# Intelligent condition monitoring of railway catenary systems
## A Bayesian Network approach

Wang, Hongrui; Nunez, Alfredo; Dollevoet, Rolf; Liu, Zhigang; Chen, Junwen

**Publication date**
2017

**Document Version**
Accepted author manuscript

**Published in**
Proceedings of the 25th International Symposium on Dynamics of Vehicles on Roads and Tracks

**Citation (APA)**
Wang, H., Nunez, A., Dollevoet, R., Liu, Z., & Chen, J. (2017). Intelligent condition monitoring of railway catenary systems: A Bayesian Network approach. In M. Spiryagin, T. Gordon, C. Cole, & T. McSweeney (Eds.), *Proceedings of the 25th International Symposium on Dynamics of Vehicles on Roads and Tracks : Rockhampton, Queensland, Australia, August 14-18, 2017*

# Intelligent condition monitoring of railway catenary systems: a Bayesian network approach

Hongrui Wang, Alfredo Núñez & Rolf Dollevoet
*Section of Railway Engineering, Delft University of Technology, Delft, South Holland, The Netherlands*

Zhigang Liu & Junwen Chen
*School of Electrical Engineering, Southwest Jiaotong University, Chengdu, Sichuan, China*

ABSTRACT: This study proposes a Bayesian network (BN) dedicated for the intelligent condition monitoring of railway catenary systems. It combines five types of measurements related to catenary condition, namely the contact wire stagger, contact wire height, pantograph head displacement, pantograph head vertical acceleration and pantograph-catenary contact force, as inputs based on their physical meanings and correlations. It outputs an integrated indicator of catenary condition level. The BN parameters are learned from historical measurement data. Preliminary results shows the applicable ability of the BN to integrate multiple types of parameter while make sense of the output to facilitate maintenance decision making.

## 1 INTRODUCTION AND MOTIVATION

High-speed and upgraded conventional railway lines are expanding worldwide due to the emerging needs of rapid and sustainable public transit. The catenary system or the overhead contact line is the dominantly employed infrastructure in the traction power supply scheme for railway rolling stocks. With the growth of total mileage and operation speed, requirements on the condition monitoring of catenary system are consequently raising in both quantity and technology. Many researchers and engineers are focus on developing new techniques, such as non-contact detections based on computer vision (Karakose et al. 2017, Cho et al. 2015) and laser projection (Liu et al. 2017), and physics-based detections based on contact force (Wang et al. 2016) and pantograph acceleration (Carnevale et al. 2016) measurements, among others. By using fast and proper methodologies for feature extraction and pattern recognition, the autonomous and intelligent condition monitoring of catenary system is desired to be achieved in the future. But so far, most of these measurements are developed and employed independently for specific purposes. In certain ways, they provide performance indicators for improving the current collection quality while reducing the life cycle cost of catenary when effective maintenance is performed. Since they all serve the same purpose, various measurements and methodologies that are considered useful can be eventually integrated and constitute a diversified catenary condition monitoring system (CCMS).

Until now, the integration of the CCMS is a relatively new topic with few considerations. While in some cases multiple types of measurement are performed simultaneously in inspections, they are seldom analysed jointly. As a starting point, let us firstly recall the available measurements that are dedicated for catenary condition monitoring. For most railway infrastructure companies, the periodical measurements for catenary system are the geometrical parameters of contact wire, including the contact wire height, stagger and sometimes thickness. This is also regarded as the first level of CCMS (Carnevale et al. 2016). In this level, threshold-based criteria are usually utilized to indicate the catenary condition and identify anomalies. It is based on the common knowledge and experience on whether the geometrical condition of contact wire is acceptable for normal operation. The second level of CCMS, as also suggested in the European Standard EN50317 (BS EN 50317), requires to measure the dynamic interaction between the catenary and pantograph. Dynamic parameters such as pantograph-catenary contact force, contact point vertical displacement, pantograph vertical acceleration etc., are considered as the source of indicators that reflect the quality of interaction. Although these measurements are not yet widely employed by industry, they are by definition more directly linked with the performance of catenary system comparing with the geometrical parameters (Vo Van et al. 2016).

