Liqiang Hou[1], Shanlin Yang[1], Xiaojia Wang[1], Jianxin Shen[1]

School of Management, Hefei University of Technology (1)

Short Term Load Forecasting Based on WLS-SVR and TGARCH Error Correction Model in Smart Grid

***Abstract**. Smart grid is the main development goal of future power grid while the short-term load forecasting is the significant premise of making management, power supply and trading plan in market circumstance. The forecasting accuracy directly determined the safety and economy of electric system. Support Vector Machines (SVM), as the new machine learning method, has applied successfully to short-termed load forecasting. However, research finds out that the singular points of the initial data have impact on forecasting accuracy. So in this paper, firstly, based on the analysis of SVM, we render Weighted Least Square and Support Vector Regression (WLS-SVR) applying to short-termed load forecasting, which overcomes the disadvantage of singular points. Secondly, we offer Threshold Generalized Autoregressive Conditional Heteroskedasticity (TGARCH) model to construct error prediction model to modify the initial predicted value. Finally, according to the PJM historical data, we get the results showing that the accuracy is greatly improving by implementing our methods which makes our methods founded.*

***Streszczenie.** W artykule przedstawiono model przewidywania krótkookresowego obciążenia sieci elektroenergetycznej. W proponowanym rozwiązaniu wykorzystano metodę SVM (ang. Support Vector Machine). W celu eliminacji istniejącego wpływu wartości syngularnych na dokładność wyniku, zastosowano regresję ze średnią ważoną. Dodatkowo wykorzystano model TGARCH w określaniu błędów predykcji. Przedstawiono wyniki badań weryfikacyjnych, przeprowadzonych na rzeczywistych danych. (**Przewidywanie krótkoterminowe obciążenia inteligentnej sieci elektroenergetycznej z wykorzystaniem modelu WLS-SVR oraz korekcji błędów modelem TGARCH**).*

Keywords: Smart grid; TGARCH；error correction；WLS-SVR；Short-term load forecasting
Słowa kluczowe: inteligentne sieci, TGARCH, korekcja błędu; WLS-SVR, przewidywanie krótkoterminowe obciążenia

1. Introduction

Smart grid, as the new electric net system, by data and information way to achieve the opening and sharing information, can highly integrate the data, optimize operation and management of net infrastructure and promote the consumption of electricity by interaction. Recently, many countries and areas have many researches on the smart grid. For instance, in USA, there are IntelliGrid [1] Grid2030 [2] GridWise [3]; In EU, there are smart gird based on European technology [4]. As the massive construction of smart gird, the market revolution even takes bigger stride. But the prediction accuracy and efficiency is on the call of improvement no matter whether on stability or the market allocation of electric net.

Short-termed load forecasting is the base of making purchasing electricity plan and arrangement of operation as to improve the quality of supplying electricity. So we need to improve the prediction accuracy. But most prediction theory and methods are based on statistical regression [5] and time-series model [6-7]. However, the Influencing factors are so many and the load curve is non-linear. In order to combine all the influencing factors, the ANN model [8], fuzzy model [9] based on smart algorithm are used wildly. But the Artificial Neural Networks (ANN) model and fuzzy model lacks the breakthrough of generalization in theory. Recently, a new learning nerve learn LS-SVR attracts lots attention.LS-SVR is a new learning methods rendered by Vapnik [10]. It based on the structural risk minimum principle by solving a quadratic programming to get rid of little sample, non-linear, high-dimension and minimum point in local condition. Suykends [11], expanded the SVR to LS-SVR, which converts the quadratic programming into linear simultaneous equations. The LS-SVR is easy in construction and simple in algorithm with better convergence speed but it has drawback too, such as the looseness. In order to overcome this problem, Suykends renders WLS-SVR.

In practice, the error prediction is a compound stochastic process. The series subtly include some information that the prediction model does not. On one hand, the state series in a model predictive ability determines the ability in certain time (higher, lower or common).On the other hand; it also reflects the predicted errors which are the actual ones. If we can find out the regular pattern of the model by the historical predicted data, we can certainly predict the error in order to modify the initial prediction so we can improve the accuracy [12,13]. As the predicted errors have autocorrelation, by digging out the inner information, we are able to predict value in the next time. Then we modify the prediction value of growth rate of the load. By means of case stimulant, we can conclude that the TGARCH model is high efficient in predict the fluctuate of the short-termed load and reflect the time-change feature of conditional predicted errors, which to some extent improve the prediction accuracy of time series.

