

Data-driven energy prediction modeling for both energy efficiency and maintenance in smart manufacturing systems

Miguel Angel Bermeo-Ayerbe^{a,b,*}, Carlos Ocampo-Martinez^a, Javier Diaz-Rozo^b

^a*Automatic Control Department, Universitat Politècnica de Catalunya, Institut de Robòtica i Informàtica Industrial (CSIC-UPC), Barcelona, Spain*

^b*Aingura IIoT, Donostia-San Sebastian, Spain*

Abstract

The optimization and monitoring of the energy consumption of machinery lead to a sustainable and efficient industry. For this reason and following a digital twin strategy, an online data-driven energy modeling approach with adaptive capabilities has been proposed and described throughout this paper. This approach is useful in developing robust energy management systems that enhance the energy efficiency of industrial machinery. In this way, the dynamic behavior of their energy consumption is modeled without using phenomenological laws. In contrast, traditional methodologies hardly consider such dynamic behavior or use an exhaustive modeling process. The proposed approach includes an adaptive mechanism to consider the natural degradation of machinery. This mechanism is based on a concept drift detector, which detects when the current consumption of the machine is not correctly represented by the model estimation and adapts the model to account for these new behaviors. The concept drift detector has broad applicability in the face of reducing maintenance costs, measuring the impact and evolution of either abnormal behaviors (e.g., failures) or degradation, and identify which elements change. The proposed methodology has been validated in an industrial testbed. An experiment with three emulated

*Corresponding author

Email addresses: miguel.angel.bermeo@upc.edu (Miguel Angel Bermeo-Ayerbe), carlos.ocampo@upc.edu (Carlos Ocampo-Martinez), jdiaz@ainguraiiot.com (Javier Diaz-Rozo)

concept drifts was carried out in the testbed. As a result, the proposed adaptive approach obtained more than doubled the fit rate of the energy prediction/estimation compared to the non-adaptive model and successfully detected these changes in energy consumption.

Keywords: Non-intrusive load monitoring, Data-driven model, Subspace identification, Energy models, Concept drift, Digital twin, Gaussian mixture models, Energy efficiency, Machine fault diagnosis

1. Introduction

Towards the optimal use of energy resources and a sustainable environment, the deployment of methodologies related to energy efficiency in the industry has been promoted in recent decades [1]. In addition, proper maintenance of industrial equipment reduces energy waste and production costs in the industrial sector [2, 3]. Load monitoring leads to an understanding of the energy consumption of specific appliances/equipment, which allows implementing energy efficiency methodologies as well as early detection of anomalies and measurement of equipment degeneration [4]. In this way, predictive maintenance systems based on energy consumption can be implemented, monitoring health status of equipment to anticipate and prevent possible failures. Within load monitoring strategies, the non-intrusive load monitoring (NILM) approach has been relevant due to its cost-effectiveness and ease of installation and replacement. NILM measures the power service entrance of an electrical circuit via a smart meter to disaggregate equipment-level data [5]. NILM has extensive development in smart homes, but nowadays, it has aroused industrial interest, monitoring the energy consumption of a production system (e.g., manufacturing machine) with a minimum amount of sensors [6].

As most machining systems have significant energy saving potential, being one of the main consumers and waste producers of factories, motivating to monitor and model their energy consumption profile to optimize their yield [7, 8]. Thereby, production efficiency is improved while energy costs are reduced. Furthermore, to cope with the high costs associated with stopping a production line due to machine failure, the machine health management for diagnosis and prognosis based on power consumption behavior has been crucial for the realization of smart manufacturers [9]. In this way, a digital twin approach that can be computed and used as a reference to check whether the real machinery functioning works as expected or experiences

any failure [10].

The traditional modeling process to characterize the energy consumption of an electromechanical system requires solid knowledge about phenomenological laws (e.g., chemistry, physics), mathematics, modeling, and particular properties of the system, being an exhaustive and exclusive procedure for a particular system [11, 12]. Therefore, with an Industry 4.0 (I4.0) perspective, this article is focused on developing a data-driven modeling approach geared towards electromechanical industrial machinery systems. The proposed approach certainly reduces modeling efforts and takes advantage of a large amount of available system data in I4.0 infrastructures.

Energy modeling of industrial machines focuses mainly on the machining process, looking for energy consumption relationships and cutting/milling/machining methods, process parameters, and tool geometries [1], e.g., Liang et al. [13] present an Artificial Neural Networks (ANN)-based energy modeling for energy-efficient machining optimization, considering dynamic and aging conditions of machines during manufacturing life cycles. Zou et al. [14] propose a data-driven stochastic manufacturing modeling method to identify and predict energy-saving opportunities and their impact on production, which are used for a distributed predictive control strategy, improving overall profit and energy efficiency on production. Xie et al. [15] develop an integrated model for predicting the specific energy consumption (SEC) of manufacturing processes, considering the dynamic characteristics of material-cutting power and the influence of the machine tools, work-piece material, and cutting parameters. Most data-driven energy models are based on well-known structures of machine learning strategies such as neuronal models, linear support vector-based on regression, fuzzy logic functions, and Gaussian mixture regression (GMM) [16, 17, 18, 19]. However, the dynamic behavior of energy consumption is hardly characterized by a short prediction horizon and few details of transient states.

The contribution of this paper is an online adaptive energy modeling approach in a NILM scheme for manufacturing machines, addressing the variability of real manufacturing systems and their impact on operational performance. This methodology is based on a system identification (SI) algorithm to build mathematical models based on data [20]. These models predict the dynamic behavior of energy consumption of industrial machinery, with high precision and are also fully applicable in the control theory. Moreover, since systems undergo a natural degradation over time that causes a loss of accuracy in the model prediction, an adaptive mechanism is proposed

to measure how that prediction deviates from the current consumption, using a concept drift strategy [21]. When this deviation exceeds a threshold, the model is adapted to include the new dynamic behavior. The proposed approach has applicability for both optimizing operations and predictive maintenance, whose adaptive feature provides robustness and reliability to end-users.

The proposed modeling methodology computes a state-space representation (SSR) to predict the instantaneous power consumption. The prediction is performed according to inputs signals, indicating the configuration (e.g., speed) and state (on/off) of each electromechanical equipment of industrial machinery. Therefore, the dynamic behavior of transient states when an equipment undergoes a change, either in its configuration or state, can be properly modeled, including the main temporal response features (e.g., overshoot, undershoot, settling time, rise time, peak time). Furthermore, the stationary states of each equipment can be represented by modeling the level of energy consumption for each configuration. These criteria are crucial to capture changes to face the detection of failures or new unexpected behaviors, compromising the efficiency of the process. However, SSR is a linear model around an operating point of a real system, which limits its use in broad operations to deal with non-linear behaviors. Whereas ANN approaches can model non-linear behaviors and can learn new behaviors. Those capabilities are necessary for the industry since the real industrial systems can show non-linear behaviors and undergo changes due to constant operations. Nevertheless, the dynamic behaviors of industrial systems have a stochastic nature, where ANN approaches do not excel by definition.

