EXTREME VALUE CONTROL AND AN ADDITIVE SEASONAL MOVING AVERAGE MODEL FOR THE EVALUATION OF DAILY PEAK HOUR TRAFFIC DATA

P. LEHTINEN
Helsinki Telephone Company, Helsinki, Finland

ABSTRACT

This paper describes the design of a traffic data evaluation program based on daily peak hour traffic values on even hour basis and extreme value statistics. The procedure is being implemented in the automatic traffic recording system AUTRAX in the Helsinki Telephone Company.

The moving average of ten daily peak hour traffic values can be used as a remarkably close approximation to the time-consistent busy hour traffic volume. Based on the extreme value theory a control limit is calculated for daily peak hours relative to the level and variance of the series of peak hour traffic values. Information about the weekly variation within a year is furnished concurrently with the moving average by estimating additive weekly seasonal effects. Continuous control can then be applied to trigger a trend alarm in the case of significant increase of the traffic level. Associated with a trend alarm a linear weekly forecast function is determined based on the yearly average level of traffic presently and at the time of the latest previous trend alarm. This makes it possible to calculate the grace period of weekly averages being below the capacity limit for producing efficient reports of immediate interest to the supervisors.

1. INTRODUCTION

Traffic load measurements constitute a basis for the design and expansion of telephone exchanges. Automatic centralized traffic recording system AUTRAX has been installed in the Helsinki Telephone Company in Finland for maintenance and traffic measurements in electromechanical switching offices. It provides a large amount of data and attention has been focused on the design of effective evaluation programs to produce reports of immediate interest for the supervisors.

The establishment of the traffic intensity objective is of real importance. Traditionally the notion of time consistent busy hour has been used to define the loading of a switching equipment group according to the average day method recommended by CCITT. However, the pattern of the hourly variation of traffic varies from day to day and depends on the type of day. Individual customer is presumably more concerned with daily peak hours if he is significantly delayed or repeatedly blocked at these times.

In recent years New Jersey Bell has made an approach to use daily peak hour traffic values based on extreme value statistics in the formulation of service level criteria /4, 1, 7/. The procedure requires a continuous measurement of traffic flow but only the highest hourly reading during the day is recorded. On the other hand, it eliminates some problems that occasionally arise with the time-consistent busy hour approach. This paper explains how the extreme value method can be used for continuous control giving an alarm in the case of significant increase of traffic level and an estimate of the grace period of average peak hour traffic value being below the capacity limit.

2. AN OVERVIEW OF THE AUTOMATIC TRAFFIC RECORDING SYSTEM AUTRAX

The automatic traffic recording system AUTRAX consists of data acquisition terminals connected to switching equipment, a central processing computer, remote report printers and data facilities which connect these system elements together (Fig. 1).

Fig. 1 The configuration of the AUTRAX traffic recording system
The Centralized Data Control (CDC) computer is connected to the terminals via dedicated multi-party telephone lines and data sets. Operating simultaneously each of the eight input ports can address up to 32 interface modules for serial polling.

The CDC performs the following operations:

1. Automatic polling of the terminals
2. Validation check of the ICUP data
3. Processing of the data
4. Generation of Quick-Look reports
5. Distribution of Quick-Look reports to remote printers
6. Printing of alarm messages on the high speed printers
7. Development and testing of the programs and the data bases
8. Processing of the real time operating system AUTRAX for the execution of action entries in updating the polling time schedule, handling the magnetic tape unit, modifying Quick-Look parameters, etc.

The system has been widely used to produce Quick-Look reports for maintenance purposes. The average holding time is a good indicator of the performance of switching equipment since deviations to either side very likely indicate misfunctioning.

Basically ICUP data is a discrete integral of the traffic process. At the CDC each terminal has its individual data base consisting of a list of the corresponding input points and defining the circuit grouping for the evaluation of the data for traffic reports. On the disc there are accumulator blocks for storing current and past traffic values during online operation.

An automatic program of eight different daily polling schedules can be applied for each terminal. The length of polling intervals can be defined on a quarter hour basis varying from 15 minutes to 24 hours. When polling on an even hour basis the capacity performance of the system is remarkably better than with 15 minute intervals. Consequently the use of hourly traffic values has been of prime interest in the acquisition and analysis of traffic data.