However, the utilizations of these measurement data treat different types of parameter as independent from each other. This makes the output performance indicators from a single type of data highly dependent on the reliability of data measurement. Thus, the output results could be too absolute and unique for engineers to make maintenance decisions and often turn out to be false alarms. In other words, the robustness of the output is low due to the lacks of data diversity and consideration of noise and other sources of stochasticity. To address this issue, the geometrical and dynamic parameters can be adopted all together as the input for CCMS. In such a case, the integration of various parameters should be carried out considering the following relationships. First, it should be based on the probabilistic relation between the data value or feature and the level of catenary condition, which is often predetermined empirically. Second, the different parameters are not independent variables anymore, but somewhat correlated with each other depending on their physical meanings. It means that, for example, the contact wire height and the pantograph-catenary contact force, not only they can both reflect the catenary condition, but also the former indicates the data reliability of the latter and thus discredit or support the output of the latter.

In this context, instead of using indicators extracted by threshold-based criteria from the measured parameters separately and directly, this paper employs the Bayesian network (BN) (Jensen 1996) to establish a physics-based connection between the parameters and assess the catenary condition. A probabilistic directed graphical model is constructed that can describe and visualize the structure of the network. Based on historical measurement data of catenary system, the physics-based probabilistic relations between the indicators extracted from different parameter are determined. The established BN can adequately utilize the various types of parameter and their correlations, as an example of CCMS integration. Meanwhile, the missing and error data problem in catenary measurements can be simultaneously addressed owing to the employment of multiple data sources.

The rest of the paper is organized as follows. Section 2 briefly introduces the BN theory and presents the proposed BN for CCMS. Section 3 gives an example of application and discusses the preliminary results. Some conclusions and further developments are drawn in Section 4.

## 2 METHODOLOGY

### 2.1 *Bayesian network*

BNs, also called directed graphical models, are a class of probabilistic graphical models. They are essentially graphs in which nodes represent random variables and the directed arcs between nodes represent conditional independence assumptions. A general probabilistic graphical model consists of $N$ variables (nodes) with a set $P$ of probability distribution functions for each variable and a set $E$ of dependencies (arcs) between variables. A BN is a directed and acyclic graph in which all the arcs must be pointing from a parent node to another node and there cannot be any closed cycle constituted by the directed arcs and nodes (Jensen 1996).

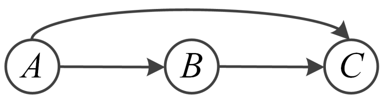


Figure 1. A BN with probabilities omitted.

A simple BN is depicted in Figure 1 when $N$=3, $E$={($A$,$B$),($A$,$C$),($B$,$C$)} and the probability functions of variables $A$, $B$ and $C$ are known. This is a BN because the arcs are directed and there are no closed cycles in the topology. As a counterexample, there will be a closed cycle if the arc pointing from variable $A$ to $C$ is reversed, resulting in a cyclic graph. The directions of arcs denote that the probability of $B$ is conditioned on $A$ and the probability of $C$ is conditioned on both $A$ and $B$, so that the joint distribution of all the variables is

$$P(A,B,C) = P(A)\cdot P(B\mid A)\cdot P(C\mid A,B) \tag{1}$$

By the defined rules, BNs with arbitrary topologies can be constructed for specific applications. In general, given variables $\mathbf{X}=X_1,\ldots,X_N$, the joint probability function for any BN is the product of the probabilities of each variable given its parents' values (if there is any parent node)

$$P(\mathbf{X}) = \prod_{i=1}^{N} P(X_i \mid parents(X_i)) \tag{2}$$

In a constructed BN, the probability of each value of a variable can be computed through exact or approximate inference when values of other variables are known based on the Bayes' theorem. However in most cases, the probability distributions of variables are unknown prior to inference. To address this issue, learning is introduced to BNs to determine the probability distributions (Heckerman 1998). The learning method employed in this paper is presented following the construction of the network topology for CCMS.

### 2.2 *Bayesian network for CCMS*

BNs are increasingly utilized for system fault diagnosis with the superiority in dealing with uncertainty problems (Cai et al. 2017). For the catenary system, there are multiple types of parameters measured for condition monitoring, while each measurement has its own uncertainty and error. These uncertainties and errors are subsequently introduced to the extracted performance indicators, which eventually influences the assessment of catenary condition. Inaccurate assessments can be misleading for decision making and thus waste many efforts and budgets. By using the BN to combine all the available measurements as inputs, an integrated performance indicator can be formed as the output. This output is theoretically more reliable than using any measurement solely, because of the inherent joint probability distribution established by arbitrary topology.

