

A Time Series Global Model Method for Random Vibration Modeling and Generation in Rotating Machinery under any Operating Speed

Xenofon D. Konstantinou, Dimitrios M. Bourdalos, Kyriakos Kritikakos, John S. Sakellariou, Spilios D. Fassois, Josef Koutsoupakis, George Karyofyllas, Dimitrios Giagopoulos

Abstract. A statistical time series global method for random vibration surrogate modeling and signal generation is postulated. Motivated by the mixed spectra nature of the signals, the method is based on a Linear Parameter Varying Generalized AutoRegressive (LPV-GAR) type statistical representation, which, unlike current Neural Network type methods, offers small model size, low complexity, identification simplicity, transparency, and physical interpretability. The LPV nature of the model, employing the rotational speed as scheduling variable, allows, for the first time, for effective signal generation under any speed within an admissible range. The method's effectiveness is assessed via Monte Carlo experiments using a physics-based ADAMS-built high-fidelity drivetrain model and proper signal similarity metrics. The postulated method's performance is superior to those reported in the literature on similar problems, while the importance of using GAR models over simpler AR alternatives is also demonstrated.

Key words: Rotating Machinery, Surrogate models, Data-Driven Random Vibration Signal Generation, Time Series Methods, Generalized AutoRegressive models, Linear Parameter Varying models

1 INTRODUCTION

Over the past decades, data-driven surrogate modeling for the generation of random vibration signals has gained significant attention in the context of design, dynamical analysis, and fault diagnosis in engineering systems and structures [1]. The motivation for this stems from the need for replacing elaborate, large-scale, and expensive physics-based simulations, or actual physical experiments, by small-scale and inexpensive simulations under various Operating

Xenofon D. Konstantinou & Dimitrios M. Bourdalos & Kyriakos Kritikakos & John S. Sakellariou & Spilios D. Fassois
Stochastic Mechanical Systems & Automation (SMSA) Laboratory, Department of Mechanical Engineering & Aeronautics,
University of Patras, 26504 Patras, Greece e-mail: xkonstantinou@ac.upatras.gr

Josef Koutsoupakis & George Karyofyllas & Dimitrios Giagopoulos
Machine Dynamics Laboratory (MDL), Department of Mechanical Engineering, Aristotle University of Thessaloniki,
Thessaloniki GR54124, Greece, e-mail: jkoutsoup@auth.gr

Conditions (OCs) and real-world scenarios, including faulty conditions [2, 1]. This is important for rotating machinery too, where physics-based models are often characterized by large size and high complexity, and physical experiments under varying OCs (such as rotational speeds and loads) or faulty conditions are typically difficult and expensive [3].

Current data-driven signal generation surrogate models for complex rotating machinery systems (such as gearboxes) are mainly based on specialized Neural Network (NN) architectures, such as Generative Adversarial Networks (GANs) [4, 5, 6, 7, 8, 9] and Variational AutoEncoders (VAEs) [10, 11]. Yet, these models are characterized by their own limitations, including dependence on a number of user-defined hyperparameters (typically more than 8, with their selection often requiring expert knowledge), size and complexity still considered relatively high, tuning requiring intensive computational procedures, and also lack of transparency and interpretability (as their weights cannot be directly associated with the underlying physical system dynamics, such as Gear Meshing Frequencies, GMFs) [12]. An additional important drawback is their inability to operate properly and generate high-fidelity random vibration signals under OCs (such as rotational speed) not specifically used during training [3].

In order to overcome these drawbacks and limitations, the present study takes a different approach and aims at exploring statistical times series methods for effectively tackling the problem. Motivated by the *mixed spectra* nature (consisting of cyclic components, associated with the GMFs, and a rational counterpart associated with the structural dynamics and transmission path effects [13]) characterizing random vibration signals in rotating machinery, the study postulates the use of Generalized AutoRegressive (GAR) representations [14]. Moreover, a Linear Parameter Varying (LPV) structure is additionally imposed on them - with the rotational speed being the *scheduling variable* - in order to provide a *global* LPV-GAR model capable of representing the random vibration signal under *any admissible* rotational speed within a prespecified range. The postulated global LPV-GAR model thus features small size, low complexity, identification simplicity, transparency, and physical interpretability.

The postulated LPV-GAR method for random vibration signal modeling and signal generation in rotating machinery is then assessed via a physics-based, ADAMS-built [15], high-fidelity drivetrain model, with the similarity between generated and reference (ADAMS-generated) vibration signals quantified by the Cosine Similarity (CS) and the Pearson Correlation Coefficient (PCC) metrics [16]. The achievable performance (in terms of both metrics) is compared to that reported in the literature for similar problems, whereas the importance of resorting on GAR models (versus their simpler AR counterparts that are commonly used in rotating machinery [17, pp. 180–183]) is also considered.

