ISSN: 2414-1895 DOI: 10.6919/ICJE.202312_9(12).0038

Research on Accuracy Prediction of DC Charging Pile Measureme -nt based on Over Threshold Distribution

Shupo Bu, Xiacheng Wang, Chen Tang, Jin Chen, Qianhua Yu, Yinlong Fan

Kunshan Dengyun College of Science and Technology, Suzhou 215300, China

Abstract

The relationship between MTTF and failure rate is analyzed.TheWeibull distribution failure rate is used to model the bathtub curve.The parameters of the Weibull distribution are estimated by the least squares method and the correlation coefficient method.The distribution curve of the reliability is obtained.The life table of the energy meter provides the theoretical basis.

Keywords

GEV; GPD; Over Threshold Distribution; DC Charging Post.

1. Introduction

With the deepening of electric vehicles, measurement accuracy has become the main application scenario for electric vehicle charging facilities. DC charging stations have become the main charging method for public charging stations due to their advantages such as high power, fast charging, and the ability to directly charge power batteries. However, the harmonic components generated by the high-power nonlinear charging machine used in DC charging during the charging and discharging process, as well as the large number of ripple waves generated by pulsating charging, seriously affect the measurement accuracy of DC charging piles, causing operational errors to increase or even exceed the error limit, leading to problems such as trade settlement. Therefore, it is of great practical significance to study the measurement accuracy such as working error of the increasingly large number of DC charging piles.

Scholars have conducted extensive theoretical research and experimental demonstrations to address this issue. The currently recognized method is to rely on model optimization and reconstruction of synchronous sampling values of current and voltage, so as to control the measurement error of active energy within a small range, and ensure that the DC charging station meets the requirements of relevant standards.

Reference [1] developed a testing strategy based on weights for the factors that affect the operational errors of electric energy metering, but did not accurately predict the operational error data after the DC charging station was put into operation.

Reference[2] uses the generalized extreme value distribution to analyze the probability distribution of the maximum electric field value, but under limited sample size conditions, only analyzing the maximum value in the measured data will result in significant deviation due to the small number of samples.

During the operation of DC charging piles, the larger the error, the greater the impact on measurement accuracy. The smaller the error, the smaller the impact, and even negligible. Therefore, measurement accuracy can depend on the larger value of the error. However, in the process of statistical analysis, the random error distribution follows a symmetric normal distribution model. For asymmetric data that follows an interval exceeding a certain value but less than the maximum, the normal distribution cannot be fitted correctly. Therefore, while ignoring small error data, this article introduces extreme value theory to describe the characteristics of extreme values and larger values that exceed threshold data, providing a reliable basis for preventing the potential risks caused by error extreme values.

The Generalized Pareto Distribution (GPD) takes all extreme data exceeding a certain threshold as the research object, and describes the extreme value model of its probability distribution law. It can also obtain good results even with limited measured samples. It is a hot topic in extreme value theory research and is increasingly valued by researchers in fields such as engineering, commerce, and medicine.

This article proposes an optimization evaluation method for the measurement accuracy of DC charging piles based on the generalized Pareto distribution. The independent and identically distributed op- erating error sequence of DC charging piles is used as a random variable, and the generalized Pareto distribution in extreme value theory is used to model data that exceeds the error threshold. This meth- od effectively evaluates the operational risk situations that the system may encounter with significant impact. This method only relies on a limited sample, It has the characteristic of fast calculation, providing meaningful reference and guidance for the trade settlement between the power supply and consumption parties in the system.

2. Generalized Extreme Value Distribution (GEV)

Set up a random sample sequence with independent and identically distributed errors, remember do nindividual the maximum/minimum value of a random sample sequence [3], then the random variable the distribution is called the probability distribution of maximum/minimum error.

The distribution function of the generalized extreme value distribution (GEV) is:

$$G(x)=\exp\left\{-\left[1+\xi(\frac{x-\mu}{\sigma})\right]^{-1/\xi}\right\} \tag{1}$$

Where in, $1+\xi(x-\mu)/\sigma>0$, μ, σ, ξ they are positional parameters, scale parameters and shape parameters. The shape parameter ξ controls the shape of the GEV distribution curve to adapt to different error sample sequences. According to ξ different values, the GEV distribution can be divided in- to three categories, as shown in Table 1.

