2018 26th Mediterranean Conference on Control and Automation (MED)
Zadar, Croatia, June 19-22, 2018

Side-slip angle estimation of autonomous road vehicles based on big data analysis

Dániel Fényes, Balázs Németh, Máté Asszonyi, Péter Gáspár

Abstract— **The paper proposes a side-slip angle estimation method for autonomous road vehicles using a big data approach. The core of the solution is that on-board signals of numerous autonomous vehicles are available, which can be used to generate the side-slip estimator. The estimation is based on the ordinary linear regression method with OLS subset selection using large amount of data collected from the car sensors. The advantage of the proposed solution is that the numerically complex data mining operations are performed off-line, while the side-slip angle estimation of the vehicle using only its own on-board signals requires low computation effort. The efficiency of the estimation is presented through several CarSim simulations, in which the parameters of the vehicle and the road are varied. Moreover, the method is compared to the simulation results of a sensor fusion based Kalman filtering method.**

I. INTRODUCTION AND MOTIVATION

The research in the field of autonomous vehicles and intelligent transportation systems has had a fast-growing tendency in the last years. Several research institutes have focused on the new challenges posed by autonomous vehicles, while in some cases the conventional vehicle dynamical problems are extended with new perspectives. For side-slip angle estimation several methods have been developed. The conventional instruments of vehicle dynamics to measure the side-slip angle are optical sensors. Although they provide high accuracy, these sensors have relatively high costs [1]. Therefore, in several research projects filtering methods and observers are designed to estimate the side-slip angle, such as the Kalman filter, the Luenberger method and sliding mode observers, see [2], [3].

Owing to the numerous available measurements of the semi-autonomous and automated vehicles the data-based approaches can also be sufficiently used in the future. Data mining techniques can be used to find relations between the various measurements and the unmeasured signals. Big data were used in the prediction of vehicle slip through the combination of individual measurements of the vehicle and database information [4]. In [5] a layered neural network was developed to compute the side-slip angle. The inputs of the neural network were the yaw-rate and the lateral acceleration. An artificial neural network method for slip estimation using acceleration, velocity, inertial and steering angle information was also proposed in [6]. Moreover, in [7] an adaptive neuro-fuzzy inference system approach was applied with various signal measurements. Another formulation of the neural networks, such as the general regression for the side-slip angle estimation, was used in [8]. The Kalman filtering based slip angle estimation method was improved with decision trees in [9]. The decision algorithm was used to classify the inertial data and the error covariance parameter of the filter to compensate for changes in disturbances and nonlinearities. Although the neural network based methods can provide efficient estimation, the implementation and the adaptive training of the networks can be difficult. Moreover, the precise estimation using Kalman filtering can require sensor fusion, which is based on accurate GPS measurements. However, the solutions with GPS modules suffer from the loss of signals in urban locations and tunnels [10].

In this paper a novel application of the linear regression method on the side-slip angle estimation problem is proposed. The approach is based on big data analysis, which requires numerous signals from the automated vehicles. As a first step, the applied estimation algorithm, i.e., the ordinary linear regression method with OLS (Ordinary Least Square) subset selection, creates several different linear models using various sets of sensor signals. In the second step this algorithm compares the created models to the true model and selects the best one by determining their predictive accuracy. The advantage of the method is that it requires little on-line computation, while the complex operations are solved off-line. Moreover, in the estimation method only the onboard signals of the vehicle are used, which are available without a loss in communication. The presented method is analyzed through several simulations and the results are compared with a Kalman filtering method based on the sensor fusion of GPS and Inertial Measurement Unit (IMU).

The structure of the paper is the following. Section II proposes the method of data collection and the mining of the linear regression information. The results are demonstrated in Section III through the variation of different parameters, such as velocity, mass and sensor noises. Moreover, Section IV presents a comparison with a Kalman filtering based

D. Fényes, B. Németh, M. Asszonyi and P. Gáspár are with Systems and Control Laboratory, Institute for Computer Science and Control, Hungarian Academy of Sciences, Kende u. 13-17, H-1111 Budapest, Hungary. E-mail: [daniel.fenyes;balazs.nemeth;peter.gaspar]@sztaki.mta.hu

This work was supported by the GINOP-2.3.2-15-2016-00002 grant of the Ministry of National Economy of Hungary.

estimation method. Finally, the contributions of the paper and the further challenges are summarized in Section V.

