ARTICLE IN PRESS

International Journal of Forecasting xxx (xxxx) xxx

Contents lists available at ScienceDirect

International Journal of Forecasting

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

A robust support vector regression model for electric load forecasting

Jian Luo [b], Tao Hong [c], Zheming Gao [a,*], Shu-Cherng Fang [d]

[a] College of Information Science and Engineering, Northeastern University, Shenyang, Liaoning 110819, China
[b] School of Management, Hainan University, Haikou, Hainan 570228, China
[c] Systems Engineering and Engineering Management Department, University of North Carolina at Charlotte, Charlotte, NC 28223, USA
[d] Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, Raliegh, NC 27695, USA

ARTICLE INFO

Keywords:
Cybersecurity
Electric load forecasting
Support vector regression
Data integrity attacks
Weight function

ABSTRACT

Electric load forecasting is a crucial part of business operations in the energy industry. Various load forecasting methods and techniques have been proposed and tested. With growing concerns about cybersecurity and malicious data manipulations, an emerging topic is to develop robust load forecasting models. In this paper, we propose a robust support vector regression (SVR) model to forecast the electricity demand under data integrity attacks. We first introduce a weight function to calculate the relative importance of each observation in the load history. We then construct a weighted quadratic surface SVR model. Some theoretical properties of the proposed model are derived. Extensive computational experiments are based on the publicly available data from Global Energy Forecasting Competition 2012 and ISO New England. To imitate data integrity attacks, we have deliberately increased or decreased the historical load data. Finally, the computational results demonstrate better accuracy of the proposed robust model over other recently proposed robust models in the load forecasting literature.

1. Introduction

Load forecasts are widely used across all segments of the power industry. Accurate load forecasts are crucial to the excellence of power system operations and planning, such as unit commitment, energy transfer scheduling, and load-frequency control (Hahn et al., 2009). In recent years, the information technologies such as Internet, communication networks, and computers have made the operations of power grids much more efficient than ever before. These technologies, however, have also enabled cyberattacks on power systems. Cybersecurity currently presents a serious challenge to the resilience of the power grid (Ericsson, 2010). The cyber attack on Ukraine's power grid (Perez, 2016), for instance, was a real threat to people's daily lives. Several other cyber attacks on power systems were discussed in (Hong & Hofmann, 2021). Data integrity attack is one form of cyber attacks. Hackers may access the supposedly protected data sets and inject misleading information to the grid measurements in a way such that the manipulations may not be easily detected by conventional operational practices. To prepare for data integrity attacks against load forecasting systems, it is imperative to develop robust load forecasting models.

Power companies count on accurate load forecasts to operate in a safe manner, to optimize operational costs, and to improve the reliability of distributional networks. In electricity markets, accurate load forecasts are also critical to support energy trading. While accurate load forecasts rely on the accurate historical information, data integrity attacks may contaminate the historical data as follows. Hackers may deliberately increase the historical

* Corresponding author.
E-mail address: gaozheming@ise.neu.edu.cn (Z. Gao).

https://doi.org/10.1016/j.ijforecast.2022.04.001
0169-2070/

Please cite this article as: J. Luo, T. Hong, Z. Gao et al., A robust support vector regression model for electric load forecasting. International Journal of Forecasting (2022), https://doi.org/10.1016/j.ijforecast.2022.04.001.

load values to lead to over-forecasts of electricity demand. Consequently, the unnecessary expenses on power generation, infrastructure maintenance and upgrade, and short sell of over-bought electricity reduce the overall economic efficiency. Hackers may also decrease the historical load values to result in under-forecasts. Consequently, the insufficient capacity expansion in the generation, transmission, and distribution systems may occur, which may lead to worse reliability indices and increase the risk for brownouts or even blackouts.

A rich literature on load forecasting has been developed during the past few decades (Hong & Fan, 2016; Weron, 2006). Most of the published studies focus on developing and implementing various load forecasting models, such as multiple linear regression (MLR) (Charlton & Singleton, 2014; Hong, Wilson, & Xie, 2014), artificial neural networks (ANN) (Hippert et al., 2001), support vector regression (SVR) (Chen et al., 2004), and fuzzy interaction regression (FIR) (Hong & Wang, 2014). The two Global Energy Forecasting Competitions, namely, GEFCom2012 and GEFCom2014 have also stimulated many novel ideas to tackle emerging problems such as hierarchical load forecasting and probabilistic load forecasting (Hong & Fan, 2016; Hong, Wilson, & Xie, 2014).

Several winners from the aforementioned competitions conducted the procedures of outlier detection and data cleansing before performing load forecasting (Charlton & Singleton, 2014; Xie & Hong, 2016). Some other papers touched on the anomaly detection with varying degrees of emphasis (Akouemo & Povinelli, 2016; Luo, Hong, & Fang, 2018; Yue et al., 2019). While most of the existing studies focused on small-scale random outliers or anomalies, how data integrity attacks may affect electric load forecasting has not been seriously investigated. As the first of its kind, the empirical study in Luo, Hong, and Yue (2018) benchmarked the robustness of four representative load forecasting models (i.e, MLR, ANN, SVR, and FIR). It clearly demonstrated that the performance of all four models w.r.t. forecast accuracy deteriorates dramatically as the level of malicious data integrity attacks increases.

Under data integrity attacks, a large portion of historical load data could be maliciously altered with large magnitudes by hackers, resulting in many observations deviating markedly from the normal levels. Consequently, the anomalies tend to greatly impact the commonly used least square estimators. To alleviate the impacts of anomalies from data attacks, the iteratively re-weighted least squares (IRLS) and L_1 regression are introduced in Luo et al. (2019) to reduce the impact of large residuals. Although the robust regression models in Luo et al. (2019) were more robust than the ones in Luo, Hong, and Fang (2018), none of them can generate accurate forecasts under large-scale data attacks, especially when the attack affects greater than 40% of the data.

The SVR model was used by the winning team in EUNITE Competition 2001 (Chen et al., 2004), but the progress of adopting SVR models to load forecasting slowed down after that. In Luo, Hong, and Yue (2018), the SVR model was shown to be more robust than the MLR, ANN, and FIR models. A possible reason is that the regression curve obtained by the SVR model is determined mainly by the underlying support vectors and thus less affected by the outliers and noise. To capture the nonlinear relationship between the load and its explanatory variables, nonlinear kernel functions are often adopted in SVR models for electric load forecasting (Ceperic et al., 2013; Hahn et al., 2009). A kernel function maps all data points from the original space to a higher dimensional feature space, where a hyperplane is generated to fit the mapped points. However, there is no universal rule to automatically choose a proper kernel function for a given data set. For electric load forecasting, the performance of SVR models relies heavily on the parameters selected for the kernel function. It may take a significant amount of computational time and effort to select a proper kernel function and its parameters. Moreover, the numerical issues related to the inverse of kernel matrices based on the large data sets used for load forecasting may substantially influence the forecast accuracy and computational efforts of the SVR model with a kernel, especially when the kernel matrix is singular. Hence, the employment of kernel function greatly limits the efficiency and accuracy of SVR models for electric load forecasting. In theory, any twice continuously differentiable nonlinear function has a Taylor approximation in quadratic form. Furthermore, the vanilla benchmark model in GEFCom2012 (Hong, Wilson, & Xie, 2014) includes the interactions between variables, of which some are also in the quadratic form. Therefore, we intend to propose a kernel-free quadratic surface SVR (QSSVR) by directly using a quadratic surface for an effective and efficient load forecasting in this paper.

The key feature of this paper is to propose a robust kernel-free weighted quadratic surface SVR (WQSSVR) model for load forecasting under data integrity attacks. The kernel-free QSSVR model directly utilizes a quadratic surface to fit the data points. We design a weight function to evaluate the relative importance of each data point in the load history to reduce the influence of manipulated observations. To imitate the data integrity attacks targeting economic losses or system blackouts of modern power grids, we conduct computational experiments in which the majority part of historical load data is deliberately decreased or increased following different normal or uniform distributions. Finally, the proposed WQSSVR model demonstrates superior accuracy compared to robust regression models (IRLS and L_1 regression) proposed in Luo et al. (2019) and two other competing models (i.e., MLR and SVR with Gaussian kernel) in the benchmark study of Luo, Hong, and Fang (2018).

The rest of the paper is arranged as follows: Section 2 briefly reviews some related SVR models for electric load forecasting, Section 3 proposes a robust kernel-free WQSSVR model for load forecasting, Section 4 conducts the computational experiments on electric load forecasting under various types of data attacks to compare the accuracy of the proposed model with its counterparts, Section 5 shows some additional numerical tests of the proposed model on a total of 30 data sets in different zones, for the discussion of robustness, and finally Section 6 concludes this paper.

2. Support vector regression models for electric load forecasting

Based on finding of the generalized portrait of patterns, the support vector (SV) algorithm (previously called the generalized portrait algorithm) was first developed in Russia (Vapnik, 1982; Vapnik & Lerner, 1963). In its current form, the support vector machine (SVM) was largely developed and extensively applied at AT&T Bell Laboratories by Vapnik and co-workers (Cortes & Vapnik, 1995; Vapnik, 1995). To overcome the drawbacks induced by employing kernels, a kernel-free quadratic surface SVM (QSSVM) model was proposed to directly utilize one quadratic surface for an effective nonlinear classification in Luo et al. (2016). Based on the QSSVM model, a semi-supervised QSSVM model with fuzzy set (Tian et al., 2017) and an unsupervised QSSVM model (Luo et al., 2020) have been developed for effective mislabeled classification and unsupervised classification, respectively. As a counterpart of the SVM model for regression problems, the SVR model has attained an excellent performance in solving many real-world forecasting problems, such as forecasting the electric loads (Chen et al., 2004) and loss given defaults (Yao et al., 2015). Electric load forecasting is particularly important in the power industry. The first notable success of the SVR model for electric load forecasting was shown at the EUNITE competition 2001 (Chen et al., 2004).

In Chen et al. (2004), a training data set $\left\{\left((x^i)^T, y^i\right)^T, i = 1, \ldots, n\right\}$ is obtained, where the input $x^i = \left(x_1^i, x_2^i, \ldots, x_m^i\right)^T \in R^m$ includes the calendar attributes, the temperature attribute, and their interactions, and the output $y^i \in R$ denotes the load value of the ith day. Given this training data set, the classic SVR model aims to find the parameters $w \in R^m$ and $c \in R$ of a fitting hyperplane $y = w^T x + c$, so that $w^T x^i + c$ is close to y^i for each training point. SVR assumes that the deviation tolerance between $w^T x + c$ and y is at most δ, which indicates that the loss is calculated when $\left|w^T x + c - y\right| > \delta$. In other words, if the training point falls within the tube $|y - (w^T x + c)| \le \delta$, which is called the insensitive tube (Chen et al., 2004), the training point is considered to be accurately predicted. Hence, the classic SVR model is formulated in the following manner (Chen et al., 2004):

$$\begin{aligned} \min_{w,c,\xi}\ & \frac{1}{2} w^T w + C_p \sum_{i=1}^{n} \xi_i \\ s.t.\ \ & \delta + \xi_i \ge y^i - (w^T x^i + c), i = 1, 2, \ldots, n, \\ & y^i - (w^T x^i + c) \ge -\delta - \xi_i, i = 1, 2, \ldots, n, \\ & \xi_i \ge 0, i = 1, 2, \ldots, n. \end{aligned} \tag{1}$$

where $\xi_i, i = 1, 2, \ldots, n$ are the errors of training points outside the insensitive tube, $\xi \triangleq (\xi_1, \xi_2, \ldots, \xi_n)^T$ for convenience in this paper, $\delta, C_p > 0$ are the given parameters. Similar to the derivation of margin between the two classes in SVM (Cortes & Vapnik, 1995; Vapnik, 1995), the vertical distance between the hyperplanes $y-(w^T x+c) = \delta$ and $y - (w^T x + c) = -\delta$ can be calculated as $\frac{\delta}{w^T w+1}$. Minimizing $\frac{1}{2} w^T w$(the same as the first term in the objective of SVM) and the constraints in the model imply that as many training points are put in the insensitive tube as possible (Chen et al., 2004). If the point is not in the tube, there is a related error ξ_i to be minimized in the objective of the SVR model. From another perspective, minimizing $\frac{1}{2} w^T w$ indicates the number of zero element in $w \in R^m$ is maximized so that the complexity of fitting hyperplane $y = w^T x + c$ is minimized (Ceperic et al., 2013). Hence, the SVR model fits the training data by minimizing both the sum of errors of training points $\sum_{i=1}^{n} \xi_i$ and the regularization term $\frac{1}{2} w^T w$ (Ceperic et al., 2013; Chen et al., 2004).

