Incremental Fast Relevance Vector Regression Model Based Multi-Pollutant Emission Prediction of Biomass Cogeneration Systems

Xiuli Wang[a], Zhifei Sun[a], Defeng He[a,*], Shaomin Wu[b], Lianna Zhao[c]

[a]*College of Information Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang 310023, China*
[b]*Kent Business School, University of Kent, Canterbury, Kent CT2 7FS, United Kingdom*
[c]*Information School, The University of Sheffield, Sheffield, South Yorkshire S10 2AH, United Kingdom*

Abstract

Exact and trusty prediction of pollutant emissions is pivotal for optimal combustion control in biomass cogeneration systems, which possess multiple variables, high-volume data streams, and dynamic characteristics. Aiming at the multivariate dynamic systems, this paper extends a classical fast relevance vector regression (FRVR) algorithm into a multivariate form to accomplish synchronous multi-pollutant prediction. Meanwhile, a flexible and effective online training strategy is proposed to solve the problems of low accuracy of multi-step prediction and lack of dynamic updating capability. First, the given dataset is divided utilizing the *k*-means clustering method to enhance the clustering of similar features and expedite the prediction process. Then, the classical FRVR algorithm is extended into a multiple-output form, enabling the simultaneous prediction of multiple pollutant emissions. Moreover, the incremental learning method is introduced into the proposed multivariate FRVR model to improve its dynamic performance and online learning ability. Finally, the proposed method's effectiveness is verified through a biomass cogeneration systems case. Experimental findings fully illustrate that the proposed method provides the lower RMSE and MAE while runtime decreases by 50% and R^2 reaches 96%. The proposed method significantly outperforms others, showing excellent potential in the pollutant prediction field.

Keywords: Extended fast relevance vector regression algorithm, incremental learning method, *k*-means clustering method, pollutant emission prediction.

1. Introduction

Compared with fossil fuels, biomass energy has gained increasing attention as renewable energy, which has a vital effect in reconstructing energy distributions (Valente et al. (2020)), optimizing the ecological environment (Xu et al. (2023)), and promoting industrial development (Li et al. (2024); Nunes et al. (2014)) . These advantages have directly contributed to the fact that biomass cogeneration systems attract broad attention in many important fields and industrial applications (Qiu (2013)). However, biomass combustion process produces large amounts of nitrogen oxide (NO_x) and sulfur dioxide (SO_2), causing ozonosphere destruction and global warming, which seriously damages the ecosystem and endangers human health (Jin et al. (2022)). Therefore, mitigating pollutant emissions has emerged as a pivotal issue for biomass cogeneration systems. Flue gas purification and combustion process monitoring are being developed to decrease greenhouse gas production and pollutant emissions (Chen et al. (2023)). However, regardless of implementing reduction strategies for emissions, exact and trusty prediction of pollutant emissions is a basic premise (Han et al. (2022)).

There are three categories of research on pollutant emissions concentration prediction. The first category involves direct monitoring of pollutant emissions using continuous emission monitoring systems (CEMS) (Nazari et al. (2010)). Nevertheless, the CEMS is difficult to ensure economic and continuous measurement due to high-cost maintenance and frequent offline calibration. (Yang et al. (2020b)). The second category is applying computational fluid dynamics (CFD) analysis models to predict pollutant emissions (Kang et al. (2017); Park et al. (2017)). However, the pollutants emission from biomass combustion is a dynamic and time-varying process that contains many complicated

*Corresponding author
Email address: hdfzj@zjut.edu.cn (Defeng He)

physical and chemical reactions (Li et al. (2022); Chen et al. (2023)). Establishing a precise model to describe this process is challenging. And even if it can be established, the model parameters are not easy to estimate. The third category is utilizing data-driven methods to learn regularities from large-scale data characteristics. It is widely applicable across multiple domains due to adaptation to complex patterns and strong generalization ability (Li et al. (2021)). Data-driven methods mainly use artificial neural networks (ANN), support vector machine (SVM), relevance vector regression (RVR), and their variations to predict pollutant emissions. ANNs have high non-linear approximation capabilities and do not require accurate modelling of the combustion process (Chen et al. (2024)). The extreme learning machine (ELM) was built to accurately predict NO_x emissions (Tan et al. (2016)), achieving a prediction model with a mean absolute error (MAE) of only 0.92%. However, the training process of ANNs requires large amounts of high-quality data, which is highly challenging to acquire in the actual industrial process. The SVM algorithm is also extensively applied in pollutant emission prediction (Lv et al. (2013)). It was applied to build a non-linear relationship between operating variables and nitrogen oxides to predict boiler nitrogen oxide emission (Zhou et al. (2012)). The least squares support vector machine (LSSVM) was established to analyze the flame image feature for pollutant emissions prediction (Li et al. (2015)). However, these algorithms can only provide point estimation and have strict restrictions on kernel functions. The RVR algorithm, similar to the SVM algorithm, has been proposed with good learning ability, easy training process, and probability distribution prediction.

The RVR algorithm is a sparse probabilistic regression model based on a Bayesian framework (Tipping (2001)). This algorithm relieves the computational effort of non-linear regression problems via kernel functions and provides good prediction accuracy and sparsity. It is widely used in industrial fields, such as the prediction of remaining useful life (RUL) of equipment, fault diagnosis, and so on (Liu et al. (2015); Wang et al. (2021)). An RVR algorithm was proposed for predicting the remaining useful life of rolling bearings (Guo & He (2022)). The RVR algorithm was also utilized to find relevant vectors for representing the battery capacity fade, where an empirical degradation model was developed for the conditional parameters (Jia et al. (2021)). However, these RVR algorithms are limited to regression from multiple input variables to a single output variable (Tipping & Faul (2003)). While multivariate outputs commonly exist in numerical industrial applications, especially biomass cogeneration systems. To overcome this disadvantage, the RVR algorithm has been extended to multiple output forms. For example, an RVR algorithm was proposed for learning a mapping from image features to state space in a one-to-many mapping (Thayananthan et al. (2006); Thayananthan et al. (2008)). However, this algorithm is only an integration of several different RVRs without realizing the simultaneous output of multivariate variables under an RVR framework. It is still a single variable RVR algorithm in essence. A multivariate RVR (MRVR) algorithm was extended based on a matrix Gaussian distribution to achieve multi-step prediction of capacitor performance degradation (Wang et al. (2021)). In this paper, the MFRVR algorithm is extended to a multivariate form, which improves the computational time and solves multivariate regression compared to the classical RVR algorithm.

An ideal model for the dynamic prediction of pollutant emission should describe the non-linear, delayed, and multivariate coupled nature of the combustion process. However, most existing data-driven algorithms only apply to the offline case and cannot achieve dynamic updates (Yang et al. (2020a)). A modified long-short term memory (LSTM) is designed to construct a dynamic prediction model for SCR systems outlet NO_x emissions (Xie et al. (2020)). However, the model suffered from error accumulation, which resulted in poor prediction accuracy. Although the above methods are outstanding in non-linear modelling, they cannot achieve satisfactory results regarding dynamic characteristics (Chen & Huang (2023)).

Motivated by the above discussions, a dynamic multivariate prediction framework is established by extending the FRVR algorithm into a multivariate form and integrating the incremental learning method. Firstly, a dataset is partitioned into a fixed number of smaller datasets by applying the k-means method since the FRVR algorithm is suitable for small samples, but the biomass systems are lengthy processes. Secondly, the FRVR algorithm is extended into a multi-output form to predict multiple pollutant emissions from biomass cogeneration systems. Finally, an incremental learning method is added to the MFRVR algorithm for online learning and dynamic training. The contributions and novelty of this paper are as follows.

1) The classical FRVR algorithm is extended into a multi-output form by introducing a matrix Gaussian distribution of weights. Compared with the classical RVR algorithm and its deformations, this paper can simultaneously achieve multipollutant

emissions prediction with a high prediction accuracy.

2) An incremental learning method is incorporated into the MFRVR algorithm for dynamic updating pollutant models and real-time monitoring of pollutant prediction in biomass cogeneration systems. Compared to similar research algorithms, the proposed method has more precise online prediction capability and low computational complexity.

The rest of the paper is arranged as follows. First, Section 2 proposes an MFRVR algorithm by extending the FRVR algorithm into a multi-output format. Then, Section 3 provides a detailed introduction to the k-means clustering method and the incremental learning method for establishing an IMFRVR model. Subsequently, in Section 4, the accuracy and complexity of the IMFRVR model are extensively compared with other existing methods through the biomass cogeneration systems case. At last, further summarizes the paper, and some related thoughts for future works are discussed in Section 5.

