Knowledge-Based Systems 253 (2022) 109477

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Robust kernel-free support vector regression based on optimal margin distribution

Jian Luo [a], Shu-Cherng Fang [b], Zhibin Deng [c,*], Ye Tian [d,*]

[a] Management School, Hainan University, Haikou 570228, China
[b] Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raleigh, NC 27695, USA
[c] School of Economics and Management, University of Chinese Academy of Sciences, MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS, Beijing 100190, China
[d] School of Business Administration and Collaborative Innovation Center of Financial Security, Faculty of Business Administration, Southwestern University of Finance and Economics, Chengdu 611130, China

ARTICLE INFO

Article history:
Received 26 August 2021
Received in revised form 15 July 2022
Accepted 15 July 2022
Available online 22 July 2022

Keywords:
Support vector regression
Optimal margin distribution
Kernel-free SVM
Robust forecasting model
Battery power consumption forecasting

ABSTRACT

Support vector machines have been proven to be useful for regression analysis and forecasting. When stochastic uncertainty is involved in the datasets, robust support vector regression (SVR) models are useful. In this study, we proposed a kernel-free quadratic surface support vector regression (QSSVR) model based on optimal margin distribution (OMD). This model minimizes the variance of the functional margins of all data points to achieve better generalization capability. When the data points exhibit stochastic uncertainty (without the assumption of any specific distribution), the covariance information of noise is employed to construct a robust OMD-based QSSVR (RQSSVR-OMD) model, with a set of probabilistic constraints to ensure its worst-case performance. Moreover, the probabilistic constraints in the proposed model are proven to be equivalently reformulated as second-order cone constraints for efficient implementation. Extensive computational experiments on public benchmark datasets were conducted to demonstrate the superior performance of the proposed RQSSVR-OMD model over other well-established SVR models in terms of forecasting accuracy and time. The proposed model was also validated to successfully handle real-life uncertain battery data for battery power-consumption forecasting.

1. Introduction

Prediction and forecasting are of vital importance in human decision-making, whereas regression analysis is widely used for prediction and forecasting. From the viewpoint of statistical learning theory [1], support vector regression (SVR) is an extension of the support vector machine (SVM) for regression analysis. The first SVM was introduced by Vapnik and Lerner in 1963 for the linear classification of a given set of input–output paired data points, which was subsequently improved by employing the kernel-function technique for nonlinear classification of such datasets [2]. Robust SVM models are used when the given data points involve uncertainty, such as randomly distributed noise and outliers [3–7]. Several SVR and robust SVR models have been proposed [8–11] for nonlinear regression of data points, possibly involving stochastic uncertainty. For practical applications, the accuracy of generalization, computational time, and restrictive assumptions of SVR are the key factors for a fair evaluation. In this study, we proposed a new robust SVR model that effectively conducts nonlinear regression analysis of datasets with minimal assumptions on uncertainty in an accurate and efficient manner.

The classic SVM determines an optimal hyperplane that maximizes the minimum margin between two distinct classes of labeled data points [2,12,13]. Subsequently, most developed SVM models followed this "maximal minimum margin" principle and its variants. Reyzin and Schapire [14] found that the poor margin distribution caused by maximizing the minimum margin, such as the boosting-style leaning algorithm "Arc-gv" in [15], would lead to a poor generalization capability in accuracy. Furthermore, Gao and Zhou [16] proved that by simultaneously maximizing the functional margin mean and minimizing the functional margin variance over all data points, could reduce the error in generalization. Motivated by this study, Zhang and Zhou [17] proposed an "optimal margin distribution learning machine" (OMDLM) for classification. Their computational experiments validated the better (more accurate) generalization capability of OMDLM over other state-of-the-art SVM models. This development clearly indicates the crucial role of the "distribution of functional margins" in achieving accurate generalization.

* Corresponding authors.
E-mail addresses: zhibindeng@ucas.edu.cn (Z. Deng), yetian@swufe.edu.cn (Y. Tian).

https://doi.org/10.1016/j.knosys.2022.109477
0950-7051/

To extend the SVM model to cases in which the data points are not completely linearly separable, the concept of "soft margin" was introduced by Cortes and Vapnik [2]. Regarding nonlinear classification, Boser et al. [18] suggested an SVM model by applying a kernel method to generate a separating hyperplane that maximizes the minimum margin in the feature space. Various SVM models with different kernel functions such as Gaussian, polynomial, and hyperbolic tangent functions have become readily available [2,19]. Because kernel-based SVM models operate on a higher-dimensional feature space and there is a lack of a universal rule to select the best kernel function with the most appropriate kernel parameters for a particular application, a significant amount of computational time and tuning efforts may be required for a successful application. Moreover, the possible singularity of a kernel matrix can significantly influence the accuracy of its generalization capability. Consequently, several kernel-free SVM models have been developed for nonlinear classification. For instance, Astorino and Fuduli [20] proposed a spherical separating surface with no kernel for semi-supervised classification, Luo et al. [21] proposed a kernel-free soft-margin quadratic surface SVM (QSSVM) for binary classification, Luo et al. [22] proposed an unsupervised non-kernel SVM approach for unsupervised classification with application to credit risk assessment, and Gao et al. [23] proposed a non-kernel least-squares twin SVM for fast and accurate multiclass classification. The computational results in [21–23] support the assumption that directly adopting a nonlinear surface for separation is an effective way to construct kernel-free SVM models with accurate generalization capability for general applications. Based on the QSSVM model framework, Ye et al. [24,25] proposed effective kernel-free quadratic surface SVR models for regression analysis.

The SVR model proposed by Drucker et al. [26] depends only on a subset of the training data, ignoring any training data points close to the model prediction in a given tube around the fitting hyperplane. This model is followed by various SVR models that appear in the literature for business and energy forecasting, such as financial forecasting [27], electric load forecasting [28], and hydropower production capacity prediction [29]. In general, SVR models perform well when the training data points are known [26]. However, for real-world applications, the datasets inevitably involve possible contamination noise, measurement errors, mislabeling, and even attacked data points, which may significantly downgrade the performance of SVR. Several kernel-based robust support vector regression (RSVR) models have been proposed to achieve robust performance by considering the uncertainties embedded in data points as stochastic noise. For instance, Trafalis and Alwazzi [9] considered input data points with normally distributed errors, whereas Abaszade and Effati [11] considered the case of uniformly distributed errors. Shivaswamy et al. [8] reduced the restrictive assumptions on the data distribution by using only the means and covariance matrices of the input data points in their RSVR model. When both the input and output data are stochastically uncertain, Huang et al. [10] proposed an RSVR model with a Gaussian kernel that requires only the means and covariances of the data points. The RSVR model demonstrated superior performance for some synthetic and real datasets. However, as stated in [10], covariance information in the feature space is required in any kernelized formulation of an RSVR model. Although this information in the original space is readily accessible, it cannot be obtained in the feature space because of an unknown nonlinear transformation. Thus, Huang et al. [10] used the estimated spectral norms of the covariance matrices in the feature space, even though adopting the estimations reduces forecasting accuracy.

The novel contribution of this paper is to propose a new robust kernel-free nonlinear regression model based on the principle of "optimal margin distribution" and probabilistic constraints (including covariance information) for datasets involving stochastic uncertainties, without knowing the actual distributions of noise, in an accurate and efficient manner. We note that the kernel-free SVRs in [24,25] are based on the commonly used principle of "maximal minimum margin". To enhance generalization accuracy, we proposed a new kernel-free quadratic surface SVR model based on the "optimal margin distribution" (QSSVR-OMD) instead of the "maximal minimum margin" principle. Moreover, for stochastically uncertain data, the use of a Gaussian kernel limits the forecasting efficiency and accuracy of the robust SVR models in [8,10] owing to the excessive computation requirement and inexact estimation of covariance matrices in the feature space. To address these challenges, we chose to build a kernel-free SVR directly using the means and covariances of the input–output data sets. For robust performance, probabilistic constraints were included in the proposed SVR for the worst situation. Specifically, a brand-new distributionally kernel-free robust QSSVR (RQSSVR-OMD) model is proposed for stochastically uncertain input–output data by incorporating covariance information and probabilistic constraints into the QSSVR-OMD model. For fast computation, the properties of the probabilistic constraints were further studied to show an equivalent formulation of the second-order cone (SOC) constraints. This allows the regression analysis of the proposed RQSSVR-OMD model to be conducted in an efficient manner. In addition, we will conduct extensive computational experiments on public benchmark datasets to demonstrate the superior performance of the proposed RQSSVR-OMD model over other well-established SVR models. The proposed model can also be applied to handle real-life uncertain battery data for accurate battery power-consumption forecasting.

The remainder of this paper is organized as follows. Section 2 briefly reviews the related SVR models in the literature. In Section 3, we construct a kernel-free QSSVR-OMD model and describe its properties. Then, the proposed RQSSVR-OMD model is introduced and investigated in Section 4. Section 5 describes the computational experiments performed on public benchmark and real-life datasets to validate the effectiveness and efficiency of the proposed model compared with other well-established robust SVR models. Finally, Section 6 concludes the paper.

2. Review of related support vector regression models

In this section, related classic SVR and RSVR models are briefly reviewed. Further details can be found in [8,26], and [10].

