Impact of the data uncertainties on the regression model: application to the trajectory-aided surface GNSS navigation

*Guoliang Zhu[1], Lionel Fillatre[2], Igor Nikiforov[1]

[1]ICD-LM2S-UTT
UMR STMR-CNRS, Troyes, France

[2]Laboratoire I3S
UMR7271-UNS CNRS, Sophia-Antipolis, France

guoliang.zhu@utt.fr, lionel.fillatre@i3s.unice.fr, igor.nikiforov@utt.fr

Résumé – L'axe ferroviaire est défini dans la base de données à bord du train par une ligne polygonale avec un certain niveau d'incertitude. Le but de ce papier est d'estimer la distance parcourue, la vitesse et l'accélération en utilisant un récepteur GNSS à bas coût et d'étudier l'impact de l'incertitude sur ces estimations. L'algorithme des moindres carrés (LS) à base d'un bloc de mesures GNSS et d'un modèle dynamique réaliste du train est conçu pour estimer la distance parcourue, la vitesse et l'accélération du train. L'erreur moyenne et le second moment de ces estimations sont calculés de façon théorique et comparés avec les résultats de simulations Monte-Carlo.

Abstract – The railway centerline is defined by a polygonal line with some level of uncertainty in the train onboard database. The goal of this paper is to estimate the travelled distance, velocity and acceleration by using a low-cost GNSS receiver and to study the impact of the centerline uncertainty on these estimations. The Least Square (LS) algorithm based on a block of GNSS measurements and a realistic dynamical train model is designed to estimate the travelled distance, velocity and the acceleration of the train. The mean error and the second order moment are theoretically calculated for these estimations and compared with the results of Monte-carlo simulations.

1 Introduction and Motivation

Let us consider the following situation often arising in the regression analysis: the matrix of regressors is not perfectly known and the responses are contaminated by an additional random noise which cannot be assumed in the theoretical model. The practical needs of signal processing for surface GNSS navigation are reduced to the solution of the weighted least squares problem in the above mentioned situation. Safe and precise train positioning is essential for maintaining the safety and efficiency of the railways operating system. Hence, some desired parameters such as the travelled distance and the speed should be estimated with a high level of accuracy [1]-[2], reliability and integrity [3]-[5].

The algorithms described in this paper are devoted to the trajectory-aided train positioning by a low-cost GNSS receiver. The railway centerline geometry provides the users with some very reliable *a priori* information on the smooth character of the train trajectory. But this information is available within an additional database of noisy measurements. To estimate the train travelled distance, speed and acceleration, an optimal integration of the database information with the real-time GNSS measurements is necessary. Two crucially important questions arise: i) what is the impact of such a railway centerline geometry imprecision on the estimation of train speed, acceleration and distance? ii) will a change of acceleration (jerk) cause imprecise estimation of the train speed, acceleration and distance? Dealing with these questions leads us to consider a non-linear regression model with random disturbances caused by data uncertainties.

This paper is organized as follows. Section 2 is devoted to the problem statement and contribution. Section 3 provides the geometric design of railway centerline model, the train dynamical model and the method of train distance, velocity and acceleration estimation. The impact of the centerline uncertainty on these three estimations is discussed in this section 4. Simulation results are shown in section 5. Finally, some conclusions are drawn in section 6.

2 Problem Statement and Contribution

Let us consider that the train runs along the train track with a variable speed. The contribution of this paper

*The authors would like to thank the China Scholarship Council (CSC) and the University of Technology of Troyes (UTT), in France, for supporting their research.

is two fold. First, assuming that the train acceleration is constant over a short time period, the train travelled distance, speed and acceleration are estimated simultaneously by using a block of GNSS measurements and the train dynamical model. Using a block of measurements is necessary to overcome the ill-posed nature of the estimation problem. Second, the negative impact of the railway centerline uncertainty on the mean error and on the second order moment of these three estimations is estimated. The equations for first two moments of these three estimations are obtained and compared with the results of Monte-Carlo simulations.

