A Hybrid Model Based on Singular Spectrum Analysis and Support Vector Machines Regression for Failure Time Series Prediction: A hybrid model for failure time series prediction

Article *in* Quality and Reliability Engineering · December 2016

DOI: 10.1002/qre.2098

CITATIONS
12

READS
330

5 authors, including:

ji wu
Beihang University (BUAA)
46 PUBLICATIONS **206** CITATIONS
SEE PROFILE

Senzhang Wang
Nanjing University of Aeronautics & Astronautics
107 PUBLICATIONS **773** CITATIONS
SEE PROFILE

Special Issue Article

Quality and Reliability Engineering International

(wileyonlinelibrary.com) DOI: 10.1002/qre.2098

Published online in Wiley Online Library

A Hybrid Model Based on Singular Spectrum Analysis and Support Vector Machines Regression for Failure Time Series Prediction

Xin Wang,[a] Ji Wu,[a*†] Chao Liu,[a] Senzhang Wang[b] and Wensheng Niu[c]

Effectively forecasting the failure data in the maintenance stage is essential in many reliability planning and scheduling activities. Although a number of data-driven techniques have been applied to cope with this issue and achieved noteworthy performance, the reliability prediction problem is still not fully explored, especially for applying the hybridization methods. In this paper, we introduce a hybrid model which integrates singular spectrum analysis (SSA) and support vector machines regression (SVR) to forecast the failure time series data gathered from the maintenance stage of the Boeing 737 aircraft. Two significant components recognized as the trend and fluctuation are extracted from the original failure time series data by using the techniques of SSA and noise test, and the two components are associated with the inherent and operational reliability, respectively. Then two models named as trend-SSA and fluctuation-SVR are individually developed to conduct the tasks of modeling and forecasting the two components. Furthermore, the optimal parameters of the hybrid model are obtained efficiently by a stepwise grid search method. The performance of the presented model is measured against other unitary models such as Holt-Winters, autoregressive integrated moving average, multiple linear regression, group method of data handling, SSA, and SVR, as well as their hybridizations. The comparison results indicate that the proposed model outperforms other techniques and can be utilized as a promising tool for reliability forecast applications. Copyright © 2016 John Wiley & Sons, Ltd.

Keywords: singular spectrum analysis; support vector machines regression; failure time series forecast; hybrid models; grid search method

1. Introduction

Data-driven technique has been performed as an effective method to failure data prediction,[1] especially for complex systems with high reliability demands. For such systems, it is difficult to analyze the numerous factors relevant with failure occurrence because of incomplete coverage of the factors and some unexpected events in practice. For example, aircraft failures occurred in a certain period might be affected by several independent aspects, such as mission intension, flight environment, human factors, and also the aircraft itself. Moreover, each aspect still includes various situations that may significantly affect the curve shape of the failure data, like policy grounding, continuous bad weather, sudden inspection, or poor assembly. To summarize, these induce the inhomogeneity of real reliability performance and the complexity of mathematical formulations and solutions that receive poor achievements.[2] Thus, establishing learning models with historical data and identifying potential rules are currently the popular failure forecast practices. Comparing with the conventional statistical methods, priori assumptions and distributions are not necessary for this kind of non-parametric technique, which is competent to perform the reliability prediction by experimental validations.[3]

There have already been such data-driven paradigms applied in reliability modeling and forecast through machine learning techniques, as well as some statistical theories. These methods conduct the forecast task by empirical regression of failure data or corresponding features mapped in new dimensional spaces. Particularly, artificial neural network (ANN) is a recommended approach for its universality and satisfactory accuracy in the reliability fields such as engine systems[4,5] and software engineering.[6–9] Al-Garni and Jamal[10] utilized the feed-forward network with back-propagation algorithm to specify the failure rate of Boeing 737 tires, which set the flight operational time and the number of landings as model inputs. The prediction by the feed-forward network is more accurate than that by the Weibull model. Zio *et al.*[11] employed infinite impulse response locally recurrent neural network (IIR-LRNN) to predict engine failures by using the data presented by Xu *et al.*[4] The experimental results illustrated that the IIR-LRNN has the superiority to radial basis function and autoregressive integrated moving average (ARIMA), and achieves a comparable performance to multilayer

[a]*School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China*
[b]*College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China*
[c]*Aeronautical Computing Technique Research Institute, Aviation Industry Corporation of China, Xi'an, China*
**Correspondence to: Ji Wu, School of Computer Science and Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China.*
†E-mail: wujizeze@gmail.com

perceptron. Another powerful tool is support vector machines (SVMs),[12,13] which has been successfully used in regression tasks (denoted as SVR) for reliability prediction.[2,14] In contrast to ANN, SVM has its superiorities due to the fact that besides empirical risk minimization, structural risk minimization is also considered to improve the prediction accuracy. Furthermore, the solution process for SVM is equivalent to solving a convex quadratic programming problem so that a global optimum can be solved.[2] On the other hand, Rocco[1] used singular spectrum analysis (SSA) to exploit the engine failure data[4] and to predict. By comparing with previous methods, SSA is comparable to SVR in the literature[2] and is a promising approach for forecasting failure data. In the aforementioned studies, various unitary models have been applied and evaluated by using the component level reliability data provided by Xu *et al.*,[4] which has limited data points and regular curves. But, for the system level reliability data, more complex and irregular curves are depicted and the unitary models could not perform very well. Thus, the hybridization idea is introduced by extracting regular components from the original data and then dealing with them separately. In the view of Rocco,[1] a hybrid approach integrating SSA and SVR was highly suggested for extending the combination idea of Vahabie *et al.*,[15] which might arouse the decomposition capability of SSA and improve the reliability forecast performance. However, to our best knowledge, this idea has not been validated currently.

Apart from the unitary models applied in failure prediction, the hybridization methods are also applied in multiple domains and have been evaluated with competitive performance, especially in the field of meteorology and hydrology.[16–20] Wu and Chau[16] employed modular soft computing models of ANN and SVR to forecast rainfall time series with data-preprocessing techniques of moving average and SSA. The comparison results revealed the benefits of modular models. Samsudin *et al.*[18] proposed a novel time series forecast model based on least squares support vector machine (LSSVM) with the input determination by using group method of data handling (GMDH). The experimental results indicated that this hybrid model (denoted as GLSSVM) provides a promising tool in time series forecast. Chau and Wu[19] developed a hybrid model for daily rainfall prediction, which applied ANN and SVR to three crisp components (low-, medium-, and high-intensity rainfall) obtained from SSA decomposition and fuzzy C-means clustering. The results exhibited considerable accuracy of SSA on improving the forecast performance. Sivapragasam *et al.*[20] introduced a simple but efficient method for rainfall and runoff forecast based on SSA and SVR, which employed SSA decompositions as the high-dimensional input space for SVR modeling, and was validated as a more accurate technique than non-linear prediction[21]. Wen *et al.*[22] applied SSA to decompose the stock price time series into different terms and then introduced their transformations into the SVR model to make the predictions, which also has shown good performance. However, the mathematical definitions for these hybrid models are seldom provided. How to obtain the optimal parameters of the hybrid model is less studied, either. Although several components are generated by decomposing the original data, the decomposition performance is not evaluated, and the inherent meaning of each component produced is not illuminated.

In spite of the relatively good prediction accuracy has achieved from the unitary methods, reliability prediction problems still involve less scientific works[3] and are rarely investigated with hybrid method. In this paper, a hybrid model based on SSA and SVR is proposed for reliability forecast. In the modeling process, the decomposition-SSA model is adopted at first to decompose the raw failure data into two significant components respectively named the trend and fluctuation, which can represent the inherent and operational reliability of system. Then the trend-SSA model and the fluctuation-SVR model are constructed for these two components, separately. Meanwhile, a stepwise grid search method is developed to search for the optimal parameters of this hybrid model, in which the sample size of the trend and fluctuation model is considered together with other parameters such as sliding window and group number. The proposed hybrid model is validated on a new failure dataset collected from the field maintenance work in an airline company. The new dataset differs from the dataset provided by Xu *et al.*[4] in the following two aspects: (i) The failure data point in the new dataset is collected with the number of failures occurred over a constant period of time, such as one month, etc. Thus, this kind of failure time series is more consistent with the actual maintenance activities and can be used directly for guidance in the planning and scheduling practice. (ii) Comparing to the 10 years' time-to-failure data with 40 data points of the turbochargers in diesel engines,[4] a larger time span of 18 years' failure time series with 216 data points is covered, which are gathered from the failure record library for the whole aircraft. The forecast result with the new dataset can provide a useful perspective to the system reliability of the aircraft. Considering the very limited reliability datasets in literature, the new dataset depicts more informative and complex curves and has more challenges for modeling and forecast. Lastly, the experiments are conducted by comparing to several previous forecast techniques and their hybridizations: Holt-Winters, ARIMA, multiple linear regression (MLR), GMDH, SSA and SVR, and the two hybrid models in the literature[20,22]. The results demonstrate the superior performance of the proposed model.

To our best knowledge, the hybrid method involving two or more techniques has not been applied to model and forecast failure time series. The two hybrid models presented in the literature[20,22] which integrate SSA and SVR, are different from our method in the following two aspects: (1) Domain features. Different application domains lead to different data patterns and problem solutions. (i) Data patterns. The rainfall time series data[19] usually has periodic and seasonal components while the stock price time series data[22] is often dominated by the trend. The failure time series data of the system commonly depicts more waved curves in which the periodism and seasonality are not significant and the fluctuation occupies the leading position. (ii) Problem solutions. The rainfall time series can be splitted into three components according to the rainfall intensity[19] while the stock price time series is divided in terms of the economic features.[22] In this paper, the failure time series is decomposed into two components based on the reliability categories. (2) Hybridization mechanism. The two studies all take SSA as a preprocessing technique and the decomposed components are used to construct the high dimensional input space of the SVR model with[22] or without[20] transformation. Whereas, in our study, SSA and SVR are applied to deal with the two components separately, represented as trend-SSA and fluctuation-SVR. The final predicted value comes only from the unitary SVR model in the two studies while it comes from the combined model in our method. We can express the proposed model as '*SSA*^{2}-*SVR*' because SSA is used not only as a decomposition model but also as a forecast model.

In summary, the key contributions are listed as follows: (i) We propose a new hybrid method integrating SSA and SVR to forecast the aircraft failure time series, which introduces the hybridization idea into reliability forecast applications by extracting the two significant

components associated with the inherent and operational reliability. The forecast accuracy is obviously improved comparing to the previous methods. (ii) We propose a novel procedure to construct the hybrid model by conducting the noise test to evaluate the decomposition performance. The algorithms to generate the hybrid model and search for the optimal parameters are also developed.

The rest of this paper is organized as follows: Section 2 introduces the related techniques used in the hybrid model. Section 3 presents the hybrid SSA-SVR model. In Section 4, we show the experiments and results. Lastly, we conclude the paper in Section 5.

2. Related work

This section provides brief introductions to the related techniques including failure rate, SSA, SVR, and other relevant work used in the hybrid model. In this paper, SSA is used to decompose the failure time series and model the trend, while SVR is used to model the fluctuation.

2.1. Failures rate

Equipment failure rate is an important indicator of reliability. Usually, failure rate function $\lambda(t)$ can be expressed as Weibull distribution,[23] in which different parameter values correspond to different stages of the bathtub curve. In practice, complex systems often use mean failure rate to describe the failure rate within a specified period.[24] The formula is shown in Equation (1), where n_i is the number of failures in the usage cycle i, t_i is the corresponded time period, C is the number of cycles, and $MTBF$ represents mean time between failure.

$$\lambda_a = \sum_{i=1}^{C} n_i / \sum_{i=1}^{C} t_i = 1/MTBF \tag{1}$$

2.2. Singular spectrum analysis

Singular spectrum analysis is a non-parametric technique developed to model time series.[25] SSA modeling process consists of two stages: decomposition and reconstruction. In the decomposition stage, the original time series is embedded into a Hankel matrix X (defined as trajectory matrix) with window length L. Equation (2) displays the embedding step in which N is the length of time series and $K = N - L + 1$. Then a set of non-zero eigenvalues λ and corresponding sub matrices E are produced by the operation of singular value decomposition (SVD) to XX^T. The decomposition results are shown in Equation (3), where U_i is the corresponding eigenvector of λ_i and V_i represents the principal component. Additionally, the contribution of E_i is defined by the proportion of λ_i in the total eigenvalues.

$$X = [X_1, \cdots, X_K] = \begin{bmatrix} x_1 & \cdots & x_K \\ \vdots & \ddots & \vdots \\ x_L & \cdots & x_N \end{bmatrix} \tag{2}$$

$$\begin{cases} X = E_1 + \cdots + E_d \\ E_i = \sqrt{\lambda_i} U_i V_i^T \\ V_i = X^T U_i / \sqrt{\lambda_i} \end{cases} \tag{3}$$

In the reconstruction stage, the sub matrices are divided into several groups according to the contributions and features of singular values and then restored to time series by diagonal averaging as presented in Equation (4), where y is the matrix element and g is the time series element with the constraints of $L^* = \min(L, K)$ and $K^* = max(L, K)$. SSA recurrent forecast approach uses the notion of linear recurrent formulae (LRF), which applies a combination of past data points as the current prediction. Golyandina and Korobeynikov[26] also refer to a vector forecast approach, which has better stability but needs more computational cost than the recurrent approach and will be used as the forecast method in this paper.

$$g_k = \begin{cases} \frac{1}{k+1} \sum_{m=1}^{k+1} y_{m,k-m+2} & 0 < k < L^* - 1 \\ \frac{1}{L^*} \sum_{m=1}^{L^*} y_{m,k-m+2} & L^* - 1 < k < K^* \\ \frac{1}{N-k} \sum_{m=k-K^*+2}^{N-K^*+1} y_{m,k-m+2} & K^* < k < N \end{cases} \tag{4}$$

In the SSA process, window length L should not be larger than half of the time series length N and divisible by the conceivable period of the time series. Moreover, a smaller L is used to extract the trend, and a larger L is used to extract the periods.[27] In the grouping process, the number of sub matrices used for modeling can be determined by the accumulated contribution of singular values, which is usually in the range of a threshold denoted as η (0.85–0.95 in general). Furthermore, the grouping scheme should consider the impacts of trend, harmonics, noise, and other factors. Some statistical and quantitative methods are efficient: Kendal non-parametric test can judge the trend.[28] Scatter plots of eigenvectors, reconstructed time series, and so forth can be used to detect the harmonic components.[26] Weighted correlation coefficient can ascertain the separability between the reconstructed

components.[29] Additionally, Alexandrov and Golyandina[30] proposed an automatic grouping method according to the periodogram or weighted correlation coefficients.

