

High-Accuracy Geometric Highway Model Derived from Multi-Track, Messy GPS Data

Stanley E. Young
Kansas Department of Transportation
2300 Van Buren Street
Topeka, KS 66611
young@ksdot.org

Rick Miller
Kansas Department of Transportation
2300 Van Buren Street
Topeka, KS 66611
rick@ksdot.org

ABSTRACT

GPS provides longitude, latitude, and elevation points and is already integrated into some highway data collection processes at the Kansas Department of Transportation. For instance, GPS tracks have been systematically collected on the 11,500 mile Kansas state highway system since 1997 as part of two separate data collection activities: videolog and pavement management. While not the primary purpose of the data collection activities, over 11 million GPS points have been collected to date. These individual geodetic points became the basis for a high-accuracy geometric highway model and for a high-quality base map for GIS purposes. Methods to process this data were developed and applied to the entire database of GPS points, resulting in a geometric model of the state highway system. In turn this geometric model can now be used as the basis for other applications. One of the first applications was the assessment of sub-standard stopping sight distance. The principles and mechanics of the processing methods that combine multi-year, multi-run GPS data into highly accurate spatial models are reviewed and summarized.

Key words: geometric highway model—GPS—sub-standard stopping sight distance

INTRODUCTION

The Kansas Department of Transportation (KDOT), like many state highway organizations, uses a variety of location referencing systems to specify position along a highway. These systems include stationing, county route logpoint, and many relative or descriptive means. However, inventory databases contain several million spatial coordinate triplets of longitude, latitude, and elevation collected using GPS technology, collected along the 11,500 miles of Kansas highways since 1997. An algorithm for estimating a three-dimensional spatial model of roadway geometry from the historic GPS data has been the research objective over the past few years. Although several concepts from digital signal processing, optimal estimation, and data reduction are applicable in estimating geometry from multi-run data series, the unique structure and error distribution of the historic GPS data prohibit direct application of established methods. The spatial error from successive GPS data is highly correlated. Even though the GPS error is widely published to be in the range of 1 to 5 meters, the relative accuracy of sequential GPS data is much greater, providing the opportunity to develop a high accuracy geometric model of the highway system for the entire state from which to assess geometric properties, such as stopping sight distance and passing sight distance.

This paper presents the basic principles of an algorithm to process the GPS data that is a variant of a robust least-squares technique. The algorithm takes advantage of the high correlation in GPS error in order to estimate grades and trajectories accurately. The algorithm utilizes low-order normalized piecewise polynomials as basis functions. Both the position data and estimates of the first derivative are used in the fitting process. Robust fitting techniques are used to iteratively re-weight data points and to estimate and subtract bias errors.

The resulting spatial equations that describe the trajectory of the highway are reduced to a set of control points, that is, a sequential set of points along the route or highway spaced closely enough to capture and recreate the horizontal and vertical curvature using a spline technique. Accuracy of the model is also estimated. The accuracy metrics enable follow-on applications to determine if the model is accurate enough to obtain valid results. Accuracy metrics are calculated for absolute position error, relative position error, and error in the first derivative of the spatial model.

In order to validate the algorithm, the elevation output from the model was compared to the elevation design profile of three highway segments. The algorithm produced elevation models with elevation standard deviations of ~1.5 feet and grade standard deviations between 0.10% and 0.17%.

GPS ERROR CHARACTERISTICS

Absolute GPS error characteristics using civilian GPS receivers have been widely published. Several factors influence the size and shape of the error distribution. Among these are the number and distribution of GPS satellites, type and fidelity of GPS equipment, the use of differential correction services, atmospheric effects, activation of selective availability, jamming of the GPS signal, and local obstructions that limit the view of the sky, such as trees, canyons, and buildings. Many of the factors listed above primarily affect the absolute error in the data.

GPS data is typically taken in one-second intervals. If successive GPS data points use the same constellation of satellites, the relative error between the two data points is minimal. Assuming absolute errors of 2 m and 5 m, respectively, for horizontal and vertical error, the relative error between successive readings is easily sub-meter in both dimensions. The error correlation between successive GPS data was estimated between the 0.99- and 0.999-level in the work performed. Improvements to the GPS system in

recent years, such as differential correction services and discontinuance of selective availability, have decreased the absolute error of the system, but have done little to change the relative error characteristics.