In this paper, subspace identification (Sub-ID) procedure with a Hammerstein modeling is proposed to characterize the non-linearities between inputs and outputs (power consumption). Additionally, with the adaptive mechanism, the model will change to consider new behaviors of the system. In this way, the proposed modeling approach provides a model with high prediction precision and a known model structure in dynamic and control systems, whose interpretation of parameters has been widely studied. On the other hand, the interpretation of the ANN weights is challenging, as well as to determine the relationships of those weights with the dynamic parameters (e.g., time constant, natural frequency, and damping ratio), similarly for other machine learning approaches. Thus, for cases where a dynamic behavior needs to be modeled in great detail, the proposed SI approach is an option and can provide features to facilitate subsequent dynamic analyses.

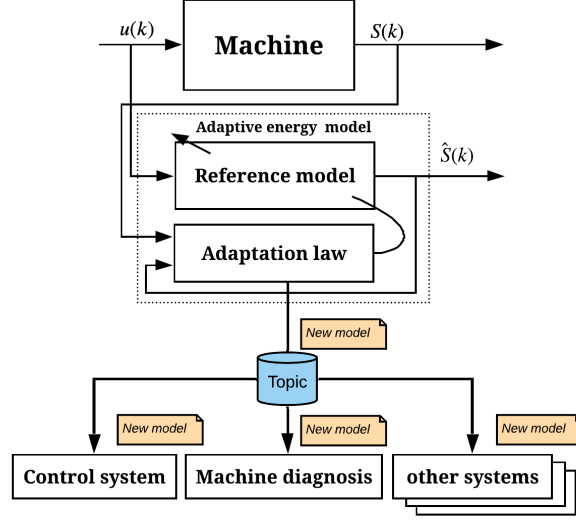


Figure 1: Adaptive energy model scheme and publish-subscribe pattern.

The structure of this paper is organized as follows: In Section 2, the proposed methodology is detailed, describing the data-driven modeling procedure and adaptive mechanism. Section 3.1 presents the validation process that was carried out in an industrial machine, whose results are shown in Section 3.2 with their corresponding analysis. Finally, in Section 4, the conclusions and further work are drawn.

2. Energy models

Manufacturing industries constantly face challenges to improve their performance and reduce costs, optimizing not only reliability and productivity but also considering different sustainability issues. Therefore, following a NILM scheme, this paper proposes a methodology to compute dynamic energy models to predict machine energy consumption with high accuracy and an adaptive mechanism to update the model when prediction accuracy deteriorates, i.e., a mechanism to assess when the model has lost accuracy, and then compute a new one, adapting to new behaviors over time. This degradation in precision occurs because the system experiences either a progressive or abrupt change in behavior during the operating time. For this reason, the model cannot correctly represent the current behavior of the system, which may be due to deterioration of the equipment from constant use, machine modification (e.g., installation of new components, change of tools), or machine failure.

The proposed data-driven energy models can be applied in both the

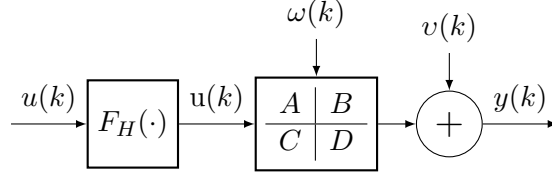


Figure 2: Block model of a Hammerstein model [23].

machine control system (e.g., to enhance energy efficiency) and machine diagnostics (e.g., predictive maintenance, fault detection). For this reason, the scheme supports the publish-subscribe pattern as shown in Figure 1, providing versatility and extensibility since it is a multi-purpose oriented approach. The adaptive energy model of the machine depends on machine control signals $u(k)$ and energy measurements $S(k)$. Control signals indicate when equipment is activated and under which conditions. Moreover, such control signals are extracted from the programmable logic controller (PLC) or computer numerical control (CNC), avoiding installing more sensors, just an energy sensor following a NILM scheme. The current power consumption prediction/estimation $\hat{S}(k)$ of the machine is used as a reference to measure the degradation of prediction accuracy (concept drift). Therefore, when the model has been updated, it would be sent to a dispatcher (topic) to distribute it to the subscribed external systems.

In this way, with a publish-subscribe pattern, the proposed approach can be scalable to the production line to monitor or improve energy efficiency, using the different energy models for each machine in production planning. Furthermore, this methodology can be easily integrated into infrastructures as the energy cloud proposed by Sequeira et al. [22]. Thus, this novel proposed approach is divided into two parts: a data-driven energy modeling methodology for electromechanical systems and an adaptive mechanism, which are outlined in Subsections 2.1 and 2.2, respectively.

2.1. Subspace identification structure for nonlinear systems

Previous works have already proposed a methodology for identifying energy models via subspace identification (Sub-ID) algorithms [20]. This methodology represents the machine as a multiple-input and multiple-output (MIMO) system; the inputs $U \in \mathcal{U}$ are activation/deactivation signals for each device, and the outputs $S \in \mathbb{R}^l$ are the total instantaneous power consumption of the machine for each line. Note that the number of outputs l may vary depending on the energy sensor or the electrical phases to consider. Often a three-phase line is used. Furthermore, every input has a domain \mathbb{U}_i

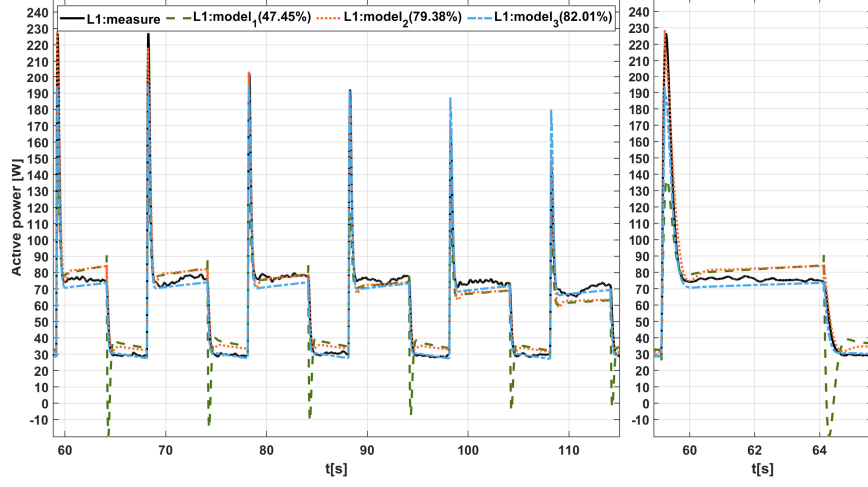


Figure 3: Hammerstein-based electric engine modeling.

for $i \in \{0, 1, \dots, m\}$, where m is the number of inputs, then $\mathcal{U} = \mathbb{U}_1 \times \dots \times \mathbb{U}_m$. The input domain \mathbb{U}_i can be binary (either on or off), or the activation can have a range of values related to power consumption, e.g., the power consumption of an engine can vary depending on speed or torque, thereby, they are candidates to be used as inputs and establish an operating range. Thus, the general form of the input domain can be expressed as

$$\mathbb{U}_i \triangleq \{u_i \in \mathbb{Z}_{\geq 0} \mid \underline{u}_i \leq u_i \leq \bar{u}_i \vee u_i = 0\},$$

being $\underline{u}_i, \bar{u}_i \in \mathbb{Z}_{\geq 1}$ the lower and upper bounds to activate i -th device, respectively.