II. Estimation method using big data approach

In this paper the ordinary linear regression method with OLS subset selection is used to produce prediction models. The bases of both methods are briefly described in this section. [11]

A. Linear regression

Let's consider a dataset with n independent instances, k input variables and an output variable. The instances are written in the form of an $n \times k$ design matrix X. Moreover, let ζ^* be the parameter vector of the true model and then the output vector y can be determined as

$$y = X\zeta^* + \epsilon \tag{1}$$

where ϵ is the noise vector whose elements are sampled from $N(0, \sigma^2)$. It is assumed that σ^2 is known or, at least, it can be estimated ($\hat{\sigma}^2$). $\mathcal{M}(\zeta)$ denotes a fitted, linear model that has an unique parameter vector ζ while the true model is denoted by $\mathcal{M}(\zeta^*)$. The aim of the modeling task is to find a model from the entire model space $\mathbb{M} = \{\mathcal{M}(\zeta) : \zeta \in \mathbb{R}^k\}$ whose predictive accuracy is the greatest on the given dataset. The models can be produced by numerous algorithms such as OLS method, OLS subset selection, shrinkage, Akaike/Bayesian/Copula/Residual Information Criterion (AIC/BIC/CIC/RIC) methods etc, see [11], [12]. These methods can reduce the dimension of the models by discarding the redundant variables. In order to evaluate a model, the distance between the current model and the true model must be known. This distance can be calculated as

$$\mathcal{D}(\mathcal{M}(\zeta^*), \mathcal{M}(\zeta)) = \frac{||y_{\mathcal{M}(\zeta^*)} - y_{\mathcal{M}(\zeta)}||^2}{\sigma^2} \tag{2}$$

where $||\cdot||$ denotes the $\mathcal{L}_2$ norm and σ^2 can be replaced by its estimated value $\hat{\sigma}^2$.
The final task is to determine a model which minimizes this expression.

$$\mathcal{D}(\mathcal{M}^*, \mathcal{M}) = min! \tag{3}$$

B. OLS subset selection method

As mentioned above, the basic concept of the OLS subset selection method is to create subset models using various sets of variables. If a dataset has k variables, $k+1$ nested models ($\mathcal{M}_j$) can be created, where $j = 0$ is the null model with zero variables and $j = k$ is the full model with all of the variables. In this case, an estimate of the parameter vector of $\mathcal{M}_j$ can be determined as

$$\hat{\zeta}_{\mathcal{M}_j} = (X'_{\mathcal{M}j} X_{\mathcal{M}j})^{-1} X'_{\mathcal{M}j} y \tag{4}$$

where $X_{\mathcal{M}_j}$ is the $n \times j$ design matrix and let $\mathcal{P}_{\mathcal{M}_j} = X_{\mathcal{M}j}(X'_{\mathcal{M}j} X_{\mathcal{M}j})^{-1} X'_{\mathcal{M}j}$ be an orthogonal projection matrix from the original space (k) onto the reduced space (j). Finally, $\hat{y}_{\mathcal{M}_j} = \mathcal{P}_{\mathcal{M}_j} y$ is the estimate of $y^*_{\mathcal{M}_j} = \mathcal{P}_{\mathcal{M}_j} y^*$
Since this method creates only $k+1$ subset models (instead of 2^k, which is computationally unfeasible at increased k), the predefined order of the variables is a crucial point of this method. There are a set of ranking algorithms, which help determine the best predefined order of the variables, see [13] [14]

III. Demonstration of the estimation methods

The first step of building an estimation model is the acquisition of the appropriate training data. In this paper, the training data are collected from simulations using the high-fidelity simulation software CarSim. In the simulations, a D-class sedan passenger car has been used, whose sprung mass is $1320kg$. The car has been driven along the Michigan Waterford Hill course several times at various longitudinal velocities. The following attributes of the passenger car have been measured:

1) Longitudinal acceleration a_x
2) Lateral acceleration a_y
3) Yaw rate $\dot{\psi}$
4) Angular velocity of wheels
   W_{ij} $(i \in \{front, rear\},\ j \in \{left, right\}$
5) Steering angle of front wheels
   δ_i, $(i \in \{left, right\})$
6) Angle of steering wheel δ_s
7) Side-slip angle β

The sample time has been $0.01s$ and in this way, more than 2 million instances have been collected.
Using various sets of the variables from the collected training data, 5 different estimation models have been built, as listed in I Table. The performances of the models are evaluated by the k-fold cross-validation technique (see [15]), whose results can also be seen in Table I.

