Mathematical Biosciences and Engineering

MBE, 19(7): 6620–6637.
DOI: 10.3934/mbe.2022311
Received: 23 March 2022
Revised: 22 April 2022
Accepted: 23 April 2022
Published: 27 April 2022

http://www.aimspress.com/journal/MBE

Research article

Generalized Cauchy process based on heavy-tailed distribution and grey relational analysis for reliability predicting of distribution systems

Jun Gao[1], Fei Wu[1,*], Yakufu Yasen[2], Wanqing Song[1] and Lijia Ren[1]

[1] School of Electronic & Electrical Engineering, Shanghai University of Engineering Science, 333# Longteng road, Songjiang District, Shanghai 201620, China
[2] State Grid Kashi Electric Power Supply Company, 156 Renmin West Road, Kashi 844099, China

* **Correspondence:** Email: fei_wu1@163.com; Tel: +18939916815; Fax: +8602167791125.

Abstract: Failure interruption often causes large blackouts in power grids, severely impacting critical functions. Because of the randomness of power failure, it is difficult to predict the leading causes of failure. ASAI, an essential indicator of power-supply reliability, can be measured from the outage time series. The series is non-stationary stochastic, which causes some difficulty in analyzing power-supply reliability. Considering that the time series has long-range dependence (LRD) and self-similarity, this paper proposes the generalized Cauchy (GC) process for the prediction. The case study shows that the proposed model can predict reliability with a max absolute percentage error of 8.28%. Grey relational analysis (GRA) has proved to be an effective method for the degree of correlation between different indicators. Therefore, we propose the method, which combines both GC and GRA to obtain the correlation coefficients between different factors and ASAI and to get the main factors based on this coefficient. The case study illustrates the feasibility of this approach, which power enterprises can employ to predict power-supply reliability and its influencing factors and help them identify weaknesses in the grid to inform employees to take protective measures in advance.

Keywords: reliability of distribution systems; grey relational coefficient; generalized Cauchy process; failure interruption; long-range dependence

Abbreviations: ASAI: electric power distribution service reliability; GC: generalized Cauchy; ACF: autocorrelation function; LRD: long-range dependence; GRA: grey relational analysis; MLE: maximum likelihood estimation; ME: maximum error; PDF: probability density function; fBM:

fractional Brownian motion; LSTM: long-short-term memory; MAE: mean absolute error; RMSE: root mean square error; SBTU: the sum of blackout time of users; MAPE: mean absolute percentage error; MAXE: max absolute percentage error

1. Introduction

The power system is one of the pillars of modern society and economy. Power outages cause considerable inconvenience to residents and extensive economic and non-economic losses. The causes of power outages are divided into failure interruption and scheduled interruption. Most of the significant power system accidents are caused by failure interruption. For example, the "8.14" power outage in the eastern United States-Canada was caused by a foreign-body short circuit; the aging equipment caused the "5.25" power outage in Moscow. These accidents have caused significant economic losses and endangered people's lives. However, the power enterprises inform consumers when the scheduled interruptions occur, and the outage losses become negligible. Failure interruptions are challenging to predict because of the randomness and abruptness. The main factors of failure interruption include product quality, climate reasons, equipment aging, foreign-body short circuit, other external factors, and construction influence. Predicting the main factors affecting power-supply reliability in the future through historical data is of great significance for power enterprises to take measures to improve reliability.

Due to the burstiness of failure interruptions, there are hardly any papers exploring how to analyze and predict power-supply reliability and factors. Milad Doostan et al. [1] proposed Holt-Winters exponential smoothing method to predict vegetation-induced outage time in the distribution systems. This method combined weather and geographic data with past vegetation-related blackout information and predicted future blackout times. Using the random forest classifier, Roope et al. [2] studied the impact of convective storms on power outage loss. The historical data obtained from weather radar, ground weather observations, lightning detectors, and corresponding outage loss data are used as inputs for classification. The obtained model can predict the outage losses under extreme weather. B Chowdhury et al. [3] predicted whether there would be zero, one or two, or more outages when thunderstorms occur by using the weighted logistic regression random forest model and taking advantage of interruption data and weather forecast data. It can be seen that the purpose of the above study is only to analyze a specific factor without a comprehensive analysis of the main factors affecting power-supply reliability, so its significance for improving reliability is not as great as this paper. In addition, the study of power-supply reliability is divided into two aspects. One is to analyze reliability by establishing a mathematical model [4,5] of the distribution systems, including the least path method, fault tree analysis, minimum path set method, failure mode, effects analysis, etc. These methods need to draw the distribution systems structure diagram, so they are very complex. The other is to explore spatial-temporal features based on historical data to predict, including the gray prediction model [6], regression prediction method [7], auto-regressive integrated moving average (ARIMA) model [8], etc. Rajeevan et al. [8] established the reliability model of Wind Farm by using the load curve of the utility grid and the capacity outage probability table developed by the ARIMA wind speed model. Chen et al. [9] proposed to use the improved logistic regression method to predict the reliability of power equipment operation. Li et al. [10] used the least absolute deviation prediction method based on the particle swarm optimization algorithm to carry out the regression model on the power-supply reliability. However, the prediction principles of these methods do not take into account the LRD and self-similarity of the time series. Compared with the GC prediction model introduced in this paper, they have lower prediction accuracy.

The GC prediction model is a statistical prediction method, based on historical data consistent with LRD. The LRD means that the changes in the subsequent data are influenced by the current and past data and can be represented by the H and D indicators. The GC process is quite different from the fractional Brownian motion (fBM) model [11], an LRD prediction model with the H and D indicators. There is a linear relationship between H and D. Nevertheless, these indicators of the GC process are two independent parameters, so they can more flexibly describe the long correlation process. The basic principle of the GC prediction model is as follows. First, calculate the time series indicators to determine whether the GC prediction model can predict the series. If it is satisfied, make use of the theory of the Ortigueira fractal linear system to generate the steady GC sequence. The increment of the GC sequence obeys Gaussian distribution, and the variance of increment in the same interval can be achieved through numerical simulation when distribution parameters are determined. Based on the Itô process and the application of Scholes and F.Black [12], along with Wang et al. [13] to fBM, the obtained Gaussian distribution can be substituted into the discrete expression for prediction. The drift coefficient and diffusion coefficient in stochastic differential equations (SDE) of the GC process can be obtained by maximum likelihood estimation (MLE) [14] because the GC process is the non-Markov process. It is not easy to get the Probability Distribution Function (PDF) directly, so the approximate PDF can be acquired using the Monte Carlo simulation method.