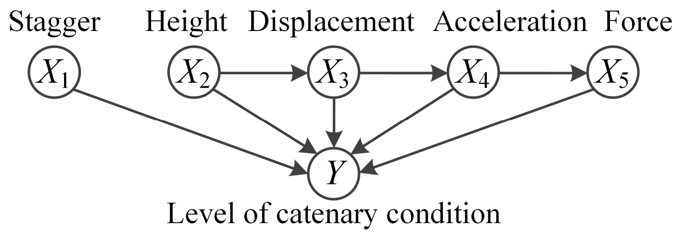


Figure 2. The BN for CCMS with probabilities omitted.

As an attempt to construct a BN for the CCMS, five types of parameter are employed in this paper, namely the contact wire stagger, contact wire height, pantograph head (pan-head) displacement, pan-head vertical acceleration and pantograph-catenary contact force. The indicators extracted from the parameters are hereafter referred as variable $X_1$ to $X_5$, respectively. According to the physical meaning of the parameters and their effects in pantograph-catenary interaction, their relationships between each other and the output $Y$, which is the level of catenary condition in this model, are constructed as depicted in the graphical model in Figure 2. The five types of parameter, denoting the inputs for catenary condition assessment, are all considered as parent nodes for the output $Y$. Among these parameters, the contact wire stagger is the only one describing the catenary geometry in the horizontal direction. It is thus regarded only as a parent node for the output but independent from the other parameters. In the vertical direction, the height, displacement, acceleration and contact force are associated according to their relevancies to the most important parameter for catenary condition monitoring, namely the pantograph-catenary contact force. Concretely, because the contact wire height is the static and initial condition for the pantograph-catenary contact, it can be regarded as the parent node of the pan-head displacement during the sliding contact. This displacement can be regarded the parent node of the pan-head acceleration, since it is directly correlated with the pan-head acceleration as a reflection of the pan-head vibration. As a main cause for the contact force fluctuation, the pan-head acceleration contributes directly to the contact force in the form of inertia force. It can thus be regarded as a parent node for the contact force. Consequently, a cascaded path that associates the parameters in the order of their relevancies with the contact force is constructed. It starts with the static height and points to the dynamic displacement, then to the pan-head acceleration, and finally to the contact force. This path is acting as a mechanism to enhance the reliability of node values, because the node values are conditioned on the corresponding and cascaded parent nodes.

To make the constructed BN inferable, the probabilities of each node should be determined. For the five types of parameter, their probability functions are unknown but can be learned from the statistics of historical measurement data. The node values are indicators extracted by threshold-based criteria dedicated for different parameters. Table 1 gives an example of threshold-based criteria that are employed for the measurement data used this paper. The indicator value from 1 to 4 describes the severity of parameter deviation comparing with the desired values that correspond to the indicator value 1. With the extracted indicators, the BN contains six variables among which five are observable from measurement data and the output variable is unobservable. In such a case with partial observability, the probability distributions of the variables can be learned based on the expectation-maximization (EM) algorithm (Lauritzen 1995). Briefly, it contains two steps. The first one is the expectation that calls an inference routine to compute the expected sufficient statistics. Then, the second step estimates the conditional probability functions similar to the case of full observability as

$$P(X_i = k \mid parents(X_i) = j) = \frac{\text{Number of samples with } parents(X_i) = j \text{ and } X_i = k}{\text{Number of samples with } parents(X_i) = j} \quad (3)$$

where $k$ and $j$ denote the possible values of $X_i$ and the parents of $X_i$, respectively. Nevertheless, the output variable must have some initial values as its boundaries before estimating the probability distribution. This can be done by considering two extreme cases in which assign $Y$=1 when all its parent nodes are 1 and assign $Y$=4 when all its parent nodes are 4, representing the best and the worst level of catenary condition, respectively.