The rest of the paper is organized as follows: The precise problem statement is presented in [section 2](#), while the postulated LPV-GAR model-based method is introduced in [section 3](#). Details on the physics-based simulation model and the Monte Carlo numerical experiments are provided in [section 4](#), and performance assessment in [section 5](#). Concluding remarks are finally summarized in [section 6](#).

2 PRECISE PROBLEM STATEMENT

The problem may be stated as follows: Given M angularly sampled N -sample-long training vibration signals $y_{\Omega_1}[k], \dots, y_{\Omega_M}[k]$ ($k = 1, 2, \dots, N$; with k designating equispaced samples characterized by angular sampling frequency f_s^θ), with each signal obtained under a specific (fixed and known) *training* rotational speed Ω_i ($i = 1, \dots, L$; $L \leq M$) with $\Omega_i \in [\Omega_{min}, \Omega_{max}] \subset \mathbb{R}^+$, do the following:

- (a) Obtain a proper stochastic global model $\mathcal{M}_\Omega$ properly representing a random vibration signal under *any* rotational speed $\Omega \in [\Omega_{min}, \Omega_{max}]$ (*Training Phase*), and,
- (b) Generate random vibration signal realizations under *any* rotational speed $\Omega \in [\Omega_{min}, \Omega_{max}]$ (*Signal Generation Phase*).

Remark: The training vibration signals are typically obtained via angular resampling of corresponding time-domain counterparts.

3 THE DATA-DRIVEN SIGNAL GENERATION METHOD

The method is based on the Generalized AutoRegressive (GAR) model structure [14] capable of representing both cyclic components and a component characterized by a rational (AR) PSD. The former ones are to represent the cyclic frequencies – such as Gear Meshing Frequencies (GMFs) and their sidebands – while the latter one to represent the part of the signal associated with the structural dynamics, transmission path effects, and other sources [13].

For any specific rotational speed Ω the model is of the form:

$$y_\Omega[k] = \sum_{i=1}^d [A_i(\Omega) \sin(2\pi f_i k) + B_i(\Omega) \cos(2\pi f_i k)] + \frac{1}{a(\mathcal{B}, \Omega)} e_\Omega[k] \quad (1a)$$

$$a(\mathcal{B}, \Omega) := 1 + a_1(\Omega)\mathcal{B} + a_2(\Omega)\mathcal{B}^2 + \dots a_n(\Omega)\mathcal{B}^n \quad (1b)$$

$$e_\Omega[k] \sim i.i.d \mathcal{N}(0, \sigma_e^2(\Omega)) \quad (1c)$$

with $y_\Omega[k]$ designating the angularly resampled vibration signal and $e_\Omega[k]$ the model innovations - which, as indicated, is a Gaussian *i.i.d.* (identically independently distributed) zero-mean signal with variance $\sigma_e^2(\Omega)$. f_i the i -th cyclical frequency (order) normalized by the angular resampling frequency f_s^θ , $A_i(\Omega)$ and $B_i(\Omega)$ ($i = 1, \dots, d$) the cyclical component amplitudes, and $a(\mathcal{B}, \Omega)$ the AR polynomial with $\mathcal{B}$ representing the backshift operator $\mathcal{B}^i \cdot y_\Omega[k] = y_\Omega[k - i]$.

With proper manipulation, the model of Equation 1 may be re-written as¹:

$$y_\Omega[k] = \sum_{i=1}^d [A_i(\Omega) \sin(2\pi f_i k) + B_i(\Omega) \cos(2\pi f_i k)] + \sum_{j=1}^n a_j(\Omega) y_\Omega[k - j] + e_\Omega[k] \quad (2)$$

¹Note that the cyclical component amplitudes are not identical to those in Equation 1; yet for simplicity the same symbols are maintained.

with $e_\Omega[k]$ as before. This is subsequently designated as a $\text{GAR}(n, d)$ model, with n designating the AR order and d the number of cyclical components.

Based on the above GAR model structure pertaining to a specific rotational speed Ω , a global model capable of representing a random vibration signal under *any* rotational speed $\Omega \in [\Omega_{\min}, \Omega_{\max}]$ may be employed. Such a model may be thought of as interpolating the individual (local) $\text{GAR}(n, d)$ models using polynomial interpolation to generate a global Linear Parameter Varying (LPV) GAR model that may be designated as $\text{LPV-GAR}(n, d)_p$ model, with p designating the polynomial degree. This model is then of the form:

$$\begin{aligned}
 y_\Omega[k] &= \underbrace{\sum_{i=1}^d [A_i(\Omega) \sin(2\pi f_i k) + B_i(\Omega) \cos(2\pi f_i k)]}_{\text{Cyclical component}} + \underbrace{\sum_{j=1}^n a_j(\Omega) y_\Omega[k-j] + e_\Omega[k]}_{\text{Rational component}} \quad (3a) \\
 e_\Omega[k] &\sim i.i.d \ N(0, \sigma_e^2(\Omega)) \quad (3b) \\
 A_i(\Omega) &= \sum_{l=0}^p A_{i,l} \cdot \Omega^l, \quad B_i(\Omega) = \sum_{l=0}^p B_{i,l} \cdot \Omega^l \quad (3c) \\
 a_j(\Omega) &= \sum_{l=0}^p a_{j,l} \cdot \Omega^l, \quad \sigma_e(\Omega) = \sum_{l=0}^p \sigma_{e,l} \cdot \Omega^l \quad (3d) \\
 \Omega &\in [\Omega_{\min}, \Omega_{\max}] \subset \mathbb{R}^+ \quad (3e)
 \end{aligned}$$

with $A_{i,l}$, $B_{i,l}$, $a_{j,l}$ and $\sigma_{e,l}$ designating the indicated polynomial coefficients (projection coefficients, with $i = 1, \dots, d$; $j = 1, \dots, n$; $l = 0, \dots, p$).

The *Training Phase* of the proposed signal generation method, which includes the identification of the $\text{LPV-GAR}(n, d)_p$ model, is based on M training vibration signals acquired at specific rotational speeds Ω_i ($i = 1, \dots, L$; $L \leq M$), as described in [section 2](#), and consists of four steps as indicated below.

Step 1: Angular Resampling. The time-domain vibration signals are initially angularly resampled via the Computed Order Tracking method [17, pp. 148–151]. This resampling removes non synchronous frequency modulation caused by minor shaft speed fluctuations, which are present even when the nominal speed is constant [17, pp. 148–151]. In order to ensure a common resampling frequency f_s^θ within the considered operational range $[\Omega_{\min}, \Omega_{\max}]$, $f_s^\theta = f_s / \Omega_{\max}$, where f_s is the time-domain sampling frequency and $\Omega_{\max}$ the maximum speed (in rev/s). For speeds Ω below $\Omega_{\max}$, the bandwidth exceeds the Nyquist limit ($f_s^\theta/2$), causing aliasing which indicates the need for low-pass filtering at $(f_s^\theta/2) \times \Omega$ cut-off frequency [18].

Step 2: Determination of the cyclical frequencies (orders) of the sinusoidal part of the local GAR models. This stage involves the selection of a subset of frequencies (orders) for the cyclical component of the employed local GAR models. Thus, a set of $\bar{d}$ cyclical frequencies (orders) $\mathcal{F}_\Omega^{in} := \{f_1, \dots, f_{\bar{d}}\}$, including Gear Meshing Frequencies (GMFs) along with a selected num-

ber of their harmonics and sidebands, is initially determined for all L training phase operational speeds $\Omega_1, \dots, \Omega_L$. Subsequently, for each training speed Ω_i ($i = 1, \dots, L$), a subset $\mathcal{F}_{\Omega_i}$ is determined by eliminating the frequencies of $\mathcal{F}_{\Omega}^{in}$ that are not manifested in the corresponding signal $y_{\Omega_i}[k]$ ². This is achieved by applying the Least Absolute Shrinkage and Selection Operator (LASSO) method only for the cyclical component of each local GAR model (as indicated in [19]) as follows³:

$$\hat{\boldsymbol{\theta}}(\Omega_i) := \arg \min_{\boldsymbol{\theta}(\Omega_i)} \left\{ \frac{1}{2N} \sum_{k=1}^N [y_{\Omega_i}[k] - \boldsymbol{\phi}^T[k] \cdot \boldsymbol{\theta}(\Omega_i)]^2 + \lambda \sum_{j=1}^{2\bar{d}} |\theta_j(\Omega_i)| \right\} \quad (4)$$

with:

$$\boldsymbol{\phi}[k] := [\sin(2\pi f_1 k) \dots \sin(2\pi f_{\bar{d}} k) : \cos(2\pi f_1 k) \dots \cos(2\pi f_{\bar{d}} k)]_{[2\bar{d} \times 1]}^T,$$

$$\boldsymbol{\theta}(\Omega_i) := [A_1(\Omega_i) \dots A_{\bar{d}}(\Omega_i) : B_1(\Omega_i) \dots B_{\bar{d}}(\Omega_i)]_{[2\bar{d} \times 1]}^T$$

and λ is the regularization factor, a design parameter of the method. Its value is selected by evaluating the following Bayesian criterion [14] over a set of distinct candidate values of λ :

$$\text{Bayesian Criterion} := (N - n) \ln \hat{\sigma}_e^2(\Omega_i) + (n + 1 + 2\bar{d}) \ln \frac{N}{2\pi} \quad (5)$$

where $n = 0$, as the AR component is omitted in Equation 4. The LASSO solution is obtained using the Coordinate Descent Algorithm [20].