GRA focuses on a dynamic system variable to judge the influence degree of its influencing factors on the system. It provides a reference for the development and changing trend of the system and has been applied in distribution systems. It is generally divided into two aspects, one is to sort the factors that affect the systems, and the other is to evaluate different schemes and select the best one. Niu et al. [15] proposed a new short-term empirical mode decomposition-grey relationship analysis-modified particle swarm optimization-least square vector machine load prediction model. The function of GRA is to find out the main influencing factors from many influencing factors and carry out load prediction. Akay et al. [16] proposed a new power quality understanding and evaluation method for low-voltage DC distribution systems through the analytic hierarchy process and entropy weighting coefficient method. GRA is used to calculate the correlation coefficient between the power quality level of the scheme to be evaluated and the ideal power quality to evaluate its power quality. Therefore, it is reasonable to use GRA in this paper to find the main factors affecting power-supply reliability.

The case study data is the power outage events of the power grid in Shanghai in 2019. By calculating the H and D indicators of outage time series, they all meet the requirements of the GC prediction model, which shows that this model has important application prospects in the power-supply reliability prediction of distribution networks. The result proves that the GC prediction model has better performance than the fBM, LSTM, and DANN (deterministic annealing neural network) [17,18]. The correlation coefficient between the reliability index ASAI of distribution systems and the influencing factors is calculated by GRA to predict the main factors affecting power-supply reliability. According to the conclusion, corresponding technical measures and management methods can be taken in advance to reduce the occurrence of fault interruptions and power outage losses. This method can improve the operation and maintenance level of the power grid and the reliability of the power supply of the distribution system.

The organization of this paper is as follows. Section 2 mainly describes the analysis of the GC process. Section 3 combines the actual situation, using the GC prediction model and GRA to predict the power-supply reliability and analyze the factors of power failure in Shanghai. The prediction results prove the effectiveness of the prediction model. The last section summarizes and analyzes the whole article.

2. The LRD and heavy-tail properties of GC process

2.1. The LRD character of GC process

If $\mathrm{Y(t)}$ is the stationary Gaussian process and its autocorrelation function (ACF) satisfies Eq (1), it is said to be the generalized Cauchy (GC) process [19].

$$F_{YY}(\tau)=\left(1+|\tau|^{4-2D}\right)^{\frac{1-H}{2-D}} \tag{1}$$

Here D is the fractal dimension parameter, H is the Hurst parameter and the time interval is denoted as τ. The fractal characteristics [20] show that D and H reveal the local irregularity and global LRD the characteristics of the GC process when $1.5 \le D < 2$ and $0.5 < H < 1$, respectively.

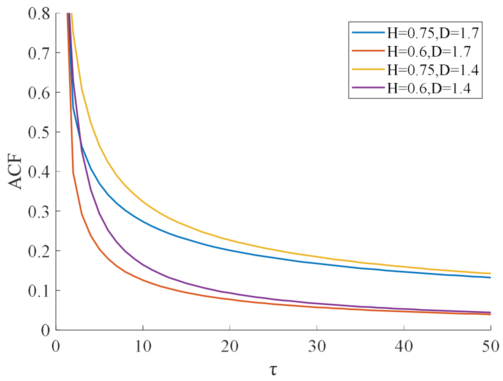

Figure 1. ACF curve of GC process under different parameters.

Figure 1 shows the ACF curves of the GC process under different parameter values, which show the influence of different H and D indicators on the correlation. It can be seen ACF curve approaches 0 within a short time interval for $H < 0.5$. This indicates that in this case, the time series does not have LRD characteristics. And when $H > 0.5$, the larger the value of H, the higher the ACF curve, which indicates that the larger the value of H, the stronger the LRD characteristics of the time series. In addition, the larger the value of D, the more the curve of ACF approaches 0, which indicates that the larger the value of D, the more complex the time series, and the weaker the LRD characteristics.

2.2. Heavy-tailed nature of the GC process

According to probability theory, a heavy-tailed distribution represents a probability distribution whose tails are not exponentially bounded. That is, the tails are “heavier” or “thicker” than the exponential distribution.

Supposing a random variable Y, its cumulative distribution function can be written in the following form:

$$F(y) = P(Y > y) \tag{2}$$

If this random variable Y satisfies the following equation:

$$\lim_{y\to\infty} e^{\rho y}(1 - F(y)) = \infty \tag{3}$$

The distribution of Y is said to be a heavy-tailed distribution, where ρ is a constant arbitrarily greater than 0. According to the properties of the cumulative distribution function,

$$0 \le F(y) \le 1 \tag{4}$$

$$0 \le 1 - F(y) \le 1 \tag{5}$$

$e^{\rho y}$is a monotonically increasing exponential function greater than 0. The following form can be obtained by transforming Eq (3).

$$\lim_{y\to\infty} \frac{1-F(y)}{e^{-\rho y}} = \infty \tag{6}$$

when $y \to \infty$, both the numerator and denominator of Eq (6) tend to 0. According to the rule of infinitesimal operation, the limit of Eq (6) is infinity, then the denominator $e^{-\rho}$ is a higher order infinitesimal of the numerator $1 - F(y)$, which means that the random variable Y obeys a distribution that decays more slowly than the exponential distribution when $y \to \infty$. This also means that the tails of the heavy-tailed distribution decay more slowly than the tails of the exponential distribution.

From the Taqqu's law [21], we can obtain that the stochastic process exhibits a heavy-tail on the PDF, which is equivalent to the LRD characteristics on the ACF. The ACF of random process $\{Y(t), 0 < t < \infty\}$ at time t_1, t_2 can be represented by PDF $f(x: t_1, t_2)$.

$$r_y(t_1, t_2) = E[y(t_1)y(t_2)] = \iint y(t_1)y(t_2)f_{GC}(y; t_1, t_2)\, dy(t_1)dy(t_2) \tag{7}$$

It follows from Eq (7), that slow decay of PDF leads to the slow decay of ACF. This enables us to state the existence of the strong internal relationship between the heavy tail of PDF analysis and the LRD characteristics of ACF analysis.