3 Description of Models

3.1 Railway Centerline Model

Let us assume that the railway centerline is approximated by a polygonal line (piecewise linear curve), which represents a connected series of line segments in the Earth-centered, Earth-fixed coordinates. More formally, the railway centerline is defined by a sequence of vertices $Z_0, Z_1, Z_2, \ldots, Z_n$, $Z_i \in \mathbb{R}^3$, so that the curve consists of the line segments connecting the consecutive vertices. It is assumed that the errors related with such an approximation of the vector function $\ell \mapsto X(\ell)$, $\ell \in \mathbb{R}$, $X \in \mathbb{R}^3$, defining the railway centerline is negligible for our study. Here and in the rest of the paper, ℓ denotes the curvilinear abscissa, or the covered distance, and $\lambda = \|Z_{j+1} - Z_j\|_2 = \text{const}$ is the distance between two adjacent vertices, respectively.

Unfortunately, the on-board database uses an imprecise information about the positions of vertices, namely : $\widetilde{Z}_0, \widetilde{Z}_1, \widetilde{Z}_2, \ldots, \widetilde{Z}_n$. The quantity $\xi_i = Z_i - \widetilde{Z}_i$ defines the knowledge uncertainty concerning the train track (see Fig. 1). To simplify the presentation, a two-dimensional train trajectory is considered.

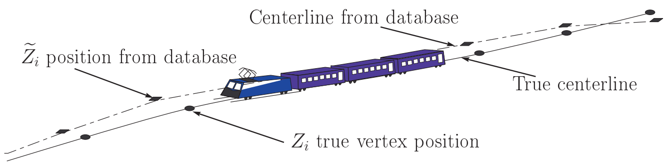

FIG. 1: Railway centerline model.

3.2 Train Dynamical Model

The train dynamical model is described by an equation formulated in terms of the covered distance, speed and acceleration (see Fig. 2). Let $\Delta t = t_k - t_{k-1}$ be the GNSS sampling interval and t_k denotes the instant of the k-th measurement. Let us consider a short time period of length $T = (q+1) \cdot \Delta t$ where q is a positive integer. Over this time period, the distance ℓ_k covered by the train, its speed v_k and its acceleration a_k at instant t_k $(1 \le k \le q)$ are given as follows

$$\begin{cases} \ell_k = \ell_{k-1} + v_{k-1} \cdot \Delta t + \dfrac{1}{2} a_{k-1} \cdot \Delta t^2 \\ v_k = v_{k-1} + a_{k-1} \cdot \Delta t \\ a_k = a_{k-1}. \end{cases} \tag{1}$$

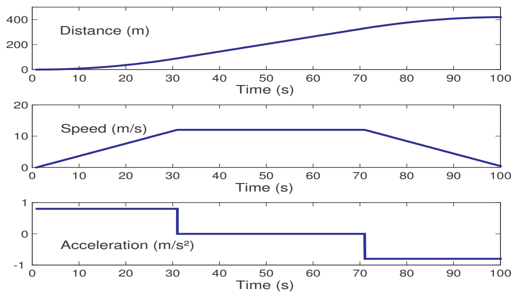

FIG. 2: Typical train motion diagram.

To simplify the notations, it is assumed that $\Delta t = 1$ (s). Let us consider the block of $q+1$ last GNSS measurements at time instant t_k. Assuming that the acceleration a_k is constant during T (s), the train position is given by

$$X(\ell_{k-q+p}) = Z_{j-1} + E_j \cdot [\ell_{k-q+p} - \lambda \cdot (j-1)], \tag{2}$$

where $p = 0, 1, \ldots, q$ and $E_j = (e_x^j, e_y^j, e_z^j)^T = \dfrac{1}{\lambda}(Z_{j+1} - Z_j)$ is the directional vector corresponding to the segment number j, $\|E_j\|_2 = 1$. The current segment number $j = j(k-q+p)$, viewed as a function of $k-q+p$, is calculated as: $j(k-q+p) = \min\{j \in \mathbb{N} | j \ge \lfloor \ell_{k-q+p}/\lambda \rfloor\}$ where $\mathbb{N}$ is the set of natural numbers. The distance ℓ_{k-q+p} is given as

$$\ell_{k-q+p} = \begin{pmatrix} 1 & (p-q) & \frac{1}{2}(p-q)^2 \end{pmatrix} \begin{pmatrix} \ell_k \\ v_k \\ a_k \end{pmatrix} = \omega_p \cdot \theta_k, \tag{3}$$

where θ_k must be estimated.

