Damp trend Grey Model forecasting method for airline industry

Rafael Bernardo Carmona Beníteza, Rafael Bernardo Carmona Paredesb, Gabriel Lodewijks, Joao Lemos Nabaisd

aUniversidad Anáhuac México Norte, School of Business and Economics, Av. Universidad Anáhuac No. 46, Col. Lomas Anáhuac, Huixquilucan, 52786 State of Mexico, Mexico
bUniversidad Nacional Autónoma de México (UNAM), Institute of Engineering, Mechanical Engineering, Torre de Ingeniería Piso 2 ala Norte, Ciudad Universitaria, 04510 Mexico City, Mexico
cDelft University of Technology (TUDELFT), Department of Transport Engineering and Logistics, Mekelweg 2, 2628CD Delft, The Netherlands
dIDMEC, Department of Informatics and Systems Engineering, Setubal School of Technology, Polytechnic Institute of Setubal, Setubal, Portugal

A R T I C L E   I N F O

Keywords:
Grey models
Times-series
GM (1, 1)
Forecasting methods
Air passenger forecasting

A B S T R A C T

This paper presents a modification of the Grey Model (GM) to forecast routes passenger demand growth in the air transportation industry. Forecast methods like Holt-Winters, autoregressive models, exponential smoothing, neural network, fuzzy logic, and Grey Model (GM) is the fact that these models tend to calculate high airlines routes pax growth for long lead-time forecasting (Grubb & Mason, 2001) (Gardner & McKenzie, 1985; Gardner & McKenzie, 1988; Gardner & McKenzie, 1989). Another problem when forecasting airlines pax flow growth for long lead-time forecasting is the quantity of data points available and needed to use any of these methods. A forecasting method able to solve both problems allows estimating reasonable airlines routes pax flow growth for relatively new routes; it is very important for airlines making decisions of network planning, network management, fleet assignment, man power planning, aircraft routing, flight scheduling, revenue management, new routes and investments.

1. Introduction

In this paper, a new version of the Grey Model (GM) forecasting method is proposed. In this version, a damping trend factor has been included to the GM model. It forecasts reasonable airlines routes passenger (pax) growth for long lead-time.

A problem when using forecasting methods such as Holt-Winters, autoregressive models, exponential smoothing, neural network, fuzzy logic, and Grey Model (GM) is the fact that these models tend to calculate high airlines routes pax growth for long lead-time forecasting (Grubb & Mason, 2001) (Gardner & McKenzie, 1985; Gardner & McKenzie, 1988; Gardner & McKenzie, 1989). Another problem when forecasting airlines pax flow growth for long lead-time forecasting is the quantity of data points available and needed to use any of these methods. A forecasting method able to solve both problems allows estimating reasonable airlines routes pax flow growth for relatively new routes; it is very important for airlines making decisions of network planning, network management, fleet assignment, man power planning, aircraft routing, flight scheduling, revenue management, new routes and investments.

Armstrong (2006) reviewed forecasting methods. He recommended the damping trend as well established forecasting method to improve accuracy in practical applications. Despite of these improvements, Armstrong (2006) explained that a damp trend factor has been added in small number of forecasting methods. Filides, Wei, and Ismail (2008) and Hyndman, Koehler, Ord, and Snyder (2008) found and concluded that using a damping trend factor is favorable for forecasting exponential smoothing method.

This paper shows that forecasting pax flow between cities origin and cities destination (O–D pairs) using a GM model without a damping trend factor does not forecast reasonable data. Thus, the authors proposed to add a damping trend parameter (c) to modify the trend component in the GM model forecasting method.

The modified GM model is proposed to estimate reasonable routes pax flow between cities/airports (O–D pairs) when having a minimum of four data points. The GM model has three main advantages. Firstly, the GM model forecasts data that have unknown parameters. Secondly, the GM model requires few data (minimum 4 data points) to approximate the behavior of unknown systems. A big advantage because there are many circumstances in which the data is not enough to perform a good forecast for long lead-times using other forecasting methods. Thirdly, the GM model has been used by other researcher, such as Hsu and Wen (2000, 2002, 2003), to create data for the design of airline networks without assessing how good is the GM forecast. Hsu and Wen (2000, 2002, 2003) did not prove if the GM model is an accurate forecasting method to estimate reasonable airlines routes pax growth for
long lead-time. This paper analyses the feasibility of using the modification of the GM model for long lead-times proposed by Carmona Benitez (2012).

