ANNUAL AVERAGE DAILY TRAFFIC FORECASTS IN HONG KONG

Yuen-fan TANG
Research Student
Department of Civil and Structural Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Fax: +852-23346389
E-mail: 99903588r@polyu.edu.hk

William H.K. LAM
Professor
Department of Civil and Structural Engineering
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong
Fax: +852-23346389
E-mail: cehklam@polyu.edu.hk

Abstract: In Hong Kong, the Annual Traffic Census report (ATC) is published in the middle of the year to provide the last year traffic statistics such as the Annual Average Daily Traffic (AADT) at the major and minor roads. In practice, there is always a need to use the most up-to-date traffic data such as AADT for transport model development and calibration. However, the current-year AADT data are always not available. This paper attempts to develop time series models based on historical ATC data and available current-year partial traffic counts to forecast the daily traffic flow for the whole current year in Hong Kong. As a result, 8 time series models are developed and validated against each of the observed daily flows for Monday to Friday, Saturday and Sunday. These models consider various combinations of the following factors: daily variation, monthly variation, historical and current-year partial daily flow and growth factor. The comparison results show that the time series model with the historical and current-year partial daily flow performs better than the other models. And this model was used to estimate the AADT for whole current year.

Key Words: Auto-Regressive Integrated Moving Average, daily variation, monthly variation, daily flow and growth factor

1. INTRODUCTION

Automatic traffic counters for the measurement of traffic volume were firstly used in Hong Kong in 1961. With gradual developments in the subsequent years, a comprehensive system was established in 1971. This system is known as the Annual Traffic Census (ATC) (Transport Department, 1965-1999). The existing system basically follows the methodology from the comprehensive review of the ATC system in Transport Department (1988).

Over 1,500 counting stations were operated in 1999 (Chan et al., 2000). The counting stations are classified as "Core" and "Coverage" stations according to the extent of data being collected. "Core" station is a randomly selected counting station located on a road link of any class providing hourly, daily and monthly factors to generalize the characteristic for its own group of links. "Coverage" stations is a counting station located on a road link of any class providing daily flow that will be factored by the corresponding scaling factor ($\dot{F}_{D,M,i}$) of the group it belongs to, to give the Average Annual Daily Traffic (AADT). The group scaling factor is defined as below:

$$F_{D,M,i} = \frac{1}{k_i} \sum_{K=1}^{k_i} \frac{DT_{K,D,M}}{AADT_K} \tag{1}$$

where $D =$ Monday, ...Sunday = 1, ...7
$M =$ January, February, ...December = 1, ...12
$K =$ Road link (for ATC core station only)
$k_i =$ Number of core stations in cluster i
$F_{D,M,i} =$ The group scaling factors for the D^{th} day of the week in the M^{th} month in cluster i
$DT_{K,D,M} =$ Observed daily flow for the D^{th} day of the week in the M^{th} month at the K^{th} road link
$AADT_K =$ Average Annual Daily Traffic at the K^{th} road link

There are 2 types of coverage stations: (1) "Coverage (B)" stations falling on cordons or screenlines; and (2) "Coverage (C)" stations not falling on cordons or screenlines. For example, Figure 1 shows the locations of the cordons or screenlines in Hong Kong Island together with the chosen core station 1003 for analysis in this paper.

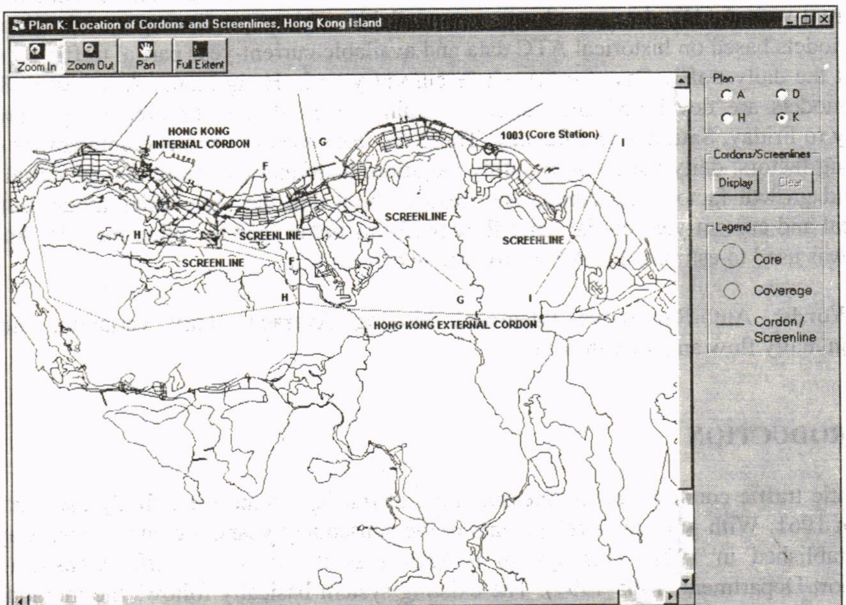

Figure 1 Locations of Cordons and Screenlines, and Core Station 1003 in Hong Kong Island

In the ATC report, AADT is one of the most useful information for the engineers/planners and is used to estimate the growth factors for the coverage stations not counted in the current year. In practice, for most of the on-going transport studies, there is always a need to use the most up-to-date traffic data such as AADT for transport model development and calibration. However, the current-year AADT data are always not available. Obviously, there is a need to update the AADT based on historical ATC data and available current-year partial traffic counts.

