Estimation of Vehicle Trajectories with Locally Weighted Regression

Tomer Toledo, Haris N. Koutsopoulos, and Kazi I. Ahmed

Vehicle trajectory data are important for calibrating driver behavior models (e.g., car following, acceleration, lane changing, and gap acceptance). The data are usually collected through imaging technologies, such as video. Processing these data may require substantial effort, and the resulting trajectories usually contain measurement and processing errors while also missing data points. An approach is presented to the processing of position data to develop vehicle trajectories and consequently speed and acceleration profiles. The approach uses local regression, a method well suited for mapping highly nonlinear functions. The proposed methodology is applied to a set of position data. The results demonstrate the value of the method to development of vehicle trajectories and speed and acceleration profiles. The conducted sensitivity analysis also shows that the method is rather robust regarding measurement errors and missing values.

The study of driving behavior, such as acceleration and lane changing, has important applications in microscopic traffic analysis as well as in modeling of safety, emissions, and several other applications. Some of the most important explanatory variables in driving behavior models are the state of the subject vehicle (e.g., position, speed, acceleration, lane changes) and its relation with other vehicles (e.g., relative speeds, time and space headways, lengths of gaps in traffic). Hence, estimation of driving behavior models relies on acquiring data that describe these variables.

Vehicle trajectory data, which consist of observations of the positions of vehicles at discrete points in time and at high fidelity (typically 1 s or less), provide useful information about these variables: speeds, accelerations, and lane changes may be inferred from the time series of positions; relations between the subject and other vehicles may be extracted from the data set of trajectories by simple arithmetic operations.

A wide range of collection and sensing technologies—such as aerial photography; video; laser, ultrasound, and microwave sensors; the Global Positioning System (GPS); and cellular location technologies—have been utilized to collect trajectory data. Regardless of the collection method, the raw data first need to be reduced and then processed to infer variables of interest. Despite recent advances in automating the data reduction task, it can be time-consuming and laborious. In particular, under congested conditions, manual processing may be required, since automated approaches fail to identify the vehicles reliably. As a result, there may be measurement errors as well as missing data points in the extracted data set.

A method is proposed here to perform the task of extracting useful information from position data efficiently while the ability to recover missing data points is retained. The method is based on smoothing of the data by using locally weighted regression (*1, 2, 3,* pp. 10–49).

PROBLEM DESCRIPTION

The raw data, usually collected through video technologies, includes observations of the positions of vehicles at discrete points in time in regular intervals. The raw data may contain measurement errors and may also miss data points. For example, it is possible to have two consecutive observations in which the position of a vehicle is decreasing. In most cases this position is not possible and if ignored it will lead to negative speeds. Missing observations are also common for various reasons (such as processing errors or occlusions), and so the position observations are not always consecutive.

The various applications discussed in the introduction, particularly estimation of microscopic traffic behavior models, require knowledge of not only the vehicles' positions but also their speeds and accelerations. For example, car-following models typically relate the instantaneous acceleration a vehicle applies at time t to a stimulus that occurred at time instant $t - \tau$, where τ is the reaction time of the associated driver, assumed to be a continuous random variable with a given distribution in the population of drivers. The stimulus is often a function of the vehicle's instantaneous speed and the spacing to the vehicle in front. Consequently, the applications of interest require as input instantaneous speeds and accelerations. Furthermore, the incorporation of reaction times in the models necessitates that the instantaneous positions, speeds, and accelerations be calculated at arbitrary points in time (and not only at the regular times when position observations are available). Estimation of these variables can be obtained by taking the derivatives of a continuous-in-time position function. Hence, the time series of position data needs to be processed to extract quantities of interest such as instantaneous speeds and accelerations.

Despite its importance, the problem of trajectory data processing has received little attention in previous applications. Speeds and accelerations are often directly extracted from the observations by subtraction of the position measurements in consecutive observations. However, these represent average values, whereas as noted earlier, most driving behavior models require instantaneous speeds and accelerations as dependent and independent variables. Furthermore, these

T. Toledo, Faculty of Civil and Environmental Engineering, Technion–Israel Institute of Technology, Haifa 32000 Israel. H. N. Koutsopoulos, Department of Civil and Environmental Engineering, Northeastern University, 437 Snell Engineering Center, Boston, MA 02115. K. I. Ahmed, Collaborative Consulting, 10 Tower Office Park, Suite 218, Woburn, MA 01801. Corresponding author: T. Toledo, toledo@technion.ac.il.

Transportation Research Record: Journal of the Transportation Research Board, No. 1999, Transportation Research Board of the National Academies, Washington, D.C., 2007, pp. 161–169.
DOI: 10.3141/1999-17

numerical differentiation methods tend to amplify measurement errors in the position data and so produce noisier speed and acceleration profiles.

