Received December 1, 2019, accepted December 22, 2019, date of publication December 27, 2019, date of current version January 6, 2020.

Digital Object Identifier 10.1109/ACCESS.2019.2962702

Nonlinear Channel Estimation for OFDM System by Wavelet Transform Based Weighted TSVR

LIDONG WANG, CHUANG GAO, XIN DENG, YANG CUI, AND XUEBO CHEN
School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114051, China

Corresponding author: Xuebo Chen (xuebochen@126.com)

This work was supported in part by the Fund of the Scientific Research Project of Liaoning Provincial Department of Education under Grant 2019LNJC02, in part by the Funds of the National Natural Science Foundation of China under Grant 71571091, Grant 71771112, Grant 51674140, and Grant 61903169, and in part by the Fund of the University of Science and Technology Liaoning Talent Project under Grant 601011507-03.

ABSTRACT An efficient nonlinear channel estimation method for pilot-aided orthogonal frequency division multiplexing system is proposed in this work. The considered channel is selective in time and frequency domain, that is doubly selective channel. Wavelet transform based weighted twin support vector regression is used for channel frequency response estimation, which is suitable for the regression of nonlinear system. Different from traditional support vector regression algorithm, the proposed algorithm gives samples different weights according to their variance calculated based on wavelet transform. The weights are added into both first and quadratic terms of the objective functions to reduce the impact of outliers, which is likely to appear in the received pilot signal polluted by noise. The proposed channel estimation algorithm has good generalization ability and can reduce the influence of overfitting problem. The results of computational tests show that the proposed algorithm is with better estimating performance compared to the classical pilot-aided channel estimation methods.

INDEX TERMS Channel estimation, orthogonal frequency division multiplexing, twin support vector regression, wavelet transform.

I. INTRODUCTION

Orthogonal frequency division multiplexing (OFDM) is a multicarrier modulation technology, which has high spectral efficiency and simple single tap equalizer structure. It is an attractive multicarrier modulation technology, because it divides the whole bandwidth into several overlapping narrow band channels with low bit rate. Furthermore, the inter symbol interference (ISI) is a common problem in physical channels. It can be eliminated by inserting cyclic prefix (CP) in front of OFDM block at the transmitter. OFDM is robust in high delay spread environments and can eliminate the need to equalize the effect of the delay spread. This feature enables the system to allow higher transmitting rates, so OFDM is selected as the standard of digital audio broadcast, digital video broadcast, some wireless local area networks and the 5-th generation cellular systems. It has been proposed to be adopted in high speed train broadband communication system [1]. Performance of the OFDM systems is affected by channel estimation, timing synchronization and mobility.

The associate editor coordinating the review of this manuscript and approving it for publication was Di Zhang.

Due to the expansion of symbol length, OFDM systems are very sensitive to inter carrier interference (ICI) [2]. In mobile communication, ICI is introduced into OFDM symbol as the channel impulse response changes with time, which further reduces the performance. With the increase of mobile speed and carrier frequency, this will become more serious [3]. Meanwhile OFDM can be demodulated by coherent or noncoherent technology. Generally, coherent detection technology uses channel state information, so coherent detection method can get better signal-to-noise ratio (SNR) gain than noncoherent method. This implies, channel state information should be obtained using channel estimation for coherent detection method. In order to achieve high reception quality in applications with large delay and severe Doppler spread, good design of channel estimator is essential.

Many methods have been adopted to estimate the selective channels. Least square (LS), linear minimum mean square error (LMMSE) [4]–[6] and discrete Fourier transform (DFT) based methods [7]– [10] are conventional algorithms to be used. In doubly selective scenarios, basis expansion model (BEM) [11]– [13] and finite state Markov model (FSM) [14] have been proposed to model the channel and the corresponding channel estimation algorithms were proposed. Zhang *et al.* [15] developed an adaptive weighted averaging estimator for OFDM systems.

In selective multipath fading channel, the channel response presents complicated nonlinearities caused by some reasons, such as the saturation of components and the dispersion of optical fiber [16]–[18], which may lower the estimation precision if using linear methods [19]. So it is necessary to use the nonlinear method for channel estimation. Yang *et al.* [20] proposed a channel estimator for doubly selective channels by deep learning method, in which large scale samples should be provided to training the estimation system. While support vector regression (SVR) developed from support vector machine (SVM) is suitable for regression of nonlinear systems and only small scale samples are needed [21]–[22]. SVM is a statistical learning method based on Vapnik-Chervonenkis (VC) dimension theory, which is widely used in classification and regression [23]. The SVM can be extended to the nonlinear situation by adopting of the kernel trick. It has become one of the most effective tools for pattern recognition and system regression [24]. So far, some SVR algorithms have been used for wireless channel estimation. Matilde *et al.* [25] developed a multiple-input multiple-output channel estimation method based on SVR, but the channel was assumed to be non-selective. Djouama *et al.* [26] and Charrada and Samet [19] proposed OFDM channel estimators using SVR for different application scenarios. However, the methods mentioned above are based on the basic SVR, there are still shortcomings in computational complexity and performance. In 2010, the twin SVR (TSVR) was proposed by Peng [27]. It is with high computational speed because it only solves two small-scale quadratic programming problems (QPPs). Later, the concept of ϵ-tube was introduced into ν-support vector regression (ν-SVR) [28]. In the objective function, some data samples are forced to be in ϵ -tube by introducing a parameter ν, which can improve the performance of standard SVR. Enlightened by the idea of ν-SVR and pinball loss [29], Yitian Xu proposed an asymmetric ν-twin SVR regression method for data regression with asymmetric noise [24]. It can be observed that these algorithms only consider the minimization of empirical risk but not that of structural risk. To solve this problem, Shao *et al.* [30] and Rastogi *et al.* [31] introduced different solutions. The common feature of these algorithms is the introduction of regularization terms into the objective functions of TSVR. However, in the most of the proposed algorithms [32]–[35], all of the training data share the same penalties, which may increase the effect of noise and degrade the performance.

In this paper, a novel nonlinear fading channel estimation in OFDM systems is proposed, which is based on wavelet transform based weighted TSVR (WTWTSVR). The contributions of this paper are summarized as follows.

1. The improved TSVR channel estimation method is proposed for the first time to estimate wireless channel parameters in OFDM system, performance of which is better than traditional estimation algorithm such as LMMSE, DFT-based and BEM-based algorithms. Compared with SVR channel estimation method, TSVR has less computational complexity and faster computational speed [27].

