Real time estimation of traffic flow and travel time  
Based on time series analysis

Wei Bao

Abstract – In this paper, the author study the traffic pattern and time series. After that, a time series analysis and estimation are made.

Keywords: Time Series, Estimation

I. Introduction
Real-time and accurate short-term traffic flow forecasting has become a critical problem in intelligent transportation systems. Based on the historical and current traffic flow data, we can to estimate the travel time and traffic flows of short-term future, like 5-30 minutes is often required. In this project, the traffic flows and travel times estimation and forecasting will be made on the statistical aspect, mainly by the time series analysis. The aim of this project is to identify and analyze given traffic data by analyzing it as a time series. Different models will be used to describe the historical data and current data, and then make some forecasting. By comparing the forecasting result to observed values, try to figure out the best model for a specific road segment.

II. Traffic patterns

A traffic system is a kind of multi-participators, constantly changing, high-complexity, non-linear system. One major characteristic of a traffic system is the high level of uncertainty with respect to traffic flows and travel times. There are a lot of factors that contribute to this uncertainty, for example, weather and seasons, from an environmental aspect; and traffic accidents and drivers’ psychology, from a human aspect. Short term traffic forecasting is easily influenced by these factors, even due to the fact that people are trying to avoid delays and incidents.

II.1. Traffic flows and travel times

According our basic knowledge, travel times and speeds are a pair of reciprocal, and there are certain relationship between traffic flows and speed. Following figure describe the relationship very well:

II.2. Season, trend and random components

When studying the traffic data in detail, it’s easy to found that the traffic data has some specialties like season, trend and random components.
Season: the way the data changed periodically;  
Trend: the way the data are likely to go;  
Random components: data changed with out season or trend attributes.

Because human life is a 24 hours cycle active, the traffic data also showed a 24 hours variety repeating. Meanwhile, the traffic data showed some longer term season attribute, like week (Monday go to work with high traffic flow in the city centre, Sunday go to relax cause high traffic flow in the rural area) and month(high traffic flow in the scenic spots on summer vacation). Trend is more easily distinguished from traffic data, like morning peak hour and afternoon peak hour, traffic flows increase with certain amount, and

These blue dots are plotted by the recorded data from the sensor, red and yellow dots are plotted by the formula as (Equation 1)

\[ q = \frac{k_f}{u_f} (u_f u - u^2) \]

\[ u_f red = 82 \quad k_f red = 96 \]
\[ u_f yellow = 62 \quad k_f yellow = 115 \]
then decreased in the night. From a longer term view, a year or decade trend will indicate the society developing situation. Random components in traffic data are most common ones we can saw. As we discussed in the first paragraph of chapter II, there are a lot of factors that contribute to this uncertainty (or we called random components).

Following figure 2 illustrate the season/trend/random clearly.

![Figure 2, one week data from Sensor 12/3 lanes, Tingstadstunneln E6, Göteborg, 19th-22nd, March, 2001](image)

Based on the figure 2, the season and trend could be distinguished as following:

**Season:** 24 hours season for every whole day, from Monday (19th, March 2001) to Thursday (22nd, March 2001).

**Trend:** Keep constantly in the middle night before 5:00 in the morning, then increase sharply from 5:00 and reach the peak at 7:30, after that, the traffic flow falls down a bit and increase slightly until reach the afternoon peak at 16:00. Then it decrease constantly until to the middle night. Both 4 days act almost the same during the 24 hours.

### III. Time Series

A time series is a set of observation $x_t$, each one being recorded at a specific time $t$. A discrete-time series is one in which the set $T_0$ of time at which observation are made is a discrete set, as is the case, for example, when observation are made at fixed time intervals. Continuous-time series are obtained when observation are recorded continuously over some time interval, eg, when $T_0=\{0,1\}$.

#### III.1. Exogenous model and Endogenous model

Exogenous models are models that include independent variables. An example of a exogenous model is:

$$ Y_t = a + bX_t + e_t $$

where $X_t$ is the independent variable, $a$ is a constant and $e_t$ is a random variable.

An endogenous model includes only endogenous variables. An example of an endogenous model with time lag one is:

$$ Y_t = a + b Y_{t-1} + e_t $$

This can be interpreted as the current results are based on the information known from the previous time step.