2. Multivariable Fast Relevance Vector Regression

2.1. Regression model establishment

Given a collected input-output dataset $\{\boldsymbol{x}_i, \boldsymbol{t}_i\}_{i=1}^{N}$, where $\boldsymbol{x}_i \in \mathbb{R}^{U\times 1}$ is sample input vector, $\boldsymbol{t}_i \in \mathbb{R}^{1\times V}$ is corresponding target vector, N represents sample size, V is the dimensions of the multivariate space, and U denotes the input variables size. A dynamic model be formulated via MFRVR as

$$\boldsymbol{t}_i = y(\boldsymbol{x}_i, \boldsymbol{W}) + \boldsymbol{\varepsilon}_i \tag{1}$$

where $\boldsymbol{W} = [\boldsymbol{w}_0, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_N]^{\mathrm{T}} \in \mathbb{R}^{(N+1)\times V}$ denotes the weight matrix; $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{1\times V}$ is assumed to be a zero mean Gaussian distributed random error vector with a covariance matrix $\boldsymbol{D} \in \mathbb{R}^{V\times V}$. The $\boldsymbol{D}$ can be decomposed as $\boldsymbol{D} = \boldsymbol{RBR}^{\mathrm{T}}$, where $\boldsymbol{R} \in \mathbb{R}^{V\times V}$; $\boldsymbol{B} = \mathrm{diag}\{\beta_1, \beta_2, \cdots, \beta_v\} \in \mathbb{R}^{V\times V}$, in which diag$(\cdot)$ describes a diagonal matrix, and β_i is the variance of random error vector $\boldsymbol{\varepsilon}_i$.

Eq. (1) can be expressed in matrix format as

$$\boldsymbol{T} = \boldsymbol{\Phi W} + \boldsymbol{E}, \tag{2}$$

where $\boldsymbol{T} = [\boldsymbol{t}_1, \boldsymbol{t}_2, \cdots, \boldsymbol{t}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times V}$ represents a pollutant emission prediction matrix; $\boldsymbol{E} = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_N]^{\mathrm{T}} \in \mathbb{R}^{N\times V}$ is a random error matrix; and $\boldsymbol{\Phi} = [\boldsymbol{\varphi}(\boldsymbol{x}_1), \boldsymbol{\varphi}(\boldsymbol{x}_2), \cdots, \boldsymbol{\varphi}(\boldsymbol{x}_N)]^{\mathrm{T}} \in \mathbb{R}^{N\times(N+1)}$ is a design matrix, in which $\boldsymbol{\varphi}(\boldsymbol{x}_i) = [1, \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_1), \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_2), \cdots, \mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_N)] \in \mathbb{R}^{1\times(N+1)}$ denotes a basis vector, $\mathcal{K}(\boldsymbol{x}_i, \boldsymbol{x}_s) \in \mathbb{R}^{(N+1)\times 1}$ is a Gaussian kernel function between the vector $\boldsymbol{x}_i$ and $\boldsymbol{x}_s (s = 1, \cdots, N)$ in this study.

Due to the neglect of $\boldsymbol{R}$ by Bishop (2006) and the supposed independent Gaussian random error, the probability density function (PDF) of $\boldsymbol{T}$ conditioned on $\boldsymbol{W}$ and $\boldsymbol{B}$ can be given by

$$p(\boldsymbol{T} \mid \boldsymbol{W}, \boldsymbol{B}) = (2\pi)^{-\frac{VN}{2}} |\boldsymbol{B}|^{-\frac{N}{2}} \exp\left(-\frac{1}{2} tr\left(\boldsymbol{B}^{-1}(\boldsymbol{T} - \boldsymbol{\Phi W})^{\mathrm{T}}(\boldsymbol{T} - \boldsymbol{\Phi W})\right)\right) \tag{3}$$

where $tr(\cdot)$ is the trace of the matrix, and $|\cdot|$ expresses the determinant of a square matrix.

As many hyperparameters in the model as training samples, the model might overly learn complex patterns. So, the direct adoption of the maximum likelihood function to estimate the hyperparameters would produce over-fitting (Bishop (2006)). To prevent overfitting of the model (2), a zero mean Gaussian prior distribution with a diagonal covariance matrix $\boldsymbol{A}$ is introduced for $\boldsymbol{W}$, where $\boldsymbol{W}$ is assumed to be related to $\boldsymbol{B}$. Following this, the prior PDF of weight $\boldsymbol{W}$ is determined as

$$p(\boldsymbol{W} \mid \boldsymbol{\alpha}, \boldsymbol{B}) = (2\pi)^{-\frac{V(N+1)}{2}} |\boldsymbol{B}|^{-\frac{N+1}{2}} |\boldsymbol{A}|^{\frac{V}{2}} \exp\left(-\frac{1}{2} tr\left(\boldsymbol{B}^{-1}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{AW}\right)\right) \tag{4}$$

where $\boldsymbol{A}^{-1} = \mathrm{diag}\left[\alpha_0^{-1}, \alpha_1^{-1}, \cdots, \alpha_N^{-1}\right] \in \mathbb{R}^{(N+1)\times(N+1)}$, $\alpha_i^{-1} (i = 0, \cdots, N)$ denotes the precision of weight $\boldsymbol{w}_i$. It means a zero mean Gaussian prior distribution with among rows inverse variances $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \cdots, \alpha_N]^{\mathrm{T}} \in \mathbb{R}^{(N+1)\times 1}$ for $\boldsymbol{W}$. Thus, each weight $\boldsymbol{w}_i$ has an independent hyperparameter associated with it, regulating the strength of the prior thereon. In general, the initial value of precision α_i^{-1} set as a large number to ensure the sparsity of the model. In other words, the prior weight $\boldsymbol{w}_i$ is more centrally distributed around zero.

2.2. Hyperparameter optimization

Based on Bayes' theorem, the connection between the prior PDF and the posterior PDF can be expressed as

$$p(\boldsymbol{W} \mid \boldsymbol{T}, \boldsymbol{\alpha}, \boldsymbol{B}) = \frac{p(\boldsymbol{T} \mid \boldsymbol{W}, \boldsymbol{B}) p(\boldsymbol{W} \mid \boldsymbol{\alpha}, \boldsymbol{B})}{p(\boldsymbol{T} \mid \boldsymbol{\alpha}, \boldsymbol{B})} \tag{5}$$

The posterior PDF of weight $\boldsymbol{W}$ is formulated as follows

$$p(\boldsymbol{W} \mid \boldsymbol{T}, \boldsymbol{\alpha}, \boldsymbol{B}) = (2\pi)^{-\frac{V(N+1)}{2}} |\boldsymbol{B}|^{-\frac{N+1}{2}} |\Sigma|^{-\frac{V}{2}} \times \exp\left(-\frac{1}{2} tr\left(\boldsymbol{B}^{-1}(\boldsymbol{W} - \boldsymbol{M})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{W} - \boldsymbol{M})\right)\right) \tag{6}$$

In Eq. (6), the mean $\boldsymbol{M}$ and the covariance $\boldsymbol{\Sigma}$ are

$$\boldsymbol{M} = \boldsymbol{\Sigma}\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{T} \tag{7}$$

and

$$\boldsymbol{\Sigma} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi} + \boldsymbol{A}\right)^{-1} \tag{8}$$

respectively.

The hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{B}$ can be estimated by Bayes' theorem in the following form

$$p(\boldsymbol{\alpha}, \boldsymbol{B} \mid \boldsymbol{T}) \propto p(\boldsymbol{T} \mid \boldsymbol{\alpha}, \boldsymbol{B}) p(\boldsymbol{\alpha}, \boldsymbol{B}) \tag{9}$$

If the priors of hyperparameters $\boldsymbol{\alpha}$ and $\boldsymbol{B}$ are relatively flat, they can be estimated by maximizing the marginal likelihood function $p(\boldsymbol{T} \mid \boldsymbol{\alpha}, \boldsymbol{B})$ in the evidence framework. $p(\boldsymbol{T} \mid \boldsymbol{\alpha}, \boldsymbol{B})$ can be derived by integrating over $\boldsymbol{W}$ as

$$\begin{aligned} p(\boldsymbol{T} \mid \boldsymbol{\alpha}, \boldsymbol{B}) &= \int p(\boldsymbol{T} \mid \boldsymbol{W}, \boldsymbol{B}) p(\boldsymbol{W} \mid \boldsymbol{\alpha}, \boldsymbol{B}) d\boldsymbol{W} \\ &= (2\pi)^{-\frac{VN}{2}} |\boldsymbol{B}|^{-\frac{N}{2}} \left|\boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\right|^{-\frac{V}{2}} \\ &\times \int \exp\left(-\frac{1}{2} tr\left(\boldsymbol{B}^{-1}\boldsymbol{T}^{\mathrm{T}}\left(\boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\right)^{-1}\boldsymbol{T}\right) d\boldsymbol{W}, \end{aligned} \tag{10}$$

The In-marginal likelihood function of Eq. (10) is represented in the following manner

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B}) &= \ln p(\boldsymbol{T} \mid \boldsymbol{\alpha}, \boldsymbol{B}) \\ &= -\frac{1}{2}\left[VN\ln(2\pi) + N\ln|\boldsymbol{B}| + V\ln|\boldsymbol{C}|\right. \\ &\left. + tr\left(\boldsymbol{B}^{-1}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{C}^{-1}\boldsymbol{T}\right)\right], \end{aligned} \tag{11}$$

Let $\boldsymbol{C} = \boldsymbol{I} + \boldsymbol{\Phi}\boldsymbol{A}^{-1}\boldsymbol{\Phi}^{\mathrm{T}} \in \mathbb{R}^{N\times N}$. Considering the dependence of $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B})$ on a single hyperparameter $\alpha_i^{-1}, \boldsymbol{C}$ is decomposed as

$$\boldsymbol{C} = \boldsymbol{I} + \sum_{m\neq i} \alpha_m^{-1}\boldsymbol{\varphi}_m\boldsymbol{\varphi}_m^{\mathrm{T}} + \alpha_i^{-1}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathrm{T}} = \boldsymbol{C}_{-i} + \alpha_i^{-1}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathrm{T}}, \tag{12}$$

where $\boldsymbol{C}_{-i}$ denotes the matrix $\boldsymbol{C}$ with the contribution of basis vector $\boldsymbol{\varphi}_i$ removed and substituting Eq. (12) into Eq. (11), there is

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B}) &= -\frac{1}{2}\left[VN\ln(2\pi) + N\ln|\boldsymbol{B}| + V\ln|\boldsymbol{C}_{-i}|\right. \\ &\left. + tr\left(\boldsymbol{B}^{-1}\boldsymbol{T}^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{T}\right)\right] - \frac{1}{2}\left[-V\ln\alpha_i + V\ln\left(\alpha_i + \boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_i\right)\right. \\ &\left. + tr\left(\boldsymbol{B}^{-1}\boldsymbol{T}^{\mathrm{T}}\frac{\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_i\boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}}{\alpha_i + \boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_i}\boldsymbol{T}\right)\right], \end{aligned} \tag{13}$$

Eq. (11) can be further written as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B}) &= \mathcal{L}(\boldsymbol{\alpha}_{-i}, \boldsymbol{B}) \\ &+ \frac{1}{2}\left(V\ln\boldsymbol{\alpha}_{-i} - V\ln(\boldsymbol{\alpha}_{-i} + s_i) + \frac{tr\left(\boldsymbol{B}^{-1}\boldsymbol{q}_i^T\boldsymbol{q}_i\right)}{\alpha_i + s_i}\right) \\ &= \mathcal{L}(\boldsymbol{\alpha}_{-i}, \boldsymbol{B}) + \ell(\alpha_i, \boldsymbol{B}) \end{aligned} \tag{14}$$

where $s_i \overset{\text{def}}{=} \boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{\varphi}_i$ is the sparsity factor, which measures the extent that basis vector $\boldsymbol{\varphi}_i$ overlaps those already present in MFRVR model. $\boldsymbol{q}_i \overset{\text{def}}{=} \boldsymbol{\varphi}_i^{\mathrm{T}}\boldsymbol{C}_{-i}^{-1}\boldsymbol{T}$ represents the quality factor, which measures the error of the model with that the basis vector $\boldsymbol{\varphi}_i$ excluded. Generally, $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B})$ has now been divided into $\mathcal{L}(\boldsymbol{\alpha}_{-i}, \boldsymbol{B})$, the marginal likelihood with $\boldsymbol{\varphi}_i$ excluded; and $\ell(\alpha_i, \boldsymbol{B})$, where terms in α_i are now conveniently isolated. This decomposition of the ln marginal likelihood function can greatly simplify the hyperparameter-solving process, thus reducing the computational load of the model and accelerating runtime speed.

$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B})$ has a unique maximum concerning α_i by taking the derivative of the hyperparameter α_i in Eq. (14) and making it zero. And there is

$$\alpha_i = \begin{cases} \dfrac{s_i^2}{\frac{tr\left(\boldsymbol{B}^{-1}\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i\right)}{V} - s_i}, & \text{if } \dfrac{tr\left(\boldsymbol{B}^{-1}\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i\right)}{V} > s_i \\ \infty, & \text{if } \dfrac{tr\left(\boldsymbol{B}^{-1}\boldsymbol{q}_i^{\mathrm{T}}\boldsymbol{q}_i\right)}{V} \leq s_i \end{cases} \tag{15}$$

Similarly, $\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{B})$ has a unique maximum concerning $\boldsymbol{B}$ by taking the derivative of the hyperparameter $\boldsymbol{B}$ in Eq. (11) and making it a zero matrix. There is

$$\boldsymbol{B} = \frac{\boldsymbol{T}^{\mathrm{T}}(\boldsymbol{T} - \boldsymbol{\Phi}\boldsymbol{M})}{N} \tag{16}$$

2.3. Prediction based on MFRVR

Given a new sample $\boldsymbol{x}_*$ the procedure for predicting the mean $\boldsymbol{y}_*$ and a variance matrix $\boldsymbol{\sigma}_*^2$ is obtained as follows:

Firstly, the initial values of $\boldsymbol{\alpha}$ and $\boldsymbol{B}$ are chosen and substituted into Eq. (7) and Eq. (8) to compute $\boldsymbol{M}$ and $\boldsymbol{\Sigma}$. Then, Eq. (15) and Eq. (16) are utilized to estimate the hyperparameters. Subsequently, Eq. (7) and Eq. (8) are applied again to re-estimate the posterior mean and covariance by substituting them back into Eq. (15) and Eq. (16). These estimation steps iterate gradually until convergence.

Secondly, the estimated hyperparameter values $\hat{\boldsymbol{\alpha}}$ and $\hat{\boldsymbol{B}}$ are substituted into Eq. (3) and Eq. (5).

Finally, the predictive distribution of $\boldsymbol{t}_*$ is joint normal distributed as

$$p\left(\boldsymbol{t}_* \mid \boldsymbol{T}, \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{B}}\right) = \mathcal{N}\left(\boldsymbol{t}_* \mid \boldsymbol{y}_*, \boldsymbol{\sigma}_*^2\right) \tag{17}$$

with

$$\boldsymbol{y}_* = \boldsymbol{\Phi}(\boldsymbol{x}_*)^{\mathrm{T}} \boldsymbol{M} \tag{18}$$

$$\boldsymbol{\sigma}_*^2 = \hat{\boldsymbol{B}}\left(1 + \boldsymbol{\Phi}(\boldsymbol{x}_*)^{\mathrm{T}} \Sigma \boldsymbol{\Phi}(\boldsymbol{x}_*)\right) \tag{19}$$

The prediction variance $\boldsymbol{\sigma}_*^2$ can be regarded as the sum of two parts of noise: the estimated random error on the training data $\hat{\boldsymbol{B}}$ and random error caused by the weight posterior uncertainty $\hat{\boldsymbol{B}}\boldsymbol{\Phi}(\boldsymbol{x}_*)^{\mathrm{T}} \boldsymbol{\Sigma}\boldsymbol{\Phi}(\boldsymbol{x}_*)$.

3. MFRVR based dynamic prediction

Most algorithms employ a retraining method to achieve dynamic prediction. With the continuous increase of samples, a novel model can be obtained by retraining all samples. However, this training method can bring a heavy computational burden, resulting in slower update speeds and an inability to adapt to new data on time. Therefore, an extended incremental MFRVR (IMFRVR) prediction model is proposed for the online prediction of pollutant emissions with high prediction accuracy. The framework of IMFRVR based dynamic prediction is shown in Fig.1, which consists of the following four steps.

Step 1: Pre-processing. The complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method is presented to remove the noise of samples. The maximal information coefficient (MIC) method is adopted to reduce the dimensionality of the sample.