In this paper, we combine the advantages of WLS and SVR to construct a WLS-SVR prediction method and use it to predict the short-termed electric load. It can diminish the influence of noise. Then we consider the predicted errors as a time series. By using TGARCH model we get the modified error series to modify the initial predicted value of load. At last, the case study compares ANN, GM (1,1), LS-SVR and the method presented in our paper, proving the superiority of the model constructed in our way.

2. The Traditional Model of LS-SVR

LS-SVR expands standard SVR by optimizing the square of relaxation factors and converting the constraints of inequality to equality, so the quadratic programming problem in traditional SVR becomes linear simultaneous equations, thus the calculating difficulty reduces a lot in company with the solution high efficiency and convergence speeding up.

The basic method of SVR:

Define $x \in R^n$ and $y \in R$, let R^n be the input space, by nonlinear transformation $\phi(\cdot)$, we let in the input space x map into a high dimensional characteristic space where we use the linear function to fit sample data while making sure the generalization.

In the characteristic space, the linear estimation function is defined as:

(1) $$y = f(x,\omega) = \omega^T \phi(x) + b$$

where ω is the weight and b is the skewness.

The aim function is:

(2) $$\min_{\omega,b,\xi} J(\omega,b,\xi)=\frac{1}{2}\omega^T\omega+\frac{1}{2}C\sum_{i=1}^{N}\xi_i^2$$

(3) $$y_i=\phi(x_i)\omega+b+\xi_i \quad i=1,\cdots,N$$

Where $\omega\in R^h$ is the weight vector and $\phi(\cdot)$ is non-linear mapping function, $\xi_i\in R^{N\times 1}$ is relaxation factor, $b\in R$ is the skewness while $C>0$ is penalty factor.

Importing factors, $\alpha_i\in R^{N\times 1}$, we can easily get the function as:

(4) $$L(\omega,b,\xi_i,\alpha_i)=\frac{1}{2}\|\omega\|^2+\frac{1}{2}C\sum_{i=1}^{N}\xi_i^2 -\sum_{i=1}^{N}\alpha_i\left[\phi(x_i)\omega+b+\xi_i-y_i\right]$$

According to the KTT we get

(5) $$\begin{cases}\dfrac{\partial L}{\partial\omega}=\omega-\sum_{i=1}^{N}\alpha_i\phi(x_i)=0\\ \dfrac{\partial L}{\partial b}=\sum_{i=1}^{N}\alpha_i=0\\ \dfrac{\partial L}{\partial\xi_i}=\alpha_i-C\xi_i=0\\ \dfrac{\partial L}{\partial\alpha_i}=\phi(x_i)+b+\xi_i-y_i=0\end{cases}$$

(6) $$\begin{bmatrix}0 & E^T\\ E & \phi\phi^T+C^{-1}I\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix}$$

Where E is the matrix whose elements are all 1, I is a $N\times N$ identity matrix.

Inner product of regression in non-linear function can be replaced by kernel function satisfied *Mercer* . Let $\Omega_{ij}=\phi\phi^T$, then

(7) $$\Omega_{ij}=\phi(x_i)^T\phi(x_j)=K(x_i,x_j)$$

We then have the $LS-SVR$ regression function model

(8) $$f(x)=\sum_{i=1}^{N}\alpha_i K(x_i,x_j)+b$$

Kernel function commonly used in practice are linear kernel, polynomial kernel, and RBF kernel, we use the RBF kernel as our kernel function for its better generalization. The form is as following:

(9) $$K(x_i,x_j)=\phi(x_i)\phi(x_j)=\exp(-\|x_i-x_j\|/2\delta^2)$$

In which the regularization parameter C and kernel breadth δ is the crucial parameters of $LS-SVR$.

3. Modified Model Building

This section presents the fundamental knowledge of LS-SVR. Suppose a set of data

(10) $$T=\{(x_i,y_i)\,|\,x_i\in X\subset R^n, y_i\in Y\subset R, i=1,...,N\}$$

Where $x_i\in R^n$ are given as inputs, $y_i\in R^n$ are the corresponding outputs. SVR theory is to find a nonlinear map from input space to output space and map the data to a higher dimensional feature space through the map, and then the following estimate function is used.