In Section 2, the damping trend factor is added to the GM model to forecast reasonable long lead-time data for the airline pax industry. In Section 3, the source of data used to prove the forecasting method is presented (DOT US Consumer Report, 2005–2008). In Section 4, the results are shown for 9 extreme study cases. Finally, Section 5 concludes this chapter.

2. Grey Model design “damping trend parameter”

This paper modified a first order one variable GM model or GM (1,1) algorithm. The GM (1,1) first order one variable model is the most common GM model in the literature. This model is a time series forecasting model with time-varying coefficients. These coefficients are renewed as the new data become available. It means, the more recent data have more influence than old data.

Pax flow data between O–D pairs are always positive. Since, all the previous data points are positive, GM models can be used to forecast pax flow data points (Kayacan, Ulutas, & Kaynak, 2010).

Based on Kayacan et al. (2010) a first order one variable GM (1,1) model algorithm, calculations for the civil aviation industry were found to increase/decrease too fast or very high. Then, their model forecast high values or negative values if the tendency goes down. This values are completely unreasonable because they are simply too high or negative. Negative values are not possible since the demand is always major or equal to zero.

In this paper, Kayacan et al. (2010) model has been modified to forecast more reasonable values for the air pax industry between O–D pairs for long lead-times. A parameter that damps the model calculations has been added. This parameter is based on the assumption that routes pax flows get more stable as their demands increase. This means pax flow growth in a slower rate than in previous years. Thus, in the first year’s routes pax flows are expected to increase faster than at the end of the forecasting year.

According with Kayacan et al. (2010) the first-order differential equation that generates the grey model is:

\[
\frac{dQ(t)}{dt} + aQ(t) = b
\]  

(1)

The GM model adapted and used to perform a long lead-time forecast, for the civil aviation industry, is as follows:

Consider a time series data \( Q^{(O)} \) that denotes the number of pax on an airline route.

\[
Q^{(O)} = \{Q^{(O)}(1), Q^{(O)}(2), \ldots, Q^{(O)}(n)\}, \quad n \geq 4
\]  

(2)

where: \( n \) is the sample size of the data, minimum four [−] and \( O \) the denotes original data point.

\( Q^{(O)} \) is a non-negative sequence. A monotonically increasing sequence, \( Q^{(C)} \), is calculated by:

\[
Q^{(C)} = \{Q^{(C)}(1), Q^{(C)}(2), \ldots, Q^{(C)}(n)\}, \quad n \geq 4
\]  

(3)

where:

\[
Q^{(C)}(k) = \sum_{i=1}^{k} Q^{(O)}(i), \quad k = 1, 2, 3, \ldots, n
\]  

(4)

\( C \) denotes the accumulative data point.

The generated mean sequence \( X \) of \( Q^{(C)} \) is the mean value of the next data point and is defined as:

\[
X = \{X(1), X(2), \ldots, X(n)\}, \quad n \geq 4
\]  

(5)

where:

\[
X(k) = 0.5Q^{(C)}(k) + 0.5Q^{(C)}(k-1), \quad k = 2, 3, \ldots, n
\]  

(6)

According with Kayacan et al. (2010), the solution by least square method (OLS) of the grey differential equation of GM (1,1) requires calculating the coefficients \( a, b \). \( [a, b]^T \) is a sequence of variables where a solve the b estimation problem and can be found as follows:

\[
[a, b]^T = (B^T B)^{-1} B^T Q^T
\]  

(7)

where:

\[
Q^T = \{Q^{(O)}(2)Q^{(O)}(3) \ldots Q^{(O)}(n)\}^T
\]  