The error term $e_t$ is included to the model to capture noise in the data. The error term should have the following properties, in order to make a stationary model:

- the error term has a mean of zero;
- the error term has a constant variance over time;
- the error terms corresponding to different points in time are not correlated.

In this project, we will make forecasts of the future traffic flows data based on the historical and current traffic flow data. In this case the endogenous model is more suitable on doing this. If we also want to include more information like weather or incomes, it would require an exogenous variable in the model.

#### III.2. Smoothening, stationary and autorecorrelation

Smoothening method to produce data set is to create a function that attempts to capture important patterns in the data, while leaving out noise. One of the most common algorithms is the "moving average", often used to try to capture important trends in repeated statistical surveys.

Stationary is a state that the joint probability distribution does not change when shifted in time. For example, the white noise is stationary process.

Autocorrelation is the strength of a relationship between observations as a function of the time separation between them. More precisely, it is the cross-correlation of a signal with itself.

#### III.3. AR model and MA model[5]

If process of error term $\{e_t\}$ is auto correlated, we can build a model for capturing this information, example:

$$ Y_t = b Y_{t-1} + e_t $$

$$ e_t = p e_{t-1} + v_t, -1 < p < 1 $$

The model $e_t = p e_{t-1} + v_t$ is called a first order auto regrasive process, AR(1).

The parameter $p$ can be estimated by:

$$ e_t = y_t - by_{t-1} $$

$$ p = \frac{\sum e_t e_{t-1}}{\sum e_{t-1}^2} $$

Auto regressive model of higher order can be constructed as

$$ e_t = p_1 e_{t-1} + p_2 e_{t-2} + v_t, \text{AR(2)} $$

Another type of model for the error term $e_t$ is a moving average model. Here, the error term is assumed to be described by a weighted sum of earlier terms $v_t$. 

F. A. Author, S. B. Author
Another model type can be constructed by combining the AR and the MA models in an ARMA model:

\[ e_t = p e_{t-4} + v_t - d v_{t-4} \]

IV. The Holt-Winter algorithm [5]

According to Holt-Winter algorithm, the predicted value \( x_t^* \) is based on the level (1), the trend (2) and the season estimate (3).

\[ x_t^* = (1 - \alpha)(x_{t-4}^* + t_{t-4}) + \alpha(x_t - s_{t-p}), 0 < \alpha < 1 \]

(1)

The new level is estimated by weighting the earlier estimated level and trend in addition to the current observation and the season estimate with parameter \( \alpha \).

\[ t_t = (1 - \beta)t_{t-1} + \beta(x_t^* - x_{t-1}^*), 0 < \beta < 1 \]

(2)

The trend is estimated by the two earlier level estimates

\[ s_t = (1 - \gamma)s_{t-p} + \gamma(x_t - x_t^*), 0 < \gamma < 1 \]

(3)

The season estimate is based on previous season estimate and the difference between the estimated and the observed level.

\[ \hat{x}_{t+p} = x_t^* + \delta_{t} + s_{t-p} + \delta \]

(4)

This gives the predicted values \( \delta_{t} \) time steps ahead in time.

V. Methodology

General approach to time series modeling: [1]

VI. Plot the series and examine the features of the graph, checking in particular whether there is

(a) a trend,
(b) a seasonal components,
(c) any apparent sharp changes in behavior,
(d) any outlying observation.

VII. Remove the trend and seasonal components to get stationary residuals. To achieve this goal it may sometimes be necessary to apply a preliminary transformation to data.

VIII. Choose a model to fit the residuals, making use of various sample statistics including the sample autocorrelation function.

IX. Forecasting will be achieved by forecasting the residuals and then inverting the transformation described above to arrive at forecasts of the original series.

X. Realization

In the realization part, the traffic data recorded by then sensor 122 in Tingstadstunneln E6, Göteborg, during 19th-22nd, March, 2001 will be used. The software “ITSM 2000 –v7.1” student version, which also be used in the reference [1] will be used for parts of the analysis.

X.1. Data Entering and checking

By looking the excel file provided by supervisor, I found that both the traffic flow and traffic speed were recorded by 5 minutes interval, which means 288 datasets in one day and 1152 datasets during the whole week period (19th-22nd, March). But the student version of ITSM2000-v7.1 has a limitation of no more than 251 datasets, so I just combine four 5 minutes intervals into one 20 minutes intervals and only use the first 3 days’ data, which makes 216 datasets altogether. (In this project, only the traffic flow data will be analysis as time series)

After combining and entering the original traffic flow data into the time series software, we can see the figure as following:

![Traffic flow data figure](image1)

By doing this, it’s very easy to point out the seasons and trends attributes: a 24 hours season and 4 major trends in one day.