Step 2: The k-means clustering method for samples. The k-means clustering method is to divide the data into k clusters, where the data points within the same cluster are as similar as possible, and the differences between different clusters are as significant as possible. First, the elbow method is adopted to determine the optimal k value. It calculates the sum of squared errors (SSE) for different k values. As the k value increases, the SSE rapidly decreases. When the k value reaches the optimal value, the decreasing of SSE slows down, forming an elbow. Then, each data calculates its distance to each cluster centre and assigns it to the cluster whose centre is closest. Each cluster recomputes the mean of all data points in it and sets these means as the new cluster centers. Next, repeat the above steps until the cluster centre no longer changes. Finally, small datasets $\boldsymbol{A}_1 \ldots \boldsymbol{A}_k$ is obtained by the k-means clustering method.

Step 3: Modeling offline training process. The clustered datasets $\boldsymbol{A}_1 \ldots \boldsymbol{A}_k$ trains separately the MFRVR model established in the second section. After completing the training, the relevance vector corresponding to each small dataset can be obtained. And the relevance vector sets of $RV_1 \ldots RV_k$ forms the $\boldsymbol{RV}$ sets.

Step 4: Incremental learning process. Composing the new sample dataset by merging the sample data set $\boldsymbol{B}$ with the $\boldsymbol{RV}$ sets, and incremental training the MFRVR model with it. Then, the prediction value datasets $\boldsymbol{C}$, along with the upper and lower boundaries of the confidence interval, which are generated by the trained MFRVR model.

In this paper, the training algorithm of the MFRVR model adopts the EM optimization algorithm. Its iterative calculations in the E-step and M-step include matrix-vector multiplication, vector inner product, matrix trace, matrix multiplication, and matrix inversion. Among them, the complexities of matrix multiplication and matrix inversion require a large amount of computation, reaching $O\left(N^3\right)$ (N is the dimension of the matrix determined by the number of samples). On the other hand, the prediction process mainly involves calculations of vector inner products and matrix-vector multiplication. Due to the high sparsity of MFRVR, only the multiplications involving non-zero values in $\boldsymbol{W}$ need to be computed, significantly improving computational efficiency. Based on the above analysis, the main factors affecting the computational efficiency of MFRVR training are matrix multiplication and matrix inversion, which are determined by the number of samples. Reducing the size of online samples can effectively improve the efficiency of online training. As a result, a simple and low-computational incremental learning strategy is proposed for the MFRVR model.

Fig. 2 illustrates the incremental learning process of the MFRVR model. First, the MFRVR model is trained offline with the clustered small datasets to obtain the relevant vector sets $\boldsymbol{RV}$; then, the new datasets $\boldsymbol{B}$ and the $\boldsymbol{RV}$ sets are combined into new datasets to train the MFRVR model. Finally, the prediction value datasets $\boldsymbol{C}$, along with the upper and lower boundaries of the confidence interval, are generated by the trained MFRVR model. This process does not require storage of historical data, which significantly reduces memory space usage and space complexity. Meanwhile, it effectively associates new samples with previous learning results, which reduces training time and improves training efficiency.

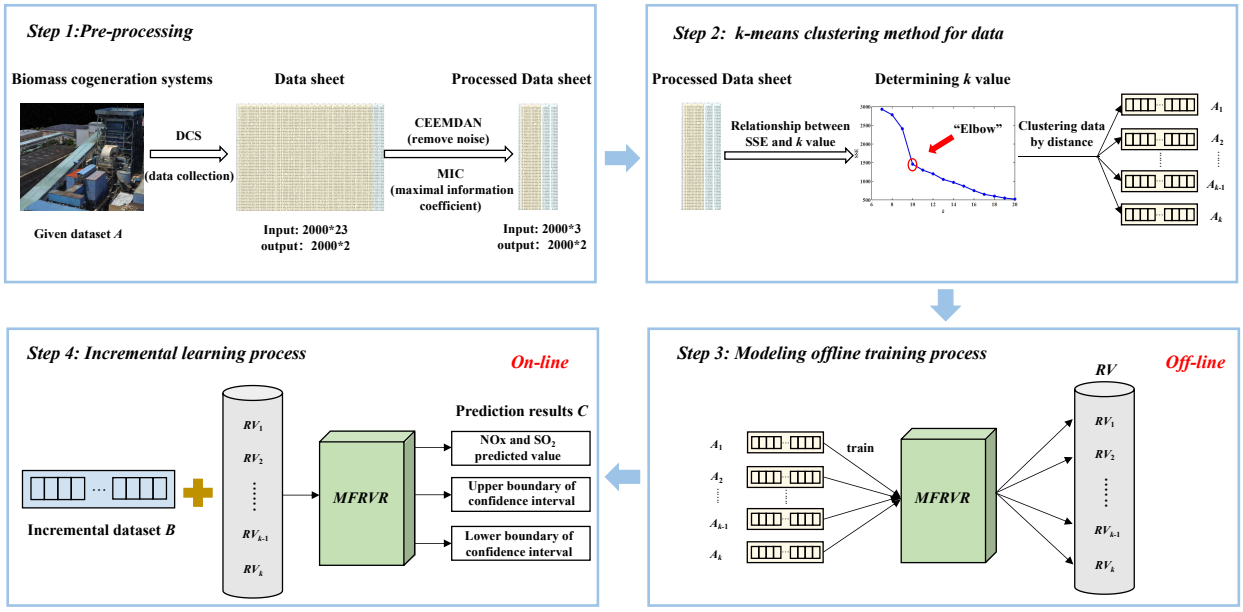

Figure 1: The framework of IMFRVR based dynamic prediction.

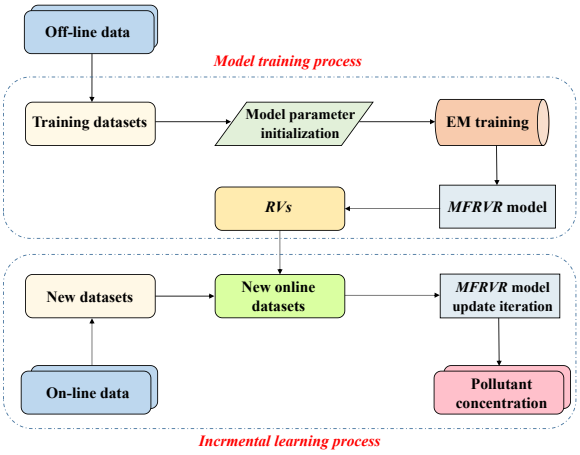

Figure 2: Incremental learning process of the MFRVR model.

4. Experimental results and discussions

4.1. Datasets and experimental setup

The experimental study is conducted on the biomass cogeneration systems developed by Jiaxing New Jies Heat & Power Co. Ltd, Zhejiang Province. As shown in Fig. 3, the biomass cogeneration systems include biomass storage and transportation, biomass combustion, water vapor circulation, steam power generation, provides water vapor and heat, flue gas treatment, and flue gas emission. These steps are interrelated and indispensable. Fig. 4 shows the distribution diagram of some devices of the biomass cogeneration systems, including several key components such as the biomass circulating fluidized bed boiler, primary air fan, secondary air fan, dust collector, and induced draft fan. Generally, biomass cogeneration systems are equipped with a distributed control system (DCS) to collect operational data of the system. Based on the pollutant formation mechanism and engineer recommendations, 23 parameters are selected as input variables shown in Table 1. And the output variables of the data are SO_2 and NO_x emission. Experimental data are collected from the operation of the biomass cogeneration system on 9 August 2021, with a sampling period of 5 s. A total of 17,000 samples are obtained.

