

Article

Twin Least Square Support Vector Regression Model Based on Gauss-Laplace Mixed Noise Feature with Its Application in Wind Speed Prediction

Shiguang Zhang ^{1,2,3}, Chao Liu ^{1,*}, Wei Wang ¹ and Baofang Chang ¹

¹ College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; zhangshiguang@htu.edu.cn (S.Z.); wangwei@htu.edu.cn (W.W.); changbaofang@htu.edu.cn (B.C.)

² School of Computer Science and Technology, Tianjin University, Tianjin 300350, China

³ Engineering Lab of Intelligence Business and Internet of Things, Xinxiang 453007, China

* Correspondence: 2015139@htu.edu.cn

Received: 10 July 2020; Accepted: 26 September 2020; Published: 29 September 2020



Abstract: In this article, it was observed that the noise in some real-world applications, such as wind power forecasting and direction of the arrival estimation problem, does not satisfy the single noise distribution, including Gaussian distribution and Laplace distribution, but the mixed distribution. Therefore, combining the twin hyperplanes with the fast speed of Least Squares Support Vector Regression (LS-SVR), and then introducing the Gauss–Laplace mixed noise feature, a new regressor, called Gauss-Laplace Twin Least Squares Support Vector Regression (GL-TLSSVR), for the complex noise. Subsequently, we apply the augmented Lagrangian multiplier method to solve the proposed model. Finally, we apply the short-term wind speed data-set to the proposed model. The results of this experiment confirm the effectiveness of our proposed model.

Keywords: Gauss-Laplace mixed noise; least squares support vector regression; twin hyperplanes; wind speed prediction

1. Introduction

In recent years, the support vector machine (SVM) [1–4] have received widespread attention as a powerful method, because support vector machines have better generalization performance than other machine learning techniques. Thanks to good generalization capabilities, SVM technology is applied to various fields. For example, SVM has been applied to face detection [5], feature selection [6], function approximation [7], financial forecasting [8], and wind turbine system [9–15]. As for the support vector regression (SVR) model [16], it uses support vector machine technology to solve the regression estimation, there are many important methods, such as ε -support vector regression (ε -SVR) [17], ν -support vector regression (ν -SVR) [18] etc. In addition, based on some advantages of SVR, SVR has been successfully applied to Biology, medicine, environmental protection, information technology, engineering technology, and other fields [19–24].

In these SVR models, when solving regression problems, the noise of the training data is considered to be the single distribution. According to the Bayesian principle, first, the Gaussian noise with square loss is the best, secondly, the Beta noise with Beta loss is the best, finally, the Laplace noise with Laplace loss is the best [25,26]. However, in some practical applications, if data are collected in a multi-source environment, then the noise distribution is complex and unknown. Therefore, a single distribution cannot clearly describe the real noise [27,28]. In general, the mixed distribution has a good approximation ability for any continuous distribution. For some actual noises, prior knowledge is difficult to obtain. At this time, mixed noise can be well adapted to unknown or complex noise. In 2017, the research and application of a new wind speed hybrid

forecasting system that is based on multi-objective optimization is proposed [27,29], the proposed hybrid model is integrated with three components, singular spectrum analysis, the firefly algorithm, and the BP neural network [30]; as compared with a single BP, the prediction effect of the hybrid prediction method is better, which shows that the prediction ability of the hybrid method is stronger. In addition, accurate prediction of wind speed is a key task for the development and utilization of wind energy, when compared with other related methods, the proposed hybrid method has satisfactory performance in terms of accuracy and stability [31]. In this literature [32], two new nonlinear regression models for single-task and multi-task problems are developed, in which the noise is composed of Gaussian mixture. When compared to some other models, the proposed model becomes a robust nonlinear regression model with strong adaptation.

However, the main disadvantage of SVR is the high cost of learning. In order to improve the calculation speed of SVR, based on twin support vector machine (TSVM) [33], Peng [34–36] proposed twin support vector regression (TSVR). Unlike SVR, TSVR generates two non-parallel upper and lower bound functions by solving a pair of smaller quadratic programming problems (QPPs). In theory, TSVR reduces the computational cost compared to standard SVR. Zhao et al. [37] extended the concept of twin hyperplanes, and combined the advantages of least squares support vector regression (LSSVR) to generate the estimated regressor, called Twin Least Squares Support Vector Regression (TLSSVR). By observing the model of Peng [34], Khemchandani et al. [38] believed that only the principle of empirical risk minimization was considered in TSVR. To overcome these difficulties, Shao et al. [39] proposed another twin regression model, called ε -TSVR, which considers the principle of structural risk minimization. Later, Rastogi et al. [40] extended ε -TSVR and proposed ν -TSVR, which can automatically optimize parameters ε_1 and ε_2 based on sample data. By using the pinball loss function, Xu et al. [41] further developed an asymmetric ν -twin support vector regression, called Asy- ν -TSVR, which can effectively reduce noise interference and improve the generalization performance. Therefore, extensive research has been conducted on the twin-type SVR. In all of these twin-type SVR models, the distribution of training data is not considered in solving regression problems. This means that, regardless of whether the samples are important or not, all of the samples play the same role in the constraint function, so it will cause regression performance to decline. Depending on the importance of the data, given different samples, the penalty is more reasonable. For this reason, various methods [42–46] have been developed in order to study this shortcoming. For example, Xu et al. [44] proposed using the local information present on the sample based on K-nearest neighbor weighted twin support vector regression to improve the prediction accuracy. By clustering based on the similarity of training data, Parastalooi et al. [45] proposed an improved twin support vector regression. Ye [46] proposed an effective weighted Lagrangian ε -twin support vector regression (WL- ε -TSVR) with quadratic loss function, in which the weight matrix D was introduced in order to reduce the outlier pair to a certain extent Regression of the influence of variables, so as to impose different penalties on samples.

Traditionally, the upper and lower regression of the twin SVR is obtained by approximate dual solutions. However, Chapelle [47] observed that, by comparing the approximate efficiency of SVR in the primal space and the dual space, the approximate dual solution may not produce a good primal approximate solution. Some related work is directly solved in the primal space [48–51]. For example, inspired by the twin SVR and Newton methods, Balasundaram et al. [49] proposed a new unconstrained Lagrangian TSVR (ULTSVR) to solve a pair of unconstrained minimization problems, thereby increasing the calculation speed. Gupta [50] and Balasundaram [51] use the generalized derivative method to obtain QPPs. Although their work is efficient and fast, they only consider empirical risk minimization and do not consider structural risks.

Inspired by the above research, we try to study the characteristics of the complex or unknown noise distribution of the Gauss–Laplace mixed noise twin least squares support vector regression (GL-TLSSVR) model. In this article, for the solution to the regression task, the augmented Lagrange

multiplier method (ALM) algorithm is used in our experiments, it can help us better to find the optimal solution.

This work mainly provides four contributions, we describe the whole methodology in the flowchart, as shown in Figure 1.

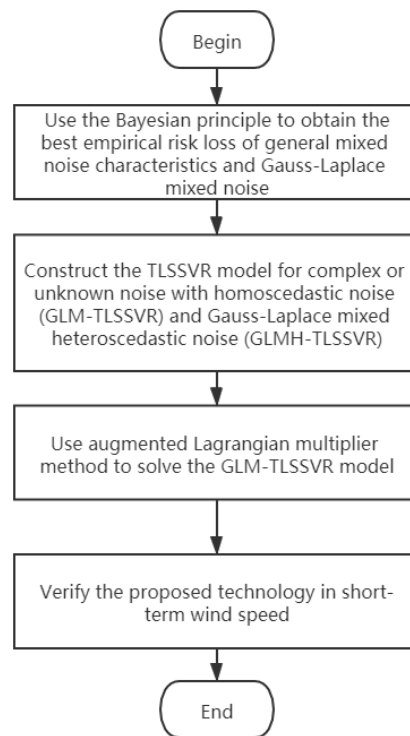


Figure 1. The whole methodology process of this article.

2. Related Work

In this section, the data-set is represented by $D_N = \{(A_i, y_i)\}, i = 1, 2, \dots, N$, where $A_i \in R^n$, $y_i \in R (i = 1, 2, \dots, N)$ is the training samples.

According to the Bayesian principle, we can derive the optimal empirical risk loss of the mixed noise characteristics [52]. The best empirical risk loss for this mixed noise distribution is shown below

$$l(\zeta) = \lambda_1 \cdot l_1(\zeta) + \lambda_2 \cdot l_2(\zeta). \quad (1)$$

where $l_1(\zeta) > 0, l_2(\zeta) > 0$ are the convex empirical risk loss of the above two noise characteristics. $\lambda_1, \lambda_2 \geq 0$ are weight factors, and $\lambda_1 + \lambda_2 = 1$.

