■ 論 文 ■

Determining Optimal Aggregation Interval Size for Travel Time Estimation and Forecasting with Statistical Models

통행시간 산정 및 예측을 위한 최적 집계시간간격 결정에 관한 연구

PARK, Dongjoo
(Assistant Professor, School of Civil Engineering, Asian Institute of Technology)

목 차

요 약

실시간 통행시간관련자료의 집계시간간격은 보다 신뢰성있는 통행시간정보제공과 교통정보센터의 효율적인 운영을 위해 매우 중요한 요소이다. 그러나 대부분의 기존 VDS 및 TCS교통정보 데이터는 통계학적·공학적 차원에서의 합리적인 연구나 검증없이 경험적 간격으로 집계되고 있다. 본 연구의 목적은 링크 및 교통축(Corridor) 통행시간 산정 및 예측시의 최적 집계 시간간격을 결정할 수 있는 통계학적 모형을 개발하고 실제 도로망에서 수집되는 통행시간자료에 적용하는 것이다. 첫째로, 본 연구는 링크 및 교통축 통행시간 산정 및 예측으로 인한 오차를 계량화하는 통계학적 모형을 제시하고, 제시된 모형의 의미를 교통류이론 측면과 통행시간정보 이용자측면에서 살펴보았다. 둘째로, 미국 Texas, Houston의 도시고속도로에서 AVI시스템을 통해 수집된 통행시간자료를 제시된 모형에 적용하였다.

적용결과 링크통행시간 산정을 위한 최적 집계시간간격보다 링크통행시간예측을 위한 최적 집계시간간격이 큰 것으로 나타났으며, 교통축 통행시간 산정 및 예측을 위한 최적 집계시간간격은 교통축을 구성하는 링크간의 상관관계(Correlation)에 큰 영향을 받는 것으로 분석되었다.

Ⅰ. Introduction

Recent advances in ITS deployment have resulted in many Traffic Management Centers having access to large amount of real-time data. This data is often used to estimate and/or forecast link and route travel times or other traffic parameters such as *volume and occupancy*. Estimation and forecasting techniques for travel time (Boyce et al. 1993; Dailey, 1993; Tarko and Rouphail, 1993; Van Arem et al. 1997; Park and Rilett, 1998/1999, Park et al. 1999; Rilett and Park, 2000) and volume (Okutani and Stephanedes, 1984; Davis and Nihan, 1991; Van Der Voort et al. 1996) have been studied extensively. In general, these approaches aggregate the raw data into intervals of set duration and then use the aggregated information as input to the estimation and/or forecasting models. The focus of this paper is on providing a statistically based approach for identifying the optimal aggregation size for use within these estimation and forecasting techniques.

There have been a number of studies which identify the required number of probe vehicles to obtain statistically reliable link travel time estimates (i.e. link travel time at some point in the past) when the aggregation interval size is defined a priori (Turner and Holdner, 1995; Srinivasan and Jovanis, 1996; Sen et al. 1997). Gajewski, et al (2000) use a cross-validated mean squared error approach to chose optimal aggregation widths in estimating speed data from loop detectors for data archiving purposes. However, these studies rely on a number of assumptions that may be problematic. The first is that they have focused on estimating travel times and therefore the results may not be optimal for travel time forecasting. Secondly, they have focused on link travel times. It has been hypothesized that route or corridor travel times are more important than link travel times for travelers(Park, 1998; Rilett et al. 1999). Because link travel times are correlated, it is hypothesized that the optimal aggregation sizes for the "link" travel time estimation and forecasting may not be the same as those for the "corridor/route" travel time estimation and forecasting. Lastly, and most importantly, the techniques that have been developed have tended to use the "variance of the sample mean" as the sole criterion for identifying the aggregation size without any consideration of "bias", which is defined in this paper as the difference between the individual travel times and the observed sample mean travel time.

The objective of this paper is to propose a methodology for identifying the optimal aggregation interval size for estimating and forecasting traffic parameters. The traffic parameters of interest are link travel time and corridor travel time although the methodology may be generalized to other parameters such as link volume. The proposed approaches will explicitly account for traffic dynamics, frequency of observations, and spatial and temporal dependence in data. The main criteria for identifying the best aggregation size is the Mean Square Error (MSE). In this paper the MSE is treated as being comprised of two main components — *the bias and precision* — and the focus of the paper is on the trade-offs between these two components. The first part of the paper develops aggregation size models for four cases:

- Link travel time estimation
- Link travel time forecasting
- Corridor/route travel time estimation
- Corridor/route travel time forecasting

The proposed models are then demonstrated using travel time data obtained from Houston, Texas which were collected as part of the Automatic Vehicle Identification (AVI) system of the Houston Transtar system.

II. Notation

1. Index Variables

- t : time of day index in second
- h : basic or smaller time period or interval
- H : total time period of interest
- $\Delta\tau$: aggregation interval size
- N : total number of smaller intervals within a larger interval (=H/$\Delta\tau$)
- P : total future time for travel time forecasting
- k : future time period ahead from time period h
- K : total number of smaller time periods required for forecasting future P time period (=P/$\Delta\tau$)
- l : link index
- L : total number of links on the corridor
- v(h) : observed number of AVI vehicles at time period h
- v(h,k) : observed number of AVI vehicles at time period h+k
- v_l(h,k) : observed number of AVI vehicles on link l at time period h+k
- v_{l1l2}(h,k) : observed number of AVI vehicles at time period h+k that travel from link 1 to 2

2. Travel Time Variables

- $X(h)$: random variable for travel time on a link at time period h
- $X_l(h)$: random variable for travel time on link l at time period h
- $x^i(h)$: observed link travel time of the i-th vehicle on a link at time period h
- $x_l^i(h)$: observed link travel time of the i-th vehicle on link l at time period t
- $E(x^i(h)) = \mu^i_{x(h)}$: expected link travel time when the i-th vehicle travels the link at time period h
- $E(x(t) = \mu_{x(t)}$: expected link travel time on a link at time t
- $\hat{E}(x^i(h)) = \hat{\mu}^i_{x(h)}$: estimated mean link travel time when the i-th vehicle travels the link at time period h (kernel estimate in this paper)
- $\hat{E}(x^i(t)) = \hat{\mu}_{x(h)}$: estimated mean link travel time on a link at time t (kernel estimate in this paper)
- $X(h, k)$: random variable for travel time on a link at time period h+k
- $x^i(h\ k)$: observed link travel time of the i-th vehicle on a link at time period h+k
- $\hat{x}^i(h\ k)$: predicted link travel time of the i-th vehicle on a link at time period h+k. (That is predicted at time h for k-time period ahead)
- $\hat{X}(h, k)$: predicted mean link travel time on a link at time period h+k. (That is predicted at time h for k-time period ahead)
- $\overline{X}(h, k)$: sample mean link travel time on a link at time period h
- $\overline{X}_c(h)$: observed mean link travel time on a corridor at time period h
- $E(\overline{X}(h)) = \mu_{\overline{X}(h)}$: expected sample mean link travel time on a link at time period h
- $\overline{X}(H)$: observed mean link travel time on a link at larger time period H
- $\overline{X}(h\ k)$: observed mean link travel time on a link at time period h+k
- $\hat{X}(h, k)$: predicted mean link travel time on a link at time period h+k; (That is, predicted at time h for k-time period ahead)
- Δh : time period gap (in integer) between two links in terms of $\Delta\tau$ due to

the travel times on the first link, $\Delta h = \lfloor \frac{\overline{X}(h)}{\Delta\tau} \rfloor$ (note that the operator $\lfloor x \rfloor$ returns the largest integer smaller than or equal to x, i.e. the floor of x)

Δh_l : time period gap between the first link and the l-th link on corridor

3. Mean Square Error Variables

$MSE(h)$: mean square error (MSE) of link travel time estimation at time period h

$M\hat{S}E(h)$: estimated MSE of link travel time estimation at time period h

$M\hat{S}E(h)_{X_1X_2}$: estimated MSE of corridor travel time estimation with two links 1 and 2 at time period h

$M\hat{S}E(h)_{X_1X_2\cdots X_L}$: estimated MSE of corridor travel time estimation with L links(1, 2,.....L link) at time period h

$M\hat{S}E(H)$: estimated MSE of travel time esti estimation at bigger time period H

$M\hat{S}E(h, k)$: estimated MSE of travel time forecasting at time period h; for k time period ahead forecasting

$M\hat{S}E(h, P)$: estimated MSE of travel time forecasting at time period t for P future time interval (i.e. as many as 1~K multiple periods forecasting)

$M\hat{S}E(H, P)$: estimated MSE of travel time forecasting at time period H for P future time interval (i.e. 1~K multiple periods forecasting for each time periods 1~N)

4. Kernel Estimation Variables

w : window size of kernel estimate

m : number of observed AVI vehicles during a time window w of kernel estimate

s(t) : number of seconds during a time window w of kernel estimate

$f(x^j)$: kernel density of the vehicle j

$\delta(x^j)$: indicator of kernel density of vehicle j

Ⅲ. Optimal Aggregation Interval Size for Travel Time Estimation

The variance of sample mean has been used as the main criteria for identifying the best aggregation interval size for the link travel time estimation (Sen et al. 1998, Turner and Holdner, 1995). In this section a more comprehensive method is proposed which takes into account not only the variance of the sample mean but also the bias.

