IEEE *Access*
Multidisciplinary : Rapid Review : Open Access Journal

RESEARCH ARTICLE

System Condition Monitoring Based on a Standardized Latent Space and the Nataf Transform

ADAITON OLIVEIRA-FILHO[1], (Member, IEEE), **RYAD ZEMOURI**[2], **FRANCIS PELLETIER**[3], **AND ANTOINE TAHAN**[1]

[1]Department of Mechanical Engineering, École de Technologie Supérieure, Montreal, QC H3C 1K3, Canada
[2]Research Center of Hydro-Québec, Varennes, QC J3X 1S1, Canada
[3]Power Factors, Brossard, QC J4Z 1A7, Canada

Corresponding author: Adaiton Oliveira-Filho (adaitonfilho@icloud.com)

ABSTRACT This work introduces a new condition monitoring approach for complex systems based on a standardized latent space representation. Latent variable models such as the variational autoencoders are widely used to analyze systems described by a high-dimensional physical space. The encoding of such space defines a low-dimensional and physically representative latent space. Of note, however, the latent space obtained for complex systems operating under multiple conditions is often difficult to exploit in defining an efficient Health Index, thanks to the non-deterministic and hyperparameter-dependent nature of the latent space. In addition, the distribution of the healthy cluster is not known a priori. The original contribution of this paper is to use the Nataf isoprobabilistic transform to map the latent space into a standardized space. This normalizes the spatial structure of the latent space and relaxes the model's sensitivity to hyperparameters during the learning process. Moreover, the characterization of the healthy condition in the standard Nataf space leads to the definition of two complementary health indices suitable for complex systems. An implementation in two case studies demonstrates the potential of the proposed approach. First, the approach was successfully applied within NASA's Commercial Modular Aero-Propulsion System Simulation dataset. The second case study consisted of analyzing multiple degradation in operating wind turbines. Encouraging results emerge from both case studies, with critical conditions being detected significantly earlier than in competing approaches. The proposed approach can be generalized to complex systems equipped with multiple sensors, and overcomes difficulties related to latent space analysis of multiple condition systems.

INDEX TERMS Condition monitoring, early detection, latent space modeling, Nataf transform, variational autoencoder.

NOMENCLATURE

cdf	Cumulative distribution function.
CM	Condition Monitoring.
DNN	Deep Neural Network.
HI	Health Index.
LHI	Latent Health Index.
LV	Latent Variable.
LVM	Latent Variable Model.

NT	Nataf Transform.
O&M	Operation and Maintenance.
pdf	Probability density function.
VAE	Variational Autoencoder.

The associate editor coordinating the review of this manuscript and approving it for publication was Haidong Shao.

NOTATION

PROBABILITY DISTRIBUTIONS

$\chi_N, f_{\chi_N}, F_{\chi_N}$	Chi distribution with N degrees of freedom and the respective pdf and cdf.

Φ cdf of the normal distribution $\mathcal{N}(0, 1)$.

φ_N N-variate normal joint cdf of the multivariate normal distribution $\mathcal{N}(\mathbf{0}, \mathbf{1})$.

P_x Percentile x of a distribution.

CONDITION MONITORING

τ, Δ_{lag} Time period, lag time.

DT_M, DT_N Detection thresholds.

I_M, I_N Health Indices defined in the standard Nataf space.

R^2 Coefficient of determination.

NATAF TRANSFORM

$\mathbf{L}$ Cholesky decomposition of the correlation matrix.

$\mathbf{R}_U$, r_{Uij} Fictive Nataf correlation matrix and correlation coefficients.

$\mathbf{R}_Z$, r_{Zij} Matrix of correlation and correlation coefficients describing the latent space variable.

$\mathcal{T}$, $\mathcal{T}_1$, $\mathcal{T}_2$ Nataf transform and its intermediary transforms.

D, D_τ, D_{HY} Random variables describing the norm of points from datasets projected in the standard Nataf space.

VAE MODEL

β_{cl}, β_{kl} Loss function coefficients.

$\mathcal{L}$, $\mathcal{L}_{\text{RE}}$, $\mathcal{L}_{\text{KL}}$, $\mathcal{L}_{\text{CL}}$ Loss function and its components.

f_ϕ, g_θ VAE parametric functions.

REPRESENTATION SPACES

$\mathbb{R}$, $\mathbb{X}$, $\mathbb{S}_{\text{healthy}}$ Datasets.

$\mathbf{S}_{\text{HY}}$, $\mathbf{S}_\tau$ Random vectors describing the distribution of datasets in the standard Nataf space.

$\mathbf{x}$, $\mathbf{z}$, $\mathbf{s}$ Data points from physical space, latent space, and Nataf space, respectively.

$\mathbf{X}$, $\mathbf{Z}$, $\mathbf{U}$, $\mathbf{S}$ Random vectors describing the distribution of points in the physical space, latent space, intermediary Nataf space, and Nataf space, respectively.

$\mathbf{x}_\tau$, $\mathbf{z}_\tau$, $\mathbf{s}_\tau$ Datasets covering the period τ of points in the physical space, latent space, and Nataf space, respectively.

$F_i(z_i)$ cdf corresponding to the random variable Z_i.

N_X, N, N_C Dimensions of the physical space, latent space, and classification space, respectively.

$p(\mathbf{x})$ Data distribution.

t, t_M, t_N, t_{Do1}, t_S, $t_{\text{ShutDown1}}$ Time variable, instants of time.

X_i, Z_i, S_i Random variables describing the distribution of coordinates in the physical space, latent space, and Nataf space, respectively.

x_i, z_i, s_i Data coordinates from physical space, latent space, and Nataf space, respectively.

I. INTRODUCTION

Modern engineering systems are increasingly complex and are expected to conform to rising standards of efficiency and reliability [1]. These challenging trends are well illustrated by the technological evolution of aeronautical systems, energy conversion systems such as wind turbines (WTs), and large industrial machines, among others [2], [3], [4]. The costs and strategic importance of these assets justify the need for advanced Condition Monitoring (CM) approaches. Such approaches aim at continuously monitoring the condition of assets, which in turn allows for optimized Operation and Maintenance (O&M) planning and enhanced reliability, in addition to providing economic benefits. Modern engineering systems are equipped with a large variety of sensors. The measures from such sensors provide a high-dimensional multi-physics description of the condition of these systems. The ever-increasing databases resulting from modern systems create new opportunities for CM. Data-driven approaches based on artificial intelligence models such as Deep Neural Networks (DNNs) receive a lot of attention from industry and academia [5], [6].

Latent space representations are low-dimensional projections of high-dimensional physical spaces achieved through an appropriate inference function. The Variational Autoencoder (VAE) model [7] is probably the most popular Latent Variable Model (LVM) combining artificial neural networks and Bayesian inference methods. Once this generative model is trained, the encoder allows to project the input physical space into a low-dimensional Latent Variable (LV) space. Most importantly, the latent space reveals relations between data points that are usually not evident in the original high-dimensional physical space. Due to these capabilities, the VAE model has proven results in a wide range of applications. Furthermore, multiple variations of the VAE model have been introduced to deal with specific data and modeling characteristics [8], [9], [10], [11], [12].

The case of systems operating under multiple distinct conditions is of particular interest for latent space representations. Indeed, variational encoding can isolate different operating conditions, and will ultimately reveal the relations existing between these conditions [13], [14], [15]. Nevertheless, LVs corresponding to complex multiple

condition systems are unsuitable for robust CM models due to several restrictions. First, the latent space representation of complex multiple condition systems is often intricate and non-deterministic. Indeed, the spatial representation of LVs is very sensitive to the model's hyperparameters and changes randomly for independent training instances. Finally, the distribution of the healthy cluster is not known a priori.

Variations of the classical unsupervised VAE model in the literature partially address the limitations of the classical VAE in analyzing complex systems [16]. In particular, supervised implementations of the VAE model exploit labels such as binary classification (healthy or abnormal condition), multiple-class categorization (healthy condition plus multiple abnormal conditions), or regression (condition indicated by a scalar) [8], [9], [11]. In [15], the authors show that the embedding of the VAE model with a classification DNN allows disentangling clusters in the latent space. In [17], the VAE is embedded with a regression DNN. These supervised learning models allow for disentangled clusters in the LV representation, but the latent space still varies between different instances of training.

The present paper aims to define a robust detection approach for complex systems, and to that end, introduces a new framework to achieve a latent space representation for high-dimensional complex engineering systems. The solution investigated herein consists in combining the LVs obtained by the VAE with the Nataf Transform (NT) [18]. The NT has been widely used among several reliability applications, such as [19], [20], and [21]. The features of the complex system are then projected from the original and intricate latent space into the standard Nataf space. The latter representation overcomes the limitations of classical latent space representations, which allows to define a robust Health Index (HI).

The following are the original contributions of the present paper:

- The NT is used to map an intricate latent space into the standard Nataf space. It is shown that the standard Nataf space overcomes the limitations of VAE-based latent space representations of complex multiple condition systems.
- A pair of complementary HI based on the standard Nataf space are introduced. The performance of the detection using the proposed HIs is demonstrated within two application case studies.
- This work introduces two new cases supporting the use of latent spaces as built-in visualization tools to enhance the interpretability of DNN models.

The remainder of this paper is organized as follows: Section II reviews latent space representations and anomaly detection based on such spaces. Section III describes the Nataf Transform. The proposed approach is introduced in Section IV. Two case studies demonstrate the proposed approach: the degradation of engines from the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) FD001 dataset, in Section V, and the degradation of WTs

from an operating North American wind farm, in Section VI. Section VII summarizes and discusses the results from the case studies. Finally, Section VIII concludes the paper.

II. PROBLEM FORMULATION

A. LATENT VARIABLE MODELS

Let $\mathbb{X} = \{\mathbf{x}^{(i)}\}_{i=1}^{N_X}$ be a dataset of N_X observed variables collected from the same system $\mathcal{S}$, whose unchanging underlying true distribution $p^*(\mathbf{x})$ is unknown. The observations $\mathbf{x}$ are assumed to be independently and identically distributed (i.i.d.) [22], and can be considered as random samples obtained from the unknown underlying process $\mathbf{x} \sim p^*(\mathbf{x})$. The objective of LVM is to describe the data space distribution through a set of unobservable variables $\mathbb{Z} = \{\mathbf{z}^{(i)}\}_{i=1}^{N}$ whose prior distribution $p(\mathbf{z})$ is assumed to be known [22], [23]. This latent space or manifold has no physical meaning, but is very suitable for revealing and disentangling the main relevant features from the entangled space of data $\mathbb{X}$. An important condition for this feature extraction process and disentanglement is that the latent space must have a lower dimension than the original space, i.e., $\dim(\mathbb{Z}) = N < N_X = \dim(\mathbb{X})$.

The objective of the LVMs is to find the posterior probability distribution function $p(\mathbf{x})$ that gives the best approximation of the true distribution of the data $p(\mathbf{x}) \approx p^*(\mathbf{x})$. The posterior data distribution $p(\mathbf{x})$ is then given through the conditional generative distribution $p_\theta(\mathbf{x}|\mathbf{z})$ by $p(\mathbf{x}) = \int p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})d\mathbf{z}$, where $p(\mathbf{z})$ is not conditioned on any observations, and is called the prior distribution over $\mathbf{z}$. The latent posterior distribution $p_\phi(\mathbf{z}|\mathbf{x})$ can be obtained using the Bayes theorem for model inference through $p_\phi(\mathbf{z}|\mathbf{x}) = p_\theta(\mathbf{x}|\mathbf{z})p(\mathbf{z})/p(\mathbf{x})$. To this end, two parametric functions f_ϕ and g_θ are defined to sample $\mathbb{X}$ to the latent space $\mathbb{Z}$, and conversely from $\mathbb{Z}$ to $\mathbb{X}$. The inference function f_ϕ, called the encoder, is used to map the data space to the latent space $\mathbf{z} = f_\phi(\mathbf{x})$ and is employed to parameterize the inference distribution $p_\phi(\mathbf{z}|\mathbf{x})$. The mapping of the data space from the latent space is obtained by the generative function $\mathbf{x} = g_\theta(\mathbf{z})$, called the decoder, which is used to parameterize the generative distribution $p_\theta(\mathbf{x}|\mathbf{z})$.

The key idea of this process is to force the latent prior distribution to follow a known distribution such as the Gaussian distribution $p(\mathbf{z}) \approx \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$. The latent Kullback-Liebler divergence loss is used to regularize the inference function in order to enforce the encoding data into the prior Gaussian latent space $D_{\mathrm{KL}}(p_\phi(\mathbf{z}|\mathbf{x}) \parallel p(\mathbf{z}))$. Then, the mapping of the data space obtained by the generative function from the known prior distribution is forced to follow the true data distribution $p^*(\mathbf{x})$. Since the latter is unknown, $p^*(\mathbf{x})$ can explicitly be estimated by forcing the decoder to reconstruct the data $\tilde{\mathbf{x}} = g_\theta(f_\phi(\mathbf{x}))$. A data reconstruction loss, such as the L_2 norm or the mean-squared error $\mathcal{L}_{\theta,\phi} = \frac{1}{2}\|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2$, is then used to guide this reconstruction process.
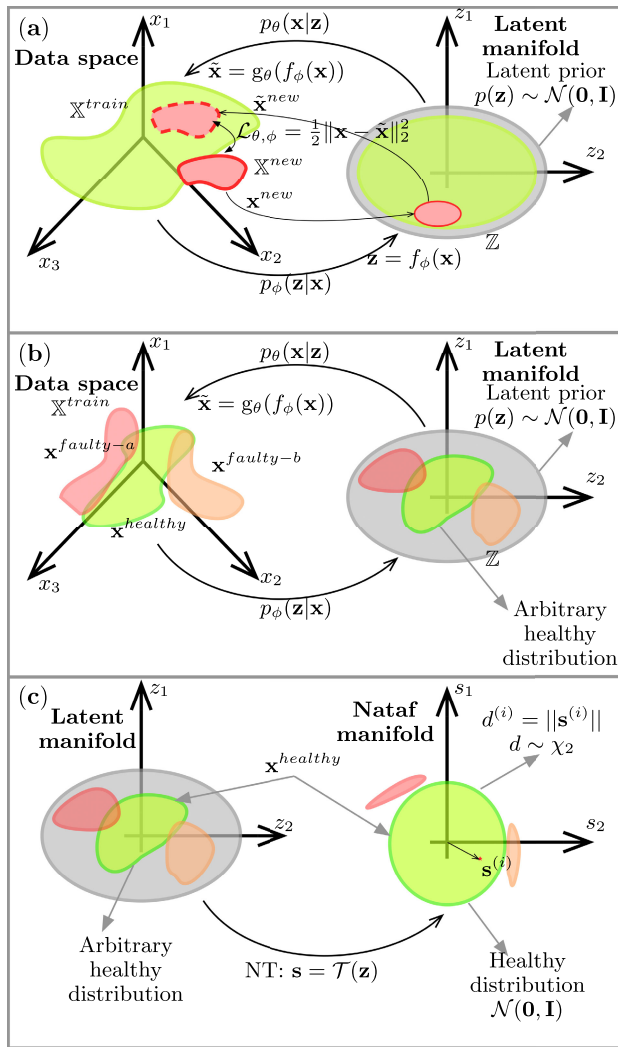
FIGURE 1. Latent variable models: (a) The reconstruction process as a suitable HI; (b) Training datasets include healthy and faulty samples; (c) The proposed Nataf inference process is defined with respect to healthy samples.

B. LATENT VARIABLE MODEL FOR FAULT DETECTION

It is important to note that when the two parametric functions (g_θ, f_ϕ) are learned on a training dataset $\mathbb{X}^{train}$ and the hyperparameters (θ, ϕ) are set, the posterior latent distribution obtained by mapping a new data sample $\mathbb{X}^{new}$ will belong to the prior manifold $p(\mathbf{z})$ even if $p(\mathbf{x}^{new}) \not\approx p(\mathbf{x}^{train})$. The reconstruction data $\tilde{\mathbf{x}}^{new} = g_\theta(f_\phi(\mathbf{x}^{new}))$ obtained by both transition functions will have a posterior probability data distribution that somewhat approximates the training probability distribution $p(\tilde{\mathbf{x}}^{new}) \approx p(\mathbf{x}^{train})$. See Fig. 1(a). This reconstruction process is very suitable for fault detection if the LVMs learn the mapping of the healthy probability distribution only.

