

A Novel Big-data-based Estimation Method of Side-slip Angles for Autonomous Road Vehicles

Dániel Fényes, Balázs Németh and Péter Gáspár

*Institute for Computer Science and Control, Hungarian Academy of Sciences,
Kende u. 13-17, H-1111 Budapest, Hungary*

Keywords: Side-slip Estimation, Regression Analysis, Big Data, Kalman Filtering.

Abstract: In the paper a novel side-slip estimation algorithm, which is based on big data approaches, is proposed. The idea of the estimation is based on the availability of a large amount of information of the autonomous vehicles, e.g. yaw-rate, accelerations and steering angles. The significant number of signals are processed through big data approaches to generate a simplified rule for the side-slip estimation using the onboard signals of the vehicles. Thus, a subset selection method for time-domain signals is proposed, by which the attributes are selected based on their relevance. Furthermore, a linear regression using the Ordinary Least Squares (OLS) method is applied to derive a relationship between the attributes and the estimated signal. The efficiency of the estimation is presented through several CarSim simulation examples, while the WEKA data-mining software is used for the OLS method.

1 INTRODUCTION AND MOTIVATION

The spread of autonomous driving is predicted to be a future tendency of intelligent transportation systems. Several research institutes have focused on the new challenges posed by autonomous vehicles, such as environment detection and the accurate estimation of vehicle states. One of these signals is the side-slip angle, which has relevance in the evaluation of the vehicle stability. In several research projects filtering methods and observers are designed to estimate the side-slip angle, see (Stephant et al., 2004; Coyte et al., 2014). The precise estimation using Kalman filtering requires sensor fusion with GPS measurements, but these solutions suffer from the loss of signals in urban locations and tunnels (Grip et al., 2009).

Therefore, several further techniques have been published in the literature. Big data were used in the prediction of vehicle slip through the combination of individual measurements of the vehicle and database information (Jeon et al., 2015). In (Sasaki and Nishimaki, 2000) a layered neural network was developed to compute the side-slip angle. An artificial neural network method for slip estimation using acceleration, velocity, inertial and steering angle information was also proposed in (Kato et al., 1994). Moreover, in (Boada et al., 2015) an adaptive neuro-fuzzy infer-

ence 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 (Wei et al., 2016).

In this paper a novel side-slip angle estimation method which is based on linear regression is presented. As a first step, a subset selection method is proposed, which is able to prioritize the attributes based on their relation to the estimated signal. In the method the time-domain measurements of the attributes are processed through probability-based computations. Secondly, the OLS method is used to express the relationship between the attributes and the estimated signal in a linear form. In this process the pace regression algorithms of the WEKA data-mining software are performed (Wang and Witten, 1999). 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 structure of the paper is the following. Section 2 provides a subset selection method, by which priorities between the attributes can be set. The results of the selection are used through linear regression, which is presented in Section 3. The results of the big-data-based method are illustrated through various simulations in Section 4. Finally, the contribu-

tions of the paper and further challenges are summarized in Section 5.

2 SUBSET SELECTION FOR THE SIDE-SLIP ESTIMATION

The purpose of subset selection is to find the attributes which have relevance in the estimation of the side-slip angle. A difficulty in the subset selection method may be the high number of attributes, which can lead to the infeasibility of the problem. Thus, in the generation of subset models, it is necessary to reduce the complexity of the big data analysis, while information loss from the data is negligible (Wang and Witten, 2002). Several ranking algorithms have been proposed for the relationships between event-based attributes (Shibata, 1981; Thompson, 1978). However, in the side-slip estimation problem the attributes contain time-domain instances. Thus, it requires the evaluation of the measurement in the time-domain, which is proposed in the following.