Given a dataset of n points $\{(\mathbf{x}^i, y^i)\}_{i=1}^n$, where $\mathbf{x}^i = [x_1^i, x_2^i, \ldots, x_m^i]^T \in \mathbb{R}^m$ and $y^i \in \mathbb{R}$, the linear SVR model intends to determine the parameters $(\mathbf{w}, b)$, $\mathbf{w} \in \mathbb{R}^m$, and $b \in \mathbb{R}$ of a hyperplane

$$f(\mathbf{x}, y) \triangleq \mathbf{w}^T\mathbf{x} + b - y = 0, \tag{1}$$

that fits the n data points while minimizing the regularization term $\|w\|^2$ and the fitting errors of data points outside the tube $|f(\mathbf{x}, y)| \le \delta$ for a given $\delta > 0$. Hence, the linear SVR model with linear constraints with respect to a given δ can be formulated as follows [26]:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 + C_1 \sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & \delta + \xi_i \ge \mathbf{w}^T\mathbf{x}^i + b - y^i \ge -\delta - \xi_i, \ i = 1, \ldots, n, \\ & \xi_i \ge 0, \ i = 1, \ldots, n, \\ & \mathbf{w} \in \mathbb{R}^m, b \in \mathbb{R}, \end{aligned} \tag{SVR}$$

where $C_1 > 0$ and $\delta > 0$ are given parameters. For nonlinear regression, a nonlinear kernel function $\phi(\mathbf{x}) : \mathbb{R}^m \to \mathbb{R}^d$ is first applied to map the data points $\mathbf{x}$ from the original space $\mathbb{R}^m$ to a

higher-dimensional feature space $\mathbb{R}^d$ (where $m \le d$ and d could be $+\infty$), and then the SVR model is applied to the transformed data in the higher-dimensional feature space $\mathbb{R}^d$. Hence, the SVR model with kernel $\phi(\mathbf{x})$ can be formulated as follows [26]:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\mathbf{v}\|_2^2 + C_1 \sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & \delta + \xi_i \ge \mathbf{v}^T \phi(\mathbf{x}^i) + b - y^i \ge -\delta - \xi_i, \ i = 1, \ldots, n, \\ & \xi_i \ge 0, \ i = 1, \ldots, n, \\ & \mathbf{v} \in \mathbb{R}^d, b \in \mathbb{R}, \end{aligned} \tag{SVR-ker}$$

where $C_1 > 0$ and $\delta > 0$ are given parameters. In general, it is difficult to identify a good kernel function $\phi(\mathbf{x})$ directly; thus, the dual problem of the SVR-ker model, which involves the preassigned expression of the kernel $K(\mathbf{x}, \mathbf{y}) = \phi(\mathbf{x})^T \phi(\mathbf{y})$, is always solved. Various kernels have been used in the literature, such as the Gaussian radial basis kernel [30], polynomial kernel [31], and hyperbolic tangent kernel [32]. Notably, the (SVR-ker) model can be reformulated as a second-order cone programming (SOCP) problem [8]; however, the performance of (SVR-ker) model depends significantly on the parameters of the kernel.

Throughout this paper, both input and output data are assumed to be disturbed by noise, that is, the observations $(\mathbf{x}^i, y^i) \in \mathbb{R}^{m+1}$, $i = 1, \ldots, n$, are random vectors. As in the setting in [10], the expectation and covariance information for each observation are assumed to be known or can be estimated in advance. The expectation and covariance matrices of each observation are denoted by

$$\bar{\mathbf{x}}^i \triangleq E(\mathbf{x}^i), \ \bar{y}^i \triangleq E(y^i), \tag{2}$$

and

$$\begin{aligned} &\Sigma^i_{\mathbf{xx}} \triangleq Cov(\mathbf{x}^i, \mathbf{x}^i), \ \Sigma^i_{yy} \triangleq Cov(y^i, y^i), \\ &\Sigma^i_{y\mathbf{x}} \triangleq Cov(y^i, \mathbf{x}^i), \ \Sigma^i_{\mathbf{x}y} \triangleq Cov(\mathbf{x}^i, y^i), \end{aligned} \tag{3}$$

respectively. Let $\mathbf{q}^i \triangleq [(\mathbf{x}^i)^T, y^i]^T \in \mathbb{R}^{m+1}$, $i = 1, \ldots, n$ then, the expectation and covariance matrix of $\mathbf{q}^i$ are

$$\bar{\mathbf{q}}^i \triangleq E([(\mathbf{x}^i)^T, y^i]^T) = [(\bar{\mathbf{x}}^i)^T, \bar{y}^i]^T \tag{4}$$

and

$$\Sigma^i_{\mathbf{qq}} \triangleq Cov(\mathbf{q}^i, \mathbf{q}^i) = \begin{bmatrix} \Sigma^i_{\mathbf{xx}} & \Sigma^i_{\mathbf{x}y} \\ \Sigma^i_{y\mathbf{x}} & \Sigma^i_{yy} \end{bmatrix}, \tag{5}$$

respectively. Let $\bar{K}$ be an $n \times n$ Gaussian kernel matrix with $\bar{K}_{ij} \triangleq e^{-\|\bar{\mathbf{x}}^i - \bar{\mathbf{x}}^j\|^2/(2\sigma^2)}$, and $\bar{K}_i \triangleq [\bar{K}_{1i}, \bar{K}_{2i}, \ldots, \bar{K}_{ni}]^T$ be the ith column of $\bar{K}$. We note that $\bar{K}$ is real, symmetric, and positive semi-definite; thus, $\bar{K}^{\frac{1}{2}}$ exists without loss of generality. Huang et al. [10] proposed a RSVR model with Gaussian kernel to handle the uncertain input and output data as follows:

$$\begin{aligned} \min \quad & t + C_1 \sum_{i=1}^{n} \epsilon_i \\ \text{s.t.} \quad & \sqrt{\|\tilde{\Sigma}^i_{\mathbf{qq}}\|(\boldsymbol{\theta}^T \bar{K} \boldsymbol{\theta} + 1) + (\boldsymbol{\theta}^T \bar{K}_i + b - \bar{y}^i)^2} \le (\delta + \epsilon_i)\sqrt{\beta}, \\ & i = 1, \ldots, n, \\ & \|\bar{K}^{\frac{1}{2}} \boldsymbol{\theta}\|_2 \le t, \\ & t, \epsilon_i \ge 0, \ i = 1, \ldots, n, \\ & \boldsymbol{\theta} \in \mathbb{R}^n, b \in \mathbb{R}, \end{aligned} \tag{RSVR-ker}$$

where $C_1 > 0$, $\delta > 0$ and $\beta \in [0, 1]$ are the given parameters, and $\tilde{\Sigma}^i_{\mathbf{qq}}$, $i = 1, \ldots, n$, are the covariance matrices of uncertain points in the feature space depending on the kernel used. However, $\tilde{\Sigma}^i_{\mathbf{qq}}$ cannot be obtained, because the kernel function $\phi(\mathbf{x})$ is usually unknown. To overcome this difficulty, Huang et al. [10] assumed that the input and output datasets are uncorrelated, that is, $\Sigma^i_{\mathbf{x}y} = \Sigma^i_{y\mathbf{x}} = \mathbf{0}$, where $\mathbf{0}$ is a vector whose elements are all zero, and used the spectral norm of $\tilde{\Sigma}^i_{\mathbf{qq}}$ in the aforementioned (RSVR-ker) model. Under this assumption, the spectral norm of $\tilde{\Sigma}^i_{\mathbf{qq}}$ for a Gaussian kernel can be approximated as follows [10]:

$$\begin{aligned} \|\tilde{\Sigma}^i_{\mathbf{qq}}\| &= \left\| \begin{matrix} \tilde{\Sigma}^i_{\mathbf{xx}} & \tilde{\Sigma}^i_{\mathbf{x}y} \\ \tilde{\Sigma}^i_{y\mathbf{x}} & \Sigma^i_{yy} \end{matrix} \right\| \approx \left\| \begin{matrix} \tilde{\Sigma}^i_{\mathbf{xx}} & \mathbf{0} \\ \mathbf{0} & \Sigma^i_{yy} \end{matrix} \right\| = \max\left\{ \|\tilde{\Sigma}^i_{\mathbf{xx}}\|, \|\Sigma^i_{yy}\| \right\} \\ &= \max\left\{ \sqrt{\frac{2}{n}\left(1 - e^{-\frac{m\|\Sigma^i_{\mathbf{xx}}\|^2}{2\sigma^2}}\right)}, \|\Sigma^i_{yy}\| \right\}, \end{aligned}$$

where $\tilde{\Sigma}^i_{\mathbf{xx}}$, $\tilde{\Sigma}^i_{\mathbf{x}y}$ and $\tilde{\Sigma}^i_{y\mathbf{x}}$ are the covariance matrices of the uncertain points in the feature space, and $\tilde{\Sigma}^i_{\mathbf{x}y} = \tilde{\Sigma}^i_{y\mathbf{x}} = 0$ because of the assumption that $\Sigma^i_{\mathbf{x}y} = \Sigma^i_{y\mathbf{x}} = 0$.

Notably, $(n + 1)$ SOC constraints exist in the RSVR-ker model. Moreover, each of the first n SOC constraints has an $(n + 2)$-dimensional SOC and the last SOC constraint has an n-dimensional SOC. As we will observe in the numerical experiment in Section 5.1, when the number of observations n is large, solving the RSVR-ker model becomes computationally difficult. This motivated us to develop a new robust SVR model with low-dimensional SOC constraints.

