IFAC

Least Squares Method Applied to Rail Vehicle Contact Condition Monitoring

Guy Charles * Roger Dixon * Roger Goodall *

* Control Systems Group, Electronic & Electrical Engineering,
Loughborough University, Loughborough, LE11 3TU, UK

Abstract: The dynamics of a railway vehicle are driven by the geometry and conditions at the wheel-rail contact. Typically the condition and shape of the wheel and rail are monitored separately and off-line. The work presented here is part of ongoing research into on-line model-based estimation of parameters in the wheel-rail contact dynamics. This paper outlines a practical approach to estimating a nonlinear function within a dynamic system by using a piecewise cubic functions. The parameters for the cubic functions are estimated with a least squared approach applied to the dynamic measurements taken from the system. A simplified plan-view wheelset and suspended mass model is introduced to use as an application of this technique. A contact geometry term, conicity, which is a nonlinear function of the relative lateral wheel-rail position, is included in the rail vehicle model. The conicity is successfully estimated using the least-squares method outlined in the paper.

Keywords: Railways; Vehicle dynamics; Least-squares estimation; Nonlinear.

1. INTRODUCTION

The work presented in this paper is the product of feasibility studies into model based condition monitoring applied to the wheel-rail interface dynamics in railway vehicles. The most important element in the dynamics of a railway vehicle is the interaction between the wheel and the rail. Any changes to the shape and condition of either the wheel or rail will have subsequent changes on the response of the vehicle. Typical monitoring approaches will inspect the shape and surface condition of the wheel and rail separately, however it is the interaction of the two in situ that really influences the system dynamics. Assumptions about how the two independently measured components combine can cause problems in identifying faults arising in operation. Since the process of monitoring and maintenance aims to address the quality and safety of the dynamic response of the vehicle, it makes sense that the condition monitoring should be based upon the dynamic response of the vehicle.

Model-based condition monitoring uses knowledge of the system in the form of a mathematical model and the measured vehicle response to track irregularities to perform real-time estimations of the system parameters of interest. The first model based condition monitoring approach commonly taken is to apply a Kalman Filter (Kalman (1960)). The use of a Kalman Filter for parameter estimation creates a nonlinear problem, even if the system of interest is linear itself. In applying the Kalman Filter to the particular application outlined in this paper, it was found that the nature of the Kalman Filter is not well suited to this problem. It does not directly consider the full time series of collected data, which is very important in this case as the contact point between the wheel and rail is always moving in response to the rail irregularities. Hence it is important to treat a whole frame of data at once to get a full picture of the condition of the wheel and rail shapes. Tuning the Kalman Filter's process and measurement noise covariance matrices will effectively force the process to place more weight on the past data, but this can have adverse effects on the estimation quality and the ability to cope with nonlinearities.

A least squares approach was seen to be more applicable to this problem, where the estimation method could be based on a complete frame of collected data. This allows the estimation process to consider the whole data set when forming a best-fit solution for the nonlinear parameter curve. Furthermore, with a slight change to the basic least squares method, the well known terms in the system equations can be included, making the technique estimate the unknown nonlinear function alone. Here we present a number of implementations of least squares estimation of the nonlinear contact profile, including a piecewise cubic function Ichida et al. (1976).

The concept of estimating wheel-rail geometry terms was inspired by previous work which looked into condition monitoring of suspension components of a rail vehicle, such as dampers Li et al. (2006). This work is part of research aimed at estimating the profile shape of the wheel-rail combination as the vehicle travels down the track. It is proposed that 'fast' changes would be due to the rail profile and 'slow' changes due to the wheel profile. To the authors' knowledge this work is the first to investigate model-based condition monitoring at the wheel-rail interface.

Phase one of the work is simulation based, hence the results shown here are all based on a well understood wheelset and half vehicle models implemented in Simulink. The emphasis is on evaluating the viability of this approach and establishing what useful information can be taken from condition monitoring in this application.
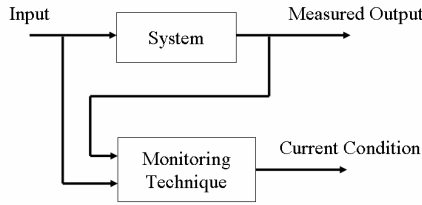
Fig. 1. Block diagrams for generic condition monitoring schemes.