Consider a simple linear equation (1) with only one variable.

$$y = c_{k,0} + c_{k,1}x_k, \quad (1)$$

where y is the output (estimated attribute), $x_k \in X$ is a selected variable from the set of the measured attributes (X), $c_{k,0}$ and $c_{k,1}$ are the fitting parameters. Since, in this case, the attributes are time-domain variables, the relative derivatives can be derived from (1).

$$\frac{d}{dt} \frac{y}{x_k} = c_{k,1}. \quad (2)$$

Then, the statistical parameters of $c_{k,1}$ are computed, such as $\sigma_{c_{k,1}}$ variance and $\hat{c}_{k,1}$ mean. It can be seen that if the linear model (1) provides the perfect fitting (the correlation between the two variables is 1)

$$\sigma_{c_{k,1}} = 0, \quad (3a)$$

$$\hat{c}_{k,1} = c_{k,1}^*, \quad (3b)$$

where $*$ denotes the parameters of the real model. Furthermore, if the linear model gives the worst fitting (the correlation between the two variables is 0)

$$\sigma_{c_{k,1}} \rightarrow \infty, \quad (4a)$$

$$\hat{c}_{k,1} = 0, \quad (4b)$$

If $y \neq 0$, $x_k \neq 0$ and $y \neq b + x_k$, where b is a constant bias parameter.

As the previous equations show, the relevance of an attribute depends on the variance and the mean of its $c_{k,1}$ parameter. If the normal distribution of $c_{k,1}$ is

assumed, then the relationship between these parameters is given by the gaussian distribution function:

$$G_k(c_{k,1}) = \frac{1}{\sigma_{c_{k,1}}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{c_{k,1}-\hat{c}_{k,1}}{\sigma_{c_{k,1}}}\right)^2} \quad (5)$$

The relevance of an attribute can be expressed by integrating the gaussian function and taking into account the sign of $c_{k,1}$.

$$\begin{aligned} D_k(c_{k,1}) &= \int_{-\infty}^{\infty} \text{sign}(c_{k,1})G(c_{k,1})dc_{k,1} = \\ &= \frac{1}{\sigma_{c_{k,1}}\sqrt{2\pi}} \int_{-\infty}^{\infty} \text{sign}(c_{k,1})e^{-\frac{1}{2}\left(\frac{c_{k,1}-\hat{c}_{k,1}}{\sigma_{c_{k,1}}}\right)^2} dc_{k,1} \end{aligned} \quad (6)$$

$D_k(c_{k,1})$ yields a value $D_k(c_{k,1}) \in [0, 1]$, which indicates the correctness of $c_{k,1}$ according to its sign, but ignores the variance $\sigma_{c_{k,1}}$. Therefore, the relevance of an attribute is formed as:

$$\mathcal{R}_k = \frac{D_k(c_{k,1})}{\sigma_{c_{k,1}}} \quad (7)$$

Finally, the attribute with the highest relevance on y is the attribute whose \mathcal{R} is the highest.

$$\max_x(\mathcal{R}) \quad (8)$$

3 ESTIMATION OF THE SIDE-SLIP ANGLE BASED ON THE OLS METHOD

The background of the estimation of the side-slip angle is the combination of the ordinary linear regression method and the subset selection procedure. In the following the most important features are summarized and a detailed description is found in (Wang and Witten, 1999).

In the method a dataset with n independent instances is considered with k input variables and one output variable. The instances are written in the form of an $n \times k$ design matrix X . Furthermore, the parameter vector of the true model is ζ^* . Using the parameter vector of the true model and the matrix of the instances, the output vector y is formed as

$$y = X\zeta^* + \varepsilon \quad (9)$$

where ε is the noise vector, of which elements have normal distribution $N(0, \sigma^2)$. Its variance is σ , and σ^2 is assumed to be known or estimated. The estimation of σ is denoted by $\hat{\sigma}^2$. (9) describes the true model $\mathcal{M}(\zeta^*)$, which is approximated with a fitted, linear model $\mathcal{M}(\zeta)$, which has a unique parameter vector ζ .

The goal of the estimation problem 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 task can be solved through several algorithms, such as the OLS method, the OLS subset selection, shrinkage, RIC, CIC methods, see (Wang and Witten, 1999). These methods can reduce the dimension of the models by discarding the redundant variables.

