Stochastic prediction of high-speed train dynamics to long-term evolution of track irregularities

N. Lestoille[a,b], C. Soize[b,*], C. Funfschilling[a]

[a]*SNCF, Innovation and Research Department, Immeuble Lumière, 40 avenue des Terroirs de France, 75611, Paris Cedex 12, France.*
[b]*Université Paris-Est, Modélisation et Simulation Multi-Échelle (MSME UMR 8208 CNRS), 5 Bd Descartes, 77454 Marne-la-Vallée, France.*

Abstract

There is a great interest to predict the long- term evolution of the track irregularities for a given track stretch of the high-speed train network, in order to be able to anticipate the start off of the maintenance operations. In this paper, a stochastic predictive model, based on big data made up of a lot of experimental measurements performed on the French high-speed train network, is proposed for predicting the statistical quantities of a vector-valued random indicator related to the nonlinear dynamic responses of the high-speed train excited by stochastic track irregularities. The long- term evolution of the vector-valued random indicator is modeled by a discrete non-Gaussian nonstationary stochastic model (ARMA type model), for which the coefficients are time-dependent. These coefficients are identified by a least-squares method and fitted on long time, using experimental measurements. The quality assessment of the stochastic predictive model is presented, which validates the proposed stochastic model.

Keywords: Railway track irregularities, train dynamics, stochastic prediction, nonstationary, non-Gaussian

Notations
A lower case letter, such as x or α, is a real deterministic variable.
A boldface lower case letter, such as $\mathbf{x}$ or $\boldsymbol{\alpha}$, is a real deterministic vector.
An upper case letter, such as X or A, is a real random variable.
A boldface upper case letter, $\mathbf{X}$ or $\mathbf{A}$, is a real random vector.
A letter between brackets such as $[x]$ or $[A]$, is a real deterministic matrix.
A boldface letter between brackets such as $[\mathbf{x}]$ or $[\mathbf{A}]$, is a real random matrix.
E: Mathematical expectation.
$\mathbf{X}$: Track irregularities random vector.
τ: Discrete long time, corresponding to the long-term evolution of a stretch of the track.

*Corresponding author
Email address:* `christian.soize@univ-paris-est.fr` (C. Soize)

C: Vector-valued random indicator of the dynamic response of the train.

$\mathbf{C}^{\mathrm{mod}}$: Vector-valued random indicator of the dynamic response of the train, obtained by computation of the train response submitted to the track irregularities that are measured and in taking into account the model uncertainties induced by modeling errors.

1. Introduction

Industrial context. The track geometry, which describes the relative positions of the rails, deteriorates over long time because of the train passages and of the environmental conditions. This degradation of the track geometry is characterized by four track irregularities and impacts the dynamic response of the train by reducing the passengers comfort and the train stability. To monitor the track conditions, SNCF company regularly measures the track geometry of the high-speed lines using a recording train, and perform maintenance works on the track when the track irregularities exceed a standardized threshold. The maintenance work s on railway tracks are very costly for the railway companies which are permanently seeking to improve the ir maintenance strategy. In particular, the prediction of the long- term evolution of the dynamic response of the train and of the track geometry would help railway companies to anticipate their maintenance operations and to improve their planning. Several approaches for improving the track maintenance strategy have been studied in [1, 2, 3, 4, 5]. Actually, the dynamic response of the train on the track being nonlinear, it is necessary to take into account both the track irregularities and the train response for the maintenance planning. The dynamic response of the train on a given stretch of the railway track is usually computed using a computational model, but no indicator currently exists in order to assess the long-term evolution of the dynamic response of the train. In this paper, the evolution of a given stretch of the track over a long time, denoted by τ, is characterized by the evolution of a vector-valued random indicator related to the dynamic responses of the train and denoted by $\mathbf{C}(\tau)$. Measurements of the track irregularities of this track stretch are performed for K discrete long times $\tau_1, \tau_2, \ldots, \tau_K$. The goal is to build and to identify a stochastic predictive model of the long- term evolution of the vector-valued random indicator $\mathbf{C}(\tau)$ by using the experimental data generated at discrete long times $\{\tau_1, \tau_2, \ldots, \tau_K\}$. This model should allow for predicting the statistics of the vector-valued random indicator $\mathbf{C}(\tau_{K+1})$ at long time τ_{K+1}.

Setting the problem to be solved and methodology used. The adopted methodology is summarized by the diagram displayed in Figure 1, which corresponds to the proposed stochastic predictive model for the long-term evolution of a vector-valued indicator of the dynamic response of the train, with respect to the long-term evolution of the track irregularities. The construction of the stochastic predictive model for the random indicator (framed part of the diagram) is the problem addressed in this paper. The computation of the dynamic
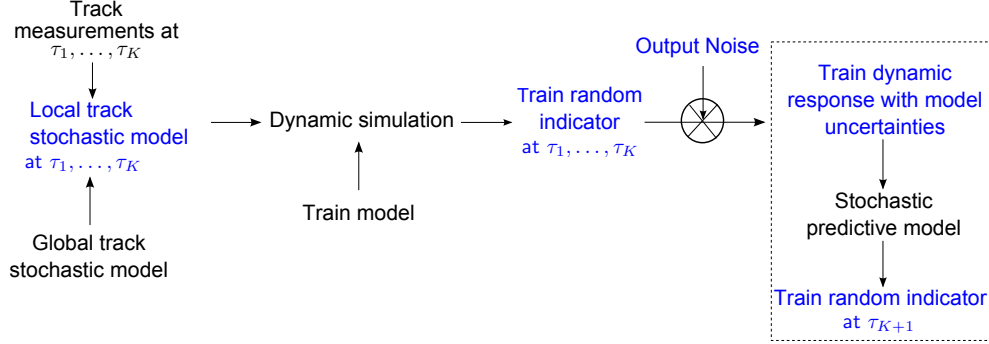
2

Figure 1: Scheme summarizing the methodology.

response of the train from the measured track irregularities is summarized in Section 2. For the construction of the stochastic predictive model, the following hypotheses are used:

- The prediction is performed for a given stretch of the railway track.

- Measurements of the track geometry of this track stretch are carried out at discrete long times $\{\tau_1, \tau_2, \ldots, \tau_K\}$.

- A stochastic computational model of the dynamic responses of the high-speed train is used to estimate the statistics of random quantities of interest, that are the vector-valued random indicator represented by the $\mathbb{R}^N$-valued time series $\mathbf{C}^{\mathrm{mod}}(\tau_1), \ldots, \mathbf{C}^{\mathrm{mod}}(\tau_K)$. The main steps for the construction of the stochastic computational model are summarized hereinafter:

  - A stochastic model of the track irregularities is constructed using experimental measurements of the track geometry as explained in [6, 7] on the base of the works published in [8, 9, 10].
  - A stochastic model of the uncertainties induced by the modeling errors introduced in the computational model of the train dynamics is constructed using simultaneously experimental measurements of the track geometry and experimental measurements of the dynamic responses of the train as explained in [11, 7].

For preserving the readability of this paper, in Section 2, we give a brief summary of these stochastic modelings, which are directly used in Section 3 for the prediction of the long- term evolution of the vector-valued random indicator.

- At each discrete long time τ_k, for $k = 1, \ldots, K$, the statistics of the vector-valued random indicator $\mathbf{C}^{\mathrm{mod}}(\tau_k)$ are calculated using the stochastic computational model of the dynamic responses of the high-speed train.

3

- Using the statistics of the vector-valued random indicators $\mathbf{C}^{\mathrm{mod}}(\tau_1)$, $\ldots, \mathbf{C}^{\mathrm{mod}}(\tau_K)$, which have been calculated for all long time s $\tau_1, \ldots, \tau_K$, a stochastic predictive model of the long- term evolution of this vector-valued indicator with respect to the track irregularities is constructed and then identified. The vector-valued random indicator, that is predicted by the stochastic predictive model, is represented by the $\mathbb{R}^N$-valued time series $\mathbf{C}^k$ (also denoted by $\mathbf{C}(\tau_k)$). If the stochastic predictive model were perfect (no error), we would have $\mathbf{C}^{\mathrm{mod}}(\tau_k) = \mathbf{C}(\tau_k)$ for $k = 1, \ldots, K$ (which is not the case).

- Using the stochastic predictive model identified with $\mathbf{C}^{\mathrm{mod}}(\tau_1), \ldots, \mathbf{C}^{\mathrm{mod}}(\tau_K)$, the statistics of the vector-valued random indicator $\mathbf{C}^{K+1}$ are calculated at the discrete long time τ_{K+1}, with $\mathbf{C}^K = \mathbf{C}^{\mathrm{mod}}(\tau_K)$ as initial condition. An estimation of the error induced by the stochastic predictive model is finally given.

Organization of the paper. In Section 2, the stochastic computational model of the dynamic response of the high-speed train is summarized. Section 3 is devoted to the construction of the stochastic predictive model. Section 4 deals with the identification procedure of the proposed stochastic predictive model. In Section 5, we present the prediction of the vector-valued indicator at long time τ_{K+1}. Finally, Section 6 is devoted to an application of the predictive model to the French railway network for the high-speed train s.

2. Stochastic computational model of the dynamic response of the train

2.1. Local stochastic model of the track irregularities

As explained in [7], a global stochastic model of the track irregularities is first constructed and then adapted to the given local stretch of the railway track. This global stochastic model has been developed in [9] and detailed in [12, 13, 14, 15]. It is independent of the track stretch and of long time τ_k, and is constructed using the measurements of the track irregularities of all the track stretches of the French high-speed lines. Due to the very large number of measured data, this model is considered as very robust.

