



# Article Virtual Sensing of Key Variables in the Hydrogen Production Process: A Comparative Study of Data-Driven Models

Yating Yao<sup>1</sup>, Yupeng Xing<sup>1</sup>, Ziteng Zuo<sup>1</sup>, Chihang Wei<sup>2</sup> and Weiming Shao<sup>1,\*</sup>

- <sup>1</sup> Department of Chemical Equipment and Control Engineering, College of New Energy, China University of Petroleum (East China), Qingdao 266580, China; s21150032@s.upc.edu.cn (Y.Y.); b22150006@s.upc.edu.cn (Y.X.); 2015010229@s.upc.edu.cn (Z.Z.)
- <sup>2</sup> School of Information Science and Technology, Hangzhou Normal University, Hangzhou 311121, China; weichihang@hznu.edu.cn

\* Correspondence: shaoweiming@upc.edu.cn

Abstract: Hydrogen is an ideal energy carrier manufactured mainly by the natural gas steam reforming hydrogen production process. The concentrations of CH<sub>4</sub>, CO, CO<sub>2</sub>, and H<sub>2</sub> in this process are key variables related to product quality, which thus need to be controlled accurately in real-time. However, conventional measurement methods for these concentrations suffer from significant delays or huge acquisition and upkeep costs. Virtual sensors effectively compensate for these shortcomings. Unfortunately, previously developed virtual sensors have not fully considered the complex characteristics of the hydrogen production process. Therefore, a virtual sensor model, called "moving window-based dynamic variational Bayesian principal component analysis (MW-DVBPCA)" is developed for key gas concentration estimation. The MW-DVBPCA considers complicated characteristics of the hydrogen production process, involving dynamics, time variations, and transportation delays. Specifically, the dynamics are modeled by the finite impulse response paradigm, the transportation delays are automatically determined using the differential evolution algorithm, and the time variations are captured by the moving window method. Moreover, a comparative study of data-driven virtual sensors is carried out, which is sporadically discussed in the literature. Meanwhile, the performance of the developed MW-DVBPCA is verified by the real-life natural gas steam reforming hydrogen production process.

**Keywords:** hydrogen production process; data-driven virtual sensor; real-time estimation; variational Bayesian principal component analysis

# 1. Introduction

Hydrogen is an ideal energy storage medium with zero carbon emissions, essential in energy systems' low-carbon transformation [1]. Currently, over 90% of the world's large-scale industrial hydrogen production process is the fossil energy reforming process, of which more than 50% is the natural gas steam reforming process [2,3]. The natural gas steam reforming hydrogen production process consists of feedstock purification, steam reforming, medium temperature conversion, and pressure swing adsorption. In this process, CH<sub>4</sub>, CO, CO<sub>2</sub>, and H<sub>2</sub> are primary process gases, and the concentrations of these gases are key variables (KVs) related to the product quality. Infrequent and inaccurate measurements of these KVs risk substantial production losses, poor control performance, and even hidden dangers to safety. Consequently, real-time measurements of these KVs are essential and desirable.

Currently, measurements of the hydrogen production process' KVs are classified into offline laboratory analyses, hardware sensors, and virtual sensors. The offline laboratory analyses give accurate measurements of the KVs, but result in significant measurement delays [4]. The hardware sensors measure the KVs in real-time, but need colossal investment and maintenance costs [5,6]. Virtual sensors are actually predictive mathematical



Citation: Yao, Y.; Xing, Y.; Zuo, Z.; Wei, C.; Shao, W. Virtual Sensing of Key Variables in the Hydrogen Production Process: A Comparative Study of Data-Driven Models. *Sensors* **2024**, *24*, 3143. https:// doi.org/10.3390/s24103143

Academic Editor: Chun Sing Lai

Received: 6 January 2024 Revised: 24 April 2024 Accepted: 13 May 2024 Published: 15 May 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). models which use explanatory variables (EVs, i.e., easily measurable variables like pressure and flow rate) as inputs and estimates of the KVs as outputs, having the benefits of no measurement delays and low costs [7,8]. Therefore, the virtual sensor effectively compensates for the shortcomings of the offline laboratory analysis and hardware sensor [9,10]. As a result, virtual sensors have been intensively studied and widely used in the hydrogen production process.

Virtual sensors can generally be grouped into first-principle-based and data-driven virtual sensors. The first-principle-based virtual sensor is established by analyzing and explaining the physicochemical mechanisms of the hydrogen production process. For example, Yang et al. established a predictive model of ammonia pyrolysis decomposition rate through theoretical derivation for the reaction in the hydrogen production process [11]. Huang et al. developed a mathematical model between the EVs and the conversion rate of methane by coupling conduction, convection, thermal radiation, and chemical reaction kinetics [12]. Zhou et al. simulated the chemical looping hydrogen production process by analyzing the process mechanisms [13]. However, due to the complex reaction dynamics, precise first-principle-based models of the hydrogen production process are challenging to obtain. Usually, laboratory-scale studies or ideal steady-state analyses are the main applications of these first-principle-based virtual sensors. On the contrary, a data-driven virtual sensor is constructed by process-measured data, and does not rely on accurate mechanisms [14,15]. Therefore, it is closer to reality, and better describes the actual operations of the hydrogen production process [16,17]. As a result, data-driven virtual sensors are widely applied in the hydrogen production process. For example, Tong et al. built the multi-layer perceptron (MLP) neural network to estimate the  $CH_4$ concentration in the hydrogen production process [18]. Zamaniyan et al. employed the MLP model to predict the concentration of  $H_2$  and CO of the hydrogen production process [19]. Ögren et al. established deep neural networks (DNNs) to estimate the concentration of  $CH_4$ , CO, CO<sub>2</sub>, and H<sub>2</sub>, which have large network layers [20]. Considering the challenges of artificial neural network models in model selection and parameter adjustment with limited data, small-data-oriented models have been applied to the hydrogen production process. Fang et al. used a support vector machine (SVM) to predict the hydrogen yield [21]. Zhao et al. proposed a predictive model based on the least squares SVM (LSSVM) employed in wind power hydrogen production [22].

Despite these achievements on virtual sensors for the KVs, the hydrogen production process is usually complicated because of practical operations. Firstly, due to the feed-stock variations and external disturbances, the hydrogen production process shows strong dynamics [23]. Secondly, the hydrogen production process involves many series-wound devices (e.g., pre-reformers and pre-heaters), which requires significant time to deliver the energies and feedstocks to the reformer [24]. That means the EVs and the KVs are recorded simultaneously and are mismatched (i.e., there are transportation delays between the EVs and the KVs). Thirdly, the hydrogen production process exhibits time-varying properties because of the process drifts caused by mechanical abrasions, catalyst deactivation, or even climatic variations [25].

The above complex properties of the hydrogen production process make it challenging to develop high-precision virtual sensors for the KVs. As far as we know, the existing virtual sensors developed for the hydrogen production process do not account for these complicated characteristics. Moreover, the hydrogen production process has multiple KVs to estimate, and the existing virtual sensors constructed multiple independent single-output virtual sensors (one for each KV), ignoring the correlations between the KVs. To this end, we first develop a multi-output virtual sensor model called "dynamic variational Bayesian principal component analysis (DVBPCA)" for real-time prediction of the KVs in the hydrogen production process. As a further contribution, a moving window-based DVBPCA (MW-DVBPCA) is developed to improve estimation performance while considering the time variations of the hydrogen production process. The main contributions we make in the article are organized as follows.

- (1) The developed DVBPCA considers both process dynamics and transportation delays of energies and materials. Concretely, the finite impulse response (FIR) method is employed to model the dynamics of the hydrogen production process. And the transportation delays related to the EVs are automatically determined by differential evolution (DE). Moreover, the DVBPCA is able to make full use of the correlations between KVs for performance enhancement.
- (2) The moving window (MW) approach is employed for updating the DVBPCA with the latest online process information, which effectively captures the time-varying characteristics of the hydrogen production process in real-time.
- (3) A comparative study of data-driven virtual sensors is implemented for the hydrogen production process, which is sparsely mentioned in the predecessors' research. Furthermore, the performance of the developed MW-DVBPCA is verified by the real-life natural gas steam reforming hydrogen production process.

This article is structured as follows. The variational inference (VI) is introduced in Section 2. Section 3 describes the developed DVBPCA and MW-DVBPCA in detail. In Section 4, the performance of the DVBPCA and MW-DVBPCA is evaluated using the operational data from the distributed control system (DCS) database of a real-life hydrogen production process. Finally, Section 5 concludes the article.

#### 2. Variational Inference

The VI and Monte Carlo–Markov Chain are common methods to study posterior distributions over RVs in probabilistic models [26]. And the VI was used to learn the developed DVBPCA, considering its advantages in convergence diagnosis and computational efficiency.

Denote the random variables (RVs) with unknown posterior distribution as  $\Theta$ . The equation shown as Equation (1) decomposes the log of model evidence, as follows:

$$\ln p(\mathbf{D}) = \mathcal{L}(q) + \mathrm{KL}(q \parallel p)$$

$$\mathrm{KL}(q \parallel p) = \int q(\mathbf{\Theta}) \ln \left\{ \frac{q(\mathbf{\Theta})}{p(\mathbf{\Theta} \mid \mathbf{D})} \right\} d\mathbf{\Theta}$$

$$\mathcal{L}(q) = \int q(\mathbf{\Theta}) \ln \left\{ \frac{p(\mathbf{D}, \mathbf{\Theta})}{q(\mathbf{\Theta})} \right\} d\mathbf{\Theta}$$
(1)

where KL( $q \parallel p$ ) means the Kullback–Leibler (KL) divergence between  $q(\Theta)$  and  $p(\Theta \mid D)$ , D is training dataset,  $q(\Theta)$  represents any form of probability distribution over  $\Theta$ ,  $p(\Theta \mid D)$  represents the actual posterior distribution over  $\Theta$ , and  $\mathcal{L}(q)$  means the variational evidence lower bound (ELBO) of the log of model evidence ln P(D) as KL( $q \parallel p$ )  $\geq 0$ .