3. THE OUTLINES OF THE TRAFFIC MEASUREMENT AND CONTROL PROCEDURE

The basic procedure is described by the flowchart in Fig. 2 including the following operations:

1. Polling traffic data from terminals on even hour basis and calculating hourly traffic intensity of each trunk group to determine daily peak hour traffic volume on even hour basis.
2. Storing the daily peak hour traffic volumes of ten consecutive days to calculate and store the corresponding moving average as a fairly good estimate of the time-consistent busy hour traffic volume.
3. Saving the value of the highest moving average that has occurred since the beginning of the current measurement season.
4. Determination of average daily variation within a week based on the two latest weeks.
5. At the end of each week the average value of five latest daily peak hour traffic volumes is stored and average weekly variation within a one year period is calculated.
6. Updating the time indexes of the latest K = 4 exceedances of the control limit after making a check on the occurrence of exceedance.

A conclusion of significant trend is made if the number of exceedances is equal to the threshold count \( K = 4 \) within the latest \( Q = 20 \) days. An alarm message is given and a readjusted control limit is calculated. In addition, a weekly forecast function is calculated to describe linear trend based on the yearly average level of traffic at the two latest alarm points.

The time index and traffic level of the latest alarm is saved to be used in calculating linear trend at the next alarm time.

An alarm message is printed and optionally any other information that is available.

Supplementary elements can be added to the general scheme above. For example, rejection test statistics must be applied to eliminate outliers due to special circumstances such as unsuccessful poll, system errors or exceptional days. In some applications hourly variation may be of special interest, too.

---

Fig. 2 Flowchart of the daily peak hour traffic data evaluation program
4. OPERATIONAL PARAMETERS

The procedure can be modified by calibrating the following parameters in a desired way:

<table>
<thead>
<tr>
<th>Description of parameters</th>
<th>Symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of week (wave length of daily observations)</td>
<td>w</td>
</tr>
<tr>
<td>Length of moving average of daily peak hours</td>
<td>m</td>
</tr>
<tr>
<td>Number of days in calculating averaged daily deviations</td>
<td>k</td>
</tr>
<tr>
<td>Length of the series of moving averages to be saved</td>
<td>h</td>
</tr>
<tr>
<td>Length of year (wave length of weekly traffic values)</td>
<td>y</td>
</tr>
<tr>
<td>Length of moving average of weekly traffic values</td>
<td>n</td>
</tr>
<tr>
<td>Number of weeks in calculating averaged weekly deviations</td>
<td>k</td>
</tr>
<tr>
<td>Langth of the series of moving averages of weekly traffic values to be saved</td>
<td>h</td>
</tr>
<tr>
<td>Exceedance count threshold</td>
<td>K</td>
</tr>
<tr>
<td>Number of consecutive days included in the exceedance count</td>
<td>Q</td>
</tr>
<tr>
<td>Capacity limit</td>
<td>R</td>
</tr>
</tbody>
</table>

Dependent upon these parameters further calculations are done as follows:

1. Determine the number of memory locations needed for storing daily observations \(d_n, d_{n-1}, \ldots, d_1\) and weekly averages \(w_n, w_{n-1}, \ldots, w_1\) for calculating moving averages, averaged deviations and one period long moving averages of averaged deviations:

\[
N = \max \{ m, k, 2y - 1 \},
\]

\[
M = \max \{ w, k, 2w - 1 \}.
\]

2. Determine the center indexes of one week and one year periods to be used in calculating the daily and weekly averages.

\[
t = \lceil \frac{w}{2} \rceil + 1 \quad \text{and} \quad \delta = \lceil \frac{y}{2} \rceil + 1.
\]

At the start of the procedure all of the data storage registers are assumed to be zero. After an initial period the extreme value control and the additive seasonal moving average model will be automatically adapted to the traffic process.

