0)}(1), \tilde{x}^{(0)}(2), ..., \tilde{x}^{(0)}(n)]
\]

5> The original sequence of Markov forecast is: \( D = (d(1), d(2), ..., d(n-1)) \),

where \( d(k) = \frac{\tilde{x}^{(0)}(k+1) - \tilde{x}^{(0)}(k)}{\tilde{x}^{(0)}(k)} \times 100, k = 1, 2, ..., n-1 \)

state division: \( Θ = [θ_{1i}, θ_{2i}] \), where \( θ_{1i} = A_i, \quad θ_{2i} = B_i \quad (i = 1, 2, ..., n) \)

6> State transition probability matrix’s calculation:

\[
R(m) = \begin{bmatrix}
P_{11} & P_{12}(m) & ... & P_{1m}(m) \\
P_{21} & P_{22}(m) & ... & P_{2m}(m) \\
... & ... & ... & ... \\
P_{n1} & P_{n2}(m) & ... & P_{nm}(m)
\end{bmatrix}
\]

where \( P_{ij} = M_j(m) / M_i \quad (i,j = 1, 2, ..., n) \), \( M_j(m) \) is the number of original data samples which is the state \( Θ \) that is transferred from \( Θ_j \) by \( m \) steps; \( M_i \) is the number of original data samples in state \( Θ_i \).

7> Through investigation of the one step transition probability matrix, we decide on the future transfer state, the fluctuation interval is \([θ_{1i}, θ_{2i}]\). The most likely forecast \( Y(k) \) can be calculated by the following formula:

\[
Y(k) = (θ_{1i} + θ_{2i}) / 2
\]

8> The value of air traffic forecast is: \( \tilde{Q} = (\tilde{q}(1), \tilde{q}(2), ..., \tilde{q}(n)) \)

where \( \tilde{q}(k) = q(k-1) \cdot [1 + Y(k)] / 100, \quad k = 1, 2, ..., n \)

5. SIMULATION OF THE FORECAST MODEL

Based on the data of East China region air traffic flow from 1993 to 2005, the result of the gray-Markov model based on residual correction is showed in table 5-1.
In the table, $X^{(0)}$ is the data sequence of East China region air traffic flow; $X^{(1)}$ is $1-AGO$ sequence, $\tilde{X}^{(0)}$ is forecast result of $X^{(0)}$ using the GM(1,1) model; $\tilde{X}^{(0)}$ is forecast result of $X^{(0)}$ using the GM(1,1) model based on residual correction; $D$ is percentage of the year increment of $\tilde{X}^{(0)}$ using the GM(1,1) model based on residual correction; $\hat{Q}$ is forecast result of Gray-Markov model.

