# Track irregularities stochastic modeling

Guillaume Perrin, Christian Soize, Denis Duhamel, Christine Fünfschilling

# Track Irregularities Stochastic Modeling

G. Perrin[a,b,c,*], C. Soize[b], D. Duhamel[a], C. Funfschilling[c]

[a]*Université Paris-Est, Laboratoire Navier (UMR 8205), CNRS, ENPC, IFSTTAR, F-77455 Marne-la-Vallée.*
[b]*Université Paris-Est, Modélisation et Simulation Multi-Échelle (MSME UMR 8208 CNRS), 5 Bd. Descartes, 77454 Marne-la-Vallée, France.*
[c]*SNCF, Innovation and Research Department, Immeuble Lumière, 40 avenue des Terroirs de France, 75611, Paris, Cedex 12, France.*

## Abstract

High speed trains are currently meant to run faster and to carry heavier loads, while being less energy consuming and still respecting the security and comfort certification criteria. To face these challenges, a better understanding of the interaction between the dynamic train behavior and the track geometry is needed. As during its lifecycle, the train faces a great variability of track conditions, this dynamic behavior has indeed to be characterized on track portions sets that are representative of the whole railway network. This paper is thus devoted to the development of a stochastic modeling of the track geometry and its identification with experimental measurements. Based on a spatial and statistical decomposition, this model allows the spatial and statistical variability and dependency of the track geometry to be taken into account. Moreover, it allows the generation of realistic track geometries that are representative of a whole railway network. First, this paper describes a practical implementation of the proposed method and then applies this method to the modeling of a particular French high speed line, for which experimental data are available.

*Keywords:*
Karhunen-Loève expansion, Polynomial Chaos Expansion, Random fields, Railway Track Geometry.

*Corresponding author
Email addresses:* `guillaume.perrin@enpc.fr` (G. Perrin), `christian.soize@univ-paris-est.fr` (C. Soize), `denis.duhamel@enpc.fr` (D. Duhamel), `christine.funfschilling@sncf.fr` (C. Funfschilling)

## 1. Introduction

The expected benefits of simulation in the railway field are multiple: robust and optimized conception, shorter and cheaper certification procedure, better knowledge of the critical situations of the track/vehicle system, optimization of the maintenance. However, if simulation is introduced in certification and conception processes, it has to be very representative of the physical behavior of the system. The model has thus to be fully validated and the simulations have to be raised on a realistic and representative set of excitations.

Two kinds of inputs are traditionally introduced in a railway simulation: the vehicle model $\mathcal{V}$ and the track geometry $\mathcal{T}$. Multibody simulations are usually employed to model the train dynamics. Carbodies, bogies and wheelsets are therefore modeled by rigid bodies and are linked with connections represented by rheologic models (damper, springs, ...).

Two description scales can then be distinguished for the track geometry. On the first hand, the track design, which corresponds to the mean line position of a perfect track is decided once for all at the building of a new track for economical and political reasons. This description is characterized by three curvilinear quantities: the vertical curvature $c_V$, the horizontal curvature $c_H$, and the track superelevation $c_L$. On the other hand, for a fixed track design, the actual positions of the rails are in constant evolution, which is mostly due to the interactions between the train, the track and the substructure. The irregularities appearing during the track lifecycle are indeed of four types (see Figure 1): lateral and vertical alignment irregularities $\mathfrak{X}_1$ and $\mathfrak{X}_2$ on the one hand, cant deficiencies $\mathfrak{X}_3$ and gauge irregularities $\mathfrak{X}_4$ on the other hand. Therefore, each rail position $\boldsymbol{R}_{\ell/r}$ ($\ell$ refers to the left rail whereas $r$ refers to the right rail) can be written as the sum of a mean position $\boldsymbol{M}_{\ell/r}$, which only depends on the curvilinear abscissa of the track, $s$, the track gauge $E$, and the three parameters of the track design, $c_H$, $c_V$ and $c_L$, and a deviation toward this mean position $\boldsymbol{I}_{\ell/r}$, which only depends on the track irregularities:

$$\boldsymbol{R}_{\ell/r}(s) = \boldsymbol{M}_{\ell/r}(s) + \boldsymbol{I}_{\ell/r}(s), \tag{1}$$

$$\boldsymbol{M}_{\ell/r}(s) = \boldsymbol{O}_{\mathrm{NT}}(s) \pm \frac{E}{2}\boldsymbol{N}(s), \tag{2}$$
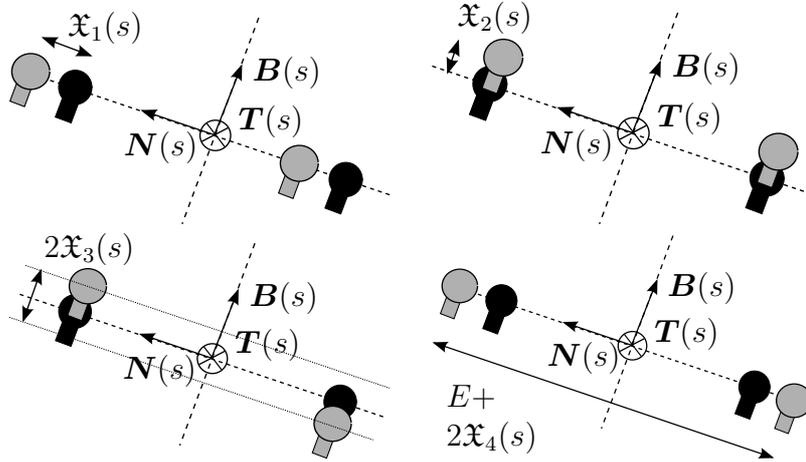
Figure 1: Parametrization of the track irregularities (for each rail, the mean position is represented in black, whereas the real position is in grey).

$$\boldsymbol{I}_{\ell/r}(s) = \left\{ \mathfrak{X}_2(s) \pm \mathfrak{X}_3(s) \right\} \boldsymbol{B}(s) + \left\{ \mathfrak{X}_1(s) \pm \mathfrak{X}_4(s) \right\} \boldsymbol{N}(s), \qquad (3)$$

where $-$ goes with the subscript $\ell$ and $+$ goes with $r$ in the symbol $\pm$, $\boldsymbol{O}_{\mathrm{NT}}(s) = (\boldsymbol{M}_\ell(s) + \boldsymbol{M}_{/r}(s))/2$ is the mean position of the two rails, and $(\boldsymbol{O}_{\mathrm{NT}}(s), \mathbf{T}(s), \mathbf{N}(s), \mathbf{B}(s))$ is the Frenet frame.

Hence, made up of straight lines and curves at its construction, the new track is gradually damaged and regularly subjected to maintenance operations. During their lifecycles, trains are therefore confronted to very different running conditions. The track-vehicle system being strongly non-linear, the dynamic behavior of trains has thus to be analyzed not only on a few track portions but on this whole realm of possibilities.

In reply to these expectations, the measurement train IRIS 320 has been running continuously since 2007 over the French railway network, measuring and recording the track geometry of the main national lines. This systematic measurement of the railway network quality and variability makes up a very useful database to analyze the complex link between the train dynamics and the physical and statistical properties of the track geometry.

Based on these experimental measurements, a complete parametrization of the track geometry and of its variability would be of great concern in specification, security and certification prospects, to monitor more systematically the appearing track irregularities and to be able to generate track geometries

that are realistic and representative of a whole railway network.

To this end, the random irregularities can be statistically characterized using a one-sided power spectral density (PSD) function of the track geometry [1, 2, 3]. In this case, the PSD functions of each track irregularity can be either estimated from the available measured track geometries or from track safety standards such as the ones given by the Federal Railroad Administration [4]. Track irregularities can then by generated from these PSD functions from a time-frequency transformation based on a spectral representation (see [5, 6] for further details). In such a method, the railway track irregularities are thus considered as four independent Gaussian random fields, which are supposed to be stationary and ergodic in space. By definition, these track irregularities are however strongly dependent, and due to the specific interaction between the train and the track, they are actually neither Gaussian nor stationary.

In this context, the present work proposes a method to identify, in inverse, from measured data, the statistical properties of this vector-valued, non-Gaussian and non-stationary track irregularity random field. These data being complete, this modeling allows us to generate numerically track geometries that are physically realistic and statistically representative of the set of available track measurements, while taking into account the dependencies between the four kinds of track irregularities.