1. Mean Square Error : Link Travel Time Esti- mation

The MSE associated with a link travel time estimate for a given time period h of size $\Delta\tau$ is shown in Equation 1.

See Equation (1) in Appendix

Note that the MSE is comprised of two components. In this paper the first component will be referred to as bias while the second component will be referred to as the precision and is a measure of variability of the estimator. It may be seen that the bias decreases as the aggregation interval size decreases. In the extreme case, where only individual travel times are used, the first term would be zero. Conversely, the precision decreases as the aggregation interval size increases. This occurs because as the aggregation interval size increases the number of observations within the interval also increases which results in more reliable mean estimates. The goal of the paper is to identify the aggregation interval size that minimizes the MSE and therefore, some trade-off between precision and bias will be required.

The precision can be calculated using individual travel times and the observed mean link travel time. However, the bias cannot be directly obtained because the true mean travel time at time period h is unknown and consequently a continuous estimate of the mean is required. In this paper a Gaussian kernel is used to estimate the true mean travel time at time t and consequently the estimated mean travel time during period h is shown in Equation 2.

See Equation (2) in Appendix

Therefore, the estimated MSE for link travel time estimation would be:

See Equation (3) in Appendix

In Equation 2 the Guassian Kernel estimator is used to estimate the mean travel time for each time t $\hat{E}(x(t))$ as follows:

See Equation (4) in Appendix

See Equation (5) in Appendix

See Equation (6) in Appendix

Note that it assumed that the standard deviation, σ, is a quarter of window size of the kernel estimate, w. The optimal value of w is identified using a Generalized Cross Validation (GCV) technique (Eubank, 1999). The GCV value is calculated for each window size w and the value of w which minimizes the GCV is chosen.

See Equation (7) in Appendix

Once the MSE can be calculated for a given aggregation size as shown in Equation 2 the next step is to identify the optimal aggregation interval size over a given time period H. The aggregation interval sizes are assumed to of equal length $\Delta\tau$ and therefore the number of time periods for which travel times estimates are required would be $N=H/\Delta\tau$. The estimated MSE over the entire time period H would be as follows:

See Equation (9) in Appendix

Therefore, the objective function required to identify the optimal aggregation size is shown in Equation 10.

See Equation (10) in Appendix

2. Mean Square Error of Corridor Travel Time Estimation

While the estimation of link travel times is important, from the view point of a driver or operator, the most important travel time information would be with respect to a corridor or route. Therefore, an optimal aggregation interval size for corridor/route travel time estimation is also proposed.

1) Two Link Case

The methodology is first motivated using a simple two link route and is subsequently expanded to the more general multi-link route. Denote the two links which comprise the route as link 1 and link 2. The estimated MSE for travel time estimation on the route is shown in Equation 11.

See Equation (11) in Appendix

There are two important points to note about this formulation. The first is that the covariance can only be estimated for vehicles that travel across both links (i.e. $v_{12}(h)$). More importantly, unlike the assumption used in most literature on this subject, it is explicitly assumed that the travel time distributions across each link are dependent and therefore need to be accounted for within the methodology.

It can be seen for the two link network that the estimated MSE is the sum of the estimated MSE of each link and twice the estimated covariance. Therefore, if the travel times on the two links are positively correlated, the total error of the corridor travel time estimation would increase.

The overall MSE for time period H would be:

See Equation (12) in Appendix

Therefore, the best aggregation interval size, $\Delta\tau$, for the two link corridor travel time estimation problem can be identified by minimizing the following objective function:

See Equation (13) in Appendix

If the estimated covariance term in Equation 11 is not available then as an alternative approach, it can be approximated. First, the cross term in Equation 11 is denoted as A and it is partioned into two components as shown below:

See Equation (14) in Appendix

The overall estimated MSE of the corridor travel time estimation can then be reformulated as shown in Equation 15.

See Equation (15) in Appendix

Note that in Equation 15 the sample covariance, rather than the estimated covariance, is used and one additional cross term (i.e. the last term) has been added. The estimated MSE for interval H would be:

See Equation (16) in Appendix

It may be seen that the last term of Equation 16 would be close to zero because the product of the difference values will i) be relatively small and ii) will tend to cancel each other. Note that this assumption will be shown to be true based on an empirical analysis later in this paper. In this situation the sample covariance can be used to approximate the estimated covariance. In essence it is assumed that the sample variance is an unbiased estimator of the population variance. When estimating the covariance between links the observed mean travel time, rather than the estimated mean travel time, is used as shown in Equation 17.

See Equation (17) in Appendix

Given the above assumptions, the following objective function can be used to identify the best aggregation interval size for travel time estimation over the interval H for a two link corridor.

See Equation (18) in Appendix

2) Multiple link case

The formulation developed above can be generalized for the multiple link situation. For any random vector $(X_1, \ldots, X_n)$, total variance can be formulated as shown in Equation 19 (Casella & Berger, pp.237):

See Equation (19) in Appendix

Therefore, if the travel times on the L consecutive links on a corridor are denoted as X_1, X_2, $X_3, \ldots X_L$, the estimated MSE of the travel time estimation for a corridor for larger time period H is shown in Equation 20. The aggregation interval size which minimizes the estimated MSE would be optimal. Note that sample covariance between all link pairs on a corridor is considered explicitly in this formulation.

See Equation (20) in Appendix

Ⅳ. Optimal Aggregation Interval Size for Travel Time Forecasting

The preceding sections focused on providing an estimate of the link travel time over some time interval at some point in the past. The past could refer to the previous five minutes or to a five minute period in the previous month. However, in a real-time situation the most useful link travel time information for a driver would be based on when the driver expects to traverse the link rather than an estimate of past conditions. Intuitively, for the real-time situation there is higher level of uncertainty because the link travel time will need to be forecast into the near future. The objective of this section is to identify the optimal aggregation size of the time periods over a given forecasting period H for travel time forecasting.

1. Mean Square Error of Link Travel Time Forecasting

The mean square error (MSE) of the link travel time forecast at time period h for k time periods ahead is given by Equation 21.

See Equation (21) in Appendix

See Equation (22) in Appendix

Note that the above equation requires a forecast of the link travel time in each period from h to h+k. There are a number of link travel time forecasting techniques that can be used for this task including time series, artificial neural networks, and Kalman filtering (Boyce et al. 1993; Tarko and Rouphail, 1993; Van Arem et al. 1997; Park and Rilett, 1998/1999, Park et al. 1999; Rilett and Park, 1999).

The MSE shown in Equation 21 can be decomposed into two components as shown below.

See Equation (23) in Appendix

The first component is associated with the variance between the observed mean travel time and individual travel time. It may be seen that the first term would decrease as the aggregation interval size decreases. The second component results from the difference between the observed mean travel time and the forecast mean travel time. In contrast to the first component, this term would decrease as the aggregation interval size increases due to smoothing effects. As would be expected the MSE for link travel time forecasting is different from that for link travel time estimation, because the goal is to forecast the future instead of estimating the past.

Given that the link travel time is forecast at time period h for one through K time periods into the future (i.e. total time forecast would be P), the overall MSE would be as follows:

See Equation (24) in Appendix

Consequently, the overall MSE for the P time periods during the entire period of interest H is shown below:

See Equation (25) in Appendix

The objective function which provides the aggregation interval size $\Delta\tau$ that minimizes the overall MSE, is shown in Equation 26 and it may be seen that this is simply the minimization of Equation 25.

