machines

MDPI

Article

Investigating the Distribution of Flatness Measurements in Battery Manufacturing through Empirical Investigation and Statistical Theory

Hangyu Li [1,2] and Sun Jin [2,*]

[1] Department of Mathematical Sciences, Shanghai Jiao Tong University, Shanghai 200240, China; lihangyu@sjtu.edu.cn
[2] Institute of Manufacturing for Thin-walled Structures, Shanghai Jiao Tong University, Shanghai 200240, China
* Correspondence: jinsun@sjtu.edu.cn

Abstract: The battery is an important part of the new energy electric vehicle, and the control of the flatness of its side plate/bottom plate is the key to quality improvement in mass production. However, there are few pieces of research on the flatness distribution form at present, and the distribution form is often assumed to be a normal distribution, which leads to a significant deviation between the tolerance design and quality control of the flatness and the reality. This paper establishes a statistical model of flatness distribution, its theoretical distribution form is deduced as a normal range distribution, and then the experimental data of the flatness distribution are collected to verify this conclusion. Determining the flatness distribution form has practical effects on improving manufacturing quality and reducing costs in battery manufacturing.

Keywords: flatness distribution; statistical model; normal range distribution; KS test

Citation: Li, H.; Jin, S. Investigating the Distribution of Flatness Measurements in Battery Manufacturing through Empirical Investigation and Statistical Theory. *Machines* **2023**, *11*, 723. https://doi.org/10.3390/machines11070723

Academic Editor: Jie Liu

Received: 31 May 2023
Revised: 27 June 2023
Accepted: 4 July 2023
Published: 8 July 2023

1. Introduction

The widespread adoption of electric vehicles (EVs) has emerged as a prominent trend within the automotive industry and is primarily propelled by the commitment of numerous nations to achieve net-zero emissions by 2050 without increasing atmospheric carbon concentrations [1]. This global shift towards EVs has been accompanied by escalating sales figures and more stringent governmental directives, driving an increase in battery production capacity on a global scale. Batteries constitute up to 30% of an EV's weight [2] and contribute to approximately 30% of the overall cost [3]. Consequently, optimizing battery manufacturing quality to align with economic objectives, such as reduced reduction, increased yield, and minimized scrap rates, becomes critical. These have a substantial influence on the overall impact and viability of EVs.

The packaging of batteries is decisive in determining key performance parameters such as lifetime, cyclability, ruggedness, safety, and cost. At the core of battery packaging technology lies the electrochemical cell, which represents the smallest unit of a battery. Packaging tailors the battery to specific applications, encompassing aspects such as sealing, form factor, temperature and charge monitoring, and overall management. In the case of high-voltage or high-capacity EV batteries, multiple electrochemical cells may be interconnected in parallel or series, or both. Furthermore, these cells are integrated into modules that monitor individual cells and control the temperature. The modules are subsequently assembled to form a complete battery pack (Figure 1).

The flatness of the side plate and the bottom plate is important as it directly impacts the successful integration of the battery into the module and, consequently, the module into the battery pack. Hence, flatness represents one of the most crucial and widely considered geometric tolerances that determine the quality of the products. The definition of flatness is the degree to which the surface of a measured plane or the median plane contains

all its elements. The definition of flatness tolerance is such an area, that is, two parallel planes, which can contain the median plane and the entire area of the measured plane. The concept of flatness pertains to a fundamental geometric primitive. The geometrical product specifications (GPS) standard describes the verification process for flatness as encompassing partitioning, extraction, association, and evaluation. The new GPS standard recognizes two types of reference planes: the Minimum Zone Reference Planes (MZPL) and the Least Squares Reference Planes (LSPL). Corresponding to these reference planes, two association methods for flatness error evaluation are used: the Minimum Zone Method (MZM) and the Least Squares Method (LSM) [4,5].

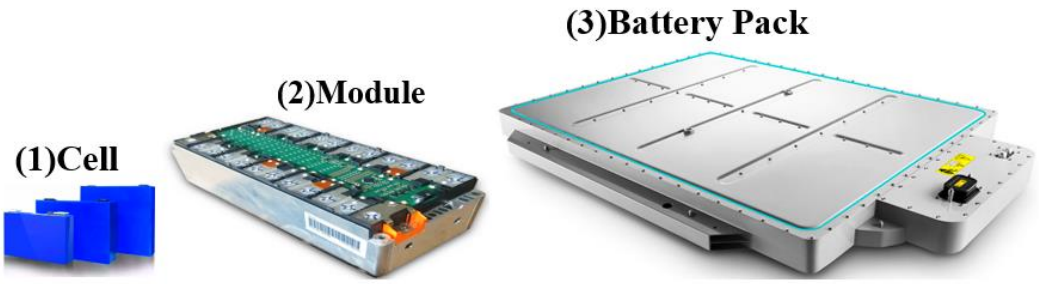

Figure 1. Illustrative example of packaging for a high-power automotive traction battery with a prismatic lithium-ion cell package (**1**). Multiple prismatic cells are combined to create modules (**2**), which are further integrated with a battery management system, thermal management systems, and electronic components to form a complete battery pack (**3**).

Several researchers have developed new techniques to improve the flatness tolerance evaluation. The literature provides insights into various techniques used to develop methodologies for obtaining minimum zone solutions. These techniques include nonlinear optimization, metaheuristics, and approximation based on linear programming [6–17]. Although a smaller area can be found eventually, the algorithm has a long computation time and is complex to implement. In modern equipment like coordinate measuring machines (CMM), the LSM is often used for flatness evaluation at present [18,19]. However, in battery manufacturing, the flatness of large planes, such as the side plates and end plates of the battery module, needs to be monitored closely, making it challenging to obtain the distribution characteristics of flatness in depth. Most classical quality monitoring techniques in parts manufacturing rely on assumptions regarding sample distribution. Hence, the theoretical distribution of flatness plays a crucial role in both design and manufacturing processes. Chelishcev et al. [20] used a CMM to measure the rotating parts and used the probability density function method to predict the size distribution of the measured parts and demonstrated the applicability of a distribution-free model in predicting the sample size, minimum content, and confidence level for CMM-based Geometric Dimensioning and Tolerancing inspection without the need for prior measurements. Berrado et al. [21] put forth a method that considers the measurement round-off and small sample sizes. They employed the Moran log spacing statistic to evaluate the suitability of a normal distribution in fitting the data. The results of their research suggest that the normal distribution adequately fits all the examined datasets. The study has demonstrated that the assumption of normality is adequately met for 20 industrial datasets encompassing four processes and three distinct material types. Ghie et al. [22] proposed an analysis method based on the Jacobian–Torsor model combined with the uncertainty of measurement and used a probability density function to simulate the distribution of part manufacturing sizes. Wang et al. [23] presented the actual measurement and statistical processing of error data to show that the distribution law of shape and position error obeys the folded normal process data (also known as the absolute normal distribution) under certain conditions. Chatterjee et al. [24] provided examples of situations commonly associated with folded normal distributions. These instances encompass measurements pertaining to flatness or