These tracks can then be used in any deterministic railway dynamic software to characterize the dynamic behavior of the train. Hence, this modeling could bring an innovative technical answer to introduce numerical methods and treatments in the maintenance and certification processes.

In Section 2, the stochastic modeling of the track geometry from experimental data and the hypothesis on which this modeling is based are described in details. Section 3 presents then the spectral and statistical validation of the method.

## 2. Track geometry stochastic modeling

The mean line of the track geometry being chosen at the building of a new line, this work is only devoted to the modeling of the track irregularity vector

$$\mathfrak{X} = (\mathfrak{X}_1, \mathfrak{X}_2, \mathfrak{X}_3, \mathfrak{X}_4), \tag{4}$$

where $\mathfrak{X}_1$, $\mathfrak{X}_2$, $\mathfrak{X}_3$ and $\mathfrak{X}_4$ are the four types of track irregularities previously

introduced. In this work, it is supposed that the track geometry of a given railway network has been measured. The track irregularity vector,

$$\left\{ \mathfrak{X}(s) = (\mathfrak{X}_1(s), \mathfrak{X}_2(s), \mathfrak{X}_3(s), \mathfrak{X}_4(s)), \ s \in [0, S^{\text{tot}}] \right\}, \tag{5}$$

is assumed to be a second-order centered random field, such that:

$$E\left[\mathfrak{X}(s)\right] = \mathbf{0}, \quad s \in [0, S^{\text{tot}}], \tag{6}$$

with $E\left[\cdot\right]$ the mathematical expectation. Due to the specific interaction between the train and the track, it is recalled that this random field is neither Gaussian nor stationary.

To characterize the statistical properties of such a random field, this work proposes a three-step method. First, a local-global approach is introduced, which allows us to focus the stochastic modeling only on the projection of $\mathfrak{X}$ on the interval $\Omega = [0, S]$, with $S \ll S^{\text{tot}}$. Then, a Karhunen-Loève (KL) expansion is carried out, which allows us to write $\mathfrak{X}$ as a weighted sum of deterministic spatial functions. The identification of the multidimensional distribution of these weights, which are non correlated but *a priori* dependent random variables, is finally performed using a polynomial chaos expansion (PCE). At last, it will be shown in what extent such a method allows the generation of realistic track geometries that are representative of the measured railway network.

*2.1. Local-global approach*

The non Gaussian and non stationary properties of track irregularity random field $\mathfrak{X}$ motivate the introduction of a local-global approach for the characterization of its distribution. This approach is based on the hypothesis that a whole railway track can be considered as the concatenation of a series of independent track portions of same length $S$, for which physical and statistical properties are the same. Length $S$ plays therefore a key role in the modeling procedure, and its value has to be carefully evaluated. In order to choose length $S$ such that its influence on the stochastic modeling is minimized, $\nu$ track portions of same length $L = 10km$, $\left\{ \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\nu)} \right\}$, have been collected from the available measurements of the railway network of interest. For any value for $S$, we denote by $\left\{ \boldsymbol{y}^{(1)}(S), \ldots, \boldsymbol{y}^{(\nu)}(S) \right\}$ the $\nu$

new track geometries of total length $L$, which are then built from the concatenation of track sub-sections of length $S$ that have been randomly chosen in $\left\{ \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\nu)} \right\}$.

Three errors functions, for which definitions can be found in Appendix A, are therefore introduced in this paper to quantify the influence of $S$:

- a covariance error, $err^2_{\mathrm{cov}}(S)$: fixing $S$ to a particular value amounts to supposing that for $|s - s'| \geq S$, $E\left[ \mathfrak{X}(s)\mathfrak{X}(s')^T \right]$ is negligible;

- a spectral error, $err^2_{\mathrm{spect}}(S)$: generating complete track geometries from the concatenation of several track portions of length $S$ introduces an artificial periodicity and is likely to degrade the low frequency characterization of $\mathfrak{X}$;

- an estimation error, $err^2_{\mathrm{est}}(S)$: the higher $S$ is, the smaller the number of independent realizations for $\mathfrak{X}$, $\nu^{\mathrm{exp}}(S)$, can be extracted from the complete measurement of the railway network of interest. This error is therefore directly related to the estimation accuracy of the covariance function of $\mathfrak{X}$, and to the identification precision of the PCE coefficients, on which the modeling will then be based. With reference to the Central Limit Theorem (see [7] for further details), we simply choose $err^2_{\mathrm{est}}(S) = 1/\sqrt{\nu^{\mathrm{exp}}(S)}$ to illustrate this phenomenon.

For the chosen railway network, based on these sets of track geometries of same lengths $L = 10km$, errors $err^2_{\mathrm{cov}}(S)$, $err^2_{\mathrm{spect}}(S)$ and $err^2_{\mathrm{est}}(S)$ are represented in Figure 2. When $S$ increases, it can be verified that $err^2_{\mathrm{cov}}(S)$ and $err^2_{\mathrm{spect}}(S)$ decrease whereas $err^2_{\mathrm{est}}(S)$ increases. Length $S$ has thus to be chosen as the right balance between these three error functions.

For confidentiality reasons, the exact value of $S$ is however not given in this work, and the spatial quantities will be normalized by length $S$ in the following.

Under this local-global hypothesis, $\nu^{\mathrm{exp}} = 1,730$ track portions of same length $S$ are now extracted from the complete railway network of total length $S^{\mathrm{tot}}$. These measurements are supposed to be $\nu^{\mathrm{exp}}$ independent realizations of random field $\mathfrak{X}$, which defines the maximum available information for the stochastic modeling. These portions are in practice discretized at a spatial
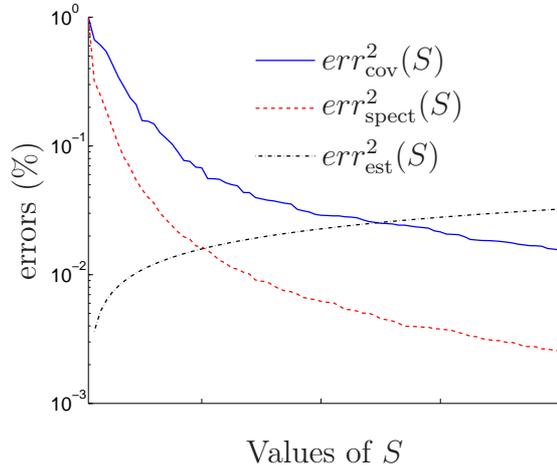
Figure 2: Graphs of errors $err^2_{\mathrm{cov}}(S)$, $err^2_{\mathrm{spect}}(S)$ and $err^2_{\mathrm{est}}(S)$ for the computation of the local-global length $S$.

step $h$ that is sufficiently small for the statistical and spectral information of $\mathfrak{X}$ to be completely taken into account. The experimental discretized data are thus assumed to be complete. Hence, random field $\mathfrak{X}$ can be identified to its $N_S$-dimension discretized projection $\boldsymbol{X}$, with $N_S = 4(S/h + 1)$, such that:

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \boldsymbol{X}_3 \\ \boldsymbol{X}_4 \end{pmatrix}, \quad \boldsymbol{X}_t = \begin{pmatrix} \mathfrak{X}_t(0) \\ \mathfrak{X}_t(h) \\ \vdots \\ \mathfrak{X}_t(S) \end{pmatrix}, \ 1 \le t \le 4. \tag{7}$$

The random vector $\boldsymbol{X}$ is still called irregularity vector, whereas measured track portions are denoted by $\left\{ \boldsymbol{x}^1, \cdots, \boldsymbol{x}^{\nu^{\mathrm{exp}}} \right\}$ and will now be considered as $\nu^{\mathrm{exp}}$ independent realizations of random vector $\boldsymbol{X}$. Therefore, the covariance matrix, $[C_{\boldsymbol{XX}}]$, of $\boldsymbol{X}$ can be estimated, for $\nu^{\mathrm{exp}}$ sufficiently large, by:

$$[C_{\boldsymbol{XX}}] = E\left[\boldsymbol{X}\boldsymbol{X}^T\right]$$
$$\approx [\widehat{C}_{\boldsymbol{XX}}] = \frac{1}{\nu^{\mathrm{exp}}} \sum_{i=1}^{\nu^{\mathrm{exp}}} \boldsymbol{x}^{(i)} \left(\boldsymbol{x}^{(i)}\right)^T. \tag{8}$$