The Cauchy distribution is also a heavy-tailed distribution, and its PDF expression is:

$$p(y; \xi, \varpi) = \frac{1}{\pi\varpi\left[1+(\frac{y-\xi}{\varpi})^2\right]} \tag{8}$$

where the ξ is position parameter, the range parameter ϖ indicates the discrete degree. When $\xi = 1$, $\varpi = 0$,the above formula can be simplified as:

$$p(y) = \frac{1}{\pi(1+y^2)} \tag{9}$$

In the description of time series, mean and variance are two important statistical characteristics. The former is used to describe the global nature of the time series, that is, the overall development trend of the sequence, similar to the H indicator in the previous section. The latter is used to describe the local nature, that is, the local concentration of the sequence, similar to the D indicator in the previous section. Let ψ and ϱ be the mean and variance of the Cauchy distribution. Their equations are:

$$\Psi = \int_{-\infty}^{\infty} yp(y)dy = \int_{-\infty}^{\infty} \frac{y}{\pi(1+y^2)}dy = \infty \tag{10}$$

$$\varrho = \int_{-\infty}^{\infty} y^2p(y)dy = \int_{-\infty}^{\infty} \frac{y^2}{\pi(1+y^2)}dy = \infty \tag{11}$$

Since the mean and variance of the Cauchy distribution are infinite, its mean and variance do not exist. This is also the reason why the *H* and *D* exponents are used to represent the LRD characteristics. Another PDF equation for the GC process is given by Carrillo et al.

$$f_{GC}(y) = \frac{\rho\Gamma(2/\rho)\varpi}{2(\Gamma(1/\rho))^2}(\varpi^\rho + |y - \xi|^\rho)^{-2/\rho} \tag{12}$$

the heavy-tailed parameter ρ indicates the heavy-tailed degree of the PDF. When $0 < \rho \le 2$, the GC process is a heavy-tailed distribution. $\Gamma(\cdot)$ represents the gamma function. In Figure 2, the influence of the parameters of the GC process on the PDF is shown. The parameters μ, γ, and ρ in Figure 2 are position parameters, range parameters, and heavy-tailed parameters, respectively. The range parameter is the symmetry axis of the GC PDF. The range parameter indicates the discrete degree of PDF. The smaller the range parameter, the more concentrated the value of the points in the PDF. The heavy-tailed parameter indicates the heavy-tailed degree of the PDF. The smaller the heavy-tailed parameter, the heavier the tail of the PDF. The GC process is a heavy-tailed distribution when $0 < \rho \le 2$.

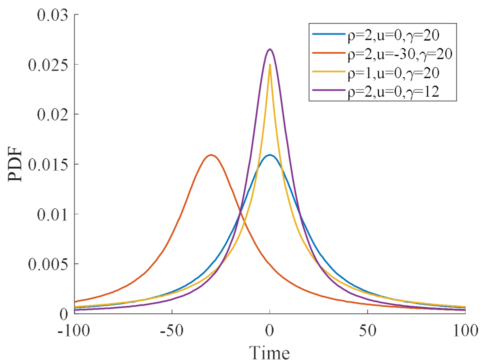

Figure 2. PDF of GC process.

3. GC prediction model

3.1. The generation of GC prediction model

The theory of Ortigueira fractal linear system [22,23] shows that steady-state time series can be obtained through white noise and filters, and the specific process is:

$$h(t)=s(t)*g(t)=\int_0^t g(t-\tau)s(\tau)d\tau \tag{13}$$

where $s(t)$ is Gaussian white noise obtained by random function, $g(t)$ is the impulse function, τ is the time interval, and $h(t)$ is the generated stationary time series. Similarly, non-stationary white noise can be passed through the linear filter to obtain non-stationary time series, so GC sequence can be obtained by combining Gaussian white noise and impulse function. The concrete form is as follows.

The power spectrum $g(w)$ of $Y(t)$ is the Fourier transform of the ACF of it, as shown in Eq (14). $g(w)$ is related to the impulse function $g(t)$ as follow.

$$g(w) = F[(1 + |\tau|^{4-2D})^{-\frac{1-H}{2-D}}] \tag{14}$$

$$g(t) = F^{-1}\left[F((1 + |\tau|^{4-2D})^{-\frac{1-H}{2-D}})^{0.5}\right] \tag{15}$$

And the power spectrum of unit white noise is $S(w)$, where $\theta(w)$ is the random function. The Fourier inverse transform is carried out to obtain the Gaussian white noise function $s(t)$ as shown in Eq (16).

$$s(t) = F^{-1}[S(w)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\theta} e^{jwt} dw \tag{16}$$

By substituting Eqs (15) and (16) into (13), the expression of GC sequence can be obtained as shown in Eq (17).

$$h(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{j\theta} e^{jwt} dw * F^{-1}\left[F((1 + |\tau|^{4-2D})^{-\frac{1-H}{2-D}})^{0.5}\right] \tag{17}$$

where $F(\cdot)$ and $F^{-1}(\cdot)$ are respectively Fourier transform and the inverse transform. The process of the procedure for generating the increment of GC sequence is shown in Figure 3. When $H = 0.55$, $D = 1.2$, the GC difference time sequence is obtained by difference and the result is shown in Figure 4.

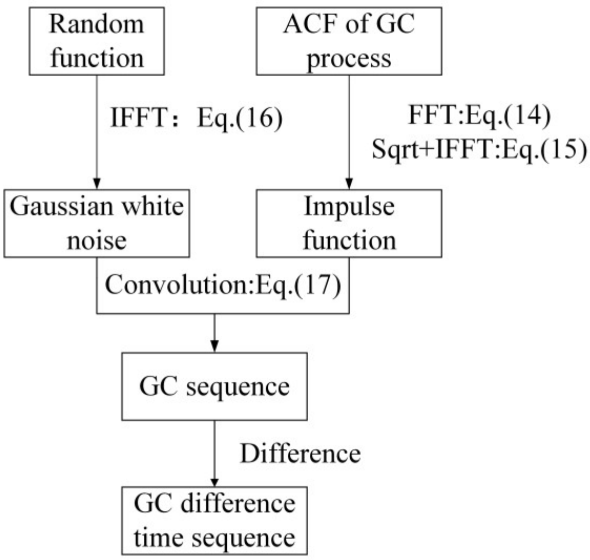

Figure 3. Flowchart of the procedure for generating the increment of GC sequence.