TABLE I
PROPERTIES OF ESTIMATION MODELS

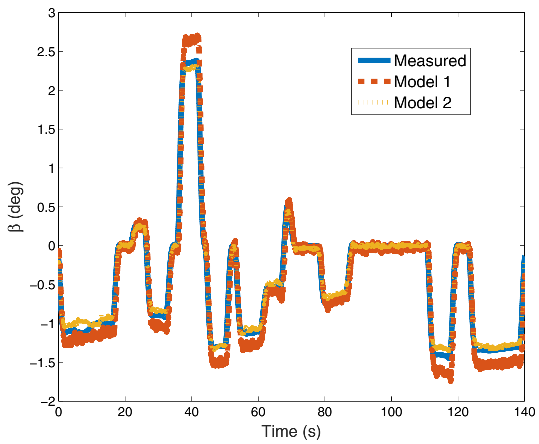
Model no.	Used variables	Corr. coeff.
1	a_x,a_y,$\dot{\psi}$	0.9013
2	a_x,a_y,$\dot{\psi}$,δ_s	0.9534
3	a_x,a_y,$\dot{\psi}$,δ_s,$W_{f,l}$,$W_{f,r}$	0.9796
4	a_x,a_y,$\dot{\psi}$,δ_s,$W_{f,l}$,$W_{f,r}$,$W_{r,l}$,$W_{r,r}$	0.9911
5	a_x,a_y,$\dot{\psi}$,δ_l,δ_r,$W_{f,l}$,$W_{f,r}$,$W_{r,l}$,$W_{r,r}$	0.9911

Table I shows that the last two models have the best predictive accuracy. Not surprisingly, these models have exactly the same accuracy, since the two steering angles (δ_l, δ_r) and the angle of the steering wheel δ_s are closely related to each other. Figure 1 illustrates a test scenario in which the car is driven along the test track at various velocities according to the Figure 2. Moreover, all of the models are used to predict the side-slip angle, see Figure 1. Table II summarizes the statistical parameters of the models, such as mean and standard deviation.

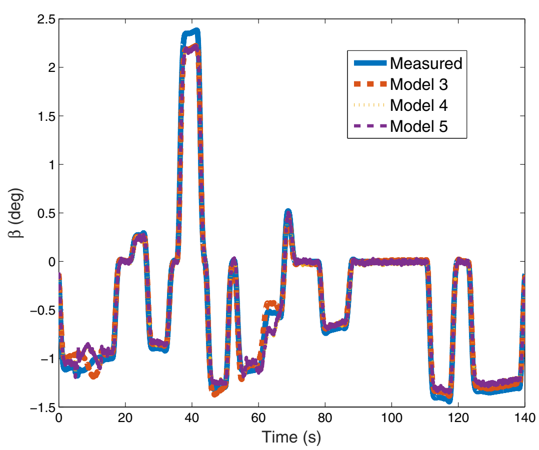
TABLE II

PREDICTIVE ACCURACY OF MODELS

Model number	Mean	Standard deviation
1	0.0135	0.3018
2	−0.0146	0.1754
3	−0.0209	0.1556
4	−0.0060	0.1549
5	−0.0079	0.1505



(a) Model 1 and Model 2



(b) Models 3, Model 4 and Model 5

Fig. 1. Estimation of the side-slip angle

It is shown that all models predict the side-slip angle accurately. Apart from a short section ($65 - 70s$), the best prediction is given by the last two models (Model 4 and Model 5), therefore the simpler one of these models is used in the rest of the paper. In the following, the accuracy and the capability of the selected predictive model is examined in different situations.