In this experiment, 2076 samples are selected, with the training and test samples divided in a ratio of 7:3. Each sample adopts the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) method to remove the noise present in the sample. And the maximal information coefficient (MIC) method is used to reduce the dimensionality of the sample. The correlation heat map for the 23 input variables is shown in Fig. 5. Previous research indicates that a MIC value exceeding 0.4 signifies a strong correlation between the two variables. Therefore, after comparing and evaluating the MIC values between variables, three variables (generator active power, primary air duct flow, and secondary air duct flow) are ultimately selected as input variables for the subsequent model. The experiments set

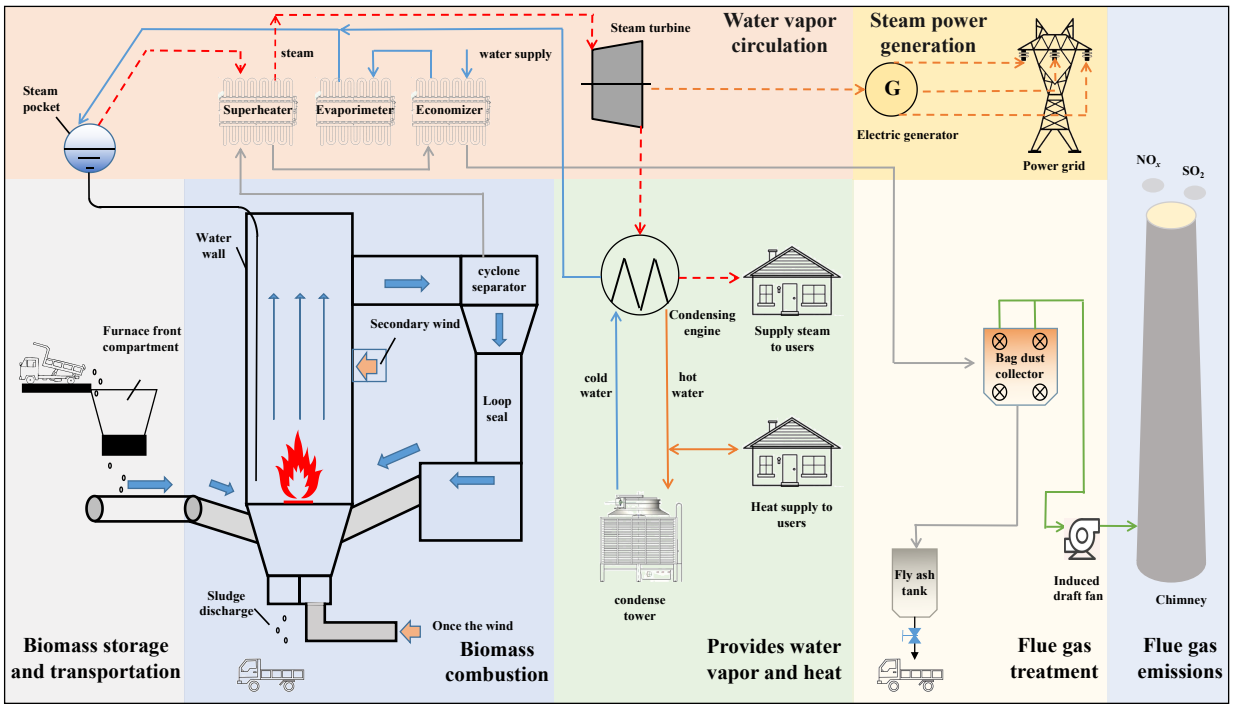

Figure 3: The schematic diagram of the biomass cogeneration systems.

incremental learning to batch mode, dividing the total test set into four incremental sets and one small validation set. Each incremental set contains 200 samples. In the incremental learning phase, the incremental sets are fed into the training prediction model every 20 seconds, for four incremental sets. It can simulate the data flow in the biomass cogeneration process. Finally, a small validation set exploit verifies the effectiveness of the proposed algorithm.

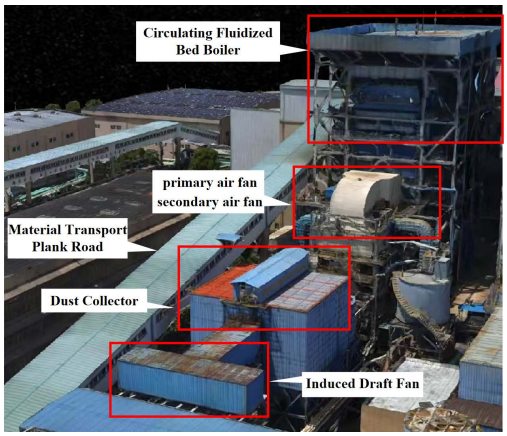

Figure 4: Distribution diagram of some devices of the biomass cogeneration systems.

Table 1: Input variables

No	Input variables
1	Garameter description
2	Instantaneous main steam flow rate
3	Main steam line outlet temperature
4	Primary air mains flow
5	Primary hot air outlet temperature
6	Secondary air mains flow
7	Secondary hot air outlet temperature
8	Smoke exhaust temperature left
9	Smoke exhaust temperature right
10	Combustion chamber sub-boiling temperature1
11	Combustion chamber sub-boiling temperature2
12	Combustion chamber boiling center temperature1
13	Combustion chamber boiling center temperature2
14	Combustion chamber boiling center temperature3
15	Combustion chamber boiling center temperature4
16	Furnace outlet temperature left
17	Furnace outlet temperature right
18	Low temperature superheater outlet oxygen left
19	Low temperature superheater outlet oxygen right
20	Cyclone outlet temperature left
21	Cyclone outlet temperature right
22	Oxygen content of air preheater outlet left
23	Oxygen content of air preheater outlet right

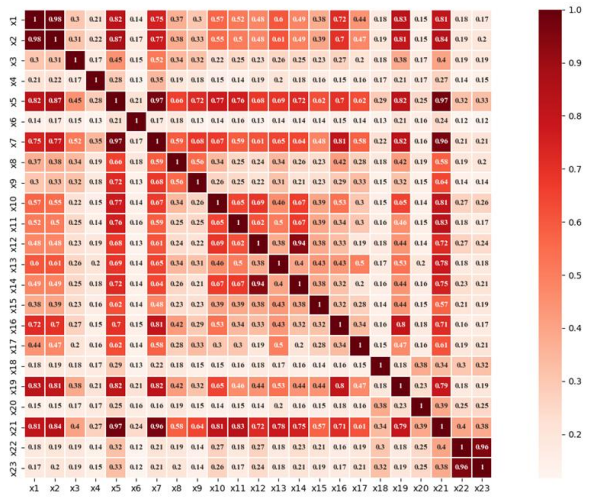

Figure 5: Heat map of correlation of 23 input variables.

4.2. Evaluation of IMFRVR model features

The IMFRVR model mainly contains two hyperparameters: kernel width and the number of clusters k value. In the following, the effect of the two hyperparameters on the model are verified. Moreover, the optimal parameters are obtained by analyzing experimentally. To assess the effectiveness of the prediction model, three evaluation criteria are employed: root mean square error (RMSE), mean absolute square error (MAE), and determination coefficient $\left(\mathrm{R}^2\right)$. These measures quantify the disparity between the actual and predicted values, delineating:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\tilde{x}_i - x_i)^2} \tag{20}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\tilde{x}_i - x_i| \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(\tilde{x}_i - x_i)^2}{\sum_{i=1}^{N}(\bar{x}_i - x_i)^2} \tag{22}$$

where N is the testing samples size; x_i represents the measured value; $\widetilde{x_i}$ represents the predicted value; and $\bar{x}_i$ represents the average of all measurements in the testing samples. RMSE quantifies the general discrepancy between the measured and predicted values. MAE assesses the resemblance between measured values and predicted values. R^2 evaluates the degree of correlation between the measured and predicted values. A reduced RMSE and MAE signify improved predictive capability, while an elevated R^2 value indicates greater precision in prediction.

4.3. Selection of k value

As an important parameter in the k-means clustering method, selecting of the k value directly determines the partition result of the original dataset into subsets and then affects the final experimental prediction effect. To this end, after data preprocessing and normalization of 2000 groups of samples, SSE is chosen as the evaluation metric, and the relationship between SSE and k values is depicted in Fig. 6.

As the k value increases, the division of samples becomes more granular, and each cluster's clustering density gradually increases, and the SSE naturally becomes smaller. When the k value is lower than the optimal value, raising the k value will significantly enhance the clustering density within each cluster, resulting in a substantial reduction in the SSE. As the k value continues to rise, the degree of clustering density reduction will become slower, and the SSE gradually approach a state of equilibrium. The relationship diagram between the SSE and the k value is an elbow shape. The optimal value can be determined by identifying the k value associated with the elbow point. The optimal k value is determined to be 10 in this paper. The clustering outcomes are depicted in Fig. 7. Each data cluster represents a different color in the figure. The number of each cluster is shown in Table 2.

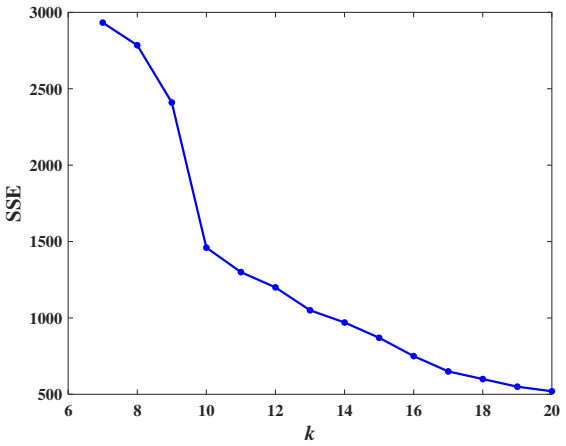

Figure 6: A coupled mass-spring-damper system.