For each stretch of the track, the track irregularities are measured by a recording train at each discrete long time τ_k for $k = 1, \ldots, K$, providing a large amount of experimental data. The measurements of the four track irregularities are carried out at $N_s + 1$ sampling points along each stretch of the track. The track irregularities vector is then modeled by a centered random variable $\mathbf{X} = (\mathbf{X}^1, \mathbf{X}^2, \mathbf{X}^3, \mathbf{X}^4)$ with values in $\mathbb{R}^{4(N_s+1)}$. For the global stochastic modeling, the centered random vector $\mathbf{X}$ is written (using a truncated principal component s decomposition) as

$$\mathbf{X} \simeq \sum_{\ell=1}^{N_\eta} \sqrt{\lambda_\ell}\, \mathbf{u}^\ell\, \eta_\ell \,, \tag{1}$$

in which $N_\eta \ll 4(N_s + 1)$, where $\lambda_1 \geq \ldots \geq \lambda_{N_\eta}$ are the largest eigenvalues, and where $\mathbf{u}^1, \ldots, \mathbf{u}^{N_\eta}$ are the associated eigenvectors of the covariance matrix $[C_{\mathbf{XX}}]$ of $\mathbf{X}$, which is estimated using the experimental measurements of the track irregularities at a fixed long time τ_k. The random coefficients $\eta_1, \ldots, \eta_{N_\eta}$ are uncorrelated, centered and normalized non-Gaussian and statistically dependent random variables such that $\eta_\ell = \lambda_\ell^{-1/2} \mathbf{X}^T \mathbf{u}^\ell$. Let $[\lambda]$ be the diagonal matrix whose diagonal entries are $\lambda_1, \ldots, \lambda_{N_\eta}$. Let $[Q]$ be the rectangular real matrix of dimension $(4(N_s + 1), N_\eta)$ such that

$$[Q] = [\mathbf{u}^1 \, \lambda_1^{1/2} \ldots \mathbf{u}^{N_\eta} \, \lambda_{N_\eta}^{1/2}] \quad , \quad [Q]^T \, [Q] = [\lambda] \,. \tag{2}$$

Then random vector of track irregularities $\mathbf{X}$ can be rewritten as

$$\mathbf{X} \simeq [Q] \, \boldsymbol{\eta} \,. \tag{3}$$

The global stochastic model has to be adapted to the given local track stretch, in constructing a local stochastic model of the track irregularities for the given track stretch. This local stochastic model takes into account uncertainties induced by (i) the measurement noise associated with local measurements $\mathbf{x}_{\tau_1}^{\mathrm{meas}}$, $\mathbf{x}_{\tau_2}^{\mathrm{meas}}, \ldots$ of the track irregularities of the given track stretch, and (ii) the local variability of the given track stretch in order to decrease the "statistical distance" between the global stochastic model and the local measurements. The local stochastic model is constructed in introducing a random field noise for which the spatial properties are driven by the global stochastic model and whose intensity of the statistical fluctuations is identified using measurement $\mathbf{x}_{\tau_1}^{\mathrm{meas}}$ at long time τ_1. It is then assumed that the optimal value computed for the random noise intensity at long time τ_1 is representative of the level of uncertainties (measurement noise and variability) for all the values of the long time τ_k of this given track stretch. For $\kappa = 1, \ldots, 4$, let $\mathbf{X}^\kappa$ be the random vector of dimension $N_s + 1$, which represents one type of track irregularities and which is written (using Eq. (3)) as

$$\mathbf{X}^\kappa = [Q^\kappa] \, \boldsymbol{\eta} \,, \tag{4}$$

in which the $((N_s + 1) \times N_\eta)$ real matrix $[Q^\kappa]$ is extracted from matrix $[Q]$. The proposed adapted stochastic model is written as

$$\widetilde{\mathbf{X}}^\kappa(\delta_{\tau_1}^\kappa) = [Q^\kappa] \left(\boldsymbol{\eta} + \delta_{\tau_1}^\kappa \, \mathbf{G}^\kappa \right) \quad , \quad \kappa = 1, 2, 3, 4 \,, \tag{5}$$

in which $\boldsymbol{\delta}_{\tau_1} = (\delta_{\tau_1}^1, \delta_{\tau_1}^2, \delta_{\tau_1}^3, \delta_{\tau_1}^4)$ is the vector-valued hyperparameter allowing the uncertainty level to be controlled, and which has to be identified for each stretch of the track using experimental data $\mathbf{x}_{\tau_1}^{\mathrm{meas}}$ at $\tau = \tau_1$. For fixed κ, $\mathbf{G}^\kappa$ is a $\mathbb{R}^{N_\eta}$-valued Gaussian second-order centered random variable, defined on the probability space $(\Theta, \mathcal{F}, \mathcal{P})$, for which the covariance matrix is the identity matrix. In the following, we will also use the notation $\mathbf{G} = (\mathbf{G}^1, \mathbf{G}^2, \mathbf{G}^3, \mathbf{G}^4)$ that is a Gaussian second-order centered random vector defined on $(\Theta, \mathcal{F}, \mathcal{P})$ with values in $\mathbb{R}^{4N_\eta}$. The optimal value $\boldsymbol{\delta}_{\tau_1}^{\mathrm{opt}} = (\delta_{\tau_1}^{1,\mathrm{opt}}, \delta_{\tau_1}^{2,\mathrm{opt}}, \delta_{\tau_1}^{3,\mathrm{opt}}, \delta_{\tau_1}^{4,\mathrm{opt}})$ of hyperparameter $\boldsymbol{\delta}_{\tau_1}$ is estimated by using the maximum log-likelihood method and the experimental measurements, as explained in [7].

5

2.2. Stochastic model of uncertainties induced by modeling errors in the computational model of the train dynamics

2.2.1. Definition of the vector-valued indicator of the high-speed train dynamics

A vector-valued indicator is defined to characterize the dynamic response of the high-speed train on a given track stretch of length S. This indicator is inspired by criteria (accelerations and forces) related to the dynamic response of the high-speed train that are described in UIC 518 leaflet [16] for the certification of railway vehicles. Experimental measurements of the components of the indicator have been performed and can be used to validate the simulation results. To compute the N components of the indicator, N quantities of interest, denoted by $y_1(s), \ldots, y_N(s)$, s in $[0, S]$, are calculated from forces and accelerations at different locations inside the train. For the given track stretch, the vector-valued indicator $\mathbf{c} = (c_1, \ldots, c_N)$ is then defined by

$$c_j = \max_{s \in [0,S]} |y_j(s)| \quad , \quad j = 1, \ldots, N. \tag{6}$$

For a given stretch of the track, the track geometry and the indicator of the train dynamics are simultaneously measured. For the ν_p track stretches for which measurements are performed, the measured indicators of the train dynamics are denoted by $\mathbf{c}^{\exp,1}, \ldots, \mathbf{c}^{\exp,\nu_p}$. Using the computational model of the train dynamics and the measured track irregularities of each measured track stretch, ν_p simulated indicators of the train dynamics, denoted by $\mathbf{c}^{\mathrm{sim},1}, \ldots, \mathbf{c}^{\mathrm{sim},\nu_p}$, are computed. The ν_p vector-valued indicators $\mathbf{c}^{\mathrm{sim},1}, \ldots, \mathbf{c}^{\mathrm{sim},\nu_p}$ are considered as ν_p independent realizations of the vector-valued random indicator denoted by $\mathbf{C}^{\mathrm{sim}}$.

2.2.2. Stochastic modeling of the random indicator taking into account model uncertainties induced by modeling errors

Let $\mathbf{C}^{\mathrm{mod}} = (C_1^{\mathrm{mod}}, \ldots, C_N^{\mathrm{mod}})$ be the random indicator deduced from $\mathbf{C}^{\mathrm{sim}}$ for which the model uncertainties induced by modeling errors (see [17]) are taken into account. The stochastic model of $\mathbf{C}^{\mathrm{mod}}$ is obtained by introducing a non-Gaussian second-order multiplicative output noise $\mathbf{B}^{\mathrm{out}} = (B_1^{\mathrm{out}}, \ldots, B_N^{\mathrm{out}})$ defined on the probability space $(\Theta', \mathcal{F}', \mathcal{P}')$ with values in $\mathbb{R}^N$, which is statistically independent of Gaussian random vector $\mathbf{G}$, and which is written as

$$C_j^{\mathrm{mod}} = C_j^{\mathrm{sim}} \exp(B_j^{\mathrm{out}}) \, , \; j = 1, \ldots, N. \tag{7}$$

Non-Gaussian second-order random vector $\mathbf{B}^{\mathrm{out}}$ is identified as explained in [7] in introducing its polynomial chaos representation (see for instance [18, 19, 17]) for which the coefficients are identified by using the maximum-likelihood method and the experimental measurements $\mathbf{c}^{\exp,1}, \ldots, \mathbf{c}^{\exp,\nu_p}$ of the indicator.

Consequently, the vector-valued random indicator $\mathbf{C}^{\mathrm{mod}}$ is defined on the product of the probability spaces $(\Theta, \mathcal{F}, \mathcal{P})$ and $(\Theta', \mathcal{F}', \mathcal{P}')$. Let $\mathbf{G}(\theta_1), \ldots, \mathbf{G}(\theta_\nu)$ be ν independent realizations of $\mathbf{G}$ and let $\mathbf{B}^{\mathrm{out}}(\theta_1'), \ldots, \mathbf{B}^{\mathrm{out}}(\theta_\nu')$ be ν independent realizations of output noise $\mathbf{B}^{\mathrm{out}}$. The family of the ν independent realizations $\{\mathbf{C}^{\mathrm{mod}}(\theta_\ell, \theta_\ell'), \ell = 1, \ldots, \nu\}$ is such that, for all $j = 1, \ldots, N$,

$$C_j^{\mathrm{mod}}(\theta_\ell, \theta_\ell') = C_j^{\mathrm{sim}}(\theta_\ell) \exp(B_j^{\mathrm{out}}(\theta_\ell')) \, , \; j = 1, \ldots, N. \tag{8}$$

3. Stochastic modeling of the long- term evolution for the vector-valued random indicator

The long-term evolution of the vector-valued random indicators $\mathbf{C}^{\mathrm{mod}}(\tau_1), \ldots,$ $\mathbf{C}^{\mathrm{mod}}(\tau_K)$, which have been computed in Section 2.2.2, is modeled by a time series denoted by $\{C(\tau_k), k \geq 1\}$ (or $\{C^k, k \geq 1\}$) which has the same properties as $\mathbf{C}^{\mathrm{mod}}(\tau_k)$. The construction and the identification of the stochastic predictive model of the vector-valued random indicator $\{C^k, k \geq 1\}$ are relatively difficult for the following reasons:

- The long-term evolution of the indicator is strongly nonstationary (and consequently statistics cannot be enriched by using time averaging estimators).

- The experimental data are very limited because only one measurement of the track geometry of the given track stretch is available at each discrete long time τ_k.

- The value K of the number of discrete long times used for predicting the statistics of the vector-valued random indicator $\mathbf{C}^{K+1}$ at discrete long time τ_{K+1} is very low (typically K is of order 10).

- The vector-valued random indicator $\mathbf{C}$ is a non-Gaussian random vector, in particular the initial value $\mathbf{C}^1 = \mathbf{C}^{\mathrm{mod}}(\tau_1)$ is non-Gaussian, which means that the time series $\{C^k, k \geq 1\}$ is non-Gaussian.

Two steps are required for predicting the statistics of the vector-valued random indicator $\mathbf{C}$ at long time τ_{K+1}. The first one is related to the choice and the construction of a stochastic predictive model and the second one consists in identifying it by solving a statistical inverse problem. The time series $\mathbf{C}(\tau_1), \ldots, \mathbf{C}(\tau_K)$, for which the discrete time-evolution stochastic model has to be constructed using time series data $\mathbf{C}^{\mathrm{mod}}(\tau_1), \ldots, \mathbf{C}^{\mathrm{mod}}(\tau_K)$, is nonstationary, and consequently the identification requires to solve a nonstationary statistical inverse problem.