Table 1. Types of GEV

Types	ξ Price	Upper/Lower Bounds
GEV I type (Gumbel distribution)	=0	No upper bound No lower bound
GEV II type (Frechet distribution)	>0	No upper bound Bounded below
GEV III type (Weibull distribution)	<0	Having an upper bound No lower bound

According to Table 1, when ξ<0 and the number of random samples is large enough, the GEV III type is simplified as a Weibull distribution, which can be used to describe distributions with upper and lower bounds on extreme values. Figure 1 shows the density function of the extreme value distribution.

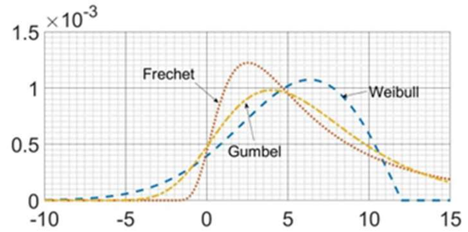

Figure 1. PDF of extremum distribution

3. POT Model and Generalization Pareto Distribution (GPD)

3.1 POT Model

POT The model, known as the Beyond Safety Threshold Model, is an important modeling method in extreme value theory, which models data in the sample data that exceeds a certain safety threshold.

Set $X_1, X_2, \cdots X_n$ as an independent identically distributed random variable, and the overall distribution function is $F(x)$, μ as distribution function $F(x)$ Upper endpoint of support, called threshold.When $X_k > \mu$, call $Y_k = X_k - \mu$ a sequence of exceeding thresholds called a threshold [4], also known as distribution $F(x)$ cloth tail data, Over threshold distribution function $F(y;\mu)$,It can be obtained from the conditional probability that:

$$\begin{aligned} F(y;\mu) &= P\{X-\mu \le y \mid X > \mu\} \\ &= \frac{F(\mu+y)-F(\mu)}{1-F(\mu)}, y > 0 \end{aligned} \tag{2}$$

The extreme value distribution function corresponding to the over threshold distribution is no longer a classical extreme value model.

3.2 Generalized Pareto Distribution (GPD)

According to Pickands limit theorem, when μ large enough, the over threshold distribution function for a certain class of distributions $F(y;\mu)$ Can be approximated as generalized Pareto distribution (GPD).

Set $X_1, X_2, \cdots X_n$ as independent and identically distributed error random sample sequence, $M_k = \max(X_1, X_2, \cdots X_n)$, then there is a certain threshold $\mu < M_k$, make all thresholds exceeded μ the sample approximately follows the generalized Pareto distribution, three parameters GPD(GP3) the distribution function and probability density function are:

$$F(x;\mu) \approx G(x;\mu,\xi,\sigma) = \begin{cases} 1-\left[1+\xi\dfrac{x-\mu}{\sigma}\right]^{-1/\xi}, \xi \ne 0 \\ 1-\exp\left[-\dfrac{x-\mu}{\sigma}\right], \xi = 0 \end{cases} \tag{3}$$

$$g(x;\mu,\xi,\sigma) = \begin{cases} \dfrac{1}{\sigma}\left(1+\xi\dfrac{x-\mu}{\sigma}\right)^{-1/\xi-1}, \xi \ne 0 \\ \dfrac{1}{\sigma}\exp\left(-\dfrac{x-\mu}{\sigma}\right), \xi = 0 \end{cases} \tag{4}$$

Volume 9 Issue 12, 2023
DOI: 10.6919/ICJE.202312_9(12).0038

wherein $\mu>0$, σ, ξ respectively refer to positional parameters, scale and shape parameters.When $\xi\le 0$, $\mu<x<\infty$; when $\xi>0$, $\mu<x<\mu+\sigma/\xi$; when $\mu=0$, GP3 degenerate to two parameters including shape and scale parameters GPD(GP2).

According to ξ different values, the generalized Pareto distribution can be divided into three categories, as shown in Table 2.