2. Wavelet transform is introduced to calculate weights for training data. Due to the time-frequency characteristics of wavelet transform, the proposed algorithm is suitable for the processing of time sequence samples such as channel parameter estimation. The weights are added into both first and quadratic terms of the objective functions to reduce the impact of outliers, which is likely to appear in the received pilot signal polluted by noise. In essence, the WTWTSVR can utilize the prior information of samples and reduce the influence of signal with big noise on regression performance.

The rest of this paper is organized as follows: Section II briefly describes OFDM system. Section III proposes wavelet transform based weighted TSVR channel estimation. Simulation results are described in Section IV to show the performance of the proposed method, and Section V gives the conclusions.

II. SYSTEM MODEL

Figure 1 shows the baseband equivalent model of OFDM system. The sequence $X(k)$ obtained from QPSK or QAM constellation is parsed into blocks of N symbols and then transformed into a time-domain sequence using an N-point inverse discrete Fourier transform (IDFT). To avoid ISI, a cyclic prefix (CP) of length M equal to or larger than the channel order L (channel consists $L+1$ discrete paths) is inserted at the head of each block. The time-domain signal $x(n)$ can be serially transmitted over the fading channel. $x(n)$ can be expressed as

$$x(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) e^{j2\pi nk/N} \tag{1}$$

where $n = -M, \ldots, N-1$, $k = 0, 1, \ldots, N-1$. After CP is removed at the receiver, the received signal in time domain $y(n)$ can be expressed as

$$y(n) = \sum_{l=0}^{L} h(n,l) x(n-l) + v(n) \tag{2}$$

where $v(n)$ is additive white Gaussian noise (AWGN) with zero-mean, variance σ_n^2 and independent with each other, i.e. $E(v(n)v(m)) = 0, \forall n \neq m$. $h(n,l)$ is the baseband-equivalent doubly selective channel impulse response of the lth path ($l = 0, 1, \ldots, L$) at time n, which includes the physical channel as well as filters at the transmitter and receiver.

The matrix form of (2) can be expressed as

$$\mathbf{y} = \mathbf{hF}^{-1}\mathbf{X} + \mathbf{v} = \mathbf{hx} + \mathbf{v} \tag{3}$$

where $\mathbf{X} = [X(0), X(1), \ldots, X(N-1)]^{\mathrm{T}}$, $\mathbf{x} = [x(0), x(1), \ldots, x(N-1)]^{\mathrm{T}}$, $\mathbf{y} = [y(0), y(1), \ldots y(N-1)]^{\mathrm{T}}$, $\mathbf{v} = [v(0), v(1), \ldots, v(N-1)]^{\mathrm{T}}$, $(\cdot)^{\mathrm{T}}$ denotes the transpose operation; $\mathbf{X}, \mathbf{x}, \mathbf{y}, \mathbf{v} \in \mathbb{C}^N$ represent vectors of transmitted signal in frequency domain, transmitted signal in time domain, received signal in time domain and white gaussian

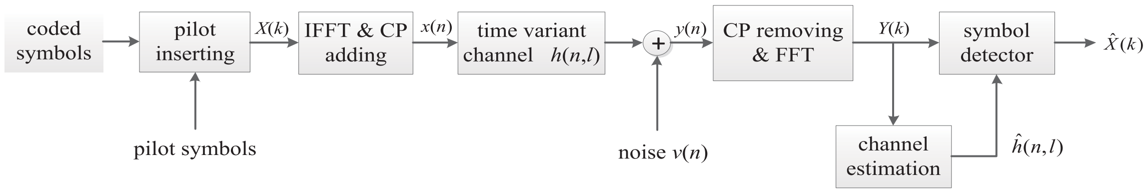


FIGURE 1. Baseband equivalent model of OFDM system.

noise in time domain respectively. $\mathbf{h} \in \mathbb{C}^{N\times N}$ is the channel matrix, the element of which can be expressed as

$$\mathbf{h}(n,p) = \begin{cases} h(n, (n-p)_N), & (n-p)_N \leq L \\ 0, & others \end{cases} \tag{4}$$

where $n, p = 0, 1, \ldots, N-1$. $(\cdot)_N$ represents module-N. $\mathbf{F}^{-1}$ is an N-point IDFT matrix, the entry of which $[\mathbf{F}^{-1}]_{n,k}$=$(1/\sqrt{N})\exp(j2\pi nk/N)$. By the way, define discrete Fourier transform (DFT) matrix $\mathbf{F}$, the entry of which $[\mathbf{F}]_{n,k}$=$(1/\sqrt{N})\exp(-j2\pi nk/N)$.

Perform Fourier transform on both sides of (3) and obtain the following equation,

$$\begin{aligned} \mathbf{Fy} &= \mathbf{FhF}^{-1}\mathbf{X} + \mathbf{Fv} \\ \mathbf{Y} &= \mathbf{HX} + \mathbf{V} \end{aligned} \tag{5}$$

where $\mathbf{Y} = \mathbf{Fy} = [Y(0), Y(1), \ldots, Y(N-1)]^T \in \mathbb{C}^N$ and $\mathbf{V} = \mathbf{Fv}$ =$[V(0), V(1), \ldots, V(N-1)]^T \in \mathbb{C}^N$ are received signal vector and AWGN vector in frequency domain respectively. $\mathbf{H} = \mathbf{FhF}^{-1} \in \mathbb{C}^{N\times N}$ is a frequency domain channel matrix with ICI induced by time variations of the channel, the elements of which can be described as

$$\mathbf{H}(m,p) = \frac{1}{N}\sum_{l=0}^{L} e^{-j2\pi ml/N} \sum_{n=0}^{N-1} h(n,l) e^{-j2\pi(m-p)n/N} \tag{6}$$

where $m, p = 0, 1, \ldots, N-1$. The off-diagonal elements of $\mathbf{H}$ is the ICI response. $\mathbf{H}$ can be divided into two parts, one part $\mathbf{H}_d \in \mathbb{C}^{N\times N}$ is to retain only the principal diagonal elements, the other one $\mathbf{H}_n \in \mathbb{C}^{N\times N}$ is to retain only the off-diagonal elements. Then (5) can be expressed as

$$\begin{aligned} \mathbf{Y} &= \mathbf{H}_d\mathbf{X} + \mathbf{H}_n\mathbf{X} + \mathbf{V} \\ &= \mathbf{H_d}\mathbf{X} + \mathbf{U} + \mathbf{V} \\ &= diag(\mathbf{X})\mathbf{H}_d^{'} + \mathbf{U} + \mathbf{V} \end{aligned} \tag{7}$$

where $\mathbf{U} = \mathbf{H}_n\mathbf{X}$ is the ICI component, $diag(\cdot)$ denotes the diagonal operators and $\mathbf{H}_d^{'} \in \mathbb{C}^N$ is a column vector, the element of which is taken from the principal diagonal element of $\mathbf{H_d}$.