For instance, let $p(\mathbf{x}^{healthy})$ be the healthy data distribution and $p(\mathbf{x}^{faulty})$ the faulty distribution for a given anomaly such as $p(\mathbf{x}^{healthy}) \not\approx p(\mathbf{x}^{faulty})$. If the LVM has learned to map only the healthy data space $\tilde{\mathbf{x}}^{healthy} \sim p(\mathbf{x}^{healthy})$,

the reconstruction process obtained for any faulty data will give $\tilde{\mathbf{x}}^{faulty} \sim p(\mathbf{x}^{healthy})$. The divergence between the original data distribution $p(\mathbf{x})$ and the reconstructed data distribution $p(\tilde{\mathbf{x}})$ is then used as a suitable HI HI $= \mathcal{L}(\mathbf{x}, \tilde{\mathbf{x}})$. This concept is illustrated in Fig. 1(a), where the distribution of the reconstructed faulty samples is shifted from the original faulty data. Thus, exploiting the reconstruction error is a simple and effective way of detecting any shift in the probability distribution of the monitoring data space caused by an anomaly.

The simplicity of reconstruction-error-based detection is thanks to the fact that the model captures the characteristics of a healthy state. Furthermore, it does not require analyzing complex failure data, or any expert-dependent labeling. This concept involving the calculation of the difference of the residual outputs between monitoring and normal states has been used as a system-wide HI in several papers [24], [25], [26], [27], [28], [29], [30], [31], [32], [33].

C. LIMITATIONS OF THE LATENT MANIFOLD FOR FAULT DETECTION

In the case where only healthy data are inferred on the latent manifold, the inference transition function is less sensitive to anomalous data than is the generative transition function. Indeed, in most papers in which the latent manifold is used for fault detection, the main HI is a combination of the data HI and the latent HI [34], [35], [36], [37], [38], [39], [40].

The main advantage of the inference process over the generative process is that the latent manifold improves the interpretability of the fault detection model. Indeed, the disentangled latent manifold provides analysts and O&M practitioners with a visual tool. The representation of the LVs within time gives a graphical description for the evolution of the system's physical condition. Improving the sensitivity of the latent manifold for fault detection is thus of foremost importance.

Several latent manifold conceptual indicators could be incorporated to improve the detection of any deviation from the learned healthy state in the latent space [41]. These latent conceptual indicators are the likelihood-based indicator, the discriminative indicator, the statistical distance-based indicator, the prototypical approaches, and the multivariate signal processing approaches [41]. In [42], Balshaw et al. conducted a comparative study evaluating several latent manifold conceptual indicators. The authors concluded that data space HIs outperform latent space HIs. It was also highlighted that the temporal structure of the data must be considered to increase the sensitivity of the latent space to the anomalous data. This temporal preservation of the latent manifold involves the use of a sliding time window W_t in the latent space from which the latent HI is developed, i.e., the Latent HI (LHI) can be written as LHI $= F(\mathbf{z}_t)$, where $F(.)$ is a latent manifold conceptual indicator and $t \in [t_1, t_2]$. The temporal preservation of the latent variables demonstrated

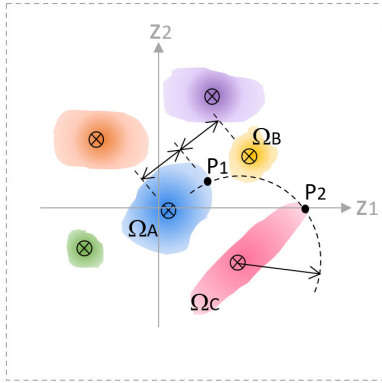**FIGURE 2.** Illustration of a 2D LVM-latent space distribution for a multiple condition database. $\otimes$: Cluster centroid.

better performance in detecting anomalies in vibration signals than in the conventional static case [43].

Furthermore, to improve the interpretability of the latent manifold, it would be more appropriate to include some failure data during the learning process. This can be done in a semi-supervised way, with the expert analyzing some failure data before the learning process, or in an unsupervised way, where the expert analyzes the clusters obtained by the inference function (after the training).

One of the restrictions on the LVM is that the inference process is unique to a given asset and very sensitive to the obtained hyperparameters set during the training. Moreover, the posterior latent distribution of the healthy data is unknown. Because of these limitations of the LVM, a universal LHI suitable for several assets cannot be defined. As shown in Fig. 1(b), all the samples of the training set (faulty and healthy) lie within the same manifold, while the healthy distribution is not known a priori.

In addition, the posterior latent distribution could be unsuitable for some of the conceptual indicators of the manifold, as depicted in Fig. 2. Indeed, in this figure, it can be noted that point $\mathbf{P}_1$ is equidistant to the centroids of the clusters Ω_A and Ω_B, but $\mathbf{P}_1 \in \Omega_A$. Also, the Euclidean distance is the same between the cluster Ω_C centroid and the points $\mathbf{P}_1$ and $\mathbf{P}_2$, even if $\mathbf{P}_2 \in \Omega_C$ and $\mathbf{P}_1 \notin \Omega_C$. Both cases suggest that the relations between points and clusters in the latent space cannot be described using solely the Euclidean distance.

D. TOWARD AN ENHANCED LATENT MANIFOLD FOR FAULT DETECTION

We introduce here an approach that leads to a standard and known posterior healthy distribution. As described in this section and illustrated further in this paper, our proposition allows to overcome the limitations of classical latent space representations. Ultimately, the enhanced representation space is used in the definition of HI for the detection of abnormal conditions in complex systems.

To achieve a standardized representation space, a second-step inference transition function is introduced to map the latent manifold into a new manifold, as shown in Fig. 1(c). The inference function is the NT [18], [44], which is presented in detail in section III.

The NT takes the healthy dataset from the latent space as a reference. Let $\mathbb{Z}_{\text{healthy}}$ be the set of healthy data points in the latent manifold. The NT maps $\mathbb{Z}_{\text{healthy}}$ into $\mathbb{S}_{\text{healthy}}$ such that $\mathbb{S}_{\text{healthy}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. Given a data point $\mathbf{s}^{(i)} \in \mathbb{S}_{\text{healthy}}$, let $d^{(i)} = ||\mathbf{s}^{(i)}||_2$, where $|| \cdot ||_2$ is the Euclidean norm, and $D = d^{(i)}$ is the corresponding random variable. Since $\mathbb{S}_{\text{healthy}} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, the random variable D follows the χ distribution with N degrees of freedom χ_N. Specifically, the norm of points in healthy condition in the standard Nataf space is characterized by the density function $f_{\chi_N}(d)$ and the cumulative distribution function (cdf) $F_{\chi_N}(d)$ [45]. This property allows to define HI based on the comparison between the Nataf space projection of datasets and the healthy reference condition. The latter is the same for any asset. Therefore, the NT provides a standard reference for defining a universal HI. Moreover, standard and universal thresholds can be set to quantify the degradation over time. Section IV presents a condition monitoring approach based on the standard Nataf space manifold.

III. NATAF ISOPROBABILISTIC TRANSFORM

The NT was introduced in [18] and is an important tool in reliability analysis [46]. Its goal is to transform an intricate probabilistic description into a more tractable one [47]. This section briefly describes the NT.

Let the random vector $\mathbf{Z} = [Z_1, \ldots, Z_N]^T$ and the correlation matrix $\mathbf{R}_Z$ be the probabilistic model describing the complex system, where each component corresponds to one specific feature or measure. The NT maps $\mathbf{Z}$ into the target description $\mathbf{S} = [S_1, \ldots, S_N]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $\mathbf{0}$ is the null vector (with length N), and $\mathbf{I}$ is the identity matrix (with size $N \times N$). Fig. 3 depicts the NT for a two-dimensional latent space.

The NT can be written as the composition of two transformations $\mathcal{T} = \mathcal{T}_1 \circ \mathcal{T}_2$ [47]. First, $\mathcal{T}_1 : \mathbf{Z} \mapsto \mathbf{U}$ is given by (1), where Φ is the cdf of the normal distribution $\mathcal{N}(0, 1)$, and $F_i : \mathbb{R} \to [0, 1]$ is the marginal cdf of Z_i. The function F_i is known a priori, or otherwise, can be estimated from samples with a kernel density estimator.

$$U_i = \Phi^{-1}(F_i(Z_i)) \qquad i \in \{1, \ldots, N\} \tag{1}$$

The random vector $\mathbf{U}$ resulting from (1) is assumed to be a Gaussian vector with marginal distributions $\mathcal{N}(0, 1)$. The components $U_1, \ldots, U_N$ are characterized by the correlation matrix $\mathbf{R}_U \in \mathbb{R}^{N \times N}$, which is referred to as the fictive correlation matrix. The transformation $\mathcal{T}_2$ maps $\mathbf{U}$ into the random vector $\mathbf{S}$ with uncorrelated components. For this, let the matrix $\mathbf{L} \in \mathbb{R}^{N \times N}$ be the Cholesky decomposition of the inverse of $\mathbf{R}_U$, as in:

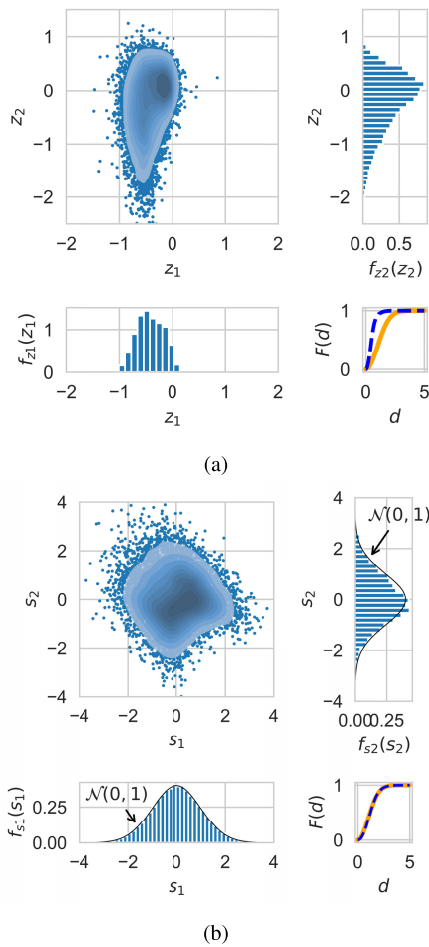$$\mathbf{L}\mathbf{L}^T = \mathbf{R}_U^{-1}. \tag{2}$$

FIGURE 3. The NT maps the latent space distribution $\mathbf{Z} = [Z_1, Z_2]^T$ in (a) into $\mathbf{S} = [S_1, S_2]^T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ in (b). The cdf $F(d)$ of the distance metric d is plotted (continuous orange line) against the χ_2 cdf (dashed blue line) for each space.

The transformation $\mathcal{T}_2 : \mathbf{U} \mapsto \mathbf{S}$ is given by (3), and the space of the variable $\mathbf{S}$ is referred to as the standard Nataf space:

$$\mathbf{S} = \mathbf{LU}. \tag{3}$$

Estimating the fictive correlation matrix components $r_{Uij} = \{\mathbf{R}_U\}_{ij}, i, j \in \{1, \ldots, N\}$, is a key step in the definition of $\mathcal{T}_2$. The appendix presents the numerical method allowing to estimate r_{Uij} from the correlation coefficients $r_{Zij} = \{\mathbf{R}_Z\}_{ij}$, $i, j \in \{1, \ldots, N\}$. Also, as discussed in the Appendix, the computation of r_{Uij} does not converge when the components of the healthy cluster distribution are highly correlated, i.e., $r_{Zij} \approx 1$. In such a case, the NT fails to map the initial distribution into a multivariate distribution with non-correlated components [44].

IV. PROPOSED CONDITION MONITORING APPROACH

This section introduces a new CM approach aimed at the early detection of abnormal conditions in complex systems. The proposed approach is based on HIs defined from the

standard Nataf space representation. The implementation of the approach comprises two phases: online and offline. The offline phase defines the building blocks that later constitute the online phase's pipeline. Fig. 4 depicts these two phases of the implementation with their respective steps, which are described in detail in the following sections.

A. DATABASE DEFINITION AND LATENT SPACE REPRESENTATION

We assume that the complex system is described by multiple measures. Although LVMs such as the VAEs are capable of processing many features and suit feature learning [7], a feature engineering analysis is recommended to select the pertinent measures. In particular, recent literature suggests that the latent space representation is enhanced by restricting the features to non-correlated measures with high informative power [48]. The offline phase of the proposed approach requires labeled data in two steps: first, training the VAE model (or similar) and then defining the NT. In the training process, the need for labels depends on the complexity of the system and on the available measures describing the state of the complex system. The final latent space representation must be such that the healthy cluster is as disentangled as possible from other condition clusters. For simpler systems or high informative features, the unsupervised VAE can eventually project the different conditions into disentangled clusters. However, the VAE is usually not sufficient to disentangle categorical clusters for complex real-life systems or relatively noisy measures. To ensure the disentanglement of clusters in the latent representation space for these cases, supervised (or semi-supervised) implementations of the VAE are recommended.

The latent space representation is strongly influenced by the model hyperparameters, and therefore, these hyperparameters can be used to adjust the distribution of clusters in that representation. Choosing suitable hyperparameters leads to disentangled clusters in the latent space. Additionally, to ensure the convergence of the NT, the healthy cluster coordinates must be weakly correlated, i.e., have small correlation coefficients $r_{Uij}, i, j \in \{1, \ldots, N\}$. This requirement is further discussed in the Appendix.

In summary, to apply our approach, the database and its latent space representation must verify three conditions to ensure that the NT is successfully defined.

- The training database must be labeled, or at least, the healthy data must be identified from the degraded conditions.
- In the latent representation space, the cluster of healthy points must be disjoint with respect to any other cluster.
- In the latent representation space, the healthy cluster latent variable coordinates must be weakly correlated.