Figure 2 shows the G-L empirical risk loss for different parameters.

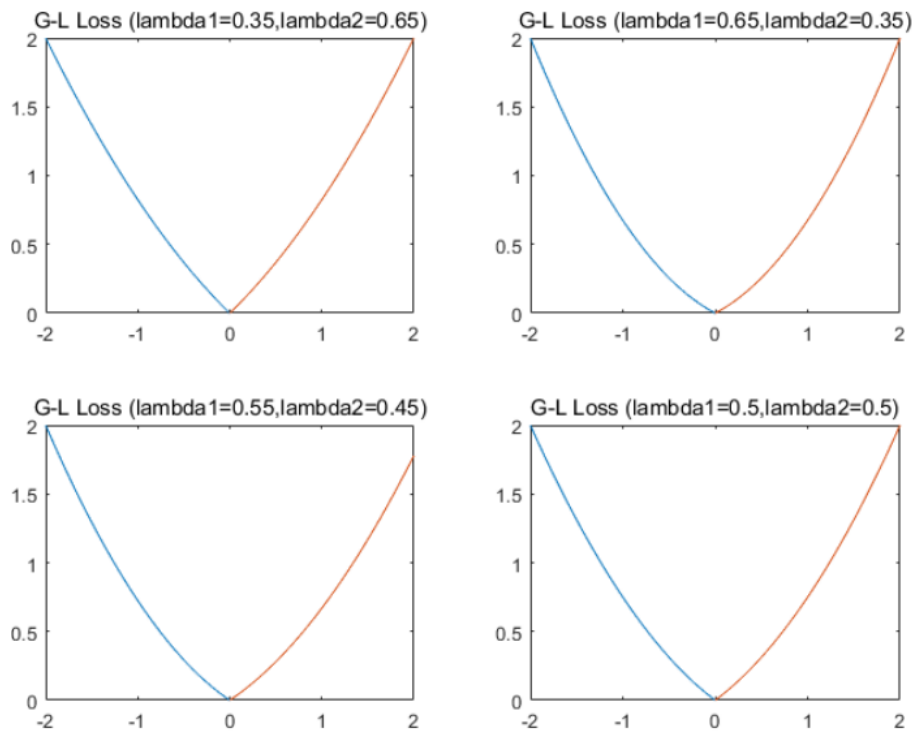


Figure 2. G-L empirical risk loss for different parameters.

3. TLSSVR Model of G-L Mixed Noise Characteristics

For the linear model, we want to find a linear regression function $f(A) = \omega^T \cdot A + b$. When dealing with some nonlinear problems, some specific methods are given ([53]): the input vector $A_i \in R^n$ is mapped by a non-linear mapping $\Phi: R^n \rightarrow H$ (take a prior distribution) to the high dimensional feature space H (H is Hilbert space), induced by the nonlinear kernel function $K(A_i, A_j) = (\Phi(A_i) \cdot \Phi(A_j))$ ($i, j = 1, 2, \dots, N$), $(\Phi(A_i) \cdot \Phi(A_j))$ is the inner product in H .

The twin least squares support vector regression model with mixed noise characteristics (M-TLSSVR) is proposed. The primal problem with model M-TLSSVR is shown below

$$\begin{aligned} \text{Min}\{g_{P_{M-TLSSVR}} &= \frac{1}{2}\omega_1^T \cdot \omega_1 + \frac{C_1}{N} \cdot [\lambda_1 \cdot \sum_{i=1}^N (l_1(\xi_i)) + \lambda_2 \cdot \sum_{i=1}^N (l_2(\xi_i))]\} \\ \text{s.t.} \quad y_i &= \omega_1^T \cdot \phi(A_i) + b_1 - \xi_i \end{aligned} \tag{2}$$

$$\begin{aligned} \text{Min}\{g_{P_{M-TLSSVR}} &= \frac{1}{2}\omega_2^T \cdot \omega_2 + \frac{C_2}{N} \cdot [\lambda_3 \cdot \sum_{i=1}^N (l_1(\xi_i^*)) + \lambda_4 \cdot \sum_{i=1}^N (l_2(\xi_i^*))]\} \\ \text{s.t.} \quad y_i &= \omega_2^T \cdot \phi(A_i) + b_2 + \xi_i^* \end{aligned} \tag{3}$$

where ω_1, ω_2 denotes the weight vector and b_1, b_2 is the bias term, $\Phi(A)$ is the nonlinear mapping that transfers the input vector to a higher-dimensional feature space. ξ_i, ξ_i^* are random slack variable at time i . $l_1(\xi_i), l_1(\xi_i^*), l_2(\xi_i), l_2(\xi_i^*) > 0 (i = 1, 2, \dots, N)$ be general convex empirical risk loss values for noise characteristic in the sample point $(A_i, y_i) \in D_N ((i, j = 1, 2, \dots, N))$. $C_1 > 0, C_2 > 0$ be the penalty parameter, weight factor $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$, and $\lambda_1 + \lambda_2 = 1, \lambda_3 + \lambda_4 = 1$.

According to the literature [28], the mixed noise model is distributed by multiple noises, and its performance is better than the single noise model. In this section, Gauss–Laplace mixed homoscedastic and heteroscedastic noise distributions are used to describe complex noise characteristics.

3.1. TLSSVR Model of G-L Mixed Homoscedastic Noise Characteristics

According to Bayesian principle, it concludes that the empirical risk loss of the homoscedastic Gaussian noise of the lower bound function is $l_1(\xi) = \frac{1}{2\sigma^2} \cdot \xi^2$, the Laplace noise is $l_2(\xi) = |\xi|$. Adopting G-L mixed homoscedastic noise distribution to fit complicated noise-characteristic, by Equation (1), the empirical risk loss about G-L mixed homoscedastic noise is $l(\xi) = \frac{\lambda_1}{2\sigma^2} \cdot \xi^2 + \lambda_2 \cdot |\xi|$. The lower bound function of the G-L mixed homoscedastic noise characteristic TLSSVR model (GLM-TLSSVR) is proposed, the primal problem of the lower bound function is depicted as

$$\begin{aligned} \text{Min}\{g_{P_{GLM-TLSSVR}} &= \frac{1}{2}\omega_1^T \cdot \omega_1 + \frac{C_1}{N} \cdot (\frac{\lambda_1}{2\sigma^2} \cdot \sum_{i=1}^N \xi_i^2 + \lambda_2 \cdot \sum_{i=1}^N |\xi_i|)\} \\ \text{s.t. } y_i &= \omega_1^T \cdot \phi(A_i) + b_1 - \xi_i \end{aligned} \tag{4}$$

Similarly, we can get that the primal problem of the upper bound function of the model GLM-TLSSVR is

$$\begin{aligned} \text{Min}\{g_{P_{GLM-TLSSVR}} &= \frac{1}{2}\omega_2^T \cdot \omega_2 + \frac{C_2}{N} \cdot (\frac{\lambda_3}{2\sigma^{*2}} \cdot \sum_{i=1}^N \xi_i^{*2} + \lambda_4 \cdot \sum_{i=1}^N |\xi_i^*|)\} \\ \text{s.t. } y_i &= \omega_2^T \cdot \phi(A_i) + b_2 + \xi_i^* \end{aligned} \tag{5}$$

Where ξ_i and ξ_i^* are the random noise and slack variables at time i . parameter vector $\omega_1, \omega_2 \in R^n$, σ^2, σ^{*2} are homoscedastic, $C_1, C_2 > 0$ are a penalty parameter, and the weight factors are $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$, $\lambda_1 + \lambda_2 = 1, \lambda_3 + \lambda_4 = 1$.

Proposition 1. The solution of primal problem (4), (5) of GLM-TLSSVR about ω_1, ω_2 exist and are unique.

Theorem 1. The dual problem of primal problem (4) of GLM-TLSSVR is

$$\begin{aligned} \text{Max}\{g_{D_{GLM-TLSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i + \beta_i) \cdot (\alpha_j + \beta_j) \cdot K(A_i, A_j) - \\ &\frac{C_1\sigma^2}{N} \cdot \frac{\lambda_2^2}{\lambda_1} \cdot \sum_{i=1}^N \frac{\alpha_i}{\beta_i} - \frac{N\sigma^2}{2C_1\lambda_1} \sum_{i=1}^N \alpha_i^2\} \\ \text{s.t. } \sum_{i=1}^N (\alpha_i + \beta_i) &= 0 \end{aligned} \tag{6}$$

The dual Problem of primal problem (5) of GLM-TLSSVR is

$$\begin{aligned} \text{Max}\{g_{D_{GLM-TLSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^* + \beta_i^*) \cdot (\alpha_j^* + \beta_j^*) \cdot K(A_i, A_j) - \\ &\frac{C_2\sigma^{*2}}{N} \cdot \frac{\lambda_4^2}{\lambda_3} \cdot \sum_{i=1}^N \frac{\alpha_i^*}{\beta_i^*} - \frac{N\sigma^{*2}}{2C_2\lambda_3} \cdot \sum_{i=1}^N \alpha_i^{*2}\} \\ \text{s.t. } \sum_{i=1}^N (\alpha_i^* + \beta_i^*) &= 0 \end{aligned} \tag{7}$$

where parameter vector $\omega_1, \omega_2 \in R^n$, σ^2, σ^{*2} are homoscedastic, $C_1, C_2 > 0$ are a penalty parameter, and the weight factors are $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$, $\lambda_1 + \lambda_2 = 1, \lambda_3 + \lambda_4 = 1$, $\alpha_i, \alpha_j, \alpha_i^*, \alpha_j^*$ are the Lagrange multiplier.

