Smoothing Methods to Minimize Impact of Global Positioning System Random Error on Travel Distance, Speed, and Acceleration Profile Estimates

Jungwook Jun, Randall Guensler, and Jennifer H. Ogle

The Georgia Institute of Technology is evaluating the feasibility and effectiveness of mileage-based pricing programs as transportation control measures. Incentives were provided to study participants who change driving behavior in response to cent per mile pricing (fixed pricing and pricing as a function of congestion level). In-vehicle Global Positioning System (GPS) devices were used to estimate distance traveled and driver behavior (e.g., speed and acceleration profiles). The accuracy of estimated mileage accrual speeds by road classification, and acceleration rates used in pricing algorithms, is paramount. Various data-smoothing techniques were applied to the instrumented vehicle GPS speed data, and performance of the algorithms was evaluated in minimizing the impact of GPS random error on speed, acceleration, and distance estimates. The conventional discrete Kalman filter algorithm was modified to enhance its ability to control GPS random errors. Each smoothing method produces different second-by-second speed and acceleration profiles (*t*-test and χ² tests) except for the Kalman filters. The techniques provided different travel distance estimates, but the modified Kalman filter was the most accurate compared with distance estimates from the onboard vehicle speed sensor monitor. The modified Kalman filter is the recommended technique for smoothing GPS data for use in pricing studies. Additional smoothing methods will be evaluated as they are identified.

Most transportation-related problems, including traffic congestion, crash frequency, energy consumption, and vehicle emissions, are directly related to vehicle usage rates and driver behavior. To encourage drivers to use vehicles more efficiently and to change driving behavior, a number of incentive programs (commute options, transit and rideshare, parking cash-out, congestion pricing, and value pricing of insurance) are being evaluated as potential transportation demand management strategies.

Among these incentive programs, pay-as-you-drive (PAYD) insurance and variable congestion tolls have received increased attention from planners and transportation policy makers because the program will likely reduce vehicle usage rates and improve driver behavior to achieve safety benefits. Plus, on the average, such pricing programs

should provide significant benefits to consumers through reduced insurance premiums. In implementing future programs, tracking of mileage and location of travel will be variables (*1*). Consequently, future use of Global Positioning System (GPS) data beyond current freight logistics applications is likely to be instrumental to implementation of the most refined pricing programs. The accuracy of estimated mileage accrual, speeds by road classification, and even acceleration rates based on GPS data becomes paramount.

PAYD insurance programs are expected to assess insurance premiums based on travel distance and driving speed. For example, the Progressive Casualty Insurance Corporation (Progressive) in the United States and Norwich Union of England currently use information on travel time, travel distance, and speed in the insurance premium structure (*2, 3*). The eventual goal of PAYD programs is to evaluate a driver's potential crash risk and to set premiums that are proportional to such risk, including both probabilities coupled with damage functions. Hence, insurance companies and customers need to ensure that reliable data are used in such programs.

To collect data on vehicle activity and driver behavior, various data measurement devices such as the distance measurement instrument, the onboard diagnostics (OBD) system, and the GPS can be used. Among these devices, the GPS has been the most common choice in transportation research (including PAYD programs), because it provides more useful data, such as travel routes, start-and-stop points of a trip, travel time, speed, and acceleration rates.

Although an accurate data measurement device, as indicated in previous studies (*4, 5*), the GPS remains subject to various systematic and random errors:

- Systematic errors may be due to a low number of satellites, a relatively high position dilution of precision (PDOP) value, which relates to satellite orientation on the horizon and the impact on position precision, and other parameters (for example, antenna placement) that affect precision and accuracy of the device used (*6*).
- Random errors may result from satellite orbit, clock and receiver issues, atmospheric and ionospheric effects, multipath signal reflection, and signal blockage (*4, 5*).

Whereas systematic errors can be readily identified and removed, random errors are more difficult to address. Depending on how GPS data will be used, and depending on the magnitude of the random error effect, it may be necessary to process the GPS data to minimize the effects of random error for some processes in which the data will be used. Although in smaller research efforts GPS errors can be

J. Jun and R. Guensler, School of Civil and Environmental Engineering, Georgia Institute of Technology, 790 Atlantic Drive Atlanta, GA 30332-0355. J. H. Ogle, Department of Civil Engineering, Clemson University, 208 Lowry Hall, Clemson, SC 29634-0911.

identified through visual inspection of the data, in deployments that yield large GPS data sets, visual inspection is not practical. Because of significant data-processing time, automated analysis techniques are required. Statistical smoothing techniques may be useful processing tools because not only are they designed to decrease the impact of random errors on the results of the study, they also require less time for detecting random errors than visual inspection.

Statistical smoothing techniques can be categorized by their statistical backgrounds into three types: (*a*) to minimize overall error terms, (*b*) to adjust the probability of occurrence, and (*c*) to perform a feedback system recursively. Although each approach is capable of detecting random errors in the GPS data profiles, given their different statistical backgrounds, each technique can result in different outputs. Thus, before adopting a specific smoothing technique for identifying random errors in the GPS data profile, researchers need to understand their characteristics better. This study describes the characteristics of three smoothing techniques that are popularly used in a variety of traffic-related research and also have different statistical algorithms or backgrounds: the least-squares spline approximation, the kernel-based smoothing method, and the Kalman filter.

• The least-squares spline approximation minimizes the residual sum-of-squared errors (RSS) and has a statistical background similar to regression-based smoothing techniques such as the local polynomial regression, cubic fits, robust exponential smoothing, and time-series models.
• The kernel-based smoothing method adjusts the probability of occurrences in the data stream to modify outliers and has the same statistical background as nearest-neighbor smoothing and locally weighted regression models.
• The Kalman filter smoothes data points by recursively modifying error values.

This study evaluates one smoothing method within each general category of smoothing techniques. Each smoothing technique was applied to a large GPS data set collected in Atlanta, Georgia, and then was comparatively evaluated for the impact on estimated speeds, accelerations, and travel distance profiles. The researchers believe the three general smoothing approaches examined are representative of each general statistical approach, although they are not exhaustive.