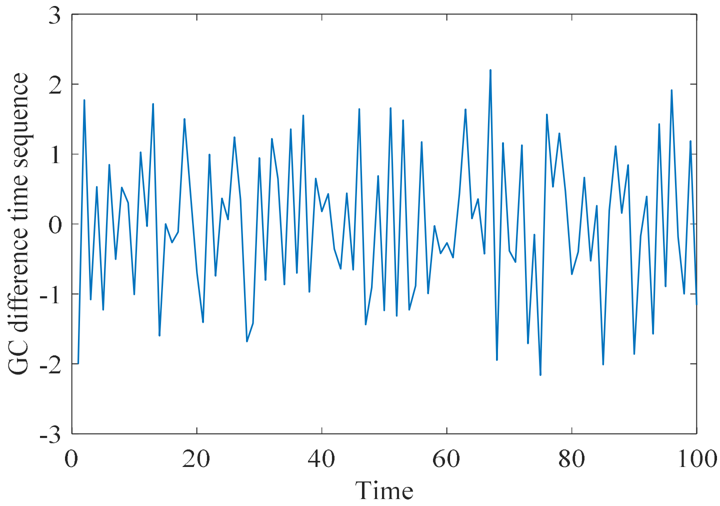

Figure 4. GC difference time sequence.

Based on the definition of the Itô process, the GC process can be regarded as the stochastic interference term with LRD property. Therefore, the stochastic differential equation (SDE) of the Ito process driven by the GC process is:

$$dY(t) = \partial(t)dt + \emptyset(t)dh(t) \tag{18}$$

$\partial(t)$ is the drift coefficient, $h(t)$ is the GC time series generated by Gaussian white noise and impulse function, and $\emptyset(t)$ is the diffusion coefficient. The drift coefficient represents the overall degradation trend of the degradation process, while the diffusion coefficient represents the uncertainty of the degradation quantity in the degradation process.

According to the improvement of the fBm model by Scholes, F. Black and Wang et al, they established the Black-Scholes model to describe the trend of the financial option B_t:

$$dB_t=\mu B_t dt+\delta B_t dF_H(t) \tag{19}$$

where $F(t)$ is the fBm. The stochastic sequence forecasting model based on the GC process is obtained by combining Eqs (18) and (19), and the form is as follow:

$$dY(t) = \partial(t)Y(t)dt + \emptyset(t)Y(t)dh(t) \tag{20}$$

The distribution of increment $\Delta h(t)$ in the GC process can be obtained by statistical reasoning. The specific solution process is as follows:

Step 1: Generate a sequence of GC process values according to Eq (17).

Step 2: Determine the time interval τ, and the difference of the two-state quantities with the interval τ in the sequence, namely,

$$\Delta h(t) = h(t+\tau) - h(t) \tag{21}$$

Step 3: Step 2 is repeated for the time series generated by Step 1, and multiple differences are made to construct an incremental set.

Step 4: Find the variance of the set distribution. When the increment interval τ is specified, the increment follows the Gaussian distribution, that is, $\Delta h(t) \sim N(0, \emptyset_\tau)$.Thus,

$$\Delta h(t) = h(t+\tau) - h(t) \sim N(0, \emptyset_\tau) \tag{22}$$

And because

$$\Delta Y(t) = Y(t+\tau) - Y(\tau) \tag{23}$$

Substituting Eqs (22) and (23) into (20), when $\tau = 1$, the GC prediction model is obtained:

$$Y(t+1) = Y(t) + \partial(t)Y(t)\Delta t + \emptyset(t)Y(t)\Delta h(t) \tag{24}$$

3.2. Parameter estimation of GC prediction model

Now there are some methods used to compute *D* and *H* roughly. For instance, box dimension, root mean square method, and spectroscopy is used to get the value of *D*. But some methods of these have a slight error, such as spectroscopy. It is better to use the method of box dimension to obtain the fractal dimension. The periodic graph method, the variance method, rescaled range method, and the absolute value method is used to get the value of *H*. The rescaled range method is commonly used to estimate the parameter value of *H*. The specific solving processes are as [24].

There are still two parameter values, including drift coefficient and diffusion coefficient in the above formula, that have not been determined, so some parameter solving method is needed to obtain the parameter values. Given the time series *Y(t)*, the maximum likelihood estimation (MLE) method can be used to calculate the above two parameters. The specific steps are shown in [25].

Therefore, the flowchart of the procedure of GC prediction model is shown in Figure 5.

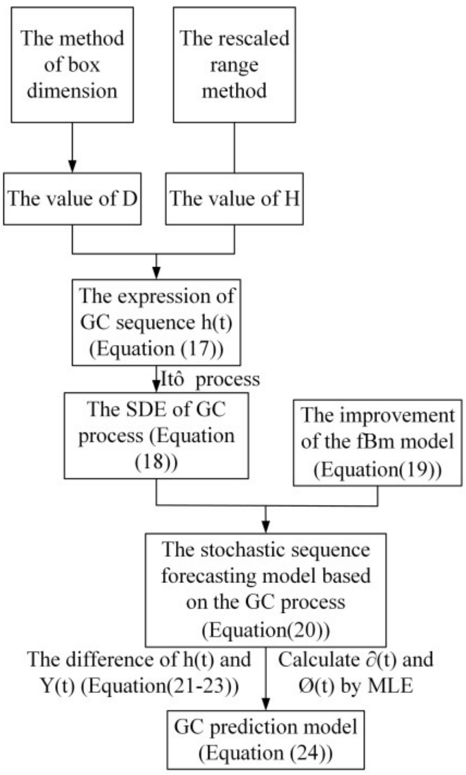

Figure 5. Flowchart of the procedure for generating the increment of GC sequence.