3. Quadratic surface support vector regression based on optimal margin distribution

In this section, the quadratic surface SVR (QSSVR) model is first introduced by directly using a quadratic surface for nonlinear regression. Then, the QSSVR-OMD model is proposed by incorporating the objective of minimizing the variance of functional margins into the QSSVR model. The proposed QSSVR-OMD model is further simplified, and the convexity of its optimization problem is proved.

3.1. Quadratic Surface Support Vector Regression (QSSVR)

Given a dataset of n points $\{(\mathbf{x}^i, y^i)\}_{i=1}^n$, the QSSVR model aims to determine the parameters $(W, \mathbf{h}, c)$ of a quadratic surface

$$g(\mathbf{x}, y) \triangleq \frac{1}{2}\mathbf{x}^T W \mathbf{x} + \mathbf{h}^T \mathbf{x} + c - y = 0, \tag{6}$$

where $W = W^T \in \mathbb{R}^{m \times m}$, $\mathbf{h} \in \mathbb{R}^m$, and $c \in \mathbb{R}$, which fit n data points without using any kernel. Similar to the definitions of the various margins of points in [21], we develop the following definitions of margins for this regression problem:

Definition 1. $\zeta_i \triangleq |g(\mathbf{x}^i, y^i)|$ is the functional margin at point $(\mathbf{x}^i, y^i)$ with respect to $g(\mathbf{x}, y) = 0$.

Definition 2. We assume that a positive (or negative) gradient direction at point $(\mathbf{x}^i, y^i)$ with respect to $g(\mathbf{x}, y) = g(\mathbf{x}^i, y^i)$ intercepts the quadratic surface $g(\mathbf{x}, y) = 0$ at point $(\mathbf{x}^B, y^B)$. The distance between points $(\mathbf{x}^i, y^i)$ and $(\mathbf{x}^B, y^B)$, denoted by $\hat{\zeta}_i$, is the geometrical margin at point $(\mathbf{x}^i, y^i)$ with respect to $g(\mathbf{x}, y) = 0$.

Definition 3. For $g(\mathbf{x}^i, y^i) \ge 0$ (or ≤ 0), given $\delta > 0$, the positive (or negative) gradient direction at point $(\mathbf{x}^i, y^i)$ with respect to $g(\mathbf{x}, y) = g(\mathbf{x}^i, y^i)$ intercepts the surfaces $g(\mathbf{x}, y) = +\delta$ (or $-\delta$) and $g(\mathbf{x}, y) = 0$ at points $(\mathbf{x}^I, y^I)$ and $(\mathbf{x}^B, y^B)$, respectively. The distance between points $(\mathbf{x}^I, y^I)$ and $(\mathbf{x}^B, y^B)$, denoted by $\bar{\zeta}_i$, is called the relative geometric margin at point $(\mathbf{x}^i, y^i)$ with respect to $g(\mathbf{x}, y) = 0$.

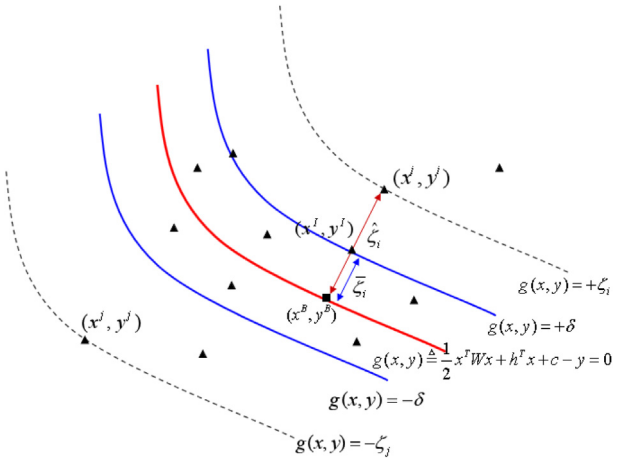

Fig. 1. Demonstration of various margins with respect to a quadratic surface $g(\mathbf{x}, y) = 0$ in two dimension.

Fig. 1 illustrates the $(\mathbf{x}^i, y^i)$, $(\mathbf{x}^I, y^I)$, $(\mathbf{x}^B, y^B)$, ζ_i, $\hat{\zeta}_i$, and $\bar{\zeta}_i$ for $m = 1$. Similar to the proof in [21], the geometrical margin $\hat{\zeta}_i$ and the relative geometrical margin $\bar{\zeta}_i$ at point $(\mathbf{x}^i, y^i)$ can be approximated by $\frac{\zeta_i}{\|W\mathbf{x}^i+\mathbf{h}\|_2}$ and $\frac{\delta}{\|W\mathbf{x}^i+\mathbf{h}\|_2}$, respectively. Therefore, with respect to the quadratic surface $g(\mathbf{x}, y) = 0$, the relative geometrical margins of different points are different in general.

Similar to the classic SVR models, the goal of the QSSVR model is to generate one "tube" and then try to include as many points in this "tube" as possible. Specifically, we first ignore the errors in the data points inside the tube $|g(\mathbf{x}, y)| \le \delta$ for a given $\delta > 0$. Then, to include as many data points in this tube as possible, we not only maximize the sum of the relative geometrical margin of each point with respect to $g(\mathbf{x}, y) = 0$ (which can be approximated by minimizing $\sum_{i=1}^n \|W\mathbf{x}^i + \mathbf{h}\|_2^2$ (refer to the formulation of QSSVM in [21]), but also minimize the deviations of data points with errors greater than δ. Therefore, the QSSVR model can be formulated as follows:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^n \|W\mathbf{x}^i + \mathbf{h}\|_2^2 + C_1 \sum_{i=1}^n \epsilon_i \\
\text{s.t.} \quad & \delta + \epsilon_i \ge \frac{1}{2}(\mathbf{x}^i)^T W\mathbf{x}^i + \mathbf{h}^T\mathbf{x}^i + c - y^i \ge -\delta - \epsilon_i, \\
& i = 1, \ldots, n, \\
& \epsilon_i \ge 0, i = 1, \ldots, n, \\
& W = W^T \in \mathbb{R}^{m\times m},\ \mathbf{h} \in \mathbb{R}^m,\ c \in \mathbb{R}., \qquad \text{(QSSVR)}
\end{aligned}
$$

where $C_1, \delta > 0$ are the given parameters. According to Definition 1, the functional margin of point $(\mathbf{x}^i, y^i)$ with respect to $g(\mathbf{x}, y) = 0$ is $\zeta_i = |\frac{1}{2}(\mathbf{x}^i)^T W\mathbf{x}^i + \mathbf{h}^T\mathbf{x}^i + c - y^i|$. From (QSSVR) model, we can observe that $\epsilon_i = \zeta_i - \delta$ for point $(\mathbf{x}^i, y^i)$ is outside the tube $|g(\mathbf{x}, y)| \le \delta$. Hence, minimizing $\sum_{i=1}^n \epsilon_i$ in the objective of (QSSVR) model minimizes the mean of the functional margins outside the tube $|g(\mathbf{x}, y)| \le \delta$. This is related to optimizing the first-order information of the functional margins of points in OMD theory [17].

3.2. Quadratic surface support vector regression based on optimal margin distribution

Given a dataset of n points $\{(\mathbf{x}^i, y^i)\}_{i=1}^n$, the QSSVR-OMD model aims to find the parameters $(W, \mathbf{h}, c)$ of a quadratic surface $g(\mathbf{x}, y) \triangleq \frac{1}{2}\mathbf{x}^T W\mathbf{x} + \mathbf{h}^T\mathbf{x} + c - y = 0$ which fits n data points by incorporating the optimal functional margin distribution into the QSSVR model. Because the objective of QSSVR model is to minimize the mean of the functional margins of points outside the tube $|g(\mathbf{x}, y)| \le \delta$, an additional objective of QSSVR-OMD is to minimize the variance of the functional margins of all points with respect to $g(\mathbf{x}, y) = 0$. This is motivated by optimizing the second-order information (i.e., minimizing the variance) of the functional margins of the points in the OMD theory. We denote the variance of the functional margins of all points by

$$
\gamma \triangleq \frac{1}{n}\sum_{i=1}^n (\zeta_i - \frac{1}{n}\sum_{j=1}^n \zeta_j)^2 = \frac{1}{n^2}\zeta^T(n\mathbf{I}_n - \mathbf{1}_n)\zeta,
$$

where $\zeta = [\zeta_1, \zeta_2, \ldots, \zeta_n]^T$, $\zeta_i = |\frac{1}{2}(\mathbf{x}^i)^T W\mathbf{x}^i + \mathbf{h}^T\mathbf{x}^i + c - y^i|$ is the functional margin at point $(\mathbf{x}^i, y^i)$, $\mathbf{I}_n$ and $\mathbf{1}_n$ denotes the $n \times n$ identity and $n \times n$ matrices, where each element is one, respectively. Then, by incorporating the term minimizing γ into the QSSVR model, we can formulate the QSSVR-OMD model as follows:

$$
\begin{aligned}
\min \quad & \sum_{i=1}^n \|W\mathbf{x}^i + \mathbf{h}\|_2^2 + C_1 \sum_{i=1}^n \epsilon_i + \frac{C_2}{n^2}\zeta^T(n\mathbf{I}_n - \mathbf{1}_n)\zeta \\
\text{s.t.} \quad & \delta + \epsilon_i \ge \frac{1}{2}(\mathbf{x}^i)^T W\mathbf{x}^i + \mathbf{h}^T\mathbf{x}^i + c - y^i \ge -\delta - \epsilon_i, \\
& i = 1, \ldots, n, \\
& \zeta_i \ge \frac{1}{2}(\mathbf{x}^i)^T W\mathbf{x}^i + \mathbf{h}^T\mathbf{x}^i + c - y^i \ge -\zeta_i, i = 1, \ldots, n, \\
& W = W^T \in \mathbb{R}^{m\times m},\ \mathbf{h} \in \mathbb{R}^m,\ c \in \mathbb{R}, \\
& \zeta_i, \epsilon_i \ge 0, i = 1, \ldots, n, \qquad \text{(QSSVR-OMD)}
\end{aligned}
$$

where $C_1, C_2, \delta > 0$ are given parameters. We note that the second and last terms in the objective of (QSSVR-OMD) OMD model are related to the first-order and second-order regression errors of points outside the tube $|g(\mathbf{x}, y)| \le \delta$ and all points, respectively.