DATA COLLECTION PROCESS

The DRIVE Atlanta Laboratory at the Georgia Institute of Technology (Georgia Tech) developed a wireless data collection system known as the Georgia Tech (GT) Trip Data Collector (TDC). The GT-TDC collects second-by-second vehicle activity data, including vehicle position (latitude and longitude via GPS) and vehicle speed. In addition, the GT-TDC collects 10 engine operating parameters from the OBD system in post-1996 model year vehicles and monitors vehicle speed at 4 Hz from the vehicle speed sensor (VSS). (Thus, the VSS and OBD systems were not installed in all vehicles in the commute Atlanta program.) The data were integrated into trip files, encrypted, and transmitted to the central server system at Georgia Tech using a wireless data transmit system via a cellular connection. Figure 1 presents the appearance of the GT-TDC and its accessories.

The GT-TDCs were installed in about 500 light-duty vehicles through the commuter choice and value pricing insurance incentive program (Commute Atlanta). To evaluate the filtering techniques, this study used GPS data gathered between October and November
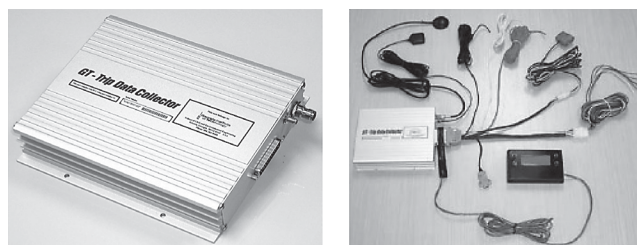

FIGURE 1   GT-TDC.

2004 from seven vehicles that generated 1,702 trips (1,497,066 data points).

CAPABILITY OF GPS RECEIVER IMPLEMENTED IN GT-TDC

The GT-TDC integrates the 12-channel SiRF Star II GPS receiver, which is designed for in-car navigation systems. This receiver was selected for the Commute Atlanta program in 2002 when a previous study conducted by Ogle et al. (*4*) found that this GPS receiver provided performance for collecting vehicle speed and acceleration similar to the DGPS receiver once selected availability was eliminated in 2000. The SiRF Star II GPS receiver calculates the vehicle location based on the C/A code communicated between satellites and the receiver and separately estimates vehicle speed using the Doppler effect (vehicle speed is independent of vehicle location). Although real-time kinematic (RTK) GPS systems can resolve uncertainty in vehicle location and speed estimates, there are four reasons why the researchers could not implement the PTK-GPS system in the GT-TDC:

• RTK-GPS equipment is too costly for use in large deployments (the Commute Atlanta program recruited about 500 vehicles).
• RTK-GPS systems require subdevices (two or more GPS antennae, a rover radio, a base station radio, a base station GPS antenna, and a rover receiver as well as additional base stations) and equipment packages that need to be small and self-contained.
• RTK-GPS systems typically require the onboard GPS receiver to be within a boundary of 6 mi (10 km) from the base station with a line of sight between the reference receiver and the rover receiver (*7*) (which would not be possible for vehicles roaming throughout the 22,000-km^2 Atlanta metropolitan area).
• Even though high-end RTK-GPS systems are accurate, the loss of satellite signal lock due to overhead obstructions will still affect position and speed data (*8*) and statistical smoothing techniques may still be required.

The research team believes that the evaluation of smoothing techniques for the GPS data and better understanding of their statistical performance are necessary for transportation researchers because inexpensive GPS receivers (non-RTK-GPS systems) will be used in large-scale deployments.

STATISTICAL SMOOTHING TECHNIQUES

The basic principle of smoothing techniques is to augment or reduce erratic data points by replacing the value of input variables (*9*). Erratic location and speed data recorded from the GPS receiver can

lead to erroneous determinations on acceleration values. Most GPS receivers, including the SiRF Star II, use a proprietary filtering algorithm to compensate for data points beyond known variances (*4, 10*). That is, the device software embedded within the receiver automatically provides some level of data correction. Additional measures of reliability are included in the data stream to help identify questionable data. Researchers have developed numerous techniques to filter the data based on these measures with some degree of success. However, regardless of these smoothing and filtering algorithms, the proprietary filtering algorithms cannot filter all outliers, as evidenced by random errors still present in the GPS output data stream.

To minimize the impact of random errors on speed, acceleration, and travel distance estimates, Georgia Tech researchers proposed a supplemental smoothing process for postcollection GPS analysis. Without the full identification and correction of random GPS errors, researchers cannot reasonably evaluate driver acceleration and deceleration behaviors and travel distance. This study evaluates three statistical smoothing techniques and compares their capabilities minimizing the GPS random errors in the data streams.

Least-Squares Spline Approximation

The least-squares spline approximation, the so-called "piecewise polynomial regression model," divides the data set (Y_i) into several pieces with a predetermined width (or interval) and estimates predictors ($\hat{Y}_i$) using the RSS (*7, 11*). The local polynomial regression model derives a regression function from each localized data set using Equations 1 and 2. Equation 2 measures RSS and estimates each parameter ($\beta_0, \ldots, \beta_{d-1}$) within each interval.

$$\hat{f}(X) = \beta_0 X^0 + \beta_1 X^1 + \beta_2 X^2 + \cdots + \beta_{d-1} X^{d-1} + \epsilon \tag{1}$$

$$\text{RSS}(\hat{f}) = \sum_{i=1}^{n} \left[Y_i - \hat{f}(x_i) \right]^2 \tag{2}$$

where d is an order (or degree) of the function, and n is the sample size within the selected interval (*9*).

