A Novel Strategy for Short-Term Prediction of Fading Channel

Yongbo Sui¹, Yigang He¹,², Luqiang Shi¹, Yuang Huang¹, Tongtong Cheng¹ and Yuting Wu¹

¹ School of Electrical and Automation Engineering, HeFei University of Technology, HeFei, Anhui Province, China, 230009.
² The School of Electrical Engineering, Wuhan University, Wuhan, China, 430072
Corresponding author: Yigang He (e-mail: 18655136887@163.com).

Abstract. In this paper, a novel prediction strategy based on firefly algorithm, echo state network and Savitzky-Golay filter for short-term prediction of fading channel is proposed and introduced in detail. This hybrid strategy includes two parts, that is, optimization part and repairing part. In the former, parameters of echo state network are optimized by firefly algorithm so that an optimal prediction echo state network is obtained, then in order to improve performances in prediction strategy, prediction sample of echo state network is repaired by repairing strategy based on Savitzky-Golay filter. Finally, the classic Rayleigh channel is tested and performances are analysed and compared with classic AR, which verified the validity and correctness of our proposed prediction strategy.

1. Introduction
Fading always exists in wireless communications when transmitting terminal sends wireless signals to receiving terminal, including long-term fading and short-term fading. Among them, the former is mainly path loss and the signals strength decreases gradually. Oppositely, the latter donates changes dynamically of wireless signals strength caused by multi-path transmission in a short time. Compared with the former, the short-term fading is more complex due to changes dynamically. Hence, how to obtain channel information (CI) of short-term fading channel is a significant issue.

When to obtain CI, channel estimation is a useful approach. Generally, channel estimation works in duplex mode, that is, the pilots are sent in uplink transmission and the CI of target channel are transmitted back to the reverse link, simultaneously [6]. However, since the millimeter and terahertz are widely applied in wireless communication, due to the unavoidable process and feedback delay, channel estimation may be outdated and channel prediction may has more advantages than it. In order to describe characters of shore-term fading channel, Eyceoz, T introduced a classic channel prediction algorithm, that is, autoregressive (AR) algorithm, which channel data in current time is regarded as basics to predict next data and in this way a series of channel data is obtained. Other well-known and effective method is classic Jakes, which utilizes superimposed sinusoids to describe short-term fading channel [2]. Although methods above have satisfactory performances, as statistical methods, they may be limited in complex scenarios. Taking this into consideration, nonlinear channel prediction is a certain more useful method. Such as Nikola M made a series of research by artificial neural network (ANN) to predict short-term fading channel in 2009[3], 2011[4] and 2014[5]. The core of his proposed way is artificial neural network with several input samples and driving signal (DS) extractor, where signal generator is utilized to improve performances of artificial neural network. Sun Jiancheng et al. completed channel prediction of short-term fading using support vector machines (SVM) and embedding phase space, which obtained
more channel state information [7]. However, how to choose a couple of proper embedding dimension
and time delay may be hard for different channel systems. Echo state network is a novel nonlinear
machine learn tool proposed by Germany scholar Jaeger H in 2002. Compared with traditional neural
network (NN), its main advantage is that can record temporary data due to vast neurons in reservoir.
Hence, it has been widely applied for various areas, including nonlinear time data prediction. Y. Zhao
et al. successfully used echo state network to predict Ricean fading channel [8]. However, whole
performance may be limited for manual parameters in prediction model. Besides, to the best of our
knowledge, other related literatures by combing echo state network and wireless channel model are few,
which motiies us to investigate deep research.

This paper focus on the short-term prediction of fading channel and proposes a series prediction
strategy based on firefly algorithm (FA), echo state network (ESN) and Savitzky-Golay filter (S-G filter),
including two parts: optimization part and repairing part. The main contributions are listed:
A: An optimal predictor based on FA-ESN is proposed. FA is drawn to optimize spectral radius and
sparsity degree in internal reservoir of ESN instead of manual set. Due to optimal parameters, the
proposed strategy can improve system performance.
B: In order to further improve performances and reduce prediction error, a repairing mechanism
based on S-G filter is introduced. Prediction channel data of ESN is detected and repaired by repairing
strategy.