Hamed et al. (1995) applied the Box-Jenkins auto-regressive integrated moving average (ARIMA) model to analyze a set of 1-minute interval traffic volume in urban roads. The ARIMA (0,1,1) was found to be the most accurate model for short-term prediction of traffic volumes on urban roads. The model can easily be implemented, is computationally tractable, and only requires the storage of the last forecasted error and current traffic observation. Ahmed and Cook (1979) investigated the application of analysis techniques developed by Box and Jenkins to traffic volume and occupancy time series data on freeway. The ARIMA models were found to perform well in analyzing freeway time-series data, in terms of mean absolute error and mean square error, than moving-average, double exponential, and Trigg and Leach adaptive models.

In United Kingdom, Moorthy and Ratcliffe (1988) reported on applications of time series analysis to produce short-term forecasts using automatic traffic counts. The comparison of the observed and modeled traffic flows indicated that over 90% of the forecasts were made with less than 5.0% error. In USA, Nihan and Holmesland (1980) explored the use of time series techniques for short-term traffic volume forecasts. The results of their study showed that time series techniques can be used to provide highly reliable and inexpensive short-term forecasts. In another study, Benjamin (1986) presented a procedure for forecasting average daily traffic volume using a time series analysis.

This paper aims to develop time series models for the short-term prediction of current-year daily traffic flow at the ATC core station. The historical and current-year partial daily flows, daily variation, monthly variation and growth factor are the factors considered for development of the time series models. In total, 8 time series models are developed with different combination of the above factors. The results of the various models are compared with the real data for validation. Then, the selected time series model is used to estimate the AADT for the year 1999.

Section 2 describes the methodology for model development using regression analysis and Box-Jenkins approach. Section 3 presents the forecasted daily flows by the different time series models, while Section 4 shows the errors of estimated AADT together with discussions. Finally, conclusions and recommendations are given in Section 5.

2. METHODOLOGY

Lo et al. (2000) reviewed the current ATC approach in the station cluster analysis for the development of group scaling factors. New methodologies for the grouping of stations and development of growth factors were developed. Based on the new grouping of ATC stations, the core station that has the largest AADT in the selected group is chosen for analysis in this paper. The chosen core station is “1003” (see Figure 1). It is located at the Island Eastern Corridor (an expressway) from Healthy Street to Taikoo Shing in Hong Kong Island. This core station was opened in 1984. The years of 1984 to 1993 were excluded in data analysis due to some missing records.

The period of historical ATC data is chosen from January 1994 to June of 1999 for model development and estimation, and from July to December of 1999 ATC data for diagnostic checking and/or model validation. The traffic data collected at the core station provide traffic information on a monthly basis. In the ATC report, AADT is mainly classified by Monday to Friday (M-F), Saturday (Sat) and Sunday (Sun). On the basis of this classification, 3 individual data groups are used to develop the time series models. The plots of data are illustrated in Figure 2 for the daily flows by “M-F”, “Sat” and “Sun”.

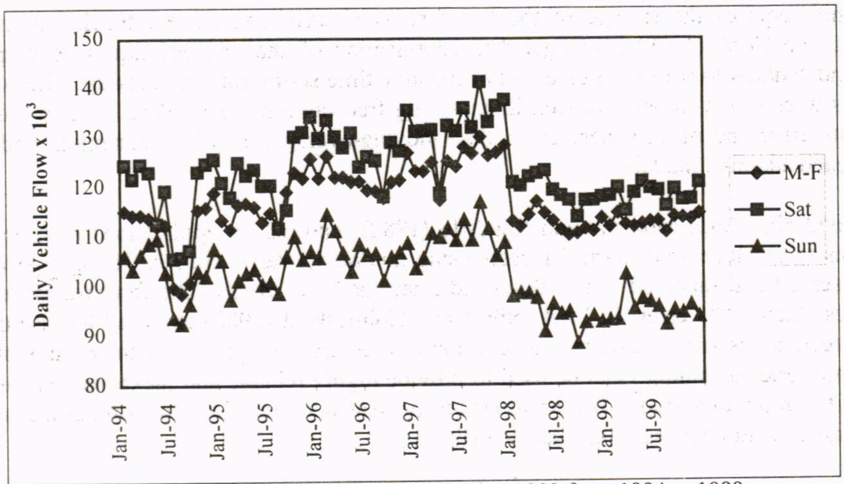

Figure 2 Observed Daily Flows on Station 1003 from 1994 to 1999

In this paper, we investigate various factors and their combination to develop the time series models for short-term prediction of daily flows at the selected ATC core station. The factors included for model developments are historical and current-year partial daily flow (DT), daily variation (DV), monthly variation (MV) and growth factor (GF). These factors are related to the traffic flow variations by day and month. The procedures of model development are illustrated by the flow chart in Figure 3. At the core station, for the D^{th} day of the week, the average daily flow ADT_D of the year can be estimated by:

$$ADT_D = \frac{1}{m}\sum_{M=1}^{m} DT_{D,M} \tag{2}$$

where $DT_{D,M}$ = Observed daily flow on the D^{th} day of the week in the M^{th} month
m = Number of weeks with observed data in the year = 12.