Based on this procedure, the set $\mathcal{F}_{\Omega_i}$ for each training speed Ω_i is obtained as follows:

$$\mathcal{F}_{\Omega_i} = \left\{ \text{all } \Omega_i \in \mathcal{F}_{\Omega}^{in} \mid \max_{j=1, \dots, \bar{d}} \{A_j(\Omega_i), B_j(\Omega_i)\} > \delta \right\} \quad (6)$$

with $\delta > 0$ designating a user-defined small positive number (threshold above which an estimated amplitude is considered to be non-zero; design parameter).

The set $\mathcal{F}$ of all (d) cyclical frequencies (orders) kept in the final LPV-GAR model of Equation 3 includes all those kept for any individual model, that is:

$$\mathcal{F} := \bigcup_{i=1, \dots, L} \mathcal{F}_{\Omega_i} \quad (7)$$

Step 3: Identification of the local GAR models. Based on the subset of cyclical frequencies (orders) $\mathcal{F}$ selected in the previous stage, the local GAR models are identified using Ordinary Least Squares (OLS) [21, pp. 204-205]. The AR order n (design parameter) is determined based on the Bayesian criterion⁴ given in Equation 5, with $\bar{d}$ substituted by d .

²If more than one training signal is available per speed ($M > L$), one signal is selected to be used in Step 2.

³Estimators/estimates are indicated via a hat and column-vectors/matrices via bold-face lower/upper case symbols.

⁴The final AR order n is determined as the mean value suggested by the criterion over all M training signals.

Step 4: Determination of the LPV-GAR model. Once the identification of the local GAR models is completed, their parameters are interpolated using standard polynomials, yielding the coefficients $A_{i,l}$, $B_{i,l}$, $a_{j,l}$, and $\sigma_{e,l}$ in Equations 3c and 3d. This leads to the determination of the LPV-GAR model, as defined in Equation 3. The polynomial order p is a design parameter of the method and is selected as the value at which the mean squared error reaches a plateau.

Once the above modeling procedure is completed, the LPV-GAR model may be used for the accurate generation of random vibration signal realizations (*Signal Generation Phase*) for any rotational speed $\Omega \in [\Omega_{min}, \Omega_{max}]$. This is achieved by substituting $e_{\Omega}[k]$ in Equation 3 with a randomly generated Gaussian white noise characterized by zero mean and variance $\sigma_e^2(\Omega)$.

4 THE ROTATING MACHINERY MULTI-BODY DYNAMICS MODEL AND THE MONTE CARLO NUMERICAL EXPERIMENTS

A high-fidelity Multi-body Dynamics (MBD) drivetrain model is developed employing the ADAMS software [15]. The 152 degree-of-freedom (DoFs) drivetrain model, is comprised of two gearboxes, that is a planetary and a two-stage gearbox. The planetary gearbox consists of a sun gear, a ring gear which is stationary, and four planet gears mounted on the carrier, as shown in Figure 1. The gears and shaft are modeled as rigid bodies, while the two gearboxes are modeled as flexible bodies. Motion is introduced via the sun gear's inlet shaft and transmitted through the carrier to the two-stage gearbox, which consists of four gears, with the pinion mounted on the carrier. The shafts mounted to the gearboxes, and the bolts fixing the gearboxes to the ground are modeled as bushings, taking into account the bearings' stiffness and preload conditions in the physical counterpart. The gears are assumed to be fixed to their respective shafts, and gear contact forces are modeled explicitly for each gear pair considering an impact

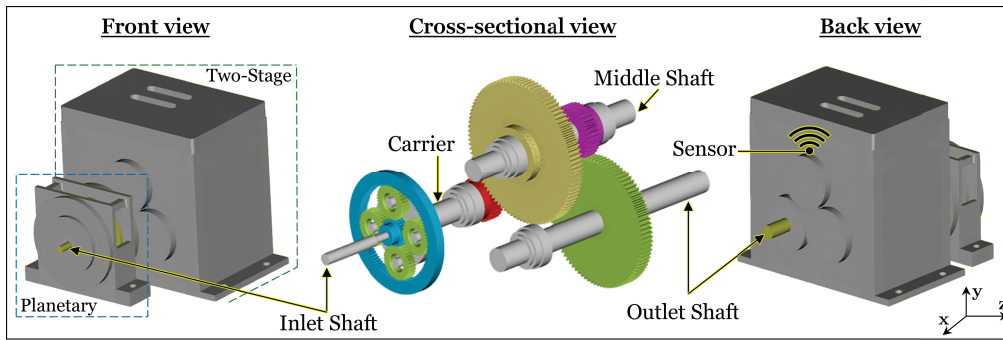


Figure 1: The MBD model of the experimental gear drivetrain, comprising a planetary gearbox with a 28-tooth sun gear, four planet gears (each with 36 teeth), and a stationary 100-tooth ring gear, followed by a two-stage gearbox consisting of a 29-tooth pinion, a 100-tooth gear, and a 36-tooth gear mounted on the middle shaft, along with a 90-tooth gear on the output shaft. A sensor (accelerometer) is positioned on the casing of the middle shaft bearing at the rear side of the gearbox.

force formulation with Hertzian contact stiffness [22], allowing for accurate meshing force estimation.