X.2. Data pre-processing

In order to achieve the stationary state, those seasons and trends attributes must be removed from the dataset. The “Classical Transform” function of ITSM2000 can be used to achieve the stationary state. Obviously the season intervals are 72(24*12/4), so enter the season period: 72 and choose “Liner Trend” in “Polynomial Fit”, we can get the figure of stationary state as:

![Stationary state](image2)

X.3. Model fitting

There 2 ways to find the suitable for the stationary dataset: by auto fit function or by looking ACF/PACF...
figure. An ARMA model of parameter \((p=4, q=4)\) is suggested by the auto fit function.

![Figure 5, Sample ACF and Sample PACF](image)

It also can be see that the Sample ACF and Sample PACF values falls down after 4 or 11. We choose an ARMA model with \((p=4, q=4)\) to fit the original data in this case.

X.4. ARMA Forecasting[4]

Once the model has been fitted, we can use the model to forecast the future value of data. Figure 6 shows the forecasted value of next 24 hours by ARMA forecasting function:

![Figure 6, 24 hours Forecast](image)

Plotting the forecast data and measured data in 22\(^{nd}\) March in a same figure, we can found the difference between them:

![Figure 7, Difference between measurements and forecast](image)

Table 1 illustrates the difference very well,

<table>
<thead>
<tr>
<th>Average Value</th>
<th>Measurement</th>
<th>Forecast</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Deviation</td>
<td>652.9722222</td>
<td>687.805556</td>
<td>94.94%</td>
</tr>
<tr>
<td>Correlation</td>
<td>464.8221575</td>
<td>478.977401</td>
<td>97.04%</td>
</tr>
</tbody>
</table>

Table 1, Difference value

From value in table 1, we believe the forecast value is quite close to the measurement data, which means a good model has been found.

X.5. Real time

So far, all the studies are still based on the historical data, in a settled period, but the basic theory and approach are almost the same as in the real time status. All we need to do in the real time status is to entering the current traffic data continuously into the original dataset and run the forecast step again, then the new forecast data will come out. In the reality work, this could be done automatically by a control computer connecting all the sensors.

XI. Travel times estimation

To most of the road users, travel times are more meaningful to them other than traffic flow or traffic speed. Once the traffic flow is forecasted by our model, we can use the Equation 1 in Chapter II to calculate the traffic speed \(V\). If we already the length of this road segment \(L\), divide \(L\) by speed \(V\), we can get a very rough local travel time \(T\).

If accurate globe travel times estimation is needed, then we must have travel flow forecasts all over the network, which means we need large mount current traffic data and build suitable model for every road. This is a really hard job, a better way to do this is build up a systemic estimations model.

XII. Conclusion

Although my result proved a nice estimation in specify road, it still has a long way to go to the practical situation.

To obtain a good model could describe the traffic patterns, abundant original data are necessary. Model fitting and choosing are the keys to result an accurate estimation or a bad one, which means people must very familiar with the ACF/PACF figure and have enough experience in model choosing.

One inevitable shortcoming of time series analysis and forecast in traffic aspect is that it can never model/forecast any traffic accidents.

Acknowledgements

This work was supported by Clas Rydergren

References

1. Peter J. Brockwell, Richard A. Davis, "Introduction to Time Series and Forecasting", second edition
2. Charles W. Ostrom, "Time Series Analysis"
3. Sun Xianghai, Liu Tanqiu, 
<<A study on urban short-term traffic flow forecasting 
based on a nonlinear time series model>>

4. HAN Chao, SONG Su, WANG Cheng-hong, 
<<A Real-time Short-term Traffic Flow Adaptive 
Forecasting Method Based on ARIMA Model>>

5. Clas Rydergren, course slide of TNK084, Traffic 
Theory

Time consume

Subjects covered and time plan: 
Statistics fundamental: 10 hours 
Time serious analysis studying: 20 hours 
ITSM2000 software studying: 5 hours 
Traffic data analysis: 5 hours 
Model built and calibration: 15 hours 
Forecasting and comparing: 10 hours 
Document: 5 hours 
Total work hours: 70 hours.