To evaluate the ability of the least-squares spline approximation as a smoothing method, researchers must decide the bandwidth representing the interval of the local data set and the order (or degree) of the regression function. The 1- and 2-s intervals have only one and two GPS data points, respectively. These intervals conceptually do not have sufficient data points for the polynomial model (one or two GPS data points cannot be smoothed by the smoothing algorithm). As bandwidths increase, they contain larger numbers of data points, and filtering may yield speed estimates for which some of the actual speed variability is smoothed away. Thus, this evaluation used a 3-s interval to avoid rapid increases and rapid decreases in acceleration rates calculated from change in speed over two consecutive seconds. In the case of order selection, because this study selected a 3-s interval, the quadratic function ($d = 3$) was selected as the order of the regression function.

Kernel-Based Smoothing Method

The kernel-based smoothing method assigns a weight (or a smoothing parameter) using the kernel density estimator (*9*). To obtain this estimator, the study used the Gaussian kernel estimator in Equation 3 (*9, 11*) and estimated the smoothing curve with the Nadaraya–Watson (NW) kernel-smoothing algorithm in Equation 4 (*9, 11*), as follows:

$$K_h(X_i, x) = K\left(\frac{|X_i - x|}{h} \right) = (2\pi h^2)^{-\frac{1}{2}} e^{\left[-0.5 \times \left(\frac{X_i - x}{h} \right)^2 \right]} \tag{3}$$

$$\hat{f}_{NW(x)} = \frac{\sum_{i=1}^{n} K_h(X_i - x)\hat{Y}_i}{\sum_{i=1}^{n} K_h(X_i - x)} \tag{4}$$

where h is the kernel bandwidth that controls the width of the localized data set and $K(t)$ is a kernel function that satisfies the following condition:

$$\int K(t)\,dt = 1 \tag{5}$$

The kernel-based smoothing method also requires bandwidth selection. Although the correct width (h) is not simply selected, and various references for selecting the kernel width exist, the normal reference rule in Equation 6 can be used in this study because of its relative simplicity (*9*). Bandwidths from the normal reference rule are between 2- and 4-s intervals based on the initial sample test. This study used a 3-s bandwidth for the kernel-based smoothing for two reasons: (*a*) the 4-s bandwidth significantly degrades the capability of the GPS data to be smoothed and (*b*) the least-squares spline approximation also uses the 3-s interval. Sin (*12*) also used the 3-s interval as the bandwidth parameter for evaluating the Epanechnikov kernel smoothing method and showed that this 3-s interval produced the best overall results.

$$h = \left(\frac{4}{3} \right)^{1/5} \sigma n^{-1/5} \approx 1.06 \sigma n^{-1/5} \tag{6}$$

where

   h = bandwidth,
   σ = standard deviation, and
   n = number of data points.

Discrete Kalman Filter

The final smoothing method in this study, the discrete Kalman filter, recursively estimates outputs using the feedback system in Figure 2 (*13*).

To perform the feedback system, the Kalman filter uses two processes—the prediction process (or the time update)—and the correction process (or the measurement update) and initially estimates a one-step predictor (a priori predictor) from the prediction process and obtains the correction (a posteriori predictor) from the correction process (*13–15*).
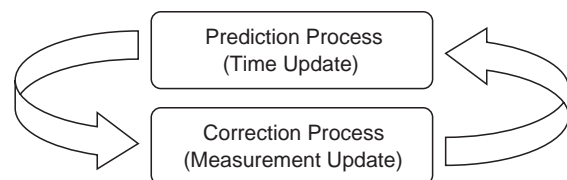

FIGURE 2   Kalman filter cycle.

The time update equations are as follows:

$$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_k \tag{7}$$

$$P_k^- = AP_{k-1}A^T + W \tag{8}$$

where

k = time step;

$\hat{x}_{k-1}$ and P_{k-1} = initial predictor and the initial error noise, respectively;

u_k = additional known-input parameter;

W = prediction error variance, which is the Gaussian noise, $N(0, Q)$; and

A and B = time transition matrices for the prediction process (12–14).

Because this study uses a GPS unit as a measurement device and separately tests the Kalman filter for smoothing speed (and therefore acceleration) and trip location points (X and Y coordinates), u_k in Equation 7 becomes zero (the one-dimensional Kalman filter). In addition, this study uses the second-by-second GPS speed data; therefore, the time transition matrix, A, is 1 s. Thus, Equations 7 and 8 are reduced to the following form:

$$\hat{x}_k^- = \hat{x}_{k-1} \tag{9}$$

$$P_k^- = P_{k-1} + W \tag{10}$$

The measurement update equations are

$$K_k = P_k^- H^T \left(HP_k^- H^T + V\right)^{-1} \tag{11}$$

$$\hat{x}_k = \hat{x}_k^- + K_k \left(z_k - H\hat{x}_k^-\right) \tag{12}$$

$$P_k = \left(I - K_k H\right)P_k^- \tag{13}$$

where

K = Kalman gain matrix,

H = time transition matrix for the observation process,

z = observed data,

P = modified error variance in the Kalman filter, and

V = measurement error variance, which is the Gaussian noise, $N(0, R)$.

Similar to the preceding reduced equations, the measurement update equations can also be reduced:

$$K_k = P_k^- \left(P_k^- + V\right)^{-1} \tag{14}$$

$$\hat{x}_k = \hat{x}_k^- + K_k \left(z_k - \hat{x}_k^-\right) \tag{15}$$

$$P_k = \left(I - K_k\right)P_k^- \tag{16}$$

Just as the least-squares spline approximation and the kernel-based smoothing method required a bandwidth value and the order of the function before the smoothing process, the Kalman filter requires values for the measurement noise (R) and the process noise (Q).