3.3 Exact and Imprecise Pseudo-range Measurement Model

Suppose that there are n satellites located at the known positions $X_i^s = (x_i, y_i, z_i)^T, i = 1, \ldots, n$. The pseudo-range r_{k-q+p}^i from the i-th satellite to the train can be written as:

$$\begin{aligned} r_{k-q+p}^i &= d_{k-q+p}^i + c b_r^{k-q+p} + \varepsilon_{k-q+p}^i \\ &= \|X(\ell_{k-q+p}) - X_i^s\|_2 + c b_r^{k-q+p} + \varepsilon_{k-q+p}^i, \end{aligned} \tag{4}$$

where b_r^{k-q+p} is a user clock bias, $c \simeq 2.9979 \cdot 10^8$m/s is the speed of light and $\varepsilon_{k-q+p}^i \sim \mathcal{N}(0, \sigma^2)$ is the pseudo-range noise at time $k-q+p$. By linearizing the pseudo-range equation around the working point $\ell_{k-q+p,0}$, we get

$$\begin{aligned} r_{k-q+p}^i - d_{k-q+p,0}^i &\simeq h_{k-q+p,0}^i \cdot (\ell_{k-q+p} - \ell_{k-q+p,0}) \\ &\quad + c b_r^{k-q+p} + \varepsilon_{k-q+p}^i, \end{aligned} \tag{5}$$

where the working point is calculated as $\ell_{k-q+p,0} = \omega_p \cdot \theta_{k,0} = \omega_p \cdot \hat{\theta}_{k-1}$, $d^i_{k-q+p,0}$ is the distance from the working point to the i-th satellite and $h^i_{k-q+p,0}$ is the coefficient of the Jacobian matrix. The above mentioned linearized measurement equations can be rewritten in the following matrix form

$$R_{k-q+p} - D_{k-q+p,0} \simeq H_{k-q+p,0} \cdot (\ell_{k-q+p} - \ell_{k-q+p,0}) + \mathbf{1}_n \cdot cb_r^{k-q+p} + \Xi_{k-q+p}, \qquad (6)$$

where $\mathbf{1}_n$ is a vector of dimension n whose each element is one. Substituting equation (3) into equation (6) and the final pseudo-range measurement model is rewritten in the matrix form as

$$R^k - D_0^k + Y_0^k \simeq H_0^k \cdot \beta_k + \Xi^k, \qquad (7)$$

where the vector $\beta_k = (\theta_k^T, cb_r^k, \cdots, cb_r^{k-q})^T$ is unknown and must be estimated.

Since the true vertex position Z_j is unknown and only its imprecise estimation $\widetilde{Z}_j$ is available, let us assume that the random vector $\xi_j = Z_j - \widetilde{Z}_j$ is assumed to be uniformly distributed in the cube $[-b, b]^3$ with $b > 0$. So an unprecise measurement model is calculated as

$$R^k - \widetilde{D}_0^k + \widetilde{Y}_0^k \simeq \widetilde{H}_0^k \cdot \beta_k + \Xi^k, \qquad (8)$$

where $\widetilde{D}_0^k$, $\widetilde{Y}_0^k$, and $\widetilde{H}_0^k$ are calculated exactly as in the above equation but with the vector $\widetilde{Z}_j$, $\widetilde{E}_j = \dfrac{\widetilde{Z}_{j+1} - \widetilde{Z}_j}{\|\widetilde{Z}_{j+1} - \widetilde{Z}_j\|_2}$ instead of Z_j, E_j.

