ν-Support Vector Regression Model Based on Gauss-Laplace Mixture Noise Characteristic for Wind Speed Prediction

Shiguang Zhang [1,2,3,*] ⓘ**, Ting Zhou** [4]**, Lin Sun** [1,3,*] ⓘ**, Wei Wang** [1]**, Chuan Wang** [1] **and Wentao Mao** [1]

[1] College of Computer and Information Engineering, Henan Normal University, Xinxiang 453007, China; wangwei@htu.edu.cn (W.W.); wangchuan@htu.edu.cn (C.W.); maowentao@htu.edu.cn (W.M.)
[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
[4] The State-Owned Assets Management Office, Henan Normal University, Xinxiang 453007, China; zhouting@htu.edu.cn
* Correspondence: zhangshiguang@htu.edu.cn (S.Z.); linsunok@gmail.com (L.S.)

check for updates

Abstract: Most regression techniques assume that the noise characteristics are subject to single noise distribution whereas the wind speed prediction is difficult to model by the single noise distribution because the noise of wind speed is complicated due to its intermittency and random fluctuations. Therefore, we will present the ν-support vector regression model of Gauss-Laplace mixture heteroscedastic noise (GLM-SVR) and Gauss-Laplace mixture homoscedastic noise (GLMH-SVR) for complex noise. The augmented Lagrange multiplier method is introduced to solve models GLM-SVR and GLMH-SVR. The proposed model is applied to short-term wind speed forecasting using historical data to predict future wind speed at a certain time. The experimental results show that the proposed technique outperforms the single noise technique and obtains good performance.

Keywords: ν-support vector regression; Gauss-Laplace mixture noise; empirical risk loss; inequality constraints; wind speed prediction

1. Introduction

Wind speed and wind power prediction is becoming increasingly important, and wind speed prediction is crucial for the control, scheduling, maintenance, and resource planning of wind energy conversion systems [1,2]. However, the volatility and uncertainty of wind speed give a fundamental challenge to power system operations. Because the basic characteristics of the wind is its intermittency and random fluctuations [3,4], the integration of wind power into power systems puts forward a series of challenges. The most effective way to resolve the challenges is to improve the prediction accuracy of wind speed and power forecasting [5–7].

In general, there are three important types in building a regression algorithm: model structures, objective functions, and optimization strategies. The model structures include linear or nonlinear functions, neural networks [8–10], etc. As for objective functions, empirical risk loss has a great effect on the performance of regression models. The selection of empirical risk loss is mostly dependent on the types of noises [11,12]. For example, squared loss is suitable for Gaussian noise [13–15], least absolute deviation loss for Laplacian noise [16], and Beta loss for Beta noise [17–19]. By the formula of the optimization method, a series of optimization algorithms are developed [20]. This work mainly studies what should be considered in the optimal architecture of the support vector regression (SVR) model in complex or unknown noise.

Recently, SVR has become an increasingly important technology. In 2000, v-SVR was introduced by Schölkopf, et al. [21] and automatically computes ϵ. Suykens et al. [22,23] constructed least squares SVR with Gaussian noise (LS-SVR). Wu [13] and Pontil et al. [24] constructed v-SVR with Gaussian noise (GN-SVR). In 2002, Bofinger et al. [25] discovered that the output of a wind turbine system is limited between zero and maximum power and that the error statistics do not follow a normal distribution. In 2007, Zhang et al. [26] and Randazzo et al. [27] proposed the estimation of coherent electromagnetic wave impact in the direction of arrival under Laplace noise environment. Bludszuweit et al. [28] explained the advantages of using Beta probability density function (PDF) instead of Gauss PDF to approximate the error distribution of wind power forecasting. According to Bayesian principle, square loss, Beta loss, or Laplacian loss are optimal when the noise is Gaussian, Beta, or Laplacian, respectively [17,18]. However, in some real-world applications, the noise distribution is complex and unknown if the data are collected in muti-source environments. Therefore, a single distribution attended to describes clearly that the real noise is not optimal and almost impossible [29,30]. Generally speaking, mixture distributions have good approximation capability for any continuous distributions. It can adapt well to unknown or complex noises when we have no prior knowledge of real noise. In 2017, the hybrid forecasting model based on multi-objective optimization [29,31], a hybrid method based on singular spectrum analysis, firefly algorithm, and BP neural network forecast the wind speed of complex noise [32]; this shows that the hybrid method has strong prediction ability. The hybrid of least squares support vector machine [33] is applied to predict the wind speed of unknown noise, which improves the forecasting performance of wind speed. Two novel nonlinear regression models [34] where the noise is fitted by mixture of Gaussian were developed, produced good performance compared with current regression algorithms, and provided superior robustness.

To address the above problem, we try to study the v-SVR model of Gauss-Laplace mixture noise characteristics for complex or unknown noise distribution. In this case, we must design a method to find the optimal solution of the corresponding regression task. Although there has been a large number of SVR algorithm implementations in the past few years, we introduced the augmented Lagrange multiplier method (ALM) method described in Section 4. Sub-gradient descent method can be used if the task is non-differentiable or discontinuous [17], or sequence minimum optimization algorithm (SMO) can be used if the sample size is large [35].

This work offers the following four contributions: (1) the optimal empirical risk loss for general mixture noise characteristic and Gauss-Laplace mixture noise by the use of Bayesian principle is obtained; (2) the v-SVR model of mixture noise, Gauss-Laplace mixture homoscedastic noise (GLM-SVR), and Gauss-Laplace mixture heteroscedastic noise (GLMH-SVR) for complex or unknown noise is constructed; (3) the augmented Lagrange multiplier method is applied to solve GLM-SVR, which guarantees the stability and validity of the solution; and (4) GLM-SVR is applied to short-term wind speed forecasting using historical data to predict future wind speed at a certain time and to verify the validity of the proposed technique.

A summary of the rest of this article is organized as follows. Section 2 derives the optimal empirical risk loss using Bayesian principle; Section 3 constructs the v-SVR model of Gauss-Laplace mixture noise characteristics; Section 4 gives the solution and algorithm design of GLM-SVR; numerical experiments are conducted out on short-term wind speed prediction in Section 5; and Section 6 summarizes this article.

2. Bayesian Principle to Empirical Risk Loss of Mixture Noise

In this section, using the theory of Bayesian principle, we obtain the optimal empirical risk loss of mixture noise characteristics.

Given the following dataset

$$D_L = \{(X_1, Y_1), (X_2, Y_2), \cdots, (X_L, Y_L)\}, \tag{1}$$

where $X_i = (x_{i1}, x_{i2}, \cdots, x_{in})^T \in R^n$ and $Y_i \in R(i = 1, 2, \cdots, L)$ are the datasets, R represents real number set, R^n is the n dimensional Euclidean space, L is the number of sample points, and superscript T denotes the matrix transpose, suppose the sample of dataset D_L is generated by the additive noise function ε; the following relationship between the measured values Y_i and predicted values $f(X_i)$ is as follows:

$$Y_i = f(X_i) + \varepsilon_i (i = 1, 2, \cdots, L) \tag{2}$$

where ε_i be random and i.i.d. means independent and identical distribution with $P(\varepsilon_i)$ of mean μ and standard deviation σ. In engineering technology, the noise density $P(\varepsilon) = P(Y - f(X))$ is unknown. We want to predict the unknown decision function $f(X)$ from the training samples $D_f \subseteq D_L$.

Following References [24,36] by the use of Bayesian principle, in maximum likelihood sense, the optimal empirical risk loss is as follows:

$$l(\varepsilon) = l(X, Y, f(X)) = -logP(Y - f(X)). \tag{3}$$

i.e., the optimal empirical risk loss $l(\varepsilon)$ is the log-likelihood of the noise model.