Overall, these requirements are easily met when the appropriate model is chosen for the latent space projection. As discussed later in this text, using the NT broadens the range of suitable hyperparameters. In other words, the use
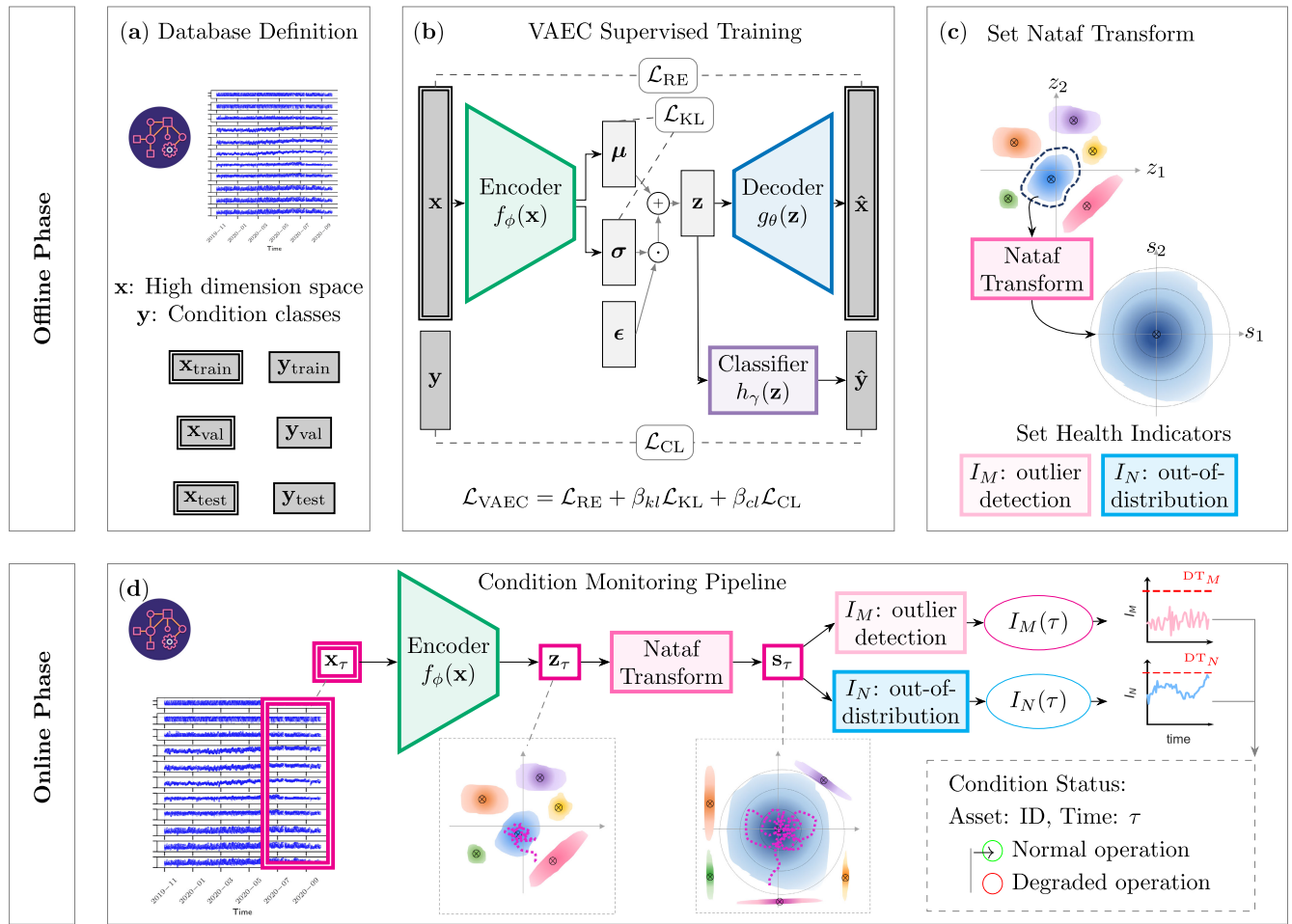
FIGURE 4. Framework of the proposed condition monitoring approach for a system operating under multiple conditions. The offline phase comprises three steps: (a) definition and labeling of the database, (b) training of the VAE embedded with Classification (VAEC) model, and (c) definition of the NT. The online phase (d) estimates the condition status from the normalized dataset $\mathbf{x}_\tau = \{\mathbf{x}(t), t \in \tau\}$.

of the NT mapping relaxes the requirements on the model defining the latent space representation.

B. STANDARD NATAF SPACE REPRESENTATION

The NT is defined with respect to the healthy condition cluster, as depicted in Fig. 4(c). It then maps any point in the latent space into the standard Nataf space. Fig. 5 illustrates a typical standard Nataf space for a multiple condition system.

From the latent space representation, the definition of the NT comprises the following steps:

1) Characterize the healthy cluster in the latent space database. Specifically, estimate the empirical cdfs corresponding to the LVs $Z_1, \ldots, Z_N$, and the correlation matrix $\mathbf{R}_Z$ between these LVs.

2) Estimate the fictive correlation matrix $\mathbf{R}_U$ from $\mathbf{R}_Z$ using the algorithm presented in the Appendix.

3) Validate the NT. In the standard Nataf space, verify that the healthy cluster approximates the multivariate normal distribution and that the norm of all healthy points approximates the χ_N distribution.
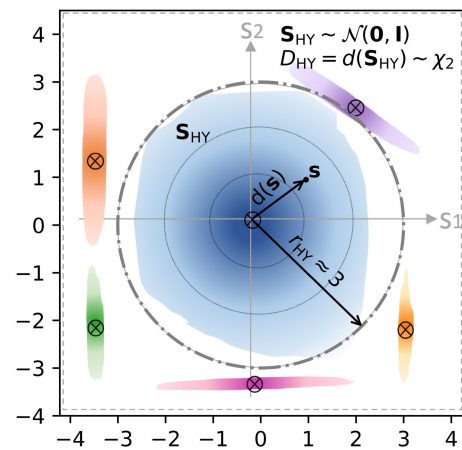

FIGURE 5. Illustration of a 2D standard Nataf space projecting a multiple condition dataset. $\mathbf{s} = [s_1, s_2]^T$ is a point and $d(\mathbf{s}) = ||\mathbf{s}||_2$. The random vector $\mathbf{S}_{HY}$ describes the distribution of the reference healthy cluster.

If the computation of the fictive correlation matrix does not converge, one should re-train the VAE-like model to obtain non-correlated latent space variables.
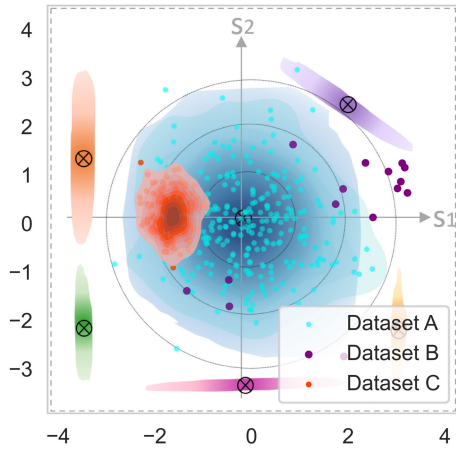
FIGURE 6. Projection of datasets in the standard Nataf space. Dataset A represents a healthy dataset, whereas B and C correspond to two patterns of degradation.

The Euclidean norm in the standard Nataf space is highly representative of the relation between data points and the reference dataset. For instance, points distributed similarly to the reference condition in the physical space are mapped by the NT within the reference cluster in the standard Nataf space. More importantly, the projection of a healthy dataset into the standard Nataf space gives a cluster distributed as the healthy reference (see Dataset A in Fig. 6). For reference, the healthy cluster radius is defined as the percentile P_{99} of the χ_N distribution. In a 2D standard Nataf space, for example, this radius is $r_{HY} = P_{99}[\chi_2] \approx 3$.

Conversely, abnormal datasets projected in the standard Nataf space are inconsistent with the normal multivariate distribution of healthy conditions. Indeed, degraded datasets can manifest in the standard Nataf space in two different patterns: a sudden and rapid degradation, or continuous and slow degradation. In the first type, illustrated by Dataset B in Fig. 6, the complex system reaches failure within a few time steps. In such a case, the norm of successive data points goes rapidly from typical values to above the radius r_{HY}. The second degradation type is illustrated by Dataset C in Fig. 6. This degradation pattern manifests as a shift in the distribution of points in the standard Nataf space. In this case, the norm of the projected points does not necessarily go beyond the reference radius r_{HY}.

Accordingly, we introduce below two complementary HIs using the Euclidean norm of points in the standard Nataf space: I_M and I_N. The HI I_M aims at detecting outliers more sensitive to rapid degradation with a few monitoring samples, whereas the HI I_N is an out-of-distribution detector more sensitive to the slow degradation, but that needs more samples within sliding time windows.

C. HEALTH INDEX I_M FOR OUTLIER DETECTION

The HI I_M is the Euclidean norm of the point $\mathbf{s}(t)$ in the standard Nataf space, as given by (4).

$$I_M(t) = d(\mathbf{s}(t)) \tag{4}$$

The base alarm criterion for I_M is defined by the threshold rule $I_M(t) > \mathrm{DT}_M$, where DT_M is a detection threshold to be set. Since the distribution of the norm statistics for healthy datasets is known to be the χ_N, a general DT_M value can be set. For example, one can set $\mathrm{DT}_M = P_{99}[\chi_N]$, which is valid for any kind of complex system. Also, for any system whose data is projected in a 2D Nataf space ($N = 2$), $\mathrm{DT}_M = P_{99}[\chi_2] = 3.0$ is a suitable value. An alternative method for defining the value of DT_M consists in using historical data from healthy systems. First, the I_M is estimated for systems known to be in healthy condition. Then, DT_M can be set as a percentile of the distribution of the estimated I_M. This work analyzes these two techniques to define DT_M.

It is recommended to associate the base detection criterion with a persistence criterion to limit the frequency of false alarms. For example, (5) establishes that an alarm is triggered when the base criterion is met four consecutive times. In practice, the alarm base and persistence criteria should be adapted to match the characteristics of the system of interest.

$$I_M(t_i) > \mathrm{DT}_M \text{ for } i \in \{k, k-1, k-2\} \tag{5}$$

By definition, I_M allows detecting degradation modes evolving within a few time steps. Depending on the domain of the complex system under analysis, the detection anticipation might not be sufficient to allow planning and action by O&M analysts and practitioners. Nevertheless, the pertinence of I_M might be studied case by case. Conversely, the HI I_N introduced in the next section is expected to detect the onset of degradation modes earlier than competing approaches.

D. HEALTH INDEX I_N FOR OUT-OF-DISTRIBUTION DETECTION

Many HIs in the literature are based on out-of-distribution detection [15], [49], [50], [51]. To our knowledge, no previous work has exploited the out-of-distribution detection in the standard Nataf space. The HI I_N introduced in this section was precisely designed to detect shifts in the probability distribution of datasets in the standard Nataf space.

Let τ be a sliding period in which the complex system of interest operates. Let $\mathbf{S}_\tau$ be a random variable describing the distribution of the projection of the dataset of interest in the standard Nataf space, within the sliding period τ. Finally, let $F_\tau(d)$ be the corresponding empirical cdf of $D_\tau = d(\mathbf{S}_\tau)$. The HI I_N is defined as the area metric between the analytical cdf $F_{\chi_N}(d)$ and the empirical cdf $F_\tau(d)$, as given by (6) and illustrated in Fig. 7. In practice, the area metric can be estimated from samples of both distributions by using the Wasserstein metric (or Kantorovich–Rubinstein metric) [52], [53]:

$$I_N(\tau) = \int_0^{+\infty} |F_{\chi_N}(x) - F_{D_\tau}(x)| dx. \tag{6}$$

The definition of I_N as the area metric between two cdfs was motivated by the following considerations: (i) the reference normal multivariate distribution exhibits a spherical symmetry, and therefore, no pertinent information is lost
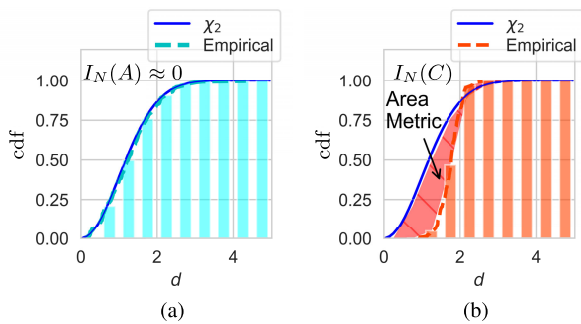
FIGURE 7. Estimation of I_N from the pdf of d(s) for the datasets (a) A and (b) C from Fig. 6.

when the analysis is based on the norm of points; (ii) the reference distribution F_{χ_N} is well known; (iii) the proposed I_N does not depend on the dimension N of the latent space, and (iv) the area metric is an interpretable HI.

The base alarm criterion for I_N is given by $I_N(\tau) > \mathrm{DT}_N$, where DT_N is a detection threshold to be defined. The value of DT_N can be defined by considering the distribution of I_N estimated from systems known to be in healthy condition. This distribution of I_N depends on the space dimension N as well as on the length of the period τ.

Combining this base criterion with a persistence criterion is important to prevent false alarms. The alarm criteria retained for the HI I_N is given in (7). Again, these criteria might be modified according to the characteristics of the system under analysis.

$$I_N(\tau_k) < \mathrm{DT}_N \text{ and } I_N(\tau_k) > I_N(\tau_{k-1}) > I_N(\tau_{k-2}) \quad (7)$$

E. ONLINE CONDITION MONITORING

The online phase is depicted in Fig. 4(d). The steps of the online phase are as follows:
1) VAE embedded with Classification (VAEC) encoding of the normalized SCADA dataset $\mathbf{x}_\tau$ into the set of latent space points $\mathbf{z}_\tau$.
2) NT-mapping of $\mathbf{z}_\tau$ into the standard Nataf points $\mathbf{s}_\tau$.
3) Estimation of the two HIs from $\mathbf{s}_\tau$.
   - The HI I_M is a point-wise index. It is estimated at each time step $t_k \in \tau$.
   - The HI I_N takes $\mathbf{s}_\tau$. It is estimated for each τ.
4) Application of alarm criteria on the estimated HIs I_M and I_N. Combine both statuses with the "OR" logical relation, i.e., an alarm is triggered when any of the alarm criteria is met.

The online pipeline calculations are repeated periodically over periods defined by partitioning the timeline. We assume that the system data is defined over the time steps $t_0, t_1, \ldots$ A period of operation τ can be identified by using its starting and final time steps: $\tau(k, k+n) = \{t_k, t_{k+1}, \ldots, t_{k+n}\}$. The corresponding normalized dataset is then $\mathbf{x}_{\tau(k,k+n)} = \{\mathbf{x}(t_k), \ldots, \mathbf{x}(t_{k+n})\}$. The partition of the timeline into datasets $\tau(k, k+n)$ must meet

two opposing requirements. On the one hand, the datasets must be big enough to allow describing the probability distribution of clusters in the latent and the Nataf spaces. On the other hand, the promptness of the detection depends on how rapidly new data points are fed into the online pipeline. Overlapping datasets can be used to satisfy both conditions. The partition of the timeline is then defined by the length of the datasets and the lag time between the beginning of two successive datasets. A sensitivity analysis is recommended for an appropriate choice.

F. PERFORMANCE OF THE CONDITION MONITORING APPROACH

To evaluate the performance of a detection approach, it is customary to implement it on a selection of reference study cases for which the ground-truth detection instants are known. If multiple case studies of reference are available, statistics such as accuracy, precision, F1 score, and Recall can be estimated (see [54]).

However, very often, only a few case studies are available. In such cases, one can prioritize case-specific evaluations. One primary performance metric for detection approaches is the anticipation of failure, i.e., the time between the instant when the detection approach triggers an alarm and the instant of failure [55]. Since the instant of failure is rarely available, the anticipation is often estimated with a reference approach. The anticipation interval of an alarm is important for the viability of the detection approach. Indeed, the alarm anticipation must be significant enough to allow in-situ interventions, be it by shutting the system down, inspecting it, or scheduling necessary repairs.

V. CASE STUDY I: ENGINE DEGRADATION FROM C-MAPSS DATABASE

This section demonstrates the proposed detection approach in the Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) FD001 dataset [56]. This dataset comprises a run-to-failure time series of commercial aircraft turbofan engines in cruise condition. C-MAPSS datasets are widely used by the PHM community, and have indeed come to constitute benchmarking datasets for multiple CM and prognosis approaches, notably the Remaining Useful Life (RUL) estimation [57]. In this case study, we apply the proposed HIs to detect the onset of abnormal conditions defined as a function of the ground-truth RUL values. The RUL estimation itself is beyond the scope of the present paper.

A. THE C-MAPSS FD001 DATASET

The C-MAPSS database comes from high-fidelity computational simulations of the operation of a large commercial turbofan. This case study considers the FD001 dataset, which comprises run-to-failure time series corresponding to the degradation of the turbofan in cruise condition [58]. Sensors indicate measures at the main turbofan sub-components, namely the fan, the Low-Pressure Compressor (LPC), the
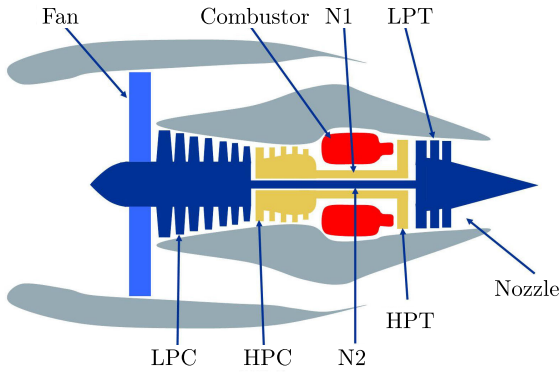
FIGURE 8. Simplified diagram of the turbofan engine from NASA's C-MAPSS. Source: [59].

High-Pressure Compressor (HPC), the combustor, the High-Pressure Turbine (HPT), the Low-Pressure Turbine (LPT), and the Nozzle. These components are depicted in Fig. 8 [58].