Smith (*4*) reports on a large-scale effort to collect vehicle trajectory data. Position measurements were collected by using aerial photography. These observations are reported without any correction even though the report characterizes vehicle movements as jerky. Average speed, which is directly inferred from the difference between two consecutive position observations, is the only other variable reported. Wei et al. (*5*) developed VEVID, a tool that automates the trajectory data collection process. However, they did not perform any processing of position measurements; average speeds and accelerations were again calculated by simple subtraction. Ervin et al. (*6, 7*) used Kalman filtering to smooth trajectory data collected by using video cameras with the SAVME system. Punzo et al. (*8*) used a similar approach with data collected by using GPS technology. The Kalman filter was applied with average speeds that were derived from the raw data. Thus, the result was smoothed average measurements rather than instantaneous ones. Although the difference between the two types of measurements may be negligible for the very short time intervals (0.1 s) reported in their application, it is expected to be more significant for longer intervals. Furthermore, with the SAVME system, the Kalman filter is based on a model of the vehicle dynamics. However, the development of such model is, in most cases, the purpose of the data collection effort.

METHODOLOGY

A standard approach to the problem of trajectory function estimation would be to fit a global polynomial curve to the position observations and use the first and second derivatives of this function as estimates of instantaneous speeds and accelerations, respectively. However, under congested traffic conditions, vehicles frequently stop, often for significant durations. Furthermore, the instances at which a vehicle is stopped cannot be directly identified from the observed trajectory, since measurement errors occur while the data are collected and reduced. Therefore, a very high order polynomial would be necessary to fit a curve to the trajectories of such vehicles. This requirement gives rise to computational and numerical difficulties since the Hessian of the objective function of such problems becomes nearly singular (for polynomials of time of order 10 or above, the powers of time-independent variables vary from tens to billions). A nearly singular Hessian makes the estimation process computationally intensive and time-consuming since the convergence rate reduces significantly. It also invokes problems related to the precision of the computer. Furthermore, even a high-order polynomial may not fit the data well during the instances when a vehicle is stopped. Finally, high-order polynomials may fit the observations correctly, but they may oscillate considerably between successive observations, leading to unrealistic behavior.

Instead of fitting a global curve to the observations, a methodology is proposed to smooth position measurements and estimate continuous trajectory, speed, and acceleration profiles based on fitting a local curve at the points of interest. The methodology consists of two steps, which are repeated for each vehicle:

1. Estimation of a smooth time-continuous trajectory function from the discrete position observations by using weighted local regression and
2. Estimation of instantaneous speeds and accelerations by differentiating the fitted trajectory function.

Step 1 consists of the application of local regression by using the data in the neighborhood of the point of interest. Local regression, which fits a local curve to each point of interest using the observations around it, is designed to replace standard regression estimates when one is dealing with data that require a flexible functional form.

Cleveland (*1*), Cleveland and Devlin (*2*), and Cleveland and Loader (*3,* pp. 10–49) discuss the concept, properties, and computational algorithms for local regression. Cleveland and Devlin (*2*) also report various application areas of the method, such as support for exploratory graphical data analysis, provision of additional regression diagnostics for testing parametric models fitted to the data, and direct use of the local regression functions in place of parametric functions.

Recently, the method has become popular in the machine-learning community. It is used as a form of memory (or instance) based learning to learn continuous nonlinear mappings in applications such as learning robot dynamics and process models (*9*). In the transportation literature, Sun et al. (*10*) applied local regression for short-term traffic forecasting. They report that local regression was superior when compared with nearest-neighborhood and kernel smoothing methods.

In the context of the current application of estimating vehicle trajectories, the local regression estimator is defined as follows: $x(t)$, $t = 1, \ldots, T,$ denotes the time series of measurements of the position of a given vehicle (the vehicle index is omitted for simplicity). At a point t_0, a local trajectory function is estimated by using only observations in the neighborhood of t_0. N denotes the number of observations in the neighborhood (window) around t_0 that are used in the estimation.

The trajectory function in the neighborhood of t_0 is assumed to be a function of time:

$$x(t) = f_{t_0}\left(t,\ \boldsymbol{\beta}_{t_0}\right) + \epsilon_{t_0,t} \tag{1}$$

where

$f_{t_0}(t, \boldsymbol{\beta}_{t_0})$ = fitted position at time t estimated by local regression function centered at time t_0,
$\boldsymbol{\beta}_{t_0}$ = vector of parameters of fitted curve to be estimated, and
$\epsilon_{t_0,t}$ = normally distributed error terms.

Local regression then uses weighted least-squares estimation of the parameters of the local function $f_{t_0}(t, \boldsymbol{\beta}_{t_0})$ with the N observations in the window around t_0. The observation weights are usually based on some measure of the time difference between the observation and t_0. Hence, the problem of applying local regression to position data in order to develop a local trajectory function centered at t_0 is formulated as a minimization problem:

$$\min_{\boldsymbol{\beta}_{t_0}} \left[\mathbf{X}_{t_0} - f_{t_0}\left(\mathbf{t}, \boldsymbol{\beta}_{t_0}\right)\right]' \mathbf{W}_{t_0} \left[\mathbf{X}_{t_0} - f_{t_0}\left(\mathbf{t}, \boldsymbol{\beta}_{t_0}\right)\right] \tag{2}$$

where

$\mathbf{X}_{t_0}$ = column vector of N position observations used to estimate a trajectory function centered on t_0,
$f_{t_0}(\mathbf{t}, \boldsymbol{\beta}_{t_0})$ = corresponding vector of fitted values, and
$\mathbf{W}_{t_0}$ = $[N \times N]$ diagonal matrix, with elements corresponding to weights of observations used for local estimation.