The matrix W is symmetric; hence, the QSSVR-OMD model can be equivalently simplified as follows: First, let Φ be the vector formulated by considering the $\frac{m^2+m}{2}$ elements of the upper triangular part of matrix W, that is, $\Phi \triangleq [w_{11}, w_{12}, \ldots, w_{1m}, w_{22}, w_{23}, \ldots, w_{2m}, \ldots, w_{mm}]^T \in \mathbb{R}^{\frac{m^2+m}{2}}$. Then, for $i = 1, \ldots, n$, we can construct an $m \times \frac{m^2+m}{2}$ matrix $\mathbf{M}_i$ for point $\mathbf{x}^i \in \mathbb{R}^m$ as follows: For the jth row of $\mathbf{M}_i$, $j = 1, 2, \ldots, m$, if the pth element of Φ is w_{jk} or w_{kj} for some $k = 1, 2, \ldots, m$, we assign the pth element of the jth row of $\mathbf{M}_i$ to be x_k^i. Alternatively, we assign it a value of zero. Furthermore, let $\mathbf{H}_i \triangleq [\mathbf{M}_i, \mathbf{I}_m] \in \mathbb{R}^{m\times(\frac{m^2+m}{2}+m)}$, $i = 1, \ldots, n$, $\mathbf{Q} \triangleq \sum_{i=1}^n \mathbf{H}_i^T\mathbf{H}_i \in \mathbb{R}^{\frac{m^2+3m}{2}\times\frac{m^2+3m}{2}}$, the vector of variables $\mathbf{z} \triangleq [\Phi^T, \mathbf{h}^T]^T = [w_{11}, \ldots, w_{1m}, w_{22}, \ldots, w_{2m}, \ldots, w_{mm}, \mathbf{h}^T]^T \in \mathbb{R}^{\frac{m^2+3m}{2}}$. Then, the first term in the objective of (QSSVR-OMD) OMD model becomes:

$$
\begin{aligned}
&\sum_{i=1}^n \|W\mathbf{x}^i + \mathbf{h}\|_2^2 = \sum_{i=1}^n \|\mathbf{H}_i\mathbf{z}\|_2^2 = \sum_{i=1}^n (\mathbf{H}_i\mathbf{z})^T(\mathbf{H}_i\mathbf{z}) \\
&= \sum_{i=1}^n \mathbf{z}^T(\mathbf{H}_i)^T\mathbf{H}_i\mathbf{z} = \mathbf{z}^T(\sum_{i=1}^n (\mathbf{H}_i)^T\mathbf{H}_i)\mathbf{z} = \mathbf{z}^T\mathbf{Q}\mathbf{z}.
\end{aligned}
$$

Moreover, let $\mathbf{s}^i \triangleq [\frac{1}{2}x_1^i x_1^i, \ldots, x_1^i x_m^i, \frac{1}{2}x_2^i x_2^i, \ldots, x_2^i x_m^i, \ldots, \frac{1}{2}x_{m-1}^i x_{m-1}^i, x_{m-1}^i x_m^i, \frac{1}{2}x_m^i x_m^i, x_1^i, x_2^i, \ldots, x_m^i]^T \in \mathbb{R}^{\frac{m^2+3m}{2}}$, the QSSVR-OMD model can be equivalently reformulated as below:

$$
\begin{aligned}
\min \quad & \mathbf{z}^T\mathbf{Q}\mathbf{z} + C_1 \sum_{i=1}^n \epsilon_i + \frac{C_2}{n^2}\zeta^T(n\mathbf{I}_n - \mathbf{1}_n)\zeta \\
\text{s.t.} \quad & \delta + \epsilon_i \ge \mathbf{z}^T\mathbf{s}^i + c - y^i \ge -\delta - \epsilon_i, i = 1, \ldots, n,
\end{aligned}
$$

$$\zeta_i \geq \mathbf{z}^T\mathbf{s}^i + c - y^i \geq -\zeta_i, i = 1, \ldots, n,$$
$$\mathbf{z} \in \mathbb{R}^{\frac{m^2+3m}{2}}, c \in \mathbb{R},$$
$$\zeta_i, \epsilon_i \geq 0, \ i = 1, \ldots, n, \qquad \text{(QSSVR-OMD')}$$

where $C_1, C_2, \delta > 0$ are given parameters. Here, parameter C_1 determines the trade-off between the relative geometrical margins of data points and the functional margins of data points whose deviations are greater than δ, and parameter C_2 determines the trade-off between the relative geometrical margins of data points and the variance of the functional margins of data points. Hence, the objective function of the QSSVR-OMD' model prevents overfitting by minimizing the first regularization term and underfitting by minimizing the second and third terms of the first-order and second-order fitting errors of the training points, respectively. The proposed model is proven to be a convex quadratic programming problem as follows:

Theorem 1. *The QSSVR-OMD' model is a convex quadratic programming problem with linear constraints.*

Proof. First, the matrix $n\mathbf{I}_n - \mathbf{1}_n$ is proved to be positive semi-definite for any positive integer n. It is easy to verify that matrix $n\mathbf{I}_n - \mathbf{1}_n$ is positive semi-definite when $n = 1$. For $n \geq 2$, we assume that the eigenvalue of matrix $n\mathbf{I}_n - \mathbf{1}_n$ is λ then, $det(n\mathbf{I}_n - \mathbf{1}_n - \lambda\mathbf{I}_n) = 0$, that is,

$$\begin{aligned} det(n\mathbf{I}_n - \mathbf{1}_n - \lambda\mathbf{I}_n) &= \begin{vmatrix} n-\lambda-1 & -1 & \cdots & -1 \\ -1 & n-\lambda-1 & \cdots & -1 \\ \vdots & & & \\ -1 & -1 & \cdots & n-\lambda-1 \end{vmatrix} \\ &= \begin{vmatrix} -\lambda & -\lambda & \cdots & -\lambda \\ -1 & n-\lambda-1 & \cdots & -1 \\ \vdots & & & \\ -1 & -1 & \cdots & n-\lambda-1 \end{vmatrix} \\ &= \begin{vmatrix} -\lambda & -\lambda & \cdots & -\lambda \\ 0 & n-\lambda & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & n-\lambda \end{vmatrix} \\ &= (n-\lambda)^{n-1}\lambda = 0 \end{aligned}$$

Hence, matrix $n\mathbf{I}_n - \mathbf{1}_n$ has only two eigenvalues, 0 and n, which shows that matrix $n\mathbf{I}_n - \mathbf{1}_n$ is positive semidefinite.

Then, for any $\tilde{\mathbf{x}} \in \mathbb{R}^{\frac{m^2+3m}{2}}$, $\tilde{\mathbf{x}}^T\mathbf{Q}\tilde{\mathbf{x}} = \tilde{\mathbf{x}}^T(\sum_{i=1}^n \mathbf{H}_i^T\mathbf{H}_i)\tilde{\mathbf{x}} = \sum_{i=1}^n \tilde{\mathbf{x}}^T\mathbf{H}_i^T\mathbf{H}_i\tilde{\mathbf{x}} = \sum_{i=1}^n \|\mathbf{H}_i\tilde{\mathbf{x}}\|_2^2 \geq 0$. Hence, the matrix $\mathbf{Q}$ is also positive semi-definite.

Therefore, the objective of the QSSVR-OMD' model is a convex quadratic function because it is a nonnegative combination of three convex quadratic functions. In summary, the QSSVR-OMD' model is a convex quadratic programming problem with linear constraints. □

Using this theorem, the QSSVR-OMD' model can be solved in polynomial time within any given accuracy by some well-established interior-point solvers [33].

4. Robust QSSVR-OMD for uncertain input and output data

In this section, we first introduce the probabilistic constraints to handle uncertain input and output data of any distribution, and then describe the sufficient and necessary conditions for these constraints. Finally, we propose a distributionally robust QSSVR-OMD model based on these constraints and data covariance matrices.