III. NONLINEAR CHANNEL PARAMETERS ESTIMATION FOR OFDM SYSTEMS

Time-varying fading channels have nonlinear characteristics, while typical channel estimation methods mostly use linear interpolation. TSVR is suitable for regression of nonlinear systems because of its kernel mapping technique, so we adopt this method for channel estimation. The TSVR is an extension of the classification tool support vector machine to regression applications, the goal of regression is to get the input-output relationship through the training of sample data. On the basis of Peng's work, many improved algorithms were proposed [29]–[36]. In this section, an improved TSVR-wavelet transform based weighted TSVR is adopted for deep fading channel estimation utilizing the advantages of TSVR regression for nonlinear systems.

Given a training set $\mathbf{Tr} = \{(t_1, z_1), (t_2, z_2), \ldots, (t_m, z_m)\}$, where $t_k \in \mathbb{R}$ and $z_k \in \mathbb{R}$, $k = 1, 2, \ldots, m$. The output can be expressed as $\mathbf{Z} = (z_1, z_2, \ldots, z_m)^T \in \mathbb{R}^m$ and the input vector as $\mathbf{T} = (t_1, t_2, \ldots, t_m)^T \in \mathbb{R}^m$. Let $\mathbf{e}$ and $\mathbf{I}$ be a vector with all elements being one and an identity matrix, respectively.

In high mobility wireless environments, channels undergo selectivity in both time and frequency domains and the doubly selective channels present very complicated nonlinearity especially in fast and deep fading situation. So linear channel estimation methods cannot obtain high performance. Now we adopt a nonlinear WTWTSVR to satisfy the estimation requirements of nonlinear channels since TSVR is superior in solving nonlinear, small training samples and high dimensional pattern recognition. The same as the classical TSVR model [27], the WTWTSVR is constituted by two nonparallel hyperplane. The down-bound is $f_1(t)$, and up-bound $f_2(t)$. Note that each hyperplane is an ϵ-insensitive bound, and the resultant regressor is $f(t) = \frac{1}{2}(f_1(t) + f_2(t))$.

A. CHANNEL ESTIMATION ALGORITHM-WAVELET TRANSFORM BASED WEIGHTED TSVR

Initially, SVR was adopted to deal with the linear regression problem, and the relationship between input and output can be regressed by training samples using hyperplane. However, the hyperplane method is only suitable for linear problems. Based on the theory of Vapnik [23], the algorithm can be extended to nonlinear cases through kernel mapping. The kernel function is used to map the input vector into a higher-dimensional space. The kernel-mapping functions are included: down-bound $f_1(t) = K(t, \mathbf{T}^T)\mathbf{w}_1 + b_1$ and up-bound $f_2(t) = K(t, \mathbf{T}^T)\mathbf{w}_2 + b_2$, where K is a selected kernel function, $\mathbf{w}_1, \mathbf{w}_2 \in \mathbb{R}^m$ and $b_1, b_2 \in \mathbb{R}$ are parameters to be regressed. Therefore, the end regressor is the average of $f_1(t)$ and $f_2(t)$, i.e. $f(t) = \frac{1}{2}, (f_1(t) + f_2(t))$. The optimization problems can be described as (8) and (9) where $v_1, v_2, c_1, c_2, c_3, c_4 > 0$ are pre-selected parameters, ϵ_1 and ϵ_2 are insensitive parameters. ξ_1 and ξ_2 are slack vectors [27], and m is the number of training samples. $\mathbf{d} \in \mathbb{R}^m$ and $\mathbf{D} \in \mathbb{R}^{m\times m}$ are weighting vector and weighting matrix, respectively, calculated based on wavelet transform theory, which will be discussed in details later.

The first term of (8) is the sum of weighted squared distances from down-bound function to the training samples. The second term is regularization term, which can make $f_1(t)$ smooth enough. The function of the third term is to make the 'ϵ tube' as narrow as possible. The combination of the three terms of the objective function in (8) can reflect the idea of structural risk minimization, which can improve the generalization ability of the proposed algorithm and reduce the influence of overfitting problem [27]. The choice of c_1, c_2, v_1 in the objective function of (8) can determine the ratio of three penalty terms. For function (9), we have similar illustrations.

The functions (8) and (9) can be changed into the dual problems by using Lagrangian multipliers, which can be easily solved by the optimization method. Define the Lagrangian function for the quadratic programming problem (QPP) problem (8) as (10), shown at the bottom of the next page, where $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_m)^T$, $\beta = (\beta_1, \beta_2, \ldots, \beta_m)^T$, and γ are nonnegative Lagrangian multipliers. By applying Karush–Kuhn–Tucker (KKT) optimization conditions, the dual problem can be described as (11), shown at the bottom of the next page, where

$$\mathbb{H} = [\, K(\mathbf{T}, \mathbf{T}^T)\ \mathbf{e}\,] \tag{12}$$

Then calculating the dual QPP problem (11), we have

$$\mathbf{u}_1 = [\,\mathbf{w}_1^T\ b_1\,]^T = (\mathbb{H}^T\mathbf{D}\mathbb{H} + c_1\mathbf{I})^{-1}\mathbb{H}^T(\mathbf{DZ} - \alpha). \tag{13}$$

Similarly, the dual problem of (9) is obtained as (14), and we can have

$$\mathbf{u}_2 = [\,\mathbf{w}_2^T\quad b_2\,]^T = (\mathbb{H}^T\mathbf{D}\mathbb{H} + c_3\mathbf{I})^{-1}\mathbb{H}^T(\mathbf{DZ} + \lambda). \tag{15}$$

B. ACQUISITION OF TRAINING SAMPLES

SVR is a supervised machine learning method, which requires input and output of training samples for parameter training. In an OFDM symbol, the transmitting pilot subcarrier positions are expressed as $[m\Delta f]$, $m = 0, 1, \ldots, N_p - 1$, where Δf is the pilot interval in frequency domain and N_p is the number of pilots in an OFDM symbol. Let the transmitting pilot matrix be $\mathbf{X}_P = \mathbf{X}(m\Delta f) \in \mathbb{C}^{N_P}$. Estimate the channel frequency response at pilot subcarriers according to (7) as

$$\hat{\mathbf{H}}_P = (diag(\mathbf{X}_P))^{-1}\mathbf{Y}_P \tag{16}$$

where $\mathbf{Y}_P = \mathbf{Y}(m\Delta f) \in \mathbb{C}^{N_P}$ is the received pilot vector and $\hat{\mathbf{H}}_P = \hat{\mathbf{H}}(m\Delta f) \in \mathbb{C}^{N_P}$ is the estimated frequency response at pilot positions $m\Delta f$. Algorithm described in (16) is the well known least square algorithm. Thus, the input ($\mathbf{T} = [m\Delta f]$) and the output ($\hat{\mathbf{H}}_P$) of training samples for TSVR can be obtained.