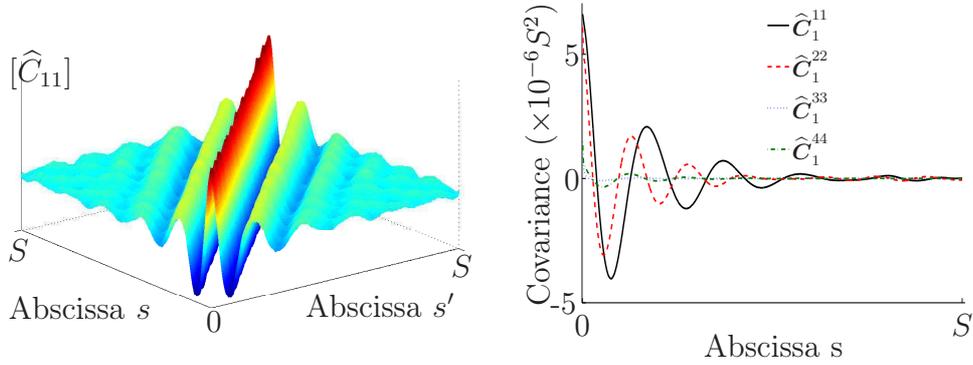Approximation $[\widehat{C}_{11}]$ of $[C_{11}] = E\left[\boldsymbol{X}_1\boldsymbol{X}_1^T\right]$ is presented in Figure 3. This

7

Figure 3: Graphs of projections of $[\widehat{C}_{\boldsymbol{XX}}]$.

figure emphasizes a quasi symmetry along the first bisector. The first column, $\widehat{\boldsymbol{C}}_1^{ij}$, of matrices $[\widehat{C}_{ij}] = [\widehat{\boldsymbol{C}}_1^{ij} \; \cdots \; \widehat{\boldsymbol{C}}_{N_S/4}^{ij}]$ can thus be used to condense the covariance information of each block. Figure 3 compares also the four first columns, $\widehat{\boldsymbol{C}}_1^{tt}$, $1 \le t \le 4$, of the diagonal blocks of $[\widehat{C}_{\boldsymbol{XX}}]$. It can be noticed that the covariance matrices are very different from one track irregularity to another.

*2.2. KL and PCE expansions*

The objective of the stochastic modeling is to identify in inverse the statistical properties of irregularity vector $\boldsymbol{X}$ from the $\nu^{\text{exp}}$ independent realizations $\{\boldsymbol{x}^1, \cdots, \boldsymbol{x}^{\nu^{\text{exp}}}\}$. A two-step approach is introduced in this work to characterize the distribution of $\boldsymbol{X}$. First, a spatial and statistical representation is achieved, which is based on a Karhunen-Loève (KL) expansion. This very efficient method, which has first been introduced by Pearson [8] in data analysis, has been applied in many works for the last decades (see, for instance [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27]). It is indeed particularly interesting as it allows the uncorrelation of the projection coefficients of $\boldsymbol{X}$ on the KL vector basis, while optimally compacting the signal energy. The Karhunen-Loève expansion (see [28, 29] for further details) and more precisely the Principle Component Analysis of centered random vector $\boldsymbol{X}$ can be written as:

$$\boldsymbol{X} = \sum_{k=1}^{N_S} \sqrt{\widehat{\lambda}_k}\,\widehat{\boldsymbol{u}}^k\,\eta_k, \qquad (9)$$

8

in which $\left\{\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_{N_S}\right\}$ are the eigenvalues and $\left\{\widehat{\boldsymbol{u}}^1, \cdots, \widehat{\boldsymbol{u}}^{N_S}\right\}$ are the associated eigenvectors, which are solutions of the eigenvalue problem:

$$[\widehat{C}_{\boldsymbol{XX}}]\widehat{\boldsymbol{u}} = \widehat{\lambda}\widehat{\boldsymbol{u}}, \tag{10}$$

in which matrix $[\widehat{C}_{\boldsymbol{XX}}]$ is defined by Eq. (8). Projection basis $\left\{\widehat{\boldsymbol{u}}^1, \cdots, \widehat{\boldsymbol{u}}^{N_S}\right\}$ is chosen orthonormal in the sense that, for all $k$ and $\ell$:

$$\left(\widehat{\boldsymbol{u}}^\ell\right)^T \widehat{\boldsymbol{u}}^k = \delta_{kl}, \tag{11}$$

where $\delta_{kl}$ is the Kronecker symbol, equal to one if $k = \ell$ and zero otherwise. In the same manner than the track irregularity vector is the concatenation of four random sub-vectors corresponding to the four track irregularities, such that $\boldsymbol{X} = (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3, \boldsymbol{X}_4)$, the components of $\widehat{\boldsymbol{u}}^k$ can be divided in four sub-vectors of same dimension $N_S/4$, which are denoted by $\widehat{\boldsymbol{u}}_1^k$, $\widehat{\boldsymbol{u}}_2^k$, $\widehat{\boldsymbol{u}}_3^k$ and $\widehat{\boldsymbol{u}}_4^k$. It has to be noticed that, for a particular eigenvector $\widehat{\boldsymbol{u}}^k$, if the norms of at least two sub-vectors, $(\widehat{\boldsymbol{u}}_t^k)^T\widehat{\boldsymbol{u}}_t^k$ and $(\widehat{\boldsymbol{u}}_r^k)^T\widehat{\boldsymbol{u}}_r^k$, for $1 \leq t < r \leq 4$, are non negligible, then, a spatial dependency is introduced in the modeling between the track irregularities $t$ and $r$. We compare in Figure 4 several graphs of eigenvectors $\widehat{\boldsymbol{u}}^k$. For a better visualization, the mean values of the different sub-vectors, which are zero, are deliberately translated.

The coefficients $\{\eta_1, \cdots, \eta_{N_S}\}$, such that:

$$\eta_k = \frac{\boldsymbol{X}^T\widehat{\boldsymbol{u}}^k}{\sqrt{\widehat{\lambda}_k}}, \tag{12}$$

are moreover uncorrelated random variables, for which mean is zero and variance is equal to one. Based on the eigenvalues decrease, random vector $\boldsymbol{X}$ can then be approximated by its truncated expansion with $N_\eta$ terms, $\boldsymbol{X}^{(N_\eta)}$, such that:

$$\boldsymbol{X} = \sum_{k=1}^{N_\eta} \sqrt{\widehat{\lambda}_k}\widehat{\boldsymbol{u}}^k \, \eta_k, \tag{13}$$