2. CONDITION MONITORING AND PARAMETER ESTIMATION

Condition monitoring aims to establish the current (ideally real-time) condition of a system, using some level of knowledge of the system of interest. This may be in the form of a model, expert system, experience, learnt behaviour, etc. Figure 1 shows the block diagram for a generic condition monitoring system. The controlled input and measured output for the system are given to the condition monitoring strategy. These direct results can then be taken into further processing algorithms to establish a condition or provide fault detection. This work is focused on the first step of condition monitoring and looks at the feasibility and scope of results that can be obtained from the estimation of wheel-rail contact parameters.

In the previous condition monitoring work mentioned in the introduction, Kalman Filter and Particle Filter approaches were used to estimate components in the secondary suspension. Data was taken from track tests on the Tyne and Wear metro and a Coradia 175 vehicle on the Chester to Holyhead line. A set of sensors housed in a single unit were attached to the bogie frame (figure 2(a)). A Particle Filter approach was used to estimate the parameters of interest, which are shown to converge to the design values in figure 2(b).

3. LEAST-SQUARES PARAMETER ESTIMATION

A complete background on the least squares estimation and regression methods can be found in Aström (1989) or Ljung (1999). The basis of the Least Squares method is to optimise the set of system parameters in such a way as to minimise the error between the measured output of the system and the regression model of the system given by:

$$\hat{y}(i) = \phi_1(i)\theta_1 + \phi_2(i)\theta_2 + ... + \phi_n(i)\theta_n \quad (1)$$
$$= \phi^T(i)\theta \quad (2)$$

where $\hat{y}$ is the estimated observed variable, ϕ is the vector of regressors, θ is the vector of unknown parameters, and i is the discrete sample number ($1 \leq i \leq t$). Nonlinear systems may be formed into this linear regression structure by using nonlinear regressors in the model.

The problem is to optimise the parameter vector, θ, in such a way that the output from the regression model given in 1 will agree as closely as possible over the sample set of observations and regressors, $\{y(i), \phi(i)\}$. The error at sample i is given by

$$\epsilon(i) = y(i) - \hat{y}(i) = y(i) - \phi^T(i)\theta \quad (3)$$
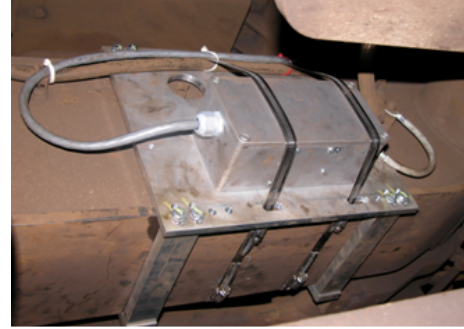
(a) Sensor box attached to vehicle bogie
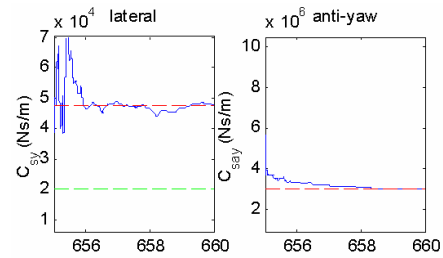
(b) Lateral and yaw damper estimates

Fig. 2. Previous railway vehicle condition monitoring research carried out at Loughborough University.

Introducing the notation for the complete sample set

$$Y(t) = [y(1)\ y(2)\ ...\ y(t)]^T \quad (4)$$
$$E(t) = [\epsilon(1)\ \epsilon(2)\ ...\ \epsilon(t)]^T \quad (5)$$
$$\Phi(t) = [\phi^T(1)\ \phi^T(2)\ ...\ \phi^T(t)]^T \quad (6)$$

allows the error to be written as:

$$E = Y - \hat{Y} = Y - \Phi\theta \quad (7)$$

The least-squares error can then be written as

$$V(\theta,t) = \frac{1}{2}\sum_{i=1}^{t}\epsilon(i)^2 = \frac{1}{2}\sum_{i=1}^{t}(y(i) - \phi^T(i)\theta)^2 \quad (8)$$

$$= \frac{1}{2}E^T E \quad = \frac{1}{2}||E||^2 \quad (9)$$

The least-squares error is minimal for parameters $\hat{\theta}$ such that

$$\Phi^T\Phi\hat{\theta} = \Phi^T Y \quad (10)$$

If the matrix $\Phi^T\Phi$ is non-singular, the minimum is unique and given by:

$$\hat{\theta} = (\Phi^T\Phi)^{-1}\Phi^T Y \quad (11)$$

Proof: The error function can be rewritten as

$$2V(\theta,t) = (Y - \Phi\theta)^T(Y - \Phi\theta)$$
$$= Y^T Y - Y^T\Phi\theta - \theta^T\Phi^T Y + \theta^T\Phi^T\Phi\theta \quad (12)$$