Then by interpolation, the frequency response of data position can be calculated and the predicted frequency response of all subcarriers in an OFDM symbol can be expressed as

$$\tilde{\mathbf{H}} = f(\hat{\mathbf{H}}_P) \tag{17}$$

C. WAVELET TRANSFORM BASED WEIGHTING PARAMETERS CALCULATION

The parameters mentioned above, $\mathbf{d} \in \mathbb{R}^m$ and $\mathbf{D} = diag(\mathbf{d})$, where $\mathbf{D} \in \mathbb{R}^{m\times m}$ are the weighting vector and matrix, respectively, and these two parameters can be determined offline according to the variance of training samples. In the classical SVR algorithms, such as functions of TSVR, all of the samples have the same weights, which means that the same penalties are given to training samples. The points with too much noise, such as outliers, will degrade the performance of the regressors. Instinctively, one should give different weights to different training samples, bigger weights

$$\begin{aligned}
&\min_{\mathbf{w}_1, b_1, \xi_1, \epsilon_1} \frac{1}{2}(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1))^T\mathbf{D}(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1)) + \frac{1}{2}c_1(\mathbf{w}_1^T\mathbf{w}_1 + b_1^2) + c_2(v_1\epsilon_1 + \frac{1}{m}\mathbf{d}^T\xi_1)\\
&\quad \text{s.t. } \mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1) \geqslant -\epsilon_1\mathbf{e} - \xi_1\ \xi_1 \geqslant 0\mathbf{e}\ \epsilon_1 \geqslant 0,
\end{aligned} \tag{8}$$

$$\begin{aligned}
&\min_{\mathbf{w}_2, b_2, \xi_2, \epsilon_2} \frac{1}{2}(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_2 + \mathbf{e}b_2))^T\mathbf{D}(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_2 + \mathbf{e}b_2)) + \frac{1}{2}c_3(\mathbf{w}_2^T\mathbf{w}_2 + b_2^2) + c_4(v_2\epsilon_2 + \frac{1}{m}\mathbf{d}^T\xi_2)\\
&\quad \text{s.t. } (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_2 + \mathbf{e}b_2) - \mathbf{Z} \geqslant -\epsilon_2\mathbf{e} - \xi_2\ \xi_2 \geqslant 0\mathbf{e}\ \epsilon_2 \geqslant 0.
\end{aligned} \tag{9}$$

$$\begin{aligned}
&L\,(\mathbf{w}_1, b_1, \xi_1, \epsilon_1, \alpha, \beta, \gamma)\\
&= \frac{1}{2}(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1))^T D(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1))\\
&\quad + \frac{1}{2}c_1(\mathbf{w}_1^T\mathbf{w}_1 + b_1^2) + c_2(v_1\epsilon_1 + \frac{1}{m}\mathbf{d}^T\xi_1)\\
&\quad - \alpha^T(\mathbf{Z} - (K(\mathbf{T}, \mathbf{T}^T)\mathbf{w}_1 + \mathbf{e}b_1) + \epsilon_1\mathbf{e} + \xi_1) - \beta^T\xi_1 - \gamma\epsilon_1,
\end{aligned} \tag{10}$$

$$\begin{aligned}
&\min \frac{1}{2}\alpha^T\mathbb{H}(\mathbb{H}^T\mathbf{D}\mathbb{H} + c_1\mathbf{I})^{-1}\mathbb{H}^T\alpha - \mathbf{Z}^T\mathbf{D}\mathbb{H}(\mathbb{H}^T\mathbf{D}\mathbb{H} + c_1\mathbf{I})^{-1}\mathbb{H}^T\alpha + \mathbf{Z}^T\alpha\\
&s.t.\ 0\mathbf{e} \leqslant \alpha \leqslant \frac{c_2}{m}\mathbf{d},\\
&\quad \mathbf{e}^T\alpha \leqslant c_2v_1.
\end{aligned} \tag{11}$$

should be assigned the smaller variance samples. Motivated by the above idea, the Gaussian function is adopted and the weighting vector $\mathbf{d}\left(=[d_1, d_2, \ldots, d_m]^T\right)$ is determined as:

$$\mathbf{d} = Ae^{\left(-\left|\mathbf{Z}-\hat{\mathbf{Z}}\right|^2/\sigma^2\right)}, \tag{18}$$

where A is the amplitude and σ is the standard deviation, $\hat{\mathbf{Z}}$ represents the estimated vector of output vector $\mathbf{Z}$. In the proposed algorithm, we will calculate $\hat{\mathbf{Z}}$ by three steps based on the wavelet transform method.

Firstly, use wavelet transform to decompose time series. Because the sample sequence is discrete, discrete wavelet transforms (DWT) is adopted. Denote $x_l^a(n)$ as a signal in the l-th decomposition step. The DWT of $x_l^a(n)$ can be calculated by low-pass and high-pass filter banks. The output of the low-pass filter with impulse response $\phi(t)$ is the low-pass parameter x_{l+1}^a, meanwhile for a high-pass filter with impulse response $\psi(t)$, we can obtain the high-pass parameter x_{l+1}^d.

$$x_{l+1}^a(n) = \sum_k \phi(k-2n)x_l^a(k) \tag{19}$$

$$x_{l+1}^d(n) = \sum_k \psi(k-2n)x_l^a(k) \tag{20}$$

The low-pass parameter x_{l+1}^a can be further decomposed into x_{l+2}^a and x_{l+2}^d.

Secondly, the decomposed coefficients $(x_1^d, x_2^d, \ldots, x_l^d, x_l^a)$ are processed by some chosen methods to reduce noise. In the proposed algorithm, we set the high frequency component of the decomposed signal to zero for noise reduction and obtain the denoised sequence $(x_1^{d\prime}, x_2^{d\prime}, \ldots, x_l^{d\prime}, x_l^{a\prime})$.

Thirdly, the estimation output signal $\hat{y}$ is reconstructed by the obtained sequence in the second step $(x_1^{d\prime}, x_2^{d\prime}, \ldots, x_l^{d\prime}, x_l^{a\prime})$.

$$x_{l-1}^{a\prime}(n) = \sum_k \phi(n-2k)x_l^{a\prime}(k) + \sum_k \psi(n-2k)x_l^{d\prime}(k), \tag{21}$$

and the reconstruction can be carried on further, and after l steps, the estimation value $\hat{\mathbf{Z}}$ can be obtained, i.e., $\hat{\mathbf{Z}} = x_0^{a\prime}$. Then $\hat{\mathbf{Z}}$ can be substituted into (18), and get weighting vector $\mathbf{d}$ and weighting matrix $\mathbf{D}$.