for which the truncation residual error, $\boldsymbol{X} - \boldsymbol{X}^{(N_\eta)}$, is evaluated by the normalized $L_2$ error, $\varepsilon_{KL}^2(N_\eta)$, such that:
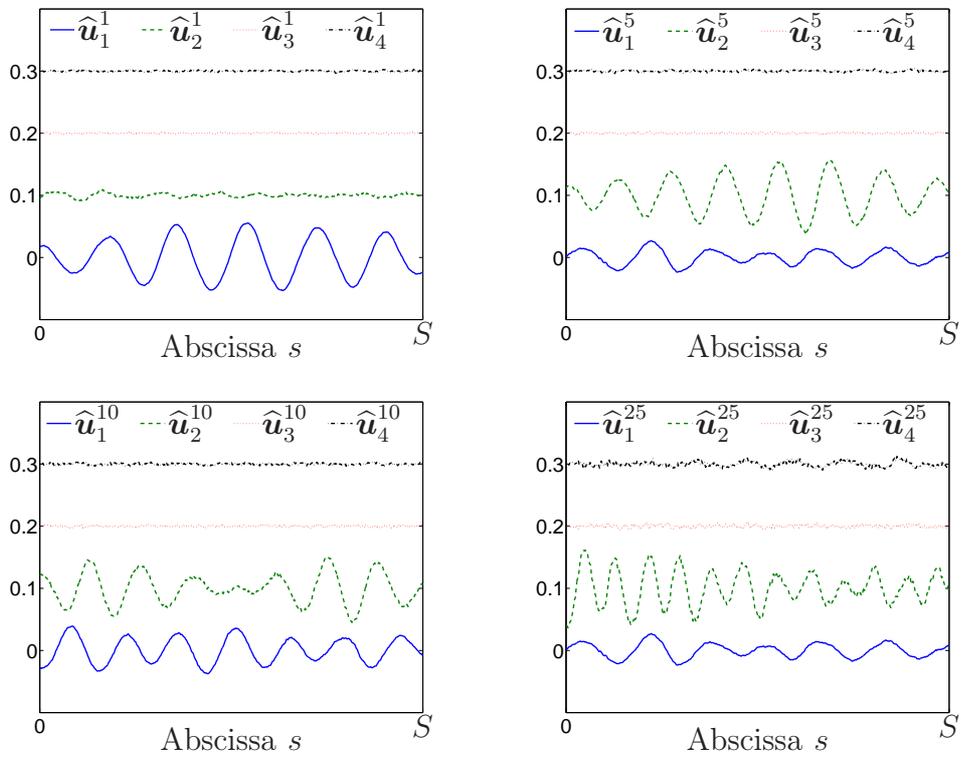
Figure 4: Graphs of four particular eigenvectors $\widehat{\boldsymbol{u}}^1$, $\widehat{\boldsymbol{u}}^5$, $\widehat{\boldsymbol{u}}^{10}$ and $\widehat{\boldsymbol{u}}^{25}$ ($\times 10^{-3}S$).
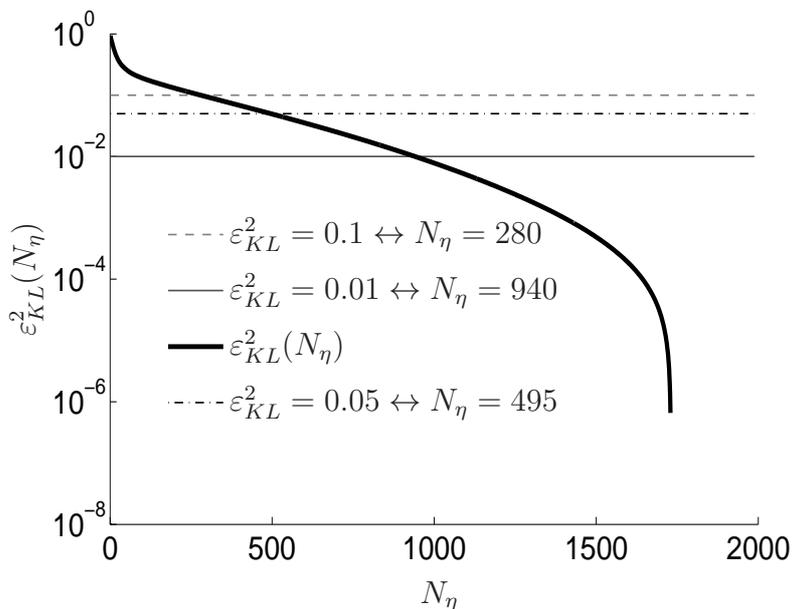
Figure 5: Graph of $\varepsilon_{KL}^2(N_\eta)$.

$$\varepsilon_{KL}^2(N_\eta) = \frac{\left\|\boldsymbol{X} - \boldsymbol{X}^{(N_\eta)}\right\|_2^2}{\|\boldsymbol{X}\|_2^2} = 1 - \frac{\sum_{k \leq N_\eta} \lambda_k}{\|\boldsymbol{X}\|_2^2}, \tag{14}$$

where for all second order random vector $\boldsymbol{Z}$,

$$\|\boldsymbol{Z}\|_2^2 = E\left[\boldsymbol{Z}^T \boldsymbol{Z}\right]. \tag{15}$$

The evolution of error $\varepsilon_{KL}^2(N_\eta)$ with respect to $N_\eta$ is plotted in Figure 5. The higher $N_\eta$, the more precise the characterization of the track geometry, but the more difficult the characterization of the random vector $\boldsymbol{\eta} = \left(\eta_1, \ldots, \eta_{N_\eta}\right)$. As a good compromise, the truncation parameter $N_\eta$ is fixed to the value 940 in the following, which corresponds to an error threshold of 1% for $\varepsilon_{KL}^2$.

The second step of the modeling of $\boldsymbol{X}$ is the characterization of the multidimensional probability density function (PDF) of $\boldsymbol{\eta}$, which is denoted by $p_{\boldsymbol{\eta}}$. Two kinds of methods can be used to build such a PDF: the direct and the indirect methods. Among the direct methods, methods based on

11

Information Theory and the Maximum Entropy Principle (MEP) have been developed (see [30] and [31]) to directly compute $p_{\boldsymbol{\eta}}$ from the only available information of random vector $\boldsymbol{\eta}$, as the solution of an optimization problem. On the other hand, the indirect methods allow the construction of the PDF $p_{\boldsymbol{\eta}}$ of the considered random vector $\boldsymbol{\eta}$ from a transformation $\boldsymbol{t}$ of a known $N_{\eta}$-dimensional random vector $\boldsymbol{\xi}$:

$$\boldsymbol{\eta} = \boldsymbol{t}(\boldsymbol{\xi}), \quad p_{\boldsymbol{\eta}} = \mathbb{T}(p_{\boldsymbol{\xi}}), \tag{16}$$

defining a transformation $\mathbb{T}$ between $p_{\boldsymbol{\eta}}$ and the known PDF $p_{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$. The construction of the transformation $\boldsymbol{t}$ is thus the key point of these indirect methods. Among these methods, the polynomial chaos expansion (PCE) methods, which have first been proposed by Wiener [32], and spread to computational sciences by Ghanem and Spanos [33, 28], are based on a direct projection of random vector $\boldsymbol{\eta}$ on a chosen polynomial Hilbertian basis $\mathcal{B}_{\mathrm{orth}} = \{ \psi_j(\boldsymbol{\xi}), \ 1 \leq j \}$ of all the second-order random vectors with values in $\mathbb{R}^{N_{\eta}}$, such that:

$$\boldsymbol{\eta} = \sum_{j=1}^{+\infty} \boldsymbol{y}^{(j)} \psi_j(\boldsymbol{\xi}). \tag{17}$$

In the last decades, this very promising method has therefore been applied in many works (see [34, 35, 36, 37, 38, 39, 40, 41, 42] for more details about the inverse PCE identification from experimental data). For practical purposes, the sum that is defined by Eq. (17) has however to be truncated. We therefore define $\{\psi_1(\xi_1, \ldots, \xi_{N_g}), \cdots, \psi_N(\xi_1, \ldots, \xi_{N_g})\}$ as the $N$-dimensional subset of $\mathcal{B}_{\mathrm{orth}}$ that gathers the polynomial functions of total degree inferior to $p$, which are normalized with respect to the PDF $p_{\xi_1, \ldots, \xi_{N_g}}$ of the $N_g$ first elements of random vector $\boldsymbol{\xi}$, such that:

$$\boldsymbol{\eta} \approx \boldsymbol{\eta}^{\mathrm{chaos}}(N) = \sum_{j=1}^{N} \boldsymbol{y}^{(j)} \psi_j(\xi_1, \ldots, \xi_{N_g}), \tag{18}$$

$$\psi_j(\xi_1, \ldots, \xi_{N_g}) = \sum_{q=1}^{N} c_{qj} \, \xi_1^{\alpha_1^{(q)}} \times \cdots \times \xi_{N_g}^{\alpha_{N_g}^{(q)}}, \quad 0 \leq \alpha_1^{(q)} + \ldots + \alpha_{N_g}^{(q)} \leq p, \tag{19}$$

$$\int_{\mathbb{R}^{N_g}} \psi_j(\boldsymbol{x}) \psi_n(\boldsymbol{x}) p_{\xi_1, \ldots, \xi_{N_g}}(\boldsymbol{x}) d\boldsymbol{x} = \delta_{jn}. \tag{20}$$

12

Given values of $N$ and $N_g$, characterizing $p_{\boldsymbol{\eta}}$ amounts finally to identifying the deterministic projection vectors $\{\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(N)}\}$ from the available information for $\boldsymbol{\eta}$. According to Eq. (12), this available information corresponds to the $\nu^{\text{exp}}$ independent realizations of $\boldsymbol{\eta}^{\text{chaos}}(N)$, $\{\boldsymbol{\eta}^1, \cdots, \boldsymbol{\eta}^{\nu^{\text{exp}}}\}$, which can be deduced from the $\nu^{\text{exp}}$ independent realizations of $\boldsymbol{X}$, $\{\boldsymbol{x}^1, \cdots, \boldsymbol{x}^{\nu^{\text{exp}}}\}$, as:

$$\eta_k^i = \frac{1}{\sqrt{\widehat{\lambda}_k}} \left(\boldsymbol{x}^i\right)^T \widehat{\boldsymbol{u}}^k, \quad 1 \leq i \leq \nu^{\text{exp}}, \quad 1 \leq k \leq N_\eta. \tag{21}$$

In [40, 41], it has been shown that a good approach to identify such coefficients in high dimension (that is to say when $N_\eta$ is high) is to search them as the arguments that maximize the likelihood of random vector $\boldsymbol{\eta}^{\text{chaos}}(N)$ at the experimental points $\{\boldsymbol{\eta}^1, \cdots, \boldsymbol{\eta}^{\nu^{\text{exp}}}\}$.