The dataset comprises three kinds of data:

- Scenario descriptors: include major flight data such as altitude, flight Mach number, and throttle-resolver angle.
- Sensors: include 14 measurements covering temperature, total pressure, mass flux, and pressure ratio at multiple sections of the turbofan.
- cycle counting from 1st cycle at the beginning of the simulation until the failure cycle $n_{failure}$.

The RUL function is set as a piecewise linear degradation curve. For each run-to-failure time series, it is defined from the cycle counting data. Fig. 9 depicts the typical RUL function $RUL(n) = \min\{n_{failure} - n, RUL_0\}$, where the constant initial value is set as $RUL_0 = 125$.
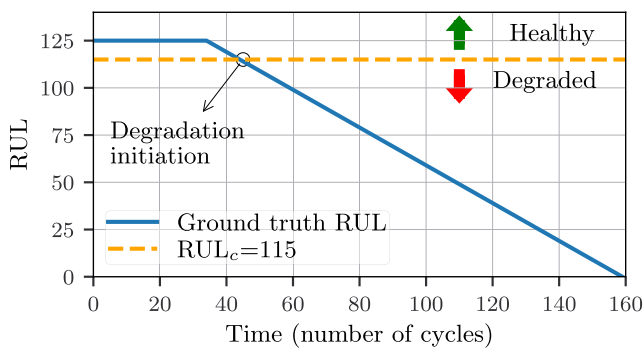

FIGURE 9. Target RUL and predicted RUL with the LSTM-VAE-Regression model.

Moreover, to define the detection problem from the C-MAPSS FD001 dataset, the critical RUL value RUL_c is introduced as the threshold defining a degraded condition. That means that the relation $RUL < RUL_c$ indicates the onset of a degrading condition. Indeed, ground-truth RUL functions decrease linearly from this critical threshold. The reference degradation initiation cycle is simply the cycle at which the

relation $RUL < RUL_c$ is observed. For example, according to this criterion, the degradation starts at the 45th cycle in Fig 9 with $RUL_c = 115$.

B. LATENT SPACE PROJECTION

The latent space representation of the C-MAPSS datasets has been analyzed in works such as [17], [60], and [61]. The latent space projection introduced by [17] is suitable for demonstrating the approach we propose in the present work. Costa and Sanchez [17] introduced a DNN model consisting of an LSTM-VAE embedded with a regression DNN. The latent space is set with dimension $N = 2$. The resulting latent space projection is such that the distribution of points is an indicator of the RUL value. We refer the reader to the original paper for details on the LSTM-VAE and the regression DNN, as well as architecture and training parameters [17]. Nevertheless, the original model led to a cluster with mostly correlated latent variable components z_1 and z_2, thus hindering the definition of the NT. To overcome this limitation, the loss function was modified by increasing the weight of the Kullback-Liebler loss component. Fig. 10 depicts the resulting latent space projection of the training database.
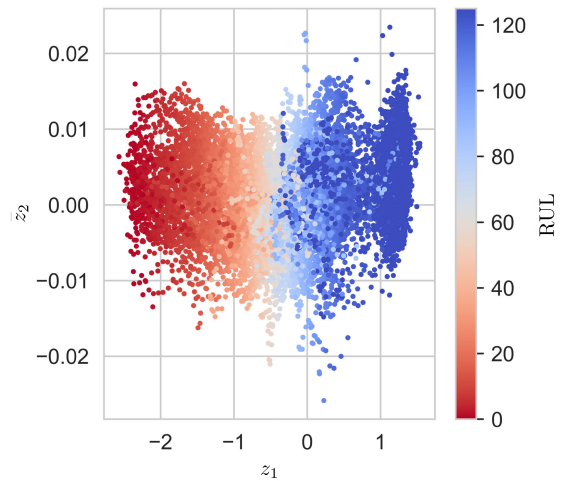

FIGURE 10. Latent space representation of the C-MAPSS FD001 dataset. This 2D latent space corresponds to the LSTM-VAE-Regression model.

C. STANDARD NATAF SPACE AND DETECTION CRITERIA

The reference healthy operational condition dataset comprises the training points for which the threshold relation $RUL > RUL_c$ is met. This dataset is used as a reference for the definition of the NT. Fig. 11 depicts the mapping of the healthy dataset into the standard Nataf space.

The distribution of the norm of all points from the reference healthy cluster in the standard Nataf space is depicted in Fig. 12. It confirms a good agreement between the empirical cdf and the χ_2 cdf F_{χ_2}.

To define the detection threshold for the two HIs, I_M and I_N, the distributions of these indices are estimated for the
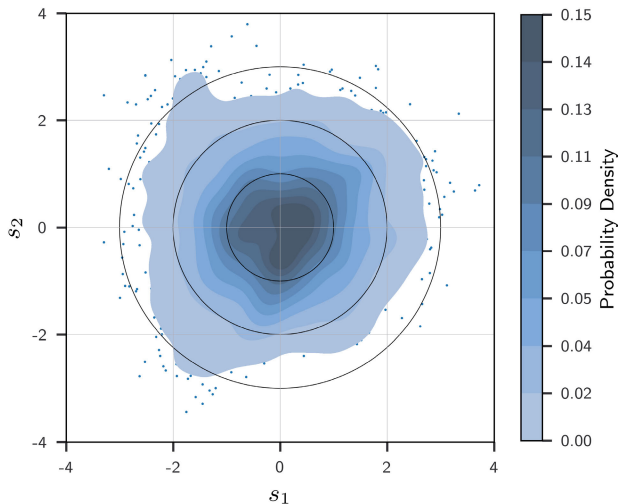
FIGURE 11. Standard Nataf space projection of the healthy condition data (RUL > RUL$_c$).
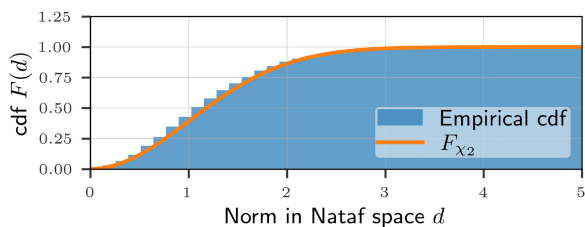

FIGURE 12. Distribution of the distance metric for the healthy condition cluster in the standard Nataf space.
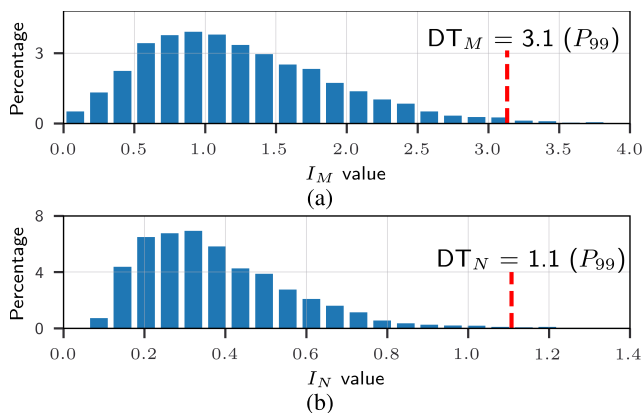

FIGURE 13. Distribution of the HIs calculated on the healthy dataset: (a) I_M and (b) I_N.

healthy data points, and the detection thresholds are set as the 99th percentile P_{99} of each distribution. Fig. 13 depicts the distributions for the two HIs obtained for the FD001 dataset. According to these experimental results, the alarm thresholds are respectively set as DT$_M = 3.1$ and DT$_N = 1.1$.

In this case, the standardization of the healthy condition in the Nataf space eases the definition of DT$_M$. For instance, since the I_M is strictly equal to the norm of points, the HI I_M evaluated on a healthy dataset follows the χ_N
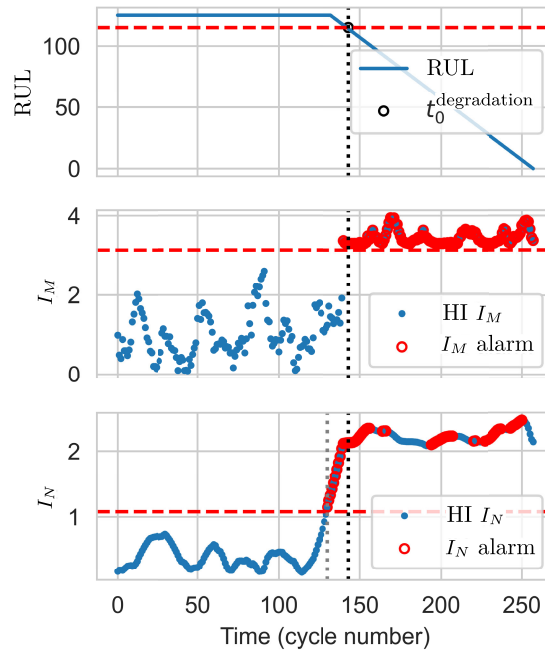

FIGURE 14. HI I_M, I_N and the corresponding alarm status for turbofan 2 of the C-MAPSS FD001 dataset.

distribution, and thus, DT$_M$ can be set as a percentile of the χ_N distribution. Given $N = 2$, the percentile P_{99} gives DT$_M = P_{99}[\chi_2] = 3.0$, which is consistent with the value from the empirical distribution of I_M.

D. EARLY DETECTION OF ABNORMAL CONDITIONS

Once the LSTM-VAER model is trained and the NT defined, new datasets can be projected into the standard Nataf space. The estimation of the two HI I_M and I_N for engine 2 is given in Fig. 14. The reference onset of degradation for this engine occurred at the 143rd cycle. The alarm criteria led to the I_M alarm at the 140th cycle and the I_N alarm was triggered at the 130th cycle. Fig. 15 depicts the curves for I_M and I_N for engine 50. The reference onset of degradation for this engine occurred at the 56th cycle. The HI I_M detected the abnormal condition at the 53rd cycle and I_N led to an alarm at the 38th cycle.

Fig. 16 depicts the I_M detection time versus ground-truth degradation onset time for all the engines of the FD001 dataset. The HI I_M led to the same detection instants as the target values (ground-truth degradation initiation) for most cases, with no significant anticipation. This result is attributable to the characteristics of the degradation function. As specified, we assumed that RUL $<$ RUL$_c$ indicated the onset of degradation, and that the ground-truth RUL function is a linear function of the cycle counting.

From Fig. 17, it can be seen that the detection by I_N anticipates the reference detection instants for almost all engines. Considering all the engines of the C-MAPSS FD001 dataset and the degradation defined above, the HI I_N led to the
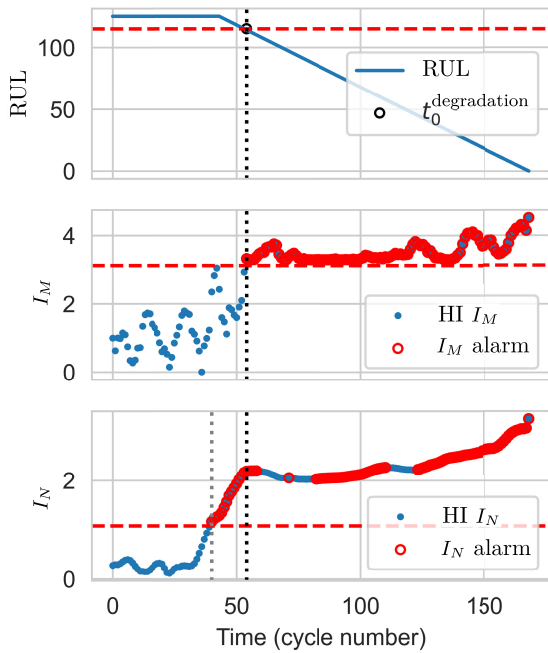
FIGURE 15. HI I_M, I_N and the corresponding alarm status for turbofan 50 of the C-MAPSS FD001 dataset.

FIGURE 16. Comparison between the reference degradation initiation instant and the HI I_M instants of detection.
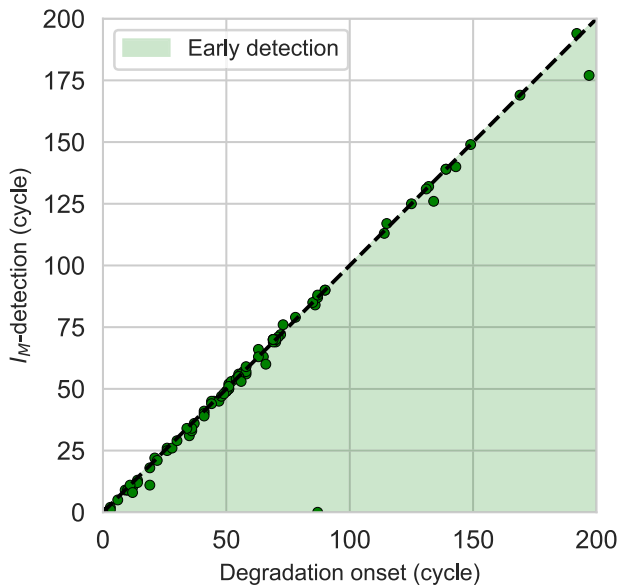

FIGURE 17. Comparison between the reference degradation initiation instant and the HI I_N instants of detection.

early detection of degraded conditions, with an anticipation of the reference degradation instant in 12 cycles on average.

VI. CASE STUDY II: WIND TURBINE CONDITION MONITORING

This second case study demonstrates the implementation of the proposed approach on a real-life database from a North American wind farm. Modern WTs are by default equipped with the Supervisory Control and Data Acquisition
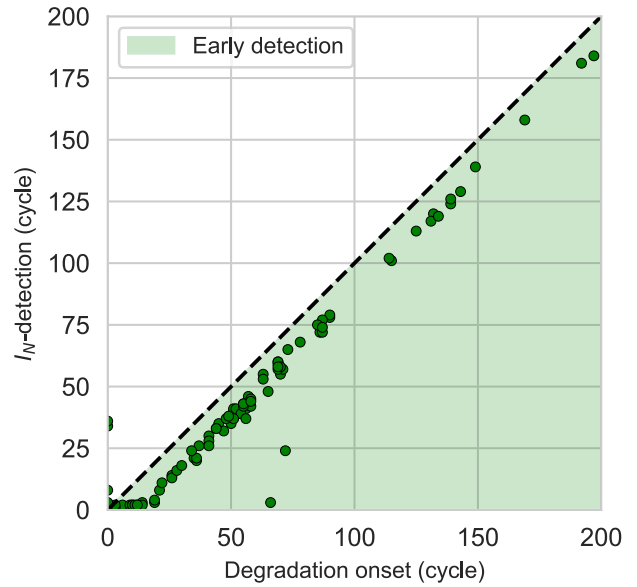
(SCADA) system, which generates multiple data describing the WT operation. Analyzing the high-dimensional and non-linear physical spaces resulting from the SCADA sensors is a challenging task. Latent space representations are widely used in the WT CM literature [13], [15], [62], [63].

In particular, the authors of the present paper analyzed the case of WT CM in [15]. Our previous proposition used the Mahalanobis distance to define an HI from a latent space representation. This definition allowed a good detection performance, but required extra effort in the definition of a convenient latent space. The selection of suitable hyperparameters can be tiresome and time-consuming for complex systems such as WTs. Also, the lack of comparability between independent training instances was pointed out as a limitation. We show in the case study below that the approach proposed in the present paper outperforms the previous one and finally achieves a standardized low-dimension representation for WTs by using the NT. This second case study demonstrates the suitability of our approach to analyze real-life industrial cases.

A. WIND TURBINE SCADA DATABASE

This case study uses a proprietary database from a North American wind farm. This database contains data on over 100 onshore WTs and covers operations spanning 2 years and four months. The following are the specifications of the WTs: horizontal axis; upwind; three-blade rotors; pitch-controlled; rated power of order 2 MW; cut-in wind speed $V_{in} = 3.5$ m/s; rated wind speed of $V_r = 13$ m/s, and cut-off wind speed $V_{out} = 20$ m/s.

The database comprises three kinds of information: measures from the built-in SCADA system, SCADA log files, and O&M reports. The SCADA system is composed

of sensors covering measures of geometrical, kinematic, thermal, and electrical characteristics. For each measure, the available data correspond to the average value over 10-minute periods, which is particularly appropriate for performance monitoring [64]. It is worth mentioning that vibration and acoustic measurements are not available for the WTs under analysis. The log files indicate warnings and alarms generated by the SCADA system usually based on simple threshold rules. The O&M reports are completed by the O&M practitioners based on in-field observations during inspections and repairs.