Table 1. Threshold-based criteria for different parameters.

| Parameter | Indicator values based on thresholds | | | |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Stagger (mm) | [-300,300] | [301,310] or [-310,-301] | [311,320] or [-320,-311] | Other values |
| Height (mm) | [5800,5900] | [5770,5799] or [5901,5930] | [5740,5769] or [5931,5960] | Other values |
| Displacement (mm) | [5950,6050] | [5925,5949] or [6051,6075] | [5900,5924] or [6076,6100] | Other values |
| Acceleration (g) | [-5,5] | [5.01,7.5] or [-7.5,-5.01] | [7.51,10] or [-10,-7.51] | Other values |
| Contact force (N) | [50, 140] | [140.1,160] or [30,49.9] | [10,29.9] or [160.1,180] | Other values |

## 3 EXAMPLE OF APPLICATION

### 3.1 *Input and output*

The employed sample data are measured from a conventional railway line in China. The data are preselected to be measured under the approximately constant operation speed 100km/h, because the thresholds for dynamics parameters are sensitive to the variation of speed, whose influence is not yet considered in this study. A total number of 7323 continuous samples of the five observed variables, with the initial values of output assigned as mentioned above, are input into the BN to learn the probability distributions of the six variables based on the EM algorithm. Figure 3 depicts a part of the sample measurement data and the corresponding indicators based on the threshold-based criteria.

The constructed BN with probabilities known can be used as an integrated CCMS by inputting indicators extracted from new measurements. As a result, the probabilities of the possible values of output, namely the level of catenary condition can be given by inference. In this case, a new continuous measurement with 1000 samples of the five types of parameter are used as input to obtain the probabilities of the output, which can be used as an integrated indicator for catenary condition assessment.

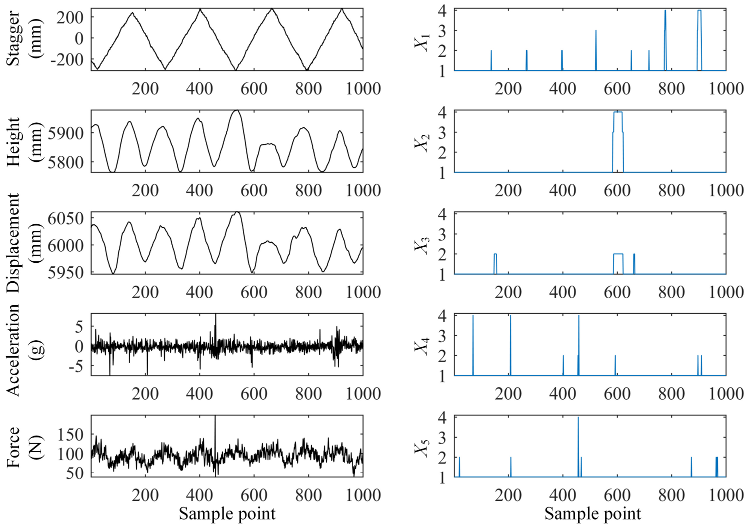


Figure 3. Partial sample data (left) and the corresponding indicators (right) used for BN learning.

## 3.2 *Result discussion*

The computation is carried out in MATLAB based on the BN toolbox developed by Murphy (Murphy 2001). The probabilities distributions of the BN are obtained based on the sample data. The distributions are the probabilities of the each value of the six variables when one or more values of the variables are known. In this study, the output is considered unobservable while the other five variables are observed as inputs. So for example, when $X_1 = X_2 = X_3 = X_4 = 1$ and $X_5 = 4$, the probabilities of that $Y$ equals to 1, 2, 3 or 4 is estimated to be 0.0069, 0.3064, 0.3382 and 0.3543 respectively. It can be seen that with only the force indicator reflects a bad condition of catenary while other indicators show the opposite, the probability of $Y$=4 is largely discredited and allocated that of $Y$ equals to 2 or 3. The probabilities in other cases are available but not shown due to the limitation of paper length.

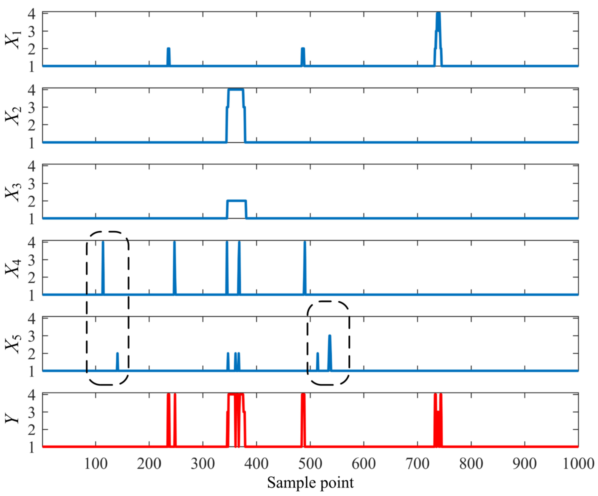


Figure 4. Indicators of new measurements ($X_1$, $X_2$, $X_3$, $X_4$ and $X_5$) and the corresponding output values with the highest probability.