The probability density function (PDF) of each single distribution model and the parameters estimation formula under Bayesian principle are summarized in Reference [16]. In particular, the noise ε in Equation (2) is Laplacian, with PDF $P(\varepsilon) = \frac{1}{2}e^{-|\varepsilon|}$. By Equation (3), the optimal empirical risk loss in the sense of maximum likelihood sense should be $l(\varepsilon) = |\varepsilon|$. If the noise in Equation (2) is Gaussian, with zero mean and homoscedastic standard deviation σ, by Equation (3), empirical risk loss about Gaussian noise is $l(\varepsilon) = \frac{1}{2\sigma^2} \cdot \varepsilon^2$. Suppose the noise ε in Equation (2) is Gaussian, with zero mean and heteroscedastic standard deviation $\sigma_i (i = 1, 2, \cdots, L)$. By Equation (3), the loss about Gaussian noise is $l(\varepsilon_i) = \frac{1}{2\sigma_i^2} \cdot \varepsilon_i^2$ $(i = 1, \cdots, L)$.

It is assumed that the noise ε in Equation (2) is the mixture distributions of two kinds of noise characteristics with the probability density functions $P_1(\varepsilon)$ and $P_2(\varepsilon)$, respectively. Suppose that $P(\varepsilon) = [P_1(\varepsilon)^{\lambda_1}] \cdot [P_2(\varepsilon)^{\lambda_2}]$, by Equation (3), the optimal empirical risk loss about the mixture noise distributions is as follows:

$$l(\varepsilon) = \lambda_1 \cdot l_1(\varepsilon) + \lambda_2 \cdot l_2(\varepsilon). \tag{4}$$

where $l_1(\varepsilon) > 0, l_2(\varepsilon) > 0$ are convex empirical risk losses of the above two kinds of noise characteristics, respectively. Weight factor $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

The Gauss-Laplace empirical risk loss for different parameter are shown in Figure 1.
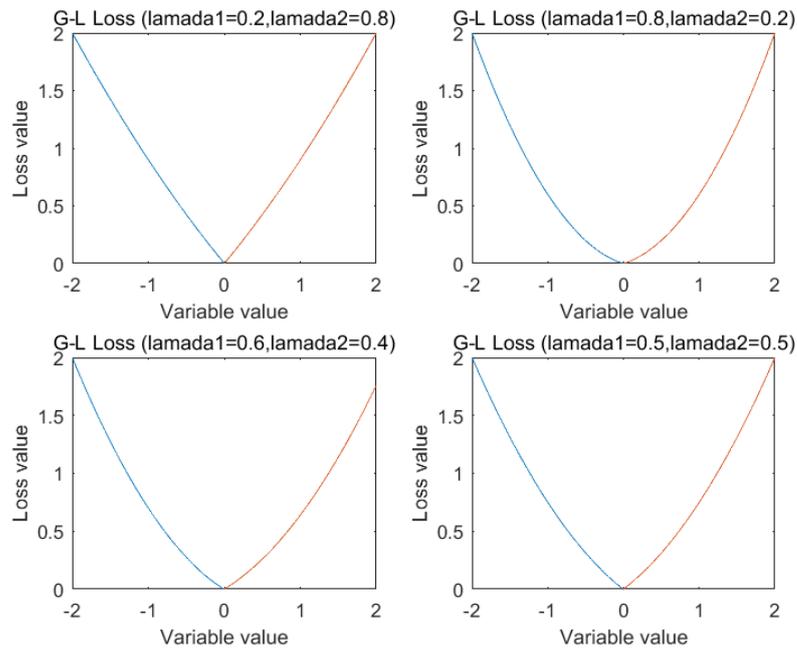
Figure 1. Gauss-Laplace empirical risk loss for different parameter.

3. Model ν-SVR of Gauss-Laplace Mixture Noise

Given dataset D_L, we build a linear regressor $f(X) = \omega^T \cdot X + b$, where ω denotes the weight vector and b is the bias term. To deal with nonlinear problems, the following summaries can be made [37,38]: the input vector $X_i \in R^n$ is mapped by a nonlinear mapping (chosen a priori) Φ: $R^n \to H$ is the high dimensional feature space H (H is Hilbert space), induced by the kernel matrix $K(X_i, X_j) = (\Phi(X_i) \cdot \Phi(X_j))$ $(i, j = 1, 2, \cdots, L)$. $(\Phi(X_i) \cdot \Phi(X_j))$ is the inner product in H, and the kernel mapping Φ may be any positive definite Mercer kernel. Therefore, we will solve the optimization problem in feature space H. The linear ν-SVR is extended to the nonlinear ν-SVR by using the kernel matrix $K(X_i, X_j)$.

We propose the uniform model ν-SVR of mixture noises (M-SVR). The primal problem of model M-SVR is described as follows:

$$
\begin{aligned}
Min\{g_{P_{M-SVR}} &= \frac{1}{2}\omega^T \cdot \omega + \frac{C}{L} \cdot [\nu \cdot \varepsilon + \lambda_1 \cdot \sum_{i=1}^{L}(l_1(\xi_i) + l_1(\xi_i^*)) \\
&+ \lambda_2 \cdot \sum_{i=1}^{L}(l_2(\xi_i) + l_2(\xi_i^*))]\} \\
Subject \quad to : &\omega^T \cdot \Phi(X_i) + b - Y_i \leq \varepsilon + \xi_i \\
&Y_i - \omega^T \cdot \Phi(X_i) - b \leq \varepsilon + \xi_i^*
\end{aligned}
\tag{5}
$$

where ξ_i and ξ_i^* are random noises and slack variable at time i. $l_1(\xi_i), l_1(\xi_i^*), l_2(\xi_i),$ and $l_2(\xi_i^*) > 0$ $(i = 1, 2, \cdots, L)$ are convex empirical risk loss function values for general noise characteristic in the sample point $(X_i, Y_i) \in D_L$ $(i, j = 1, 2, \cdots, L)$. $C > 0$ is the penalty parameter, $\varepsilon \geq 0$, and $\nu \in (0, 1]$. Weight factor $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

As a function approximation machine, the objection is to estimate an unknown function $f(X)$ from the training samples $D_f \subseteq D_L$. In the field of practical application, most of the distributions do not satisfy the Gauss distribution, and it also does not obey the Laplace distribution. The noise distribution is unknown or complex; a single distribution intended to describe the real noise is almost impossible. Generally, mixture distributions (as Gauss-Laplace mixed distribution) have good approximation capabilities for any continuous distributions, and it can fit the unknown or complex noise. Therefore,

we will use the Gauss-Laplace mixed homoscedastic and heteroscedastic noise distribution to fit the unknown or complex noise characteristics in the next section.

3.1. Model ν-SVR of Gauss-Laplace Mixture Homoscedastic Noise

If the noise in Equation (2) is Gaussian, with zero mean and the homoscedastic standard deviation σ, by Equation (3), the empirical risk loss of homoscedastic Gaussian noise is $l_1(\xi_i, \xi_i^*) = \frac{1}{2\sigma^2}(\xi_i^2 + (\xi_i^*)^2)$ and the empirical risk loss of Laplace noise is $l_2(\xi_i, \xi_i^*) = |\xi_i| + |\xi_i^*|$. We adopt the Gauss-Laplace mixture homoscedastic noise distribution to fit the unknown noise characteristics. By Equation (4), the loss function corresponding to Gauss-Laplace mixture homoscedastic noise characteristics is $l(\xi_i, \xi_i^*) = \frac{\lambda_1}{2} \cdot (\xi_i^2 + (\xi_i^*)^2) + \lambda_2 \cdot (|\xi_i| + |\xi_i^*|)$. We put forward a technique of ν-SVR model for Gauss-Laplace mixture homoscedastic noise characteristics (GLM-SVR). The primal problem of model GLM-SVR be described as follows:

$$
\begin{aligned}
Min\{g_{P_{GLM-SVR}} &= \frac{1}{2}\omega^T \cdot \omega + \frac{C}{L} \cdot [\nu \cdot \varepsilon + \frac{\lambda_1}{2} \cdot \sum_{i=1}^{L}(\xi_i^2 + (\xi_i^*)^2) \\
&+ \lambda_2 \cdot \sum_{i=1}^{L}(\xi_i + \xi_i^*)]\} \\
Subject \quad to : \omega^T &\cdot \Phi(X_i) + b - Y_i \leq \varepsilon + \xi_i \\
Y_i - \omega^T &\cdot \Phi(X_i) - b \leq \varepsilon + \xi_i^*
\end{aligned}
\tag{6}
$$

where $\xi_i, \xi_i^* \geq 0 (i = 1, 2, \cdots, L)$ are random noises and slack variables at time i. $C > 0$ is the penalty parameter, $\varepsilon \geq 0$, and $\nu \in (0, 1]$. Weight factor $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proposition 1. *The solution of the primal problem Equation (6) of model GLM-SVR about ω exists and is unique.*