Therefore, the posterior distributions over  $\Theta$  can be found by maximizing the ELBO  $\mathcal{L}(q)$  as shown in Equation (2), i.e.,

$$p(\mathbf{\Theta} \mid \mathbf{D}) = q^*(\mathbf{\Theta}) = \underset{q(\mathbf{\Theta})}{\arg \max} \mathcal{L}(q(\mathbf{\Theta}))$$
(2)

where  $q^*(\Theta)$  represents the variational solution of  $q(\Theta)$ .

Assume  $q(\Theta)$  can factorize into the product  $\prod_{j=1}^{J} q_j(\theta_j)$ , where  $\Theta = \{\theta_1, \theta_2, \dots, \theta_j\}$ . The general rule to find the variational distribution  $q_j(\theta_j)$  is obtained as Equation (3):

$$\ln q_j^*(\boldsymbol{\theta}_j) = \int \ln p(\boldsymbol{D}, \boldsymbol{\Theta}) \prod_{i \neq j} q_i^*(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i + \mathbb{C} = \mathbb{E}_{q_i(\boldsymbol{\theta}_{i \neq j})}[\ln p(\boldsymbol{D}, \boldsymbol{\Theta})] + \mathbb{C}$$
(3)

where  $\mathbb{E}_{q(\theta)}[f(\theta)]$  is the expectation of a function  $f(\theta)$  of the RV  $\theta$  with respect to the distribution  $q(\theta)$  over  $\theta$ ,  $\theta_{i\neq j}$  means factors in  $\Theta$  except  $\theta_j$ , and  $\mathbb{C}$  means constant terms.

# 3. Dynamic Variational Bayesian Principal Component Analysis Based on Moving Window

# 3.1. Time-Delayed Moving Average Model

Due to the flexibility of combining various regression models, the FIR paradigm is generally employed for capturing dynamics. Moreover, the KVs are measured at a lower frequency than the EVs in the hydrogen production process, which implies that the model structure based on autoregression from the FIR family is unsuitable. More importantly, as mentioned earlier, the transportation delays between the EVs and the KVs are unknown, and require consideration. Based on these considerations, a time-delayed moving average (MA) model structure is embedded in the virtual sensor, which is given by

$$\boldsymbol{y}(n) = f \begin{bmatrix} x_1(n-l_1), x_1(n-l_1-1), \cdots, x_1(n-l_1-p_1), \\ x_2(n-l_2), x_2(n-l_2-1), \cdots, x_2(n-l_2-p_2), \\ \vdots \\ x_S(n-l_S), x_S(n-l_S-1), \cdots, x_S(n-l_S-p_S) \end{bmatrix}$$
(4)

where  $y(n) \in \mathbb{R}^{K_y}$  are the *n*-th measurements of the  $K_y$ -dimensional KVs,  $x_i(n)$  is *n*-th measurement of the *i*-th EV, *S* is the number of the EVs,  $l_i$  and  $p_i$  refer to the transportation delay and the order of  $x_i$ , respectively, and both are non-negative integers, and  $f(\cdot)$  means the mathematical relationship between the EVs and KVs (but there may exist strong colinearities between the EVs of the model  $f(\cdot)$ ). Denote  $l = (l_1, l_2, \dots, l_S)$  and  $p = (p_1, p_2, \dots, p_S)$  for conciseness.

# 3.2. Dynamic Variational Bayesian Principal Component Analysis

Variational Bayesian principal component analysis (VBPCA) is a widely used model to deal with the colinearities between the EVs, owing to its advantage in the automatic determination of appropriate model dimensionality [27]. Therefore, it is ideally suited as the model  $f(\cdot)$ . In this article, name the VBPCA combined with the model structure shown in Equation (4) as the dynamic VBPCA (DVBPCA) due to the capability of capturing process dynamics. Moreover, the DVBPCA can handle the overfitting resulting from highly dimensional variable augmentation in Equation (4) through penalizing parameter values, which are detailed as follows.

Denote the observed sequential dataset as  $D = (d_1, d_2, \dots, d_N)^T$ , where  $d_n = \begin{pmatrix} x_n \\ y_n \end{pmatrix} \in \mathbb{R}^K$  means the *n*-th set of samples,  $x_n = (v_1^{l_1, p_1}(n), v_2^{l_2, p_2}(n) \dots, v_{K_x}^{l_{K_x}, p_{K_x}}(n))^T \in \mathbb{R}^{K_k}$  and  $y_n = y(n) \in \mathbb{R}^{K_y}$  are inputs and outputs, respectively, and  $v_i^{l_i, p_i}(n) = (x_i(n-l_i), x_i(n-l_i-1), \dots, x_i(n-l_i-p_i))$ . Figure 1 graphically represents the DVBPCA. where,  $W_x \in \mathbb{R}^{K_x \times M}$  is the loading matrix of the inputs  $x_n, W_y \in \mathbb{R}^{K_y \times M}$  is the loading matrix of the outputs  $y_n, W = \begin{pmatrix} W_x \\ W_y \end{pmatrix}$ ,  $z_n \in \mathbb{R}^M$  are the *M*-dimensional hidden variables corresponding to the sample at time instant  $n, \mu_x \in \mathbb{R}^{K_x}$  are the mean values of the inputs  $x_n, \mu_y \in \mathbb{R}^{K_y}$  are the mean values of the outputs  $y_n, \mu = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}$ ,  $\tau$  is the accuracy parameter of the noise, and  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_M)^T$  is the accuracy parameter set of the loading matrix. Moreover, denote each column of *W* as  $w_m$ , the transpose of each row of *W* as  $w_k^*$ , and  $Z = (z_1, z_2, \dots, z_N)^T$ . As illustrated in Equation (5), the observed data point  $d_n$  is assumed to be generated by

$$d_n = W z_n + \mu + \varepsilon_n \tag{5}$$

where  $\varepsilon_n$  is the measurement noise.



Figure 1. Probability graph of the DVBPCA.

As shown in Equation (6), assign conjugate prior to the hidden variables

$$p(z_n) = \mathcal{N}(z_n \mid \mathbf{0}_M, \mathbf{I}_M) \tag{6}$$

where  $\mathcal{N}(\cdot)$  represents the normal distribution probability density function,  $\mathbf{0}_M$  represents an *M*-dimensional column vector with element 0, and  $\mathbf{I}_M$  represents an *M*-dimensional unit matrix.

According to Figure 1, which shows that the samples are independently collected, the conditional distributions of the observed samples are designed as shown in Equations (7); that is,

$$p(\boldsymbol{D} \mid \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\mu}, \tau) = \prod_{n=1}^{N} \mathcal{N} \Big( \boldsymbol{d}_n \mid \boldsymbol{W} \boldsymbol{z}_n + \boldsymbol{\mu}, \tau^{-1} \mathbf{I}_K \Big)$$
(7)

where  $I_K$  represents a *K*-dimensional unit matrix.

To conform with common sense and to simplify subsequent calculations, the distributions over all RVs are selected as conjugate priors as detailed in Equation (8)–(11); that is,

$$p(\mathbf{W} \mid \boldsymbol{\alpha}) = \prod_{m=1}^{M} \mathcal{N}\Big(\boldsymbol{w}_m \mid \mathbf{0}_K, \boldsymbol{\alpha}_m^{-1} \mathbf{I}_K\Big)$$
(8)

$$p(\boldsymbol{\mu}) = \mathcal{N}\left(\boldsymbol{\mu} \mid \mathbf{0}_{K}, \boldsymbol{\beta}^{-1}\mathbf{I}_{K}\right)$$
(9)

$$p(\boldsymbol{\alpha}) = \prod_{m=1}^{M} \mathcal{G}(\alpha_m \mid a_{\alpha}, b_{\alpha})$$
(10)

$$p(\tau) = \mathcal{G}(\tau \mid a_{\tau}, b_{\tau}) \tag{11}$$

where  $\mathbf{0}_K$  represents a *K*-dimensional column vector with element 0,  $\beta$  is the accuracy parameter of the mean,  $a_{\alpha}$ ,  $b_{\alpha}$ ,  $a_{\tau}$ ,  $b_{\tau}$  are the hyper-parameters of the distributions over these RVs, and  $\mathcal{G}(\cdot)$  is the gamma distribution probability density function. Particularly, when little or no prior information is obtained for these priors, it is suggested to set noninformative priors for the RVs to minimize the effect on the posterior distribution. That is, the values of hyperparameters  $a_{\alpha}$ ,  $b_{\alpha}$ ,  $a_{\tau}$ , and  $b_{\tau}$  are set as small as possible so that the posteriors do not rely on the priors, but only on the data.

**Remark 1.** To facilitate model training, rewrite Equation (8) as  $p(\mathbf{W} \mid \mathbf{\alpha}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{w}_{k}^{*} \mid \mathbf{0}, \mathbf{A}^{-1})$ , where  $\mathbf{A} = diag\{\alpha_{1}, \alpha_{2}, ..., \alpha_{M}\}$ . Each inverse variance of the  $w_{m}$  is controlled by the corresponding  $\alpha_{m}$ . Thus, if the posterior distribution of a particular  $\alpha_{m}$  is focused on larger values, the corresponding  $w_{m}$  will converge to be tiny, effectively deactivating this particular direction in the latent space. Therefore, the effective dimensionality of the potential space is determined automatically.