(8)

\[
B^T = \begin{bmatrix}
-X(2) & -X(3) & \ldots & -X(n)
\end{bmatrix}^T
\]  

(9)

The forecast value for \( Q \) sequence at time \( k \) is:

\[
Q^{(C)}_{est}(k + 1) = \left[Q^{(O)}(1) + \frac{b}{a}\right] \exp^{-ak} + \frac{b}{a},
\]  

(10)

\[
Q^{(C)}_{est}(1) = Q^{(O)}(1)
\]  

(11)

where: \( K \) is the counter number of years until the forecast wants to be performed i.e. from 2005 to 2050, \( K = 45 [-] \) and \( est \) the denotes the forecast data point [−].

The GM model forecast high \( Q_{est}(t) \) values for long lead-time. For this reason, a damping trend parameter was introduced to the GM model to forecast more reasonable routes \( Q_{est}(t) \). The damping trend parameter intends to minimize the exponential increase of the GM model calculations, because as time increases the GM model calculations increase exponential. Then, the damping trend parameter reduces the GM model calculations as distance increases. The \( Q_{est}(t) \) estimation at time \( k \) is equal to:

\[
Q^{(O)}_{est}(k) = (Q^{(O)}_{est}(k) - Q^{(O)}_{est}(k-1)), \quad k = 2, 3, \ldots, n
\]  

(12)

\[
Q^{(O)}_{est}(k) = (Q^{(O)}_{est}(k) - Q^{(O)}_{est}(k-1)) \frac{1}{\exp^{5(\pi - a)K}},
\]  

(13)

where: \( \pi \) is the damping trend parameter [−].

The damping trend parameter works under the assumption that any route pax flow will not grow more than a maximum possible increment. In this paper, the maximum possible grow is determined by the International Air Transport Association (IATA) forecast, from 2006 to 2050.

In 2006, 760 million passengers traveled around the world (IATA, 2007). IATA has forecasted 3.3 billion air pax by 2014 (IATA, 2011a) and 16 billion air pax by 2050 (IATA, 2011b). Then, the maximum possible increment from 2006 to 2050 is expected to be 21.05 times. This allows forecasting the pax flow from 2009 to 2050 by using the GM model at the last time \( k \), in this case 2050.

The damping trend parameter is calculated as follows:

\[
\zeta = \frac{1}{K} \ln \left(\frac{Q^{(C)}_{est}(K) - Q^{(C)}_{est}(K-1)}{Q^{(O)}_{est}(2) - Q^{(O)}_{est}(1)}\right)
\]  

(14)

Eq. (15) is the constraint that allows calculating the damping trend parameter with Eq. (14):

\[
Q^{(O)}_{est}(K) = AQ \times Q^{(O)}_{est}(2)
\]  

(15)

where: \( AQ \) is the The expected growth from \( Q^{(O)}_{est} \) to \( Q^{(O)} \). IATA expected growth, 21.05

In order to improve the accuracy of the model predictions, an error modification of the GM based on the Fourier series is explained. This method is described by Kayacan et al. (2010) and modified as follows.
Considering Eq. (10) and the predicted values given by the GM (1,1) model the error sequence of $Q^{(0)}$ can be determined as:

$$e^{(0)} = (e^{(0)}(2), e^{(0)}(3), \ldots, e^{(0)}(n))$$ (16)

where:

$$e^{(0)}(k) = Q^{(0)}_{\text{est}}(k) - Q^{(0)}_{\text{act}}(k), \quad k = 2, 3, \ldots, n$$ (17)

Now, expressing the residual error in Eq. (17) as Fourier series:

$$\text{Error}_k = e^{(0)}(k) \approx 0.5a_0 + \sum_{r=1}^{\infty} \left[a_r \cos \left(\frac{2\pi r}{T}k\right) + b_r \sin \left(\frac{2\pi r}{T}k\right)\right],$$ (18)

where:

$$T = n - 1$$ (19)

$$z = \left(\frac{n-1}{2}\right) - 1$$ (20)