Sub-ID is a state-space identification procedure based on the Hankel matrix, which is created with an input-output identification dataset. This matrix performs as a regressive matrix that solves a least-squares problem, given a linear combination of subspace matrices that describe the data. The state-space matrices are computed through the decomposition of the subspace matrices, using reliable, widely known and available numerical algorithms [20]. The result is a discrete-time linear time-invariant (LTI) state-space representation,

$$x(k+1) = Ax(k) + Bu(k) + \omega(k), \quad (1a)$$

$$y(k) = Cx(k) + Du(k) + v(k), \quad (1b)$$

with

$$\mathbf{E} \begin{bmatrix} \omega^T \\ v^T \end{bmatrix} \begin{bmatrix} \omega^T & v^T \end{bmatrix} = \begin{pmatrix} Q & S \\ S^T & R \end{pmatrix} \delta_{pq} \geq 0, \quad (1c)$$

being $\mathbf{E}[\bullet]$ the expected value, $u \in \mathcal{U}$, $y \in \mathbb{R}^l$ and $x \in \mathbb{R}^n$ vectors at discrete-time instant k of the m inputs, l outputs and n states (i.e., the model order) of the system, respectively. Moreover, $A \in \mathbb{R}^{n \times n}$ is the (dynamical) system matrix, $B \in \mathbb{R}^{n \times m}$ is the input matrix that represents a linear transformation of the current input in the contribution to the next state, $C \in \mathbb{R}^{l \times n}$ is the output matrix that describes the effect of current states over outputs, $D \in \mathbb{R}^{l \times m}$ is the feedthrough (or feedforward) matrix that allows modeling the direct effect of the input on the measurements/outputs, $\omega \in \mathbb{R}^n$ and $v \in \mathbb{R}^l$ are non-measurable vector signals that affect the states and measures, respectively. Additionally, $Q \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{n \times l}$ and $R \in \mathbb{R}^{l \times l}$ are the covariance matrices of the noise signals $\omega(k)$ and $v(k)$ [24].

The linear model (1) is capable of predicting the future behavior of the system (with high accuracy) around an operating point, with direct proportionality between input and output. It is suitable for applications that require the system to operate in a narrow range. Although for other cases, nonlinear models are an option, allowing to expand the operating range and describe complex behaviors. The *block-oriented* model is one of the most studied classes of nonlinear models, which is based on interconnecting static nonlinearities (memoryless) to an LTI model. Within this class, the *Hammerstein* model is studied, which consists of removing the nonlinearities from inputs to outputs of the system with a nonlinear function (see Figure 2), thus, new inputs are established, having a linear relationship with the outputs and is used to identify an LTI model [23].

Most manufacturing machinery relies on electric motors, which may have a variable-frequency drive (VFD) connected to control their speed. With that in mind, a systems identification methodology is proposed to model the dynamics of electric peaks based on Sub-ID, e.g., the peak generated by the electric motor when it is turned on. Consider a three-phase motor with a constant load that is modeled as a black box, with speed as input and power consumption of each line as output. Using a dataset with impulse responses at different speeds, the model can be identified, but the obtained model has some problems, e.g., the transient peak is shorter than the real one and when it turns off there is a negative peak. The previous discussion can be further performed by using Figure 3 (for the first phase/line) with model 1 (whose order n is equal to three) that follows the structure (1) and has a fit rate of 47.54%. The fit rate was computed as the complement of the normalized root mean squared errors, also known as the *Coefficient of*

Determination (R^2) [25], i.e.,

$$R^2(S, \hat{S}) = 100 \max \left(1 - \frac{\|S - \hat{S}\|_2}{\|S - \mu_S\|_2}, 0 \right) \%, \quad (2)$$

where S is the energy consumption measurement, μ_S the mean of S and \hat{S} the estimated consumption. Note that the estimation \hat{S} is computed using just the input signal, i.e., S measurements are not fed back to compute \hat{S} .

For improving the previous result, it is recommended to create an artificial input with a delay from the original input and when the real input is off (with null value) the artificial one should be 0 as well, i.e.,

$$u_a(k) = \begin{cases} 0 & \text{if } u(k) = 0 \\ u(k - \delta) & \text{otherwise} \end{cases}, \quad (3)$$

being δ the delay. Consequently, a new input vector is defined as $u(k) = [u(k) \ u_a(k)]^T$ and used to generate a new LTI model.

Following the strategy described above, model 2 shown in Figure 3 is obtained, giving a fit rate of around 79%. This result demonstrates an improvement compared to model 1, having proper peak identification and eliminating the negative peak. However, models 1 and 2 have a notable steady-state error, which is due to the polynomial increase in consumption while the engine speed increases linearly. For this reason, the LTI model is not able to describe this behavior. This problem is solved by mapping the input from linear increasing to polynomial increasing, i.e.,

$$u_p(k) = w_1 u(k) + w_2 u^2(k), \quad (4)$$

where w_1 and w_2 are weights that can be determined with a polynomial regression between the speed and the mean of the consumption in the stationary part. In this way, a new LTI model is computed using a new input vector $u_p(k) = [u_p(k) \ u_{p,a}(k)]^T$, being $u_{p,a}$ a delayed input from u_p .

Applying the previous procedure, model 3 is obtained with a fit rate of 82% (see Figure 3). Although model 3 fit rate is 3% over model 2 and may seem unrepresentative, this strategy can be applied for cases with a high stationary error, resulting in a significant improvement. Furthermore, this methodology can be extrapolated to different devices with a similar energy consumption profile. Thereby, a proper characterization of the nonlinearities of the system leads to a reduction in model order or an increase in accuracy, being essential for the identification and elimination of delays between input and output.

2.2. Adaptation mechanism

Machine power consumption profiles can change over time due to many reasons, e.g., degradation resulting from constant use in an industrial environment or the installation of new equipment. Figure 1 depicts an online adaptation scheme to fit the model estimation with current machine energy consumption behaviors. The reference model is initialized using a model identified offline and predicts the machine consumption, which is compared with real energy consumption in the adaptation-law block. This block tracks the error between current and estimated consumption to measure accuracy degradation, and when the model cannot accurately estimate consumption, a new model is generated that replaces the current reference model.

Based on the methodology presented by Diaz-Rozo et al. [21], the adaptation law has been addressed by detecting a concept drift, which is defined as a change over time in unforeseen ways of the statistical properties of the target variables [26]. Diaz-Rozo et al. [21] proposed a concept drift detector to track the outliers in an adaptive sliding window, discriminating data drifting from a scattered outlier.