4 Impact of Centerline Uncertainty on the LS Estimator

The goal of this section is to study the impact of the railway centerline uncertainty ξ_j on the first and second moments of the LS estimator $\hat{\beta}_k$. The measurement equation (8) can be written as follows:

$$Y^k + \Delta Y^k \simeq (H_0^k + \Delta H^k) \cdot \beta_k + \Xi^k,$$

where $Y^k = R^k - D_0^k + Y_0^k$ are the responses, $\Delta Y^k = D_0^k - \widetilde{D}_0^k - Y_0^k + \widetilde{Y}_0^k$ and $\Delta H^k = \widetilde{H}_0^k - H_0^k$ denote the data uncertainties in the regression model. We follow here the analysis of the data uncertainty impact on the LS estimators developed in [6]. The LS estimator is given by

$$\widehat{\beta}_k = \left[\left(H_0^k + \Delta H^k\right)^T \left(H_0^k + \Delta H^k\right)\right]^{-1} \left(H_0^k + \Delta H^k\right)^T \cdot \left(Y^k + \Delta Y^k\right). \qquad (9)$$

Since the random vector ΔY^k acts in the same way as the pseudo-range noise Ξ^k, the two errors can be considered together. After expanding $\left[\left(H_0^k + \Delta H^k\right)^T \left(H_0^k + \Delta H^k\right)\right]^{-1}$ around H_0^k and computing the expectation of (9), the mean error is

$$\mathbb{E}(\widehat{\beta}_k - \beta_k) = \left(\overline{B}^k\right)^{-1} \left[\left(\overline{H}_0^k\right)^T C - F + G\right] \beta_k, \qquad (10)$$

where $\overline{B}^k = \left(\overline{H}_0^k\right)^T \overline{H}_0^k$. The matrices $\overline{H}_0^k$ are calculated exactly as in equation (7) but with the working point $\ell_{k-q+p,0} = \omega_p \cdot \theta_{k-1}$. The matrix functions of second moments $G = \mathbb{E}\left[\left(\Delta H^k\right)^T \overline{H}_0^k \left(\overline{B}^k\right)^{-1} \left(\overline{H}_0^k\right)^T \Delta H^k\right]$, $C = \mathbb{E}\left[\Delta H^k \left(\overline{B}^k\right)^{-1} \left(\overline{H}_0^k\right)^T \Delta H^k\right]$ and $F = \mathbb{E}\left[\left(\Delta H^k\right)^T \Delta H^k\right]$ underline only the impact of data uncertainties in the matrix of regressors on the mean error of the LS estimator.

Let us now define $\gamma_{\ell,m} = \mathbb{E}(\Delta H_\ell \Delta H_m^T)$. The matrix C is given by

$$C = \begin{pmatrix} \sum\limits_{u=0}^{q} \gamma_{k,k-u} \chi_{1,u+1}^T \omega_q^T \omega_{q-u} & \mathbf{0}_{n,(q+1)} \\ \sum\limits_{u=0}^{q} \gamma_{k-1,k-u} \chi_{1,u+1}^T \omega_{q-1}^T \omega_{q-u} & \mathbf{0}_{n,(q+1)} \\ \vdots & \vdots \\ \sum\limits_{u=0}^{q} \gamma_{k-q,k-u} \chi_{1,u+1}^T \omega_0^T \omega_{q-u} & \mathbf{0}_{n,(q+1)} \end{pmatrix},$$

where $\chi_{1,u+1}^T$ and $\mu_{2,u+1}^T$ are two matrices of size $3 \times n$ and $(q+1) \times n$, respectively, extracted from the following matrix :

$$\left(\overline{B}^k\right)^{-1} \left(\overline{H}_0^k\right)^T = \begin{pmatrix} \chi_{1,1} & \chi_{1,2} & \cdots & \chi_{1,(q+1)} \\ \mu_{2,1} & \mu_{2,2} & \cdots & \mu_{2,(q+1)} \end{pmatrix}.$$