4. Cast study

4.1. *Prediction of power supply reliability based on GC prediction model*

To evaluate the predictive performance of the GC prediction model, we used the outage data of the Shanghai power grid in 2019 to predict and analyze the power-supply reliability indicator (ASAI) of the power grid. The formulas of ASAI are as Eqs (25) and (26). The total number of users in Shanghai in 2019 is 166,554, and the total number of hours in one year is 8760. It is easy to see that the ASAI for the same period is almost uniform because of the large base numbers. The calculation results have shown that the ASAI is about 99.9999% every day, and the values are very close to each other. If the ASAI is predicted directly, the prediction model must have a very high degree of accuracy, leading to significant error. However, according to the formula, if we get the sum of blackout time of users (SBTU), we can also calculate ASAI straightforwardly. Therefore, predicting ASAI by forecasting SBTU will make the results more accurate. According to statistics, the SBTU of every day in Shanghai in 2019 is shown in Figure 6.

$$AIHC = SBTU/8760 \tag{25}$$

$$ASAI = (1 - AIHC/166554) * 100\% \tag{26}$$

where the total number of customers is 166,554, the average customer outage time is named AIHC and the total number of hours in one year is 8760.

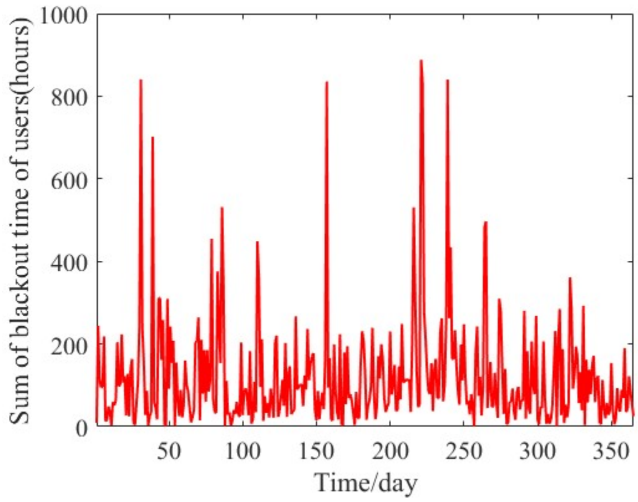

Figure 6. Daily blackout time of users in Shanghai in 2019.

As shown in Figure 6, the curve fluctuates wildly after the 220th day, which is an abnormal phenomenon. Suppose the GC prediction model can be used to make predictions for extreme situations. In that case, this shows the accuracy of the GC prediction model, which will also enlighten the power sector to improve the power supply reliability in the distribution systems. Therefore, this article uses the data of the first 220 days as the training set to predict the data of 10, 20, and 30 days. By calculating the *H* and *D* indicators of SBTU data in the first 220 days, the result is that *H* is 0.5341 and *D* is 1.4709.

It satisfies the application requirements of the GC prediction model. The historical data were brought into this model. MAE (mean absolute error), MAPE (mean absolute percentage error), MAXE (max absolute percentage error), RMSE (mean square root error), and ME (maximum error) generated by different forecasting days have been calculated. The results are shown in Table 1.

Table 1. The errors of different forecast days.

Prediction days	MAE	MAPE	RMSE	ME	MAXE
10	15.06	5.75	20.59	42.3	8.28
20	19.72	7.84	26.45	67.24	10.87
30	22.41	8.94	29.75	78.46	12.92

From the data in Table 1, it can be seen that when the number of prediction days is 10 days, the prediction effect of the GC prediction model is better. The prediction result is shown in Figure 7.

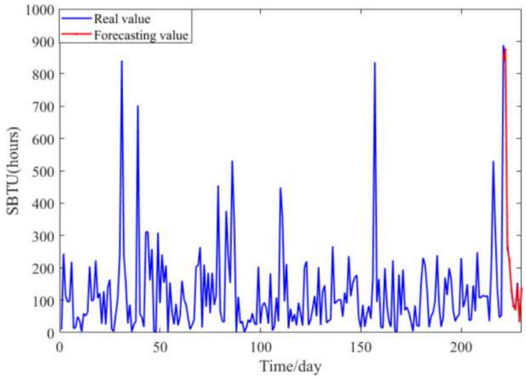

Figure 7. The prediction effects of GC.

We compared this method with LSTM, DANN, and fBM. The principle of these models is shown below. LSTM is a special RNN (recurrent neural network), which is different from adding three structures to RNN: forget gate, input gate, and output gate. It is proposed to solve the problem of gradient disappearance and gradient explosion in the long sequence training process. The basic principle of DANN is as follow. The barrier and Lagrange functions with neural network are integrated to obtain this method. The annealing technique is used to control the barrier parameter and the Lagrange function and neural network are applied to control the solution parameters. However, given the common feature of annealing algorithms, the performance of DANN partly depends on the selected parameters. Thus, its prediction effect is not stable. fBM contains the Hurst exponent, which is a parameter that describes the LRD, so fBM has LRD characteristics. The principle of this model is to use the drift term to describe the trend term of the degradation process and use the fBM-driven diffusion term to describe the randomness and LRD of the degradation process.

The prediction results of the SBTU are shown in Figure 8. MAE, MAPE, RMSE, and ME generated by different forecasting methods have been calculated. The results are shown in Table 2. It can be seen that the prediction accuracy of the GC prediction model is significantly higher than that of LSTM, fBM and DANN. These models have some drawbacks compared to GC prediction model.

Table 2. The errors of different forecast models.

Prediction model	MAE	MAPE	RMSE	ME	MAXE
GC	15.06	5.75	20.59	42.3	8.28
fBM	21.54	8.15	25.72	59.46	11.57
LSTM	41.96	15.81	32.76	108.72	85.47
DANN	35.51	22.61	37.69	52.29	40.10

LSTM does not fully take advantage of the LRD of random sequences. And the GC prediction model fully considers this property, so its prediction accuracy is higher than LSTM. Besides, because there are four fully connected layers in each LSTM cell tuple, if the network is profound or the time is considerable, it requires a lot of training data and training time. The optimal values of DBNN model parameters cannot be calculated theoretically, but can only be obtained from experience and experiments. Therefore, the calculation of this method is very cumbersome, and its prediction accuracy is lower than that of GC prediction model. fBM has a linear relationship between Hurst exponent and fractal dimension, so only one parameter is used to describe the LRD characteristics of random sequences. Compared with the fBM model, the advantage of the GC model is that the Hurst exponent and fractal dimension are independent parameters, so the model has higher flexibility and accuracy.