(11) $$f(x)=\omega^T\cdot\phi(x)+b$$

where $\phi(x)$ maps the input data to a higher dimensional feature space, ω is a weight vector, and b is the threshold value. $f(x)$ is the regression estimate function which constructed through learning of the sample set. In the LS-SVR for function estimation, the objective function of optimization problem, is defined as

(12) $$\min_{\omega,b,\xi} J(\omega,\xi_i)=\frac{1}{2}\|\omega\|^2+\frac{1}{2}C\sum_{i=1}^{N}\xi_i^2$$

Subject to the constraints

(13) $$f(x)=\omega^T\cdot\phi(x_i)+b+\xi_i\,,\quad i=1,2,...,N$$

where $\|\omega\|^2$ is the weights vector norm, which is used to constrain the model structure capacity in order to obtain better generalization performance. C is the user-defined regularization constant which balances the model's complexity and approximation accuracy, and ξ_i is the approximation error.

Estimation of support values in the LS-SVR is optimal only when there is a Gaussian distribution of error variables. When, however, a Gaussian assumption for error variables is not realistic, it may lead to less robust estimates. This is because the SSE cost function of the LS-SVR, which assigns an equal weight to error at all times, treating all data equally, gives less precisely measured points more influence than they should have and highly precise points too little influence. To obtain a robust estimate when the distribution is not a normal Gaussian one, a correction must the distribution is not a normal Gaussian one, a correction must be made by defining weights based on the error distribution; the so-called weighted LS-SVR method.

To modify these weights to obtain a robust estimate based on the previous LS-SVR solution. In the main space R^n, a new objective function of the optimization problem shown as follows.

(14) $$\min_{\omega,b,\xi,v} J(\omega,b,\xi,\upsilon)=\frac{1}{2}\|\omega\|^2+\frac{1}{2}C\sum_{i=1}^{N}\upsilon_i\xi_i^2$$

(15) $$s.t.\quad y_i=\phi(x_i)\,\omega+b+\xi_i,\quad i=1,...,N$$

Where J is loss function, $\xi_i\in R^{N\times 1}$ is relaxation factor, $\phi(\cdot)$ is the nonlinear mapping function,

$\phi(\cdot)$ put the input vectors $x\in X$ of system input space $X\in R^n$ map to the input vectors $\phi(x)\in H$ of the high dimensional feature space H .

With the *Lagrange* multipliers $\alpha_i\in R^{N\times 1}$ are introduced , the *Lagrange* function is given by

(16) $$L(\omega,b,\xi_i,\alpha_i,\upsilon_i)=\frac{1}{2}\|\omega\|^2+\frac{1}{2}C\sum_{i=1}^{N}\upsilon_i\xi_i^2 -\sum_{i=1}^{N}\alpha_i\left[\phi(x_i)\omega+b+\xi_i-y_i\right]$$

The Karush-Kuhn-Tucker (KTT) conditions for optimality are given by

$$\begin{cases}\dfrac{\partial L}{\partial\omega}=\omega-\sum_{i=1}^{N}\alpha_i\phi(x_i)=0\\ \dfrac{\partial L}{\partial b}=\sum_{i=1}^{N}\alpha_i=0\\ \dfrac{\partial L}{\partial\xi_i}=\alpha_i-\dfrac{C\upsilon_i\xi_i}{N}=0\\ \dfrac{\partial L}{\partial\alpha_i}=\omega\cdot\phi(x_i)+b+\xi_i-y_i=0\end{cases}$$

After elimination of ξ_i and ω , it reduction to a matrix form

(17) $$\begin{bmatrix}0 & E^T\\ E & \Omega+V_C\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix}$$

There $V_C = diag\{1/C\upsilon_1,...,1/C\upsilon_N\}$, $\alpha = [\alpha_1, \alpha_2, ..., \alpha_N]^T$, $y = [y_1, y_2, ..., y_N]^T$, E is the weight value variable, and E is an $N \times 1$ identity matrix, Ω is $N \times N$ $Hessian$ matrix, and Ω follows Mercer's condition.

(18) $$\Omega_{ij} = \phi(x_i)^T \cdot \phi(x_j) = K(x_i, x_j), \quad i, j = 1, 2, ..., N$$

Esq. (10) and (15) provide the final result of the LS-SVR model for function estimation, and then the following estimate function is the regression function model of WLS-SVR.

(19) $$f(x) = \sum_{i=1}^{N} \alpha_i K(x_i, x_j) + b$$

As choices of kernel function, there are several possibilities, Kernel function commonly used in practice are linear kernel, polynomial kernel, and RBF kernel. The kernel selected in this paper is the RBF kernel as it has better generalization. Its expression form is as following.