Then classic Rayleigh channel is tested and some comparisons are made to verify the correctness and
validity with classic AR.

So the rest of our paper is as followed: Section 2 introduces basic related theories, including ESN,
FA and S-G filter. Then, our proposed short-term prediction method of fading channel is explained in
detail. Taking classic Rayleigh channel in consideration, some comparisons are made to verify the
correctness and validity with classic AR.

2. Background and Related Theories

2.1. Classic Rayleigh Channel

Due to various scenarios, different channel models exist and are studied by scholars. In our paper, the
classic Rayleigh channel is regarded as the target channel model. So its model between transmitting
terminal and receiving terminal are

\[ r(t) = h(t)s(t) + \varepsilon(t) + \xi(t), t > 0 \]  

Where \( s(t) \) is the transmitting signal, \( r(t) \) is the receiving signal, \( h(t) \) is the fading model and \( \varepsilon(t) \) is
the extra Gauss white noise. Generally, Rayleigh channel is statistically described as low pass fading
process, that is, Jakes. And in receiving terminal, the receiving signal \( r(t) \) has various amplitudes and
phases due to multipath effect, even if those come from the same transmitting signal. The channel model
is defined by

\[ h(t) = \sum_{n=1}^{N} A_n e^{j(2\pi f_c t + \phi_n)} \]  

Where \( N \) is the number of valid transmission paths, \( A_n \) is the amplitudes in the \( n \)-th transmission
path, \( f_c \) is the Doppler frequency shift and the \( \phi_n \) donates the phase. \( f_c \) Is calculated by

\[ f_c = \frac{V}{C} f_d \cos \phi \]  

Where \( V \) is the movement speed, \( C \) is the speed of electromagnetic wave, \( f_d \) is the maximum
Doppler frequency shift and \( \phi \) is the angle between transmitter and receiver.

For Rayleigh channel, its probability density function is
And cumulative distribution function (CDF) can be expressed by

\[ F(z) = 1 - e^{-\frac{z^2}{2\sigma^2}} \]  

Where \( \sigma \) donates the power of scatter components in channel and \( z \) is the channel envelope amplitude. It is noted that the couple parameters above are two basic aspects to estimate channel model and the PDF yields Rayleigh distribution in classic Rayleigh channel.

### 2.2. Echo State Network (ESN)

Proposed by Jaeger H in 2002, echo state network is a novel and outstanding modified neural network [8]. Unlike traditional neural network, its hidden layer is filled with vast random neurons, hence, it has ability to record short-term data. Fig.1 is the typical structure of echo state network with multiple input and multiple output.

As we can see, echo state network includes three layers, that is, input layer, hidden layer and output layer. It is noted that due to existing vast random neurons, the hidden layer is also called reservoir. Hence, its special three characters are 1: its core is the random and immobile reservoir; 2: the output weight is the only parameter that needs to adjust; 3: the training process can be completed by simple linear regression calculation. So its state variables are [9]

\[
\begin{align*}
    &u(t) = [u_1(t), u_2(t), \ldots, u_M(t)]^T \\
    &x(t) = [x_1(t), x_2(t), \ldots, x_N(t)]^T \\
    &y(t) = [y_1(t), y_2(t), \ldots, y_L(t)]^T
\end{align*}
\]  

Where \( u(t) \in \mathbb{C}^{M \times 1}, x(t) \in \mathbb{C}^{N \times 1} \) and \( y(t) \in \mathbb{C}^{L \times 1} \) are input decrease time signal, internal neurons and output signals, respectively. And \( M, N \) and \( L \) are the dimensional of input layer, internal reservoir and output layer.