Daily variation (DV) factor is the proportion of average daily flow with respect to the overall average daily flow.

$$DV(D) = \frac{ADT_D}{\frac{1}{n}\sum_{M}\sum_{D} DT_{D,M}} \times 100\% \tag{3}$$

where n = Total number of Monday, ...Sunday selected = 84 (12 months by 7 weekdays).

Monthly Variation (MV) factor is the proportion of daily flow out of the average daily flow.

$$MV(W) = \frac{DT_{W,M}}{ADT_W} \times 100\% \tag{4}$$

where W = Monday to Friday, Saturday, Sunday = 1, 2, 3

Note that the daily flow for "M-F" is the mean of daily flows for Monday to Friday at the selected station.

Growth Factor (GF) is the growth rate on daily flow from the pervious month to the current month.

$$GF(W) = \frac{DT_{W,M-1} - DT_{W,M}}{DT_{W,M-1}} \tag{5}$$

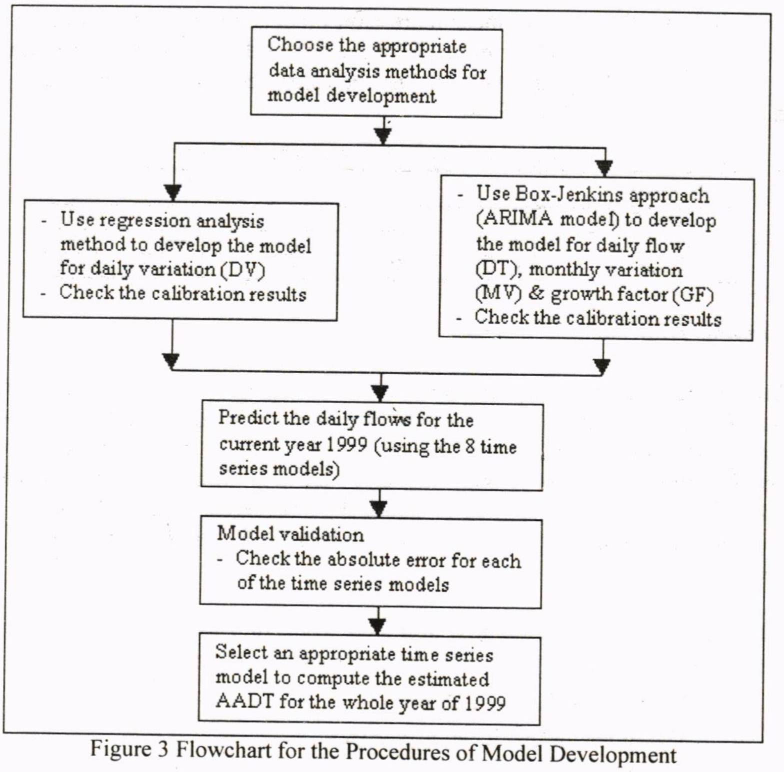

Figure 3 Flowchart for the Procedures of Model Development

2.1 Regression Analysis

In the ATC report, the "DV" is only provided as one point for a day of week in the whole year. Subsequently, the "DV" consists of 6 observation points (5 points used for specification and 1 point used for verification) from year 1994 to 1999 for a day of week. They are obtained by using Equation (3). However, the "DV" does not have enough observation points to develop the time series model by using the Box-Jenkins approach. Therefore, the regression analysis method is used to predict the "DV" for the year 1999. For the regression analysis, it can be used to describe and predict a dependent variable on the basic of one or more independent variables. The following linear, quadratic and cubic regression equations for "DV" are examined.

$$DV(W) = A + By \tag{6}$$

$$DV(W) = A + By + Cy^2 \tag{7}$$

$$DV(W) = A + By + Cy^2 + Dy^3 \tag{8}$$

where $DV(W) =$ The estimated value of the dependent variable, i.e. daily variation "DV" for the W^{th} day of week.

$y =$ The period of time ($1 \le y \le 5$, year 1994 is 1 and so on).

$A, B, C, D =$ The regression coefficients.

The results of the regression analysis are given in Table 1. Since the R Square for "Sat" is not large enough in the linear function, both the quadratic function and cubic function have been examined. The cubic regression model performed better than the other regression models for

"Sat". For "M-F" and "Sun", the linear regression models are adequate as the R Square of "M-F" and "Sun" are around 0.8 or more. With the use of these 3 models, the "DV" for "M-F", "Sat" and "Sun" in the year 1999 can then be estimated and used as a factor for short-term prediction of the daily traffic flow.

Table 1 Results of the Regression Analysis

Daily Variation	A	B	C	D	R^2
Monday to Friday	100.107	0.268	-	-	0.794
Saturday	105.971	0.225	0.969	-0.0896	0.996
Sunday	93.492	-1.564	-	-	0.874

2.2 Box-Jenkins Approach

Box-Jenkins model is one of the commonly used techniques that can be used for forecasting either discrete data or continuous data (Bowerman and O'Connell, 1987). However the data must be measured at equally spaced, discrete time intervals, for example: hourly, daily, weekly, or monthly. The Box-Jenkins model is used to forecast both non-seasonal and seasonal data, and can only be applied to stationary time series. If the time series were non-stationary, then it is required to convert the data to become stationary time series by using the differences or transformation.