We assume that both input and output data are disturbed by random noises. That is, the observations $(\mathbf{x}^i, y^i) \in \mathbb{R}^{m+1}, i = 1, \ldots, n$, are random vectors. Then, $\mathbf{s}^i$ and $i = 1, \ldots, n$, are also random vectors. Similar to Eqs. (2)–(5) in Section 2, the expectation and covariance matrices of $\mathbf{s}^i$ and y^i are denoted by

$$\bar{\mathbf{s}}^i \triangleq E(\mathbf{s}^i), \ \bar{y}^i \triangleq E(y^i), \tag{7}$$

and

$$\Sigma_{\mathbf{ss}}^i \triangleq Cov(\mathbf{s}^i, \mathbf{s}^i), \ \Sigma_{yy}^i \triangleq Cov(y^i, y^i), \tag{8}$$
$$\Sigma_{\mathbf{sy}}^i \triangleq Cov(\mathbf{s}^i, y^i), \ \Sigma_{\mathbf{ys}}^i \triangleq Cov(y^i, \mathbf{s}^i), \tag{9}$$

respectively. Let $\mathbf{u}^i \triangleq [(\mathbf{s}^i)^T, y^i]^T \in \mathbb{R}^{\frac{m^2+3m}{2}+1}, \ i = 1, \ldots, n$, then, the random variables $\mathbf{z}^T\mathbf{s}^i + c - y^i, \ i = 1, \ldots, n$, become $[\mathbf{z}^T, -1]\mathbf{u}^i + c$. The expectation and covariance matrices of $\mathbf{u}^i$ are:

$$\bar{\mathbf{u}}^i \triangleq E([(\mathbf{s}^i)^T, y^i]^T) = [(\bar{\mathbf{s}}^i)^T, \bar{y}^i]^T, \tag{10}$$

and

$$\Sigma_{\mathbf{uu}}^i \triangleq Cov(\mathbf{u}^i, \mathbf{u}^i) = \begin{bmatrix} \Sigma_{\mathbf{ss}}^i & \Sigma_{\mathbf{sy}}^i \\ \Sigma_{\mathbf{ys}}^i & \Sigma_{yy}^i \end{bmatrix}, \tag{11}$$

respectively. We denote $\mathbf{u}^i \sim (\bar{\mathbf{u}}^i, \Sigma_{\mathbf{uu}}^i)$ as a family of distributions with expectation $\bar{\mathbf{u}}^i$ and covariance matrix $\Sigma_{\mathbf{uu}}^i$. Subsequently, for uncertain input and output data of any distribution, the probabilistic constraints for the worst situation are defined as follows:

$$\sup_{\mathbf{u}^i \sim (\bar{\mathbf{u}}^i, \Sigma_{\mathbf{uu}}^i)} \Pr\left\{\left|[\mathbf{z}^T, -1]\mathbf{u}^i + c\right| \geq \delta + \epsilon_i\right\} \leq \beta, \ i = 1, \ldots, n, \tag{12}$$

where $\Pr\{A\}$ is the probability of an event A, $\delta > 0$ and $0 < \beta \leq 1$ are constants, and $\epsilon_i \geq 0$ is a slack variable that will be penalized for large values in the objective. Clearly, small values of ϵ_i and β yield observed prediction errors close to zero. To develop a distributionally robust QSSVR-OMD model, we incorporate the probabilistic constraints (12) into the QSSVR-OMD model and replace the random vectors $\mathbf{x}^i, \mathbf{u}^i$ with their expectations $\bar{\mathbf{x}}^i, \bar{\mathbf{u}}^i$ in other constraints and the objective function. This provides the following formulation for a distributionally robust QSSVR-OMD:

$$\begin{aligned} \min \quad & \mathbf{z}^T\bar{\mathbf{Q}}\mathbf{z} + C_1\sum_{i=1}^n \epsilon_i + \frac{C_2}{n^2}\zeta^T(n\mathbf{I}_n - \mathbf{1}_n)\zeta \\ \text{s.t.} \quad & \sup_{\mathbf{u}^i \sim (\bar{\mathbf{u}}^i, \Sigma_{\mathbf{uu}}^i)} \Pr\left\{\left|[\mathbf{z}^T, -1]\mathbf{u}^i + c\right| \geq \delta + \epsilon_i\right\} \leq \beta, i = 1, \ldots, n, \\ & \zeta_i \geq [\mathbf{z}^T, -1]\bar{\mathbf{u}}^i + c \geq -\zeta_i, i = 1, \ldots, n, \\ & \mathbf{z} \in \mathbb{R}^{\frac{m^2+3m}{2}}, c \in \mathbb{R}, \\ & \zeta_i, \epsilon_i \geq 0, \ i = 1, \ldots, n, \end{aligned} \qquad \text{(RQSSVR-OMD)}$$

where $C_1, C_2, \delta > 0$ and $\beta \in [0, 1]$ are the given parameters, and $\bar{\mathbf{Q}}$ is generated by replacing the random vector $\mathbf{x}^i$ in $\mathbf{Q}$ with its expectation $\bar{\mathbf{x}}^i$ for $i = 1, \ldots, n$. It can be easily verified that $\bar{\mathbf{Q}}$ remains positive semidefinite. However, this model is difficult to solve owing to probabilistic constraints. To implement this model efficiently, we study the sufficient and necessary conditions for (12) in the next theorem.

Theorem 2. *Let $\mathbf{z} \in \mathbb{R}^{\frac{m^2+3m}{2}}$, $c \in \mathbb{R}$, and $\bar{\mathbf{s}}^i, \bar{y}^i, \Sigma_{\mathbf{uu}}^i$ be defined in (7) and (11); then, the sufficient and necessary conditions for (12) are*

$$\sqrt{[\mathbf{z}^T, -1]\Sigma_{\mathbf{uu}}^i[\mathbf{z}^T, -1]^T + ([\mathbf{z}^T, -1]\bar{\mathbf{u}}^i + c)^2} \leq (\delta + \epsilon_i)\sqrt{\beta}, \quad i = 1, \ldots, n. \tag{13}$$

Proof. The proof consists of two parts: First, we prove that

$$\sup_{\mathbf{u}\sim(\bar{\mathbf{u}}^i,\Sigma^i_{\mathbf{uu}})} \Pr\left\{\left|[\mathbf{z}^T,-1]\mathbf{u}^i+c\right|\geq\delta+\epsilon_i\right\} = \frac{E[([\mathbf{z}^T,-1]\mathbf{u}^i+c)^2]}{(\delta+\epsilon_i)^2},\ i=1,\ldots,n. \tag{14}$$

Furthermore, using the Chebyshev inequality, we have that

$$\Pr\left\{\left|[\mathbf{z}^T,-1]\mathbf{u}^i+c\right|\geq\delta+\epsilon_i\right\} \leq \frac{E[([\mathbf{z}^T,-1]\mathbf{u}^i+c)^2]}{(\delta+\epsilon_i)^2},\ i=1,\ldots,n. \tag{15}$$

We assume that each random variable $[\mathbf{z}^T,-1]\mathbf{u}^i+c, i=1,\ldots,n$, is discretely distributed, considering the values $-\delta-\epsilon_i$ with probability p_1, 0 with probability p_2, and $\delta+\varepsilon_i$ with probability $1-p_1-p_2$. That is

$$[\mathbf{z}^T,-1]\mathbf{u}^i+c \sim \begin{pmatrix} -\delta-\epsilon_i & 0 & \delta+\epsilon_i \\ p_1 & p_2 & 1-p_1-p_2 \end{pmatrix}. \tag{16}$$

Then, $\Pr\left\{\left|[\mathbf{z}^T,-1]\mathbf{u}^i+c\right|\geq\delta+\epsilon_i\right\} = p_1+1-p_1-p_2 = 1-p_2$ and

$$\frac{E[([\mathbf{z}^T,-1]\mathbf{u}^i+c)^2]}{(\delta+\epsilon_i)^2} = \frac{p_1(-\delta-\epsilon_i)^2+(1-p_1-p_2)(\delta+\epsilon_i)^2}{(\delta+\epsilon_i)^2} = 1-p_2,\ i=1,\ldots,n.$$

Hence, a specific distribution (16) is found, such that the equality in (15) holds. This indicates that (14) is true.

Second, we prove that

$$E[([\mathbf{z}^T,-1]\mathbf{u}^i+c)^2] = [\mathbf{z}^T,-1]\Sigma^i_{\mathbf{uu}}[\mathbf{z}^T,-1]^T+([\mathbf{z}^T,-1]\bar{\mathbf{u}}^i+c)^2. \tag{17}$$

To observe this, we note that

$$\begin{aligned} &E[([\mathbf{z}^T,-1]\mathbf{u}^i+c)^2] \\ &= E[([\mathbf{z}^T,-1]\mathbf{u}^i)^2-([\mathbf{z}^T,-1]\bar{\mathbf{u}}^i)^2]+([\mathbf{z}^T,-1]\bar{\mathbf{u}}^i+c)^2 \\ &= [\mathbf{z}^T,-1]\Sigma^i_{\mathbf{uu}}[\mathbf{z}^T,-1]^T+([\mathbf{z}^T,-1]\bar{\mathbf{u}}^i+c)^2, \end{aligned}$$

where the last equality follows from the definition of covariance matrix $\Sigma^i_{\mathbf{uu}}$. Substituting (14) and (17) into (12) and calculating the square root on both sides yields inequality (13) in the statement. □

Theorem 2 shows that probabilistic constraints (12) can be equivalently reformulated as SOC constraints (13). Then, for uncertain input and output data of any distribution with given expectations and covariance matrices, the distributionally robust QSSVR-OMD model can be formulated as follows:

$$\begin{aligned} \min\ & \mathbf{z}^T\bar{\mathbf{Q}}\mathbf{z}+C_1\sum_{i=1}^n\epsilon_i+\frac{C_2}{n^2}\zeta^T(n\mathbf{I}_n-\mathbf{1}_n)\zeta \\ \text{s.t.}\ & \sqrt{[\mathbf{z}^T,-1]\Sigma^i_{\mathbf{uu}}[\mathbf{z}^T,-1]^T+([\mathbf{z}^T,-1]\bar{\mathbf{u}}^i+c)^2}\leq(\delta+\epsilon_i)\sqrt{\beta}, \\ & i=1,\ldots,n, \\ & \zeta_i\geq[\mathbf{z}^T,-1]\bar{\mathbf{u}}^i+c\geq-\zeta_i, i=1,\ldots,n, \\ & \mathbf{z}\in\mathbb{R}^{\frac{m^2+3m}{2}},\ c\in\mathbb{R}, \\ & \zeta_i\geq 0,\ \epsilon_i\geq 0,\ i=1,\ldots,n, \end{aligned} \tag{RQSSVR-OMD}$$

where $C_1, C_2, \delta>0$ and $\beta\in[0,1]$ are the given parameters. This model is a convex SOCP problem; thus, it can be efficiently solved using convex programming solvers.