![Figure 1. Typical structure of echo state network](image)

Then its internal reservoir is updated by

\[ x(t + 1) = \tanh(W_{in}u(t) + Wx(t) + W_{back}y(t)) + \mathbb{N}(t) \]  

And the output data is obtained by

\[ y(t + 1) = f_{out} \times \{W_{out} \times (u(t + 1), x(t + 1))\} \]  

Where \( f(\bullet) \) and \( f_{out} \) are the internal active function and output active function, respectively. It is noted that \( f(\bullet) \) is defined as hyperbolic tangent function (tanh). \( W_{in} \in \mathbb{R}^{N \times M}, W \in \mathbb{R}^{N \times N}, W_{out} \in \mathbb{R}^{L \times N} \) and \( W_{back} \in \mathbb{R}^{N \times L} \) are input weight, updated weight, output weight and echo weight, respectively. \( \mathbb{N}(t) \) is added Gauss white noise (AGWN).
It is noticed that four parameters can determine whole system performances in internal reservoir for ESN, which are spectral radius in internal reservoir $SR$, the size of reservoir $Z \in \mathbb{Z}^+$, spectral radius in input layer $IS \in (0,1]$ and sparsity degree in internal reservoir $SD \in (0,1)$. And only if $0 < SR < 1$, ESN has echo character.

2.3. Firefly Algorithm (FA)
As a novel and valid swarm intelligent algorithm, since is proposed by Xin-She Yang in 2008[13], firefly algorithm is applied for various areas, such as data optimization [10], optimal coordination [11] and image process [12]. Its main expressions are

$$\chi_i(t + 1) = \chi_i(t) + \beta(\chi_j(t) - \chi_i(t)) + \alpha \cdot \text{rand} - 0.5 \quad (9)$$

$$I = I_o \times e^{-\gamma_i \cdot \chi_i}, \beta = \beta_o \times e^{-\gamma_j} \quad (10)$$

Where $\chi_i(t)$ is the space vector of the $i$-th firefly in $t$-th iteration, $\alpha$ is the random coefficient in $(0, 1)$, $\text{rand}$ generates the random value in $(0, 1)$. $I$ and $\beta$ are the lightness and attraction between the $i$-th and $j$-th firefly.

2.4. Savitzky-Golay filter (S-G filter)
Savitzky-Golay filter is a useful tool for data filter and since it was proposed and introduced by Savitzky-Golay, it is widely applied to denoise data steam [14]. Unlike traditional filters, due to only polynomial least squares fitting, less computations and less process time are required in S-G filter with same precision.

Assuming that sample interval of data is constant, marked as $\Delta \sigma$, and choose some samples $n_i$ and $n_r$ of nearby $\sigma_i$ to obtain polynomial, that is

$$g_i = \sum_{j=0}^{k} b_j (\sigma - \sigma_i) j \quad (11)$$

Where $k = n_i + n_r$ and $g_i$ is the fitting data corresponding to $\sigma_i$. Set samples $\varsigma_j$, so a proper $b_j$ exists to obtain

$$\min \sum_{j=n_i}^{i+n_r} [p_j(\sigma_j) - \varsigma_j]^2 \to 0 \quad (12)$$

So define $A$, $B$ and $Y$ by

$$A = \begin{pmatrix} -\sigma_i^k & \cdots & -\sigma_i & \cdots & 1 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^{(n_i+n_r+1)}, \quad B = \begin{bmatrix} b_k \\ \vdots \\ b_1 \\ b_0 \end{bmatrix} \in \mathbb{R}^{k+1}, \quad Y = \begin{bmatrix} \varsigma_{j-n_i} \\ \vdots \\ \varsigma_j \\ \varsigma_{j+n_r} \end{bmatrix} \in \mathbb{R}^{(n_i+n_r+1)} \quad (13)$$

So we can get

$$A^T AB = A^T Y \min \sum_{j=n_i}^{i+n_r} [p_j(\sigma_j) - \varsigma_j]^2 = \min \| AB - Y \| \quad (14)$$

Assuming that the equation above is optimal, $AB = Y$, that is, $A^T AB = A^T Y$. Due to positive definite matrix $A^T A$, so

$$B = (A^T A)^{-1} A^T Y \quad (15)$$
Finally, the estimated value is

\[ Y_{\text{estimate}} = A(A^T A)^{-1} A^T Y \]  

(16)

3. Short-Term Prediction Strategy of Fading Channel Based on Fa, ESN And S-G Filter

In this section, the proposed short-term prediction strategy of fading channel based on FA, ESN and S-G filter is explained. Although satisficing prediction performances exists in ESN, related parameters in internal reservoir can determine whole system performances, hence, FA is drawn into our proposed strategy to obtain optimal parameters in optimization part. Due to unavoidable errors, a repairing mechanism based on S-G filter is proposed to further improve prediction accuracy in repairing part.