The error in the spatial model obtained from GPS data can be estimated using statistical sampling theory. The error in the estimate of the mean of a population is inversely proportional to the square root of the number of samples, as shown in Equation 1.

$$\delta_{\bar{x}} = \sqrt{\frac{\frac{1}{n} \sum_{i=1}^n (\delta_{x_i})^2}{n}} \quad (1)$$

Applying this concept to the absolute positioning of highways yields the results shown in Figure 1, which shows the hypothetical reduction in the absolute position error of the three-dimensional model as a function of the number of observations, assuming that each observation has a 2-m random error in horizontal position and a 5-m random error in elevation. Assuming 2-m and 5-m absolute position accuracies of the individual GPS data run, the absolute accuracy of a spatial model is expected to reach sub-meter horizontal accuracy and ~2-m vertical accuracy by averaging a handful of GPS tracks. Equation 1 assumes independent observations. Figure 1 assumes that the data points are separated in time sufficiently to assume independent observations.

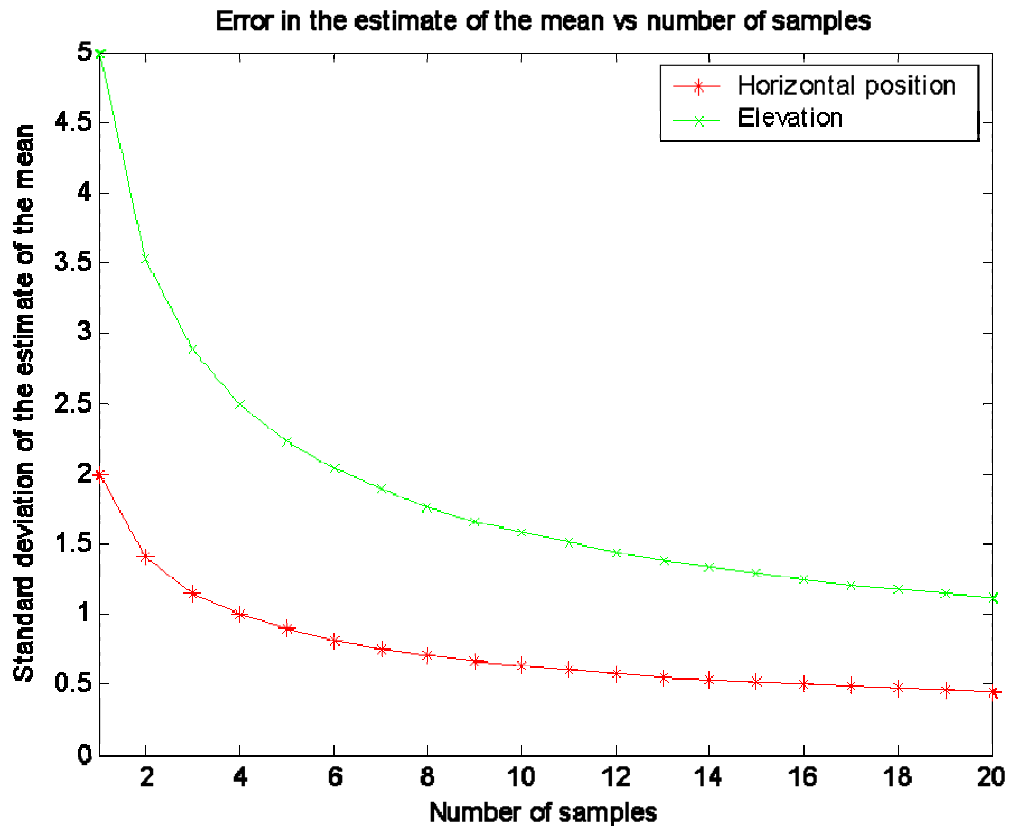


Figure 1. Hypothetical reduction in the absolute position error of the three-dimensional model as a function of the number of observations