In summary, the application of local regression requires selection of three basic elements:

1. Function specification through selection of the form of $f_{t_0}(\cdot)$, which defines the shape of the locally fitted curves;
2. Window size, which determines the number of neighboring points used in fitting each measurement; and
3. Weight assignment for each point within the local regression window.

The use of a polynomial function to specify the local regression curve is common in the literature. With polynomial fitting, the specification question becomes one of choosing the polynomial order. The function specification and the window size, *N,* affect the bias and variance of the estimated positions in opposite directions (*11*): the bias increases with increasing window size and decreases with the polynomial order. Inversely, the variance decreases with the window size and increases with the polynomial order. The mean squared error of the estimates combines both these statistics and may be minimized to obtain optimal window size and polynomial order choices.

Various weight functions, $w(u)$, where u is a function of the time difference between the point of interest and the observation that is used in the regression analysis, have been proposed in the literature. Cleveland and Loader (*3,* pp. 10–49) point out that smooth weight functions lead to smoother estimates. The chosen function should also assign higher weights to observations that are closer to the point of interest and so has to satisfy the following conditions:

$$
\begin{aligned}
&w(u) \geq 0 \\
&w(u) = 0 \text{ for } u \geq 1 \\
&w(u) \text{ is nonincreasing for } u \geq 0
\end{aligned} \tag{3}
$$

Assuming that the foregoing requirements are met, the choice of the exact functional form does not seem to have a significant impact on the results (*3,* pp. 10–49). Cleveland et al. (*11*) recommend a tricube weight function. According to this function, the weight assigned to each observation t depends on the normalized time difference, u, between t and the point of interest t_0:

$$w(t_0, t) = \left(1 - u(t_0, t)^3\right)^3 \tag{4}$$

where

$w(t_0, t)$ = weight assigned to the observation at time t in fitting a curve centered at t_0,

$u(t_0, t)$ = normalized measure of the time difference between t and t_0 given by

$$u(t_0, t) = \frac{|t - t_0|}{d} \tag{5}$$

and

d = distance from t_0 to the nearest point outside the window of N points to be considered in fitting the curve.

The shape of the tricube weight function is shown in Figure 1. It should be noted that $w(t_0, t)$ decreases as the time difference between t and t_0 increases and that $w(t_0, t_0) = 1$. For a symmetric window, N, about t_0, $d = (N + 1)/2$.

In the second step of the process, the fitted value at time t_0 is used as an estimate of the position at that time. The first and second derivatives of the fitted polynomial are used as estimates to the instantaneous speed and acceleration, respectively:

$$\hat{x}(t_0) = f_{t_0}\left(t_0, \hat{\boldsymbol{\beta}}_{t_0}\right) \tag{6}$$

$$\hat{v}(t_0) = \left.\frac{df_{t_0}\left(t, \hat{\boldsymbol{\beta}}_{t_0}\right)}{dt}\right|_{t=t_0} \tag{7}$$

$$\hat{a}(t_0) = \left.\frac{d^2 f_{t_0}\left(t, \hat{\boldsymbol{\beta}}_{t_0}\right)}{dt^2}\right|_{t=t_0} \tag{8}$$

where $\hat{x}(t_0)$, $\hat{v}(t_0)$, and $\hat{a}(t_0)$ are the estimated position, instantaneous speed, and acceleration at time t_0, respectively, and $\hat{\boldsymbol{\beta}}_{t_0}$ is the parameter estimates of the local regression curve fitted around t_0.

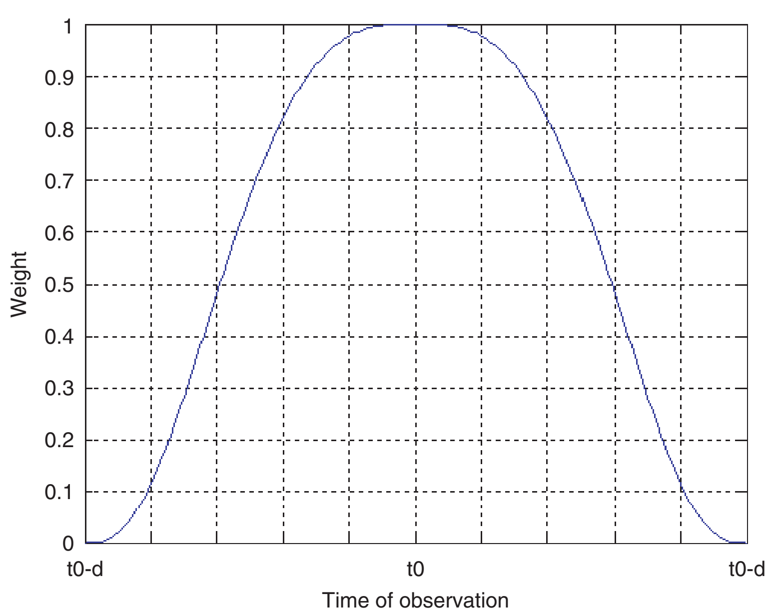

FIGURE 1 Tricube weight function.

APPLICATION

The locally weighted regression procedure was applied to a trajectory data set collected by Hasan et al. (*12*) from a section of I-93 in Boston (Figure 2).