Proof. On the lower bound function of the GLM-TLSSVR model, for any vector u , If we set $u_{\pm} \geq 0$ to $u = u_+ - u_-$, then $\min |u| = \min\{u_+ + u_-\}$ will be established [54]. Therefore, by setting $\xi_i = p_i - r_i$, $r_i, p_i \geq 0$, the primal problem of the lower bound function of GLM-TLSSVR is simplified, as follows

$$\begin{aligned}
 \text{Min}\{g_{\text{PGLM-TLSSVR}} &= \frac{1}{2}\omega_1^T \cdot \omega_1 + \frac{C_1}{N} \cdot [\frac{\lambda_1}{2\sigma^2} \cdot \sum_{i=1}^N \xi_i^2 + \lambda_2 \cdot \sum_{i=1}^N (r_i + p_i)]\} \\
 \text{s.t. } y_i - (\omega_1^T \phi(A_i) + b_1) - r_i + p_i &= 0 \\
 y_i = \omega_1^T \phi(A_i) + b_1 - \xi_i & \\
 r_i, p_i \geq 0 (i = 1, \dots, N) &
 \end{aligned}
 \tag{8}$$

We introduce the Lagrange function and KKT(Karush–Kuhn–Tucker) condition [55]. □

We get the solution of the lower bound function

$$\begin{aligned}
 \omega_{1i} &= \sum_{i=1}^N (\alpha_i + \beta_i) \cdot \phi(A_i), \\
 b_1 &= \sum_{i=1}^N [y_i - \sum_{j=1}^N (\alpha_i + \beta_i) \cdot K(A_i, A_j) - \frac{1}{\lambda_1} \cdot \frac{N \cdot \sigma^2 \cdot \alpha_i}{C_1}].
 \end{aligned}$$

Thus, the lower function of model TLSSVR with Gauss–Laplace mixture homoscedastic noise characteristic (GLM-TLSSVR) can be written as

$$f_1(A) = \omega_1^T \cdot \phi(A) + b_1 = \sum_{i=1}^N (\alpha_i + \beta_i)K(A_i, A) + b_1$$

Theorem 2. The primal problem of the upper bound function of GLM-TLSSVR is simplified, as follows

$$\begin{aligned}
 \text{Min}\{g_{\text{PGLM-TLSSVR}} &= \frac{1}{2}\omega_2^T \cdot \omega_2 + \frac{C_2}{N} \cdot (\frac{\lambda_3}{2\sigma^{*2}} \cdot \sum_{i=1}^N \xi_i^{*2} + \lambda_4 \cdot \sum_{i=1}^N (r_i + p_i))\} \\
 \text{s.t. } y_i - (\omega_2^T \phi(A_i) + b_2) + r_i - p_i &= 0 \\
 y_i = \omega_2^T \phi(A_i) + b_2 + \xi_i^* & \\
 r_i, p_i \geq 0 (i = 1, \dots, N) &
 \end{aligned}
 \tag{9}$$

Similarly, we introduce the Lagrange function and KKT conditions again.

We obtain the solution of the upper bound function

$$\begin{aligned}
 \omega_{2i} &= \sum_{i=1}^N (\alpha_i^* + \beta_i^*) \cdot \phi(A_i), \\
 b_2 &= \sum_{i=1}^N [y_i - \sum_{j=1}^N (\alpha_i^* + \beta_j^*) \cdot K(A_i, A_j) - \frac{1}{\lambda_3} \cdot \frac{N \cdot \sigma^{*2} \cdot \alpha_i^*}{C_2}].
 \end{aligned}$$

Thus, the upper function of model TLSSVR with Gauss–Laplace mixture homoscedastic noise characteristic (GLM-TLSSVR) can be written as

$$f_2(A) = \omega_2^T \cdot \phi(A) + b_2 = \sum_{i=1}^N (\alpha_i^* + \beta_i^*)K(A_i, A) + b_2$$

At last, the estimated regressor of GLM-TLSSVR is written, as follows

$$f(A) = \frac{\omega_1^T + \omega_2^T}{2} \cdot \phi(A) + \frac{b_1 + b_2}{2} = \sum_{i=1}^N \frac{\alpha_i + \beta_i + \alpha_i^* + \beta_i^*}{2} \cdot K(A_i, A) + \frac{b_1 + b_2}{2}$$

where, parameter vector $\omega_1, \omega_2 \in R^n, \phi : R^n \rightarrow H, (\phi(A_i) \cdot \phi(A_j))$ is the inner product of $H, K(A_i, A_j) = (\phi(A_i) \cdot \phi(A_j))$ is the kernel function.

3.2. TLSSVR Model of G-L Mixed Heteroscedastic Noise Characteristics

If the noise is Gaussian noise with zero mean and variance of heteroscedasticity, and these variance are $\sigma_i^2, (\sigma_i^*)^2$, where $\sigma_i \neq \sigma_j, \sigma_i^* \neq \sigma_j^*, i \neq j(i, j = 1, \dots, N)$. Similarly, by Equation (1), the empirical risk loss is $l(\xi_i) = \frac{\lambda_1}{2\sigma_i^2} \cdot \xi_i^2 + \lambda_2 \cdot |\xi_i|, (i = 1, \dots, N)$. A TLSSVR model for Gauss–Laplace mixed heteroscedastic noise characteristics is established, and we named it GLMH-TLSSVR. A pair of optimization problems of GLMH-TLSSVR can be depicted:

$$\begin{aligned} \text{Min}\{g_{P_{GLM-TLSSVR}} &= \frac{1}{2}\omega_1^T \cdot \omega_1 + \frac{C_1}{N} \cdot (\frac{\lambda_1}{2\sigma_i^2} \cdot \sum_{i=1}^N \xi_i^2 + \lambda_2 \cdot \sum_{i=1}^N |\xi_i|)\} \\ \text{s.t. } y_i &= \omega_1^T \cdot \phi(A_i) + b_1 - \xi_i \end{aligned} \tag{10}$$

$$\begin{aligned} \text{Min}\{g_{P_{GLM-TLSSVR}} &= \frac{1}{2}\omega_2^T \cdot \omega_2 + \frac{C_2}{N} \cdot (\frac{\lambda_3}{2\sigma_i^{*2}} \cdot \sum_{i=1}^N \xi_i^{*2} + \lambda_4 \cdot \sum_{i=1}^N |\xi_i^*|)\} \\ \text{s.t. } y_i &= \omega_2^T \cdot \phi(A_i) + b_2 + \xi_i^* \end{aligned} \tag{11}$$

where ξ_i and ξ_i^* are random noise and slack variables at time i . These heteroscedastic variables are $\sigma_i^2, (\sigma_i^*)^2 (i = 1, 2, \dots, N)$, $C_1, C_2 > 0$ are the penalty parameter, weight factor $\lambda_1, \lambda_2, \lambda_3, \lambda_4 \geq 0$, and $\lambda_1 + \lambda_2 = 1, \lambda_3 + \lambda_4 = 1$.

