

Digital Filters for Maintenance Management

Fausto Pedro García Márquez and Diego José Pedregal Tercero
*Ingenium Research Group, University of Castilla-La Mancha
Spain*

1. Abstract

Faults in mechanisms must be detected quickly and reliably in order to avoid important losses. Detection systems should be developed to minimize maintenance costs and are generally based on consistent models, but as simple as possible. Also, the models for detecting faults must adapt to external and internal conditions to the mechanism. The present chapter deals with three particular maintenance algorithms for turnouts in railway infrastructure by means of discrete filters that comply with these general objectives. All of them have the virtue of being developed within a well-known and common framework, namely the State Space with the help of the Kalman Filter (KF) and/or complementary Fixed Interval Smoother (FIS) algorithms. The algorithms are tested on real applications and thorough results are shown.

2. Introduction

Faults in any important mechanisms must be detected quickly and reliably if the information is to be useful. Generally such mechanisms may be modeled as discrete dynamic systems, where data must be processed on line. When feasible, the detection system should use a model as simple as possible for detecting faults quickly by analyzing data in real time. The models for detecting faults must adapt to external and internal conditions to the mechanism, since both of them may affect the system as a whole.

The present chapter deals with maintenance systems for turnouts in railway infrastructure by means of discrete filters. Turnouts are assembled from switches and a crossing where the moving parts are often described as the “points” move by the point mechanism. The standard railway point mechanism is a complex electro-mechanical device with many potential failure modes.

Several approaches for maintenance of such devices are shown in this chapter and briefly described in this introduction. All of them have the virtue of being developed within a well-known common framework, namely the State Space (SS) with the help of the Kalman Filter (KF) and/or complementary Fixed Interval Smoother (FIS) algorithms, exposed in general terms in the following section.

Based on this common framework, the following subsections in this introduction show the particular applications shown in later sections of the chapter.

2.1. Filtering with Integrated Random Walks (IRW)

One possible way to analyze faults on line is to work with a reference dynamic system for their analysis. If the absolute value of the difference between the actual data and the reference data (i.e. the profile without any fault) is analyzed, the majority of faults may be detected by means of a simplified univariate dynamic system, like the one explored in [9]. The dynamic system and the use of the SS framework and the KF in this study allow increasing the reliability of the model presented that is the basic input to a rule-based decision mechanism. When applied to the linear discrete data filtering problem, the KF is a powerful algorithm, because it supports estimations of past, present and, most importantly, future states. It can therefore be used in predictive maintenance applications where data collected from sensors is affected by measurement and transmission line noise [12].

The previous approach may be exploited by setting up a bivariate model composed of two time series, i.e. the reference curve on one hand and each one of the empirical curves obtained on line on the other hand. More specifically (see section 4.2 below) a tentative model consists of a bivariate trend plus noise structure. The correlation between either trends or signals free from noise is considered as an indication of similarity between the curves and therefore the inexistence of failures. As long as the new incoming data is free from fault, the correlation parameter is close to one, but as a failure starts to develop this parameter tends to differ from one. The cut-off value of the correlation coefficient relevant to discern 'good' and 'bad' curves is selected on practical grounds based on past experience with this kind of data, but refined formal statistical criteria may be used as well [19]. Even forecasts of the curve that is being studied may be produced at any point in time, based on the current parameter values and the future data of the reference curve [14]. Therefore the fault may be detected ahead of time.

2.2. Random Walks and smoothing

Similar measurement data were collected from sensors mounted on a UK type M63 point machine at the Carillion Rail (formerly GTRM) Training Centre in Stafford (UK). It is difficult to compare the measurements taken during induced failure conditions with those from the fault-free condition because of noise in the measurements. The measurement data needed to be filtered in order to reduce the noise before comparisons may be made. Filtering using a SS model and the KF was an option (like in [9], [19] and [20]). Assuming the noisy data is a signal plus noise model, the KF reduces the power of the 100 and 200 Hz interfering signals. Rather than augmenting the SS models to express the additional knowledge of the interfering signals, a much simpler smoothing seems more convenient because of the relationship between the sample rate and the frequencies of the interfering signals, and provides excellent results for the data collected during this series of experiments [10].

2.3. Advance Dynamic Harmonic Regression (DHR)

A different case study was based on data collected from point mechanisms at Abbotswood Junction (UK). Three electro-mechanical and four electro-hydraulic point machines were

monitored by a RCM system. Processed information was sent remotely from the trackside data-collection units to a personal computer located in a local relay room.

A fault is detected by comparing the forecasts of the model, considered as the expected signal in the case of no faults, with the actual data coming from the point mechanism when a movement is in progress. If the error is too large, measured by its standard deviation, a fault alarm is issued. The limit at which an error is considered too large is a design parameter that is fixed by experimentation. The system adapts to the changes experienced by the point machine. There are internal alterations (like friction, wear, etc.) and external as well (like environmental conditions, impacts, obstacles, etc.). The adaptability of the system is accomplished by continuous estimation of the models as new information becomes available and by discarding the oldest information. Models are always estimated on fault-free data [13].

The key point in this application is that the expected shape is computed as the forecast of a combination of two models that work interactively on historical data coming from signals free from any fault. The first of the models forecasts the time span a movement would take in case of absence of faults (an appropriate model used in this case was of the VARMA class or a local level plus noise but set up in continuous time). The second model is run to forecast the signal itself (due to the nature of the data a pertinent class is a Dynamic Harmonic Regression, DHR, similar to a Fourier analysis, but with advanced features included to incorporate a time varying period observed in the data).

The outline of the chapter is as follows. Section 3 reports a brief explanation of the general framework on which all the models in this chapter are set up, namely the State Space systems. Section 4 shows the first of the applications, i.e. in the point mechanisms. Finally section 5 shows how a fault detection algorithm may be implemented on seven point machines at Abbotswood junction (UK).

3. State Space systems

The general framework on which all models in this chapter are cast, is the so called State Space systems, that have experienced a remarkable attention during the last decades, as the extended literature about it reveals [3], [7], [13], [15], [16], [17], [21], [24], [26] and [27].

A stochastic discrete-time State Space system (SS) is a model composed of two sets of equations, the *Observation Equations*, and *State Equations*. The former relates the output to the states of the system, while the latter reflects the dynamic behavior of the system by relating the current value of the states to their past values. There are a number of different formulations of these equations, but one fairly general representation is given by equations (1) (see [3] and [21]). In general, much simpler models are sufficient, as later case studies show.

$$\begin{aligned} \text{State Equations} & \quad : \quad \mathbf{x}_{t+1} = \mathbf{\Phi}_t \mathbf{x}_t + \mathbf{E}_t \mathbf{w}_t & \text{(i)} \\ \text{Observation Equations} & \quad : \quad \mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{C}_t \mathbf{v}_t & \text{(ii)} \end{aligned} \tag{1}$$

In (1) \mathbf{z}_t is the m dimensional vector of observed variables for $t=1,2,\dots,N$; \mathbf{x}_t is an n dimensional stochastic state vector; \mathbf{w}_t is an r dimensional vector of (to be Gaussian) system disturbances, i.e. zero mean white noise inputs with a covariance matrix \mathbf{Q}_t ; and \mathbf{v}_t is a s dimensional vector of zero mean white noise variables (measurement noise: again assumed to be Gaussian) with a covariance matrix \mathbf{R}_t . In general, the vector \mathbf{v}_t is assumed to be independent of \mathbf{w}_t (not necessarily), and these two noise vectors are independent of the initial state vector \mathbf{x}_0 . Φ_t , \mathbf{E}_t , \mathbf{H}_t , \mathbf{C}_t , \mathbf{Q}_t , and \mathbf{R}_t are, respectively, the $n \times n$, $n \times r$, $m \times n$, and $m \times s$, $r \times r$ and $s \times s$ system matrices, some elements of which are known and others that need to be estimated in some way.