There exist several methods for constructing and identifying a nonstationary stochastic predictive model:

- The first one belongs to the class of the ARMA models [20, 21]. Since the vector-valued random indicator is nonstationary and non-Gaussian, a nonstationary ARMA model has to be used (see for instance, [22, 23, 24]), and the non-Gaussian character of the output is induced by the non-Gaussian random initial condition $\mathbf{C}^1 = \mathbf{C}^{\mathrm{mod}}(\tau_1)$.

- Since the probability distribution of the non-Gaussian random initial value $\mathbf{C}(\tau_1) = \mathbf{C}^{\mathrm{mod}}(\tau_1)$ is known (estimated), the second method would consist in using the Bayesian filtering [25, 26] based on one-order or higher-order Markov chain models. A first choice could consist in choosing a one-order Markov chain model and to directly identify the nonhomogeneous transition kernel. Such an approach has been tested and has not given

7

sufficiently robust predictions. A second one would be to use a Kalman filter or an extended Kalman filter [25, 27, 28] taking into account the non-Gaussian property of the vector-valued random indicator. Such an approach cannot directly be used due to the lack of data.

The long-term evolution model $\mathbf{C}^k = \mathbf{C}(\tau_k)$, for $k = 2, \dots, K$, with the initial condition $\mathbf{C}^1 = \mathbf{C}^{\mathrm{mod}}(\tau_1)$, of the vector-valued random indicator $\mathbf{C}$, is thus chosen as a nonstationary non-Gaussian one-order Markov chain. The long-term evolution model is written as

$$\mathbf{C}^k = ([I_N] - \Delta\tau_k \, [A])\mathbf{C}^{k-1} + \Delta\tau_k \, \mathbf{g}^k + [h^k] \, \Delta\mathbf{W}^k \quad , \quad k = 2, \dots, K \,, \quad (9)$$

with the initial condition

$$\mathbf{C}^1 = \mathbf{C}^{\mathrm{mod}}(\tau_1) \,, \quad (10)$$

in which

- $[I_N]$ is the identity matrix in $\mathbb{R}^{N \times N}$.

- $[A]$ is a matrix in $\mathbb{R}^{N \times N}$, which has to be identified.

- $\Delta\tau_k = \tau_k - \tau_{k-1}$ represents the time step, that is not constant and depends on k, and in which τ_k, $k = 2, \dots, K$, is the measurement time of the track irregularities.

- $\Delta\mathbf{W}^k = \sqrt{\Delta\tau_k} \, \boldsymbol{\mathcal{N}}^k$, in which $\boldsymbol{\mathcal{N}}^2, \dots, \boldsymbol{\mathcal{N}}^K$ are independent Gaussian normalized random vectors defined on a third probability space $(\Theta'', \mathcal{F}'', \mathcal{P}'')$, with values in $\mathbb{R}^N$ ($E\{\boldsymbol{\mathcal{N}}^k\} = 0$, $E\{\boldsymbol{\mathcal{N}}^k (\boldsymbol{\mathcal{N}}^k)^T\} = [I_N]$). Therefore, $E\{\Delta\mathbf{W}^k \otimes \Delta\mathbf{W}^k\} = \Delta\tau_k \, [I_N]$. The family of random vectors $\{\boldsymbol{\mathcal{N}}^k, k \geq 2\}$ is statistically independent of random vectors $\mathbf{G}$ and $\mathbf{B}^{\mathrm{out}}$, and consequently, is independent of $\mathbf{C}^{\mathrm{mod}}(\tau_1)$.

- $\{\mathbf{g}\} = \{\mathbf{g}^2, \dots, \mathbf{g}^K\}$ is a family of $K - 1$ vectors in $\mathbb{R}^N$, which has to be identified.

- $\{[h]\} = \{[h^2], \dots, [h^K]\}$ is a family of $(K - 1)$ real matrices in $\mathbb{R}^{N \times N}$, which has to be identified. For all k, $[h^k]$ is chosen as a lower triangular matrix with positive diagonal entries.

- in the initial condition, $\mathbf{C}^{\mathrm{mod}}(\tau_1)$ is a non-Gaussian second-order $\mathbb{R}^N$-valued random variable whose probability distribution is known (estimated using the stochastic computational model of the high-speed train dynamics described in Section 2).

With respect to an ARMA model or to a Kalman filter, it should be noted that:

- The nonstationarity property is induced by the coefficients $\mathbf{g}^k$ and $[h^k]$ that depend on discrete time τ_k, represented by index k.

- The matrix $[A]$ does not depend on discrete time k. It should be noted that, if matrix $[A]$ had been chosen as a function of k, the time series $[A^k]$, $\mathbf{g}^k$, and $[h^k]$ could not be identified because of the lack of data.

8

- If the initial condition $\mathbf{C}^1 = \mathbf{C}^{\mathrm{mod}}(\tau_1)$ were a deterministic vector or a Gaussian random vector, then the nonstationary time series $\{C^k, k \geq 1\}$ would be Gaussian. Nevertheless, since random vector $\mathbf{C}^{\mathrm{mod}}(\tau_1)$ is not Gaussian, time series $\{C^k, k \geq 1\}$ generated by the predictive model above is a non-Gaussian nonstationary time series.

With such a proposed stochastic predictive model, $[A]$, $\mathbf{g}^k$, and $[h^k]$ have to be identified using data $\mathbf{C}^{\mathrm{mod}}(\tau_1), \ldots, \mathbf{C}^{\mathrm{mod}}(\tau_K)$, for which the joint probability distribution is known (by estimation). Such an identification can be done using the first- and second-order moments equations associated with Eq. (9).

Once the stochastic predictive model is identified, it can be used for predicting the statistics of the random vector $\mathbf{C}^{K+1}$ at long time τ_{K+1}. However, the values of $\mathbf{g}^{K+1}$ and $[h^{K+1}]$, which are needed for the estimation of $\mathbf{C}^{K+1}$, are unknown. We then propose to represent $\mathbf{g}^k$ and $[h^k]$ as the values $\mathbf{g}^{\mathrm{aff}}(\tau_k)$ and $[h^{\mathrm{aff}}(\tau_k)]$ at τ_k, $k \geq 1$, of affine functions $\mathbf{g}^{\mathrm{aff}}$ and $[h^{\mathrm{aff}}]$ (a more higher-degree representations for $\mathbf{g}$ and $[h]$ could be introduced, but it has been seen that no gain is obtained with respect to the one-degree (affine) representation). For the identification of the parameters of such affine representations $\mathbf{g}^{\mathrm{aff}}$ and $[h^{\mathrm{aff}}]$, two approaches can *a priori* be used. The first one consists in introducing the affine representations in Eq. (9) before the identification (as proposed, for instance, in [22]), and then to identify $[A]$ and the parameters involved in the affine representations $\mathbf{g}^{\mathrm{aff}}$ and $[h^{\mathrm{aff}}]$ of $\mathbf{g}^k$ and $[h^k]$. The second one consists in identifying the coefficients $\mathbf{g}^k$ and $[h^k]$ (as explained before) and then approximating them by the values $\mathbf{g}^{\mathrm{aff}}(\tau_k)$ and $[h^{\mathrm{aff}}(\tau_k)]$. The identification of the parameters of affine functions $\mathbf{g}^{\mathrm{aff}}$ and $[h^{\mathrm{aff}}]$ is performed using the least-squares method with constraints. These two approaches have been tried and their results show that the second one is more robust than the first one. Using the second approach, the prediction $\mathbf{C}^{K+1}$ will be denoted by $\mathbf{C}^{\mathrm{aff},K+1}$.

4. Identification of the parameters of the stochastic model for the long- term evolution

For the identification of the model parameters $[A]$, $\{\mathbf{g}\}$ and $\{[h]\}$, the classical least-squares method with weights is used. The cost function of the least-squares method is constructed using the first-order vector moment equation, the second-order tensor moment equation, and the second-order cross-tensor moment equations associated with Eq. (9).

4.1. First-order vector moment equation with weights used for constructing the cost function

Taking the mathematical expectation of Eq. (9) yields the following $K-1$ first-order vector moment equations,

$$\mathbf{f}^k([A], \{\mathbf{g}\}) = \mathbf{0}_N \quad , \quad k = 2, \ldots, K \,, \tag{11}$$

in which, for all k fixed in $\{2, \ldots, K\}$, the $\mathbb{R}^N$-valued function $\mathbf{f}^k$ is defined by

$$\mathbf{f}^k([A], \{\mathbf{g}\}) = [\mathbb{N}_f^k] \left(E\{\mathbf{C}^k\} - ([I_N] - \Delta\tau_k[A]) \, E\{\mathbf{C}^{k-1}\} - \Delta\tau_k \, \mathbf{g}^k \right), \quad (12)$$

in which, for all i and j in $\{1, \ldots, N\}$, the weights $[\mathbb{N}_f^k]_{ij}$ are introduced for normalization reasons inside the cost function and for being in capability to give a more important weight α_j to certain components C_j^k of the vector-valued indicator $\mathbf{C}^k$ in the cost function. The matrix $[\mathbb{N}_f^k]$ of these weights is thus defined by

$$[\mathbb{N}_f^k]_{ij} = \frac{\alpha_i}{E\{C_i^k\}} \delta_{ij}, \quad (13)$$

in which δ_{ij} is the Kronecker symbol, where $C_1^k, \ldots, C_N^k$ are the components of the random vector $\mathbf{C}^k$, and where $\{\alpha_1, \ldots, \alpha_N\}$ is the family of weight co-efficients that have to be arbitrarily chosen (see Section 6.2) and belong to the admissible set $\mathcal{C}_1$ defined by

$$\mathcal{C}_1 = \left\{ \alpha_1 \geq 0, \ldots, \alpha_N \geq 0, \, \sum_{j=1}^{N} \alpha_j = 1 \right\}. \quad (14)$$

4.2. Second-order tensor moment equation with weights used for constructing the cost function

For each $k = 2, \ldots, K$, Eq. (9) is right-tensorized by $\mathbf{C}^k$ and the mathematical expectation is applied, which gives the following second-order tensor moment equation,

$$E\{\mathbf{C}^k \otimes \mathbf{C}^k\} = ([I_N] - \Delta\tau_k[A]) \, E\left\{\mathbf{C}^{k-1} \otimes \mathbf{C}^k\right\} + \Delta\tau_k \, \mathbf{g}^k \otimes E\{\mathbf{C}^k\}$$
$$+ [h^k] \, E\{\Delta\mathbf{W}^k \otimes \mathbf{C}^k\}. \quad (15)$$