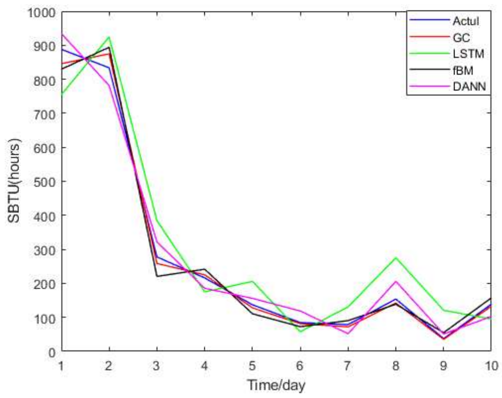

Figure 8. The prediction effects of different models.

4.2. *Analysis in influencing factors of power-supply reliability based on GRA*

By comparing the prediction results of the GC model with that of LSTM, DANN, and fBM in Figure 8, it can be seen that the prediction effect of the GC prediction model is much better than them. By bringing the prediction results of SBTU into Eqs (25) and (26), the prediction data of ASAI can be obtained. The results are shown in Figure 9.

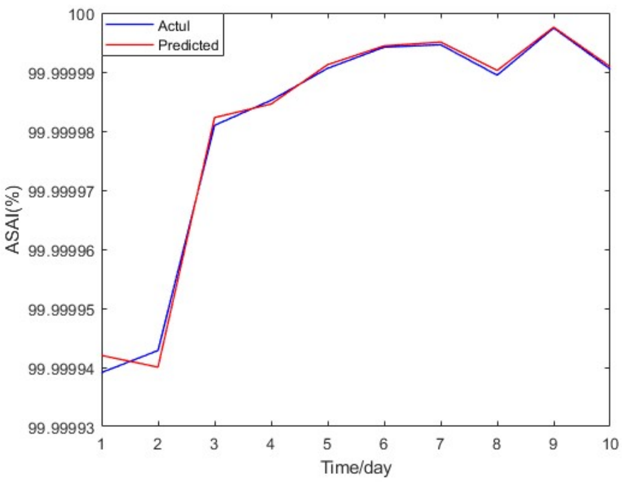

Figure 9. The prediction results of ASAI.

As mentioned above, we have predicted the changing trend of ASAI. Suppose we can get the significant factors affecting ASAI in the future. In that case, it will be of great significance for regulating the distribution network and improving power-supply reliability. According to the classification of responsibility causes in power industry standard DLT836.2-2016, the fault causes can be divided into these categories: product quality, equipment aging, operation maintenance, vehicle damage, animal reason, foreign-body short circuit, construction influence, other external factors, failure of power transmission and transformation facilities, climate reasons and user impact. According to statistics, SBTU is caused by different causes, as shown in Figure 10.

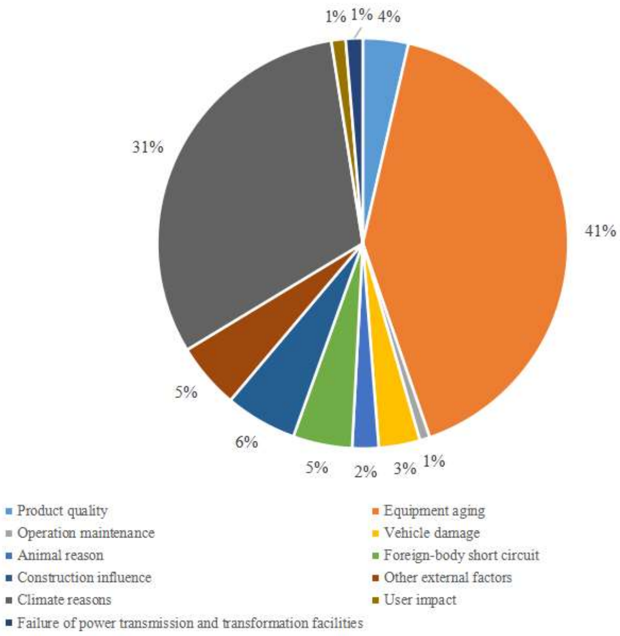

Figure 10. SBTU for different factors.

According to the data in this figure, the factors that have a significant impact on power-supply reliability can be selected from many influencing factors, which are distributed as follows: product quality, equipment aging, foreign-body short circuit, construction influence, other external factors, and climate reasons. This paper uses the GC prediction model to predict SBTU for these reasons. GRA is used to analyze SBTU and ASAI data obtained above to determine the main influencing factors of future power-supply reliability. By calculating the *H* and *D* values of SBTU data for different reasons in the first 220 days, the results are shown in Table 3.

Table 3. H value and D value under different factors.

Factors \ Indicator	*H*	*D*
Product quality	0.5231	1.4303
Other external factors	0.5743	1.3983
Climate reasons	0.5378	1.3814
Equipment aging	0.5168	1.4253
Foreign-body	0.638	1.3854
Construction influence	0.5013	1.3861

It can be seen from Table 3 that both *H* and *D* under different factors meet the requirements of the GC prediction model. Bring the historical data of SBTU under different factors into the model, and the prediction results are shown in Figure 11. We analyze the relationship between the prediction results of ASAI and SBTU under different outage factors by GRA, so as to obtain the correlation coefficient between different outage factors and ASAI. The prediction results and correlation coefficients of different factors are shown in Table 4. The essence of GRA is to consider the correlation degree of changes between curves of different numbers. The prediction results in Figure 11 are mean normalized, and the results are shown in Figure 12.