According to Figure 1, the equation shown as Equation (12) represents the joint probability density function between training data D and the RVs  $\Theta$  as follows:

$$p(\boldsymbol{D},\boldsymbol{\Theta}) = p(\boldsymbol{D} \mid \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\mu}, \tau) p(\boldsymbol{Z}) p(\boldsymbol{W} \mid \boldsymbol{\alpha}) p(\boldsymbol{\mu}) p(\boldsymbol{\alpha}) p(\tau).$$
(12)

The VI is employed to train the DVBPCA to obtain the posterior distributions  $q(\Theta)$  of the RVs  $\Theta = \{Z, W, \mu, \alpha, \tau\}$ .

Assume the posterior distribution over  $\boldsymbol{\Theta}$ ,  $q(\boldsymbol{\Theta})$ , can factorize into

$$q(\mathbf{\Theta}) = q(\mathbf{Z})q(\mathbf{W})q(\boldsymbol{\mu})q(\boldsymbol{\alpha})q(\tau)$$
(13)

According to Equation (2), the posterior distribution  $q^*(z_n)$  over  $z_n$  is detailed as

$$\ln q^{*}(\mathbf{Z}) = \mathbb{E}_{\mathbf{W},\boldsymbol{\mu},\boldsymbol{\alpha},\tau}[\ln p(\mathbf{D},\boldsymbol{\Theta})] + \mathbb{C}$$

$$= \mathbb{E}_{\mathbf{W},\boldsymbol{\mu},\tau}[\ln p(\mathbf{D} \mid \mathbf{Z},\mathbf{W},\boldsymbol{\mu},\tau)] + \mathbb{E}[p(\mathbf{Z})] + \mathbb{C}$$

$$= \sum_{n=1}^{N} \mathbb{E}_{\mathbf{W},\boldsymbol{\mu},\tau}[-\frac{\tau}{2}(d_{n} - \mathbf{W}z_{n} - \boldsymbol{\mu})^{\mathrm{T}}(d_{n} - \mathbf{W}z_{n} - \boldsymbol{\mu})] - \sum_{n=1}^{N} \mathbb{E}[\frac{1}{2}z_{n}^{\mathrm{T}}z_{n}] + \mathbb{C}$$

$$= \sum_{n=1}^{N} \{\langle \tau \rangle (d_{n} - \langle \boldsymbol{\mu} \rangle)^{\mathrm{T}} \langle \mathbf{W} \rangle z_{n} - \frac{1}{2}(\langle \tau \rangle \langle \mathbf{W}^{\mathrm{T}}\mathbf{W} \rangle + \mathbf{I}_{M})z_{n}\} + \mathbb{C}$$

$$= \sum_{n=1}^{N} \ln \mathcal{N}(z_{n} \mid m_{z}^{(n)}, \mathbf{\Sigma}_{z})$$
(14)

where  $m_z^{(n)} = \langle \tau \rangle \Sigma_z \langle W^T \rangle (d_n - \langle \mu \rangle)$ ,  $\Sigma_z = (\langle \tau \rangle \langle W^T W \rangle + \mathbf{I}_M)^{-1}$ , and  $\langle \cdot \rangle$  is the expectation.

Similarly, by using the general rule given by Equation (2), we can find the variational posterior distributions over other RVs, as shown on the right side of Equation (13) one by one, which are given as follows.

The posterior distribution  $q^*(W)$  over W is obtained as

$$q^*(\mathbf{W}) = \prod_{k=1}^{K} \mathcal{N}\left(\boldsymbol{w}_k^* \mid \boldsymbol{m}_w^{(k)}, \boldsymbol{\Sigma}_w\right)$$
(15)

where  $\Sigma_w = \left(\sum_{n=1}^N \langle \tau \rangle \langle z_n z_n^{\mathrm{T}} \rangle + A \right)^{-1}$ , and  $m_w^{(k)} = \sum_{n=1}^N \langle \tau \rangle \Sigma_w \langle z_n \rangle (d_{nk} - \langle \mu_k \rangle)$ . The posterior distribution  $q^*(\mu)$  over  $\mu$  is obtained as

$$J^*(\boldsymbol{\mu}) = \mathcal{N}(\boldsymbol{\mu} \mid \boldsymbol{m}_{\boldsymbol{\mu}}, \boldsymbol{\Sigma}_{\boldsymbol{\mu}})$$
(16)

where  $\Sigma_{\mu} = (\langle \tau \rangle N + \beta)^{-1} \mathbf{I}_{K}$ , and  $m_{\mu} = \sum_{n=1}^{N} \langle \tau \rangle \Sigma_{\mu} (d_{n} - \langle W \rangle \langle z_{n} \rangle)$ . The posterior distribution  $q^{*}(\alpha)$  over  $\alpha$  is obtained as

$$q^*(\boldsymbol{\alpha}) = \prod_{m=1}^M \mathcal{G}(\alpha_m \mid c_\alpha, d_\alpha)$$
(17)

where  $c_{\alpha} = \frac{K}{2} + a_{\alpha}$ , and  $d_{\alpha} = b_{\alpha} + \frac{1}{2} \langle w_m^{\mathrm{T}} w_m \rangle$ .

The posterior distribution  $q^*(\tau)$  over  $\tau$  is obtained as

$$q^*(\tau) = \mathcal{G}(\tau \mid c_\tau, d_\tau) \tag{18}$$

where  $d_{\tau} = b_{\tau} + \frac{1}{2} \sum_{n=1}^{N} \{ d_n^{\mathsf{T}} d_n - 2d_n^{\mathsf{T}} (\langle \mathbf{W} \rangle \langle \mathbf{z}_n \rangle + \langle \boldsymbol{\mu} \rangle) + \langle \mathbf{z}_n^{\mathsf{T}} \mathbf{W}^{\mathsf{T}} \mathbf{W} \mathbf{z}_n \rangle + 2 \langle \mathbf{z}_n^{\mathsf{T}} \rangle \langle \mathbf{W}^{\mathsf{T}} \rangle \langle \boldsymbol{\mu} \rangle + \langle \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{\mu} \rangle \},$ and  $c_{\tau} = \frac{NK}{2} + a_{\tau}.$ 

The ELBO  $\mathcal{L}(q)$  is calculated using the updated posterior distributions, i.e.,

$$\mathcal{L}(q) = \mathbb{E}[\ln p(\boldsymbol{D}, \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \tau)] - \mathbb{E}[\ln q^*(\boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\mu}, \boldsymbol{\alpha}, \tau)]$$
  

$$= \mathbb{E}[\ln p(\boldsymbol{D} \mid \boldsymbol{Z}, \boldsymbol{W}, \boldsymbol{\mu}, \tau)] + \mathbb{E}[\ln p(\boldsymbol{Z})] + \mathbb{E}[\ln p(\boldsymbol{W} \mid \boldsymbol{\alpha})] + \mathbb{E}[\ln p(\boldsymbol{\mu})]$$
  

$$+ \mathbb{E}[\ln p(\boldsymbol{\alpha})] + \mathbb{E}[\ln p(\tau)] - \mathbb{E}[\ln q^*(\boldsymbol{Z})] - \mathbb{E}[\ln q^*(\boldsymbol{W})] - \mathbb{E}[\ln q^*(\boldsymbol{\mu})]$$
  

$$- \mathbb{E}[\ln q^*(\boldsymbol{\alpha})] - \mathbb{E}[\ln q^*(\tau)]$$
(19)

The convergence of the training procedure can be diagnosed with ELBO. The termination criteria for the parameter learning process are defined as detailed in Equation (20), as follows:

$$\left|\frac{\mathcal{L}^{iter+1} - \mathcal{L}^{iter}}{\mathcal{L}^{iter+1}}\right| < \delta \tag{20}$$

where  $\mathcal{L}^{iter+1}$  and  $L^{iter}$  are the values of ELBO at the (*iter* + 1)-th and *iter*-th iterative steps, respectively, and  $\delta$  is the threshold.

The main steps in training the DVBPCA model are summarized in Algorithm 1.

### Algorithm 1 Pseudocode for the DVBPCA.

- 1: **Input:** The observed dataset *D*, *M*,  $\beta$ , tol, maximum number of iterations *Iter*<sub>max</sub>, prior hyper-parameters ( $a_{\alpha}, b_{\alpha}, a_{\tau}, b_{\tau}$ ).
- 2: **Initialization:** initialize  $m_z$ ,  $\Sigma_z$ ,  $m_w$ ,  $\Sigma_w$ ,  $m_\mu$ ,  $\Sigma_\mu$ ,  $c_\alpha$ ,  $d_\alpha$ ,  $c_\tau$ ,  $d_\tau$ .
- 3: **for** iter=1 to  $Iter_{max}$  **do**
- 4: VB-E-Step
- 5: Calculate  $m_z$ ,  $\Sigma_z$  using Equation (14).
- 6: VB-M-Step
- 7: Calculate  $m_w$ ,  $\Sigma_w$  using Equation (15).
- 8: Calculate  $m_{\mu}$ ,  $\Sigma_{\mu}$  using Equation (16).
- 9: Calculate  $c_{\alpha}$ ,  $d_{\alpha}$  using Equation (17).
- 10: Calculate  $c_{\tau}$ ,  $d_{\tau}$  using Equation (18).
- 11: Calculate  $\mathcal{L}(q)$  using Equation (19).
- 12: if the termination condition is satisfied then
- 13: **Break**
- 14: **end if**
- 15: end for
- 16: **Output:** the posterior distributions of RVs  $\Theta$ .