For nonlinear fitting, all training points are first mapped to a higher dimensional space by a nonlinear kernel function, where a hyperplane is found to fit the mapped points, similar to the idea in the classic SVR. Below is the SVR model with a commonly used Gaussian kernel (denoted as "SVR_Gau" in this paper) (Chen et al., 2004):

$$\begin{aligned} \min_{w,c,\xi}\ & \frac{1}{2} w^T w + C_p \sum_{i=1}^{n} \xi_i \\ s.t.\ \ & \delta + \xi_i \ge y^i - (w^T \phi(x^i) + c), i = 1, 2, \ldots, n, \\ & y^i - (w^T \phi(x^i) + c) \ge -\delta - \xi_i, i = 1, 2, \ldots, n, \\ & \xi_i \ge 0, i = 1, 2, \ldots, n. \end{aligned} \tag{2}$$

Here, $\delta, C_p > 0$ are the given parameters and $\phi: R^m \to R^d$(where $d > m$) is the Gaussian kernel function. These SVR models won the competition in 2001 organized by the EUNITE network (Chen et al., 2004). Since then, SVR has evolved to some variants and emerged as a relatively new technique for load forecasting (Ghelardoni et al., 2013).

3. A weighted quadratic surface SVR model

In this section, we first introduce the underlying variables for electric load forecasting. We then introduce a kernel-free QSSVR model by directly utilizing a quadratic surface for regression. By incorporating the weights of training points, we propose a kernel-free WQSSVR model. Finally, some theoretical properties of the proposed WQSSVR model are derived.

3.1. The vanilla model

Thousands of models have been published in the load forecasting literature (Hong, 2010). As a frequently cited model, the following vanilla model is utilized in GEFCom2012 (Hong, Wilson, & Xie, 2014) to benchmark the load forecast accuracy:

$$\begin{aligned} E(y_l) = {} & r_0 + r_1 x_{tr} + r_2 x_h + r_3 x_w + r_4 x_m + r_5 x_t + r_6 (x_t)^2 \\ & + r_7 (x_t)^3 + r_8 x_{h*} x_w + r_9 x_{t*} x_h + r_{10} (x_t)_*^2 x_h \\ & + r_{11} (x_t)_*^3 x_h + r_{12} x_{t*} x_m + r_{13} (x_t)_*^2 x_m + r_{14} (x_t)_*^3 x_m, \end{aligned} \tag{3}$$

where y_l is a variable of electric loads; x_{tr} is a variable of the increasing integers representing a linear trend; x_h is a vector including 24 dummy variables representing 24 h in a day; x_w is a vector including 7 dummy variables

representing 7 days in a week; x_m is a vector including 12 dummy variables representing 12 months in a year; and x_t is a variable representing the coincident temperature. Hence, this vanilla model has totally 289 variables, which works effectively for electric load forecasting, (Hong & Fan, 2016; Hong, Pinson, & Fan, 2014; Hong, Wilson, & Xie, 2014). In this paper, we intend to utilize these variables in the vanilla model as the underlying variables for the proposed WQSSVR model.

3.2. Quadratic surface SVR model

To reach a high accuracy using SVR with a kernel, forecasters may have to invest a significant amount of computational time and effort to select a proper kernel function and its kernel parameters. When the kernel matrix is singular, the inverse of kernel matrices based on the load forecasting data, which are typically large, can heavily influence the forecast accuracy and computational efforts of the SVR with a kernel. To save computational time and improve the classification accuracy, the kernel-free SVM models proposed in Luo et al. (2016), Tian et al. (2017) and Luo et al. (2020) directly use the quadratic surfaces for classification without using any kernel. Similarly, for nonlinear fitting of training data set $\left\{\left((x^i)^T, y^i\right)^T, i = 1, \ldots, n\right\}$, where $x^i = \left(x_1^i, x_2^i, \ldots, x_m^i\right)^T \in R^m$ and $y^i \in R$, we introduce the QSSVR model to find the parameters (W, b, c) of a quadratic surface

$$y = q(x) \triangleq \frac{1}{2}x^T W x + b^T x + c,$$

$$W = W^T = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{12} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ w_{1m} & w_{2m} & \cdots & w_{mm} \end{bmatrix} \in R^{m\times m},$$

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \in R^m, c \in R,$$

which fits the n training points without utilizing any kernel function. And let $w_{ij} = 0, i \neq j$ for high efficiency on large hourly load data with 289 independent variables.

Analogous to the classic SVR models, the goal of QSSVR model is to generate one "tube" and then try to include as many training points as possible in this "tube". More specifically, we first ignore the errors of the training points inside the tube $|y - (0.5x^T Wx + b^T x + c)| \leq \delta$ for a given δ. To include many training points in this tube, we try to maximize the margin between the upper and the lower bounds of the tube. This objective can be characterized by maximizing the relative geometrical margin at each training point. In Appendix A, the relative geometrical margin at point $((x^i)^T, y^i)^T$ is defined and approximated as $\frac{\delta}{\|(Wx^i+b,-1)\|_2}$. Hence, one objective of QSSVR can be formulated as minimizing $\sum_{i=1}^n \|(Wx^i+b,-1)\|_2^2 = \sum_{i=1}^n \|Wx^i + b\|_2^2 + n$, which is equivalent to minimizing $\sum_{i=1}^n \|Wx^i + b\|_2^2$. This term in the objective can also be regarded as one regularization term to avoid over-fitting, which is similar to the first regularization term in the QSSVM (Luo et al., 2016) to maximize the margin between two classes. Another objective of QSSVR is to minimize the deviations of training points with errors larger than δ. Thus, the QSSVR model can be formulated as

$$\begin{aligned} \min_{W,b,c,\xi} \ & \sum_{i=1}^n \|Wx^i + b\|_2^2 + C_p \sum_{i=1}^n \xi_i \\ s.t. \ & \delta + \xi_i \geq y^i - (\frac{1}{2}(x^i)^T W x^i + b^T x^i + c), i = 1, 2, \ldots, n, \\ & y^i - (\frac{1}{2}(x^i)^T W x^i + b^T x^i + c) \geq -\delta - \xi_i, i = 1, 2, \ldots, n, \\ & \xi_i \geq 0, i = 1, 2, \ldots, n. \end{aligned} \tag{4}$$

where $\delta, C_p > 0$ are the given parameters and the constant $C_p > 0$ determines the trade-off between the relative geometrical margins and the amount up to which the deviations larger than δ are tolerated.

3.3. Weighted quadratic surface SVR model

When a load forecasting system is under data integrity attacks, the training data may be contaminated. Consequently, the QSSVR model may suffer the loss of forecast accuracy, because the model assumes that every training point makes the same contribution to parameter estimation. To properly address these issues, we design the following weight function to efficiently calculate the weights of all training points $\left\{\left((x^i)^T, y^i\right)^T, i = 1, \ldots, n\right\}$ for characterizing their relative contributions:

$$\beta_i = e^{-|u_i|}, i = 1, 2, \ldots, n$$

where $u_i = |\gamma_i - \overline{\gamma}| / MED$, $\gamma_i \triangleq \left|y^i - \hat{y}^i\right|$ ($\hat{y}^i = w^T x^i + c$, where w and c are generated by using the training points and L_1 regression Luo et al., 2019), $\overline{\gamma}$ is the median of $\gamma_i, i = 1, 2, \ldots, n$, MED is the median of $|\gamma_i - \overline{\gamma}|, i = 1, 2, \ldots, n$. The weight $\beta_i \in [0, 1]$ is calculated by using the exponential function, which is similar to Welsch function in the robust M-estimates (Basu & Paliwal, 1989). The weight of the point with smaller $|u_i|$ is larger than that of the point with larger $|u_i|$. As $|u_i|$ increases from 0 to infinity, β_i first decreases quickly from 1 and then decreases slowly to 0. Hence, if a training point is manipulated, then $|u_i|$ of this point becomes large and the weight for this point is near 0. The weakness of this weight function is that the weights for some points without attacks (which leads to small $|u_i|$) may not be large since β_i decreases quickly from 1 at the first stage of increasing $|u_i|$.

To reduce the contributions of attacked points, we propose a WQSSVR model by incorporating the calculated weights $\beta_i, i = 1, \ldots, n$ into the two terms of the

objective function of QSSVR as the following:

$$\begin{aligned}\min_{W,b,c,\xi}\ & \sum_{i=1}^{n}\beta_i\|Wx^i+b\|_2^2+C_p\sum_{i=1}^{n}\beta_i\xi_i\\ s.t.\quad & \delta+\xi_i\geq y^i-(\frac{1}{2}(x^i)^TWx^i+b^Tx^i+c),\ i=1,2,\ldots,n,\\ & y^i-(\frac{1}{2}(x^i)^TWx^i+b^Tx^i+c)\geq-\delta-\xi_i,\ i=1,2,\ldots,n,\\ & \xi_i\geq 0,\ i=1,2,\ldots,n.\end{aligned}\tag{5}$$

where $\delta, C_p > 0$ are the given parameters. Similarly, the terms $\beta_i\xi_i$ and $\beta_i\left\|Wx^i+b\right\|_2^2$ can be deemed as measuring the error ξ_i and $\left\|Wx^i+b\right\|_2^2$(i.e., the relative geometrical margin of point) with the weight β_i, respectively. If the point x^i is more likely to be an attacked point, which indicates that x^i is less important, the related β_i is expected to be small to reduce the effect of ξ_i and $\left\|Wx^i+b\right\|_2^2$ in the WQSSVR model. Hence, the main advantage of introducing the weights into the QSSVR model is to reduce the contributions of attacked points to minimize the errors and maximize the relative geometrical margins of training points. Moreover, in the objective of the proposed WQSSVR model, minimizing the regularization term $\sum_{i=1}^{n}\beta_i\|Wx^i+b\|_2^2$ and the term of errors $\sum_{i=1}^{n}\beta_i\xi_i$ help avoid over-fitting and under-fitting the training data, respectively.

Since W is a diagonal matrix, the WQSSVR model (5) can be equivalently reformulated for one smaller-sized WQSSVR' optimization model by reducing the matrix variable W to a smaller-sized vector variable, similar to the reformulations in Dagher (2008) and Luo et al. (2016) as the following:

$$\begin{aligned}\min_{z,c,\xi}\ & z^TGz+C_p\sum_{i=1}^{n}\beta_i\xi_i\\ s.t.\quad & \delta+\xi_i\geq y^i-(s_i^Tz+c)\geq-\delta-\xi_i,\ i=1,2,\ldots,n,\\ & \xi_i\geq 0,\ i=1,2,\ldots,n.\end{aligned}\tag{6}$$

where $\delta, C_p > 0$ are the given parameters. This is derived in detail in Appendix B with the definitions of z, s_i, and G.

3.4. Theoretical properties of WQSSVR

For any formulated model in the form of an optimization problem, it is possible that there is no feasible solution. In this subsection, we investigate the existence of optimal solution to the proposed WQSSVR′ model (6) as follows, which is similar to proving the theoretical properties of soft QSSVM model in Luo et al. (2016).