Finally, let us define the following matrices

$$F = \begin{pmatrix} \sum\limits_{u=0}^{q} \operatorname{tr} \gamma_{k-u,k-u} \omega_{q-u}^T \omega_{q-u} & \mathbf{0}_{3,(q+1)} \\ \mathbf{0}_{(q+1),3} & \mathbf{0}_{(q+1),(q+1)} \end{pmatrix},$$

$$G = \begin{pmatrix} \sum\limits_{u=0}^{q} \sum\limits_{v=0}^{q} \operatorname{tr} Q_{u+1,v+1} \gamma_{k-v,k-u} \omega_{q-u}^T \omega_{q-v} & \mathbf{0}_{3,(q+1)} \\ \mathbf{0}_{(q+1),3} & \mathbf{0}_{(q+1),(q+1)} \end{pmatrix},$$

where $Q_{u+1,v+1}$ is a block of size $n \times n$ of the following $(q+1)n \times (q+1)n$ matrix $Q = \overline{H}_0^k \left(\overline{B}^k\right)^{-1} \left(\overline{H}_0^k\right)^T$:

$$Q = \begin{pmatrix} Q_{1,1} & Q_{1,2} & \cdots & Q_{1,(q+1)} \\ Q_{2,1} & Q_{2,2} & \cdots & Q_{2,(q+1)} \\ \vdots & \vdots & \ddots & \vdots \\ Q_{(q+1),1} & Q_{(q+1),2} & \cdots & Q_{(q+1),(q+1)} \end{pmatrix}.$$

After expanding and ignoring the terms of order $(\Delta H^k)^2$ and under the assumption that the uncertainties ΔH^k are reasonably small, the second moment of $\hat{\beta}_k$ is given by

$$\mathbb{E}(\widehat{\beta}_k - \beta_k)(\widehat{\beta}_k - \beta_k)^T = \sigma^2 I + \left(\overline{B}^k\right)^{-1} \left(\overline{H}_0^k\right)^T (\Sigma_Y - N - N^T + M) \overline{H}_0^k \left(\overline{B}^k\right)^{-1}, \qquad (11)$$

where the matrix functions $\Sigma_Y = \mathbb{E}\left[\Delta Y^k \left(\Delta Y^k\right)^T\right]$, $M = \mathbb{E}\left[\Delta H^k \beta_k \beta_k^T \left(\Delta H^k\right)^T\right]$ and $N = \mathbb{E}\left[\Delta Y^k \beta_k^T \left(\Delta H^k\right)^T\right]$ underline the impact of the data uncertainties in the matrix of regressors and responses on the second order moment of the LS estimator. After calculating, we get

$$M = \begin{pmatrix} M_{1,1} & M_{1,2} & \cdots & M_{1,(q+1)} \\ M_{2,1} & M_{2,2} & \cdots & M_{2,(q+1)} \\ \vdots & \vdots & \ddots & \vdots \\ M_{(q+1),1} & M_{(q+1),2} & \cdots & M_{(q+1),(q+1)} \end{pmatrix}$$

and

$$N = \begin{pmatrix} N_{1,1} & N_{1,2} & \cdots & N_{1,(q+1)} \\ N_{2,1} & N_{2,2} & \cdots & N_{2,(q+1)} \\ \vdots & \vdots & \ddots & \vdots \\ N_{(q+1),1} & N_{(q+1),2} & \cdots & N_{(q+1),(q+1)} \end{pmatrix},$$

where

$$M_{u+1,v+1} = \ell_{k-u}\ell_{k-v}\gamma_{k-u,k-v}$$

and

$$N_{u+1,v+1} = \ell_{k-u}\ell_{k-v,0}\gamma_{k-u,k-v} - \ell_{k-u}\mathbb{E}(\Delta H_{k-u}\Delta D_{k-v}^T)$$

are two blocks of size $n \times n$ of matrices M and N, respectively, ℓ_{k-u}, ℓ_{k-v} are calculated exactly as in equation (3) and $u, v = 0, ..., q$.