See Equation (26) in Appendix

2. Mean Square Error of Corridor Travel Time Forecasting

Similar to the travel time estimation problem, the most useful information from an individual drivers perspective would be to provide the forecast route travel time rather than the forecast travel time for the individual links which consist of that route. This section discusses how to identify the best aggregation interval size for forecasting route travel times where the decision metric is the MSE.

1) Two Link Case

The methodology will be motivated using a two link route and subsequently generalized to the n-link route situation. It is assumed that at time period "*h*" the route travel time will be forecast for "*k*" time periods ahead and this process will involve two steps. The forecast route travel time is modeled as the sum of link 1's travel time (given the appropriate starting period) plus link 2's travel time based on when the vehicle arrives at link 2. The first step is to forecast link 1's travel time "*k*" time periods ahead. Because it may be possible for a vehicle to arrive at link two in a time period that is greater than t+k, it may be necessary to forecast the link travel time on link two for more than k periods into the future. Therefore, the second step is to forecast the second links travel time "$k+h_1$" time periods ahead. Once the travel times on the links have been forecast the route travel time is calculated by summing the appropriate link travel times. The MSE associated with forecasting the route travel time "*h*" time periods ahead is:

See Equation (27) in Appendix

The first and second terms in Equation 27 may be obtained by Equation 23. Denote the cross term of Equation 27 as B. Using the same logic that was applied to Equation 14, B can be decomposed as shown in Equation 28:

See Equation (28) in Appendix

From Equation 28 it can be seen that B is the sum of the sample covariance between links 1 and 2 and the cross term of the difference between observed mean travel time and predicted mean travel time on links 1 and 2. Accordingly the overall MSE of the corridor travel time forecasting can be summarized as follows:

See Equation (29) in Appendix

It has been shown empirically in previous research that the cross term in Equation 29 cancels out when forecasting route travel times for the "*P*" future time period (i.e. one through *K* time periods ahead) (Rilett and Park, 2000). Based on this assumption the overall estimated MSE can be written as follows:

See Equation (30) in Appendix

Lastly, if the MSE is calculated for the entire time period of interest, "*H*", the overall estimated MSE of the corridor travel time forecasting would be as follows:

See Equation (31) in Appendix

2) Multiple Link Case

Similar to Equation 20, the overall estimated MSE of the corridor travel time forecasting would be the sum of the MSE of the link travel time forecasting and the sample covariance of all link pairs on the corridor.

V. Data Collection and Study Design

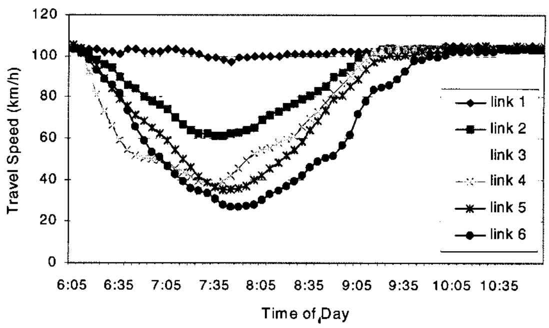
1. Study Freeway and Data Collection

1) Study Freeway Corridor : US 290 in Houston, Texas

The test bed for this study was US-290 which is a radial six-lane urban freeway in Houston as shown in ⟨Figure 1⟩. It has a barrier-separated HOV lane that runs along the center line of the freeway for approximately nineteen kilometers and the data utilized was from the non-HOV section of the freeway.

2) Data Collection

Travel time data were collected over a 27.6 kilometers stretch of US-290 from seven AVI reader stations (yielding six links) as shown in ⟨Figure 1⟩. The data were collected over a twenty-four hour period each weekday in both directions of travel for eighteen months from January, 1996 to June, 1997 yielding 342 weekdays. Only the Eastbound AM peak travel time data were employed in this study because these links experienced higher congestion levels than the Westbound links. Based on a visual inspection of the travel time patterns, the study time period was defined as lasting from 6:00 AM to 11:00 AM and the data collected over this time period were used for this study. ⟨Figure 2⟩ shows the peak period space mean speeds averaged

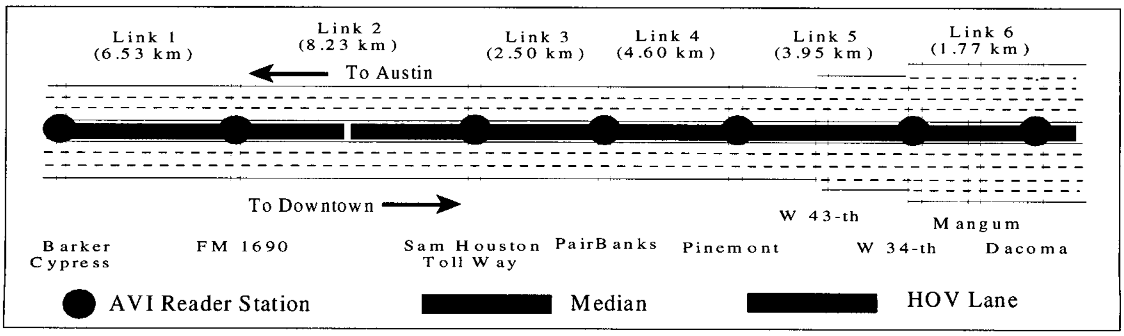


⟨Figure 2⟩ Mean Travel Speed vs Time of Day

over a five-minute period for 342 days on the six links. It may be seen that various levels of congestion occur during both the peak and non-peak periods. During this peak period, an average of approximately seventeen to thirty five vehicles per five minute period traversed links 3, 4, 5, and 6 while an average of approximately eight to twelve vehicles traversed links 1 and 2 per five-minute period.

3) Space Mean Speed

For visual comparison purposes, the link travel speeds are better indicators of conditions on the highway than link travel times because the links have different lengths. Because of the nature of the AVI data, the "space" mean speed rather than the "time" mean speed is obtained for any given time period and will be referred to as the mean speed in this paper. The mean speed on link 3 experiences traffic congestion earlier than the other



⟨Figure 1⟩ Study Urban Freeway Corridor US-290 in Houston, Texas links. It may be seen that after 6:00 AM, the travel time on link 3 decreases rapidly to approximately 40 km/h at 6:30 AM. The mean speed on links 4 and 5 ranges between approximately 30 km/h and 40 km/h during the 7:10 AM to 7:40 AM time period. The mean speed on link 6 is approximately 30 km/h between 7:30 AM and 8:10 AM, indicating that link 6 experiences a more severe and later period of congestion than the other links. Link 2 has a comparatively lower level of congestion as evidenced by the fact it only experiences a 40 km/h drop in average speed during the AM peak. In contrast to the other links, link 1 experiences an average travel speed pattern which is almost equal to the free-flow travel speed. The lowest speeds on links 2, 3, 4, and 5 occur at approximately 7:30 AM while that of link 6 occurs at approximately 7:50 AM.

2. Study Design

Based on the above analysis links 3, 4 and 5 were selected as the test bed for studying the optimal aggregation interval size problem. The corridor consisting of links 3, 4 and 5 was used for the corridor analysis. Links 1 and 2 were not included in the corridor because those links do not have a significant number of AVI vehicles during each interval. Link six was not examined because it had a bimodal travel time distribution caused by the fact that the travel time is a function of whether vehicles exit onto the eastbound or southbound section of IH 610.

The period of study was the AM peak period from 6 AM to 11 AM. The major time period of interest was one hour and therefore there were five major periods analyzed. These time periods are denoted as 6~7, 7~8, 8~9, 9~10, and 10~11. Within each hour period seven aggregation sizes, $\Delta\tau$, were analyzed for travel time estimation and forecasting: 1, 2, 3, 5, 10, 15, and 30 minutes. The MSE results presented in this paper corresponds to the aggregate MSE values for each of the one-hour periods unless mentioned otherwise.

The prediction period, P, used in the travel time forecasting analysis was set to 30 minutes. Therefore, the number of time periods studied was $N=P/\Delta\tau$ where N is always an integer value. For example, if a five minute aggregation interval was examined, then the link travel time was forecast for six consecutive time periods while if a thirty minute aggregation interval was examined then the travel time was forecast for only one time period ahead. The link travel time was forecast using a spectral basis neural network (SNN) because previous research had shown this technique to have the lowest average absolute percent error (Park et al., 1999). Half of the 342 days' travel time data was randomly chosen as a training data set and the other half was used as a testing data set. Therefore, the analysis of the results was based on the overall MSE for the 171 testing days. In addition, it was assumed that the input and output travel time has the same aggregation size. The "optimal" number of recent travel times that were used as input data for the SNN for each aggregation interval size was determined through a preliminary analysis.