Theorem 1. *For any given training data set* $((x^i)^T, y^i)^T$, $i = 1,\ldots,n$ *and* $C_p > 0$, *there exists an optimal solution to the WQSSVR′ model with a finite objective value.*

Proof. Take any $(\widetilde{z}^T,\widetilde{c})^T$ and let $\widetilde{\xi}_i \triangleq \max 0, |y^i-(s_i^T\widetilde{z}+\widetilde{c})|-\delta$, $i = 1,\ldots,n$. Then, $\widetilde{\xi}_i \geq 0$. Moreover, for $i = 1,\ldots,n$, $\widetilde{\xi}_i = 0$ if $|y^i-(s_i^T\widetilde{z}+\widetilde{c})|-\delta \leq 0$, and $\widetilde{\xi}_i = |y^i-(s_i^T\widetilde{z}+\widetilde{c})|-\delta$ if $|y^i-(s_i^T\widetilde{z}+\widetilde{c})|-\delta > 0$. Hence, $|y^i-(s_i^T\widetilde{z}+\widetilde{c})|-\delta \leq \widetilde{\xi}_i$ and $\widetilde{\xi}_i \geq 0$, which infers that $(\widetilde{z}^T,\widetilde{c},\widetilde{\xi}^T)^T$ is one feasible solution of the WQSSVR′ model. Notice that the objective function is continuous and the feasible domain is a closed convex set defined by linear inequalities. Moreover, for any z and $\xi_i \geq 0$, $i = 1,\ldots,n$, $z^TGz+C_p\sum_{i=1}^{n}\beta_i\xi_i = \sum_{i=1}^{n}(\beta_i\|H_iz\|_2^2+C_p\beta_i\xi_i) \geq 0$ (where H_i is defined in Appendix B), which indicates that the objective value is bounded below by 0 over the feasible domain. Therefore, there exists an optimal solution to WQSSVR′ model with a finite objective value.

Let $F^* \triangleq (z^T,c,\xi^T)^T \in \mathbb{R}^{2m}\times\mathbb{R}^1\times\mathbb{R}^n|(z^T,c,\xi^T)^T$ is an optimal solution to the WQSSVR′ model}, then F^* is a set including all optimal solutions to the WQSSVR′ model. Similar to proving the theoretical properties of soft QSSVM model in Luo et al. (2016), we can verify that the optimal solution of WQSSVR′ model is unique with respect to the variable z if G is positive definite. Moreover, if G is positive definite, there exist constants c and $\bar{c}$ such that $c \leq c \leq \bar{c}$, for any $(z^T,c,\xi^T)^T \in F^*$. Therefore, for any given training data set, if G is positive definite, the main shape of the fitting quadratic surface is uniquely determined by the optimal solution of WQSSVR′ model.

Notice that if the matrix G in WQSSVR' model is only positive semi-definite, we can always append a perturbation such that the matrix $G+\varepsilon I$ ($\varepsilon > 0$, I is the identity matrix) becomes positive definite. Then, consider the following perturbed WQSSVR-eps model:

$$\begin{aligned}\min_{z,c,\xi}\ & z^T(G+\varepsilon I)z+C_p\sum_{i=1}^{n}\beta_i\xi_i\\ s.t.\quad & \delta+\xi_i\geq y^i-(s_i^Tz+c)\geq-\delta-\xi_i,\ i=1,2,\ldots,n,\\ & \xi_i\geq 0,\ i=1,2,\ldots,n.\end{aligned}\tag{7}$$

where $\delta, C_p > 0$ are the given parameters. Similar to the proof of Theorem 1, the WQSSVR-eps model has at least one optimal solution with a finite optimal value since the constraints of WQSSVR′ model are the same as those of WQSSVR-eps model. Let $((z^\varepsilon)^T,c^\varepsilon,(\xi^\varepsilon)^T)^T$ be an optimal solution of WQSSVR-eps model, then $((z^\varepsilon)^T,c^\varepsilon,(\xi^\varepsilon)^T)^T$ is feasible to the WQSSVR′ model. Moreover, similar to proving the properties of soft QSSVM model in Luo et al. (2016), the WQSSVR′ model and its perturbed WQSSVR-eps model can be related by the next Lemma 2 and Theorem 3.

Lemma 2. *For any given training data set* $((x^i)^T, y^i)^T$, $i = 1,\ldots,n$ *and* $C_p > 0$, *if the optimal value of WQSSVR′ model is* v *and the optimal value of WQSSVR-eps model is* v_ε, *for given* $\varepsilon > 0$, *then* $v_\varepsilon \to v$ *as* $\varepsilon \to 0$.

Proof. Let $(\widetilde{z}^T,\widetilde{c},\widetilde{\xi}^T)^T \in F^*$. Then we prove this lemma by considering two separate cases as follows.

Case 1: If $\|\widetilde{z}\| \neq 0$, for $((z^\varepsilon)^T,c^\varepsilon,(\xi^\varepsilon)^T)^T$ and any $\eta > 0$, there exists $\varepsilon_0 \triangleq \frac{\eta}{(\widetilde{z}^T)(\widetilde{z})}$ such that when $0 < \varepsilon < \varepsilon_0$, $v \leq (z^\varepsilon)^TGz^\varepsilon+C_p\sum_{i=1}^{n}\beta_i\xi_i^\varepsilon \leq (z^\varepsilon)^TGz^\varepsilon+C_p\sum_{i=1}^{n}\beta_i\xi_i^\varepsilon+\varepsilon(z^\varepsilon)^Tz^\varepsilon = v_\varepsilon$ since $((z^\varepsilon)^T,c^\varepsilon,(\xi^\varepsilon)^T)^T$ is optimal to the WQSSVR-eps model and feasible to the WQSSVR′ model, $v_\varepsilon \leq \widetilde{z}^T(G+\varepsilon I)\widetilde{z}+C_p\sum_{i=1}^{n}\beta_i\widetilde{\xi}_i$ since $(\widetilde{z}^T,\widetilde{c},\widetilde{\xi}^T)^T$ is feasible to the WQSSVR-eps model, $\widetilde{z}^T(G+\varepsilon I)\widetilde{z}+C_p\sum_{i=1}^{n}\beta_i\widetilde{\xi}_i = \widetilde{z}^TG\widetilde{z}+C_p\sum_{i=1}^{n}\beta_i\widetilde{\xi}_i+\varepsilon(\widetilde{z})^T(\widetilde{z}) = v+\varepsilon(\widetilde{z})^T(\widetilde{z}) < v+$

$\varepsilon_0(\widetilde{z})^T(\widetilde{z}) = v + \eta$due to $(\widetilde{z}^T, \widetilde{c}, \widetilde{\xi}^T)^T \in F^*$. Hence, these derivations infer that $|v_\varepsilon - v| < \eta$.

Case 2: If $\|\widetilde{z}\| = 0$, similar to that in Case 1, we have $v \le (z^\varepsilon)^T G z^\varepsilon + C_p \sum_{i=1}^n \beta_i \xi_i^\varepsilon \le v_\varepsilon \le \widetilde{z}^T(G+\varepsilon I)\widetilde{z} + C_p \sum_{i-1}^n \beta_i \widetilde{\xi}_i = v + \varepsilon(\widetilde{z})^T(\widetilde{z}) = v$, which infers that $v = v_\varepsilon$.

Therefore, $v_\varepsilon \to v$ as $\varepsilon \to 0$. □

Remark. For any given $C_p > 0$ and $0 < \varepsilon_1 < \varepsilon_2$, let $((z^{\varepsilon_1})^T, c^{\varepsilon_1}, (\xi^{\varepsilon_1})^T)^T$ and v_{ε_1} be the optimal solution and optimal value of the following WQSSVR-eps1 model, respectively:

$$\min_{z,c,\xi} \; z^T(G+\varepsilon_1 I)z + C_p \sum_{i=1}^n \beta_i \xi_i$$

$$s.t. \quad \delta + \xi_i \ge y^i - (s_i^T z + c) \ge -\delta - \xi_i, \; i = 1, 2, \ldots, n,$$

$$\xi_i \ge 0, \; i = 1, 2, \ldots, n.$$

Moreover, let $((z^{\varepsilon_2})^T, c^{\varepsilon_2}, (\xi^{\varepsilon_2})^T)^T$ and v_{ε_2} be the optimal solution and optimal value of the following WQSSVR-eps2 model, respectively:

$$\min_{z,c,\xi} \; z^T(G+\varepsilon_2 I)z + C_p \sum_{i=1}^n \beta_i \xi_i$$

$$s.t. \quad \delta + \xi_i \ge y^i - (s_i^T z + c) \ge -\delta - \xi_i, \; i = 1, 2, \ldots, n,$$

$$\xi_i \ge 0, \; i = 1, 2, \ldots, n.$$

and the WQSSVR-eps1 and WQSSVR-eps2 models have the same constraints. Then we have $v_{\varepsilon_1} \le (z^{\varepsilon_2})^T G z^{\varepsilon_2} + \varepsilon_1 (z^{\varepsilon_2})^T (z^{\varepsilon_2}) + C_p \sum_{i=1}^n \beta_i \xi_i^{\varepsilon_2} < (z^{\varepsilon_2})^T G z^{\varepsilon_2} + \varepsilon_2 (z^{\varepsilon_2})^T (z^{\varepsilon_2}) + C_p \sum_{i=1}^n \beta_i \xi_i^{\varepsilon_2} = v_{\varepsilon_2}$. Therefore, the sequence $\{v_\varepsilon\}$ monotonically decreases to v as $\varepsilon \to 0$.

Theorem 3. *For any given training data set $((x^i)^T, y^i)^T$, $i = 1, \ldots, n$ and $C_p > 0$, if the sequence $\left\{((z^\varepsilon)^T, c^\varepsilon, (\xi^\varepsilon)^T)^T\right\}$ converges to $((z^0)^T, c^0, (\xi^0)^T)^T$ as $\varepsilon \to 0$, then $((z^0)^T, c^0, (\xi^0)^T)^T \in F^*$ and $(z^0)^T z^0 \le z^T z$, for any $(z^T, c, \xi^T)^T \in F^*$.*

Proof. While $\left\{((z^\varepsilon)^T, c^\varepsilon, (\xi^\varepsilon)^T)^T\right\} \to ((z^0)^T, c^0, (\xi^0)^T)^T$ as $\varepsilon \to 0$, $((z^0)^T, c^0, (\xi^0)^T)^T$ is feasible to the WQSSVR′ model since the constraints of WQSSVR′ model are the same as those of WQSSVR-eps model. By Lemma 2, $v_\varepsilon \to v$ as $\varepsilon \to 0$. Then we have $((z^0)^T, c^0, (\xi^0)^T)^T \in F^*$. For any $\varepsilon > 0$ and any $(z^T, c, \xi^T)^T \in F^*$, it should be noted that $(z^T, c, \xi^T)^T$ is feasible to the WQSSVR-eps model and $((z^\varepsilon)^T, c^\varepsilon, (\xi^\varepsilon)^T)^T$ is optimal to the WQSSVR-eps model. Hence, we have $(z^\varepsilon)^T(G+\varepsilon I)z^\varepsilon + C_p \sum_{i=1}^n \beta_i \xi_i^\varepsilon = (z^\varepsilon)^T G z^\varepsilon + \varepsilon (z^\varepsilon)^T z^\varepsilon + C_p \sum_{i=1}^n \beta_i \xi_i^\varepsilon \le z^T(G+\varepsilon I)z + C_p \sum_{i=1}^n \beta_i \xi_i = z^T G z + \varepsilon z^T z + C_p \sum_{i=1}^n \beta_i \xi_i$.

Since $(z^T, c, \xi^T)^T \in F^*$, we have $z^T G z + C_p \sum_{i=1}^n \beta_i \xi_i \le (z^\varepsilon)^T G z^\varepsilon + C_p \sum_{i=1}^n \beta_i \xi_i^\varepsilon$. Consequently, $(z^\varepsilon)^T G z^\varepsilon + \varepsilon (z^\varepsilon)^T z^\varepsilon + C_p \sum_{i=1}^n \beta_i \xi_i^\varepsilon \le z^T G z + \varepsilon z^T z + C_p \sum_{i=1}^n \beta_i \xi_i \le (z^\varepsilon)^T G z^\varepsilon + C_p \sum_{i=1}^n \beta_i \xi_i^\varepsilon + \varepsilon z^T z$, which infers that $(z^\varepsilon)^T z^\varepsilon \le z^T z$ for any $\varepsilon > 0$. Hence, as $\varepsilon \to 0$, $(z^0)^T z^0 \le (z)^T z$ for any $(z^T, c, \xi^T)^T \in F^*$ □

For any training data set, G is positive semi-definite, and G can be set as positive definite in order to derive some good properties of the WQSSVR′ model. If G is not positive definite by default, we can always replace G by the positive definite matrix $G + \varepsilon I$ ($\varepsilon > 0$, I is the identity matrix). Then, by Theorem 3, the perturbed WQSSVR-eps model (7) with a sufficiently small $\varepsilon > 0$ can be solved to generate a fitting quadratic surface, which is an optimal solution of the WQSSVR′ model. Therefore, G is considered as positive definite when deriving the following properties of the WQSSVR′ model.