Data were collected under heavily congested conditions. Vehicle speeds ranged from 10.5 km/h to 54.3 km/h with a mean speed of 22.6 km/h and standard deviation of 7.7 km/h. Traffic density in the section was 48.5 vehicles/km/lane [Level of Service F, on the basis of *Highway Capacity Manual* criteria (*13*)]. The data set consists of measurements of the positions of 653 vehicles at a 1-s resolution. The total number of observations in the data set is 20,795 (average of 31.8 observations per vehicle). These observations were obtained from the analysis of video with the automatic and manual features of the traffic image-processing software ViVAtraffic (*14*). This case study is used to illustrate the potential benefits of the proposed method and to examine the sensitivity of the local regression estimates to the choices of window size and polynomial order. In addition, the consistency of the estimates and the ability of the proposed method to recover missing observations and estimate data values for these points is examined.

The locally fitted trajectory function is assumed to be a polynomial function in time:

$$f_{t_0}\left(t,\ \boldsymbol{\beta}_{t_0}\right)=\mathbf{Z}(\mathbf{t})\boldsymbol{\beta}_{t_0}=\sum_{m=0}^{M}\boldsymbol{\beta}_{t_0,m}(t)^m \tag{9}$$

where

- $\mathbf{Z}(\mathbf{t})$ = vector of independent variables corresponding to the observation at time t, including the polynomial in time-independent variables, $\mathbf{Z}(\mathbf{t}) = [1 \quad t \quad t^2 \quad t^3 \quad \dots \quad t^M]$;
- M = order of the polynomial to be estimated; and
- $\boldsymbol{\beta}_{t_0} = [\beta_{t_0,0} \quad \beta_{t_0,1} \quad \beta_{t_0,2} \quad \dots \quad \beta_{t_0,M}]$ = vector of the $M + 1$ parameters of the polynomial function estimated around time t_0.

As mentioned earlier, the local regression procedure serves not only to smooth the data but also to estimate a continuous trajectory function, which may be used subsequently to derive the subject's instantaneous speeds and accelerations at arbitrary points in time. However, because of measurement errors, the observed position of a vehicle at two successive time periods may be decreasing. Hence, a curve fitted to these points may yield an unrealistic (negative) speed or acceleration estimate, or both. In order to ensure that the estimated speeds and accelerations are acceptable, suitable constraints may be added to the weighted least-squares formulation (Equation 2). These constraints include

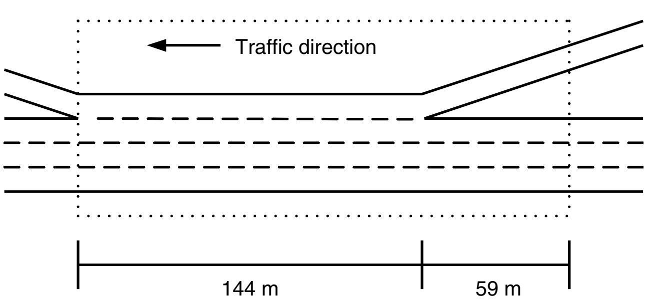

FIGURE 2 I-93 data collection site.

1. Nonnegativity speed constraints,
2. Upper bounds on the speed, and
3. Upper (acceleration) and lower (deceleration) bounds on the vehicle's acceleration to reflect vehicle performance characteristics.

Choice of Window Size and Polynomial Order

The choice of the window size, N, and the polynomial order, M, affect the bias and variance of the estimated trajectory. However, in this application a closed-form solution for estimating the bias (or variance) does not exist since the curve fit is constrained. Instead, a sensitivity analysis can be conducted to evaluate the impact of the window size and polynomial order on the quality of the results. In addition, two other considerations should be taken into account for the determination of the window size:

1. The window size should include an odd number of points to ensure that it is symmetric about t_0.
2. The minimum window size should be sufficient to adequately estimate the derived speeds and accelerations. The window size N bounds the order of the polynomial, $M \le N - 1$. In the case here, the minimum polynomial degree is dictated by the interpretation of the variables of interest. If, for example, the window size is equal to 5, the order of the polynomial cannot exceed 4. As a result, the order of the polynomial representing the acceleration profile would be 2, since the second derivative of the trajectory function is the acceleration function. This implies that the curvature of the acceleration profile (its second derivative) is restricted to be a constant, which may not be realistic. Therefore, in the discussion that follows, $N = 7$ was used as the minimum window size.

Different combinations of window size and polynomial order were used to test the sensitivity of the quality of the results to these factors. The differences between the original and estimated positions were quantified by using the mean absolute error (MAE) and root-mean-squared error (RMSE) statistics:

$$\text{MAE}=\frac{\sum_{l=1}^{L}\left|x_l-\hat{x}_l\right|}{L} \tag{10}$$

$$\text{RMSE}=\sqrt{\frac{\sum_{l=1}^{L}\left(x_l-\hat{x}_l\right)^2}{L}} \tag{11}$$

where x_l and $\hat{x}_l$ are the observed and estimated positions for observation l, and L is the number of observations in the data set.