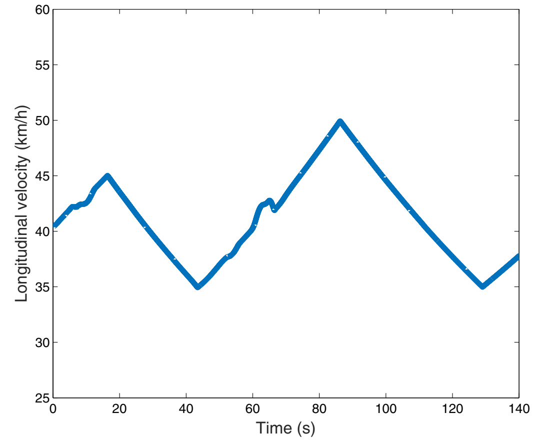


Fig. 2. Velocity during the test scenario

A. Increased noise

In the first situation the impact of the increasing noise on the predictive accuracy is examined. In practice, the inertial sensors and the gyroscope are significantly affected by sensor noises while the angular velocity of the wheels and the angle of the steering wheel can be relatively well measured. The first two sensors are disturbed by white noises whose initial parameters are the following:

- Inertial sensor: $bias = 0.15\frac{m}{s^2}$, $variance = 0.1^2$
- Gyroscope: $bias = 0.01\frac{rad}{s}$, $variance = 0.01^2$

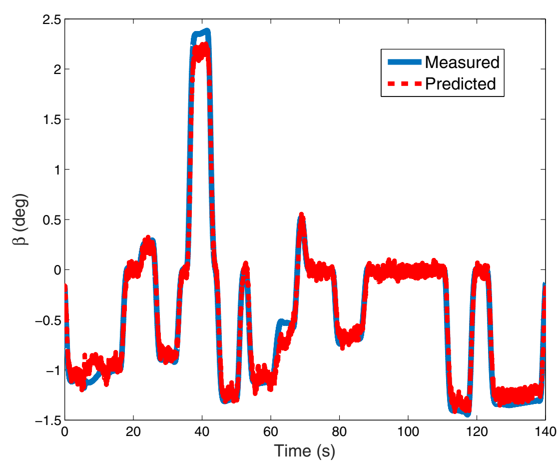


Fig. 3. Effect of the increased noise on the prediction

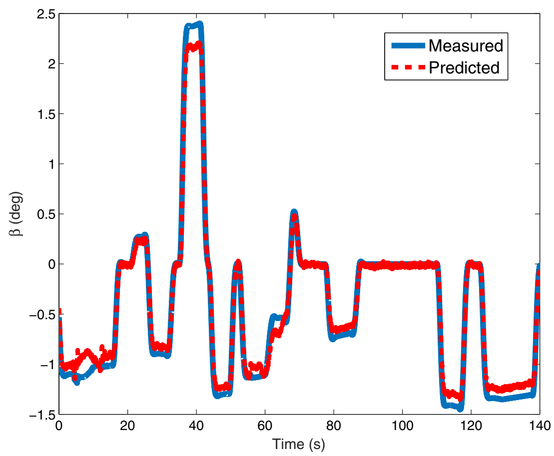
In this scenario these parameters are changed to:

- Inertial sensor: $bias = 0.3\frac{m}{s^2}$, $variance = 0.2^2$
- Gyroscope: $bias = 0.02\frac{rad}{s}$, $variance = 0.02^2$

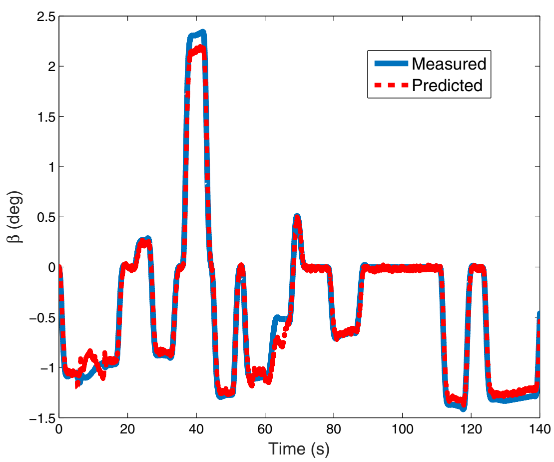
Figure 3 shows the result of the simulation. It can be seen that the variation of the noise has only a slight effect on the prediction. The estimated model predicts the side-slip angle accurately despite the increased noises.