We can formulate the dual problem of the WQSSVR' model as follows. First, the Lagrangian function is written as below by introducing three groups of dual variables $\alpha \triangleq (\alpha_1, \alpha_2, \ldots, \alpha_n)^T$, $\hat{\alpha} \triangleq (\hat{\alpha}_1, \hat{\alpha}_2, \ldots, \hat{\alpha}_n)^T$, $\mu \triangleq (\mu_1, \mu_2, \ldots, \mu_n)^T$:

$$\begin{aligned} &L(z, c, \xi, \alpha, \hat{\alpha}, \mu) \\ &= z^T G z + C_p \sum_{i=1}^n \beta_i \xi_i + \sum_{i=1}^n \alpha_i \left(s_i^T z + c - y^i - \delta - \xi_i\right) \\ &\quad + \sum_{i=1}^n \hat{\alpha}_i \left(y^i - s_i^T z - c - \delta - \xi_i\right) - \sum_{i=1}^n \mu_i \xi_i \end{aligned} \tag{8}$$

the first-order partial derivative of Lagrangian function be 0, the following formulas can be obtained:

$$\frac{\partial L}{\partial z} = 0 \Rightarrow z = \frac{1}{2} \sum_{i=1}^n \left(\hat{\alpha}_i - \alpha_i\right) G^{-1} s_i,$$

$$\frac{\partial L}{\partial c} = 0 \Rightarrow \sum_{i=1}^n \left(\hat{\alpha}_i - \alpha_i\right) = 0,$$

$$\frac{\partial L}{\partial \xi} = 0 \Rightarrow C_p \beta_i = \hat{\alpha}_i + \alpha_i + \mu.$$

Finally, to replace the primal variables $(z^T, c, \xi^T)^T$ with the dual variables $(\alpha^T, \hat{\alpha}^T)^T$, we replace the corresponding variables in the formula (8) with the above formulas to formulate the Lagrangian dual problem of WQSSVR' model as the following D-WQSSVR' model:

$$\begin{aligned} \max_{\alpha,\hat{\alpha}} \; &\sum_{i=1}^n \left[\left(\hat{\alpha}_i - \alpha_i\right) y^i - \delta \left(\hat{\alpha}_i + \alpha_i\right)\right] \\ &- \frac{1}{4} \left(\sum_{i=1}^n \left(\hat{\alpha}_i - \alpha_i\right) s_i\right)^T G^{-1} \left(\sum_{i=1}^n \left(\hat{\alpha}_i - \alpha_i\right) s_i\right) \end{aligned} \tag{9}$$

$$s.t. \quad \sum_{i=1}^n \left(\hat{\alpha}_i - \alpha_i\right) = 0,$$

$$0 \le \alpha_i + \hat{\alpha}_i \le C_p \beta_i.$$

Notice that the WQSSVR′ model is a linearly constrained quadratic minimization problem (Boyd & Vandenberghe, 2004). And the matrix G in the objective of WQSSVR′ model is positive semi-definite so that the objective function is convex. The constraint functions are linear. A quadratic minimization problem with a convex objective function and linear constraint functions is a linearly constrained convex quadratic programming problem. Moreover, the dual of a convex quadratic programming problem is convex (Boyd & Vandenberghe, 2004). Then both of WQSSVR' and D-WQSSVR' models are convex linearly constrained quadratic programming problems. And there is no duality gap between WQSSVR' model (8) and D-WQSSVR' model (9) by the strong duality theory (Boyd & Vandenberghe, 2004). Hence, the optimal conditions for these convex models are the following KKT

conditions:

$$\begin{aligned}&\alpha_i(s_i^T z + c - y^i - \delta - \xi_i) = 0,\\ &\hat{\alpha}_i(y^i - s_i^T z - c - \delta - \xi_i) = 0,\\ &\left(C_p\beta_i - \alpha_i - \hat{\alpha}_i\right)\xi_i = 0,\\ &\alpha_i \cdot \hat{\alpha}_i = 0.\end{aligned} \tag{10}$$

From these KKT optimal conditions, the optimal solutions $((z^*)^T, c^*, (\xi^*)^T)^T$ and $\left(\alpha_i^*, \hat{\alpha}_i^*, i = 1, \ldots, n\right)$ of WQSSVR' and D-WQSSVR' models can be obtained, respectively. For each training point, $-\delta - \xi_i \le y^i - s_i^T z - c \le \delta + \xi_i$ and only one of the two inequalities can be the equality since $\delta > 0$ and $\xi_i \ge 0$. Hence, from the KKT conditions (10), at least one of α_i and $\hat{\alpha}_i$ would be 0 for each training points, and $\alpha_i = \hat{\alpha}_i = 0$ for each point falling within the tube since $-\delta - \xi_i < y^i - s_i^T z - c < \delta + \xi_i$. Moreover, from KKT conditions (10), the following result can be obtained to find the primal optimal solution z^* from the dual optimal solution $\left(\alpha_i^*, \hat{\alpha}_i^*, i = 1, \ldots, n\right)$,by Lagrangian duality theory (Boyd & Vandenberghe, 2004):

$$z^* = \frac{1}{2}\sum_{i=1}^{n}\left(\hat{\alpha}_i^* - \alpha_i^*\right) G^{-1} s_i \tag{11}$$

Hence, from the optimal conditions (10) and formula (11), only the training points falling outside the tube (i.e., either α_i^*or $\hat{\alpha}_i^*$ would be non-zero) contribute to the shape of fitting quadratic surface (i.e., given by z^*) so that these points are called the support vectors. Moreover, by utilizing any support vector (e.g., the jth training point), the intercept parameter can be calculated as the following Eq. (12):

$$c^* = y_j + \delta - \frac{1}{2}\sum_{i=1}^{n}\left(\hat{\alpha}_i^* - \alpha_i^*\right) G^{-1} s_i^T s_j \tag{12}$$

Therefore, from formulas (11) and (12), the optimal solution of WQSSVR′ model is the expansion of support vectors. Then the fitting quadratic surface obtained by the WQSSVR model is determined mainly by the support vectors. Notice that this WQSSVR model can be solved efficiently by a primal–dual interior-point method, which has a simple structure, excellent theoretical properties, and good practical performance for convex quadratic programming (Vanderbei, 1999; Wright, 1997).

4. Experiments and results

The experiments in this section are based on data from GEFCom2012 (Hong, Wilson, & Xie, 2014). We first present the benchmark performance of five models using the original and uncontaminated data. We then present the results of the same five models under two data attack scenarios, with one increasing the historical load values and the other with decreasing the historical load values.

4.1. Computational experiments

The GEFCom2012 data for the load forecasting track includes 3.5 years of hourly load and temperature information for 21 zones, where the 21st zone is the sum of the other 20 zones. Following the practices reported in Hong, Wilson, and Xie (2014) and Hong and Wang (2014), we picked two years (2005 and 2006) and one year (2007) of hourly load and temperature information as the training and testing periods, respectively. The vanilla benchmark model is effective for forecasting the loads in the residential zones, which represent the majority zones of GEFCom2012 data excluding zones 4 and 9.

For fair comparisons, the iteratively re-weighted least squares (IRLS) model with the weight function of "bisquare" function (denoted as 'IRLS_bis') (Luo et al., 2019), L_1 regression (denoted as 'L_1'), MLR, SVR_Gau, and WQSSVR models share exactly the same variables as described in the vanilla model. All computational experiments were performed using MATLAB (R2019a) software on a desktop equipped with an Intel Xeon Processor 2.99 GHz CPU, 31.3 GB usable RAM and Microsoft Windows 10 Enterprise. The IRLS_bis, L_1, MLR, SVR_Gau, and WQSSVR models were implemented using the modules "robustfit", "linprog", "robustfit", "fitrsvm", and "quadprog" of MATLAB, respectively.

Following many studies in the literature of electric load forecasting (Hong, 2010; Hong & Fan, 2016; Hong, Pinson, & Fan, 2014; Hong et al., 2016; Hong & Wang, 2014; Hong et al., 2015; Luo, Hong, & Fang, 2018; Luo et al., 2019; Luo, Hong, & Yue, 2018; Xie & Hong, 2016), we do not normalize the load data. Moreover, to tune the parameter C_p for WQSSVR model, we divided the training data set into three equivalent parts, with the first two parts and last part as the pre-training data and validation data to select C_p in the set of $\{2^{28}, 2^{29}, \ldots, 2^{36}, 2^{37}\}$. The parameter δ in the constraints of WQSSVR′ model in Section 3.3 is selected as $(\overline{\gamma} - MED)h$, where $\overline{\gamma}$ and MED are defined at the start of Section 3.4 and also calculated for obtaining the weights of training points, $h = 4$ or 0.5.

4.2. Benchmarking performance without data integrity attacks

Table 1 records the mean absolute percentage error (MAPE) values of all five models without data integrity attacks. A smaller MAPE value indicates the more accurate load forecasts yielded by the corresponding model. Overall, the MLR and SVR_Gau models produce the most and least accurate load forecasts among all five models, respectively. The performance of WQSSVR is as close as that of L_1 regression or IRLS_bis model in terms of load forecast accuracy. Since the results at low level zones do not add much information nor change our final conclusions and findings, we focus on the aggregated zone Z_{21} for experiments of load forecasting under data integrity attacks to avoid verbose presentation.

4.3. Data integrity attacks targeting economic losses

When the values of the majority part of the load history are increased, the load forecasts would likely be higher than the nominal loads. These over-forecasts may lead to the economic losses due to the over-capacity and opportunity costs. Following the setup as those in Luo et al. (2019), Luo, Hong, and Yue (2018), Sobhani et al.

Table 1
MAPE (%) of hourly load forecast without data attacks.

Zone	IRLS_bis	L_1	MLR	SVR_Gau	WQSSVR
21	5.30	5.33	**5.22**	6.31	5.38
1	7.08	7.08	7.01	8.34	**6.93**
2	5.56	**5.52**	5.62	7.39	5.59
3	5.56	**5.52**	5.62	7.39	5.59
5	9.69	**9.64**	9.88	10.51	9.69
6	5.56	**5.53**	5.55	7.24	5.59
7	5.56	**5.52**	5.62	7.46	5.60
8	7.59	7.59	7.50	8.74	**7.39**
10	6.70	6.79	6.70	9.53	**6.64**
11	7.97	8.20	7.70	9.63	**7.64**
12	6.95	6.99	**6.78**	8.35	7.31
13	7.48	7.44	**7.39**	8.11	7.54
14	9.41	9.40	**9.38**	10.52	9.92
15	**7.38**	7.40	7.44	7.96	7.76
16	8.13	**8.11**	8.12	9.10	8.23
17	5.31	5.30	**5.26**	6.93	5.32
18	6.77	6.73	**6.72**	7.71	6.75
19	7.88	**7.87**	7.90	8.86	7.94
20	5.73	5.68	5.74	7.44	**5.66**
Avg	7.02	7.02	**7.00**	8.40	7.06
4	**15.83**	15.89	16.08	15.86	15.90
9	164.05	153.48	**139.16**	157.52	159.93

Table 2
Averages of MAPE (%) and SE of MAPE averages (%) for under normally distributed data attacks targeting economic losses.

		p%		
	k	$N(0.25, 0.5^2)$	$N(0.5, 0.5^2)$	$N(0.75, 0.5^2)$
IRLS_bis	40	5.23/0.01	5.23/0.01	**5.24/0.01**
L_1		**5.16/0.02**	**5.19/0.02**	5.50/0.03
MLR		9.23/0.18	17.41/0.27	27.50/0.40
SVR_Gau		6.39/0.05	6.91/0.06	8.45/0.13
WQSSVR		5.28/0.05	5.27/0.05	5.31/0.05
IRLS_bis	50	5.46/0.01	10.44/0.29	28.22/0.46
L_1		**5.24/0.01**	5.97/0.09	9.63/0.18
MLR		11.03/0.25	22.66/0.40	35.01/0.47
SVR_Gau		6.84/0.06	8.87/0.14	14.27/0.22
WQSSVR		5.34/0.07	**5.60/0.09**	**5.84/0.09**
IRLS_bis	60	7.90/0.11	21.93/0.45	38.60/0.42
L_1		**5.47/0.03**	10.35/0.12	25.27/0.34
MLR		13.29/0.39	27.36/0.43	42.27/0.41
SVR_Gau		7.66/0.10	12.93/0.27	25.13/0.39
WQSSVR		5.64/0.08	**5.81/0.06**	**11.97/0.31**
IRLS_bis	70	11.84/0.21	29.19/0.35	46.26/0.39
L_1		6.80/0.14	20.56/0.39	41.61/0.91
MLR		15.22/0.42	32.52/0.38	48.59/0.36
SVR_Gau		9.12/0.07	19.77/0.30	38.45/0.29
WQSSVR		**5.86/0.06**	**9.58/0.35**	**26.00/0.61**
IRLS_bis	80	15.96/0.36	35.27/0.46	56.23/0.35
L_1		10.67/0.31	31.18/0.75	57.01/0.28
MLR		17.89/0.42	37.20/0.45	57.55/0.31
SVR_Gau		11.87/0.14	28.73/0.29	51.86/0.19
WQSSVR		**7.96/0.26**	**19.28/0.52**	**47.90/0.19**
IRLS_bis	90	19.02/0.43	40.97/0.49	63.60/0.40
L_1		16.10/0.62	39.62/0.86	64.38/0.84
MLR		19.76/0.42	41.98/0.48	64.19/0.41
SVR_Gau		15.84/0.25	37.25/0.34	**60.69/0.26**
WQSSVR		**12.39/0.95**	**32.14/2.28**	62.97/2.47

(2020), and Zheng et al. (2020), the data integrity attack on the training data set targeting economic losses is mathematically simulated by randomly picking k% of all observations with their load values being deliberately increased by p%. These selected observations become outliers, a.k.a. the attacked points.