straightness, as well as the distance between two objects. A variety of planes are involved in the manufacture of battery modules/PACK, among which the plane is an important geometric element that constitutes mechanical parts [25]. It is often used as the datum plane for testing. To control the quality of the plane, it is necessary to measure the flatness. Investigate the distribution of flatness. And flatness also has an important application in mechanical design, such as in modern technologies like CAD/CAM/CAE [4].

The novelty of this paper is related to three aspects: Based on the definition of flatness and the principle of multivariate statistics, flatness statistics are constructed, and their theoretical distribution is deduced.

The theoretical distribution of flatness is validated against experimental flatness datasets using the maximum likelihood estimation (MLE) and Kolmogorov–Smirnov (KS) test.

The application scope of the distribution form of flatness is extended to large sample data acquired on a long plane. Then, the tolerances related to flatness are optimized.

The subsequent sections of this article are organized as follows: Section 2 introduces a statistical model of flatness. The theoretical distribution of flatness is deduced as a normal range distribution, and the normal range distribution function and density function are calculated using MLE and numerical integration. Section 3 describes the implementation of experiments, while Section 4 presents a comprehensive summary and discussion of the key points and findings presented in the article.

2. Materials and Methods

2.1. Statistical Model of Flatness

This investigation primarily focuses on the flatness distribution. A flowchart to study the form of the distribution is shown in Figure 2. Form tolerances typically define a designated area in which a consideration plane, line, axis, or center plane should be located. In accordance with the measurement standard [26], flatness tolerance specifies that all points on the surface must fall within the regions defined by two parallel planes. Specifically, the tolerance value represents the distance between these two parallel planes.

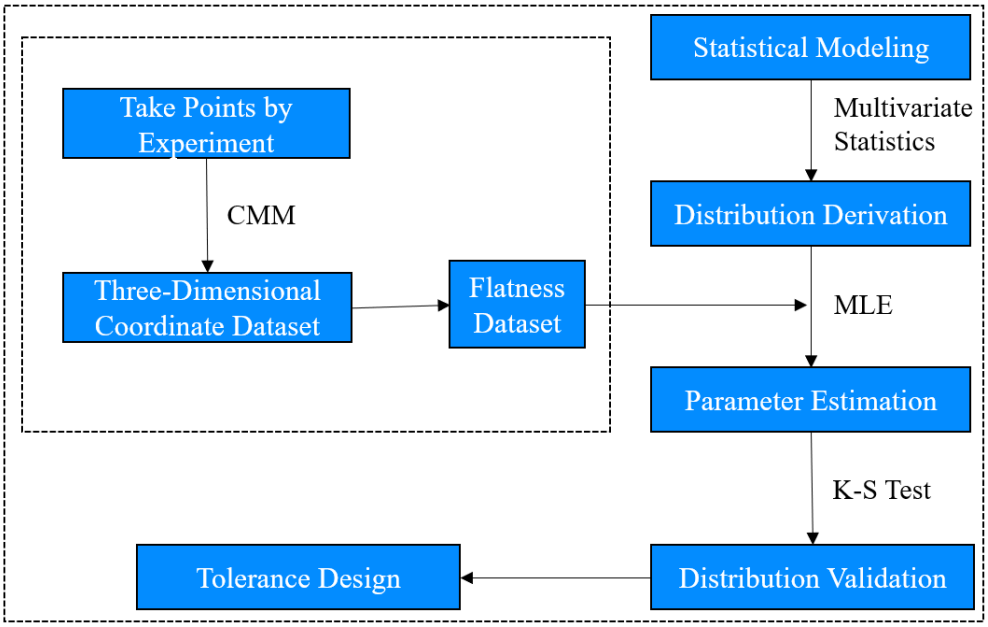

Figure 2. Flatness Distribution Study Flowchart.

In practical applications, a limited number of coordinate points on the plane are measured to determine the flatness value. It is necessary for all measurement points to fall within the specified area defined by the measurement criteria. The measurement of

flatness can be represented by the distance between two parallel planes that contain all the measurement points. One effective approach to meet this requirement is to fit the set of measurement points with a regression plane and calculate the flatness using the residual. Figure 3 provides a visual representation of this concept, depicting a cross-section of a panel. In the figure, the symbol L represents the region enclosing the fitting plane, and it indicates the distance between the parallel planes that enclose all the measurement points. The symbol h represents the distance between the measurement point and the fitting plane. The symbol d represents the vertical distance between the measurement point and the horizontal plane. The dots represent the set of points taken during measurement. The blue line represents the actual shape of the plane. The yellow line represents the fitted plane. Mathematically, it is feasible to evaluate the flatness distribution by the LSM. This method has excellent premise assumptions:

1. $z_i = \beta_0 + \beta_1 x_i + \beta_2 y_i + \varepsilon_i, i = 1, \ldots, m$; the independent variable is currently correlated with the dependent variable;
2. $\varepsilon_i \sim N(0, \sigma^2), i = 1, \ldots, m$; the mean value of the error ε_i is 0; the variance is fixed and obeys the normal distribution;
3. The errors are independent of each other; $\sum_{i=1}^{m} \varepsilon_i = 0$.

Therefore, to simplify the process of determining the flatness, we make the following assumptions: the fitting plane is fixed, and the measured points $X_i = (x_i, y_i, z_i)$ on the plane obey the three-dimensional normal distribution with a fixed mean value of 0 and a variance $\sigma^2 (X_i \sim N_3(\mu, \sum), \mu = (0,0,0)^T, \sum = \mathrm{diag}\{\sigma^2\})$.