The error function is quadratic in θ. By completing the square the following result can be found:

$$2V(\theta,t) = Y^T(I - \Phi(\Phi^T\Phi)^{-1}\Phi^T)Y$$
$$+(\theta-(\Phi^T\Phi)^{-1}\Phi^T Y)^T\Phi^T\Phi(\theta-(\Phi^T\Phi)^{-1}\Phi^T Y) \quad (13)$$

The first term on the right hand side is independent of θ, and the second term is always positive. Therefore the minimum is obtained for V when

$$\theta = \hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T Y \qquad (14)$$

Alternative solutions to minimising the least squared error can be constructed to account for singular matrices in equation 3 Ljung (1999). No problems with singular matrices were experienced in this work.

3.1 Known Terms in the Regression Model

The regression model in 1 is shown to be made up of a linear combination of unknown parameters and known regressors. In many cases some of these parameters will be known, and it would be more convenient if only the unknown parameters are estimated. In the railway vehicle application presented in this paper this is the case. Better results are obtained if the least squares method estimates the nonlinear function alone, rather than all the parameters in the regression model.

A new regression model is defined as:

$$\hat{y}(i) = \phi(i)\theta + \omega(i) \qquad (15)$$

where ω are the combined known parameter and regressor terms. The complete set of these known terms can be now be defined as

$$\Omega = [\omega(1)\omega(2)... \quad \omega(t)]^T \qquad (16)$$

Now the same least squares error, $V(\theta, t)$, is minimal with the parameters $\hat{\theta}$ such that

$$\Phi^T(\Phi\hat{\theta} + \Omega) = \Phi^T Y \qquad (17)$$

If the matrix $\Phi^T \Phi$ is non-singular, the minimum is unique and given by:

$$\hat{\theta} = (\Phi^T \Phi)^{-1}(\Phi^T Y - \Phi^T \Omega) \qquad (18)$$

The proof can be shown using a similar approach as given in the previous section.

3.2 Estimating Complex Nonlinear Functions

Typically any nonlinear function that is to be estimated by this method needs to be made up of an analytical function, for example, a cubic function of one of the regressors. This is shown to give large errors in the application presented in this paper. The nonlinear function is too complex to be estimated with a simple function. One approach is to use higher order equations to match the nonlinear function, but this adds complexity and can give false results. An alternative approach is to use a piecewise cubic function which fits the data in a least squares sense over each piece of the function whilst preserving the continuity of function and gradient between each piece of the function. The process is briefly described here, however full details can be found in Ichida et al. (1976).

The piecewise cubic function is given by:

$$S_i(x) = m_i a_i(x) + m_{i+1} b_i(x) + y_i c_i(x) + y_{i+1} d_i(x) \qquad (19)$$
$$a_i(x) = (x_{i+1} - x)^2 (x - x_i)/h_i^2 \qquad (20)$$

$$b_i(x) = -(x_{i+1} - x)(x - x_i)^2/h_i^2 \qquad (21)$$
$$c_i(x) = (x_{i+1} - x)^2 \{2(x - x_i) + h_i\}/h_i^3 \qquad (22)$$
$$d_i(x) = (x - x_i)^2 \{2(x_{i+1} - x) + h_i\}/h_i^3 \qquad (23)$$
$$h_i = x_{i+1} - xi \qquad (24)$$

where $S_i(x)$ is the cubic function for piece i, (x_i, y_i) is the coordinates of knot point i, m_i is the gradient at knot point i.

Let the least and largest x data in the interval $[x_i, x_{i+1}]$ be x_{pi} and x_{qi}, the the problem becomes one of minimising the sum square of the errors for the n sections in the piecewise function:

$$E = \sum_{i=1}^{n} \left[\sum_{k=pi}^{qi} \{S_i(x_k) - f_k\}^2 \right] \qquad (25)$$

where f_k is the data for the function to be estimated. Differentiating E w.r.t. the cubic function parameters and setting to zero reduces the minimisation problem to the solution of the following relationship:

$$Az = g \qquad (26)$$

where z is a vector of the cubic function parameters, A is a matrix containing sum terms of x_i, x_k and g is a vector containing sum terms of x_i, x_k, f_k. Details on the specific structure of these matrices can be found in Ichida et al. (1976).