Proposition 2. *The solution of primal problem (10), (11) of GLMH-TLSSVR about ω_1, ω_2 exist and are unique.*

Theorem 3. *The dual problem of GLMH-TLSSVR in primal problem (10), (11) be*

$$\begin{aligned} \text{Max}\{g_{D_{GLM-TLSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i + \beta_i) \cdot (\alpha_j + \beta_j) \cdot K(A_i, A_j) - \\ &\frac{C_1 \sigma_i^2}{N} \cdot \frac{\lambda_2^2}{\lambda_1} \cdot \sum_{i=1}^N \frac{\alpha_i}{\beta_i} - \frac{N \sigma_i^2}{2C_1 \lambda_1} \sum_{i=1}^N \alpha_i^2\} \\ \text{s.t. } \sum_{i=1}^N (\alpha_i + \beta_i) &= 0 \end{aligned} \tag{12}$$

$$\begin{aligned} \text{Max}\{g_{D_{GLM-TLSSVR}} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i^* + \beta_i^*) \cdot (\alpha_j^* + \beta_j^*) \cdot K(A_i, A_j) - \\ &\frac{C_2 \sigma_i^{*2}}{N} \cdot \frac{\lambda_4^2}{\lambda_3} \cdot \sum_{i=1}^N \frac{\alpha_i^*}{\beta_i^*} - \frac{N \sigma_i^{*2}}{2C_2 \lambda_3} \cdot \sum_{i=1}^N \alpha_i^{*2}\} \\ \text{s.t. } \sum_{i=1}^N (\alpha_i^* + \beta_i^*) &= 0 \end{aligned} \tag{13}$$

Proof. Similar to Theorems 1 and 2, an appendix to the proof of Theorem 3. □

We can obtain the solution of the lower bound function

$$\begin{aligned} \omega_{1i} &= \sum_{i=1}^N (\alpha_i + \beta_i) \cdot \phi(A_i), \\ b_1 &= \sum_{i=1}^N [y_i - \sum_{j=1}^N (\alpha_i + \beta_j) \cdot K(A_i, A_j) - \frac{1}{\lambda_1} \cdot \frac{N \cdot \sigma_i^2 \cdot \alpha_i}{C_1}]. \end{aligned}$$

Thus, the lower function of model TLSSVR with Gauss–Laplace mixture heteroscedastic noise characteristics (GLMH-TLSSVR) can be written as

$$f_1(A) = \omega_1^T \cdot \phi(A) + b_1 = \sum_{i=1}^N (\alpha_i + \beta_i) K(A_i, A) + b_1$$

We also get the solution of the upper bound function

$$\omega_{2i} = \sum_{i=1}^N (\alpha_i^* + \beta_i^*) \cdot \phi(A_i), b_2 = \sum_{i=1}^N [y_i - \sum_{j=1}^N (\alpha_i^* + \beta_i^*) \cdot K(A_i, A_j) - \frac{1}{\lambda_3} \cdot \frac{N \cdot \sigma_i^{*2} \cdot \alpha^*}{C_2}].$$

The upper function of model TLSSVR with Gauss–Laplace mixture heteroscedastic noise characteristics (GLMH-TLSSVR) can be written as

$$f_2(A) = \omega_2^T \cdot \phi(A) + b_2 = \sum_{i=1}^N (\alpha_i^* + \beta_i^*) K(A_i, A) + b_2$$

At last, the estimated regressor of GLMH-TLSSVR is written, as follows

$$f(A) = \frac{\omega_1^T + \omega_2^T}{2} \cdot \phi(A) + \frac{b_1 + b_2}{2} = \sum_{i=1}^N \frac{\alpha_i + \beta_i + \alpha_i^* + \beta_i^*}{2} \cdot K(A_i, A) + \frac{b_1 + b_2}{2}$$

If this noise characteristic is Gaussian with the homoscedasticity, we can use Theorem 3 in order to derive Theorems 1 and 2.

4. ALM Method Analysis

In this section, we apply the augmented Lagrange multiplier method (ALM) [56] to solve the duality problems in Equations (6) and (7) by applying gradient descent or Newton’s method to equality-constrained sequences. By eliminating the equality constraints, any equality constraints can be reduced to the equivalent unconstrained problem [57,58]. When we deal with some large-scale data sets, some rapid optimizations can combine these techniques with the proposed model. For example, the sequential minimum optimization (SMO) algorithm [59] and stochastic gradient appropriate (SDG) algorithm [60].

From Theorems 1–3, we can find that this ALM method can help us to effectively identify the GLM-TLSSVR and GLMH-TLSSVR models. In this section, the lower bound function and upper bound function of the GLM-TLSSVR model can be solved by the ALM method. Similarly, the lower bound function and upper bound function of the GLMH-TLSSVR model can also be solved by the ALM method. The specific algorithm steps are as follows

- (1) Set data-set be $D_N = \{(A_1, y_1), (A_2, y_2), \dots, (A_L, y_N)\}$, where $A_i \in R^n, y_i \in R, i = 1, \dots, N$.
- (2) Select the appropriate kernel function through the 10-fold cross-validation strategy and obtain the appropriate parameters $C_1, C_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4$ of the lower and upper bound function of the model GLM-TLSSVR.
- (3) When the optimization problem is solved in Equations (6) and (7), we can obtain the optimal solution $\alpha = (\alpha_1, \dots, \alpha_N), \alpha^* = (\alpha_1^*, \dots, \alpha_N^*), \beta = (\beta_1, \dots, \beta_N), \beta^* = (\beta_1^*, \dots, \beta_N^*)$.
- (4) The decision function is established, as shown below

$$f(A) = \omega^T \cdot \phi(A) + b = \sum_{i=1}^N \frac{\alpha_i + \alpha_i^* + \beta_i + \beta_i^*}{2} \cdot K(A_i, A) + b$$

5. Experiments and Discussion

In the section, to check the performance of the proposed model GLM-TLSSVR, we compared it with ν -SVR, LS-SVR, and TSVR on actual data-set from Heilongjiang, China. This part mainly includes three contents: G-L mixed noise characteristics of wind speed in Section 5.1; the criteria for algorithm evaluation in Section 5.2; and, application on predicting the short-term wind speed in Section 5.3.

5.1. G-L Mixed Noise Characteristics of Wind Speed

What we collected consists of one-year wind speed data-set from Heilongjiang Province, China. These data record the wind speed value every 10 min in order to better analyze the characteristics of mixed noise in the wind speed forecast error. In the above wind speed data, we found that some noise is a mixture of Gauss–Laplace. Some of the researchers have found that turbulence is the main cause of the uncertainty of strong random fluctuations in wind speed. From the perspective of wind energy, the most significant feature of wind resources is its variability.

We adopted the persistence method, which is often used to study the distribution of wind speed forecast errors, in order to analyze the wind speed data set of a one-month time series [54]. This experiment shows that the error variable ζ does not satisfy a single noise distribution, but approximately obeys the Gauss–Laplace mixed noise distribution, and the PDF of ζ is $P(\zeta) = \frac{1}{2}e^{-|\zeta|} \cdot \frac{1}{2\sigma^2}\zeta^2$, we show the forecast error of Gauss–Laplace mixed wind speed distribution in Figure 3. It is found that this is a regression learning task about mixed noise.

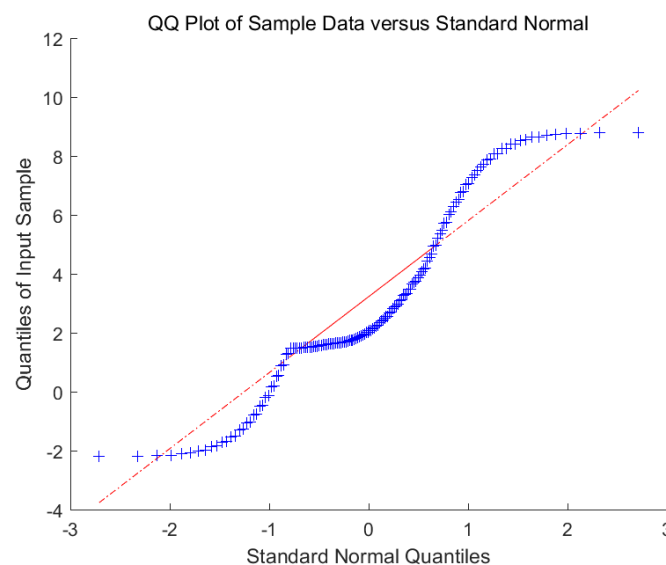


Figure 3. The wind speed forecast error distribution with mixed Gauss-Laplace noise. (This red line is used as a reference. It is determined by the quarter point and the third quarter point. These two points just determine the line in the QQ plot. These blue distribution points are the error between the actual value of wind speed and the predicted value of wind speed.).

5.2. The Criteria for Algorithm Evaluation

We specified evaluation criteria before introducing the experimental results in order to compare the performance of various models. The evaluation criteria are, as follows: the mean absolute error (MAE), the root mean square error (RMSE), sum of squared regression (SSR), sum of squared deviation of testing (SST), sum of squared error of testing (SSE), and teTime are used to evaluate the predictive performance of models ν -SVR, LS-SVR, TSVR, and GLM-TLSSVR. The five criteria are defined, as follows [34,37].

In Table 1, L is the number of testing samples, y_i is the i th the real value, y_i^* represents the predicted value, and \bar{y} is the mean of the testing data-set. teTime(in seconds) represents the testing time of constructing a regressor.

Table 1. Evaluation criteria for short-term wind speed prediction.