The results for MAE and RMSE are summarized in Table 1. Given that the measurement accuracy in the original data was estimated at ±1.4 m (*12*), the magnitude of errors is within a reasonable range. For a given polynomial order, the deviation from the observed positions increases with the window size. Similarly, for a given window size, a larger polynomial order more closely matches the observations. The combined effect of these impacts results in very small differences among regression line estimates for the various window sizes when the maximum possible polynomial size was used. For the larger polynomial order, additional inaccuracies may be introduced as the objective function Hessian approaches singularity. When a window

TABLE 1 MAE and RMSE for Different Window Sizes and Polynomial Order Values

Window Size / Polynomial Order	7	9	11	13
6	0.033 0.190	0.284 0.403	0.339 0.475	0.370 0.517
8	—	0.037 0.196	0.273 0.388	0.327 0.461
10	—	—	0.039 0.200	0.263 0.375
12	—	—	—	0.042 0.206

NOTE: Top row in each cell: MAE (m); bottom row: RMSE (m).

size is selected, it is also important to consider that the number of observations for which the full window is available decreases with the window size.

The estimated position, speed, and acceleration profiles were also evaluated. Figure 3 compares the estimated positions for different window sizes and the observed positions. The results illustrate a good fit between the estimated and observed positions in all cases. The estimated curves eliminate the negative change in position in the observed data at Time Period 31. The vehicle was also stopped for a few seconds (Time Period 10 to 15 approximately). The maximum polynomial order was used for each window size.

Figures 4 and 5 show the corresponding estimated speed and acceleration profiles, respectively. The results indicate that there is only slightly more variability in the speed and acceleration profiles for the different window sizes. Computational inaccuracies due to the high-order polynomial may have also contributed to these differences. For polynomials of order 14 and higher, near-singular Hessian functions led to numerical inaccuracies, and therefore these estimates were omitted from further analysis.

The observed and estimated positions of another vehicle (Figure 6) demonstrate the ability of the locally weighted regression process to recover from measurement errors. The observed positions in Time Periods 15, 16, and 17 (90.47, 90.45, and 88.39 m, respectively) suggest that the vehicle is moving backward; hence it has a negative average speed. Given the nature of the location at which the data were collected, this behavior is not possible. The application of local regression, with estimated positions (89.73, 89.85, and 91.34 m, respectively), eliminated the problem. The corresponding instantaneous speeds are estimated at 1.55, 0.35, and 3.88 m/s. A window of Size 9 was used in this case.

Consistency of Fitted Values

Punzo et al. (*8*) raise the issue of consistency of the positions, speeds, and accelerations estimated from the trajectory data. They point out that the observations made over time should satisfy basic equations of motion. In their application they examine the consistency of the spacing between two vehicles and their speeds. However, in more general settings, where more than two vehicles are present, and considering that reaction time is an important factor in driving models, it may be more useful to evaluate the internal consistency of the trajectory of a vehicle:

$$\hat{x}(t) = x(0) + \int_0^t \hat{v}(s)ds \qquad (12)$$

$$\hat{v}(t) = v(0) + \int_0^t \hat{a}(s)ds \qquad (13)$$

Tables 2 and 3 present the MAE and RMSE, respectively, of consistency errors for the vehicle position with respect to its speed and for the vehicle speed with respect to its acceleration. The position and speed consistency errors are calculated as the difference between the right-hand and the left-hand sides of Equations 12 and 13,

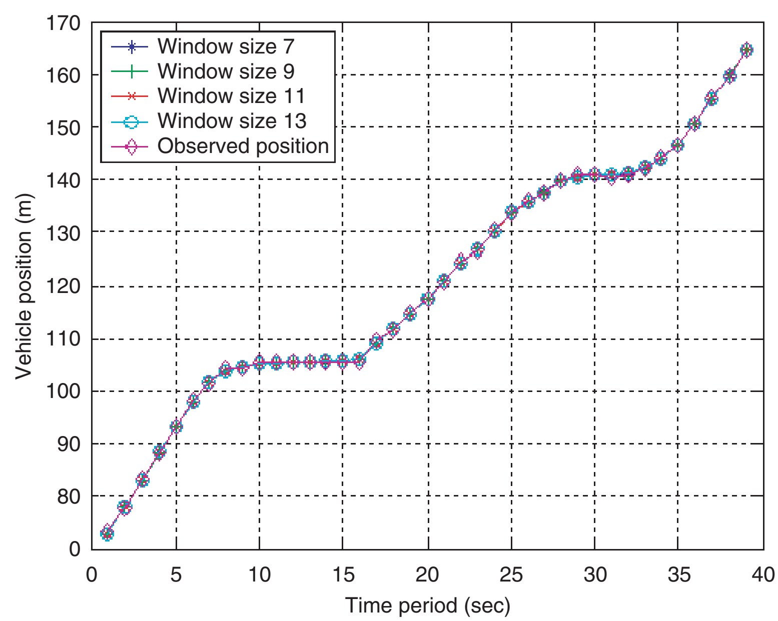

FIGURE 3 Observed and estimated position profiles with different window sizes.