Rewriting Eq. (18) as follows:

$$e^{(0)}_{\text{est}} \approx PC$$ (21)

where:

$$P = \begin{bmatrix}
0.5 \cos \left(\frac{2\pi}{T}\right) \sin \left(\frac{2\pi}{T}\right) & \cos \left(\frac{2\pi}{T}\right) \sin \left(\frac{2\pi}{T}\right) & \cdots & \cos \left(\frac{2\pi}{T}\right) \sin \left(\frac{2\pi}{T}\right) \\
0.5 \cos \left(\frac{3\pi}{T}\right) \sin \left(\frac{3\pi}{T}\right) & \cos \left(\frac{3\pi}{T}\right) \sin \left(\frac{3\pi}{T}\right) & \cdots & \cos \left(\frac{3\pi}{T}\right) \sin \left(\frac{3\pi}{T}\right) \\
\vdots & \vdots & \ddots & \vdots \\
0.5 \cos \left(\frac{n\pi}{T}\right) \sin \left(\frac{n\pi}{T}\right) & \cos \left(\frac{n\pi}{T}\right) \sin \left(\frac{n\pi}{T}\right) & \cdots & \cos \left(\frac{n\pi}{T}\right) \sin \left(\frac{n\pi}{T}\right)
\end{bmatrix}$$ (22)

$$C = [a_0 a_1 b_1 a_2 b_2 \cdots a_r b_r]^T$$ (23)

$$C \approx \left(P^T P\right)^{-1}P^T e^{(0)}$$ (24)

Finally, the Fourier series correction can be solved as follows:

$$Q^{(0)}_{\text{ef}}(k) = Q^{(0)}_{\text{est}}(k) - e^{(0)}_{\text{est}}(k), \quad k = 2, 3, \ldots, n$$ (25)

Then, the final pax flow estimated is $Q_{\text{ef}}(k)$.


The Airline Fares Consumer Report is published by the US Department of Transportation Office of Aviation Analysis. It includes information of approximately 18,000 routes operated by various airlines in the US economy. The reports include non-directional market passenger number, revenue, nonstop and direct mileage broken down by competitor. Only those carriers with a 10% or greater market share are listed. The total number of passenger flow is calculated for each route connecting two cities. In this paper, the data includes the total number of passenger traveling between cities from 2005 to 2008.

The GM model can forecast airlines routes passenger growth using the DOT US Consumer Report because data exist for 4 years. The passenger growth percentage change reflects how the demand of passengers fluctuates from 2005 to 2008 per route. The passenger growth percentage change allows studying the performance of the GM model for small, normal and high fluctuations. In this paper, it is possible to study 9 study cases. These cases are set up by the possible number of combination, positive and negative fluctuations, between years. Table 1 shows 9 routes that represent one of each of the possible scenarios. For example, the route case scenario from Long Beach to Chicago is $(-, +, -)$ because the increment was negative from 2005 to 2006, the increment was positive from 2006 to 2007, and the increment was negative from 2007 to 2008.

### 4. Simulation results

Fig. 1 illustrates the passenger flow forecasting values by using the GM (1,1) without the $\zeta$ parameter (left side) and with the $\zeta$ parameter (right side). The GM (1,1) forecasts without the $\zeta$ parameter forecast extremely high and unreasonable passenger flow. The GM (1,1) with the $\zeta$ parameter modification calculates reasonable airlines routes passenger flows. In Fig. 1, the route connects Long Beach with Chicago. In this route, the passenger flow increased too much from 2006 to 2007. This is the reason why the model forecast high values without using the $\zeta$ parameter. The route scenario is $(-, +, -)$. The model memorizes this fluctuation pattern to forecast future pax flow.

Fig. 2 shows the route case scenario $(+, +, -)$. In this case, the route is from Atlanta to Corpus Christi. In this route, the pax flow decreased too much from 2006 to 2008. The model memorizes this decrease to forecast future passenger flow. This is the reason why the GM (1,1) without the $\zeta$ parameter modification forecast negative values. On the right side, the GM (1,1) forecasts without the $\zeta$ parameter modification forecast more reasonable values. Although, the pax flow growths in future years; during the first years the pax forecasting decrease, and after some years start increasing.