Parameter	Mathematical Expression
MAE	$\frac{1}{L} \sum_{i=1}^N y_i^* - y_i $
RMSE	$\sqrt{\frac{1}{L} \sum_{i=1}^N (y_i^* - y_i)^2}$
SSE	$\sum_{i=1}^L (y_i^* - y_i)^2$
SSR	$\sum_{i=1}^L (y_i^* - \bar{y})^2$
SST	$\sum_{i=1}^L (y_i - \bar{y})^2$
SSE/SST	$\sum_{i=1}^L (y_i^* - y_i)^2 / (y_i - \bar{y})^2$
SSR/SST	$\sum_{i=1}^L (y_i^* - \bar{y})^2 / (y_i - \bar{y})^2$

5.3. Application on Predicting the Short-Term Wind Speed

In the section, we confirmed the feasibility and effectiveness of the proposed model GLM-TLSSVR on the short-term wind speed data set of Heilongjiang Province, China. The source of the wind speed data set is a related wind farm under the Meteorological Bureau of Heilongjiang Province, and a lightning imager measures the wind speed. This wind speed data set has been recorded for more than a year, and the average wind speed is recorded every 10 min. In general, we collected a total of 62,466 samples, which have four attributes, namely variance, mean, maximum, and minimum. We use 1440 uninterrupted data samples (from 1 to 1440, the time span is 10 days) as training samples. We also use 720 uninterrupted data samples (from 1441 to 2160, the time span is five days), and 80 consecutive data as the testing samples. As for the original sequence, we need to transform it into a multiple regression task by using mode $\vec{X}_i = (X_{i-11}, X_{i-10}, \dots, X_{i-1}, X_i)$ as an input vector to predict X_{i+step} , where the vector orders of wind speed is determined by the chaotic operator network method. Where X_j is the real value of wind speed at time $j(j = i - 11, i - 10, \dots, i)$. In the experiments, we try step = 1, 3, and 5. In other words, we predicted the wind speed of every point X_i after 10, 30, and 50 min, respectively.

These four models (ν -SVR, LS-SVR, TSVR, and GLM-TLSSVR) have been implemented in Python 3.7 on Windows 10 running on a PC with system configuration Intel i7 processor (3.19 GHz) with 8 GB of RAM. The initial parameters be $C_1, C_2 \in 2^i | i = -9, -8, \dots, 10, \lambda_1, \lambda_2, \lambda_3, \lambda_4 \in [0, 1]$. $C_1, C_2, \lambda_1, \lambda_2, \lambda_3, \lambda_4$ are some tuned parameters by virtue of the 10-fold cross validation technique, where the cross validation technique is explained in detail in [61,62]. This technique can help us to find the optimal parameters. In this article, in order to reduce the computational burden of the GLM-TLSSVR model, the parameter assignments are, as follows: $C_1 = C_2, \lambda_1 = \lambda_2 = \frac{1}{2}, \lambda_3 = \lambda_4 = \frac{1}{2}$. As for the choice of kernel function, many experiments show that polynomial kernel function and Gaussian kernel function have good performance. In this experiment, we apply Gaussian kernel functions and polynomial kernel function to these four models (ν -SVR, LS-SVR, TSVR, and GLM-TLSSVR), as below [63].

$$K(X_i, X_j) = ((X_i, X_j) + 1)^d,$$

$$K(X_i, X_j) = e^{-\frac{\|X_i - X_j\|^2}{\sigma^2}},$$

where d is a positive integer, and σ is positive.

The dual problem of ν -SVR, LS-SVR, and TSVR are as follows.

ν -SVR: the authors of [18,61] define the dual problem of ν -SVR, as

$$\begin{aligned} \text{Max}\{g_{D_{\nu\text{-SVR}}}\} &= -\frac{1}{2} \sum_{i \in \text{RSV}} \sum_{j \in \text{RSV}} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j) \cdot K(A_i, A_j) - \sum_{i \in \text{RSV}} (\alpha_i^* - \alpha_i) \cdot y_i \\ \text{s.t. :} \quad & \sum_{i=1}^N (\alpha_i^* - \alpha_i) = 0, \\ & 0 \leq \alpha_i^* \leq \frac{C}{N}, \\ & \sum_{i=1}^N (\alpha_i + \alpha_i^*) \leq C \cdot \nu, i = 1, \dots, N. \end{aligned} \quad (14)$$

LS-SVR: the authors of [64] define the dual problem of LS-SVR as

$$\begin{aligned} \text{Max}\{g_{D_{\text{LS-SVR}}}\} &= -\frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (\alpha_i \cdot \alpha_j \cdot K(A_i, A_j)) + \sum_{i=1}^N (\alpha_i \cdot y_i) - \frac{N}{2C} \cdot \sum_{i=1}^N \alpha_i^2 \\ \text{s.t. :} \quad & \sum_{i=1}^N \alpha_i = 0. \end{aligned} \quad (15)$$

TSVR: the authors of [34] define the dual problem of TSVR, as

$$\begin{aligned} \text{Max}\{g_{D_{\text{TSVR}}}\} &= -\frac{1}{2} \alpha^T H (H^T H)^{-1} H^T \alpha + f^T H (H^T H)^{-1} H^T \alpha - f^T \alpha \\ \text{s.t. :} \quad & 0 \leq \alpha \leq C_1 e. \end{aligned} \quad (16)$$

$$\begin{aligned} \text{Max}\{g_{D_{\text{TSVR}}}\} &= -\frac{1}{2} \gamma^T H (H^T H)^{-1} H^T \gamma - h^T H (H^T H)^{-1} H^T \gamma + h^T \gamma \\ \text{s.t. :} \quad & 0 \leq \gamma \leq C_2 e. \end{aligned} \quad (17)$$

where, $H = [K(A, A^T)e]$.

In Figure 4, wind-speed forecasting-results at A_i -point of the above four models are presented after 10 min. Figure 5 shows the error statistic of wind-speed prediction using the above four models after 10 min. In Figure 6, wind-speed forecasting-results at A_i -point of the above four models are presented after 30 min. Figure 7 shows the error statistic of wind-speed prediction using the above four models after 30 min. In Figure 8, wind-speed forecasting-results at A_i -point of the above four models are presented after 50 min. Figure 9 shows the error statistic of wind-speed prediction using the above four models after 50 min. Tables 2–4 display the statistical criteria of MAE, RMSE, SSE/SST, SSR/SST, and teTime.

Table 2. Error statistics of four short-term wind speed forecasting models after 10 min.

Model	MAE (m/s)	RMSE (m/s)	SSE/SST	SSR/SST	teTime (s)
ν -SVR	0.4797	0.6799	0.2603	0.4552	0.68
LS-SVR	0.4434	0.6366	0.2282	0.5064	0.66
TSVR	0.4182	0.6161	0.2137	0.5270	0.56
GLM-TLSSVR	0.4091	0.6069	0.2074	0.5384	0.55

Table 3. Error statistics of four short-term wind speed forecasting models after 30 min.

Model	MAE (m/s)	RMSE (m/s)	SSE/SST	SSR/SST	teTime (s)
<i>v</i> -SVR	0.7596	1.0041	0.4378	0.2365	0.71
LS-SVR	0.7131	0.9466	0.3891	0.2932	0.68
TSVR	0.6167	0.8546	0.3171	0.3793	0.59
GLM-TLSSVR	0.5787	0.8204	0.2923	0.4197	0.57

Table 4. Error statistics of four short-term wind speed forecasting models after 50 min.

Model	MAE (m/s)	RMSE (m/s)	SSE/SST	SSR/SST	teTime (s)
<i>v</i> -SVR	0.7781	0.9877	0.4333	0.2227	0.77
LS-SVR	0.7252	0.9202	0.3761	0.2714	0.69
TSVR	0.6566	0.8485	0.3198	0.3287	0.65
GLM-TLSSVR	0.6121	0.8005	0.2847	0.3702	0.58

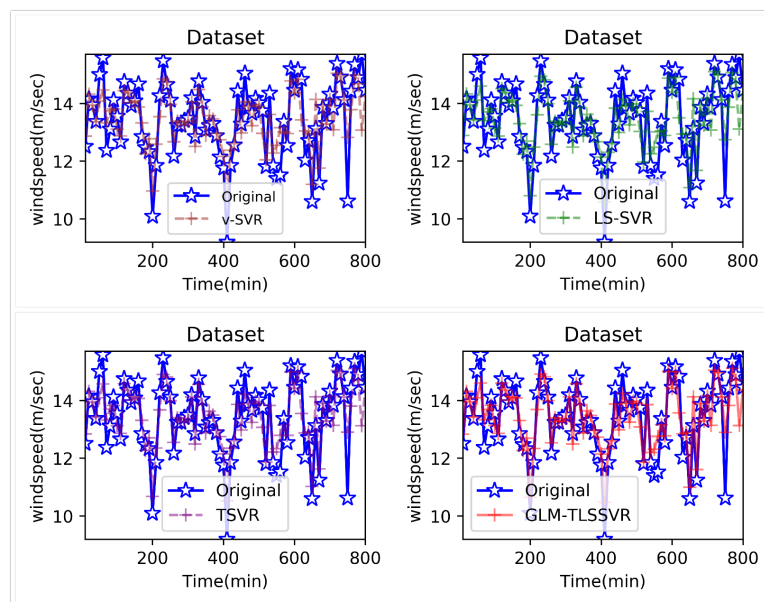


Figure 4. Result of four short-term wind speed forecasting models after 10 min.