Theorem 1. *The dual problem of the primal problem of Equation (6) of model GLM-SVR is as follows:*

$$
\begin{aligned}
Max\{g_{D_{GLM-SVR}} &= -\frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}(\alpha_i^* - \alpha_i) \cdot (\alpha_j^* - \alpha_j) \cdot K(X_i, X_j) \\
&+ \sum_{i=1}^{L}(\alpha_i^* - \alpha_i) \cdot Y_i - \frac{L}{2C \cdot \lambda_1}\sum_{i=1}^{L}[(\alpha_i - C \cdot \lambda_2)^2 \\
&+ (\alpha_i^* - C \cdot \lambda_2)^2]\} \\
Subject \quad to \quad &\sum_{i=1}^{L}(\alpha_i^* - \alpha_i) = 0 \\
&0 \leq \alpha_i, \alpha_i^* \leq \frac{C}{L}, i = 1, 2, \cdots, L \\
&\sum_{i=1}^{L}(\alpha_i^* + \alpha_i) \leq C \cdot \nu.
\end{aligned}
\tag{7}
$$

where $\alpha_i, \alpha_i^*(i = 1, 2, \cdots, L)$ are Lagrange multipliers and $K(X_i, X_j)$ $(i, j = 1, 2, \cdots, L)$ is a kernel matrix. $C > 0$ is the penalty parameter, $\nu \in (0, 1]$. Weight factor $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proof. Let us take Lagrange functional $L_\omega(\omega, b, \alpha, \alpha^*, \xi, \xi^*, \eta, \eta^*, \varepsilon, \gamma)$ as

$$
\begin{aligned}
L_\omega &= \frac{1}{2}\omega^T \cdot \omega + \frac{C}{L} \cdot [\nu \cdot \varepsilon + \frac{\lambda_1}{2} \cdot \sum_{i=1}^{L}(\xi_i^2 + (\xi_i^*)^2) + \lambda_2 \cdot \sum_{i=1}^{L}(\xi_i + \xi_i^*)] - \gamma\varepsilon \\
&- \sum_{i=1}^{L}(\eta_i\xi_i + \eta_i^*\xi_i^*) - \sum_{i=1}^{L}\alpha_i(\xi_i + Y_i - \omega^T \cdot \Phi(X_i) - b + \varepsilon) \\
&- \sum_{i=1}^{L}\alpha_i^*(\xi_i^* - Y_i + \omega^T \cdot \Phi(X_i) + b + \varepsilon).
\end{aligned}
$$

To minimize L_ω, let us find partial derivatives ω, b, ξ, ξ^*, and ε. On the basis of KKT (Karush–Kuhn–Tucker) conditions, we get

$$
\nabla_\omega(L_\omega) = 0, \nabla_b(L_\omega) = 0, \nabla_\varepsilon(L_\omega) = 0, \nabla_\xi(L_\omega) = 0, \nabla_{\xi^*}(L_\omega) = 0.
$$

and have

$$
\omega_i = \sum_{i=1}^{L}(\alpha_i^* - \alpha_i) \cdot \Phi(X_i),
$$

$$\sum_{i=1}^{L}(\alpha_i^* - \alpha_i) = 0,$$

$$\frac{C}{L} \cdot \nu - \gamma - \sum_{i=1}^{L}(\alpha_i^* + \alpha_i) = 0,$$

$$\frac{C}{L} \cdot (\lambda_1 \cdot \xi_i + \lambda_2) - \eta_i - \alpha_i = 0,$$

$$\frac{C}{L} \cdot (\lambda_1 \cdot \xi_i^* + \lambda_2) - \eta_i^* - \alpha_i^* = 0,$$

By substituting the above extreme conditions into L_ω and by seeking maximum α, α^*, we obtain the dual problem of Equation (7) of the primal problem of Equation (6) of model GLM-SVR. □

Then, we obtain

$$\omega_i = \sum_{i=1}^{L}(\alpha_i^* - \alpha_i) \cdot \Phi(X_i),$$

$$b = \frac{1}{L} \sum_{i=1}^{L}[Y_i - \sum_{i \in RSV}(\alpha_i^* - \alpha_i) \cdot K(X_i, X_j) - \frac{(L \cdot \alpha_i - C \cdot \lambda_2)}{C \cdot \lambda_1}].$$

To estimate ε, we have

$$\varepsilon = \frac{1}{L} \sum_{j=1}^{L}(\sum_{i \in RSV}(\alpha_i^* - \alpha_i) \cdot K(X_i, X_j) + b - Y_j).$$

Thus, the decision function of model GLM-SVR can be written as

$$f(X) = \omega^T \cdot \Phi(X) + b = \sum_{i \in RSV}(\alpha_i^* - \alpha_i)K(X_i, X) + b,$$

where RSVs are samples about $\alpha_i^* - \alpha_i \neq 0$ (called support vectors), $K(X_i, X_j) = (\Phi(X_i) \cdot \Phi(X_j))$ $(i, j = 1, 2, \cdots, L)$ is the Kernel function, $\Phi : R^n \to H$ (R^n is the n dimensional Euclidean space, and H is the Hilbert space), and $(\Phi(X_i) \cdot \Phi(X_j))$ is the inner product of H.

3.2. Model ν-SVR of Gauss-Laplace Mixture Heteroscedastic Noise

If the noise in Equation (2) is Gaussian, with zero mean and the heteroscedastic variance $\sigma_i^2, (\sigma_i^*)^2$, that is $\sigma_i \neq \sigma_j, \sigma_i^* \neq \sigma_j^*$, where $i \neq j(i, j = 1, \cdots, L)$, by Equation (3), the empirical risk loss of heteroscedastic Gaussian noise characteristic is $l_1(\xi_i, \xi_i^*) = \frac{1}{2\sigma_i^2}\xi_i^2 + \frac{1}{2(\sigma_i^*)^2}(\xi_i^*)^2$, $(i = 1, \cdots, L)$ and the empirical risk loss of Laplace noise characteristic is $l_2(\xi_i, \xi_i^*) = |\xi_i| + |\xi_i^*|$. We utilize the Gauss-Laplace mixture heteroscedastic noise distribution to predict the unknown noise characteristics. By Equation (4), the empirical risk loss about Gauss-Laplace mixture heteroscedastic noise is $l(\xi_i, \xi_i^*) = \frac{\lambda_1}{2} \cdot (\frac{1}{\sigma_i^2}\xi_i^2 + + \frac{1}{(\sigma_i^*)^2}(\xi_i^*)^2) + \lambda_2 \cdot (|\xi_i| + |\xi_i^*|)$ $(i = 1, \cdots, L)$. We propose a novel technique of ν-SVR model for Gauss-Laplace mixture heteroscedastic noise characteristics (GLMH-SVR). The primal problem of model GLMH-SVR can be formulated as follows:

$$Min\{g_{P_{GLMH-SVR}} = \frac{1}{2}\omega^T \cdot \omega + \frac{C}{L} \cdot [\nu \cdot \varepsilon + \frac{\lambda_1}{2} \cdot \sum_{i=1}^{L}(\frac{1}{\sigma_i^2} \cdot \xi_i^2 +$$

$$\frac{1}{(\sigma_i^*)^2} \cdot (\xi_i^*)^2) + \lambda_2 \cdot \sum_{i=1}^{L}(\xi_i + \xi_i^*)]\} \tag{8}$$

$$Subject \quad to : \omega^T \cdot \Phi(X_i) + b - Y_i \leq \varepsilon + \xi_i$$

$$Y_i - \omega^T \cdot \Phi(X_i) - b \leq \varepsilon + \xi_i^*.$$

where $\xi_i, \xi_i^* \geq 0, i = 1, 2, \cdots, L$ are random noises and slack variables at time i, the variance $\sigma_i^2, (\sigma_i^*)^2 (i = 1, 2, \cdots, L)$ is heteroscedastic, $C > 0$ is the penalty parameter, $\varepsilon \geq 0$, and $\nu \in (0, 1]$. Weight factor $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proposition 2. *The solution of the primal problem of Equation (8) of GLMH-SVR about ω exists and is unique.*