Given the general SS form (1), the estimation problem consists of finding the first and second order moments (mean and covariance) of the state vector, conditional on all the data in a sample. Provided that all disturbances in the model are Gaussian, a *Kalman Filter* (KF) produces the optimal estimates of such moments in the sense of minimizing the Mean Squared Errors (MSE). An algorithm that is used in parallel with the KF and is not so well-known in certain contexts is the *Fixed Interval Smoothing* (FIS) algorithm, which allows for an operation similar to that of the KF but with a different set of information. The KF used in this chapter is:

$$\begin{aligned}\mathbf{F}_t &= [\mathbf{C}_t \mathbf{R}_t \mathbf{C}_t^T + \mathbf{H}_t \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_t^T] \\ \mathbf{K}_t &= [\Phi_{t+1} \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_t^T] \mathbf{F}_t^{-1} \\ \hat{\mathbf{x}}_{t+1|t} &= [\Phi_{t+1} - \mathbf{K}_t \mathbf{H}_t^T] \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \mathbf{z}_t \\ \hat{\mathbf{P}}_{t+1|t} &= \Phi_{t+1} \hat{\mathbf{P}}_{t|t-1} \Phi_{t+1}^T - \mathbf{K}_t [\Phi_{t+1} \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_t^T]^T + \mathbf{E}_t \mathbf{Q}_t \mathbf{E}_t^T\end{aligned}$$

The backward FIS recursions are:

$$\begin{aligned}\hat{\mathbf{x}}_{t|N} &= \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{P}}_{t|t-1} \mathbf{s}_{t-1} \\ \hat{\mathbf{P}}_{t|N} &= \hat{\mathbf{P}}_{t|t-1} - \hat{\mathbf{P}}_{t|t-1} \mathbf{S}_{t-1} \hat{\mathbf{P}}_{t|t-1} \\ \mathbf{s}_{t-1} &= \mathbf{H}_t^T \mathbf{F}_t^{-1} (\mathbf{z}_t - \mathbf{H}_t \hat{\mathbf{x}}_{t|t-1}) + \overline{\Phi}_t^T \mathbf{s}_t \quad \text{with } \mathbf{s}_N = \mathbf{0} \\ \mathbf{S}_{t-1} &= \mathbf{H}_t^T \mathbf{F}_t^{-1} \mathbf{H}_t + \overline{\Phi}_t^T \mathbf{S}_t \overline{\Phi}_t \quad \text{with } \mathbf{S}_N = \mathbf{0} \\ \overline{\Phi}_t &= \Phi_t - \Phi_t \hat{\mathbf{P}}_{t|t-1} \mathbf{H}_t^T \mathbf{F}_t^{-1} \mathbf{H}_t\end{aligned}$$

This general SS formulation is capable of handling many nonstationary linear dynamical systems; also it can model nonlinear systems but conditionally Gaussian; general heteroscedastic systems; time-varying systems; etc. In addition, many kinds of extensions of model have been proposed in the literature, such as linear approximations of functionally nonlinear dynamic systems; non-Gaussian disturbances; etc. Missing data is not a problem given the recursive nature of the algorithms, because such data are replaced by their

expectations based on the model and the data. Then, if such data is at the end of the sample the KF produces forecasts of the signal, while if they are in the middle or at the beginning both algorithms produce interpolation or forecasts from the beginning of the series backwards.

The application of the recursive KF/FIS algorithms requires values for all the system matrices Φ_t , E_t , H_t , C_t , Q_t , and R_t . Most of the elements of these matrices must be estimated by efficient methods. The Maximum Likelihood (ML) method in the time domain by means of 'prediction error decomposition' ([24] and [15]) is the most common because of its generality and good theoretical properties.

4. Filtering with Integrated Random Walks (IRW)

4.1. Data

Approximately 55 % of railway infrastructure component failures on high speed lines are due to signalling equipment and turnouts. "Signalling equipment" covers signals, track circuits, interlockings, automatic train protection (ATP) or LZB (track loop based ATP), and the traffic control centre. From another point of view, the annual cost of maintaining points is rather high compared to other infrastructure elements, about 3.4 million UKP (United Kingdom Pound) per year for about 1000 km of railway. TC-TCR trade circuits, for example, cost 2.1 million UKP per year for the same area. Of the points expenditure, 1.2 million UKP is for clamp lock type (hydraulic) turnout and 1.4 UPK million for electrically operated turnouts (data provided by a British asset manager). Turnouts can also be used to implement flank protection for a train route allocated to another train. This is achieved by positioning the blades of the turnout in such a way that a train driving through the turnout is not directed into a track segment belonging to the route of another train.

Most standard point machines (see Fig. 1) contain a switch actuating and a locking mechanism which includes a hand-throw lever and a selector lever to allow operation by power or hand. The mechanism is normally divided into three major subsystems: (i) the motor unit which may include a contactor control arrangement and a terminal area; (ii) a gearbox comprising spur-gears and a worm reduction unit with overload clutch; and (iii) the dual control mechanism as well as a controller subsystem with motor cut-off and detection contacts. Generally, there are also mechanical linkages for the detection and locking of the point. The standard railway point is therefore a complex electro-mechanical device with many potential failure modes.

The circuit controller includes detection switches and a pair of snap-action switches to stop the machine at the end of its stroke and to brake the motor electrically so that the mechanism is not subject to impacts. The detection switches have high pressure wiping contacts made of silver/cadmium oxide or gold and they are operated by both the lockbox and the detection rod. The detection switches have additional contacts to allow mid-stroke short circuiting of the detection relays to avoid wrong indications in the signal box or electronic interlocking.

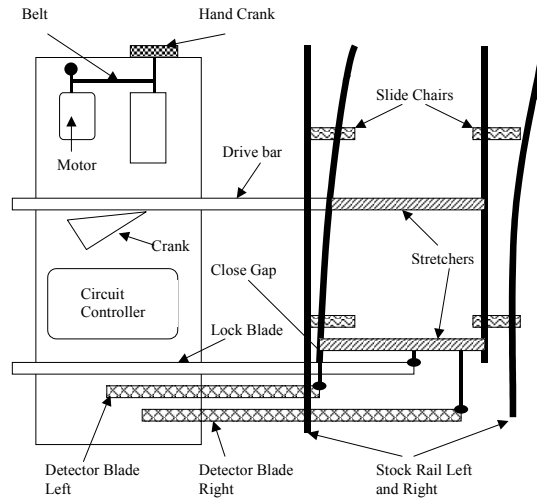


Fig. 1. Point Mechanism

476 experiments (point moves or attempted point moves) were carried out while collecting time, force and operating current data. The data from the point mechanism is initially classified in terms of direction of movement, i.e., either reverse to normal direction or normal to reverse direction. For both directions, faults have been detected with “current (A) vs. time (s)” curves and “force (N) vs. (s)” curves (see some examples in Fig. 2(a) and 2(b)). It was observed that “current (A) vs. time (s)” curves are not the best choice for detecting faults in point mechanisms. The final classification of faults employs only the magnitude and the moment when they change with respect to the reference curves.

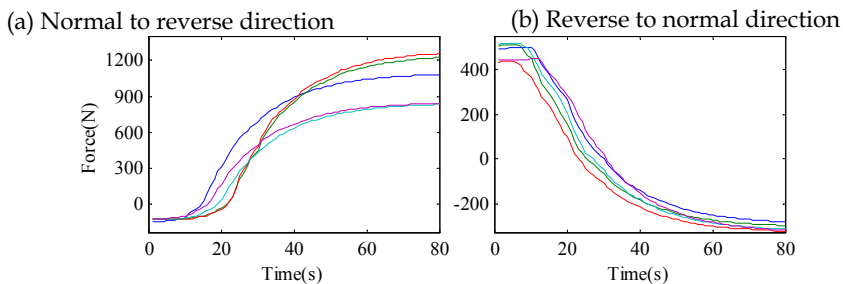


Fig. 2. Operating force curves for a point mechanism

For detecting faults in point mechanisms, a model was employed that can determine the dynamic character of the system. For instance, the reference signals or curves for detecting faults depend on the environmental conditions (temperature, humidity, etc.), and on the in service time of the system, because the friction forces are larger at the beginning than once the system has worn in. The available data consists of 79 curves for the reverse to normal direction, including 4 curves “as commissioned”, and 72 curves for the normal to reverse

direction, with 3 curves “as commissioned” (some of them may be seen in Fig. 2). A reference dynamic system has to be applied to all of these variables. The data collected refers to force (N) versus time (s). The first conclusion after studying these curves is that we can detect only a few faults by analyzing the signal directly but, if we analyze the differences between the current data x^j and the reference data x^i in the form of absolute values $d_i(1)$, we can detect the majority of faults as they develop.

$$d_t^j = |x_t^j - x_t^i|, \quad \forall t \quad (1)$$

Some of these curves are shown in Fig. 3(a) and 3(b) for reverse to normal direction and normal to reverse direction respectively. The ‘x’ axis is time [s] and the ‘y’ axis is the difference between the dynamic mean geometric and the current curve as an absolute value [N].