Table 2. Types of GPD

Types	ξ Price	Distribution Function $G(x)$	Probability Density Function $g(x)$
Pareto I type	$\xi=0$	$G_{\mathrm{I}}=\begin{cases}1-\exp(-\frac{x-\mu}{\sigma}),x\ge\mu\\ 0,x<\mu\end{cases}$	$g_{\mathrm{I}}=\frac{1}{\sigma}\exp(-\frac{x-\mu}{\sigma}),x\ge\mu$
Pareto II type	$\xi>0$ $\alpha=1/\xi$ $\alpha>0$	$G_{\mathrm{II}}=\begin{cases}1-\left(\frac{x-\mu}{\sigma}\right)^{-\alpha},x\ge\sigma+\mu\\ 0,\ x<\sigma+\mu\end{cases}$	$g_{\mathrm{II}}=\frac{\alpha}{\sigma}\left(\frac{x-\mu}{\sigma}\right)^{-\alpha-1},x\ge\mu+\sigma$
Pareto III type	$\xi<0$ $\alpha=-1/\xi$ $\alpha>0$	$G_{\mathrm{III}}=\begin{cases}1-\left(-\frac{x-\mu}{\sigma}\right)^{\alpha},\mu-\sigma\le x\le\mu\\ 1,\ x>\mu\end{cases}$	$g_{\mathrm{III}}=\frac{\alpha}{\sigma}\left(-\frac{x-\mu}{\sigma}\right)^{\alpha-1},\mu-\sigma\le x\le\mu$

When $\mu=0$, $\sigma=1$ is the standard Pareto distribution, G_{I} Exponential distribution, G_{II} Pareto distribution, Beta distribution.

Based on ξ the size of the values, corresponding evaluations can be made on the tail data of the extreme value distribution. The larger the size, the slower the convergence speed of the tail of the distribution, and the more severe the heavy-tailed phenomenon, indicating a higher number of sequences exceeding the threshold, as shown in the probability density function in Figure 2.

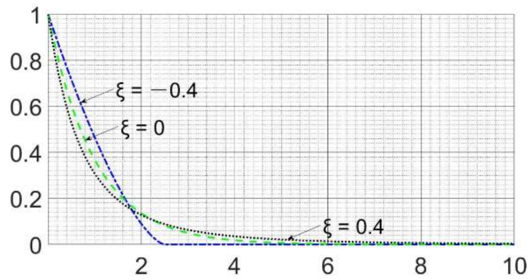

Figure 2. PDF of Pareto distribution

When ξ is a positive number, the distribution belongs to the heavy-tailed distribution; When ξ is a negative number, the distribution belongs to the light tailed distribution; When ξ is zero, the distribution is exponential.

According to Pickands' theorem, for any $\xi\in R$,distribution function $F(x;\mu)$ in $G(x;\mu,\xi,\sigma)$ the maximum attractive field, when the threshold μ approaches the right endpoint, there exists a positive real number $\sigma(\mu)$ [5].

$$\lim_{\mu\to x_F}\sup_{0<x<x_F-\mu}\left|F_\mu(x)-G(x;\mu,\xi,\sigma)\right|=0 \tag{5}$$

From the theorem, it can be concluded that in the maximum attractive field, the over threshold distribution can be approximately described by the generalized Pareto distribution, and its parameters are also estimated by the generalized Pareto distribution.

According to the POT model and the generalized Pareto distribution, μ for a given threshold $Y_k = X_k - \mu$, GP3 degenerates into Y_k a GP2 distribution that is a over threshold sequence. For this article, the error extremum specifically refers to random samples with errors greater than the threshold.

The generalized Pareto distribution can effectively utilize a certain number of extreme observations and estimate parameters for observations that exceed the threshold.

It should be pointed out that for the same random sample sequence, if the maximum value approximately follows the generalized extreme value distribution (GEV), then the excess approximately follows the generalized Pareto distribution (GPD) and has the same shape parameters.

4. Selection of Optimal Threshold for Exceeding Limit

The selection of the overrun threshold μ for the POT model, if the threshold is too large, the number of overrun sequences will decrease, leading to an increase in the variance of the parameter estimation. If the threshold is too small, the number of out of range sequences will increase, and the fitting effect on the tail of the data will deteriorate. It cannot guarantee the asymptotic nature of the POT model and the generalized Pareto distribution, and the parameter estimation is biased.

