Ann Oper Res (2009) 166: 109–124
DOI 10.1007/s10479-008-0403-5

An algorithmic approach for maintenance management based on advanced state space systems and harmonic regressions

Diego J. Pedregal · Fausto P. García · Clive Roberts

Published online: 10 September 2008

Abstract Point mechanisms are special track elements which failures results in delays and increased operating costs. In some cases such failures cause fatalities. A new robust algorithm for fault detection of point mechanisms is developed. It detects faults by comparing what can be considered the 'normal' or 'expected' shape of some signal with respect to the actual shape observed as new data become available. The expected shape is computed as a forecast of a combination of models. The proposed system deals with complicated features of the data in the case study, the main ones being the irregular sampling interval of the data and the time varying nature of the periodic behaviour. The system models are set up in a continuous-time framework and the system has been tested on a large dataset taken from a point mechanism operating on a commercial line.

Keywords Maintenance · Reliability · Condition monitoring · Point mechanisms · State space models

There have been many advances in rail-based transport over the past few years, such as higher speeds, increasing numbers of trains, shorter intervals between trains, and greater axle loads. All these changes result in the need for higher track quality standards and reduced maintenance time. The most important performance characteristics required are increased safety, greater reliability and reduced maintenance costs. Today, improvements are

D.J. Pedregal · F.P. García (✉)
Institute of Applied Mathematics in Science and Engineering (IMACI), E.T.S.I. Industriales, Universidad de Castilla-La Mancha, Campus universitario s/n, 13071 Ciudad Real, Spain
e-mail: FaustoPedro.Garcia@uclm.es

D.J. Pedregal
e-mail: Diego.Pedregal@uclm.es

C. Roberts
Department of Electronic, Electrical and Computer Engineering, University of Birmingham, Edgbaston, B15 2TT, Birmingham, UK
e-mail: C.Roberts.20@bham.ac.uk

often provided by the use of condition monitoring systems to aid the planning of predictive maintenance and operational policies.

Investment in these systems has mainly been justified through cost benefit analyses anticipating a significant reduction of in-service failures and an increased efficiency in maintenance and fault finding tasks. However, many of the systems deployed to date lack sufficient intelligence to provide the anticipated payback.

Points are amongst the last remaining 'fail hard' elements of the system railway. There are many situations where failure of an individual point can cause a total system shutdown with potentially very high costs, both in terms of operational difficulties and consequential costs. Another important consequence is its effect on operational schedules when track conditions are poor. Preventing such problems is therefore important for a reliable passenger service.

An artificially intelligent condition monitoring system for failure-detection in level crossings has been previously developed (Roberts et al. 2002). In that system, the hydraulic pressure, pump motor current, voltage and displacement were measured and analysed, together with signals from the level crossing relay controls. The condition monitoring described in this paper evolved from that work and is applied to Point Mechanisms in Abbotswood junction. This junction is located at the South West of Norton Junction (Norton Halt), Wadborough, in the South Western region of the UK rail network. The junction was created during the construction of the Oxford Worcester & Wolverhampton Railway line in 1850, connecting this to the Bristol and Gloucester line.

The system developed in this paper for the condition monitoring described above 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. Other ways to perform the same task may be seen in García Márquez et al. (2007a) and Pedregal (2003). The expected shape is computed as a forecast of a statistical model based only on fault-free movements of the point mechanism observed in the past. These forecasts are compared to the new incoming actual data and if the errors between the two are above an a priori defined critical level, the system issues a fault warning.

One important feature of this system is that it adapts gradually to the changes experienced in the state of the point mechanism. Forecasts are computed in such a way as to include in the estimation sample the data from the most recent point movements while discarding the older ones. In this way time varying properties of the system caused by factors such as wear are included, and hence the forecasts are adaptive.

The proposed fault detection algorithm involves estimation of various advanced models. In the first step, a model known as a Local Linear Trend, set up in a continuous-time State Space framework (SS), is used to obtain the duration of the next points movement (Harvey 1989). Continuous time is mandatory for this step since the observations on which the model is based are sampled at an irregular rate. The well-known Kalman Filter provides the required forecasts, based on the data and the State Space model (Pedregal and Young 2002). The second step involves the estimation of Harmonic Regression models (similar to a Fourier series) but with peculiarities imposed by the data, namely (i) the fundamental period changes in time (and is forecast based on the model of the first step), and (ii) the sampling interval of the data is irregular, due to the way the measurements are recorded (Young et al. 1999).

1 Monitoring system

Monitoring systems can be categorized into three distinct types: Event Loggers; condition monitoring; and fault detection and diagnosis.

Event loggers are used to record digital relay switching associated with assets such as level crossings or railway junctions. These devices do not provide on-line analysis of data, but simply provide means of recording data and inspecting events after an incident to attempt to provide diagnosis information.

Condition monitoring systems are used to collect both digital and analogue signals within a location room utilising distributed transducers connected to either point-to-point or digital bus communication links. These systems often contain some kind of alarm system based around thresholding techniques, the limits of which are set by local maintenance personnel. In order to gain a benefit from such systems, maintenance personnel have to inspect the recorded signatures daily to attempt to anticipate any failure. The systems provide no diagnosis capabilities. A basic Remote Condition Monitoring System is showed in Fig. 1.

Fault detection and diagnosis systems are a more advanced version of condition monitoring systems, incorporating 'intelligent' algorithms capable of detecting faults prior to failure, diagnosing the incipient fault and providing some indication of the criticality of the detected fault.

Over the recent years, academic research has been undertaken in order to develop Condition Monitoring algorithms. An array of techniques have been considered, varying from reasoned approaches based on the experience of maintenance operators to more sophisticated models based on artificial intelligence and signal processing techniques. The model-based techniques using analytical redundancies in the system have been presented in Frank (1990), Gertler (1988), Isermann (1984).