2.3. Support vector machines regression

Support vector machines regression is a machine learning method that performs structural risk minimization to improve the generalization capability of the created model with a real-valued mapping between multidimensional inputs and the output.[31,32] This mapping can be denoted as an underlying function and is nonlinear when a nonlinear kernel function $K(\mathbf{x}_i, \mathbf{x})$ is introduced, which transforms the input space into a high-dimensional feature space where the regression function is linear. The nonlinear SVR regression function[31] using ε-insensitive loss function[33] is given by

$$f(\mathbf{x}) = \sum_{SVs}(\bar{a}_i - \bar{a}_i^*)K(\mathbf{x}_i, \mathbf{x}) + \bar{b} \tag{5}$$

with the objective function in dual form represented as

$$\text{Minimize } L_d(\alpha, \alpha^*) = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)(\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)y_i + \sum_{i=1}^{l}(\alpha_i + \alpha_i^*)\varepsilon \tag{6}$$

subject to

$$0 \le \alpha_i, \alpha_i^* \le C \qquad i = 1, \cdots, l \tag{7}$$

$$\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0 \tag{8}$$

where α_i and α_i^* are the Lagrange multipliers in solving the objective function, $\bar{a}_1$, $\bar{a}_1^*$, and $\bar{a}$ represent optimality, C is a pre-specified value in order to modulate the balance of empirical and generalization errors. In Equation (5), the observed input vectors $\mathbf{x}_i$ are support vectors which have indeed contribution to the regression function, and correspond to positive α_i and α_i^*. In this paper, the popular Gaussian radial basis function is chosen as the kernel function, which can be denoted as $K(\mathbf{x}, \mathbf{y}) = \exp(-(x-y)^2/2\sigma^2)$, where σ is the parameter related to the global basis function width. The advantage of this kernel consists in the capability of handling infinite dimension space with less parameters.[34]

In the previous SVR process, there are three parameters, that is, ε, C, and σ, to be estimate. There are already some intelligent methods that can be used for parameter selection, such as genetic algorithms[35], particle swarm optimization[36], simulated annealing algorithms[37], and so forth. Here, we adopt a practical heuristic method to calculate the parameters directly from the training data.[38]

2.4. Other techniques

Because of several other related techniques are involved in this paper, we briefly introduce them in the following. The first is the noise test to a time series. This is a classical issue in time series analysis and various statistical approaches can be exploited to solve the problem, such as observing the charts of auto correlation function (ACF) and partial ACF, or conducting a hypothesis test.[39] This paper employs the generalized variance portmanteau test to evaluate the randomness of time series data, which is proposed by Esam and Ian[40] based on an asymptotic chi-square distribution. The second is the decomposition technique for time series. The main idea is smoothing or filtering by using wavelets,[41] empirical mode decomposition (EMD),[42] seasonal-trend decomposition procedure based on loess (STL),[43] and so forth. Here, we apply singular spectrum analysis as the decomposition approach for it is free of parameters and used widely. Another technique used in this paper is the search algorithm to search for the optimal parameters. Numerous studies have been done to build or improve the search strategy in the modeling process, such as gradient descent algorithm,[44] genetic algorithm,[45] SCE-UA algorithm,[46] and so forth. Among them, grid search is a practical and not too much time consuming method to identify good parameters, and can be easily parallelized.[47] In this paper, to achieve the goals of effectiveness and practicability, we apply the grid search algorithm to solve the optimization problem of the hybrid model.

3. The SSA-SVR based hybrid model

The framework of the proposed SSA-SVR method is shown in Figure 1. The core idea is that the two components of the trend and fluctuation decomposed from the original failure time series by SSA are individually processed by SSA and SVR. As shown in Figure 1, except for the preparation step, there are three steps to construct the hybrid model, that is, decomposition, modeling, and combination. The assistant techniques involving data partition, noise test, and parameter optimization for constructing the hybrid model are also illustrated and will be detailed in the following subsections. The related notations are listed in Table I.

3.1. Raw data decomposition with singular spectrum analysis

First, the original failure time series data (*TFS*) is divided into training set (*TR*) and test set (*TE*) as defined in Equations (9)–(13), where f_t is the number of failures observed during time period t, and f_{tr} and f_{te} refers to the f_t. The n, l, and m are the sample size of the three datasets, respectively.

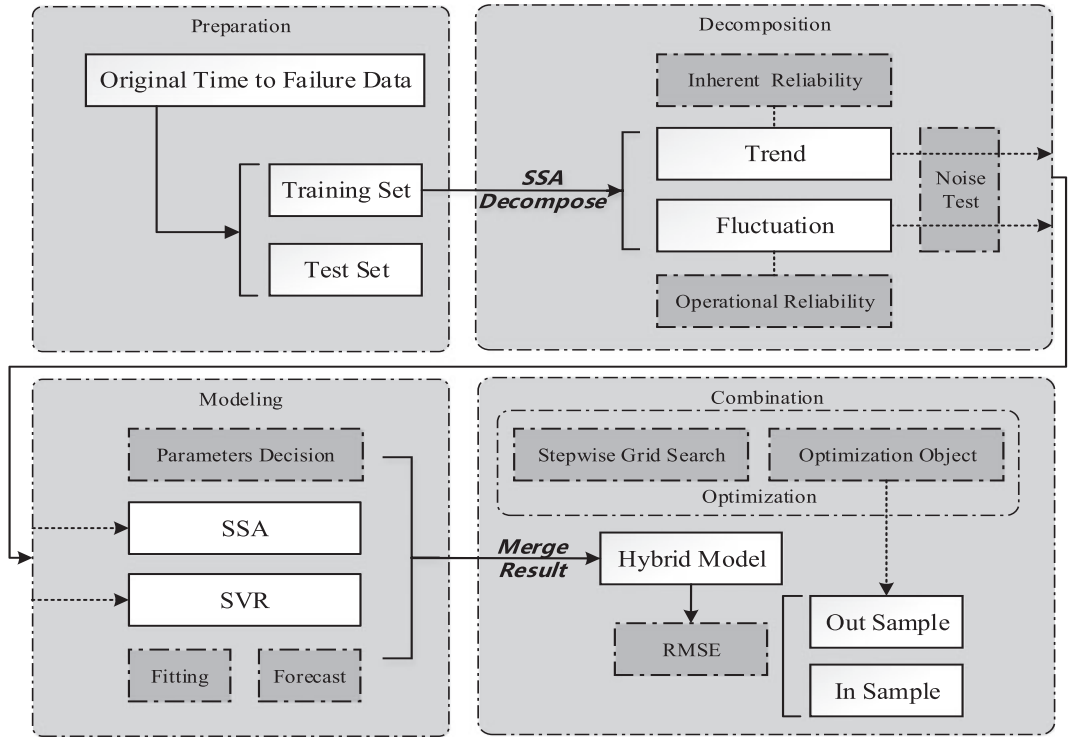

Figure 1. Framework of the proposed singular spectrum analysis (SSA)-support vector machines regression (SVR) hybrid model. RMSE, root mean square error

Table I. Notations

Notation	Explanation
F	Trajectory matrix of *TFS*
d	The number of non-zero eigenvalues in FF^T
g	The number of groups
RC_j	jth reconstructed time series
L	The length of slide window in the processed model
G	The number of groups in SSA reconstruction model
P	The conceivable period for failure time series
$\mathscr{C}_{ssa}$	SSA decomposition model
$\mathscr{R}_{ssa}$	SSA reconstruction model
$\mathscr{D}_{ssa}$	SSA model for extracting the two components
$\mathscr{M}_{ssa}$	Trend-SSA model
$\mathscr{F}_{ssa}$	Trend-SSA forecast model
$\mathscr{M}_{svr}$	Fluctuation-SVR model
$\mathscr{F}_{svr}$	Fluctuation-SVR forecast model
$\mathscr{M}_{com}$	SSA-SVR hybrid model
$\mathscr{F}_{com}$	SSA-SVR hybrid forecast model
$\mathscr{V}_{RMES}$	The function to calculate RMSE values
$\mathscr{O}_{trend-ssa}$	Optimization function for trend-SSA model
$\mathscr{O}_{fluctuation-svr}$	Optimization function for fluctuation-SVR model

$$TFS = \{f_{t_1}, f_{t_2}, \cdots, f_{t_n}\} \tag{9}$$

$$TR = \{f_{tr_1}, f_{tr_2}, \cdots, f_{tr_l}\} \tag{10}$$

$$TE = \{f_{te_1}, f_{te_2}, \cdots, f_{te_m}\} \tag{11}$$

Subject to:

$$TR \subset TFS; TE \subset TFS; TR \cap TE = \varnothing \tag{12}$$

$$l + m \leq n; l, m \neq 0;\ \forall t \in \mathbb{N} \tag{13}$$

Thus, *TR* can be considered as the input of the decomposition model denoted as $\mathscr{D}_{ssa}$, which includes two steps named as SSA decomposition ($\mathscr{C}_{ssa}$), as defined in Equation (14), and SSA reconstruction ($\mathscr{R}_{ssa}$), as defined in Equation (15), based on Section 2.2. Through the steps of $\mathscr{C}_{ssa}$ and $\mathscr{R}_{ssa}$, *TR* is transformed into the summation of the reconstructed sub-series (TR_{RC_j}) by filtering out other useless constituents (*e*). These sub-series have their own features and can be divided into several groups by applying the principles introduced in Section 2.2. However, in the reliability field, some domain knowledge can be considered for a reasonable grouping process. Intuitively, the system's reliability consists of two aspects: the inherent and operational. The inherent reliability usually performs as a stable trend and depicts a smooth curve when using the indicator of failure time series, whereas the operational reliability usually performs as an unstable fluctuation and depicts a waved curve. Hence, we can aim to seek these two components defined as the trend and fluctuation in the previous sub-series.

$$\mathscr{C}_{ssa}: TR \overset{Embedding}{\rightarrow} F = \begin{pmatrix} f_{tr_1} & \cdots & f_{tr_{l-L+1}} \\ \vdots & \ddots & \vdots \\ f_L & \cdots & f_{tr_l} \end{pmatrix} \overset{SVD}{\rightarrow} F = \sum_{i=1}^{d} E_i \tag{14}$$

$$\mathscr{R}_{ssa}: \overset{Grouping}{\rightarrow} F \approx \sum_{j=1}^{g} F_j \overset{Diagonal\ averaging}{\rightarrow} TR = \sum_{j=1}^{g} TR_{RC_j} + e \tag{15}$$

$$TR_{trend} = TR_{RC_1} \tag{16}$$

$$TR_{fluctuation} = \sum_{j=2}^{g} TR_{RC_j} \tag{17}$$

In the decomposition step, the trend component is extracted with a small window length, which is set as the conceivable period (*P*) of the original failure time series. In this situation, the first singular value of *SVD* usually has a large contribution and can be reconstructed as the trend component. After filtering the trend, the remainder reconstructed time series perform as certain regular waves and can be recognized as the fluctuation component. Equations (16) and (17) give the definitions to the trend and fluctuation, and the mechanism to extract the two components are shown in Figure 2.

3.2. Noise test

To evaluate the performance of SSA decomposition, the statistical technique of noise test is introduced before building the hybrid model. As discussed in Section 2.4, the generalized variance portmanteau test is adopted to conduct the task. The training set (*TR*) and the two extracted components (TR_{trend} and $TR_{fluctuation}$) are tested, separately. Though, to a certain extent, it is opinionated to declare that a time series cannot be modeled or forecasted by applying machine learning method when it is recognized as the white noise by the test. However, when a time series is not considered as the random sequence by the test (with low *p*-values), it is credible to regard that this time series is more suitable for modeling and forecast. So we can use this idea to evaluate the effect of SSA decomposing. The expected result is that *TR* is tested as the noise but TR_{trend} and $TR_{fluctuation}$ are nonrandom sequences by the tests. That is to say, the forecast would be conducted on TR_{trend} and $TR_{fluctuation}$ rather than on *TR*. In general, the trend component decomposed from SSA method has a growing or declining curve and will not be tested as the noise. The key point is the fluctuation. If this component is tested as the noise, it is necessary to adjust the grouping arrangement of SSA decompositions for a better separation.

3.3. Hybrid modeling

Next, we proceed with the steps of modeling and combination, as shown in the lower two parts of Figure 1. SSA is still applied to TR_{trend} for its superiority of dealing with the time-to-failure data according to the research by Rocco,[1] which related to the turbochargers failure data with a growing trend. To the component of $TR_{fluctuation}$, SVR is adopted as the modeling approach to control the generalization errors of the waved curve. The practice has proved that the fluctuation component plays a critical role to improve the forecast accuracy and SVR can well track it.