Theorem 2. *The dual problem of GLMH-SVR in the primal problem of Equation (8) is as follows:*

$$
\begin{aligned}
Max\{g_{D_{GLMH-SVR}} &= -\frac{1}{2}\sum_{i=1}^{L}\sum_{j=1}^{L}(\alpha_i^* - \alpha_i)\cdot(\alpha_j^* - \alpha_j)\cdot K(X_i, X_j) \\
&+ \sum_{i=1}^{L}(\alpha_i^* - \alpha_i)\cdot Y_i - \frac{L}{2C\cdot\lambda_1}[\sum_{i=1}^{L}(\sigma_i^2\cdot\alpha_i - C\cdot\lambda_2)^2 \\
&+ \sum_{i=1}^{L}((\sigma_i^*)^2\cdot\alpha_i^* - C\cdot\lambda_2)^2]\} \\
Subject \quad to \quad &\sum_{i=1}^{L}(\alpha_i^* - \alpha_i) = 0 \\
&0 \leq \alpha_i \leq \frac{C}{L\cdot\sigma_i^2}, i = 1, 2, \cdots, L \\
&0 \leq \alpha_i^* \leq \frac{C}{L\cdot(\sigma_i^*)^2}, i = 1, 2, \cdots, L \\
&\sum_{i=1}^{L}(\alpha_i^* + \alpha_i) \leq C\cdot\nu, i = 1, 2, \cdots, L
\end{aligned}
\tag{9}
$$

where $\sigma_i^2, (\sigma_i^*)^2 (i = 1, 2, \cdots, L)$ is heteroscedastic, $C > 0$ is the penalty parameter, and $\nu \in (0, 1]$. Weight factor $\lambda_1, \lambda_2 \geq 0$ and $\lambda_1 + \lambda_2 = 1$.

Proof. An Appendix A to the proof of Theorem 2. □

We get the following:

$$
\omega_i = \sum_{i=1}^{L}(\alpha_i^* - \alpha_i)\cdot\Phi(X_i),
$$

$$
b = \frac{1}{L}\sum_{i=1}^{L}[Y_i - \sum_{i\in RSV}(\alpha_i^* - \alpha_i)\cdot K(X_i, X_j) - \frac{(L\cdot\sigma_i^2\cdot\alpha_i - C\cdot\lambda_2)}{C\cdot\lambda_1}].
$$

To estimate ε, we get use the following:

$$
\varepsilon = \frac{1}{L}\sum_{j=1}^{L}(\sum_{i\in RSV}(\alpha_i^* - \alpha_i)\cdot K(X_i, X_j) + b - Y_j).
$$

Thus, the decision function of model GLMH-SVR can be written as follows:

$$
f(X) = \omega^T\cdot\Phi(X) + b = \sum_{i\in RSV}(\alpha_i^* - \alpha_i)K(X_i, X) + b,
$$

where RSVs are samples about $\alpha_i^* - \alpha_i \neq 0$ (called support vectors), parameter vector $\omega \in R^n$, $\Phi : R^n \rightarrow H$, and $K(X_i, X_j)$ is the Kernel function.

If the noise in Equation (2) is Gaussian, with zero mean and homoscedasticity, Theorem 1 can be derived by Theorem 2.

4. Solution Based on the Augmented Lagrange Multiplier Method

The augmented Lagrange multiplier method (ALM) method [39–41] is a class of algorithms for solving equality- and inequality-constrained optimization problems. It solves the dual problem of Equation (7) of model GLM-SVR by applying Newton's method to a series of constrained problems. By eliminating equality and inequality constraints, the optimization problem of Equation (7) can be reduced to an equivalent unconstrained problem. Gradient descent method or Newton method can be used to solve above problems [24,42,43]. If there are large-scale training samples, some fast

optimization techniques can also be combined with the proposed objective function, such as stochastic gradient decent [44].

In this section, we apply Newton's method to the sequence of inequality and equality constraints and use ALM method to solve model GLM-SVR. Theorem 1 and Theorem 2 provide the algorithms for effectively identifying models GLM-SVR and GLMH-SVR, respectively. The solution based on ALM and algorithm design of model GLM-SVR is given. Similarly, model GLMH-SVR can be solved by the use of ALM.

(1) Let training samples $D_L = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_L, Y_L)\}$, where $X_i \in R^n$, $Y_i \in R$, $i = 1, \ldots, L$.

(2) The 10-fold cross-validation strategy is adopted to search the optimal parameters C, ν, λ_1, and λ_2 and to select the appropriate kernel function $K(\bullet, \bullet)$.

(3) Solve the optimization problem of Equation (7), and get the optimal solution $\alpha = (\alpha_1, \cdots, \alpha_L)$, $\alpha^* = (\alpha_1^*, \cdots, \alpha_L^*)$.

(4) Construct the decision function as

$$f(X) = \omega^T \cdot \Phi(X) + b = \sum_{i \in RSV} (\alpha_i^* - \alpha_i) K(X_i, X) + b,$$

where $b = \frac{1}{L} \sum_{i=1}^{L} [Y_i - \sum_{i \in RSV} (\alpha_i^* - \alpha_i) \cdot K(X_i, X_j) - \frac{(L \cdot \alpha_i - C \cdot \lambda_2)}{C \cdot \lambda_1}]$
and estimate ε as

$$\varepsilon = \frac{1}{L} \sum_{j=1}^{L} (\sum_{i \in RSV} (\alpha_i^* - \alpha_i) \cdot K(X_i, X_j) + b - Y_j).$$

Parameter vector $\omega \in R^n$, $\Phi : R^n \to H$ (H is Hilbert space), where $(\Phi(X_i) \cdot \Phi(X_j))$ is the inner product of H and $K(X_i, X_j) = (\Phi(X_i) \cdot \Phi(X_j))$ is the Kernel function.

5. Case Study

In this section, a case study is implemented to demonstrate the effectiveness of the proposed model GLM-SVR through comparisons with other techniques for training-set D_f from Heilongjiang, China. This case study includes three subsections: Data collection and analysis in Section 5.1; evaluation criteria for forecasting performance in Section 5.2; and short-term wind speed forecasting of a real dataset in Section 5.3.

5.1. Analysis of Wind Speed Mixture Noise Characteristics

In order to analyze the mixture noise characteristics of wind speed forecasting error, we collected wind speed dataset from Heilongjiang, China. The dataset consists of one-year wind speed data, recording the wind speed values every 10 min. We first found the Gauss-Laplace mixture noise in the above data. The researchers have found that turbulence is the major cause of the wind speed's strong random fluctuation uncertainty. From wind energy perspective, the most striking characteristic of the wind resource is its variability. Now we display the distributions of wind speed. We obtain a value for wind speed after every 10 min and compute the histograms of wind speed in one or two hours. Two typical distributions are given as follows: one was computed when the wind speed was higher and the other was computed when the wind speed was lower, as shown in Figure 2 and Figure 3, respectively.
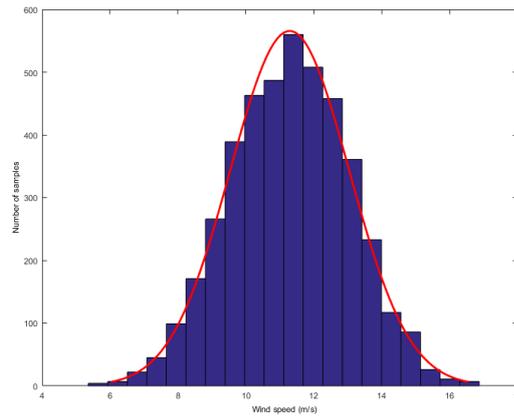
Figure 2. The distribution of high wind speed.
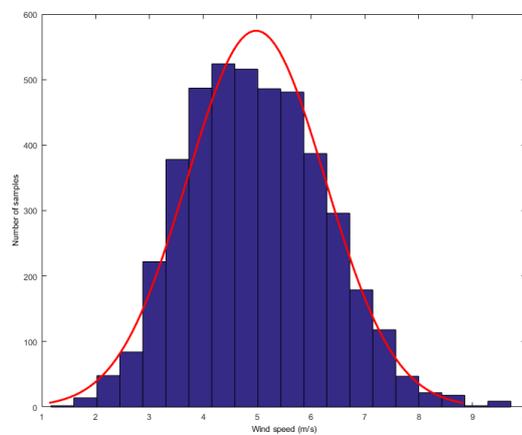
Figure 3. The distribution of low wind speed.