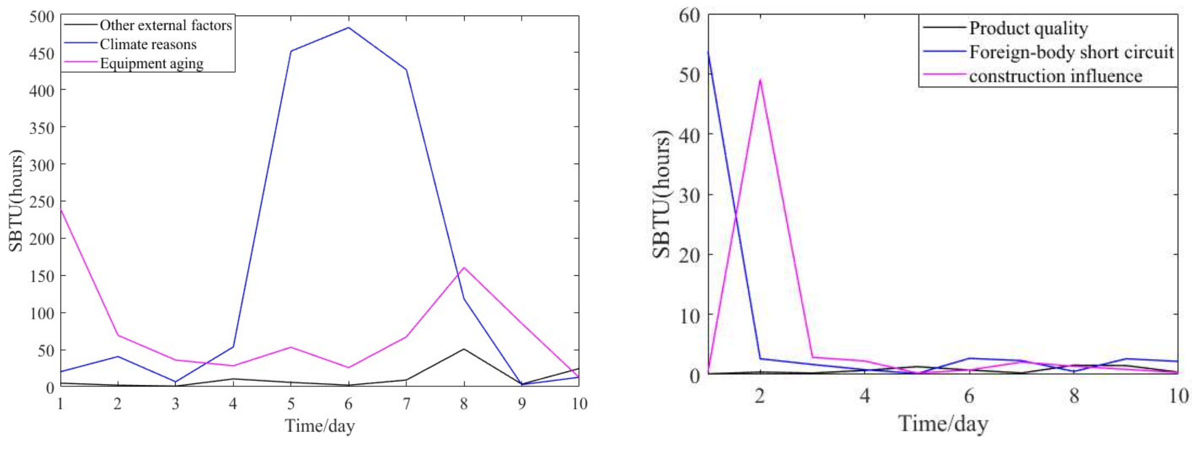

Figure 11. SBTU prediction results under different factors in a state of great fluctuation.

According to the calculation results of correlation the coefficient in Table 4, the correlation coefficient of predicted main factors affecting ASAI is sorted from large to small: equipment aging, foreign-body short circuit, other external factors, construction influence, climate reasons, and product quality. Furthermore, the correlation coefficient between actual data ASAI and SBTU with different outage causes is calculated in Table 5. It can be seen that the actual main factors affecting ASAI are sorted from large to small: equipment aging, other external factors, foreign-body short circuits, construction influence, product quality, and climate reasons.

Table 4. The correlation coefficient between predicted values of ASAI and different factors.

	ASAI (%)	Product quality	Other external factors	Climate reasons	Equipment aging	Foreign-body short circuit	Construction influence
1	99.99994203	0.1	4.79	20.16	240.01	53.73	0.3
2	99.99994005	0.4	2.09	40.65	69.25	2.62	49.05
3	99.99998228	0.24	0.81	6.73	35.93	1.67	2.86
4	99.99998457	0.68	10.57	53.65	28.21	0.8	2.24
5	99.99999124	1.3	5.89	451.84	53.17	0.16	0.24
6	99.99999442	0.75	2.11	483.57	25.71	2.69	0.75
7	99.99999505	0.25	9.04	426.85	67.15	2.31	2.07
8	99.99999026	1.55	50.82	118.21	160.51	0.51	1.36
9	99.99999758	1.48	3.56	2.85	85.32	2.62	0.85
10	99.9999909	0.38	24.56	12.82	12.2	2.18	0.32
correlation coefficients		0.7597	0.9165	0.7808	0.9607	0.9498	0.7813

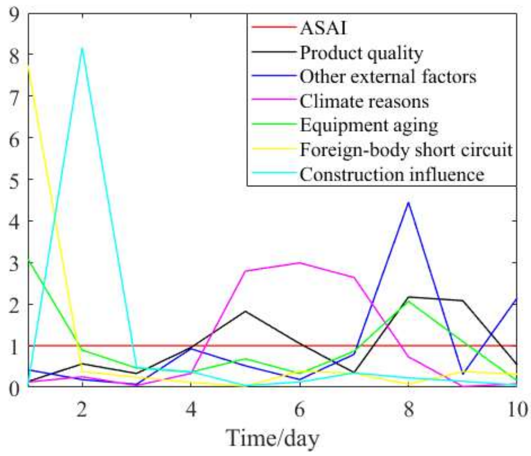

Figure 12. Normalized results of data.

Table 5. Correlation coefficient between predicted values of ASAI and different factors.

Failure factors	Correlation coefficients
Product quality	0.918
Other external factors	0.9254
Climate reasons	0.7652
Equipment aging	0.9534
Foreign-body short circuit	0.9234
Construction influence	0.9231

It can be seen from the results that the top three factors affecting reliability obtained by using forecast data are the same as the conclusion calculated from the actual data. Therefore, it can be considered that it is feasible to find the main influencing factors of ASAI in the future through the GRA of the predicted values.

5. Conclusions

This paper describes the GC prediction model with LRD features and GRA and applies it to the reliability prediction field of distribution networks. Based on the case study results, this model is used to predict ASAI with good accuracy of the prediction results. In addition, it is used to predict the outage time for different reasons and combined with the power supply reliability index for GRA, which can effectively predict the main factors affecting power supply reliability in the future. This method is vital for power companies to improve the reliability of power supply in distribution networks.

It is worth to note that the primary purpose of this paper is to analyze the factors affecting the reliability of the power supply network by correlation analysis between the predicted data. However, the GC prediction model can be used as long as it is a random sequence that satisfies the LRD characteristics. For example, in industrial equipment, bearing failure is a common factor leading to a generator failure, so the model can simulate the degradation trend of bearings and thus predict the time of generator failure. In the field of energy and power, wind power generation is related to real-time wind speed. The time series of wind speed is stochastic, so a GC prediction model can predict the wind speed and thus obtain the power generation. From the above cases, we can find that the GC prediction model is also widely used in other fields.

Acknowledgments

The work was supported by major project in Ministry of Science and Technology of the People's Republic of China, Grand number: 2020AAA0109301. The name of this major project is Science and technology innovation 2030 "new generation of AI".

Conflict of interest

The authors declare there is no conflict of interest.