Modified Kalman Filter

Although the correct value of the measurement noise for the Kalman filter is not easily determined, previous studies (14, 15) suggested using the square of the mean error value from a manufacturer's technical specification. For smoothing vehicle location (X and Y coordinates), researchers used 100 ft (10^2 ft) as the measurement noise (R) (14–16). However, researchers should understand that this mean error in the manufacturer's technical specification was estimated in the perfect GPS condition: this value does not truly indicate the mean of errors in real-world conditions. In the case of speed profiles, researchers compared 1,171,496 GPS-measured speeds and corresponding VSS-derived speeds over a 2-month period and estimated the mean delta speed to be 0.5 mph. Thus, researchers used 0.25 mph (0.5^2 mph) as the GPS speed measurement noise. Given a 1-Hz data capture rate, the process noise of locations was the same as the measurement noise (1^2 s × 10^2 ft), and the process noise of speeds was also same as the measurement noise of speeds (1^2 s × 0.5^2 mph).

Here, another critical problem occurs when researchers use the measurement noise associated with location and speed data. The quality of the GPS data strongly depends on the GPS signal condition, usually represented by the number of satellites and PDOP values. When the condition of the GPS signal does not reach the level of minimum requirement, such as at least four satellites in view and PDOP values less than or equal to eight, the measurement errors are much greater than the preceding estimates. In addition, the most important component of the Kalman filter is the measurement error because the measurement error determines how much random GPS random error should be reduced. Thus, this study modified the conventional discrete Kalman filter by using two measurement errors based on the GPS quality criteria, the number of satellites, and PDOP values. Researchers estimated the first measurement error in the conditions of at least four satellites in view and PDOP values less than or equal to eight and the second measurement error from the other GPS signal conditions.

With this approach, this study used 10^2 degree (690^2 mi) as the measurement error of X and Y coordinates based on the result of preliminary evaluations and also used 10^2 mph of the measurement error for the speed profiles in the bad GPS signal conditions such as the loss of GPS signal lock.

ANALYSES AND RESULTS

This research evaluated three smoothing techniques to discern their effect on minimizing random GPS errors before calculating speed, acceleration, and distance profiles. Because reliable acceleration profiles can be derived from reliable speed profiles, both speed and acceleration profiles were tested by each smoothing technique. For evaluating travel distance profiles, this study conducted smoothing techniques to second-by-second X and Y coordinates and estimated travel distances. To compare all outputs produced by each technique and to verify their effectiveness, this study used speeds, accelerations, and distance profiles derived by the vehicle speed sensor as supplemental measurements.

Results of Speed and Acceleration Evaluations

Most previous studies of smoothing techniques generally tended to compare the original GPS data with the filtered GPS data esti-

mated by smoothing techniques, primarily because they did not have an alternative source of data, or ground truth. This research compares speed profiles obtained by the GT-TDC from the GPS receiver, the vehicle speed sensor monitor, and the OBD system [note that speed values from the VSS and OBD originate from the same source (*10*), transaxle rotation sensors, but are monitored and processed at different frequencies].

With the main objective of eliminating or reducing unrealistic acceleration data (or "outliers") from driving profiles, researchers carefully examined the results of the smoothing process to determine the effects of the smoothing. Researchers visually inspected the characteristics of speed and acceleration results from each smoothing technique with the original GPS-recorded speeds and accelerations and statistically compared the speed and acceleration estimates with the VSS-derived speeds and accelerations. Researchers also investigated how the smoothing algorithms actually dealt with these outliers. It is important to examine this effect for the following reasons:

- Given that acceleration profiles are derived from sequential GPS speed data points, the impact of each smoothing technique on the original speed profile results in different acceleration profiles.
- Given that random GPS errors in the speed profile provide unrealistic accelerations, extremely high acceleration or deceleration values must be eliminated by the smoothing technique.
- The smoothing techniques do not generally estimate much higher accelerations (or decelerations) than the original accelerations (or decelerations).

After running each smoothing technique with the original GPS-measured speed profile, researchers estimated three statistics: the mean of the errors (ME), the variance of the errors (VE), and the mean of the absolute errors (MAE), using the following equations:

$$\text{ME} = \text{mean}\left(Y_i - \hat{Y}_i\right) \tag{17}$$

$$\text{VE} = \text{Var}\left(Y_i - \hat{Y}_i\right) \tag{18}$$

$$\text{MAE} = \frac{\sum_i^n \text{abs}\left(Y_i - \hat{Y}_i\right)}{n} \tag{19}$$

The results of the comparative analysis are presented in Table 1. For the impact of each smoothing technique on all GPS speed data, all techniques provided a similar mean of delta speeds, but the modified Kalman filter provided the smallest mean of delta speeds when the signal of the GPS system indicated poor quality, such as fewer than four satellites. This result indicates that it is superior to other smoothing techniques. In the case of accelerations, it also provided the smallest difference from the VSS-derived accelerations across all metrics.

To verify whether means of delta speed and delta acceleration between those estimates derived by each smoothing technique and the VSS-derived speed are significantly different, this study performed the *t*-test ($\alpha = 0.05$). The hypothesis for testing the homogeneity is formulated as follows:

$$H_0 : \mu(x) = \mu(y)$$

$$H_1 : \mu(x) \neq \mu(y)$$

Table 2 indicates that all delta speeds and delta accelerations did significantly differ; this indicates that each smoothing method except the conventional Kalman filter and the modified Kalman filter overall provided a different error distribution even though the means of delta speeds and accelerations are similar.

In addition, because the statistical background of each smoothing method is different, each provides a unique output. For example, the kernel-based smoothing method often negatively affected speed accuracy estimates, although it did decrease outliers (large error-contained speeds). In contrast, the least-squares spline approximation, which minimizes the RSS between the original data profile and the estimated output profile, also affects reliable speed points near suspected outliers. In contrast to these two methods, the Kalman filter does have as significant an impact on those GPS speed points with low fluctuations between the sequential points but instead affects those sequential speed points with large speed differences (see Figure 3*a*).