Consider an error signal between measures and estimations

$$e(k) = S(k) - \hat{S}(k),$$

whose probability density function can be described by a Gaussian Mixture Model (GMM) with k_c clusters

$$p(\mathbf{x} = e(k); \Psi) = \sum_{i=1}^{k_c} \phi_i \mathcal{N}(\mathbf{x}; \theta_i), \quad (5)$$

being $\Psi = (\phi_1, \dots, \phi_{k_c}, \theta_1, \dots, \theta_{k_c})$ the mixture parameters, $\phi_i \in \mathbb{R}$ a mixture weight that satisfies $\sum_{i=1}^{k_c} \phi_i = 1$ and $\phi_i \geq 0$, \mathcal{N} a multivariable normal distribution defined as

$$\mathcal{N}(\mathbf{x}; \theta) = \frac{\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_i)\right)}{\sqrt{(2\pi)^l |\boldsymbol{\Sigma}|}}, \quad (6)$$

with parameters $\theta = (\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} \in \mathbb{R}^l$ is a vector of means and $\boldsymbol{\Sigma} \in \mathbb{R}^{l \times l}$ a covariance matrix for l features. In this sense, Assumption 1 is established.

Assumption 1. *The error signal $e(k)$ has stochastic nature and it is composed of a mixture of a finite number of Gaussian distributions with unknown parameters.*

For a given number of clusters k_c , the unknown parameters Ψ can be estimated by the Expectation-maximization (EM) algorithm with identification data (or training data) [27]. The EM algorithm is an iterative method for finding maximum likelihood solutions, which initialize the parameter (either using the k -means algorithm or randomly) and then alternate two steps called E-step and M-step.

The *Expectation* step (E-step) evaluates the training data with current parameters to compute the posterior probabilities or the expectation of the log-likelihood. The *Maximization* step (M-step) re-estimates the parameter by maximizing the expected log-likelihood obtained in the E-step. These new estimated parameters are used in the next E-step and so on, guaranteeing a progressive increase in the log-likelihood until the parameter converges (within a given tolerance) or reaches the maximum number of iterations.

Several criteria and methodologies have been proposed to select the number of clusters k_c [28, 29]. The criteria reported by Christopher Bishop in [27] is followed for developing this paper. This strategy only needs to perform a single training, given a relatively large number k_c of components, inspecting the mixture weights $\phi = (\phi_1, \dots, \phi_{k_c})$ to select the number of most relevant mixtures. Thereby, a threshold λ_k is defined, and then k_c will be the number of weights greater than λ_k . Once determined k_c , a new GMM will be trained with this number of components.

Having the basic concepts related to GMM, the concept drift will be detected when the GMM of error loses accuracy, i.e., the error distribution has changed. This strategy consists of two main steps: 1) detect outliers and 2) count the number of outliers within a sliding window and check whether or not there is a concept drift.

2.2.1. Outlier detection

By using a sequential analysis technique called the Page-Hinkley test [30], changes in the probability distribution can be detected. To measure the goodness of fit from the GMM to a data sample, the log-likelihood criterion is used, defined as

$$\log \mathcal{L}(e(k); \Psi) = \log (p(\mathbf{x} = e(k); \Psi)),$$

where $\log \mathcal{L}(e(k); \Psi) \leq 0$ for any value of $e(k)$ since the likelihood function p is defined within the range $[0, 1]$. Therefore, log-likelihood values close to 0 mean that the samples are more likely.

As the signal error can have noise, disturbances, or error peaks during transitions (device switching) that produce false negatives in log-likelihood, then the log-likelihood should be smoothed out by filtering out sudden spikes. The filtering procedure used in this paper was the Slope Comparing Adaptive Repeated Median (SCARM) [31] – an adaptive online repeated median filter – since it is a robust filter that does not need explicit parameters. In this way, it is not required to predefined a suitable width of the sliding window as conventional approaches, e.g., moving average filter.

From smoothed log-likelihood

$$\hat{\bar{l}}(k) = \text{scarm}(\log \mathcal{L}(e(k); \Psi)),$$

the Page-Hinkley test [30] is performed to detect when the probability distribution deviates from normal behavior. This test measures the point of change in the mean of a sequence of a normal random variable using a cumulative sum test scheme, i.e.,

$$\text{cum}(k) = \text{cum}(k-1) + \hat{\bar{l}}(k) - \bar{l}(k) - \omega, \quad (7a)$$

$$\bar{l}(k) = \frac{1}{n_{\text{cum}}} \sum_{i=0}^{n_{\text{cum}}-1} \hat{\bar{l}}(k-i), \quad (7b)$$

being $\bar{l}(k)$ a moving average with windows of $n_{\text{cum}} = \min\{n_w, k\}$, n_w the window width for $k \geq n_w$ and $\omega \in \mathbb{R}_{\geq 0}$ a tolerance that defines the maximum accepted change. Typically, ω can have a value close to $\frac{\bar{l}(k) - \bar{l}^*}{2}$, where \bar{l}^* is known as *Rejectable quality level* [30]. Thus, to assess the rejection of the no-change hypothesis, a ph signal is defined as

$$\text{ph}(k) = \text{cum}_{\text{max}}(k) - \text{cum}(k), \quad (8a)$$

with

$$\text{cum}_{\text{max}}(k) = \max\{\text{cum}(k-i), i = 0, \dots, n_{\text{cum}} - 1\}, \quad (8b)$$

which is used to determine the reject signal that detects outliers as follows:

$$\text{reject}(k) = \begin{cases} 1, & \text{if } \text{ph}(k) \geq \lambda \\ 0, & \text{otherwise} \end{cases}, \quad (9)$$

with a threshold $\lambda \in \mathbb{R}_{\geq 0}$ to label the current sample as an outlier with value 1. Then, the no-change hypothesis would be rejected.

2.2.2. Concept drift detection

In order to discriminate outliers produced by noise, brief disturbances, transient states, or model degradation (concept drift), a threshold s is defined to establish the minimum number of non-outliers (within a defined window) required to indicate whether the model is suitable. Therefore, a concept drift is detected when the following inequality is satisfied:

$$r = \sum_{i=0}^{n_w-1} (1 - \text{reject}(k-i)) < s, \quad (10)$$

being r the number of non-outliers. Note that r starts to be calculated when $k \geq n_w$.

For establishing the values of n_w and s , the criteria of an adaptive window proposed by Diaz-Rozo et al. [21] was followed. Based on Batch Sampling, Chernoff bound and let $r(k - n_w), \dots, r(k)$ be independent trials, Assumption 2 is established.

Assumption 2. Let $r(k_w)$ be a discrete random variable $\forall k_w \in \{k - n_w, \dots, k\}$, which follows a Bernoulli distribution with parameter p , i.e.,

$$r(k_w) \sim \text{Bern}(p).$$

Remark 1. Let $X = \{x_1, \dots, x_{n_{\text{Bern}}}\}$ be a set of independent trials that follows the Bernoulli distribution with parameter $p \in [0, 1]$, such that: $\forall j \in \{1, \dots, n_{\text{Bern}}\}$, $x_j \in \{0, 1\}$, and $\forall x_j$, $\Pr[x_j = 1] = p$ or $\Pr[x_j = 0] = 1 - p$. Besides, X satisfies that $E[X] = p$ and $\text{Var}(X) = p(1 - p)$.