Table 1
Information of tested data sets.

Data set	# of points	# of dimensions
slump	103	8
autoprice	159	15
machine	209	7
forestfire	270	9
mpg	392	8
housing	506	14
abalone	4177	9
Parkspe	1040	27
traffic	2101	48
SML	4137	19
super	5315	51

5. Computational experiments

Some public benchmark datasets were utilized to investigate the performance of the proposed RQSSVR-OMD model. The proposed model was then applied to battery power consumption forecasting with real-life uncertain data. For fair comparisons, the SVR model with Gaussian kernel (denoted by "SVR-Gker"), the SVR model with quadratic kernel (denoted by "SVR-Qker"), the robust SVR model with Gaussian kernel proposed by Shivaswamy et al. [8] (denoted by "RSVR-Gker_S"), the robust SVR model with Gaussian kernel proposed by Huang et al. [10] (denoted by "RSVR-Gker_H"), and the cutting-edge kernel-free soft QSSVR proposed by Ye et al. [25] (denoted by "SQSSVR") were also tested using the same data sets. For all the methods tested in this study, the same 10-fold cross-validation and grid methods were used to select the best parameters of C_1, C_2, β, δ, and the Gaussian kernel parameter σ^2 from the ranges of $\log_2 C_1 \in \{3,4,\ldots,28,29\}$, $\log_2 C_2 \in \{0,1,\ldots,13,14\}$, $\beta \in \{0.1,0.2,\ldots,0.8,0.9\}$, $\log_2\delta\in\{-5,-4,\ldots,3,4\}$, and $\log_2\frac{1}{2\sigma^2}\in\{-15,-14,\ldots,4,5\}$, respectively. Notably, the variation in β does not affect the forecasting accuracy of the proposed RQSSVR-OMD model for all computational experiments in this section. All numerical tests in this study were performed using MATLAB (R2019a) software on a desktop equipped with an Intel Xeon processor, 2.99 GHz CPU, 32 GB of RAM, and Microsoft Windows 10 Enterprise. The SVR model with Gaussian or quadratic kernel is implemented using the "fitrsvm" and "predict" modules of MATLAB, the SQSSVR model is implemented using the "quadprog" module of MATLAB, whereas the RSVR-Gker_S, RSVR-Gker_H and RQSSVR-OMD models are implemented using the "sedumi" module of CVX MATLAB toolbox [33].

5.1. Numerical experiments on public benchmark data sets

For computational tests, 11 public benchmark datasets were obtained from the UCI machine learning repository [34]. The information of all tested data sets is summarized in Table 1. We note that because a few noise points exist in these public benchmark datasets, 50 Gaussian distributed noise with a zero mean and 0.01 standard deviation are added to each dimension of every input and output point in each dataset to generate 50 disturbed points, similar to the computational procedures in [11]. Hence, the actual number of points used in the tested datasets ranged from 5150 to 265,750. In addition, the covariance matrix of every uncertain point utilized in the RSVR-Gker_S, RSVR-Gker_H, and proposed RQSSVR-OMD models was calculated from the generated 50 disturbed points, similar to the computational procedure in [10].

To obtain statistically meaningful results, we conducted ten tests for each public benchmark dataset. In each test, we randomly selected $k\%$ of the total points as the training dataset.

Table 2
Forecasting errors of tested models on benchmark data.

Data set	SVR-Gker		SVR-Qker		RSVR-Gker_S		RSVR-Gker_H		SQSSVR		RQSSVR-OMD	
	MAPE	ER	MAPE	ER	MAPE	ER	MAPE	ER	MAPE	ER	MAPE	ER
slump	18.79%	12.69	18.74%	12.68	7.20%	4.70	6.46%	4.21	7.36%	4.50	1.30%	0.47
autoprice	32.22%	4111	40.66%	4407	15.62%	2084	15.14%	2056	18.88%	2069	12.44%	1605
machine	120.64%	76.40	228.83%	111.94	48.72%	42.51	48.59%	42.02	49.16%	45.83	40.75%	32.19
forestfire	585.82%	29.87	977.72%	28.60	316.48%	22.52	302.77%	22.24	217.27%	23.15	196.23%	22.26
forestfire_1	656.36%	31.23	1544.72%	29.53	387.36%	23.86	337.10%	22.65	1004.94%	37.40	224.01%	22.41
mpg	12.63%	3.42	24.05%	15.23	11.36%	2.94	10.64%	2.91	10.87%	2.93	8.31%	1.96
mpg_1	26.60%	10.80	50.31%	17.40	24.59%	6.79	24.17%	6.75	25.73%	7.77	8.47%	1.98
housing	19.55%	4.26	39.36%	12.16	15.98%	4.22	15.55%	4.11	16.24%	3.91	11.35%	2.30
housing_1	39.88%	10.59	42.91%	11.03	32.78%	6.65	32.33%	6.62	62.59%	21.39	13.92%	3.78
abalone	14.96%	1.62	26.35%	6.19	14.47%	1.58	14.36%	1.57	14.51%	1.56	14.23%	1.53
abalone_1	27.25%	4.62	69.67%	8.92	26.18%	2.62	25.95%	2.61	112.00%	16.99	22.74%	2.59
Parkspe	370.24%	12.27	339.41%	11.15	315.24%	11.89	311.13%	11.77	576.11%	13.86	184.18%	10.21
traffic	30.30%	0.04	64.42%	0.07	102.16%	0.11	99.87%	0.10	110.89%	0.08	18.66%	0.02
SML	10.27%	6.56	10.42%	31.78	9.92%	2.15	9.90%	2.10	0.84%	0.17	0.79%	0.16
super	1147.35%	36.49	1269.64%	37.34	776.16%	34.06	758.28%	33.98	657.35%	33.93	124.86%	33.34

Table 3
Diebold–Mariano test.

Data set	SVR-Gker		SVR-Qker		RSVR-Gker_S		RSVR-Gker_H		SQSSVR	
	DM of MAPEs	DM of ERs	DM of MAPEs	DM of ERs	DM of MAPEs	DM of ERs	DM of MAPEs	DM of ERs	DM of MAPEs	DM of ERs
slump	−6.56***	−27.66***	−6.60***	−27.56***	−5.18***	−12.40***	−4.17**	−12.81***	−5.62***	−11.96***
autoprice	−5.84***	−6.07***	−6.76***	−6.91***	−3.48**	−3.94**	−3.30**	−2.28*	−4.26**	−2.87*
machine	−3.82**	−3.26**	−8.66***	−8.52***	−2.99*	−2.73*	−2.76*	−2.58*	−3.15*	−3.18*
forestfire	−4.41**	−2.38*	−5.16**	−3.04*	−2.62*	−0.32	−2.37*	−0.16	−1.07	−2.31*
forestfire_1	−3.52**	−3.72**	−3.60**	−2.91*	−2.51*	−0.92	−2.29*	−0.55	−3.96**	−3.77**
mpg	−6.63***	−8.56***	−11.62***	−15.27***	−4.53**	−9.93***	−4.31**	−9.80***	−4.39**	−9.83***
mpg_1	−12.88***	−16.52***	−15.36***	−23.35***	−13.69***	−20.18***	−13.73***	−19.81***	−13.65***	−20.98***
housing	−9.21***	−5.66***	−17.63***	−13.38***	−8.16***	−7.89***	−8.00***	−7.61***	−19.23***	−18.38***
housing_1	−18.53***	−12.62***	−18.91***	−13.89***	−9.68***	−9.09***	−9.58***	−8.80***	−20.19***	−19.48***
abalone	−3.07*	−4.01**	−29.56***	−16.78***	−2.75*	−3.62**	−2.31*	−3.37**	−2.98*	−3.88**
abalone_1	−33.86***	−6.05***	−36.78***	−23.67***	−6.63***	−4.66**	−6.26***	−2.74*	−24.65***	−41.33***
Parkspe	−8.16***	−8.05***	−8.59***	−4.58**	−5.98***	−19.93***	−5.94***	−19.89***	−8.73***	−6.17**
traffic	−2.74*	−6.60***	−2.96*	−6.98***	−5.12***	−47.87***	−5.07***	−46.63***	−4.03***	−77.09***
SML	−12.91***	−4.35**	−13.62***	−8.95***	−6.29***	−7.21***	−6.22***	−7.16***	−2.46*	−2.87*
super	−6.26***	−7.37***	−6.72***	−7.69***	−3.47*	−5.86***	−3.45*	−5.81***	−2.28*	−5.06***