This article uses the sample average exceedance function graph to select the optimal exceedance threshold. According to the definition of conditional expectation, there are:

$$e(\mu) = E\left[X - \mu \middle| X > \mu\right] \tag{6}$$

It is called the average transfinite function of all samples with $e(\mu)$ random variables x exceeding the threshold μ.

$$e(\mu) = \frac{\sum_{i=1}^{N}(x_i - \mu)}{\sum_{i=1}^{N}(K_i)} = \frac{\sum_{i=1}^{N}(x_i - \mu)}{N_\mu}, \mu > 0 \tag{7}$$

Select a series of thresholds and construct a sample average exceedance function graph from a point set $(\mu, e(\mu))$. If a certain threshold μ_0, the excess distribution approximately follows the generalized Pareto distribution with parameter $\sigma(\mu_0), \xi$, then for values μ greater than μ_0, the sample average excess function fluctuates near a straight line, indicating that the selected threshold is appropriate.

Due to the need to draw a graph and determine whether the curve is a straight line for the transfinite mean function, it has great subjectivity. Therefore, this article uses the kurtosis coefficient method to supplement it.

The kurtosis coefficient method reflects the sharpness of an image, and the larger the kurtosis, the sharper the center point on the image. In the case of the same variance, the variance of most values in the middle is very small. In order to achieve the same goal as the normal distribution variance, some values must be further away from the center point, which is known as the heavytailed phenomenon. When the kurtosis value is greater than 3, it indicates that the distribution has a heavy tailed feature, when the kurtosis value is less than 3, it has a light tailed feature, and when the kurtosis value is equal to 3, it indicates a normal distribution.

Defined based on kurtosis values:

Volume 9 Issue 12, 2023
DOI: 10.6919/ICJE.202312_9(12).0038

$$Kurt[X] == \frac{E\left[\left(x_i - \bar{x}\right)^4\right]}{\left\{E\left[\left(x_i - \bar{x}\right)^2\right]\right\}^2} \tag{8}$$

Among them, $\bar{x}$ is the sample mean.

Criterion for kurtosis coefficient method: If the kurtosis value is greater than 3, the one with $|x_i - \bar{x}|$ the highest absolute value x_i will be removed from the sample sequence until the kurtosis value is less than 3. At this point, the threshold is the maximum value in the remaining sequence [6], which is:

$$\mu = \max\left(\left|x_{n-k}\right|\right), k \in (1,n) \text{ .}$$

5. Maximum Likelihood Estimation of Generalized Pareto Distribution

After the threshold is determined, the logarithmic likelihood function for the error exceeding the limit $y_i = x_i - \mu = x_i - x_{n-k}$ of the DC charging station following the generalized Pareto distribution is:

$$\ln L(y;\xi,\sigma) = \begin{cases} -k\ln\sigma - \sum_{i=1}^{k}\frac{y_i}{\sigma}, \xi = 0 \\ -k\ln\sigma - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{k}\log\left(1+\frac{\xi y_i}{\sigma}\right), \xi \neq 0 \end{cases} \tag{9}$$

When $\xi = 0$, let $\frac{\ln L(\sigma; y,\xi)}{d\sigma} = \frac{-k}{\sigma} + \left(\frac{1}{\sigma}\right)^2 \sum_{i=1}^{k} y_i = 0$ Decipher: $\hat{\sigma} = \frac{1}{k}\sum_{i=1}^{k} y_i$.

$$\text{When } \xi \neq 0, \begin{cases} \frac{\partial \ln L(\xi; y_i,\sigma)}{\partial \xi} = \sum_{i=1}^{k}\left[\left(\frac{1}{\xi}\right)^2 \frac{\left(\xi y_i + \sigma\right)\ln\left(1+\frac{\xi y_i}{\sigma}\right) - \xi\left(1+\xi\right) y_i}{\xi y_i + \sigma}\right] = 0 \\ \frac{\partial \ln L(\sigma; y_i,\xi)}{\partial \sigma} = \sum_{i=1}^{k}\left[\frac{y_i - \sigma}{\sigma\left(\sigma + \xi y_i\right)}\right] = 0 \end{cases} \tag{10}$$

After solving, the estimated values of the parameters can be obtained.