Temperature-related abnormal conditions such as the overtemperature of critical components were reported in the database and are commonly investigated in the literature using temperature measures from the SCADA system ([65], [66]). Table 1 lists the temperature-related conditions analyzed in the present paper. The color codes are defined for subsequent use in this text. The reader is referred to [55] and [63] for illustrations of some of these degradation modes.

TABLE 1. Temperature-related WT conditions with the respective color codes.

ID	Description	Color code
HY	Reference healthy condition	
BEA	Main bearing overtemperature	
GBX	Gearbox oil overtemperature	
GEN	Generator overtemperature	
ICE	Ice accretion on blades	

In all, 35 SCADA measures were made available in the database. The statistical analysis of the time series corresponding to these measures revealed some variables with little or no informative power, e.g., time series with mostly non-numerical values, constant-value time series, as well as sets of highly correlated time series. This work retained the 11 measures listed in Table 2. The selection includes multiple temperature measures that provide key information on the characterization of the abnormal behaviors listed in Table 1. The preprocessing of the SCADA measures consists of clipping and min-max normalization of data using the lower and upper bounds indicated in Table 2. These values follow from the statistical analysis of measures from all the WTs taken together. Data points with measures falling outside the interval [LB, UB] are removed from the database. The features are normalized into the [0, 1] interval.

A labeled dataset is required for the supervised learning of the retained LVM, the VAEC model. The labeling step consists of a semi-manual selection of data points based on the indications from the SCADA log files. This step was guided by industry experts and used considerations of the physical nature of each temperature-related condition. For example, to build the dataset representing the GBX (Gearbox oil overtemperature) condition, we analyzed the evolution of the key measures WS, n_{ROTOR}, P, $T_{GBX-BEA}$, $T_{GBX-OIL}$, T_{AMB}, and T_{NAC}. The starting point to select data points

TABLE 2. SCADA measures with the respective lower (LB) and upper (UB) bounds.

Measure (Unit)	Description	LB	UB
WS (m/s)	Wind speed	0	31
n_{ROTOR} (rpm)	Rotor speed	0	18
P (kW)	Active power	0	2000
T_{AMB} (°C)	Ambient Temp.	-25	45
T_{NAC} (°C)	Nacelle Temp.	-20	70
T_{BEA} (°C)	Main Bearing Temp.	-20	70
$T_{GBX-BEA}$ (°C)	Gearbox Bearing Temp.	0	100
$T_{GBX-OIL}$ (°C)	Gearbox Oil Temp.	0	100
T_{GEN1} (°C)	Generator Temp. Pos. 1	-10	140
T_{GEN2} (°C)	Generator Temp. Pos. 2	-10	140
$T_{GEN-COOL}$ (°C)	Generator Cooling Temp.	-10	120

corresponding to the GBX condition is the time period of gearbox overtemperature reported in the SCADA log files. Often, $T_{GBX-BEA}$ and or $T_{GBX-OIL}$ become increasingly high even before the starting instant indicated in the SCADA log files. Thus, the selected intervals were usually larger than the intervals of abnormal conditions indicated in the SCADA log files.

Given the large dispersion of the SCADA measures, many data points are required to build statistically representative datasets. Moreover, given an operating WT, only a few degradation cases, if any, are observed. Thus, it is not possible to build datasets for all conditions using data from only one unit. To overcome this limitation, we build datasets by gathering data points from multiple WTs. These WTs are supposed to have similar degradation modes since they are of the same model, are subject to the same manufacturing standards, and operate under similar wind and environmental conditions. In practice, the CM approach can be built for a subset of units from a large wind farm. Dataset augmentation techniques can be used if the number of data points for any dataset is still insufficient. Fig. 18 displays the datasets of interest in the normalized power curve. The ICE cluster is well separated from the other conditions and, therefore, can be detected directly from the power curve [67]. Nevertheless, the datasets corresponding to the other critical conditions are mostly superposed, and their detection is not as
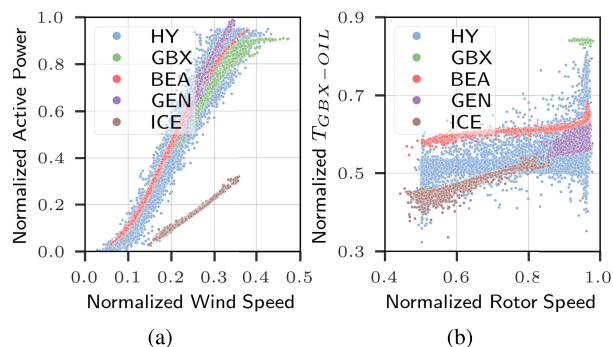

FIGURE 18. Distribution of the datasets of interest in two plots: (a) normalized power curve, (b) normalized gearbox temperature versus normalized rotor rotation speed.

straightforward as for the ICE condition. The ICE condition is included in this case study for validation of the proposed approach.

B. VAEC MODEL ARCHITECTURE AND TRAINING

The proposed approach was implemented in Python (ver. 3.10). The VAEC model is built using the TensorFlow library and the Keras API [68]. The supervised VAEC model was adopted since the classical (and unsupervised) VAE is not sufficient to obtain a latent space representation with disentangled clusters [15]. In accordance with the data, the input (and output) dimension is $N_X = 11$ and the classifier DNN has its output dimension $N_C = 5$. The latent space dimension is set to $N = 2$. This choice allows to easily display the representation spaces in plots, thus defining a built-in visualization. The retained architecture is described as follows:

- Encoder: three hidden layers and one dropout layer. The number of nodes per layer is, successively, 11 (input layer), 32, 16, 8, and 2 (output layer). The input layer is set with the ReLU activation function, and tanh is used in the remaining layers of the encoder. Moreover, a 10% dropout layer is placed after the 32-node layer to prevent overfitting.
- Decoder: 2, 8, 16, and 32 nodes per layer. The decoder output layer is set with the linear activation function. The other decoder layers are set with the tanh activation function.
- Classification DNN: the input of the classification DNN is the latent space with dimension 2. The successive hidden layers have a decreasing number of nodes: 128, 64, 32, and 16. The tanh activation function is used in the input and in the hidden layers. Finally, the classification output is a five-node layer using the Softmax activation function.

The supervised training uses the Adam algorithm with a learning rate of 0.0001, clip value of 0.3, number of epochs 1024, and batch size set to 128 [69].

C. SUITABLE MODEL HYPERPARAMETERS AND NATAF TRANSFORM

One of the advantages of the proposed approach is that it relaxes the requirements for the training hyperparameters. To demonstrate this claim, this section describes the influence of the loss function hyperparameters. The VAEC loss function is given by (8), where $\mathcal{L}_{RE}$ is the reconstruction error loss, $\mathcal{L}_{KL}$ is the Kullback-Liebler loss, and $\mathcal{L}_{CL}$ is the classifier loss. The coefficients $\beta_{kl} \geq 0$ and $\beta_{cl} \geq 0$ allow to adjust the weight of each loss component in the VAEC training:

$$\mathcal{L}_{VAEC} = \mathcal{L}_{RE} + \beta_{kl}\mathcal{L}_{KL} + \beta_{cl}\mathcal{L}_{CL}. \tag{8}$$

The loss function coefficients are hyperparameters that strongly affect the latent space distribution. For instance, Fig. 19 depicts three training instances of the VAEC latent
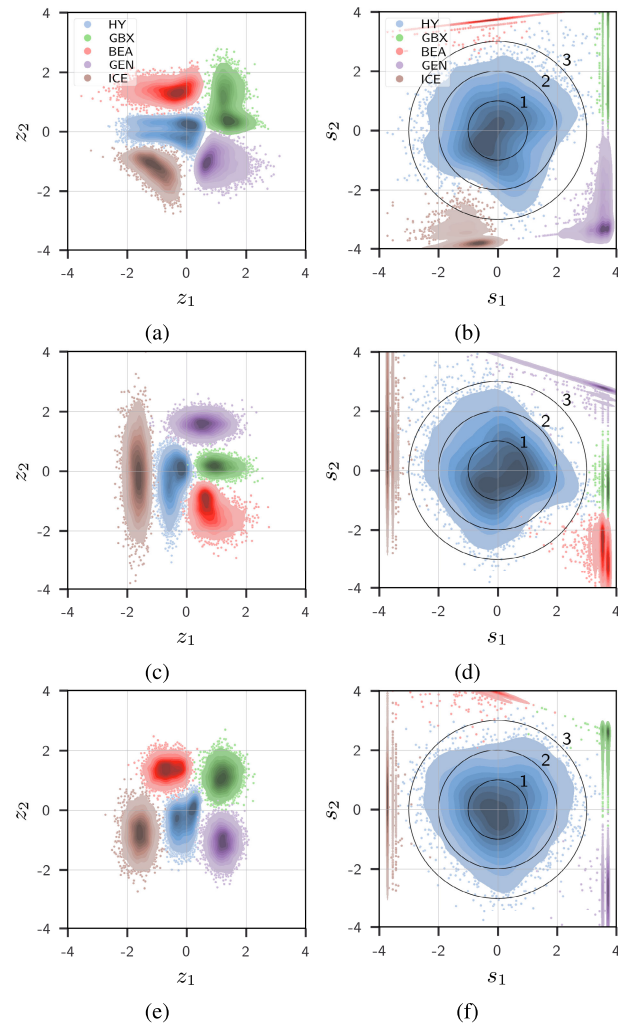

FIGURE 19. Latent space and standard Nataf space for independent training instances: (a-b) $\beta_{kl} = 0.001$, $\beta_{cl} = 0.01$, (c-d) $\beta_{kl} = 0.01$, $\beta_{cl} = 0.1$, (e-f) $\beta_{kl} = 0.01$, $\beta_{cl} = 0.1$. In each cluster, the color gradient indicates the probability density of points, with darker shades corresponding to higher density.

space and the respective standard Nataf spaces defined taking the healthy cluster HY as the reference distribution. Note in Fig. 19(a), 19(c), and 19(e) that the latent space representations are highly variable, and this is true even when the coefficients are the same. On the contrary, the standard Nataf space gives a standardized representation of the WT healthy condition.

As a corollary, the range of suitable training hyperparameters is broader when using the NT, as compared to the latent representation based solely on the variational encoding of the data. It was indeed shown that different loss function coefficient choices lead to very different latent space outcomes, whereas in the Nataf space, the reference data invariably follows the multivariate normal distribution. The coefficients retained for the VAEC model training are $\beta_{kl} = 0.01$ and $\beta_{cl} = 0.1$. Hereafter, we use the reference standard Nataf space depicted in Fig. 19(e-f).

D. DETECTION THRESHOLDS

The point-wise estimation of $d(\mathbf{s}(t))$ gives a time series with a 10-minute time step that is highly dispersed. In the WT industry, it is common to use daily averages as a means to regularize HIs [55]. The definition of the I_M is adapted accordingly, as follows:

$$I_M(t) = \frac{1}{n_{\text{DAY}}} \sum_{t \in \text{DAY}} d(\mathbf{s}(t)). \qquad (9)$$

Regarding the estimation of the out-of-distribution HI I_N, a sliding period $\tau = 4$ days is used with a lag time $\Delta_{\text{lag}} = 1$ day (overlap of three days between two successive periods). The detection thresholds DT_M and DT_N were set from a selection of ten datasets covering the operation of WTs known to be in healthy condition. Fig. 20 depicts the distribution for both HIs I_M and I_N estimated from the selection of healthy datasets. As for the C-MAPSS experiments, the detection thresholds were set as the percentile P_{99} of each distribution, which corresponds to $DT_M = 3.0$ and $DT_N = 1.2$, respectively.
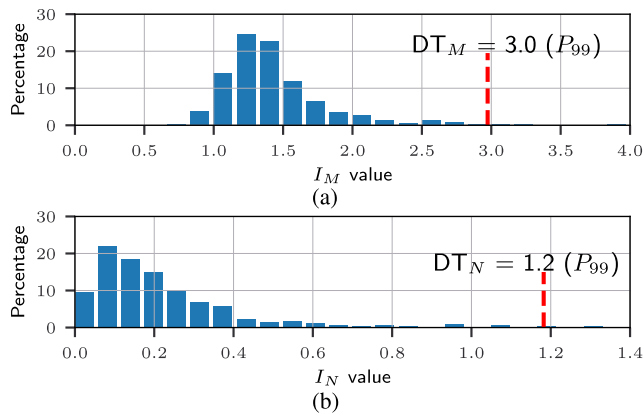

FIGURE 20. Distribution of the HIs calculated on a selection of ten WTs in healthy condition: (a) I_M. (b) I_N.

Fig. 21(a) and 21(b) depict the HI I_M and I_N for one of the ten datasets considered in the estimation of the healthy distribution. Both HI time series remain below the respective detection thresholds. Thus, as expected, no alarm is triggered for this healthy WT.

E. CONDITION MONITORING OF A WIND TURBINE IMPACTED BY MULTIPLE ABNORMAL CONDITIONS

This case study covers the operation of one WT within the period $[t_0 = 2019 \text{ Oct 1}, t_f = 2020 \text{ Aug 3}]$. The specific WT was selected from the wind farm under analysis because it had been reported to be suffering from multiple critical abnormal conditions; in particular, blade ice accretion (ICE) and main bearing overtemperature (BEA). Therefore, this WT allows to demonstrate the capacity of the proposed approach in detecting multiple types of abnormal conditions. The timeline of events was reconstructed with industry experts based on information from the SCADA log files and O&M reports. The following are the main events in the timeline:

- $t_0 = 2019$ Oct 1: Beginning of the period of analysis.
- $t^*_{\text{ICE}} = 2020$ Feb 11: SCADA alarm ICE condition.
- $t_{\text{Down1}} = 2020$ Feb 24: Shutdown labeled as "Grid voltage fault", and then as "Communication Loss". Three days later, an unsuccessful start-up attempt is labeled by the SCADA system with multiple failure modes, including the BEA condition. The unit is then kept down with "Repair" status for more than 40 days.
- $t_{\text{StartUp}} = 2020$ Apr 09: WT start-up.
- $t^*_{\text{BEA}} = 2020$ Jun 22: SCADA alarm BEA condition.
- $t_{\text{Down2}} = 2020$ Aug 3: Shutdown due to BEA-related failure.
- $t_f = 2020$ Aug 3: End of the period of analysis.

Fig. 21(c) depicts the I_M time series estimated from the SCADA data using (9). The alarm criteria (given by (5) with $DT_M = 3.0$) lead to a first alarm at $t = 2020$ Feb 11. This alarm superposes the occurrence of ice accretion on blades that were indicated by the SCADA system. An I_M alarm is then triggered days later, at $t = 2020$ Feb 24, suggesting the WT was in abnormal condition just before it was shut down for repair. The SCADA system log files report multiple possible causes for the degradation. This unit remained shut down for 45 days with "Repair" status. The actions undertaken during this period were not disclosed. The WT starts operating again at $t_{\text{StartUp}} = 2020$ Apr 09. After this date, the I_M values go above the threshold DT_M erratically. It is only after $t = 2020$ Jun 01 that I_M becomes persistent and, thus, triggers an alarm. The I_M alarm at $t = 2020$ Jun 01 is 28 days before the SCADA alarm at t^*_{BEA}.

The estimation of the HI I_N over the period of interest gives the $I_N(\tau_k)$ time series depicted in Fig. 21(d). As previously mentioned, each τ_k covers four consecutive days of operation, and the lag time between two successive datasets is of one day ($\tau = 4$ days, $\Delta_{\text{lag}} = 1$ day). The alarm criteria (given by (7) with $DT_N = 1.2$) led to an alarm at $t = 2020$ May 22, which is 31 days before t^*_{BEA} and more than 2 months before t_{Down2}.

Regarding the ICE episode in February 2020, although the base criterion $I_N > DT_N$ was met multiple times in the freezing days around t^*_{ICE}, the persistence criterion filtered out any alarm during this month. Eventually, one could adjust the alarm criteria or use specific control charts to enable alarms for rapidly evolving degradation modes. Nevertheless, in this work, the I_N alarm criteria were set to prioritize the detection of degradation evolving over the long term, i.e., within multiple days, in opposition to the HI I_M alarm, which focuses on degradation modes evolving over the short term. The asset condition status results from the combination of the two HIs with the OR logical relation. Fig. 22 indicates the resulting alarm time series raised by the proposed approach for the WT under analysis.