3.1. Optimization Part

In optimization part, the spectral radius in internal reservoir \( SR \) and sparsity degree in internal reservoir \( SD \) are optimized with FA.

Hence, the process of optimization part are followed:

Step1: Obtain short-term fading channel samples. At time \( t \), \( m \) samples are regarded as input data in ESN and the output data \( y(n) \) is defined as \( h((t + m + 1)T_s) \), that is

\[
\begin{bmatrix}
u(t) = [h(tT_s), h((t + 1)T_s), \ldots, h((t + m)T_s)]^T \\
y(t) = h((t + m + 1)T)
\end{bmatrix}
\]  

(17)

Where \( T_s \) is the sample rate. In this way, a series of samples are generated and divided for training samples \( x_t \) and testing samples \( x_c \).

Step2. According to Eq.(5) and Eq.(16), initialize parameters of ESN: the number of input variables \( M \), the dimension of internal reservoir \( L \), the location of valid recording and updating internal reservoir \( L_c \), the number of neurons in internal reservoir \( \infty \), the input weight \( W_{in} \) and the internal updating weight \( W \), which are initialized by \( W_{in} = \text{rand}^{M \times (M+1)} - 0.5 \), \( W = \text{rand}^{M \times M} - 0.5 \).

It is noted that like traditional neural network, a unit bias is needed when to initialize output weight.

Step3: Initialize parameters of FA: the size of fireflies \( \eta \), the dimension of firefly \( D \), the maximum iterations \( n \), the random coefficient \( \alpha \) and the reference accuracy \( \varepsilon \).

Step4: Generate randomly space vectors of all fireflies \( X \) (SD, SR).

Step5: Calculate lightness and attraction degree of all fireflies.

Step6: Update internal variables \( x \) according to Eq. (6).

Step7: Judge whether the current iteration number \( t \) is equal to the location of valid recording and updating internal reservoir \( L_c \) or not. If not, \( t = t + 1 \) and back to Step6.

Step8: Record and update internal variables \( x \).

Step9: Calculate output weight \( W_{out} \) by training output samples.

Step10: Calculate root-mean-square error (RMSE) by

\[
\text{RMSE} = \sqrt{\frac{1}{D} \sum_{i=1}^{D} (\hat{y}_i - y_i)^2}, i = 1, 2, 3, \ldots, D
\]  

(18)

Where \( \hat{y}_i \) and \( y_i \) are prediction samples and testing target samples.

Step11. Update global optimal firefly.

Step12. Judge whether the current accuracy \( r \) is equal or less reference accuracy \( \varepsilon \), whether the current iteration \( T \) is equal to the maximum iterations \( n \). If not, \( T = T + 1 \) and back to Step4.

Step13. Output optimal internal reservoir \( SR \) and sparsity degree \( SR \) in internal reservoir

Step14. Output prediction channel samples \( \hat{y} \).

3.2. Repairing Part

After optimization process and prediction process, optimal internal reservoir \( SR \) and sparsity degree \( SD \) in internal reservoir are obtained in turn. However, due to unavoidable prediction errors exist, hence,
prediction short-term fading channel samples can be further improved. In this subsection, a repairing mechanism based on S-G filter is proposed.

The processes of repairing mechanism is followed:

Step1: Obtain prediction samples in optimal ESN.

Step2: Initialize basic parameters: samples window length $\ell$ and total samples window number $\Lambda$, reference abnormal rate $\tau$ and maximum repairing iterations $\bar{\mathcal{N}}$. It is noted that

$$\lambda = \Gamma(\ell / \ell) + 1$$  \hspace{1cm} (19)

Where $\Gamma(\bullet)$ is rounding function and $L$ is the length of obtained prediction channel samples $\hat{y}$ in ESN.