6. Upper Limit Estimation of Error Extremum

6.1 Upper Limit Estimation of Error Extremum

The generalized Pareto distribution and the generalized extreme value distribution have the same shape parameters, and the error distribution of the DC charging station is within a limited range between the threshold and the extreme value. Therefore, the shape parameter must be less than zero.

According to the generalized extreme value distribution type in Table 1, the extreme values of the sample sequence meet the GEV III type distribution conditions, and there is an upper bound on the extreme values, namely:

$$\begin{cases} x = u - \tilde{\sigma}/\xi \\ \tilde{\sigma} = \sigma + \xi(u - \mu) \end{cases} \tag{11}$$

6.2 Estimation of Hazard Rate Function

If the distribution function of a random variable X is $F(x)$ and the probability density function is $f(x)$, then the hazard rate function of the generalized Pareto distribution is:

$$\lambda(x) = \frac{f(x)}{1 - F(x)} \tag{12}$$

The hazard rate function is a monotonic function x, when $\xi < 0$, $\lambda(x)$ monotonically decreasing; when $\xi = 0$, $\lambda(x)$ was a constant; when $\xi > 0$, $\lambda(x)$ monotonically increasing.

The meaning of the hazard rate function referred to in this article is the probability that the operating error of the DC charging station will develop in the direction of exceeding the threshold μ when the error random variable sequence is greater than the threshold μ. Obviously, the lower the danger rate, the higher the measurement accuracy of the DC charging station, and vice versa, the poorer the measurement accuracy.

7. Example Analysis

In order to verify the effectiveness of the method proposed in this article, all detection data were obtained through on-site detection equipment. Based on the data of 459 DC charging stations on site, the collected data includes: working error, indication error, and payment amount error. The error warning indicators are developed using the method described in this article. To simplify the calculation, this article only takes the working error as an example.

To ensure the accuracy of threshold selection, the average exceedance function graph is used to determine the threshold range, and the kurtosis coefficient method is used to supplement it.

According to formulas (6) and (7), extract the detection data of the working error (positive value) of the on-site charging station to calculate the mean value of the exceeding limit. The function diagram of the average exceeding amount is shown in Figure 3.

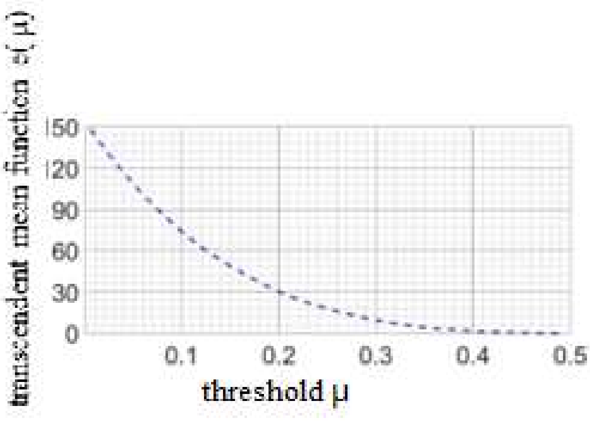

Figure 3. Over limit mean graph

From Figure 2, it can be seen that the threshold range of the working error (positive value) is approximately linear between regions [0.3,0.5]. Enlarge this interval as shown in Figure 4. After observation, it is determined that the threshold is selected within $e(\mu)$ and μ the linearly changing interval [0.41,0.48]. However, the optimal threshold interval still has a certain degree of subjectivity, and further judgment of the threshold needs to be made using the kurtosis coefficient method.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202312_9(12).0038

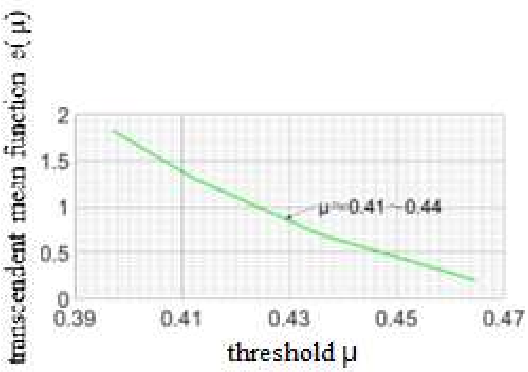

Figure 4. Enlarge the over limit mean

Based on the kurtosis coefficient method to determine accuracy and the actual situation of the sample data, the threshold for working error (positive value) is selected as 0.4878, and the threshold for working error (negative value) is selected as 1.1454.