As shown in Figure 3, h_i represents the distance from the ith point to the fitted plane (points are positive above the plane and negative below the plane). We consider the R (flatness) statistic of the form.

$$R = \max_{1 \leqslant i \leqslant m} \{h_i\} - \min_{1 \leqslant i \leqslant m} \{h_i\}$$

In the next section, we aim to derive the distribution form of the statistic R.

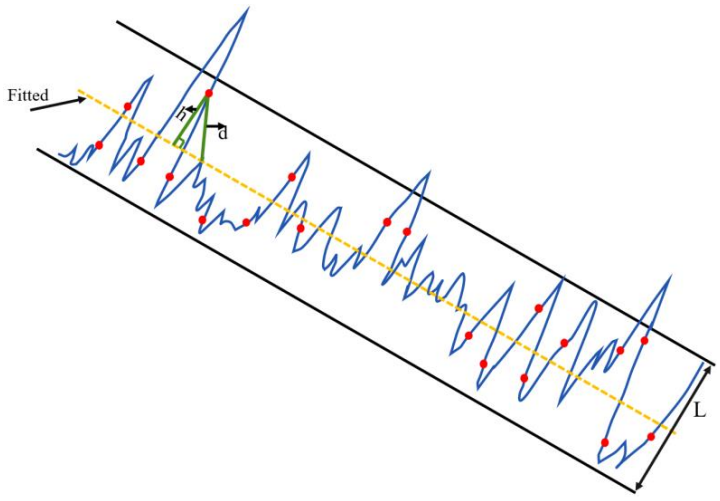

Figure 3. A cross-sectional diagram depicts the theoretical flatness value on a surface of a part, represented as 'L'.

2.2. Exact Distribution of R (Flatness)

This section shows that R statistic can be expressed as the range form of a multivariate independent normal distribution. Therefore, its exact distribution can be obtained easily from the result of Anderson [27] on multivariate statistics. Taking the points on the long plane is mainly performed by CMM, and the coordinates of each point are $(x_i, y_i, z_i), i = 1, \ldots, m$.

According to the model established in the previous section, let d_i be the vertical distance between the ith point and the plane. Geometrically, the difference between h_i and d_i is only a constant k (that is, the cosine of the angle between the fitting plane and the horizontal plane), where

$$H = \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_m \end{pmatrix} = \begin{pmatrix} k & 0 & \cdots & 0 \\ 0 & k & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & k \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix} = kD, \text{ where } K = \begin{pmatrix} k & 0 & \cdots & 0 \\ 0 & k & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & k \end{pmatrix}_{m \times m} \tag{1}$$

Due to the randomness of the points, H_i and D_i can be regarded as a random variable and h_i, d_i as the sample point. The joint density function of n-dimensional random variables $(H_1, \ldots, H_m)$, $(D_1, \ldots, D_m)$ are $f_H(h)$, $f_D(d)$, respectively. According to the Jacobian transformation of multivariate random variables, we have $f_H(h) = f_D\left(\frac{h}{k}\right) J(d \to h)$, where

$$J(d \to h) = \left| \frac{\partial d^T}{\partial h} \right| = |K|^{-1} = k^{-m} \tag{2}$$

Therefore, to simplify the derivation of the flatness distribution, it is only necessary to derive the distribution of the distance d between the point and the vertical direction of the fitting plane. According to the assumption that the error ε_i is a normal distribution that is independent and homoscedastic, it can be seen that D is an n-dimensional normal distribution given by

$$D = \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix} = \begin{pmatrix} z_1 - \beta_1 x_1 - \beta_2 y_1 - \beta_0 \\ z_2 - \beta_1 x_2 - \beta_2 y_2 - \beta_0 \\ \vdots \\ z_m - \beta_1 x_m - \beta_2 y_m - \beta_0 \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{pmatrix}, N_n \sim \left(0, \sum\right) \tag{3}$$

where

$$\sum = \begin{pmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{pmatrix}_{n \times n} \tag{4}$$

The joint density function of the n-dimensional normal distribution, denoted as D, can be expressed as the following:

$$f_D(d) = \frac{1}{(2\pi)^{m/2} \left| \sum \right|^{1/2}} \exp\left[-\frac{d^T \sum^{-1} d}{2} \right] \tag{5}$$

Then, the joint density function of the n-dimensional random variable H through a Jacobian transformation is

$$\begin{aligned} f_H(h) &= k^{-m} \frac{1}{(2\pi)^{m/2} \left| \sum \right|^{1/2}} \exp\left[-\tfrac{1}{2} K^{-1} h^T \sum^{-1} (Kh) \right] \\ &= \frac{1}{(2\pi)^{m/2} \left| k^2 \sum \right|^{1/2}} \exp\left[-\tfrac{1}{2} h^T k^2 \sum^{-1} (h) \right] \end{aligned} \tag{6}$$

Hence, $H \sim N_n\left(0, k^2 \sum\right)$, $H_1, \ldots, H_n$ are independent normal distributions with the same variance $\left(N\left(0, k^2 \sigma^2\right)\right)$. Let the distribution function be $F(h_i)$, and its density function be $f(h_i)$. Next, the distribution of flatness $R = \max\limits_{1 \leqslant i \leqslant m} \{h_i\} - \min\limits_{1 \leqslant i \leqslant m} \{h_i\}$ is solved by deriving the range distribution H.