As $\Delta\mathbf{W}^k$ and $\mathbf{C}^{k-1}$ are statistically independent, and as $\Delta\mathbf{W}^k$ is centered, the term $E\{\Delta\mathbf{W}^k \otimes \mathbf{C}^k\}$ can be expressed as

$$\begin{aligned} E\{\Delta\mathbf{W}^k \otimes \mathbf{C}^k\} &= E\{\Delta\mathbf{W}^k \otimes (([I_N] - \Delta\tau_k \, [A])\mathbf{C}^{k-1} + \Delta\tau_k \, \mathbf{g}^k \\ &\quad + [h^k] \, \Delta\mathbf{W}^k)\} \\ &= E\{\Delta\mathbf{W}^k \otimes \Delta\mathbf{W}^k\} \, [h^k]^T \\ &= [h^k]^T \Delta\tau_k. \end{aligned} \quad (16)$$

For all k in $\{2, \ldots, K\}$, Eq. (15) can be rewritten as

$$E\{\mathbf{C}^k \otimes \mathbf{C}^k\} = ([I_N] - \Delta\tau_k[A]) \, E\{\mathbf{C}^{k-1} \otimes \mathbf{C}^k\} + \Delta\tau_k \, \mathbf{g}^k \otimes E\{\mathbf{C}^k\}$$
$$+ [h^k] \, [h^k]^T \Delta\tau_k. \quad (17)$$

From Eq. (17), it can then be deduced that, for all k in $\{2, \ldots, K\}$, the second-order tensor moment equation with weights is written as

$$[F^k([A], \{\mathbf{g}\}, \{[h]\})] = [0_{N,N}], \quad (18)$$

in which the function $[F^k]$ with values in $\mathbb{R}^{N \times N}$ is defined by

$$
\begin{aligned}
[F^k([A], \{\mathbf{g}\}, \{[h]\})] = \\
\mathbb{T}^k : \big(E\{\mathbf{C}^k \otimes \mathbf{C}^k\} - ([I_N] - \Delta\tau_k[A]) \, E\{\mathbf{C}^{k-1} \otimes \mathbf{C}^k\} \\
- \Delta\tau_k \, \mathbf{g}^k \otimes E\{\mathbf{C}^k\} - [h^k] \, [h^k]^T \Delta\tau_k \big) \,, \quad (19)
\end{aligned}
$$

in which the symbol " $:$ " denotes the double contraction tensor operation. For all i, j, i', j' in $\{1, \ldots, N\}$, the weights $\{\mathbb{T}^k\}_{iji'j'}$ are introduced for the same reasons as those given for Eq. (13). The fourth-order tensor $\mathbb{T}^k$ of these weights is thus defined by

$$
\{\mathbb{T}^k\}_{iji'j'} = \frac{\sqrt{\alpha_i \alpha_j}}{E\{C_i^k C_j^k\}} \delta_{ii'} \delta_{jj'} \,, \quad (20)
$$

in which $\{\alpha_1, \ldots, \alpha_N\}$ are the weights defined in Section 4.1 and where $C_1^k, \ldots, C_N^k$ are the components of the random vector $\mathbf{C}^k$.

4.3. Second-order cross-tensor moment equation with weights used for constructing the cost function

For each $k = 2, \ldots, K$ and for $\mu = 1, \ldots, k-1$, Eq. (9) is right-tensorized by $\mathbf{C}^{k-\mu}$ and taking the mathematical expectation yield

$$
\begin{aligned}
E\{\mathbf{C}^k \otimes \mathbf{C}^{k-\mu}\} = \\
([I_N] - \Delta\tau_k[A]) \, E\{\mathbf{C}^{k-1} \otimes \mathbf{C}^{k-\mu}\} + \Delta\tau_k \, \mathbf{g}^k \otimes E\{\mathbf{C}^{k-\mu}\} \\
+ [h^k] \, E\{\Delta\mathbf{W}^k \otimes \mathbf{C}^{k-\mu}\} \,. \quad (21)
\end{aligned}
$$

As $\Delta\mathbf{W}^k$ and $\mathbf{C}^{k-\mu}$ are statistically independent for all k in $\{2, \ldots, K\}$ and μ in $\{1, \ldots, k-1\}$, and as $\Delta\mathbf{W}^k$ is centered, it can be deduced that

$$
E\{\Delta\mathbf{W}^k \otimes \mathbf{C}^{k-\mu}\} = 0 \,, \quad (22)
$$

$$
E\{\mathbf{C}^k \otimes \mathbf{C}^{k-\mu}\} = ([I_N] - \Delta\tau_k[A]) \, E\{\mathbf{C}^{k-1} \otimes \mathbf{C}^{k-\mu}\} + \Delta\tau_k \, \mathbf{g}^k \otimes E\{\mathbf{C}^{k-\mu}\} \,. \quad (23)
$$

From Eq. (23), it can then be deduced that, for all k in $\{2, \ldots, K\}$ and for all μ in $\{1, \ldots, k-1\}$, the second-order cross-tensor moment equation with weights can be written as

$$
[H^{k,k-\mu}([A], \{\mathbf{g}\})] = [0_{N,N}] \,, \quad (24)
$$

in which function $[H^{k,k-\mu}]$ with values in $\mathbb{R}^{N \times N}$ is defined by

$$
\begin{aligned}
[H^{k,k-\mu}([A], \{\mathbf{g}\})] = \mathbb{T}^k : \big(E\{\mathbf{C}^k \otimes \mathbf{C}^{k-\mu}\} \\
- ([I_N] - \Delta\tau_k[A]) \, E\{\mathbf{C}^{k-1} \otimes \mathbf{C}^{k-\mu}\} - \Delta\tau_k \, \mathbf{g}^k \otimes E\{\mathbf{C}^{k-\mu}\} \big) \,, \quad (25)
\end{aligned}
$$

in which the components of fourth-order tensor $\mathbb{T}^k$ are defined by Eq. (20).

4.4. Formulation of the least-squares optimization problem for the identification of the parameters

(i) *Definition of the admissible sets for the optimization problem.* In the framework of the long- term evolution of the track irregularities, one or several components of the vector-valued indicator are not bounded functions when time k is increasing. Consequently, for any fixed value of K, it is not necessary to introduce a convergence condition on $\max_k\{\Delta\tau_k\}$ as a function of the spectral radius of $[A]$. Therefore, no constraints are introduced for matrix-valued parameter $[A]$ in the optimization problem, and the admissible space for $[A]$ is then defined by

$$\mathcal{C}_A = \{\,[A] \in \mathbb{R}^{N \times N}\,\}. \tag{26}$$

The discrete time-dependent diffusion matrix of the Markov chain defined by Eq. (9) is written as $[\sigma^k] = [h^k][h^k]^T$ that is a positive-definite matrix taking into account the hypothesis introduced in Section 3. Consequently, the family $\{[h]\} = \{[h^k], k = 2, \ldots, K\}$, of lower triangular matrices $[h^k]$ must belong to the following admissible set

$$\mathcal{C}_h = \{\,[h^k] \in \mathbb{M}_N^L \quad \text{for} \quad k = 2, \ldots, K\,\}, \tag{27}$$

in which $\mathbb{M}_N^L$ is the set of all the lower triangular $(N \times N)$ real matrices with positive diagonal entries. No constraints are needed for the family of $\mathbb{R}^N$-values vectors $\{\mathbf{g}\} = \{\mathbf{g}^k, k = 2, \ldots, K\}$. The admissible space for $\{\mathbf{g}\}$ is thus

$$\mathcal{C}_g = \{\,\mathbf{g}^k \in \mathbb{R}^N \quad \text{for} \quad k = 2, \ldots, K\,\}. \tag{28}$$

(ii) *Definition of the target quantities for the optimization problem.* The target quantities introduced in Eqs. (11), (18) and (24) are, for $k = 2, \ldots, K$, the vectors $E\{\mathbf{C}^k\}$ and $E\{\mathbf{C}^{k-1}\}$, and, for $\mu = 0, \ldots, k-1$, the tensors $E\{\mathbf{C}^k \otimes \mathbf{C}^{k-\mu}\}$ and $E\{\mathbf{C}^{k-1} \otimes \mathbf{C}^{k-\mu}\}$. For computing the cost function in the least-squares problem, these target quantities are approximated by $E\{\mathbf{C}^{\text{mod}}(\tau_k)\}$, $E\{\mathbf{C}^{\text{mod}}(\tau_{k-1})\}$, $E\{\mathbf{C}^{\text{mod}}(\tau_k) \otimes \mathbf{C}^{\text{mod}}(\tau_{k-\mu})\}$ and $E\{\mathbf{C}^{\text{mod}}(\tau_{k-1}) \otimes \mathbf{C}^{\text{mod}}(\tau_{k-\mu})\}$, that are estimated using the stochastic computational model of the high-speed train dynamics (see Section 2).

(iii) *Formulation of the optimization problem for the identification of the parameters.* Using Eqs. (11), (18) and (24), the optimal values $[A^{\text{opt}}]$, $\{\mathbf{g}^{\text{opt}}\}$, and $\{[h^{\text{opt}}]\}$ for parameters $[A]$, $\{\mathbf{g}\}$, and $\{[h]\}$ of the stochastic predictive model described by Eq. (9) are estimated using the least-squares method with constraints, for which the cost function is defined by

$$J([A], \{\mathbf{g}\}, \{[h]\}) = \sum_{k=2}^{K} (\|\mathbf{f}^k([A], \{\mathbf{g}\})\|_2^2 + \|[F^k([A], \{\mathbf{g}\}, \{[h]\})]\|_F^2$$

$$+ \sum_{\mu=1}^{k-1} \|[H^{k,k-\mu}([A], \{\mathbf{g}\})]\|_F^2), \quad (29)$$

12

in which $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^N$ and $\|\cdot\|_F$ is the Frobenius norm in $\mathbb{R}^{N \times N}$. The optimal values $[A^{\mathrm{opt}}]$, $\{\mathbf{g}^{\mathrm{opt}}\}$, and $\{[h^{\mathrm{opt}}]\}$ are constructed as the solution of the following optimization problem.

$$\{[A^{\mathrm{opt}}], \{\mathbf{g}^{\mathrm{opt}}\}, \{[h^{\mathrm{opt}}]\}\} = \arg \min_{[A] \in \mathcal{C}_A, \, \{\mathbf{g}\} \in \mathcal{C}_g, \, \{[h]\} \in \mathcal{C}_h} J([A], \{\mathbf{g}\}, \{[h]\}). \quad (30)$$

For this optimization problem, the number of variables is $N^2 + N(K-1) + (K-1)N(N+1)/2$ (and will be 675, in the application presented after). This optimization problem is solved with the trust-region-reflective algorithm with constraints, for which the initial values, denoted by $[A^0]$, $\{\mathbf{g}^0\}$, and $\{[h^0]\}$, are computed by solving the simplified optimization problem described in Appendix A.