It has been found that one of the most important aspects of any Condition Monitoring scheme is using a combination of both qualitative and quantitative knowledge. This approach allows maintenance operators to understand what the condition monitoring equipment is doing, whilst providing numerical algorithms that a computer is able to process. The theory of observer-based fault detection and isolation techniques for linear time invariance system has been described in Patton et al. (1989), Tylee (1983). In the next section some of the methods for fault detection in railway assets are presented.

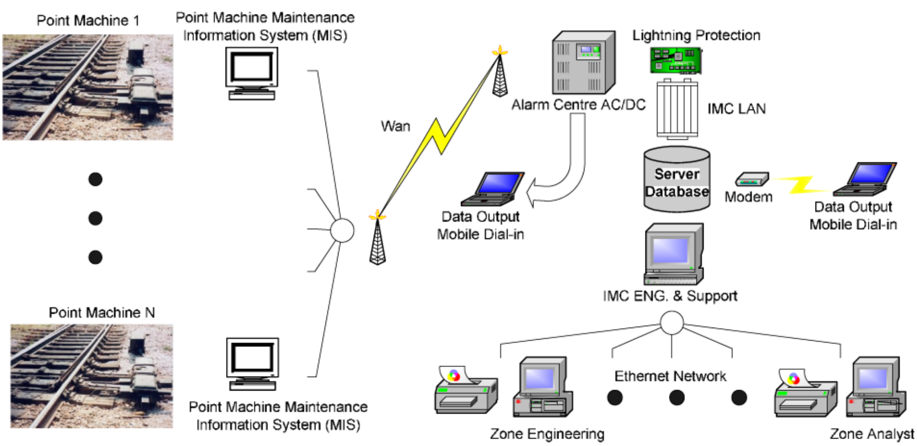


Fig. 1 A basis remote condition monitoring system

2 The state of the art in railway condition monitoring

Various methods for condition monitoring of point mechanisms have appeared in recent years. These methods can be divided into two classes according to the details of the locking mechanism used to secure the points in position following movement of the points. In the first type of locking system the force in the actuating bar is allowed to fall to zero after movement of the points. This is known as the ‘steady-state’ system. In the second type of locking system, a non-zero force is maintained in the actuating bar after point movement. This is called ‘non steady state’.

The methods employed for the steady state locking type examine signals measured during points operation. The most common method of this type is pattern recognition using the curve shape. Many examples of this type of work have been reported, for example, Shimonae et al. (1991) analysed the current signal in an electrical-mechanical point mechanism. Zhou et al. (2001, 2002a, 2002b) used 16 sensors to measure quantities such as displacement, motor torque, driving current and voltage, electrical noise, temperature and state changes and Pattern recognition of the curve shape signal with internal and external lock was developed for an electric-hydraulic point mechanism by Pabst (1998) where voltage, power and current signals were analysed. This type of point mechanism was also studied by Fry (1999) using pressure transducers to monitor the pressure characteristics and pump activity. The method used by Oyebande and Renfrew (2002) primarily analysed the power spectral density and net energy of the voltage, motor speed, switchblade positions and its movement signals. Abed et al. (1998) and Roberts and Chen (2005) investigated Single Throw Mechanical Equipment (STME). In their work the air pressure on the main air supply, linear position and current sensors are used for condition monitoring. On the other hand, Zattoni (2006) used the H_2-norm to examine the current and voltage signals of an electrical mechanical point machine. The electric-hydraulic point mechanism was considered by Rouvray et al. (1998). They studied directional change and position of the points, as well as the electro valve. Finally, a method based on artificial intelligence was studied by Roberts and Chen (2005) in which the authors developed a fuzzy algorithm for electric-pneumatic point mechanism failure monitoring. All of these methods are used offline, in others words, they cannot detect the fault in real time.

Condition monitoring methods have also been developed for electric-mechanical point machines with non steady state locking systems. McHutchon et al. (2005) used principal component analysis (PCA), statistical functions and wavelets transform. In another approach, García Márquez et al. (2003, 2007b) and García Márquez and Schmidt (2007) applied pattern recognition statistical data analysis, using a Kalman filter for parameter estimation. Finally, Pedregal et al. (2004) employed unobserved component models. All these methods can be used either in situations where the systems requires rapid and automatic adaptation as new data becomes available (on-line); or in situations where a dataset is fixed and there is no need for such rapid adaptation (off-line).

3 Case study

Following successful implementation on a level crossing mechanism (Roberts et al. 2002), the authors adapted the methods to detect faults in seven point machines at Abbotswood junction, shown in Fig. 2 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

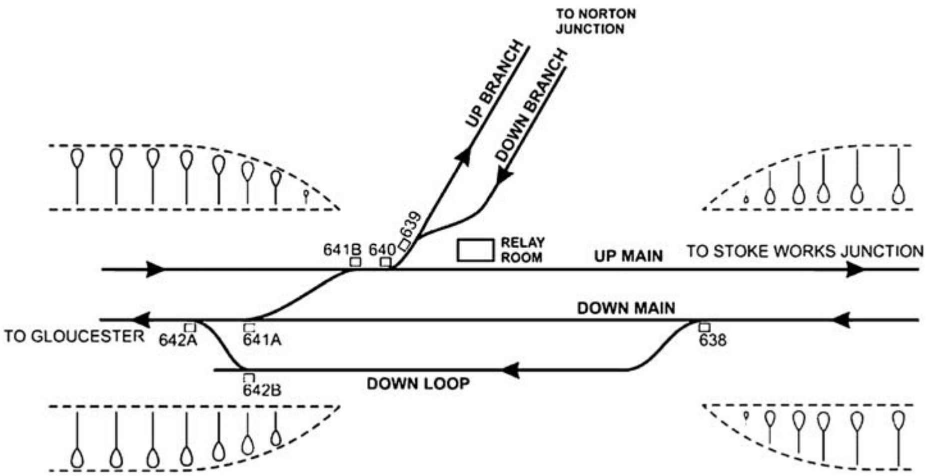


Fig. 2 Abbotswood junction

Table 1 Sensors used in points mechanisms

	Load pin	Normal pressure	Motor current	Reverse pressure	Oil level	Pump current
638	☑	☐	☑	☐	☐	☐
639	☐	☑	☐	☑	☑	☑
640	☐	☑	☐	☑	☑	☑
641B	☐	☑	☐	☑	☑	☑
641A	☐	☑	☐	☑	☑	☑
642A	☑	☐	☑	☐	☐	☐
642B	☑	☐	☑	☐	☐	☐