Let H be the statistical population. Take a sample $h_1, \ldots, h_m$ and obtain the corresponding order statistic $h_{(1)} \le h_{(2)} \le \cdots \le h_{(m)}$, where $h_{(i)}$ is the ith order statistic ($i = 1, \ldots, m$). $h_{(1)} = \min(h_1, \ldots, h_m)$ is the minimum order statistic. $h_{(m)} = \max(h_1, \ldots, h_m)$ is the maximum order statistic. $R = h_{(m)} - h_{(1)}$ is the sample range. The joint density function of $h_{(1)}$ and $h_{(m)}$ is

$$f\left(h_{(1)}, h_{(m)}\right) = \frac{m!}{(m-2)!} f\left(h_{(1)}\right) f\left(h_{(m)}\right) \left[F\left(h_{(m)}\right) - F\left(h_{(1)}\right)\right]^{m-2}, \; h_{(1)} \le h_{(m)} \tag{7}$$

Let

$$\begin{cases} R = h_{(m)} - h_{(1)} \\ U = h_{(1)} \end{cases} \tag{8}$$

Then, its inverse transformation is

$$\begin{cases} h_{(1)} = U \\ h_{(m)} = U + R \end{cases} \tag{9}$$

Its Jacobian determinant $|J| = 1$. The joint density function of R and U is

$$f_{(R,U)}(r, u) = m(m-1) f(u) f(r+u) [F(r+u) - F(u)]^{m-2} \tag{10}$$

The distribution function $F_R(r)$ of the flatness R is

$$\begin{aligned} F_R(r) &= \int_0^r \int_{-\infty}^{+\infty} m(m-1) f(u) f(t+u) [F(t+u) - F(u)]^{m-2} du dt \\ &= \int_{-\infty}^{+\infty} \int_0^r m(m-1) f(u) f(t+u) [F(t+u) - F(u)]^{m-2} dt du \\ &= \int_{-\infty}^{+\infty} m(m-1) f(u) \int_0^r [F(t+u) - F(u)]^{m-2} d[F(t+u) - F(u)] du \\ &= \int_{-\infty}^{+\infty} [F(r+u) - F(u)]^{m-1} m f(u) du \\ &= m \int_{-\infty}^{+\infty} f(x) [F(r+x) - F(x)]^{m-1} dx \end{aligned} \tag{11}$$

On computation, the density function $f_R(r)$ of the flatness R is

$$f_R(r) = [F_R(r)]' = m(m-1) \int_{-\infty}^{+\infty} f(x) f(r+x) [F(r+x) - F(x)]^{m-2}, \; r > 0 \tag{12}$$

where

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp -\frac{x^2}{2\sigma_0^2}, \sigma_0 = k\sigma \tag{13}$$

$$F(x) = \int_{-\infty}^{x} f(t) dt \tag{14}$$

Figure 4 shows the density and distribution function plots of the normal range distribution under different m values.

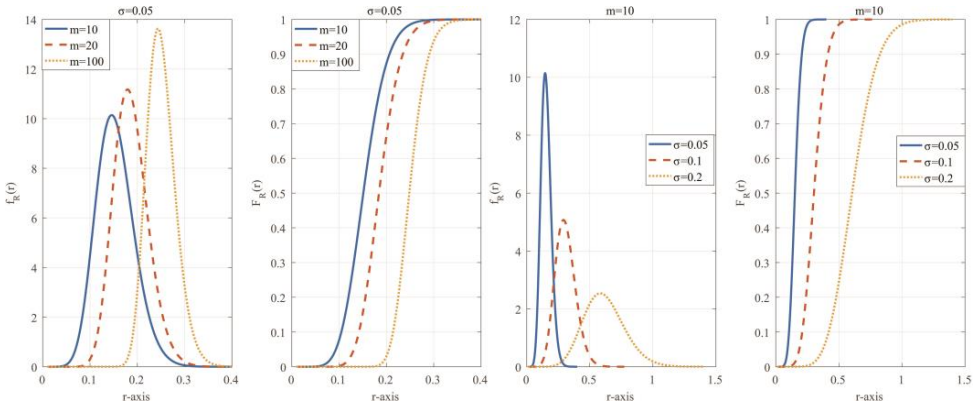

Figure 4. Density and distribution function plot of the normal range distribution.

2.3. Parameter Estimation of the Normal Range Distribution

The parameter σ_0 is estimated using the maximum likelihood estimation (MLE) method. Because the normal range distribution and density function are extraordinarily complex and have no analytical solution, a simplified approach is required to maximize the function. Many scholars have tried to solve the problem of seeking the optimal value of the objective function and achieved good results [28,29]. The grid search method is a well-established and uncomplicated approach for identifying extreme values, which can be used to solve extreme value problems [30]. It is a fundamental algorithm for optimizing parameters.

Let $r_i(i = i, \ldots, n)$ be n simple random samples from a normal range distribution, then the log-likelihood function based on these samples is

$$\log L(\sigma_0; r) = \sum_{i=1}^{n} \log(f_R(r_i)) \tag{15}$$

Subsequently, we need an estimate of σ_0, as

$$\widehat{\sigma_0} = \underset{\sigma}{\operatorname{argmax}} \left(\sum_{i=1}^{n} \log(f_R(r_i)) \right) \tag{16}$$

Since $f_R(r)$ has no analytical solution, it is necessary to use the Simpson integration method and truncate it first:

$$\begin{array}{ll} f_R(r) & = m(m-1) \int_{-a}^{a} f(x) f(x+r) [F(x+r) - F(x)]^{m-2} dx, \; r > 0 \\ & = \int_{-a}^{a} G(x) dx \end{array} \tag{17}$$

There is a truncation error.

$$\begin{array}{ll} \Delta_1 & = m(m-1) \int_{a}^{+\infty} f(x) f(x+r) [F(x+r) - F(x)]^{m-2} dx \\ & + m(m-1) \int_{-\infty}^{-a} f(x) f(x+r) [F(x+r) - F(x)]^{m-2} dx \\ & \leq m(m-1) \int_{a}^{+\infty} f(x) f(x+r) dx + m(m-1) \int_{-\infty}^{-a} f(x) f(x+r) dx \\ & = m(m-1) \frac{\sqrt{(2)}}{2} \sigma_0 \exp\left(-\frac{r^2}{4\sigma_0} \right) \end{array} \tag{18}$$

According to the precondition, since the plane shows relatively small fluctuations in the vertical direction, the standard deviation of its values along the Z axis is much smaller than that in the X and Y axes and is generally between 0.001 and 0.1. Moreover, the inclination angle between the fitted plane and the actual plane is relatively small, the k value is close to 1, and the flatness value is generally not greater than 1. The number of points m on the plane generally does not exceed 100. Let $a = 10$, $k = 1$, $m = 100$, so the truncation error

$$\Delta_1 \leq 100 \times 99 \times 9.9818 \times 10^{-12}$$

When using numerical integration for approximate calculations, such as $\max\left|f_R^{(4)}(r)\right| \leq M$, we have the computational errors $R_n[f_R] \leq \frac{2a}{2880}\left(\frac{2a}{N}\right)^4 M$ obtained by ignoring the truncation error and calculation errors in numerical integration. The Simpson integration formula is

$$f_R(r) \approx \int_{-a}^{a} G(x) dx \approx \frac{l}{3}(G_0 + 4G_1 + G_2) \tag{19}$$