Fig. 3 shows the route case scenario $(+, +, -)$. In this case, the route is from Honolulu to Ontario California. On the right side, the GM (1,1) calculations are simple wrong. The GM model cannot forecast routes with this type of case scenario. The passenger flow was increasing during the first years and it decreased between the last years. On the right side, Fig. 3 confirms that the GM (1,1) with the $\zeta$ parameter modification forecast more reasonable values. The GM model calculations are similar during the first years. After 2016, the GM model predictions start increasing.

Fig. 4 shows the route case scenario $(+, +, +)$. In this case, the route is from Westchester County to Orlando/Kissimmee. On the right side, the GM (1,1) calculations demonstrate the impossibility of the GM (1,1) to forecast airlines passenger growth. On the right side, Fig. 4 confirms that the GM (1,1) with the $\zeta$ parameter modification forecast more reasonable values. The GM model calculation for 2050 is 2.5 million passengers what is realistic.

Fig. 5 shows the route case scenario $(+, -)$. In this case, the route is from Atlanta to Pittsburgh. On the left side, the GM (1,1) simulation results suggest that the passenger flow between both cities will decrease in future years. This is hardly difficult to happen because Atlanta is a hub of Delta Airlines and both cities are highly populated. The results suggest that relations, such as economic and social, between Atlanta and Pittsburgh will decrease until any relation exists. On the right side, Fig. 5 suggests that the passenger flow between both cities will increase. The model estimates few pax flow increase during the first years. The model estimates pax flow growth exponentially, but the estimation for the last year is over 6 million pax a year what is a realistic value. It is more logic since both are big cities that have an important role in the US economy.

Fig. 6 shows the route case scenario $(-, +, +)$. In this case, the route is from Allentown to Detroit. On the right side, the GM (1,1) calculations demonstrate the impossibility of the GM (1,1) to forecast airlines passenger growth. On the right side, Fig. 6 confirms that the GM (1,1) with the $\zeta$ parameter modification forecast more reasonable values. The GM (1,1) calculation for 2050 is 55 thousand passengers what is a realistic value.

Fig. 7 shows the route case scenario $(-, +, +)$. In this case, the route is from Minneapolis/St. Paul to Oklahoma City. On the right side, the GM (1,1) calculations demonstrate the impossibility of the GM (1,1) to forecast airlines passenger growth. On the right side, Fig. 7 confirms that the GM (1,1) with the $\zeta$ parameter modification forecast more reasonable values.
parameter modification forecast more reasonable values. The GM model calculations show a slow increase in passenger flow for the first 20 years. By 2050, the calculation is 224,191 passengers. It is 623 passengers per day. This number of passengers could be operated by a B737-900 operating two flights from origin to destination and two flights from destination to origin per day.

Table 1
GM model simulation forecasting routes.

<table>
<thead>
<tr>
<th>Origin</th>
<th>Destination</th>
<th>Number of passengers transported</th>
<th>Percentage change</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long Beach</td>
<td>Chicago</td>
<td>14,390</td>
<td>4960</td>
</tr>
<tr>
<td>Atlanta</td>
<td>Corpus Christi</td>
<td>14,390</td>
<td>14,380</td>
</tr>
<tr>
<td>Honolulu</td>
<td>Ontario</td>
<td>7450</td>
<td>51,730</td>
</tr>
<tr>
<td>Westchester County</td>
<td>Orlando/Kissimmee</td>
<td>7180</td>
<td>57,970</td>
</tr>
<tr>
<td>Atlanta</td>
<td>Pittsburgh</td>
<td>282,790</td>
<td>288,920</td>
</tr>
<tr>
<td>Allentown</td>
<td>Detroit</td>
<td>19,470</td>
<td>4020</td>
</tr>
<tr>
<td>Minneapolis/St. Paul</td>
<td>Oklahoma City</td>
<td>32,360</td>
<td>17,040</td>
</tr>
<tr>
<td>New York</td>
<td>Los Angeles</td>
<td>2,426,710</td>
<td>2,409,280</td>
</tr>
<tr>
<td>Houston</td>
<td>Islip/Long Island</td>
<td>4,600</td>
<td>20,620</td>
</tr>
</tbody>
</table>

Fig. 1. Grey Model forecast without (left side) and with (right side) the $\zeta$ for the Long Beach – Chicago route from 2009 to 2050.