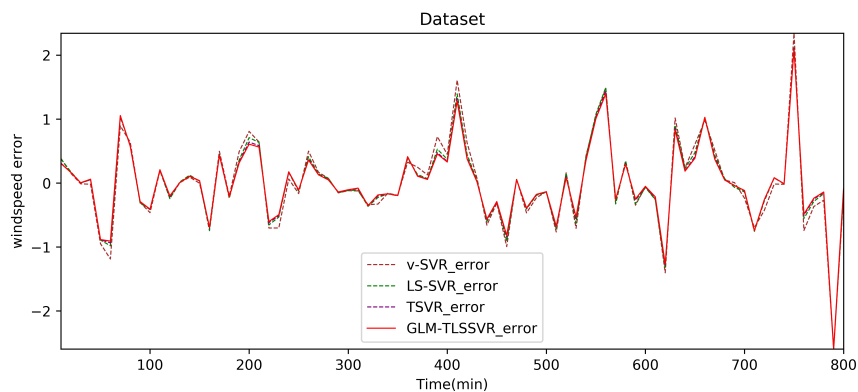


Figure 5. Error of four short-term wind speed forecasting models after 10 min.

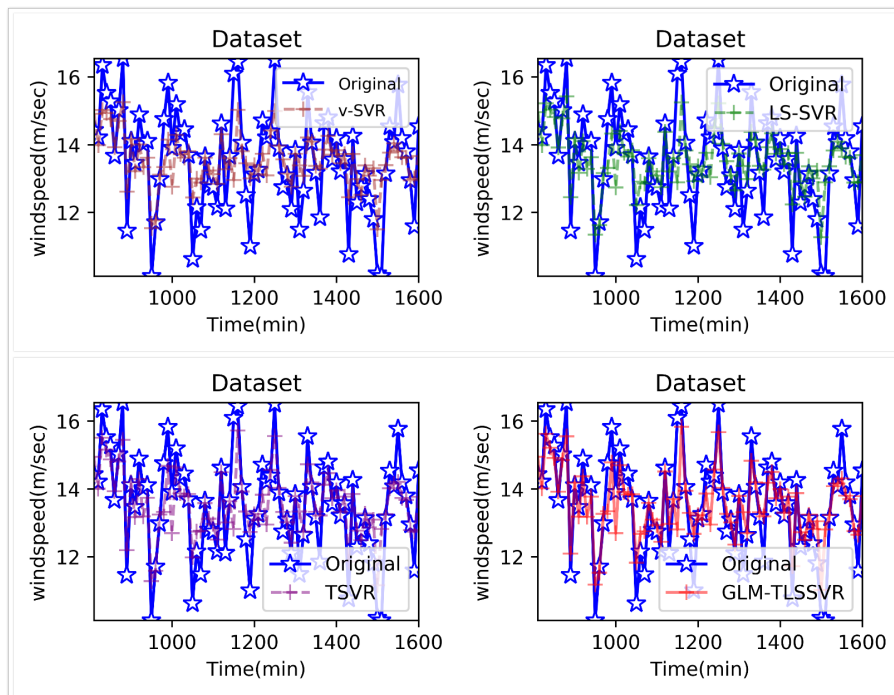


Figure 6. Result of four short-term wind speed forecasting models after 30 min.

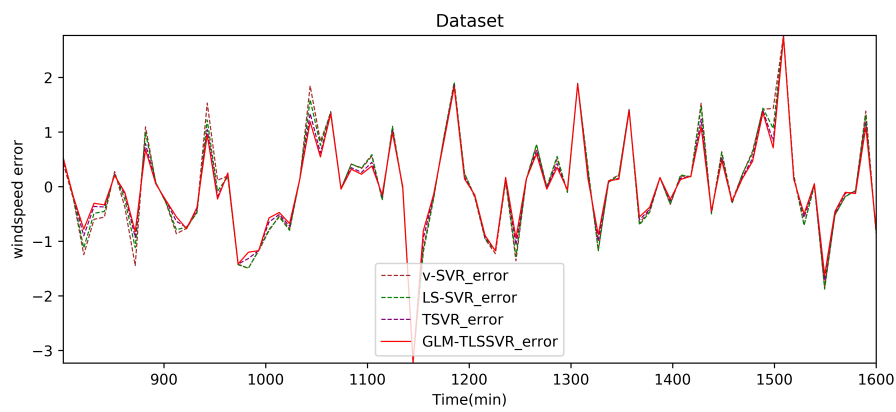


Figure 7. Error of four short-term wind speed forecasting models after 30 min.

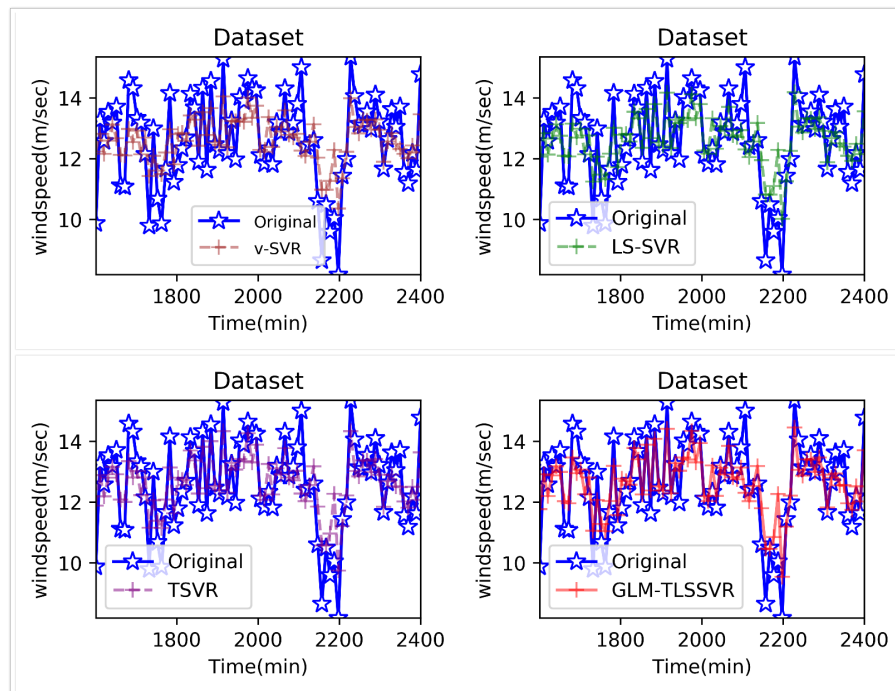


Figure 8. Result of four short-term wind speed forecasting models after 50 min.

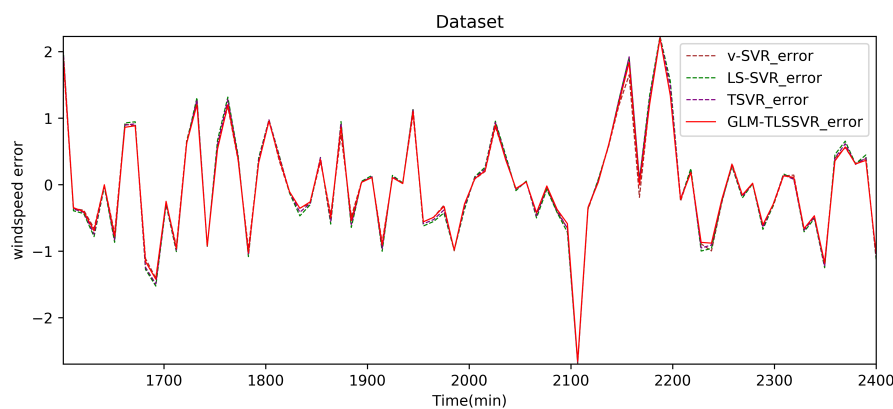


Figure 9. Error of four short-term wind speed forecasting models after 50 min.