#### 3.3. Moving Window-Based Dynamic Variational Bayesian Principal Component Analysis

The moving window (MW) method is used to trace time variations and reject disturbances [28]. It captures the latest process information by discarding the oldest sample after the latest sample is obtained and rebuilding a model with all data in the window. Specially, for the first local model, as mentioned above, set noninformation priors for the RVs. For subsequent local modeling, the prior distributions of the RVs in the MW-DVBPCA are updated one by one and replaced with the trained posterior distributions of the RVs at the previous moment. The MW-DVBPCA is detailed in this subsection.

Firstly, the training dataset (i.e., historical data) is preset in the MW, and the DVBPCA is trained on the training dataset to obtain the posterior distributions of the RVs  $\Theta$ . When the test samples (i.e., online samples) are acquired, the DVBPCA generated earlier is used to estimate the KVs, and the posterior distributions of RVs are used as the prior for the new DVBPCA model, which is elaborated as follows.

Let  $x_{n'}$  and  $y_{n'}$  denote the observations of inputs and outputs at certain online time instant n', respectively, where  $y_{n'}$  is unknown. The hidden RVs  $z_{n'}$  are first introduced.

Since the outputs  $y_{n'}$  are unknown, the posterior distribution over  $z_{n'}$  is calculated with the observation of  $x_{n'}$  which, based on Equation (14), is given by the equation as shown in Equation (21) as follows:

$$q^*(\boldsymbol{z}_{n'}) = \mathcal{N}\left(\boldsymbol{z}_n \mid \boldsymbol{m}_z^{(n')}, \boldsymbol{\Sigma}_z^{n'}\right)$$
(21)

$$p(\boldsymbol{y}_{n'} \mid \boldsymbol{x}_{n'}) = \int p(\boldsymbol{y}_{n'} \mid \boldsymbol{z}_{n'}, \boldsymbol{W}_{y}, \boldsymbol{\mu}_{y}, \tau) p(\boldsymbol{z}_{n'} \mid \boldsymbol{x}_{q}) q^{*}(\boldsymbol{W}_{y}) q^{*}(\tau) dz dW d\mu d\tau$$

$$= \int \mathcal{N} \Big( \boldsymbol{y}_{n'} \mid \boldsymbol{W}_{y}^{\mathrm{T}} \boldsymbol{z}_{n'} + \boldsymbol{\mu}_{y}, \tau^{-1} \mathbf{I}_{K} \Big) \mathcal{N} \Big( \boldsymbol{z}_{n'} \mid \boldsymbol{m}_{z}^{(n')}, \boldsymbol{\Sigma}_{z}^{n'} \Big) \mathcal{N} \Big( \boldsymbol{W}_{y} \mid \boldsymbol{m}_{w}^{y}, \boldsymbol{\Sigma}_{w} \Big) \mathcal{N} \Big( \boldsymbol{\mu}_{y} \mid \boldsymbol{m}_{\mu}^{y}, \boldsymbol{\Sigma}_{\mu}^{yy} \Big) \mathcal{G}(\tau \mid c_{z}, d_{z}) dz dW d\mu d\tau$$

$$(22)$$

where  $m_w^y$  and  $\Sigma_w$  are the mean and covariance calculated from the posterior distribution  $q^*(W_y)$ , respectively, and  $m_\mu^y$  and  $\Sigma_\mu^{yy}$  are the mean values and covariance matrix calculated from the posterior distribution  $q^*(\mu_y)$ , respectively. Equation (22) can be transformed into

 $p(\mathbf{y}_{n'} \mid \mathbf{x}_{n'}) = \int \mathcal{N}\Big(\mathbf{y}_{n'} \mid \mathbf{m}_{w}^{y\mathrm{T}} \mathbf{m}_{z}^{(n')} + \mathbf{m}_{\mu}^{y} \operatorname{tr}\Big[\mathbf{\Sigma}_{w} \mathbf{\Sigma}_{z}^{n'} + \mathbf{\Sigma}_{w} \mathbf{m}_{z}^{(n')} \mathbf{m}_{z}^{(n')\mathrm{T}} + \mathbf{m}_{w}^{y} \mathbf{m}_{w}^{y\mathrm{T}} \mathbf{\Sigma}_{z}^{n'}\Big] + \mathbf{\Sigma}_{\mu}^{yy} + \tau^{-1} \mathbf{I}_{K}\Big) \mathcal{G}(\tau \mid c_{\tau}, d_{\tau}) d\tau$ (23)

where  $tr[\cdot]$  represents the trace.

The mean vector and covariance matrix of Gaussian distribution in Equation (23) are obtained as illustrated in Equation (24) and Equation (25), respectively, as follows:

$$\mathbb{E}_{p\left(\boldsymbol{y}_{n'}|\boldsymbol{x}_{n'},\tau\right)}[\boldsymbol{y}_{n'}] = \mathbb{E}[\boldsymbol{W}_{y}^{\mathrm{T}}\boldsymbol{z}_{n'} + \boldsymbol{\mu}_{y} + \boldsymbol{\varepsilon}_{y}] = \boldsymbol{m}_{w}^{y\mathrm{T}}\boldsymbol{m}_{z}^{(n')} + \boldsymbol{m}_{\mu}^{y}$$
(24)

$$\begin{aligned} \mathbb{V}_{p(y_{n'}|x_{n'},\tau)}[y_{n'}] &= \mathbb{V}[W_{y}^{T}z_{n'} + \mu_{y} + \varepsilon_{y}] \\ &= \mathbb{V}[W_{y}^{T}z_{n'}] + \mathbb{V}[\mu_{y}] + \mathbb{V}[\varepsilon_{y}] \\ &= \mathbb{E}[W_{y}^{T}z_{n'}z_{n'}^{T}W_{y}] - m_{w}^{yT}m_{z}^{(n')}m_{z}^{(n')T}m_{w}^{y} + \Sigma_{\mu}^{yy} + (\tau)^{-1}\mathbf{I}_{K} \\ &= \operatorname{tr}(\mathbb{E}[W_{y}W_{y}^{T}z_{n'}z_{n'}^{T}] - m_{w}^{y}m_{w}^{yT}m_{z}^{(n')}m_{z}^{(n')T}) + \Sigma_{\mu}^{yy} + (\tau)^{-1}\mathbf{I}_{K} \\ &= \operatorname{tr}\left[\Sigma_{w}\Sigma_{z}^{n'} + \Sigma_{w}m_{z}^{(n')}m_{z}^{(n')T} + m_{w}^{y}m_{w}^{yT}\Sigma_{z}^{n'}\right] + \Sigma_{\mu}^{yy} + \Sigma_{\mu}^{yy} + (\tau)^{-1}\mathbf{I}_{K} \end{aligned}$$

where  $p(\boldsymbol{\varepsilon}_y) = \mathcal{N}(\boldsymbol{\varepsilon}_y \mid 0, (\tau)^{-1} \mathbf{I}_K)$ .

By replacing the distribution of  $\tau$  with its expectation, the probability distribution of the predicted KVs is approximated as

$$p(\boldsymbol{y}_{n'} \mid \boldsymbol{x}_{n'}) \approx \mathcal{N}\left(\boldsymbol{y}_{n'} \mid \boldsymbol{m}_{w}^{y\mathrm{T}} \boldsymbol{m}_{z}^{(n')} + \boldsymbol{m}_{\mu}^{y}, \mathrm{tr}\left[\boldsymbol{\Sigma}_{w} \boldsymbol{\Sigma}_{z}^{n'} + \boldsymbol{\Sigma}_{w} \boldsymbol{m}_{z}^{(n')} \boldsymbol{m}_{z}^{(n')\mathrm{T}} + \boldsymbol{m}_{w}^{y} \boldsymbol{m}_{w}^{y\mathrm{T}} \boldsymbol{\Sigma}_{z}^{n'}\right] + \boldsymbol{\Sigma}_{\mu}^{yy} + \left(\frac{c_{\tau}}{d_{\tau}}\right)^{-1} \mathbf{I}_{K}\right).$$
(26)

Therefore, based on Equation (26), the estimations  $\hat{y}_{n'}$  of  $y_{n'}$  are given by

$$\hat{y}_{n'} = \mathbb{E}_{p(y_{n'}|x_{n'})}[y_{n'}] = m_w^{\text{yT}} m_z^{(n')} + m_\mu^{\text{y}}$$
(27)

and the prediction uncertainty is quantified by the variance matrix of  $y_{n'}$  denoted by  $\mathbb{V}_{p(y_{n'}|x_{n'})}[y_{n'}]$ , which is given by Equation (28), i.e.,

$$\mathbb{V}_{p\left(\boldsymbol{y}_{n'}|\boldsymbol{x}_{n'}\right)}[\boldsymbol{y}_{n'}] = \operatorname{tr}\left[\boldsymbol{\Sigma}_{w}\boldsymbol{\Sigma}_{z}^{n'} + \boldsymbol{\Sigma}_{w}\boldsymbol{m}_{z}^{(n')}\boldsymbol{m}_{z}^{(n')\mathrm{T}} + \boldsymbol{m}_{w}^{y}\boldsymbol{m}_{w}^{y\mathrm{T}}\boldsymbol{\Sigma}_{z}^{n'}\right] + \boldsymbol{\Sigma}_{\mu}^{yy} + \left(\frac{c_{\tau}}{d_{\tau}}\right)^{-1}\mathbf{I}_{K}.$$
 (28)

Once the KVs predictions for the new sample are completed, the historical data window is slid down to include the latest measured sample set  $\{x_{n'}, y_{n'}\}$  and to eliminate the oldest sample. The DVBPCA is then retrained according to the current window of the historical dataset for online prediction. This process repeats once new samples are obtained online.