5. Prediction of the long- term evolution with the stochastic model

As it has been explained in Section 3, the stochastic model identified in Section 4 has to be adapted in order to be able to predict the long- term evolution $\mathbf{C}^{K+1} = \mathbf{C}(\tau_{K+1})$ of the indicator of the train dynamics at discrete long time τ_{K+1}. In this section, the families $\{\mathbf{g}\}$ and $\{[h]\}$ are considered as the values of the functions $\tau \mapsto \mathbf{g}(\tau)$ and $\tau \mapsto [h(\tau)]$ at discrete long times τ_k, $k = 2, \ldots, K$, such that $\mathbf{g}(\tau_k) = \mathbf{g}^k$ and $[h(\tau_k)] = [h^k]$. Functions $\mathbf{g}$ and $[h]$ are then approximated (see Section 3) by

$$\mathbf{g}^{\mathrm{aff}}(\tau_k) = \mathbf{a}_g \, \tau_k + \mathbf{b}_g \quad , \quad [h^{\mathrm{aff}}(\tau_k)] = [a_h] \, \tau_k + [b_h], \quad (31)$$

in which $\mathbf{a}_g$ and $\mathbf{b}_g$ are vectors in $\mathbb{R}^N$, where $[b_h]$ is a matrix in $\mathbb{M}_N^L$, and where $[a_h]$ is a lower triangular $(N \times N)$ real matrix such that $[a_h] \, \tau_k + [b_h]$ is in $\mathbb{M}_N^L$ for all k in $\{2, \ldots, K\}$. These conditions define the admissible set for $[a_h]$ and $[b_h]$ denoted by $\mathcal{C}_{a_h b_h}$ and ensure that the values of $[h^{\mathrm{aff}}]$ are in $\mathcal{C}_h$.

- For $i = 1, \ldots, N$, the optimal value $(a_{g_i}^{\mathrm{opt}}, b_{g_i}^{\mathrm{opt}})$ of (a_{g_i}, b_{g_i}) is given as the solution of the following optimization problem,

$$(a_{g_i}^{\mathrm{opt}}, b_{g_i}^{\mathrm{opt}}) = \arg \min_{(a_{g_i}, b_{g_i}) \in \mathbb{R}^2} \sum_{k=2}^{K} \Delta \tau_k \, |g_i^{\mathrm{opt},k} - (a_{g_i} \, \tau_k + b_{g_i})|^2. \quad (32)$$

  This unconstrained linear least-squares problem is solved with a classical algorithm.

- The optimal value $([a_h^{\mathrm{opt}}], [b_h^{\mathrm{opt}}])$ of $([a_h], [b_h])$ is given as the solution of the following optimization problem,

$$([a_h^{\mathrm{opt}}], [b_h^{\mathrm{opt}}]) =$$

$$\arg \min_{([a_h], [b_h]) \in \mathcal{C}_{a_h b_h}} \sum_{k=2}^{K} \Delta \tau_k \, |[h^{\mathrm{opt},k}] - ([a_h] \, \tau_k + [b_h])|^2. \quad (33)$$

  This constrained linear least-squares problem is solved using the trust-region-reflective algorithm.

The values $\mathbf{g}^{\mathrm{aff}}(\tau_{K+1})$ and $[h^{\mathrm{aff}}(\tau_{K+1})]$ are then calculated by

$$\mathbf{g}^{\mathrm{aff}}(\tau_{K+1}) = \mathbf{a}_g^{\mathrm{opt}}\,\tau_{K+1} + \mathbf{b}_g^{\mathrm{opt}}\ ,\ \ [h^{\mathrm{aff}}(\tau_{K+1})] = [a_h^{\mathrm{opt}}]\,\tau_{K+1} + [b_h^{\mathrm{opt}}]\,. \qquad (34)$$

The prediction $\mathbf{C}^{\mathrm{aff},K+1}$ of the vector-valued random indicator $\mathbf{C}^{K+1} = \mathbf{C}(\tau_{K+1})$ at long time τ_{K+1} , given $\mathbf{C}^{\mathrm{mod}}(\tau_K)$, is then estimated using Eq. (9) that yields

$$\mathbf{C}^{\mathrm{aff},K+1} = ([I_N] - \Delta\tau_{K+1}\,[A^{\mathrm{opt}}])\mathbf{C}^{\mathrm{mod}}(\tau_K) + \Delta\tau_{K+1}\,\mathbf{g}^{\mathrm{aff}}(\tau_{K+1})$$
$$+ [h^{\mathrm{aff}}(\tau_{K+1})]\,\Delta\mathbf{W}^{K+1}\,. \qquad (35)$$

Remark. As explained in Section 1, Eq. (35) corresponds to Eqs. (9) and (10) for which the initial condition is given at time τ_K:

$$\mathbf{C}^{\mathrm{aff},k} = ([I_N] - \Delta\tau_k\,[A^{\mathrm{opt}}])\mathbf{C}^{\mathrm{aff},k-1} + \Delta\tau_k\,\mathbf{g}^{\mathrm{aff}}(\tau_k)$$
$$+ [h(\tau_k)^{\mathrm{aff}}]\,\Delta\mathbf{W}^k\ \ ,\ \ k = K+1\,, \qquad (36)$$

with the initial condition

$$\mathbf{C}^{\mathrm{aff},K} = \mathbf{C}^{\mathrm{mod}}(\tau_K)\,, \qquad (37)$$

in which $\Delta\tau_{K+1} = \tau_{K+1} - \tau_K$.

6. Application to a high-speed line of the French railway network

6.1. Constructing the vector-valued random indicator $\mathbf{C}^{\mathrm{mod}}$

The components of the vector-valued random indicator have been selected by using the expertise of SNCF, are related to the certification standards for railway vehicles [16], and correspond to quantities that are effectively measured for characterizing the dynamic response of the train. The number of components of the vector-valued indicator is $N = 9$. For $j = 1, \ldots, 9$, these components are

- C_1: the lateral acceleration of the first bogie in the train.

- C_2: the vertical acceleration of the first bogie in the train.

- C_3: the lateral acceleration of the third bogie in the train.

- C_4: the lateral acceleration of the second coach in the train.

- C_5: the sum of lateral forces on the ninth train wheelset.

- C_6: the sum of vertical forces on the first train wheelset.

- C_7: the sum of vertical forces of the second train wheelset.

- C_8: the sum of vertical forces of the tenth train wheelset.

- C_9: the difference between right-wheel and left-wheel vertical forces of the tenth train wheelset.

In order to assess the quality of the dynamic response of the high-speed train, for each component C_j, a value of the threshold c_j^* is chosen according to the expertise of the railway company and this value is close to the limit value used for the certification of the vehicles given in [16]. The values of the components of vector $\mathbf{c}^*$ can be adapted depending on the maintenance policy chosen by the railway company. All the numerical results presented in the figures displayed in Section 6 are dimensionless concerning the values of the components of the indicator (named as "dimensionless indicator"). The discrete long times $\tau_1, \ldots, \tau_{K+1}$ are transformed in dimensionless discrete long times (named as dimensionless long time τ). A reference time-increment denoted by $\Delta\tau_{\mathrm{ref}}$ is chosen as

$$\Delta\tau_{\mathrm{ref}} = \Delta\tau_{K+1}\,, \tag{38}$$

and the dimensionless long times are written as $\tau_k/\Delta\tau_{\mathrm{ref}}$, for $k = 1, \ldots, K+1$.

All the results presented in Section 6 are related to a given stretch of the track that corresponds to a stretch in alignment (the horizontal curvature is null). The measurements of the track irregularities performed at the discrete long times $\tau_1, \ldots, \tau_K$ show that the track irregularities and their evolution are mainly in the vertical direction. This is the reason why the components of the vector-valued indicator $\mathbf{C}^{\mathrm{mod}}$ which are related to the vertical direction (namely $C_2^{\mathrm{mod}}, C_6^{\mathrm{mod}}, C_7^{\mathrm{mod}}, C_8^{\mathrm{mod}}$) will present a significant long-term evolution, while the long-term evolution of those related to the horizontal direction (namely $C_1^{\mathrm{mod}}, C_3^{\mathrm{mod}}, C_4^{\mathrm{mod}}, C_5^{\mathrm{mod}}, C_9^{\mathrm{mod}}$) will not be significant.

The construction of the vector-valued indicator is carried out as explained in Section 2. For this track stretch and for discrete long times τ_k, $k = 1, \ldots, K$, track irregularities vector $\mathbf{x}_{\tau_k}^{\mathrm{meas}} = (\mathbf{x}_{\tau_k}^{\mathrm{meas},1}, \mathbf{x}_{\tau_k}^{\mathrm{meas},2}, \mathbf{x}_{\tau_k}^{\mathrm{meas},3}, \mathbf{x}_{\tau_k}^{\mathrm{meas},4})$ is measured by a recording train ($K = 12$ measures are available for the given track stretch). The optimal value $\boldsymbol{\delta}_{\tau_1}^{\mathrm{opt}} = (\delta_{\tau_1}^{1,\mathrm{opt}}, \delta_{\tau_1}^{2,\mathrm{opt}}, \delta_{\tau_1}^{3,\mathrm{opt}}, \delta_{\tau_1}^{4,\mathrm{opt}})$ of hyperparameter $\boldsymbol{\delta}_{\tau_1}$ for this track stretch is $\boldsymbol{\delta}_{\tau_1}^{\mathrm{opt}} = (0.15, 0.9, 0.85, 0.8)$. The local stochastic model of the track stretch is constructed using Eq. (5) that yields, for $\kappa = 1, 2, 3, 4$ and for $k = 1, \ldots, K$,

$$\widetilde{\mathbf{X}}_{\tau_k}^{\kappa}(\delta_{\tau_1}^{\kappa,\mathrm{opt}}) = [Q^{\kappa}] \left(\boldsymbol{\eta}_{\tau_k}^{\mathrm{meas}} + \delta_{\tau_1}^{\kappa,\mathrm{opt}} \mathbf{G}^{\kappa}\right)\,, \tag{39}$$

in which, from Eqs. (2) and (3), it can be deduced that

$$\boldsymbol{\eta}_{\tau_k}^{\mathrm{meas}} = [\lambda]^{-1} [Q]^T \mathbf{x}_{\tau_k}^{\mathrm{meas}}\,, \tag{40}$$

which is the projection of the measurement $\mathbf{x}_{\tau_k}^{\mathrm{meas}}$ on the global stochastic model.

The Monte-Carlo method is performed with $\nu = 2000$ independent realizations of the stochastic model, and is summarized hereinafter.