B. Variation of mass

Secondly, the effect of the variation of the car mass on the accuracy of the prediction is investigated. The initial mass of the passenger car is $m = 1320kg$. Firstly, the mass of the car is reduced to $m = 1000kg$ and then it is increased to $m = 1740kg$. The results of the changes can be seen in Figure 4. Apart from a short section, the applied model has high predictive accuracy, which means that the variation of the mass has no significant influence on the prediction. Therefore, the calculated model can resist the change in the mass.



(a) $m = 1000kg$

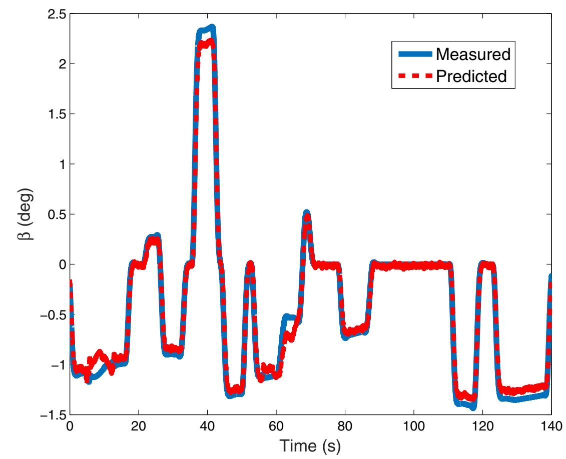


(b) $m = 1740kg$

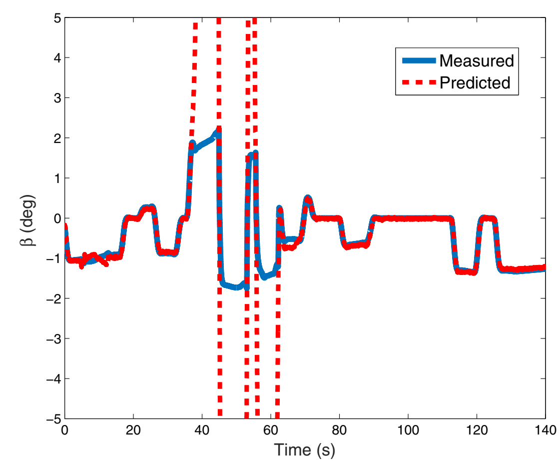
Fig. 4. Effect of the variation of mass on the prediction

C. Variation of the adhesion coefficient

In the third case the variation of the adhesion coefficient μ is simulated. The initial value of the adhesion coefficient is $\mu = 1$. In the simulations its value is decreased to $\mu = 0.7$ and then to $\mu = 0.4$. The results of the simulations are shown in Figure 5. In the case of $\mu = 0.7$ it can be seen that the model predicts the side-slip angle as accurately as in the normal case, see Figure 1. In the case of $\mu = 0.4$ the model also operates appropriately apart from a section between $40-60s$. In this section the vehicle loses its stability, since its longitudinal velocity is too high and the adhesion coefficient is too low to follow the defined path. Nevertheless, the applied model operates accurately in all other sections.



(a) μ=0.7



(b) μ=0.4

Fig. 5. Effect of the variation of the adhesion coefficient on the prediction

D. Melbourne Grand Prix Circuit

Finally, the predictive model is tested on another track. Since all of the regression models can be easily overfitted, it is important to guarantee that the calculated model operates in other cases. Therefore, in this case the passenger car is driven along the Melbourne Grand Prix Circuit at various velocities. Figure 6 shows the variation of the velocity during the simulation. The side-slip angle and its prediction are shown in Figure 7. The simulation shows that the applied model is able to predict the side-slip angle accurately. Its predictive accuracy is close to the normal case. It means that the proposed predictive model is generally able to predict the side-slip angle.