5 Numerical Simulations

The simulation scenario is shown in Fig. 2. The comparison of the theoretical moments for the estimated distance, speed and acceleration with the results of a 10^4-repetition Monte-Carlo simulation, is shown in Fig. 3-4. The standard GNSS constellation has been used with $n = 6$ visible satellites and with the pseudo-range SD $\sigma = 2$ (m). The distance between two adjacent vertices has been chosen $m = 50$ (m). The centerline uncertainty b and the sampling period q have been chosen of 2 (m) and 20 samples, respectively. The true acceleration during the acceleration, free-running and braking period is 0.8 m/s^2, 0 m/s^2 and -0.8 m/s^2.

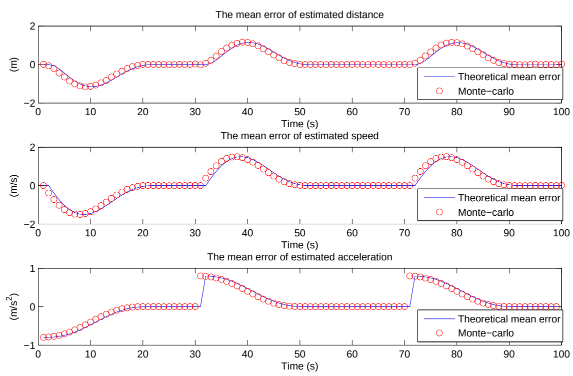

FIG. 3: The mean error of the estimated distance, speed and acceleration for the centerline uncertainty of $\xi_j \in [-2, 2]^2$.

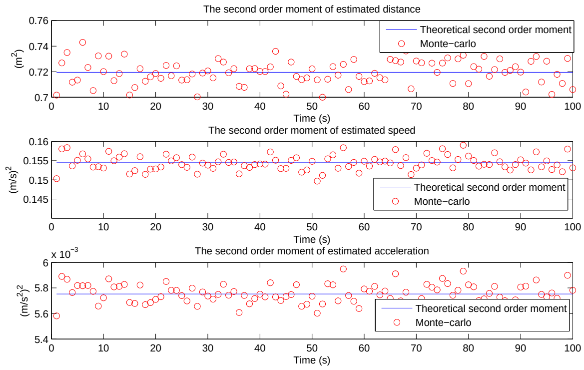

FIG. 4: The second order moment of the estimated distance, speed and acceleration for the centerline uncertainty of $\xi_j \in [-2, 2]^2$.

6 Conclusions

The results show that the mean error of estimated distance, speed and acceleration obtained by GNSS is always practically unbiased, even with an imprecise geometric model of the railway centerline. They also show that the change of acceleration causes an imprecise estimation of the travelled distance, speed and acceleration only for a short time period. Due to the existence of jerk, the mean error is obviously biased during this short time period, but the second order moment remains almost unchanged.

References

[1] Zhu, G., Fillatre, L., Nikiforov, I. *Impact of the railway centerline geometry uncertainties on the train velocity estimation by GPS.* 5th International Conference on Personal Satellite Services, Toulouse, France, 2013.

[2] Zhu, G., Fillatre, L., Nikiforov, I. *Estimation of the train travelled distance, velocity and acceleration by using GPS signals.* The European Navigation Conference, Vienna, Austria, 2013.

[3] Nikiforov, I., Choquette, F. *Integrity Equations for Safe Train Positioning Using GNSS.* Istituto Italiano di Navigazione, vol.171 pp. 52-77, 2003.

[4] Lacresse, H., Grall, A., and Nikiforov, I. *Statistical fault detection with linear or nonlinear nuisance parameters.* Safeprocess 2003, IFAC Conference, Washington, D.C. 6 p, 9-11 June 2003.

[5] Lacresse, H., Grall, A., and Nikiforov, I. *Fault detection with non-linear nuisance parameters and safe train navigation.* Proc. 16th IFAC World Congress, Prague. 1-6 p, 3-8 July 2005.

[6] Hodges, S.D., Moore, P.G. *Data Uncertainties and Least Squares Regression.* JRSS, Series C (Applied Statistics), vol.21 pp.185-195, 1972.