5. MOVING AVERAGE MODEL AND EXTREME VALUE LIMIT FOR CONTINUOUS CONTROL

The daily peak hour traffic values constitute a stochastic process and the moving average model provides a simple method for measuring its level. It can be called a filter when used to remove constituent elements of high frequency (short wavelength) and leave untouched those of low frequency (long wavelength) /6/. This is achieved by including \(m = 5, 10, 15\) or 20 daily observations in the moving average. In this study \(m = 10\) has been chosen:

\[
d_1 = \left( d_1 + d_{1-1} + \ldots + d_{1} \right) / \mu.
\]

This definition of traffic intensity is a remarkably close approximation to the time-consistent busy hour traffic volume.

Despite of periodic or irregular fluctuation the occurrence of significant trend of the process can be indicated by frequent exceedances of observations over an extreme value limit that is calculated with respect to the level and variance of the process. In the formulation of an adaptive model these can be defined by means of the moving average value and the variance \(s^2\) of daily peak hour observations included in it. In the case of a trend alarm a new control limit \(r\) is calculated based on the latest moving average \(d_1\) and the respective variance \(s^2\):

\[
\begin{align*}
\hat{d}_1 &= \frac{1}{\mu} d_1, \\
\hat{s}^2 &= \frac{1}{\mu - 1} \left( d_1 - \hat{d}_1 \right)^2.
\end{align*}
\]

E.J. Gumbel has shown /5/ that the distribution function for the largest value \(x\) in a sample of \(Q\) observations can be expressed asymptotically as

\[
G(x) = \exp \left[ -e^{-\beta(x-a)} \right]
\]

provided that the distribution of the underlying observations converges with increasing \(x\) towards unity at least as quickly as an exponential function and \(x\) is unlimited. As a matter of fact, the daily peak hour observations have such an exponential distribution being considered extreme values themselves. The location parameter \(a\) and the measure of dispersion \(\beta\) are independent of \(Q\) and they may be estimated by the method of moments as follows:

\[
\hat{a} = \frac{\bar{x}}{\sqrt{6}}
\]

\[
\hat{\beta} = \frac{\bar{x} - \chi}{\sqrt{6}} \quad (\gamma = 0, 57721 = \text{Euler's constant}).
\]

The control limit \(r\) is calculated as the characteristic largest value \(x_Q\) of the extreme value distribution \(G(x)\) as follows

\[
G(x_Q) = 1 - \frac{1}{\hat{Q}},
\]

\[
r = \hat{a} - \frac{1}{\hat{\beta}} \ln \left( -\ln \left( \frac{1}{\hat{Q}} \right) \right).
\]

In \(Q\) observations the expected number of values equal to or larger than \(r\) is unity.

The trend alarm will be based on frequent occurrence of exceedances. The number \(C\) of exceedances within \(Q\) observations is approximately Poisson distributed with parameter \(\theta = 1\).

\[
P(C = c) = \frac{1 e^{-\theta} \theta^c}{c!}.
\]

Accordingly \(P(C \geq 4) = 1 - P(C \leq 3) = 0, 02\) and the risk of uncertainty is 2 percent when asserting that four exceedances within the peak hour observations on \(Q\) consecutive days indicates significant trend. This can be visualized informally by observing that one exceedance per week on the average will trigger an alarm inasmuch as only one per roughly a month is considered normal.

![Graph](https://via.placeholder.com/150)

**Fig. 3** An example of exceedances over the extreme value control limit for daily peak hour traffic values.
Update time index of the Jth latest exceedance:
\[ \text{EXC}(J) := \text{EXC}(J) + 1 \]

No
\[ \text{J} := \text{K} \]

Yes
Exceedance?
\[ \text{NEWOBS} \geq r \]
\[ \text{J} := \text{K} \]

No
\[ \text{J} := 0 \]

Yes
Exceedance time indeces are shifted:
\[ \text{EXC}(	ext{J}+1) := \text{EXC}(	ext{J}) \]

No
\[ \text{J} := \text{K} \]

Yes
Time index of the current exceedance is stored: \[ \text{EXC}(1) := 1 \]

Alarm:
\[ \text{EXC}(\text{K}) \leq Q \]
\[ \text{EXCEED} := \text{TRUE} \]

End of \text{EXCEED}

Fig. 4 The flow chart of subroutine \text{EXCEED} for updating the time indeces \text{EXC(J)}, \( J = 1,2, \ldots, K \), of the latest \( K \) exceedances over the control limit \( r \) and making a check on the occurrence of present exceedance and consequent trend alarm.