It is defined a distance \mathcal{D} between the candidate model and the true model for the evaluation of the estimation method. \mathcal{D} is defined in the form of

$$\mathcal{D}(\mathcal{M}(\zeta^*), \mathcal{M}(\hat{\zeta})) = \frac{\|y - \hat{y}\|^2}{\sigma^2}, \quad (10)$$

where $\|\cdot\|$ denotes the \mathcal{L}_2 norm and σ^2 is replaced by its estimated value $\hat{\sigma}^2$. Moreover, \hat{y} is the prediction output vector.

Thus, the estimation problem is to minimize the distance between the models, which is represented by the following expression

$$\min_{\zeta} \mathcal{D}(\mathcal{M}(\zeta^*), \mathcal{M}(\zeta)) \quad (11)$$

In the OLS regression the prediction is expressed in the following way

$$\hat{y} = X\hat{\zeta} \quad (12)$$

where $\hat{\zeta}$ is the parameter vector of the prediction model. The distance (10) can be reformulated as

$$\mathcal{D}(\mathcal{M}(\zeta^*), \mathcal{M}(\hat{\zeta})) = \frac{\|y - X\hat{\zeta}\|^2}{\sigma^2}, \quad (13)$$

which can be expressed by the following form of the training sets:

$$\frac{(y - X\hat{\zeta})^2}{\sigma^2} = \sum_{i=1}^{\omega} \frac{\|y_i - X_i\hat{\zeta}\|^2}{\sigma^2} \quad (14)$$

where y_i and X_i are one of the training sets, while ω is the number of the training sets.

The solution of the optimization task (11) can be achieved by the partial derivation of (13) according to $\hat{\zeta}$. The solution is the following:

$$\hat{\zeta} = (X^T X)^{-1} X^T y. \quad (15)$$

Thus, the predicted output \hat{y} in the vector form \hat{y} can be expressed by using the orthogonal projection matrix

$\mathcal{P} = X(X^T X)^{-1} X^T$ such as

$$\hat{y} = \mathcal{P}y. \quad (16)$$

The estimation of the side-slip angle $\beta(k)$ is based on the optimization. The prediction model is formed

as

$$\begin{aligned} \hat{\beta}(k) = & \zeta_1 + \sum_{i=k-j}^k \zeta_{2,i} a_x(i) + \sum_{i=k-j}^k \zeta_{3,i} a_y(i) + \\ & + \sum_{i=k-j}^k \zeta_{4,i} \psi(i) + \sum_{i=k-j}^k \zeta_{5,i} \delta_s(i) \end{aligned} \quad (17)$$

in which past signals are applied in the estimation to take into consideration the possible regressions. Although their incorporation may improve the accuracy of the prediction, the increase in j results in a more complex model.

The number of the past elements can be assisted through a spectrum analysis of $\beta(k)$, with which the dominant frequencies can be determined. It results in the sampling time T and the entire time horizon of the past is $j \cdot T$.

The measured elements of X_k are $X_k = [a_x(k) \ a_y(k) \ \psi(k) \ \delta_s(k)]^T$ with

- longitudinal acceleration a_x ,
- lateral acceleration a_y ,
- yaw-rate ψ ,
- angle of the steering wheel δ_s .

Furthermore, the side-slip angle $y_k = \beta(k)$ as a reference for the OLS regression analysis is also measured. The result of the optimization is the vector

$$\hat{\zeta} = [\zeta_1 \ \zeta_{2,k-j} \dots \zeta_{2,k} \ \zeta_{3,k-j} \dots \zeta_{3,k} \ \dots \ \zeta_{4,k-j} \dots \zeta_{4,k} \ \zeta_{5,k-j} \dots \zeta_{5,k}]^T. \quad (18)$$

4 SIMULATION RESULTS OF THE ESTIMATION METHODS

The efficiency of the proposed estimation method is demonstrated through simulation examples, using the high-fidelity vehicle dynamic software CarSim.

Several simulations have been performed, which resulted in a training set for big data analysis. In the simulations a D-class sedan passenger car has been used, whose sprung mass is 1320kg, during the collection of the signals the noises of the sensors were considered.