The MBD model parameters are fine-tuned through an optimization process using experiments from the physical counterpart. This is based on the Covariance Matrix Adaptation Evolution Strategy (CMAES), selected for its ability to effectively improve the fidelity of highly complex and nonlinear models [23]. For each Monte Carlo numerical experiment, a lateral (y) acceleration signal is measured using a single sensor mounted on the casing of the middle shaft bearing, as shown in Figure 1. The sampling frequency is $f_s = 10\,240$ Hz, while the signal length is $N = 40\,960$ samples (4 s).

5 PERFORMANCE ASSESSMENT OF THE SIGNAL GENERATION METHOD

The performance of the proposed LPV-GAR model-based signal generation method is comprehensively assessed using two quality indices: The Pearson Correlation Coefficient (PCC), and Cosine Similarity (CS) which quantify the similarity (D) between the generated signals ($y_\Omega[k]$) and the reference signals ($x_\Omega[k]$) obtained from MBD simulations (see section 4). These are defined as follows [16]:

$$\text{Pearson Correlation Coefficient (angular domain): } D(\dot{\mathbf{y}}_\Omega, \dot{\mathbf{x}}_\Omega) := \frac{\dot{\mathbf{y}}_\Omega^T \cdot \dot{\mathbf{x}}_\Omega}{\|\dot{\mathbf{y}}_\Omega\| \cdot \|\dot{\mathbf{x}}_\Omega\|} \times 100\% \quad (8)$$

$$\text{Cosine Similarity (frequency domain): } D(\tilde{\mathbf{y}}_\Omega, \tilde{\mathbf{x}}_\Omega) := \frac{\tilde{\mathbf{y}}_\Omega^T \cdot \tilde{\mathbf{x}}_\Omega}{\|\tilde{\mathbf{y}}_\Omega\| \cdot \|\tilde{\mathbf{x}}_\Omega\|} \times 100\% \quad (9)$$

with $\dot{\mathbf{y}}_\Omega, \dot{\mathbf{x}}_\Omega$ denoting sample-mean-corrected versions of the vectors $\mathbf{y}_\Omega := [y_\Omega[1], \dots, y_\Omega[N]]^T$ and $\mathbf{x}_\Omega := [x_\Omega[1], \dots, x_\Omega[N]]^T$ containing all signal samples, $\tilde{\mathbf{y}}_\Omega, \tilde{\mathbf{x}}_\Omega$ their (uncorrected) counterparts containing their $N/2$ DFT magnitudes, and $\|\cdot\|$ the Euclidean norm.

The importance of employing GAR models over the simpler and commonly used AR alternatives is also explored via a corresponding LPV-AR(n) _{p} alternative representation. All details on the vibration signals used in the *Training Phase* and the *Signal Generation Phase* are summarized in Table 1.

5.1 The Training Phase

The selected design parameters and estimation details for the proposed (LPV-GAR) and the alternative (LPV-AR) model-based signal generation methods are summarized in Table 2.

5.2 The Signal Generation Phase - Performance Assessment

The previously identified LPV-GAR(100, 266)₃ and LPV-AR(1280)₃ models are now employed to generate 20 random vibration signal realizations for each of the 11 rotational speeds specified in Table 1. This results in a total of 220 generated signals, which are directly compared with the corresponding reference (real) signals obtained from the MBD simulations.

The average DFT amplitude over all 20 LPV-GAR(100, 266)₃ and LPV-AR(1280)₃ model-based generated realizations for an indicative non-training speed (1 110 rpm) is presented in

Table 1: Details on the vibration signals used in the Training and Signal Generation Phases.

Phase	Rotational Speed (rpm)	#Speeds	#Signals/Speed	Total #Signals
Training	{900, 960, ..., 1 200} (step of 60 rpm)	6	1	6 ^(a)
Signal Generation	{900, 930, ..., 1 200} ^(c) (step of 30 rpm)	11	20	220

Resampling frequency: $f_s^\theta = 2\,340$ (samples/rev)^(b); Signal length: $N = 30\,420$ samples.

^(a)Each training signal corresponds to a distinct speed, hence $M = L = 6$.

^(b)Given the planetary gearbox ratio of approximately 4.57, the maximum rotational speed of the carrier shaft is $\Omega_{max} = 4.38$ rev/s.

^(c)The non-training speeds are: {930, 990, 1 050, 1 110, 1 170} rpm.

Table 2: Details on design parameter selection.