Controlled Solutions Ltd. The junction consists of four electro-mechanical M63 and three electro-hydraulic point machines, shown in Fig. 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), 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.

The data studied in this paper were collected between May and November 2001. Table 1 shows which signals were analysed for each point machine.

4 Experiments

Each signal is composed of long periods of inactivity interspersed with other short periods where the point mechanism is moving. Figure 3 shows a typical signal where the inactive time has been compressed to show the movements of the signal. The real periods of inactivity

Fig. 3 Signal used by the fault detection algorithm

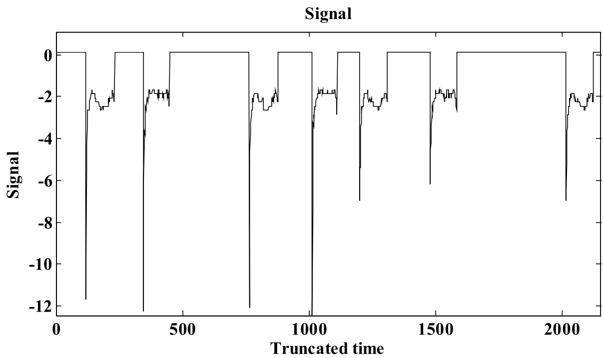
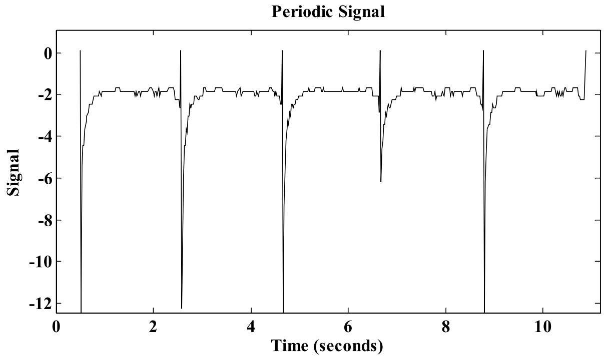


Fig. 4 Signal obtained by concatenation of the normal to reverse movements



are much longer that the figure suggests. The periods of activity alternate between normal and reverse movements.

A new signal can be composed by retaining only those times when the point mechanism is active. It can be seen in Fig. 3 that even movements (normal to reverse move) have a slightly different pattern than the odd numbered movements (reverse to normal). Therefore, two signals are formed, one by concatenating the normal to reverse movements, and the other by linking the sequential reverse to normal movements. An example of the first of these two signals is shown in Fig. 4.

The signal in Fig. 5 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 the time it takes for the point mechanism to complete a normal to reverse movement. A model of this signal must take into account two difficulties: (i) there are small variations in the time intervals at which the signals are measured; and (ii) the period of the signal varies over time. In fact, the main changes of the period may be considered as a measurement of the wear in the system, as illustrated in Fig. 5.

Figure 5 shows the period (the time to complete a normal to reverse movement) from 380 operations that make up the data set used in the case study below. The time axis is the accumulated normal to reverse operating time. The period shows evidence of a slow variation with additional random variation superimposed. A similar behaviour is observable in the reverse to normal signal. Modelling and forecasting this signal is part of the fault detection system implementation. This part of the system may be considered as an alarm system in itself, since forecast values above a critical level fixed a priori based on previous experience would be an indication of a failure, in a similar way as it is usually considered

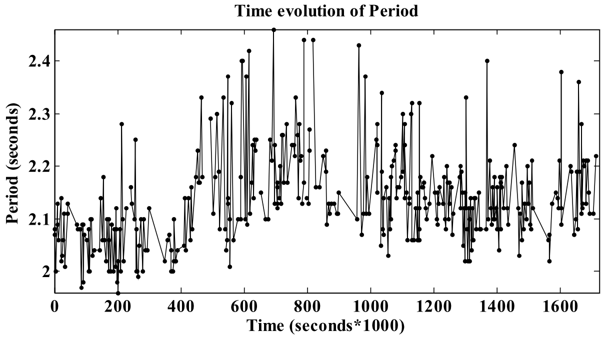


Fig. 5 Time periods in seconds of full movements

in other areas of maintenance, like vibration analysis (Christer et al. 1997; Pedregal and Carnero 2006; and references therein).

The proposed fault detection algorithm comprises the following steps:

1. Forecasting the time length or period of the next move on the basis of the signal in Fig. 5 and a local level model set up in continuous-time (described below). Confidence intervals will be considered, instead of point forecasts (see below).
2. Forecasting the signal in Fig. 4 by a Harmonic Regression model that uses the period forecast in the previous step.
3. Assessing forecasts by comparing the forecast of step 2 with the actual signal coming from the sensors installed in the point mechanism. A fault is detected when the variance of the difference between the forecast generated in step 2 and the observed signals exceeds a threshold. The threshold is chosen according to the type of point mechanism.