D. SUMMARY OF THE ALGORITHM AND COMPUTATIONAL COMPLEXITY ANALYSIS

The channel estimation algorithm can be summarized as:

Input: Pilot series matrix at the transmitter $\mathbf{X}_P$ and its position vector $\mathbf{T} = [m\Delta f]$, $m = 0, 1, \ldots, N_p - 1$; the received pilot vector $\mathbf{Y}_P$; the appropriate parameters $c_1, c_2, c_3, c_4, \nu_1$ and ν_2, in (8) and (9); Gaussian function parameters E, σ^2 in (18).

Output: The predicted frequency response of all subcarriers $\tilde{\mathbf{H}}$.

Process:

1. Estimate channel response at the pilot carries $\hat{\mathbf{H}}_P$ by (16).
2. Let $\mathbf{Z} = \text{real}(\hat{\mathbf{H}}_P)$, where real$(\cdot)$ means taking real part, and preprocess $\mathbf{Z}$ by the wavelet transform method described in subsection (III-C) and get $\hat{\mathbf{Z}}$. The parameter $\mathbf{d}$ can be calculated by (18).
3. In (11) and (14), $\mathbb{H} = [\,K(\mathbf{T}, \mathbf{T}^T)\ \mathbf{e}\,]$, where $\mathbf{T} = [m\Delta f]$, $m = 0, 1, \ldots, N_p - 1$. By Solving the QPP problems in equation (11) and (14), α and λ can be obtained respectively.
4. Calculate $\mathbf{u}_1$ and $\mathbf{u}_2$ by (13) and (15), respectively.
5. Compute $h_{\text{real}}(t) = \frac{1}{2}K(t, \mathbf{T}^T)(\mathbf{w}_1 + \mathbf{w}_2)^T + \frac{1}{2}(b_1 + b_2)$, $t = 0, 1, \ldots, N - 1$.
6. Let $\mathbf{Z} = \text{imag}(\hat{\mathbf{H}}_P)$, where imag$(\cdot)$ means taking imaginary part, repeat 2-5, $h_{\text{imag}}(t)$ can be obtained .
7. Frequency response of all subcarriers in an OFDM symbol can be expressed as $(\tilde{\mathbf{H}})_t = h_{\text{real}}(t) + h_{\text{imag}}(t)j$, where $(\tilde{\mathbf{H}})_t$ means the t-th element of $\tilde{\mathbf{H}}$.

The computational complexity of the proposed method needs to be estimated. Since the calculation amount of QPP and inverse matrices is dominant, the computational complexity is mainly determined by them. Let the number of pilots is N_p, then the calculation of dual QPPs in the WTWTSVR is $O(2N_p^3)$ meanwhile that of the traditional SVR is $O(8N_p^3)$. It can be seen that the calculation burden of SVR is about three times more than that of the proposed algorithm. The matrices inversion in QPPs need computational cost $O(N_p^3)$. The wavelet transform based signal processing also needs some calculation. If Db-3 wavelet with a filter length of 6 is adopted, the computational cost of the wavelet transform is no more than $12N_p$, which can be ignored comparing with that of the QPPs and the matrices inversion. Considering the real part and imaginary part data should be calculated separately, the computational complexity need to be doubled. In summary, the amount of computational complexity of the algorithm is $O(6N_p^3)$.

IV. EXPERIMENTAL RESULTS

In this section, some simulation results of the proposed channel estimation algorithm based on the novel WTWTSVR will be shown. SVR is effective for channel estimation, and its regression performance has been verified [19], [25], so we compare performance of the proposed algorithm with LS estimation with linear interpolation, TSVR [27] and perfect estimation. Consider an OFDM system with doubly

$$\begin{aligned}
&\min \frac{1}{2}\lambda^T\mathbb{H}(\mathbb{H}^T\mathbf{D}\mathbb{H} + c_3\mathbf{I})^{-1}\mathbb{H}^T\lambda + \mathbf{Z}^T\mathbf{D}\mathbb{H}(\mathbb{H}^T\mathbf{D}\mathbb{H} + c_3\mathbf{I})^{-1}\mathbb{H}^T\lambda - \mathbf{Z}^T\lambda \\
&s.t.\ 0\mathbf{e} \leqslant \lambda \leqslant \frac{c_4}{m}\mathbf{d}, \\
&\quad \mathbf{e}^T\lambda \leqslant c_4\nu_2.
\end{aligned} \tag{14}$$

TABLE 1. Default channel parameters.

Parameter	value
Modulation mode	16-QAM
Carrier frequency f_c	2.15 GHz
Number of subcarriers in one symbol N	512
Length of CP L	50
Sampling interval T_s	72 μs
Mobile speed v	120/350 km/h

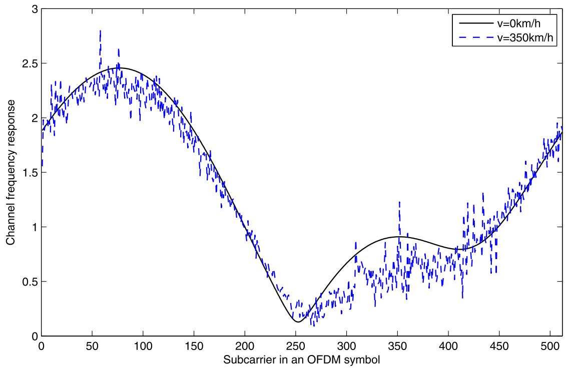


FIGURE 2. Channel frequency response under multipath number being 2.

selective channel. The multipath number $L+1$=51, and assume that the channel taps are independent and identically distributed (i.i.d.) and correlate in time. The correlation function according to Jakes' model [37] can be described as $E[h(n_1, l_1)h^*(n_2, l_2)]=\sigma_h^2 J_0(2\pi f_{\max} T_s(n_1-n_2))\delta(l_1-l_2)$, where $E(\cdot)$ means expected value, $(\cdot)^*$ denotes conjugate, n_k is time index, l_k is channel path index, J_0 is the first kind zeroth-order Bessel function, Ts is the sampling interval in time domain, and σ_h^2 is the variance of the channel. The parameters of the selective channel in the OFDM system follow the default setting as in Table 1.

In order to demonstrate the effects of multipath and moving speed on channel frequency response, two scenarios simulations are performed. Figure 2 shows the channel frequency response at subcarriers in an OFDM symbol under multipath number being 2 for mobile speed equal to 0 and 350 km/h respectively. Figure 3 shows that of multipath number being 50. From Figures 2 and 3 we can see that multipath can cause frequency-domain fading and the more the number of multipaths, the deeper the fading. At the same time, mobile can cause ICI, the faster the movement, the bigger the ICI. According to (7) we know that the ICI is influenced by the product of channel response $\mathbf{H}_n$ and data to be transmitted $\mathbf{X}$. So, if data to be transmitted is random, the ICI is like noise, which can be reflected in the simulations.