More complex approaches can use smoothing spline methods (Schimek (2000); Seber and Wild (2003)) which also maintain second derivative continuity, but the proposed method above provides a relatively simple solution that gives good results.

4. RAILWAY VEHICLE APPLICATION

The stability and guidance dynamics of a railway vehicle are driven by the constrained contact between the a solid axle wheelset and rail. The steering and stability effects are partly provided by the changes in wheelset-rail contact geometry as the wheelset moves laterally on the railhead and the corresponding changes in normal contact force and rolling radius at each wheel. The remainder of the contact forces are caused by creepages (or slip) in the contact patch. The wheel-rail contact patch is typically an ellipse of the order of 10mm in diameter, but its rolling friction is non-trivial to analyse and calculate. Commercial railway simulation packages use well known computationally fast approximations to this problem (Kalker (1980) and O.Polach (2005)). These calculate a lateral and longitudinal creep force at the contact patch, plus a spin creep moment about the normal axis of the contact patch.

At low creepages, the creep force behaviour can be treated as linear (see figure 3). This is an acceptable assumption for dry conditions where the contact forces are sufficient to prevent easy saturation. For a given contact patch size and shape, the gradient of the creep-force curve (known as the creep coefficient) is constant.

A linearised quarter vehicle model, consisting of a single wheelset and a suspended mass is used to simulate the vehicle (see figure 4). The plan view dynamics (yaw and
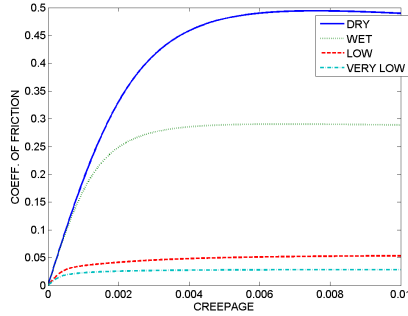
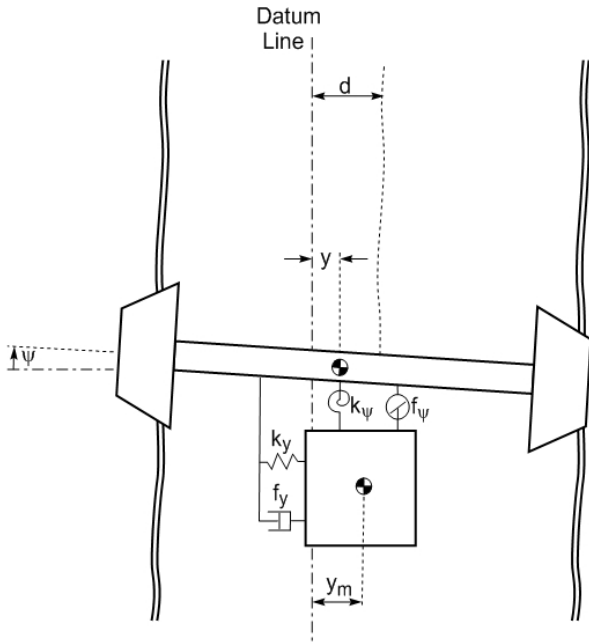Fig. 3. Creep curves for dry, wet, low and very low adhesion conditions (wheel load of 4000kg, 20m/s)

f_{11}	longitudinal creep coefficient	7.44e6	N
f_{22}	lateral creep coefficient	6.79e6	N
f_{23}	spin creep coefficient	13.7e3	N
f_{33}	spin creep coefficient	0	N
f_y	lateral damper coefficient	50e3	Ns/m
f_ψ	yaw damper coefficient	0	Ns/rad
I	wheelset yaw inertia	700	kgm^2
l	half wheelset width	0.7452	m
m	wheelset mass	1250	kg
m_m	suspended mass	8000	kg
k_y	lateral suspension stiffness	0.23e6	N/m
k_ψ	yaw suspension stiffness	2.5e6	N/rad
r_0	rolling radius at zero lateral position	0.45	m
v	velocity	20	m/s

Table 1. Parameter values used in the wheel-rail profile estimation model.