The high correlation of GPS data error provides in essence a high quality estimate of heading in the horizontal plane and grade in the vertical plane. Additional error reduction arises because successive estimates of slope are highly independent, unlike position estimates. These GPS data attributes combined with basic knowledge of road design geometry allows for further error reduction. This is best illustrated with a simple example. An 800-foot crest vertical curve was constructed with a curvature of 247 ft per percent grade (%G). Samples of the grade were taken approximately every 12 feet to simulate seven GPS datasets. Random noise was added to the samples consistent with the error characteristics for a correlation coefficient, $p = 0.99$, which is approximately 2.6% grade, or 0.026 ft/ft. Since the grade of a crest vertical curve varies linearly with distance, a straight line was fit to the data using a covariance-weighted least-squares technique capable of producing one-sigma confidence bounds. Figure 2 illustrates the results. Simulated samples with a normally distributed error of 2.6% grade were fit with a first-degree polynomial (straight line) as shown. The one-sigma confidence bounds are illustrated in the dashed lines.

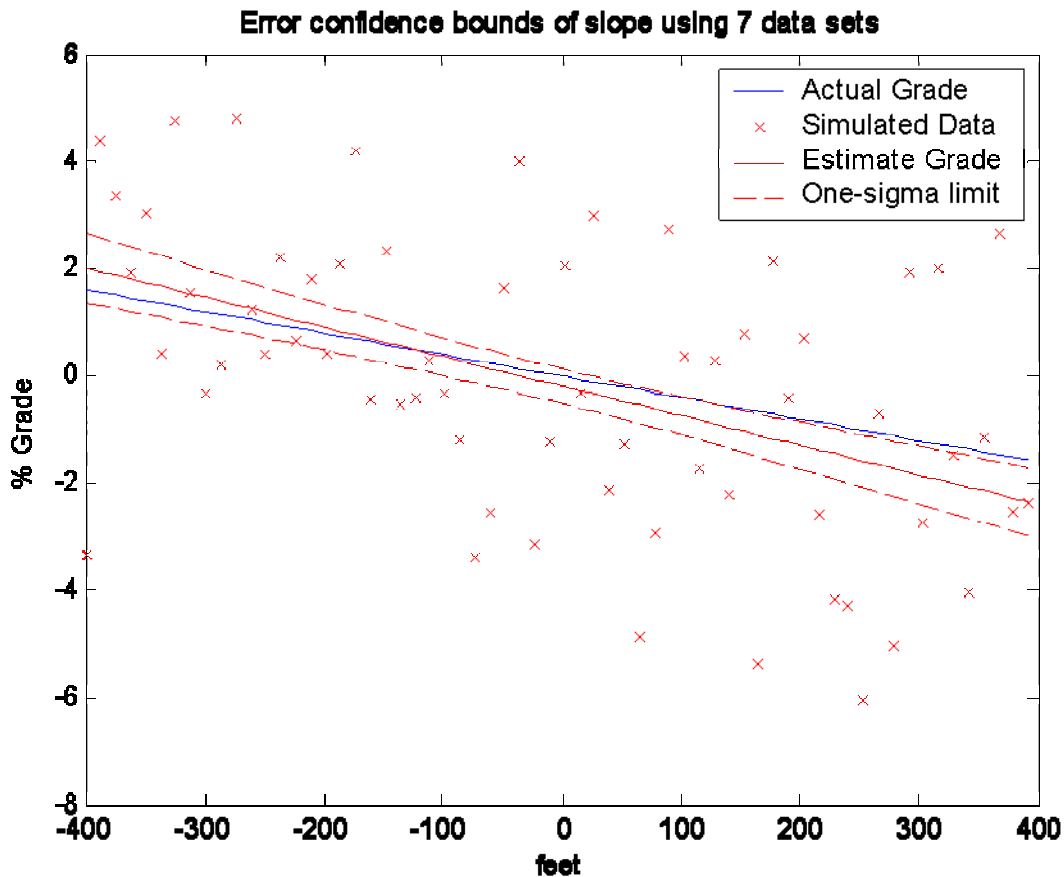


Figure 2. Slope profile for an 800 foot crest vertical curve with curvature of 247 ft/%G

Using this method, the expected error in slope from the spatial model as a function of the number of datasets can be calculated assuming typical values for GPS error and the correlation coefficient. The result of this calculation, assuming 5-m absolute vertical accuracy, a 0.99 correlation coefficient, and a simple 800-ft crest vertical curve are shown in Figure 3. Grade accuracies between 0.3% and 0.4% grade should be attainable with five or more GPS tracks. A similar analysis for heading indicates an expected one-sigma accuracy of 0.2 to 0.4 degrees after using five GPS tracks to obtain the model.