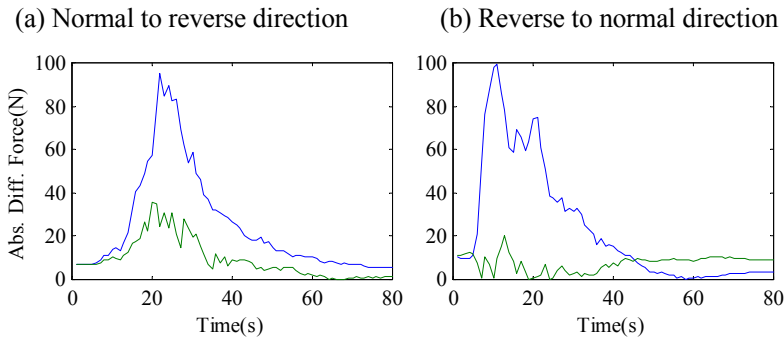


Fig. 3. Difference between the reference signal for the point and the newly acquired data in absolute values

4.2. The model

One feasible model written in SS form (1) for this application is of the type local mean plus noise for two signals simultaneously, where the local means are modeled by the dynamics implied by the state equations, i.e.

$$\begin{aligned} \mathbf{x}_{t+1} &= \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \mathbf{x}_t + \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix} \\ \mathbf{z}_t = \text{signal} + \text{noise} &= (\mathbf{I} \quad \mathbf{0}) \mathbf{x}_t + \mathbf{v}_t \end{aligned} \quad (2)$$

$$\mathbf{Q} = \begin{pmatrix} \sigma_{w_1}^2 & \rho \sqrt{\sigma_{w_1}^2 \sigma_{w_2}^2} \\ \rho \sqrt{\sigma_{w_1}^2 \sigma_{w_2}^2} & \sigma_{w_2}^2 \end{pmatrix}; \quad \mathbf{R} = \begin{pmatrix} \sigma_{v_1}^2 & \sigma_{v_1 v_2} \\ \sigma_{v_1 v_2} & \sigma_{v_2}^2 \end{pmatrix}$$

In model (2) all the system matrices are time invariant: \mathbf{I} is a two dimensional identity matrix; $\mathbf{0}$ is a two by two matrix of zeros; σ_{\cdot}^2 are the variances of the noise signals or disturbances either in the state or observation equations; $\sigma_{\cdot\cdot}$ is the covariance between two disturbances; and ρ is the correlation coefficient between the two noise signals in the state equation.

By comparing systems (2) and (1) it is easy to see the system matrices values in this particular case, i.e.

$$\Phi_t = \begin{pmatrix} \mathbf{I} & \mathbf{I} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}; \mathbf{E}_t = \begin{pmatrix} \mathbf{0} \\ \mathbf{I} \end{pmatrix}; \mathbf{w}_t = \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}; \mathbf{H}_t = (\mathbf{I} \ \mathbf{0}); \mathbf{C}_t = 1$$

The unknown *hyper-parameters* to be estimated by ML in this model are \mathbf{Q} and \mathbf{R} . It should be noted that \mathbf{Q} is parameterized in the way shown above in order to force the appearance of the correlation coefficient between the state disturbances explicitly. The following points must be taken into account when interpreting model (2):

- The observation equation implies that the series are composed of a local mean level or trend with added noise.
- The first two states in the model are the local mean level (or trends) of each series. In other words, they are the signals free from noise;
- Given the structure of the model, it is easy to show that the third and fourth states are the gradients of the trends. The slopes are modelled here as stochastic and therefore changing as a function of time according to the variance of the state disturbances;
- If the correlation coefficient is 1, both trends are proportional to each other, meaning that the dynamic behaviour of both trends is the same. This is an important point that the authors wanted to test later;
- By definition, σ^2 must be positive; $-1 \leq \rho \leq 1$; and \mathbf{R} must be positive definite. Since all these are parameters to be estimated, it may be advantageous constrained search algorithms;
- The asymptotic distribution of the ML estimates are Gaussian if all the disturbances in model (2) are Gaussian. Then, since ρ is estimated explicitly, the confidence intervals and statistical hypothesis tests for this parameter may be easily constructed.

In fact, the parameter ρ is proposed here as a way to discriminate between “faulty” and “as commissioned” curves (see below), where the “faulty” curve is caused by wear as described above. Strictly speaking, the two curves are behaving in the same way when $\rho=1$, but previous experience with point mechanisms of a similar kind must be incorporated here, because it is, difficult, in general to find those values in practical situations. Then, a cut-off value of ρ must be considered in order to discriminate between ‘good’ and ‘bad’ curves.

The modeling strategy outlined above may be applied to both off-line and on-line situations. In this latter case it would be possible to get an estimated time series for ρ (with confidence bands) and the time of wear assessment detected on-line very quickly when parameters start to move away from their initial values. Even forecasts of the current curve may be produced at any point in time, based on the current parameter values and the future data of the reference curve.

Very fast algorithms have been developed for ML estimation of SS systems in which all the unknowns are some elements of the covariance matrices \mathbf{Q} and \mathbf{R} , such as in model (4).

The problem of initializing the KF and hence ML needs to be resolved. One of the most important tools is the use of the *exact likelihood function* [5] and [6].

4.3. Experimental Results

The model described in the previous subsection was employed in an off-line mode with data collected during laboratory tests (see Fig. 2). The model output (shown in Fig. 4, based in signals from Fig. 3) was then used to classify the curves as either “as commissioned” or “faulty”. This step may be achieved several ways. The approach compares ρ with the individual points in time with a relating high threshold value. A value of ρ below the threshold is an indication of a lack of correlation with the current reference curve and therefore is classified as “faulty”ⁱ. A more refined and somewhat more formal criterion is based on such single point estimate and its 95% confidence band. In this case, a curve is considered to be “as commissioned” if the upper limit of the confidence band is close to target value or equal to 1.

For point operation in both directions, with a value of $\rho = 0.99$ the totality of faulty curves could be detected. In the NR direction, since the highest value of ρ for faulty curves was 0.92 and the 95% confidence interval uses (0.77, 0.98). In the RN direction, the highest value of ρ for faulty curves was 0.97 and the 95% confidence interval was (0.93, 0.99).

The results achieved with the same reference curve, but different test results are shown in Fig. 4, one “as commissioned” curve (top panel) and one faulty curve (bottom).

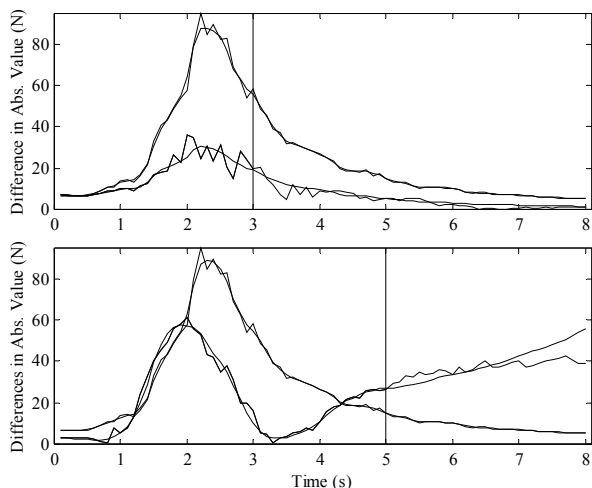


Fig. 4. Two examples of forecasts based on model (4) at different forecast origins. One “as commissioned” curve (top) and one “faulty” curve (bottom). Forecast origins are marked by the vertical line.

In both cases the reference curve was available for the whole time span (based on previous curves taken from the system) and the information to test each curves was set up to the

ⁱ Alternatively, the estimated correlation coefficient may be tuned so that the number of curves correctly classified is maximised.

forecast origin (vertical line). The objective of obtaining a forecast for the behavior of the system based on such incomplete information was thus using model (4). In an on-line situation, the parameters and the forecasts are updated each time a new observation is available.

Fig. 5 shows the recursive estimate of ρ with its 95% confidence intervals (assuming gaussian noises) for an “as commissioned” curve (top) and a “faulty” one (bottom). In both cases the confidence on the estimate tends to increase as more information becomes available.

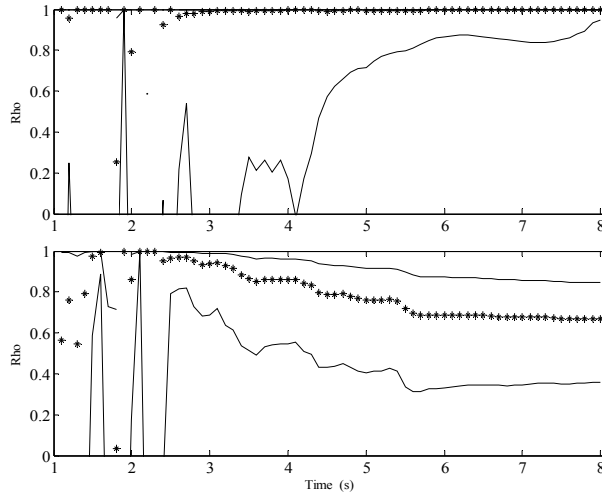


Fig. 5. Recursive estimation of ρ (stars) and 95% confidence bands (solid) for one “as commissioned” curve (top) and one “faulty” curve (bottom).