6. \text{CALCULATION OF ADDITIVE SEASONAL EFFECTS CONCURRENTLY WITH THE MOVING AVERAGE}

Additive seasonal effects can be associated with the calculation of moving averages. The seasonal adjustment procedure is described in the following.

Assume that the daily time series \( d_t \) can be represented locally by
\[ d(t+u) = a(t) + b(t)u + \sum_{p=1}^{w} \delta P(t) \nu(t+u) + z(t+u), \]
where \( a(t) \), \( b(t) \) and \( \delta P(t) \) are parameters and \( w \) is the length of one week period,
\[ \nu(t) = \begin{cases} 1 & \text{if } d(t) \text{ is observed on day } p \\ 0 & \text{otherwise}, \end{cases} \]
and \( z(t) \) is a stationary time series with zero mean. In this representation \( a(t) + b(t)u \) is a trend component and \( \sum_{p=1}^{w} \delta P(t) \nu(t+u) \) is the seasonal component.

The sum of daily seasonal effects is zero over a one week long period of \( w = 5 \) observations and it follows that a moving average such as
\[ \tilde{a}(t) = \frac{1}{5} \{ d(t-2) + d(t-1) + d(t) + d(t+1) + d(t+2) \} \]
will not contain a seasonal component. Letting
\[ \psi_0 = \frac{4}{5}, \quad \psi_1 = \frac{1}{5}, \quad i = -2, -1, 0, 1, 2, \]
the differenced "daily deviation"
\[ \tau(t) = d(t) - \tilde{a}(t) = \frac{1}{5} \sum_{p=1}^{w} \delta P(t) \nu(t) \]
\[ + \frac{2}{5} \psi_1 z(t+i) \]
constructed from the original series contains the daily seasonal component but no trend.

Since the variation of daily peak hour traffic values within a week changes from month to month an estimate of the daily seasonal component for time \( t \) can be calculated as an averaged daily deviation on only a few weeks of data:
\[ \lambda(t) = \frac{1}{\kappa} \sum_{i=0}^{\kappa-1} \tau(t-iw) \]
with \( \kappa = 2 \), for example.

The averaged daily deviations computed in this way during a one week long period do not necessarily sum to zero. Therefore one may adjust them to achieve a zero sum by defining
\[ \tilde{\lambda} = \frac{1}{\kappa} \sum_{i=0}^{\kappa-1} \tau(t-iw) - \tilde{\lambda}, \quad \xi = 1, 2, \ldots, 5, \]
\[ \delta P(\xi)(t) = \lambda(t+1-\xi) - \tilde{\lambda}, \quad \xi = 1, 2, 3, 4, 5, \]
The adjusted averaged daily deviations \( \delta P(\xi)(t) \), \( \xi = 1, 2, 3, 4, 5 \), are associated with the five days from time \( t \) backwards.

The accomplishment of the additive seasonal moving average model of the daily peak hour series can be done by the following procedure. Index one has been assigned to the newest observation and the latest value in the result-
ant series \( d_1 \), \( t_1 \), \( \lambda_1 \) and the sequences of past values are designated by indices \( i \) in increasing order.

Step 0. Get new daily observation.

Step 1. Update the sequence of daily observations \( d_i \) by shifting the list and storing the new observation \( \text{NEWOBS} \) in \( d_1 \).

Step 2. Update the sequence of moving averages \( d_i' \) by shifting the list and setting
\[
d_i' = \frac{1}{w} \sum_{t=0}^{i-w} d_t.
\]

Step 3. Update the time index of the maximum moving average \( d_i' \) that has occurred since the beginning of the current measurement season and make a substitution if the present value \( d_i' \) supersedes the previous maximum.

Step 4. Calculate the one period moving average \( W = \frac{1}{w} \sum_{i=1}^{i-w} d_i' \).