The hybrid model defined as $\mathscr{M}_{com}$ can be directly obtained by summing the two fitted models of TR_{trend} and $TR_{fluctuation}$ on account of the additive property of SSA decomposition, as well as the hybrid forecast model $\mathscr{F}_{com}$ by summing the two forecast models. The procedure to construct the hybrid model is summarized in Algorithm 1, where the parameters for the three models

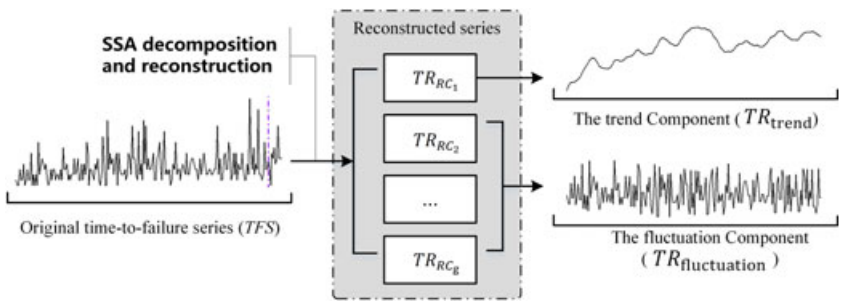

Figure 2. The decomposition mechanism for extracting the two components. SSA, singular spectrum analysis

$\mathscr{D}_{ssa}$, $\mathscr{M}_{ssa}$, and $\mathscr{M}_{svr}$ are listed as inputs of the hybrid model. S_d, S_t, and S_f are the sample sizes. L_d, L_t, and L_f are the window lengths. G_t is the groups of $\mathscr{M}_{ssa}$ and $\mathscr{F}_{ssa}.type \in \{recurrent, vector\}$ indicates the type of SSA forecast approach.

Input. *TFS*, *l*, *m*, $S_d, S_t, S_f, L_d, L_t, L_f, G_t, \mathscr{F}_{ssa}.type$
Output. SSA-SVR models and corresponding *RMSEs* with different lags
1. **get** *TR and TE by TFS, l, m //divide the original data into training set and test set*
2. **execute** $\mathscr{D}_{ssa}(TR, S_d, L_d)$ *//extract the two components from the training set*
3. $s_1 \leftarrow \mathscr{C}_{ssa}(TR, S_d, L_d)$ *//build the SSA decomposition model*
4. $r_1 \leftarrow \mathscr{R}_{ssa}(s_1, list(1))$ *//build the SSA reconstruction model*
5. $TR_{trend} \leftarrow r_1\$F1$ *//define the trend component*
6. $TR_{fluctuation} \leftarrow TR - TR_{trend}$ *//define the fluctuation component*
7. **noise test** $TR, TR_{trend}, TR_{fluctuation}$ *//inspect the three time series by noise test*
8. **execute** $\mathscr{M}_{ssa}(TR_{trend}, S_t, L_t)$ *//build trend-SSA model*
9. **execute** $\mathscr{F}_{ssa}(\mathscr{M}_{ssa}, G_t, \mathscr{F}_{ssa}.type, m)$ *//build trend-SSA forecast model*
10. **execute** $\mathscr{M}_{svr}(TR_{fluctuation}, S_f, L_f)$ *//build fluctuation-SVR model*
11. **execute** $\mathscr{F}_{svr}(\mathscr{M}_{svr}, m)$ *//build fluctuation-SVR forecast model*
12. $\mathscr{M}_{com} \leftarrow \mathscr{M}_{ssa} + \mathscr{M}_{svr}$ *//build the hybrid model*
13. $\mathscr{F}_{com} \leftarrow \mathscr{F}_{ssa} + \mathscr{F}_{svr}$ *//build the hybrid forecast model*
14. **execute** $\mathscr{V}_{RMES}(\mathscr{M}_{com}, \mathscr{F}_{com}.output, TE)$ *//calculate the RMSE values*
15. **return** $\mathscr{M}_{com}, \mathscr{F}_{com}, \mathscr{V}_{RMES}$ *//return the outputs*

Algorithm 1. Generate SSA-SVR hybrid model

3.4. Parameter optimization

In the SSA-SVR modeling process, except for the parameter of sample size, there are parameters to be fitted for all the three models, namely, decomposition-SSA model ($\mathscr{D}_{ssa}$), trend-SSA model ($\mathscr{M}_{ssa}$, $\mathscr{F}_{ssa}$), and fluctuation-SVR model ($\mathscr{M}_{svr}$, $\mathscr{F}_{svr}$). However, to separate the trend and fluctuation reasonably, $\mathscr{D}_{ssa}$ is relatively constant with the unique model parameters. Thus, we only need to seek the optimal model parameters of the trend and fluctuation in terms of RMSEs to achieve the best performance on the test data. Considering the independence of the two components and numerous parameter combinations of the corresponding models, and less time consumption, the optimization procedure is implemented individually by stepwise grid search method, which includes five search variables: S_t, S_f, L_t, L_f, and G_t. Equation (18) displays the object function and the constraints are shown in Equations (19)–(22).Object function:

$$\text{Minimize } RMSE = \sqrt{\frac{1}{m}\sum_{te=1}^{m}(y_{te} - f_{te})^2}_{\forall f_{te} \in TE, \forall y_{te} \in \mathscr{F}_{com}.output} \tag{18}$$

Subject to:

$$S_t, S_f \le S_d \le l; \forall S_t, S_f \in \mathbb{N} \tag{19}$$

$$2 \le L_t, L_f \le \frac{S_t}{2}, \frac{S_f}{2};\ \forall L_t, L_f \in \mathbb{N} \tag{20}$$

$$1 \le G_t \le L_t;\ \forall G_t \in \mathbb{N} \tag{21}$$

$$\mathscr{F}_{ssa}.type = vector \tag{22}$$

where y_{te} is the prediction value of the hybrid forecast model $\mathscr{F}_{com}$ at time *te* and f_{te} is the corresponding observation. We set the length of sliding windows L_t and L_f to be not smaller than 2 and not larger than half of the sample sizes S_t and S_f. The maximum of G_t cannot be larger than the number of singular values which equals to L_t. The procedure to optimize the hybrid model is mainly divided into two modules, namely, $\mathscr{O}_{trend-ssa}$ and $\mathscr{O}_{fluctuation-svr}$, which respectively correspond to the tasks of multilayer grid search for trend optimization and monolayer grid search for fluctuation optimization. The optimization procedure in detail is presented in Algorithm 2, where the symbol with superscript * denotes a constant or optimized value and implies the output of the grid search.

Input. *TFS*, *l*, *m*, $S_d, S_t, S_f, L_d, L_t, L_f, G_t, \mathscr{F}_{ssa}.type$
Output. Optimal SSA-SVR models and corresponding *RMSEs* with different lags
1. **execute** $\mathscr{O}_{trend-ssa}$ $(TFS, S_d^*, S_f^*, L_f^*, step_t)$ *//optimize trend-SSA model*
2. **for** all S_t : **seq**$(P, S_d^*, step_t)$ *//traverse the value range for sample size* (S_t)
//traverse the value range for window length (L_t)
3. **for** all L_t : **seq**(**min**$(P, S_t/2)$, $S_t/2$, $step_t$)
//traverse the value range for groups (G_t)
4. **for** all G_t : **seq**(**min**(P, L_t), **min**$(L_t, 50)$, $step_t$)
//build the hybrid model with the specified trend-SSA parameters
5. **execute** $\mathscr{M}_t \leftarrow \mathscr{M}_{com}(TFS, S_d^*, S_f^*, L_f^*, S_t, L_t, G_t)$
//build the hybrid forecast model with the specified trend-SSA parameters
6. **execute** $\mathscr{F}_t \leftarrow \mathscr{F}_{com}(\mathscr{M}_t, \mathscr{F}_{ssa}.type, m)$

//calculate the RMSE values with the specified trend-SSA parameters
7. **execute** $v_t \leftarrow \mathscr{V}_{RMES}(\mathscr{M}_t, \mathscr{F}_t.output, TE)$
//save the models and RMSE values with the specified trend-SSA parameters
8. **save** $\mathscr{M}_t, \mathscr{F}_t, v_t$
9. **end(for); end(for); end(for)** *//end the loops*
//return the sorted results based on the RMSE values for the trend component
10. **return** $\mathscr{M}_t, \mathscr{F}_t, v_t$ **rank by** v_t
11. **execute** $\mathscr{O}_{fluctuation-svr}$ $(TFS, S_d^*, S_t^*, L_t^*, G_t^*, step_f)$ *//optimize fluctuation-SVR model*
12. **for** all S_f : **seq**$(P, S_d^*, step_f)$ *//traverse the value range for sample size* (S_f)
//traverse the value range for window length (L_f)
13. **for** all L_f: **seq**(**min**$(P, S_f/2), S_f/2, step_f)$
//build the hybrid model with the specified fluctuation-SVR parameters
14. **execute** $\mathscr{M}_f \leftarrow \mathscr{M}_{com}(TFS, S_d^*, S_t^*, L_t^*, G_t^*, S_f, L_f)$
//build the hybrid forecast model with the specified fluctuation-SVR parameters
15. **execute** $\mathscr{F}_f \leftarrow \mathscr{F}_{com}(\mathscr{M}_f, m)$
//calculate the RMSE values with the specified fluctuation-SVR parameters
16. **execute** $v_f \leftarrow \mathscr{V}_{RMES}(\mathscr{M}_f, \mathscr{F}_f.output, TE)$
// save the models and RMSE values
SVR parameters
17. **save** $\mathscr{M}_f, \mathscr{F}_f, v_f$
18. **end(for); end(for)** *//end the loops*
//return the sorted results based on the RMSE values for the fluctuation component
19. **return** $\mathscr{M}_f, \mathscr{F}_f, v_f$ **rank by** v_f
//calculate the outputs by using the ranked parameters
20. **calculate** $\mathscr{M}^*_{com}, \mathscr{F}^*_{com}, v^*_{com}$ by $S_t^*, L_t^*, G_t^*, S_f^*, L_f^*$
//return the sorted outputs based on the RMSE values
21. **return** $\mathscr{M}^*_{com}, \mathscr{F}^*_{com}, \boldsymbol{v}^*_{com}$ **rank by** v^*_{com}

Algorithm 2. Stepwise grid search for the hybrid SSA-SVR model

4. Experiments

This section provides the implementations of the SSA-SVR hybrid model and the parameter optimization proposed in Section 3 on a real industrial case. Besides, the experimental preparations and discussions are illustrated.

4.1. Preparations

We first introduce the experiment setup as follows, that is, the experimental data, comparison models, performance metrics, and some support tools and environments.

4.1.1. Experimental data. The raw experimental data was gathered from the failure record library of an airline's operating Boeing 737 aircraft (Aircraft A). After collecting, sorting, and checking the raw data, we obtained five kinds of monthly time series data: Fight Hours (FH), Fight Cycles (FC), Failures (FA), Failure Rate (FR), and Mean Time between Failure (MTBF), referring to 216 data points, during 1997–2014, a total of 18 years. Figure 3 is the scatter diagrams with regression fitting lines between the five time series and the diagonals present density distributions for each time series. We can see positive correlations in pairs of FH & FC and FA & FR, and a negative correlation in pairs of FR & MTBF. But there is no significant correlation between FH & FA. According to Section 2.1, FR and MTBF can be calculated through FH and FA, whereas FH can be usually acquired by flight plans as well as FC. So we choose the key time series FA as the experimental data. The 12 data points in 2014 are used as the test set.

To reduce the adverse effects of the abnormal points in the raw data, we filter the FA time series based on a predefined threshold. When the data point is larger (or smaller) than the predefined upper (or lower) value, it is replaced by the threshold. Considering the features of FA time series, we set the upper and lower thresholds as 2 and 0.1 times of the average value, respectively. After the preprocessing, the final FA time series (Aircraft A) is illustrated in Figure 4, in which 17 upper outliers are eliminated. The reason to apply this preprocessing method is that the number of outliers is relatively small (only one upper outlier in the test set), and the threshold points can also demonstrate that the fitted or prediction values should be at a high or low level and can be used to make the suitable maintenance strategies.