The car has been driven along the Michigan Waterford Hill Race Course (Figure 1) several times at various longitudinal velocities (Figure 2), which has resulted in the collection of more than 2 million instances.

The prediction model also contains the past values of the attributes, see the relationship (17). The

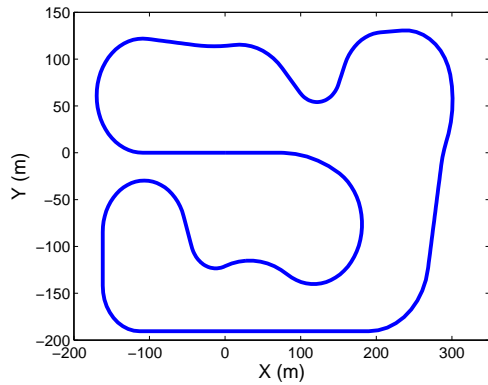


Figure 1: Simulation scenario.

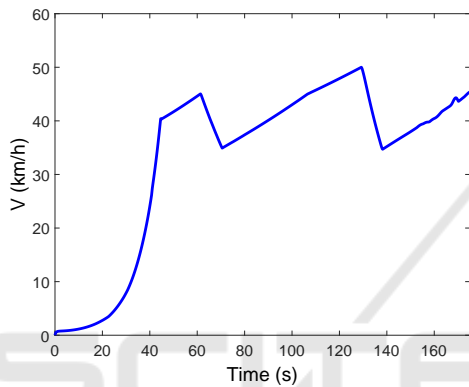


Figure 2: Velocity during the scenario.

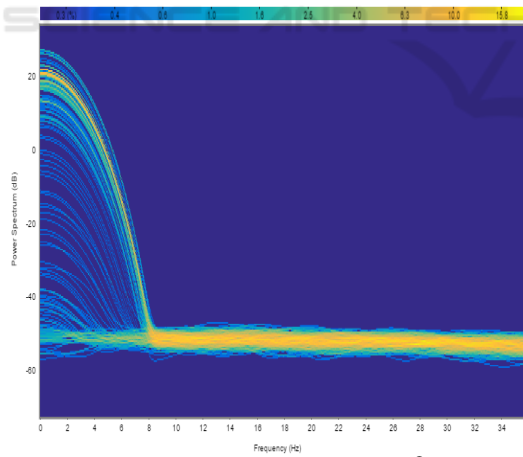


Figure 3: Spectrum analysis of β .

T sampling time of the signals from the past is computed through the Fourier spectrum analysis, whose results are found in Figure 3.

They show that the frequency domain of the power spectral density of the side-slip signal is between 0–8 Hz. Thus, the sampling time of the past values, according to the Nyquist-Shannon sampling criteria, has been chosen at 16 Hz. Moreover, 6 past values of all

attributes have been considered in the model estimation, which corresponds to 0.375 sec horizon backwards. The relevance of the selected attributes is evaluated by the algorithm presented in Section 2.

Figure 4 shows the relative derivatives of the measured signals. It can be seen that the means of the acceleration signals (a_x, a_y) are close to zero and their variances are relatively high, which indicates low relevance on the estimated variable (β). The mean of the yaw rate ($\dot{\psi}$) and the steering angle (δ) are above zero and their variances are significantly smaller, which implies that these signals have high relevance on the selected output.