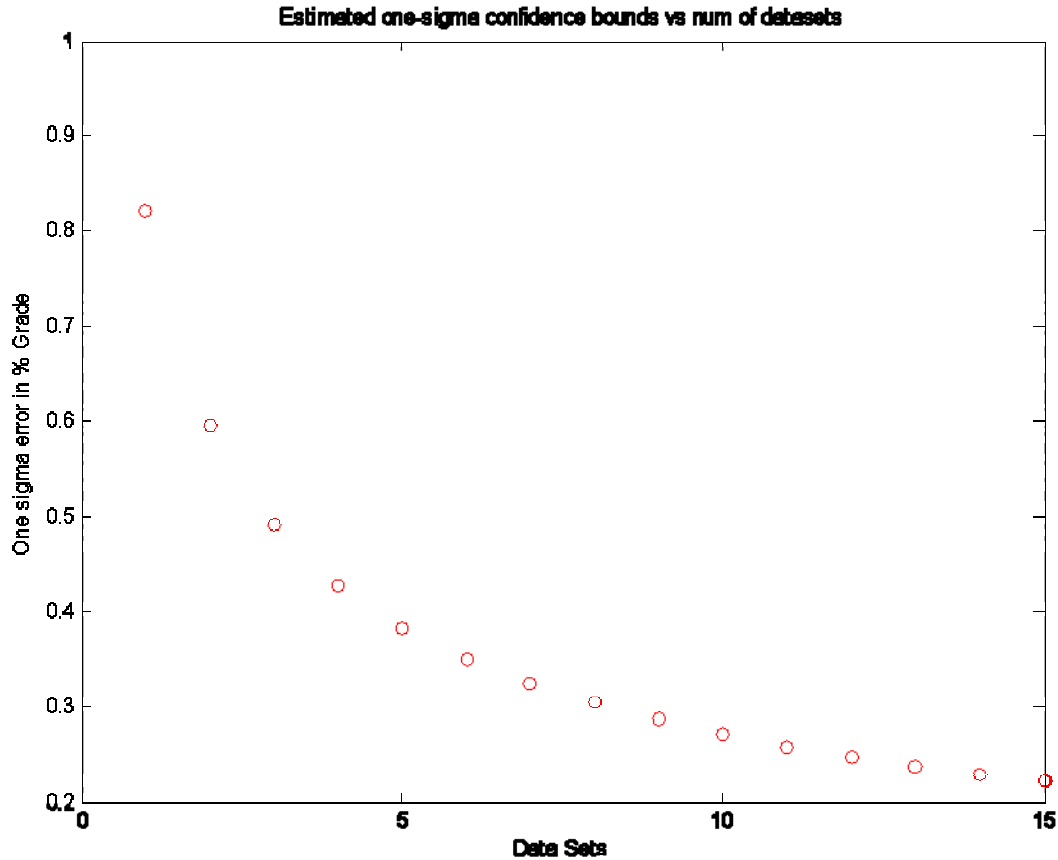


Figure 3. One-sigma confidence bounds for the estimate of slope on an 800-foot crest vertical curve

MODELING PRINCIPLES

The modeling methodology is based on a robust, iterative, linear least-squares methodology with covariance weighting. Each contiguous section of roadway is broken into 0.2- to 0.4-mile sections. Within each section, the coordinates are transformed into a normalized coordinate system. The independent axis is aligned with the general direction of the route and scaled between -1 and 1. The dependent axis is scaled accordingly in order to preserve first derivative information (heading and grade). The normalization of the data within each segment combined with the low-order polynomial basis functions provides a stable computation environment that closely resembles the Legendre polynomials.