To illustrate the rationale of the HI I_N in a real-life case, the estimation of this HI is visually represented in Fig. 23 (See page 17). The steps for the estimation of I_N are depicted for the periods highlighted in Fig. 23(a), where the periods $\tau_k, k \in \{A, B, C, D\}$, are:
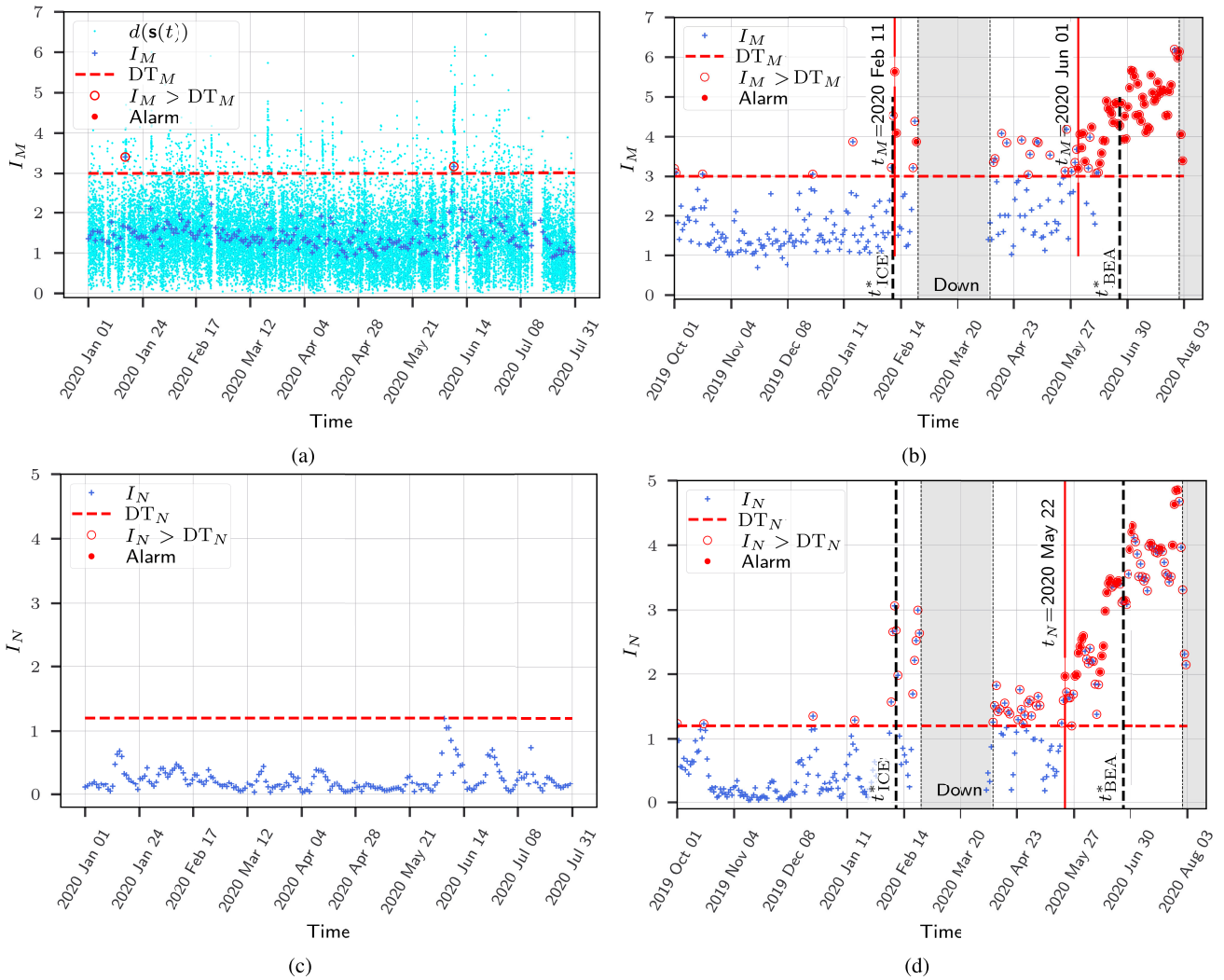
FIGURE 21. For a healthy WT: estimation of (a) I_M and (b) I_N. For a WT with reported main bearing overtemperature: (c) I_M and (d) I_N with and the respective alarms.
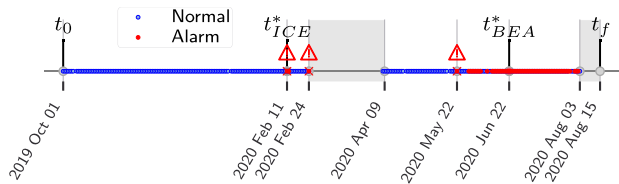

FIGURE 22. Timeline indicating dates of events and alarms in the case study. Gray fill indicates WT shut down. The proposed approach detected abnormal conditions starting at $t = 2020$ Feb 11 (coincides with t^*_{ICE}), $t = 2020$ Feb 24 (which equals t_{Down1}), and $t = 2020$ May 22 (which is 31 days before t^*_{BEA}).

- $\tau_A = \tau(2019$ Oct 31, 2019 Nov 04).
- $\tau_B = \tau(2020$ Jan 10, 2020 Jan 14).
- $\tau_C = \tau(2020$ Apr 18, 2020 Apr 22).
- $\tau_D = \tau(2020$ Jun 07, 2020 Jun 11).

Each row of Fig. 23 shows the evolution of the abnormal condition over the four selected datasets. In the latent space representation in Fig. 23(b), it is remarkable that the dataset

distribution evolves from the healthy cluster toward the main bearing overtemperature cluster (BEA). This progression is also visually evidenced in the standard Nataf space in Fig. 23(c), and even more so in the area metric between the empirical cdf and the χ_2 cdf in Fig. 23(d).

Finally, Fig. 24(a) and 24(b) depict the time evolution of the WT projection into the latent and Nataf spaces, respectively. Each point in these plots is the centroid of the projection of a four-day dataset. The timeline is indicated with the color map, which gives the visualization of the WT condition as a trajectory in the representation spaces. The blue cluster corresponds to the healthy HY condition and the red cluster to the BEA condition. Note that the transition from the HY cluster toward the BEA cluster is evidenced in both plots.

F. COMPARATIVE ANALYSIS OF THE PERFORMANCE FOR DETECTING MAIN BEARING OVERTEMPERATURE

This section focuses on the detection of the BEA condition that raised the SCADA alarm at t^*_{BEA} and led to the WT
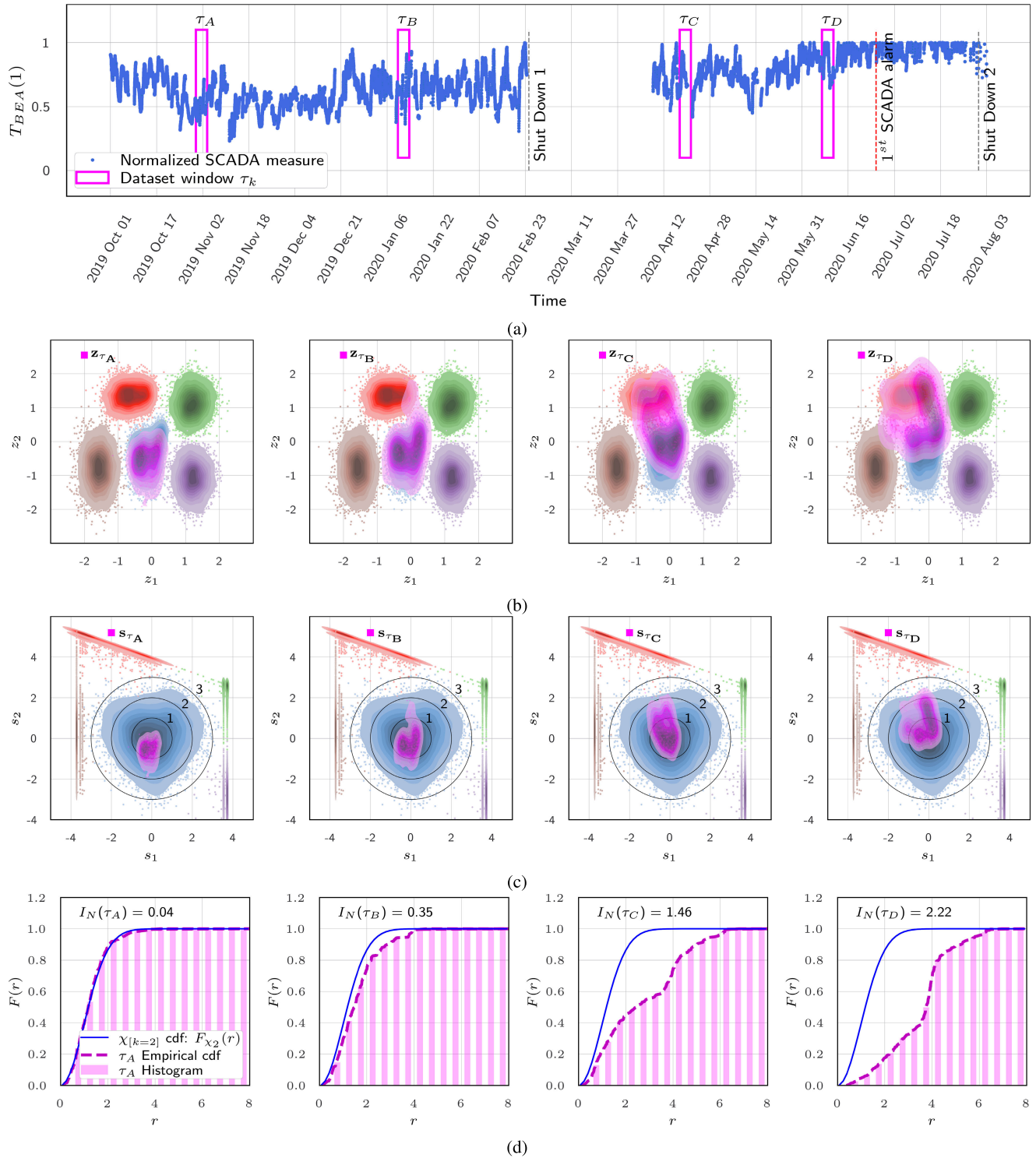
FIGURE 23. Estimation of the HI $I_N(\tau_k)$ for 4-day periods τ_k, $k \in \{A, B, C, D\}$ indicated (a) in the T_{BEA} time series; (b) Encoding of the datasets of interest τ_k (in magenta) on the latent space; (c) NT of the datasets of interest τ_k on the standard Nataf space; (d) Estimation of the HI I_N from the empirical cdf.

shutdown at t_{Down2}. To evaluate the performance, Table 3 summarizes competing detection approaches and gives the respective detection dates for the dataset analyzed above. The approaches from [55] and [70] detect main bearing

degradation exclusively. In [71], a classical unsupervised VAE model is used.

From Table 3, the anticipation obtained from the proposed approach outperforms that estimated using previous

FIGURE 24. Projection of the case study main bearing overtemperature as a trajectory in (a) the latent space and (b) the standard Nataf space. (c) Timeline color-map.

TABLE 3. WT main bearing overtemperature detection approaches.

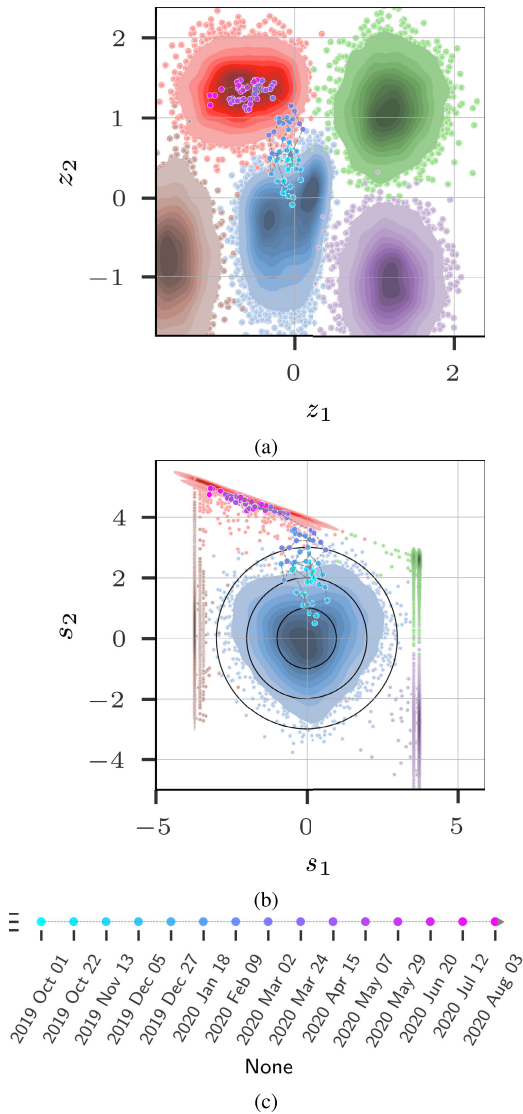| Detection approach reference and description	Detection date
SCADA system alarm [64] The SCADA system raises alarms based on threshold rules for the main bearing temperature.	$t^*_{BEA} = 2020$ Jun 22
Residue of linear multivariate model [55] (i) Detection of WT main bearing overtemperature based on the residue between predicted and actual daily averages of the main bearing temperature $\hat{e} = \hat{T}_{BEA} - T_{BEA}$; (ii) $\hat{T}_{BEA}$ is estimated from a linear multivariate model whose parameters are identified using healthy condition data; (iii) The Exponentially Weighted Moving Average (EWMA) control chart is used to detect the BEA condition.	$t_{D1} = 2020$ Jun 21
Mahalanobis distance in the VAEC latent space [15] (i) multiple condition detection for WT combining the VAEC encoding and the Mahalanobis distance; (ii) A labeled database comprising multiple conditions is used to train the VAEC model. The VAEC encoder then projects the WT SCADA measures into a disentangled latent space; (iii) The Mahalanobis distance is used to measure the distance between the distributions in the latent space of new datasets of interest and the reference healthy condition dataset; (iv) The EWMA control chart is used to detect trends in the daily average of the Mahalanobis distance.	$t_{D2} = 2020$ Jun 15
Approach introduced in this paper (i) Multiple condition detection for complex systems combining a suitable LVM and the NT; (ii) An LVM is trained with the pertinent database such that the latent space representation comprises disentangled clusters and the reference healthy condition is well characterized; (iii) The NT is defined using the healthy cluster as reference, thus defining the standard Nataf space projection; (iv) Two HIs follow from the norm of data points projected into the Nataf space: outlier HI I_M and the out-of-distribution HI I_N; the two HIs are combined to produce the system condition status.	$t_{D3} = 2020$ May 22

C-MAPSS FD001 dataset into a 2D latent space. Turbofan engines in this study case were considered as healthy or degraded. For the second case study, WTs SCADA data were projected into a 2D latent space using the VAEC model. Its supervised training used a labeled database comprising five conditions - the healthy condition plus four abnormal conditions.

It is worth noting that the VAE model alone might be sufficient to obtain a disentangled LV representation for simpler systems or systems with highly informative features. As long as the healthy condition is identified in the latent space and the LVs are not highly correlated, the NT can be defined. In most real-life complex systems, however, supervised implementations of LVM might be necessary to obtain a suitable LV representation, i.e., with disentangled clusters. The complexity of the systems analyzed in the case studies and the nature of the available measures guided the choice of supervised LVMs. Fortunately, the labels are required only in the offline phase. Once the LVM model and the NT are set, the online phase relies only on the updated features to give the condition output for the complex system being monitored.

works [15], [55], [70]. The gain in performance by I_M is mostly due to the use of the labeled database to train the VAEC model. This supervised training allowed to have a disentangled LVM latent space that was then mapped into the standard Nataf space. The performance of the proposed approach is further discussed in the next section.