**Preprocessing:**

Step3: Obtain window samples $\hat{y}_{\text{win}}$.

Step4: Limit clearly abnormal samples by upper limitation threshold $u_\delta$ and lower limitation threshold $L_\delta$.

**Detection and processing:**

Step5: Estimate window samples $\hat{y}_{\text{esti,win}}$ using Eq. (15) based on S-G filter.

Step6: Judge current window number $i$ is equal to defined $\lambda$. If not, $i = i + 1$, then back to Step3.

Step7: Detect abnormal prediction samples in order to repair them. For abnormal prediction samples, a standard value is utilized, that is

$$C_k = \frac{\tau}{\ln(\hat{y}_k + \epsilon)}$$  \hspace{1cm} (20)

And bias can be calculated by

$$\Delta_k = \frac{|\hat{y}_{\text{esti,k}} - \hat{y}_k|}{\hat{y}_k}$$  \hspace{1cm} (21)

Where $\tau$ is the standard coefficient in $(0, 1)$, and $\hat{y}_{\text{esti,k}}$ is the $k$-th estimated samples, $k \in [1, 2, \ldots, L]$ is the index number of $\hat{y}$. When $\Delta_k > C_k$, the $k$-th prediction sample is defined as abnormal point and is repaired in next process.

Step8: Repair abnormal prediction samples. When abnormal prediction samples are detected, the repairing samples are calculated by

$$\hat{y}_{\text{repair,k}} = \lambda \hat{y}_k + (1 - \lambda) \hat{y}_k$$  \hspace{1cm} (22)

Where $\lambda$ is the repairing coefficient in $(0, 1)$ and $\hat{y}_k$ is the average value of two samples nearby the $k$-th sample. In a short, the repairing value is the sum of weighted prediction value and weighted average value of its nearby two samples. It is noted that the more training samples are, the higher the reliability of samples are, and the larger the repairing coefficient is.

Step9: Update prediction samples with repairing values and obtain repairing prediction samples.

Step10: Calculate abnormal rate $P$.

Step11: Judge whether current abnormal rate $P$ is less than reference value $\tau$ or whether the current iteration $j$ is equal to maximum iterations $\bar{\mathcal{N}}$. If not, initialize window sample number $i = 1$ and $j = j + 1$, back to Step3.

Step12: Output repairing channel prediction samples.

**4. Experiments and Discussion**

In order to estimate performances of our proposed prediction mechanism, the related parameters are defined as follows: Classic Rayleigh channel: $f_d = 500$Hz , $T_r = 5*10^{-5}$s , $f_s = 20$kHz , $C = 3*10^8$ m/s ; ESN: $x_c = 10000$ , $x_s = 8000$ , $m = 6$ , $L = 10000$ , $L_x = 100$ , $N = 500$ ; FA: $h = 30$ , $I_o = 1$ , $\beta_o = 1$ , $n = 50$ , $\alpha=0.3$ , $\epsilon=1*10^{-4}$ ; S-G filter: $\ell = 5$ , $\sigma=1*10^{-4}$ , $\mathcal{N}=10$ , $\lambda=0.3$ , $\tau=0.3$ , $\delta_o=4$ and $\delta_i=0$. 

6
4.1. Input Variables and Training Samples.
When to predict channel samples, some samples are regarded as basic data to predict next sample, that is, the \((m+1)\)-th sample is obtained by foregoing \(m\) samples in Eq. (16). It is noted that performance of the whole system is determined by basic data in a way. So in order to investigate and analyze this section, RMSE values with various basic data \(m\) is showed in Fig.2, where training samples \(c_x = 10000\), testing samples \(s_x = 8000\) and the standard coefficient \(\tau = 0.1\). From Fig.2, RMSE has a declining trend when to increase \(m\), and when \(m\) is greater than 3, RMSE comes to be stable gradually, which indicates enough channel information are given to predicate next channel sample data when \(m \geq 3\).