<table>
<thead>
<tr>
<th>Year</th>
<th>$X^{(0)}$</th>
<th>$X^{(1)}$</th>
<th>$\tilde{X}^{(0)}$</th>
<th>Error</th>
<th>$\hat{X}^{(0)}$</th>
<th>Error</th>
<th>$D$</th>
<th>$\hat{Q}$</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>194118</td>
<td>194118</td>
<td>194118</td>
<td>0%</td>
<td>194118</td>
<td>0%</td>
<td>194118</td>
<td>0%</td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td>222460</td>
<td>416578</td>
<td>238119</td>
<td>7%</td>
<td>238119</td>
<td>7%</td>
<td>22.67</td>
<td>218383</td>
<td>1.8%</td>
</tr>
<tr>
<td>1995</td>
<td>272957</td>
<td>689535</td>
<td>267540</td>
<td>2%</td>
<td>267540</td>
<td>2%</td>
<td>12.36</td>
<td>279790</td>
<td>2.5%</td>
</tr>
<tr>
<td>1996</td>
<td>328193</td>
<td>1017728</td>
<td>300597</td>
<td>8.4%</td>
<td>317951</td>
<td>3.1%</td>
<td>18.84</td>
<td>314360</td>
<td>4%</td>
</tr>
<tr>
<td>1997</td>
<td>363785</td>
<td>1381513</td>
<td>337738</td>
<td>7.2%</td>
<td>374562</td>
<td>4%</td>
<td>17.8</td>
<td>335441</td>
<td>7.8%</td>
</tr>
<tr>
<td>1998</td>
<td>402448</td>
<td>1783961</td>
<td>379468</td>
<td>5.7%</td>
<td>378617</td>
<td>6.9%</td>
<td>1.08</td>
<td>417637</td>
<td>3.8%</td>
</tr>
<tr>
<td>1999</td>
<td>421159</td>
<td>2205120</td>
<td>426354</td>
<td>1.2%</td>
<td>421862</td>
<td>0.2%</td>
<td>11.42</td>
<td>425944</td>
<td>1.1%</td>
</tr>
<tr>
<td>2000</td>
<td>460829</td>
<td>2665949</td>
<td>479033</td>
<td>4%</td>
<td>474919</td>
<td>4.1%</td>
<td>12.58</td>
<td>474595</td>
<td>3%</td>
</tr>
<tr>
<td>2001</td>
<td>533375</td>
<td>3199324</td>
<td>538222</td>
<td>1%</td>
<td>534455</td>
<td>0.2%</td>
<td>12.54</td>
<td>534284</td>
<td>0.2%</td>
</tr>
<tr>
<td>2002</td>
<td>597541</td>
<td>3796865</td>
<td>604723</td>
<td>1.2%</td>
<td>601273</td>
<td>1.6%</td>
<td>12.5</td>
<td>601262</td>
<td>0.6%</td>
</tr>
<tr>
<td>2003</td>
<td>623611</td>
<td>4420476</td>
<td>679441</td>
<td>9%</td>
<td>641937</td>
<td>2.9%</td>
<td>6.76</td>
<td>634343</td>
<td>1.7%</td>
</tr>
<tr>
<td>2004</td>
<td>792472</td>
<td>5212948</td>
<td>763391</td>
<td>3.7%</td>
<td>763391</td>
<td>3.7%</td>
<td>18.92</td>
<td>754276</td>
<td>4.8%</td>
</tr>
<tr>
<td>2005</td>
<td>892965</td>
<td>6105913</td>
<td>857713</td>
<td>0.6%</td>
<td>857713</td>
<td>0.6%</td>
<td>12.36</td>
<td>858815</td>
<td>3.8%</td>
</tr>
<tr>
<td>Mean error</td>
<td>3.9%</td>
<td>2.8%</td>
<td>2.7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

State divided of $D$ using the Markov model is: $\Theta_1 = [0,11)$, $\Theta_2 = [11,12)$, $\Theta_3 = [12,13)$, $\Theta_4 = [13,15)$, $\Theta_5 = [15,20)$, $\Theta_6 = [20,25)$

Transition probability matrix is:

\[
R^{(1)} = \begin{bmatrix}
0 & 0.5 & 0 & 0 & 0.5 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0.25 & 0 & 0.5 & 0 & 0 & 0.25 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0.5 & 0 & 0.5 & 0 \\
0 & 0 & 1 & 0 & 0 & 0
\end{bmatrix}
\]

So: the value of East China region air traffic flow forecast in 2006 is: 100459 sorties
the value of East China region air traffic flow forecast in 2010 is: 1609152 sorties
the value of East China region air traffic flow forecast in 2015 is: 2899744 sorties
6. CONCLUSIONS

In this paper, we focus on the characteristics of data of air traffic flow; present the gray-Markov model based on residual correction. The $GM(1,1)$ model is used to reflect development trend of data and the Markov process is used to reflect transition rules of states. So the Gray-Markov model not only makes sufficient use of the information of data, but also improves the precision of forecasting sequence. Through forecasting the value of East China region air traffic flow, it shows that this model can be used in medium-term and long-term forecast of air traffic flow, and it has an excellently accuracy.

This model is based on the statistical analysis of historical data, so the more data we have, the more excellently accuracy will be obtained. In addition, forecast accuracy is related with the state division, while the division of states and number of states are determined according to the request of information and the problem.

REFERENCE