Model	Design Parameter	Search Interval	Selected Value
LPV-GAR ^(a)	AR order (n)	[10, 1 500]	100
	Polynomial order (p) ^(b)	[1, 5]	3
	LASSO tuning parameter (λ)	$[10^{-4}, 10^{-2}]$	$2.35 \times 10^{-3(c)}$
LPV-AR	AR order (n)	[10, 1 500]	1 280
	Polynomial order (p) ^(b)	[1, 5]	3

^(a)The design parameter δ has been selected equal to 10^{-4} .

^(b)All polynomial coefficients up to degree p are included.

^(c)Corresponding to $d = 266$ cyclical frequencies. The initial set $\mathcal{F}_\Omega^{in}$ comprises $\bar{d} = 416$ cyclical frequencies, including all GMFs, their harmonics, and associated sidebands.

Figure 2 and directly compared with that of the reference signals obtained from the MBD simulations. Evidently, the close agreement between the average DFT of the LPV-GAR model-based generated signals and that of the reference signals demonstrates the high accuracy of the proposed signal generation method. This performance is clearly superior to that of the alternative LPV-AR model-based method, as it fails to accurately reproduce certain cyclical frequencies (orders), either omitting them entirely or generating them with incorrect amplitudes (see Figure 2(b)).

The cumulative results, over all Generation Phase speeds and signals, are presented in Figure 3 in terms of sample mean values of the PCC and CS similarity metrics for the LPV-GAR(100, 266)₃ and LPV-AR(1280)₃ model-based methods. It is evident that the LPV-GAR model-based generation method outperforms its LPV-AR counterpart. As expected, the performance of both methods deteriorates at speeds not included in the Training Phase (non-training speeds). This deterioration is more evident in the PCC metric (angular domain), indicating a phase mismatch between the generated and reference signals, an issue that has also been reported in the literature [5]. It is important to note that the achievable performance is superior to the best results reported in the literature for similar problems, in which the PCC reaches 85.4%

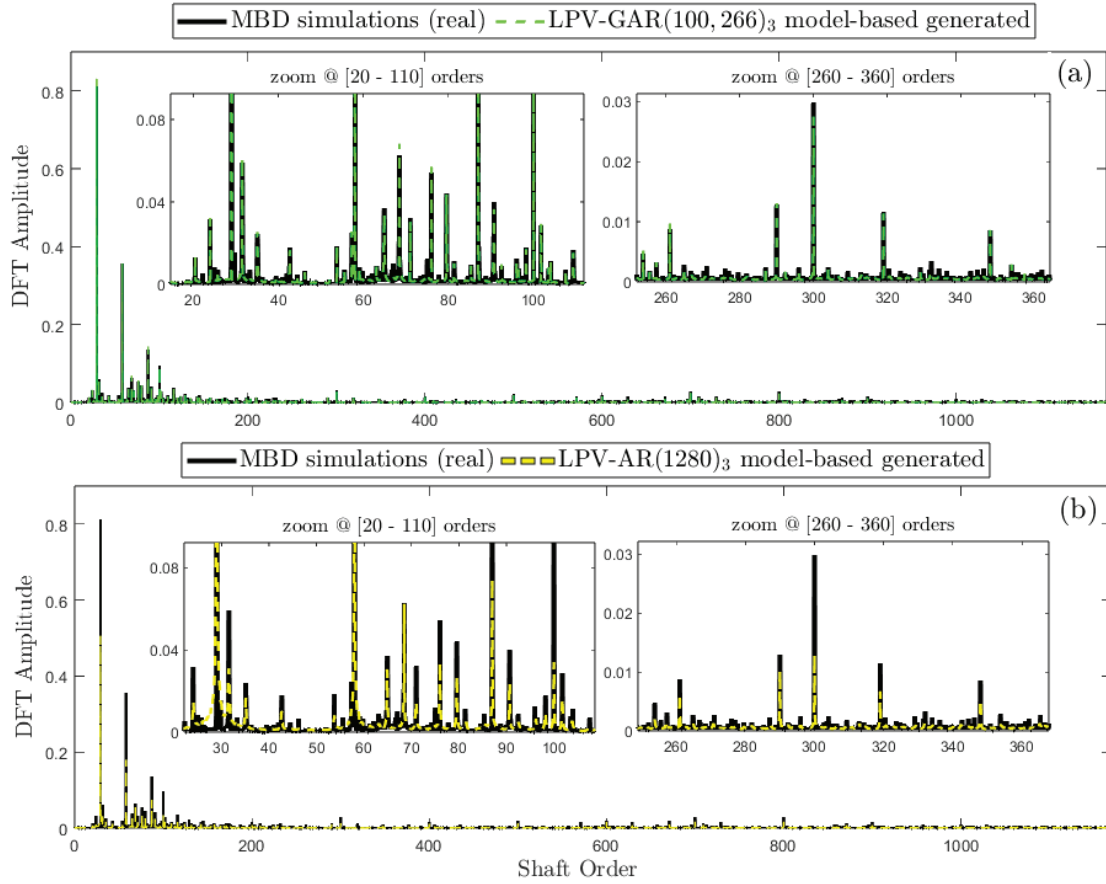


Figure 2: Random vibration signal generation performance for a non-training rotational speed (1110 rpm): Average DFT magnitude of 20 generated signals (green or yellow) versus that of the actual MBD-provided signal (black): Signal generation based on (a) the LPV-GAR(100, 266)₃ model (in green), and (b) the LPV-AR(1280)₃ model (in yellow).