- $\nu = 2000$ independent realizations $\mathbf{G}(\theta_1), \ldots, \mathbf{G}(\theta_\nu)$ of $\mathbf{G} = (\mathbf{G}^1, \mathbf{G}^2, \mathbf{G}^3, \mathbf{G}^4)$ are generated.

- These ν independent realizations of $\mathbf{G}$ are used for constructing, for each $k = 1, \ldots, K$, the ν independent realizations of random vector $\widetilde{\mathbf{X}}_{\tau_k} = (\widetilde{\mathbf{X}}_{\tau_k}^1, \widetilde{\mathbf{X}}_{\tau_k}^2, \widetilde{\mathbf{X}}_{\tau_k}^3, \widetilde{\mathbf{X}}_{\tau_k}^4)$ by using Eq. (39). For each realization of the track

irregularities $\widetilde{\mathbf{X}}_{\tau_k}$, the deterministic dynamic response of the train induced by this realization of the track irregularities is computed with a multibody commercial software (*Vampire*), and the ν corresponding independent realizations $\mathbf{C}^{\mathrm{sim}}(\tau_k; \theta_1), \ldots \mathbf{C}^{\mathrm{sim}}(\tau_k; \theta_\nu)$ of the vector-valued random indicator $\mathbf{C}^{\mathrm{sim}}(\tau_k)$ are computed.

- Then, ν independent realizations of the $\mathbb{R}^N$-valued non-Gaussian second-order random vector $\mathbf{B}^{\mathrm{out}}$ introduced in Section 2.2.2 are generated.

- The ν corresponding independent realizations of the family $\boldsymbol{\mathcal{C}}^{\mathrm{mod}}$ of the vector-valued random indicators $\{\mathbf{C}^{\mathrm{mod}}(\tau_1), \ldots, \mathbf{C}^{\mathrm{mod}}(\tau_K)\}$ are $\boldsymbol{\mathcal{C}}^{\mathrm{mod}}(\theta_1, \theta_1'), \ldots, \boldsymbol{\mathcal{C}}^{\mathrm{mod}}(\theta_\nu, \theta_\nu')$ that are computed using Eq. (7), which is rewritten as $C_j^{\mathrm{mod}}(\tau_k; \theta_\ell, \theta_\ell') = C_j^{\mathrm{sim}}(\tau_k; \theta_\ell) \exp(B_j^{\mathrm{out}}(\theta_\ell'))$, with $j = 1, \ldots, N$, $k = 1, \ldots, K$, and $\ell = 1, \ldots, \nu$.

- As explained in Section 4.4-(ii), we need to estimate the target quantities in order to perform the identification of the parameters (see Section 6.3). Consequently, for $k = 2, \ldots, K$ and for $\mu = 0, \ldots, k-1$, the vectors $E\{\mathbf{C}^{\mathrm{mod}}(\tau_k)\}$ and $E\{\mathbf{C}^{\mathrm{mod}}(\tau_{k-1})\}$, and the tensors $E\{\mathbf{C}^{\mathrm{mod}}(\tau_k) \otimes \mathbf{C}^{\mathrm{mod}}(\tau_{k-\mu})\}$ and $E\{\mathbf{C}^{\mathrm{mod}}(\tau_{k-1}) \otimes \mathbf{C}^{\mathrm{mod}}(\tau_{k-\mu})\}$, are estimated by using the classical statistical estimator of the mathematical expectation with independent realizations $\boldsymbol{\mathcal{C}}^{\mathrm{mod}}(\theta_1, \theta_1'), \ldots, \boldsymbol{\mathcal{C}}^{\mathrm{mod}}(\theta_\nu, \theta_\nu')$.

In order to analyze the long- term evolution of the vector-valued random indicator $\mathbf{C}^{\mathrm{mod}}(\tau_k)$ as a function of $k = 1, \ldots, K$, the mean function and the confidence region at 90% of its components $C_j^{\mathrm{mod}}(\tau_k)$, are calculated as a function of discrete long time τ_k. The mean function and the confidence region, which depend on τ_k, are compared to the threshold level c_j^*. For $j = 2, 6, 7, 8$ (components related to the vertical direction), a time evolution of the mean value and of the confidence region can be observed. For the other values of j, there is no significant time evolution. In order to limit the number of figures, only the dimensionless results related to $j = 2, 6, 7, 8$ are plotted in Figure 2. In this figure, the values of the components of the indicator rises faster and faster, showing that the dynamic response of the train in the vertical direction deteriorates, and that the deterioration is accelerating in long time. A maintenance operation is required at the end of the observed period. The confidence region remains nearly constant for the observed components, except for the second component, for which it increases and it is very large. Nevertheless the observation of the confidence region in global shows that the stochastic model for the vector-valued indicator is well adapted.

6.2. Choice of the family of the weight coefficients

The family $\{\alpha_1, \alpha_2, \ldots, \alpha_N\}$ of the weight coefficients (introduced in Section 4) is chosen in order to increase or to decrease the weight of each component in the cost function: high values of α_j are given for components $j = 2, 6, 7, 8$ and low values are given for components $j = 1, 3, 4, 5, 9$. The condition appearing in Eq. (14) has also to be fulfilled. The chosen values for $\boldsymbol{\alpha}$ are $\alpha_1 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_9 = 0.01$, $\alpha_2 = 0.35$, and $\alpha_6 = \alpha_7 = \alpha_8 = 0.2$.
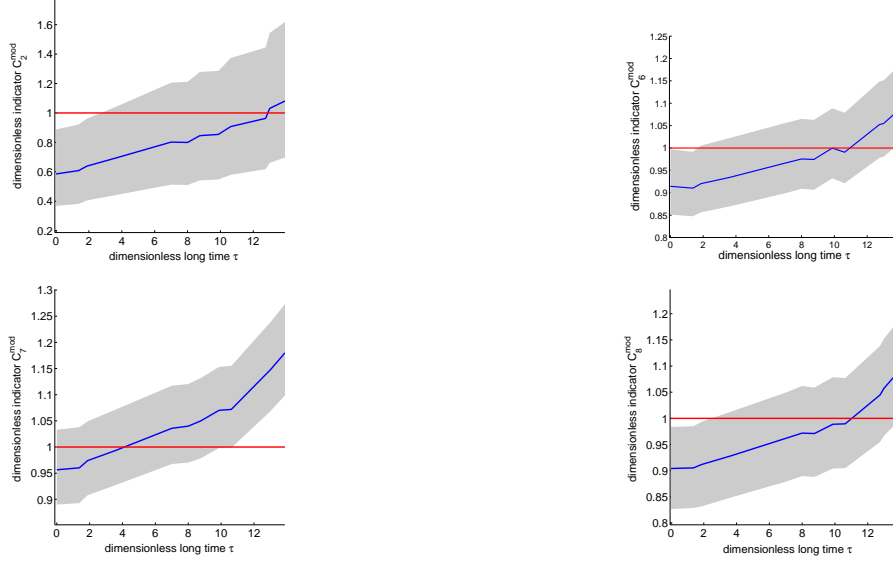
Figure 2: Dimensionless long- term evolution of dimensionless random indicator C_2^{mod}(up), C_6^{mod}(up middle), C_7^{mod}(down middle), C_8^{mod}(down). Confidence region with probability 90% (grey region), threshold (horizontal solid line), mean value (solid line).

6.3. Identification of the parameters of the stochastic model for the long- term evolution

The optimal values $\{[A^{\mathrm{opt}}], \{\mathbf{g}^{\mathrm{opt}}\}, \{[h^{\mathrm{opt}}]\}\}$ of the parameters of Eq. (9) are computed by using the methodology presented in Section 4. The initial values for the trust-region-reflective algorithm are computed as explained in Appendix A. For the computation of the optimal values of the 675 unknown parameters that have to be identified (see Section 4.4), the trust-region-reflective algorithm stops when the number of iterations is greater than 200 times the number of variables $(135,000)$, for which the norm of the residuals is 2.7×10^{-5}.

6.4. Comparison of $\mathbf{C}^k$ with $\mathbf{C}^{\mathrm{mod}}(\tau_k)$

In this section, for $k = 2, \ldots, K$, we present a comparison of the vector-valued random indicator $\mathbf{C}^{\mathrm{mod}}(\tau_k)$ constructed in Section 6.1, with $\mathbf{C}^k$ computed with Eqs. (9) and (10) using the optimal parameters $\{[A^{\mathrm{opt}}], \{\mathbf{g}^{\mathrm{opt}}\}, \{[h^{\mathrm{opt}}]\}\}$ identified in Section 6.3. In order to limit the number of figures, we restrict the presentation to the component $j = 6$ of the vector-valued random indicator. The results for the other components $j = 2, 7, 8$ are similar. For each quantity, $\mathbf{C}^{\mathrm{mod}}(\tau_k)$ and $\mathbf{C}^k$, denoted hereinafter as $D(\tau_k)$, the time evolution of the mean value $E\{D(\tau_k)\}$ and of the standard deviation $E\{(D(\tau_k) - E\{D(\tau_k)\})^2\}^{1/2}$ are estimated using the classical statistical estimator and are displayed in Figure 3. The time evolution of the probability density
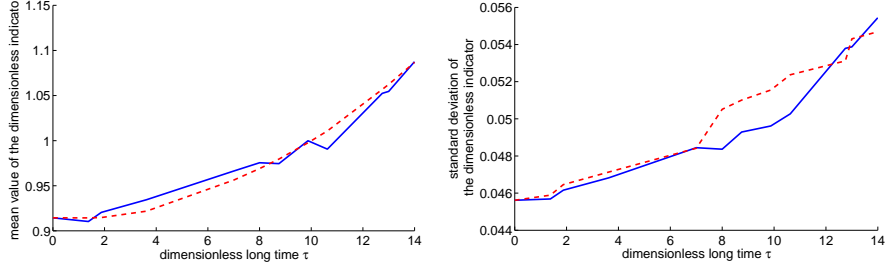
17

Figure 3: Second-order moments of the dimensionless long- term evolution for the dimensionless indicator $D(\tau_k) = C_6^{\mathrm{mod}}(\tau_k)$ (solid line) and $D(\tau_k) = C_6^k$ (dashed line): mean value (left) and standard deviation (right).
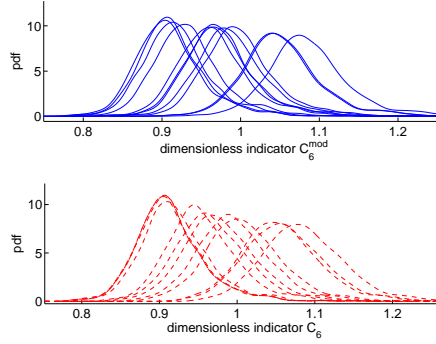

Figure 4: Probability density function (pdf) of the dimensionless long- term evolution for the dimensionless indicator $D(\tau_k) = C_6^{\mathrm{mod}}(\tau_k)$ (up, solid line) and $D(\tau_k) = C_6^k$ (down, dashed line). Each curve corresponds to a given time τ_k. When k is increasing, the curves move to the right.

function (pdf) $p_{D(\tau_k)}$ is estimated by the Gaussian kernel density method and is shown in Figure 4. Figure 5 displays the graph $k \mapsto d^i(\tau_k)$ in which $d^i(\tau_k)$ is such that $\mathrm{Prob}\{D(\tau_k) \leq d^i(\tau_k)\} \geq q^i$, in which q^i is the quantile varying in the interval $[0.4, 0.98]$. Figures 3, 4, and 5 show that the stochastic model of the long- term evolution defined by Eqs. (9) and (10) is a good approximation for representing the long- term evolution of vector-valued random indicator $\mathbf{C}^{\mathrm{mod}}$, in particular for the time-evolution of the pdf (Figure 4) and the time-evolution of the quantiles (Figure 5), because the long-term evolution of $\mathbf{C}$ is very close to the long-term evolution of $\mathbf{C}^{\mathrm{mod}}$.