5. Random Walks and smoothing

5.1. Device and data

Following successful implementation on a level crossing mechanism (Roberts 2002) [23], the authors adapted the methods to detect faults in seven point machines at Abbotswood junction, shown in Fig. 6 as boxes 638, 639, 640, 641A, 641B, 642A and 642B.

The configuration deployed at Abbotswood junction was developed in collaboration with Carillion Rail (formerly GTRM), Network Rail (formerly RailTrack) and Computer Controlled Solutions Ltd. The junction consists of four electro-mechanical M63 and three electro-hydraulic point machines, shown in Figure 2. Each M63 machine is fitted with a load pin and Hall-effect current clamps. The electric-hydraulic point machines are instrumented with two hydraulic pressure transducers, namely an oil level transducer and a current transducer. A 1 Mb/sec WorldFIP network, compatible with the Fieldbus standard EN50170 (CENELEC EN50170 2002) [4], connects the trackside data-collection units to a PC located in the local relay room. Data acquisition software was written to collect data with a sampling rate of 200 Hz. Processed results can be observed on the local PC and also remotely.

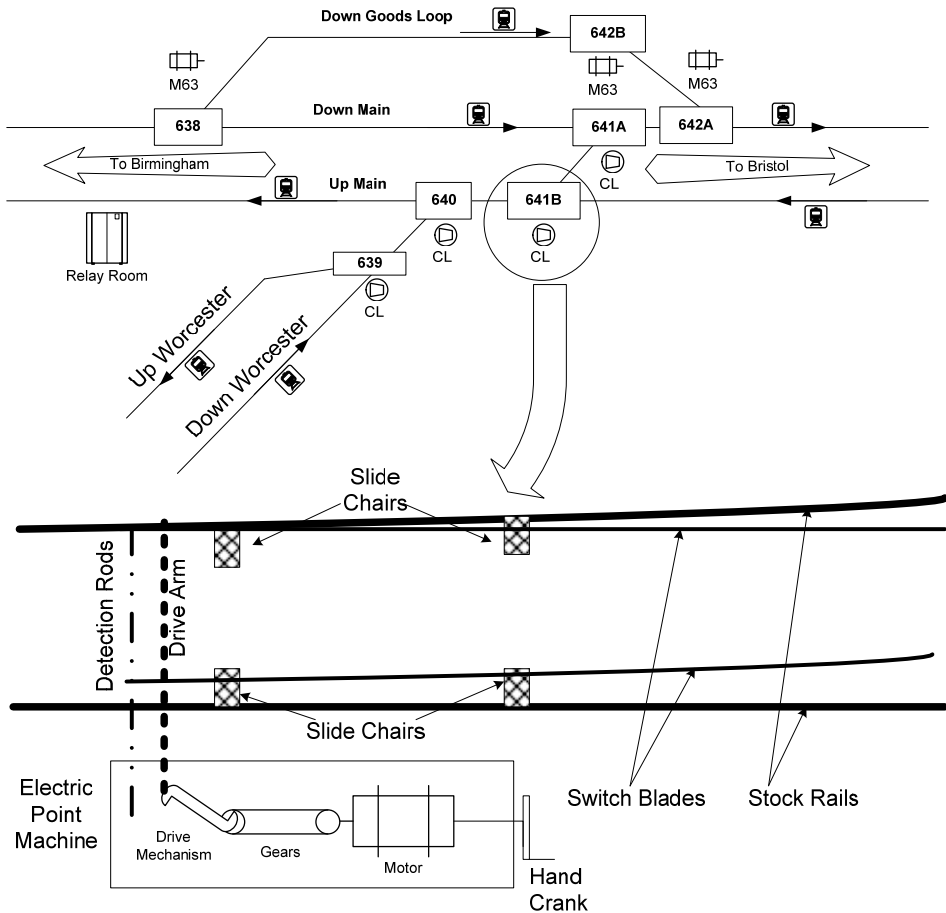


Fig. 6. Set of points and the relevant components/sub-units at Abbotswood junction.

The supply voltage of the point machine was measured (Fig. 7a), as well as the current drawn by the electric motor (Fig. 7b) and the system as a whole (Fig. 7d). In addition, the force in the drive bar was measured with a load pin introduced into the bolted connection between the drive bar and the drive rod (Fig. 7c). Fig. 7 shows the raw measurement signals taken in the fault-free (control or “as commissioned”) condition for normal to reverse and reverse to normal operation, respectively. Note that the currents and voltages begin and end at zero for both directions of operation, but a static force remains following the reverse to normal throw and a different force remains after the normal to reverse throw.

It is difficult to compare the measurements taken during induced failure conditions with those from the fault-free condition because of noise in the measurements.

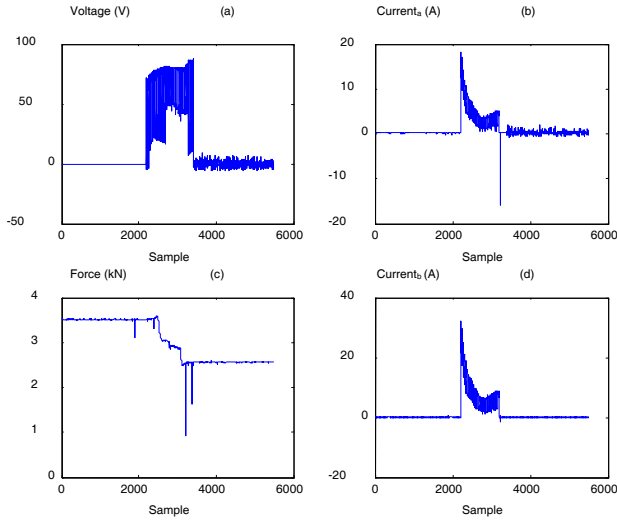


Fig. 7. 'As commissioned' measured signals for the normal to reverse throw

5.2. Filtering the signal

One possibility to reduce the noise is by using the SS formulation in (1) as a digital filter capable of reducing observation noise when the measured quantity varies slowly, but additive measurement noise covers a broad spectrum [8], [9]. In this particular case the signal being measured is modeled as a random walk, i.e. it tends to change by small amounts in a short time but can change by larger amounts over longer periods of time. The SS model used for each signal is described by equations (3).

$$\left. \begin{aligned} x_{t+1} &= x_t + w_t \\ z_t &= x_t + v_t \end{aligned} \right\} \quad (3)$$

$$Q = E(w_t^2), \quad R = E(v_t^2)$$

Comparing with the general SS equations (1) we have:

- Variables x_t , z_t , Q , R , w_t and v_t are all scalars.
- $\Phi_t = 1$; $E_t = 1$; $w_t = w_t$; $H_t = 1$; $C_t = 1$.
- The initial value given to \hat{x}_0 is: $\hat{x}_0 = 0$.
- The initial value of P_0 is chosen to reflect uncertainty in the initial estimate. Here P_0 is initialised as $P_0 = 10^6$.
- The remaining quantities to be specified are Q , the variance of the noise driving the random walk, and R , the variance of the observation noise.

By empirical methods using simulation, the best filtering is achieved with $Q = 0.03$ and $R = 0.5$. Note that the ratio Q/R defines the filter behavior.

The power spectral density of the filtered motor current (computed only while the motor is running) shows significant energy peaks at 100 and 200 Hz (Fig. 8, where the normalized frequency of 1 corresponds to a frequency of 250 Hz).

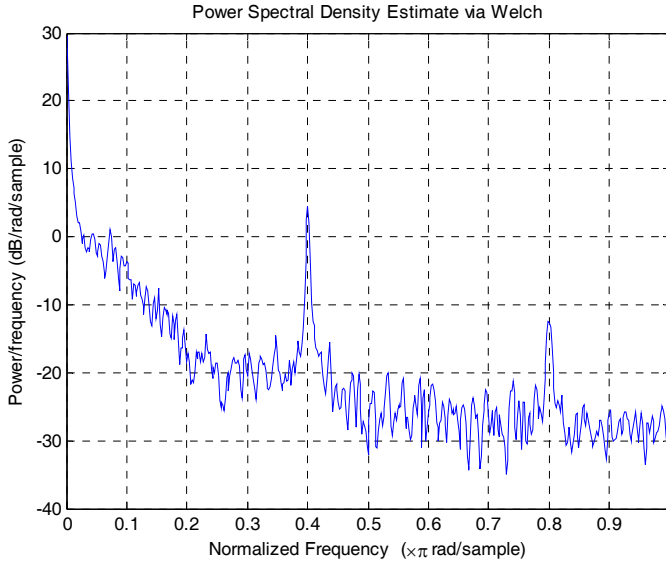


Fig. 8. Motor current power spectral density following Kalman filtering

The dynamic model used can be augmented to model the observed interfering signals as narrow band disturbances centred at 100 and 200 Hz. The spectrum of the motor current signal is examined next before a decision on the most appropriate filtering is taken.