VII. SUMMARY AND DISCUSSION

This section summarizes and discusses the results from case studies I and II.

The success of the proposed approach in obtaining a standardized LVM representation for complex systems was evidenced within the two case studies analyzed in this paper, which used latent space representations based on supervised LVMs to achieve disentangled latent representations. The LSTM-VAE-Regression model was chosen to project the

The analysis in this paper used 2D representation spaces ($N = 2$). This setting facilitates plotting and visualization. It was shown that the resulting standard Nataf spaces verified the expected properties in both case studies. Indeed, in both cases, the healthy cluster follows the 2D normal distribution and the norm of its points follows the χ_2 distribution. The HIs I_M and I_N were defined from the norm in the 2D standard Nataf space. One could consider setting higher latent space dimensions if needed. For higher N values, the properties of the standard Nataf space and the definition of the two HIs are essentially the same, except for the degree of liberty N of the χ_N distribution.

As expected, the HIs I_M and I_N gave complementary outcomes. The HI I_M allowed to detect outliers of the normal distribution. It was proven to be particularly suitable for detecting abnormal conditions evolving within a few time steps, such as the ICE condition in case study 2. In practice, the margin of action for rapidly evolving degradation modes is limited, and the detection of such events is of limited interest. Nevertheless, the HI I_M is retained in the proposed approach to ensure the robustness of the combined HI I_M OR I_N.

The two case studies revealed that the HI I_N is suitable for degradation modes evolving in the long term, and was able to detect the onset of abnormal conditions long before these conditions became critical. In the C-MAPSS FD001 engines database, I_N predicted the reference detection 12 cycles in advance, on average, whereas I_M gave results similar to the ground-truth times of degradation initiation. In the WT case study, I_N anticipated the SCADA alarm for the BEA condition in 31 days, outperforming the I_M detection and multiple competing approaches. The earlier detection of abnormal conditions by I_N is a promising result given that the longer anticipations give more time for O&M planning and interventions.

The choice of detection criteria might be adapted according to the application of interest and the accuracy and false alarm rate requirements. Moreover, a sensitivity analysis is recommended for optimal adjustments of settings such as τ and Δ_{lag}. The persistence criteria chosen for I_N in this paper were such that the HI focused on degradation modes evolving in the long term. That explains why no I_N alarm was raised for the rapidly evolving ICE condition in February 2020. The association of the two HIs allowed to maintain relatively simple base and persistence criteria. Regarding the detection thresholds, the results from the two case studies suggest that the definition of the threshold DT_M from the theoretical distribution χ_N is pertinent. Indeed, the theoretical value $\text{DT}_M = P_{99}[\chi_2] = 3.03$ is very close to the values obtained from the empirical probability distribution for healthy engines ($\text{DT}_M = 3.1$) and for healthy WTs ($\text{DT}_M = 3.0$). Analogously, similar DT_N values were estimated for the two case studies: $\text{DT}_N = 1.1$ for the healthy engines and $\text{DT}_N = 1.2$ for the healthy WTs.

The enhanced performance of the detection based on the standard Nataf space can be associated with two main aspects of the proposed approach. First, supervised learning was used in both case studies as a strategy to obtain disentangled clusters in the LVM latent spaces. In this regard, the proposed approach requires more information than the competing approaches analyzed in section VI-F. Second, the NT can be interpreted as a regularized zoom-in on the reference healthy condition. As a result, shifts in the original physical space are amplified in the standard Nataf space, leading to stronger indications of the onset of degradation modes.

Plots of the 2D LVM latent space and of the standard Nataf space supported the arguments and analysis presented in this paper. It was shown that, for the C-MAPSS FD001 data, the position in the latent space is related to the level of degradation of the engine. For the VAEC latent space trained to reproduce the WT SCADA data, the position in the latent space suggested the condition to which a data point belongs. The case studies showed that the standard Nataf space preserves the capacity to visually represent the evolution of the condition of complex systems. Finally, in Fig. 24(a) and 24(b), it was shown that the evolution of the system condition can be represented as a trajectory in the LV and Nataf spaces. Where the LVM latent space identifies multiple conditions, that can potentially provide diagnosis information since abnormal points move toward the corresponding abnormal condition clusters. Exploring this aspect, as well as a quantitative assessment of the improvements in interpretability, represents potential avenues for future research.

VIII. CONCLUSION

This work addressed the LVM-based representation of complex systems. The main contribution of this paper is the introduction of a standardized latent space representation for complex systems by using the Nataf isoprobabilistic transform. It was shown that, contrary to the highly variable LVM representation, the standard Nataf space preserves the probability distribution of the healthy condition, namely, the multivariate normal distribution. The proposed CM framework can be applied to a variety of complex systems. This versatility was demonstrated within two case studies covering distinct technological domains.

The standardization of the LVM representation allows to compare representation spaces corresponding to multiple assets. Most importantly, two HIs, I_M and I_N, were defined from the characterization of the standard Nataf space. The proposed approach combines the two HIs to achieve robust detection. I_N allows to detect the onset of abnormal conditions significantly earlier than competing approaches. Complementarily, I_M is a point-wise HI that captures degradation modes evolving within a few time steps. For example, in case study 2, I_M detected ice accretion on blades, which was not detected by I_N. The two case studies demonstrated the performance and potential of the proposed detection approach.

As a secondary contribution, this paper introduced two examples of applications for the emergent field of inter-

TABLE 4. 7-Point Gauss–Hermite quadrature.

Roots (x_i)	Weights (w_i)
± 3.75043971768	0.000548268858737
± 2.36675941078	0.0307571239681
± 1.1544053948	0.240123178599
0.0	0.457142857143

pretability of DNN models. For instance, the two case studies illustrate how the LVM latent spaces and the standard Nataf space can be used to visually represent the evolution of complex systems. These visualizations complement the condition status information and can, therefore, enhance the interpretability and trust of the outcomes among O&M analysts and practitioners.

APPENDIX. NATAF TRANSFORM

This Appendix presents the numerical method allowing to estimate the fictive correlation coefficients $r_{Uij} = \{\mathbf{R}_U\}_{ij}$ of the NT.

A. ITERATIVE COMPUTATION OF THE NATAF CORRELATION COEFFICIENTS

Let $r_{Zij} = \{\mathbf{R}_Z\}_{ij}$ be the correlation coefficient between the physical space variables Z_i and Z_j, $i, j \in \{1, \ldots, N\}$. The relation between the correlation coefficients r_{Uij} and r_{Zij} for given $i, j \in \{1, \ldots, N\}$ is given by (10), as shown at the bottom of the page. In this integral equation, μ_{Zi} and σ_{Zi} are the average and standard deviation of Z_i, respectively, and φ_2 is the probability density for a bi-variate Gaussian distribution with correlation r_{Uij}, as given by (11), as shown at the bottom of the page, [44].

The m-points Gauss–Hermite quadrature is employed for the approximation of (10), which is given by (12), as shown at the bottom of the page. The integral equation can then be recursively solved by using Algorithm 1. It takes r_{Zij} as input to estimate r_{Uij}. The settings are the error tolerance $\epsilon_{\text{tol}} > 0$, the maximum number of iterations $N_{\max} \gg 1$, and the Gauss–Hermite quadrature order m with the respective roots and weights. For example, Table 4 gives the Gauss–Legendre quadrature roots and weights for the order $m = 7$ [44].

Algorithm 1 Nataf Fictive Correlation Coefficient Evaluation

Input: r_{Zij}

Settings: ϵ_{tol}, $N_{\max}$, Gauss–Hermite quadrature order m with respective roots $\{s_1, \ldots, s_m\}$ and weights $\{p_1, \ldots, p_m\}$

Output: r_{Uij}

1: $t \leftarrow 0$
2: $r_{Uij}^{(t)} \leftarrow r_{Zij}$
3: $\epsilon \leftarrow \infty$
4: **while** $|\epsilon| > \epsilon_{\text{tol}}$ and $t < N_{\max}$ **do**
5:     $t \leftarrow t + 1$
6:     $R^{(t)} \leftarrow \begin{pmatrix} 1 & r_{Uij}^{(t-1)} \\ r_{Uij}^{(t-1)} & 1 \end{pmatrix}$
7:     $\hat{L}^{(t)} \leftarrow$ Cholesky decomposition of $R^{(t)}$
8:     $\begin{pmatrix} u_{i,\ell}^{(t)} \\ u_{j,k}^{(t)} \end{pmatrix} \leftarrow \hat{L}^{(t)} \begin{pmatrix} s_{i,\ell}^{(t)} \\ s_{j,k}^{(t)} \end{pmatrix}, \forall \ell, k \in \{1, \ldots, m\}$
9:     $\begin{pmatrix} z_{j,k}^{(t)} \\ z_{j,k}^{(t)} \end{pmatrix} \leftarrow \hat{L}^{(t)} \begin{pmatrix} F_i^{-1}[\Phi(u_{i,\ell}^{(t)})] \\ F_j^{-1}[\Phi(u_{j,k}^{(t)})] \end{pmatrix}, \forall \ell, k \in \{1, \ldots, m\}$
10:     $r_{Zij}^{(t)} \leftarrow \sum_{\ell=1}^{m} \sum_{k=1}^{m} p_\ell p_k \left(\frac{z_{i,\ell} - \mu_{Zi}}{\sigma_{Zi}} \right) \left(\frac{z_{j,k} - \mu_{Zj}}{\sigma_{Zj}} \right)$
11:     $\epsilon \leftarrow r_{Zij} - r_{Zij}^{(t)}$
12:     $r_{Uij}^{(t)} \leftarrow r_{Uij}^{(t-1)} + \epsilon$
13: $r_{Uij} \leftarrow r_{Uij}^{(t)}$
14: **if** $t \geq N_{\max}$ **then**
15:     $r_{Uij}^{(t)} \leftarrow NaN$
16:     **print** "Did not converge."
17: **end**

B. CONVERGENCE CONDITIONS

The joint probability density function (pdf) $\varphi_2(u_i, u_j, r_{Uij})$ is not defined when $|r_{Uij}| = 1$, i.e., when Z_i and Z_j are correlated variables (See (11)). More generally, the integral equation given by (10) is not guaranteed to have a solution when the correlation coefficient is close to one [47]. In practice, Algorithm 1 does not converge in such cases, which means that there is no uncorrelated set of variables that is equivalent to the original set in terms of probability description.

$$r_{Zij} = \mathbb{E}\left[\left(\frac{Z_i - \mu_{Zi}}{\sigma_{Zi}} \right) \left(\frac{Z_j - \mu_{Zj}}{\sigma_{Zj}} \right) \right] = \mathbb{E}\left[\left(\frac{F_i^{-1}(\Phi(U_i)) - \mu_{Zi}}{\sigma_{Zi}} \right) \left(\frac{F_j^{-1}(\Phi(U_j)) - \mu_{Zj}}{\sigma_{Zj}} \right) \right]$$

$$= \frac{1}{\sigma_{Zi}\sigma_{Zj}} \int \int_{\mathbb{R}^2} \left\{ [F_i^{-1}(\Phi(u_i)) - \mu_{Zi}][F_j^{-1}(\Phi(u_j)) - \mu_{Zj}]\varphi_2(u_i, u_j, r_{Uij}) \right\} du_i du_j \tag{10}$$

$$\varphi_2(u_i, u_j, r_{Uij}) = \frac{1}{2\pi\sqrt{1 - r_{Uij}^2}} \exp\left(-\frac{u_i^2 - 2r_{Uij}u_i u_j + u_j^2}{2(1 - r_{Uij}^2)} \right) \tag{11}$$

$$r_{Zij} \approx \sum_{\ell=1}^{m} \sum_{k=1}^{m} p_\ell p_k \left(\frac{z_{i,\ell} - \mu_{Zi}}{\sigma_{Zi}} \right) \left(\frac{z_{j,k} - \mu_{Zj}}{\sigma_{Zj}} \right) \tag{12}$$

C. PARTICULAR CASE: NATAF TRANSFORM IN A 2D SPACE

Setting the representation space dimension as $N = 2$ allows to easily plot these spaces, therefore defining a built-in visualization tool. Moreover, the 2D NT requires the estimation of only one correlation coefficient r_{U12}. In this paper, the NT is defined with respect to the healthy cluster projected into a 2D LVM space. Let Z_1^{HY} and Z_2^{HY} be the two coordinates associated with the latent space distribution of the healthy cluster. From Z_1^{HY} and Z_2^{HY}, one can estimate the empirical cdfs $F_1(z_1)$ and $F_2(z_2)$, and the Pearson correlation coefficient r_{Z12}^{HY}. Then, r_{U12}^{HY} is estimated using Algorithm 1. For notation simplicity, let $\rho = r_{U12}^{\text{HY}}$. Then, provided that $\rho \neq \pm1$, the correlation matrix $[1, \rho; \rho, 1]$ is invertible. Then, the NT is completely defined by (13) and (14), where Φ is the cdf of the normal distribution $\mathcal{N}(0, 1)$.

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} \Phi^{-1}[F_1(z_1)] \\ \Phi^{-1}[F_2(z_2)] \end{pmatrix} \tag{13}$$

$$\begin{pmatrix} s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ \rho u_1 + \sqrt{1 - \rho^2}u_2 \end{pmatrix} \tag{14}$$

ACKNOWLEDGMENT

REFERENCES

[1] J. E. Breneman, C. Sahay, and E. E. Lewis, *Introduction to Reliability Engineering*. Hoboken, NJ, USA: Wiley, 2022.

[2] V. Smil, "The two prime movers of globalization: History and impact of diesel engines and gas turbines," *J. Global Hist.*, vol. 2, no. 3, pp. 373–394, Nov. 2007.

[3] V. Smil, *Energy and Civilization: A History*. Cambridge, MA, USA: MIT Press, 2018.

[4] G. J. Herbert, S. Iniyan, E. Sreevalsan, and S. Rajapandian, "A review of wind energy technologies," *Renew. Sustain. Energy Rev.*, vol. 11, no. 6, pp. 1117–1145, 2007.

[5] M. Azimi, A. Eslamlou, and G. Pekcan, "Data-driven structural health monitoring and damage detection through deep learning: State-of-the-art review," *Sensors*, vol. 20, no. 10, p. 2778, May 2020.

[6] G. Niu, "Data-driven technology for engineering systems health management," *Springer Singap.*, vol. 10, pp. 978–981, Jan. 2017.

[7] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Int. Conf. Learn. Represent.*, 2014, pp. 1–14.

[8] D. P. Kingma, S. Mohamed, D. J. Rezende, and M. Welling, "Semi-supervised learning with deep generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[9] K. Sohn, H. Lee, and X. Yan, "Learning structured output representation using deep conditional generative models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 1–9.

[10] J. Zhu, G. Peng, and D. Wang, "Dual-domain-based adversarial defense with conditional VAE and Bayesian network," *IEEE Trans. Ind. Informat.*, vol. 17, no. 1, pp. 596–605, Jan. 2021.

[11] A. Proteau, R. Zemouri, A. Tahan, and M. Thomas, "Dimension reduction and 2D-visualization for early change of state detection in a machining process with a variational autoencoder approach," *Int. J. Adv. Manuf. Technol.*, vol. 111, nos. 11–12, pp. 3597–3611, Dec. 2020.

[12] R. Zemouri, M. Lévesque, É. Boucher, M. Kirouac, F. Lafleur, S. Bernier, and A. Merkhouf, "Recent research and applications in variational autoencoders for industrial prognosis and health management: A survey," in *Proc. Prognostics Health Manage. Conf. (PHM- London)*, May 2022, pp. 193–203.

[13] C. M. A. Roelofs, M.-A. Lutz, S. Faulstich, and S. Vogt, "Autoencoder-based anomaly root cause analysis for wind turbines," *Energy AI*, vol. 4, Jun. 2021, Art. no. 100065.

[14] L. Zhang, H. Zhang, and G. Cai, "The multiclass fault diagnosis of wind turbine bearing based on multisource signal fusion and deep learning generative model," *IEEE Trans. Instrum. Meas.*, vol. 71, pp. 1–12, 2022.