The least-squares spline approximation provides higher speed estimates and lower speed estimates than original speeds (sometimes, the least-squares spline approximation provides negative speed estimates). The kernel-based smoothing method simultaneously smoothes the large range of speed data points around the out-

TABLE 1   Speed and Acceleration Smoothing Results

Speed Comparison	Mean of Delta Speeds (mph)	
	From All GPS Data	From GPS Data with Bad-Quality Signal
Least-squares spline approximation	−0.50	4.4
Kernel-based smoothing method	−0.49	4.4
Discrete Kalman filter	−0.49	4.4
Modified Kalman filter	−0.50	4.0
Acceleration Comparison	Mean (mph)	Variance (mph)	MAE (mph)
Least squares spline approximation	−0.00179	1.9669	0.77372
Kernel-based smoothing method	−0.00158	1.6287	0.69836
Discrete Kalman filter	−0.00133	1.4388	0.63735
Modified Kalman filter	−0.00047	1.4173	0.63222

TABLE 2   Results of *t*-Test for the Mean of Delta Speed

Delta Speed	(VSS—Spline)		(VSS—Kernel)		(VSS—Kalman)	
	Result	*p*-Value	Result	*p*-Value	Result	*p*-Value
(VSS—spline)	—	—	—	—	—	—
(VSS—Kernel)	Reject	0	—	—	—	—
(VSS—Kalman)	Reject	1.68E-28	Reject	7.45E-156	—	—
(VSS—modified Kalman)	Reject	2.08E-22	Reject	7.42E-172	Accept	0.22181
Delta acceleration						
(VSS—spline)	—	—	—	—	—	—
(VSS—Kernel)	Reject	7.49E-15	—	—	—	—
(VSS—Kalman)	Reject	1.40E-13	Reject	1.69E-50	—	—
(VSS—modified Kalman)	Reject	3.19E-18	Reject	1.22E-59	Accept	0.22397

liers; this results in larger speed errors between the original and smoothed speed profiles.

Figure 3*b* illustrates how each smoothing method produces different acceleration profiles. As expected, the least-squares spline approximation frequently provides higher accelerations (or decelerations) than the original accelerations (or decelerations), which is not a desirable result in the smoothing process. On the basis of these results, the Kalman filter is the preferred smoothing method.

Distance Estimates

In addition to the speed and acceleration profiles, travel distance profiles were also compared. Travel distances could be estimated from either the GPS speed data or the GPS *X* and *Y* coordinates. This study used *X* and *Y* coordinates instead of GPS speed data, as the latter were already investigated in the previous section and because distance errors were expected to be larger when calculated with sequential position data. This study examined each smoothing technique for its ability to minimize the impact of erroneous GPS data points on the estimates of travel distance per trip. Table 3 presents the results of the distance smoothing process. Similar to speed and acceleration, the groups of the Kalman filter provided the lowest delta distance (Table 3). The modified Kalman filter provided almost the same travel distances as the VSS-derived travel distances (Figure 4).

This study performed a χ^2 test to verify whether travel distance estimates were homogeneous with the VSS-derived distance. A contingency table (1-mi interval) with estimated χ^2 statistics was created in Table 4, and the hypothesis for testing the homogeneity was formulated as follows:

$$H_0 : F(x) = F(y)$$

$$H_1 : F(x) \neq F(y)$$

For 40 degrees of freedom, the critical value is $\chi^2_{40,0.05} = 55.76$. Table 4 indicates that all χ^2 statistics were significantly greater than the critical value except that of the modified Kalman filter, which indicated that only travel distance estimated from the modified Kalman filter did not differ from the VSS-derived distance. The

result of a *t*-test also indicates that travel distances filtered by the modified Kalman filter are not significantly different from those derived from the VSS data ($p = .75$) and that travel distances filtered by other techniques are significantly different.

CONCLUSIONS

GPS data contain random errors that have the potential to affect speed, acceleration, and travel distance estimates based on instrumented vehicle data. To use vehicle-based GPS data for insurance pricing, emissions analyses, and other modeling, GPS data smoothing may be required. This study selected three smoothing techniques that are popularly used in various traffic-related research and that are also characterized as different statistical background groups and evaluated their capabilities of minimizing the impact of error-containing GPS data while estimating driving speeds, accelerations, and travel distances. In addition, this study modified the conventional discrete Kalman filter algorithm to apply better to the GPS data-smoothing process.

The study found that the modified Kalman filter provided the smallest differences from the VSS-derived speed, acceleration, and travel distance estimates across all statistical metrics. In addition, through visual inspection of impacts of each smoothing technique on the second-by-second data streams, the modified Kalman filter was superior to the other smoothing techniques because it controlled outliers more effectively. Furthermore, the Kalman filter required less computational time than the other techniques, which indicates that this technique can be applied for the real-time smoothing algorithm.