The general model for non-stationary time series is Auto-Regressive Integrated Moving Average (ARIMA), $ARIMA(p,d,q)$, which can be expressed as:

$$(1-\phi_1 B-\ldots-\phi_p B^p)z_t=(1-\theta_1 B-\ldots-\theta_q B^q)a_t \tag{9}$$

where
- $p=$ The order for the differenced series.
- $q=$ The order for the white noise series.
- $d=$ The order of differencing.
- $\phi_i=$ An auto-regressive parameter, $1\le i\le p$.
- $\theta_j=$ A moving average parameter, $1\le j\le q$.
- $z_t=$ The original time series differenced and transformed to achieve stationary.
- $a_t=$ A series of errors (called white noise).
- $B=$ The backward shift operator, i.e. $z_{t-1}=Bz_t$.

Note that an $ARIMA(p,d,q)\times(P,D,Q)$ model, generally known as multiplicative model, is used to represent a seasonal series. P, D and Q have the same definitions in the seasonal model as p, d and q in the non-seasonal model, respectively.

The data series of "DT" and "MV" are ranged from January 1994 through June 1999 for calibration identification and estimation, and from July to December 1999 for diagnostic checking. The data series of "GF" are ranged from February 1994 through June 1999 for calibration and estimation, and from July to December 1999 for diagnostic checking. The models should give a good prediction in the short run, where the short run is referred to as the next six successive periods or next six months. Several models were examined including a multiplicative model with seasonality. Table 2 shows the time series parameter estimates. All parameter estimates were significantly different from zero at 0.05 level of significance.

Table 2 Summary Results of ARIMA Models

Series	Daily Flow	Monthly Variation	Growth Factor
Monday to Friday	(1,0,0), $\phi_1 = 0.7951$ t-test = 10.754	(1,0,0), $\phi_1 = 0.4637$ t-test = 4.213	(1,1,0), $\phi_1 = -0.5698$ t-test = -5.561
Saturday	(1,0,0), $\phi_1 = 0.6866$ t-test = 7.660	(1,0,0), $\phi_1 = 0.3912$ t-test = 3.414	(1,0,0), $\phi_1 = -0.2797$ t-test = -2.334
Sunday	(1,0,0), $\phi_1 = 0.7653$ t-test = 9.613	(1,0,0), $\phi_1 = 0.3506$ t-test = 3.002	(0,0,1), $\theta_1 = -0.3407$ t-test = 2.894

2.3 Time Series Models

In the pervious section, traffic data from January 1994 to June 1999 were used to develop the time series models. Now we use the traffic data from July to December 1999 to check the accuracy of the developed time series models. The models proposed for the diagnostic checking are as follows:

$$\overline{\overline{DT_M}} = DT_{M-1} \times DV_y \times MV_M \times (1 + GF_M) + \varepsilon_M \tag{10}$$

where
$\overline{\overline{DT_M}}$ = The daily flow at month M estimated by the following 4 factors.
DT_{M-1} = The observed daily flow at month $M-1$.
DV_y = The daily variation at year y estimated by time series models.
MV_M = The monthly variation at month M estimated by time series models.
GF_M = The growth factor at month M estimated by time series models.
ε_M = The error term at month M.

In Equation (10), the pervious month ($M-1$) observed daily flow is used to forecast the current month (M) daily flow. And this equation is aimed for checking the model with the captioned 4 factors but excluding "DT". In addition, Equation (10) is also used to assess the effects of each of these 4 factors (or their combinations) on the dependent variable (i.e. $\overline{\overline{DT_M}}$). In the following Equation (11), the daily flow ($\overline{DT_M}$) at month M estimated by time series analysis is used to predict the daily flow at month M.

$$\overline{\overline{DT_M}} = \overline{DT_M} \times DV_y \times MV_M \times (1 + GF_M) + \varepsilon_M \tag{11}$$

where $\overline{DT_M}$ = The daily flow at month M estimated by time series models.

All the combinations of the four factors for model development are listed in the Table 3. As the performance of the time series "DV" model is not very satisfactory in terms of the large percentage of error, the "DV" factor will not be considered to combine with the other factors for development of the time series models. However, the validation results of the time series "DV" model are still shown in the following for comparison with the selected time series models. In total, there are the 8 time series models chosen for validation tests (including the "DV" model).

Table 3 Combinations of Factors for Consideration in the Time Series Models

Number of Factors	1	2	3	4
Combinations of the Factors	*DT** *DV** *MV** *GF**	DT×DV, *DT×MV** DV×MV, *DT×GF** DV×GF, *MV×GF**	DT×DV×MV DT×DV×GF *DT×MV×GF** DV×MV×GF	DT×DV×MV×GF

"*" = Selected time series models for validation

3. MODEL VALIDATION

The validation results of the 8 time series models are presented in the following. Comparisons of the absolute error are made for "M-F" series and are summarized in Tables 4 and 5. Apart from the "DV" model, it can be seen in Table 4 that all the daily flow forecasts were made with less than ± 5% error. However, the errors are similar in the remaining 7 models. The performance of these 7 time series models should further be assessed by the following mean absolute error (MAE) and mean square error (MSE):

$$MAE = \frac{\sum_{M}\left|DT_{M}-\overline{\overline{DT_{M}}}\right|}{N} \tag{12}$$

$$MSE = \frac{\sum_{M}\left(DT_{M}-\overline{\overline{DT_{M}}}\right)^{2}}{N} \tag{13}$$

where N = Number of observations, 70.
M = Month from March 1994 to December 1999.