(20) $$K(x_i, x_j) = \phi(x_i)\phi(x_j) = \exp(-\|x_i - x_j\| / 2\delta^2)$$

Where the regularization parameter C and kernel breadth δ are the crucial parameters of the WLS-SVR.

The WLS-SVR reflects the behavior of the random errors in the model, through introduce weight factor $\upsilon_i (i = 1,...,N)$ to correct the approximation error vector ξ of the LS-SVR, and then its algorithm has better robust. One common choice for υ_i has been given by Suykens et al.

(21) $$\upsilon_i = \begin{cases} 1 & if \quad |\xi_i/\hat{s}| \le c_1 \\ \dfrac{c_2 - |\xi_i/\hat{s}|}{c_2 - c_1} & if \quad c_1 \le |\xi_i/\hat{s}| \le c_2 \\ 10^{-4} & otherwise \end{cases}$$

(22) $$\hat{s} = 1.483 MAD(\xi)$$

Where $\hat{s}$ is a robust of the standard deviation of the LS-SVR error variable $\xi_i = \alpha_i / C$, which denotes how much the estimated error distribution deviates from a Gaussian distribution. MAD is the middle of error absolute value. The constants c_1, c_2 typically are chosen to be $c_1 = 2.5$ and $c_2 = 3$.

4. Testing the Adequacy of the GARCH Model

Consider a time series model with the following structure:

(23) $$y_t = f(w_t; \varphi) + \varepsilon_t$$

Where f is at least twice continuously differentiable with respect to $\varphi \in \phi$, for all $w_t = (y_{t-1}^{'}, u_t^{'})^{'}$ with $y_{t-1} = (1 - y_{t-1}, \cdots, y_{t-n})^{'} \in R^{n+1}$ and exogenous $u_t = (u_{1t}, \cdots, u_{kt})^{'} \in R^k$, everywhere in ϕ .The error is parameterized as

(24) $$\varepsilon_t = \xi_t h_t^{1/2}$$

Where $\{\xi_t\}$ is a sequence of independent identically distributed random variables with mean zero, unit variance and $E\xi_t^3 = 0$. The conditional variance $h_t = \eta^{'} s_t$ such that $s_t = (1, \varepsilon_{t-1}^2, \cdots, \varepsilon_{t-q}^2, h_{t-1}, \cdots, h_{t-p})^{'}$ and $\eta = (\alpha_0, \alpha_1, \cdots, \alpha_q, \beta_1, \cdots, \beta_p)^{'}$ with $\alpha_0 > 0$,whereas $\alpha_1, \cdots, \alpha_q, \beta_1, \cdots, \beta_p$ satisfy the conditions in Nelson and Cao [14]that ensure the positivity of h_t .These conditions allow some of the parameters to be negative, unless $p = q = 1$ .Esq.(24) is thus the standard $GARCH(p,q)$ model. We assume regularity conditions hold such that the central limit theorem and the law of large numbers apply whenever required. For such conditions in the multi-vitiate $GARCH(p,q)$ case; see Comte and Lieberman [15]. In the univariate case, their conditions require the density of ξ_t to be absolutely continuous with respect to the Lévesque measure and positive in a neighborhood of the origin. Further-more, it is required that $E\xi_t^8 < \infty$, which of course implies further restrictions on the density of ξ_t .

The assumption $E\xi_t^3 = 0$ that Comte and Lieberman do not need guarantees block diagonality of the information matrix of the log-likelihood function. However, it is not just a technical simplification. We shall consider, among other things, a test that has power against asymmetric response to shocks. In deriving such a test it is appropriate to assume that the conditional distribution of ε_t given h_t is not skewed.

In order to consider the adequacy of the GARCH model, we formulate a parametric alternative to the model. Assume that in Esq. (24).

(25) $$\xi_t = z_t g_t^{1/2}$$

Where $\{z_t\}$ is a sequence of independent, identically distributed random variables with zero mean, unit variance and $Ez_t^3 = 0$ .Furthermore, $g_t = 1 + \pi^{'} v_t$,

Where $v_t = (\xi_{t-1}^2, \cdots, \xi_{t-m}^2)^{'}$ and $\pi = (\pi_1, \cdots, \pi_m)^{'}$, $\pi_j \ge 0,$ $j = 1, \cdots, m$ .Eq. (25) may thus be written as