4.4. Kernel width selection

After specifying the total number of samples and the data for each small cluster, each cluster's most appropriate kernel width should be determined to generate the optimal set of relevance vectors. The kernel width is selected by the size approximation method, specifically

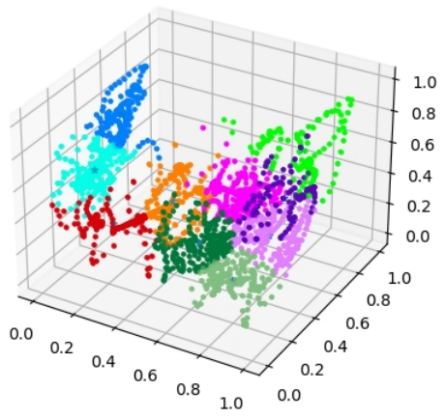

Figure 7: Diagram of the clustering result.

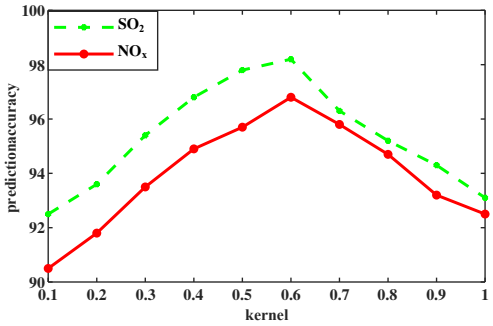

Figure 8: Relationship between kernel width and prediction accuracy.

within the range of [0-1]. Taking class 0 data as an example, the relationship between pollutant emission prediction accuracy and kernel width is illustrated in Fig. 8. When the kernel width is 0.6, the prediction accuracy for both pollutants reaches its peak. Therefore, the kernel width for class 0 is set to 0.6. The prediction effect for class 0 is depicted in Fig. 9. The method described above is applied to ascertain the optimal kernel width and relevance vector set for each class. The final results are shown in Table 2.

Table 2: Clustering results

Category	Samples size	Training/test sample size	kernel width	RV size
0	114	57	0.6	16
1	308	154	0.3	25
2	176	88	0.5	26
3	200	100	0.4	24
4	172	86	0.5	13
5	116	58	0.6	22
6	172	86	0.5	21
7	270	135	0.3	36
8	248	124	0.3	26
9	218	109	0.4	21

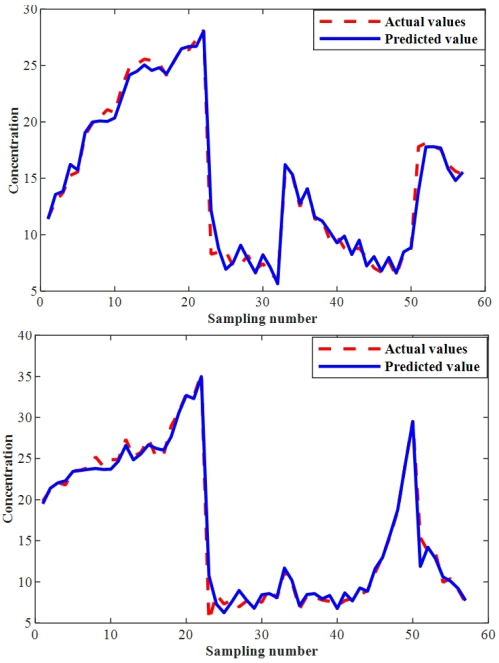

Figure 9: Class 0 prediction results of SO_2 and NO_x.

4.5. Prediction and comparative experiments

After confirming the number of clusters method and the kernel width of the algorithm through the above experimental part, the effectiveness of the IMFRVR model is illustrated by a comparison experiment. IMFRVR model is compared with four algorithms: direct incremental multivariable fast relevance vector regression (DIMFRVR), incremental fast relevance vector regression (IFRVR), retrain multivariable fast relevance vector regression (RMFRVR), and multi-output Gaussian process regression (MOGPR).

4.5.1. Performance of different prediction methods

The IFRVR algorithm is chosen to show the superior performance of the proposed algorithm. The IFRVR algorithm predicts the pollutant emission in the systems separately and cannot achieve simultaneous predictions. In the experiment, the same data preprocessing approaches and clustering methods are applied for both algorithms to complete the preliminary data preparation. The incremental data is input in the form of a data stream. Finally, the proposed algorithm's effectiveness is illustrated by comparing MAE, RMSE, R^2, and running time.

The prediction results of the two algorithms for pollutant emission are shown in Fig. 10. and Fig. 11. The pollutant emissions predicted by the IMFRVR model

are close to the actual values. However, for the IFRVR algorithm, there are significant differences between the predicted and actual values at the sharp point of data. According to Tab.3. and Tab.4., it can be observed that the IMFRVR model exhibits better prediction accuracy with the lower RMSE and MAE and higher R^2 in contrast to the IFRVR algorithm. This shows that simultaneous prediction of pollutants not only achieves better prediction efficiency, but also improves prediction accuracy, and the complexity of the algorithm itself does not change significantly.

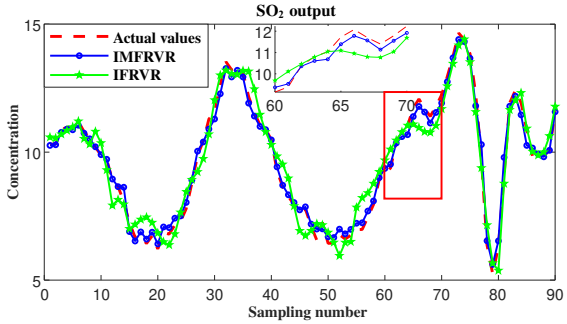

Figure 10: IMFRVR and IFRVR predict SO_2.

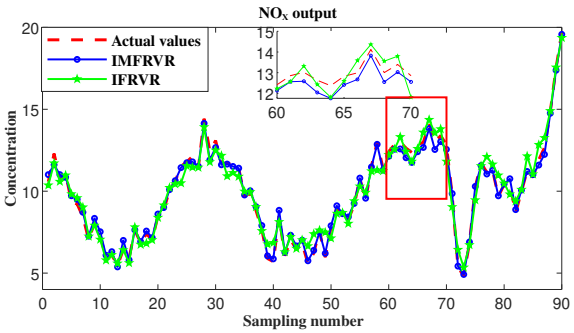

Figure 11: IMFRVR and IFRVR predict NO_x.

4.5.2. *Effect of different dynamic process learning approaches*

In this study, the RMFRVR model is compared to demonstrate the effectiveness of the incremental learning idea in the proposed algorithm. The new addition samples in the RMFRVR model are directly combined with the original sample data to form a new sample set for training the prediction model. In the experiments, the same method of data processing and input of the same form of incremental data are used for both algorithms. Finally, the proposed algorithm's effectiveness is illustrated by MAE, RMSE, R^2, and running time. The prediction results of the two algorithms for pollutant emission are shown in Fig. 12. and Fig. 13.

By analyzing Table 3 and Table 4, it is evident that the RMSE, MAE, and R^2 of the proposed model are better than the RMFRVR model. In particular, the training time is shortened by 2.7 times, thereby addressing the need for real-time predicting pollutants in the biomass cogeneration process. The inferior accuracy of the RMFRVR model can be attributed to its learning of all training samples when inputting in incremental samples. This results in the model's inability to effectively utilize its sparse characteristics as more incremental data sets are added, consequently diminishing its capacity to learn from large sample data and leading to a decline in accuracy. Furthermore, the continuous addition of data sets increases the model's computational time, resulting in prolonged running times.

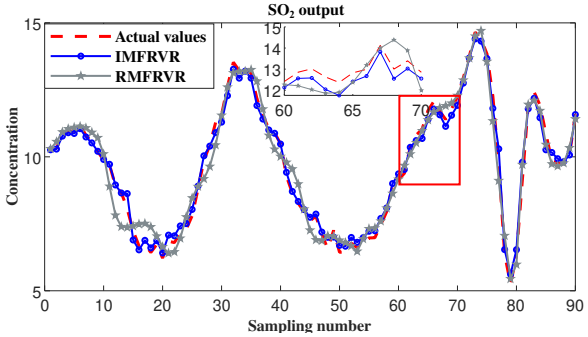

Figure 12: IMFRVR and RMFRVR predict SO_2.