Note that the MAE and MSE results of "DV" model were also shown in Table 5 for comparison. On the other hand, the "GF" time series model cannot reproduce the estimated daily flow in January and February of 1994. Therefore, the data sets for calculations of MAE and MSE of the time series models are ranged from March 1994 to December 1999 so as to provide consistent comparison.

Table 5 shows that the MAE and MSE of the 8 time series models for "M-F". It can be seen that the smallest values of MAE and MSE of the "DV" model are ***2,716*** and ***14.848×10⁶*** respectively. Therefore, the time series model using "DT" factor as shown in Equation (14) is recommended for short-term prediction of daily flows from Monday to Friday. The comparison of the observed and forecasted daily flows for "M-F" is illustrated in Figure 4. Comparison results of the time series models for "Sat" and "Sun" are similar to the results of "M-F".

$$\overline{\overline{DT_{M}}} = \overline{DT_{M}} + \varepsilon_{M} \tag{14}$$

Table 4 Diagnostic Checking of the 8 Time Series Models for Monday to Friday

Month (1999)	Actual Daily Flow	DV Forecast	DV Error (%)	MV Forecast	MV Error (%)	GF Forecast	GF Error (%)	MV×GF Forecast	MV×GF Error (%)
Jul	113,052	114,853	-1.59	112,679	0.33	112,403	0.57	112,167	0.78
Aug	110,743	116,822	-5.49	112,674	-1.74	112,399	-1.49	112,158	-1.28
Sep	113,866	118,826	-4.36	112,688	1.03	112,401	1.29	112,174	1.49
Oct	113,705	120,864	-6.30	112,711	0.87	112,400	1.15	112,195	1.33
Nov	113,380	122,936	-8.43	112,737	0.57	112,401	0.86	112,223	1.02
Dec	114,537	125,045	-9.17	112,766	1.55	112,401	1.86	112,252	2.00

Month (1999)	Actual Daily Flow	DT Forecast	DT Error (%)	MV×DT Forecast	MV×DT Error (%)	DT×GF Forecast	DT×GF Error (%)	MV×DT×GF Forecast	MV×DT×GF Error (%)
Jul	113,052	113,702	-0.58	113,055	0.00	112,779	0.24	112,542	0.45
Aug	110,743	114,327	-3.24	113,697	-2.67	113,181	-2.20	113,176	-2.20
Sep	113,866	114,824	-0.84	114,341	-0.42	113,806	0.05	113,820	0.04
Oct	113,705	115,219	-1.33	114,847	-1.00	114,299	-0.52	114,322	-0.54
Nov	113,380	115,533	-1.90	115,246	-1.65	114,693	-1.16	114,720	-1.18
Dec	114537	115,782	-1.09	115,562	-0.89	115,006	-0.41	115,035	-0.43

Table 5 Absolute Errors of Various Models on Forecasted Daily Flows, for Monday to Friday

Month (1999)	DV	MV	DT	GF	MV×DT	MV×GF	DT×GF	MV×DT×GF
Jul	1,801	373	650	649	3	885	273	510
Aug	6,079	1,930	3,583	1,655	2,954	1,414	2,438	2,433
Sep	4,960	1,178	958	1,465	475	1,692	60	46
Oct	7,158	995	1,513	1,305	1,142	1,510	594	617
Nov	9,557	642	2,153	978	1,866	1,156	1,314	1,341
Dec	10,508	1,771	1,245	2,136	1,025	2,285	468	498
MAE	3,424	3,262	***2,716***	3,342	3,873	3,738	4,370	4,786
MSE×10^6	22.705	22.838	***14.848***	22.682	33.905	28.493	41.577	50.688

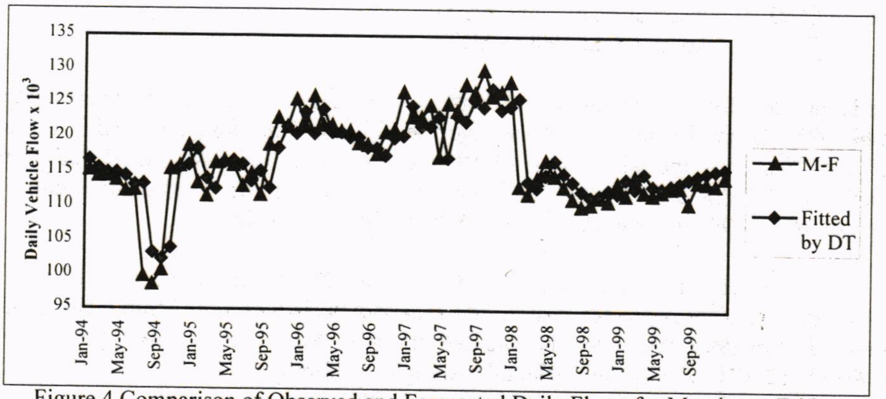

Figure 4 Comparison of Observed and Forecasted Daily Flows for Monday to Friday

4. COMPARISON OF THE ERRORS OF ESTIMATED 1999 AADT

For the model application, traffic flow data from January 1994 to June 1999 were used to develop the "DT" time series model. In this section, we use the recommended time series model to forecast the unknown daily flows in various months of the year 1999. The time series model should be updated with the available daily flows between January and December in 1999.