(26) $$\varepsilon_t = z_t (h_t g_t)^{1/2}$$

It could be called an "ARCH nested in GARCH" model as $\xi_{t-j}^2 = \varepsilon_{t-j}^2 / h_{t-j}$, $j = 1, \cdots, m$ .We want to test $H_0 : \pi = 0$ against $\pi \ne 0$ and thus follow the standard practice of choosing a two-sided alternative although the elements of π are constrained to be non-negative. Under this hypothesis, $g_t \equiv 1$, and the model collapses into a $GARCH(p,q)$ model. (For ways of testing $H_0 : \pi = 0$ against $\pi > 0$ when $h_t \equiv \alpha_0$,see Lee and King [16] and Demos and Sentana [17])

We introduce the following notation. Let $\hat{\varepsilon}_t$ and $\hat{h}_t$ be the error ε_t and the conditional variance h_t , respectively, estimated under H_0 ,And $\hat{\xi}_t^2 = \hat{\varepsilon}_t^2 / \hat{h}_t$.

Furthermore,

let $\hat{x}_t = \hat{h}_t^{-1} \partial \hat{h}_t / \partial \eta^{'}$ and $\hat{v}_t = (\hat{\xi}_{t-1}^2, \cdots, \hat{\xi}_{t-m}^2)^{'}$.

The quasi maximum likelihood approach leads to the following result:

Theorem . Consider the Esq. (26) where $g_t = 1 + \pi^{'} v_t$ and $\{z_t\}$ is a sequence of independent identically distributed random variables with zero mean; unit variance and $Ez_t^3 = 0$ .Under $H_0 : \pi = 0$, the statistic

$$LM_\pi = (1/4T)\{\sum_{t=1}^{T} (\hat{\varepsilon}_t^2 / \hat{h}_t - 1)\hat{v}_t^{'}\} V(\hat{\eta})^{-1} \{\sum_{t=1}^{T} (\hat{\varepsilon}_t^2 / \hat{h}_t - 1)\hat{v}_t\}$$

where

$$V(\hat{\eta})^{-1} = (4T/\hat{k})\{\sum_{t=1}^{T} \hat{v}_t \hat{v}_t^{'} - \sum_{t=1}^{T} \hat{v}_t \hat{x}_t^{'} (\sum_{t=1}^{T} \hat{x}_t \hat{x}_t^{'})^{-1} \sum_{t=1}^{T} \hat{x}_t \hat{v}_t\}^{-1}$$

with $\hat{h}_t\hat{k} = (1/T)\sum_{t=1}^{T}(\hat{\varepsilon}_t^2/\hat{h}_t - 1)^2$ is a consistent estimator of the inverse of the covariance matrix of the partial score under the null hypothesis; has an asymptotic χ^2 distribution with m degrees of freedom.

5. TGARCH Error Correction Model Building

As the returns of load is a time-series, in clear illustration, we let the time space as $t \in D^- + D^+$,

Where the D^- represents the past time space, so the returns are the foundation in constructing the model as they are known to all.

D^+ represents the prediction time space, where we have no idea of the rates. Assumed that the historical data can reflect the time change feature of returns, we are able to construct a prediction model that reporting the history data variation tendency.

Thus we get predicted value series $\{\hat{y}_t : t = 1, \cdots, N\}$ and corresponding error value series $\{\varepsilon_t : t = 1, \cdots, N\}$ .What we care is the key factors while ignoring the secondary factors, which leads to prediction error. In order to fix it, we bring in the modification process. In fact, we can let the product error be a series $\{\varepsilon_t : t = 1, \cdots, N\}$ and consider ε_t to be a new stochastic process, so we can build the prediction model for ε_t and get its residual error series $\{\hat{\varepsilon}_t : t = 2, \cdots, N+1\}$, the adjusted initial predicted value are

(27) $$y'_t = \hat{y}_t + \hat{\varepsilon}_t$$

The model of TGARCH is used to describe the stable stochastic time series. By logarithmic transformation， conditional variance may turns out to be negative. Besides, during the calculation there are no parameter constraints, so the complex of calculation is trimmed while the efficiency goes up. Meanwhile, it can easily used in auto-correlation elimination of fluctuation ration in asymmetric information circumstance.

This paper applies the TGARCH model to construct an error forecast model based on the returns of load predicted error time series.