Using the training data set under the simulated data integrity attack targeting economic losses, we estimate the parameters for all five models. The original testing data set without data attacks is used to calculate the MAPE values of forecasted load. To comprehensively evaluate the performance of all five models, we conducted three groups of computational experiments under normally distributed or uniformly distributed data attacks following the practice in Luo, Hong, and Fang (2018), Sobhani et al. (2020), Spiliotis et al. (2019), and Zheng et al. (2020):

(1) vary k from 40 to 90 with the increment of 10, p% is generated by the normal distribution $N(\mu, \sigma^2)$, where μ is varied from 0.25 to 0.75 with the increment of 0.25 and σ is 0.5;
(2) k is 70, p% is generated by the normal distribution $N(\mu, \sigma^2)$, where μ is 0.5 and σ is varied from 0.25 to 1.5 with the increment of 0.25;
(3) k is 70, p% is generated by the uniform distribution $U(a,b)$.

For each (k, p) pair, we repeated the test with randomly selected k% observations for 10 times. For each model tested in the three groups of experiments, the averages of MAPE values of 10 experiments are reported in Tables 2, 3, and 4, respectively. The standard errors (SE) for the averages of MAPE values are also recorded in these tables to quantify the uncertainty in the reported MAPE averages. Notice that the standard error is calculated as dividing the square root of 10 into the standard deviation of 10 MAPE values. From these three tables, we have the following observations to make:

1. For most tested computational experiments, the WQSSVR model produces more accurate forecasts than the other four models, especially for large k and large mean of p. This is mainly because the relative importance (i.e., weights) of attacked points are greatly reduced in WQSSVR model, which mainly utilizes the information of normal points.

2. For small-scale data attacks (such as $k = 40$) or small mean of p % (such as $N(0.25, 0.5^2)$, $U(-0.9, 0.9)$, and so forth), the L_1 regression and IRLS_bis models perform well. Similar observations can be found in Jian Luo et al. (2019). However, as k increases or the mean of p increases, the WQSSVR model shows increasing dominance over other tested models.

3. From Table 3, as the standard deviation of normally distributed data attacks increases, the MAPE averages of WQSSVR model and L_1 regression decrease. This is mainly because the ratio of training points with reduced loads increases closer to the ratio of training points with increased loads.

4. From Table 4, as the mean of uniformly distributed data attacks increases (i.e., the ratio of the number of points with increased loads to that of points with decreased loads increases as: 5/5, 6/4, 7/3, 8/2, 9/1, and 10/0), the WQSSVR model shows the increasing advantage over other four models.

Table 3
Averages of MAPE (%) and SE of MAPE averages (%) under normally distributed data attacks targeting economic losses.

	f k	$p\%$					
		$N(0.5, 0.25^2)$	$N(0.5, 0.5^2)$	$N(0.5, 0.75^2)$	$N(0.5, 1^2)$	$N(0.5, 1.25^2)$	$N(0.5,1.5^2)$
IRLS_bis	70	31.50/0.21	29.19/0.35	25.72/0.41	23.27/0.67	23.73/1.00	23.74/1.08
L_1		31.77/0.43	20.56/0.39	13.08/0.57	9.53/0.27	7.97/0.22	6.82/0.15
MLR		32.43/0.26	32.52/0.38	32.12/0.60	31.40/0.96	33.37/1.46	32.30/1.56
SVR_Gau		28.52/0.16	19.77/0.30	16.31/0.29	14.26/0.35	14.80/0.32	15.05/0.53
WQSSVR		**28.25/1.74**	**9.58/0.35**	**7.03/0.21**	**6.70/0.23**	**6.64/0.17**	**6.59/0.06**

Table 4
Averages of MAPE (%) and SE of MAPE averages (%) under uniformly distributed data attacks targeting economic losses.

	k	$p\%$					
		$U(-0.9, 0.9)$	$U(-0.72, 1.08)$	$U(-0.54, 1.26)$	$U(-0.36, 1.44)$	$U(-0.18, 1.62)$	$U(0, 1.8)$
IRLS_bis	70	7.06/0.05	9.45/0.46	19.04/0.37	31.52/0.86	44.34/0.51	58.25/0.77
L_1		**5.58/0.02**	**5.61/0.06**	8.25/0.10	16.73/0.70	30.51/0.85	49.05/1.26
MLR		6.91/0.06	11.13/0.40	23.04/0.37	34.76/0.71	46.89/0.45	60.14/0.71
SVR_Gau		9.79/0.15	8.50/0.17	11.24/0.24	18.26/0.46	31.21/0.30	46.38/0.61
WQSSVR		6.79/0.21	6.86/0.18	**6.13/0.06**	**7.37/0.43**	**13.32/0.68**	**27.07/1.29**

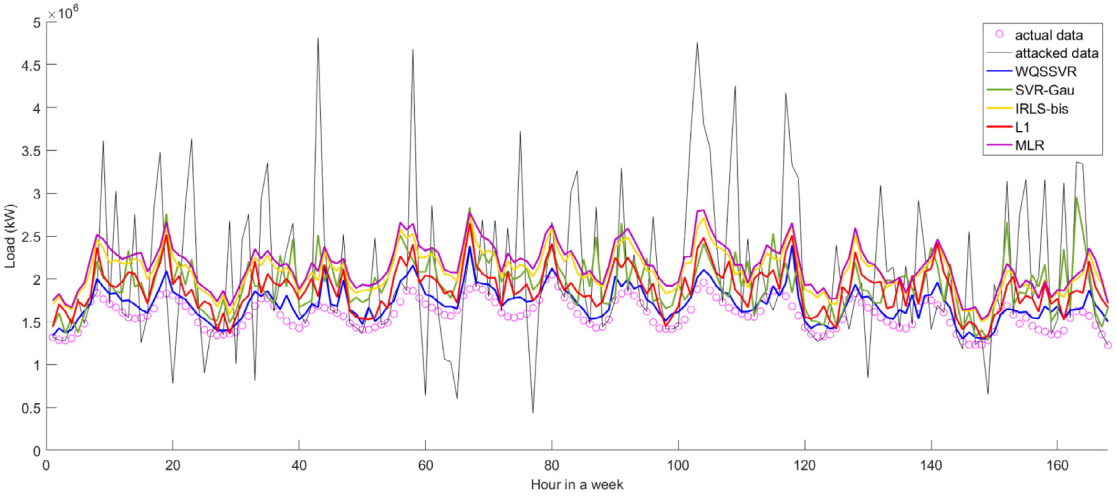

Fig. 1. Fitted (2005/1/7–2005/1/13) hourly load profile under normally distributed data attacks targeting economic losses.

5. The SE of all tested models are much smaller than the related MAPE averages. For most tested computational experiments, the SE of MAPE averages of WQSSVR are smaller than those of other tested models, which indicates that the performance of WQSSVR is more stable in terms of load forecast accuracy.

The fitted hourly load profiles of all tested models for one representative period under normally distributed and uniformly distributed data integrity attacks are shown in Figs. 1 and 2 with $k = 70$ and $p\%$ following $N(0.5, 0.5^2)$ and $U(-0.36, 1.44)$, respectively. The period of one week is in the winter of year 2005 (a training year). From these two figures, we can observe that all five models are more or less over-fitting the actual load for the training period, due to the data integrity attacks targeting economic losses (i.e., the mean of $p\%$ is positive so that the loads of most attacked points are larger than those of the original points). The fitting curve provided by the WQSSVR model is much closer to the actual data than that provided by each of other four models.

Moreover, the forecasted hourly load profiles of all tested models for one representative period under normally distributed and uniformly distributed data integrity attacks are shown in Figs. 3 and 4 with $k = 70$ and $p\%$ following $N(0.5, 0.5^2)$ and $U(-0.36, 1.44)$, respectively. The period of one week is in the winter of year 2007 (a testing year). From these two figures, we can observe that all five models are more or less over-predicting the actual load for the testing period, due to the data integrity attacks targeting economic losses (i.e., the mean of $p\%$ is positive). The predicted loads provided by the WQSSVR model are much closer to the actual loads than those provided by other four models. In Fig. 5, the curve representing the weights of points during two days in the summer of training year 2006 under data integrity attacks with $k = 70$ and $p\%$ following $N(0.5, 0.5^2)$ for WQSSVR model is also plotted. Notice that the loads of attacked points increased by the percentage following a normal distribution with the mean of 50% and standard deviation of 50%, so the loads of most attacked points are larger

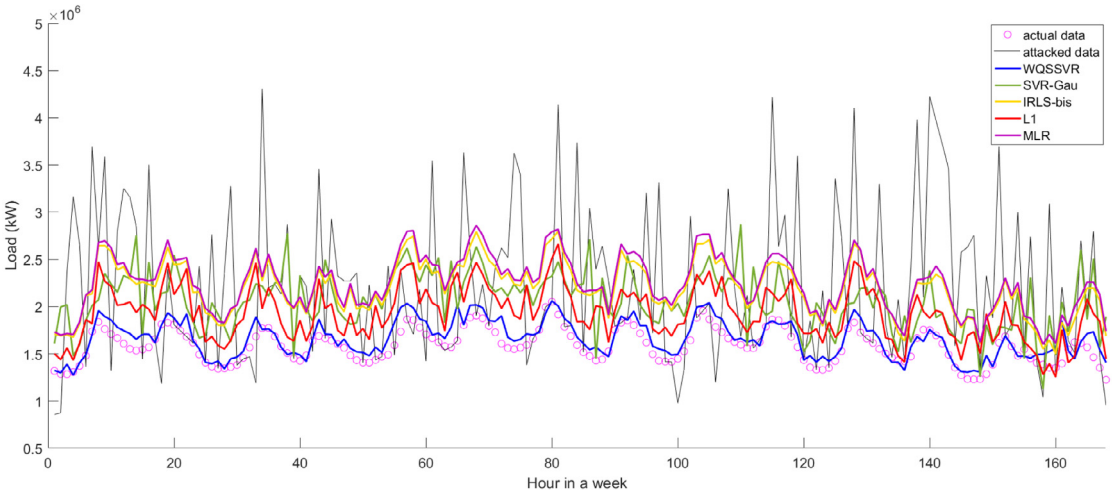

Fig. 2. Fitted (2005/1/7–2005/1/13) hourly load profile under uniformly distributed data attacks targeting economic losses.