![Figure 2. RMSEs before and after repairs with various m](image1)

![Figure 3. RMSE before and after repairs with various training samples.](image2)

Similarly, training samples is another factor to affect whole performances, and insufficient training samples may be reduce prediction precision. In this part, RMSE values with various training samples are showed in Fig.3, where testing samples \(c_x = 1000\) and the standard coefficient \(\tau = 0.1\). As we can see, RMSE comes be stable when training samples \(c_x\) is more than 4000. So the training samples \(c_x\) is suggested for more than 4000. Moreover, it is noted that no matter how bias data \(m\) and training samples \(c_x\) are set, RMSE after repairs are lower than those before repairs, which indicates the availability of our proposed repairing way.

4.2. Abnormal Rate and Root Mean Square Error
In order to show the availability of repairing process based on S-G filter, the abnormal rate and root-mean-square error in repairing iterations are showed in Fig.4 and Fig.5. In our paper, the maximum repairing number is set as 10, and the abnormal rate is 0.013, which indicates 104 samples are detected in first repairing process. Then with repairing processes, abnormal sample rate is convergent and become to be steady in 7-th repairing process. The root mean square error (RMSE) between repairing channel prediction samples and actual testing samples in repairing iterations is figured in Fig.5. The RMSE value is 0.023 in case of no repairing process. When to add repairing mechanism, the RMSE is also convergent gradually, and become be steady in 6-th repairing process. And maximum error value and errors before and after repairs are showed in Fig.6. As we can see, the maximum error value is 0.4 and the RMSE is 0.022 and after repairing process, the maximum error value is 0.3 and the RMSE is 0.0185, which indicates that after repairing process, the prediction channel samples are more close to actual samples data.
It is noted that the abnormal rate is a relative parameter determined by the standard coefficient $\tau$. Abnormal samples rate and RMSE with various the standard coefficient $\tau$ are showed in Fig.7. According to Fig.7, when to decrease $\tau$, the detected abnormal samples rate are lesser resulting to lower repairing precision, that is higher RMSE, yet lesser computations. On the contrary, higher repairing prediction channel samples are obtained sacrificing more computations when to set smaller standard coefficient $\tau$. Moreover, RMSE is still lower than 0.022 before repairing process, no matter how the standard coefficient $\tau$ is set, which indicates the effectiveness of our proposed repairing mechanism based on S-G filter.

### 4.3. Abnormal Rate and Root Mean Square Error

PDF and CDF are two statistical properties to estimate quality of wireless communication channel. It is noted that various channels yield different PDF and CDF. AR is a classic prediction strategy of short-term fading channel, so in this subsection, AR is implemented with the order 100 of AR filter. PDF curves and CDF curves of theory, FA-ESN-S-G filter and AR are showed in Fig.8. According

| Algorithm   | RMSE(PDF) | RMSE(CDF) | RMSE(LCR) | RMSE(ADF) |
|-------------|-----------|-----------|-----------|-----------|
| AR(100)     | 0.0183    | 0.0028    | 85.0653   | 0.0013    |
| Proposed way| 0.0258    | 0.0057    | 28.6324   | 0.0017    |
Expansions above, for extra project margin, the basic data \( m \) is set as 6 and training samples \( x_c \) is set as 10000. So RMSEs of PDF and CDF values are calculated in Table.1.

As we can see, two curves are more close to theoretical curves, no matter PDF curve in Fig.8(a) and CDF curve in Fig.8(b) and RMSEs of PDF and CDF in FA-ESN-S-G filter are a bit worse than them of AR, which prove envelope of short-term predicted fading channel is close to theoretical envelope and guarantee directly the correctness and validity of our proposed prediction mechanism.

4.4. Level Crossing Rate (Lcr) and Average Duration of Fades (Adf)

LCR and ADF are two statistical parameters in estimate fading channel, which indicate sensibility over certain time. The former presents fading envelop crosses certain level with positive direction, and the latter donates the average duration crossing below certain threshold in receiver. LCR and ADF curves are explained in Fig.9, where the bias data \( m = 6 \) and training samples \( x_c = 10000 \) for extra project margin. RMSEs of LCR and ADF in AR and proposed method are showed in Table.1. As we can see, compared with AR with order 100 of filter, curves of LCR and ADF in FA-ESN-S-G filter are in excellent agreement with theoretical values and although RMSEs of ADF in AR are less to ours in a way, RMSE of LCR in FA-ESN-S-G filter is more superior to ARs, which validates the validity of our proposed prediction mechanism in way.