5 Fault detection system

The algorithm proposed here has to capture all the features seen in the data, see Figs. 3 to 5. In particular, it has to deal with a number of technical problems, like handling data sampled at irregular sampling intervals and the time varying period (or duration of the movements). In this section, the specific models used in each one of the steps of the algorithm are described in detail and the full algorithm is then stated more formally. Both models use methods that are infrequently seen in the literature of fault detection.

5.1 Forecasting the period of next movement

The model used for forecasting the period of the next movement (in a particular direction) represents the observation from Fig. 5, 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 model is given by

$$P(t) = l(t) + v(t) \tag{1}$$

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. Equation (1) represents the *observation equation* of a continuous-time SS system in which the *state equation* can be used to define the dynamics of the local level. One possible formulation of a linear time-invariant continuous-time SS system related to the application below is

$$\begin{aligned} &\text{State equation:} && \frac{d}{dt}[\mathbf{x}(t)] = \mathbf{A}\mathbf{x}(t) + \mathbf{w}(t) \\ &\text{Observation equation:} && P(t) = \mathbf{H}\mathbf{x}(t) + v(t) \end{aligned} \tag{2}$$

where $\mathbf{x}(t)$ is the state vector (of which $l(t)$ will be one state variable); $\mathbf{A}$ and $\mathbf{H}$ are matrices of appropriate dimensions; and vector $\mathbf{w}(t)$ and $v(t)$ are independent white noise sources with zero means and constant diagonal covariance matrices $\mathbf{Q}$ and R, respectively.

Many possibilities of dynamical behaviour may be used for the local mean level in (1) and (2), but the one preferred here is a *Local Linear Trend*, commonly found in the literature (Harvey 1989). The continuous-time SS representation is

$$\begin{aligned} &\frac{d}{dt}\begin{bmatrix} l(t) \\ s(t) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} l(t) \\ s(t) \end{bmatrix} + \begin{bmatrix} w_1(t) \\ w_2(t) \end{bmatrix} \\ &P(t) = l(t) + v(t) \end{aligned} \tag{3}$$

with

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

The parameters of the model are the diagonal covariance matrix $\mathbf{Q}$ and the scalar variance of the observational noise R. One way to handle system (3) is by finding a *discrete-time* SS equivalent to it (see e.g. Harvey 1989), by means of the solution to the differential equation implied by the system. A change in notation is necessary to convert system (3) to discrete-time: denote the kth observation of the series z_k (for $k = 1, 2, \ldots, N$) 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 (4).

$$\begin{aligned} &\begin{bmatrix} l_k \\ s_k \end{bmatrix} = \begin{bmatrix} 1 & \delta_k \\ 0 & 1 \end{bmatrix}\begin{bmatrix} l_{k-1} \\ s_{k-1} \end{bmatrix} + \begin{bmatrix} w_{1,k} \\ w_{2,k} \end{bmatrix} \\ &P_k = l_k + v_k \end{aligned} \tag{4}$$

In order to make systems (4) and (3) equivalent, the variances of observational noise is unchanged as R, but the covariance matrix of the process noise in the state equations becomes

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

(see Harvey 1989, p. 487). 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 (3) involved system matrices that are all constant and the state noises were all independent of each other with constant variances. Its discrete-time counterpart, system (4), involves a time variable transition matrix $\mathbf{A}_k$ and time variable variance noises that are correlated to each other according to the expression of $\mathbf{Q}_k$.

The unknown parameters q_1, q_2 and R have to be estimated. In fact, although one elegant way to choose these parameters in engineering is taking advantages of physical laws or fixed based on experimentation, some difficulties arise in the present context, where little information is available about the dynamics of the system and experimentation is rather expensive. Instead the determination of the unknowns will be based on statistical methods. The usual method to estimate variances q_1, q_2 and R is by Maximum Likelihood (ML) (Harvey 1989), though alternative frequency domain techniques are also possible (Pedregal and Young 2002; Young et al. 1999).

The final optimal estimates of the local level $l(t)$, as well as their forecasts are finally computed by well-known recursive algorithms, namely the Kalman Filter and Fixed Interval Smoother, see (Bryson and Ho 1969; Harvey 1989; Jarque and Bera 1980; Kalman 1960; Pedregal and Young 2002). In the present context, only forecasting is necessary so the Fixed Interval Smoother is not used. The particular Kalman Filter for system (4) used in this paper is presented in the appendix. One important point here is that the recursive algorithm provides also an estimation of the covariance matrix of the states and the output ($\hat{\mathbf{V}}_k$, and $\hat{f}_k$ in the appendix, respectively), thus allowing for estimating confidence intervals to the forecasts, something that will be exploited later. All the computations have been carried out in the SSPACE toolbox written for MATLAB™ (Pedregal 2003).

5.2 Forecasting the signal: harmonic regression

Once the period or the time length of the next movement of the point mechanism is forecast by the SS system (3) or (4), as described in the previous section, it is necessary to produce the forecast for the next occurrence of the signal. This is done using a Harmonic Regression model set up as described below. This model is very convenient in the present situation because it can easily handle two features of the data: the signal is sampled at irregular time intervals and the period of the waves is time varying.

The formula of a Harmonic Regression with the required properties is shown in (5).

$$z(t) = \sum_{i=1}^{M} \left[a_i \sin(\omega_{i,t} t^*) + b_i \cos(\omega_{i,t} t^*) \right] + e(t) \tag{5}$$

Here $z(t)$ is the signal in Fig. 4; M is the number of harmonics that should be included in the regression in order to achieve an adequate representation of the signal $z(t)$; a_i and b_i are $2M$ unknown parameters that modulate the deterministic sinusoidal waves; $\omega_{i,t}$ are frequencies at which the sinusoids are evaluated, with $\omega_{i,t} = 2i\pi/P(t)$ for $i = 1, 2, \ldots, M$ and $M \geq P(t)/2$; $e(t)$ is a random white noise with constant variance.