Although only three smoothing methods were evaluated in this study, the researchers recommend use of the modified discrete Kalman filter for smoothing GPS speed and position data. This recommendation derives from analytical results and because the general statistical nature of the Kalman filter has a smaller impact on the accurate data points that reside near erroneous data points. The researchers will continue to evaluate additional smoothing methods over the next year and plan to deploy an RTK-GPS system to assess whether the modified Kalman filter will prove useful in smoothing data streams collected with high-end GPS systems.
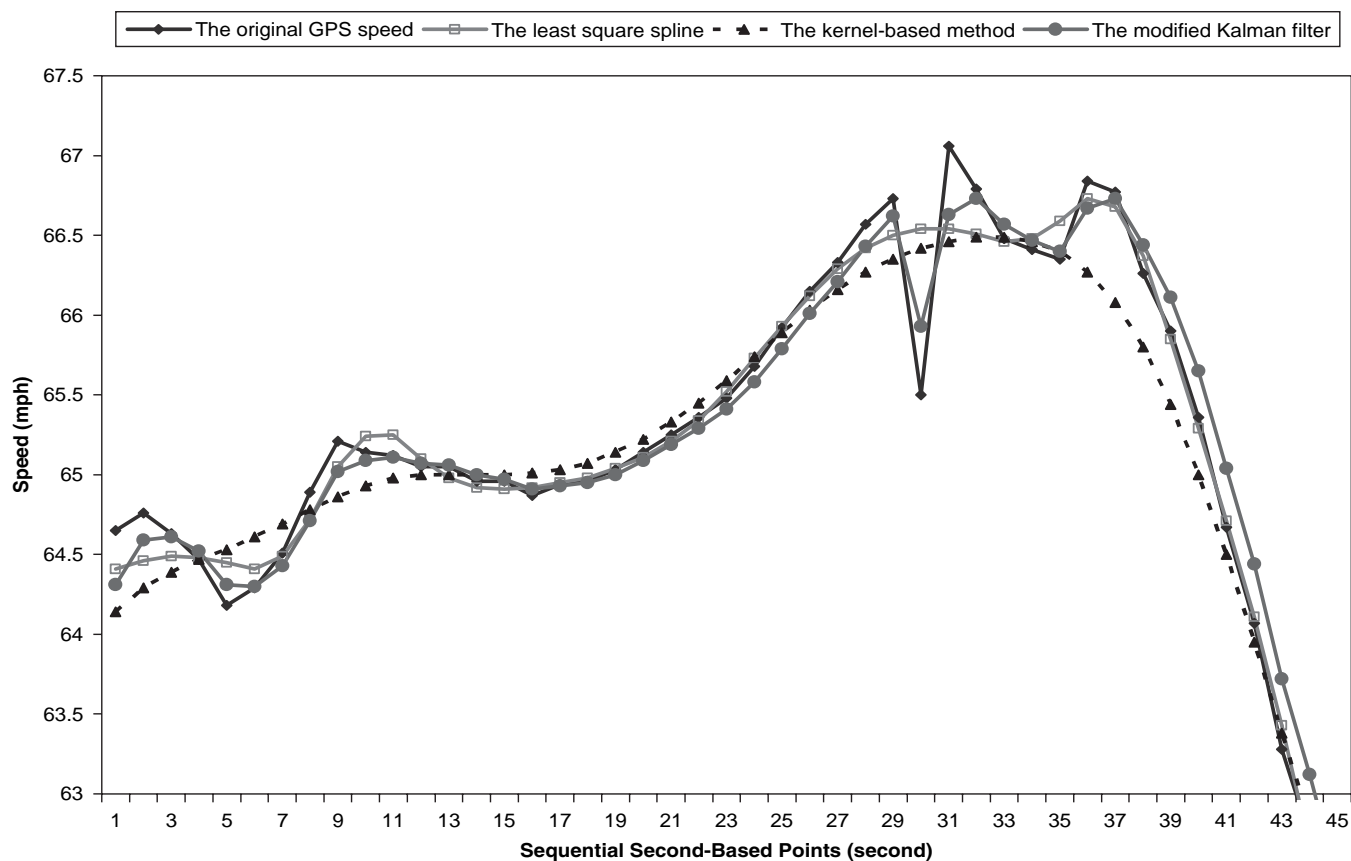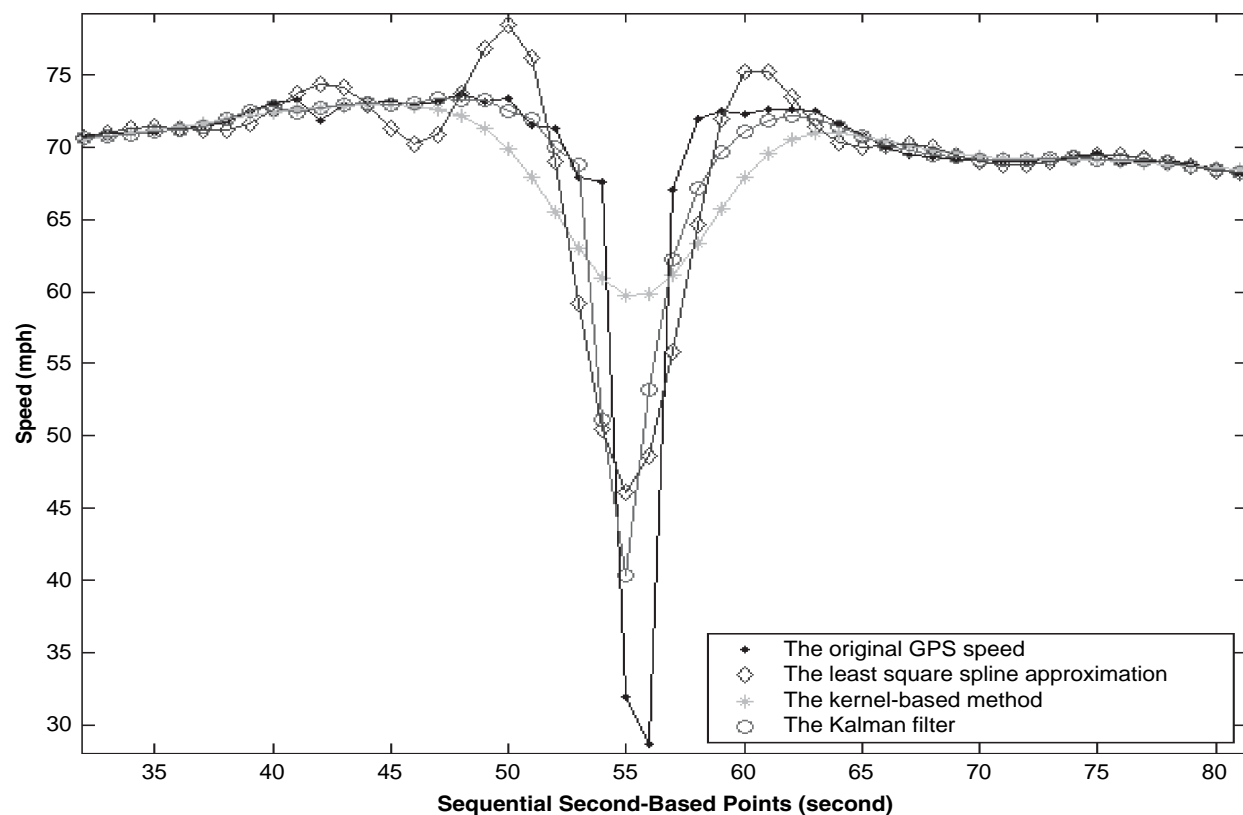
FIGURE 3   Smoothing impacts of outliers.