Thus, the classical sample size problem can be posed to determine the smallest size that satisfies a probability, which is defined as follows:

$$\Pr[|\tilde{p} - p| \leq \epsilon p] \geq 1 - \frac{\delta}{2}, \quad (11)$$

where $\epsilon \in (0, 1)$ is the margin of absolute error and $\delta \in (0, 2)$ is the confidence parameter. Inequality (11) was developed by Osamu Watanabe [32], resulting in adaptive window size bounded by

$$n_w \leq \frac{3(1 + \epsilon)}{(1 - \epsilon)\epsilon^2 p} \ln\left(\frac{2}{\delta}\right), \quad (12)$$

knowing p in advance, which according to Diaz-Rozo et al. [21], p is estimated every cycle as r/n_w , as long as it satisfies with $r \geq s$, otherwise $p = 1$.

Algorithm 1 The adaptive mechanism algorithm

Precondition: \mathbf{U}_{iden} and \mathbf{S}_{iden} , identification input-output data

Precondition: n , model order

Precondition: ω and λ , Page-Hinkley test parameters

Precondition: δ and ϵ , Batch sampling parameters

- 1: Compute the initial state-space model (1) and train the initial GMM (5). Using $p = 1$, s and n_w are initialized with Eqs. (13) and (12), respectively
 - 2: Start-up of the machine and control system
 - 3: Sensor and reference model generate $S(k)$ and $\hat{S}(k)$, respectively, adaptation law block receives them, and compute $e(k)$
 - 4: *Outlier detection step:*
Smooth the log-likelihood $\hat{l}(k)$, and compute $\text{cum}(k)$, $\text{ph}(k)$ and $\text{reject}(k)$ signals with Eqs. (7), (8), and (9), respectively
 - 5: *Concept drift detection step:*
Calculate the $r(k)$ signal using Eq. (10) If $r(k) < s$ then a concept drift is detected and go to step 6, else $p = r/n_w$ and go to step 7
 - 6: Using the last measures of S and u , a new state-space model is calculated, as well as a new GMM is trained and compute n_w with $p = 1$, updating the reference model
 - 7: If the process has finished then exit, else go to step 3
-

Futhermore, s is a defined window

$$s = \frac{3(1 + \epsilon)}{\epsilon^2} \ln \left(\frac{2}{\delta} \right), \quad (13)$$

which does not depend on p .

When a concept drift is detected, a new model is generated with the stored measurements from the last machine cycle and replaced by the current reference model. Algorithm 1 gathers all the main steps that the adaptive mechanism block must follow.

3. Adaptive energy model in a real industrial setup

In this section, both the validation procedure and the performance results of the proposed approach are presented. First, the used industrial

experimental testbed is described, together with an explanation of the emulated concept drifts. Next, the results are shown and discussed.

3.1. Experiments

The performance of the proposed approach was validated using the industrial testbed shown in Figure 5. This testbed is certificated by the Industrial Internet Consortium (IIC) [33], whose objective is to validate Industrial Internet of Things (IIoT) technologies for smart factories, having the typical components of a machine tool: two tool axis (spindles), an X-axis of translation (servomotor), a robot to change tools, a pneumatic system, a coolant system, and an industrial control system (e.g., PLC, CNC). In order to monitor the power consumption of the machine, an energy meter called Aingura Insights¹ (AI) is installed at the electrical source entrance of the machine, following a NILM scheme [5]; AI acquires energy samples with a frequency of 8 kHz and is equipped with a Zynq® Ultrascale+™ MPSoC to process them. In this way, the energy consumption of the whole machine is measured, whereas the states (e.g., speed, torque) of every element in the machine are extracted from PLC and CNC, without affecting any of its processes.

For the validation test, a machine cycle was established where the robot installs and removes the tools on the spindles at the beginning and end of the cycle, respectively, and during the cycle, the spindles and the X-axis are activated at different speeds, while the peripheral equipment (i.e., pneumatic system and coolant system) are activated automatically to satisfy machine conditions. An experiment was carried out with this machine cycle, with a duration of four hours and 43 minutes. Moreover, three concept drifts were emulated by modifying the settings of the axis and the peripheral equipment to affect their consumption profile.

The collected data of both smart energy meter and control system were post-processed, resampling the data with a digital filter at 100 Hz and synchronizing them to generate a dataset that establishes the system inputs as machine states and the outputs as power consumption measures. The following inputs were established from the five most consuming devices of the machine: the angular speeds of the two spindles and the X-axis servo motor, and activation/deactivation signals of both cooling system motor and pneumatic system pump. The first machine cycle of the dataset was used

¹<https://www.ainguraiiot.com>

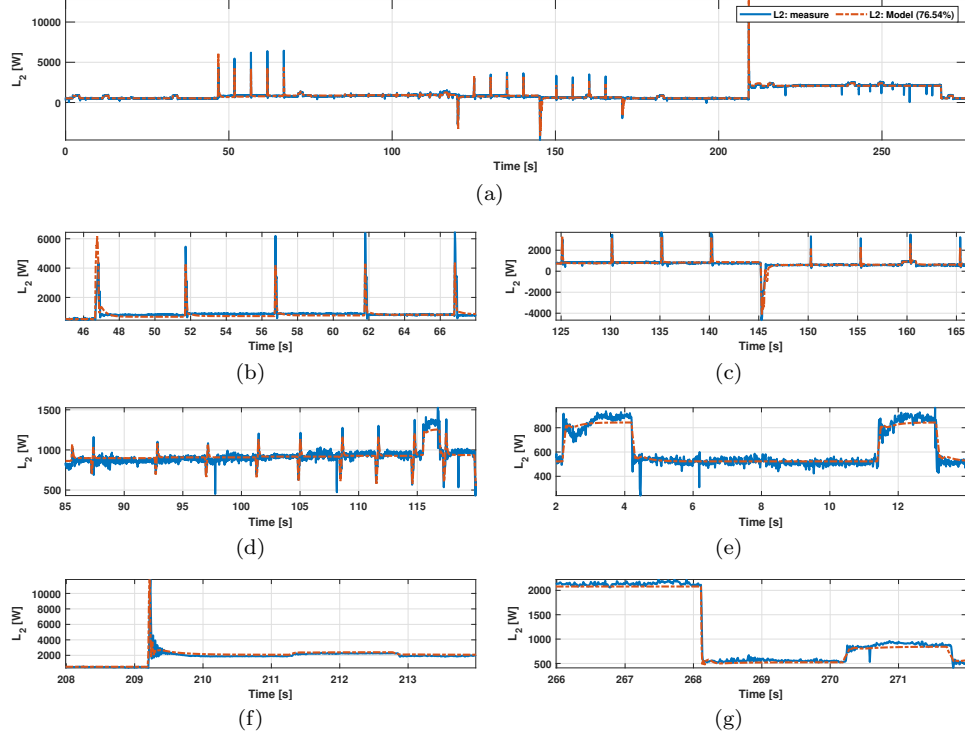


Figure 4: Initial reference model performance. Consumption profile and estimation of a machine cycle, with a close view of the main consumers. (a) Comparison between the energy consumption profile and the estimate of the second power phase. (b) Red spindle. (c) Blue spindle. (d) The X-axis servo motor. (e) The pneumatic system pump. (f) Cooling system motor started. (g) Cooling System motor shutdown.

as training data to compute the initial model (with order n equal to six). This initial model was tested with training data and yielded an average fit rate of 77.2% for the three phases. Figure 4 displays the energy profile of the second power phase since the performance of the other phases is similar.