#### 3.4. Differential Evolutionary-Based Model Selection

The DE is an effective and simple intelligent optimization algorithm commonly used for solving NP-hard problems in real number fields [29]. Associated studies have demonstrated that DE has fast convergence performance as an excellent global optimization algorithm. Therefore, the DE is utilized for model selection (i.e., selecting the best parameter) of the DVBPCA and MW-DVBPCA, involving the dynamic orders p and time delays lgiven in Equation (4). Equation (29) shows the optimization problem, as follows:

$$\min g(p, l)$$
s.t.  $p_{\min} \le p \le p_{\max}$ 

$$l_{\min} \le l \le l_{\max}$$

$$(29)$$

where  $g(\cdot)$  is the fitness function,  $p_{\min}$  and  $p_{\max}$  are the lower and upper bounds of the dynamic orders p, respectively, and  $l_{\min}$  and  $l_{\max}$  are the lower and upper bounds of the delays l, respectively.

Based on Equation (4),  $\{p, l\}$  are non-negative integers. Therefore, the individuals generated by the DE need to be rounded. And the DE algorithm contains three core evolutionary operations, i.e., mutation, crossover and selection, which are introduced in detail as follows.

*Initialization.* Set the population size *NP*, dimension of individual (2*S*), and the upper bounds ( $\rho_1^{(max)}, \rho_2^{(max)}, \dots, \rho_{2S}^{(max)}$ ) and the lower bounds ( $\rho_1^{(min)}, \rho_2^{(min)}, \dots, \rho_{2S}^{(min)}$ ) of the dynamic orders and time delays, where individuals ( $\rho_1, \rho_2, \dots, \rho_{2S}$ ) in the population are randomly generated in integer field by Equation (30), i.e.,

$$\boldsymbol{\rho}_{e} = \boldsymbol{\rho}_{e}^{(min)} + (\boldsymbol{\rho}_{e}^{(max)} - \boldsymbol{\rho}_{e}^{(min)}) \cdot \operatorname{rand}(NP, 1)$$
(30)

*Mutation*. In the *r*-th generation, DE generates a mutation vector  $\mathbf{v}_e$  for the *e*-th individual, given by Equation (31); that is,

$$\mathbf{v}_{e}^{r+1} = \boldsymbol{\rho}_{e_{0}}^{r} + F \cdot \left( \boldsymbol{\rho}_{e_{1}}^{r} - \boldsymbol{\rho}_{e_{2}}^{r} \right)$$
(31)

where  $\rho_{e_0}^r$ ,  $\rho_{e_1}^r$ , and  $\rho_{e_2}^r$  are individuals that randomly selected from population with  $e_0 \neq e_1 \neq e_2 \neq e_1$ , and *F* is the mutation rate.

*Crossover*. A trial individual vector  $u_e$  is generated by the original individual vector or the mutation vector based on the crossover rate *CR* ranged in (0, 1), which is shown in Equation (32), i.e.,

$$u_{i,e}^{r+1} = \begin{cases} v_{i,e}^{r+1}, & \text{if } \operatorname{rand}(0,1) \le CR \text{ or } e = \operatorname{rand}_e \\ \rho_{i,e'}^r, & \text{otherwise} \end{cases}$$
(32)

where rand<sub>*e*</sub> is a random integer ranged in (1, 2S).

*Selection.* According to the fitness function  $g(\cdot)$ , select the individual with a lower value as detailed in Equation (33) as follows:

$$\boldsymbol{\rho}_{e}^{r+1} = \begin{cases} \boldsymbol{u}_{e}^{r+1}, & \text{if } g(\boldsymbol{u}_{e}^{r+1}) \leq g(\boldsymbol{\rho}_{e}^{r}) \\ \boldsymbol{\rho}_{e'}^{r}, & \text{otherwise} \end{cases}$$
(33)

*Termination.* The evolution continues until it reaches the given maximum generation or the change in the fitness value is small. Otherwise, the above three operators should be performed repeatedly.

Figure 2 shows the flowchart of the DVBPCA and MW-DVBPCA combined with the DE for parameter optimization. Note that for the multi-output models, DVBPCA and MW-DVBPCA, the average prediction error of four KVs is set as the fitness function.



Figure 2. The flowchart of the DVBPCA and MW-DVBPCA.

# 4. Case Studies and Comparisons

A comparative study of data-driven virtual sensors upon a real-life natural gas steam reforming hydrogen production process is provided in this section. Meanwhile, a case study is useful for understanding the hydrogen production processes and data features. The state-of-the-art virtual sensors and developed models (DVBPCA and MW-DVBPCA) are investigated, including partial least squares (PLS) [30], VBPCA, long short-term memory (LSTM) [31], echo state networks (ESN) [32], and dynamic PLS (DPLS) [33]. The PLS is a classic static model with the advantages of modeling data colinearities and simple data structure. The VBPCA is a static probabilistic model. Different from the PLS, the VBPCA determines the number of principal components automatically, which is also the basis of our developed models (DVBPCA and MW-DVBPCA). The LSTM and ESN are advanced dynamic neural network models that consider the dynamics. The DPLS is the improvement of the PLS based on the FIR paradigm, which also accounts for the dynamics of the hydrogen production process.

# 4.1. Natural Gas Steam Reforming Hydrogen Production Process

The natural gas steam reforming hydrogen production process is the largest industrial source of hydrogen, and it consists of four main processes: feedstock purification, steam reforming, medium temperature conversion, and pressure swing adsorption. Among these processes, steam reforming is the dominant reaction process, which is schematically shown in Figure 3 [34]. Processed gases are mixed with steam in the pre-reformer, where all hydrocarbons and some  $CH_4$  are converted to CO,  $CO_2$  and  $H_2O$ . The temperature then decreases, which is not conducive to promoting hydrogen generation. The pre-reformer output is heated in a pre-heater and then continuously fed to the primary reformer for complete reforming.

The main chemical reactions of this process are shown as follows:

$$C_{n}H_{2n+2} + nH_{2}O \stackrel{\Delta}{\longleftrightarrow} nCO + (2n+1)H_{2}$$

$$CH_{4} + H_{2}O \stackrel{\Delta}{\longleftrightarrow} CO + 3H_{2}$$

$$CO + H_{2}O \stackrel{\Delta}{\longleftrightarrow} CO_{2} + H_{2}.$$
(34)

According to Equation (34), the exit gases of the hydrogen production process consist of  $CH_4$ , CO,  $CO_2$ , and  $H_2$ , and the concentrations of these gases are KVs (as labeled "Y" in

Figure 3) related to product quality. Therefore, these KVs need to be strictly monitored. In practice, offline laboratory analysis and hardware sensors are traditional methods for measuring the KVs, but have delays and high investment. Meanwhile, series-wound devices (such as pre-reformers and pre-heaters) introduce considerable transportation delays, which must be considered in virtual sensor modeling. Therefore, for giving real-time predictions of the KVs, a virtual sensor considering transportation delays is desirable.



Figure 3. The flowchart of steam reforming process.

# 4.2. Explanatory Variable Selection and Data Collection

A total of 13 EVs for the virtual sensor development are selected based on the first principles and the operation experience of the engineers. These EVs are easily measured variables, closely linked to the KVs and highly correlated with process variations, as shown by labels U1–U13 in Figure 3. The descriptions of these EVs are listed in Table 1.

Table 1. EVs in the hydrogen production process.

EV	Description		
U1	Flow of fuel natural gas into primary reformer		
U2	Flow of fuel off gas into primary reformer		
U3	Pressure of fuel off gas at the exit of heat exchanger 3		
U4	Pressure of furnace flue gas at primary reformer's exit		
U5	Temperature of fuel off gas at the exit of heat exchanger 3		
U6	Temperature of fuel natural gas at pre-heater's exit		
U7	Temperature of process gas at primary reformer's entrance		
U8	Temperature of furnace flue gas at primary reformer's top left		
U9	Temperature of furnace flue gas at primary reformer's top right		
U10	Temperature of mixed furnace flue gas at primary reformer's top		
U11	Temperature of transformed gas at primary reformer's left exit		
U12	Temperature of transformed gas at primary reformer's right exit		
U13	Temperature of transformed gas at primary reformer's exit		

The samples used to develop the virtual sensor of the KVs were obtained from the DCS database. The KVs' sampling rate is 10 min, and 2500 samples were collected. These were divided into three parts in order of collection time: the training set, the validation set, and the testing set, which consist of 1500 samples, 300 samples, and 700 samples, respectively. Figures 4 and 5 show sample division and the changing trend of these EVs and four KVs, respectively. In this article, we use scaled dimensionless data for modeling.



Figure 4. The diagram of data partition of 13 EVs.



Figure 5. The diagram of data partition of four KVs.

### 4.3. Evaluation Metrics

The root mean squares error (RMSE), the coefficient of determination ( $R^2$ ), and the mean absolute error (MAE) are chosen to quantify different models' performance, which are defined as

$$RMSE = \sqrt{\frac{1}{N'} \sum_{n'=1}^{N'} (\hat{y}_{n'} - y_{n'})^2}$$

$$R^2 = 1 - \sum_{n'=1}^{N'} (\hat{y}_{n'} - y_{n'})^2 / \sum_{n'=1}^{N'} (y_{n'} - \bar{y})^2$$

$$MAE = \frac{1}{N'} \sum_{n'=1}^{N'} |\hat{y}_{n'} - y_{n'}|$$
(35)

where  $\hat{y}_{n'}$  and  $y_{n'}$  are the estimated values and true values of the *n*'th testing sample, respectively, *N*' is the size of the testing dataset, and  $\bar{y} = \sum_{n'=1}^{N'} y_{n'}/N'$  are the mean of the KVs in the testing dataset. The RMSE and MAE indexes characterize the average and largest prediction errors on the test set, respectively, and the R<sup>2</sup> index characterizes the correlation between the predicted and true values. Therefore, the smaller the RMSE and MAE or the bigger the R<sup>2</sup>, the higher the predictive accuracy.