A spectral analysis of the motor current signal against time (or sample) shows that the characteristic of the noise varies with the operating condition of the motor. From the spectrogram one can identify a small 50 Hz interference signal before the motor begins to turn (samples 1 to 1100). In the second stage, where the motor is turning, the interfering signal has strong 100 Hz and 200 Hz components but no 50 Hz component. In the final stage, the motor current does not have identifiable 50, 100, or 200 Hz components, but is affected by general wideband noise.

Power spectral densities (psds) were computed for data selected from each of the three distinct operating regions. There is a 50 Hz interference signal during the first region and wideband noise during the last. Fig. 9 shows the psd for the middle phase, which is the noisiest region. It is possible to augment the SS model to describe the observed interfering signals, using different models for each of the three distinct phases. However, a simpler yet effective smoothing scheme exists, as described in the next section.

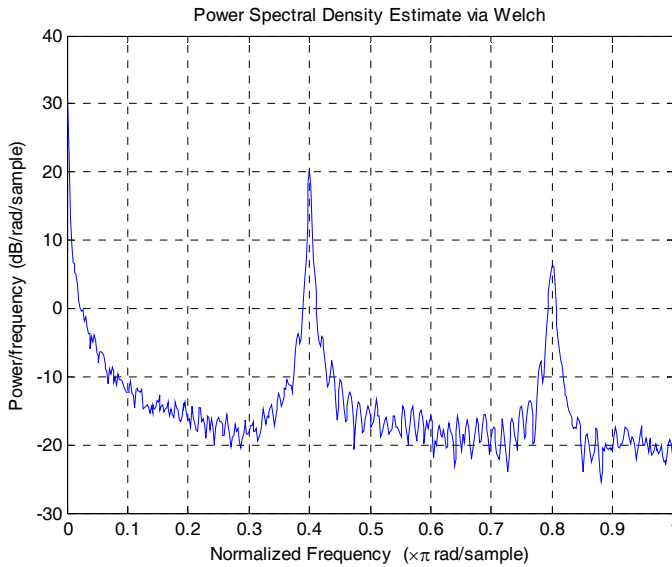


Fig. 9. Power Spectral Density estimate (samples 1000 to 4000).

5.3. Smoothing

Noting that the sampling rate is 500 Hz and the interfering signals appear at 50, 100 and 200 Hz, an alternative filtering method, or, more correctly, smoothing method, is to compute a moving average of the original signal over a suitable number of samples. For example, computing the moving average with 10 samples has zero response to signals at 50 Hz. However, a 100 Hz signal, with only 5 samples per cycle, is not necessarily removed, depending on the relative phase of the 100 Hz signal and the samples. Removal of the 50 Hz, 100 Hz and 200 Hz interfering signals is guaranteed by computing a moving average over 40 samples, i.e. over a time window of 80 ms. This moving average also spreads an instantaneous motor current change over 80 ms, but this is not a problem in practice as the motor current does not change instantaneously. A moving average computed over 40 samples (80 ms) removes information at 12.5 Hz (and integer multiples thereof) and in addition acts as a general first-order low pass filter with a -3 dB point at 5.5 Hz. Losing information around 12.5 Hz is not important as long as comparisons are made between identically processed signals. By suitable alignment of the moving average result, filtering becomes smoothing. The smoothed signals are delayed by 40 ms, but this is of no concern for comparison with similarly processed fault-free signals. There is still some residual 100 and 200 Hz interference, but it is much reduced. Identical smoothing has been applied to all measurement channels, even though they are not equally affected by 50 Hz noise and its harmonics. A comparison of the smoothed signals with the corresponding signals obtained in the fault-free condition is now possible.

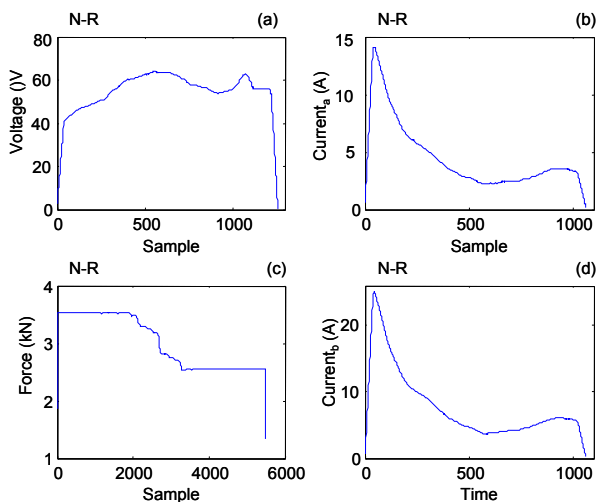


Fig. 10. Average control curves. N-R: Normal to Reverse Direction

5.4. Results

The failure modes listed are identified using a pattern recognition method. The signals obtained in the fault-free condition, smoothed as described above and averaged over five throws, are shown in Fig. 10. The smoothed signals obtained under induced failure modes have been compared to the reference (or control) signals.

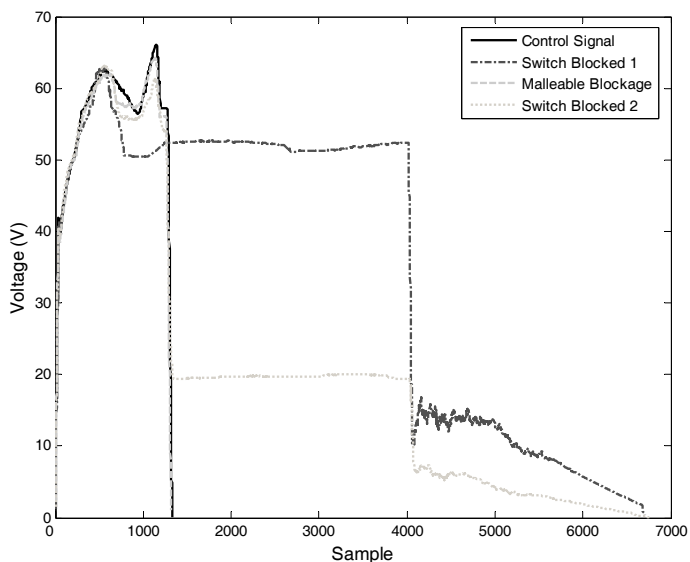


Fig. 11. A Control signal and Switch Blocked and Malleable Blockage failure modes signals

Fig. 11 shows the voltage signals for the failure modes Switch Blocked 1, Switch Blocked 2 and Malleable Blockage, in the normal to reverse direction.

Every failure can potentially be detected from signals a, b and c for normal to reverse transitions, and using signals b and c for reverse to normal transitions. Therefore, employing only signal b or c it potentially is possible to detect every fault in both operating directions.

6. Advanced Dynamic Harmonic Regression (DHR)

The system developed in this section detects faults by means of comparing what can be considered a “normal” or “expected” shape of a signal with respect to the actual shape observed as new data become available. One important feature of this system is that it adapts gradually to the changes experienced in the state of the point mechanism. The forecasts are always computed by including into the estimation sample the last point movements and discarding the older ones. In this way, time varying properties of the system due to a number of factors, like wear, are included, and hence the forecasts are adaptive.

The data is a signal with long periods of inactivity, mixed up with other short periods where a point movement is being produced. Fig. 12 shows one small part of the dataset in the later case study, where the time axis has been truncated in order to show the movements of the signal. The real picture is one in which the inactivity periods are much longer than those shown in the figure, in a way that the movement periods would appear as thin lines.

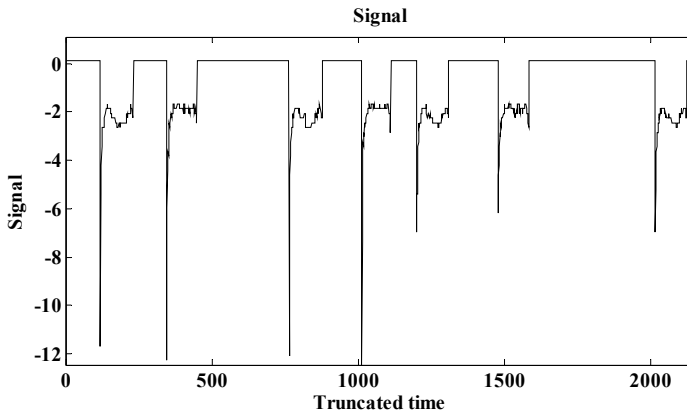


Fig. 12. Signal used by the fault detection algorithm.