Finally, the last step of this identification is the justification of the values of the truncation parameters $N$ and $N_g$. In this prospect, the log-error function $err(N, N_g)$ is introduced to quantify the amplitude of the residual error of the PCE truncation, $\boldsymbol{\eta} - \boldsymbol{\eta}^{\text{chaos}}(N)$, such that:

$$err(N, N_g) = \sum_{k=1}^{N_\eta} err_k(N, N_g), \tag{22}$$

$$err_k(N, N_g) = \int_{\text{BI}_k} \left| log_{10}\left(p_{\eta_k}(x_k)\right) - log_{10}\left(p_{\eta_k^{\text{chaos}}}(x_k)\right) \right| dx_k, \tag{23}$$

where $\text{BI}_k$ is the domain bounding the experimental values of $\eta_k$, $p_{\eta_k}$ and $p_{\eta_k^{\text{chaos}}}$ are the PDFs of $\eta_k$ and $\eta_k^{\text{chaos}}(N)$ respectively. Truncation parameters $N$ and $N_g$ can thus be chosen with respect to a given error threshold for $err(N, N_g)$.

For our study, $\boldsymbol{\xi}$ is a $N_\eta$-dimension random vector, whose components are independent and uniformly distributed between -1 and 1. According to Figure 6, where $err(N, N_g)$ is plotted against $N$ for different values of $N_g$, truncation parameters $N_g$ and $N$ are chosen equal to 3 and 3,276 respectively, which corresponds to the maximal polynomial order $p = 25$ for the reduced polynomial basis, $\{\psi_1(\xi_1, \ldots, \xi_{N_g}), \cdots, \psi_N(\xi_1, \ldots, \xi_{N_g})\}$.

### 2.3. Generation of a whole track geometry

Once truncation parameters $N_\eta$, $N$, $N_g$ have been identified according to convergence analysis, and PCE projection coefficients $\{\boldsymbol{y}^{(j)}, \ 1 \leq j \leq N\}$
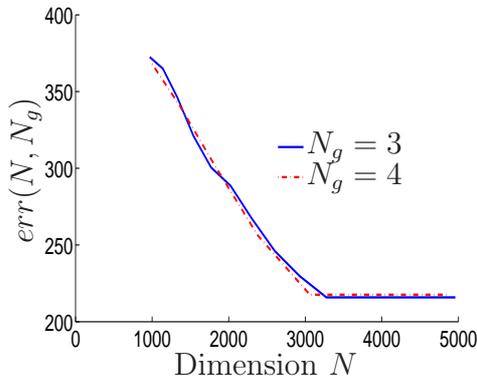
Figure 6: Convergence analysis for the truncated PCE of $\boldsymbol{\eta}$.

have been computed with the advanced algorithms described in [41] and [40], the track irregularity random field is completely characterized and can finally be estimated from Eqs. (13) and (18) as:

$$\boldsymbol{X} \approx \sum_{k=1}^{N_\eta} \sqrt{\widehat{\lambda}_k} \widehat{\boldsymbol{u}}^k \sum_{j=1}^{N} y_k^{(j)} \psi_j(\xi_1, \ldots, \xi_{N_g}). \tag{24}$$

For each realization of random vector $(\xi_1, \ldots, \xi_{N_g})$, a representative and realistic track geometry of length $S$ can finally be generated. Thanks to the local-global approach, described in Section 2.1, a whole track geometry of length $S^{\text{tot}} = N_{\mathcal{T}} S$ ($N_{\mathcal{T}}$ can be smaller or greater than $\nu^{\text{exp}}$), $\boldsymbol{X}^{\text{tot}}$, can therefore be constructed from $N_{\mathcal{T}}$ independent copies $\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(N_{\mathcal{T}})}$ of track irregularity vector $\boldsymbol{X}$, such that $\boldsymbol{X}^{\text{tot}} = (\boldsymbol{X}^{(1)}, \cdots, \boldsymbol{X}^{(N_{\mathcal{T}})})$.

Therefore, $\nu$ independent realizations $\left\{ \boldsymbol{X}^{\text{tot}}(\theta_1), \cdots, \boldsymbol{X}^{\text{tot}}(\theta_\nu) \right\}$ of $\boldsymbol{X}^{\text{tot}}$ can be generated from $\nu N_{\mathcal{T}}$ realizations of irregularity vector $\boldsymbol{X}$. However, for each particular realization $\boldsymbol{X}^{\text{tot}}(\theta_m)$ of $\boldsymbol{X}^{\text{tot}}$, a particular attention has to be paid to the interface between the different realizations of $\boldsymbol{X}$. Indeed, these junctions have to guarantee the continuity of the track irregularity vector and at least the continuity of its first and second order spatial derivatives in order to avoid an artificial perturbation of the train dynamics. Spline interpolations on a length corresponding to the minimal wavelength of the measured irregularities are then used to fulfill these continuity conditions.

Therefore, the proposed stochastic modeling allows us to generate realistic
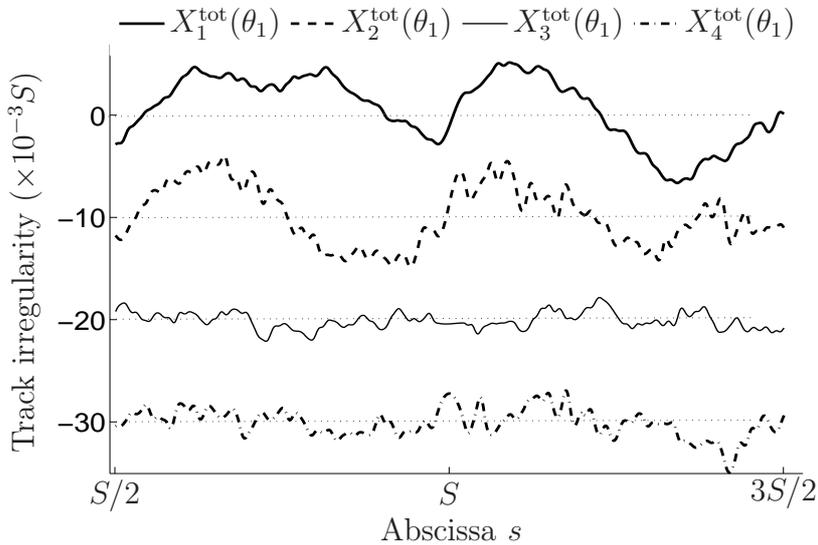
14

Figure 7: Extract of a simulated track geometry.

track geometries of length $S^{\mathrm{tot}} = N_{\mathcal{T}} S$ that are representative of the whole network, and which take into account the spatial and statistical dependencies between the different track irregularities. As an illustration, a particular extract of length $S$ of complete track geometry $\boldsymbol{X}^{\mathrm{tot}}(\theta_1)$ is represented in Figure 7. This graph is centered at abscissa $s = 3S/2$, that is to say at a junction between the two first realizations of random vector $\widetilde{\boldsymbol{X}}$. The four components of $\boldsymbol{X}^{\mathrm{tot}}(\theta_1)$ are represented in the same graph, but their values are translated to allow a better visualization of the results.