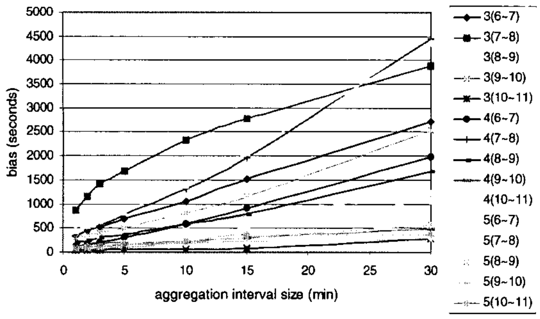
Ⅵ. Analysis of Results

1. Optimal Aggregation Interval Size for Travel Time Estimation

1) Link Travel Time Estimation

(1) Bias of MSE

〈Figure 3〉 shows the first component of the link travel time estimation MSE from Equation 10, referred to in this paper as the bias, for each time period over the entire 172 day testing data set on links 3, 4, and 5, respectively. It may be seen that, in general the bias increases as the aggregation interval size increases. For example,



〈Figure 3〉 Bias of Link Travel Time Estimation MSE

the bias for the 7~8 time period on link 3 for aggregation interval sizes of 1, 2, 3, 5, 10, 15, and 30 minutes were 881, 1162, 1408, 1689, 2337, and 3891 seconds, respectively. Other links have a similar pattern except for the 9~10 and 10~11 time periods where the bias term was relatively constant regardless of the aggregation interval sizes.

In order to examine the reason that the bias term is a function of time of day can be explained by decomposing the first component of the MSE in Equation 1 as shown in Equation 32.

See Equation (32) in Appendix

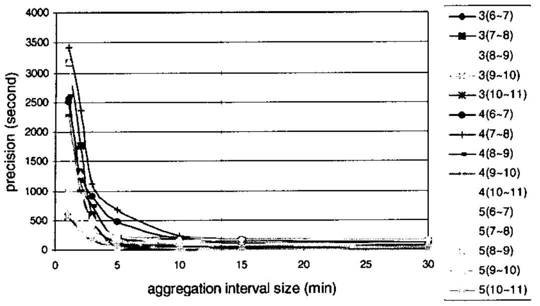
It may be noted that the middle component consists of two terms. The first term represents the relative difference between an individual vehicle i's travel time during time period period h and the expected mean speed during time period h. The second term represents the expected mean travel time during period h and the average travel time over the given aggregation interval. In other words, the first term is a function of the individual driver's characteristics while the second term is a function of the traffic dynamics or travel time fluctuation. From a traffic flow perspective it would be reasonable to assume that these two random variables are independent and that the expected value of the cross term is zero. Based on an empircial analysis of the test bed data it was found that on average the absolute value of the cross term of Equation 32 on link 4 during the entire study time period ranges between 9.1 % and 18.5 % of the first and third terms, respectively. Accordingly Equation 32 can be approximated by:

See Equation (33) in Appendix

Note that if there is no systematic fluctuation in average travel time during an interval, the second term of Equation 33 would be zero and consequently the first component of the MSE would be approximately equal to the first term of Equation 33 and would be constant regardless of the aggregation interval size. This relationship exists because the difference in travel times between vehicles traveling a link at the same instant would be similar because drivers' characteristics do not change as a function of the level of congestion. Given that the time periods 9~10 and 10~11 are relatively uncongested this would explain why the bias-aggregation interval functions tend to be flat. However, when travel time fluctuations exist, such as under congested conditions, the first component of the MSE increases as the aggregation interval size increases because the mean travel time of an interval and the travel time for each time instant would change continuously.

(2) Precision of MSE

〈Figure 4〉 represents the second component of the MSE (i.e. precision) on links 3, 4, and 5 over different time periods as a function of aggregation interval size. It can be seen that as the aggregation interval increases, the second component of MSE decreases and after some threshold value remains stable. These results are similar to that found in previous study (Sen et al., 1997). It may be seen that the exact threshold values were found to be dependent on the level of congestion in that as congestion levels increased the threshold values increased as well. In order to explain



〈Figure 4〉 Precision of Link Travel Time Estimation MSE

this phenomenon the second component of the MSE is decomposed into two terms as follows:

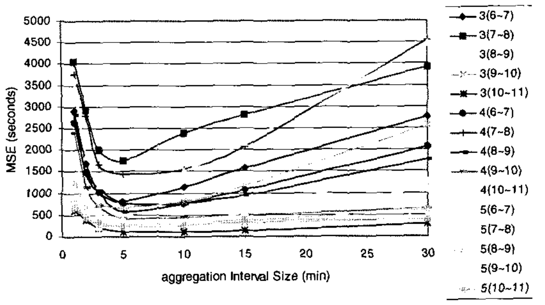
See Equation (34) in Appendix

It may be seen in Equation 34 that the first term corresponds to the sum of the square error divided by square of the observed vehicles within an interval. It can be noted that this term would decrease as the aggregation interval size increases because it's denominator would increase at a higher rate than it's numerator. From the perspective of traffic flow, it is hypothesized that two sources may be responsible for this phenomnon. The first source is the driver's characteristics with respect to aggressiveness and the other source represents the ability of a driver to choose a travel speed greater than the mean traffic speed. If it is assumed that these two possible sources of variance remain constant, the numerator would increase linearly as the aggregation interval size increases. However, the first term would decrease as the aggregation interval increases because the denominator increases exponentially. Conversely, it would be expected that as the level of congestion increases the numerator would increase only slightly because of the drivers inability to travel at the speed they would like as as the congestion level increases. Regardless, the magnitude of the increase in the difference squared would be smaller than that of the squared number of probe vehicles.

A similar conceptual interpretation can be applied to the second term in Equation 34. From a traffic flow viewpoint, the second term "represents" the correlation or covariance between vehicles of a link within a given time interval. Therefore, the term would be positive if a vehicle travels a link faster (or slower) than the mean travel time of an instant or time of day (i.e. kernel estimate for each instant) while the other vehicles starting the same link at the different time of day travel the link faster (or slower) than the mean travel time of the time of day. In contrast to the first term in Equation 34, in the second term the relative size of the numerator is close to the denominator (i.e. $v(h)(v(h)-1)$ vs. $v(h)^2$). Therefore in general the second term would be non-zero even for large aggregation interval size as long as there is a significant correlation between vehicles. Therefore, the second component of the MSE for link travel time estimation is essentially constant after some threshold value.

(3) Overall MSE

〈Figure 5〉 shows the overall MSE of the link travel time estimation over the different analysis periods as a function of aggregation interval size for links 3, 4, and 5. As would be expected based on the preceding discussion the MSE relationships are convex in shape. It may be seen that the best aggregation interval size for all of the links and all analyses periods is five minutes. The only exceptions to this result are the 6~7 and 10~11 time periods on link 4 where the optimal interval size is ten minutes. However, in these two cases the MSE for the five minute aggregation interval was very similar to the MSE for the ten minute aggregation interval. It may also be seen in 〈Figure 5〉 that as the congestion level increases, the convexity of the MSE curve also increases. For example, during lower congestion periods the overall MSE did not vary much for the different aggregation interval sizes. It is hypothesized that during non-congested



〈Figure 5〉 Overall Link Travel Time Estimation MSE

peak hour periods, the mean travel time for each instant (i.e. $E(x^i(t))$) is approximately equal to the interval estimate of the travel time which would account for this result.

2) Corridor Travel Time Estimation

As discussed earlier, the overall corridor travel time estimation error consists of two components: one is the link travel time estimation error and the other is the covariance between links on the corridor. Because the MSE for the link travel time estimation has already been examined the focus of this section is on the analysis of the covariance between links. The estimated covariance for all aggregation interval sizes was approximated using Equation 18. Note that, the one minute interval size was not analyzed because there were not enough AVI observations to identify the covariance and because the best size of aggregation interval would not be less than 2 minutes based on the best link travel time estimation aggregation interval size.