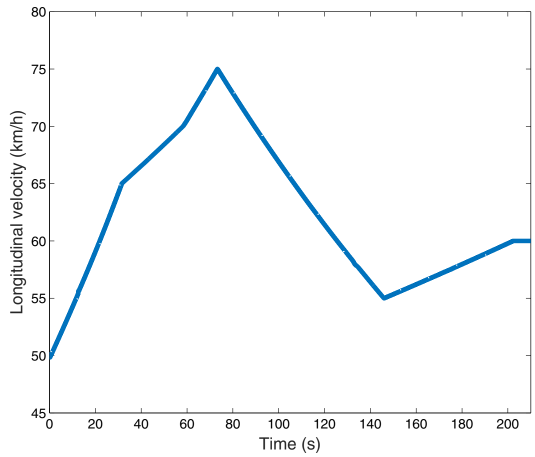


Fig. 6. Longitudinal velocity on the Melbourne track

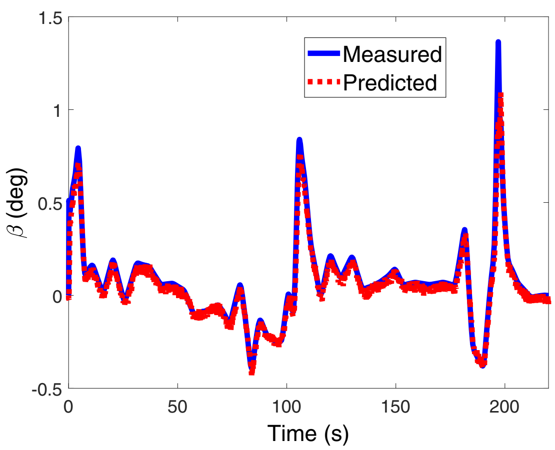


Fig. 7. Side-slip angles on the Melbourne course

IV. Comparison of the results with Kalman filtering

In this section the efficiency of the proposed estimation method is compared with the results of a Kalman filtering technique, which is based on the sensor fusion of GPS and IMU signals. The goal of the comparison is to illustrate that the big data based estimation can be an alternative method together with the conventional techniques.

The purpose of the sensor fusion is to use the precise longitudinal and lateral velocity information of the GPS module, whose sampling time is relatively high. Therefore, in the fusion the less precise longitudinal and lateral acceleration signals $a_{x,IMU}, a_{y,IMU}$ of the IMU with low sampling time are also incorporated.

The state-space representation of the system results from the following kinematic relation of the vehicle [16]

$$a_{x,IMU} = \dot{u}_x - \dot{\psi} u_y + a_{x,bias} + w_{x,IMU}, \quad (5a)$$
$$a_{y,IMU} = \dot{u}_y + \dot{\psi} u_x + a_{y,bias} + w_{y,IMU}, \quad (5b)$$

where u_x, u_y are the longitudinal and lateral velocities, $a_{x,bias}, a_{y,bias}$ represent the sensor bias and w_x, w_y are the noises on the acceleration measurements. Moreover, $\dot{\psi}$ is the yaw-rate of the vehicle, which is assumed to be calculated through another Kalman filter. Using the approximation $\dot{u}_x = (u_x(k) - u_x(k-1))/(\Delta t)$ and $\dot{u}_y = (u_y(k) - u_y(k-1))/(\Delta t)$, where Δt is the sampling time of acceleration signals, the relationships (5) are transformed in a discrete form

$$\begin{aligned} u_x(k) =& u_x(k-1) - a_{x,bias}(k)\Delta t + a_{x,IMU}(k)\Delta t + \\ &+ \dot{\psi}(k) u_y(k-1)\Delta t + w_{x,IMU}\Delta t, \end{aligned} \quad (6a)$$
$$\begin{aligned} u_y(k) =& u_y(k-1) - a_{y,bias}(k)\Delta t + a_{y,IMU}(k)\Delta t + \\ &+ \dot{\psi}(k) u_x(k-1)\Delta t + w_{y,IMU}\Delta t. \end{aligned} \quad (6b)$$

Thus, the state transition model is formed as

$$x(k) = F(k)x(k-1) + B_1 w(k) + B_2 u(k), \quad (7)$$

where the state vector is $x(k) = \begin{bmatrix} u_x(k) & a_{x,bias}(k) & u_y(k) & a_{y,bias}(k) \end{bmatrix}^T$, $F(k)$ is the state transition matrix and B_1, B_2 are incorporated in impacts of the disturbances $w(k) = \begin{bmatrix} w_{x,IMU} & w_{y,IMU} \end{bmatrix}^T$ and the measurements $u(k) = \begin{bmatrix} a_{x,IMU}(k) & a_{y,IMU}(k) \end{bmatrix}^T$ on the estimation.