Fig. 4. Diagram of the wheelset and suspended mass simplified railway vehicle model.

lateral displacement) are sufficient to describe the stability and guidance response to lateral track irregularities. This model includes nonlinear wheel and rail geometries in the form of a nonlinear conicity term and is given by:

$$\ddot{y} = \frac{1}{m}\left\{-\frac{2f_{22}}{v}\left(\dot{y} + \frac{r_0\lambda}{l}\dot{y} - v\psi\right)\right.$$
$$\left. -\frac{2f_{23}}{v}\dot{\psi} - \frac{W\lambda}{l}(y - d) + F_{sy}\right\} \tag{27}$$
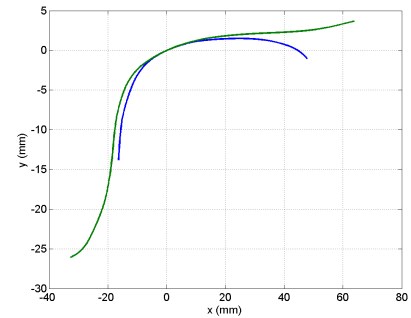
$$\ddot{\psi} = \frac{1}{I}\left\{-2f_{11}\left(\frac{l\lambda}{r_0}(y-d) + \frac{l^2}{v}\dot{\psi}\right) - \frac{2f_{33}}{v}\dot{\psi}\right.$$
$$\left. -\frac{I_{wy}v\lambda}{lr_0}\dot{y} + \frac{f_{23}}{v}\left(\dot{y} + \frac{r_0\lambda}{l}\dot{y} - v\psi\right)\right.$$
$$\left. +W\lambda l\psi + M_{s\psi}\right\} \tag{28}$$

$$\ddot{y_m} = \frac{1}{m_m}\left\{-F_{sy}\right\} \tag{29}$$
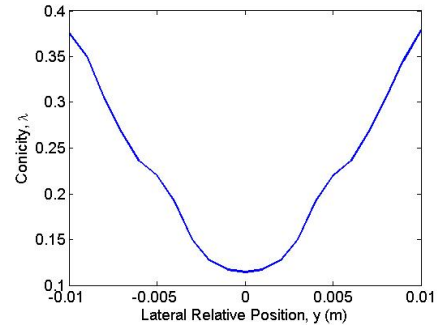
where y is the lateral position of the wheelset, y_m is the lateral position of the suspended mass, ψ is the yaw angle, W is the wheel load and d is the lateral track irregularity

(a) Wheel and rail profiles

(b) Equivalent conicity function

Fig. 5. Nonlinear geometry of the wheel and rail used in this vehicle model.

input. λ is the nonlinear conicity function, the estimation of which is the focus of this study. F_{sy} and $M_{s\psi}$ are the suspension lateral force and yaw moment respectively, given by:

$$F_{sy} = k_y(y_m - y) + f_y(\dot{y}_m - \dot{y}) \tag{30}$$
$$M_{s\psi} = -k_\psi\psi - f_\psi\dot{\psi} \tag{31}$$

All other terms and their values used in this work are given in table 1.

The wheel-rail dynamic interaction is particularly nonlinear with complexities arising from the geometry, contact and creep. Typically these dynamics are analysed with a linearised conicity term (λ). However, here some degree of the nonlinearities must be included, which in this initial case was achieved by making conicity a nonlinear function of relative lateral position.
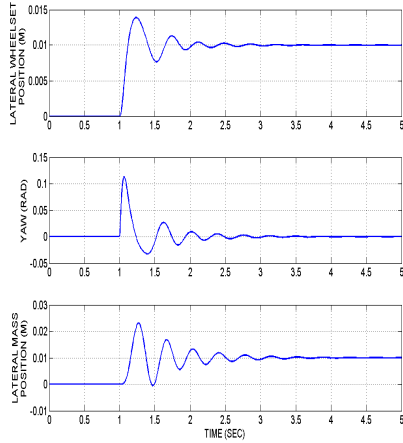
Fig. 6. Single wheelset and lateral mass step response.

Conicity is a linearised term to describe the nonlinear wheel-rail contact shape. It is effectively a secant gradient of the rolling radius difference:

$$\lambda(y) = \frac{r_L - r_R}{2(y - d)} \tag{32}$$

where r_L and r_R respectively are the left and right rolling radii at the contact point, which depend upon the combination of the wheel and rail profile shapes.