6.5. Prediction of the vector-valued random indicator $\mathbf{C}^{K+1}$

The prediction of vector-valued random indicator $\mathbf{C}^k$ for $k = K + 1$ is estimated by $\mathbf{C}^{\mathrm{aff},K+1}$, constructed with Eq. (35), for which the parameters are $[A^{\mathrm{opt}}]$, $\mathbf{g}^{\mathrm{aff}}(\tau_{K+1}) = \mathbf{a}_g^{\mathrm{opt}} \tau_{K+1} + \mathbf{b}_g^{\mathrm{opt}}$ and $[h^{\mathrm{aff}}(\tau_{K+1})] = [a_h^{\mathrm{opt}}] \tau_{K+1} + [b_h^{\mathrm{opt}}]$ as
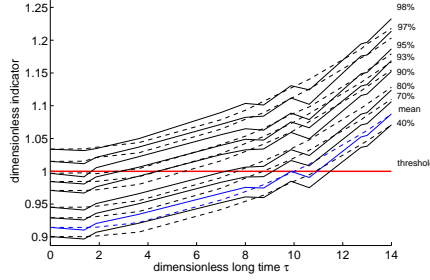
18

Figure 5: Quantiles of the dimensionless long- term evolution for the dimensionless indicator $D(\tau_k) = C_6^{\mathrm{mod}}(\tau_k)$ (solid line) and $D(\tau_k) = C_6^k$ (dashed line), graphs $k \mapsto d^i(\tau_k)$, in which $d^i(\tau_k)$ is such that $\mathrm{Prob}\{D(\tau_k) \leq d^i(\tau_k)\} \geq q^i$ for different values in percent of the quantile q^i belonging to the interval $[40\%, 98\%]$. The horizontal line corresponds to the threshold level c_6^*.
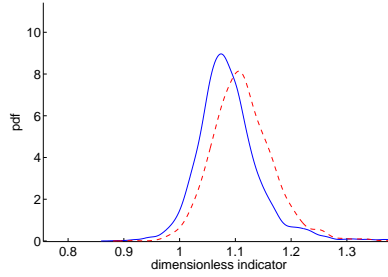

Figure 6: Probability density function of the prediction for dimensionless long-term evolution of the dimensionless indicator $C_6^{\mathrm{aff},K+1}$ at long time τ_{K+1} (dashed line) given $C_6^{\mathrm{mod}}(\tau_K)$ at long time τ_K (solid line).

explained in Section 5 (see Eq. (34)). Given the computed indicator $\mathbf{C}^{\mathrm{mod}}(\tau_K)$ at long time τ_K, the prediction of the probability density function of $\mathbf{C}^{\mathrm{aff},K+1}$ at long time τ_{K+1} is displayed in Figure 6 for the component $j = 6$ of the vector-valued dimensionless indicator. This figure shows a significant evolution of the pdf of $C_6^{\mathrm{aff},K+1}$ at long time τ_{K+1}, which has been predicted from the known pdf of $C_6^{\mathrm{mod}}(\tau_K)$ at long time τ_K. Such a pdf evolution is coherent with the time evolution of the pdf that has been displayed in Figure 4. It should be noted that the values predicted for the other statistical quantities such as the mean value, the standard deviation, and the quantiles, show a similar significant evolution.

6.6. Relevance of the stochastic predictive model

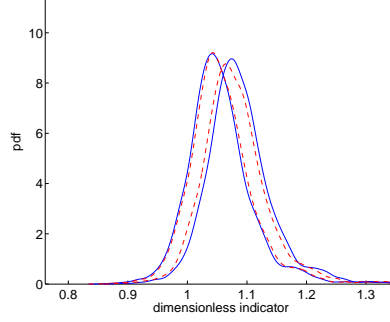Two criteria are used in order to assess the relevance of the stocastic predictive model that has been identified:

19

Figure 7: Quality assessment of the stochastic predictive model comparing the pdf of $C_6^{\mathrm{aff},k}$ (dashed line) with the pdf of $C_6^{\mathrm{mod}}(\tau_k)$ (solid line) for $k = K - 1$ and for $k = K$.

- For the quality assessment, the predictions $\mathbf{C}^{\mathrm{aff},K-1}$ and $\mathbf{C}^{\mathrm{aff},K}$ are performed given the known random vector $\mathbf{C}^{\mathrm{mod}}(\tau_{K-2})$. This prediction is carried out using Eq. (34) and Eq. (35) that is rewritten as

$$\mathbf{C}^{\mathrm{aff},k} = ([I_N] - \Delta\tau_k \, [A^{\mathrm{opt}}])\mathbf{C}^{\mathrm{aff},k-1} + \Delta\tau_k \, \mathbf{g}^{\mathrm{aff}}(\tau_k)$$
$$+ [h^{\mathrm{aff}}(\tau_k)] \, \Delta\mathbf{W}^k \quad , \quad k = K - 1, K, \quad (41)$$

with the initial condition

$$\mathbf{C}^{\mathrm{aff},K-2} = \mathbf{C}^{\mathrm{mod}}(\tau_{K-2}). \qquad (42)$$

The quality assessment is then evaluated in comparing $\mathbf{C}^{\mathrm{aff},K-1}$ with $\mathbf{C}^{\mathrm{mod}}(\tau_{K-1})$, and then in comparing $\mathbf{C}^{\mathrm{aff},K}$ with $\mathbf{C}^{\mathrm{mod}}(\tau_K)$.

For the probability density function, the quality assessment of the stochastic predictive model is displayed in Figure 7, in which the pdf of $C_6^{\mathrm{aff},k}$ is compared with the pdf of $C_6^{\mathrm{mod}}(\tau_k)$ for $k = K - 1$ and for $k = K$. For both long times $K - 1$ and K, the curves of both probability density functions are very close, which shows that the stochastic predictive model is good.

For the quantiles of the dimensionless indicator $D(\tau_k) = C_6^{\mathrm{mod}}(\tau_k)$ and $D(\tau_k) = C_6^{\mathrm{aff},k}$ for $k = K - 1$ and for $k = K$, the quality assessment can be viewed in Figure 8 that displays the graphs $k \mapsto d^i(\tau_k)$, in which $d^i(\tau_k)$ is such that $\mathrm{Prob}\{D(\tau_k) \leq d^i(\tau_k)\} \geq q^i$ for different values in percent of the quantile q^i belonging to the interval $[40\%, 98\%]$. Again, for each quantile, the curve describing C_6 is very close to the curve describing C_6^{mod}, which confirms that the stochastic predictive model is very good.

- The relevance of the stochastic predictive model can be obtained in computing the modeling error induced by both the stochastic predictive model itself (as described in Eq. (9)) and the approximations of functions $\mathbf{g}$ and
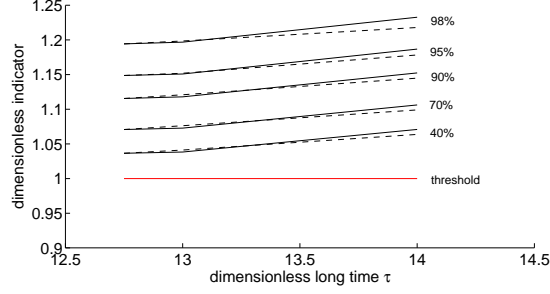
20

Figure 8: Quality assessment of the stochastic predictive model comparing the quantiles of $C_6^{\mathrm{aff},k}$ (dashed line) with the quantiles of $C_6^{\mathrm{mod}}(\tau_k)$ (solid line) for $k = K - 1$ and for $k = K$.
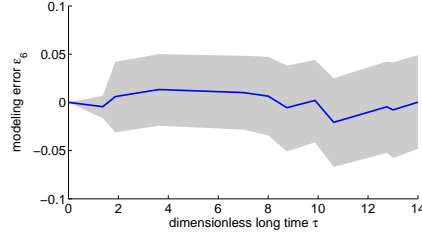

Figure 9: Modeling error of the stochastic predictive model for C_6

[h] by the affine functions $\mathbf{g}^{\mathrm{aff}}$ and [h^{aff}]. For each $k = 2, \ldots, K$, $\mathbf{C}^{\mathrm{aff},k}$ is computed with Eq. (36), using $\mathbf{C}^{\mathrm{mod}}(\tau_{k-1})$ as initial value, that is to say by using the following equation

$$\mathbf{C}^{\mathrm{aff},k} = ([I_N] - \Delta\tau_k\,[A^{\mathrm{opt}}])\mathbf{C}^{\mathrm{mod}}(\tau_{k-1}) + \Delta\tau_k\,\mathbf{g}^{\mathrm{aff}}(\tau_k) \\ + [h^{\mathrm{aff}}(\tau_k)]\,\Delta\mathbf{W}^k\,. \quad (43)$$

For component j of the vector-valued indicator and at time τ_k, the random modeling error denoted by ε_j^k is defined by

$$\varepsilon_j^k = \frac{C_j^{\mathrm{mod}}(\tau_k) - C_j^{\mathrm{aff},k}}{C_j^{\mathrm{mod}}(\tau_k)}\,. \quad (44)$$

For the component $j = 6$, the mean value of ε_j^k and its confidence interval at 90% are plotted in Figure 9 for $k = 1, \ldots, K$. It should be noted that the dispersion is 0 for $k = 1$ corresponding to the initial time τ_1 for which $C_j^{\mathrm{aff},1} = C_j^{\mathrm{mod}}(\tau_1)$. Then, the dispersion lightly increases with long time τ (that is coherent) but stays very small (smaller than 5%).

- For the long-term evolution, the quality assessment of the stochastic predictive model in terms of the pdf, in terms of the quantiles, and in terms

of the confidence region of the modeling errors, shows that the stochastic predictive model is good and can be used to predict the long-term evolution of the vector-valued random indicator **C**.