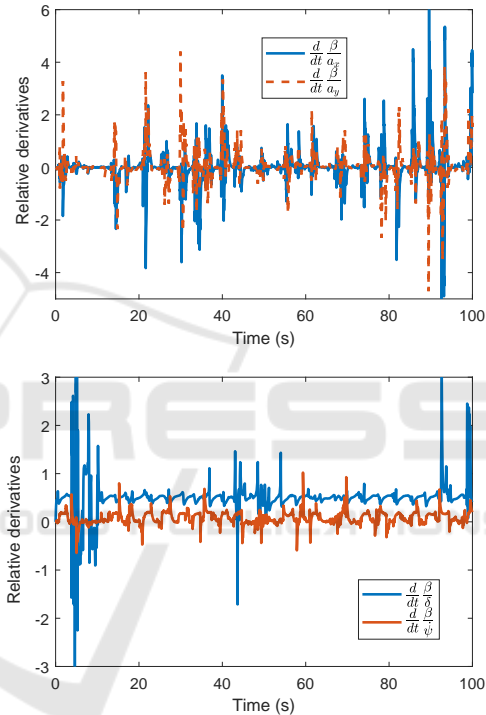


Figure 4: Relative derivatives of the selected attributes.

Table 1 summarizes the computed relevance of all signals. In the following the predefined order of the attributes is determined by their calculated relevance. Furthermore, the OSL subset selection method has created a model using the predefined order of the attributes, whose correlation coefficient is 0.911.

Figure 5 illustrates a scenario of the slip estimation, in which the car is driven along the test track at various velocities. In the example the bias and the variances of the sensors are selected in the following way:

- inertial sensor: bias = 0.15 m/s^2 , variance = 0.1²
- qyroscope: bias = 0.01 rad/s , variance = 0.01²

The error of the estimation is very low, although the velocity of the vehicle varies significantly, see Fig-

Table 1: Relevance of attributes.

T	\mathcal{R}_{a_x}	\mathcal{R}_{a_y}	$\mathcal{R}_{\dot{\psi}}$	\mathcal{R}_{δ}
t	0.00013	0.00449	0.14336	0.58502
t-1	0.0002	0.00266	0.01528	0.00082
t-2	0.00095	0.00074	0.00103	0.00086
t-3	0.00216	0.00003	0.00236	0.00592
t-4	0.001	0.0007	0.00022	0.01225
t-5	0.00284	0.00217	0.00156	0.00455
t-6	0.00044	0.00287	0.00165	0.00133

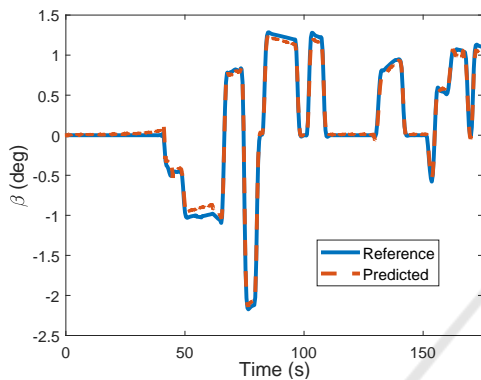


Figure 5: Estimation of the side-slip angle.

ure 2. In the following, the accuracy and the capability of the generated predictive model are examined in various situations.

4.1 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 angle of the steering wheel can be relatively well measured. In this scenario the parameters of the sensors are significantly modified to:

- inertial sensor: bias = 0.3 m/s^2 , variance = 0.4^2
- gyroscope: bias = 0.02 rad/s , variance = 0.04^2

Figure 6 shows the result of the simulation with the modified sensor parameters. 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.

4.2 Variation of Vehicle Mass

Secondly, the effect of the car mass variation on the accuracy of the prediction is investigated. The nominal sprung mass of the passenger car is $m = 1320 \text{ kg}$, which is modified in two ways: the mass is reduced to

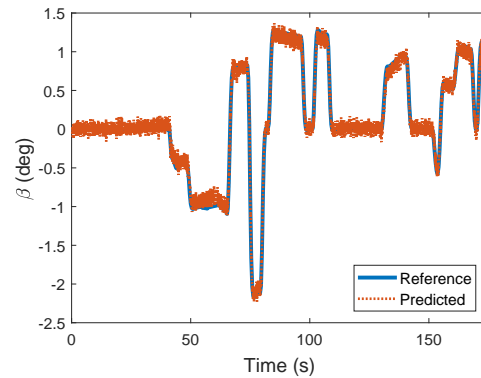


Figure 6: Effect of the increased noise on the prediction.