A generic wheel-rail geometry has been used in this work. Figure 5(a) shows the wheel and rail profiles used and figure 5(b) shows the corresponding nonlinear effective conicity. Figure 6 shows the response of this single wheelset to a lateral step in track, which demonstrates the guidance effect that the solid axle wheelset provides to the railway vehicle dynamics.

A full background and description of the dynamics of railway vehicles can be found in Garg and Dukkipati (1984) or Wickens (2003).
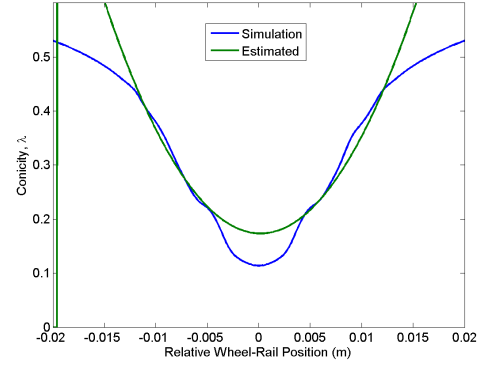
4.1 Least Squared Conicity Estimation

The first step in applying the least squares estimation method is to determine the regression model and regressors. The observed variable taken here is $\ddot{\psi}$, and the assumption that all the state and system parameters other than that being estimated are available to construct the regressors. Track irregularities are also assumed to be known at this point in the work. As in all model based condition monitoring approaches, it is necessary to select an estimation model that is sufficiently complex to account for all the dynamics of interest. Here, a reduced model is sufficient to capture these dynamics, given by:
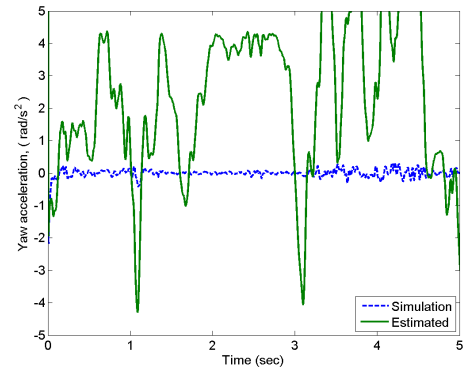
$$\ddot{\psi} = \frac{1}{I} \left\{ -\frac{2f_{11}l\lambda}{r_0}(y - d) - \frac{2l^2 f_{11}}{v}\dot{\psi} - k_\psi\psi - f_\psi\dot{\psi} \right\} \tag{33}$$

This can be converted into a regression sense by considering the conicity function to be estimated using a cubic function of the relative wheel-rail position:

$$\ddot{\psi} = A_1\lambda(y - d) + A_2\psi + A_3\dot{\psi} \tag{34}$$
$$= A_1(a_1(y - d)^2 + a_2(y - d) + a_3)(y - d)$$

(a) Estimated conicity function

(b) Regression model response

Fig. 7. Cubic least squares estimation.

$$+A_2\psi + A_3\dot{\psi} \tag{35}$$
$$= B_1(y - d)^3 + B_2(y - d)^2 + B_3(y - d)$$
$$+A_2\psi + A_3\dot{\psi} \tag{36}$$

where A_1, A_2 and A_3 are known constants, a_1, a_2 and a_3 are the cubic function coefficients, B_1, B_2 are B_3 are the regression parameters. Hence the vector of regressors, vector of unknown parameters and known terms are given by:

$$\phi = [(y - d)^3 \quad (y - d)^2 \quad (y - d)]^T \tag{37}$$
$$\theta = [B_1 \quad B_2 \quad B_3]^T \tag{38}$$
$$\omega = A_2\psi + A_3\dot{\psi} \tag{39}$$