Step 5. Update the sequence of daily deviations \( t_i \) by shifting the list and setting
\[
t_i = d_i - W
\]
where \( t_i \) indicates the center index \( \lceil \frac{w}{2} \rceil + 1 \) of the one period long moving average.

Step 6. Update the sequence of averaged daily deviations \( \lambda_i \) by shifting the list and setting
\[
\lambda_i = \frac{1}{w} \sum_{t=1}^{w} t_{i+(i-1)w}.
\]

Step 7. Calculate the one period long moving average of averaged daily deviations \( \bar{X} = \frac{1}{w} \sum_{i=1}^{i-w} \lambda_i \).

Step 8. Calculate the adjusted daily deviations \( \delta_i = \lambda_i - \bar{X} \).

The indices \( \xi \) of daily seasonal components above are not explicitly related to the type of day. It can be deduced that the index of \( \delta_i \) associated with the next coming day is \( \rho = \lfloor \frac{w+1}{2} \rfloor \).

\[
W(t+\omega) = A(t) + B(t)\omega + \sum_{P=1}^{\omega} \frac{A^P(t)\Omega^P(t+\omega)}{\omega} + Z(t+\omega)
\]
where \( A(t) \), \( B(t) \), \( A^P(t) \) are parameters and \( \omega \) is the length of one year period,

\[
\Omega^P(t) = \begin{cases} 1 & \text{if } W(t) \text{ is observed on week } \rho, \\ 0 & \text{otherwise}, \end{cases}
\]

and \( Z(t) \) is a stationary time series with zero mean.

For forecasting purposes it is more desirable to use weekly averages \( W_1 \) of daily peak hour traffic values and estimate additive seasonal effects within a one year long period. The procedure is analogous to the analysis of the daily series above assuming that the weekly series can be represented locally by

\[
\bar{X}(t) = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} T(t-jy) \text{ with } k = 2
\]
and modifying them to achieve zero sum as follows

\[
\bar{X} = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} A(\ell+1-\xi)
\]

\[
\delta^P(\ell+\xi) = A(\ell+1-\xi) - \bar{X}, \quad \xi = 1, 2, \ldots, 52.
\]
These adjusted averaged weekly deviations \( \delta^P(t+\xi) \), \( \xi = 1, 2, \ldots, 52 \), are associated with the 52 weeks from week 1 backwards.

The procedure of analyzing the sequence of weekly averages \( W_1 \) can be constructed similarly to that of daily observations. The most recent weekly average is denoted by \( W_1 \) and the resultant latest values in the concomitant series by \( W_1', T_1 \), and \( A_1 \). Past values of these sequences are designated by indices \( j \) in increasing order.

Step 0. Get new weekly average of daily peak hour values.

Step 1. Update the sequence of weekly observations \( W_j \) by shifting the list and storing the new observation in \( W_1 \).

Step 2. and Step 3. are reserved for supplementary use.

Step 4. Calculate the one period moving average \( Y = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} W_j \).

Step 5. Update the sequence of weekly deviations \( T_j \) by shifting the list and setting \( T_1 = W_1 - Y \) where \( k \) indicates the center index \( \lceil \frac{\omega}{2} \rceil + 1 \) of the one period long moving average.

Step 6. Update the sequence of averaged weekly deviations \( A_j \) by shifting the list and setting
\[
A_j = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} T(t+1-jy)
\]

The procedure is analogous to the analysis of the weekly series above assuming that the weekly series can be represented locally by

\[
\bar{X}(t) = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} T(t-jy) \text{ with } k = 2
\]

\[
\delta^P(\ell+\xi) = A(\ell+1-\xi) - \bar{X}, \quad \xi = 1, 2, \ldots, 52.
\]

The procedure of analyzing the sequence of weekly averages \( W_1 \) can be constructed similarly to that of daily observations. The most recent weekly average is denoted by \( W_1 \) and the resultant latest values in the concomitant series by \( W_1', T_1 \), and \( A_1 \). Past values of these sequences are designated by indices \( j \) in increasing order.

Step 0. Get new weekly average of daily peak hour values.

Step 1. Update the sequence of weekly observations \( W_j \) by shifting the list and storing the new observation in \( W_1 \).