3.2. Results

The adaptive energy model was validated with the following settings: the outlier detection step was tuned to classify outliers greater than 20% of the maximum value of ph signal obtained with the training data, using a $\omega = 0$, giving a value of $\lambda = 3000$. The concept-drift detection step was configured with a maximum absolute marginal error of 5 % and a confidence parameter greater than 80 %, i.e., $\epsilon = 0.05$ and $\delta = 0.4$. With these parameters, the width of the windows to detect concept drift are $s = 2027$ and $n_w = 2134$



Figure 5: Industrial Internet Consortium Testbed.

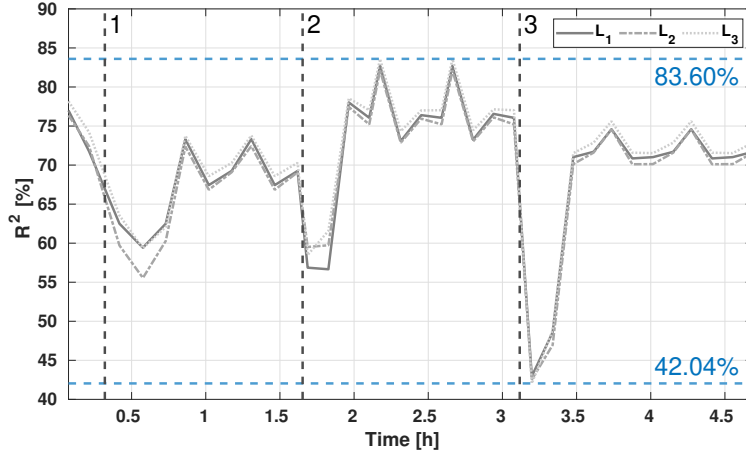


Figure 6: Evolution model fit-rate.

when $p = 1$.

Using the generated dataset as a data stream, the proposed approach was evaluated, resulting in the evolution of the fit rate per machine cycle shown in Figure 6, being labeled the three emulated concept drifts as vertical dashed lines. Figure 6 demonstrates the ability of the adaptive mechanism to recover the fit rate against concept drifts, achieving a maximum fit rate of 82 %. Moreover, analyzing the amount of drop in the fit rate after each concept drift indicates how abrupt the concept drift is, or in a context of failures, the impact of a failure on power consumption and how it evolves, e.g., in this experiment, it is notable that the drops of the concept drift increase exponentially and if it is due to failure, then the failure impact increases exponentially.

Figure 7a shows a comparison between the energy measurements and the prediction signals, which was filtered (with a mean filter) to improve the

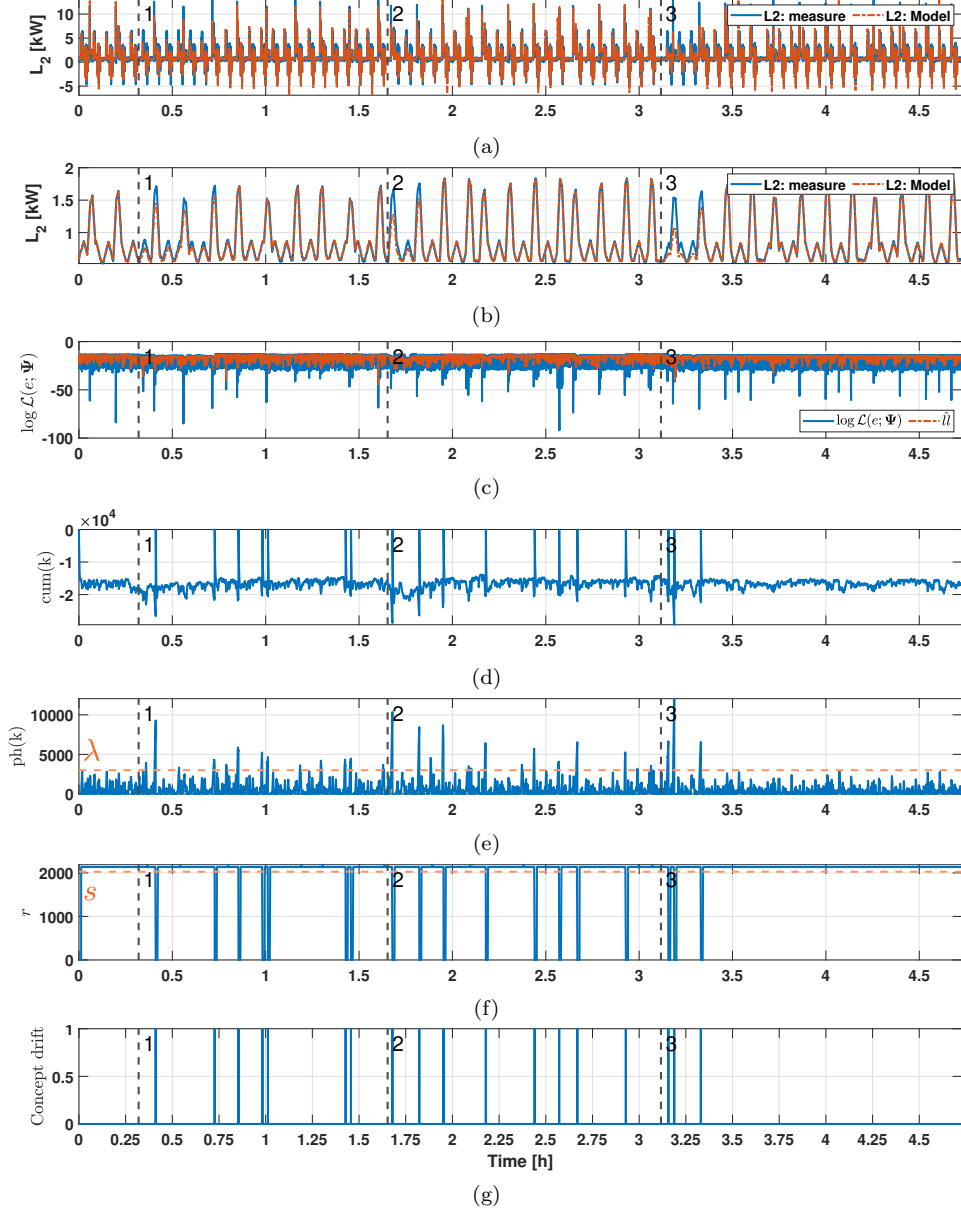


Figure 7: Adaptive energy model signals obtained during the experiment. (a) Profiles of measured and predicted energy consumption of the second power phase. (b) Energy consumption profiles with a moving average. (c) Log-likelihood signal. (d) cum signal. (e) Page-Hinkley signal. (f) r signal. (g) Concept drift signal.