To analyze one-month time series of wind speed dataset, the persistence method is used to investigate the distribution of wind speed prediction errors [28]. The result indicates that the error ξ does not satisfy the single distribution but approximately obeys the Gauss-Laplace mixed distribution and that the PDF of ξ is $P(\xi) = \frac{1}{2}e^{-|\xi|} \cdot \frac{1}{2\sigma^2}\xi^2$, as shown in Figure 4. This is a regression learning task of mixture noise.
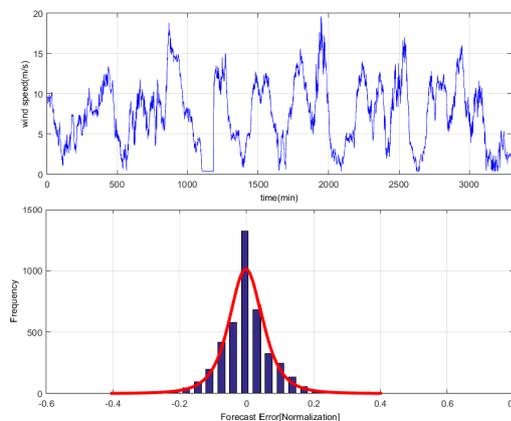
Figure 4. Gauss-Laplace mixture distribution of wind speed prediction error.

5.2. Evaluation Criteria for Forecasting Performance

As we all know, no prediction model forecasts perfectly. There are also certain criteria, such as mean absolute error (MAE), the root mean square error (RMSE), mean absolute percentage error

(MAPE), and standard error of prediction (SEP), which are used to evaluate the predictive performance of models v-SVR, GN-SVR, and GLM-SVR. The four criteria are defined as follows:

$$MAE = \frac{1}{L} \sum_{i=1}^{L} |Y_i' - Y_i|, \tag{10}$$

$$MAPE = \frac{1}{L} \sum_{i=1}^{L} \frac{|Y_i' - Y_i|}{Y_i} \times 100\%, \tag{11}$$

$$RMSE = \sqrt{\frac{1}{L} \sum_{i=1}^{L} (Y_i' - Y_i)^2}, \tag{12}$$

$$SEP = \frac{RMSE}{\overline{Y}} \times 100\%, \tag{13}$$

Among them, L is the size of the training samples, Y_i is the ith actual measured data, Y_i' is the ith forecasted result, and $\overline{Y}$ is the mean value of observations of all selected samples in the training-set D_L [45–47]. The MAE reveals how similar the predicted values are to the observed values, whereas the RMSE measures the overall deviation between the predicted and observed values. MAPE is the ratio between errors and observed values, and SEP is the ratio between RMSE and mean values of observations. The indicators MAPE and SEP are unit-free measures of accuracy for predicting wind series and are sensitive to small changes.

5.3. Short-Term Wind Speed Prediction of Real Dataset

In this subsection, we demonstrate the validity of the proposed model by conducting experiments on wind speed dataset from Heilongjiang Province, China. The data records more than one year of wind speeds. The average wind speed in 10 min are stored. As a whole, 62,466 samples with 4 attributes: mean, variance, minimum, and maximum. We first extracted 2160 consecutive data points (from 1 to 2160; the time length is 15 days) as the training set and 720 consecutive data points (from 2161 to 2880, the time length is 5 days) as the testing set. We transform the original sequence into a multivariate regression task using mode $\overrightarrow{X_i} = (X_{i-10}, X_{i-9}, \cdots, X_{i-1}, X_i)$ as an input vector to predict X_{i+step}, in which the vector orders of wind speed is determined by the chaotic operator network method [48], where X_j is the real value of wind speed at time $j(j = i - 10, i - 9, \cdots, i), step = 1, 3, 5$, that is to say, using the above mode to predict the wind speed at each point X_i after 10-min, 30-min, and 50-min, respectively.

Models v-SVR, GN-SVR, and GLM-SVR have been implemented in Matlab 7.8 programming language. The initial parameters are $C \in [1, 201]$, $v \in (0, 1]$, λ_1, and $\lambda_2 \in [0, 1]$. We use the 10-fold cross validation strategy to find optimal positive parameters C, v, λ_1, and λ_2, of which the parameters selection technology is studied in detail in References [49,50]. In this article, the parameter assignments are as follows: $C = 181, v = 0.5, \lambda_1 = 0.5$, and $\lambda_2 = 0.5$. Many practical applications display that polynomial and Gaussian kernels perform well under general smooth assumptions. In this case study, polynomial and Gaussian kernel functions are utilized in models v-SVR, the v-SVR model of Gauss homoscedastic noise (GN-SVR), and GLM-SVR as below [51,52].

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

and

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

where d is a positive integer and σ is positive.