# 4.4. Parameter Selection

All models' optimal parameters need to be chosen to obtain different models' best prediction performance. Select the DE algorithm to minimize the average RMSE (i.e., the mean of the RMSE of the four KVs) on the validation set for parameter optimization of different models. For the PLS, the number of principal components and the time delays are the parameters to be optimized. For the DPLS, the number of principal components, the time delays and the dynamic orders are the parameters to be optimized. For the VBPCAbased models, the time delays and the dynamic orders are the parameters to be optimized. For the LSTM and ESN, the time delays are the parameters to be optimized.

For the VBPCA-based models, the key issue of automatic parameter determination of the principal component number is whether each column of the loading matrix W is either insignificant or significant, i.e.,  $w_m = 0$  or  $w_m \neq 0$ , which can be controlled by  $\alpha$ according to Equation (8). Take DVBPCA for example, the fact that the variance of each  $w_m$ can be quantified by  $1/\mathbb{E}(\alpha_m)$  for m = 1, ..., M, which is displayed in Figure 6. The color distinction of the points in Figure 6 means the value of  $1/\mathbb{E}(\alpha_m)$  is either insignificant or significant by setting threshold, where red means significant and blue means insignificant. As shown in Figure 6, the appropriate dimensionality of the principal component subspace is selected as 39.



**Figure 6.** Variance of each  $w_m$  by DVBPCA.

For the ESN, set the input regulation scale to 0.1, set the reservoir size to 50, and set the spectral radius to 0.8. Considering that the ESN involves the random weights in the reservoir computing step, 20 modeling tests were performed. For each trial, the random weights are saved and fixed, and the DE algorithm is used to optimize the model delays further. Then, the best results from 20 tests are picked for model performance comparison. Note that the ESN is a single-output model; we therefore construct four ESN models, one for each KV. For the LSTM, based on the debugging experience, set the time step to 1, set the learning rate to 0.1, set the hidden layer to 1, and set the neuron number in the hidden layer to 100, which performs favorably empirically in each of the replicated experiments of this work.

For the MW-DVBPCA, the impact of MW size is detailed in Figure 7. The MW size is in units of 10 mins, i.e., setting the MW size to 50 means that the MW contains 500 min of data. As shown in Figure 7, for small MW sizes, the data in the window may not appropriately represent the relationship between process variables. In contrast, excessive MW size covers too much outdated sample data so that the MW-DVBPCA fails to track the process change adequately. Therefore, according to Figure 7, the MW size was selected as 50.



**Figure 7.** Impact of MW size for: (a)  $CH_4$  concentration, (b) CO concentration, (c)  $CO_2$  concentration, and (d)  $H_2$  concentration.

# 4.5. Results and Analysis

The estimations on the test set of the KVs obtained by the investigated seven virtual sensors are visualized in Figures 8–10. In Figure 8, obviously, the PLS and VBPCA have the poor estimation performance. Because of the significant dynamics of the hydrogen production process, the estimation accuracy of the static models, PLS and VBPCA, is not as satisfactory as that of the other five dynamic models (such as around the 120-th sample of the  $CO_2$  concentration). Figure 9 shows that the estimated values of the DVBPCA tracks real values better than the other three models (such as around the 300-th sample of the CO concentration and around the 300-th sample of the  $CO_2$  concentration). That is because, on the one hand, the DVBPCA can deal with the colinearities between the EVs compared with the LSTM and ESN. On the other hand, the DVBPCA can tackle the overfitting results from high-order variable augmentation compared with the DPLS. Moreover, the ESN constructs four virtual sensors, one for each KV for estimation, but the DVBPCA is a multi-output model that considers the inherent relationships between the KVs. Figure 10 illustrate that the estimations of the four KVs by the MW-DVBPCA match the real values much better (particularly in the localized area of the  $CH_4$  concentration around the 440-th sample) than other models, revealing the importance of considering time variation properties in the virtual sensor modeling of the hydrogen production process. Moreover, although the overall predicted values do match well with the true ones when the proposed MW-DVBPCA method was used, some discrepancies can be observed between test sample number 200 and 300 for all the four KVs. The possible reasons for these differences in the predicted and true values are as follows. Firstly, the characteristics of the samples between test sample number 200 and 300 are changing rapidly, and model learning does not accommodate such changes in time. Secondly, the samples between test sample number 200 and 300 have nonlinearities, but the local model constructed by MW-DVBPCA is linear. Overall, it is recognized that the proposed models show noticeable advantages over the benchmark models.



**Figure 8.** Estimations of four KVs by the PLS, VBPCA, and LSTM: (a)  $CH_4$  concentration, (b) CO concentration, (c)  $CO_2$  concentration, and (d)  $H_2$  concentration.



**Figure 9.** Estimations of four KVs by the ESN, DPLS, and DVBPCA: (a)  $CH_4$  concentration, (b) CO concentration, (c)  $CO_2$  concentration, and (d)  $H_2$  concentration.



**Figure 10.** Estimations of four KVs by the MW-DVBPCA: (a) CH<sub>4</sub> concentration, (b) CO concentration, (c) CO<sub>2</sub> concentration, and (d) H<sub>2</sub> concentration.

Figure 11 compares the seven models in terms of scatter plots. Based on Figure 11, a further comparison of the prediction of the data-driven models can be made. As shown in Figure 11a, the predictions of the PLS for  $CH_4$  component deviate obviously from the real values. Figure 11b,c shows that the VBPCA model improves the prediction accuracy of CO and CO<sub>2</sub> concentrations somewhat compared to the PLS model. However, the overall prediction accuracy is still very low. Figure 11d reveals that all models have relatively poor prediction accuracy for the H<sub>2</sub> concentration, but the MW-DVBPCA presents a better result. As indicated in Figure 11, the scatters by MW-DVBPCA are more closely and clearly located around the diagonal line than those of other models, thus illustrating better performance.



**Figure 11.** Scatter plot comparisons between various models for: (a)  $CH_4$  concentration, (b) CO concentration, (c)  $CO_2$  concentration, and (d)  $H_2$  concentration.

The estimation performance of all data-driven virtual sensors is quantitatively tabulated in Tables 2 and 3. For a further comparison, Tables 2 and 3 show the estimations of the data-driven virtual sensors not considering transportation delays. These models' parameter selection is consistent with the corresponding model considering transportation delays. Overall, the estimation results in Tables 2 and 3 provide an initial validation of the effectiveness of the data-driven virtual sensors. However, the performances of the different data-driven virtual sensors vary considerably. The performance of the models that consider delays is better than that of the corresponding models that do not, as shown in Tables 2 and 3. Take the DVBPCA as an instance. The predictive performance based on the RMSE index of the four KVs by the model accounting for the time delays is improved by 18.6%, 13.0%, 2.3%, and 5.5%, respectively, compared to the model ignoring time delays. This is because the hydrogen production process has substantial time delays, and it has been proven that ignoring the delays could result in significantly deteriorated performance [29]. The R<sup>2</sup> values for two static models, i.e., the PLS and VBPCA, are as low as below 0.5; in contrast, the dynamic models, such as LSTM, ESN, and DPLS, show significantly better performance than the PLS and VBPCA, indicating the dynamic model

better fits the data features of the actual hydrogen production process. Moreover, due to the capability of dealing with overfitting, the DVBPCA performs better than the DPLS. Concretely, compared to the DPLS, the RMSEs of the four KVs obtained by the DVBPCA are decreased by 13.4%, 1.8%, 3.2%, and 4.8%, respectively. Moreover, the MW-DVBPCA further improves the estimations of the four KVs. The R<sup>2</sup>s of the CH<sub>4</sub> and H<sub>2</sub> concentrations reach as high as up to 0.9. Compared with the DVBPCA, the predictive performance on the four KVs by the MW-DVBPCA improves by 1.3%, 10.3%, 2.8%, and 33.4%, respectively, in terms of the MAE index.

		CH <sub>4</sub>			СО		
		RMSE	<b>R</b> <sup>2</sup>	MAE	RMSE	<b>R</b> <sup>2</sup>	MAE
	PLS	0.1915	-0.1329	0.1643	0.1944	-0.0975	0.1483
	VBPCA	0.2061	-0.2566	0.1703	0.2054	-0.2259	0.1592
	LSTM	0.1567	0.2733	0.1373	0.1351	0.5370	0.1080
$(\mathbb{T})$	ESN	0.1150	0.6088	0.0965	0.1148	0.6173	0.0968
0	DPLS	0.0942	0.7373	0.0770	0.0956	0.7348	0.0764
	DVBPCA	0.0827	0.7975	0.0704	0.0826	0.8018	0.0662
	MW-DVBPCA	0.0525	0.9186	0.0388	0.0709	0.8538	0.0533
	PLS	0.1348	0.4623	0.1077	0.1352	0.4687	0.1067
2	VBPCA	0.1339	0.4694	0.1092	0.1315	0.4976	0.1052
	LSTM	0.1227	0.5543	0.1051	0.0829	0.8128	0.0606
	ESN	0.0955	0.7301	0.0759	0.0912	0.7582	0.0730
	DPLS	0.0777	0.8215	0.0647	0.0732	0.8444	0.0587
	DVBPCA	0.0673	0.8659	0.0536	0.0719	0.8497	0.0574
	MW-DVBPCA	0.0511	0.9228	0.0368	0.0665	0.8717	0.0515

Table 2. Quantitative estimation results of various models for the concentrations of CH<sub>4</sub> and CO.