Step 2. and Step 3. are reserved for supplementary use.

Step 4. Calculate the one period moving average \( Y = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} W_j \).

Step 5. Update the sequence of weekly deviations \( T_j \) by shifting the list and setting \( T_1 = W_1 - Y \) where \( k \) indicates the center index \( \lceil \frac{\omega}{2} \rceil + 1 \) of the one period long moving average.

Step 6. Update the sequence of averaged weekly deviations \( A_j \) by shifting the list and setting
\[
A_j = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} T(t+1-jy)
\]

The procedure is analogous to the analysis of the weekly series above assuming that the weekly series can be represented locally by

\[
\bar{X}(t) = \frac{1}{\omega} \sum_{j=-\omega}^{\omega} T(t-jy) \text{ with } k = 2
\]

\[
\delta^P(\ell+\xi) = A(\ell+1-\xi) - \bar{X}, \quad \xi = 1, 2, \ldots, 52.
\]
Step 7. Calculate the one period long moving average of averaged weekly deviations

$$\bar{A}_t = \frac{1}{y} \sum_{j=1}^{y} A_j.$$  

Step 8. Calculate the adjusted weekly deviations

$$A'_t = A_t - \bar{A}_t, \quad \zeta = y/y' - 1, \ldots, 1,$$

which add up to zero.

The index of $\zeta$, associated with the next entire week is $\sigma = \{y' - 1\}$. This method of indexing the weekly seasonal components is appropriate for adjusting a linear forecast of weekly averages with regard to the variation within a year.

7. LINEAR TRAFFIC FORECAST ASSOCIATED WITH THE TRENDS CONTROL

Attention has been recently focused on the methods that produce local traffic forecasts from past data. Straightforward extrapolation can be misleading unless variations with the season of the year, unusual peaks and depressions, missing records, etc. are taken into account. Associated with the trend control a simple linear forecasting technique has been developed, which will make allowances for unusual variations including those due to season without being sensitive to deficient data.

The essence of the forecasting technique is the following. In the case of significant trend indicated by the extreme value method a linear weekly forecast function is determined based on the yearly average level of traffic presently and at the time of the latest previous trend alarm. As a result the real time online traffic recording system can automatically on a trunk group basis print information about the grace period as a theoretical busy hour since the hourly variation of telephone traffic varies from day to day and depends heavily on the type of community.

8. AN INTRODUCTORY ANALYSIS OF THE COHERENCE BETWEEN THE TIME CONSISTENT BUSY HOUR TRAFFIC VOLUME AND THE AVERAGE DAILY PEAK HOUR

The prime objective of traffic engineering is customer service applying to all hours of the day. The time consistent busy hour determined by the average day method has been compared with daily peak hours in Fig. 7. There is no such period as a theoretical busy hour since the hourly variation of telephone traffic varies from day to day and depends heavily on the type of community.

In a real time traffic recording system it is hardly possible to develop a transfer function between the series of time consistent busy hour traffic volumes and average daily peak hour values on a trunk group basis. This is why an analysis of linear correlation of those two sets of traffic values has been made for arbitrary aggregations of the trunk groups in an office (Fig. 8).

The positive correlation is extremely high since the two sets of values are almost identical. The only inconsistency is due to the variability in the pattern of hourly variation on successive days. The regression line can be considered an elementary transfer function visualizing the day-repeatability. It can be concluded that generally the average daily peak hour traffic value is slightly larger than the time consistent busy hour traffic volume. The difference is typically one or two Erlang units. Further studies are needed to determine the coherence in more detail.
In an automatic traffic measurement system the moving average of daily peak hour traffic values can be used as a close approximation to the time-consistent busy hour traffic volume. Consequently the system can automatically calculate an extreme value limit for the daily peak hours relative to the level and variance of the series of peak hour traffic values. This feature is good for continuous overload control. In addition, the weekly variation can be estimated concurrently with the moving average. As a result, continuous control can be applied to trigger a trend alarm in the case of significant increase of the traffic level giving a message of the grace period of weekly averages being below the capacity. All the above elements constitute a remarkably simple model based on real traffic and very few theoretical assumptions.

REFERENCES