4.1.2. Comparison models. To validate the performance of the proposed SSA-SVR model, the comparison experiments are conducted with four typical forecast techniques. Here, we provide a brief description to each technique.
4.1.2.1. Holt-Winters. Holt-Winters is a kind of exponential smoothing method for time series modeling and can be categorized into two types: additive and multiplicative.[48] Additive Holt-Winters model can be written as Equation (23), where α, β, and γ are smoothing coefficients in the range of 0 to 1 (more closer to 1 implies more recent observations relied), and can be determined according to the minimum mean square error (MSE). L_t, b_t, and S_t respectively correspond to the component of level, trend slope, and seasonality at

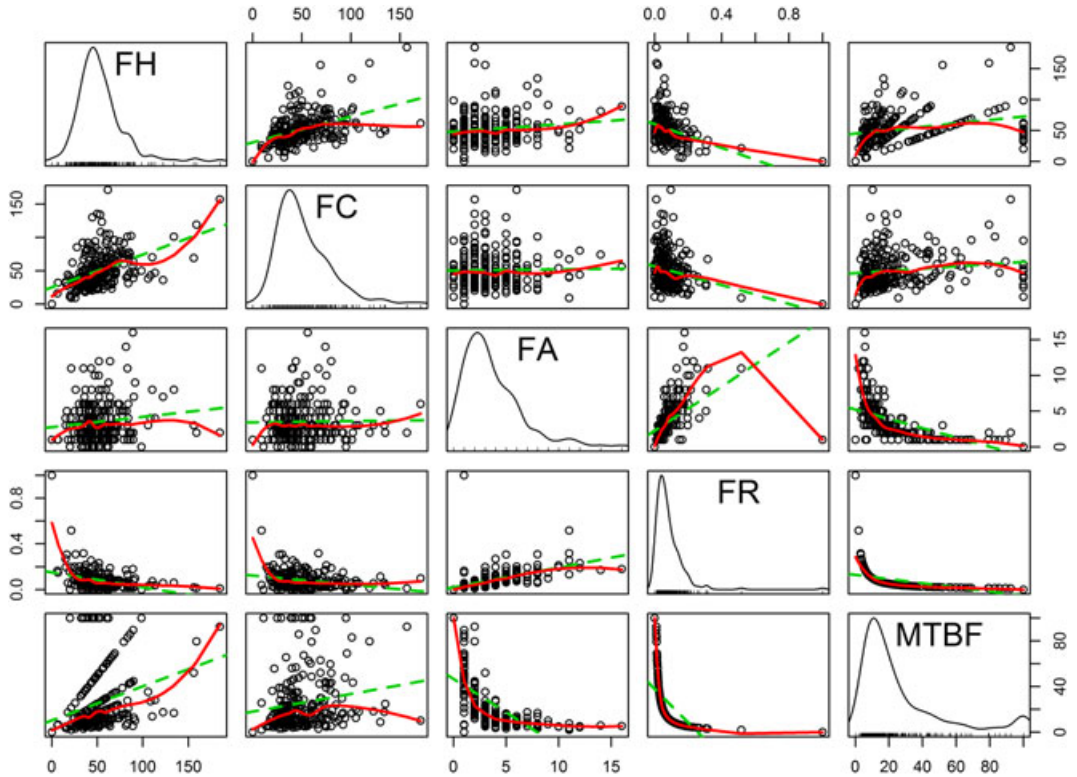

Figure 3. The scatter diagrams with the analysis of regression and density distribution on five time series (Aircraft A). FH, Fight Hours; FC, Fight Cycles; FA, Failures; FR, Failure Rate; MTBF, Mean Time between Failure

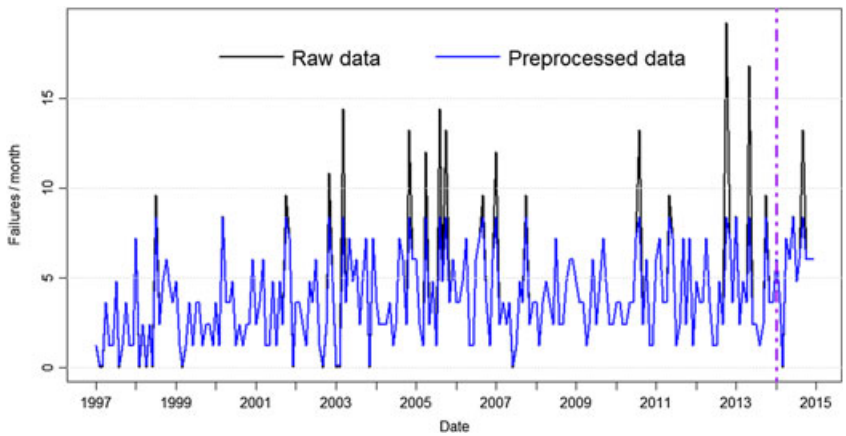

Figure 4. Failures time series after preprocessing (Aircraft A)

time t, which require to calculate the initial values in the actual use. The subscript s is the length of seasonal cycle. k is the step length of prediction, and F_{t+k} is the prediction value of step k at time t.

$$\begin{cases} L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \\ b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \\ S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s} \\ F_{t+k} = L_t + kb_t + S_{t+k-s} \end{cases} \tag{23}$$

4.1.2.2. Autoregressive integrated moving average. Autoregressive integrated moving average is a classical method for time series analysis proposed by Box and Jenkins in 1970s.[49] ARIMA model can be denoted as ARIMA(p, d, q) with orders p, d, and q. When d = 0, the model can be written as Equation (24), where y_{t-i} is the actual value at time $t-i$, ε_{t-j} is the random error at time $t-j$, ϕ_i and θ_j are the coefficients which can be determined by the estimation method.[50] The order d represents the times of differencing for

transforming the time series into a stationary one. When the time series includes seasonal components, the model can be denoted as SARIMA(p, d, q)(P, D, Q)$^{[s]}$ where s is the length of seasonal cycle. A detailed description of SARIMA model is provided by Jeong *et al*.[51] In the ARIMA modeling process, the orders p, d, and q can be determined by observing the charts of ACF and partial ACF, or calculating the values of Akaike information criterion (AIC) and Bayesian information criterion (BIC). In addition, Hyndman and Khandakar[52] presented an automatic approach for building the ARIMA model.

$$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{24}$$

4.1.2.3. Multiple linear regression. Multiple linear regression is a practical method and is widely used in forecast tasks.[53] MLR model exploited for time series forecast can be represented as Equation (25), where Y_t is the prediction value at time t, $a_0, a_1, \cdots, a_k$ are the regression coefficients which can be commonly estimated by least square method, and e denotes the error.[54] When the continuous lags are selected as the factor variables, k also corresponds to the window length.

$$Y_t = a_0 + a_1 Y_{t-1} + \cdots + a_k Y_{t-k} + e \tag{25}$$

4.1.2.4. Group method of data handling. Group method of data handling is one kind of artificial neural network with the capability of self-organizing and was first introduced by Ivakhnenko.[55] This network starts with only inputs and continually adds the selected neurons to the hidden layers in the training process. The general GMDH adopts quadratic polynomial as the transfer function, which is in the form of Equation (26), where y is the output, x_i and x_j are the two input variables, and $a_0, a_1, \cdots, a_5$ are the coefficients.

$$y = a_0 + a_1 x_i + a_2 x_j + a_3 x_i x_j + a_4 x_i^2 + a_5 x_j^2 \tag{26}$$

4.1.3. Performance metrics. The performance of various forecast models is mainly measured by root MSE (RMSE) defined by Equation (27), where y_t is the actual value, f_t is the calculated value, and T is the number of data points. RMSE is more sensitive to large errors than other metrics, such as mean absolute error and mean absolute percentage error.[56] In this paper, the RMSE values are calculated to evaluate the fitting and forecasting accuracy separately.

$$\mathrm{RMSE} = \sqrt{\sum_{t=1}^{T} (y_t - f_t)^2 / T} \tag{27}$$

4.1.4. R & packages. In this paper, R is applied as the development tool to model and forecast the failure time series by using several R packages, such as *stats*, *forecast*, *Rssa*, *rminer*, *GMDH*, and so forth. *Rssa* is developed by Golyandina *et al*.[57] and includes a collection of methods for SSA, like *ssa*(), *reconstruct*(), and so forth. But one thing to note here is that the *ssa*() function has an upper limit of 50 singular values. In addition, *rminer* is a data mining package developed by Cortez[58] and is particularly suitable for neural networks and support vector machines.

4.2. Application of the proposed model to forecast FA time series

After the preparations, the preprocessed FA time series is used to set up the SSA-SVR hybrid model proposed in Section 3.3. In the hybrid modeling, the components of trend and fluctuation extracted from the FA time series are separately learnt by SSA and SVR. The performance of the hybrid model is evaluated by comparing with other forecast models.

4.2.1. Decomposition and component selection. The decomposition of FA time series aims to obtain the different components, which usually imply some domain knowledge in reliability. First, we can analyze the trend. The trend of FA time series can reflect the stable level and shape of the original failure data, and corresponds to the inherent reliability of the aircraft. In order to extract the trend component, the window length L of SSA decomposition model is set to a value of 12 (potential period for monthly data) according to the principle described in Section 2.2. The sample size is set to the maximum of the value range for adequate data points.

The results of SVD approach to FA time series (Aircraft A) are presented in Figure 5. It can be seen that obviously the first singular value has a large contribution (74.85%) and the other singular values are overall at a lower level. Figure 6 displays the reconstructed time series of the twelve singular values, in which only the first one depicts a smooth and growing shape. Furthermore, the weighted correlation coefficient matrix (shown in Figure 7) presents a clear separate boundary between the first and the other sub-series. Thus, we confirm that the first reconstructed time series can be extracted to represent the trend component. For the other 11 sub-series, they all contain certain frequencies and perform as oscillatory waves around the zero axial line. Besides, it is difficult to detect any remarkable period. Figure 7 also demonstrates that there are mixed correlations among these time series and should not be partitioned off arbitrarily. Therefore, we can aggregate them into one group and define it as the fluctuation component. To some degree, the fluctuation component implies more meanings and is more informative than the trend component although it accounts for a small contribution (a total of 25.15%). The reason is that the fluctuation component can reflect the impact of external factors on the failure time series, and corresponds to the operational reliability of the aircraft. By comparison, analysis to the operational reliability is more valuable in guiding the real field activities such as making maintenance scheduling.

Now, two components defined as the trend and fluctuation are extracted by grouping the reconstructed sub-series decomposed by SVD approach (as shown in Figure 8). Next, it is necessary to consider the noise for each component. The traditional statistical

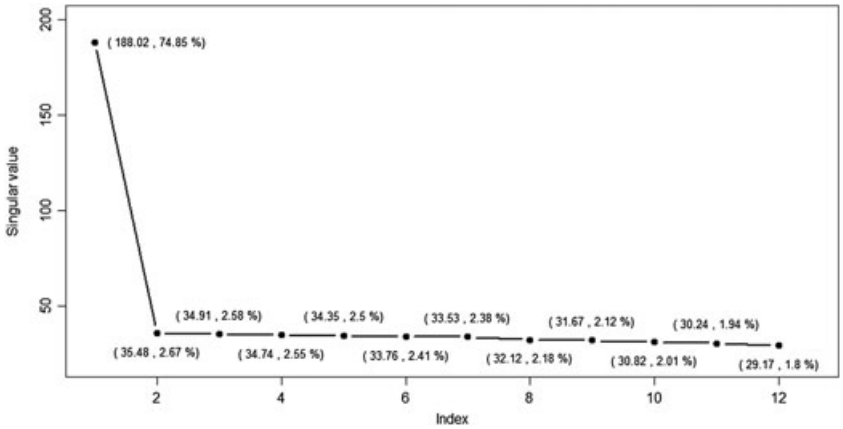

Figure 5. Twelve singular values of Failures time series by singular value decomposition approach (Aircraft A)

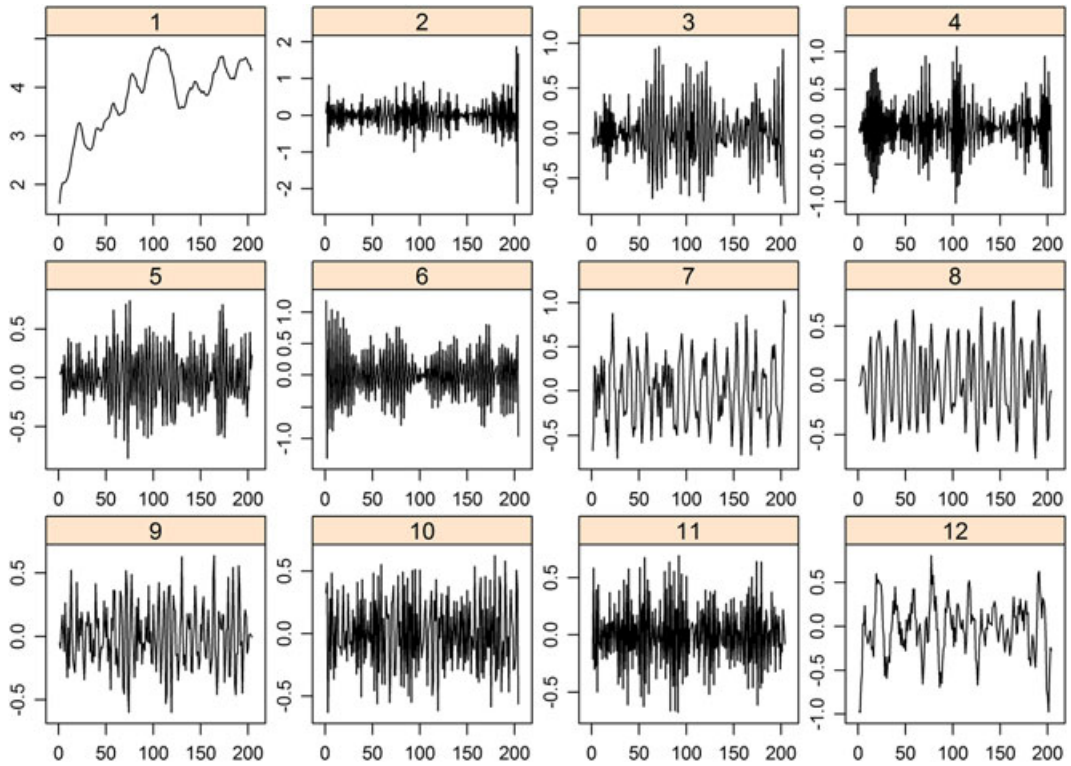

Figure 6. Reconstructed time series corresponding to the twelve singular values (Aircraft A)

hypothesis tests can be adopted to judge whether a time series can be considered as noise. Here, we apply the generalized variance portmanteau test. Figure 9 displays the test results with *p*-values on different lags. It is of interest that the original FA time series (Aircraft A) is deemed to be random on account of the higher *p*-values (close to 1) while the two decomposed components are greatly not random on account of the lower *p*-values (less than 0.05). Although we cannot deny each component still includes the noise, the decomposition to FA time series has vastly improved the adaptiveness of statistical modeling.