In other words, we would continuously update the model on month-by-month basis and then calculate the estimated AADT for the whole year of 1999. As a result, the "M-F", "Sat" and "Sun" models would provide 12 estimates for 1999 AADT respectively. The equations for calculation of 1999 AADT are shown as follows:

$$AADT(W) = \frac{1}{n_W}\sum_{M=1}^{12} n_{W,M} DT_{W,M} \tag{15}$$

where $W =$ Monday to Friday, Saturday, Sunday = 1, 2, 3

$n_{W,M} =$ Total number of W^{th} days in the M^{th} month.

$n_W =$ Total number of W^{th} day in the year.

Then the estimated AADT for "M-F", "Sat" and "Sun" are used to calculate the estimated AADT for the whole year of 1999.

$$AADT(All) = \frac{1}{n_{All}}\left(\sum_{W=1}^{3} n_W ADT_W\right) \tag{16}$$

where $n_{All} =$ Total number of day in the year, 365.

The estimation performance is measured by the following absolute error (AE) and absolute percentage error (APE):

$$AE = \left| AADT - \overline{AADT} \right| \tag{17}$$

$$APE = \frac{AE}{AADT} \times 100\% \tag{18}$$

where $AADT =$ Observation of 1999 AADT.
$\overline{AADT} =$ Estimation of 1999 AADT.

The comparison of estimated AADT for "M-F", "Sat" and "Sun" by the different number of months with observed data in 1999 are given in Table 6. The observed AADT of "M-F", "Sat" and "Sun" for the whole year of 1999 are ***112,910***, ***118,380*** and ***95,080*** vehicles, respectively. All APEs of "M-F" are less than 3%. When the number of months with observed data in 1999 is greater than 7, the APE of "M-F" are below 1%.

The APEs of "Sat" and "Sun" are around or below 2% as the number of months with observed data in 1999 is greater than 7. The largest APE is 6.18% on "Sun" with 3 months observed data in 1999. It is noted that the daily flow of "Sun" is extremely high in March of 1999 (see Figure 2), which would affect the accuracy of forecast.

Table 6 Summary Results of Estimation for AADT

No. of Month(s) with Observed Data in 1999	Monday to Friday			Saturday			Sunday		
	$\overline{AADT}$	AE	AE%	$\overline{AADT}$	AE	AE%	$\overline{AADT}$	AE	AE%
0	116,154	3,234	2.86%	122,860	4,483	3.79%	100,086	5,006	5.26%
1	115,313	2,393	2.12%	122,499	4,122	3.48%	99,309	4,228	4.45%
2	115,889	2,969	2.63%	122,443	4,066	3.43%	98,555	3,474	3.65%
3	114,998	2,077	1.84%	121,275	2,898	2.45%	100,956	5,875	6.18%
4	114,556	1,636	1.45%	121,473	3,095	2.61%	99,048	3,968	4.17%
5	114,400	1,480	1.31%	121,608	3,231	2.73%	99,102	4,021	4.23%
6	114,291	1,370	1.21%	121,174	2,797	2.36%	98,642	3,561	3.75%
7	114,124	1,203	1.07%	120,874	2,497	2.11%	98,117	3,037	3.19%
8	113,438	517	0.46%	120,148	1,771	1.50%	97,003	1,923	2.02%
9	113,806	885	0.78%	120,341	1,964	1.66%	97,160	2,080	2.19%
10	113,701	780	0.69%	119,976	1,599	1.35%	96,878	1,797	1.89%
11	113,621	701	0.62%	119,867	1,489	1.26%	96,869	1,789	1.88%

Figure 5 shows that absolute percentage errors for "M-F", "Sat" and "Sun" are decreasing as the number of months with observed data in 1999 is increasing. The APE of "M-F" is smaller when it is compared with the APEs of "Sat" and "Sun" because the number of data point for "M-F" are more than that for "Sat" and "Sun".

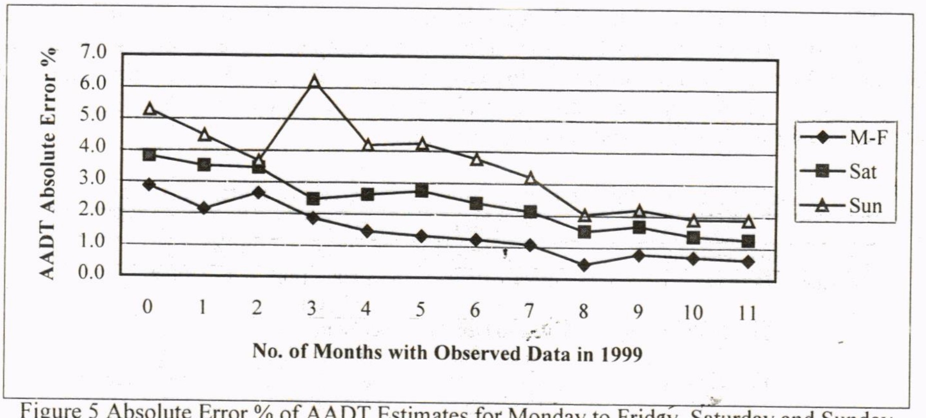

Figure 5 Absolute Error % of AADT Estimates for Monday to Friday, Saturday and Sunday

The estimated AADT for "M-F", "Sat" and "Sun" are used to calculate the estimated AADT for the whole year of 1999. Table 7 shows the comparison of AE and APE of the estimated 1999 AADT against various number of months with observed data in 1999. The observed AADT for all days in the whole year of 1999 is ***110,251*** vehicles. It was found that the APEs of the AADT estimates become stable (with APE about 1%) when the number of months with observed data in 1999 is greater than 7 (see Figure 6).