In Figures 5–7, the results of wind speed prediction at the X_i point for models v-SVR, GN-SVR, and GLM-SVR are obtained after 10 min, 30 min, and 50 min, respectively.
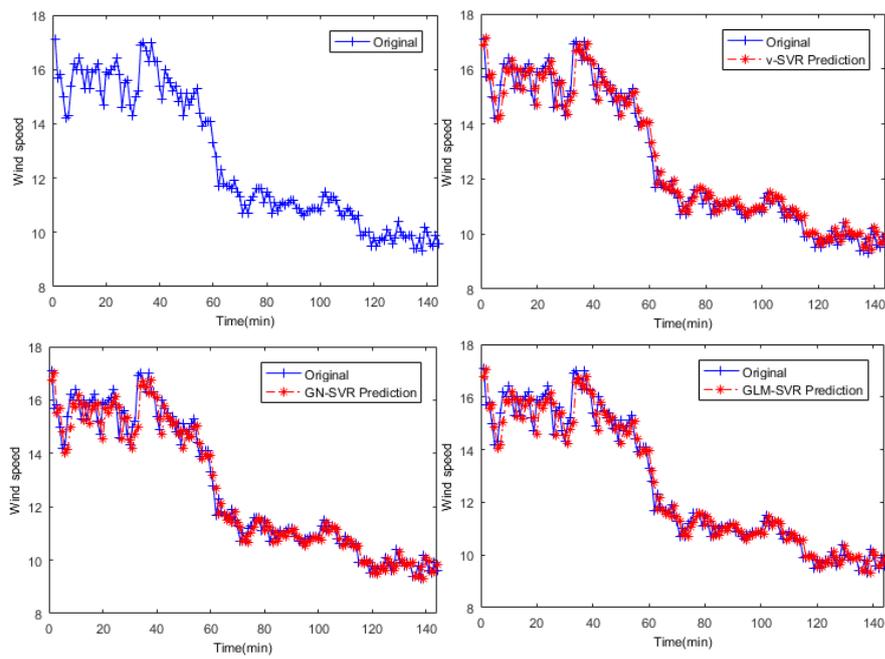
Figure 5. Result of wind speed prediction at X_i-point after 10 min.
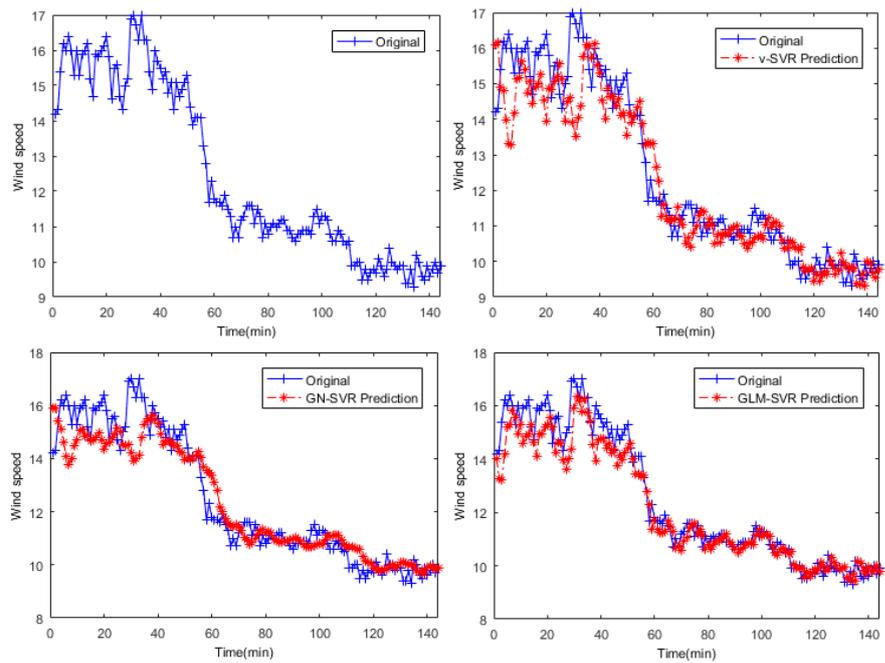
Figure 6. Result of wind speed prediction at X_i-point after 30 min.
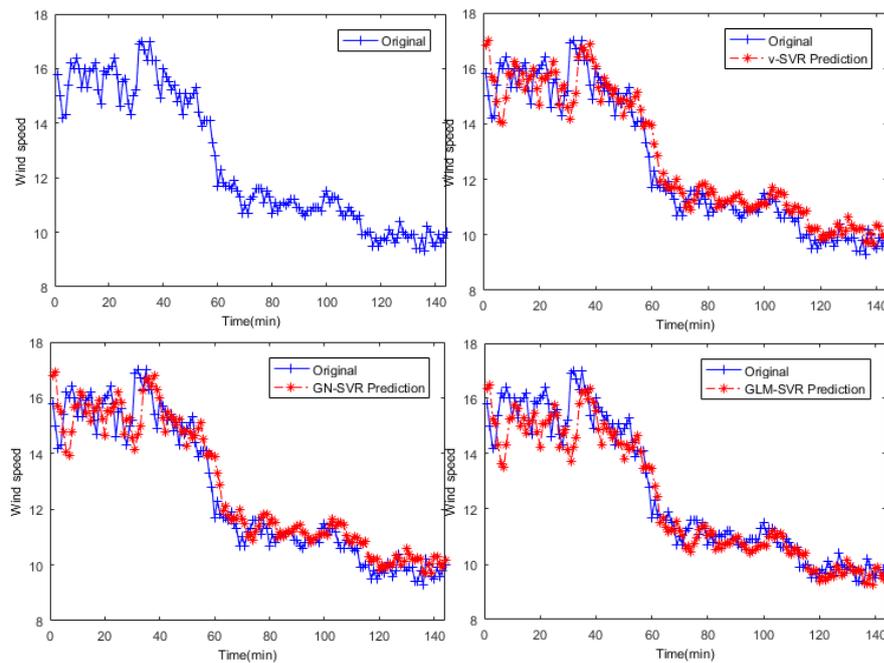
Figure 7. Result of wind speed prediction at X_i-point after 50 min.

In Tables 1–3 and Figures 8–11, indicators *MAE*, *MAPE*, *RMSE*, and SEP of wind speed prediction at X_i-point for models ν-SVR, GN-SVR, and GLM-SVR are obtained after 10 min, 30 min, and 50 min, respectively.

Table 1. Error statistic of wind speed prediction at X_i-point after 10 min.

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	SEP (%)
ν-SVR	0.3671	0.4854	2.84	3.84
GN-SVR	0.3641	0.4845	2.77	3.83
GLM-SVR	0.3616	0.4813	2.76	3.81

Table 2. Error statistic of wind speed prediction at X_i-point after 30 min.

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	SEP (%)
ν-SVR	0.5655	0.7631	4.44	6.08
GN-SVR	0.5577	0.7589	4.36	6.04
GLM-SVR	0.5468	0.7773	4.08	6.19

Table 3. Error statistic of wind speed prediction at X_i-point after 50 min.

Model	MAE (m/s)	RMSE (m/s)	MAPE (%)	SEP (%)
ν-SVR	0.6281	0.9142	4.70	7.32
GN-SVR	0.6146	0.8711	4.61	6.98
GLM-SVR	0.5920	0.8734	4.52	7.07
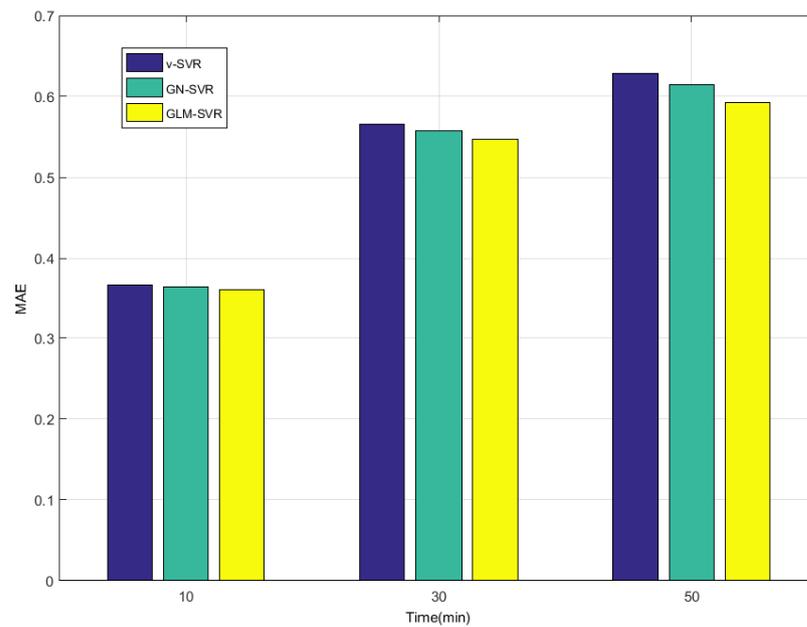
Figure 8. Error statistical histograms of index MAE for wind speed prediction at X_i-point.
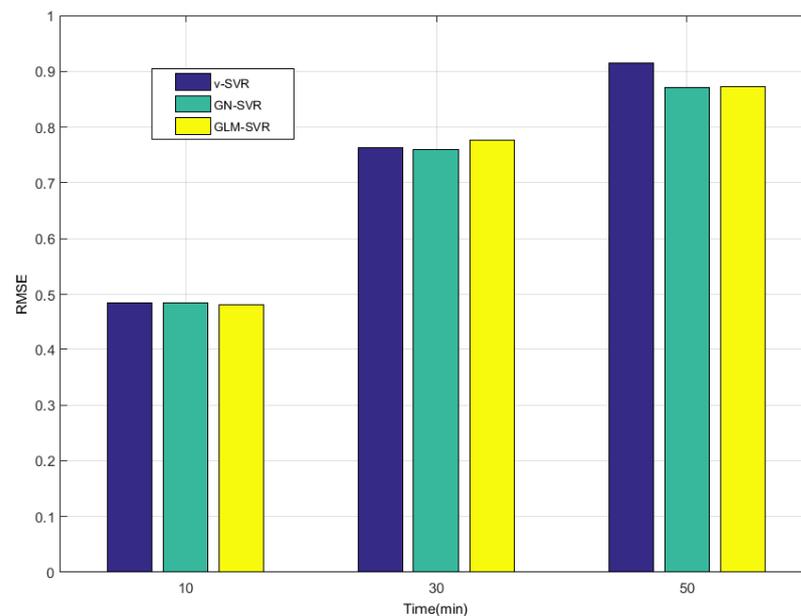
Figure 9. Error statistical histograms of index RMSE for wind speed prediction at X_i-point.