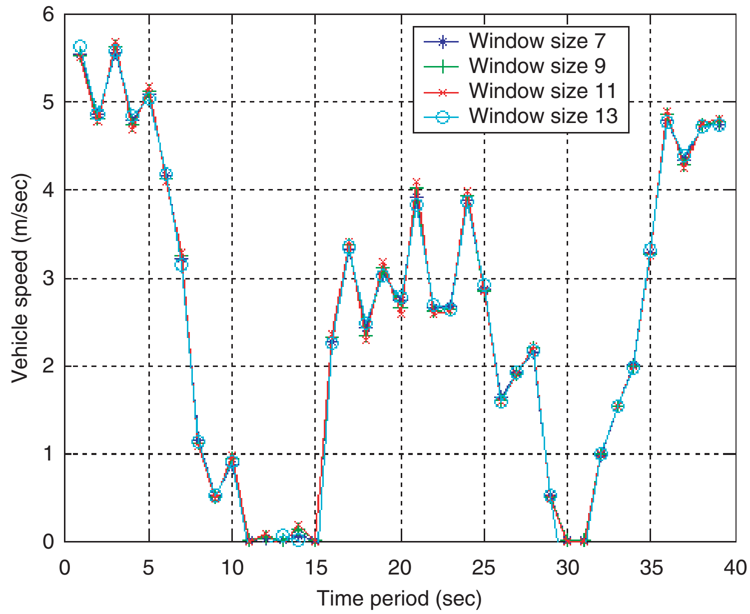

FIGURE 4 Estimated speed profiles with different window sizes.

respectively. The results indicate absolute inconsistencies of up to 0.223 m in the positions and 0.345 m/s in the speeds. These values are significantly smaller compared with the magnitude of the measurement errors. The inconsistencies decrease with an increase in the window size, which enables the fitted curve to better represent the temporal evolution of the trajectory. However, the inconsistencies generally increase with higher orders of the fitted polynomial, which provide more flexibility for sharp changes in the fitted curve and for discontinuities in the transitions from one local curve to the next. The inconsistencies reported here are lower by one or two orders of magnitude compared with those reported by Punzo et al. (*8*), who evaluated the consistency of the spacing between two vehicles and their speeds that were obtained with this method for another data set. In addition to differences in the characteristics of the raw data and in the way the consistency was evaluated, two other factors may be contributing to the difference in consistency results. First, the smoothed instantaneous speeds and accelerations were incorrectly used as if they were average values. Second, the consistency was evaluated compared with positions estimated by another approach and not with the values estimated with the local regression method.

Sensitivity to Missing Observations

As discussed in previous sections, a significant fraction of observations may be missing from the data set. For example, some observations

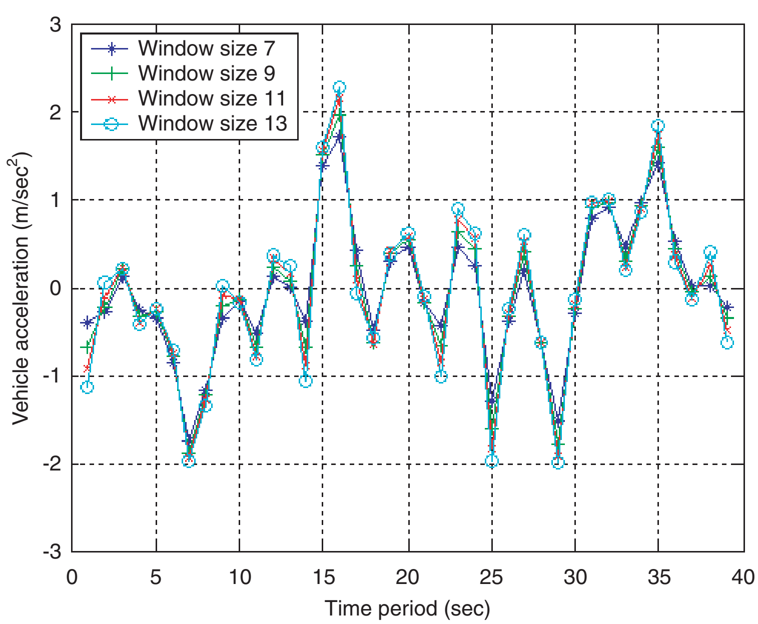

FIGURE 5 Estimated acceleration profiles with different window sizes.

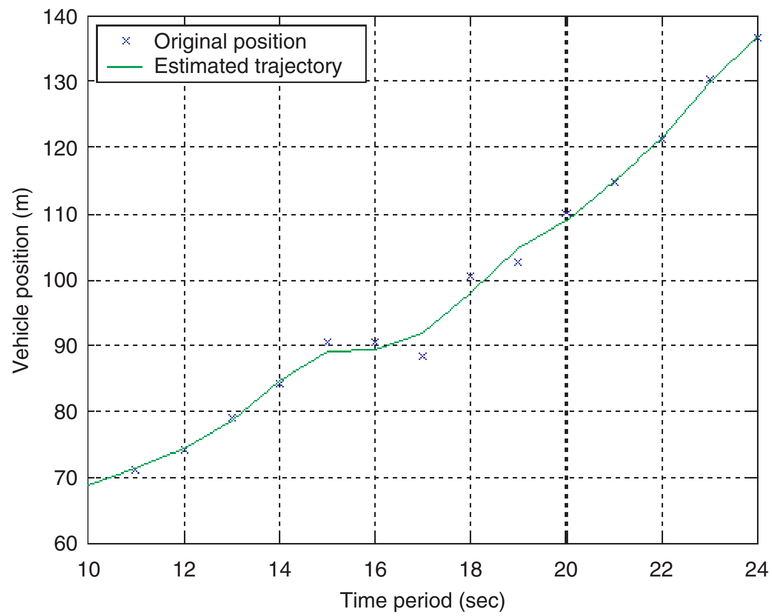

FIGURE 6 Estimated position profile for observations with measurement errors.

may be lost because of oversight in the, often manual, process of reduction of video (or other) images. In some cases a vehicle may not be visible—if, for example, it is obscured by another vehicle or by elements of the infrastructure (e.g., an overpass)—and therefore its position cannot be ascertained. Local regression has the potential to be used to estimate values of variables of interest for points where observations are missing.