Table 7 Summary Results of Estimated AADT for the Year 1999

No. of Month(s) with Observed Data in 1999	$\overline{AADT}$	Absolute Error	Absolute Error %
0	113,980	3,730	3.38%
1	113,213	2,962	2.69%
2	113,456	3,205	2.91%
3	113,151	2,901	2.63%
4	112,515	2,265	2.05%
5	112,436	2,186	1.98%
6	112,219	1,968	1.79%
7	111,967	1,717	1.56%
8	111,195	944	0.86%
9	111,501	1,250	1.13%
10	111,329	1,078	0.98%
11	111,258	1,008	0.91%

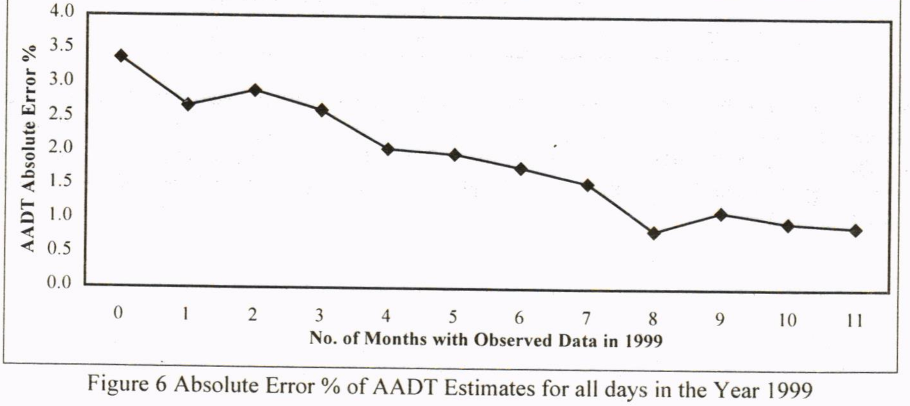

Figure 6 Absolute Error % of AADT Estimates for all days in the Year 1999

The traffic-forecast method with ARIMA model is very useful. This method is not only valid for Hong Kong (HK) but also for other areas. For example, Table 8 shows the relevant information for model comparison with the past-related studies in UK and USA. The results of our study are also compared with the results of UK (Moorthy and Ratcliffe, 1988) and USA (Nihan and Holmesland, 1980) in the following.

Table 8 Relevant Information for Model Comparison of Different Areas

		HK (2001)	UK (1988)	USA (1980)
Model Selection		Box-Jenkins Technique (ARIMA model)		
Period	Between	1994 – 1999 (6 years)	1978 – 1985 (8 years)	1968 – 1977 (10 years)
	For Calibration	Jan 94 to Dec 98 (60 months)	Jan 78 to Dec 84 (84 months)	Jan 68 to Dec 76 (108 months)
	For Validation	Jan to Dec 99 (12 months)	Jan to Dec 85 (12 months)	Jan to Dec 77 (12 months)
Seasonal Pattern		No	No	No
Time Series		Monthly average weekday (Mon-Fri) traffic		

Table 9 compares the estimation errors on weekday (Mon-Fri) traffic flows by month of the year for the above 3 study areas. All of the errors of HK (2001) are around 3% or less, except for the month of August. The errors for the month of December and June in UK (1988) are largest that are greater than 10%. On the other hand, the smallest errors occur for the month of September and October, which are close to zero. In USA (1980), the largest error occurs for the month of September. All of the other errors are around 5% or less.

Table 9 Estimation Errors (%) on Weekday (Mon-Fri) Traffic Flows by month of the year for the 3 Different Areas

	HK (2001)	UK (1988)	USA (1980)
Jan	2.23%	2.9%	4.66%
Feb	0.29%	3.7%	5.16%
Mar	2.76%	3.3%	3.22%
Apr	3.48%	0.7%	2.06%
May	3.28%	0.1%	4.00%
Jun	3.01%	11.5%	1.53%
Jul	3.04%	6.6%	-
Aug	**5.32%**	0.3%	-
Sep	2.53%	0%	**7.44%**
Oct	2.75%	0%	5.14%
Nov	3.11%	1.0%	5.70%
Dec	2.12%	**18.1%**	1.68%

Figure 7 illustrates the comparison of absolute percentage errors of weekday (Mon-Fri) traffic flows estimated for the 3 different areas. Because of the missing data on the month of July and August in USA (1980), the results of these 2 months cannot be shown. It can be seen in Figure 7 that the variations of the errors for HK (2001) and USA (1980) are more stable than that for UK (1988).