The SVR-Gker, SVR-Qker, RSVR-Gker_S, RSVR-Gker_H, SQSSVR, and RQSSVR-OMD models were then trained using the selected training dataset to generate the parameters of the corresponding regressors. Then these regressors were used to predict the output of the remaining $1-k\%$ points in the dataset, and the forecasting errors were calculated. Following a computational procedure similar to that in [8,10], k was set to 90 for fair comparisons. We note that, for k being 90, the ratio of the number of training points to that of forecasting points is the same as that of ten-fold cross-validation [12,25,35], which is a common practice in the field of machine learning. For each model, the average of the mean absolute percentage errors (MAPE) and expected residuals (ER) (defined in [10]) is reported in Table 2. Smaller MAPE and ER values indicate that the corresponding model produces more accurate and robust forecasts. For the four public benchmark data sets: "forestfire", "mpg", "housing", and "abalone", the Gaussian distributed noise with zero mean and one standard deviation is added to each dimension to generate the data sets "forestfire_1", "mpg_1", "housing_1", and "abalone_1", respectively, as shown in Table 2. The aim of this study is to investigate the performance of all tested models on datasets with noise of a relatively larger magnitude. The Diebold–Mariano statistical test was performed between MAPEs (or ERs) of the proposed method and those of every tested method, with the statistical differences being recorded in Table 3. Table 3 shows DM statistics with *** denoting the p-value <0.001, ** denoting the p-value <0.01, and * denoting the p-value <0.05, which means that the statistical difference between MAPEs (or ERs) of the proposed model and those of the benchmark model is at a significance level of less than 0.001, 0.01, and 0.05, respectively. Moreover, the average computational times of all tested models are reported in Table 4.

From the numerical results listed in Tables 2–4, we obtain the following observations:

- From Table 3, the statistical difference between MAPEs (or ERs) of the proposed model and those of each tested SVR model is at a significant level of less than 0.05 for most cases.
- For most computational experiments on tested data sets, the proposed RQSSVR-OMD model performs best in accuracy, whereas the SVR-Gker, SVR-Qker, and SQSSVR models perform worst. The RSVR-Gker_S and RSVR-Gker_H models tie for the middle position. There are three possible reasons for this result. First, the RQSSVR-OMD and RSVR models utilize the information of full covariance matrices of quadratic forms and approximated spectral norms of covariance matrices in the feature space, respectively, whereas the information of covariance matrices is not considered in the SVR-Gker, SVR-Qker, and SQSSVR models. Hence, the RQSSVR-OMD and RSVR models produce more accurate forecasts than the SVR-Gker, SVR-Qker, and SQSSVR models for these uncertain data sets with noise. Second, optimizing the functional margin distribution (particularly minimizing the variance of functional margins) in the proposed RQSSVR-OMD model contributes to improving the forecasting accuracy. Hence, the kernel-free RQSSVR-OMD model yields more accurate forecasts than the kernel-free SQSSVR model, particularly for the data sets with a small magnitude of noise. Third, the two RSVR models assume that the uncertain input data is uncorrelated with the uncertain output data (i.e., $\Sigma^i_{\mathbf{xy}} = \Sigma^i_{y\mathbf{x}} = \mathbf{0}$) and then the estimated spectral norms of covariance matrices is adopted, although a

Table 4
Average computational time (in seconds) of tested models on benchmark data.

Data set	SVR-Gker	SVR-Qker	RSVR-Gker_S	RSVR-Gker_H	SQSSVR	RQSSVR-OMD
slump	0.15	9.76	0.54	0.58	0.24	0.33
autoprice	0.68	28.72	1.36	1.43	0.96	0.67
machine	1.52	2.93	2.47	2.96	1.62	1.30
forestfire	2.51	37.63	7.29	7.39	2.77	1.39
forestfire_1	5.33	39.73	7.40	7.49	6.67	1.33
mpg	5.96	10.31	35.47	36.22	6.54	3.58
mpg_1	10.63	28.31	36.09	36.55	12.33	3.62
housing	17.80	48.66	77.47	20.64	25.77	8.19
housing_1	20.39	56.73	77.81	78.19	24.84	8.37
abalone	261.06	2433.13	2 496.19	2 878.56	353.94	249.95
abalone_1	280.57	2542.67	2 620.96	2 993.72	384.10	257.97
Parkspe	118.90	122.65	2 894.36	2 793.75	261.40	113.62
traffic	136.65	283.65	3 623.38	3 980.75	2931.47	1537.35
SML	1523.80	3757.75	12 633.18	12 954.05	5180.50	1221.75
super	2399.25	5256.60	17 251.64	17 562.80	7030.20	2603.25

correlation between the uncertain input and output data always exist, the inexact estimation of covariance matrices for uncertain points reduces the forecasting accuracy of RSVR models. The RQSSVR-OMD model has no such restricted assumptions, and the covariance matrices of quadratic forms for uncertain points can be exactly obtained. Hence, the RQSSVR-OMD model outperforms the RSVR models in terms of forecasting accuracy.

- From Table 2, as the magnitude of noise in the benchmark data sets (i.e., "forestfire", "mpg", "housing", and "abalone") increases, the superiority of the proposed RQSSVR-OMD model becomes more evident. This observation indicates that the RQSSVR-OMD model is much more robust than the other five tested SVR models. The main reason is that the SOC constraints in the RQSSVR-OMD model are equivalently reformulated from probabilistic constraints, which directly use the exact information of means and covariance matrices of the input–output uncertain data sets.
- From Table 4, the required computational time of the SVR-Gker and RQSSVR-OMD models is comparable, whereas the two state-of-the-art RSVR models are much less computationally efficient. The main reason is that, the dimensionality of the SOC constraints in the RQSSVR-OMD model is much less than that of each tested RSVR model for these data sets, whereas the number of SOC constraints in the RQSSVR-OMD is one fewer than that in each tested RSVR model. For the non-commercial SOCP solver used in our experiments, the number and dimensionality of SOC constraints primarily impact the computational time of solving the SOCP reformulations of tested RQSSVR-OMD and RSVR models.

5.2. Application to battery power consumption forecasting

The proposed model was applied to a real-life problem to predict battery power consumption. In this subsection, we first describe the battery condition data and the prepossessing procedure and then show the comparison results between the proposed model and other tested SVR models.

Data-driven modeling is a promising route for the diagnostics and prognostics of lithium-ion batteries and enables emerging applications in the development, manufacturing, and optimization of lithium-ion batteries [36]. The accurate prediction of battery power consumption using early cycle data may open new avenues in battery production, use, fault detection, repair, and optimization. The real-life battery condition data used in this study were collected from the batteries in unmanned vehicles in a large unmanned warehouse operated by a Chinese corporation. In this dataset, there are 15,058 data points of 16 features regarding the working and charging conditions of 339 batteries. The features include battery usage time, average current, battery fluctuation, charging time, business working time, system mode time, number of charging procedures, number of business tasks and errors, average vehicle traveling distance, and battery power consumption.

Moreover, to reflect real-life circumstances, all nominal variables without the logic connection were transferred into one or several categories accordingly, and then one dummy variable for each category was utilized to represent the specific state of the related battery (i.e., one indicates yes and zero indicates no). For nominal variables with logic or ranking connections, each category was transformed into an integer according to the logic or ranking connection. For the preprocessed data points $\{\mathbf{x}^i, i = 1, \ldots, n\}$ with all numerical features, where $\mathbf{x}^i = [x_1^i, x_2^i, \ldots, x_m^i]^T \in \mathbb{R}^m$, all input features are z-scored to prevent the dominance of input features with greater numerical values than those with smaller values, that is,

$$x_j^i = \frac{x_j^i - \rho_j}{\sqrt{\upsilon_j}}, \ i = 1, \ldots, n, j = 1, \ldots, m,$$

where ρ_j and υ_j are the mean and variance of $\{x_j^i, i = 1, \ldots, n\}$, respectively. In general, there are many fluctuations, errors, and noise in the measurements of battery conditions, and the working and charging conditions of the battery are stochastically uncertain. Hence, for each battery, we collected the corresponding battery condition data points from 15,058 data points and considered them as uncertain data points, from which the expectation and covariance matrix were calculated for this battery. In this manner, we generated a new dataset including 339 uncertain points of 16 features.

The proposed RQSSVR-OMD model was applied to battery power consumption forecasting by testing it on a new real-life uncertain battery dataset. For fair comparisons, well-established SVR models (including the SVR-Gker, SVR-Qker, RSVR-Gker_S, SQSSVR, and RSVR-Gker_H models) were also tested on the same battery dataset, following the experimental settings described at the start of Section 5. Following the same experimental procedures as described in Section 5.1, the average forecasting error (i.e., MAPE) and computational time (i.e., CPU time) of all the tested methods are listed in Table 5. The Diebold–Mariano statistics between MAPEs of the proposed method and those of every tested method are also recorded in Table 5. In contrast to tests on synthetic and public benchmark data, the covariance matrices utilized in the RSVR and RQSSVR-OMD models for this battery data were calculated from the real-life uncertain 15,058 points.

From Table 5, in addition to the similar observations in the previous subsection, we observe that the average forecasting error (i.e., MAPE) of the proposed RQSSVR-OMD model is reduced by at least 85% when compared to the well-established SVR and

Table 5
Numerical results of tested methods on battery data.