Moreover, the measurement of the GPS velocity signals is used to update the state estimation:

$$u_{x,GPS} = u_x + w_{x,GPS}, \quad (8a)$$
$$u_{y,GPS} = u_y + w_{y,GPS}, \quad (8b)$$

where $u_{x,GPS}, u_{y,GPS}$ are the measured velocity values and $w_{x,GPS}, w_{y,GPS}$ are the noises.

To sum up the state estimation briefly, the following three steps result in the side-slip angle of the vehicle

1) The predicted state vector is estimated through the IMU measurement, such as

$$\hat{x}(k) = F(k)\hat{x}(k-1) + B_2 u(k), \quad (9a)$$
$$P(k) = F(k)P(k-1)F^T(k) + Q(k) \quad (9b)$$

where $\hat{x} = \begin{bmatrix} \hat{u}_x(k) & \hat{a}_{x,bias}(k) & \hat{u}_y(k) & \hat{a}_{y,bias}(k) \end{bmatrix}^T$ is the estimated vector, $P(k)$ is the predicted estimate covariance and $Q(k)$ is the covariance of the process noise.

2) The actual Kalman gain for the update is computed as

$$K(k) = P(k)H^T(k)\left(HP(k)H^T + R\right)^{-1}, \quad (10)$$

where R is the covariance matrix of the observation noise and H is the observation matrix.

3) If the GPS measurement is available, the $\hat{x}$ prediction from the IMU measurement is updated through the GPS measurement, such as

$$\hat{x}_{up}(k) = \hat{x}(k) + K(k)\left(z(k) - H\hat{x}(k)\right), \quad (11a)$$
$$P_{up}(k) = P(k)(I - K(k)H), \quad (11b)$$

where $z(k) = \begin{bmatrix} u_{x,GPS} & u_{y,GPS} \end{bmatrix}^T$ contains the GPS measurements. The results of the computation $\hat{x}_{up}(k) = \begin{bmatrix} \hat{u}_{x,up}(k) & \hat{u}_{y,up}(k) \end{bmatrix}^T$, $P_{up}(k)$ in (11) are the estimated state and the predicted covariance, which are used as $\hat{x}(k-1)$ and $P(k-1)$ in the next computation cycle. Finally, the side-slip angle of the vehicle is computed as

$$\hat{\beta}(k) = \arctan \frac{\hat{u}_{y,up}(k)}{\hat{u}_{x,up}(k)}. \quad (12)$$

If the GPS measurement is unavailable, the prediction cannot be updated. Thus, in this case the side-slip angle is computed as

$$\hat{\beta}(k) = \arctan \frac{\hat{u}_{y}(k)}{\hat{u}_{x}(k)}. \quad (13)$$

Finally, the comparison of the Kalman filter and the proposed predictive model is illustrated in Figure 8. It is shown that the Kalman filtering method has a significantly lower accuracy compared to the predictive method. Therefore, the proposed side-slip angle prediction method can be used in commercial vehicles combined together with of the Kalman filtering method.