4.2.2. Modeling and combination. According to the hybrid modeling defined in Section 3.3, the trend and fluctuation components are separately processed by SSA and SVR. First, we construct the trend-SSA model. The sample size S_t is set to the maximum of the value range (i.e., 204) for learning more information from the historical data. Considering the principles mentioned in Section 2.2, the window length L_t is selected as the multiples of 12, and it should not be larger than 102 (half of the entire training set with the sample size of 204). The groups G_t should be set to the maximum of the value range (1 to 50) because the trend component has already been a pure reconstructed sub-series from SSA decomposition. SSA vector forecast approach is selected for a more stable performance.

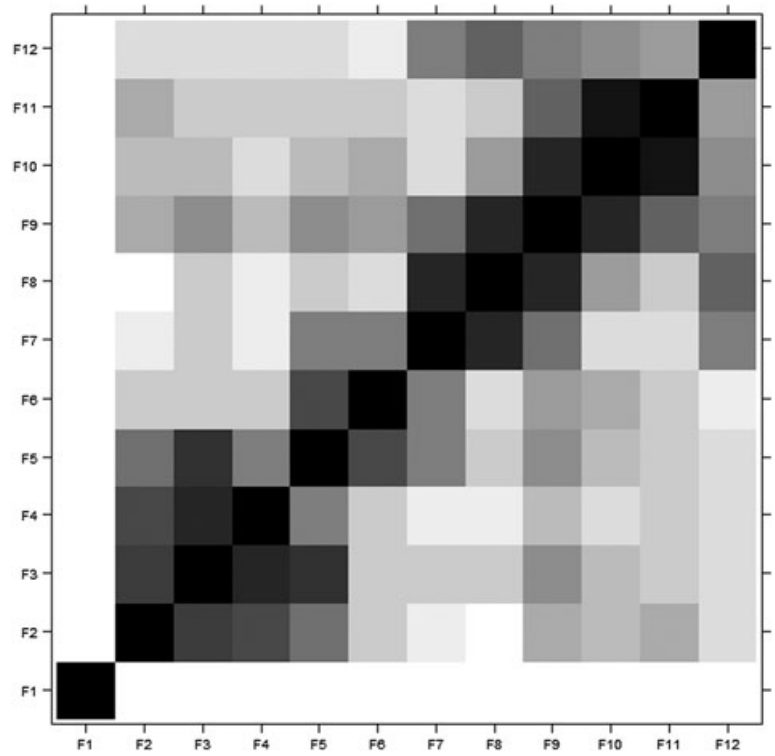

Figure 7. Weighted correlation coefficient matrix of the twelve reconstructed time series (Aircraft A)

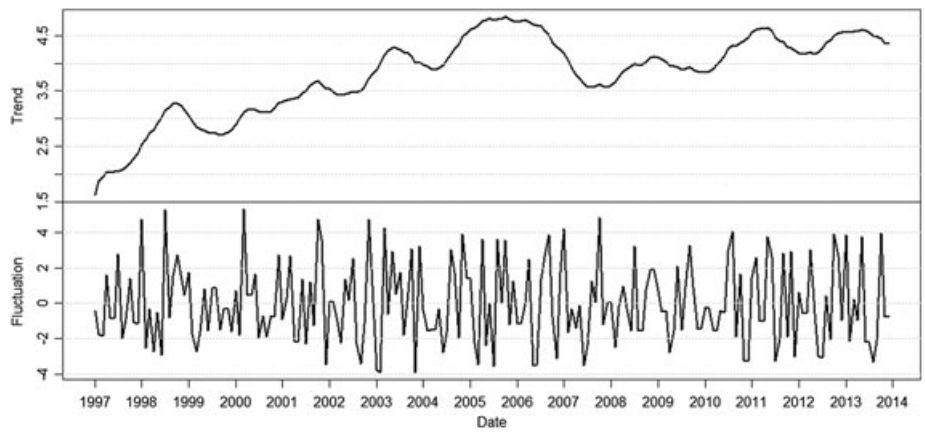

Figure 8. The two selected components: trend and fluctuation (Aircraft A)

Figure 10 presents the predictions (Aircraft A) with different combinations of L_t and G_t. As shown in the figure, a smaller L_t receives an increasing curve which is more consistent with the test set and has a smaller RMSE value. When L_t is larger than 50 (the maximum of G_t), the prediction curve presents a decreasing trend and has a worse accuracy. This is due to the lower contribution generated from the partial grouping.

Secondly, we construct the fluctuation-SVR model. The sample size S_f is also set to 204 for the similar consideration with the trend. Based on Section 2.3, the Gaussian kernel's parameter σ and the error bound ε are calculated using the heuristics. The regularization parameter C is set to 3 for a standardized output. Another parameter to be set is the window length (L_f). There are numerous situations for the selection of L_f. Here, we just consider the continuous situations such as $\{1, 2, \cdots, 12\}$ to coincide with the features of FA time series. A reasonable selection to the upper limit of L_f is 12 or its multiples for the monthly data. Figure 11 presents the forecast performance with different upper limits of L_f. It is obvious that the index of 24 receives the best result.

At last, we establish the hybrid model by summing the trend-SSA and fluctuation-SVR models directly. Table II demonstrates the comparison results of model performance between the proposed SSA-SVR model and other representative models in terms of RMSE metric. It can be seen that the proposed SSA-SVR model reveals a rising improvement in both fitting and forecasting accuracy,

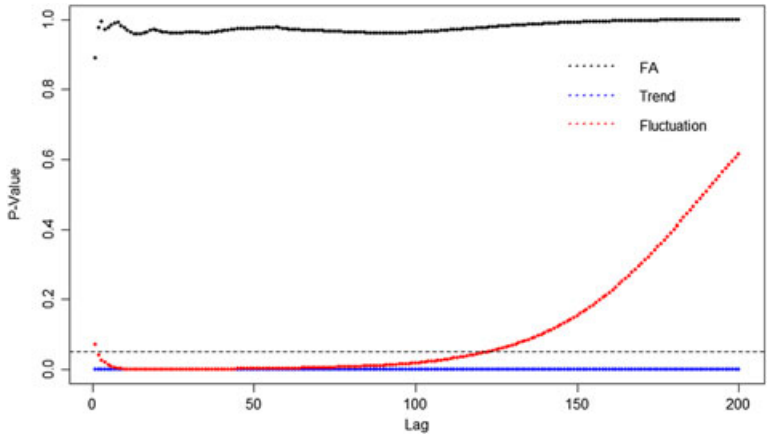

Figure 9. Generalized variance portmanteau tests for three time series: Failures (FA), trend, and fluctuation (Aircraft A)

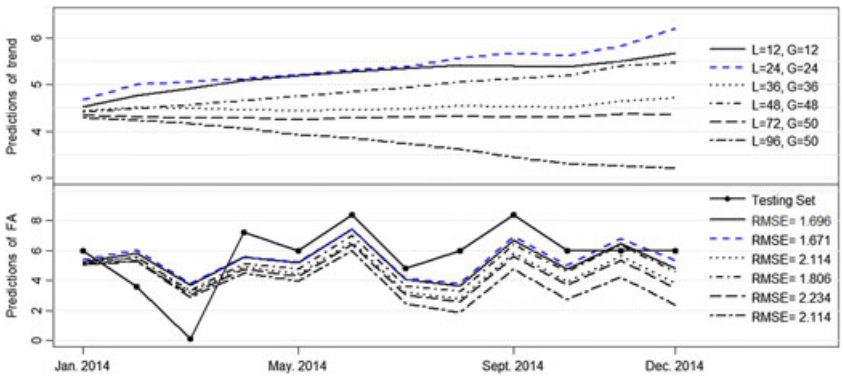

Figure 10. Trend-singular spectrum analysis predictions with different windows and groups: singular spectrum analysis forecast approach is selected as '*vector*'.The Fluctuation-support vector machines regression model is fitted with $Lf = 24$. (Aircraft A). FA, Failures

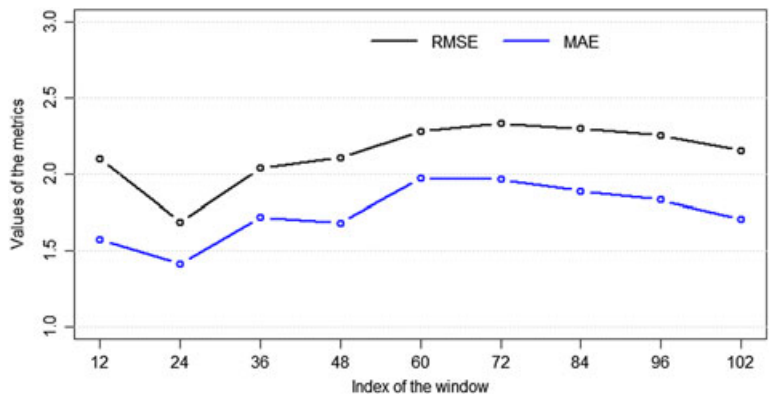

Figure 11. Fluctuation-support vector machines regression forecast performances with different windows: The trend-singular spectrum analysis model is fitted with $Lt = Gt = 24$ and '*vector*' forecast approach. (Aircraft A). RMSE, root mean square error; MAE, mean absolute error

especially at 6 and 12 forecast points, and the related computational cost is at a low level. It also implies that a larger sample size achieves better performance. Furthermore, in order to validate the proposed model, we also implemented various kinds of other hybrid models, which applied different unitary models in the trend and fluctuation. The comparison results are shown in Table III. Table IV presents the major parameters for each model. We can see that the proposed SSA-SVR model overall has a satisfactory prediction accuracy with a small computational cost, and performs best at 12 forecast points.

Moreover, in order to validate the generalization of the proposed method, we carried out the further experiments on another aircraft (Aircraft B) in the same airline company. The failure time series gathered from Aircraft B is similar to that of Aircraft A but has the variation on curve shape (as shown in Figure 12). The experimental procedure is the same with Aircraft A and the results are presented in Tables V, VI, and VII. It is shown that the proposed method also performs best at 12 forecast points with a small computational cost.

Table II. Comparisons between the proposed model and 8 time series models at various forecast points (Aircraft A)

Sample size	Parameters			In-RMSE	Out-RMSE					Time cost(s)
					1	2	3	6	12	
Holt-Winters[a]	α	β	γ							
180[a]	0.007	0.435	0.223	2.606	0.118	0.287	2.591	2.775	2.416	0.06
204[a]	0.044	0.073	0.223	2.617	0.088	0.069	2.278	2.883	2.595	0.06
ARIMA	*p*	*d*	*q*							
180	2	1	3	2.317	1.444	1.211	2.710	2.844	2.529	0.50
204	0	1	1	2.364	1.412	1.224	2.792	2.759	2.342	0.36
MLR	*L*									
180	24			2.240	2.511	1.786	2.640	2.736	2.436	0.11
204	24			2.221	2.490	1.773	2.602	2.617	2.381	0.11
GMDH	*L*	*Layers*								
180	24	3		1.956	2.056	1.572	2.136	2.584	2.664	10.37
204	24	3		1.939	2.704	1.949	2.188	2.381	2.710	10.40
SSA[v]	*L*	*G*								
180[v]	84	list(1:50)		0.660	2.311	1.741	2.280	2.888	2.872	0.33
204[v]	96	list(1:50)		0.770	2.509	1.843	2.175	2.500	2.295	0.36
SVR	*L*,*C*	ε	σ							
180	24,3	0.273	0.021	1.268	1.969	1.672	2.056	2.238	2.320	0.14
204	24,3	0.259	0.022	1.255	1.735	1.514	1.944	2.146	2.336	0.14
SSA[D]-SVR[20]	SSA[D]	SVR								
180	*L* = 12, *G* = list(1)	*L* = 24		0.036	0.778	1.567	2.369	2.351	2.345	0.48
204	*L* = 12, *G* = list(1)	*L* = 24		0.028	0.907	1.489	2.114	2.042	2.195	0.53
SSA[D]-SVR[22]	SSA[D]	SVR								
180	*L* = 12, *G* = list(1)	*L* = 24		0.814	0.259	2.215	2.526	2.297	2.418	0.48
204	*L* = 12, *G* = list(1)	*L* = 24		0.770	0.326	1.818	2.025	1.832	2.280	0.52
SSA[D]-SSA[v]-SVR	SSA[D]	SSA[v]	SVR							
180	*L* = 12, *G* = list(1)	*L* = 24, *G* = list(1:24)	*L* = 24	0.919	0.434	1.973	2.984	2.316	1.805	0.36
204	*L* = 12, *G* = list(1)	*L* = 24, *G* = list(1:24)	*L* = 24	0.910	0.558	1.720	2.582	2.005	1.666	0.39

[a]indicates Holt-Winters '*additive*' model.
[v]indicates SSA '*vector*' forecast approach.
[D]indicates SSA model for extracting the two components.
The best RMSE values at various forecast points are marked with the dark background.
Configurations of the computer to conduct the experiments are as follows: CPU: Intel(R) Core(TM)2 Duo CPU E8400 @3.00GHz 3.00GHz; RAM: 4.00GB (3.00GB available); OS: Windows 10 Professional 32bit.
RMSE, root mean square error; ARIMA, autoregressive integrated moving average; MLR, multiple linear regression; GMDH, group method of data handling; SSA, singular spectrum analysis; SVR, support vector machines regression.