Method	MAPE (%)	CPU time (s)	DM of MAPEs
SVR-Gker	87.91	17.38	−10.93***
SVR-Qker	95.00	99.81	−11.77***
RSVR-Gker_S	87.30	364.56	−10.46***
RSVR-Gker_H	84.95	386.36	−10.17***
SQSSVR	86.22	123.93	−10.64***
RQSSVR-OMD	12.15	85.68	N/A

RSVR models, as it requires less than one-fourth of the CPU time of the RSVR models. For real-life uncertain battery data, the assumption adopted by RSVR models (i.e., $\Sigma^i_{\mathbf{x}y} = \Sigma^i_{y\mathbf{x}} = \mathbf{0}$) may be violated such that the two RSVR models do not perform well. Hence, for this uncertain battery data, the three features of the kernel-free quadratic surface, OMD, and full information of noise covariances dominate the performance of the proposed model in terms of forecasting accuracy. Moreover, because all the input features of the uncertain battery dataset are scaled into their z-scores, the mathematical expression of the optimal fitting quadratic surface $\frac{1}{2}\mathbf{x}^T W^*\mathbf{x} + \mathbf{h}^{*T}\mathbf{x} + c^* = y$ can be obtained after solving the RQSSVR-OMD model. The coefficients $(W^*, \mathbf{h}^*)$ of the optimal quadratic surface indicate that variables as well as the interactions between them play a key role in predicting real-life battery power consumption. The accurate forecasting results of the proposed RQSSVR-OMD model also increase the interpretability of the obtained regressor.

6. Conclusions

In this study, we proposed a new support vector regression model, named RQSSVR-OMD, which can robustly conduct a nonlinear regression analysis for datasets involving stochastic uncertainties, without knowing the actual distributions of noise, in an accurate and efficient manner. The proposed method presents several good features: (i) it is kernel-free for saving tuning efforts; (ii) it adopts the idea of optimal margin distribution to enhance the generalization accuracy; (iii) it fully utilizes the covariance information of stochastically uncertain data points without knowing their actual distributions for robust performance; and (iv) it converts probabilistic constraints to second-order cone constraints for fast computation.

Extensive computational experiments were conducted using public benchmark datasets. The results clearly show that the proposed model outperforms the commonly used SVR model with a Gaussian kernel (SVR-Gker) or quadratic kernel (SVR-Qker) and SQSSVR models in terms of accuracy and the other two state-of-the-art robust SVR models (RSVR-Gker_S and RSVR-Gker_H) in terms of both accuracy and computational time. In particular, the successful application of the proposed model to a real-life uncertain battery dataset has demonstrated its effectiveness and efficiency in battery power consumption forecasting. This result indicates the potential of RQSSVR-OMD to address robust forecasting problems in the real world.

The proposed RQSSVR-OMD model provides details about future research directions. We are currently interested in exploring this new model for other uncertain sets (such as those using the Wasserstein metric [37]), robust multiclass classification [6], and clustering analysis [38].

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.

Acknowledgments

J. Luo's research was supported by the National Natural Science Foundation of China Grant #71701035. Z. Deng's research was supported by the National Natural Science Foundation of China Grant #12171151, the Fundamental Research Funds for the Central Universities, and a grant from the MOE Social Science Laboratory of Digital Economic Forecasts and Policy Simulation at UCAS. Y. Tian's research was supported by the Fundamental Research Funds for Central Universities #JBK2203005.

References

[1] V.N. Vapnik, An overview of statistical learning theory, IEEE Trans. Neural Netw. 10 (1999) 988–999.
[2] C. Cortes, V.N. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.
[3] H. Xu, C. Caramanis, S. Mannor, Robustness and regularization of support vector machines, J. Mach. Learn. Res. 10 (2009) 1485–1510.
[4] A. Ben-Tal, E. Hazan, T. Koren, S. Mannor, Oracle-based robust optimization via online learning, Oper. Res. 63 (2015) 628–638.
[5] X. Wang, N. Fan, P.M. Pardalos, Robust chance-constrained support vector machines with second-order moment information, Ann. Oper. Res. 263 (2018) 45–68.
[6] L. Tang, Y. Tian, W. Li, P. Pardalos, Valley-loss regular simplex support vector machine for robust multiclass classification, Knowl. Based Syst. 216 (2021) 106801.
[7] J. Ma, S. Zhou, D. Li, Robust multiclass least squares support vector classifier with optimal error distribution, Knowl. Based Syst. 215 (2021) 106652.
[8] P.K. Shivaswamy, C. Bhattacharyya, A.J. Smola, Second order cone programming approaches for handling missing and uncertain data, J. Mach. Learn. Res. 7 (2006) 1283–1314.
[9] T.B. Trafalis, S.A. Alwazzi, Support vector regression with noisy data: A second order cone programming approach, Int. J. Gen. Syst. 36 (2007) 237–250.
[10] G. Huang, S. Song, C. Wu, K. You, Robust support vector regression for uncertain input and output data, IEEE Trans. Neural Netw. Learn. Syst. 23 (2012) 1690–1700.
[11] M. Abaszade, S. Effati, Support vector regression with random output variable and probabilistic constraints, Iran. J. Fuzzy Syst. 14 (2017) 43–60.
[12] Y.H. Shao, C.N. Li, L.W. Huang, Z. Wang, N.Y. Deng, Y. Xue, Joint sample and feature selection via sparse primal and dual lssvm, Knowl. Based Syst. 185 (2019) 104915.
[13] X. Wang, S. Wang, Y. Du, Z. Huang, Minimum class variance multiple kernel learning, Knowl. Based Syst. 208 (2020) 106469.
[14] L. Reyzin, R.E. Schapire, How boosting the margin can also boost classifier complexity, in: Proc. 23rd Int. Conf. Mach. Learn. 2006, pp. 753–760.
[15] L. Breiman, Prediction games and arcing algorithms, Neural Comput. 11 (1999) 1493–1517.
[16] W. Gao, Z.H. Zhou, On the doubt about margin explanation of boosting, Artificial Intelligence 203 (2013) 1–18.
[17] T. Zhang, Z.H. Zhou, Optimal margin distribution machine, IEEE Trans. Knowl. Data Eng. 32 (2020) 1143–1156.
[18] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Proc. 5th Annu. Workshop Comput. Learn. Theory (COLT), 1992, pp. 144–152.
[19] B. Haasdonk, Feature space interpretation of svms with indefinite kernels, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 482–492.
[20] A. Astorino, A. Fuduli, Semisupervised spherical separation, Appl. Math. Model. 39 (2015) 6351–6358.
[21] J. Luo, S.C. Fang, Z. Deng, X. Guo, Soft quadratic surface support vector machine for binary classification, Asia Pac. J. Oper. Res. 33 (2016) 1650046.
[22] J. Luo, X. Yan, Y. Tian, Unsupervised quadratic surface support vector machine with application to credit risk assessment, European J. Oper. Res. 280 (2020) 1008–1017.
[23] Z. Gao, S.C. Fang, X. Gao, J. Luo, N. Medhin, A novel kernel-free least squares twin support vector machine for fast and accurate multi-class classification, Knowl. Based Syst. 226 (2021) 107123.
[24] J. Ye, Z. Yang, Z. Li, Quadratic hyper-surface kernel-free least squares support vector regression, Intell. Data Anal. 25 (2021) 265–281.
[25] J. Ye, Z. Yang, M. Ma, Y. Wang, X. Yang, ϵ-Kernel-free soft quadratic surface support vector regression, Inform. Sci. 594 (2022) 177–199.
[26] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola, V.N. Vapnik, Support vector regression machines, Adv. Neural Inf. Process. Syst. 9 (1997) 155–161.
[27] J. Ayala, M. Garcia-Torres, J. Noguera, F. Gomez-Vela, F. Divina, Technical analysis strategy optimization using a machine learning approach in stock market indices, Knowl. Based Syst. 225 (2021) 107119.

[28] J. Luo, T. Hong, S.C. Fang, Benchmarking robustness of load forecasting models under data integrity attacks, Int. J. Forecast. 34 (2018) 89–104.
[29] C. Condemia, D. Casillas-Pérez, L. Mastroeni, S. Jiménez-Fernádez, S. Salcedo-Sanz, Hydro-power production capacity prediction based on machine learning regression techniques, Knowl. Based Syst. 222 (2021) 107012.
[30] Y.W. Chang, C.J. Hsieh, K.W. Chang, M. Ringgaard, C.J. Lin, Training and testing low-degree polynomial data mappings via linear svm, J. Mach. Learn. Res. 11 (2010) 1471–1490.
[31] Y. Goldberg, M. Elhadad, SplitSVM: fast, space-efficient, non-heuristic, polynomial kernel computation for NLP applications, in: Proc. ACL-08: HLT, Short Papers, 2008, pp. 237–240.
[32] M. Sellathurai, S. Haykin, The separability theory of hyperbolic tangent kernels and support vector machines for pattern classification, in: Proc. IEEE Int. Conf. Acou. Speech Sig. Proc. (ICASSP), 1999, pp. 1021–1024.
[33] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, 2014, http://cvxr.com/cvx.
[34] K. Bache, M. Lichman, UCI machine learning repository, 2013, http://archive.ics.uci.edu/ml.
[35] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley and Sons, 2001.
[36] K. Severson, P. Attia, N. Jin, N. Perkins, et al., Data-driven prediction of battery cycle life before capacity degradation, Nature Energy 4 (2019) 383–391.
[37] P.M. Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the wasserstein metric: performance guarantees and tractable reformulations, Math. Program. 171 (2018) 115–166.
[38] L. Bai, Y.H. Shao, Z. Wang, C.N. Li, Clustering by twin support vector machine and least square twin support vector classifier with uniform output coding, Knowl. Based Syst. 163 (2019) 227–240.