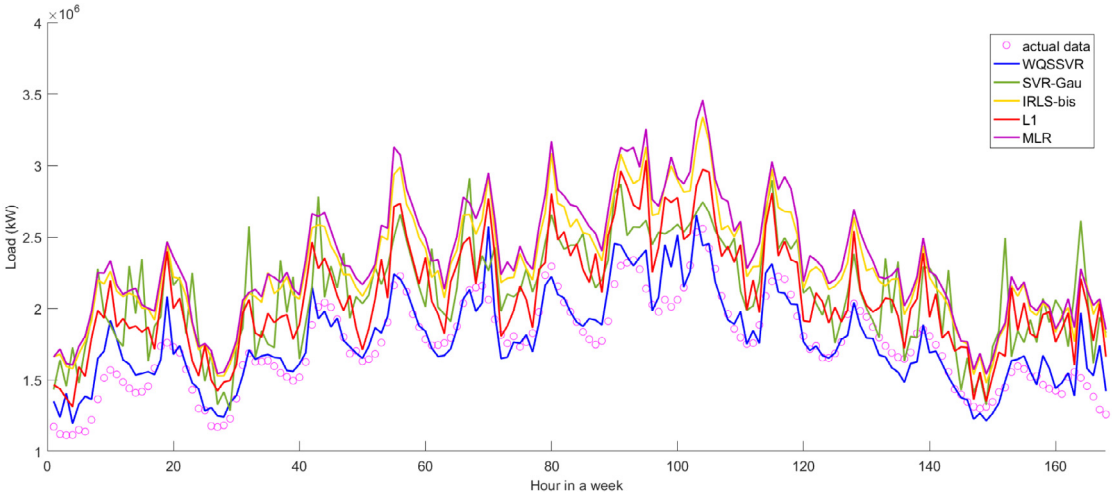

Fig. 3. Forecasted (2007/1/7–2007/1/13) hourly load profile under normally distributed data attacks targeting economic losses.

than those of the original points. We can observe that the small weights are assigned to the attacked observations with large perturbation magnitude (i.e., outliers or noise). For example, the attacked observations at hours 5, 6, 17, 19, 21, 23, 25, 29, 33, 39, and 41 are assigned the weights very close to 0.

4.4. Data integrity attacks targeting system blackouts

While increasing the historical load may result in over-forecasts, decreasing the historical load may lead to the under-forecasts so that the risk for brownouts or even blackouts is increased. Following Section 4.3, we can create a different type of data integrity attacks targeting system blackouts, by randomly picking $k\%$ of the observed load data in the training period and then decreasing them by $p\%$. Two groups of computational experiments under such data integrity attacks are conducted as the following two steps:

(1) Let $k = 40$, and vary $p\%$ from 10% to 60% with the increment of 10%.
(2) Let $k = 40$, $p\%$ is generated by the uniform distribution $U(a,b)$.

For each (k, p) pair, we repeat the test with randomly selected $k\%$ observations for 10 times. Here we primarily conduct representative experiments to avoid verbose presentation. For each model tested in the two groups of experiments, the averages of MAPE values and SE of MAPE averages of 10 experiments are reported in Tables 5 and 6, respectively. From these two tables, we have the following similar observations:

(1) For most tested computational experiments, the WQSSVR model produces more accurate forecasts than other tested models, especially for large mean of p.
(2) From Table 6, as the mean of p increases (i.e., the ratio of the number of points with decreased loads

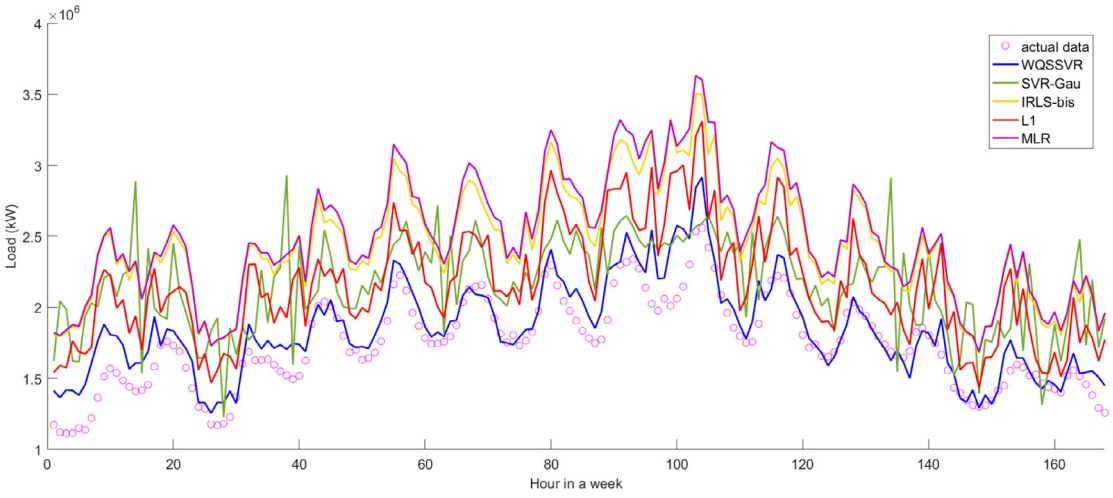

Fig. 4. Forecasted (2007/1/7–2007/1/13) hourly load profile under uniformly distributed data attacks targeting economic losses.

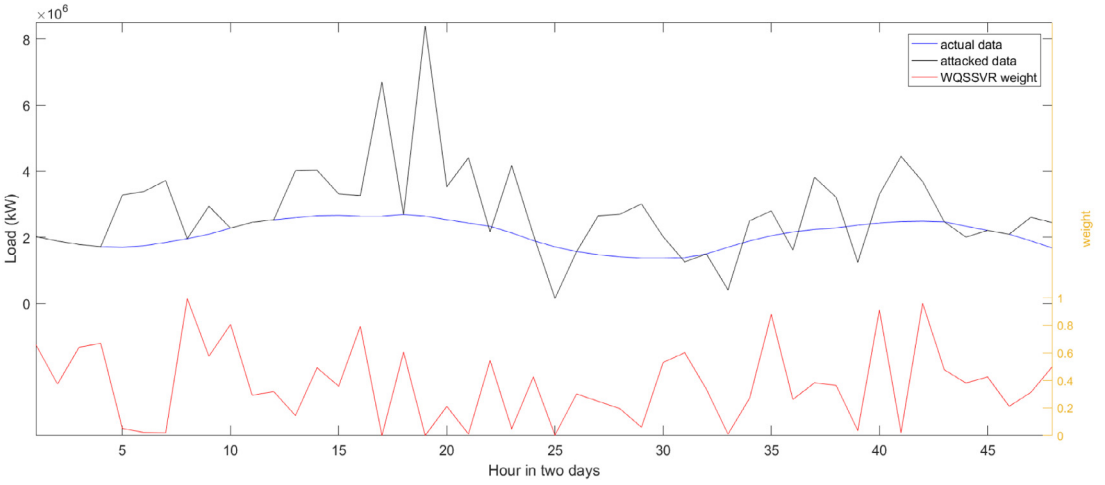

Fig. 5. Hourly load profile (2006/8/4–2006/8/5) and data weights for WQSSVR under data attacks targeting economic losses.

Table 5
Averages of MAPE (%) and SE of MAPE averages (%) under various levels of data integrity attacks targeting system blackouts.

		p%					
	k	0.1	0.2	0.3	0.4	0.5	0.6
IRLS_bis	40	6.97/0.03	9.77/0.09	13.08/0.14	16.62/0.19	20.37/0.24	23.88/0.28
L_1		6.64/0.04	7.40/0.08	7.63/0.10	7.82/0.10	8.05/0.11	8.15/0.12
MLR		6.97/0.03	9.90/0.08	13.36/0.13	17.04/0.17	20.93/0.22	24.59/0.26
SVR_Gau		8.50/0.03	10.76/0.05	12.17/0.09	13.34/0.09	14.70/0.12	16.00/0.23
WQSSVR		**6.37/0.02**	**6.88/0.03**	**6.17/0.05**	**6.12/0.09**	**6.22/0.10**	**6.23/0.12**

Table 6
Averages of MAPE (%) and SE of MAPE averages (%) under uniformly distributed data attacks targeting system blackouts.

		p%					
	k	$U(-0.5, 0.5)$	$U(-0.4, 0.6)$	$U(-0.3, 0.7)$	$U(-0.2, 0.8)$	$U(-0.1, 0.9)$	$U(0, 1)$
IRLS_bis	70	5.88/0.05	8.64/0.14	14.04/0.21	20.60/0.23	28.01/0.27	34.88/0.26
L_1		**5.47/0.04**	**6.30/0.05**	8.29/0.11	12.86/0.23	20.84/0.41	29.63/0.38
MLR		5.84/0.04	9.40/0.16	15.54/0.22	22.02/0.22	29.21/0.26	35.81/0.26
SVR_Gau		7.58/0.03	9.32/0.06	12.21/0.10	16.95/0.11	23.91/0.20	32.13/0.22
WQSSVR		6.29/0.17	6.37/0.24	**6.21/0.11**	**7.66/0.25**	**11.38/0.35**	**18.61/0.71**

Table 7
Forecast error in MAPE (%) and MSE (10^{10}) under normally distributed data attacks targeting economic losses.

		p%						
	k	$N(0.5, 0.5^2)$	$N(0.75, 0.5^2)$	$N(1, 0.5^2)$	$N(1.25, 0.5^2)$	$N(1.5, 0.5^2)$	$N(1.75, 0.5^2)$	$N(2, 0.5^2)$
IRLS_bis	5	5.29/1.47	5.29/1.47	5.29/1.47	5.29/1.47	5.29/1.47	5.29/1.47	5.29/1.47
L_1		5.29/1.48	5.27/1.46	5.27/1.46	5.27/1.46	5.27/1.46	5.27/1.46	5.27/1.46
MLR		5.31/1.44	5.80/1.50	6.30/1.77	6.63/1.98	7.61/2.51	8.67/3.26	9.96/4.16
SVR_Gau		6.26/2.21	6.24/2.20	6.24/2.20	6.24/2.20	6.24/2.20	6.24/2.20	6.24/2.20
WQSSVR		**5.24/1.41**	**5.24/1.41**	**5.24/1.41**	**5.24/1.41**	**5.24/1.41**	**5.24/1.41**	**5.24/1.41**

Table 8
Forecast error in MAPE (%) and MSE (10^{10}) under normally distributed data attacks targeting economic losses.

		p%					
	k	$N(0.025, 0.05^2)$	$N(0.05, 0.1^2)$	$N(0.075, 0.15^2)$	$N(0.1, 0.2^2)$	$N(0.125, 0.25^2)$	$N(0.15, 0.3^2)$
IRLS_bis	100	5.09/1.25	5.71/1.46	6.99/2.02	8.81/3.04	10.80/4.40	12.72/5.89
L_1		5.11/1.27	5.71/1.46	7.06/2.09	8.85/3.09	10.76/4.36	12.63/5.93
MLR		5.09/1.24	5.75/1.50	7.05/2.05	8.88/3.07	10.98/4.50	12.80/5.96
SVR_Gau		5.92/1.96	6.29/2.10	7.23/2.41	8.54/3.20	10.13/4.22	12.07/5.77
WQSSVR		**5.07/1.23**	**5.51/1.38**	**6.55/1.85**	**8.31/2.78**	**9.94/3.92**	**11.31/5.13**

to that of points with increased loads increases as: 5/5, 6/4, 7/3, 8/2, 9/1, and 10/0), the WQSSVR model shows the increasing advantage over other four models.

5. Discussion

In this section, we first test the load forecasting models for other possible normally distributed data attacks targeting economic losses and then test them on different data sets under uniformly distributed data attacks targeting system blackouts. Finally, the robustness of the proposed WQSSVR is discussed.

5.1. Other possible normally distributed data attacks targeting economic losses

To further investigate the performance of the proposed WQSSVR model for two possible types of data attacks, we simulated the normally distributed data integrity attacks targeting economic losses as the following two steps: (1) k is 5, $p\%$ is generated by the normal distribution $N(\mu, \sigma^2)$, where μ is varied from 0.5 to 2 with the increment of 0.25 and σ is 0.5. (2) $k = 100$, $p\%$ being generated by the normal distribution $N(\mu, \sigma^2)$, where μ is varied from 0.025 to 0.15 with the increment of 0.025 and $\sigma = 2\mu$. Then the five load forecasting models were tested on these data sets and the computational results for these two types of data attacks are recorded in Tables 7 and 8, respectively. The mean square error (MSE), an L_2 norm-based measurement of errors, was also used to investigate the forecast accuracy of tested models for fair comparisons.