TABLE 3   Distance Smoothing Results

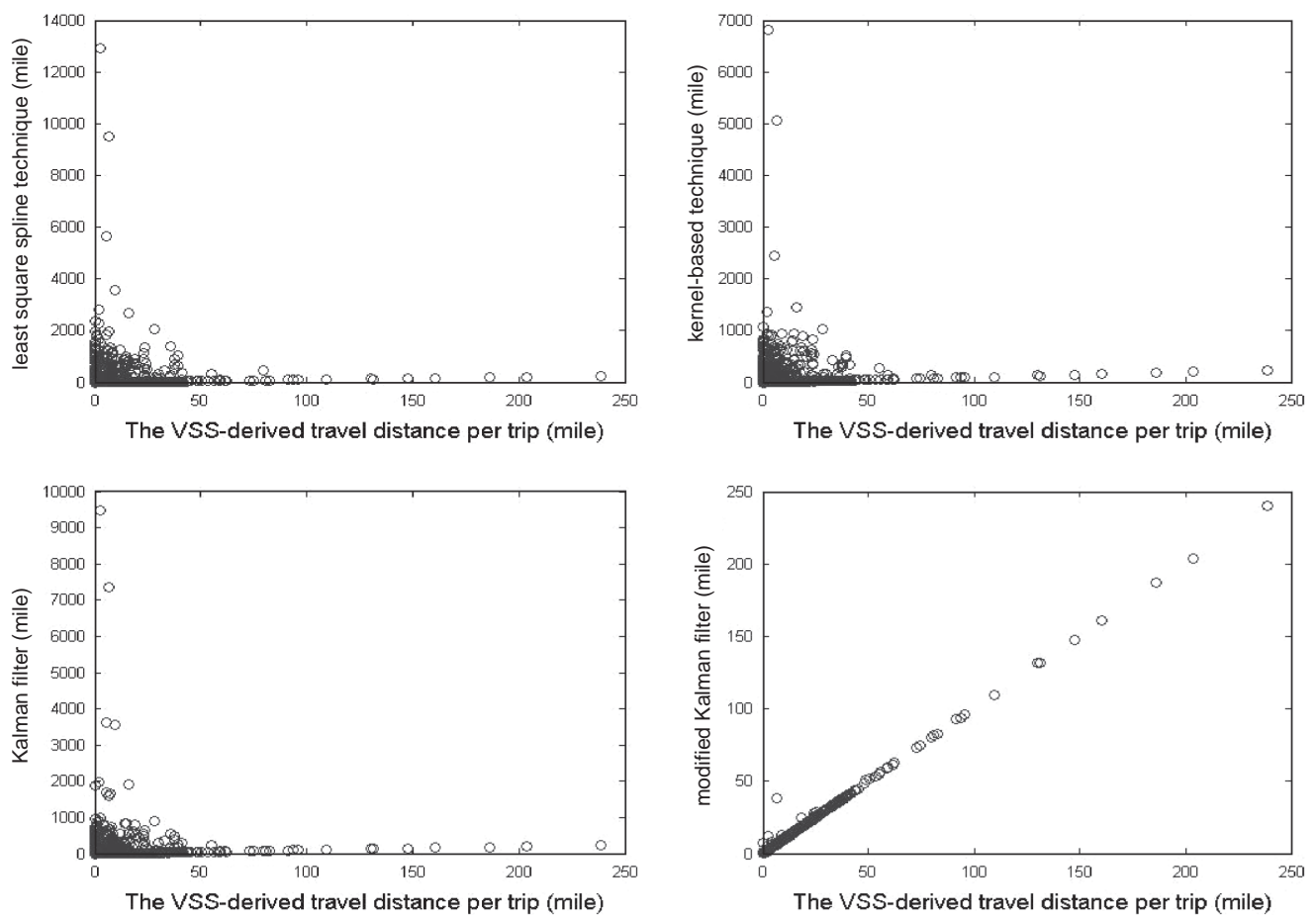| Distance Comparison	Mean of Travel Distance per Trip (mile)	MAE of Travel Distance per Trip (mile)
Least-squares spline approximation	−97.414	97.904
Kernel-based smoothing method	−56.604	57.127
Discrete Kalman filter	−52.919	53.537
Modified Kalman filter	0.179	0.192

FIGURE 4   Travel distance comparisons.