A new signal can be composed exclusively of those time intervals where the point mechanism is actually working. Looking at Fig. 12 it can be devised that even movements (normal to reverse move) have a slightly different pattern than uneven movements (reverse to normal). Therefore, two signals may be formed by concatenating the normal to reverse movements of the point mechanism in one hand, and the reverse to normal moves in the other. Fig. 13 shows one portion of the normal to reverse signal.

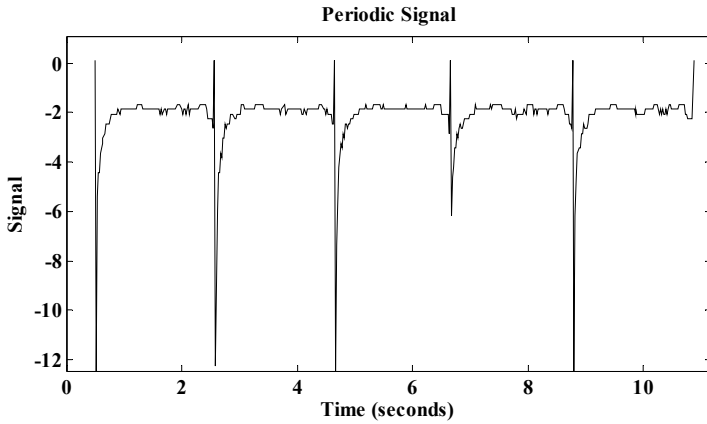


Fig. 13. Signal obtained by concatenation of portions of data where the point mechanism is working.

As it is clearly shown in Figure 13, the signal to analyse has strong periodicity and can be then modelled and forecast by a statistical model capable of replicating such behaviour. The period of the signal is exactly the time it takes to the point mechanism to produce a complete movement. Two difficulties arise that should be considered by the model: (i) the sampling interval of the data is not constant, it has small variations produced by the measurement equipment that should be taken into account; and (ii) the frequency or period of the waves changes over time. As a matter of fact, the changes of the period may be considered as a measurement of the wear in the system, as illustrated in Figure 14.

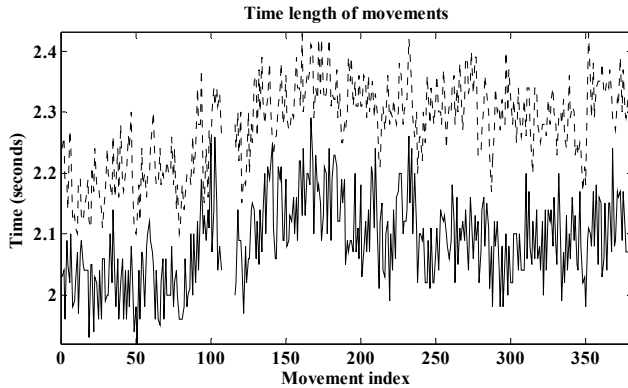


Fig. 14. Time the point mechanism spend to produce movements in normal to reverse direction (solid) and reverse to normal (dotted).

Fig. 14 shows the 380 time varying periods (or time to produce a complete movement of the mechanism) for the "normal to reverse" and "reverse to normal" signals (the first five data points corresponds to the signal shown in Fig. 13) that constitutes the full data set in the

later case study. There were several sudden increases of the period at some points in time due to faults that have been removed from the figure, in order to avoid distortions of the vertical axis. The time axis is on an irregular sampling interval, in order to take into account the moment at which each movement has taken place. It is clear that the period is lower at the beginning of the sample with a rapid increase that tends to come down from the middle of the sample. A similar behaviour is devised in the reverse to normal signal.

The fault detection algorithm proposed here in essence would be composed of the following steps:

1. Forecasting next period on the basis of the signal in Figure 14.
2. Forecasting the signal in Figure 13 by a Dynamic Harmonic Regression model that uses the period forecast of the previous step.

Assessing forecasts by comparing the forecast of step 2 with the actual signal coming from the sensors installed in the point mechanism. If the forecasts generated in step 2 are too bad (measured by the variance of the forecast error), a fault is detected. The way to assess whether a failure has been produced is by checking the variance of the forecast error above a certain level fixed for each specific point mechanism.

6.1. Step 1: Modeling and forecasting the period

Two procedures have been considered: i) VARMA models in discrete time with two signals (the periods for normal to reverse and reverse to normal) modeled jointly; ii) once again a univariate local level model plus noise, but in continuous time.

6.1.1. VARMA model

The VARMA (Vector Auto-Regressive Moving-Average) models (see e.g. [1], [18] and [25]) are natural extensions of the ARIMA (Auto-Regressive Integrated Moving Average) models to the multivariate case. One of the simplest but general formulations of a VARMA(p, q) model is

$$\mathbf{P}_t = \boldsymbol{\varphi}_1 \mathbf{P}_{t-1} + \dots + \boldsymbol{\varphi}_p \mathbf{P}_{t-p} + \mathbf{v}_t + \boldsymbol{\Theta}_1 \mathbf{v}_{t-1} + \boldsymbol{\Theta}_q \mathbf{v}_{t-q} \quad (4)$$

where $\mathbf{P}_t = [p_{1,t} \quad p_{2,t}]^T$ is a bivariate signal; \mathbf{v}_t is a bivariate white noise, i.e. purely random signal with no serial correlation and covariance matrix \mathbf{R} ; and $\boldsymbol{\varphi}_i$ ($i=1,2,\dots,p$) and $\boldsymbol{\Theta}_j$ ($j=1,2,\dots,q$) are squared blocks of coefficients of dimension 2×2 .

VARMA models admit several SS representation according to equation (1). The one preferred here is (with $r = \max(p, q)$)

$$\mathbf{x}_{t+1} = \begin{pmatrix} \boldsymbol{\varphi}_1 & \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} \\ \boldsymbol{\varphi}_2 & \mathbf{0} & \mathbf{I} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\varphi}_{r-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} \\ \boldsymbol{\varphi}_r & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{pmatrix} \mathbf{x}_t + \begin{pmatrix} \boldsymbol{\varphi}_1 + \boldsymbol{\Theta}_1 \\ \boldsymbol{\varphi}_2 + \boldsymbol{\Theta}_2 \\ \vdots \\ \boldsymbol{\varphi}_{r-1} + \boldsymbol{\Theta}_{r-1} \\ \boldsymbol{\varphi}_r + \boldsymbol{\Theta}_r \end{pmatrix} \mathbf{v}_t$$

$$\mathbf{z}_t = (\mathbf{I} \ \mathbf{0} \ \mathbf{0} \ \cdots \ \mathbf{0}) \mathbf{x}_t + \mathbf{v}_t$$

The model orders p and q can be identified using multivariate autocorrelation and multivariate partial autocorrelation functions. The block parameters, as well as the covariance matrix of the noise, are estimated using Maximum Likelihood. Forecasts are then computed on the basis of the actual data and the estimates of the model parameters, once the model passes a validation process. One of the most important validation tests is the absence of serial correlation in the perturbation vector noise \mathbf{v}_t (see e.g. [1], [18] and [25]).

It is vital that the signals \mathbf{P}_t on which all the VARMA methodology is applied should have stationary mean and variance.

6.1.2. Local level model in continuous time

The model used for forecasting the period of the next movement (in a particular direction) in this case represents the observation, i.e. the period drifts over time, as wear varies simply because of usage (increases) or by preventive maintenance (decreases). Since the point movements are not produced at equally spaced intervals of time, a continuous-time model should be used. Formally, the continuous time SS model is given by

$$\frac{d}{dt} \begin{bmatrix} l(t) \\ s(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} l(t) \\ s(t) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} \quad (5)$$

$$P(t) = l(t) + v(t)$$

with

$$\mathbf{Q} = \begin{bmatrix} q_1 & 0 \\ 0 & q_2 \end{bmatrix}$$

where $P(t)$ stands for the time varying period that is decomposed into the local level $l(t)$ and a noise term $v(t)$ assumed to be white Gaussian noise; $w_1(t)$ and $w_2(t)$ are independent white noises.