4.3. Optimizing the model parameters using grid search method

In the previous experiments, the parameters of SSA-SVR model are determined empirically and receive good performance. Now the grid search method mentioned in Section 3.4 is applied to optimize parameters systematically. Here, we just take Aircraft A as the example.

First, we conduct the grid search with the fluctuation-SVR model. Here the parameters of the trend-SSA model are set to the same values presented in Table IV. Except the heuristic parameters (ε, σ) and regularization parameter (C) in the SVR model, two parameters of samples size (S_f) and window length (L_f) are required to determine and are used to construct the search space. This is a two-dimensional parameter space, and we can obtain the optimal parameters based on a monolayer grid search algorithm. The search object is the RMSE value on 12 prediction points corresponding to the test set. Figure 13 illustrates the search results. We can see that good parameters mainly fall in two regions where L_f is at a low lever (less than 36) and S_f is close to the maximum 204 as well as 72.

Secondly, we conduct the grid search with the trend-SSA model by using the parameters $S_f = 204$ and $L_f = 24$ for the fluctuation-SVR model. Here, the parameters S_t, L_t, and G_t constitute a three-dimensional search space, and we can process it by a multilayer grid

Table III. Comparisons of different hybrid models at various forecast points (Aircraft A)

Trend	Fluctuation	In-RMSE	Out-RMSE					Time Cost(s)
			1	2	3	6	12	
Holt-Winters[a]	Holt-Winters[a]	2.398	0.250	0.195	2.163	2.971	2.885	0.40
	ARIMA	1.957	0.272	1.239	2.873	2.946	2.693	0.62
	MLR	1.757	0.704	1.252	2.719	2.577	2.588	0.28
	GMDH	1.814	2.482	1.887	2.522	2.832	2.818	10.83
	SSA[v]	0.620	2.739	2.003	2.460	2.596	2.517	0.32
	SVR	0.940	0.866	1.431	2.109	2.070	2.375	0.33
ARIMA	Holt-Winters[a]	2.392	0.275	0.235	2.115	3.002	2.822	2.48
	ARIMA	1.979	0.297	1.192	2.811	2.961	2.618	2.92
	MLR	1.762	0.728	1.213	2.659	2.586	2.509	2.84
	GMDH	1.816	2.507	1.885	2.481	2.861	2.726	13.28
	SSA[v]	0.652	2.764	2.034	2.438	2.633	2.457	2.75
	SVR	0.945	0.890	1.393	2.049	2.084	2.287	2.94
MLR	Holt-Winters[a]	2.382	0.263	0.205	2.181	2.919	2.703	0.28
	ARIMA	1.860	0.285	1.238	2.889	2.900	2.513	0.72
	MLR	1.762	0.717	1.253	2.735	2.539	2.402	0.40
	GMDH	1.819	2.495	1.895	2.540	2.783	2.603	11.28
	SSA[v]	0.607	2.752	2.013	2.478	2.549	2.342	0.42
	SVR	0.966	0.893	1.458	2.131	2.029	2.165	0.49
GMDH	Holt-Winters[a]	2.387	0.275	0.224	2.147	2.951	2.771	11.06
	ARIMA	1.859	0.297	1.213	2.848	2.921	2.573	11.52
	MLR	1.761	0.729	1.232	2.696	2.554	2.462	11.24
	GMDH	1.821	2.507	1.893	2.511	2.815	2.681	21.83
	SSA[v]	0.609	2.764	2.028	2.460	2.582	2.406	11.25
	SVR	0.963	0.901	1.432	2.090	2.049	2.238	11.30
SSA[v]	Holt-Winters[a]	2.391	0.042	0.340	2.566	2.701	2.211	0.18
	ARIMA	1.979	0.020	1.645	3.362	2.828	2.160	0.66
	MLR	1.761	0.412	1.596	3.191	2.530	2.048	0.35
	GMDH	1.813	2.190	1.908	2.853	2.578	2.025	11.48
	SSA[v]	0.651	2.447	1.733	2.641	2.280	1.897	0.35
	SVR	1.020	0.627	1.861	2.620	2.029	1.685	0.39
SVR	Holt-Winters[a]	2.388	0.293	0.236	2.157	2.933	2.725	0.33
	ARIMA	1.866	0.315	1.214	2.859	2.899	2.515	0.78
	MLR	1.767	0.749	1.233	2.702	2.545	2.432	0.47
	GMDH	1.818	2.532	1.906	2.515	2.816	2.652	11.53
	SSA[v]	0.607	2.797	2.057	2.462	2.596	2.392	0.44
	SVR	1.035	0.969	1.523	2.129	2.056	2.215	0.52

RMSE, root mean square error; ARIMA, autoregressive integrated moving average; MLR, multiple linear regression; GMDH, group method of data handling; SSA, singular spectrum analysis; SVR, support vector machines regression.

Table IV. Major parameters of the models in hybridizations (Aircraft A)

Model	Trend parameters	Fluctuation parameters
Holt-Winters[a](α, β, γ)	(0.955,0.549,1)	(0.004,0.736,0.23)
ARIMA(p, d, q)(P, D, Q)[12]	(2,1,0)(2,0,2)[12]	(2,0,1)
MLR(L)	24	24
GMDH(L, *Layers*)	(24,3)	(24,3)
SSA[v](L, G)	(24,list(1:24))	(96,list(1:50))
SVR(L, C, ε, σ)	(24,3,0.015,0.091)	(24,3,0.251,0.024)

The sample size of each model is set to 204.
ARIMA, autoregressive integrated moving average; MLR, multiple linear regression; GMDH, group method of data handling; SSA, singular spectrum analysis; SVR, support vector machines regression.

search algorithm in terms of the RMSE value on the 12 test points. Figure 14 provides the details of search results, which correspond to the four grouping levels ($G_t = \{12, 24, 36, 48\}$), respectively. From the figure, we can see that the smaller RMSE values are roughly distributed on the diagonal of each plot.

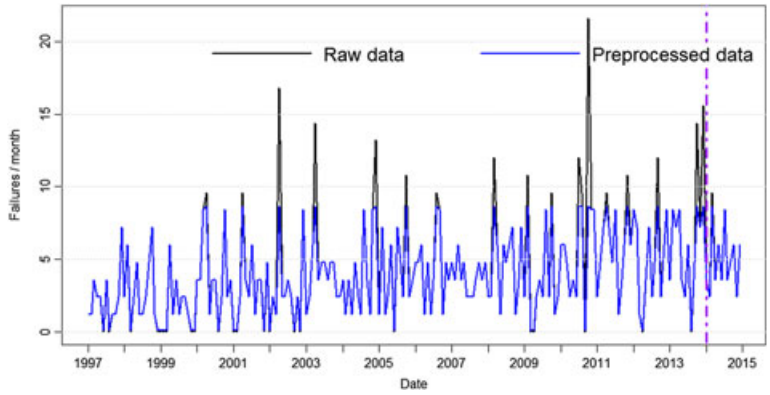

Figure 12. Failures time series after preprocessing (Aircraft B)

Table V. Comparisons between the proposed model and 8 time series models at various forecast points (Aircraft B)

Sample size	Parameters			In-RMSE	Out-RMSE					Time Cost(s)
					1	2	3	6	12	
Holt-Winters[a]	α	β	γ							
180[a]	0.045	0.091	0.274	3.046	1.621	3.541	3.352	2.586	2.439	0.07
204[a]	0.011	0.21	0.191	2.907	1.749	3.605	3.353	2.609	2.474	0.08
ARIMA	*p*	*d*	*q*							
180	0	1	1	2.690	2.433	3.107	2.936	2.543	2.381	0.43
204	0	1	1	2.572	2.448	3.12	2.943	2.545	2.375	0.47
MLR	*L*									
180	24			2.501	1.297	1.938	2.655	2.273	2.316	0.12
204	24			2.547	1.729	2.026	2.912	2.418	2.360	0.14
GMDH	*L*	*Layers*								
180	24	3		2.211	2.586	3.580	3.584	2.571	2.364	10.89
204	24	3		2.268	1.122	2.114	2.837	2.08	2.178	11.47
SSA[v]	*L*	*G*								
180[v]	84	list(1:50)		0.676	0.747	0.913	1.647	3.081	3.074	0.37
204[v]	96	list(1:50)		0.853	0.730	1.686	1.987	2.904	3.161	0.33
SVR	*L*,*C*	ε	σ							
180	24,3	0.269	0.024	1.309	0.449	1.388	2.764	2.234	2.145	0.14
204	24,3	0.252	0.023	1.374	0.243	1.383	2.898	2.280	2.123	0.14
SSA[D]-SVR[20]	SSA[D]	SVR								
180	$L=12$, $G=$list(1)	$L=24$		0.839	0.322	2.723	3.170	2.913	2.342	0.53
204	$L=12$, $G=$list(1)	$L=24$		0.871	0.869	2.642	3.290	2.750	2.438	0.54
SSA[D]-SVR[22]	SSA[D]	SVR								
180	$L=12$, $G=$list(1)	$L=24$		0.051	1.209	2.615	2.922	2.468	2.106	0.50
204	$L=12$, $G=$list(1)	$L=24$		0.036	1.419	2.605	3.038	2.484	2.236	0.58
SSA[D]-SSA[v]-SVR	SSA[D]	SSA[v]	SVR							
180	$L=12$, $G=$list(1)	$L=76$, $G=$list(1:48)	$L=16$	1.420	1.588	2.353	2.720	2.128	1.977	0.38
204	$L=12$, $G=$list(1)	$L=96$, $G=$list(1:48)	$L=16$	1.306	1.654	2.372	2.740	2.109	1.951	0.39

[a]indicates Holt-Winters '*additive*' model.
[v]indicates SSA '*vector*' forecast approach.
[D]indicates SSA model for extracting the two components.
The best RMSE values at various forecast points are marked with the dark background.
Configurations of the computer to conduct the experiments are as follows: CPU: Intel(R) Core(TM)2 Duo CPU E8400 @3.00GHz 3.00GHz; RAM: 4.00GB (3.00GB available); OS: Windows 10 Professional 32bit.
RMSE, root mean square error; ARIMA, autoregressive integrated moving average; MLR, multiple linear regression; GMDH, group method of data handling; SSA, singular spectrum analysis; SVR, support vector machines regression.

Finally, the first five groups of the optimal parameters for the proposed SSA-SVR model are listed in Table VIII. Comparing to Table III, the RMSE values are remarkably smaller in both fitting and forecasting. Generally speaking, for the trend-SSA model, the optimal sample size (S_t) is around 72, and window length (L_t) and groups (G_t) are close to the maximum of the

Table VI. Comparisons of different hybrid models at various forecast points (Aircraft B)

Trend	Fluctuation	In-RMSE	Out-RMSE					Time Cost(s)
			1	2	3	6	12	
Holt-Winters[a]	Holt-Winters[a]	2.724	1.882	3.739	3.437	2.658	2.471	0.14
	ARIMA	2.161	1.956	2.594	2.669	2.304	2.207	2.00
	MLR	2.114	2.099	2.723	2.879	2.485	2.235	0.25
	GMDH	2.160	1.696	3.170	3.214	2.436	2.256	5.58
	SSA[v]	0.762	0.500	1.376	1.472	2.763	3.005	0.27
	SVR	1.344	1.756	2.551	2.751	2.222	2.066	0.29
ARIMA	Holt-Winters[a]	2.702	1.879	3.757	3.419	2.688	2.539	2.72
	ARIMA	2.163	1.953	2.610	2.637	2.368	2.302	3.89
	MLR	2.118	2.095	2.738	2.845	2.555	2.323	2.51
	GMDH	2.144	1.693	3.188	3.185	2.473	2.332	7.43
	SSA[v]	0.790	0.503	1.396	1.441	2.875	3.053	2.51
	SVR	1.309	1.760	2.554	2.706	2.263	2.140	2.99
MLR	Holt-Winters[a]	2.704	1.883	3.734	3.423	2.672	2.512	0.23
	ARIMA	2.157	1.958	2.589	2.652	2.344	2.270	2.04
	MLR	2.120	2.100	2.718	2.861	2.531	2.293	0.37
	GMDH	2.148	1.698	3.165	3.197	2.461	2.307	5.27
	SSA[v]	0.750	0.498	1.370	1.454	2.838	3.033	0.36
	SVR	1.241	1.786	2.516	2.711	2.218	2.094	0.36
GMDH	Holt-Winters[a]	2.707	1.871	3.728	3.417	2.655	2.449	5.05
	ARIMA	2.156	1.945	2.582	2.644	2.308	2.202	6.64
	MLR	2.119	2.087	2.711	2.854	2.489	2.230	5.06
	GMDH	2.148	1.685	3.159	3.190	2.430	2.250	9.79
	SSA[v]	0.749	0.511	1.369	1.450	2.776	2.995	4.93
	SVR	1.326	1.747	2.533	2.720	2.214	2.056	5.16
SSA[v]	Holt-Winters[a]	2.703	1.771	3.577	3.392	2.579	2.206	0.24
	ARIMA	2.166	1.845	2.430	2.658	2.181	2.044	2.21
	MLR	2.177	2.593	2.585	2.659	2.184	2.097	0.32
	GMDH	2.147	1.585	3.007	3.199	2.358	2.166	5.25
	SSA[v]	0.792	0.611	1.262	1.505	2.570	2.848	0.37
	SVR	1.340	1.645	2.385	2.750	2.120	1.955	0.39
SVR	Holt-Winters[a]	2.707	1.782	3.554	3.396	2.582	2.293	0.25
	ARIMA	2.163	1.856	2.412	2.674	2.197	2.077	2.11
	MLR	2.126	2.000	2.544	2.894	2.370	2.118	0.36
	GMDH	2.151	1.596	2.985	3.210	2.371	2.174	5.36
	SSA[v]	0.750	0.600	1.234	1.520	2.593	2.882	0.39
	SVR	1.308	1.655	2.342	2.756	2.133	1.963	0.43

RMSE, root mean square error; ARIMA, autoregressive integrated moving average; MLR, multiple linear regression; GMDH, group method of data handling; SSA, singular spectrum analysis; SVR, support vector machines regression.