Slope information is estimated using a difference formula on successive GPS points. The spatial data (longitude, latitude, and elevation) and first derivative data (heading and slope) are used to estimate the parameters of the low-order polynomial on each section. Boundary constraints are introduced at the intersection of successive sections. The position, heading, grade, and curvature are maintained across the section boundaries. Lagrange multipliers provide a convenient method to enforce the boundary constraints within a linear (matrix) formulation.

The estimates of the polynomial parameters are based on least-squares minimization of the residual. The formulation of the linear problem includes weighting coefficients for each data point. The weighting coefficients are introduced in the form of a covariance matrix. The covariance matrix provides both the correct relative weight for each data point, as well as a method to encode data correlation on the off-

diagonal elements. On the first iteration of the algorithm, the covariance matrix is initialized using expected values for variance and correlation. Each subsequent iteration uses the residuals from the previous fit to re-estimate the variance of each data point and the correlation between data points. This iterative approach enables detection of outliers, which are then de-weighted in subsequent calculations.

In addition to adjusting the covariance weighting matrix, the residuals are used to estimate the bias of each GPS track. The estimated bias is then subtracted from each respective track before the next iteration. The bias is estimated not only for the dependent coordinates, but also for the independent coordinate axis. Through a simplified correlation procedure, each GPS track is de-biased (or shifted) along the independent axis until it coincides with the model.

The procedure is repeated and usually converges within five to six cycles. The final solution provides not only an optimal estimate of the trajectory of the highway, but also significant information concerning the quality and consistency of the GPS data, from which an estimate of model accuracy can be derived. The covariance matrix from the final iteration provides a good estimate of the uncertainty in each data point. This covariance matrix in turn determines the confidence in the coefficients in the polynomial model, which in turn can be transformed into uncertainty estimates of slope and curvature of the final model through similarity transformations. In a similar manner, the final bias adjustments for each GPS track provide a measure of the absolute position uncertainty of the resulting model.

In order to obtain immunity from outliers, which is a commonly cited drawback of least-squares minimization, the algorithm relies primarily on percentile-based methods for determining sampling statistics. Common methods for calculating mean and variance are predicated on a well-behaved and consistent distribution of data about the mean. In the presence of outliers, which are quite common in GPS data, standard sampling statistics are greatly affected and tend to reflect characteristics of the outliers rather than capture the trend in the majority of the data. Percentile-based methods provide a robust approach to ignore data points that fall significantly outside the general bounds of the data, a situation commonly referred to as a long-tailed distribution. Robust percentile-based methods are used on each cycle to estimate the bias, variance, and correlation of each GPS track.

RESULTS

Elevation profiles obtained using the algorithm described above were compared with elevation profiles that had been manually transcribed from archived design plans on three Kansas highway sections; K-177 and K-113 in Riley County and a portion of US 283 in Hodgeman County. K-177 is a four-lane divided facility over slowly rolling hills. The highway geometry is comparable to higher class roadways, such as interstates where the effect of the local terrain is minimized by major cuts and grading. K-113 in Riley County is a two-lane structure built on rolling terrain. The road follows the local terrain more closely than higher class facilities. U-283 in Hodgeman County is a rural two-lane road built in an area of relatively flat terrain. All three highways run south to north. The GPS data obtained from the north-bound lanes are compared to the vertical design profiles. The number of equivalent full-length GPS tracks for each highway was 8, 3, and 5 for routes 177, 113, and 283, respectively. The actual numbers of tracks for each highway are greater; however, each individual GPS track may not have spanned the entire highway section being modeled. The highways were modeled with approximately 0.25-mile intervals and fifth-order polynomials.

Figure 4 depicts the elevation data for K-177 and the model obtained from the algorithm. The maximum distance between the GPS elevation traces is over 150 feet. Although the span between the traces is large, it is not wholly unexpected. Civilian GPS receivers are known to have built-in elevation biases (Wilson 2004). The data appear to have two clusters of elevation traces. This may be data from two different

receivers, each with a different built-in bias. Many receivers have user-defined fields for antenna height. No data exists to determine how these user-defined fields were programmed during data collection. However, Figure 4 indicates that the relative shape of the roadway is consistently captured in the GPS data, despite the differences in absolute elevation. The reason for the large spread in elevation bias is unknown. The data appears bimodal, suggesting that possibly two different GPS receivers, each with a different bias, were used to collect the elevation data.