〈Table 1〉 shows the overall sample covariance between links for each one hour time period. It may be seen that the overall sample covariance between links for each one hour time period is relatively small compared to the MSE of the link travel time estimation. The absolute values of *sample covariance ranges between 0 and 276 seconds* and there does not appear to be a systematic pattern. It is important to note that the average

〈Table 1〉 Sample Covariance between Links vs Aggregation Interval Size

Links	Time Period	Aggregation Interval Size (min.)					
		2	3	5	10	15	30
3 and 4	6~7	-1	10	-44	0	73	276
	7~8	8	32	47	-48	-244	-32
	8~9	51	52	41	52	46	218
	9~10	44	50	51	57	73	91
	10~11	54	54	58	60	61	70
3 and 5	6~7	12	-5	16	-32	19	134
	7~8	23	18	33	-10	26	219
	8~9	25	35	47	39	41	-102
	9~10	41	44	61	45	79	-56
	10~11	29	31	55	66	36	72
4 and 5	6~7	29	29	10	22	-15	201
	7~8	53	62	23	-93	-232	-152
	8~9	44	27	36	1	6	194
	9~10	67	54	72	72	73	89
	10~11	74	67	76	85	88	102

of the sum of the MSE of link travel time estimation error for the best aggregation interval size is approximately 630 seconds and therefore the travel time estimation error of the test bed corridor would be at least 1890 (630×3) seconds. Therefore, from the overall corridor travel time estimation point of view, for each hour time period the sample covariance does not have a significant effect.

The smaller absolute value of sample covariance between links results from the fact that the sample covariance for each time period (e.g. 5 minutes) effectively cancels out each other. For example, in the case of 5 minute aggregation for the 7~8 time periods between links 3 and 4 for 171 days, the biggest variance of the sample covariance bet ween twelve 5 minute intervals among 171 days is 18,775,683 and the average sample covariance over twelve intervals of the day is -434. For that day the sample covariance for one through twelve 5 minute time periods were 234, 3924, 455, 1478, 578, 919, -2622, -277, 5263, -12238, -106, -2199, and -383, respectively. On the other hand, the smallest variance of the sample covariance for each 5 minute interval among 171 days is 376 and the average sample covariance is 29. The sample covariance for one through twelve 5 minute intervals of that day are 37, 24, 3, 53, 4, 53, 50, 9, 36, 14, 18, 48, and 29 respectively. As the aggregation interval size increases, the variance of the sample covariance between intervals were found to decrease.

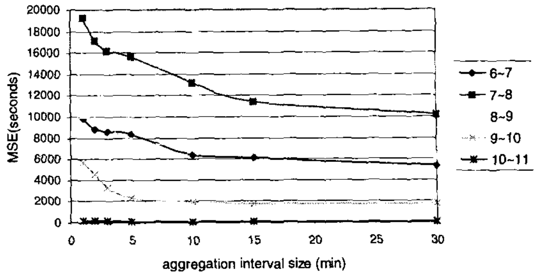
To summarize, there are two important implications of the covariance between links. Firstly, the covariance between links fluctuates not only between intervals but also between days. Secondly, in contrast to the covariance for each one hour time period H which is relatively small compared with the link travel time estimation error, the covariance between links for each time interval is significantly large. Given these two findings, unless the covariance between links for each time interval can be accurately predicted, it is unlikely that the effect of the covariance can be taken into account from an individual interval perspective.

On the other hand, regarding the approximation of the estimated covariance using the sample covariance (i.e. using Equation 18 instead of Equation 13 or Equation 17), we found that for the 91.2 percent of the possible 855 one-hour intervals (i.e. 171 days×5 one-hour time period per day) the absolute differences between these two covariance were within the 5 percent of the estimated covariance and for the other 8.8 percent the absolute differences were within 10 percent.

2. Optimal Aggregation Interval Size for Travel Time Forecasting

1) Link Travel Time Forecasting

As shown in Equation 23, the link travel time forecasting MSE consists of two components. The first is the bias between the observed mean link travel time and individual travel times and the other is the difference between observed and pre-



〈Figure 6〉 Second Component of Link Travel Time Forecasting MSE

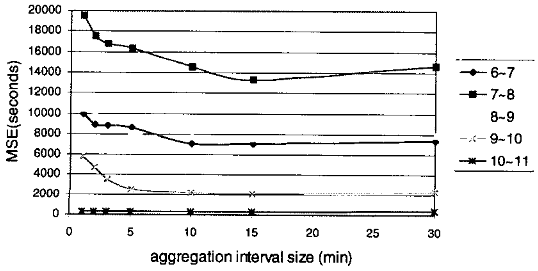
dicted mean link travel times. This section focuses on the second term because the first term was discussed in the previous section. In addition, link 4 was selected as a test bed link for the travel time forecasting problem because the traffic dynamics on links 3, 4, and 5 during the study period were similar as shown in the link travel time estimation problem. 〈Figure 6〉 shows the second component of link travel time forecasting MSE for link 4. In general as the aggregation interval size increases, the second component of MSE decreases. It is hypothesized that the smoothing effect and smaller number of forecasting periods associated with using a larger interval size are the main causes for this behavior. For example, if a fifteen minute aggregation interval (i.e. $\Delta\tau=15$ min.), was used then the travel time for two time periods need to be forecast (i.e. K=2) while if a one minute aggregation period is used then the travel time for thirty time periods would need to be forecast.

It may also be seen that each one-hour time period (i.e. H=60 min.) had different values of the second component error. For example, the second component of the forecasting MSE for the 7~8 time period ranged from 19,235 to 10,172 for the one through thirty minutes aggregation interval size, while those of 10~11 time period ranged from 192 to 97. This result would be expected because the more the link travel times fluctuate, the more difficult they are to predict (See also Park and Rilett 1998/1999, Park et al.

1999, Rilett and Park 2000). Another interesting result is that after a threshold value is recorded the second component of the MSE remains relatively constant and the time period with less traffic congestion had the smaller threshold values (e.g. the second component of MSE for 5 minute aggregation for 9~10 and 10~11 time periods and 15 minutes aggregation for 7~8 time period). More importantly, except for the 9~10 and 10~11 time periods, the second components of MSE are significantly larger than those of the first component of the MSE.

⟨Figure 7⟩ illustrates the overall MSE of the link travel time forecasting. The overall MSE decreases and then increases slightly after a minimum is reached. In addition the overall MSE around the threshold values are similar. For instance, for the 7~8 time period the overall MSE decreased by 14,568 and 13,301 when 10 and 15 minute aggregation intervals were used, respectively, and then increased up to 14,625 when a 30 minutes aggregation interval were used. The best size of the aggregation interval for the 6~7, 7~8, 8~9, 9~10 and 10~11 time periods were 10, 15, 15, 15, 5 minutes, respectively. Note, however, that around these values the overall MSE are very similar. For example, the 10~11 time period, when there is no traffic congestion, had approximately the same overall MSE regardless of the aggregation interval size.

Compared with the best aggregation interval sizes of 3~5 minutes for the link travel time estimation, the best aggregation interval sizes for



⟨Figure 7⟩ Overall Link Travel Time Forecasting MSE

travel time forecasting range from 5 to 15 minutes. It is hypothesized that this occurs because of the difficulty of multiple period link travel time forecasting, particularly under conditions of severe traffic congestion. Note that as the congestion level increased the difference between the best aggregation interval sizes for link travel time estimation and forecasting increased as well.

2) Corridor Travel Time Forecasting

Based on the previous results it is hypothesized that the best aggregation interval sizes for forecasting corridor travel times are equivalent to the best aggregation interval size for link travel time forecasting. This is because the MSE of the corridor travel time forecasting consists of the link travel time forecasting MSE of the links on the corridor and the (sample) covariance between links, and the covariance between links for each one-hour time periods were relatively small compared with the MSE for the link travel time forecasting.

Ⅶ. Concluding Remarks

In this paper statistical models were proposed which identify the best aggregation interval sizes for link and corridor/route travel time estimation and forecasting under various types of traffic dynamics and frequency of observations. The proposed models are based on the mean square error between travel time estimate or forecasts and the individual realizations. They take into account not only the precision but also the bias of the travel time estimate and forecasts and emphasize the tradeoff between them. The models were demonstrated using the actual AVI travel time data from Houston, Texas.