From Tables 2–4 and Figures 4–9, these evaluation criteria can indicate that the error statistic of GLM-TLSSVR model is better than that of models ν -SVR, LS-SVR, and TSVR. As the forecast time interval increases from 10-min. to 30-min. and 50-min., the forecasting error of the four models increases and the relative error decreases. Therefore, in these cases, it is not so important. However, as can be seen from Tables 2–4, under all conditions of MAE, RMSE, SSE/SST, and SSR/SST, the model GLM-TLSSVR with Gaussian–Laplace mixed noise characteristics is slightly better than the other three classical ν -SVR, LS-SVR and TSVR models. In general, a lower value of MAE, RMSE, and SSE/SST reflects the consistency between the predicted values and true values, while the higher values of SSR/SST indicate that the regressor accounts for higher statistical information. Further, the performance indices indicate that GLM-TLSSVR outperforms ν -SVR, LS-SVR, and TSVR for short-term wind speed data set in terms of SSE/SST, RMSE, and MAE. The ratio of SSR/SST can estimate the goodness of fit of the predictive model and extract the maximum information from the data set. Therefore, the proposed model GLM-TLSSVR is considered to be the best regression indicator among all of the models. The SSE/SST is lower for GLM-TLSSVR when compared to other methods that imply good estimation between real values and predictive values from Tables 2–4. In addition, among all of the models, the computational cost of testing model GLM-TLSSVR is the

lowest, which indicates that our proposed iterative methods are the efficient algorithm for regression, on the other hand, the reason is that this proposed model GLM-TLSSVR combines the spirit with the fast speed of LS-SVR yields a new regressor. In addition, the generalization performance of the proposed model GLM-TLSSVR is best, i.e., it owns the smallest and largest evaluation criteria, respectively, viz. RMSE, and SSR/SST from Tables 2–4; this is mainly due to the idea of twin hyperplanes.

6. Conclusions

Many regression techniques today assume that this model is a single noise characteristic. Wind speed prediction is complicated by its volatility and uncertainty, so it is difficult to model with a single noise distribution. This section summarizes our main work: (1) we use the Bayesian principle to derive the best empirical risk loss of G-L mixed noise characteristics; (2) the TLSSVR model of G-L mixed homoscedastic noise (GLM-TLSSVR) and G-L mixed heteroscedastic noise (GLMH-TLSSVR) for complicate noise is developed; (3) use the Lagrange function and obtain the dual problem of GLM-TLSSVR and GLMH-TLSSVR according to KKT conditions; (4) solve the GLM-TLSSVR by the ALM method, ensuring the stability and effectiveness of the algorithm; (5) use the proposed technique to predict the future short-term wind speed, calculate wind speed based on past data, and then predict wind speed at some time after 10, 30, and 50 min, respectively. Based on our results, it is observed that GLM-TLSSVR outperforms ν -SVR, LS-SVR, and TSVR for the short-term wind speed data-set, as shown in the experiment. Further, the ratio of SSR/SST can estimate the goodness of fit of the predictive model and extract the maximum information from the data set. Therefore, the proposed model GLM-TLSSVR is considered to be the best regression indicator among all of the models. A low ratio of SSE/SST implies good estimation between real values and predictive values. In addition, the computational time for all the models is evaluated and it is found that GLM-TLSSVR is the lowest, owing to its smaller sized constrained optimization. These results also bring many benefits to the industrial sector, such as better statistical analysis of the relationship between wind speed characteristics and power generation.

There are uncertainties in the data in some actual regression problems. Uncertainty, like this accident, is mainly reflected in the uncertain time of the accident, the uncertain situation of the accident, and the uncertain direction of the accident. We should study the regression algorithm of fuzzy uncertainty with mixed noise characteristics models. In addition, our work only discusses the problem of regression models with Gaussian–Laplace mixed noise characteristics. In fact, we can develop similar problems to classification learning. In a similar idea, we can still study the classification problems with Gaussian–Laplace mixed noise characteristics in the future.

Author Contributions: Conceptualization, S.Z.; Formal analysis & Methodology, S.Z. and C.L.; Writing—original draft, S.Z. and C.L.; Writing—review & editing, W.W. and B.C. All authors have read and approved the final published version. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Key Scientific Research Project of Colleges and Universities of Henan Province (No. 21A520020) and National Natural Science Foundation of China (NSFC) (Nos. 11702087 and 62076089).

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

Abbreviations

The following abbreviations are used in this manuscript:

ν -SVR	ν -Support vector regression
LS-SVR	Least squares support vector regression model
TSVR	Twin support vector regression model
GLM-TLSSVR	Twin LS-SVR model of Gaussian-Laplacian mixed homoscedastic-noise
ALM	Augmented Lagrange multiplier method

References

1. Vapnik, V. *The Natural of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; pp. 521–576.
2. Debnath, R.; Muramatsu, M.; Takahashi, H. An efficient support vector machine learning method with second-order cone programming for large-scale problems. *Appl. Intell.* **2005**, *23*, 219–239. [[CrossRef](#)]
3. Nie, F.; Huang, H.; Cai, X.; Ding, C.H. Efficient and robust feature selection via joint ℓ_2 , 1-norms minimization. In *International Conference on Neural Information Processing Systems*; Curran Associates Inc.: New York, NY, USA, 2010; pp. 1813–1821.
4. Claesen, M.; De Smet, F.; Suykens, J.A.; De Moor, B. A robust ensemble approach to learn from positive and unlabeled data using SVM base models. *Neurocomputing* **2015**, *160*, 73–74. [[CrossRef](#)]
5. Osuna, E.; Freund, R.; Girosi, F. Training support vector machines: An application to face detection. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Juan, PR, USA, 17–19 June 1997; pp. 130–136.
6. Lee, S.; Park, C.; Koo, J.-Y. Feature selection in the Laplacian support vector machine. *Comput. Stat. Data Anal.* **2011**, *55*, 567–577. [[CrossRef](#)]
7. Chuang, C.-C.; Su, S.-F.; Jeng, J.-T.; Hsiao, C.-C. Robust support vector regression networks for function approximation with outliers. *IEEE Trans. Neural Net-Works* **2002**, *13*, 1322–1330. [[CrossRef](#)]
8. Ince, H.; Trafalis, T.B. Support vector machine for regression and applications to financial forecasting. In *Proceedings of the International Joint Conference on Neural Networks, IEEE-INNS-ENNS*, Como, Italy, 27–27 July 2000.
9. Pandit, R.; Infield, D. Comparative analysis of binning and support vector regression for wind turbine rotor speed based power curve use in condition monitoring. In *Proceedings of the 2018 53rd International Universities Power Engineering Conference (UPEC)*, Glasgow, UK, 4–7 September 2018; pp. 1–6. [[CrossRef](#)]
10. Pandit, R.K.; Infield, D.; Kolios, A. Comparison of advanced non-parametric models for wind turbine power curves. *IET Renew. Power Gener.* **2019**, *13*, 1503–1510. [[CrossRef](#)]
11. Prasetyowati, A.; Sudiana, D.; Sudiby, H. Comparison Accuracy W-NN and WD-SVM Method In Predicted Wind Power Model on Wind Farm Pandansimo. In *Proceedings of the 2018 4th International Conference on Nano Electronics Research and Education (ICNERE)*, Hamamatsu, Japan, 27–29 November 2018; pp. 1–4. [[CrossRef](#)]
12. Yang, X.; Cui, Y.Q.; Zhang, H.S.; Tang, N.N. Research on modeling of wind turbine based on LS-SVM. In *Proceedings of the 2009 International Conference on Sustainable Power Generation and Supply*, Nanjing, China, 6–7 April 2009. [[CrossRef](#)]
13. Mohandes, M.A.; Halawani, T.O.; Rehman, S.; Hussain, A.A. Support vector machines for wind speed prediction. *Renew. Energy* **2003**, *29*, 939–947. [[CrossRef](#)]
14. Wang, Y.; Hu, Q.H.; Li, L.H.; Foley, A.M.; Srinivasan, D. Approaches to wind power curve modeling: A review and discussion. *Renew. Sustain. Energy Rev.* **2019**, *116*, 109422. [[CrossRef](#)]
15. Moustiris, K.P.; Zafirakis, D.; Kavvadias, K.A.; Kaldellis, J.K. Wind power forecasting using historical data and artificial neural networks modeling. In *Proceedings of the Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion (MedPower 2016)*, Belgrade, Serbia, 6–9 November 2016; pp. 105–106.
16. Cai, X.; Nan, X.; Gao, B. Oxygen supply Prediction model based on IWO-SVR in bio-oxidation pretreatment. *Eng. Lett.* **2015**, *23*, 173–179.
17. Burges, C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **1998**, *2*, 121–167. [[CrossRef](#)]
18. Schölkopf, B.; Smola, A.J.; Williamson, R.C.; Bartlett, P.L. New support vector algorithms. *Neural Comput.* **2000**, *12*, 1207–1245. [[CrossRef](#)]
19. Deng, N.Y.; Tian, Y.J.; Zhang, C.H. *Support Vector Machines: Optimization Based Theory, Algorithms, and Extensions*; CRC Press: Boca Raton, FL, USA, 2012.
20. Tian, Y.; Shi, Y.; Liu, X. Recent advances on support vector machines research. *Technol. Econ. Dev. Econ.* **2012**, *18*, 5–33. [[CrossRef](#)]
21. Sfetsos, A. A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renew. Energy* **2008**, *21*, 23–35. [[CrossRef](#)]