Two are the key issues in model (5) in order to be an adequate representation of $z(t)$ in Fig. 4:

1. $P(t)$ or $\omega_{i,t}$ are time varying. The nature of such variation is dependent on the signal itself. For one full movement of the point mechanism $P(t)$ is maintained constant and is equal to the time it takes to produce the full movement. But this value will be different in the next movement and should be adjusted accordingly.
2. The time index t^* is a variable linked to $P(t)$ that varies from 0 to $P(t)$ in each movement. Therefore, this variable is reset to 0 as soon as a movement finishes. This time index also takes into account the irregularities in the sampling intervals at which the signal is observed.

Figure 6 shows the key features of the data and model (5). Two full movements of the mechanism are shown $z(t)$, together with their correspondent time varying period $P(t)$ and their associated time index variable, t^*. It is clear that the full movement is longer in the second case, and all the variables are adjusted accordingly.

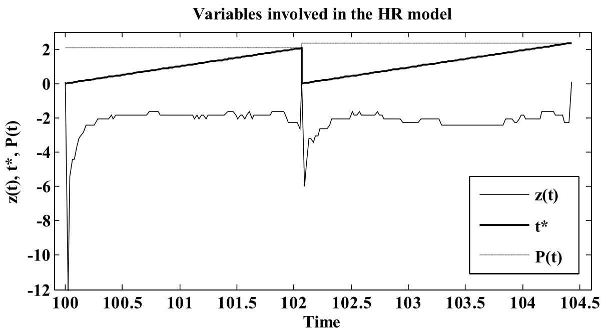


Fig. 6 Two full movements of the point mechanism, with their associated period and time index according to model (5)

Model (5) is then a regression of a signal on a set of deterministic functions of time and therefore all the standard regression theory applies without any complication, in particular estimation and forecasting with this model are made rather quickly. Model (5) could have been generalised by allowing parameters a_i and b_i vary over time, producing a much more flexible model, known as a Dynamic Harmonic Regression (DHR; see Young et al. 1999 and Pedregal and Young 2002), but this would include complications that are not necessary in the later case study.

5.3 The fault detection algorithm

The full algorithm comprises the following steps:

1. Determine historical data to use. In the later case study the last 50 free-from-faults movements of the point mechanism at each point in time are used to estimate models (3) and (5).
2. Once the forecast origin is known (there is a new movement about to be produced), a point forecast of the local level for this next movement is produced by means of model (3), together with its 95% confidence interval. In this way, a range of lengths or periods of the next movement is considered.
3. A different forecast of the signal $z(t)$ is produced for each period forecast in the previous step based on model (5). Then a full set of forecasts become available at a time sampling intervals of 5 ms, for a time horizon long enough to cover a full movement of the point mechanism.
4. As new data points of the next movement become available, they 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 the best forecast of the signal.
5. If the best forecast is systematically bad, a fault warning is issued. If the best errors are always low, no faults are detected. The boundary between 'bad' and 'good' is measured in terms of standard deviation of the errors and such a value has to be adjusted for each particular point mechanism. In the later case study the limit is 0.4 signal units.
6. If a fault is detected the historical data to perform step 1 for the next movement is the same as before. If no fault is detected, then this last movement is incorporated into the historical data next time it is going to start from step 1 and the first movement is dropped off the sample.

The algorithm can be used in on-line or off-line contexts. The use on-line would imply that step 5 is repeated every time a single data point becomes available. An off-line usage

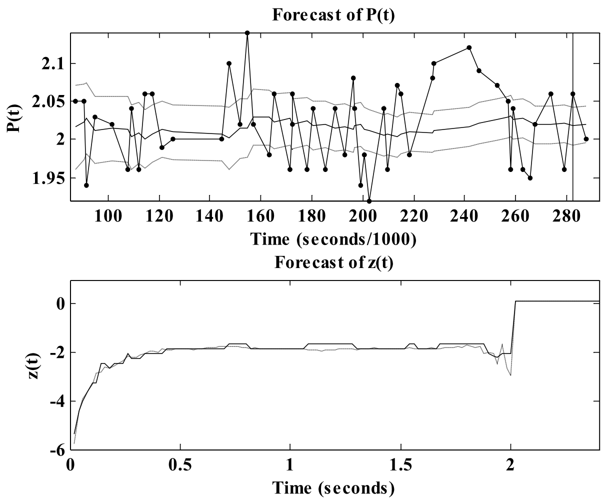


Fig. 7 *Top panel*: forecasts of model (3). *Dotted-solid line* are the actual data; *solid* is the output of model and the forecast (*last point*); *dotted lines* represent the 95% confidence interval of forecast. *Bottom panel*: actual signal $z(t)$ (*solid*) and forecast (*dotted*)

would mean that the algorithm is applied to a full movement of the mechanism, once it has been recorded.

6 Results

There are a couple of details in the system that must be fixed by experimentation, namely the alarm limit in terms of standard deviation of signal $z(t)$ and the number of harmonics to include in the Harmonic Regression (M in model (5)). Many experiments have been performed to set these design parameters of the models. The final settings are 0.4 for the standard deviation, because this is a value that produces the best discrimination between faulty and non-faulty events; and $M = 62$ harmonics for model (5) produces accurate fit and forecasts to/of the signal.

The top panel of Fig. 7 shows the output of model (3) and the forecast of the length of next mechanism movement, with the 95% confidence interval. The point forecast was 2.02 with a standard deviation of 0.012, giving a 95% confidence interval [1.996 2.044]. This interval was divided into 21 equally spaced divisions and a forecast of signal $z(t)$ was produced for each one of them. The best of all possible forecasts is the one shown at the bottom of Fig. 7 with a very small error (standard deviation is 0.096). Therefore this shape is considered as fault free.