[15] A. Oliveira-Filho, R. Zemouri, P. Cambron, and A. Tahan, "Early detection and diagnosis of wind turbine abnormal conditions using an interpretable supervised variational autoencoder model," *Energies*, vol. 16, no. 12, p. 4544, Jun. 2023.

[16] M. Tschannen, O. Bachem, and M. Lucic, "Recent advances in autoencoder-based representation learning," 2018, *arXiv:1812.05069*.

[17] N. Costa and L. Sánchez, "Variational encoding approach for interpretable assessment of remaining useful life estimation," *Rel. Eng. Syst. Saf.*, vol. 222, Jun. 2022, Art. no. 108353.

[18] A. Nataf, "Détermination des distributions de probabilités dont les marges sont données," *Comput. Rendus de l'Académie des Sci.*, vol. 225, no. 3, pp. 42–43, 1962.

[19] Y. Zhao, Z. Liao, and S. Dong, "Estimation of characteristic extreme response for mooring system in a complex ocean environment," *Ocean Eng.*, vol. 225, Apr. 2021, Art. no. 108809. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0029801821002444

[20] Z.-Q. Chen, S.-X. Zheng, J. Zhang, and H. Jia, "Seismic reliability analysis of high-pier railway bridges with correlated random parameters via an improved maximum entropy method," *Structures*, vol. 33, pp. 4538–4555, Oct. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S235201242100655X

[21] V. Valamanesh, A. T. Myers, and S. R. Arwade, "Multivariate analysis of extreme metocean conditions for offshore wind turbines," *Struct. Saf.*, vol. 55, pp. 60–69, Jul. 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167473015000296

[22] D. Kingma, "Variational inference & deep learning: A new synthesis," Ph.D. dissertation, Fac. Sci. (FNWI), Inform. Inst. (IVI), Univ. Amsterdam, Amsterdam, The Netherlands, Oct. 2017. [Online]. Available: https://hdl.handle.net/11245.1/8e55e07f-e4be-458f-a929-2f9bc2d169e8

[23] X. Kong, X. Jiang, B. Zhang, J. Yuan, and Z. Ge, "Latent variable models in the era of industrial big data: Extension and beyond," *Annu. Rev. Control*, vol. 54, pp. 167–199, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1367578822000992

[24] V. Pandiyan, J. Prost, G. Vorlaufer, M. Varga, and K. Wasmer, "Identification of abnormal tribological regimes using a microphone and semi-supervised machine-learning algorithm," *Friction*, vol. 10, no. 4, pp. 583–596, Apr. 2022.

[25] L. Yang and Z. Zhang, "Wind turbine gearbox failure detection based on SCADA data: A deep learning-based approach," *IEEE Trans. Instrum. Meas.*, vol. 70, pp. 1–11, 2021.

[26] T. Chen, X. Liu, B. Xia, W. Wang, and Y. Lai, "Unsupervised anomaly detection of industrial robots using sliding-window convolutional variational autoencoder," *IEEE Access*, vol. 8, pp. 47072–47081, 2020.

[27] D. Hong, S. Bang, and B. Kim, "Unsupervised condition diagnosis of linear motion guide using generative model based on images," *IEEE Access*, vol. 9, pp. 80491–80499, 2021.

[28] Á. D. Hallgrímsson, H. H. Niemann, and M. Lind, "Improved process diagnosis using fault contribution plots from sparse autoencoders," *IFAC-PapersOnLine*, vol. 53, no. 2, pp. 730–737, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2405896320311472

[29] Á. D. Hallgrímsson, H. H. Niemann, and M. Lind, "Unsupervised isolation of abnormal process variables using sparse autoencoders," *J. Process Control*, vol. 99, pp. 107–119, Mar. 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0959152421000056

[30] F. Cheng, W. Cai, H. Liao, B. Wu, and S. Dubey, "Fault detection and isolation for chiller system based on deep autoencoder," in *Proc. IEEE 16th Conf. Ind. Electron. Appl. (ICIEA)*, Aug. 2021, pp. 1702–1706.

[31] X. Liang, F. Duan, I. Bennett, and D. Mba, "A sparse autoencoder-based unsupervised scheme for pump fault detection and isolation," *Appl. Sci.*, vol. 10, no. 19, p. 6789, Sep. 2020. [Online]. Available: https://www.mdpi.com/2076-3417/10/19/6789

[32] E. M. Coraça, J. V. Ferreira, and E. G. O. Nóbrega, "An unsupervised structural health monitoring framework based on variational autoencoders and hidden Markov models," *Rel. Eng. Syst. Saf.*, vol. 231, Mar. 2023, Art. no. 109025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832022006408

[33] W. Booyse, D. N. Wilke, and S. Heyns, "Deep digital twins for detection, diagnostics and prognostics," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106612. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327019308337

[34] R. Zemouri, R. Ibrahim, and A. Tahan, "Hydrogenerator early fault detection: Sparse dictionary learning jointly with the variational autoencoder," *Eng. Appl. Artif. Intell.*, vol. 120, Apr. 2023, Art. no. 105859. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S095219762300043X

[35] X. Liu, W. Teng, S. Wu, X. Wu, Y. Liu, and Z. Ma, "Sparse dictionary learning based adversarial variational auto-encoders for fault identification of wind turbines," *Measurement*, vol. 183, Oct. 2021, Art. no. 109810. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0263224121007612

[36] K. Wang, M. G. Forbes, B. Gopaluni, J. Chen, and Z. Song, "Systematic development of a new variational autoencoder model based on uncertain data for monitoring nonlinear processes," *IEEE Access*, vol. 7, pp. 22554–22565, 2019.

[37] F. Cheng, Q. P. He, and J. Zhao, "A novel process monitoring approach based on variational recurrent autoencoder," *Comput. Chem. Eng.*, vol. 129, Oct. 2019, Art. no. 106515. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0098135418309219

[38] S. Lee, M. Kwak, K.-L. Tsui, and S. B. Kim, "Process monitoring using variational autoencoder for high-dimensional nonlinear processes," *Eng. Appl. Artif. Intell.*, vol. 83, pp. 13–27, Aug. 2019. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0952197619300983

[39] M. Zheng, J. Man, D. Wang, Y. Chen, Q. Li, and Y. Liu, "Semi-supervised multivariate time series anomaly detection for wind turbines using generator SCADA data," *Rel. Eng. Syst. Saf.*, vol. 235, Jul. 2023, Art. no. 109235. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0951832023001503

[40] P. Tang, K. Peng, J. Dong, K. Zhang, and S. Zhao, "Monitoring of nonlinear processes with multiple operating modes through a novel Gaussian mixture variational autoencoder model," *IEEE Access*, vol. 8, pp. 114487–114500, 2020.

[41] R. Balshaw, P. S. Heyns, D. N. Wilke, and S. Schmidt, "Latent indicators for temporal-preserving latent variable models in vibration-based condition monitoring under non-stationary conditions," *Mech. Syst. Signal Process.*, vol. 199, Sep. 2023, Art. no. 110446. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327023003540

[42] R. Balshaw, P. S. Heyns, D. N. Wilke, and S. Schmidt, "Importance of temporal preserving latent analysis for latent variable models in fault diagnostics of rotating machinery," *Mech. Syst. Signal Process.*, vol. 168, Apr. 2022, Art. no. 108663. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0888327021009870

[43] R. Ibrahim, R. Zemouri, B. Kedjar, A. Merkhouf, A. Tahan, K. Al-Haddad, and F. Lafleur, "Non-invasive detection of rotor inter-turn short circuit of a hydrogenerator using AI-based variational autoencoder," *IEEE Trans. Ind. Appl.*, pp. 1–10, 2023.

[44] H. Li, Z. Lü, and X. Yuan, "Nataf transformation based point estimate method," *Sci. Bull.*, vol. 53, no. 17, pp. 2586–2592, Sep. 2008.

[45] P. R. Krishnaiah, P. Hagis Jr., and L. Steinberg, "A note on the bivariate chi distribution," *SIAM Rev.*, vol. 5, no. 2, pp. 140–144, Apr. 1963.

[46] A. Der Kiureghian and P. Liu, "Structural reliability under incomplete probability information," *J. Eng. Mech.*, vol. 112, no. 1, pp. 85–104, Jan. 1986.

[47] R. Lebrun and A. Dutfoy, "An innovating analysis of the Nataf transformation from the copula viewpoint," *Probabilistic Eng. Mech.*, vol. 24, no. 3, pp. 312–320, Jul. 2009.

[48] R.-C. Cheng and K.-S. Chen, "Ball bearing multiple failure diagnosis using feature-selected autoencoder model," *Int. J. Adv. Manuf. Technol.*, vol. 120, nos. 7–8, pp. 4803–4819, Jun. 2022.

[49] M. Hemmer, A. Klausen, H. V. Khang, K. G. Robbersmyr, and T. I. Waag, "Health indicator for low-speed axial bearings using variational autoencoders," *IEEE Access*, vol. 8, pp. 35842–35852, 2020.

[50] S. Ramakrishna, Z. Rahiminasab, G. Karsai, A. Easwaran, and A. Dubey, "Efficient out-of-distribution detection using latent space of β-VAE for cyber-physical systems," *ACM Trans. Cyber-Physical Syst.*, vol. 6, no. 2, pp. 1–34, Apr. 2022.

[51] V. K. Sundar, S. Ramakrishna, Z. Rahiminasab, A. Easwaran, and A. Dubey, "Out-of-distribution detection in multi-label datasets using latent space of β-VAE," in *Proc. IEEE Secur. Privacy Workshops (SPW)*, May 2020, pp. 250–255.

[52] M. De Angelis and A. Gray, "Why the 1-Wasserstein distance is the area between the two marginal CDFs," 2021, *arXiv:2111.03570*.

[53] A. Ramdas, N. Trillos, and M. Cuturi, "On Wasserstein two-sample testing and related families of nonparametric tests," *Entropy*, vol. 19, no. 2, p. 47, Jan. 2017.

[54] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and F-score, with implication for evaluation," in *Proc. Eur. Conf. Inf. Retr.* Berlin, Germany: Springer, 2005, pp. 345–359.

[55] P. Cambron, A. Tahan, C. Masson, and F. Pelletier, "Bearing temperature monitoring of a wind turbine using physics-based model," *J. Quality Maintenance Eng.*, vol. 23, no. 4, pp. 479–488, Oct. 2017.

[56] A. Saxena, K. Goebel, D. Simon, and N. Eklund, "Damage propagation modeling for aircraft engine run-to-failure simulation," in *Proc. Int. Conf. Prognostics Health Manag.*, Oct. 2008, pp. 1–9.

[57] E. Ramasso and A. Saxena, "Performance benchmarking and analysis of prognostic methods for CMAPSS datasets," *Int. J. Prognostics Health Manag.*, vol. 5, no. 2, pp. 1–15, Nov. 2020.

[58] M. Arias Chao, C. Kulkarni, K. Goebel, and O. Fink, "Aircraft engine run-to-failure dataset under real flight conditions for prognostics and diagnostics," *Data*, vol. 6, no. 1, p. 5, Jan. 2021.

[59] D. K. Frederick, J. A. DeCastro, and J. S. Litt, "User's guide for the commercial modular aero-propulsion system simulation (CMAPSS)," NASA, Glenn Res. Center, Cleveland, OH, USA, Tech. Rep., 2007.

[60] A. González-Muñiz, I. Díaz, A. A. Cuadrado, and D. García-Pérez, "Health indicator for machine condition monitoring built in the latent space of a deep autoencoder," *Rel. Eng. Syst. Saf.*, vol. 224, Aug. 2022, Art. no. 108482.

[61] S. Cofre-Martel, E. L. Droguett, and M. Modarres, "Remaining useful life estimation through deep learning partial differential equation models: A framework for degradation dynamics interpretation using latent variables," *Shock Vibrat.*, vol. 2021, pp. 1–15, May 2021.

[62] H. Zhao, H. Liu, W. Hu, and X. Yan, "Anomaly detection and fault analysis of wind turbine components based on deep learning network," *Renew. Energy*, vol. 127, pp. 825–834, Nov. 2018.

[63] M. Rezamand, M. Kordestani, R. Carriveau, D. S.-K. Ting, M. E. Orchard, and M. Saif, "Critical wind turbine components prognostics: A comprehensive review," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 12, pp. 9306–9328, Dec. 2020.

[64] *Wind Energy Generation Systems—Part 12–1: Power Performance Measurement of Electricity Producing Wind Turbines*, IEC, International Electrotechnical Commission, Geneva, Switzerland, Standard, Sep. 2022.

[65] M. Beretta, A. Julian, J. Sepulveda, J. Cusidó, and O. Porro, "An ensemble learning solution for predictive maintenance of wind turbines main bearing," *Sensors*, vol. 21, no. 4, p. 1512, Feb. 2021.

[66] H. Badihi, Y. Zhang, B. Jiang, P. Pillay, and S. Rakheja, "A comprehensive review on signal-based and model-based condition monitoring of wind turbines: Fault diagnosis and lifetime prognosis," *Proc. IEEE*, vol. 110, no. 6, pp. 754–806, Jun. 2022.

[67] L. Gao, T. Tao, Y. Liu, and H. Hu, "A field study of ice accretion and its effects on the power production of utility-scale wind turbines," *Renew. Energy*, vol. 167, pp. 917–928, Apr. 2021.

[68] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, and M. Isard, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement.*, 2016, pp. 265–283.

[69] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[70] A. M. De Oliveira-Filho, P. Cambron, and A. Tahan, "Condition monitoring of wind turbine main bearing using SCADA data and informed by the principle of energy conservation," in *Proc. Prognostics Health Manage. Conf. (PHM- London)*, May 2022, pp. 276–282.

[71] N. Renström, P. Bangalore, and E. Highcock, "System-wide anomaly detection in wind turbines using deep autoencoders," *Renew. Energy*, vol. 157, pp. 647–659, Sep. 2020.

ADAITON OLIVEIRA-FILHO (Member, IEEE) received the M.Sc. degree (Hons.) in mechanics–modelling and simulation in mechanics of structures and coupled systems from Université Paris-Saclay, Gif-sur-Yvette, Île-de-France, France, in 2019, and the dual B.Sc. degree in engineering from the University of São Paulo, São Carlos, São Paulo, Brazil, and École Centrale Paris, Châtenay-Malabry, Île-de-France, in 2021. He is currently pursuing the Ph.D. degree with École de Technologie Supérieure, Montreal, QC, Canada. His research focuses on data-driven approaches for condition monitoring, diagnosis, and prognosis of complex multiphysics systems, particularly wind turbines.

RYAD ZEMOURI was born in Algiers, Algeria, in 1974. He received the Engineering degree in electronics and control science from Université Mouloud Mammeri de Tizi Ouzou, Tizi Ouzou, Algeria, in 1998, and the Post-Graduate Diploma and Ph.D. degrees from the University of Franche-Comté, Besançon, France, in 2000 and 2003, respectively.

In 2003, he joined the Conservatoire National des Arts et Metiers, Paris, France, as an Associate Professor. In 2022, he joined the Research Center of Hydro-Québec, Varennes, QC, Canada, as a Data Scientist Researcher. His research interests include machine learning for safety-critical prognosis, diagnosis, and health management.

FRANCIS PELLETIER received the M.Eng. and Ph.D. degrees in mechanical engineering from École de Technologie Supérieure, in 2002 and 2014, respectively.

He is a senior renewable energies consultant with 25 years of industry experience. Over this time frame, he has worked in wind project management, noise impact analysis, wind turbine performance testing, wind farm layout design, and energy yield estimation.

ANTOINE TAHAN was born in Syria, in 1966. He received the B.Sc.A., M.Sc.A., and Ph.D. degrees from the University of Laval, QC, Canada, in 1991, 1992, and 1998, respectively.

Since 2004, he has been a Full Professor with the Department of Mechanical Engineering, École de Technologie Supérieure, and an Active Member of the Products, Processes and Systems Engineering Laboratory (LIPPS). He is specialized in dimensional metrology and geometrical characterization, tolerance management, and propagation of uncertainties. His research focuses on the aerospace and energy industries. He is working on the propagation of uncertainties in engineering models and diagnostic and prognostic tools for complex electromechanical systems. He contributes to the development of probabilistic approaches to detect the state of degradation and damage to improve reliability and design.

• • •