The extreme value solution of (16) can be obtained using the grid method to obtain the MLE σ_0. The estimation performance of model parameters is evaluated using the root mean squared error (RMSE). As depicted in Table 1, the RMSE values are minimal, and the

parameters estimated by MLE are accurate (one hundred simulations were conducted with different values of m and n).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{\sigma_0} - \sigma_0\right)^2} \tag{20}$$

Table 1. Parameter estimation of the normal range distribution.

		m = 10		**m = 20**		**m = 30**	
σ_0	n	**Mean**	**RMSE**	**Mean**	**RMSE**	**Mean**	**RMSE**
0.01	50	0.00997	0.00034	0.01003	0.00032	0.00999	0.00025
	100	0.01002	0.00025	0.00998	0.00019	0.00998	0.00015
	500	0.00999	0.00012	0.01001	0.00008	0.00999	0.00009
	1000	0.01001	0.00009	0.01000	0.00006	0.01094	0.00094
0.05	50	0.04975	0.00171	0.05009	0.00133	0.04995	0.00128
	100	0.04976	0.00141	0.05004	0.00090	0.05001	0.00073
	500	0.04752	0.00253	0.05001	0.00050	0.04998	0.00039
	1000	0.05003	0.00036	0.05006	0.00030	0.04997	0.00028
0.1	50	0.0999	0.00291	0.10010	0.00275	0.09962	0.00234
	100	0.09975	0.00266	0.09991	0.00193	0.09983	0.00151
	500	0.10003	0.00128	0.09998	0.00094	0.10015	0.00085
	1000	0.09991	0.00089	0.09990	0.00071	0.09996	0.00058
0.5	50	0.50213	0.01974	0.50065	0.01372	0.50003	0.01175
	100	0.50027	0.01279	0.50090	0.01013	0.49994	0.00777
	500	0.50071	0.00697	0.49941	0.00425	0.50007	0.00329
	1000	0.49973	0.00329	0.49960	0.00295	0.49988	0.00277

2.4. Kolmogorov–Smirnov Test

The two-sided, one-sample Kolmogorov–Smirnov (KS) statistic is widely used as a test-of-fit statistic to assess the consistency between the distribution of a random sample and a predefined theoretical distribution [31–34]. For a given ordered sample $(x_1, \ldots, x_n)$ of size n from the theoretical distribution $F_0(x)$, its empirical distribution is $F_n(x)$. $F_0(x)$ is the probability distribution function of the theoretical distribution pre-specified under the null hypothesis H_0; that is, the two data distributions are consistent. The two-sided K-S statistic is defined as $D_n = \sup|F_n(x) - F_0(x)|$, and $F_n(x)$ can be represented as follows:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{x_i \le x} = \begin{cases} 0, & x < x_1 \\ \frac{k}{n}, & x_k \le x \le x_k + 1,\ k = 1, 2, \ldots, n-1 \\ 1, & x_n \le x \end{cases} \tag{21}$$

When $D_n > D_{n,1-\alpha}$ is calculated from observations, H_0 will be rejected; otherwise, accept the H_0 hypothesis. Figure 5 is a KS test chart of normal distribution and normal range distribution for a measurement data group, respectively, where $n = 500$ and $m = 20$. The KS test value D_n of the normal distribution is 0.0685 (p-value = 0.01842 < 0.05), which means that the null hypothesis is rejected, and the data do not obey the normal distribution.

The KS test value D_n of the normal range distribution is 0.0354 (p-value = 0.5461 > 0.05), which means that the null hypothesis can be accepted and the data obeys the normal range distribution.

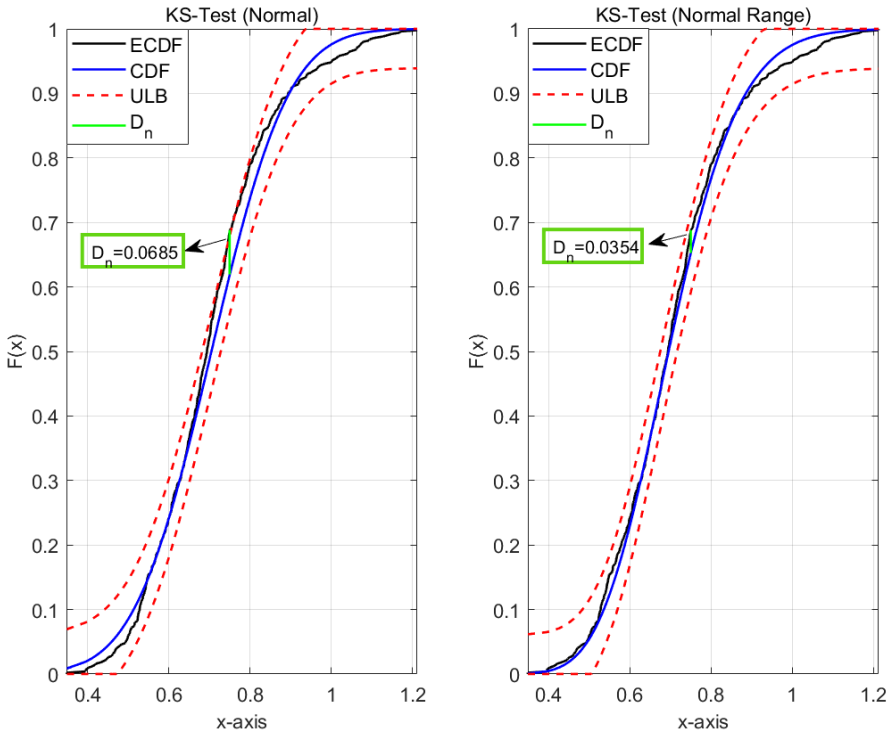

Figure 5. KS Test for normal distribution and normal range distribution (ULB = Upper and lower boundaries).