From Tables 1–3 and Figures 5–11, in most cases, it can be concluded that the error calculation of model GLM-SVR is better than that of models ν-SVR and GN-SVR. As the prediction horizon increases to 30 min and 50 min, the errors obtained by different models rise and the relative difference decreases. However, as can be seen from Tables 1–3, the Gauss-Laplace mixture noise model is slightly superior to the classical model in terms of all indicators: MAE, MAPE, RMSE, and SEP.
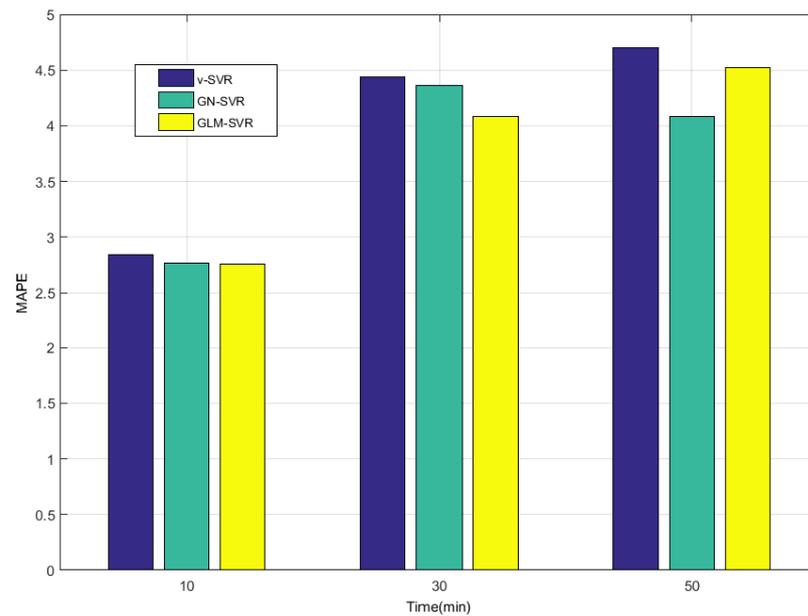
Figure 10. Error statistical histograms of index MAPE for wind speed prediction at X_i-point.
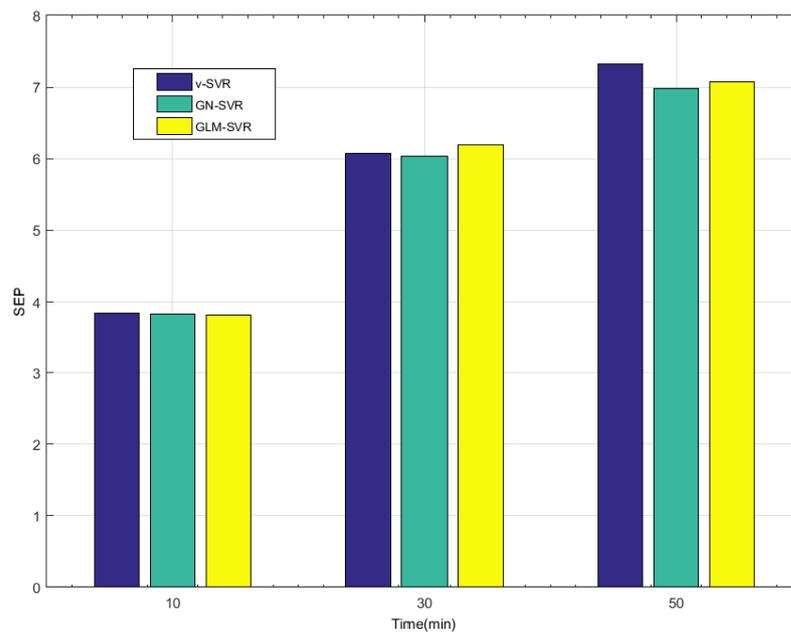
Figure 11. Error statistical histograms of index SEP for wind speed prediction at X_i-point.

6. Conclusions

The noise distribution is complex or unknown in the real world; it is almost impossible for a single distribution to describe real noise. This article describes the main results: (1) optimal empirical risk loss for mixture noise model is derived by the Bayesian principle; (2) model ν-SVR of the Gauss-Laplace mixture homoscedastic noise (GLM-SVR) and Gauss-Laplace mixture heteroscedastic noise (GLMH-SVR) for complex or unknown noise is developed; (3) the dual problems of GLM-SVR and GLMH-SVR are derived by introducing Lagrange functional L_ϖ; (4) the ALM method is applied to solve model GLM-SVR, which guarantees the stability and validity; and (5) model GLM-SVR is applied to short-term wind speed forecasting using historical data to predict future wind speed at a

certain time. The experimental results on real-world data of wind speed confirm the effectiveness of the proposed technique.

Analogously, we can study the Gauss-Laplace mixture noise model of classification, which will be successfully used to solve the classification problem for complex or unknown noise characteristics.

Abbreviations

The following abbreviations are used in this manuscript:

SVR	Support vector regression
GLM-SVR	SVR model of Gauss-Laplace mixture heteroscedastic noise
GLMH-SVR	SVR model of Gauss-Laplace mixture homoscedastic noise
GN-SVR	SVR model of Gauss homoscedastic noise
ALM	augmented Lagrange multiplier method

Appendix A

Proof of Theorem 2. Let's take Lagrange functional $L_\varpi(\varpi, b, \alpha, \alpha^*, \xi, \xi^*, \eta, \eta^*, \varepsilon)$ as:

$$L_\varpi = \frac{1}{2}\varpi^T \cdot \varpi + \frac{C}{L} \cdot [v \cdot \varepsilon + \frac{\lambda_1}{2} \cdot \sum_{i=1}^{L} \frac{1}{\sigma_i^2} \cdot \xi_i^2 + \frac{1}{(\sigma_i^*)^2} \cdot (\xi_i^*)^2)$$
$$+ \lambda_2 \cdot \sum_{i=1}^{L}(\xi_i + \xi_i^*)] - \gamma\varepsilon - \sum_{i=1}^{L}(\eta_i\xi_i + \eta_i^*\xi_i^*)$$
$$- \sum_{i=1}^{L} \alpha_i(\xi_i + y_i - \varpi^T \cdot \Phi(x_i) - b + \varepsilon)$$
$$- \sum_{i=1}^{L} \alpha_i^*(\xi_i^* - y_i + \varpi^T \cdot \Phi(x_i) + b + \varepsilon).$$

To minimize L_ϖ, let's find partial derivative $\varpi, b, \xi, \xi^*, \varepsilon$, respectively. On the basis of KKT(Karush-Kuhn-Tucker) conditions, get

$$\nabla_\varpi(L) = 0, \nabla_b(L) = 0, \nabla_\varepsilon(L) = 0, \nabla_\xi(L) = 0, \nabla_{\xi^*}(L) = 0.$$

And have

$$\varpi_i = \sum_{i=1}^{L}(\alpha_i^* - \alpha_i) \cdot \Phi(x_i), \sum_{i=1}^{L}(\alpha_i^* - \alpha_i) = 0,$$

$$\frac{C}{L} \cdot v - \gamma - \sum_{i=1}^{L}(\alpha_i^* + \alpha_i) = 0,$$

$$\frac{C}{L} \cdot (\frac{\lambda_1 \cdot \xi_i}{\sigma_i^2} + \lambda_2) - \eta_i - \alpha_i = 0,$$

$$\frac{C}{L} \cdot (\frac{\lambda_1 \cdot \xi_i^*}{(\sigma_i^*)^2} + \lambda_2) - \eta_i^* - \alpha_i^* = 0,$$

Substituting extreme conditions into L_ϖ and seeking maximum of α, α^*, Dual Problem (9) of Primal Problem (8) be derived. □