The maximum likelihood method is used to estimate the shape and scale parameters of the GP2 distribution, and according to formula (3), the distribution function of the working error can be obtained as:

$$G(y;\xi,\sigma)=\begin{cases}1-\left[1-0.39498\dfrac{y}{0.21501}\right]^{1/0.39498}, & y>0\\ 1-\left[1-0.59713\dfrac{|y|}{0.69137}\right]^{1/0.59713}, & y<0\end{cases} \tag{13}$$

The distribution function is shown in Figure 5, and the specific parameters are shown in Table 3.

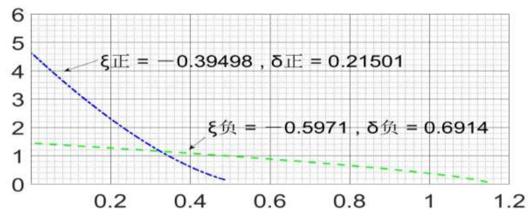

Figure 5. Generalized Pareto distribution of working error

When the shape parameter of the working error (positive value) is within the 95% confidence interval of [-0.5461, -0.2439], it belongs to the light tailed distribution, so the working error (positive value) has a limited right endpoint, which means there is an upper limit. Figure 6 shows the hazard rate function of the generalized Pareto distribution under different threshold conditions.

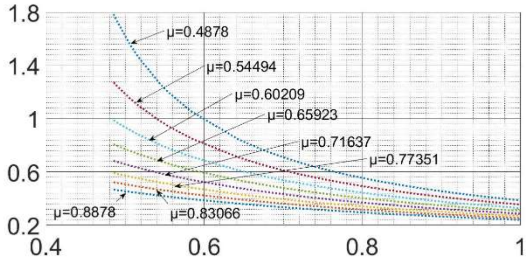

Figure 6. Hazard rate functions with different thresholds

ISSN: 2414-1895 DOI: 10.6919/ICJE.202312_9(12).0038

It can be seen From Figure 6,when $\xi < 0$, as the threshold increased, the risk rate gradually decreased, but the downward trend gradually slowed down. When the operating error (positive value) of the DC charging pile follows the generalized Pareto distribution and belongs to the light tailed distribution, individual larger errors have little impact on the overall measurement accuracy of the batch of charg- ing piles. However, compared to a certain charging station sample individual, if it is greater than the upper limit of GEV III extreme value, it should be taken seriously.

8. Conclusion

The results indicate that the error data of DC charging stations can be well fitted using the generalized Pareto distribution, and the probability of exceeding the threshold risk is given, which adds a new approach to determining the accuracy prediction of DC charging station measurement.

References

[1] Boundary value and extremum problems for generalized Oberbeck–Boussinesq model[J]. Sibirskie Èlektronnye Matematicheskie Izvestiya [Siberian Electronic Mathematical Reports],2019,16.

[2] Robson W.S. Pessoa,Felipe Mendes,Tiago Roux Oliveira,Karla Oliveira-Esquerre,Miroslav Krstic. Numerical optimization based on generalized extremum seeking for fast methane production by a modified ADM1[J]. Journal of Process Control,2019,84.

[3] Fanghua Lin. Extremum problems of Laplacian eigenvalues and generalized Polya conjecture[J]. Chinese Annals of Mathematics, Series B,2017,38(2).

[4] Mathias Raschke. Parameter Estimation for the Tail Distribution of a Random Sequence[J]. Communications in Statistics - Simulation and Computation,2013,42(5).

[5] Krzysztof Dębicki,Enkelejd Hashorva. Approximation of Supremum of Max-Stable Stationary Processes & Pickands Constants[J]. Journal of Theoretical Probability,2020,33(1).

[6] V. Evstigneev,G. V. Milyutin. Effect of the energy spectrum of an electron source on the attraction of the electrons into the accelerating mode in a betatron[J]. Soviet Physics Journal,1975,17(6).