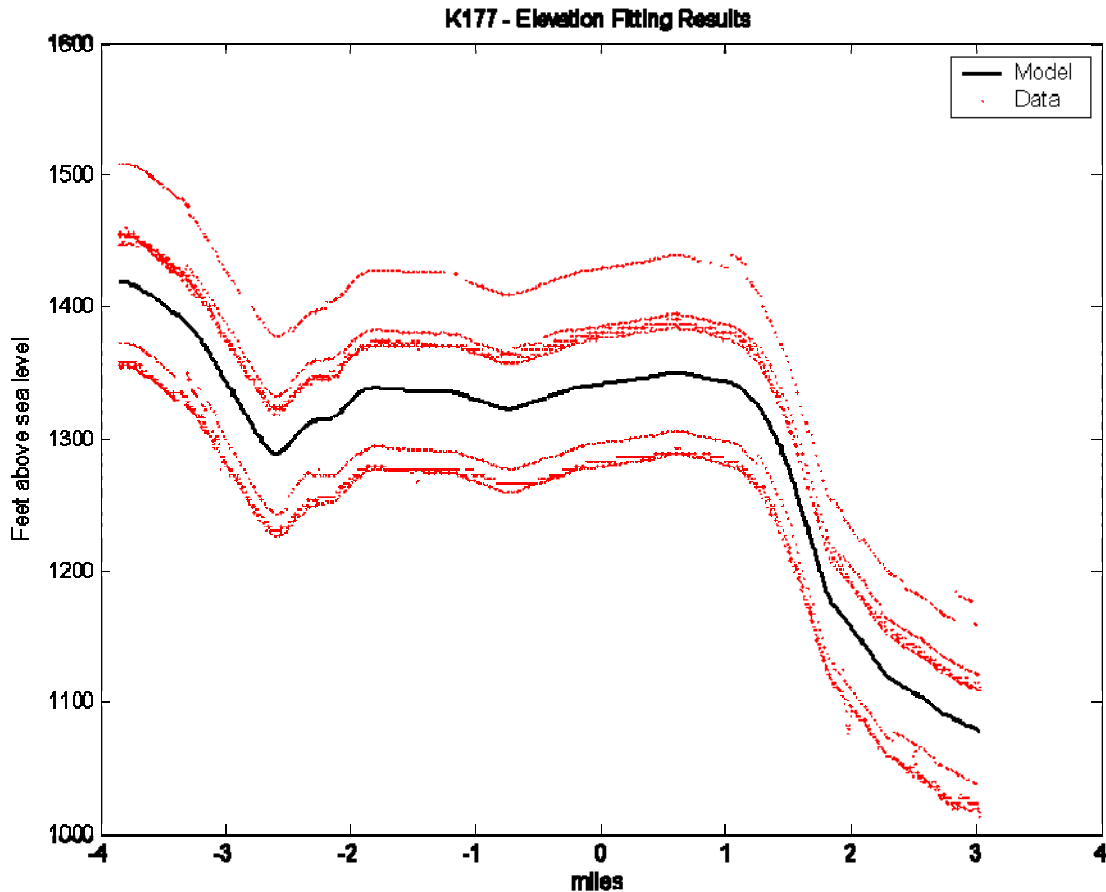


Figure 4. GPS elevation for K-177, in red, and the model obtained from the 3-D algorithm, in black

Despite the elevation biases depicted in Figure 4, a high-quality geometric model was obtained from the data and compared with the design profile, as shown in Figure 5. The red and blue traces in Figure 5 are a direct comparison of the elevation model to the design elevation profile, respectively. The difference between the model and the design profile are magnified and graphed in black on the lower portion of Figure 5, with the appropriate scale indicated on the right axis.