Fig. 2. Grey Model forecast without (left side) and with (right side) the $\zeta$ for the Atlanta – Corpus Christi route from 2009 to 2050.

Fig. 8 shows the route case scenario (–, –, –). In this case, the route is from New York to Los Angeles. These cities are between the most populated and economically active cities in the US and in the world. On the left side, the GM (1, 1) simulation results suggest that the passenger flow between both cities will decrease in future years. The results indicate that the economic relation
Fig. 3. Grey Model forecast without (left side) and with (right side) the $\gamma$ for the Honolulu – Ontario California route from 2009 to 2050.

Fig. 4. Grey Model forecast without (left side) and with (right side) the $\gamma$ for the Westchester County – Orlando/Kissimmee route from 2009 to 2050.

Fig. 5. Grey Model forecast without (left side) and with (right side) the $\gamma$ for the Atlanta – Pittsburgh route from 2009 to 2050.
Fig. 6. Grey Model forecast without (left side) and with (right side) the $\zeta$ for the Allentown – Detroit route from 2009 to 2050.

Fig. 7. Grey Model forecast without (left side) and with (right side) the $\zeta$ for the Minneapolis/St. Paul – Oklahoma City route from 2009 to 2050.

Fig. 8. Grey Model forecast without (left side) and with (right side) the $\zeta$ for the New York – Los Angeles route from 2009 to 2050.
between New York and Los Angeles will decrease if this would happen. This is very difficult, not to say impossible to happen. On the right side, Fig. 8 suggests that the passenger flow between both cities will increase. Although, it is more logic, the GM (1, 1) with the parameter calculations could be considered as high, especially after year 2020. The fact that the GM (1, 1) with the parameter calculations growth faster for this route than for the other routes under study demonstrates that the model is able to identify routes that are expected to show high increments of pax flow in the future.

Fig. 9 shows the route case scenario (+, −, +). In this case, the route is from Houston to Islip/Long Island. On the left side, the GM (1, 1) simulation results suggest that the passenger growth will decrease. By 2050, this route would have approximately the same number of passengers than it had during 2005. On the right side, Fig. 9 confirms that the GM (1, 1) with the parameter modification forecast more reasonable values. By 2050, the calculation is 417,159 passengers. It is 1159 passengers per day. This number of passengers could be operated by a B737-900 operating four flights from origin to destination and four flights from destination to origin per day. This is a reasonable estimation value.

The GM (1, 1) proposed in this paper has been run for all routes in the DOT US Consumer Report approximately 18,000 routes from airport origin to airport destination. These routes are equal to 5857 routes, operated by different airlines, from city origin to city destination. The model has been found to work properly for all airports and cities connections.

5. Conclusion

The modification to the GM (1, 1) is able to estimate more realistic results for long lead-time forecasts when the original data is little, 4 measures or 4 data points in the case of this study. The proposed model routes pax flows forecasts are more reasonable than using the GM prediction algorithm. However, it is important to understand that the GM (1, 1) could calculate good results when a major number of measures are used. It is because the GM (1, 1) will have more historical data per time t. Then, the GM model can memorize the behavior of the air passenger demand with more than 4 data points. For databases with a good amount of historic data, it may be possible that the GM (1, 1) without damping trend parameter estimates logic values. In that case, the GM (1, 1) with and without damping trend parameter need to be compared.

Acknowledgment

I thank my sponsor Consejo Nacional de Ciencia y Tecnología (CONACyT) Mexican Government, for giving me the opportunity to study a PhD at Delft University of Technology.

References