References

1. M. Doostan, R. Sohrabi, B. Chowdhury, A data-driven approach for predicting vegetation-related outages in power distribution systems, *Int. Trans. Electr. Energy Syst.*, **30** (2020), e12154. https://doi.org/10.1002/2050-7038.12154
2. R. Tervo, J. Karjalainen, A. Jung, Short-term prediction of electricity outages caused by convective storms, *IEEE Trans. Geosci. Remote Sens.*, **57** (2019), 8618–8626. https://doi.org/10.1109/TGRS.2019.2921809
3. M. Doostan, B. Chowdhury, Predicting lightning-related outages in power distribution systems: A statistical approach, *IEEE Access*, **8** (2020), 84541–84550. https://doi.org/10.1109/ACCESS.2020.2991923

4. W. Li, J. Zhou, K. Xie, X. Xiong, Power system risk assessment using a hybrid method of fuzzy set and monte carlo simulation, *IEEE Trans. Power Syst.*, **23** (2008), 336–343. https://doi.org/10.1109/TPWRS.2008.919201
5. P. Wang, B. Chen, C. Tian, B. Sun, M. Zhou, J. Yuan, A novel neutral electromagnetic hybrid flexible grounding method in distribution networks, *IEEE Trans. Power Delivery*, **32** (2016), 1350–1358. https://doi.org/10.1109/TPWRD.2016.2526054
6. M. A. Mahmoudi, M. Kharazmi, M. Rashidinejad, M. Iranmanesh, P. Aghaie, The effect of cooling loads management on electric power supply system of Kerman province by the year 2031, *Environ. Prog. Sustainable Energy*, **35** (2016), 1177–1189. https://doi.org/10.1002/ep.12302
7. Y. Xie, C. Li, Y. Lv, C. Yu, Predicting lightning outages of transmission lines using generalized regression neural network, *Appl. Soft Comput.*, **78** (2019), 438–446. https://doi.org/10.1016/j.asoc.2018.09.04
8. A. K. Rajeevan, P. V. Shouri, U. Nair, ARIMA based wind speed modeling for wind farm reliability analysis and cost estimation, *J. Electr. Eng. Technol.*, **11** (2016), 869–877. https://doi.org/10.5370/JEET. 2016. 11.4.869
9. X. Chen, J. Tang, Q. Chang, W. Li, A data-driven method for operational reliability prediction on electric devices considering multiple meteorological factors, in *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, 2018. doi: 10.1109/PMAPS.2018.8440321
10. M. Johansson, T. Olofsson, Bayesian model selection for Markov, hidden Markov, and multinomial models, *IEEE Signal Process Lett.*, 14 (2007), 129–132. https://doi.org/10.1109/LSP. 2006.882094
11. R. Zeineddine, Fluctuations of the power variation of fractional Brownian motion in Brownian time, *Bernoulli*, (2015), 760–780. https://doi.org/10.3150/13-BEJ586
12. F. Black, M. Scholes, The pricing of options and corporate liabilities, *J. Political Econ.*, **81** (1973). https://doi.org/10.1086/260062
13. X. T. Wang, W. Y. Qiu, F. Y. Ren, Option pricing of fractional version of the Black-Scholes model with Hurst exponent H being in (13, 12), *Chaos Solitons Fractals*, **12** (2001), 599–608. https://doi.org/10.1016/S0960-0779(00)00028-X
14. H. Konno, F. Watanabe, Maximum likelihood estimators for generalized Cauchy processes, *J. Math. Phys.*, **48** (2007), 103303. https://doi.org/10.1063/1.28 00162
15. D. Niu, S. Dai, A short-term load forecasting model with a modified particle swarm optimization algorithm and least squares support vector machine based on the denoising method of empirical mode decomposition and grey relational analysis, *Energies*, **10** (2017), 408. https://doi.org/10.3390/en10030408
16. D. Akay, F. E. Boran, M. Yilmaz, M. Atak, The evaluation of power plants investment alternatives with grey relational analysis approach for Turkey, *Energy Sources, Part B*, **8** (2013), 35–43. https://doi.org/10.1080/15567249.2010.493917
17. Z. Wu, Q. Gao, B. Jiang, H. R. Karimi, Solving the production transportation problem via a deterministic annealing neural network method, *Appl. Math. Comput.*, **411** (2021), 126518. https://doi.org/10.1016/j.amc.2021.126518
18. Z. Wu, H. R. Karimi, C. Dang, An approximation algorithm for graph partitioning via deterministic annealing neural network, *Neural Networks*, **117** (2019), 191–200. https://doi.org/10.1016/j.neunet. 2019.05.010

19. L. Yan, G. Shen, K. He, Itô's formula for a sub-fractional Brownian motion, *Commun. Stochastic Anal.*, **5** (2011). https://doi.org/10.31390/cosa.5. 1.09
20. H. Liu, W. Song, Y. Zhang, A. Kudreyko, Generalized Cauchy degradation model with long-range dependence and maximum Lyapunov exponent for remaining useful life, *IEEE Trans. Instrum. Meas.*, **70** (2021). https://doi.org/70(4):3512812
21. S. Duan, W. Song, E. Zio, C. Cattani, M. Li, Product technical life prediction based on multi-modes and fractional Lévy stable motion, *Mech. Syst. Sig. Process.*, **161** (2021), 107974. https://doi.org/ 10.1016/j.ymssp.2021.107974
22. H. Liu, W. Song, E. Zio, Fractional Lévy stable motion with LRD for RUL and reliability analysis of li-ion battery, *ISA Trans.*, 2021. https://doi.org/10.1016/j.isatra.2021.07.002
23. W. Song, H. Liu, E. Zio, Long-range dependence and heavy tail characteristics for remaining useful life prediction in rolling bearing degradation, *Appl. Math. Modell.*, **102** (2022), 268–284. https://doi.org/10.1016/j.apm.2021.09.041
24. H. Liu, W. Song, E. Zio, Metabolism and difference iterative forecasting model based on long-range dependent and grey for gearbox reliability, *ISA Trans.*, **122** (2022), 486–500. https://doi.org/10.1016/j.isatra.2021.05.002
25. R. E. Carrillo, T. C. Aysal, K. E. Barner, A generalized Cauchy distribution framework for problems requiring robust behavior, *EURASIP J. Adv. Signal Process.*, **2010** (2010), 1–19. https://doi.org/10.1155/2010/312989