One way to treat the continuous system above is by finding a *discrete-time* SS equivalent to it (see e.g. Harvey 1989) [15], by means of the solution to the differential equation implied by the system. A change in notation is necessary to convert the system to discrete-time: denote the k th observation of the series z_k (for $k=1,2,\dots,N$) and assume that this observation is made at time t_k . Let $t_0 = 0$ and $\delta_k = t_k - t_{k-1}$, i.e. the time interval between two consecutive measurements. System (3) may be represented by the *discrete-time* SS system in (5).

$$\begin{bmatrix} l_k \\ s_k \end{bmatrix} = \begin{bmatrix} 1 & \delta_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} l_{k-1} \\ s_{k-1} \end{bmatrix} + \begin{bmatrix} w_{1,k} \\ w_{2,k} \end{bmatrix} \quad (6)$$

$$P_k = l_k + v_k$$

In order to make systems (6) and (5) equivalent, the variances of observational noise is unchanged as R , but the covariance matrix of the process noise in the state equations becomes

$$\mathbf{Q}_k = \delta_k \begin{bmatrix} 1/3\delta_k^2 q_2 + q_1 & 1/2q_2\delta_k \\ 1/2q_2\delta_k & q_2 \end{bmatrix}$$

(see Harvey 1989, page 487) [15]. If all the data are sampled at regular time intervals, then $\delta_k = \delta$ and the noise variances are all constant; but if the data is irregularly spaced, as it is in our case, δ_k would take into account the irregularities of the sampling process. It is worth noting that the continuous-time model (5) involved system matrices that are all constant and the state noises were all independent of each other with constant variances. Beware that system (6) is written in form (1) and is the only case in this chapter that involves a time variable transition matrix Φ_k and time variable variance noises that are correlated to each other according to the expression of \mathbf{Q}_k .

6.2. Step 2: Modeling and forecasting the signal

Once the period or the time length of the next movement of the point mechanism is forecast by any of the models in section 5.1., it is necessary to produce the forecast of the signal itself for the next occurrence, in order to produce what should be expected in case of no faults.

This is done by a Dynamic Harmonic Regression model (DHR) set up as described below. This model is very convenient in the present situation because it can easily handle the time-varying nature of the movement period. Obviously, the model can also be written in the form of a SS system as in (1).

The formula of a DHR with the required properties is shown in equation (7).

$$z_{k,t} = \sum_{i=1}^M [a_{i,k} \sin(\omega_{i,k,t} t^*) + b_{i,k} \cos(\omega_{i,k,t} t^*)] + e_{k,t} \quad (7)$$

Here, $z_{k,t}$ is the periodic signal in which the subscript k indicates whether the normal to reverse ($k = 1$) or the reverse to normal ($k = 2$) signals are being considered; M is the number of harmonics that should be included in the regression to achieve an adequate representation of the signal $z_{k,t}$; $a_{i,k}$ and $b_{i,k}$ are $2M$ parameters to be estimated, representing the amplitudes of the co-sinusoidal waves; $\omega_{i,k,t}$ are frequencies at which the

sinusoids are evaluated, with $\omega_{i,k,t} = 2i\pi/p_{k,t}$ for $i=1,2,\dots,M$ and $M \geq p_{k,t}/2$ and $k=1,2$; $e_{k,t}$ is a pure random white noise with constant variance. Separate Harmonic Regression models are used for the normal to reverse and reverse to normal signals.

There are two key points for the model (7) to be an adequate representation of $z_{k,t}$:

1. $p_{k,t}$ and $\omega_{i,k,t}$ have time varying period/frequency. The nature of such variation is dependent on the signal itself. For one full movement of the point mechanism $p_{k,t}$ is maintained constant and is equal to the time it takes to produce the full movement. This value will be different in the next movement and is modified accordingly.
2. The time index t^* is a variable linked to $p_{k,t}$ that varies from 0 to $p_{k,t}$ in each movement. Therefore, this variable is reset to 0 as soon as a movement finishes (see Fig. 15).

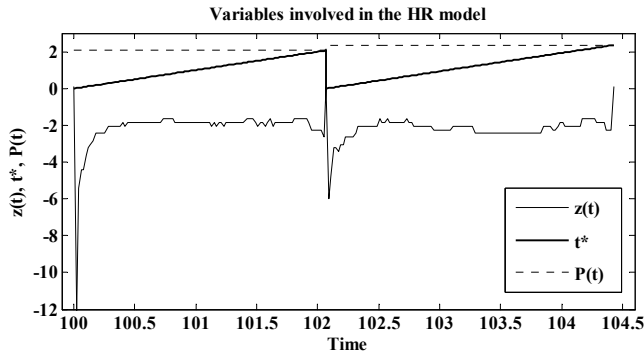


Fig. 15. Two full movements of the point mechanism, with their associated period and time index according to model (7).

Model (7) is then a regression of a signal on a set of deterministic functions of time and therefore all the standard regression theory can be applied, in particular estimates and forecasts can be made quickly. Model (7) have been generalized further by allowing parameters $a_{i,k}$ and $b_{i,k}$ to be time varying, producing a more flexible model, known as a Dynamic Harmonic Regression (DHR; see [21] [26]), but such complications are not found necessary in the case study described later.

6.3. The full fault detection algorithm

The full algorithm for fault detection comprises the following steps:

1. Determine which historical data to use. In the later case study the previous 50 free-from-faults movements of the point mechanism are used to estimate models (4) (5) and (7) at each new movement.

2. A point forecast of the time that it would take the next movement is produced by means of model (4) or (5), together with its 95% confidence interval. In this way, a range of lengths or periods of the next movement are considered. Then, a different forecast of the signal $z_{k,t}$ is produced for each period forecast in the previous step. Following this a full set of forecasts become available for a time horizon long enough to cover a full movement of the point mechanism.
3. The new data points measured by the system are compared to all the forecasts produced in the previous step. The forecast closer to the actual data measured by the minimum of the standard deviation of the error is then considered to be the best forecast of the signal.
4. If the best forecast is systematically bad, a fault has occurred and the system issues a warning. If the best errors are always low, no faults are detected. The boundary is measured in terms of standard deviation of the errors and such a value has to be adjusted for each particular point mechanism.
5. If no fault is detected, then the data of the latest movement is incorporated into the historical data to be used next time, the oldest movement data being dropped. However, if a fault is detected, the historical data used to perform step 1 for the next movement is unchanged for the next movement.

The algorithm can be used in on-line or off-line contexts. For on-line use, step 3 can be repeated as each measurement data point becomes available. For off-line use the algorithm is applied to all the data collected for a full movement of the mechanism.

The system requires a couple of values to be fixed by experimentation, namely the alarm limit that can be calculated from the standard deviation of signal $z_{k,t}$, and also the number of harmonics to include in the Harmonic Regression (M in model (7)). Experiments carried out on logged data have been performed to set these two design parameters of the algorithm. The final setting for the standard deviation is 0.4 for the standard deviation, found to give the best discrimination between faulty and non-faulty events; and $M = 62$ harmonics for model (7) produces accurate fit and forecasts to the signal.

6.4. Results

Standard identification techniques on VARMA models suggested a VARMA(0, 1). Estimation of such a model for the full data set was

$$\begin{bmatrix} p_1 \\ p_2 \end{bmatrix}_t = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix}_{t-1} + \begin{bmatrix} -0.89 & 0.13 \\ 0 & -0.71 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_{t-1} + \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}_t$$

$$\mathbf{R} = \begin{bmatrix} 3.3 & 0.9 \\ 0.9 & 2.5 \end{bmatrix} \times 10^{-3}$$

The correlation between the components of the noise vector is 0.3. The relation between the output variables can be more easily seen if the model is written in the form of difference equations,

$$\begin{aligned}
 p_{1,t} &= p_{1,t-1} - 0.89v_{1,t-1} + 0.13v_{2,t-1} + v_{1,t} \\
 p_{2,t} &= p_{2,t-1} - 0.71v_{2,t-1} + v_{2,t}
 \end{aligned}$$

The correlation of each variable with its own past is more important than the relation to each other, judging by the coefficients relating both variables and the correlation of noises. Nevertheless, the relation between them is significant and should be taken into account in order to forecast the output variables. The model is adequate in the sense that no serial correlation left in the residuals.

One example is shown in Fig. 16. The top panels show the forecast of the periods to use in the DHR models, with the 95% confidence intervals. Such period is the expected length of the next movement, that is the value introduced into the DHR model to forecast the signal itself. The forecast of the signal is shown in the bottom panels, where the dotted lines are the actual values and the solid lines are the final forecast of the system. It is clear that the left case is free from any fault, since the forecast matches perfectly the actual data, while the expected behavior in the right panel is very different to the actual data, implying that a fault has occurred.