① The time delays are not considered, ② the time delays are considered.

		CO <sub>2</sub>			H <sub>2</sub>		
		RMSE	R <sup>2</sup>	MAE	RMSE	R <sup>2</sup>	MAE
(1)	PLS	0.1930	0.0095	0.1536	0.1882	-0.0729	0.1612
	VBPCA	0.1913	0.0264	0.1521	0.1973	-0.1792	0.1660
	LSTM	0.1562	0.3949	0.1293	0.1302	0.4863	0.1118
	ESN	0.1152	0.6471	0.0946	0.1057	0.6616	0.0895
	DPLS	0.1346	0.5185	0.1099	0.0955	0.7237	0.0779
	DVBPCA	0.1145	0.6514	0.0906	0.0796	0.8080	0.0676
	MW-DVBPCA	0.1103	0.6765	0.0858	0.0548	0.9090	0.0425
٢	PLS	0.1724	0.2094	0.1400	0.1320	0.4722	0.1090
	VBPCA	0.1648	0.2777	0.1361	0.1289	0.4964	0.1065
	LSTM	0.1352	0.5137	0.1094	0.1292	0.4943	0.1015
	ESN	0.1230	0.5978	0.1007	0.0823	0.7948	0.0678
	DPLS	0.1156	0.6448	0.0934	0.0790	0.8107	0.0669
	DVBPCA	0.1119	0.6670	0.0859	0.0752	0.8285	0.0605
	MW-DVBPCA	0.1056	0.7036	0.0835	0.0531	0.9146	0.0403

Table 3. Quantitative estimation results of various models for the concentrations of CO<sub>2</sub> and H<sub>2</sub>.

① The time delays are not considered, ② the time delays are considered.

Furthermore, to check whether the MW-DVBPCA's performance is significantly different from that of other models, the Wilcoxon test is employed for statistical testing [35]. The Wilcoxon test is a non-parametric testing method which is used to examine whether there is significant difference in the median values of the squared estimation errors obtain by the two virtual sensors. In Wilcoxon's test, the likelihood that the corresponding hypothesis will be accepted is measured by calculating the *p*-value. The smaller the value of *p*-value, the lower the probability that the corresponding hypothesis will be accepted. Typically, the hypothesis should be rejected if the *p*-value is less than the given significance level  $\eta$ , the hypothesis should be rejected; that is, statistically the median values of two virtual sensors are different.

The Wilcoxon test results are given in Table 4, where  $\overline{E}_{PLS}^2$ ,  $\overline{E}_{VBPCA}^2$ ,  $\overline{E}_{LSTM}^2$ ,  $\overline{E}_{ESN}^2$ ,  $\overline{E}_{DPLS}^2$ ,  $\overline{E}_{DVBPCA}^2$ , and  $\overline{E}_{MW-DVBPCA}^2$  mean the median values of squared estimated errors obtained by the PLS, VBPCA, LSTM, ESN, DPLS, DVBPCA, and MW-DVBPCA, respectively. Additionally, set the significance level  $\eta$  at 5%. As shown in Table 4, all hypothesized *p*-values are far below  $\eta$ . Hence, all hypotheses are rejected. In other words, there is statistical significance in comparing the MW-DVBPCA with other virtual sensors in the hydrogen production process.

KV	Hypothesis	<i>p</i> -Value	Decision
	$\mathcal{H}_0: \overline{E}^2_{MW-DVBPCA} = \overline{E}^2_{PLS}$	$2.79\times 10^{-70}$	Reject $\mathcal{H}_0$
CH concontration	$\mathcal{H}_0: \overline{E}^2_{MW-DVBPCA} = \overline{E}^2_{VBPCA}$	$1.48  imes 10^{-71}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}_{MW-DVBPCA}^2 = \overline{E}_{LSTM}^2$	$1.15 imes10^{-81}$	Reject $\mathcal{H}_0$
CI14 concentration	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{ESN}}$	$5.39\times10^{-42}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{MW-DVBPCA} = \overline{E}^2_{DPLS}$	$6.32  imes 10^{-30}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{\text{MW-DVBPCA}} = \overline{E}^2_{\text{DVBPCA}}$	$1.03\times10^{-56}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{MW-DVBPCA} = \overline{E}^2_{PLS}$	$7.67\times10^{-45}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{MW-DVBPCA} = \overline{E}^2_{VBPCA}$	$3.15\times10^{-44}$	Reject $\mathcal{H}_0$
CO concentration	$\mathcal{H}_0: \overline{E}_{MW-DVBPCA}^2 = \overline{E}_{LSTM}^2$	$8.1 imes10^{-3}$	Reject $\mathcal{H}_0$
Concentration	$\mathcal{H}_0:\overline{E}^2_{MW-DVBPCA}=\overline{E}^2_{ESN}$	$1.43\times10^{-18}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{\mathrm{MW-DVBPCA}} = \overline{E}^2_{\mathrm{DPLS}}$	$2.3 imes10^{-3}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{\text{MW-DVBPCA}} = \overline{E}^2_{\text{DVBPCA}}$	$2.58 imes10^{-11}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{\mathrm{MW-DVBPCA}} = \overline{E}^2_{\mathrm{PLS}}$	$3.00\times10^{-35}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{VBPCA}}$	$1.69\times10^{-36}$	Reject $\mathcal{H}_0$
CO <sub>2</sub> concentration	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{LSTM}}$	$8.67 imes10^{-15}$	Reject $\mathcal{H}_0$
eo <sub>2</sub> concentration	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{ESN}}$	$1.19\times 10^{-5}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{DPLS}}$	$9.95  imes 10^{-4}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{\mathrm{MW-DVBPCA}} = \overline{E}^2_{\mathrm{DVBPCA}}$	$2.78  imes 10^{-4}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW} ext{-}\mathrm{DVBPCA}}=\overline{E}^2_{\mathrm{PLS}}$	$6.52  imes 10^{-74}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0:\overline{E}^2_{ ext{MW-DVBPCA}}=\overline{E}^2_{ ext{VBPCA}}$	$4.60 imes10^{-69}$	Reject $\mathcal{H}_0$
$H_2$ concentration	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{LSTM}}$	$3.04 imes10^{-60}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0:\overline{E}^2_{\mathrm{MW-DVBPCA}}=\overline{E}^2_{\mathrm{ESN}}$	$2.97\times10^{-29}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}^2_{MW-DVBPCA} = \overline{E}^2_{DPLS}$	$4.25\times10^{-30}$	Reject $\mathcal{H}_0$
	$\mathcal{H}_0: \overline{E}_{MW-DVBPCA}^2 = \overline{E}_{DVBPCA}^2$	$3.19 imes10^{-50}$	Reject $\mathcal{H}_0$

Table 4. Results of the Wilcoxon test.

# 4.6. Computational Efficiency Analysis

Since this article is concerned with real-time estimation, examining the runtime of the model is desirable. The offline and online computational efficiency of the virtual sensors is evaluated using the average over 10 independent simulations, including the CPU time consumed offline for parameter optimization (CPT<sup>opt</sup>) and the CPU time consumed online (CPT<sup>online</sup>). All experiments were computed on a Core i5 (2.90 GHz  $\times$  2) with 8 GB RAM, Windows 10 and R2021a version of MATLAB.

Table 5 lists the time taken by each virtual sensor on parameter determination. As can be observed, the CPT<sup>opt</sup> index for the DVBPCA model is much smaller than that for the LSTM, due to its more concise structure. The CPT<sup>opt</sup> indices for other dynamic global models are almost the same as those for the DVBPCA, but other dynamic global models have lower accuracy than the DVBPCA given in Table 2 and Table 3. Note that the CPT<sup>opt</sup> index for the MW-DVBPCA is much larger than that for the DVBPCA, which is because

the MW-DVBPCA needs to rebuild the model each time it predicts a new valid sample. Fortunately, the parameter determination processes are carried out offline. In other words, this process hardly affects the online calculative efficiency of the MW-DVBPCA. The last column of Table 5 illustrates the online computation time of MW-DVBPCA. Consequently, the online computational efficiency of the developed models is not an issue. In practice, the CPT<sup>opt</sup> indices for all virtual sensors are less than 0.1 s/sample, significantly faster than the minimum sampling period for the KVs in the hydrogen production process. The results show that all data-driven virtual sensors meet the time requirements for real-time estimation, including the developed DVBPCA and MW-DVBPCA.

	CPT <sup>opt</sup>	CPT <sup>online</sup>
PLS	3.44	<0.01
VBPCA	11.49	< 0.01
LSTM	3893.01	0.04
ESN	264.38	< 0.01
DPLS	255.83	< 0.01
DVBPCA	210.16	< 0.01
MW-DVBPCA	1877.91	0.06

Table 5. Comparison of computational efficiency of various models (seconds).

### 5. Conclusions and Outlook

Considering the complicated properties of the hydrogen production process, we develop virtual sensors for the KVs of the hydrogen production process in this article. The MW-DVBPCA is developed to model complicated properties of the hydrogen production process, including dynamics, time variations, and transportation delays. To be specific, the FIR paradigm and MW technique are employed to extract process dynamics and to deal with time variations, respectively. The time delays are determined automatically by the DE. A comparative study of developed virtual sensors and other state-of-the-art virtual sensors is carried out. Meanwhile, the performance of MW-DVBPCA is demonstrated by the real-life natural gas steam reforming hydrogen production process.

From the industrial point of view, the online estimations of the KVs of the hydrogen production process need further research. Some future works are given to further improve the estimation performance of the KVs.