[24] and the CS 92.2% [4] under operating conditions (such as rotational speeds and loads) included in the Training Phase, while PCC drops to 10% [5] for non-training conditions.

6 CONCLUDING REMARKS

A statistical time series global method for random vibration surrogate modeling and signal generation has been postulated and comprehensively assessed. Motivated by the mixed spectral nature of the signals, the method is based on a Linear Parameter Varying Generalized AutoRegressive (LPV-GAR) statistical representation. By employing the rotational speed as a scheduling variable, it enables, for the first time, effective signal generation under any speed

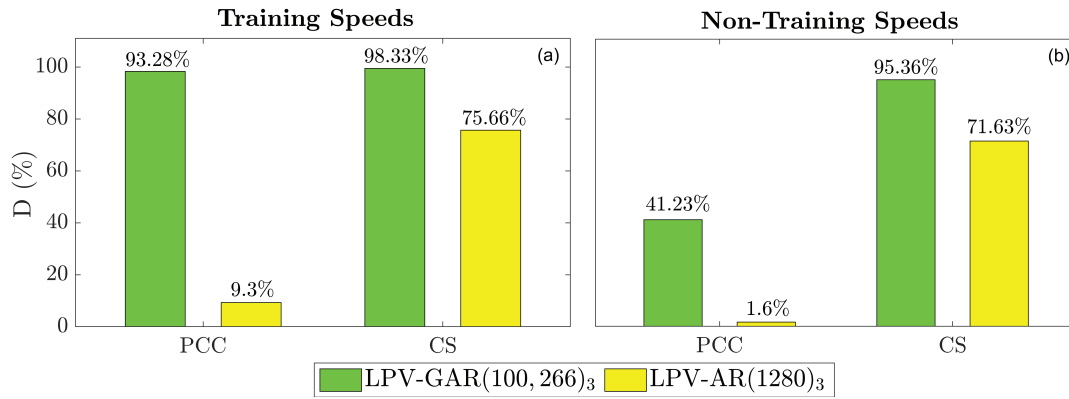


Figure 3: Random vibration signal generation overall performance assessment in terms of sample mean values of the PCC and CS similarity metrics for the LPV-GAR(100, 266)₃ (in green) and LPV-AR(1280)₃ (in yellow) model-based methods for (a) all training speeds (left subplot) and (b) all non-training speeds (right subplot). [The mean values of the 120 Generation Phase signals corresponding to the 6 training speeds and the 100 signals corresponding to the 5 non-training speeds (20 signals per speed) are presented; details in Table 1.]

within an admissible range.

The postulated method has been systematically assessed via a physics-based, ADAMS-built, high-fidelity drivetrain model, with the similarity between generated and reference (ADAMS-generated) vibration signals quantified by the CS and the PCC metrics. The achievable performance (in terms of both metrics) is compared to that reported in the literature for similar problems, whereas the importance of resorting on GAR models (versus their simpler AR counterparts) is also examined.

The main conclusions drawn from the study may be summarized as follows:

1. The LPV-GAR model-based signal generation method demonstrates high performance, providing high-fidelity signals for any rotational speed within the operational range, even for speeds not employed during training.
2. The achievable performance is superior to that of the alternative LPV-AR model-based method and the best results reported in the literature for similar problems. Unlike current Neural Network type methods, the proposed method offers small size, low complexity, identification simplicity, transparency, and physical interpretability.
3. The employed GAR models offer enhanced model parsimony and identification accuracy compared to the simpler AR counterparts that are commonly used in rotating machinery [17, pp. 180–183].

Acknowledgment

This study has been carried out as part of the ‘EEDRIVEN’ project implemented in the framework of the H.F.R.I. call ‘Basic Research Financing (Horizontal support of all Sciences)’

under the National Recovery and Resilience Plan ‘Greece 2.0’ funded by the European Union – NextGenerationEU (H.F.R.I. Project Number: 16381).