The best aggregation interval sizes for the link travel time estimation and forecasting were different and the function of the traffic dynamics. For the best aggregation interval sizes for the corridor/ route travel time estimation and forecasting, the covariance between links had an important effect. It seems that the proposed approaches are applicable for other transportation areas such as other traffic parameter estimation and forecasting (e.g. traffic volume) and calibration and validation of traffic simulation models.

Important directions for future study of this paper would be to extend the proposed approaches to i) arterial streets which have more complex and unique traffic characteristics and ii) a problem which identifies the best spatial size for the travel time estimation and forecasting. A methodology which estimates mean and variance for a bigger aggregation interval using the summary statistics from smaller aggregation intervals without keeping individual travel time realizations would also be valuable for the efficient management of the ATIS.

References

1. Boyce, D., N. Rouphail, and A. Kirson, (1993) "Estimation and Measurement of Link Travel Times in the ADVANCE Project", *IEEE-IEE Vehicle Navigation & Information Systems Conference*, Ottawa, Canada, October 12-15, pp.62~66.
2. Dailey, D. J., (1993) "Travel Time Estimation Using Cross-Correlation Techniques", *Transportation Research -Part B*, Vol. 27, No. 2, pp.97~107.
3. Davis, G. A. and N. L. Nihan, (1991) "Nonparametric Regression and Short-Term Freeway Traffic Forecasting", *Journal of Transportation Engineering*, Vol. 177, No. 2, pp.178~188.
4. Eubank, R.L. (1999), *Nonparametric Regression and Spline Smoothing*, Second Edition, Marcel Dekker, New York, New York.
5. Fu, L. and L.R. Rilett, Expected Shortest Paths in Dynamic and Stochastic Traffic Networks, *Transportation Research-Part B*, Vol. 32, No. 7, 1998, pp.499~514.
6. Gajewski, B., S. Turner, B. Eisele, and C. Spiegelman, (2000), "ITS Data Archiving: Statistical Techniques for Determining Optimal Aggregation Widths for Inductance Loop Detectors", Presented at 2000 TRB Meeting and submitted to the Journal of Transportation Research Board for publication.
7. Hall, R. (1986), The Fastest Path through a Network with Random Time-Dependent Travel Time, *Transportation Science*, Vol. 20, No. 3, 1986, pp.182~188
8. Miller-Hooks, E.D. and Hani, S. Mahmassani (1999), Least Expected Time Paths in Stochastic, Time-Varying Transportation Networks, *Forthcoming in Transportation Science*.
9. Okutani, I. and Y. J. Stephanedes, (1984) "Dynamic Prediction of Traffic Volume Through Kalman Filtering Theory", *Transportation Research-Part B*, Vol. 18, No. 1, pp.1~11.
10. Park, D. (1998), Multiple Path Based Vehicle Routing in Dynamic and Stochastic Transportation Networks, *Ph.D. Dissertation*, Department of Civil Engineering, Texas A&M University, Texas.
11. Park, D. and L.R. Rilett(1998), Forecasting Multiple-Period Freeway Link Travel Time using Modular Neural Networks, *TRR 1617*, pp.163~170.
12. Park, D. and L.R. Rilett (1999), Forecasting Freeway Link Travel Times with a Multilayer Feedforward Neural Network, *Computer-Aided Civil and Infrastructure Engineering*, Vol. 14, pp.357~367.
13. Park, D., L.R. Rilett, and G.Han (1999), Spectral Basis Neural Networks for Real-Time Travel Time Forecasting, *ASCE Journal of Transportation Engineering*, Vol. 125, No. 6, pp.515~523.
14. Rilett, L.R. and D. Park (2000), Direct Forecasting of Freeway Corridor Travel Times Using Spectral Basis Neural Networks, Forthcoming in *TRR*.
15. Rilett, L.R., D. Park, and B.J. Gajewski(1999), Estimating Confidence Interval for Freeway Corridor Travel Time Forecasts, Presented at

the 6th Annual World Congress on ITS, Toronto, Canada, November, 1999.

16. Srinivasan, K.K. and P. P. Jovanis (1996), Determination of Number of Probe Vehicles Required for Reliable Travel Time Measurement in Urban Network, *TRR* 1537, pp.15~22.
17. Sen A., P. Thakuriah, X.-Q. Zhu, and A. Karr (1997), Frequency of Probe Reports and Variance of Travel Time Estimates, *ASCE Journal of Transportation Engineering*, Vol. 123, NO. 4, pp.290~299.
18. Turner, S.M. and D.J. Holdner (1995), Probe Vehicle Sample Sizes for Real-Time Information: The Houston Experience, *IEEE VNIS Conference.*
19. Tarko, A. and N. M. Rouphail, (1993) "Travel Time Data Fusion in ADVANCE", *ASCE Third International Conference on Applications of Advanced Technologies in Transportation Engineering*, Washington, D.C., July, pp.36~42.
20. Van Arem, B., M. J. M. Van Der Vlist, M. R. Muste, and S. A. Smulders, (1997) "Travel Time Estimation in the GERDIEN Project", *International Journal of Forecasting*, Vol. 13, pp.73~85.
21. Van Der Voort, M., M. Dougherty, and S. Watson, (1996) ACombing Kohonen Maps with "RIMA Time Series Models to Forecast Traffic Flow", *Transportation Research - Part C*, Vol. 4, No. 5, pp.307~318.

〔Appendix〕

$$MSE(h) = \frac{\sum_{i=1}^{v(h)} (x^i(h) - \mu_{\overline{X}(h)})^2}{v(h)}$$

$$= \frac{\sum_{i=1}^{v(h)} (x^i(h) - \overline{X}(h) + \overline{X}(h) - \mu_{\overline{X}(h)})^2}{v(h)}$$

$$= \frac{\sum_{i=1}^{v(h)} (x^i(h) - \overline{x}(h))^2 + 2\sum_{i=1}^{v(h)} (x^i(h) - \overline{X}(h))(\overline{X}(h) - \mu_{\overline{X}(h)}) + \sum_{i=1}^{v(h)} (\overline{X}(h) - \mu_{\overline{X}(h)})^2}{v(h)}$$

$$= \frac{\sum_{i=1}^{v(h)} (x^i(h) - \overline{X}(h))^2}{v(h)} + (\overline{X}(h) - \mu_{\overline{x}(h)})^2 (\textit{cross term is zero}) \quad (1)$$

$$\hat{E}(\overline{X}(h)) = \hat{\mu}_{\overline{X}(h)} = \frac{\sum_{t=1}^{s(t)} \hat{E}(x(t))}{s(t)} \quad (2)$$

where

$\hat{E}(x(t))$: expected travel time at time t (Gaussian kernel)

w : window size of kernel estimate

$s(t)$: duration of time window w (seconds)

$$M\hat{S}E(h) = \frac{\sum_{i=1}^{v(h)} (x^i(h) - \overline{X}(h))^2}{v(h)} + (\overline{X}(h) - \hat{\mu}_{\overline{X}(h)})^2 \quad (3)$$

$$\hat{E}(x(t)) = \frac{\sum_{j=1}^{m} \delta(x^j) f(x^j) x^j}{\sum_{j=1}^{m} \delta(x^j) f(x^j)} \quad (4)$$

where:

x^j : j-th vehicle's travel time

m : number of vehicles observed during time window w

$\delta(x^j)$:

$$\delta(x^j) = \begin{bmatrix} 1, & \text{if } |t^j - t| \le \frac{w}{2} \\ 0, & \textit{otherwise} \end{bmatrix} \quad j = 1, \cdots, m \quad (5)$$

t^j : time of day when the j-th vehicle travels a link

$$f(x^j) = \frac{1}{\sqrt{2\pi\sigma}} \exp -(1/2)[t^j - t)/\sigma]^2 \quad (6)$$

$$GCV(w) = \frac{m \sum_{j=1}^{m} (x^j - \hat{E}(x^j))^2}{\left(\sum_{j=1}^{m} ((1 - z^j) \right)^2} \tag{7}$$

where:

$\hat{E}(x^j) = \hat{E}(x(t))$: time of day when the j-th vehicle travels the link is t

and

$$z^j \frac{f(x^j)}{\sum_{j=1}^{m} \delta(x^j) f(x^j)} \tag{8}$$

$$M\hat{S}E(H) = \frac{\sum_{h=1}^{N} M\hat{S}E(h) \cdot v(h)}{\sum_{h=1}^{N} v(h)} \tag{9}$$

$$Minimize_{\Delta\tau} = \frac{\sum_{t=1}^{N} M\hat{S}E(t) \cdot v(t)}{\sum_{t=1}^{N} v(t)} \tag{10}$$

$$\begin{aligned} M\hat{S}E(h)_{X_1X_2} &= \frac{\sum_{i=1}^{v_{12}(h)} \left(x_1^i(h) + x_2^i(h + \Delta h) - \hat{\mu}_{\overline{X}_1(h)} - \hat{\mu}_{\overline{X}_2(h+\Delta h)} \right)^2}{v(h)} \\ &\approx \frac{\sum_{i=1}^{v_1(h)} \left(x_1^i(h) - \hat{\mu}_{\overline{X}_1(h)} \right)^2}{v_1(h)} + \frac{\sum_{i=1}^{v_2(h)} \left(x_2^i(h) - \hat{\mu}_{\overline{X}_2(h)} \right)^2}{v_2(h)} + \frac{2 \sum_{i=1}^{v_{12}(h)} \left(x_1^i(h) - \hat{\mu}_{\overline{X}_1(h)} \right) \left(x_2^i(h + \Delta h) - \hat{\mu}_{\overline{X}_2(h+\Delta h)} \right)}{v_{12}(h)} \\ &= M\hat{S}E(h)_{X_1} + M\hat{S}E(h + \Delta h)_{X_2} + \frac{2 \sum_{i=1}^{v_{12}(h)} \left(x_1^i(h) - \hat{\mu}_{\overline{X}_1(h)} \right) \left(x_2^i(h + \Delta h) - \hat{\mu}_{\overline{X}_2(h+\Delta h)} \right)}{v_{12}(h)} \\ &= M\hat{S}E(h)_{X_1} + M\hat{S}E(h + \Delta h)_{X_2} + 2C\hat{o}v(X_1(h),\ X_2(h + \Delta h)) \end{aligned} \tag{11}$$

where:

$\hat{\mu}_{\overline{X}_i(h)}$: the kernel estimator for $\mu \overline{X}_i(h)$

$$M\hat{S}E(H)_{X_1X_2} \approx \frac{\sum_{h=1}^{N} M\hat{S}E(h)_{X_1} \cdot v_1(h)}{\sum_{h=1}^{N} v_1(h)} + \frac{\sum_{h=1}^{N} M\hat{S}E(h + \Delta h)_{X_2} \cdot v_2(h)}{\sum_{h=1}^{N} v_2(h)} + \frac{2 \sum_{h=1}^{N} C\hat{o}v(X_1(h),\ X_2(h + \Delta h)) \cdot v_{12}(h)}{\sum_{h=1}^{N} v_{12}(h)} \tag{12}$$

$$Minimize_{\Delta\tau} \frac{\sum_{h=1}^{N} M\hat{S}E(h)_{X_1} \cdot v_1(h)}{\sum_{h=1}^{N} v_1(h)} + \frac{\sum_{h=1}^{N} M\hat{S}E(h + \Delta h)_{X_2} \cdot v_2(h)}{\sum_{h=1}^{N} v_2(h)} + \frac{2 \sum_{h=1}^{N} C\hat{o}v(X_1(h),\ X_2(h + \Delta h)) \cdot v_{12}(h)}{\sum_{h=1}^{N} v_{12}(h)} \tag{13}$$

$$A=\frac{\sum_{i=1}^{v_{12}(h)}\left(x_1^i(h)-\overline{X}_1(h)+\overline{X}_1(h)-\hat{\mu}_{\overline{X}_1(h)}\right)\left(x_2^i(h+\Delta h)-\overline{X}_2(h+\Delta h)+\overline{X}_2(h+\Delta h)-\hat{\mu}_{\overline{X}_2(h+\Delta h)}\right)}{v_{12}(h)}$$

$$=\frac{\sum_{i=1}^{v_{12}(h)}(x_1^i(h)-\overline{X}_1(h))(x_2^i(h+\Delta h)-\overline{X}_2(h+\Delta h))}{v(h)}+\left(\overline{X}_1(h)-\hat{\mu}_{\overline{X}_1(h)}\right)\left(\overline{X}_2(h+\Delta h)-\hat{\mu}_{\overline{X}_2(h+\Delta h)}\right)$$

$$=S_{X_1(h)X_2(h+\Delta h)}+\left(\overline{X}_1(h)-\hat{\mu}_{\overline{X}_1(h)}\right)\left(\overline{X}_2(h+\Delta h)-\hat{\mu}_{\overline{X}_2(h+\Delta h)}\right) \tag{14}$$

$$M\hat{S}E(h)_{X_1X_2}\approx M\hat{S}E(h)_{X_1}+M\hat{S}E(h+\Delta h)_{X_2}+2S_{X_1(h)X_2(h+\Delta h)}$$
$$+2\left(\overline{X}_1(h)-\hat{\mu}_{\overline{X}_1(h)}\right)\left(\overline{X}_2(h+\Delta h)-\hat{\mu}_{\overline{X}_2(h+\Delta h)}\right) \tag{15}$$

$$M\hat{S}E(H)_{X_1X_2}\approx\frac{\sum_{h=1}^{N}M\hat{S}E(h)_{X_1}\cdot v_1(h)}{\sum_{h=1}^{N}v_1(h)}+\frac{\sum_{h=1}^{N}M\hat{S}E(h+\Delta h)_{X_2}\cdot v_2(h)}{\sum_{h=1}^{N}v_2(h)}+\frac{2\sum_{h=1}^{N}S_{X_1(h)X_2(h+\Delta h)\cdot v_{12}(h)}}{\sum_{h=1}^{N}v_{12}(h)}$$
$$+\frac{2\sum_{h=1}^{N}\left(\overline{X}_1(h)-\hat{\mu}_{\overline{x}_1(h)}\right)\left(\overline{X}_2(h+\Delta h)-\hat{\mu}_{\overline{X}_2(h+\Delta h)}\right)v_{12}(h)}{\sum_{h=1}^{N}v_{12}(h)} \tag{16}$$

$$\frac{2\sum_{h=1}^{N}C\hat{o}v(X_1(h),\ X_2(h+\Delta h))\cdot v_{12}(h)}{\sum_{h=1}^{N}v_{12}(h)}\approx\frac{2\sum_{h=1}^{N}S_{X_1(h)X_2(h+\Delta h)}\cdot v_{12}(h)}{\sum_{h=1}^{N}v_{12}(h)} \tag{17}$$

$$Minimize_{\Delta\tau}\frac{\sum_{h=1}^{N}M\hat{S}E(h)_{X_1}\cdot v_1(h)}{\sum_{h=1}^{N}v_1(h)}+\frac{\sum_{h=1}^{N}M\hat{S}E(h+\Delta h)_{X_2}\cdot v_2(h)}{\sum_{h=1}^{N}v_2(h)}+\frac{2\sum_{h=1}^{N}S_{X_1(h),\ X_2(h+\Delta h)}\cdot v_{12}(h)}{\sum_{h=1}^{N}v_{12}(h)} \tag{18}$$

$$Var\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}Var(X_i)+2\sum_{1\le i<j\le n}Cov(X_i,\ X_j) \tag{19}$$

$$M\hat{S}E(H)_{X_1X_2\cdots X_L}\approx\frac{\sum_{h=1}^{N}M\hat{S}E(h)_{X_1}\cdot v_1(h)}{\sum_{h=1}^{N}v_1(h)}+\cdots\cdots+\frac{\sum_{h=1}^{N}M\hat{S}E(h+\Delta h_{L-1})_{X_L}\cdot v_L(h+\Delta h_{L-1})}{\sum_{h=1}^{N}v_L(h+\Delta h_{L-1})}$$
$$+\frac{2\sum_{h=1}^{N}S_{X_1(h)X_2(h+\Delta h_1)}\cdot v_{12}(h)}{\sum_{h=1}^{N}v_{12}(h)}+\cdots\cdots+\frac{2\sum_{h=1}^{N}S_{X_1(h)X_L(h+\Delta h_{L-1})}\cdot v_{1L}(h)}{\sum_{h=1}^{N}v_{1L}(h)}\cdots\cdots$$
$$+\frac{2\sum_{h=1}^{N}S_{X_{L-1}(h+\Delta h_{L-2}),X_L(h+\Delta h_{L-1})}\cdot v_L(h)}{\sum_{h=1}^{N}v_{(L-1)L}(h)}\cdots\cdots \tag{20}$$