The situation is rather different in the case of Fig. 8, where one of the faulty signals is analysed. The top panel shows the forecast of the period to include in the Harmonic Regression. The mechanism took more than 7 seconds to produce the full movement (not shown in the figure to avoid a distortion of the vertical scale), while the point forecast is 2.09 seconds and the confidence interval is [2.04 2.16]. It is obvious that in such situation the best of all the forecasts using the Harmonic Regression with any of the periods in the confidence interval is very poor, as shown in the bottom panel of Fig. 8, where it is clearly shown that, according to the historical data the mechanism should have finished much sooner. The standard deviation of the error between forecast and actual data is 1.47, much higher than in the previous case. It is worth noting that the time to produce a full movement experienced an increase of a 5% in seconds from the beginning of the sample with respect to the final

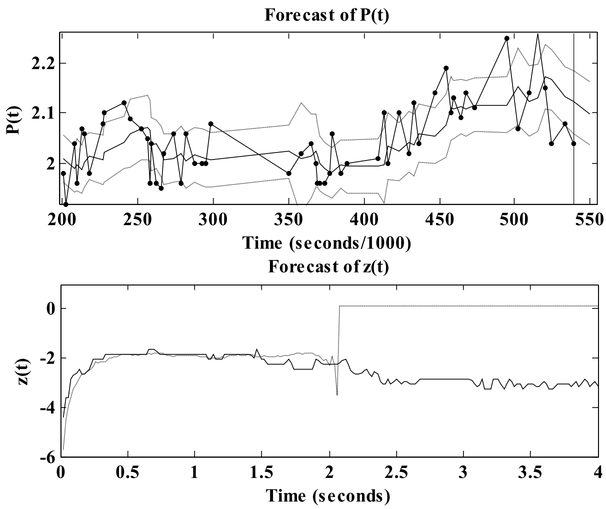


Fig. 8 *Top panel*: forecasts of model (3). *Dotted-solid* lines are the actual data; *solid* is the output of model and the forecast (*last point*); *dotted lines* represent the 95% confidence interval of forecast. *Bottom panel*: actual signal $z(t)$ (*solid*) and forecast (*dotted*)

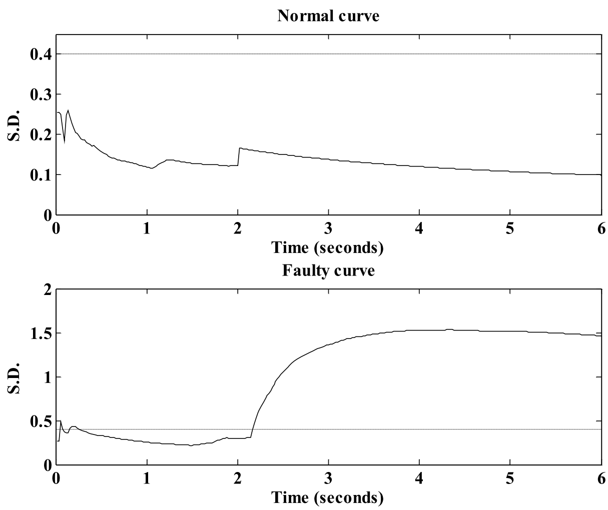


Fig. 9 The evolution of the standard deviation on which the fault assessment is based of situations in Figs. 7 and 8

observations. However all these movements where free from faults, it is simply due to wear increase.

Figure 8 highlights the on-line aspects of the algorithm. Taking the situations in Figs. 8 (free-from-fault situation) and 9 (faulty curve), the standard deviation of the errors produced with the best forecast available have been calculated as the movement is being produced. The top panel shows the situation where no faults have occurred, while the bottom panel shows evidence of a fault.

For a normal situation the standard deviation is well below the proposed alarm limit of 0.4, indicating no evidence of faults. While there is some evidence from the very beginning that something is happening in the case of the faulty situation, the situation remains normal for a while, until a sudden increase is produced in the variance, when the mechanism should have finished the work (about two seconds).

Table 2 Mean BJ and Q tests applied to all the movements in the dataset and the two statistical models involved in the algorithm (P-values in parenthesis)

	Mean of BJ	Mean of $Q(k=1)$	Mean of $Q(k=4)$	Mean of $Q(k=8)$
State space model (3)	3.57	1.08	3.91	9.24
	(0.16)	(0.29)	(0.41)	(0.32)
Harmonic regression model (5)	0.96	2.13	6.93	9.59
	(0.61)	(0.14)	(0.13)	(0.29)

This algorithm was applied to a dataset containing 380 movements in the normal to reverse direction and 380 in from reverse to normal. Eight normal to reverse movements were abnormal due to faults similar to the one shown in Fig. 8. No faults were registered in the reverse to normal direction. Selecting a standard deviation of 0.4 as the boundary of faults detection, all the faults were detected and no false alarms were produced in the test dataset.

No faults attributed to wear or other problems were detected, most likely because several preventive maintenance interventions were carried out during the normal maintenance operations on the point mechanism.

One aspect that is very important when developing statistical models is the validation tests. The most important test is by far that the model serves for the purpose it was built, as it has been demonstrated in the previous paragraphs, but further evidence from a formal point of view is usually required. In that regard, two of the most important tests are that the residuals from models (3) and (5) are both serially independent or uncorrelated and Gaussian. The classical Ljung-Box test of serial correlation (Ljung and Box 1978), based on the squares of the first k autocorrelation coefficients (r_j) on a set of T observations. It is based on the statistic

$$Q(k) = T(T+2)\sum_{j=1}^{k}\frac{r_j^2}{T-j}$$

with a chi-squared distribution with k degrees of freedom. Another standard test is the Bera-Jarque normality test (Jarque and Bera 1980), whose statistic is

$$\mathrm{BJ} = T\left(\frac{A^2}{6} + \frac{(K-3)^2}{24}\right)$$

where A and K are the skewness and kurtosis coefficients, respectively. The distribution is a Chi-squared with two degrees of freedom.