The chosen TGARCH model is:

(28) $$\begin{cases} y_t = \theta_0 + \theta_1 x_{1t} + \cdots + \theta_k x_{kt} + r_t \\ r_t = \sigma_t \varepsilon_t \\ \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i r_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2 + \theta d_t r_{t-1}^2 \\ \varepsilon_t \sim I.I.D \end{cases}$$

The model of TGARCH is used to anal the stability and not pure stochastic series. Thus, after we get the error series of returns of load, what we do first is undertaking stability test of root of utility. If unstable, then use difference method to make it stable. Next, We can get the TGARCH order (p,d,q) after knowing Lagrange multiplier LM and Q statistic for correlation test, which take step after ACF and $PACF$. We estimated the parameters by some methods to choose the best TGARCH model to fit error series.

The specific model constructing and prediction flow path are showed in figure 1.

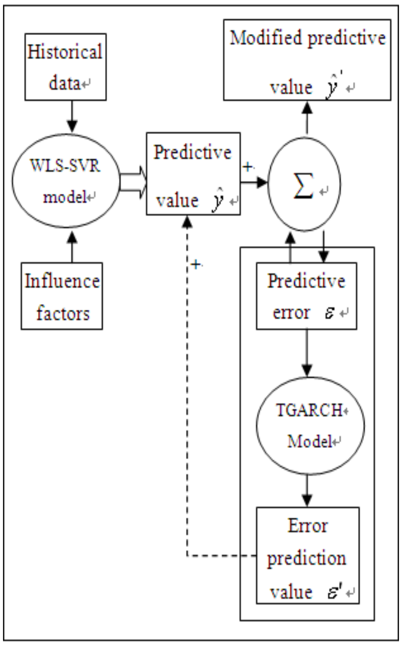

Fig.1. The flow path of constructed model including error correction

6. Case Studies

6.1 Selection of Trained Sample Data

In this paper, we choose hourly load data from May 1, 2007 to July 31, 2007 of Pennsylvania—New Jersey—Maryland (PJM) as the training sample data. There are 2193 valid sample data. We choose hourly data as the sample data series to analyze. And then, we predict the change rate of the 168 hourly time point. After getting $\hat{y}_t, t = 1,2,\cdots,168$, we can use those predicted error $\varepsilon_t, t = 1,2,\cdots,168$ in according with $\hat{y}_t, t = 1,2,\cdots,168$ to construct time series model, meanwhile we predict the afterward 24 errors and get $\hat{\varepsilon}_t, t = 169,170,\cdots,192$ .Finally, and we get the eventual predicted value by means of the formula (27).

We suppose hourly electricity load as k_t， and define the change rate of electricity load as y_t， whose method of calculating is as below:

(29) $$y_t = \ln(k_t/k_{t-1})$$

6.2 WLS-SVR Prediction Model Using Trained Sample Data.

Here we implement WLS-SVR model to the historical data and meanwhile to get the predicted error. The effect is showed as the below figure.

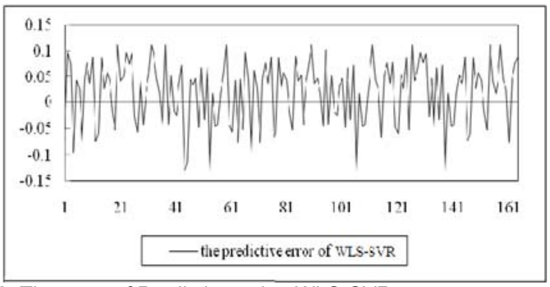

Fig. 2. The error of Prediction using WLS-SVR

6.3 Error Predictive Model and Analysis

As shown in figure 2, we need to have a further analysis on the predictive error ε_t

Table .1. Output of descriptive statistics of Predictive error ε_t

N	168
Mean	0.02673
Std Deviation	2.24623
Skewness	-0.33733
Kurtosis	5.37425
Maximum	0.08575
Minimum	-0.09346
Jarque-Bera	479.3341
Probability	0

We can see from the Table 1 that the kurtosis is 5.37425, much bigger than the kurtosis of normal distribution, namely 3. The skewness is-0.33733,J-B statistics is 479.3341 and P value is so much close to 0.It 's clear that the time series of predictive error of returns don't obey the normal distribution, for it equipped with' High Kurtosis and Fat Tail', antisymmetric, Mean 0 and left skewness.

Table .2. The Stability Test Result of predictive error ε_t

	ADF	Magnificent Level
$\ln \varepsilon_t$	-39.3018	(1%) 3.426985
		(5%) 2.876217
		(10%) 2.567311

We can see for Table 2 that the ADF statistic of ε_t is -39.30182, smaller then the marginal value -3.426985 in significance level 1%. Absolutely it is smaller than the other two marginal values. So the assumption that the time series of ε_t is stable is of sound ground.