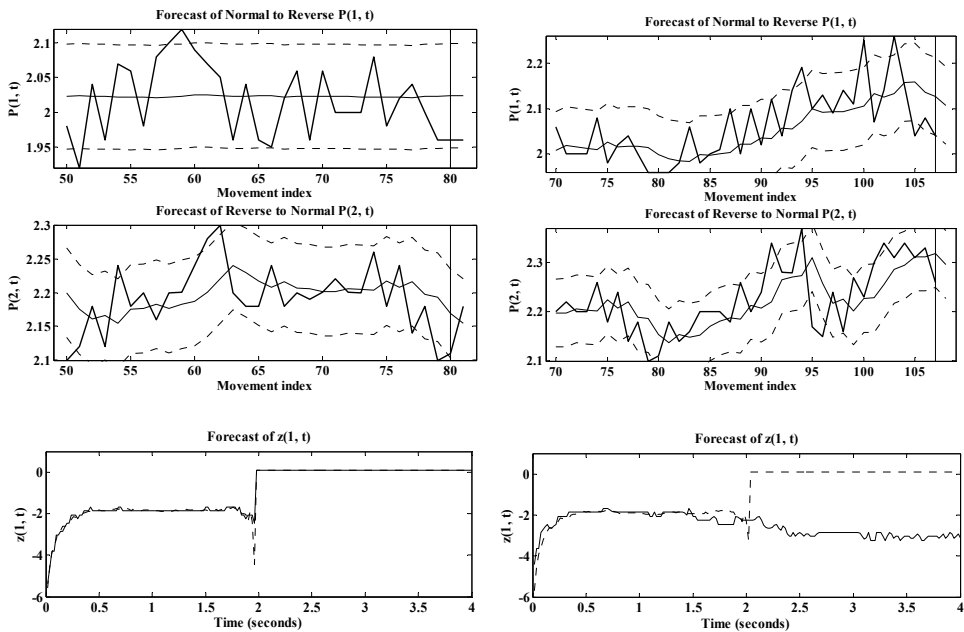


Fig. 16. Left panels shows results for fault free data. Right panels show results for a faulty signal. Panels in the two first rows show the forecast of VARMA model (from the vertical line on); solid lines show the actual periods and the forecast (smoother line). Panels in bottom row show the forecast of the DHR model with the period forecast in the top panels; solid lines are the actual data, dashed lines are the forecast.

Similar results are achieved when the local level model set up in continuous time is used instead (see Fig. 17).

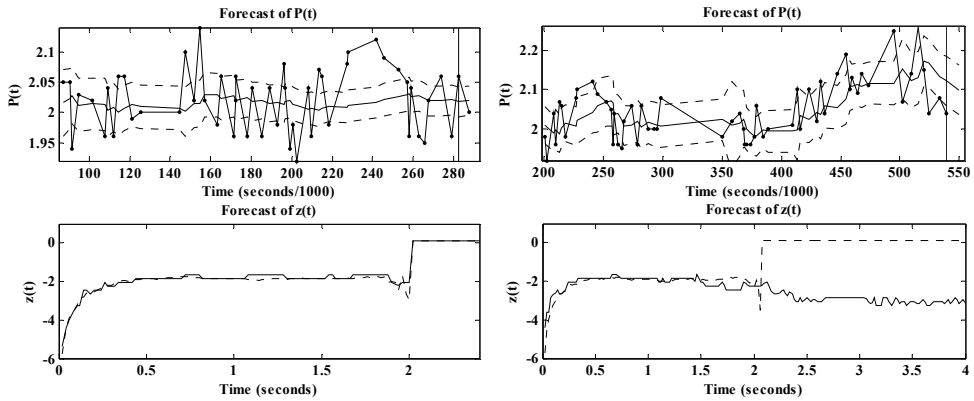


Fig. 17. Left panels shows results for fault free data. Right panels show results for a faulty signal. Panels in the first row show the forecast of the local level model (from the vertical line on); solid lines show the actual periods and the forecast (smoother line). Panels in bottom row show the forecast of the DHR model with the period forecast in the top panels; solid lines are the actual data, dashed lines are the forecast.

This algorithm was applied to the full dataset (380 movements in either directions). From normal to reverse movements 8 were abnormal due to faults similar to the one shown in Figure 17. No faults were registered in the reverse to normal direction data. Selecting a standard deviation of 0.4 as the boundary of faults detection we get that all the faults were detected and not a single false alarm was produced in any of the cases.

7. References

- [1] Box G.E.P., Jenkins G.M., Reinsel G.C. 1994. Time Series Analysis, Forecasting and Control. Englewood Cliffs, New Jersey, Prentice Hall International.
- [2] Bryson A.E., Ho Y.C. (1969). Applied optimal control, optimization, estimation and control. Waltham, Mass.: Blaisdell Publishing Company.
- [3] Casals J., Jerez M., Sotoca S., Exact Smoothing for stationary and non-stationary time series, International Journal of Forecasting, 16 (2000), 59-69.
- [4] CENELEC EN50170 (2002), General purpose field communication system.
- [5] de Jong P., Stable algorithms for the state space model, Journal of Time Series Analysis, 12, (2)(1991) 143-157.
- [6] de Jong P., The likelihood for a state space model, Biometrika, 75, (1)(1988) 165-169.
- [7] Durbin J., Koopman S.J., Time series analysis by state space methods. Oxford University Press, Oxford, 2001.
- [8] García Márquez F.P., Schmid F. and Collado J.C., 2003. "Wear Assessment Employing Remote Condition Monitoring: A Case Study". Wear, Vol. 255, Issue 7-12, pp. 1209-1220.

- [9] García Márquez F.P., Schmid F. and Conde J.C., 2003. A Reliability Centered Approach to Remote Condition Monitoring. A Railway Points Case Study. *Reliability Engineering and System Safety*, Vol. 80 No. 1, pp 33-40.
- [10] Garcia Marquez, F.P and Pedregal D.J. (2004). Failure Analysis and Diagnostics for Railway Trackside Equipment. *Engineering Failure Analysis*, Vol. 14(8), pp. 1411-1426.
- [11] Garcia Marquez, F.P and Pedregal D.J. (2007). Applied RCM² Algorithms Based on Statistical Methods. *International Journal of Automation and Computing*, Vol. 4, pp. 109-116.
- [12] Garcia Marquez, F.P and Schmid F. (2007). Digital Filter Based Approach to the Remote Condition Monitoring of Railway Turnouts. *Reliability Engineering & System Safety*, Vol. 92, pp. 830-840.
- [13] Garcia Marquez, F.P, Pedregal D.J. and Roberts C. (2010). Time Series Methods Applied to Failure Prediction and Detection. *Reliability Engineering & System Safety*. Vol. 95(6), pp. 698-703.
- [14] Garcia Marquez, F.P, Pedregal D.J. and Schmid F. (2007). Unobserved Component Models Applied To The Assessment Of Wear In Railway Points: A Case Study. *European Journal of Operational Research*, Vol. 176, pp. 1703-1702.
- [15] Harvey, A.C. (1989). Forecasting structural time series models and the Kalman filter. Cambridge: Cambridge University Press.
- [16] Kalman R.E., A new approach to linear filtering and prediction problems, *ASME Trans., Journal Basic Eng.*, 83-D (1960) 95-108.
- [17] Koopman S.J., Disturbance smoother for state-space models, *Biometrika*, 76 (1993) 65-79.
- [18] Lütkepohl H. 1991. Introduction to Multiple Time Series Analysis. Berlin, Springer-Verlag.
- [19] Pedregal D.J., Garcia Marquez, F.P and Schmid F. (2004). Predictive Maintenance of Railway Systems Based on Unobserved Components Model. *Reliability Engineering & System Safety*, Vol. 8(1), pp. 53-62.
- [20] Pedregal D.J., Garcia Marquez, F.P, Roberts C. (2009). An Algorithmic Approach for Maintenance Management". *Annals of Operations Research*. Vol. 166, pp. 109-124.
- [21] Pedregal D.J., Young P.C., Statistical approaches to modelling and forecasting time series. In Clements M., Hendry D. (eds.), *Companion to Economic Forecasting*, Blackwell Publishers, 2002.
- [22] Proctor P., Infrastructure Risk Modelling - Electric Machine Point Operating Mechanism: HW Type. EE&CS Railtrack H.Q. 2000.
- [23] Roberts, C., Dassanayake, H.P.B., Lehasab, N., Goodman, C.J. (2002). Distributed quantitative and qualitative fault diagnosis: railway junction case study. *Control Engineering Practice*, 10, 419-429.
- [24] Schweppe F., Evaluation of likelihood function for Gaussian signals, *I.E.E.E. Trans. on Inf. Theory*, 11 (1965) 61-70.
- [25] Tiao G.C., Box G.E.P., 1981, Modelling multiple time series with applications, *Journal of the American Statistical Association*, 76, 802-816.
- [26] Young P.C., Pedregal D.J., Tych W., Dynamic harmonic regression, *Journal of Forecasting*, 18, (1999) 369-394.
- [27] Young P.C., Recursive estimation and time-series analysis, Berlin: Springer-Verlag, 1984.