Table 1: Fit rates for the entire experiment.

Lines		\mathbf{L}_1	\mathbf{L}_2	\mathbf{L}_3
Fit rate (\mathbf{R}^2) [%]	Sub-ID	34.28	32.12	33.73
	ASub-ID	68.54	67.96	69.42

Table 2: Fit rate of the first and last machine cycle by model.

Fit rate (\mathbf{R}^2) [%]	Machine cycle	ASub-ID	Sub-ID	ARX	ANN
	Initial	77.17	75.57	65.15	58.32
	Final	72.05	13.57	16.05	0

visualization and perceive the similarity of both signals (see Figure 7b). In this way, the performance of the estimated energy consumption is perceived and illustrates that it can follow the measurements. After each concept drift, the prediction loses accuracy but the adaptive mechanism manages to recover it. From the error signal between measures and prediction, the computed log-likelihood signal is displayed in Figure 7c, along with its filtered signal \hat{l} to reduce noise, avoiding false positives when detecting outliers. The cum signal (7) was computed using \hat{l} signal and is shown in Figure 7d. Note that the cum signal captures the relevant variations of the \hat{l} signal, demonstrating that it is a suitable feature to insight the behavior behind the \hat{l} signal. Thus, the obtained ph signal (8) from the cum signal is presented in Figure 7e. With the λ established, the outliers were correctly detected after each concept drift (see Figure 7g).

With the ph signal, the r signal (10) is calculated by counting the number of non-outliers within a window with a variable width n_w and is shown in Figure 7f. When the r signal is less than s , a concept drift is detected. Therefore, after every true concept drift, the adaptive approach detects concept drifts, although it is not instantaneous since the concept drift is artificially introduced and labeled when all devices are turned off, then the outlier was detected after the first equipment was turned on. Among the emulated concept drifts, more than one concept drifts were detected. This because the adaptive methodology has a transient state until reaching the proper model that presents the new behavior.

3.3. Comparative assessment

The performance of the proposed adaptive energy model (ASub-ID) was compared with the non-adaptive model (Sub-ID), using the established ini-

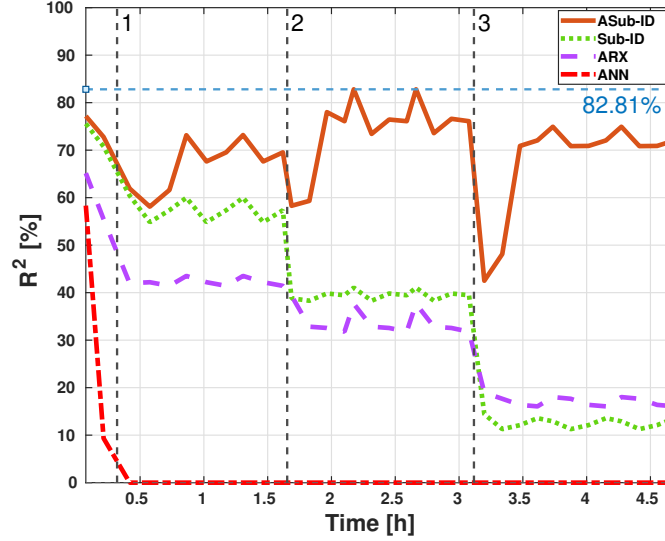


Figure 8: Comparison of different evolution fit rate per machine cycle for each type of model.

tial model for ASub-ID. As a result, ASub-ID obtained a fit rate higher than 68% (see Table 1) for the entire experiment, which represents an improvement of up to 50 % compared to a non-adapted model. Furthermore, with training data, an autoregressive-exogenous (ARX) model and an ANN were computed to compare the accuracy of each type of model. The parameters of the models were tuned to obtain the best fit rate, resulting in the following: an ARX model of 10th order and an ANN with two layers and 50 hidden neurons.

Averaging the fit rate of model prediction of each line, the fit rate per machine cycle for each approach is presented in Figure 8. Furthermore, Table 2 shows the fit rate values at the beginning and end of the experiment by model. ASub-ID had the best energy prediction over time, followed by Sub-ID, although it is overpassed by the ARX model after the third concept drift. ANN performed poorly as it needs more training data to find better weights to improve its predictions. Therefore, to model dynamic behaviors, system identification approaches (such as Sub-ID and ARX) manage to calculate suitable models with fewer model parameters and training data than ANN approaches. On the other hand, models without an adaptive mechanism lose precision over time, with greater drops after each concept drift. Thus, regardless of the structure of the model, the integration of an adaptive mechanism in the digital twin paradigms enhances the prediction accuracy

and makes it possible to monitor the changes that the system undergoes.

4. Conclusion

Through the optimization and monitoring of energy consumption in factories, this research is motivated to increase the efficiency and sustainability of the manufacturing process. For this reason, this work contributes to the literature with an online energy modeling methodology, which generates dynamic models of energy consumption of industrial machines. This approach has high accuracy and adaptive capabilities, i.e., it is capable of detecting new energy consumption behaviors (concept drift) and generating a new energy model that considers them. The combination of a system identification approach with a concept drift detection method represents a novelty that provides model adaptability capabilities in digital twin applications. Thus, this data-driven methodology allows generating multipurpose energy models that can be used by multiple external systems (e.g., energy management systems, diagnostic systems, prediction systems), providing reliability by sharing the new model to those systems that are interested, keeping them up-to-date. Therefore, the proposed approach innovatively enhances the performance of applications aimed at energy efficiency on industrial systems and also aids in their diagnostic to detect early failures in electromechanical systems.

This modeling strategy was validated on an industrial testbed, which has allowed to design and carry out an experiment with three concept drifts artificially introduced. As a result, the performance of the adaptive approach revealed an improvement in prediction by doubling the fit rate compared to the non-adaptive model. The concept-drift detection enables manufacturers to measure how the system is changing, which large consumer is changing, and to give early warning when the system behaves abnormally or experiences a failure. On the other hand, the proposed approach was compared with other widely known modeling methods, the autoregressive-exogenous model and artificial neural networks. The adaptive approach proposed outperformed and obtained a maximum fit rate per machine cycle of 82.81%, while the other models had less than 66%.

The meaningful findings obtained demonstrate the proper performance of the proposed methodology, innovatively combining dynamic models with machine learning algorithms to contribute to optimal management of energy resources. In this way, the proposed data-driven modeling incorporates an

adaptive feature, which can be individually enhanced. Therefore, as further works, automation energy modeling processes need to be improved, as well as to include more input characterizations to increase the fit rate and robustness of the solution. The adaptation mechanism needs to reduce the transient time after detecting a true concept drift and eliminate false concept drift. On the other hand, with the proposed approach, a strategy to measure the energy degradation of electromechanical systems must be developed, which helps in the early detection of faults or anomalies.