$m = 1000 \text{ kg}$ and then it is increased to $m = 1740 \text{ kg}$. The results of the changes can be seen in Figure 7. 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 vehicle mass.

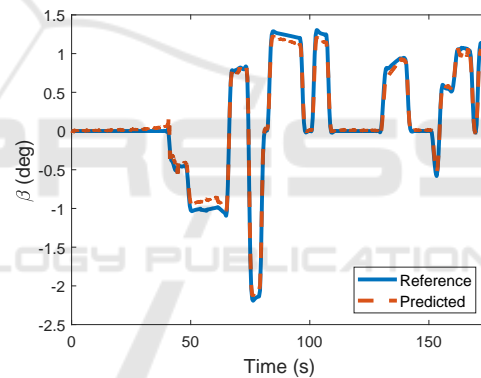


Figure 7: Effect of the variation of mass on the prediction ($m = 1000 \text{ kg}$).

4.3 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 8. 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 5. In the case of $\mu = 0.4$ the model also operates appropriately apart from two short sections between 80 – 90s and 110 – 120s. In these sections the vehicle reaches its stability boundary, and therefore, the effect of the steering angle on the lateral dynamics decreases. Nevertheless, the applied model operates accurately in all other sections.

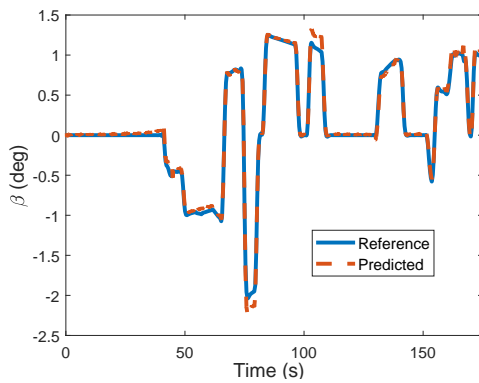


Figure 8: Effect of the variation of the adhesion coefficient on the prediction ($\mu = 0.4$).

4.4 Highway Analysis

Below, the predictive model is tested on a highway route. In the example the highway between Ulm and Stuttgart in Germany, which is one of the most hilly highway sections in Europe, is selected. Since all of the regression models can be easily overfitted, it is important to guarantee that the calculated model operates in other cases. In this case the passenger car is driven at high velocity $v_x \sim 130 \text{ km/h}$. The side-slip angle and its prediction are shown in Figure 9. The simulation shows that the applied model is able to predict the side-slip angle accurately. Its predictive accuracy is still relatively high despite the low side-slip values. It means that the proposed predictive model is generally able to predict the side-slip angle.

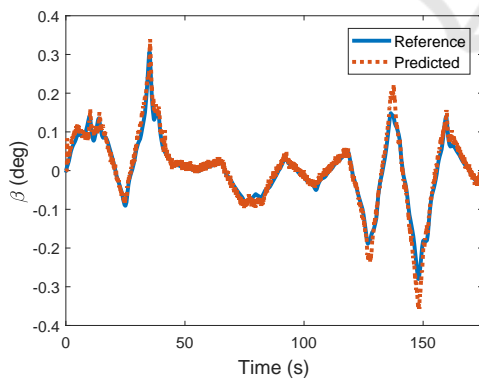


Figure 9: Side-slip angles on the highway.

4.5 Comparison of the Results with Kalman Filtering

Finally, the OLS estimation is compared with the results of a Kalman filtering technique. Although Kalman filtering and the presented OLS method are close to each other (Otter, 1978), the design technique of the filter and the computation algorithm with the

huge number of data can lead to slightly different results. For example, in case of the Kalman filter design the selection of the weights has an important role. Moreover, in this application of the Kalman filtering a sensor fusion algorithm is used, which incorporates in the signals of GPS and IMU. The purpose of 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 (Ryu et al., 2002). The goal of the comparison is to illustrate that the big data based estimation can also be an acceptable method for the estimation problem of vehicle side-slip angle.

The simulation results are illustrated in Figure 10. It can be seen that both methods can provide accurate results. However, the Kalman filter has higher precision at increased side-slip signals, while the proposed big-data-based approach is more efficient at low β values.