## 3. Validation of the track geometry stochastic modeling

As presented in Introduction, a complete parametrization of the physical and statistical properties of the track geometry would be of great interest in conception, certification and maintenance prospects. Several validations of the proposed stochastic modeling are thus presented in this section, in order to allow a relevant investigation of the dynamic interaction between the train and the track.
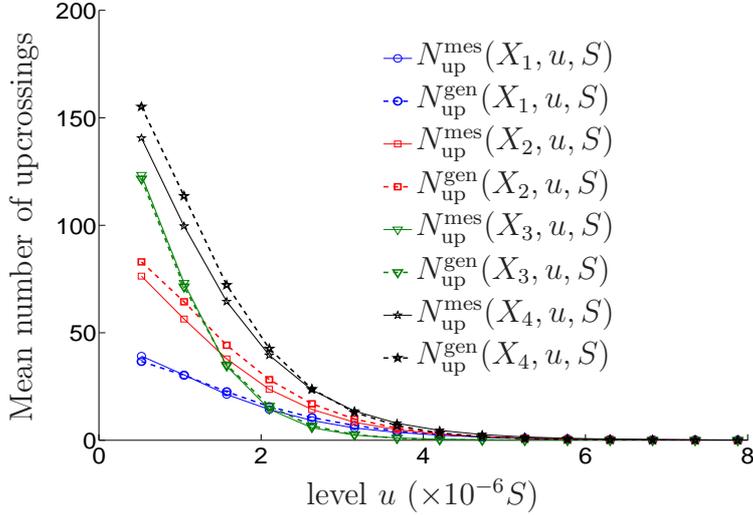
15

Figure 8: Validation of the generation of track geometries with respect to the number of upcrossings.

### 3.1. Statistical and frequency validation of the local track geometry stochastic modeling

In this section, $5,000$ track irregularities vectors of total length $S$ are generated from the track stochastic modeling developed in Section 2. For $1 \leq i \leq 4$, we use $N_{\mathrm{up}}^{\mathrm{mes}}(X_i, u, S)$ and $N_{\mathrm{up}}^{\mathrm{gen}}(X_i, u, S)$ to denote the mean numbers of upcrossings (see [43] for more details about the upcrossings) of the level $u$ by $X_i$ over the length $S$, which have been evaluated from the $\nu^{\mathrm{exp}} = 1,730$ available measured track geometries and from the $5,000$ generated track geometries of length $S$ respectively. In the same manner, for $1 \leq i \leq 4$, let $PSD_i^{\mathrm{mes}}$ and $PSD^{\mathrm{gen}}(X_i)$ be the mean power spectral densities of $X_i$ that have been computed from the former $1,730$ measured and $5,000$ generated track geometries of length $S$.

A good agreement between the quantities corresponding to the measured and to the generated track geometries can be seen in Figures 8 and 9. This allows us to validate the local stochastic modeling over a length $S$ from a statistical and frequency point of view.
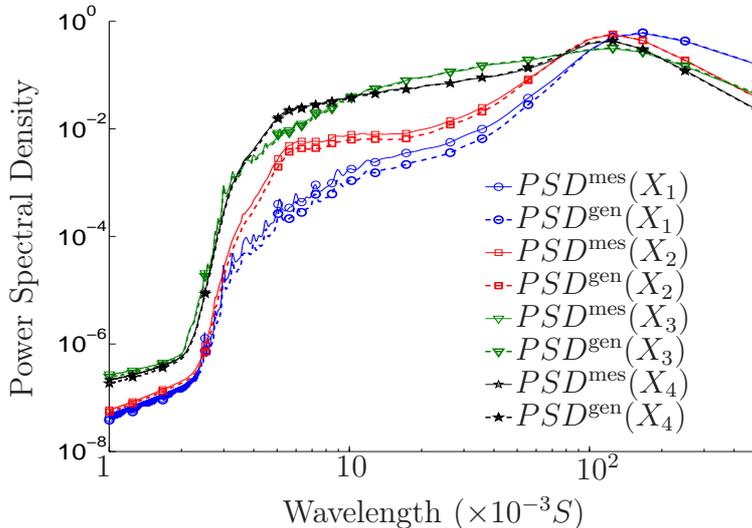
16

Figure 9: Validation of the generation of track geometries with respect to the frequency content.

## 3.2. Dynamical validation of the global track geometry stochastic modeling

This section aims now at validating the complete modeling from a dynamical point of view. In this prospect, a measured track design of length $5km$ around a horizontal curve of a high speed line is considered, for which horizontal and vertical curvatures, $c_H$ and $c_V$, and track superelevation, $c_L$, are represented in Figure 10. $\nu$ track irregularities of total length $5km$, $\left\{ \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\nu)} \right\}$, are then extracted from the available measured track geometry. These track irregularities have also been chosen around a horizontal curve. In parallel, $\nu$ track irregularities, $\left\{ \boldsymbol{X}^{\text{tot}}(\theta_1), \cdots, \boldsymbol{X}^{\text{tot}}(\theta_\nu) \right\}$, of total length $S^{\text{tot}} = 5km$ are generated from the local-global stochastic modeling described in Section 2.

Coupled to the model of a train, these two sets of track conditions can now be used in any rigid-multibodies railway software to characterize the train dynamic behavior. For our study, it is supposed that a normalized model of a TGV train is available, for which mechanical parameters have been accurately identified, and a commercial code, which is called Vampire, has been used to perform the railway simulations.

The validation of the global stochastic modeling is then based on the frequency and the statistical analysis of eight dynamical quantities of interest:
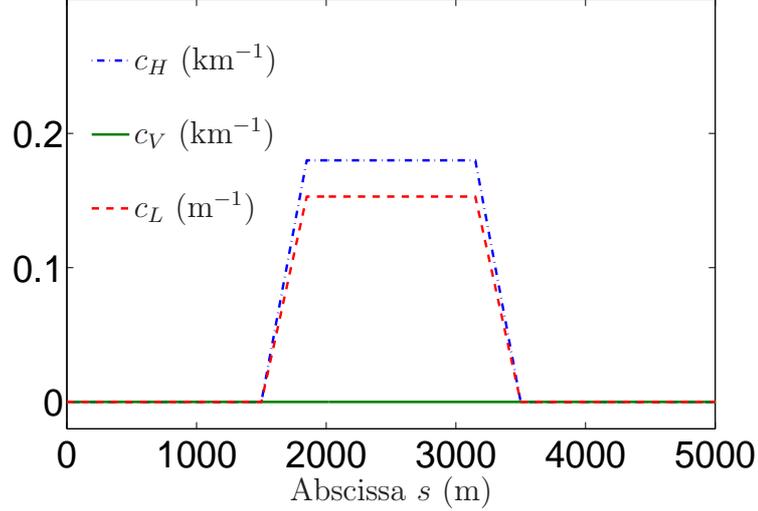
17

Figure 10: Track Design of the simulated railway track.

- the vertical and horizontal accelerations of the first bogie of the motor car, $Q_1 = \ddot{z}_{MC}$ and $Q_2 = \ddot{y}_{MC}$;

- the vertical and horizontal accelerations of the second bogie of the second passenger car, $Q_3 = \ddot{z}_{PC}$ and $Q_4 = \ddot{y}_{PC}$;

- the left and right transverse contact forces at the first wheelset of the first bogie of the motor car, $Q_5 = Y_{MC}^{\ell}$ and $Q_6 = Y_{MC}^{r}$;

- the left and right transverse contact forces at the second wheelset of the second bogie of the second passenger car, $Q_7 = Y_{PC}^{\ell}$ and $Q_8 = Y_{PC}^{\ell}$.

In the same manner than in Section 3.1, for $1 \leq i \leq 8$, we are interested in the mean power spectral densities of $Q_i$ and the mean numbers of upcrossings of the level $u$ by $Q_i$ over the length $S^{\mathrm{tot}} = 5km$, which are respectively denoted by $PSD^{\mathrm{mes}}(Q_i)$ and $N_{\mathrm{up}}^{\mathrm{mes}}(Q_i, u, S^{\mathrm{tot}})$ when these quantities are computed from the measured track geometries and $PSD^{\mathrm{gen}}(Q_i)$ and $N_{\mathrm{up}}^{\mathrm{gen}}(Q_i, u, S^{\mathrm{tot}})$ when these quantities are computed from the generated track geometries. The comparisons between these quantities are represented in Figure 11. In can therefore be deduced from these figures that the stochastic modeling method used in this work can make sure that the frequency and

18

statistical contents of generated track geometries with a local-global length of $S$ are similar to measured ones from the train response point of view.