References

1. European Wind Energy Association, Wind Force 12. Available online: http://www.ewea.org/doc/WindForce12 (accessed on 14 March 2011).
2. Sfetsos, A. A comparison of various forecasting techniques applied to mean hourly wind speed time series. *Renew. Energy* **2008**, *21*, 23–35. [CrossRef]
3. Calif, R.; Schmitt, F. Modeling of atmospheric wind speed sequence using a lognormal continuous stochastic equation. *J. Wind Eng. Inst. Aerodyn.* **2012**, *109*, 1–8. [CrossRef]
4. Calif, R.; Schmitt, F. Multiscaling and joint multiscaling of the atmospheric wind speed and the aggregate power output from a wind farm. *Nonlinear Process. Geophys.* **2014**, *21*, 379–392. [CrossRef]
5. Jung, J.; Broadwater, R.P. Current status and future advances for wind speed and power forecasting. *Renew. Sustain. Energy Rev.* **2014**, *31*, 762–777. [CrossRef]
6. Zhang, C.; Wei, H.; Zhao, J.; Liu, T.; Zhu, T.; Zhang, K. Short-term wind speed forecasting using empirical mode decomposition and feature selection. *Renew. Energy* **2016**, *96*, 727–737. [CrossRef]
7. 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]
8. Wang, J.Z.; Zhang, N.; Lu, H.Y. A novel system based on neural networks with linear combination framework for wind speed forecasting. *Energy Convers. Manag.* **2019**, *181*, 425–442. [CrossRef]
9. Sun, L.; Liu, R.N.; Xu, J.C.; Zhang, S.G. An adaptive density peaks clustering method with Fisher linear discriminant. *IEEE Access.* **2019**, *7*, 72936–72955. [CrossRef]
10. Sun, L.; Wang, L.Y.; Qian, Y.H.; Xu, J.C.; Zhang, S.G. Feature selection using Lebesgue and entropy measures for incomplete neighborhood decision systems. *Knowl.-Based Syst.* **2019**, *2019*, 104942. [CrossRef]
11. Sun, L.; Xu, J.C.; Liu, S.W.; Zhang, S.G.; Li, Y.; Shen, C.A. A robust image watermarking scheme using Arnold transform and BP neural network. *Neural Comput. Appl.* **2018**, *30*, 2425–2440. [CrossRef]
12. Liu, Z.T.; Li, C.G. Censored regression with noisy input. *IEEE Trans. Signal Process.* **2015**, *63*, 5071–5082. [CrossRef]
13. Sun, L.; Zhang, X.Y.; Qian, Y.H.; Xu, J.C.; Zhang, S.G.; Tian, Y. Joint neighborhood entropy-based gene selection method with fisher score for tumor classification. *Appl. Intell.* **2019**, *49*, 1245–1259. [CrossRef]
14. Wu, Q. A hybrid-forecasting model based on Gaussian support vector machine and chaotic particle swarm optimization. *Expert Syst. Appl.* **2010**, *37*, 2388–2394. [CrossRef]
15. Wu, Q.; Law, R. The forecasting model based on modified SVRM and PSO penalizing gaussian noise. *Expert Syst. Appl.* **2011**, *38*, 1887–1894. [CrossRef]
16. Meng, D.Y.; Torre, F.D.L. Robust matrix factorization with unknown noise. In Proceedings of the IEEE International Conference on Computer Vision (ICCV 2013), Sydney, Australia, 1–8 December 2013.
17. 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] [PubMed]
18. 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]
19. Hu, Q.H.; Zhang, S.G.; Yu, M.; Xie, Z.X. Short-term wind speed or power forecasting with heteroscedastic support vector regression. *IEEE Trans. Sustain. Energy* **2016**, *7*, 241–249. [CrossRef]
20. Ma, C.F. *Optimization Method and the Matlab Programing Design*; Science Press: Beijing, China, 2010; pp. 121–131.
21. Schölkopf, B.; Smola, A.J.; Williamson, R.C.; Bartlett, P.L. New support vector algorithms. *Neural Comput.* **2000**, *12*, 1207–1245.
22. Suykens, J.A.K.; Lukas, L.; Vandewalle, J. Sparse approximation using least square vector machines. In Proceedings of the IEEE International Symposium on Circuits and Systems, Genvea, Switzerland, 28–31 May 2000; pp. 757–760.
23. Suykens, J.A.K.; Van Gestel, T.; Brabanter, J.D.; Moor, B.D.; Vandewalle, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.
24. Pontil, M.; Mukherjee, S.; Girosi, F. *On the Noise Model of Support Vector Machines Regression*; A.I. Memo 1651; Center for Biological and Computational Learning: Cambridge, MA, USA, 2000; pp. 316–324.
25. Bofinger, S.; Luig, A.; Beyer, H.G. Qualification of wind power forecasts. In Proceedings of the Global Wind Power Conference (GWPC 2002), Paris, France, 2–5 April 2002.

26. Zhang, Y.; Wan, Q.; Zhao, H.P.; Yang, W.L. Support vector regression for basis selection in Laplacian noise environment. *IEEE Signal Lett.* **2007**, *14*, 871–874. [CrossRef]

27. Randazzo, A.; Abou-Khousa, M.A.; Pastorino, M.; Zoughi, R. Direction of arrival estimation based on support Vector regression: Experimental Validation and Comparison with Music. *IEEE Antennas Wirel. Propag. Lett.* **2007**, *6*, 379–382. [CrossRef]

28. Bludszuweit, H.; Antonio, J.; Llombart, A. Statistical analysis of wind power forecast error. *IEEE Trans. Power Syst.* **2008**, *23*, 983–991. [CrossRef]

29. 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]

30. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.

31. 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]

32. 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.

33. 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]

34. 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]

35. Shevade, S.K.; Keerthi, S.S.; Bhattacharyya, C.; Murthy, K.R.K. Improvements to the SMO Algorithm for SVM Regression. *IEEE Trans. Neural Netw.* **2000**, *11*, 1188–1193. [CrossRef]

36. Chu, W.; Keerthi, S.S.; Ong, C.J. Bayesian support vector regression using a unified loss function. *IEEE Trans. Neural Netw.* **2004**, *22*, 29–44. [CrossRef]

37. Klaus-Robert Sebastia, M.M. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw.* **2001**, *12*, 181–202.

38. Sun, L.; Liu, R.N.; Xu, J.C.; Zhang, S.G.; Tian, Y. An affinity propagation clustering method using hybrid kernel function with LLE. *IEEE Access* **2018**, *6*, 68892–68909. [CrossRef]

39. Rockafellar, R.T. The multiplier method of Hestenes and Powell applied to convex programming. *J. Optim. Theory Appl.* **1973**, *12*, 555–562. [CrossRef]

40. Rockafellar, R.T. Augmented Lagrange Multiplier Functions and Duality in Nonconvex Programming. *SIAM J. Control* **1974**, *12*, 268–285. [CrossRef]

41. Sun, L.; Chen, S.S.; Xu, J.C.; Tian, Y. Improved Monarch Butterfly Optimization algorithm based on opposition-based learning and random local perturbation. *Complexity* **2019**, *2019*, 4182148. [CrossRef]

42. Boyd, S.; Vandenberghe, L. *Convex Optimization*; Cambridge University Press: Cambridge, UK, 2004; pp. 521–620.

43. Wang, S.X.; Zhang, N.; Wu, L.; Wang, Y.M. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [CrossRef]

44. Léon, B. Large-Scale Machine Learning with Stochastic Gradient Descent. In Proceedings of the 19th International Conference on Computational Statistics (COMPSTAT'2010), Paris, France, 22–27 August 2010.

45. Fabbri, A.; Román, T.G.S.; Abbad, J.R.; Quezada, V.H.M. Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market. *IEEE Trans. Power Syst.* **2005**, *20*, 1440–1446. [CrossRef]

46. Guo, Z.H.; Zhao, J.; Zhang, W.Y.; Wang, J.Z. A corrected hybrid approach for wind speed prediction in Hexi Corridor of China. *Energy* **2011**, *36*, 1668–1679. [CrossRef]

47. Wang, J.Z.; Hu, J.M. A robust combination approach for short-term wind speed forecasting and analysis-Combination of the ARIMA, ELM, SVM and LSSVM forecasts using a GPR model. *Energy* **2015**, *93*, 41–56. [CrossRef]

48. Xiu, C.B.; Guo, F.H. Wind speed prediction by chaotic operator network based on Kalman Filter. *Sci. China Technol. Sci.* **2013**, *56*, 1169–1176. [CrossRef]

49. Abdoos, A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **2016**, *203*, 111–120. [CrossRef]

50. 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]

51. Cherkassky, V.; Ma, Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Netw.* **2004**, *17*, 113–126. [CrossRef]

52. Sun, L.; Zhang, X.Y.; Qian, Y.H.; Xu, J.C.; Zhang, S.G. Feature selection using neighborhood entropy-based uncertainty measures for gene expression data classification. *Inf. Sci.* **2019**, *502*, 18–41. [CrossRef]