Table VII. Major parameters of the models in hybridizations (Aircraft B)

Model	Trend parameters	Fluctuation parameters
Holt-Winters[a](α, β, γ)	(1,0,1)	(0.005,1,0.208)
ARIMA(p, d, q)(P, D, Q)$^{[12]}$	(2,1,0)(2,0,1)$^{[12]}$	(2,0,2)(1,0,1)$^{[12]}$
MLR(L)	16	16
GMDH(L, *Layers*)	(16,3)	(16,3)
SSA[v](L, G)	(96,list(1:48))	(96,list(1:48))
SVR(L, C, ε, σ)	(16,3,0.013,0.374)	(16,3,0.223,0.037)

The sample size of each model is set to 204.
ARIMA, autoregressive integrated moving average; MLR, multiple linear regression; GMDH, group method of data handling; SSA, singular spectrum analysis; SVR, support vector machines regression.

value range. Whereas, for the fluctuation-SVR model, the optimal sample size (S_f) is close to the maximum of the value range, and window length (L_f) is around 24. Also, with the similar procedure, the optimal parameters for Aircraft B are listed in Table IX.

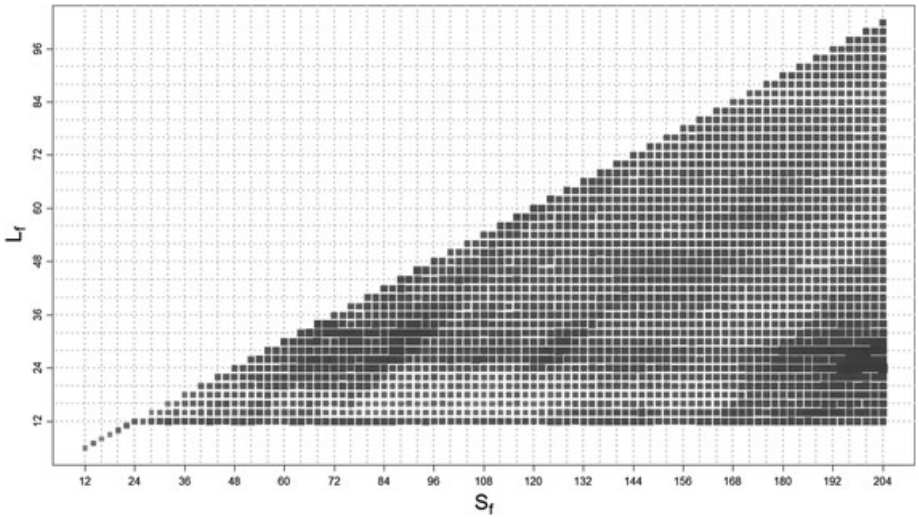

Figure 13. Grid search results for fluctuation-support vector machines regression model's samples size (*Sf*) and window length (*Lf*): The sample size (*St*) of trend-singular spectrum analysis model is 204 and window length (*Lt*) is 24 as well as groups (*Gt*). The step size is 2. The deeper color and bigger size of the blocks indicate the smaller RMSE values. (Aircraft A).

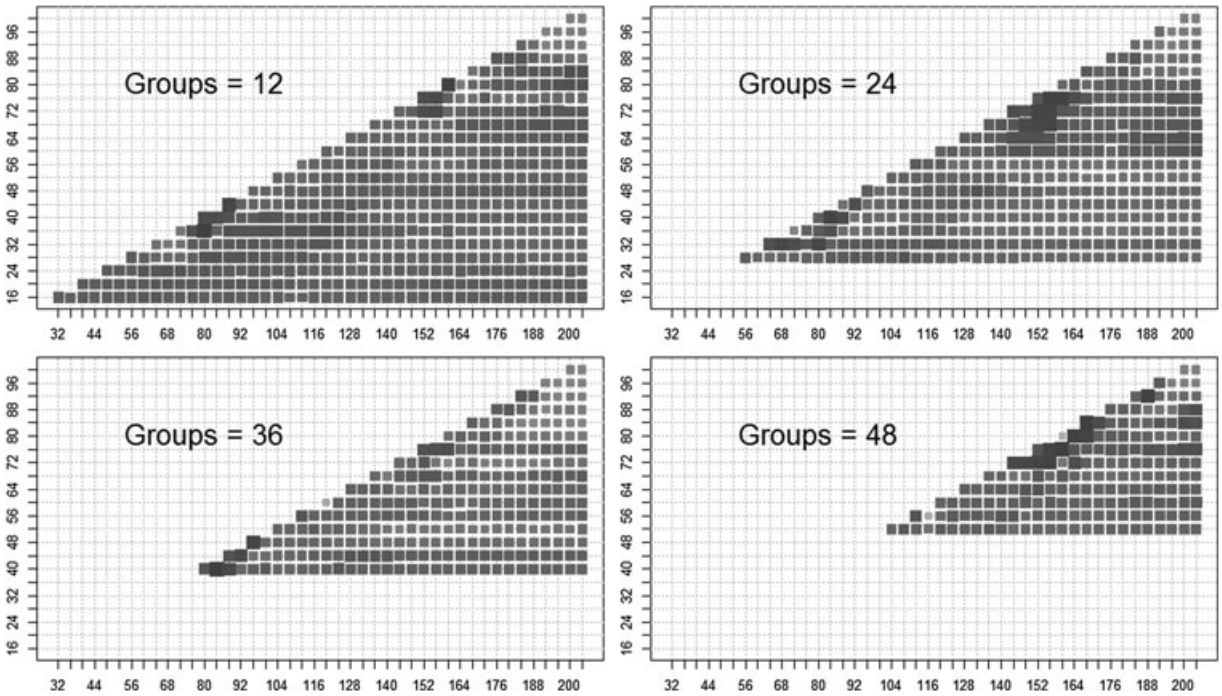

Figure 14. Grid search results for trend-singular spectrum analysis model's sample size (*St*) and window length (*Lt*) at 4 groups: The sample size (*Sf*) of fluctuation-support vector machines regression model is 204 and window length (*Lf*) is 24. The step size is 4. The deeper color and bigger size of the blocks indicate the smaller root mean square error values. (Aircraft A)

Table VIII. First five groups of optimal parameters for the proposed model and the experimental results by grid search, the step size is 4 for trend-SSA model and 2 for fluctuation-SVR model. (Aircraft A)

Rank	SSA^{v}	SVR	In-RMSE	Out-RMSE					Time Cost(s)
	(S_t, L_t, G_t)	(S_f, L_f, C, ε, σ)		1	2	3	6	12	
1	(64,32,28)	(200,28,3,0.256,0.019)	0.959	2.137	1.989	2.429	1.963	1.533	0.39
2	(72,36,28)	(204,24,3,0.251,0.024)	0.955	0.831	1.486	2.235	1.868	1.568	0.36
3	(64,24,20)	(198,24,3,0.250,0.024)	0.954	0.896	1.555	2.252	1.858	1.583	0.35
4	(84,40,36)	(198,28,3,0.255,0.019)	0.969	1.996	1.866	2.393	2.018	1.588	0.41
5	(80,36,12)	(194,24,3,0.251,0.026)	0.836	0.692	1.403	2.218	1.913	1.595	0.31

SSA, singular spectrum analysis; SVR, support vector machines regression; RMSE, root mean square error.

Table IX. First five groups of optimal parameters for the proposed model and the experimental results by grid search, the step size is 4 for trend-SSA model and 2 for fluctuation-SVR model. (Aircraft B)

Rank	SSA[v]	SVR	In-RMSE	Out-RMSE					Time Cost(s)
	(S_t, L_t, G_t)	(S_f, L_f, C, ε, σ)		1	2	3	6	12	
1	(172,72,28)	(172,50,3,0.309,0.010)	0.846	0.685	1.745	1.513	1.852	1.879	0.50
2	(172,72,28)	(174,50,3,0.303,0.011)	0.818	0.671	1.766	1.540	1.857	1.882	0.58
3	(172,72,28)	(178,54,3,0.288,0.010)	0.780	0.176	1.378	1.194	1.759	1.884	0.51
4	(172,72,28)	(176,54,3,0.295,0.010)	0.767	0.007	1.320	1.155	1.732	1.885	0.52
5	(172,72,28)	(182,54,3,0.288,0.010)	0.756	0.113	1.388	1.211	1.760	1.889	0.56

SSA, singular spectrum analysis; SVR, support vector machines regression; RMSE, root mean square error.

4.4. *Discussions*

4.4.1. Effect of sample size on the trend-SSA prediction. In the optimization process of the trend-SSA model, the optimal sample size S_t is around the value of 72. It is of interest to investigate how the forecast accuracy differs in various trend sample sizes. As is shown in Figure 15, the trend predictions are sensitive with the transformation of S_t values. The black line is the practical trend extracted from SSA decomposition to the whole sample data, which shows a clear rising trend on the test set. This is consistent with the experimental result of simple linear regression to the test data. However, the trend-SSA model with $S_t = 204$ depicts a downward curve on the test set. This is not only because SSA grouping becomes incomplete when S_t is larger than 100 (the upper limit of 50 singular values) but also the trend component has its own features. As shown in the figure, the trend time series has a structure change since 2008 and the subsequent curve presents the regular patterns. It is commonly known that the substance of SSA forecast is a kind of extrapolation method. Therefore, it is reasonable to select the latest 72 data points from 2008 to 2013 as the training set for the trend component and the experimental results also proved this.

4.4.2. Decomposition advantages in hybrid modeling. A certain method may be not suitable for the original failure time series, but it is probably suitable for a component decomposed from the original data. This is the advantage of hybrid methods, which gives full play to the strength of each modeling method. For instance, SSA performs well in trend prediction while SVR performs better in fluctuation. In the modeling process of aircraft failure time series, various forecast models can be constructed to focus on different objectives, such as long or short term prediction, trend extraction, trend prediction, and so forth. Choosing the right model will achieve the better performance.

On the other hand, the original data sometimes cannot be processed according to some statistical tests (such as noise test), but the decomposed components are probably suited to conduct these tests (see Figure 9). This is not incredible. Mostly, the decomposed components can imply certain physical meanings in the real world, like inherent and operational reliability in this paper. These components usually have their own features and can be depicted as more stable and regular curves in graphics with more statistical significances.

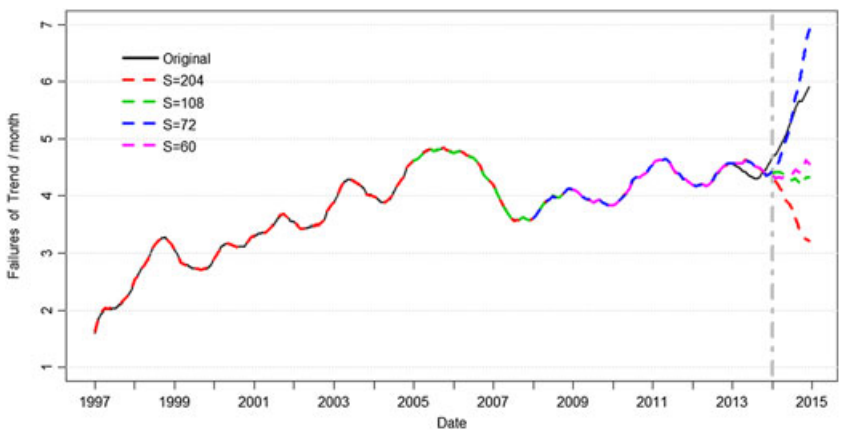

Figure 15. Trend-singular spectrum analysis predictions with various sample sizes (*St*): Window length (*Lt*) and groups (*Gt*) are set to the maximum of the value range. (Aircraft A)

5. Conclusions

The performance of a hybrid model in conjunction with SSA and SVR has been investigated for forecasting the failure time series from two Boeing 737 aircrafts. Although the unitary models including SSA and SVR provided relatively good accuracy in both fitting and forecasting stages, the performance was obviously improved when the hybridization technique was introduced. This might be a strategy of 'divide-and-conquer'. The raw failure data was decomposed into two significant components denoted as the trend and fluctuation on the basis of reliability categories. Then the two components were separately processed by using appropriate models (SSA and SVR). In this paper, we can see that SVR is good at coping with the fluctuation and is superior to other models, particularly in forecast accuracy. Moreover, the parameter optimization for this hybrid model was performed by using a stepwise gird search method and achieved the satisfying results (The RMSE value on 12 predictions reached 1.533).

Overall, the hybrid model proposed in this paper gains a noteworthy performance in reliability prediction. Yet, several related problems should be solved in the future. We can note that some extreme points exist in both the raw data and the predicted data (Figures 4, 10, and 12). These data points are significant in practice for establishing appropriate maintenance strategies and reliability analysis such as fault diagnosis. But they greatly affect the fitting and forecasting accuracy of the prediction model. Hence, it is reasonable to construct the specific forecast model for these extreme points. In addition, the parameters of the hybrid model are complex. Although the present grid search method reached the goal of seeking optimal parameters, it is still not perfect. Therefore, in the future we will focus on pursuing more optimization techniques to improve the search efficiency for the hybrid model.