22. Zhao, N.; Ouyang, X.; Gao, C.; Zang, Z. Training an Improved TSVR Based on Wavelet Transform Weight Via Unconstrained Convex Minimization. *IAENG Int. J. Comput. Sci.* **2019**, *46*, 264–274.
23. Doulamis, N.D.; Doulamis, A.D.; Varvarigos, E. Virtual associations of prosumers for smart energy networks under a renewable split market. *IEEE Trans. Smart Grid* **2017**, *9*, 6069–6083. [[CrossRef](#)]
24. Vergados, D.J.; Mamounakis, I.; Makris, P.; Varvarigos, E. Prosumer clustering into virtual microgrids for cost reduction in renewable energy trading markets. *Sustain. Energy Grids Netw.* **2016**, *7*, 90–103. [[CrossRef](#)]
25. Hu, Q.H.; Zhang, S.G.; Xie, Z.X.; Mi, J.S.; Wan, J. Noise model based ν -Support vector regression with its application to short-term wind speed forecasting. *Neural Netw.* **2014**, *57*, 1–11. [[CrossRef](#)]
26. Zhang, S.G.; Hu, Q.H.; Xie, Z.X.; Mi, J.S. Kernel ridge regression for general noise model with its application. *Neurocomputing* **2015**, *149*, 836–846. [[CrossRef](#)]
27. Jiang, P.; Li, P.Z. Research and Application of a New Hybrid Wind Speed Forecasting Model on BSO algorithm. *J. Energy Eng.* **2017**, *143*, 04016019. [[CrossRef](#)]
28. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
29. Du, P.; Wang, J.Z.; Guo, Z.H.; Yang, W.D. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. *Energy Convers. Manag.* **2017**, *150*, 90–107. [[CrossRef](#)]
30. Jiang, Y.; Huang, G.Q. A hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network for short-term wind speed forecasting. *Energies* **2016**, *9*, 757.
31. Jiang, Y.; Huang, G.Q. Short-term wind speed prediction: Hybrid of ensemble empirical mode decomposition, feature selection and error correction. *Energy Convers. Manag.* **2017**, *144*, 340–350. [[CrossRef](#)]
32. Wang, H.B.; Wang, Y.; Hu, Q.H. Self-adaptive robust nonlinear regression for unknown noise via mixture of Gaussians. *Neurocomputing* **2017**, *235*, 274–286. [[CrossRef](#)]
33. Khemchandani, R.; Chandra, S. Twin support vector machines for pattern classification. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 905–910.
34. Peng, X.J. TSVR: An efficient twin support vector machine for regression. *Neural Netw.* **2010**, *23*, 365–372. [[CrossRef](#)]
35. Peng, X.J. Efficient twin parametric insensitive support vector regression model. *Neurocomputing* **2012**, *79*, 26–38. [[CrossRef](#)]
36. Peng, X.J.; Xu, D.; Shen, J.D. A twin projection support vector machine for data regression. *Neurocomputing* **2014**, *138*, 131–141. [[CrossRef](#)]
37. Zhao, Y.P.; Zhao, J.; Zhao, M. Twin least squares support vector regression. *Neurocomputing* **2013**, *118*, 225–236. [[CrossRef](#)]
38. Khemchandani, R.; Goyal, K.; Chandra, S. TWSVR: Regression via twin support vector machine. *Neural Netw.* **2016**, *74*, 14–21. [[CrossRef](#)]
39. Shao, Y.H.; Zhang, C.H.; Yang, Z.M.; Jing, L.; Deng, N.Y. An ϵ -twin support vector machine for regression. *Neural Comput. Appl.* **2013**, *23*, 175–185. [[CrossRef](#)]
40. Rastogi, R.; Anand, P.; Chandra, S. A ν -twin support vector machine based regression with automatic accuracy control. *Appl. Intell.* **2017**, *46*, 670–683. [[CrossRef](#)]
41. Xu, Y.; Li, X.; Pan, X.; Yang, Z. Asymmetric ν -twin support vector regression. *Neural Comput. Appl.* **2017**, *2*, 1–16. [[CrossRef](#)]
42. Xu, Y.; Wang, L. A weighted twin support vector regression. *Knowl. Based Syst.* **2012**, *33*, 92–101. [[CrossRef](#)]
43. Matei, O.; Pop, P.C.; Vălean, H. Optical character recognition in real environments using neural networks and k-nearest neighbor. *Appl. Intell.* **2013**, *39*, 739–748. [[CrossRef](#)]
44. Xu, Y.; Wang, L. K-nearest neighbor-based weighted twin support vector regression. *Appl. Intell.* **2014**, *41*, 299–309. [[CrossRef](#)]
45. Parastalooi, N.; Amiri, A.; Aliherdari, P. Modified twin support vector regression. *Neurocomputing* **2016**, *211*, 84–97. [[CrossRef](#)]
46. Ye, Y.F.; Bai, L.; Hua, X.Y.; Shao, Y.H.; Wang, Z.; Deng, N.Y. Weighted Lagrange ϵ -twin support vector regression. *Neurocomputing* **2016**, *197*, 53–68. [[CrossRef](#)]
47. Chapelle, O. Training a support vector machine in primal. *Neural Comput.* **2007**, *19*, 1155–1178. [[CrossRef](#)]
48. Peng, X.J. Primal twin support vector regression and its sparse approximation. *Neuro Comput.* **2010**, *73*, 2846–2858. [[CrossRef](#)]

49. Balasundaram, S.; Gupta, D. Training Lagrangian twin support vector regression via unconstrained convex minimization. *Knowl. Based Syst.* **2014**, *59*, 85–96. [[CrossRef](#)]
50. Gupta, D. Training primal K-nearest neighbor based weighted twin support vector regression via unconstrained convex minimization. *Appl. Intell.* **2017**, *47*, 962–991. [[CrossRef](#)]
51. Balasundaram, S.; Meena, Y. Training primal twin support vector regression via unconstrained convex minimization. *Appl. Intell.* **2016**, *44*, 931–955. [[CrossRef](#)]
52. Zhang, S.; Zhou, T.; Sun, L.; Wang, W.; Chang, B. LSSVR Model of G-L Mixed Noise-Characteristic with Its Applications. *Entropy* **2020**, *22*, 629. [[CrossRef](#)]
53. Zhang, S.; Zhou, T.; Sun, L.; Wang, W.; Wang, C.; Mao, W. ν -Support Vector Regression Model Based on Gauss-Laplace Mixture Noise Characteristic for Wind Speed Prediction. *Entropy* **2019**, *21*, 1056. [[CrossRef](#)]
54. Rastogi, R.; Anand, P.; Chandra, S. L1-norm Twin Support Vector Machine-based Regression. *Optimization* **2017**, *66*, 1895–1911. [[CrossRef](#)]
55. Zhang, S.; Liu, C.; Zhou, T.; Sun, L. Twin Least Squares Support Vector Regression of Heteroscedastic Gaussian Noise Model. *IEEE Access* **2020**, *8*, 94076–94088. [[CrossRef](#)]
56. Rockafellar, R.T. Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming. *SIAM J. Control* **1974**, *12*, 268–285. [[CrossRef](#)]
57. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004; pp. 521–620.
58. Wang, S.; Zhang, N.; Wu, L.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [[CrossRef](#)]
59. Shevade, S.; Keerthi, S.S.; Bhattacharyya, C.; Murthy, K. Improvements to the SMO algorithm for SVM regression. *IEEE Trans. Neural Netw.* **2000**, *11*, 1188–1193. [[CrossRef](#)]
60. Bordes, A.; Bottou, L.; Gallinari, P. SGD-QN: Careful quasiNewton stochastic gradient descent. *J. Mach. Learn. Res.* **2009**, *10*, 1737–1754.
61. Chalimourda, A.; Schölkopf, B.; Smola, A.J. Experimentally optimal ν in support vector regression for different noise models and parameter settings. *Neural Netw.* **2004**, *17*, 127–141. [[CrossRef](#)]
62. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* **2004**, *17*, 113–126. [[CrossRef](#)]
63. Kwok, J.T.; Tsang, I.W. Linear dependency between and the input noise in ϵ -support vector regression. *IEEE Trans. Neural Netw.* **2003**, *14*, 544–553. [[CrossRef](#)] [[PubMed](#)]
64. Suykens, J.; Lukas, L.; Vandewalle, J. Sparse approximation using least square vector machines. In Proceedings of the IEEE International Symposium on Circuits and Systems, Geneva, Switzerland, 28–31 May 2000; pp. 757–760.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).