To evaluate the ability of the method to recover missing observations, a reduced data set was created by eliminating a fraction of the observations. Local regression was then applied to the incomplete data set. Observations were eliminated in two ways: randomly and deterministically. In random elimination, a prespecified fraction, f, of the observations was removed in a random fashion. In the deterministic elimination, observations were removed in a systematic way: one every $1/f$ observations, where f is the fraction of observations to be eliminated. It should be noted that situations in which several consecutive observations are missing may arise in data collection. The approach used here, as well as by other local estimators, may not be appropriate for such cases, which require larger window sizes.

The impact of the fraction of missing observations is quantified by the deviation of the estimated positions in the reduced data set from the observed positions (raw data set) and from the estimates obtained with the full data set. Figures 7 and 8 summarize the results for various fractions of missing observations. A window of Size 9 was used with the maximum available polynomial order for each observation.

The results indicate that the local regression procedure can successfully recover missing observations. The mean absolute deviation from the raw data is 12 cm when the fraction of missing observations is 10%. For the same fraction, the deviation from the estimates with the full data set is less than 10 cm and about 45 cm with 50% missing observations. These values are well within the ±1.4-m measurement error associated with the data collection. Moreover, deviations from both the raw and the estimated data are significantly lower when observations are omitted systematically.

The results suggest another way in which the local regression procedure may be utilized. Extraction of position data from video images is an expensive, time-consuming task and in many cases it limits the amount of data that can be practically processed. However, through the local regression procedure it is possible to still obtain reliable data for various applications by processing only a fraction of the available frames.

TABLE 2 MAE and RMSE of Consistency for Different Window Sizes and Polynomial Order Values

Window Size Polynomial Order	7	9	11	13
6	0.223 0.345	0.074 0.201	0.035 0.139	0.023 0.116
8	—	0.178 0.306	0.073 0.198	0.034 0.137
10	—	—	0.150 0.278	0.072 0.196
12	—	—	—	0.003 0.040

TABLE 3 RMSE and MAE of Inconsistency for Different Window Sizes and Polynomial Order Values

Window Size Polynomial Order	7	9	11	13
6	0.148 0.311	0.107 0.270	0.089 0.245	0.082 0.235
8	—	0.147 0.304	0.186 0.281	0.096 0.254
10	—	—	0.147 0.297	0.126 0.285
12	—	—	—	0.094 0.262

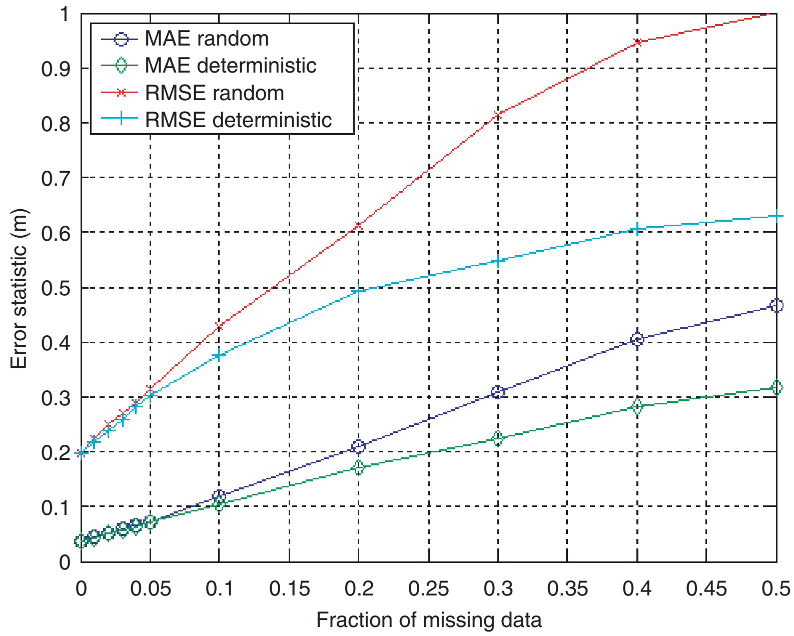

FIGURE 7 Impact of missing observations on error relative to raw observations.