From Tables 7 and 8, we can see that the MAPE and MSE values of WQSSVR model are lower than those of other tested models, respectively. These computational results indicate the superior performance of the proposed WQSSVR model over the robust regression models in Luo et al. (2019) and other commonly used load forecasting methods (i.e., MLR and SVR_Gau). From Table 7, for 5% attacked points, when the mean of $p\%$ steadily increases, the performance of WQSSVR and robust regression models stay the same. This indicates the strong robustness of WQSSVR and robust regression models for small-scale data attacks of large magnitudes. From Table 8, all tested models cannot produce very accurate load forecasts as μ increases. For future research, we are interested in studying this type of data attacks comprehensively by combining the robust forecasting models with some attack detection methods.

5.2. Different data sets under uniformly distributed data attacks targeting system blackouts

To further test the proposed WQSSVR model on other data sets (from other zones in GEFCom2012) under data integrity attacks, we first simulated the uniformly distributed data integrity attacks targeting system blackouts as the following two steps: (1) The percentage of the attacked observations in the training data was fixed to be 70%, i.e., $k = 70$. (2) Ten test cases were created by decreasing the magnitude of randomly selected load values by $p\%$, where $p\%$ is generated following a uniform distribution $U(-0.2, 0.8)$. Then the five load forecasting models were tested on these data sets and the computational results are recorded in Table 9.

Moreover, to further investigate the performance of the proposed model, the load data sets of the ISO New England control area and its eight wholesale load zones (called the ISONE data, publicly available from its website[1]) were utilized. In ISONE data, two years (2013 and 2014) and one year (2015) of hourly load and dry bulb temperature information (for the weather station corresponding to the load zone or Trading Hub) were selected as the training and testing periods, respectively. The used hourly load is the real-time demand for wholesale market settlement from revenue quality metering and defined as the sum of non-dispatchable load assets, station service load assets, and unmetered load assets. Moreover, the hourly loads in the control area (CA) zone are the sum of those in the other 8 zones, which are West/Central

[1] https://www.iso-ne.com/isoexpress/web/reports/pricing/-/tree/zone-info

Table 9
MAPE (%) of zones from GEFCom 2012 under data attacks targeting system blackouts.

Zone	IRLS_bis	L_1	MLR	SVR_Gau	WQSSVR
1	19.97	13.97	21.51	19.56	**8.63**
2	20.96	13.46	22.30	17.57	**7.79**
3	20.96	13.22	22.30	17.60	**8.07**
5	15.82	11.32	17.23	15.07	**10.11**
6	20.68	13.11	22.05	17.43	**7.77**
7	20.96	13.18	22.30	17.57	**7.89**
8	22.58	15.74	23.98	19.62	**8.84**
10	22.59	15.27	23.88	18.84	**10.05**
11	23.90	17.26	25.33	21.25	**8.26**
12	21.26	15.08	22.80	19.65	**11.70**
13	18.62	**13.31**	19.96	16.77	14.24
14	19.62	**15.81**	20.66	19.23	18.42
15	20.31	14.88	21.60	16.52	**13.74**
16	20.61	15.54	22.08	18.86	**12.11**
17	19.88	12.64	21.46	17.56	**7.02**
18	20.37	14.49	21.81	18.21	**9.48**
19	19.64	15.01	21.07	18.71	**11.81**
20	21.18	13.99	22.60	17.95	**9.56**
4	22.54	17.26	23.89	21.36	**16.35**
9	117.00	111.85	117.56	124.94	**110.11**

Table 10
MAPE (%) of zones from ISONE under data attacks targeting system blackouts.

Zone	IRLS_bis	L_1	MLR	SVR_Gau	WQSSVR
WCMASS	19.17	11.48	20.46	14.38	**8.84**
VT	20.52	12.35	21.82	13.97	**8.73**
SEMASS	19.91	12.46	21.35	13.58	**10.46**
RI	19.80	12.44	21.23	13.81	**11.78**
NH	19.90	12.54	21.18	16.00	**10.61**
NEMASS	20.11	11.99	21.51	14.24	**9.16**
ME	16.25	8.30	17.82	13.14	**5.10**
CT	20.80	12.94	22.24	14.55	**9.38**
Aggregate	17.64	9.41	19.06	13.18	**4.86**

Massachusetts (WCMASS), Vermont (VT), Southeast Massachusetts (SEMASS), Rhode Island (RI), New Hampshire (NH), Northeast Massachusetts (NEMASS), Maine (ME), and Connecticut (CT). Then we simulated the similar uniformly distributed data integrity attacks targeting system blackouts for ISONE data by setting k to be 70 and generating $p\%$ following the uniform distribution $U(-0.2, 0.8)$. Similarly, the load forecasting models were tested on these data sets and the results are recorded in Table 10. From Tables 9 and 10, we can have the observations similar to the ones in Section 4.

5.3. Robustness

In these computational experiments, the normally distributed or uniformly distributed data integrity attacks were designed against the load data mainly from three aspects, i.e., different percentage ($k\%$) of load data being perturbed maliciously, varied mean ($\mu\%$) or varied standard deviation ($\sigma\%$) of the perturbation magnitude (i.e., varied $p\%$), and different types of data attacks (targeting economic losses or system blackouts). We ranked the overall performance of five tested load forecasting models under data integrity attacks from the most accurate one to the least accurate one as: WQSSVR, L_1 regression, SVR_Gau, IRLS_bis, and MLR. Under no data integrity attacks, the overall performance of WQSSVR, L_1 regression, IRLS_bis, and MLR models are close to each other, while SVR_Gau produces the least accurate load forecasts.

Besides the detailed robustness analysis of IRLS_bis, L_1 regression, and MLR models in Luo et al. (2019), several observations about the robustness of these tested models can be made: (1) On average, the WQSSVR model is the most robust one among all five models. This is mainly because the WQSSVR model assigns small and large weights to attacked and normal points after calculating the ℓ_1–normed residuals of all points, which greatly reduces the impact of attacked points. (2) The L_1 regression and SVR_Gau models are more robust than IRLS_bis and MLR models, largely because the L_1 regression and SVR_Gau models utilize ℓ_1–norm (instead of ℓ_2–norm in the other two models) to measure the fitting errors. (3) the L_1 regression model outperforms the SVR_Gau model regardless of whether the load data is under data integrity attacks or not, largely because the SVR_Gau model becomes over-fitted by utilizing the 289 variables to predict the electric loads.

6. Conclusions

This paper focuses on robust machine learning models, which can be utilized for load forecasting with or without data attacks. In this paper, the data integrity attacks on the historical load data have been addressed from three perspectives, namely, the percentage ($k\%$) of data being perturbed, the magnitude ($p\%$) of the normally distributed or uniformly distributed perturbations, and the type of data attacks (targeting economic losses or system blackouts). Under these types of data integrity attacks, we show that the robust load forecasting models including the L_1 regression, IRLS, and SVR with Gaussian kernel may easily fail to provide reliable load forecasts under large-scale data integrity attacks (i.e., $k \geq 40$), while the proposed WQSSVR model is capable of producing much more accurate and robust load forecasts. Especially when more observations (such as 70% of whole data set) are attacked with a large mean of perturbation magnitude, the WQSSVR model demonstrates much stronger robustness than other electric load forecasting models. The computational results indicate that the load forecasting MAPE provided by the WQSSVR model remains under 10% even with 70% of the historical load data being maliciously decreased by 30% or increased by 50% on average.

These types of data integrity attacks represent only a small portion of potential attacks that a load forecasting system may encounter. Other types of data integrity attacks (such as introducing slowly increasing bias to all data points, data integrity attacks on weather data, calendar data, and peak periods) need to be further investigated. All existing methods, including the proposed WQSSVR model, may or may not perform well in the face of other cyber attacks such as the data integrity attacks on temperature and dates. This study may lead to the investigation of new theory and methodologies for load forecasting under other types of data integrity attacks.

Another approach to addressing data integrity attacks would involve detecting attacks, identifying attacked data, cleansing and recovering attacked data, and finally electric load forecasting. This study paves the way for further research on anti-attack methods for electric load forecasting. The anomaly detection (or data attack detection) and similar methods, such as detecting the inconsistencies in the relationship between the external variables and the load data (Sobhani et al., 2020), can be incorporated with the robust WQSSVR model to improve the load forecast accuracy. We are also interested in studying the impact of these data attacks on probabilistic forecasting (Hong & Fan, 2016) for future research.

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.

Acknowledgment

Jian Luo's research has been supported by the National Natural Science Foundation of China Grants # 71701035. Zheming Gao's research has been supported in part by the Fundamental Research Funds for the Central Universities under Grant N2204017.

Appendix A

The relative geometrical margin at point $((x^i)^T, y^i)^T$ is defined and approximated as follows.

Definition 1. For $q(x^i) \geq y^i$(or $\leq y^i$), the negative or positive gradient direction at point $((x^i)^T, y^i)^T$ with respect to $q(x) - y = q(x^i) - y^i$ intercepts the surface $q(x) = y$ at point $((x^B)^T, y^B)^T$. The distance between points $((x^i)^T, y^i)^T$ and $((x^B)^T, y^B)^T$, denoted as ζ_i, is called the geometrical margin at point $((x^i)^T, y^i)^T$ with respect to $q(x) = y$.

Definition 2. For $q(x^i) \geq y^i$(or $\leq y^i$), given $\delta > 0$, the negative or positive gradient direction at point $((x^i)^T, y^i)^T$ with respect to $q(x)-y = q(x^i)-y^i$ intercepts the surfaces $q(x) - y = +\delta$ (or $-\delta$) and $q(x) = y$ at points $((x^l)^T, y^l)^T$ and $((x^B)^T, y^B)^T$, respectively. The distance between points $((x^l)^T, y^l)^T$ and $((x^B)^T, y^B)^T$, denoted as $\bar{\zeta}_i$, is called the relative geometrical margin at point $((x^i)^T, y^i)^T$ with respect to $q(x) = y$.

Without loss of generality, take the training point $((x^i)^T, y^i)^T$ satisfying $q(x^i) \geq y^i + \delta$ as an example, Fig. A.1 illustrates the $(x^i, y^i)^T$, $(x^l, y^l)^T$, $(x^B, y^B)^T$, ζ_i, and $\bar{\zeta}_i$ for $m = 1$, where m is the number of dimensions of x^i. Notice that, as shown in Fig. A.1, $\frac{(\nabla q(x^i), -1)}{\|(\nabla q(x^i), -1)\|_2}$ is calculated as the positive gradient direction at point $((x^i)^T, y^i)^T$ with respect to $q(x) - y = q(x^i) - y^i$. The relative geometrical margins at different points are quite different due to the quadratic nature of fitting surface. Similar to that in Luo et al. (2016), the relative geometrical margin $\bar{\zeta}_i$ at point $((x^i)^T, y^i)^T$ is then approximated as follows: Let $((x^0)^T, y^0)^T$ be the origin of R^{m+1}, $\overrightarrow{(x^0, y^0)(x^B, y^B)}$, $\overrightarrow{(x^0, y^0)(x^l, y^l)}$ and $\overrightarrow{(x^l, y^l)(x^B, y^B)}$ are supposed to be the vectors from points $((x^0)^T, y^0)^T$, $((x^0)^T, y^0)^T$ and $((x^l)^T, y^l)^T$ to points $((x^B)^T, y^B)^T$, $((x^l)^T, y^l)^T$, and $((x^B)^T, y^B)^T$, respectively. Then we can infer from Definition 2 that $\overrightarrow{(x^0, y^0)(x^B, y^B)} = \overrightarrow{(x^0, y^0)(x^l, y^l)} + \overrightarrow{(x^l, y^l)(x^B, y^B)}$ and $\overrightarrow{(x^l, y^l)(x^B, y^B)} = -\bar{\zeta}_i \frac{(\nabla q(x^i), -1)}{\|(\nabla q(x^i), -1)\|_2}$. Hence, $x^B = x^l - \bar{\zeta}_i \frac{\nabla q(x^i)}{\|(\nabla q(x^i), -1)\|_2}$ and $y^B = y^l - \bar{\zeta}_i \frac{-1}{\|(\nabla q(x^i), -1)\|_2}$. Taylor's expansion says that $q\left(x^B\right) \approx q\left(x^l\right) + \nabla q\left(x^l\right)^T \left(x^B - x^l\right)$. Notice that, $q\left(x^B\right) = y^B$ and $q\left(x^l\right) = y^l + \delta$, then we have

$$
\begin{aligned}
y^B &\approx y^l + \delta + \nabla q\left(x^l\right)^T \left(x^B - x^l\right) \\
&= y^l + \delta + \nabla q\left(x^l\right)^T \left(-\bar{\zeta}_i \frac{\nabla q\left(x^i\right)}{\|(\nabla q\left(x^i\right), -1)\|_2}\right).
\end{aligned}
$$