Table .3. Testing the ARCH impact on ε_t

F-statistic	18.996	P value	0.000
R^2	71.468	P value	0.000

From the Table 3, we can see the feature aspect of ε_t

The fluctuation of the series has the gregarious and consistency. So we can infer that the fluctuation has ARCH effect. However, up to now we still have to test in quantity aspect. In the context, we apply LM to test its ARCH effect. When the $q = 4$, we get the result show as in Table 3.

The concomitant probability of both F and R^2 are very low which indicating that only if $q \geq 4$, there exists high order ARCH effect. In a result, it's not a smart idea to apply $ARCH(q)$ model when studying ε_t we should consider using GARCH to construct our model and predict values.

Given by estimation of parameters of TGARCH model, we use the estimation into our TGARCH regression function.

(1) Mean value function:

$$\ln \varepsilon_t = 1.00375 \cdot \ln \varepsilon_{t-1}$$

(2)Variance function:

$$\sigma_t^2 = -0.312587 + 0.147986 \times r_{t-1}{}^2 + 0.861578 \times \sigma_{t-1}^2 - 0.128 r_{t-1}{}^2$$

$R^2 = 0.9976$, $DW = 1.97$, Log likely hood =2569.3, $AIC = -4.96$, $SC = -4.96$.

The above results indicate that every parameter is significant. The R^2 of the model is so big as 0.9976 with $\gamma = 0.001725 \neq 0$. That proves there is exist asymmetric error information. From all the above results, we can safely believe the success of fitting data by the method rendered in this paper.

After the further step on the prediction of error ε_t by using TGARCH model, we make a comparison between the modified $\hat{\varepsilon}_t$ and the original ε_t in the following figure.

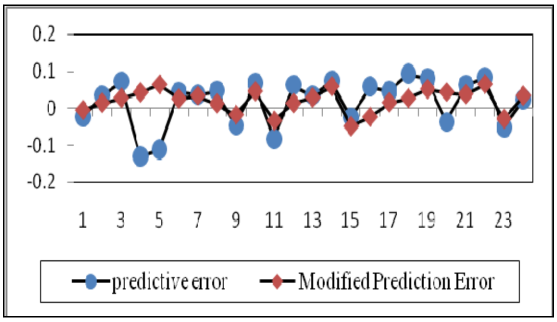

Fig. 3. Modified $\hat{\varepsilon}_t$ vs. original ε_t

The modified errors have the tendency of whole shrinking as we can see obviously from Figure 3.

6.4 Select the Two Ratios Below to Justify the Superiority of the Methods in This Paper

(30) (a): $MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|(x_t - \hat{x}_t)/x_t\right|$

(31) (b): $MSPE = \frac{1}{N}\sqrt{\sum_{i=1}^{N}[(x_t - \hat{x}_t)/x_t]^2}$

x_t is the real value of t moment while $\hat{x}_t$ is the predicted value by some method. We list different results using different methods in the Table 4

Table .4. Compare Different methods' ratio of $MAPE$ and $MSPE$.

ration	$MAPE$	$MSPE$
Our model	0.0356	0.0175
$LS-SVR$ model	0.0798	0.0581
GM model	0.1135	0.0816
$ARMA$ model	0.1247	0.0847

From Table 4, we can see the two ratios resulted from our method are less than other methods, which justify the fact that the model we presented here is effect in improving accuracy.

7. Conclusions

Short-termed load forecasting and the predicted error model is one of the key problems that require attention and solution in the working of power system in smart grid situation. Load has stochastic and stable time features, so the single traditional forecasting model fails to satisfy the need of decision. In this paper, we analyze the selection of SVR parameters and render to apply the WLS-SVR model to construct the short-termed load forecasting, which overcomes the drawback of singular points in traditional SVR. The new model shows that it has advantage in non-linear data disposal and generalization of unknown sample data. Meanwhile, we model and analysis the predicted error series by using TGARCH model which results in modified error series. The modified one is better in accuracy than the initial one. The prediction method in our paper reveals the discipline of load historical data and the correlation between the prediction methods and their predicted variance, which opens a brand-new way to forecast short-termed load with bright application future.

Acknowledgements

This paper was supported by the National Natural Scienc e Foundation of China Grant No.71101041 and No.7107104 5, National 863 project Grant No. 2011AA05A116, Foundati on of Higher School Outstanding Talents Grant No. 2012SQ RL009 and National Innovative Experiment Program No.11 1035954.