$$MSE(h,\ k)=\frac{\sum_{i=1}^{v(h,k)}(x^i(h,\ k)-\hat{X}(h,\ k))^2}{v(h,\ k)} \tag{21}$$

where

$$\hat{X}(h, k) = f(\overline{X}(h, 0), \overline{X}(h, -1), \overline{X}(h, -2), \dots) \tag{22}$$

$$MSE(h, k) = \frac{\sum_{i=1}^{v(h,k)} (x^i(h, k) - \hat{X}(h, k))^2}{v(h, k)}$$

$$= \frac{\sum_{i=1}^{v(h,k)} (x^i(h, k) - \overline{X}(h, k) + \overline{X}(h, k) - \hat{X}(h, k))^2}{v(h, k)}$$

$$= \frac{\sum_{i=1}^{v(h,k)} (x^i(h, k) - \overline{X}(h, k))^2}{v(h, k)} + \frac{\sum_{i=1}^{v(h,k)} (x^i(h, k) - \overline{X}(h, k))(\overline{X}(h, k) - \hat{X}(h, k))}{v(h, k)} + \frac{\sum_{i=1}^{v(h,k)} (\overline{X}(h, k) - \hat{X}(h, k))^2}{v(h, k)}$$

$$= \frac{\sum_{i=1}^{v(h,k)} (x^i(h, k) - \overline{X}(h, k))^2}{v(h, k)} + (\overline{X}(h, k) - \hat{X}(h, k))^2 \quad (\textit{cross term is zero}) \tag{23}$$

$$MSE(h, P) = \frac{\sum_{k=1}^{K} MSE(h, k) \cdot v(h, k)}{\sum_{k=1}^{K} v(h, k)} \tag{24}$$

$$MSE(H, P) = \frac{\sum_{h=1}^{N} MSE(h, P) \cdot v(h, P)}{\sum_{h=1}^{N} v(h, P)} = \frac{\sum_{h=1}^{N} \sum_{k=1}^{K} MSE(h, k) \cdot v(h, k)}{\sum_{h=1}^{N} \sum_{k=1}^{K} v(h, k)} \tag{25}$$

$$Minimize_{\Delta\tau} \frac{\sum_{h=1}^{N} \sum_{k=1}^{K} MSE(h, k) \cdot v(h, k)}{\sum_{h=1}^{N} \sum_{k=1}^{K} v(h, k)} \tag{26}$$

$$MSE(h, k)_{X_1X_2} = \frac{\sum_{i=1}^{v_{12}(h,k)} (x^i(h, k) + x_2^i(h, k+\Delta h_1) - \hat{X}_1(h, k) - \hat{X}_2(h, k+\Delta h_1))^2}{v_{12}(h, k)}$$

$$\approx \frac{\sum_{i=1}^{v_1(h,k)} (x_1^i(h, k) - \hat{X}_1(h, k))^2}{v_1(h, k)} + \frac{2\sum_{i=1}^{v_{12}(h,k)} (x_1^i(h, k) - \hat{X}_1(h, k))(x_2^i(h, k+\Delta h_1) - \hat{X}_2(h, k+\Delta h_1))}{v_{12}(h, k)}$$

$$+ \frac{\sum_{i=1}^{v_2(h,k)} (x_2^i(h, k) - \hat{X}_2(h, k))^2}{v_2(h, k)}$$

$$= MSE(h, k)_{X_1} + MSE(h, k+\Delta h_1)_{X_2} + \frac{2\sum_{i=1}^{v_{12}(h,k)} (x_1^i(h, k) - \hat{X}_1(h, k))(x_2^i(h, k+\Delta h_1) - \hat{X}_2(h, k+\Delta h_1))}{v_{12}(h, k)} \tag{27}$$

$$B = s_{X_1(h,k)X(h,k+\Delta h_1)} + (\overline{X}_1(h, k) - \hat{X}_1(h, k))(\overline{X}_2(h, k+\Delta h_1) - \hat{X}_2(h, k+\Delta h_1)) \tag{28}$$

$$MSE(h, k)_{X_1X_2} = MSE(h, k)_{X_1} + M\hat{S}E(h, k+\Delta h_1)_{X_2}$$
$$+ 2S_{X_1(h,k)X_2(h,k+\Delta h_1)} + 2(\overline{X}_1(h, k) - \hat{X}_1(h, k))(\overline{X}_2(h, k+\Delta h_1) - \hat{X}_2(h, k+\Delta h_1)) \tag{29}$$

$$M\hat{S}E(h,P)_{X_1X_2} \approx \frac{\sum_{k=1}^{K} M\hat{S}E(h,k)_{X_1} \cdot v_1(h,k)}{\sum_{k=1}^{K} v_1(h,k)} + \frac{\sum_{k=1}^{K} M\hat{S}E(h,k+\Delta h_1)_{X_2} \cdot v_2(h,k)}{\sum_{k=1}^{K} v_2(h,k)} + \frac{2\sum_{k=1}^{K} S_{X_1(h,k)X_2(h,k+\Delta h_1)} \cdot v_{12}(h,k)}{\sum_{k=1}^{K} v_{12}(h,k)} \tag{30}$$

$$M\hat{S}E(H,P)_{X_1X_2} \approx \frac{\sum_{h=1}^{N}\sum_{k=1}^{K} MSE(h,k)_{X_1} \cdot v_1(h,k)}{\sum_{h=1}^{N}\sum_{k=1}^{K} v_1(h,k)} + \frac{\sum_{h=1}^{N}\sum_{k=1}^{K} MSE(h,k+\Delta h_1)_{X_2} \cdot v_2(h,k)}{\sum_{h=1}^{N}\sum_{h=1}^{K} v_2(h,k)} + \frac{2\sum_{h=1}^{N}\sum_{k=1}^{K} S_{X_1(h,k)X_2(h,k+\Delta h_1)} \cdot v_{12}(h,k)}{\sum_{h=1}^{N}\sum_{k=1}^{K} v_2(h,k)} \tag{31}$$

$$\frac{\sum_{i=1}^{v(h)}(x^i(h)-\overline{X}(h))^2}{v(h)} = \frac{\sum_{i=1}^{v(h)}(x^i(h)-E(x^i(h))+E(x^i(h))-\overline{X}(h))^2}{v(h)}$$
$$= \frac{\sum_{i=1}^{v(h)}(x^i(h)-E(x^i(h)))^2}{v(h)} + \frac{\sum_{i=1}^{v(h)}(x^i(h)-E(x^i(h)))(E(x^i(h))-\overline{X}(h))}{v(h)} + \frac{\sum_{i=1}^{v(h)}(E(x^i(h))-\overline{X}(h))}{v(h)} \tag{32}$$

$$\frac{\sum_{i=1}^{v(h)}(x^i(h)-\overline{X}(h))^2}{v(h)} \approx \frac{\sum_{i=1}^{v(h)}(x^i(h)-E(x^i(h)))^2}{v(h)} + \frac{\sum_{i=1}^{v(h)}(E(x^i(h))-\overline{X}(h))^2}{v(h)} \tag{33}$$

$$(\overline{X}(h)-E(\overline{X}(h)))^2 = \left[\frac{\sum_{i=1}^{v(h)} x^i(h)}{v(h)} - \frac{\sum_{j=1}^{s(h)} E(x^j(h))}{s(h)}\right]^2 \approx \left[\frac{\sum_{i=1}^{v(h)} x^i(h)}{v(h)} - \frac{\sum_{j=1}^{v(h)} E(x^i(h))}{v(h)}\right]^2$$
$$= \left[\frac{\sum_{i=1}^{v(h)}(x^i(h)-E(x^i(h)))}{v(h)}\right]^2$$
$$= \frac{\sum_{i=1}^{v(h)}(x^i(h)-E(x^i(h)))^2}{v(h)^2} + \frac{\sum_{i,j\,i\neq j}(x^i(h)-E(x^i(h))(x^j(h)-E(x^j(h)))}{v(t)^2}$$
$$= \frac{\sum_{i=1}^{v(h)} var(x^i(x^i(h))}{v(h)^2} + \frac{\sum_{i,j\,i\neq j} cov(x^i(h),x^j(h))}{v(t)^2} \tag{34}$$