By using $y^B = y^l - \bar{\zeta}_i \frac{-1}{\|(\nabla q(x^i), -1)\|_2}$, we can infer that $\bar{\zeta}_i \approx \frac{\delta \|(\nabla q\left(x^i\right), -1)\|_2}{\nabla q\left(x^l\right)^T \nabla q\left(x^i\right) + 1}$. Similarly,

$$
q\left(x^l\right) \approx q\left(x^i\right) + \nabla q\left(x^i\right)^T \left(x^l - x^i\right)
$$

$$
q\left(x^i\right) \approx q\left(x^l\right) + \nabla q\left(x^l\right)^T \left(x^i - x^l\right)
$$

and $x^l - x^i = -\frac{(\zeta_i - \bar{\zeta}_i)\nabla q\left(x^i\right)}{\|(\nabla q(x^i), -1)\|_2}$, inferred by $\overrightarrow{(x^0, y^0)(x^l, y^l)} - \overrightarrow{(x^0, y^0)(x^i, y^i)} = \overrightarrow{(x^i, y^i)(x^l, y^l)}$. Then $\nabla q\left(x^l\right)^T \nabla q\left(x^i\right) \approx \nabla q\left(x^i\right)^T \nabla q\left(x^i\right)$. Hence, for point $((x^i)^T, y^i)^T$,

$$
\begin{aligned}
\bar{\zeta}_i &= \left\|((x^B)^T, y^B)^T - ((x^l)^T, y^l)^T\right\|_2 \\
&\approx \frac{\delta \|(\nabla q\left(x^i\right), -1)\|_2}{\nabla q\left(x^l\right)^T \nabla q\left(x^i\right) + 1} \\
&\approx \frac{\delta}{\|(\nabla q\left(x^i\right), -1)\|_2} \\
&= \frac{\delta}{\|(Wx^i + b, -1)\|_2}.
\end{aligned}
$$

Appendix B

The WQSSVR model is equivalently reformulated for one smaller-sized optimization problem as the following four steps:

(1) Define Ψ be the vector formed by taking the m elements in the diagonal of matrix W, i.e.,

$$
\Psi \triangleq (w_{11}, w_{22}, \ldots, w_{mm})^T \in R^m
$$

(2) Construct an $m \times m$ matrix M_i for each training point $x^i = (x_1^i, x_2^i, \ldots, x_m^i)^T \in R^m$ as follows. For the diagonal of M_i, let the jth element in the jth row of matrix M_i be $x_j^i, j = 1, 2, \ldots, m$. The other elements of M_i are set to be 0.

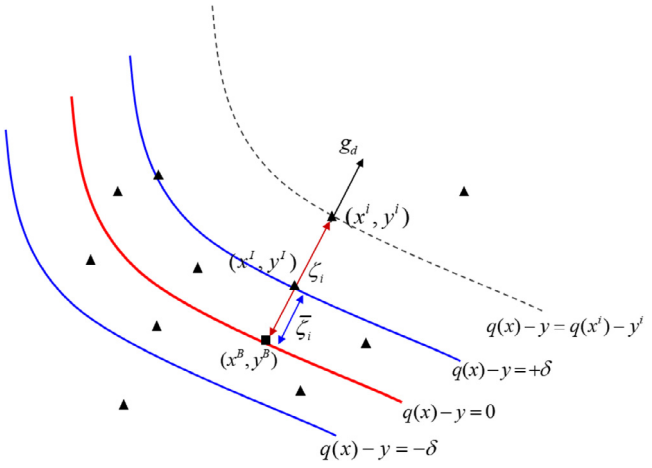

Fig. A.1. Demonstration of various margins in two dimensions.

(3) Denote $I_{m\times m}$ as the m-dimensional identity matrix, let $H_i \triangleq (M_i, I_{m\times m}) \in R^{m\times 2m}, i = 1, 2, \ldots, n,$

$$z \triangleq \left(\Psi^T, b^T\right)^T \in R^{2m}$$

$$s_i \triangleq \left(\frac{1}{2}x_1^i x_1^i, \frac{1}{2}x_2^i x_2^i, \ldots, \frac{1}{2}x_{m-1}^i x_{m-1}^i, \frac{1}{2}x_m^i x_m^i, x_1^i, x_2^i, \ldots, x_m^i\right)^T \in R^{2m}.$$

Then the first term in the objective of model (WQSSVR) becomes

$$\sum_{i=1}^{n} \beta_i \|Wx^i + b\|_2^2$$
$$= \sum_{i=1}^{n} \beta_i \|H_i z\|_2^2$$
$$= \sum_{i=1}^{n} \beta_i (H_i z)^T (H_i z)$$
$$= \sum_{i=1}^{n} z^T (\beta_i H_i^T H_i) z,$$

and the constraints in the WQSSVR model becomes

$$\left|y^i - (\frac{1}{2}(x^i)^T W x^i + b^T x^i + c)\right|$$
$$= \left|y^i - (s_i^T z + c)\right|$$
$$\leq \delta + \xi_i,$$

$i = 1, 2, \ldots, n.$

(4) Let $G \triangleq \sum_{i=1}^{n} \beta_i H_i^T H_i \in R^{2m\times 2m}$, then the WQSSVR model can be equivalently reformulated as the following WQSSVR' model:

$$\min_{z,c,\xi} \quad z^T G z + C_p \sum_{i=1}^{n} \beta_i \xi_i$$

$$s.t. \quad \delta + \xi_i \geq y^i - (s_i^T z + c) \geq -\delta - \xi_i, i = 1, 2, \ldots, n,$$
$$\xi_i \geq 0, i = 1, 2, \ldots, n,$$

where $\delta, C_p > 0$ are the given parameters. And the size of reformulated WQSSVR' model is smaller than the WQSSVR model.

References

Akouemo, H. N., & Povinelli, R. J. (2016). Probabilistic anomaly detection in natural gas time series data. *International Journal of Forecasting*, *32*(3), 948–956.

Basu, A., & Paliwal, K. K. (1989). Robust M-estimates and generalized M-estimates for autoregressive parameter estimation. In *Fourth IEEE region 10 international conference*.

Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

Ceperic, E., Ceperic, V., & Baric, A. (2013). A strategy for short-term load forecasting by support vector regression machines. *IEEE Transactions on Power Systems*, *28*, 4356–4364.

Charlton, N., & Singleton, C. (2014). A refined parametric model for short term load forecasting. *International Journal of Forecasting*, *30*(2), 364–368.

Chen, B.-J., Chang, M.-W., & Lin, C.-J. (2004). Load forecasting using support vector machines: a study on EUNITE competition 2001. *IEEE Transactions on Power Systems*, *19*(4), 1821–1830.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*, 273–297.

Dagher, I. (2008). Quadratic kernel-free non-linear support vector machine. *Journal of Global Optimization*, *41*, 15–30.

Ericsson, G. N. (2010). Cyber security and power system communicationessential parts of a smart grid infrastructure. *IEEE Transactions on Power Delivery*, *25*(3), 1501–1507.

Ghelardoni, L., Ghio, A., & Anguita, D. (2013). Energy load forecasting using empirical mode decomposition and support vector regression. *IEEE Transactions on Smart Grid*, *4*, 549–556.

Hahn, H., Meyer-Nieberg, S., & Pickl, S. (2009). Electric load forecasting methods: tools for decision making. *European Journal of Operational Research*, *199*(3), 902–907.

Hippert, H. S., Pedreira, C. E., & Souza, R. C. (2001). Neural networks for short-term load forecasting: a review and evaluation. *IEEE Transactions on Power Systems*, *16*(1), 44–55.

Hong, T. (2010). *Short term electric load forecasting*. North Carolina State University.

Hong, T., & Fan, S. (2016). Probabilistic electric load forecasting: a tutorial review. *International Journal of Forecasting*, *32*(3), 914–938.

Hong, T., & Hofmann, A. (2021). Data integrity attacks against outage management systems. *IEEE Transactions on Engineering Management*, 1–8.

Hong, T., Pinson, P., & Fan, S. (2014). Global energy forecasting competition 2012. *International Journal of Forecasting*, *30*(2), 357–363.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., & Hyndman, R. J. (2016). Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, *32*(3), 896–913.

Hong, T., & Wang, P. (2014). Fuzzy interaction regression for short term load forecasting. *Fuzzy Optimization and Decision Making*, *13*(1), 91–103.

Hong, T., Wang, P., & White, L. (2015). Weather station selection for electric load forecasting. *International Journal of Forecasting*, *31*(2), 286–295.

Hong, T., Wilson, J., & Xie, J. (2014). Long term probabilistic load forecasting and normalization with hourly information. *IEEE Transactions on Smart Grid*, *5*(1), 456–462.

Luo, J., Fang, S.-C., Deng, Z., & Guo, X. (2016). Soft quadratic surface support vector machine for binary classification. *Asia-Pacific Journal of Operational Research*, *33*(6).

Luo, Jian, Hong, T., & Fang, S.-C. (2018). Benchmarking robustness of load forecasting models under data integrity attacks. *International Journal of Forecasting*, *34*(1), 89–104.

Luo, Jian, Hong, T., & Fang, S.-C. (2019). Robust regression for load forecasting. *IEEE Transactions on Smart Grid*, *10*(5), 5397–5404.

Luo, J., Hong, T., & Yue, M. (2018). Real-time anomaly detection for very short-term load forecasting. *Journal of Modern Power Systems and Clean Energy*, *6*(2), 235–243.

Luo, Jian, Yan, X., & Tian, Y. (2020). Unsupervised quadratic surface support vector machine with application to credit risk assessment. *European Journal of Operational Research*, *280*(3), 1008–1017.

Perez, E. (2016). First on CNN: U.S. investigators find proof of cyberattack on Ukraine power grid.

Sobhani, M., Hong, T., & Martin, C. (2020). Temperature anomaly detection for electric load forecasting. *International Journal of Forecasting*, *36*, 324–333.

Spiliotis, E., Nikolopoulos, K., & Assimakopoulos, V. (2019). Tales from tails: On the empirical distributions of forecasting errors and their implication to risk. *International Journal of Forecasting*, *35*, 687–698.

Tian, Y., Sun, M., Deng, Z., Luo, J., & Li, Y. (2017). A new fuzzy set and non-kernel SVM approach for mislabeled binary classification with applications. *IEEE Transactions on Fuzzy Systems*.

Vanderbei, R. J. (1999). LOQO: an interior point code for quadratic programming. *Optimization Methods & Software*, *11*, 451–484.

Vapnik, V. (1982). *Estimation of dependences based on empirical data*. Springer Verlag.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, Inc: Springer-Verlag.

Vapnik, V., & Lerner, A. (1963). Pattern recognition using generalized portrait method. *Automation and Remote Control*, *24*, 774–780.

Weron, R. (2006). *Modeling and forecasting electricity loads and prices: a statistical approach*. John Wiley & Sons.

Wright, S. J. (1997). *Primal–dual interior-point methods*. society for industrial and applied mathematics.

Xie, J., & Hong, T. (2016). GEFCom2014 probabilistic electric load forecasting: an integrated solution with forecast combination and residual simulation. *International Journal of Forecasting*, *32*(3), 1012–1016.

Yao, X., Crook, J., & Andreeva, G. (2015). Support vector regression for loss given default modelling. *European Journal of Operational Research*, *240*(2), 528–538.

Yue, M., Hong, T., & Wang, J. (2019). Descriptive analytics-based anomaly detection for cybersecure load forecasting. *IEEE Transactions on Smart Grid*, *10*(6), 5964–5974.

Zheng, R., Gu, J., Jin, Z., Peng, H., & Zhu, Y. (2020). Load forecasting under data corruption based on anomaly detection and combined robust regression. *International Transactions on Electrical Energy Systems*, *30*(7).