## 4. Conclusion and prospects

A complete parametrization of the track geometry, which takes into account its physical properties and its variability are nowadays of great interest to be able to face always more challenging railway issues. In this prospect, this paper has presented a general method to model a discretized $\mathbb{R}^4$-valued random field indexed by $s \in [0, S]$ thanks to a double projection, which can be applied to many other mechanical systems. First, a Karhunen-Loève expansion is used to decompose the projection of the discretized random field as a deterministic matrix and a high dimension random vector. The distribution of this high dimension random vector is then characterized thanks to a truncated PCE. This paper moreover describes in details how to control and justify the different truncation parameters. Then, complete track geometries that are realistic and representative of a whole railway network can be generated from a local-global approach. At last, a double validation of this stochastic model is presented, in order to make sure that the frequency and statistical contents of generated and measured track geometries are similar. These geometries can finally be used in any railway software to characterize the dynamic behavior of trains.

## Acknowledgment

## Appendix A: definition of the local-global error functions

It is assumed that $\nu$ track portions of same length $L = 10 km$, $\left\{ \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\nu)} \right\}$, have been collected from the available measurements of the railway network of interest. For any value for $S$, $\nu$ new track geometries, $\left\{ \boldsymbol{y}^{(1)}(S), \ldots, \boldsymbol{y}^{(\nu)}(S) \right\}$, of total length $L$, are then built from the concatenation of track sub-sections of length $S$ that have been randomly chosen in $\left\{ \boldsymbol{z}^{(1)}, \ldots, \boldsymbol{z}^{(\nu)} \right\}$.

For $(s, s')$ in $[0, L]^2$ and $f \geq 1/L$, let $(s, s') \mapsto [R_{zz}(s, s')]$, $(s, s') \mapsto [R_{yy}(s, s', S)]$, $f \mapsto \boldsymbol{\Sigma}_z(f)$ and $f \mapsto \boldsymbol{\Sigma}_y(f, S)$ be the following quantities:
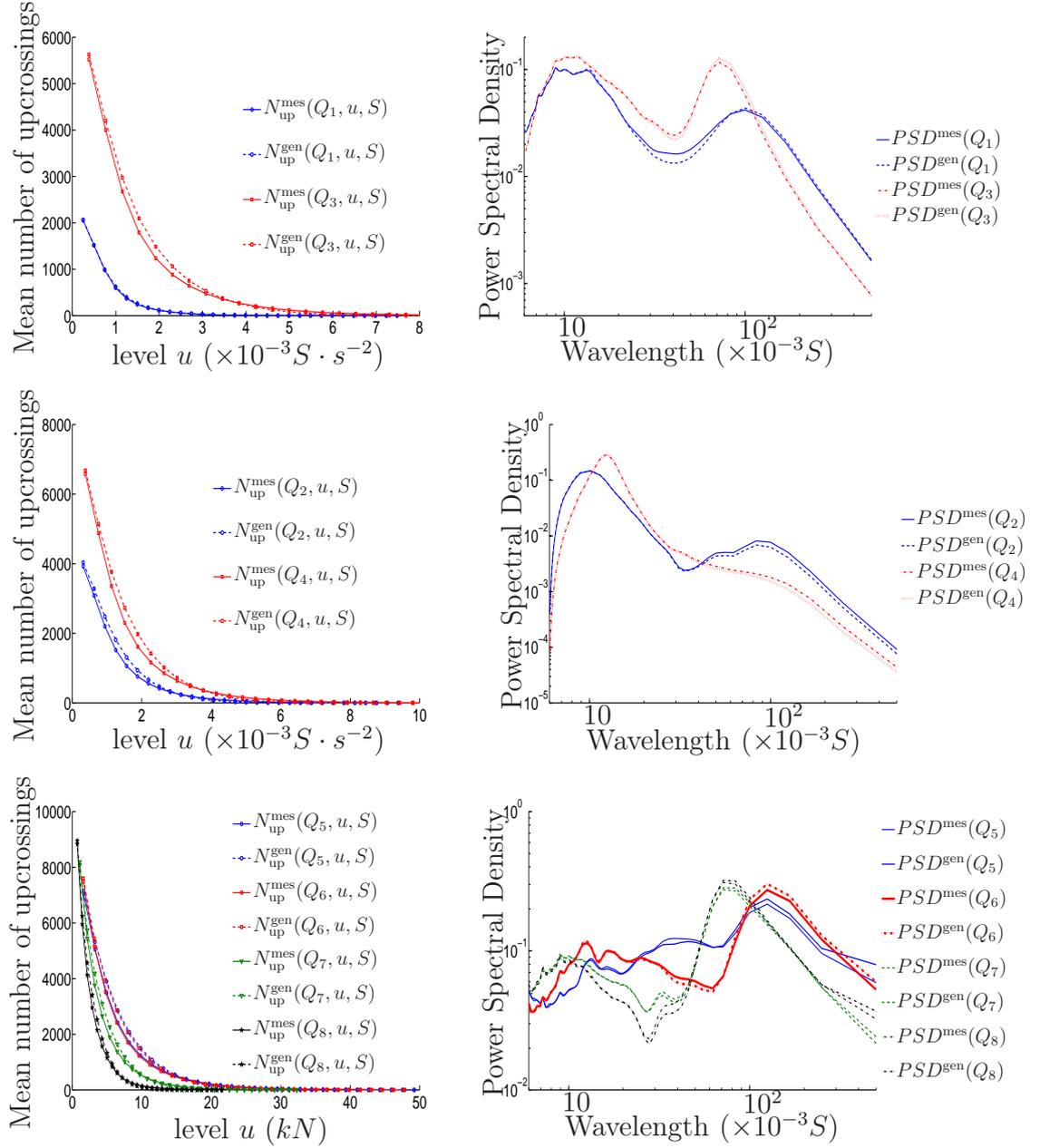
Figure 11: Spectral and statistical analysis of the dynamical quantities of interest $Q_1, Q_2, \ldots, Q_8$.

$$[R_{zz}(s, s')] = \frac{1}{\nu} \sum_{n=1}^{\nu} \boldsymbol{z}^{(n)}(s) \boldsymbol{z}^{(n)}(s')^T, \tag{25}$$

$$[R_{yy}(s, s', S)] = \frac{1}{\nu} \sum_{n=1}^{\nu} \boldsymbol{y}^{(n)}(s, S) \boldsymbol{y}^{(n)}(s', S)^T, \tag{26}$$

$$\boldsymbol{\Sigma}_z = \sqrt{\frac{1}{\nu} \sum_{n=1}^{\nu} PSD\left(\boldsymbol{z}^{(n)}\right)}, \quad \boldsymbol{\Sigma}_y(S) = \sqrt{\frac{1}{\nu} \sum_{n=1}^{\nu} PSD\left(\boldsymbol{y}^{(n)}(S)\right)}, \tag{27}$$

where $PSD\left(\boldsymbol{z}\right) = (PSD(z_1), \ldots, PSD(z_P))$ is the power spectral density estimation of any $\mathbb{R}^P$-valued function $\boldsymbol{z} = (z_1, \ldots, z_P)$. For any value of $S$ in $[0, L]$, errors $err_{\mathrm{cov}}^2(S)$ and $err_{\mathrm{spect}}^2(S)$, which have been introduced in Section 2.1 are then defined by:

$$err_{\mathrm{cov}}^2(S) = \|[R_{zz}] - [R_{yy}(S)]\|_M^2 \, / \, \|[R_{zz}]\|_M^2, \tag{28}$$

$$err_{\mathrm{spect}}^2(S) = \|\boldsymbol{\Sigma}_z - \boldsymbol{\Sigma}_y(S)\|_V^2 \, / \, \|\boldsymbol{\Sigma}_z\|_V^2, \tag{29}$$

where, for all $(P \times P)$ matrix-valued function $[R]$, and for all $\mathbb{R}^P$-valued function $\boldsymbol{\Sigma}$,

$$\|[R]\|_M^2 = \int_0^L \int_0^L \mathrm{Tr}\left([R(s, s')][R(s, s')]^T\right) ds ds', \tag{30}$$

$$\|\boldsymbol{\Sigma}\|_V^2 = \int_{1/L}^{+\infty} \boldsymbol{\Sigma}(f)^T \boldsymbol{\Sigma}(f) df. \tag{31}$$

Hence, on the first hand, $err_{\mathrm{cov}}^2(S)$ corresponds to a covariance error, which quantifies the approximation introduced by the asumption that $E\left[\boldsymbol{\mathfrak{X}}(s)\boldsymbol{\mathfrak{X}}(s')^T\right]$ is negligible. On the other hand, $err_{\mathrm{spect}}^2(S)$ can be seen as a spectral error, which characterizes the impact of $S$ on the frequency content of the track irregularities.