REFERENCES

- [1] J. Kudela and R. Matousek, “Recent advances and applications of surrogate models for finite element method computations: a review,” *Soft Computing*, vol. 26, no. 24, pp. 13 709–13 733, Dec 2022.
- [2] T. Pan, J. Chen, T. Zhang, S. Liu, S. He, and H. Lv, “Generative adversarial network in mechanical fault diagnosis under small sample: A systematic review on applications and future perspectives,” *ISA Transactions*, vol. 128, pp. 1–10, 2022.
- [3] Z. Ju, Y. Chen, Y. Qiang, X. Chen, C. Ju, and J. Yang, “A systematic review of data augmentation methods for intelligent fault diagnosis of rotating machinery under limited data conditions,” *Measurement Science and Technology*, vol. 35, no. 12, p. 122004, 2024.
- [4] T. Zheng, L. Song, J. Wang, W. Teng, X. Xu, and C. Ma, “Data synthesis using dual discriminator conditional generative adversarial networks for imbalanced fault diagnosis of rolling bearings,” *Measurement*, vol. 158, p. 107741, 2020.
- [5] T. Zhang, J. Chen, J. Xie, and T. Pan, “Sasln: Signals augmented self-taught learning networks for mechanical fault diagnosis under small sample condition,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [6] Y. Liao, R. Huang, J. Li, Z. Chen, and W. Li, “Deep semisupervised domain generalization network for rotary machinery fault diagnosis under variable speed,” *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 8064–8075, 2020.
- [7] Z. Li, T. Zheng, Y. Wang, Z. Cao, Z. Guo, and H. Fu, “A novel method for imbalanced fault diagnosis of rotating machinery based on generative adversarial networks,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–17, 2021.
- [8] J. Luo, J. Huang, and H. Li, “A case study of conditional deep convolutional generative adversarial networks in machine fault diagnosis,” *Journal of Intelligent Manufacturing*, vol. 32, no. 2, pp. 407–425, 2021.
- [9] R. Wang, Z. Chen, and W. Li, “Gradient flow-based meta generative adversarial network for data augmentation in fault diagnosis,” *Applied Soft Computing*, vol. 142, p. 110313, 2023.
- [10] Y. Liu, H. Jiang, R. Yao, and H. Zhu, “Interpretable data-augmented adversarial variational autoencoder with sequential attention for imbalanced fault diagnosis,” *Journal of Manufacturing Systems*, vol. 71, pp. 342–359, 2023.
- [11] Z. Zhao, Y. Xu, J. Zhang, R. Zhao, Z. Chen, and Y. Jiao, “A semi-supervised gaussian mixture variational autoencoder method for few-shot fine-grained fault diagnosis,” *Neural Networks*, vol. 178, p. 106482, 2024.

- [12] D. Neupane, M. R. Bouadjenek, R. Dazeley, and S. Aryal, “Data-driven machinery fault diagnosis: A comprehensive review,” *Neurocomputing*, vol. 627, p. 129588, 2025.
- [13] Y. Chen, M. Rao, K. Feng, and M. J. Zuo, “Physics-informed LSTM hyperparameters selection for gearbox fault detection,” *Mechanical Systems and Signal Processing*, vol. 171, p. 108907, 2022.
- [14] C. Chatterjee, R. Kashyap, and G. Boray, “Estimation of close sinusoids in colored noise and model discrimination,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 328–337, 1987.
- [15] Hexagon, “Adams: Multibody dynamics simulation software,” 2025, accessed: 2025-04-02.
- [16] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, “A comparison study on similarity and dissimilarity measures in clustering continuous data,” *PLoS ONE*, vol. 10, no. 12, p. e0144059, 2015.
- [17] R. Randall, *Vibration-Based Condition Monitoring: Industrial, Automotive and Aerospace Applications*. Hoboken, NJ: John Wiley & Sons, Ltd, 2021.
- [18] D. Bourdalos and J. Sakellariou, “Vibration-based robust detection of gear incipient faults under different speeds and loads via a cloud of functional models,” in *Proceedings of the 11th ECCOMAS Thematic Conference on Smart Structures and Materials (SMART 2025)*, Linz, Austria, July 2025.
- [19] T. Li, *Time Series with Mixed Spectra*. Taylor & Francis, 2014, ch. 5.4, pp. 133–135.
- [20] J. H. Friedman, T. Hastie, and R. Tibshirani, “Regularization paths for generalized linear models via coordinate descent,” *Journal of Statistical Software*, vol. 33, no. 1, p. 1–22, 2010.
- [21] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice Hall PTR, 1999.
- [22] J. Koutsoupakis, D. Giagopoulos, and I. Chatziparasidis, “AI-based condition monitoring on mechanical systems using multibody dynamics models,” *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106467, 2023.
- [23] J. Koutsoupakis, P. Seventekidis, and D. Giagopoulos, “Machine learning based condition monitoring for gear transmission systems using data generated by optimal multibody dynamics models,” *Mechanical Systems and Signal Processing*, vol. 190, p. 110130, 2023.
- [24] Q. Guo, Y. Li, Y. Liu, S. Gao, and Y. Song, “Data augmentation for intelligent mechanical fault diagnosis based on local shared multiple-generator gan,” *IEEE Sensors Journal*, vol. 22, no. 10, pp. 9598–9609, 2022.