7. Conclusion and Prospects

For the companies in charge of the maintenance of railway networks, there is a great interest to predict the long- term evolution of the track irregularities for a given track stretch of the network, in order to be able to anticipate the start off of the maintenance operations. In this work, the long- term evolution of the track irregularities of the given track stretch has been evaluated through a vector-valued random indicator related to the dynamic response of the train induced by the random track irregularities. The long- term evolution of this vector-valued random indicator is modeled by a discrete non-Gaussian nonstationary stochastic model (ARMA type model), for which the coefficients are time-dependent. These coefficients have been identified by a least-squares method and fitted on long time, using experimental measurements.

The proposed stochastic predictive model, based on big data made up of a lot of experimental measurements performed for the French high-speed train network, allows for predicting the statistical quantities of the vector-valued random indicator for long times for which no measurements have been performed yet. It has been demonstrated that this proposed stochastic predictive model is very good. This proposed model can help to determine the best time to start off the maintenance operations as a function of a chosen threshold for the indicator.

Acknowledgments

Appendix A. Simplified optimization problem for computing the initial values

Because there is no guarantee for the optimization problem introduced in Section 4.4 to be convex, the initial values of the parameters in the optimization algorithm have to be chosen close to the optimal values. That is why a simplified formulation is proposed in order to compute the initial values of the parameters. This simplified formulation consists in using Eqs. (12), (19), and (25), in which full matrix $[A]$ is replaced by a diagonal matrix with real diagonal entries $A_1, \ldots, A_N$, and for $k = 2, \ldots, K$, the lower triangular matrix $[h^k]$ is replaced by a diagonal matrix with positive diagonal entries $h_1^k, \ldots, h_N^k$. For fixed j in $\{1, \ldots, N\}$ and k in $\{2, \ldots, K\}$, Eqs. (12), (19), and (25) become

$$f_j^k(A_j, \{g_j\}) = [\mathbb{N}_f^k]_{jj} \left(E\{C_j^k\} - (1 - \Delta\tau_k A_j)E\{C_j^{k-1}\} - \Delta\tau_k\, g_j^k \right), \quad (45)$$

$$F_j^k(A_j, \{g_j\}, \{h_j\}) =$$
$$\{\mathbb{T}^k\}_{jjjj} \left(E\{(C_j^k)^2\} - (1 - \Delta\tau_k\, A_j)\, E\{C_j^{k-1}\, C_j^k\} \right.$$
$$\left. -\Delta\tau_k\, g_j^k E\{C_j^k\} - (h_j^k)^2 \Delta\tau_k \right) , \quad (46)$$

and, for $\mu = 1, \ldots, k-1$,

$$H_j^{k,k-\mu}(A_j, \{g_j\}) =$$
$$\{\mathbb{T}^k\}_{jjjj} \left(E\{C_j^k C_j^{k-\mu}\} - (1 - \Delta\tau_k\, A_j) E\{C_j^{k-1} C_j^{k-\mu}\} \right.$$
$$\left. -\Delta\tau_k\, g_j^k\, E\{C_j^{k-\mu}\} \right) , \quad (47)$$

in which $[\mathbb{N}_f^k]$ and $\mathbb{T}^k$ are defined by Eqs. (13) and (20). Let $\mathcal{C}_{g_j}$ and $\mathcal{C}_{h_j}$ be the admissible sets for $\{g_j\}$ and $\{h_j\}$ defined by

$$\mathcal{C}_{g_j} = \{g_j^k \in \mathbb{R} \quad , \quad \text{for} \quad k = 2, \ldots, K\}, \quad (48)$$

$$\mathcal{C}_{h_j} = \{h_j^k > 0 \quad , \quad \text{for} \quad k = 2, \ldots, K\}. \quad (49)$$

For all j fixed in $\{1, \ldots, N\}$, the cost function J_j is introduced such that

$$J_j(A_j, \{g_j\}, \{h_j\}) = \sum_{k=2}^{K} (|f_j^k(A_j, \{g_j\})|^2 + |F_j^k(A_j, \{g_j\}, \{h_j\})|^2$$
$$+ \sum_{\mu=1}^{k-1} |H_j^{k,k-\mu}(A_j, \{g_j\})|^2). \quad (50)$$

In the computation of the cost function J_j, for $j = 1, \ldots, N$, $k = 2, \ldots, K$ and for $\mu = 0, \ldots, k-1$, the target quantities $E\{C_j^k\}$, $E\{C_j^{k-1}\}$, $E\{C_j^k C_j^{k-\mu}\}$, and $E\{C_j^{k-1} C_j^{k-\mu}\}$ are replaced by $E\{C_j^{\mathrm{mod}}(\tau_k)\}$, $E\{C_j^{\mathrm{mod}}(\tau_{k-1})\}$, $E\{C_j^{\mathrm{mod}}(\tau_k) C_j^{\mathrm{mod}}(\tau_{k-\mu})\}$ and $E\{C_j^{\mathrm{mod}}(\tau_{k-1}) C_j^{\mathrm{mod}}(\tau_{k-\mu})\}$. The optimal values A_j^0, $\{g_j^0\}$, and $\{h_j^0\}$ of parameters A_j, $\{g_j\}$, and $\{h_j\}$ are the solution of the following optimization problem,

$$\{A_j^0, \{g_j^0\}, \{h_j^0\}\} = \arg \min_{A_j \in \mathbb{R}, \{g_j\} \in \mathcal{C}_{g_j}, \{h_j\} \in \mathcal{C}_{h_j}} J_j(A_j, \{g_j\}, \{h_j\}). \quad (51)$$

This optimization problem is solved with the trust-region-reflective algorithm with constraints, for which the initial values are $A_j^{\mathrm{ini}} = 0$, $g_j^{\mathrm{ini}} = 0$, and $h_j^{\mathrm{ini}} = 10^{-12}$. The solution $\{[A^0], \{\mathbf{g}^0\}, \{[h^0]\}\}$ of the simplified problem is used as initial value to solve the optimization problem described in Eq. (30).

References

[1] Y. Mizuno, Y. Fujino, K. Kataoka, Y. Matsumoto, Development of a Mobile Sensing Unit and its prototype implementation, Tsinghua Science & Technology 13 (2008) 223–227.

[2] P. A. Ferreira, Modelling and Prediction of the Dynamic Behaviour of Railway Infrastructures at very high Speeds, Ph.D. thesis, Universidade Tecnica de Lisboa (2010).

[3] J. Chaolong, X. Weixiang, W. Futian, W. Hanning, Track Irregularity Time Series Analysis and Trend Forecasting, Discrete Dynamics in Nature and Society 2012. `doi:10.1155/2012/387857`.

[4] J. Andrews, D. Prescott, F. D. Rozières, A stochastic model for railway track asset management, Reliability Engineering and System Safety 130 (2014) 76–84. `doi:10.1016/j.ress.2014.04.021`.

[5] P. A. Ferreira, A. Lopez-Pita, Numerical modelling of high speed train/track system for the reduction of vibration levels and maintenance needs of railway tracks, Construction and Building Materials 79 (2015) 14–21.

[6] N. Lestoille, C. Soize, G. Perrin, C. Funfschilling, Long time evolution of train dynamics with respect to track geometry, in: The Second International Conference on Railway Technology: Research, Development and Maintenance, 2014.

[7] N. Lestoille, C. Soize, C. Funfschilling, Sensitivity of train stochastic dynamics to long-term evolution of track irregularities, Vehicle System Dynamics 54 (5) (2016) 545–567. `doi:10.1080/00423114.2016.1142095`.

[8] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Identification of polynomial chaos representations in high dimension from a set of realizations, SIAM Journal on Scientific Computing 34 (6) (2012) A2917–A2945.

[9] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Track irregularities stochastic modeling, Probabilistic Engineering Mechanics 34 (2013) 123–130.

[10] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Quantification of the influence of the track geometry variability on the train dynamics, Mechanical Systems and Signal Processing (60-61) (2015) 945–957. `doi:10.1016/j.ymssp.2015.01.004`.

[11] N. Lestoille, C. Soize, C. Funfschilling, Stochastic modeling of train dynamics under effect of track irregularities and experimental comparisons, in: International Conference on Uncertainty in Structural Dynamics, 2014.

[12] C. Soize, Identification of high-dimension polynomial chaos expansions with random coefficients for non-Gaussian tensor-valued random fields using partial and limited experimental data, Computer Methods in Applied Mechanics and Engineering 199 (33-36) (2010) 2150–2164. `doi:10.1016/j.cma.2010.03.013`.

[13] G. Perrin, Random fields and associated statistical inverse problems for uncertainty quantifications - Applications to railway track geometries for high-speed trains dynamical responses and risk assessment, Ph.D. thesis, Universite Paris Est (2013).

[14] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Karhunen-Loève expansion revisited for vector-valued random fields: scaling, errors and optimal basis, Journal of Computational Physics 242 (2013) 607–622.

[15] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, A posteriori error and optimal reduced basis for stochastic processes defined by a finite set of realizations, SIAM/ASA Journal of Uncertainty Quantification 2 (2014) 745–762.

[16] UIC, Testing and approval of railway vehicles from the point of view of their dynamic behaviour - Safety - Track fatigue - Running behaviour, UIC Leaflet 518.

[17] C. Soize, Stochastic Models of Uncertainties in Computational Mechanics., American Society of Civil Engineers, Reston, VA, USA, 2012.

[18] R. Ghanem, P. D. Spanos, Stochastic Finite Elements: A Spectral Approach, rev. ed., Dover Publications, New York, 2003.

[19] O. Le Maitre, O. Knio, Spectral Methods for Uncertainty Quantification, Springer, 2010.

[20] M. B. Priestley, Spectral Analysis and Time Series, Academic Press, New York, 1981.

[21] J. D. Hamilton, Time Series Analysis, Princeton University Press, Princeton, USA, 1994.

[22] M. B. Priestley, Non-linear and non-stationary time series analysis, Academic Press, London, 1988.

[23] R. Dahlhaus, Fitting time series models to nonstationary processes, The annals of statistics 25 (1) (1997) 1–37.

[24] B. Pfaff, Analysis of integrated and cointegrated time series with R, 2nd Edition, Springer, New York, 2008.

[25] J. Kaipo, E. Somersalo, Statistical Computational Inverse problems, Springer, New York, 2005.

[26] B. P. Carlin, T. A. Louis, Bayesian Methods for Data Analysis, Thrid Edition, CRC Press, Boca Raton, 2009.

[27] A. C. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1989.

[28] G. Evensen, Data Assimilation: The Ensemble Kalman Filter, 2nd Edition, Springer-Verlag, Berlin, Heidelberg, 2009.