5. Conclusion

Due to complex characters in fading channel, prediction of short-term fading channel is a significant issue to obtain channel state information. In our paper, a novel prediction strategy based on FA-ESN...
and S-G filter is proposed and explained in detail, including two parts: optimization part and repairing part. Firstly, the spectral radius $SR$ and sparsity degree $SD$ in internal reservoir are optimized by swarm intelligence algorithm FA to obtain optimal prediction ESN model in optimization part. Due to existing unavoidable prediction errors, a repairing mechanism based on S-G filter is introduced to further improve accuracy of prediction short-term fading channel. In repairing part, window samples data is fitted and estimated by S-G filter and abnormal prediction samples are detected. Finally, they are repaired by repairing mechanism. It is noted that some factors, such as basic data, training samples and standard coefficient may affect whole performance of system, so related investigations are given and discussed based on classic Rayleigh channel. Simulations indicates that channel short-term states with high accuracy can be obtained by our proposed prediction algorithm of short-term fading channel based on FA-ESN and S-G filter and compared with classic prediction algorithm of channel AR, proposed way has more excellent advantages to some extent.

6. Acknowledgments
The research is supported by National Key Research and Development Plan "Important Scientific Instruments and Equipment Development", Grant Number: 2016YFF0102200; Equipment research project in advance, Grant Number: 41402040301; State Key Program of National Natural Science Foundation of China, Grant Number: 51637004; National Natural Science Foundation of China, Grant Number: 51577046.

7. References
[1] Eyceoz T, Duel-Hallen A and Hallen H 1997 Prediction of fast fading parameters by resolving the interference pattern The Thirty-First Asilomar Conference on Signals, Systems and Computers. 1 167
[2] Pop M F and Beaulieu N C 2001 Limitations of sum-of-sinusoids fading channel simulators IEEE Transactions on Communications. 49 699
[3] Tomasevic N M, Neskovic A M and Neskovic N J 2009 Short-term fading simulator based on artificial neural networks IEEE EUROCON 1681
[4] Nikola T, Aleksandar N and Natasa N 2011 Short-term fading simulation using artificial neural networks AEU - International Journal of Electronics and Communications 65 641
[5] Nikola T, Aleksandar N and Natasa N 2014 Artificial neural network based simulation of correlated short-term fading AEU - International Journal of Electronics and Communications. 68 301
[6] Peng W, Zou M and Jiang T 2017 Channel Prediction in Time-Varying Massive MIMO Environments IEEE Access 5 23938
[7] Sun J, Zhang T and Liu F 2005 Novel nonlinear prediction algorithm for fast fading channel. Journal of Beijing University of Aeronautics and Astronautics. 31 409
[8] Jaeger H 2002 Tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the echo state network approach.
[9] Zhao Y, Gao H, Beaulieu N C, Chen Z and Ji H 2017 Echo State Network for Fast Channel Prediction in Ricean Fading Scenarios IEEE Communications Letters 21 672
[10] Wang H, Wang W, Cui L, Sun H, Zhao J, Wang Y and Xue Y 2017 A hybrid multi-objective firefly algorithm for big data optimization Applied Soft Computing 69 806
[11] Tjahjono A 2017 Adaptive modified firefly algorithm for optimal coordination of overcurrent relays IET Generation, Transmission & Distribution 11 2575
[12] Su H, Cai Y and Du Q 2017 Firefly-Algorithm-Inspired Framework With Band Selection and Extreme Learning Machine for Hyperspectral Image Classification IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 10 309
[13] Yang X S 2008 Nature-Inspired Metaheuristic Algorithms Luniver Press
[14] Gorry A and Peter 1990 General least-squares smoothing and differentiation by the convolution (Savitzky-Golay) method Analytical Chemistry - ANAL CHEM 62 570