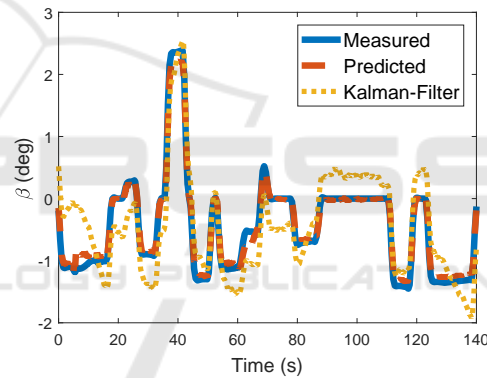


Figure 10: Comparison of the Kalman filter method and the prediction method.

5 CONCLUSIONS

In the paper a big-data-based algorithm for vehicle side-slip estimation has been proposed. Vehicle dynamic simulations have illustrated that the proposed method can be an efficient alternative algorithm to the conventional Kalman-filtering techniques. First, a subset selection method has been proposed, which provided priorities between the different attributes, such as longitudinal and lateral accelerations, the front wheel steering angle and the yaw rate. Second, a linear regression fitting based on the OLS method has been performed. As a future challenge in the research, the proposed method will be demonstrated through hardware-in-the-loop simulations and real test vehicle measurements.

ACKNOWLEDGEMENTS

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

REFERENCES

- Boada, B. L., Boada, M. J. L., Gauchía, A., Olmeda, E., and Díaz, V. (2015). Sideslip angle estimator based on ANFIS for vehicle handling and stability. *Journal of Mechanical Science and Technology*, 24(4):1473–1481.
- Coyte, J., Li, B., Du, H., Li, W., Stirling, D., and Ros, M. (2014). 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.
- Grip, H., L.I. Inslan, Johansen, T., Kalkkuhl, J., and Sussisa, A. (2009). Vehicle sideslip estimation. design, implementation, and experimental validation. *IEEE Control Systems Magazine*, 29(5):36–52.
- Jeon, J., Lee, W., Cho, H. J., and Lee, H. (2015). A big data system design to predict the vehicle slip. In *2015 15th International Conference on Control, Automation and Systems (ICCAS)*, pages 592–596.
- Kato, M., Isoda, K., and Yuasa, H. (1994). Estimation of vehicle side slip angle with artificial neural network. *SAE Review*, 15:79–82.
- Otter, P. W. (1978). The discrete Kalman filter applied to linear regression models: statistical considerations and an application. *Statistica Neerlandica*, 32(1):41–56.
- Ryu, J., Rossetter, E. J., and Gerdes, J. C. (2002). Vehicle sideslip and roll parameter estimation using GPS. In *6th Int. Symposium on Advanced Vehicle Control, Hiroshima, Japan*.
- Sasaki, H. and Nishimaki, T. (2000). A side-slip angle estimation using neural network for a wheeled vehicle.
- Shibata, R. (1981). An optimal selection of regression variables. *Biometrika*, 68:45–54.
- Stephant, J., Charara, A., and Meizel, D. (2004). Experimental validation of vehicle sideslip angle observers. In *IEEE Intelligent Vehicles Symposium, 2004*, pages 150–155.
- Thompson, M. L. (1978). Selection of variables in multiple regression. *International Statistical Society B*, 46:1–21 and 129–146.
- Wang, Y. and Witten, I. H. (1999). *Pace Regression*. (Working paper 99/12). Hamilton, New Zealand: University of Waikato, Department of Computer Science.
- Wang, Y. and Witten, I. H. (2002). Modeling for optimal probability prediction. In *Proceedings of the Nineteenth International Conference in Machine Learning*, pages 650–657, Sydney, Australia.
- Wei, W., Shaoyi, B., Lanchun, Z., Kai, Z., Yongzhi, W., and Weixing, H. (2016). Vehicle sideslip angle estimation based on general regression neural network. *Mathematical Problems in Engineering*.