## References

[1] L. Fryba, Dynamics of railway bridges, T. Telford, 1996.

[2] F. Au, J. Wang, Y. Cheung, Impact study of cable-stayed railway bridges with random rail irregularities, Engineering Structures 24 (2002) 529–541.

[3] M. Majka, M. Hartnett, Dynamic response of bridges to moving trains: A study on effects of random track irregularities and bridge skewness, Computers and Structures 87 (2009) 1233–1252.

[4] V. Garg, R. Dukkipati, Dynamics of railway vehicle systems, Academic Press, New York, USA, 1984.

[5] R. Fries, B. Coffey, A state-space approach to the synthesis of random vertical and crosslevel rail irregularities, J. Dyn. Sys., Meas., Control 112 (1) (1990) 83–87.

[6] H. Claus, W. Schiehlen, Modeling and simulation of railway bogie structural vibrations, Vehicle System Dynamics 28 (Suppl) (1997) 538–552.

[7] R. Y. Rubinstein, D. P. Kroese, Simulation and the Monte Carlo Method, Second Edition, Wiley, 2007.

[8] K. Pearson, On lines and planes of closest fit to systems of points in space, Philosophical Magazine 2 (6) (1901) 559–572.

[9] C. Allery, A. Hambouni, D. Ryckelynck, N. Verdon, A priori reduction method for solving the two-dimensional burgers' equations, Applied Mathematics and Computation 217 (2011) 6671–6679.

[10] J. Atwell, B. King, Proper orthogonal decomposition for reduced basis feedback controllers for parabolic equations, Math. Comput. Modell. 33 (1-3) (2001) 1–19.

[11] G. Berkooz, P. Holmes, J. Lumley, The proper orthogonal decomposition in the analysis of turbulent flows, Annu. Rev. Fluid Mech. 25 (1993) 539–575.

[12] G. P. Brooks, J. M. Powers, A karhunen-loève least-squares technique for optimization of geometry of a blunt body in supersonic flow, Journal of Computational Physics 195 (2004) 387–412.

[13] E. Christensen, M. Brons, J. Sorensen, Evaluation of proper orthogonal decomposition-based decomposition techniques applied to parameter dependent nonturbulent flows, SIAM J. Sci. Comput 21 (4) (2000) 1419–1434.

[14] S. Huang, S. Quek, K. Phoon, Convergence study of the truncated karhunen-loève expansion for simulation of stochastic processes, Int J Num Meth Engng 52 (9) (2001) 1029–43.

[15] K. Kunisch, S. Volkwein, Galerkin proper orthogonal decomposition methods for parabolic problems, Numer. Math. 90 (1) (2001) 117–148.

[16] L. Li, K. Phoon, S. Quek, Comparison between karhunen-loève expansion and translation-based simulation of non-gaussian processes, Computers and Structures 85 (2007) 264–76.

[17] X. Ma, N. Zabaras, Kernel principal component analysis for stochastic input model generation, Comptes rendus de l'Académie des sciences de Paris 220 (1945).

[18] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of bayesian inference in inverse problems, J. Comput. Phys. 228 (2009) 1862–1902.

[19] A. Nouy, O. L. Maître, Generalized spectral decomposition method for stochastic non-linear problems, J. Comput. Phys. 228 (1) (2009) 202–235.

[20] K. Phoon, S. Huang, S. Quek, Implementation of karhunen-loeve expansion for simulation using a wavelet-galerkin scheme, Probabilistic Engineering Mechanics 17 (2002) 293–303.

[21] K. Phoon, S. Huang, S. Quek, Simulation of strongly non-gaussian processes using karhunen-loeve expansion, Probabilistic Engineering Mechanics 20 (2005) 188–198.

[22] C. Schwab, R. A. Todor, Karhunen-loeve approximation of random fields by generalized fast multipole methods, Journal of Computational Physics 217 (2006) 100–122.

[23] P. Spanos, B. Zeldin, Galerkin sampling method for stochastic mechanics problems, Journal of Engineering Mechanics 120 (5) (1994) 1091–1106.

[24] P. Spanos, M. Beer, J. Red-Horse, Karhunen -loève expansion of stochastic processes with a modified exponential covariance kernel, Journal of Engineering Mechanics 133 (7) (2007) 773–779.

[25] B. Wen, N. Zabaras, A multiscale approach for model reduction of random microstructures, Computational Materials Science 63 (2012) 269–285.

[26] M. Williams, The eigenfunctions of the karhunen-loeve integral equation for a spherical system, Propabilistic Engineering Mechanics 26 (2011) 202–207.

[27] S. Q. Wu, S. S. Law, Statistical moving load identification including uncertainty, Probabilistic Engineering Mechanics 29 (2012) 70–78.

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

[29] O. Le Maître, O. Knio, Spectral Methods for Uncertainty Quantification, Springer, 2010.

[30] E. T. Jaynes, Information theory and statistical mechanics, The Physical Review 106 (4) (1963) 620–630.

[31] C. Soize, Construction of probability distributions in high dimension using the maximum entropy principle. applications to stochastic processes, random fields and random matrices., International Journal for Numerical Methods in Engineering 76(10) (2008) 1583–1611.

[32] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (1938) 897–936.

[33] R. Ghanem, P. Spanos, Polynomial chaos in stochastic finite elements, Journal of Applied Mechanics Transactions of teh ASME 57 (1990) 197–202.

[34] M. Arnst, R. Ghanem, C. Soize, Identification of bayesian posteriors for coefficients of chaos expansions, Journal of Computational Physics 229 (9) (2010) 3134–3154.

[35] S. Das, R. Ghanem, S. Finette, Polynomial chaos representation of spatio-temporal random field from experimental measurements, J. Comput. Phys. 228 (2009) 8726–8751.

[36] C. Desceliers, R. Ghanem, C. Soize, Maximum likelihood estimation of stochastic chaos representations from experimental data, Internat. J. Numer. Methods Engrg. 66 (2006) 978–1001.

[37] C. Desceliers, C. Soize, R. Ghanem, Identification of chaos representations of elastic properties of random media using experimental vibration tests, Comput. Mech. 39 (2007) 831–838.

[38] R. Ghanem, A. Doostan, On the construction and analysis of stochastic models: Characterization and propagation of the errors associated with limited data, J. Comput. Phys. 217 (2006) 63–81.

[39] Y. M. Marzouk, H. N. Najm, L. A. Rahn, spectral methods for efficient bayesian solution of inverse problems, J. Comput. Phys. 224 (2007) 560–586.

[40] G. Perrin, C. Soize, D. Duhamel, C. Funfschilling, Identification of polynomial chaos representations in high dimension from a set of realizations, SIAM J. Sci. Comput. 34(6) (2012) 2917–2945.

[41] C. Soize, Identification of high-dimension polynomial chaos expansions with random coefficients for non-gausian tensor-valued random fields using partial and limited experimental data., Computer Methods in Applied Mechanics and Engineering 199 (2010) 2150–2164.

[42] C. Soize, Generalized probabilistic approach of uncertainties in computational dynamics using random matrices and polynomial chaos decompositions., Internat. J. Numer. Methods Engrg. 81 (2010) 939–970.

[43] P. Krée, C. Soize, Mathematics of random phenomena, D. Reidel Publishing Company, Dordrecht, 1986.