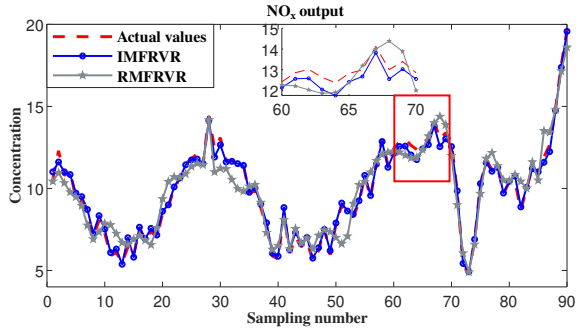

Figure 13: IMFRVR and RMFRVR predict NO_x.

4.5.3. *Impact of different data-processing methods*

The DIMFRVR model is applied to analyze the effect of clustering in the proposed model. The DIMFRVR model does not cluster the train data set but instead learns directly from the 2000 samples input into the algorithm. In the experiments, both methods keep

the same data processing process except for whether the data is clustered. Finally, the proposed algorithm's effectiveness is illustrated by MAE, RMSE, R^2, and running time. The prediction results of the two algorithms for pollutant emission are shown in Fig. 14. and Fig. 15.

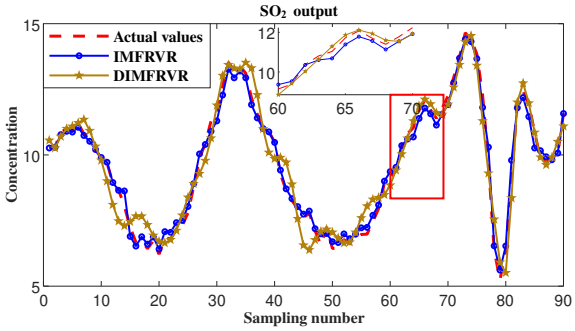

Figure 14: IMFRVR and DIMFRVR predict SO_2.

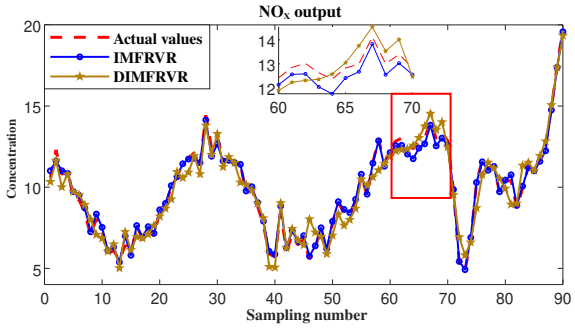

Figure 15: IMFRVR and DIMFRVR predict NO_x.

Compared with the proposed algorithm, the DIMFRVR model has a larger MAE and RMSE, and a smaller R^2. This indicates that the clustering of training data can help enhance prediction accuracy. The primary reason for the suboptimal prediction accuracy of the DIMFRVR model is that sizeable unclustered data sets are input into the model, causing the model to learn some useless information and decreasing prediction accuracy. Conversely, the segmentation of large datasets enables the clustering of data with similar characteristics, thereby facilitating the extraction of more pertinent information by the prediction model. The prolonged computational time of the DIMFRVR model can be attributed to the direct input of large unclustered datasets into the prediction model, which increases the computational complexity of the EM algorithm optimization hyperparameter process. Simultaneously, extensive sample data makes the DIMFRVR model more susceptible to noise interference. This results in the sparsity of weights being less evident than in small samples, directly affecting operating efficiency. Deformation methods based on the RVR principle demonstrate superior performance on small sample datasets. Therefore, clustering sample data is imperative to improve the accuracy and efficiency of pollutant emission predictions.

4.5.4. Impact of different multi-prediction methods

The multi-output Gaussian process regression (MOGPR) method is chosen to reveal the superior performance of the RVRbased method. During the experimental phase, the two methods utilized the same data preprocessing and sample dimensionality reduction method. The training data used for the MOGPR method was not clustered and the incremental data set was input as a data stream. Finally, the proposed algorithm's effectiveness is illustrated by comparing MAE, RMSE, R^2, and running time.

Fig. 16 and Fig. 17 clearly show that the fitting effect of the MOGPR method is significantly worse than the proposed method. Although the MOGPR method appears to capture the prediction trend, its accuracy is notably inferior. The RMSE and MAE of the MOGPR method are much larger than those of the proposed method, and the R^2 is notably smaller. This is due to the complex correlations and dependencies between the two pollutants emitted from the biomass cogeneration systems, which were not learned well by the MOGPR method under the current data conditions, making the method's prediction performance degrade on the test set.

Apart from the unsatisfactory prediction accuracy of the MOGPR method, its running time is also difficult to guarantee. After each incremental dataset is added, it trains the entire new dataset. With the continuous addition of incremental datasets, it takes longer and longer time to construct the covariance matrix for the whole training data. It directly leads to the long running time of the MOGPR method. This is unacceptable for the biomass cogeneration systems requiring real-time pollutant concentration prediction. Meanwhile, the relatively high computational complexity of the MOGPR method makes it have higher requirements on the processor, which will invariably increase the application cost. In contrast, when the proposed method performs incremental learning, the incremental data is only combined with the relevance vectors obtained from the previous training to form a new data set to train the model. The proposed method's sparse characteristics are fully utilized, the computational complexity is significantly

reduced, and the running time is guaranteed. To the greatest extent, it meets the real-time demand for predicting pollutant emissions from biomass cogeneration systems.

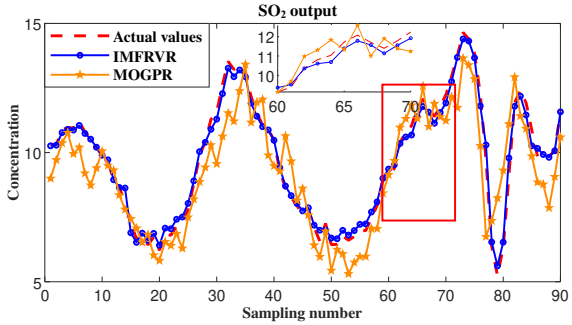

Figure 16: IMFRVR and MOGPR predict SO_2.

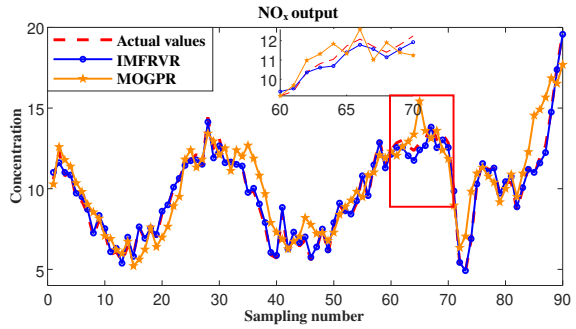

Figure 17: IMFRVR and MOGPR predict NO_x.

Table 3: Algorithm performance comparison table (SO_2).

Algorithm type	RMSE	MAE	R^2	Time(s)
IMFRVR	0.15	0.08	0.96	81.5
IFRVR	2.54	1.30	0.89	160.7
RMFRVR	2.88	1.80	0.81	300.4
DIMFRVR	3.68	2.43	0.70	632.2
MOGPR	10.63	3.91	0.64	1132

Table 4: Algorithm performance comparison table (NO_x).

Algorithm type	RMSE	MAE	R^2	Time(s)
IMFRVR	0.23	0.17	0.95	81.5
IFRVR	2.89	1.39	0.85	162.3
RMFRVR	2.79	1.48	0.76	300.4
DIMFRVR	3.87	2.82	0.69	632.2
MOGPR	11.97	3.1	0.6	1132

5. Conclusions

This study proposes an extended incremental multivariate fast relevance vector regression model for pollutant emission prediction. The effectiveness of the model is assessed through the pollutant emission of biomass cogeneration systems case. The primary contribution of this paper can be succinctly outlined as (1) regardless of NO_x or SO_2, the proposed method provides the lowest RMSE and MAE while the highest R^2 by comparing different prediction models. Among them, the prediction accuracy of NO_x and SO_2 reached $R^2 = 0.95$ and $R^2 = 0.96$, respectively; (2) the introduction of the incremental learning idea shortened the running time of the model by 72.8% compared to the retraining method, in which low compute complexity is achieved; (3) multiple pollutants synchronous prediction can fully consider the coupling relationship between output variables, improving the prediction accuracy by nearly 10%; (4) the proposed method's prediction accuracy can be more guaranteed on small sample data due to the sparsity of the RVR model itself.