These tests were applied to the all the movements registered in the dataset and the mean results are summarised in Table 2. Regarding Gaussianity (first column of Table 1), the tests indicates very clearly that there is not any problem, since the hypothesis of normality is not rejected by a very wide margin in both cases, as indicated by the P-values (16% and 61%, respectively). With respect to serial correlation, the $Q(k)$ statistic has been presented for all the movements in both models and three values of the k parameter (1, 4 and 8). Once more, the absence of serial correlation has not been rejected by wide margins, ranging from 13% to 41%.

7 Conclusions

Following successful implementation on a level crossing mechanism in a previous research in Abbotswood junction, the authors developed a method for detecting faults in point machines.

Two hydraulic pressure transducers, namely an oil level transducer and a current transducer, were employed for the instrumentation of three electric-hydraulic point machines. This research also has studied four electro-mechanical type M63 point machines. The data were collected from a PC placed in a local relay room connected by a 1 Mb/s WorldFIP network to the instrumentation.

The fault detection system is based on the comparison of the expected signal shape with the actual shape observed from the sensors installed in the point mechanism. 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:

1. The first of the models forecast the time span a movement would take in case of absence of faults. The model is set up in continuous-time to take into account the irregular sampling interval of the data. A State Space framework is used and forecasts are computed by means of the Kalman Filter.
2. The second model is run to forecast the periodic signal itself. It is a Harmonic Regression similar to a Fourier analysis, but with advanced features included to incorporate the irregular sampling interval and a time varying period observed in the data.

A fault is detected by comparing the forecasts of the model, considered as the expected signal in case of no faults, and the actual data coming from the point mechanism when a movement is being produced. If the error is too large, as measured by its standard deviation, a fault alarm is issued. The limit at which an error is considered too big is a design parameter determined by experimentation.

The system has been tested on a point mechanism working on a commercial train line. All the faults were detected and no false alarms were produced if the design parameter (standard deviation of errors) is set correctly. The on-line capabilities of the algorithm are also shown in the paper.

The algorithm in the form presented in this paper is very general and therefore can be applied without any modification to other pieces of equipment. There is only one specific requirement, the system analysed should produce signals with a typical shape that changes stochastically from one event to the next. The extrapolation to these other systems is straightforward because the case study considered here and the models implemented to tackle the problem are very general. This means, for example, that if the data is evenly spaced in time the continuous model considered to forecast the duration of a movement is exactly equivalent to a discrete time, simpler model. Obviously, some testing experience with a new system should be undertaken in order to tune the design parameter of the algorithm, namely the alarm level at which fault warnings are issued in terms of standard deviations of the forecasting errors.

Acknowledgements The authors gratefully acknowledge Dr. Paul Weston for his careful reading of an earlier draft. We also would like to thank two anonymous referees and the editor comments' on a previous version of this paper.

Appendix: The Kalman filter

Given the discrete-time State Space in (4) in the main text, i.e.

$$\begin{bmatrix} l_k \\ s_k \end{bmatrix} = \begin{bmatrix} 1 & \delta_k \\ 0 & 1 \end{bmatrix} \begin{bmatrix} l_{k-1} \\ s_{k-1} \end{bmatrix} + \begin{bmatrix} w_{1,k} \\ w_{2,k} \end{bmatrix}$$
$$P_k = [1 \quad 0] \begin{bmatrix} l_k \\ s_k \end{bmatrix} + v_k$$
$$\mathrm{VAR}\left(\begin{bmatrix} w_{1,k} \\ w_{2,k} \end{bmatrix} \right) = \mathbf{Q}_k = \delta_k \begin{bmatrix} 1/3\delta_k^2 q_2 + q_1 & 1/2 q_2 \delta_k \\ 1/2 q_2 \delta_k & q_2 \end{bmatrix}; \qquad \mathrm{VAR}(v_k) = R$$

the recursive algorithms used for optimal state estimation and Maximum Likelihood estimation of parameters can be found in many references (Harvey 1989; Young et al. 1999; Pedregal and Young 2002). The system may be written in a compact format, i.e.

$$\mathbf{x}_k = \mathbf{A}_k \mathbf{x}_{k-1} + \mathbf{w}_k$$
$$P_k = \mathbf{H}\mathbf{x}_k + v_k$$

The Kalman Filter is an algorithm that works recursively in time direction. It computes the optimal value of the states ($\hat{\mathbf{x}}_k$) and their covariance matrices ($\hat{\mathbf{V}}_k$) based on information up to the current sample in two steps: forecasting and updating. In our particular case, the Kalman Filter is the following (Kalman 1960):

1. Prediction equations:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_k \hat{\mathbf{x}}_{k-1}$$
$$\hat{\mathbf{V}}_{k|k-1} = \mathbf{A}_k \hat{\mathbf{V}}_{k-1} \mathbf{A}_k^T + \mathbf{Q}_k$$

2. Updating equations:

$$\hat{f}_k = R + \mathbf{H}\hat{\mathbf{V}}_{k|k-1}\mathbf{H}^T$$
$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \hat{\mathbf{V}}_{k|k-1}\mathbf{H}^T(P_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1})/\hat{f}_k$$
$$\hat{\mathbf{V}}_k = \hat{\mathbf{V}}_{k|k-1} - \hat{\mathbf{V}}_{k|k-1}\mathbf{H}^T\mathbf{H}\hat{\mathbf{V}}_{k|k-1}/\hat{f}_k$$

References

Abed, S. K., Fararooy, S., & Allan, J. (1998). PC-based condition monitoring system to verify reliability-centred maintenance of electro-pneumatic point machines on London Underground. In *Developments in mass transit systems, international conference on* (*conf. publ. no*. 453) (pp. 222–227).