3. Results

3.1. Verifying Flatness Distribution Based on Experimental Data

Enterprises often use the normal distribution to fit flatness data, calculate dimensional quantities, and process capability indices. However, the flatness data measured on this long plane often cannot pass the KS test when fitting the normal distribution; that is, the flatness distribution is not normal. Hence, calculating the process capability index, such as CPK/PPK, will lead to deviations in the result, distortion of product quality, and an increase in the monitoring cost for product quality. The data set for verifying the flatness distribution comes from two parts: CMM measures the coordinate points on the plane of the aluminum plate and calculates the flatness data; the dimensions of the aluminum plate (specification 6061) is 500 mm in length, 200 mm in width, and 3 mm in height. As shown in Figure 6, in the experiment, the flatness data sets with sample sizes of $n = 50$, $n = 100$, $n = 500$, and $n = 1000$ were obtained when the number of points $m = 10$, and $m = 30$ were taken (points were taken at equal intervals in the X direction and Y direction of the aluminum plate, respectively).

Firstly, the first data set obtained through experiments is processed. The $\hat{\sigma_0}$ of the corresponding data set is obtained through the parameter estimation method in the second chapter. Subsequently, the KS test is performed for the normal distribution and the normal range distribution. The results are shown in Table 2. The results show that for three different values of m, there are the following conclusions: when the sample size is small (less than or equal to 100), both the normal distribution and normal range distribution pass the power KS test; however, when the sample size exceeds 500, the normal distribution cannot pass the KS test (p-value < 0.05).

In contrast, the normal range distribution always fits the flatness data (p-value > 0.05), and the surface flatness data is not normally distributed. Therefore, the larger the sample size, the closer the flatness data is to the true distribution form, that is, the normal range distribution.

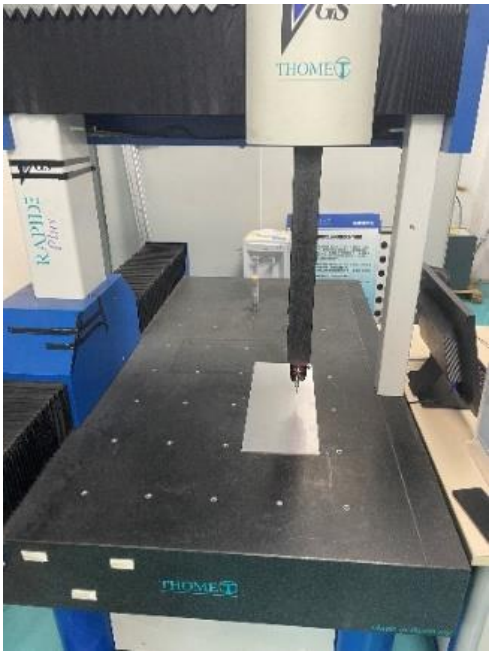

Figure 6. CMM measuring aluminum plate (CMM: Thome Rapid Plus, resolution 0.0005 mm).

Table 2. KS test for normal distribution and normal range distribution of experimental data set.

			Normal		Normal Range	
m	n	$\hat{\sigma}$	D_n	p-Values	D_n	p-Values
10	50	0.1708	0.0998	0.6646	0.1053	0.5990
	100	0.1688	0.0721	0.6758	0.0656	0.7567
	500	0.1724	0.0636	0.0348	0.0518	0.1319
	1000	0.1724	0.0536	0.0064	0.0397	0.0832
20	50	0.1873	0.1012	0.6482	0.0795	0.8844
	100	0.1941	0.0809	0.5302	0.0610	0.8279
	500	0.1886	0.0685	0.01842	0.0354	0.5461
	1000	0.1885	0.0474	0.0222	0.0196	0.8311
30	50	0.1964	0.1195	0.4388	0.0953	0.7178
	100	0.1899	0.0893	0.4028	0.0613	0.8239
	500	0.1927	0.0649	0.0295	0.0453	0.2481
	1000	0.1916	0.0488	0.0170	0.0222	0.6985

3.2. Application of Flatness Distribution Form in Dimensional Chain

Tolerance runs through the two stages of product design and manufacturing. It determines whether the product can meet the technical (quality) requirements of the product while enabling the product to be processed most economically. Dimensional chain calculation and tolerance analysis have significant applications in the design and manufacture of products. During the product design process, the accuracy of the assembly of the product can be analyzed and calculated using the dimensional chain, and the accumulation of part tolerances in the product assembly process can be analyzed. For the mechanical manufacturing process, the accumulation and synthesis of manufacturing errors during processing can be analyzed using dimensional chain calculations. The probability method is commonly used for these calculations. The probability method expresses dimensional changes as probability distributions, which requires knowing the true distribution form of shape tolerances such as flatness. The statistical tolerance of the closed loop can be obtained by the following formula when calculated by the probability method:

$$L_0 = \frac{1}{k_0}\sqrt{\sum_{i=1}^{m} \xi_i^2 k_i^2 L_i^2} \tag{22}$$

In the formula,

L_0: Closed ring tolerance;
L_i: Tolerance of the ith component ring;
k_i: Relative distribution coefficient of the size of the ith constituent ring;
ξ_i: Transfer coefficient of the ith constituent ring;
k_0: Relative distribution coefficient of the closed loop, generally approximating 1.

When the size chain of a particular product comprises flatness values, assuming that the assembly tolerance requirement is ± 0.5, the normal distribution and the normal range distribution are used to distribute the tolerance by the probability method. Figure 7 shows the influence of the two distribution forms on the differences between the parts in terms of dimensional tolerance requirements (assuming that each dimension in the dimensional chain is assigned the same tolerance):