Acknowledgements

This work is supported by the Technology Foundation Program (JSZL2014601B008) of the National Defense Technology Industry Ministry, and by NSFC (no. 61602237). The authors are grateful to the editors and the two anonymous referees for their insightful comments, which have significantly helped to enhance the quality of the paper. The authors also would like to thank Ren Jian, Zhang Kui, and Zhang Hua for their invaluable contributions during the collaboration.

References

1. Rocco SCM. Singular spectrum analysis and forecasting of failure time series. *Reliability Engineering & System Safety* 2013; **114**:126–136. doi:10.1016/j.ress.2013.01.007.
2. Moura MDC, Zio E, Lins ID, Droguett E. Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety* 2011; **96**(11):1527–1534. doi:10.1016/j.ress.2011.06.006.
3. Sapankevych NI, Sankar R. Time series prediction using support vector machines: A survey. *IEEE Computational Intelligence Magazine* 2009; **4**(2):24–38. doi:10.1109/MCI.2009.932254.
4. Xu K, Xie M, Tang LC, Ho SL. Application of neural networks in forecasting engine systems reliability. *Applied Soft Computing* 2003; **2**(4):255–268. doi:10.1016/S1568-4946(02)00059-5.
5. Su CT, Tong L, Leou CM. Combining time series and neural network approaches for modeling reliability growth. *Journal of the Chinese Institute of Industrial Engineers* 1997; **14**:419–430. doi:10.1080/10170669.1997.10432936.
6. Cai KY, Cai L, Wang WD, Yu ZY, Zhang D. On the neural network approach in software reliability modeling. *Journal of Systems & Software* 2001; **58**(1):47–62. doi:10.1016/S0164-1212(01)00027-9.
7. Karunanithi N, Whitley D, Malaiya KY. Prediction of software reliability using connectionist models. *IEEE Transactions on Software Engineering* 1992; **18**(7):563–574. doi:10.1109/32.148475.
8. Liang T, Noore A. Evolutionary neural network modeling for software cumulative failure time prediction. *Reliability Engineering & System Safety* 2005; **87**(1): 45–51. DOI: doi:10.1016/j.ress.2004.03.028
9. Hu QP, Xie M, Ng SH, Levitin G. Robust recurrent neural network modeling for software fault detection and correction prediction. *Reliability Engineering & System Safety* 2007; **92**(3):332–340. doi:10.1016/j.ress.2006.04.007.
10. Al-Garni AZ, Jamal A. Artificial neural network application of modeling failure rate for Boeing 737 tires. *Quality and Reliability Engineering International* 2011; **27**(2):209–219. doi:10.1002/qre.1114.
11. Zio E, Broggi M, Golea LR, Pedroni N. Failure and reliability predictions by infinite impulse response locally recurrent neural networks. *Chemical Engineering Transactions* 2012; **26**:117–122. doi:10.3303/CET1226020.
12. Gu B, Sheng VS, Tay KY, Romano W, Li S. Incremental support vector learning for ordinal regression. *IEEE Transactions on Neural Networks & Learning Systems* 2014; **26**(7):1403–1416. doi:10.1109/TNNLS.2014.2342533.
13. Gu B, Sheng VS, Wang Z, Ho D, Osman S, Li S. Incremental learning for ν-support vector regression. *Neural Networks the Official Journal of the International Neural Network Society* 2015; **67**(C):140–150. doi:10.1016/j.neunet.2015.03.013.
14. Pai PF. System reliability forecasting by support vector machines with genetic algorithms. *Mathematical & Computer Modelling* 2006; **43**(3–4): 262–274. DOI: doi:10.1016/j.mcm.2005.02.008
15. Vahabie AH, Yousefi MMR, Arrabi BN, Lucas C, Barghinia S. Combination of singular spectrum analysis and autoregressive model for short term load forecasting. In *Proceedings of Power Tech*, 2007 *IEEE Lausanne*, Lausanne, CH, 2007;1090-1093. DOI: 10.1109/PCT.2007.4538467
16. Wu CL, Chau KW. Prediction of rainfall time series using modular soft computingmethods. *Engineering Applications of Artificial Intelligence* 2013; **26**(3):997–1007. doi:10.1016/j.engappai.2012.05.023.
17. Wu CL, Chau KW. Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. *Journal of Hydrology* 2011; **399**(3–4):394–409. doi:10.1016/j.jhydrol.2011.01.017.
18. Samsudin R, Saad P, Shabri A. A hybrid GMDH and least squares support vector machines in time series forecasting. *Neural Network World* 2011; **21**(3):251–268 10.14311/NNW.2011.21.015.
19. Chau KW, Wu CL. A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. *Journal of Hydroinformatics* 2010; **12**(4):458–473. doi:10.2166/hydro.2010.032.
20. Sivapragasam C, Liong SY, Pasha MFK. Rainfall and runoff forecasting with SSA-SVM approach. *Journal of Hydroinformatics* 2001; **3**(3):213–217.

21. Islam MN, Liong SY, Phoon KK, Liaw CY. Forecasting of river flow data with a general regression neural network. In *Proceedings of International Symposium on Integrated Water Resources Management*, California, USA, 2000; **272**: 285–290.
22. Wen FH, Xiao JH, He ZF, Gong X. Stock price prediction based on SSA and SVM. *Procedia Computer Science* 2014; **31**:625–631. doi:10.1016/j.procs.2014.05.309.
23. Al-Garni AZ, Tozan M, Al-Garni AM, Jamal A. Failure forecasting of aircraft air-conditioning/cooling pack with field data. *Journal of Aircraft* 2007; **44**(3):996–1002. doi:10.2514/1.26561.
24. Su C, Jin Q, Fu YQ. Correlation analysis for wind speed and failure rate of wind turbines using time series approach. *Journal of Renewable and Sustainable Energy* 2012; 4(3). DOI: 10.1063/1.4730597
25. Vautard R, Yiou P, Ghil M. Singular-spectrum analysis: A toolkit for short, noisy chaotic signals. *Physica D: Nonlinear Phenomena* 1992; **58**(1–4):95–126. doi:10.1016/0167-2789(92)90103-T.
26. Golyandina N, Korobeynikov A. Basic singular spectrum analysis and forecasting with R. *Computational Statistics & Data Analysis* 2014; **71**:934–954. doi:10.1016/j.csda.2013.04.009.
27. Golyandina N. On the choice of parameters in singular spectrum analysis and related subspace-based methods. *Statistics and Its Interface* 2010; **3**(3):259–279.
28. Hipel KW, McLeod AI. Time Series Modelling of Water Resources and Environmental Systems. Elsevier: Amsterdam, NED, 1994. doi:10.1016/S0167-5648(08)70651-8.
29. Hassani H. Singular spectrum analysis: Methodology and comparison. *Journal of Data Science* 2007; **5**(2):239–257.
30. Alexandrov T, Golyandina N. Automatic extraction and forecast of time series cyclic components within the framework of SSA. in *Proceedings of the 5th St. Petersburg Workshop on Simulation*, St. Petersburg, Russia, 2005; 45–50.
31. Brereton RG, Lloyd GR. Support vector machines for classification and regression. *Analyst* 2010; **135**(2):230–67. doi:10.1039/b918972f.
32. Kecman V. Support vector machines – an introduction, support vector machines: Theory and applications. Springer Berlin Heidelberg: Berlin, GER, 2005; **1–47**. DOI: 10.1007/10984697_1
33. Vapnik VN. The Nature of Statistical Learning Theory. Springer: New York, USA, 1995. doi:10.1007/978-1-4757-2440-0.
34. Schölkopf B, Burges CJC, Smola AJ. Advances in Kernel Methods: Support Vector Learning. MIT Press: Cambridge, MA, USA, 1999.
35. Wu CL, Chau KW, Li YS. River stage prediction based on a distributed support vector regression. *Journal of Hydrology* 2008; **358**(1–2):96–111. doi:10.1016/j.jhydrol.2008.05.028.
36. Bratton D, Kennedy J. Defining a standard for particle swarm optimization. In *Proceedings of* 2007 *IEEE Swarm Intelligence Symposium*, Honolulu, HI, 2007; 120–127. *DOI*: 10.1109/SIS.2007.368035
37. Pai PF, Hong WC. Software reliability forecasting by support vector machines with simulated annealing algorithms. *Journal of Systems & Software* 2006; **79**(6):747–755. doi:10.1016/j.jss.2005.02.025.
38. Cherkassky V, Ma Y. Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* 2004; **17**(1):113–126. doi:10.1016/S0893-6080(03)00169-2.
39. Ljung GM, Box GEP. On a measure of lack of fit in time series models. *Biometrika* 1978; **65**(2):297–303. doi:10.1093/biomet/65.2.297.
40. Esam M, Ian MA. Improved multivariate portmanteau test. *Journal of Time* 2012; **33**(2):211–222. doi:10.1111/j.1467-9892.2011.00752.x.
41. Grossmann A, Morlet J. Decomposition of hardy functions into square integrable wavelets of constant shape. *Siam Journal on Mathematical Analysis* 2006; **15**(4):723–736. doi:10.1137/0515056.
42. Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society a Mathematical Physical & Engineering Sciences* 1998; **454**(197):903–995. doi:10.1098/rspa.1998.0193.
43. Cleveland RB, Cleveland WS. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics* 1990; **6**(1):3–33.
44. Chapelle O, Vapnik V, Bousquet O, Mukherjee S. Choosing multiple parameters for support vector machines. *Machine Learning* 2001; **46**(1–3):131–159. doi:10.1023/A:1012450327387.
45. Chen PW, Wang JY, Lee HM. Model selection of SVMs using GA approach. In *Proceedings of IEEE International Joint Conference on Neural Networks*, Budapest, HUN, 2004; (3): 2035–2040. DOI: 10.1109/IJCNN.2004.1380929
46. Lin JY, Cheng CT, Chau KW. Using support vector machines for long-term discharge prediction. *Hydrological Sciences Journal/journal Des Sciences Hydrologiques* 2006; **51**(4):599–612. doi:10.1623/hysj.51.4.599.
47. Hsu CW, Chang CC, Lin CJ. A practical guide to support vector classification. T*echnical report*, *Department of Computer Science*, *National Taiwan University* 2003; 67(5).
48. Chatfield C. The Holt-Winters forecasting procedure. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 1978; **27**(3):264–279. doi:10.2307/2347162.
49. Bartholomew DJ. Time series analysis forecasting and control. *Journal of the Operational Research Society* 1971; **22**(2):199–201. doi:10.1057/jors.1971.52.
50. Kumar U, Jain VK. ARIMA forecasting of ambient air pollutants (O-3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment* 2010; **24**(5):751–760. doi:10.1007/s00477-009-0361-8.
51. Jeong K, Koo C, Hong T. An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network). *Energy* 2014; **71**:71–79. doi:10.1016/j.energy.2014.04.027.
52. Hyndman RJ, Khandakar Y. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 2008; **27**(3):1–22.
53. Nikolopoulos K, Goodwin P, Patelis A, Assimakopoulos V. Forecasting with cue information: A comparison of multiple regression with alternative forecasting approaches. *European Journal of Operational Research* 2007; **180**(1):354–368. doi:10.1016/j.ejor.2006.03.047.
54. Bianco V, Manca O, Nardini S. Electricity consumption forecasting in Italy using linear regression models. *Energy* 2009; **34**(9):1413–1421. doi:10.1016/j.energy.2009.06.034.
55. Ivakhnenko AG. The group method of data handling-a rival of the method of stochastic approximation. *Soviet Automatic Control* 1968; **1**(3):43–55.
56. Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. *Applied Energy* 2010; **87**(7):2313–2320. doi:10.1016/j.apenergy.2009.12.013.
57. The website of Caterpillar-SSA. http://www.gistatgroup.com/cat/index.html.
58. Cortez P. Data Mining with Neural Networks and Support Vector Machines Using the R/rminer Tool, Vol. **572–583**. Springer Berlin Heidelberg: Berlin, GER, 2010. doi:10.1007/978-3-642-14400-4_44.

Authors' biographies

Xin Wang is currently pursuing his PhD degree at Beihang University. He received his MS and BS degrees in the School of Mechanical Engineering and Automation at Beihang University. His research is focused on reliability forecast and data-driven techniques.

Dr Ji Wu is an associate professor and assistant dean of the School of Computer Science and Engineering (SCSE) at Beihang University. He received his PhD degree from Beihang University in 2003 and MS degree from the Second Research Institute of the China Aerospace Science and Industry Group in 1999. He focuses on the industry-oriented researches. His research interests include embedded system and software modeling and verification, software requirement and architecture modeling and verification, safety and reliability assessment, and software testing. He was invited to visit Simula Research Laboratory for 1 year in 2012.

Professor Chao Liu is the Director of Software Engineering Institute (SEI), Beihang University, Beijing, China. His research interests include software quality engineering, software testing, as well as software process improvement. He received his PhD and MS degrees in Computer Software and Theory at Beihang University and his BS degree in Mathematics at Beijing University of Posts and Telecommunication.

Senzhang Wang received his PhD degree from Beihang University in 2015. He currently is an assistant professor at the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics. His research focus is on data mining, social computing, big data, and urban computing. He has published more than 30 papers in premier conferences and journals in computer science including SIGKDD, AAAI, SDM, ACM SIGSPAIAL, ECML-PKDD, ACM Transactions on Intelligent Systems and Technology, Knowledge and Information Systems, IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Multimedia, and so on.

Wensheng Niu received his BS degree in Computer Science and Technology from Beihang University in 1987 and his PhD degree in Computer Architecture from Xi'an Jiao Tong University in 1996. He is Chief Expert of Aviation Industry of China (AVIC) and is Research Professor of Beihang University. His research interests include anti-harsh environment computer, high reliability computer, and embedded high-speed computer.