REFERENCES

[1] HAASE P. IntelliGrid: a smart network of power [EB/OL]. [2008-0721].http://mydocs.epri.com/docs/CorporateDocuments/EPRI_Journal/2005-Fall/1012885_IntelliGrid.pdf

[2] U.S. Department of Energy. Grid2030: a national vision for electricity's second 100 years [EB/OL]. [2008-07-21]. http://www.oe.energy.gov/Documents and Media/Electric_Vision Document.pdf.

[3] The Grid Wise Alliance. Gridwise action plan [EB/OL]. [2008-04-01]. http://www.gridwise. org/pdf/actionplan.pdf.

[4] European Commission. European technology platform smart Grids: vision and strategy for Europe's electricity networks of the future [EB/OL]. [2008-07-21]. http://www. smart grids.eu/ documents/vision.pdf.

[5] Aldo Goia, Caterina May, Gianluca Fusai, Functional clustering and linear regression for peak load forecasting, International Journal of Forecasting, 22 (2010), 1-11.

[6] Sp.Pappas, L.Ekonomuou, P.Karampelas, Electricity demand load forecasting of the Hellenic power system using an ARMA model, Electric Power Systems Research, 80(2010)，256-264.

[7] J.F.Liu, Z.L.Deng, Self-Tuning Weighted Measurement Fusion Kalman Filter for ARMA Signals with Colored Noise, Applied Mathematics & Information Sciences, 6 (2012)， 1-7.

[8] Nima Amjady, Farshid Keynia, A New Neural Network Approach to Short Term Load Forecasting of Electrical Power System, Energies, 4(2011),488-503.

[9] P A Mastorocostas, J B Theocharis, Vassilios S Petridis, A constrained orthogonal least-squares method for generating TSK fuzzy models: Application to short-term load forecasting, Fuzzy Sets and Systems, 118(2001), 215-233.

[10] Cortes C, Vapnik V. Support-vector Networks，Machine Learning，20(1995)，273- 297.

[11] Suykems J A K, Validewalle J, Least squares support vector machine classifiers, Neural processing letters，9(1999)，293-300.

[12] M.Zhou, Z.Yan, et al, A novel ARIMA approach on electricity price forecasting with the improvement of predicted error, Proceedings of the CSEE, 6(2004),63-68.

[13] W.M.Liu,K,Yang, Day-ahead Electricity Price Forecasting with Error Calibration by Hidden Markov Model, Automation of Electric Power Systems, 4(2009)，34-37.

[14] Nelson, D.B., Cao, C.Q., Inequality constraints in the univariate GARCH model, Journal of Business and Economic Statistics, 10(1992), 229–235.

[15] Comte, F., Lieberman, O., Asymptotic theory for multivariate GARCH processes, Journal of Multivariate Analysis, forthcoming (2000).

[16] Lee, J., King, M., A locally most mean powerful based score test for ARCH and GARCH regression disturbances, Journal of Business and Economic Statistics, 11(1993), 17–27.

[17] Demos, A., Sentana, E., Testing for GARCH ：a one-sided approach Journal of Econometrics, 86(1998), 97–127.

[18] Henrique S, James W. Taylor, An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-load forecasting, Neural Networks, 5(2010),386-395.

[19] J.Z.Wang, S.L. Zhu et al, Combined modelling for electric load forecasting with adaptive particle swarm optimization,Energy, (2010)，1671-1678.

[20] Diyar Akay, Mehmet Atak, Grey prediction with rolling mechanism for electricity demand forecasting of Turkey. Energy, (2007)，1670-1675.

[21] Ehab E. Elattar, John Goulermas,et al, Electric Load Forecasting Based on Locally. Weighted Support Vector Regression. IEEE Transactions on systems, (2010)，438-447.

[22] P.F. Pai, K.P. Lin et al, Time series forecasting by a seasonal support vector, regression model, Expert Systems with Applications, (2010)，4261-4265.

[23] Henrique S, James W. Taylor An evaluation of Bayesian techniques for controlling model complexity and selecting inputs in a neural network for short-load forecasting, Neural Networks, (2010),386-395.

[24] Diyar Akay, Mehmet AtakGrey prediction with rolling mechanism for electricity demand forecasting of Turkey . Energy, (2007),1670-1675.

***Authors**: Li qiang Hou, Shan lin Yang, Xiao jia Wang ，Jian xin Shen are with the School of Management, Hefei University of Technology, Anhui Hefei 230009 P. R. China.*
(Email: liqianghou1020@gmail.com;tonyson800@sina.com)