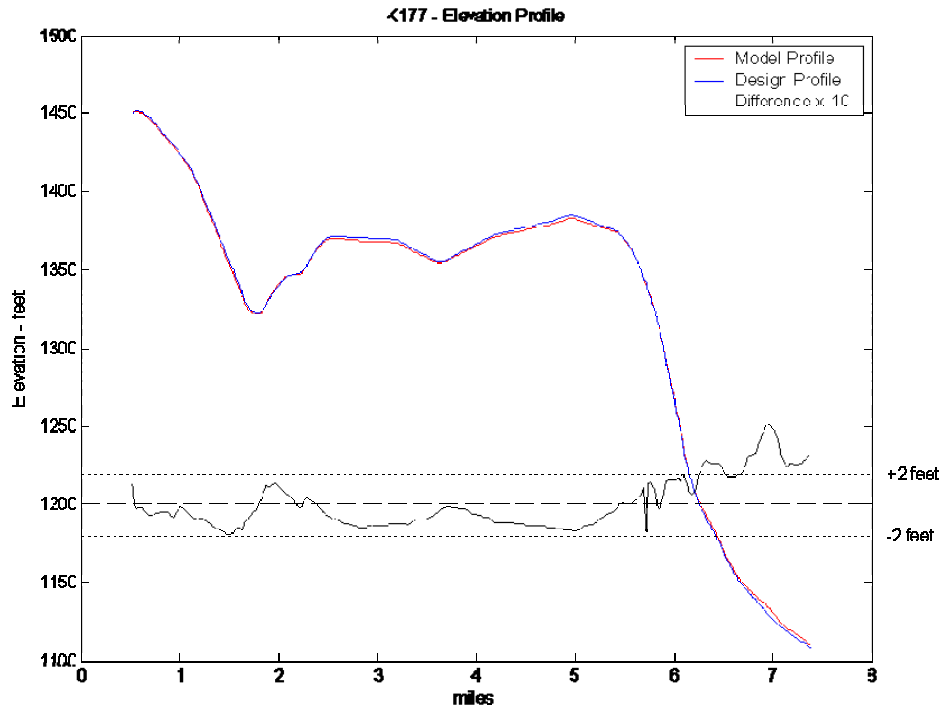


Figure 5. Elevation model for K-177

Figure 6 depicts the model grade compared to the design grade. The grade profile obtained from the 3D algorithm for K-177 is shown in red. The design grade, shown in blue, was obtained by applying a simple difference formula to the design elevation profile. Note the obvious error in the design profile at mile 5.8. The abrupt changes in the design grade at approximately 5.8 and 6.1 miles indicate transcription errors in the design grade obtained from archived plans. Recall that highways are designed primarily with parabolas and tangents (straight lines). This results in a continuous, but non-differentiable grade profile. Linear models work best on continuously differentiable curves. The abrupt changes in the rate of change of grade at the transition between parabolic and straight line segments causes the linear model to overshoot and undershoot, commonly referred to as ringing in classical control theory. This is most evident toward the end of the profile in Figure 6 between mile 6.4 and 7.5.