Figure 7(a) shows the results from applying this cubic best fit to the railway vehicle simulation. The correlation is reasonable, and gives a rough approximation for monitoring the condition of the wheel-rail geometry. However, this is not accurate enough to gain a complete insight into the wheel-rail contact conditions. Notice that, although the simulation conicity is symmetric about zero relative wheel-rail position, the estimated curve is biased slightly to one side. This is a result of the sampled data being not evenly distributed. The simulation model is driven by track irregularities, which result in the contact patch varying in position across the wheel and rail. Figure 7(b) shows that the regression model output given the same track irregularities is very different from the rail vehicle simulation. This is a result of the large effect that the
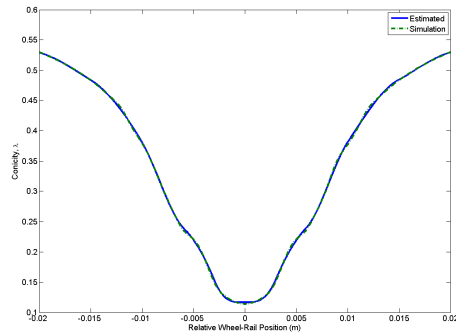
(a) Estimated conicity function
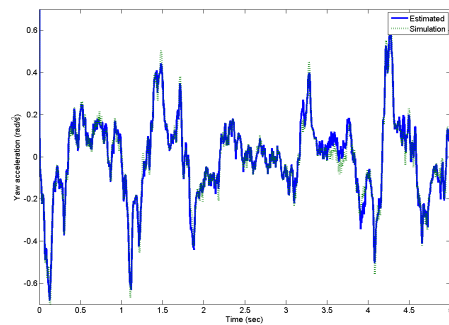
(b) Regression model response

Fig. 8. Piecewise cubic least squares estimation.

wheel-rail contact shape and condition have upon the vehicle dynamics. Even slight differences in the estimated conicity function from the 'real' simulation conicity causes large errors in the model response.

The complexity in the conicity function shape means that a single cubic fit is not good enough. The application of the piecewise cubic function aims to allow more flexibility to making the least squares fit of the data. Effectively a cubic function is fitted to each section of the conicity data, between designated knot points. Figure 8(a) shows the result of this approach and it can be seen that the correlation is very good to the simulation conicity function. Furthermore, figure 8(b) shows that the dynamic response of the regression model matches well with the simulation response, in spite of the simplified model used in the estimation process.

5. CONCLUSION

Feasibility studies into model based condition monitoring applied to the wheel-rail interface in railway vehicle dynamics has led to the use of a least squares parameter estimation approach. This general approach is particularly applicable to estimating a nonlinear system parameter as it takes into account a block of sampled data and hence the full extent of the nonlinear system parameter. An approach to estimating a complex nonlinear function within a system has been presented, using a least squared error piecewise cubic function.

The piecewise cubic least squares estimation was successfully applied to a quarter railway vehicle model in order to estimate a nonlinear wheel-rail geometry function (conicity).

This work is part of ongoing research into condition monitoring of wheel-rail contact dynamics, looking more specifically at estimating the wheel-rail profile shape and the contact adhesion conditions. Future developments include developing these methods to account for the track input as a coloured noise source and applying these methods to estimating nonlinear parameter functions within more complex models (e.g. rolling radius and contact angle). The use of real measurement data from track testing is also intended for the future of this project.

REFERENCES

K.J. Aström. *Adaptive Control.* Addison-Wesley, 1989.

Vijay K. Garg and Rao V. Dukkipati. *Dynamics of Railway Vehicle Systems.* Academic Press, 1984.

K. Ichida, F. Yoshimoto, and T.Kiyono. Curve fitting by a piecewise cubic polynomial. *Computing*, 16:329–338, 1976.

J.J. Kalker. A fast algorithm for the simplified theory of rolling contact. Internal report, Delft University of Technology, 1980.

R.E. Kalman. A new approach to linear filtering and prediction. *Transactions of ASME - Journal of Basic Engineering*, pages 35–45, 1960.

Ping Li, Roger Goodall, Paul Weston, Chung Seng Ling, Colin Goodman, and Clive Roberts. Estimation of railway vehicle suspension parameters for condition monitoring. *Control Engineering Practice*, 15:43–55, 2006.

L. Ljung. *System Identification, Theory For The User.* Prentice Hall, 1999.

O.Polach. Creep forces in simulations of traction vehicles running on adhesion limit. *Wear*, 258:992–1000, 2005.

M.G. Schimek. *Smoothing and Regression: approaches, Computation, and Application.* Wiley, 2000.

G.A.F. Seber and C.J. Wild. *Nonlinear Regression.* Wiley, 2003.

A.H. Wickens. *Fundamentals of Rail Vehicle Dynamics: Guidance and Stability.* Swets & Zeitlinger, 2003.