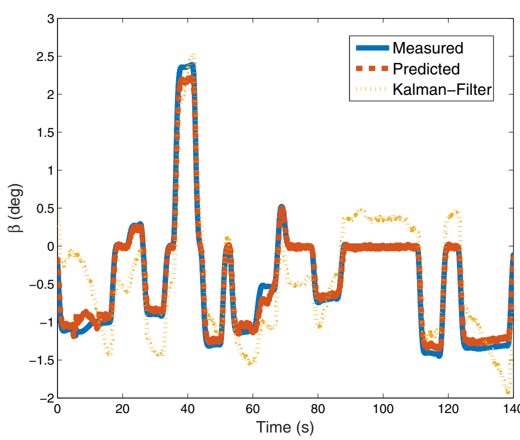


Fig. 8. Comparison of the Kalman filter method and the prediction method

V. Conclusions

The paper has proposed a data-based side-slip angle estimation method for autonomous vehicles with numerous signal measurements. In the processing of the big data from the vehicle signals the linear regression algorithm has been used. The contribution of the paper is an easily implementable algorithm with a small amount of on-line computation. The proposed method can guarantee acceptably small errors in the estimation of the side-slip angle. The efficiency of the method has been analyzed through various scenarios using the high-fidelity CarSim simulator. Moreover, the results of the method are compared to the results of a sensor fusion based Kalman filtering method. In the future the proposed method will be verified through experimental scenarios using a real test vehicle.

References

[1] M. Doumiati, A. C. Victorino, A. Charara, and D. Lechner, "Onboard real-time estimation of vehicle lateral tire road forces and sideslip angle," *IEEE/ASME Transactions on Mechatronics*, vol. 16, no. 4, pp. 601–614, Aug 2011.

[2] J. Stephant, A. Charara, and D. Meizel, "Experimental validation of vehicle sideslip angle observers," in *IEEE Intelligent Vehicles Symposium, 2004*, June 2004, pp. 150–155.

[3] P. J. Venhovens and K. Naab, "Vehicle dynamics estimation using kalman filters," *Vehicle System Dynamics*, vol. 32, no. 2-3, pp. 171–184, 1999.

[4] J. Jeon, W. Lee, H. J. Cho, and H. Lee, "A big data system design to predict the vehicle slip," in *2015 15th International Conference on Control, Automation and Systems (ICCAS)*, Oct 2015, pp. 592–596.

[5] H. Sasaki and T. Nishimaki, "A side-slip angle estmation using neural network for a wheeled vehicle," 2000.

[6] M. Kato, K. Isoda, and H. Yuasa, "Estimation of vehicle side slip angle with artificial neural network," *SAE Review*, vol. 15, pp. 79–82, 1994.

[7] B. L. Boada, M. J. L. Boada, A. Gauchía, E. Olmeda, and V. Díaz, "Sideslip angle estimator based on ANFIS for vehicle handling and stability," *Journal of Mechanical Science and Technology*, vol. 24, no. 4, pp. 1473–1481, 2015.

[8] W. Wei, B. Shaoyi, Z. Lanchun, Z. Kai, W. Yongzhi, and H. Weixing, "Vehicle sideslip angle estimation based on general regression neural network," *Mathematical Problems in Engineering*, 2016.

[9] J. Coyte, B. Li, H. Du, W. Li, D. Stirling, and M. Ros, "Decision tree assisted EKF for vehicle slip angle estimation using inertial motion sensors," in *IEEE International Joint Conference on Neural Network (IJCNN 2014)*, Beijing, China, 2014.

[10] H. Grip, L.I.Imsland, T. Johansen, J. Kalkkuhl, and A. Suissa, "Vehicle sideslip estimation. design, implementation, and experimental validation," *IEEE Control Systems Magazine*, vol. 29, no. 5, pp. 36–52, 2009.

[11] Y. Wang and I. H. Witten, *Pace Regression*. (Working paper 99/12). Hamilton, New Zealand: University of Waikato, Department of Computer Science., 1999.

[12] C. C. Rodriguez, "The ABC of model selection: AIC, BIC and the New CIC," in *AIP Conference Proceedings*, vol. 803, no. 80, 2005.

[13] R. Shibata, "An optimal selection of regression variables," *Biometrika*, vol. 68, pp. 45–54, 1981.

[14] M. L. Thompson, "Selection of variables in multiple regression," *International Statistical Society B*, vol. 46, pp. 1–21 and 129–146, 1978.

[15] K. Klauer, "Model testing and selection, theory of," *International Encyclopedia of the Social and Behavioral Sciences*, pp. 9927–9931, 2001.

[16] J. Ryu, E. J. Rossetter, and J. C. Gerdes, "Vehicle sideslip and roll parameter estimation using GPS," in *6th Int. Symposium on Advanced Vehicle Control, Hiroshima, Japan*, 2002.