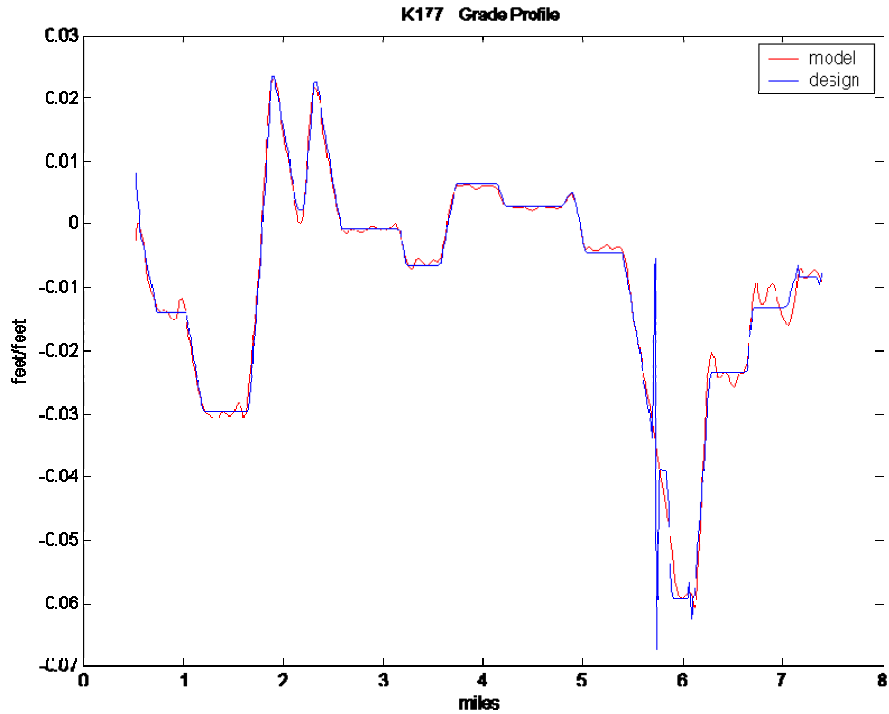


Figure 6. The grade profile obtained from the 3D algorithm for K-177

The agreement between the model grade and the design grade for the three highway sections used for verification are characterized in Table 1. The first row, labeled Standard Deviation, indicates the one-sigma error value between the model and the design profile obtained using standard sampling statistics. It is heavily influenced by outliers, such as the differences in grade due to transcription error of the grade profile. The Robust Sigma statistic label is a percentile-based method of estimating the one-sigma error boundary. It ignores the influence of outliers and is a better estimate of the agreement between the two profiles. The Model Prediction label indicates the expected one-sigma error in the model derived from the standard error of the coefficients. Note that the one-sigma error estimates in grade are less than the 0.3% to 0.4% bounds predicted from basic statistical principles and conservatively estimating the correlation in GPS error distributions. This is an indication that the correlation coefficient (a measure of the relative accuracy of GPS data) is greater than the 0.99 assumed in the preliminary calculations.

Table 1. One-sigma estimates of grade error between geometric model and design plans

	K-177	K-113	U-283
Standard Deviation	0.27	0.20	0.53
Robust Sigma	0.094	0.15	0.17
Model Prediction	0.049	0.09	0.07

(all values are one-sigma estimate of error in % grade)

CONCLUSIONS

KDOT has successfully developed procedures to use GPS data collected in existing highway inventorying systems to generate mathematical functions of highway geometry for which the accuracy can be reasonably estimated. The error in the geometric model will be reduced as the number of GPS data points grows. Subsequently, KDOT has developed a sample application built on the geometric highway models

to evaluate vertical sight distance. The advantage of such an application is the ability to evaluate the entire state highway system for potential vertical sight distance problems under a variety of input parameters for driver eye height and/or object height. These parameters have been changed in the recent past and, if adopted by KDOT, would have required a significant undertaking to understand the impact. Again using the geometric models, similar applications could be developed for horizontal sight distance analysis, sight distance at requested access points, or, for that matter, as a means to maintain a GIS basemap. KDOT plans to maintain and continue to update the geometric models as additional data become available. Additionally, KDOT hopes to build additional applications based on the models to take advantage of now having the entire state highway system located in a three-dimensional, contiguous mathematical model with estimates of accuracy.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Margaret Rys, Shing Chang, and David Ben-Arieh, professors in the Department of Industrial and Manufacturing Systems Engineering at Kansas State University, for their collaboration and support in the project, and to Jennifer Distlehorst, researcher in the Bureau of Materials and Research at the Kansas Department of Transportation, for her assistance refining and editing the manuscript.

REFERENCES

- Wilson, D.L. 2004. Assorted articles, including “Horizontal GPS Accuracy,” “Horizontal error when returning to a position (earlier fix),” “Vertical GPS Accuracy,” “Using HDOP to more precisely estimate horizontal error,” “Error when position is averaged to improve accuracy,” “GPS accuracy with WAAS,” “A look at correlations of errors,” “A discussion of weighting by HDOP when attempting to improve the horizontal accuracy by averaging,” “A look at whether DGPS (differential GPS) improves accuracy now that SA has been removed,” “A comparison of horizontal accuracy with when SA was on.” <http://users.erols.com/dlwilson/gps.htm>.
- Young, S.E. 2004. Development of a Continuous Three-Dimensional Geometric Highway Model using Historical Global Positioning System Data. Doctoral dissertation, Kansas State University.