Bryson, A. E., & Ho, Y. C. (1969). *Applied optimal control, optimization, estimation and control*. Waltham: Blaisdell Publising Company.

CENELEC EN50170 (2002). General purpose field communication system.

Christer, A. H., Wang, W., & Sharp, J. M. (1997). A state space Condition Monitoring model for furnace erosion prediction and replacement. *European Journal of Operational Research*, *101*(1), 1–14.

Frank, P. M. (1990). Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—a survey and some new results. *Automatica*, *26*, 459–474.

Fry, K. (1999). The importance of condition monitoring in providing fault-free infrastructure. *Fault Free Infrastructure, IMechE*, 17–25.

García Márquez, F. P., & Schmidt, F. (2007). A digital filter-based approach to the remote condition monitoring of railway turnouts. *Reliability Engineering and System Safety*, *92*, 830–840.

García Márquez, F. P., Schmid, F., & Collado, J. C. (2003). A reliability centered approach to remote condition monitoring. A railway points case study. *Reliability Engineering & System Safety*, *80*(1), 33–40.

García Márquez, F. P., Pedregal, D. J., & Schmid, F. (2007a). Unobserved Components models applied to the assessment of wear in railway points: a case study. *European Journal of Operational Research*, *176*, 1703–1712.

García Márquez, F. P., Roberts, C., & Watson, P. (2007b). Failure analysis and diagnostics for railway trackside equipment. *Engineering Failure Analysis*, *14*(8), 1411–1426.

Gertler, J. (1988). Survey of model-based failure detection and isolation in complex plants. *IEEE Control System Magazine*, *8*, 3–11.

Harvey, A. C. (1989). *Forecasting structural time series models and the Kalman filter*. Cambridge: Cambridge University Press.

Isermann, R. (1984). Process fault detection based on modelling and estimation methods: a survey. *Automatica*, *20*, 387–404.

Jarque, C. M., & Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economic Letters*, *6*, 255–259.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. ASME Trans. *Journal Basic Engineering*, *83*(D), 95–108.

Ljung, G. M., & Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, *66*, 67–72.

McHutchon, M. A., Staszewski, W. J., & Schmid, F. (2005). Signal processing for remote condition monitoring of railway points. *International Journal of Strain Measurement*, *41*(2), 71–77.

Oyebande, B. O., & Renfrew, A. C. (2002). Condition monitoring of railway electric point machines. Electric Power Applications. *IEE Proceedings*, *149*(6), 465–473.

Pabst, M. (1998). Remote monitoring of points based on effective power dissipation. In *IEE seminar on condition monitoring for rail transport systems* (*ref. no*. 1998/501) 12/1–12/3.

Patton, R., Frank, P., & Clark, R. (1989). *Fault diagnosis in dynamic systems, theory and application*. New York: Prentice Hall.

Pedregal, D. J. (2003). *SSpace, a flexible Matlab toolbox for state space modelling*. Universidad Castilla-La Mancha.

Pedregal, D. J., & Carnero, M. C. (2006). State space model for condition monitoring: a case study. *Reliability Engineering and System Safety*, *91*, 171–180.

Pedregal, D. J., & Young, P. C. (2002). Statistical approaches to modelling and forecasting time series. In M. Clements & D. Hendry (Eds.), *Companion to economic forecasting*. Blackwell Publishers: Oxford.

Pedregal, D. J., García Márquez, F. P., & Schmid, F. (2004). RCM^2 predictive maintenance of railway systems based on unobserved components models. *Reliability Engineering & System Safety*, 103–110.

Roberts, C., & Chen, J. (2005). Effective condition monitoring of railway line-side assets. In *International conference on railway engineering (ICRE), maintenance and condition monitoring*, Hong Kong, A3-1.

Roberts, C., Dassanayake, H. P. B., Lehrasab, N., & Goodman, C. J. (2002). Distributed quantitative and qualitative fault diagnosis: railway junction case study. *Control Engineering Practice*, *10*, 419–429.

Rouvray, P., Hallam, P., Danaher, S., & Thorpe, M. G. (1998). The application of Matlab to railway signalling system fault modelling. In *IEE colloquium on the use of systems analysis and modelling tools: experiences and applications* (*ref. no*. 1998/413), 7/1–7/8.

Shimonae, T., Kawakami, T., Miki, H., Matsuda, O., & Tekeuchi, H. (1991). Development of a monitoring system for electric point machines. In *IRSE aspect int. conf.* (pp. 395–401).

Tylee, J. L. (1983). On-line failure detection in nuclear power plant instrumentation. *IEEE Transactions on Automatic Control*, *AC-28*(3), 406–415.

Young, P. C., Pedregal, D. J., & Tych, W. (1999). Dynamic harmonic regression. *Journal of Forecasting*, *18*, 369–394.

Zattoni, E. (2006). Detection of incipient failures by using an H_2-norm criterion: application to railway switching points. *Control Engineering Practice*, *14*(8), 885–895.

Zhou, F. B., Duta, M., Henry, M., Baker, S., & Burton, C. (2001). Condition monitoring and validation of railway point machines. In *Intelligent and self-validating instruments—sensors and actuators* (*ref. no.* 2001/179), *IEE seminar on*, 6/1-6/7.

Zhou, F. B., Duta, M. D., Henry, M. P., Baker, S., & Burton, C. (2002a). Remote condition monitoring for railway point machine. In *Railroad conference, ASME/IEEE joint*, 103–108.

Zhou, F. B., Archer, N., Bowles, J., Duta, M., Henry, M., Tombs, M., Zamora, M., Baker, S., & Burton, C. (2002b). Remote condition monitoring and validation of railway points. *Computing & Control Engineering Journal*, *13*(5), 221–230.