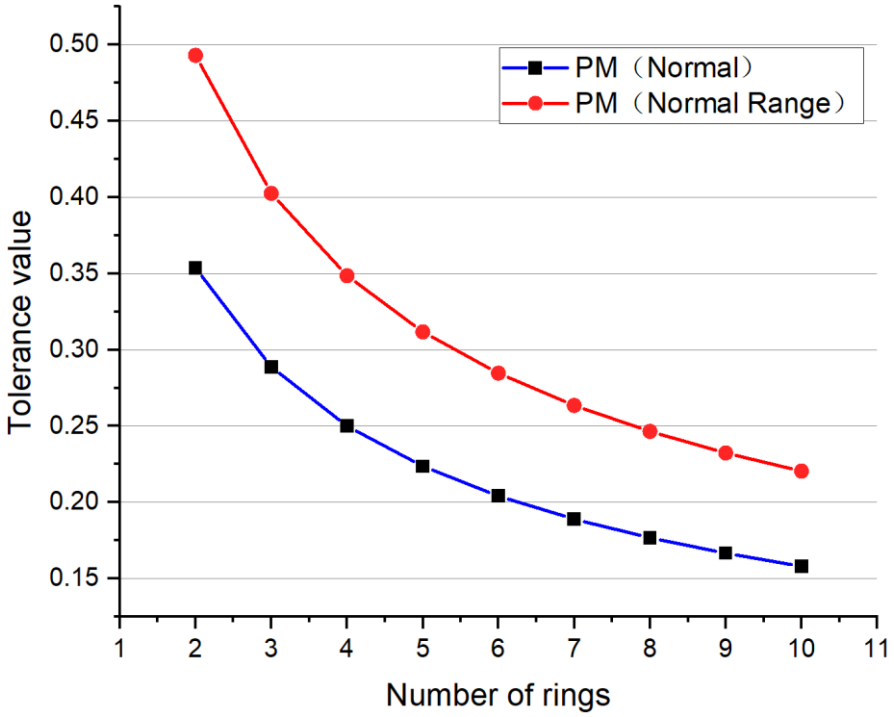

Figure 7. Comparison of tolerance distribution between the normal distribution and normal range distribution in the probability method.

Where $k_i = \lambda_i/\lambda_n$ is related to the form of the distribution, λ_i is the relative standard deviation, which is equal to the ratio of the standard deviation to half of the tolerance, that is $\lambda_i = \sigma/(T/2)$; λ_n is the relative standard deviation of the normal distribution, take the tolerance $T = 6\sigma$ (corresponding to a confidence level of 99.73%) when $\lambda_n = \sigma/(T/2) = \sigma/3\sigma = 1/3$, so $k_i = 3\lambda_i$. For a normal range distribution R, its expectation $E(R)$ and variance $D(R)$ are, respectively,

$$E(R) = \int_0^{+\infty} r dF_R(r) \tag{23}$$

$$D(R) = E(R^2) - [E(R)]^2 = \int_0^{+\infty} r^2 dF_R(r) - [\int_0^{+\infty} r dF_R(r)]^2 \tag{24}$$

The tolerance T can be obtained by $F_R(T) = 0.9973$. To facilitate the calculation, we can normalize the subsample $h_1, \ldots, h_m, (h_i \sim N(0, \sigma_0^2))$, let $h_i' = h_i/\sigma$, $i = 1, 2, \ldots, m$, then $h_i' \sim N(0, 1)$; let $R' = R/\sigma_0 = h'_{(m)} - h'_{(1)} = h_{(m)}/\sigma_0 - h_{(1)}/\sigma_0$, so there are $E(R) = \sigma_0 E(R')$, $D(R) = \sigma_0^2 D(R')$, $T = \sigma_0 T'$.

In the example in Figure 7, $m = 20$, $D(R') = 0.530984$, $T' = 6.095459$. And finally, it can be found that

$$k_i = 3\lambda_i = 3 \times \frac{\sigma_0\sqrt{0.530984}}{\sigma_0 T'/2} = 0.717274 \quad (25)$$

The tolerance values for different numbers of rings are obtained by Formula (22).

Using the normal range distribution for flatness will relax the tolerance of some rings. In the high-value scenario of battery manufacturing, this can make the processing of parts more economical and suitable for mass production.

4. Discussion

As mentioned earlier, flatness is a form tolerance critical to ensuring the quality of long straight plane assemblies. Out-of-flatness issues can result in substantial and expensive assembly rejection problems. Consequently, monitoring the flatness data corresponding to the different number of points on the plane is crucial. In this paper, we deduce that flatness has a normal range distribution based on a statistical theoretical model; the flatness data with different m values gained by the experiment all fit the normal range distribution, which has an excellent fitting effect. And it can be explained that the distribution pattern of flatness is dependent on the number of sampling points m. In the experiment results, we have shown that a normal distribution can be fitted under small samples, which is consistent with the conclusions derived by Berrado et al. [3]. However, the normal distribution cannot be fitted for large samples.

As a large amount of data were required for the analysis, we used multiple CMM measurements of a certain number of aluminum plates to study the flatness. We anticipate that these findings can also apply to flatness measurements in other processes. To enhance the credibility of this statistical theoretical model, it should be validated further by utilizing additional flatness datasets collected from diverse components. This expanded validation process will help generate more substantial evidence and strengthen the applicability of the model across various scenarios. Further research is also required for numerical calculation and parameter optimization to increase compliance with high product quality standards and reduce errors in monitoring the manufacturing process.

Author Contributions: Conceptualization, S.J.; methodology, H.L.; resources, S.J.; software, H.L.; supervision, S.J.; validation, H.L.; writing—original draft, H.L.; writing—review and editing, S.J. All authors have read and agreed to the published version of the manuscript.

Funding: This paper was supported by National Natural Science Foundation of China (51975369).

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. *ISO 1101-2017*; Geometrical Product Specifications (GPS)—Geometrical Tolerancing—Tolerances of Form, Orientation, Location and Run-Out. ISO: Geneva, Switzerland, 2017.
2. *ISO 12781-1*; Geometrical Product Specifications (GPS)-Flatness-Part 1: Vocabulary and Parameters of Flatness. ISO: Geneva, Switzerland, 2011.
3. Wen, X.-L.; Zhu, X.-C.; Zhao, Y.-B.; Wang, D.-X.; Wang, F.-L. Flatness error evaluation and verification based on new generation geometrical product specification (GPS). *Precis. Eng.* **2012**, *36*, 70–76. [CrossRef]
4. Guo, H.; Zhang, Z.; Xiao, M.; Liu, H.; Zhang, Q. Tolerance optimization method based on flatness error distribution. *Int. J. Adv. Manuf. Technol.* **2021**, *113*, 279–293. [CrossRef]
5. Mikó, B. Assessment of flatness error by regression analysis. *Measurement* **2021**, *171*, 108720. [CrossRef]
6. Mikó, B.; Szabó, S.M.; Drégelyi-Kiss, Á. Application of a Genetic Algorithm for Minimum Zone Method of Flatness. *Acta Polytech. Hung.* **2021**, *18*, 107–126. [CrossRef]
7. Mikó, B. Measurement and evaluation of the flatness error of a milled plain surface. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *448*, 012007. [CrossRef]
8. Ricci, F.; Scott, P.J.; Jiang, X. A categorical model for uncertainty and cost management within the geometrical product specification (GPS) framework. *Precis. Eng.* **2013**, *37*, 265–274. [CrossRef]