TABLE 4   Contingency Table for Travel Distance per Trip

Distance Interval (mile)	VSS Freq.	Spline Freq.	χ^2	Kernel Freq.	χ^2	Kalman Freq.	χ^2	Modified Kalman Freq.	χ^2
0 ~ 1	302	270	1.79	286	0.44	279	0.91	284	0.55
1 ~ 2	162	162	0.00	159	0.03	163	0.00	169	0.15
2 ~ 3	126	94	4.65	93	4.97	93	4.97	124	0.02
3 ~ 4	80	59	3.17	60	2.86	59	3.17	84	0.10
4 ~ 5	62	67	0.19	60	0.03	64	0.03	55	0.42
5 ~ 6	75	58	2.17	64	0.87	65	0.71	76	0.01
6 ~ 7	53	59	0.32	59	0.32	56	0.08	51	0.04
7 ~ 8	104	110	0.17	110	0.17	112	0.30	102	0.02
8 ~ 9	131	67	20.69	78	13.44	67	20.69	135	0.06
9 ~ 10	59	53	0.32	42	2.86	53	0.32	60	0.01
10 ~ 11	33	44	1.57	42	1.08	42	1.08	41	0.86
11 ~ 12	33	34	0.01	33	0.00	34	0.01	28	0.41
12 ~ 13	21	22	0.02	23	0.09	23	0.09	23	0.09
13 ~ 14	63	33	9.38	31	10.89	34	8.67	66	0.07
14 ~ 15	37	27	1.56	28	1.25	29	0.97	41	0.21
15 ~ 16	20	13	1.48	12	2.00	13	1.48	15	0.71
16 ~ 17	10	8	0.22	10	0.00	6	1.00	14	0.67
17 ~ 18	5	11	2.25	9	1.14	11	2.25	5	0.00
18 ~ 19	14	3	7.12	2	9.00	2	9.00	15	0.03
19 ~ 20	8	9	0.06	8	0.00	8	0.00	6	0.29
20 ~ 21	10	9	0.05	10	0.00	10	0.00	10	0.00
21 ~ 22	7	8	0.07	7	0.00	8	0.07	9	0.25
22 ~ 23	9	10	0.05	16	1.96	10	0.05	9	0.00
23 ~ 24	34	37	0.13	44	1.28	39	0.34	34	0.00
24 ~ 25	65	38	7.08	35	9.00	39	6.50	61	0.13
25 ~ 26	6	19	6.76	15	3.86	21	8.33	9	0.60
26 ~ 27	8	9	0.06	9	0.06	7	0.07	7	0.07
27 ~ 28	4	10	2.57	11	3.27	10	2.57	5	0.11
28 ~ 29	31	19	2.88	22	1.53	20	2.37	25	0.64
29 ~ 30	14	15	0.03	9	1.09	13	0.04	20	1.06
30 ~ 31	6	5	0.09	7	0.08	5	0.09	5	0.09
31 ~ 32	1	4	1.80	1	0.00	4	1.80	4	1.80
32 ~ 33	9	8	0.06	8	0.06	9	0.00	7	0.25
33 ~ 34	7	8	0.07	7	0.00	7	0.00	7	0.00
34 ~ 35	6	4	0.40	3	1.00	3	1.00	6	0.00
35 ~ 36	12	4	4.00	6	2.00	6	2.00	12	0.00
36 ~ 37	3	10	3.77	11	4.57	8	2.27	4	0.14
37 ~ 38	13	10	0.39	11	0.17	11	0.17	13	0.00
38 ~ 39	8	9	0.06	5	0.69	9	0.06	8	0.00
39 ~ 40	13	8	1.19	9	0.73	9	0.73	14	0.04
40 ~	38	255	160.71	247	153.27	241	147.70	39	0.01
Total	1,702	1,702	249.38	1,702	236.04	1,702	231.91	1,702	9.90

REFERENCES

1. Guensler, R., A. Amekudzi, J. Williams, S. Mergelsberg, and J. Ogle. Current State Regulatory Support for Pay-As-You-Drive Automobile Insurance Options. *Journal of Insurance Regulation,* Vol. 21, No. 3, 2003, p. 31.
2. Progressive Casualty Insurance Corporation. tripsense.progressive.com. Accessed April 2005.
3. Norwich Union. www.norwich-union.co.uk. Accessed April 2005.
4. Ogle, J., R. Guensler, W. Bachman, M. Koutsak, and J. Wolf. Accuracy of Global Positioning System for Determining Driver Performance Parameters. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1818,* Transportation Research Board of the National Academies, Washington, D.C., 2002, pp. 12–24.
5. Zito, R., G. D'Este, and M. A. P. Taylor. Global Positioning Systems in the Time Domain: How Useful a Tool for Intelligent Vehicle-Highway Systems? *Transportation Research C,* Vol. 3, No. 4, 1995, pp. 193–209.
6. Gates, T. J., S. D. Schrock, and J. A. Bonneson. Comparison of Portable Speed Measurement Devices. In *Transportation Research Record: Journal of the Transportation Research Board, No. 1870,* Transportation Research Board of the National Academies, Washington, D.C., 2004, pp. 139–146.
7. *AgGPS 214 High-Accuracy Receiver, Centimeter Precision in Agriculture.* Technical Note. Trimble, Sunnyvale, Calif., 2000.
8. Lin, L.-S. Application of GPS RTK and Total Station System on Dynamic Monitoring Land Use. *Proc., XXth ISPRS Congress, Geo-Imagery Bridging Continents,* Istanbul, Turkey, July 12–23, 2004.
9. Hastie, T., R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer, New York, 2001.
10. Ogle, J. H. *Quantitative Assessment of Driver Speeding Behavior Using Instrumented Vehicles.* Ph.D. dissertation. School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, 2005.
11. Martinez, W. L., and A. R. Martinez. *Computational Statistics Handbook with MATLAB.* Chapman & Hall/CRC, Boca Raton, Fla., 2002.
12. Sin, H. G. *Field Evaluation Methodology for Quantifying Network-Wide Efficiency, Energy, Emission, and Safety Impacts of Operational-Level Transportation Projects.* Ph.D. dissertation. School of Civil and Environmental Engineering, Virginia Polytechnic Institute and State University, Blacksburg, 2001.
13. Welch, G., and G. Bishop. *An Introduction to the Kalman Filter.* The University of North Carolina and ACM, Inc., SIGGRAPH, Chapel Hill, N.C., 2001. www.cs.unc.edu/~welch/kalman/kalmanIntro.html. Accessed June 2005.
14. Simon, D. Kalman Filtering. *Embedded Systems Programming,* Vol. 14, No. 6, June 2001, pp. 72–79.
15. Bashi, A. S. *A Practitioner's Short Guide to Kalman Filtering.* Jan. 1998. www.uno.edu/~SAGES/publications/PractitionersKalman.PDF. Accessed June 2005.
16. *Technical Specification for SiRF Star II Architecture.* SiRF, Basingstoke, Hampshire, United Kingdom www.tdc.co.uk/gps/gps_receivers_leadtek.htm. Accessed July 2005.