In the simulation, pilots are inserted in frequency domain with insertion interval $\Delta f = 2$. The software Matlab R2014a is adopted as the tool of computer simulations The Gaussian nonlinear kernel is used for the proposed WTWTSVR and TSVR [27], that is

$$K(\mathbf{a}^T, \mathbf{b}^T) = exp(-\frac{\|\mathbf{a}-\mathbf{b}\|^2}{e}), \tag{22}$$

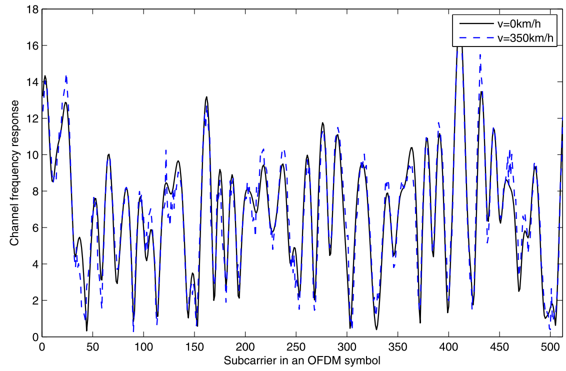


FIGURE 3. Channel frequency response under multipath number being 50.

TABLE 2. WTWTSVR parameters.

	120 km/h	350 km/h
$c_1 = c_3$	0.1	0.1
$c_2 = c_4$	0.1	1
ν	100	10
e	100	10

where e is the variance of the Gaussian function, which can be used to adjust the width of Gaussian function. The selection of parameters is very important for the analysis of system performance. The performance of algorithms is sensitive to some parameters, which need to be carefully selected. In this simulation, parameters are chosen by grid searching method from the set of $\{10^k | k = -4, -3, \ldots, 5\}$. Considering the complexity of calculation, the search dimension of parameters are reduced and let $c_1 = c_3$, $c_2 = c_4$, and $\nu_1 = \nu_2$ in the proposed WTWTSVR. The parameters selected in the WTWTSVR are shown in Table 2.

Figure 4 illustrates the regression of WTWTSVR. In this simulation, SNR=10dB and mobile speed v=120km/h. Other parameters have been given before. We can see that our WTWTSVR fits the channel frequency response well, which can confirm the generalization ability of the proposed algorithm. Additionally, the outliers (pilot samples with too much noise) are ignored and there is no overfitting phenomenon in the regression curve, which shows the robustness of the proposed algorithm.

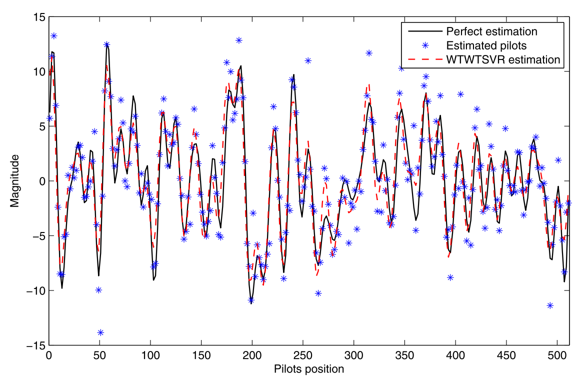


FIGURE 4. Performance of WTWTSVR estimation method.

TABLE 3. SSE values of channel frequency response by three estimation methods.

SNR(dB)	0	5	10	15	20	25	30
				v=120km/h			
Linear	0.7426	0.3973	0.2481	0.1068	0.1160	0.0456	0.0275
TSVR	0.7045	0.3397	0.2058	0.0993	0.1123	0.0446	0.0269
BEM	0.7518	0.4002	0.2505	0.1123	0.1172	0.0466	0.0281
DFT	0.7222	0.3625	0.2297	0.1043	0.1155	0.0453	0.0273
LMMSE	0.7045	0.3399	0.2111	0.1008	0.1128	0.0450	0.0270
WTWTSVR	0.6680	0.3346	0.2008	0.0986	0.1094	0.0431	0.0217
				v=350km/h			
Linear	0.9449	0.4255	0.3336	0.1672	0.1243	0.1078	0.0953
TSVR	0.8834	0.3616	0.2629	0.1534	0.1202	0.1058	0.0937
BEM	0.9662	0.4410	0.3433	0.1697	0.1260	0.1093	0.0962
DFT	0.9286	0.4156	0.3216	0.1655	0.1235	0.1076	0.0950
LMMSE	0.9057	0.3937	0.2928	0.1565	0.1212	0.1066	0.0941
WTWTSVR	0.8573	0.3580	0.2492	0.1520	0.1124	0.1019	0.0903

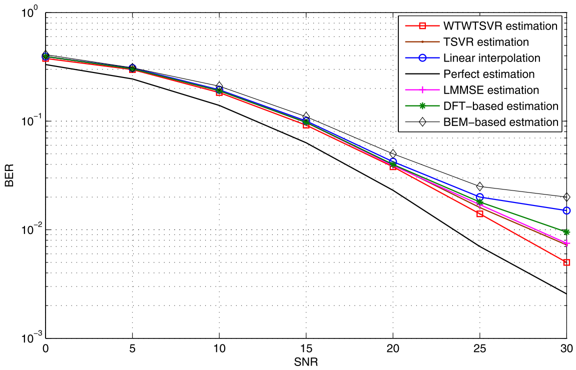


FIGURE 5. BER versus SNR for mobile speed v=120km/h.

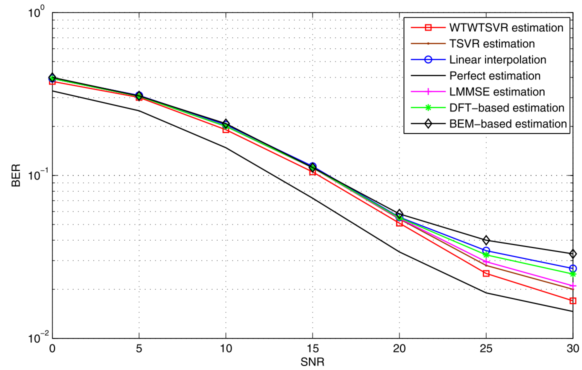


FIGURE 6. BER versus SNR for mobile speed v=350km/h.