- Robust methods. The probabilistic model in this article is based on the traditional Gaussian distribution assumption, which is susceptible to outliers. Therefore, the training set must be cleaned to remove outliers. However, some outliers are indistinctive and challenging to detect and remove. To this end, finding a probability distribution insensitive to the noise and outliers can help improve the generalization performance of predictive models. Typically, Student's t distribution with heavier tails is a candidate choice. As a result, designing a robust virtual sensor based on Student's t distribution is worth investigating.
- Data-driven approaches fused with process knowledge. In fact, the states (or the hidden variables) of the system are influenced by variables characterizing materials and energies fed into the process. Conventional virtual sensors take all observed variables as inputs and the KVs as outputs, which makes it difficult to describe the true causality between variables of the hydrogen production process, weakening the interpretability and generalization abilities. A causal virtual sensor can better reflect the process mechanism and thus estimate the KVs more accurately. Therefore, equipping the MW-DVBPCA model with the causality of process variables of the hydrogen production process is desirable.

**Author Contributions:** Software, Y.Y. and Y.X.; Supervision, C.W. and W.S.; Writing—original draft, Y.Y. and Z.Z.; Writing—review & editing, Y.Y., C.W. and W.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the National Natural Science Foundation of China under Grants 62173344 and 62103364.

**Data Availability Statement:** Data are available for reasonable reasons by contacting the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

#### References

- Lee, C.; Li, S.; Chen, C.; Huang, Y.; Wang, Y. Real-time microscopic monitoring of flow, voltage and current in the proton exchange membrane water electrolyzer. *Sensors* 2018, *18*, 867. [CrossRef] [PubMed]
- Ji, M.; Wang, J. Review and comparison of various hydrogen production methods based on costs and life cycle impact assessment indicators. *Int. J. Hydrogen Energy* 2021, 46, 38612–38635. [CrossRef]
- 3. Chen, Y. Technical progress and development trend of hydrogen production from natural gas. Coal Chem. Ind. 2020, 43, 130–133.
- Yao, L.; Ge, Z. Deep learning of semisupervised process data with hierarchical extreme learning machine and soft sensor application. *IEEE Trans. Ind. Electron.* 2017, 65, 1490–1498. [CrossRef]
- Cao, Q.; Chen, S.; Zhang, D.; Xiang, W. Research on soft sensing modeling method of gas turbine's difficult-to-measure parameters. J. Mech. Sci. Technol. 2022, 36, 4269–4277. [CrossRef]
- 6. Dogan, A.; Birant, D. Machine learning and data mining in manufacturing. *Expert Syst. Appl.* 2021, 166, 114060. [CrossRef]
- 7. Suthar, K.; Shah, D.; Wang, J.; He, Q.P. Next-generation virtual metrology for semiconductor manufacturing: A feature-based framework. *Comput. Chem. Eng.* 2019, 127, 140–149. [CrossRef]
- Liu, C.; Wang, K.; Wang, Y.; Xie, S.; Yang, C. Deep nonlinear dynamic feature extraction for quality prediction based on spatiotemporal neighborhood preserving SAE. *IEEE Trans. Instrum. Meas.* 2021, 70, 1–10. [CrossRef]
- 9. Sadeghian, A.; Jan, N.M.; Wu, O.; Huang, B. Robust probabilistic principal component regression with switching mixture Gaussian noise for soft sensing. *Chemom. Intell. Lab. Syst.* 2022, 222, 104491. [CrossRef]
- Shao, W.; Ge, Z.; Song, Z. Semisupervised Bayesian Gaussian mixture models for non-Gaussian soft sensor. *IEEE Trans. Cybern.* 2019, *51*, 3455–3468. [CrossRef]
- 11. Yang, F.; Liang, Q.; Zhao, J.; Qiao, R. Hydrogen production from ammonia pyrolysis reforming and prediction model of ammonia decomposition rate. *J. Ordnance Equip. Eng.* **2022**, *43*, 277–280,285.
- 12. Huang, X.; Zhao, B.; Zhang, H.; Zhang, R.; Wang, B.; Liu, H. Parameters research for hydrogen production of methane steam reforming under concentrated radiation. *Chem. Eng. Oil Gas.* **2021**, *50*, 58–65.
- Zhou, H.; Ma, Y.; Wang, K.; Li, H.; Meng, W.; Xie, J.; Li, G.; Zhang, D.; Wang, D.; Zhao, Y. Optimization and analysis of coal-to-methanol process by integrating chemical looping air separation and hydrogen technology. *Chem. Ind. Eng. Prog.* 2022, 41, 5332–5341.
- 14. Yin, S.; Rodriguez-Andina, J.J.; Jiang, Y. Real-Time Monitoring and Control of Industrial Cyberphysical Systems: With Integrated Plant-Wide Monitoring and Control Framework. *IEEE Ind. Electron. Mag.* **2019**, *13*, 38–47. [CrossRef]
- 15. Ye, Y.; Ren, J.; Wu, X.; Ou, G.; Jin, H. Data-driven soft-sensor modelling for air cooler system pH values based on a fast search pruned-extreme learning machine. *Asia-Pac. J. Chem. Eng.* **2017**, *12*, 186–195. [CrossRef]
- 16. Yuan, X.; Qi, S.; Shardt, Y.A.; Wang, Y.; Yang, C.; Gui, W. Soft sensor model for dynamic processes based on multichannel convolutional neural network. *Chemom. Intell. Lab. Syst.* **2020**, 203, 104050. [CrossRef]
- 17. Wei, C.; Song, Z. Real-time forecasting of subsurface inclusion defects for continuous casting slabs: A data-driven comparative study. *Sensors* **2023**, *23*, 5415. [CrossRef]
- 18. Tong, W.; Wei, B. Soft sensing modeling of methane content in conversion reaction based on GA-BP. *Automat. Instrum.* **2016**, *31*, 7–10.
- Zamaniyan, A.; Joda, F.; Behroozsarand, A.; Ebrahimi, H. Application of artificial neural networks (ANN) for modeling of industrial hydrogen plant. *Int. J. Hydrogen Energy* 2013, 38, 6289–6297. [CrossRef]
- Ögren, Y.; Tóth, P.; Garami A.; Sepman A.; Wiinikka H. Development of a vision-based soft sensor for estimating equivalence ratio and major species concentration in entrained flow biomass gasification reactors. *Appl. Energy* 2018, 226, 450–460. [CrossRef]
- 21. Fang, X.; Ding, Z.; Shu, X. Hydrogen yield prediction model of hydrogen production from low rank coal based on support vector machine optimized by genetic algorithm. *J. China Coal Soc.* **2010**, *35*, 205–209.
- Zhao, Y.; Song, H.; Guo, Y.; Zhao, L.; Sun, H. Super short term combined power prediction for wind power hydrogen production. Energy Rep. 2022, 8, 1387–1395. [CrossRef]
- Koc, R.; Kazantzis, N.K.; Ma, Y.H. A process dynamic modeling and control framework for performance assessment of Pd/alloybased membrane reactors used in hydrogen production. *Int. J. Hydrogen Energy* 2011, 36, 4934–4951. [CrossRef]
- 24. Pan, F.; Cheng, X.; Wu, X.; Wang, X.; Gong, J. Thermodynamic design and performance calculation of the thermochemical reformers. *Energies* **2019**, *12*, 3693. [CrossRef]

- Sharma, Y.C.; Kumar, A.; Prasad, R.; Upadhyay, S.N. Ethanol steam reforming for hydrogen production: Latest and effective catalyst modification strategies to minimize carbonaceous deactivation. *Renew. Sustain. Energy Rev.* 2017, 74, 89–103. [CrossRef]
- Blei, D.M.; Kucukelbir, A.; McAuliffe, J.D. Variational inference: A review for statisticians. J. Am. Stat. Assoc. 2017, 112, 859–877.
   [CrossRef]
- 27. Oh, H.S.; Kim, D.G. Bayesian principal component analysis with mixture priors. J. Korean Stat. Soc. 2010, 39, 387–396. [CrossRef]
- 28. Liu, Y.; Hu, N.; Wang, H.; Li, P. Soft chemical analyzer development using adaptive least-squares support vector regression with selective pruning and variable moving window size. *Ind. Eng. Chem. Res.* **2009**, *48*, 5731–5741. [CrossRef]
- 29. Yao, L.; Ge, Z. Refining data-driven soft sensor modeling framework with variable time reconstruction. *J. Process Control* 2020, *87*, 91–107. [CrossRef]
- 30. Geladi, P.; Kowalski, B.R. Partial least-squares regression: A tutorial. Anal. Chim. Acta 1986, 185, 1–17. [CrossRef]
- Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. D* 2020, 404, 132306. [CrossRef]
- 32. Chen, Q.; Zhang, A.; Huang, T.; He, Q.; Song, Y. Imbalanced dataset-based echo state networks for anomaly detection. *Neural Comput. Appl.* **2020**, *32*, 3685–3694. [CrossRef]
- 33. Ricker, N.L. The use of biased least-squares estimators for parameters in discrete-time pulse-response models. *Ind. Eng. Chem. Res.* **1988**, *27*, 343–350. [CrossRef]
- Shao, W.; Ge, Z.; Yao, L.; Song, Z. Bayesian nonlinear Gaussian mixture regression and its application to virtual sensing for multimode industrial processes. *IEEE Trans. Automat. Sci. Eng.* 2019, 17, 871–885. [CrossRef]
- Rodríguez-Fdez, I.; Canosa, A.; Mucientes, M.; Bugarín, A. STAC: A web platform for the comparison of algorithms using statistical tests. In Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Istanbul, Turkey, 2–5 August 2015.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.