References

- [1] Jenny L Diaz and Carlos Ocampo-Martinez. Energy efficiency in discrete-manufacturing systems: insights, trends, and control strategies. *Journal of Manufacturing Systems*, 52:131–145, 2019.
- [2] Simone Baldi, Fan Zhang, Thuan Le Quang, Petr Endel, and Ondrej Holub. Passive versus active learning in operation and adaptive maintenance of heating, ventilation, and air conditioning. *Applied Energy*, 252:113478, 2019.
- [3] Anh Hoang, Phuc Do, and Benoît Iung. Energy efficiency performance-based prognostics for aided maintenance decision-making: Application to a manufacturing platform. *Journal of Cleaner Production*, 142:2838–2857, 2017.
- [4] Jianjun Zhang, Xuanqun Chen, Wing WY Ng, Chun Sing Lai, and Loi Lei Lai. New appliance detection for nonintrusive load monitoring. *IEEE Transactions on Industrial Informatics*, 15(8):4819–4829, 2019.
- [5] Hui Liu. Detection of transient events in time series. In *Non-intrusive Load Monitoring*, pages 23–44. Springer, 2020.
- [6] JP Seevers, J Johst, T Weiß, H Meschede, and J Hesselbach. Automatic time series segmentation as the basis for unsupervised, non-intrusive load monitoring of machine tools. *Procedia CIRP*, 81:695–700, 2019.
- [7] John W Sutherland, David A Dornfeld, and Barbara S Linke. *Energy Efficient Manufacturing: Theory and Applications*. John Wiley & Sons, 2018.
- [8] Anna Carina Roemer and Steffen Strassburger. A review of literature on simulation-based optimization of the energy efficiency in production.

In *2016 Winter Simulation Conference (WSC)*, pages 1416–1427. IEEE, 2016.

- [9] Gil-Yong Lee, Mincheol Kim, Ying-Jun Quan, Min-Sik Kim, Thomas Joon Young Kim, Hae-Sung Yoon, Sangkee Min, Dong-Hyeon Kim, Jeong-Wook Mun, Jin Woo Oh, et al. Machine health management in smart factory: A review. *Journal of Mechanical Science and Technology*, 32(3):987–1009, 2018.
- [10] Chiara Cimino, Elisa Negri, and Luca Fumagalli. Review of digital twin applications in manufacturing. *Computers in Industry*, 113:103130, 2019.
- [11] Raunak Bhinge, Jinkyoo Park, Kincho H Law, David A Dornfeld, Mon-eer Helu, and Sudarsan Rachuri. Toward a generalized energy prediction model for machine tools. *Journal of manufacturing science and engineering*, 139(4), 2017.
- [12] Wei Liu, Li Li, Wei Cai, Congbo Li, Lingling Li, Xingzheng Chen, and John W Sutherland. Dynamic characteristics and energy consumption modelling of machine tools based on bond graph theory. *Energy*, 212: 118767, 2020.
- [13] YC Liang, X Lu, WD Li, and S Wang. Cyber physical system and big data enabled energy efficient machining optimisation. *Journal of cleaner Production*, 187:46–62, 2018.
- [14] Jing Zou, Qing Chang, Jorge Arinez, and Guoxian Xiao. Data-driven modeling and real-time distributed control for energy efficient manufacturing systems. *Energy*, 127:247–257, 2017.
- [15] Jun Xie, Fei Liu, and Hang Qiu. An integrated model for predicting the specific energy consumption of manufacturing processes. *The International Journal of Advanced Manufacturing Technology*, 85(5-8): 1339–1346, 2016.
- [16] Zaiwu Mei, Jianwan Ding, Liping Chen, Ting Pi, and Zaidao Mei. Hybrid multi-domain analytical and data-driven modeling for feed systems in machine tools. *Symmetry*, 11(9):1156, 2019.
- [17] Seung-Jun Shin, Jungyub Woo, Sudarsan Rachuri, and Prita Meilani-tasari. Standard data-based predictive modeling for power consumption in turning machining. *Sustainability*, 10(3):598, 2018.

- [18] HAN, Yongming. Energy management and optimization modeling based on a novel fuzzy extreme learning machine: Case study of complex petrochemical industries. In *Energy conversion and management*, volume 165, pages 163–171, 2018.
- [19] Neethu Mohan, KP Soman, and S Sachin Kumar. A data-driven strategy for short-term electric load forecasting using dynamic mode decomposition model. *Applied energy*, 232:229–244, 2018.
- [20] Miguel Bermeo-Ayerbe and Carlos Ocampo-Martinez. Energy consumption dynamical models for smart factories based on subspace identification methods. In *2019 IEEE 4rd Colombian Conference on Automatic Control (CCAC)*, pages 1–6. IEEE, 2019.
- [21] Javier Diaz-Rozo, Concha Bielza, and Pedro Larrañaga. Clustering of data streams with dynamic gaussian mixture models: an iot application in industrial processes. *IEEE Internet of Things Journal*, 5(5):3533–3547, 2018.
- [22] Hugo Sequeira, Paulo Carreira, Thomas Goldschmidt, and Philipp Vorst. Energy cloud: Real-time cloud-native energy management system to monitor and analyze energy consumption in multiple industrial sites. In *2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing*, pages 529–534. IEEE, 2014.
- [23] Juan C Gómez and Enrique Baeyens. Subspace-based identification algorithms for hammerstein and wiener models. *European Journal of Control*, 11(2):127–136, 2005.
- [24] Peter Van Overschee and Bart De Moor. *Subspace Identification for Linear Systems : Theory - Implementation - Applications*. Springer US, Boston, MA, 1996. ISBN 9781461304654.
- [25] A Colin Cameron and Frank AG Windmeijer. An r-squared measure of goodness of fit for some common nonlinear regression models. *Journal of econometrics*, 77(2):329–342, 1997.
- [26] João Gama, Indrė Žliobaitė, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, 46(4):1–37, 2014.
- [27] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

- [28] Lei Xu. A trend on regularization and model selection in statistical learning: a bayesian ying yang learning perspective. In *Challenges for Computational Intelligence*, pages 365–406. Springer, 2007.
- [29] Xuelei Hu and Lei Xu. A comparative study of several cluster number selection criteria. In *International Conference on Intelligent Data Engineering and Automated Learning*, pages 195–202. Springer, 2003.
- [30] David V Hinkley. Inference about the change-point from cumulative sum tests. *Biometrika*, 58(3):509–523, 1971.
- [31] Matthias Borowski and Roland Fried. Online signal extraction by robust regression in moving windows with data-adaptive width selection. *Statistics and Computing*, 24(4):597–613, 2014.
- [32] Osamu Watanabe. Simple sampling techniques for discovery science. *IEICE Transactions on Information and Systems*, 83(1):19–26, 2000.
- [33] Industrial Internet Consortium. Smart factory machine learning for predictive maintenance testbed, 2022. URL <http://www.iiconsortium.org/smart-factory-machinelearning.htm>.