The performance of the proposed WTWTSVR and the comparison methods are evaluated by selected criteria. The number of testing samples is denoted as l, y_i is the real value of a testing sample point and $\hat{y}_i$ denotes the predicted value of testing sample point. The criteria are specified as follows.

SNR: Signal-to-noise ratio, defined as SNR $= 10\log(\sigma_x^2/\sigma_v^2)$, where the signal power $\sigma_x^2 = E(|x(k)|^2)$, σ_v^2 is variance of AWGN.

SSE: Sum squared error of testing samples, that is SSE$=\sum_{i=1}^{l} |y_i - \hat{y}_i|^2$.

BER: bit error rate, defined as BER$=N_e/N$, where N_e and N are the number of error signals and all signals in binary respectively.

SSE shows the fitting precision. The SSE value is not the smaller the better, too small value may be caused by overfitting.

Figure 5 shows the BER performance of the proposed algorithm, TSVR-based, linear interpolation, LMMSE, DFT-based, BEM-based methods and perfect estimation in the presence of additive Gaussian noise for mobile speed at 120km/h, and Figure 6 shows that of 350km/h. Table 3 shows the SSE performance of channel frequency response for various algorithms. The results are obtained from the average of 100 tests.

It can be observed from Figures 5 and 6 that the BER performance of all methods is improved as SNR increases, while the performance deteriorates with the increase of moving speed due to the influence of ICI, where we can detect the influence of ICI and noise on system performance. In the same channel environment, the WTWTSVR and TSVR channel estimation algorithms get better performance than others, which means that SVR is suitable for the regression problems based on training samples especially in the nonlinear cases. While the WTWTSVR improved from TSVR outperforms the TSVR, which shows the advantage of the proposed algorithm and demonstrates the effectiveness of utilizing prior information. Although the performance of WTWTSVR is better than that of other methods, the performance improvement in the case of low SNR is not as obvious as that in the case of high SNR, which shows that the larger noise has a greater impact on WTWTSVR, and the wavelet transform preprocessing has limited effect on the improvement of system performance in the case of low SNR. The SSE performance of all algorithms listed in Table 3 presents the similar results as the BER performance. These simulation results demonstrate the good regression performance and effectiveness of the proposed method.

V. CONCLUSION

A novel WTWTSVR based channel estimator is proposed in this paper. The channel estimation is performed in the frequency domain using the inserted pilots. Firstly, the channel frequency response is calculated at the pilot positions and the training samples for the WTWTSVR can be obtained. Then channel frequency response at data positions can be estimated by the proposed algorithm. Unlike the case in TSVR, the proposed algorithm gives samples in different positions different weights according to their variance calculated based on wavelet transform. Compared with the existing methods, the effectiveness of this method is confirmed by computational tests, especially in the case of nonlinearity. In addition, as one of the theoretical basis of this method is wavelet theory, which is suitable for dealing with time series signal denoising, so it is adopted to estimate the channel impulse response, which is time series. Furthermore, due to the influence of computational complexity, SVR method is suitable for data sets with small number of samples, but it will bring huge computational burden for the case with large number of training samples. One can also notice that the proposed channel estimation is performed in frequency domain and the proportion of pilots is large. Therefore, for future work, it is necessary to study the small-scale pilot insertion scheme in the time-domain and frequency-domain.

REFERENCES

[1] X. Yu, Y. Luo, and X. Chen, "An optimized seamless dual-link handover scheme for high-speed rail," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8658–8668, Oct. 2016.

[2] J. Wu and P. Fan, "A survey on high mobility wireless communications: Challenges, opportunities and solutions," *IEEE Access*, vol. 4, pp. 450–476, 2016.

[3] Z. Sheng, H. D. Tuan, H. H. Nguyen, and Y. Fang, "Pilot optimization for estimation of high–mobility OFDM channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 10, pp. 8795–8806, Oct. 2017.

[4] A. Riadi, M. Boulouird, and M. M. Hassani, "Performance of massive–MIMO OFDM system with M–QAM Modulation based on LS Channel Estimation," in *Proc. Int. Conf. Adv. Syst. Emergent Technol. (ICASET)*, Mar. 2019.

[5] D. Neumann, T. Wiese, and W. Utschick, "Learning the MMSE channel estimator," *IEEE Trans. Signal Process.*, vol. 66, no. 11, pp. 2905–2917, Jun. 2018.

[6] K.-C. Hung and D. Lin, "Pilot–based LMMSE channel estimation for OFDM systems with power–delay profile approximation," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 150–159, Jan. 2010.

[7] X. Ye, G. Zheng, A. Zhang, L. You, and X. Gao, "DFT–based low–complexity channel estimation method for millimeter–wave MIMO systems," *Wireless Pers Commun*, vol. 107, no. 1, pp. 205–216, Jul. 2019.

[8] P. Ge and B. Hu, "An efficient channel estimation algorithm for PDM CO-OFDM WDM system," in *Proc. IEEE 3rd Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Mar. 2019, pp. 2527–2530.

[9] F. Gu, Y. Fan, L. Wang, X. Tan, and J. Wei, "A universal channel estimation algorithm based on DFT smoothing filtering," *IEEE Access*, vol. 7, pp. 129883–129891, 2019.

[10] Y. Kang, K. Kim, and H. Park, "Efficient DFT-based channel estimation for OFDM systems on multipath channels," *IET Commun.*, vol. 1, no. 2, p. 197, 2007.

[11] J. K. Tugnait, S. He, and H. Kim, "Doubly selective channel estimation using exponential basis models and subblock tracking," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1275–1289, Mar. 2010.

[12] J. T. Dias, R. C. de Lamare, and Y. V. Zakharov, "BEM-based channel estimation for 5G multicarrier systems," in *Proc. WSA 22nd Int. ITG Workshop Smart Antennas*, Bochum, Germany, Mar. 2018, pp. 1–5.

[13] N. M. Idrees, X. Yu, and A. Springer, "Optimal tracking of doubly-selective radio channels for OFDM based modern wireless systems," *Physical Commun.*, vol. 35, Aug. 2019, Art. no. 100739.

[14] A. Jain, R. Upadhyay, P. D. Vyavahare, and L. D. Arya, "Stochastic modeling and performance evaluation of fading channel for wireless network design," in *Proc. Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Feb. 2007.

[15] M. Zhang, X. Zhou, and C. Wang, "A novel noise suppression channel estimation method based on adaptive weighted averaging for OFDM systems," *Symmetry*, vol. 11, no. 8, p. 997, Aug. 2019.

[16] B. Clerckx and J. Kim, "On the beneficial roles of fading and transmit diversity in wireless power transfer with nonlinear energy harvesting," *IEEE Trans. Wireless Commun.*, vol. 17, no. 11, pp. 7731–7743, Nov. 2018.