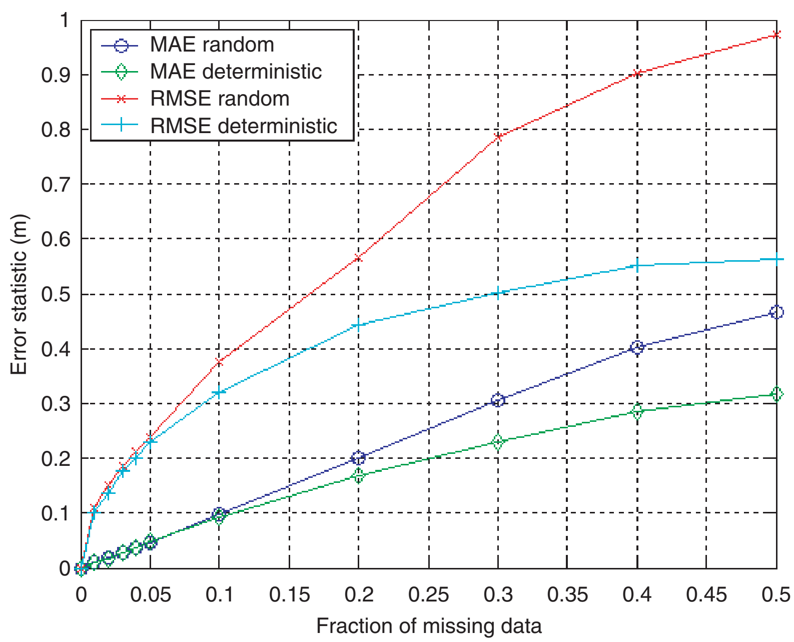

FIGURE 8 Impact of missing observations on error relative to estimated measurements with full data set.

CONCLUSION

Trajectory data, which consist of observations of the positions of vehicles at discrete points in time, are useful to infer variables that may explain driving behavior. A methodology to improve the quality of trajectory data and to estimate instantaneous speeds and accelerations was presented. The methodology is based on estimating a trajectory function by using locally weighted regression. Local regression is particularly useful to map highly nonlinear functions and allows for estimating continuous position, speed, and acceleration profiles at arbitrary points in time.

The method was applied to a data set of second-by-second position observations extracted from video. The results of the case study demonstrate the usefulness of the method. Furthermore, it is robust with respect to errors in the data and missing observations.

ACKNOWLEDGMENTS

The first author is a Horev Fellow supported by the Taub and Shalom Foundations. His work was partially supported by the European Commission under a Marie Curie grant. The authors thank the anonymous referees for their comments, which helped improve the quality of this paper.

REFERENCES

1. Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association,* Vol. 74, 1979, pp. 829–836.
2. Cleveland, W. S., and S. J. Devlin. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting. *Journal of the American Statistical Association,* Vol. 83, 1988, pp. 596–610.
3. Cleveland, W. S., and C. Loader. Smoothing by Local Regression: Principles and Methods. In *Statistical Theory and Computational Aspects of Smoothing* (W. Hardle, and M. G. Schimek, eds.), Springer, New York. 1995.
4. Smith, S. *Freeway Data Collection for Studying Vehicle Interactions.* Technical Report FHWA/RD-85/108. FHWA, U.S. Department of Transportation, 1985.
5. Wei, H., C. Feng, E. Meyer, and J. Lee. Video-Capture-Based Methodology for Extracting Multiple Vehicle Trajectories for Microscopic Simulation Modeling. Presented at 78th Annual Meeting of the Transportation Research Board, Washington, D.C., 1999.
6. Ervin, R. D., C. C. MacAdam, K. Gilbert, and P. Tchoryk, Jr. Quantitative Characterization of the Vehicle Motion Environment (VME). *Proc., 2nd Vehicle Navigation and Information Systems Conference,* 1991, pp. 1011–1029.
7. Ervin, R., C. MacAdam, J. Walker, S. Bogard, M. Hagan, A. Vayda, and E. Anderson. *System for Assessment of the Vehicle Motion Environment (SAVME).* Research Report UMTRI-2000-21-1. Transportation Research Institute, University of Michigan, Ann Arbor, 2000.
8. Punzo, V., D. J. Formisano, and V. Torrieri. Nonstationary Kalman Filter for Estimation of Accurate and Consistent Car-Following Data. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1934,* Transportation Research Board of the National Academies, Washington, D.C., 2005, pp. 3–12.
9. Atkenson, C. G., A. Moore, and S. Schaal. Locally Weighted Learning. *AI Review,* Vol. 11, 1997, pp. 11–73.
10. Sun, H., H. X. Liu, H. Xiao, R. R. He, and B. Ran. Use of Local Linear Regression Model for Short-Term Traffic Forecasting. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1836,* Transportation Research Board of the National Academies, Washington, D.C., 2003, pp. 143–150.
11. Cleveland, W. S., S. J. Devlin, and E. Grosse. Regression by Local Fitting: Methods, Properties and Computational Algorithms. *Journal of Econometrics,* Vol. 37, 1988, pp. 87–114.
12. Hasan, M., D. Cuneo, and A. Chachich. Analysis of Traffic Video to Develop Driver Behavior Models for Microscopic Traffic Simulation. Presented at IEEE Conference on Intelligent Transportation Systems ITSC 97, Boston, Mass., 1997.
13. *Highway Capacity Manual.* TRB, National Research Council, Washington, D.C., 2000.
14. *ViVAtraffic: The Video Traffic Analysis System.* 2003. http://transport.arubi.uni-kl.de/ViVAtraffic.

The Traffic Flow Theory and Characteristics Committee sponsored publication of this paper.