In future research, the IMFRVR model will be improved to address the challenge of a lack of robustness caused by the Gaussian distribution. Moreover, predicting pollutant emissions in biomass cogeneration systems under more complex operating conditions is necessary. In addition, the over-fitting problem of the IMFRVR model is also worthy of our attention. The IMFRVR model effectively avoids over-fitting problems by introducing a sparse prior, parameter sparsity, and the utilization of flatter prior distributions. However, this does not mean that the IMFRVR model does not have overfitting problems at all. Once the prior distribution of the model is set unreasonably and there is a large amount of noise or outliers in the data, the IMFRVR model may be overfitting. In our future work, we plan to study different situations where the IMFRVR model has overfitting problems in a targeted manner so that these situations can be avoided in the application process.

Credit author statement

Xiuli Wang: Conceptualization, Data curation, Methodology, Supervision., Zhifei Sun: Conceptualization, Methodology, Writing-original draft, Writing-review & editing, Supervision., Defeng He: Formal analysis, Writing-review & editing, Supervision., Shaomin Wu: Writing-review & editing, Supervision., Lianna Zhao: Writing-review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgments

This work was supported in part by the Central Guidance Project for Local Scientific and Technological Development under Grant 2023ZY1045, National Natural Science Foundation of China under Grant 62303414, Natural Science Foundation of Zhejiang Province under Grant LQ23F030016.

References

Bishop, C. M. (2006). Pattern recognition and machine learning. *Springer google schola*, *2*, 1122–1128.

Chen, H., & Huang, B. (2023). Fault-tolerant soft sensors for dynamic systems. *IEEE Transactions on Control Systems Technology*, *31*(6), 2805–2818.

Chen, H., Liu, Z., Alippi, C., Huang, B., & Liu, D. (2024). Explainable intelligent fault diagnosis for nonlinear dynamic systems: From unsupervised to supervised learning. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(5), 6166–6179.

Chen, H., Luo, H., Huang, B., Jiang, B., & Kaynak, O. (2023). Transfer learning-motivated intelligent fault diagnosis designs: A survey, insights, and perspectives. *IEEE Transactions on Neural Networks and Learning Systems*, *35*(3), 2969–2983.

Guo, W., & He, M. (2022). An integrated method for bearing state change identification and prognostics based on improved relevance vector machine and degradation model. *IEEE Transactions on Instrumentation and Measurement*, *71*, 1–14.

Han, Z., Li, J., Hossain, M. M., Qi, Q., Zhang, B., & Xu, C. (2022). An ensemble deep learning model for exhaust emissions prediction of heavy oil-fired boiler combustion. *Fuel*, *308*, 121975.

Jia, S., Ma, B., Guo, W., & Li, Z. S. (2021). A sample entropy based prognostics method for lithium-ion batteries using relevance vector machine. *Journal of Manufacturing Systems*, *61*, 773–781.

Jin, Y., Scherer, L., Sutanudjaja, E. H., Tukker, A., & Behrens, P. (2022). Climate change and ccs increase the water vulnerability of china's thermoelectric power fleet. *Energy*, *245*, 123339.

Kang, M. S., Jeong, H. J., Farid, M. M., & Hwang, J. (2017). Effect of staged combustion on low nox emission from an industrial-scale fuel oil combustor in south korea. *Fuel*, *210*, 282–289.

Li, N., Lu, G., Li, X., & Yan, Y. (2015). Prediction of pollutant emissions of biomass flames through digital imaging, contourlet transform, and support vector regression modeling. *IEEE Transactions on Instrumentation and Measurement*, *64*(9), 2409–2416.

Li, Y., Fei, M., Jia, L., Lu, N., Kaynak, O., & Zio, E. (2024). Novel outlier-robust accelerated degradation testing model and lifetime analysis method considering time-stress-dependent factors. *IEEE Transactions on Industrial Informatics*, (pp. 1–11).

Li, Y., Wang, X., Lu, N., & Jiang, B. (2021). Conditional joint distribution-based test selection for fault detection and isolation. *IEEE Transactions on Cybernetics*, *52*(12), 13168–13180.

Li, Y., Xu, S., Chen, H., Jia, L., & Ma, K. (2022). A general degradation process of useful life analysis under unreliable signals for accelerated degradation testing. *IEEE Transactions on Industrial Informatics*, (pp. 7742–7750).

Liu, D., Zhou, J., Pan, D., Peng, Y., & Peng, X. (2015). Lithium-ion battery remaining useful life estimation with an optimized relevance vector machine algorithm with incremental learning. *Measurement*, *63*, 143–151.

Lv, Y., Liu, J., Yang, T., & Zeng, D. (2013). A novel least squares support vector machine ensemble model for nox emission prediction of a coal-fired boiler. *Energy*, *55*, 319–329.

Nazari, S., Shahhoseini, O., Sohrabi-Kashani, A., Davari, S., Paydar, R., & Delavar-Moghadam, Z. (2010). Experimental determination and analysis of co2, so2 and nox emission factors in iran's thermal power plants. *Energy*, *35*(7), 2992–2998.

Nunes, L., Matias, J., & Catalão, J. (2014). A review on torrefied biomass pellets as a sustainable alternative to coal in power generation. *Renewable and Sustainable Energy Reviews*, *40*, 153–160.

Park, J. K., Park, S., Ryu, C., Baek, S. H., Kim, Y. J., & Park, H. Y. (2017). Cfd analysis on bioliquid co-firing with heavy fuel oil in a 400 mwe power plant with a wall-firing boiler. *Applied Thermal Engineering*, *124*, 1247–1256.

Qiu, G. (2013). Testing of flue gas emissions of a biomass pellet boiler and abatement of particle emissions. *Renewable energy*, *50*, 94–102.

Tan, P., Xia, J., Zhang, C., Fang, Q., & Chen, G. (2016). Modeling and reduction of nox emissions for a 700 mw coal-fired boiler with the advanced machine learning method. *Energy*, *94*, 672–679.

Thayananthan, A., Navaratnam, R., Stenger, B., Torr, P. H., & Cipolla, R. (2006). Multivariate relevance vector machines for tracking. In *Computer Vision–ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7-13, 2006, Proceedings, Part III 9* (pp. 124–138). Springer.

Thayananthan, A., Navaratnam, R., Stenger, B., Torr, P. H., & Cipolla, R. (2008). Pose estimation and tracking using multivariate regression. *Pattern Recognition Letters*, *29*(9), 1302–1310.

Tipping, M. E. (2001). Sparse bayesian learning and the relevance vector machine. *Journal of machine learning research*, *1*(Jun), 211–244.

Tipping, M. E., & Faul, A. C. (2003). Fast marginal likelihood maximisation for sparse bayesian models. In *International workshop on artificial intelligence and statistics* (pp. 276–283). PMLR.

Valente, L., Tarelho, L., & Costa, V. (2020). Emissions mitigation by control of biomass feeding in an industrial biomass boiler. *Energy Reports*, *6*, 483–489.

Wang, X., Jiang, B., Wu, S., Lu, N., & Ding, S. X. (2021). Multivariate relevance vector regression based degradation modeling and remaining useful life prediction. *IEEE Transactions on Industrial Electronics*, *69*(9), 9514–9523.

Xie, P., Gao, M., Zhang, H., Niu, Y., & Wang, X. (2020). Dynamic modeling for nox emission sequence prediction of scr system outlet based on sequence to sequence long short-term memory network. *Energy*, *190*, 116482.

Xu, S., Chen, X., Liu, F., Wang, H., Chai, Y., Zheng, W. X., & Chen, H. (2023). A novel adaptive smo-based simultaneous diagnosis method for igbt open-circuit faults and current sensor incipient faults of inverters in pmsm drives for electric vehicles. *IEEE transactions on instrumentation and measurement*, *72*, 1–15.

Yang, G., Wang, Y., & Li, X. (2020a). Prediction of the nox emissions from thermal power plant using long-short term memory neural network. *Energy*, *192*, 116597.

Yang, T., Ma, K., Lv, Y., & Bai, Y. (2020b). Real-time dynamic prediction model of nox emission of coal-fired boilers under variable load conditions. *Fuel*, *274*, 117811.

Zhou, H., Zhao, J. P., Zheng, L. G., Wang, C. L., & Cen, K. F. (2012). Modeling nox emissions from coal-fired utility boilers using support vector regression with ant colony optimization. *Engineering Applications of Artificial Intelligence*, *25*(1), 147–158.