[17] H. Hu, M. Senkaya, and Y. Zheng, "A novel measurement of the surface wave dispersion with high and adjustable resolution: Multi-channel nonlinear signal comparison," *J. Appl. Geophys.*, vol. 160, pp. 236–241, Jan. 2019.

[18] Z. Jiang, S. Chen, A. F. Molisch, R. Vannithamby, S. Zhou, and Z. Niu, "Exploiting Wireless Channel State Information Structures Beyond Linear Correlations: A Deep Learning Approach," *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 28–34, Mar. 2019.

[19] A. Charrada and A. Samet, "Joint interpolation for LTE downlink channel estimation in very high-mobility environments with support vector machine regression," *IET Commun.*, vol. 10, no. 17, pp. 2435–2444, Nov. 2016.

[20] Y. Yang, F. Gao, X. Ma, and S. Zhang, "Deep learning–based channel estimation for doubly selective fading channels," *IEEE Access*, vol. 7, pp. 36579–36589, 2019.

[21] D. Gupta, "Training primal K-nearest neighbor based weighted twin support vector regression via unconstrained convex minimization," *Appl. Intell.*, vol. 47, no. 3, pp. 962–991, Oct. 2017.

[22] Y. Xu, Z. Yang, Y. Zhang, X. Pan, and L. Wang, "A maximum margin and minimum volume hyper-spheres machine with pinball loss for imbalanced data classification," *Knowl.-Based Syst.*, vol. 95, pp. 75–85, Mar. 2016.

[23] V. N. Vapnik, *The Natural of Statistical Learning Theroy*. New York, NY, USA: Springer, 1995.

[24] Y. Xu, X. Li, X. Pan, and Z. Yang, "Asymmetric ν-twin support vector regression," *Neural Comput. Appl.*, vol. 2, pp. 1–16, Dec. 2018.

[25] M. Sanchez-Fernandez, M. de-Prado-Cumplido, J. Arenas-Garcia, and F. Perez-Cruz, "SVM multiregression for nonlinear channel estimation in multiple–input multiple–output systems," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2298–2307, Aug. 2004.

[26] A. Djouama, M.-S. Lim, and F. Y. Ettoumi, "Channel estimation in long term evolution uplink using minimum mean square error–support vector regression," *Wireless Pers. Commun.*, vol. 79, no. 3, pp. 2291–2304, Dec. 2014.

[27] X. Peng, "TSVR: An efficient twin support vector machine for regression," *Neural Netw.*, vol. 23, no. 3, pp. 365–372, Apr. 2010.

[28] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, "New support vector algorithms," *Neural Comput.*, vol. 12, no. 5, pp. 1207–1245, May 2000.

[29] X. L. Huang, L. Shi, K. Pelckmansb, and J. A. K. Suykens, "Asymmetric ν-tube support vector regression," *Comput. Statist. Data Anal.*, vol. 77, no. 9, pp. 371–382, 2014.

[30] R. Khemchandani, and S. Chandra, "TWSVR: Twin support vector machine based regression," in *Proc. Twin Support Vector MachinesStud. Comput. Intell.*, 2017, pp. 63–101.

[31] R. Rastogi, P. Anand, and S. Chandra, "A ν-twin support vector machine based regression with automatic accuracy control," *Appl. Intell.*, vol. 46, no. 3, pp. 670–683, Apr. 2017.

[32] S. Balasundaram and Y. Meena, "Training promal twin support vector regression via unconstrained convex minimization," *Appl. Intell.* , vol. 44, pp. 931–955, 2016.

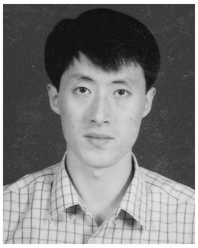
[33] X. Peng, D. Chen, L. Kong, and D. Xu, "Interval twin support vector regression algorithm for interval input-output data," *Int. J. Mach. Learn. Cyber.*, vol. 6, no. 5, pp. 719–732, Oct. 2015.

[34] R. Khemchandani, K. Goyal, and S. Chandra, "TWSVR: Regression via Twin Support Vector Machine," *Neural Netw.*, vol. 74, pp. 14–21, Feb. 2016.

[35] N. Parastalooi, A. Amiri, and P. Aliheidari, "Modified twin support vector regression," *Neurocomputing*, vol. 211, pp. 84–97, Oct. 2016.

[36] Y. Xu and L. Wang, "K-nearest neighbor-based weighted twin support vector regression," *Appl. Intell.*, vol. 41, no. 1, pp. 299–309, Jul. 2014.

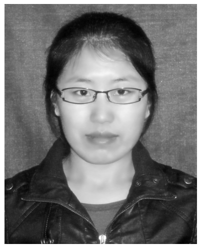
[37] G. L. Stuber, *Principles of Mobile Communication*. Boston, MA, USA: Kluwer, 1996.



LIDONG WANG was born in Liaoning, China, in 1971. He received the B.S. degree from the Harbin Institute of Technology, in 1993, and the M.S. degree from the University of Science and Technology Liaoning, China, in 2004, and the Ph.D. degree from Gyeongsang National University, South Korea, in 2008. Since 1999, he has been with the University of Science and Technology Liaoning, where he is currently an Associate Professor of communication engineering. He has authored over 30 research articles. His research interests include pattern recognition, machine learning, and digital signal processing.

CHUANG GAO was born in Liaoning, China, in 1982. He received the B.S. degree from Warwick University, U.K., in 2005, and the M.S. degree from King's College London, U.K., in 2007. He is currently pursuing the Ph.D. degree with the University of Science and Technology Liaoning, China. He has authored over ten research articles indexed by SCI and EI. His research interests include nonlinear system control, machine learning, and intelligent control.

XIN DENG received the B.S. degree in automation, the M.S. degree in control theory and control engineering, and the Ph.D. degree in iron and steel metallurgy form Northeastern University, Shenyang, China, in 2003, 2007, and 2013, respectively. He is currently an Associate Professor with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China. His research interests include modeling and optimal control of complex industrial processes, adaptive control, and nonlinear control.



YANG CUI received the B.S. degree in information and computing science and the M.S. degree in applied mathematics form the Liaoning University of Technology, Jinzhou, China, in 2009 and 2012, respectively, and the Ph.D. degree in control theory and control engineering from Northeastern University, Shenyang, China, in 2018. She is currently a Lecturer with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China. Her research interests include dynamic surface control, adaptive control, nonlinear control, and multiagent control.

XUEBO CHEN received the D.Eng. from the University of Belgrade, Yugoslavia, in 1994. He is currently working as a Professor with the University of Science and Technology Liaoning. His research interests include theory of large scale systems and control theory.