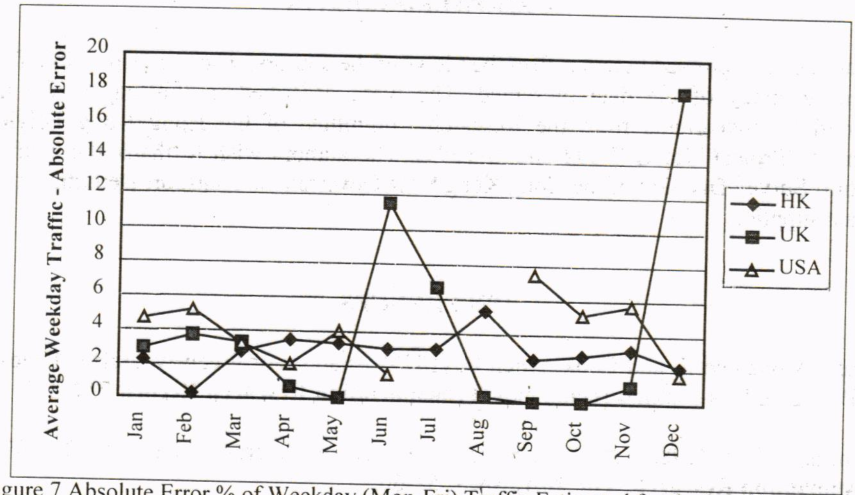

Figure 7 Absolute Error % of Weekday (Mon-Fri) Traffic Estimated for the 3 Different Areas

5. CONCLUSIONS AND RECOMMENDATIONS

In this paper, time series models were investigated for short-term prediction of current-year daily traffic flow at the ATC core station and forecast of the Annual Average Daily Traffic (AADT) in the current year. Although the forecast period for our case study was limited to a one-year forecast, the length of the forecast period can be extended to several years or shortened to several days in the near future, depending upon the time interval chosen. We will validate the developed models for the longer forecast period in the further study.

The data series were ranged from January through June 1999 for model calibration and estimation, and from July to December 1999 for diagnostic checking. Among the various combinations of the 4 factors (i.e., daily variation, monthly variation, historical and current-year partial daily flow and growth factor), 8 time series models were examined. As a result, the time series model using the historical and current-year partial daily flow was recommended for short-term prediction of daily flows from Monday to Sunday and subsequent estimation of the current-year AADT.

For most of on-going transport studies, there was always a need to use the most up-to-date traffic data such as AADT for transport model development and calibration. In the existing Annual Traffic Census report, the AADT of ATC core station was calculated on the basis of the last year (12-month) traffic flow data, i.e. these were collected at each month in the last year. In this paper, we recommended a time series model to forecast the unknown daily flows in the current year. As shown in the case study for the estimation of 1999 AADT, we have updated the recommended model with the available daily flow data between January and December in the year 1999. In other words, we continuously updated the model on month-by-month basis and then calculated the estimated AADT for the whole year of 1999. The absolute percentage error (APE) of the estimates become stable (with APE about 1%) when the number of months with observed data in 1999 was greater than 7.

ACKNOWLEDGMENTS

The comments expressed herein reflect the views of the authors, who are responsible for facts and the accuracy of the data presented. The work described in this paper was mainly supported by two grants from the Research Committee of the Hong Kong Polytechnic University (Project No. G-W002 and G-T299). The authors wish to thank the Traffic and Transport Survey Division of the Hong Kong SAR Government Transport Department for the raw data supplied.

REFERENCES

Ahumed, Mohamed S. and Cook, Allen R. (1979) Analysis of Freeway Traffic Time-Series Data By Using Box-Jenkins Techniques. **Transportation Research Record 722**, 1-9.

Benjamin, Julian (1986) A Time-Series Forecast of Average Daily Traffic Volume. **Transportation Research, Vol. 20A, No. 1**, 51-60.

Bowerman, B.L. and O'Connell, R.T. (1987) **Time Series Forecasting – Unified Concepts and Computer Implementation**, Duxbury Press, Boston

Chan, Y.M., Lee, B.S., Wong, C.H., Kwok, W.S. and Kwan, W.K. (2000) Traffic Census – Today's Practice and Tomorrow's Challenges. **Proceedings of the Fifth Conference of Hong Kong Society for Transportation Studies**, 2 December, Hong Kong, 186-194.

Hamed, Mohammad M., Al-Masaeid, Hashem R. and Bani Said, Zahi M. (1995) Short Term Prediction of Traffic Volume in Urban Arterial. **Journal of Transportation Engineering, Vol.121, No. 3**, 249-254.

Lo, H.P., Cheung, K.M., Tong, C.O. and Wong, S.C. (2000) Use of Cluster Analysis in the Annual Traffic Census. **Proceedings of the Fifth Conference of Hong Kong Society for Transportation Studies**, 2 December, Hong Kong, 442-450.

Moorthy, C. K. and Ratcliffe, B. G. (1988) Short Term Traffic Forecasting Using Time Series Methods. **Transportation Planning and Technology, Vol. 12**, 45-56.

Nihan, Nancy L. and Holmesland, Kjell O. (1980) Use of the Box and Jenkins Time Series Technique in Traffic Forecasting. **Transportation, Vol. 9, No. 2**, 125-143.

Transport Department (1965-1999) **The Annual Traffic Census**. Traffic and Transport Survey Division, Transport Department, Government of the Hong Kong SAR.

Transport Department (1988) **The Annual Traffic Census (New Methodology) in Hong Kong Island**. Technical Report No. 334, Traffic and Transport Survey Division, Transport Department, Hong Kong Government.