The indicators of new measurements depicted in Figure 4 are input into the BN. The estimated values of output $Y$ with the highest probabilities among the four possible values are depicted at the bottom of Figure 4. From the values of output, it can be seen that there are four regions with high values indicate a bad condition of catenary. All of them are supported by at least two input variables with high values, except for the one located between 700 and 800, which is only supported by the indicator value of stagger. This is because that the stagger is in the horizontal direction and not associated with other inputs. Thus, when the stagger is abnormal, the likelihood of bad catenary condition is high, which matches experiences in practice. Meanwhile, due to the lack of support from other inputs, the abnormal values circled by the dashed lines are ignored with the output still being 1. It can be seen that the BN in a way integrates the five types of indicator while considers the significance and relevancy of them.

## 4 CONCLUSION AND OUTLOOK

This study proposes a Bayesian network to build an integrated CCMS using five types of catenary parameter. By associating the parameters based on their physical meanings and relevancies, the topology of the BN is built with an output representing the level of catenary condition. The probability distributions of the BN are learned from sample measurement data based on the EM algorithm. Preliminary results show that the model can adequately integrate the five types of parameter and output the level of catenary condition to facilitate maintenance decision making with certain robustness.

There are several aspects to improve the BN in the future. The accuracy of the BN can be improved by expanding the sample size used for learning. The phase differences of dynamic parameters may cause time differences in the data series, which can be addressed by using time translations. Also, indicators computed from threshold-based criteria can be optimized by considering the influence of speed variation during measurements.

## REFERENCES


Cai, B., Huang, L., & Xie, M. (2017). Bayesian Networks in Fault Diagnosis. *IEEE Transactions on Industrial Informatics*. In press.

Carnevale, M., & Collina, A. (2016). Processing of collector acceleration data for condition-based monitoring of overhead lines. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 230(2), 472-485.

Cho, C. J., & Ko, H. (2015). Video-based dynamic stagger measurement of railway overhead power lines using rotation-invariant feature matching. *IEEE Transactions on Intelligent Transportation Systems*, 16(3), 1294-1304.

Heckerman, D. (1998). A tutorial on learning with Bayesian networks. In *Learning in graphical models* (pp. 301-354). Springer Netherlands.

Jensen, F. V. (1996). *An introduction to Bayesian networks*. London: UCL press.

Karakose, E., Gencoglu, M. T., Karakose, M., Aydin, I., & Akin, E. (2017). A New Experimental Approach Using Image Processing-Based Tracking for an Efficient Fault Diagnosis in Pantograph–Catenary Systems. *IEEE Transactions on Industrial Informatics*, 13(2), 635-643.

Lauritzen, S. L. (1995). The EM algorithm for graphical association models with missing data. *Computational Statistics & Data Analysis*, 19(2), 191-201.

Liu, Z., Liu, W., & Han, Z. (2017). A High-Precision Detection Approach for Catenary Geometry Parameters of Electrical Railway. *IEEE Transactions on Instrumentation and Measurement*, 66(7), 1798-1808.

Murphy, K. (2001). *The bayes net toolbox for matlab*. Computing science and statistics, 33(2), 1024-1034.

Railway Applications - Current Collection Systems - Requirements for and Validation of Measurements of the Dynamic Interaction Between Pantograph and Overhead Contact Line, BS EN 50317, European Committee for Electrotechnical Standardization, 2012.

Vo Van, O., Balmes, E., Capitaine, A., & Lorang, X. (2016). Sensitivity analysis of catenary geometry on current collection quality. *The Third International Conference on Railway Technology: Research, Development and Maintenance*, Cagliari, Sardinia, Italy.

Wang, H., Liu, Z., Song, Y., Lu, X., Han, Z., Zhang, J., & Wang, Y. (2016). Detection of contact wire irregularities using a quadratic time–frequency representation of the pantograph–catenary contact force. *IEEE Transactions on Instrumentation and Measurement*, 65(6), 1385-1397.