9. Colosimo, B.M.; Moroni, G.; Petrò, S. A tolerance interval based criterion for optimizing discrete point sampling strategies. *Precis. Eng.* **2010**, *34*, 745–754. [CrossRef]
10. Damodarasamy, S.; Anand, S. Evaluation of minimum zone for flatness by normal plane method and simplex search. *IIE Trans.* **1999**, *31*, 617–626. [CrossRef]
11. Cheraghi, S.; Lim, H.S.; Motavalli, S. Straightness and flatness tolerance evaluation: An optimization approach. *Precis. Eng.* **1996**, *18*, 30–37. [CrossRef]
12. Liu, C.-H.; Chen, C.-K.; Jywe, W.-Y. Evaluation of straightness and flatness using a hybrid approach—Genetic algorithms and the geometric characterization method. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2001**, *215*, 377–382. [CrossRef]
13. Kovvur, Y.; Ramaswami, H.; Anand, R.B.; Anand, S. Minimum-zone form tolerance evaluation using particle swarm optimisation. *Int. J. Intell. Syst. Technol. Appl.* **2008**, *4*, 79–96. [CrossRef]
14. Weber, T.; Motavalli, S.; Fallahi, B.; Cheraghi, S. A unified approach to form error evaluation. *Precis. Eng.* **2002**, *26*, 269–278. [CrossRef]
15. Zhu, X.; Ding, H. Flatness tolerance evaluation: An approximate minimum zone solution. *Comput. Des.* **2002**, *34*, 655–664. [CrossRef]
16. Samuel, G.; Shunmugam, M. Evaluation of straightness and flatness error using computational geometric techniques. *Comput. Des.* **1999**, *31*, 829–843. [CrossRef]
17. Zhang, Q.; Jin, X.; Zhang, Z.; Zhang, Z.; Liu, Z. An Evaluation Method for Spatial Distribution Uniformity of Plane Form Error for Precision Assembly. *Procedia CIRP* **2018**, *76*, 59–62. [CrossRef]
18. Gosavi, A.; Cudney, E.A. Form Errors in Precision Metrology: A Survey of Measurement Techniques. *Qual. Eng.* **2012**, *24*, 369–380. [CrossRef]
19. Radlovački, V.; Hadžistević, M.; Štrbac, B.; Delić, M.; Kamberović, B. Evaluating minimum zone flatness error using new method—Bundle of plains through one point. *Precis. Eng.* **2016**, *43*, 554–562. [CrossRef]
20. Chelishchev, P.; Popov, A.; Sørby, K. Robust Estimation of Optimal Sample Size for CMM Measurements with Statistical Tolerance Limits. *MATEC Web Conf.* **2018**, *220*, 04001. [CrossRef]
21. Berrado, A.; Hubele, N.; Gel, E. An Empirical Investigation into the Distribution of Flatness Measurements. *Qual. Eng.* **2006**, *18*, 351–357. [CrossRef]
22. Ghie, W. Statistical analysis tolerance using jacobian torsor model based on uncertainty propagation method. *Int. J. Multiphysics* **2009**, *3*, 11–30. [CrossRef]
23. Lin, H.C. The measurement of a process capability for folded normal process data. *Int. J. Adv. Manuf. Technol.* **2004**, *24*, 223–228. [CrossRef]
24. Chatterjee, M.; Chakraborty, A.K. A simple algorithm for calculating values for folded normal distribution. *J. Stat. Comput. Simul.* **2016**, *86*, 293–305. [CrossRef]
25. Tseng, H.Y. A genetic algorithm for assessing flatness in automated manufacturing systems. *J. Intell. Manuf.* **2006**, *17*, 301. [CrossRef]
26. American Society of Mechanical Engineers. *Dimensioning and Tolerancing Y14.5.1M*; American Society of Mechanical Engineers: New York, NY, USA, 1994.
27. Anderson, T.W. *An Introduction to Multivariate Statistical Analysis (No. 519.9 A53)*; Wiley: New York, NY, USA, 1962.
28. Mathur, M.; Karale, S.B.; Priye, S.; Jayaraman, V.K.; Kulkarni, B.D. Ant Colony Approach to Continuous Function Optimization. *Ind. Eng. Chem. Res.* **2000**, *39*, 3814–3822. [CrossRef]
29. Isiet, M.; Gadala, M. Self-adapting control parameters in particle swarm optimization. *Appl. Soft Comput.* **2019**, *83*, 105653. [CrossRef]
30. Grediac, M.; Sur, F.; Blaysat, B. The Grid Method for In-plane Displacement and Strain Measurement: A Review and Analysis. *Strain* **2016**, *52*, 205–243. [CrossRef]
31. Liu, Z.Y.; Wang, X.G.; Ren, W.L.; Long, Y.P. The deducing forms of Simpson formula and its application. *J. Shandong Univ. Technol. Nat. Sci. Ed.* **2014**, *28*, 29–31. [CrossRef]
32. Dimitrova, D.S.; Kaishev, V.K.; Tan, S. Computing the Kolmogorov-Smirnov Distribution When the Underlying CDF is Purely Discrete, Mixed, or Continuous. *J. Stat. Softw.* **2020**, *95*, 1–42. [CrossRef]
33. Gleser, L.J. Exact Power of Goodness-of-Fit Tests of Kolmogorov Type for Discontinuous Distributions. *J. Am. Stat. Assoc.* **1985**, *80*, 954. [CrossRef]
34. Massey, F.J., Jr. The Kolmogorov-Smirnov Test for Goodness of Fit. *J. Am. Stat. Assoc.* **1951**, *46*, 68–78. [CrossRef]