Analysis and Research Based on Instrument Drift Data

GUANGSHENG LUO¹, (Student Member, IEEE), CHUNYANG RUAN¹, YU YANG², AND WENWEI LI³

¹Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University, Shanghai 200083, China
²Aishu Information Technology Company Ltd., Shanghai, China
³School of Computer, Fudan University, Shanghai 200433, China

Corresponding author: Chunyang Ruan (cyruan16@fudan.edu.cn)

This work was supported in part by the Intelligent Information Processing Laboratory of School of Computer, Fudan University, Shanghai, China, and in part by the company in Shanghai.

ABSTRACT

Instruments and meters are used to inspect, measure, control, analyze, calculate and display the physical quantity, chemical quantity, biomass, electrical parameters, geometric quantity, and movement status of the object under test. Its applications cover a wide range of industries, agriculture, transportation, science, and technology, etc., and can be divided into electrical instruments, optical instruments, laboratory instruments, analytical instruments, testing machines, industrial instruments, etc. It assumes the task of gatekeeper and guide in the construction of the national economy. The analysis and research techniques of this article are based on the professional data of power plant industrial instrumentation. This paper presents an overview of various analysis methods of instrument data, and in-depth analysis of time correlation and coverage analysis, outlier analysis, etc. We give the statistical analysis process of drift data, focusing on the process of the AFAL method; We also described the drift algorithm of random behavior and deviation behavior and explored and discussed the analysis tools and data analysis results of the power plant industry professional data. We have designed a set of application systems for instrument data import, export, storage, and analysis, which can store instrument calibration data onto a Hadoop database in a prescribed format. Which provides great reference values for the calibration of instrument data. Finally, it looks forward to the analysis and research direction and development trend of industrial big data.

INDEX TERMS

Time correlation, coverage analysis, outlier analysis, AFAL, random behavior, deviation behavior.

I. OVERVIEW OF INSTRUMENT DRIFT CHARACTERISTICS

In recent years, industrial big data has attracted much attention as the key technical support of my country’s “intelligent manufacturing” and “Industrial Internet” and the important foundation for the integration of the two industrializations [1]. With the advancement of national strategies such as intelligent manufacturing and industrial Internet, individuality New models such as customized customization, network extension, and intelligent design, production, and services to continue to emerge [2], and the demand for industrial big data technology, products, and platforms continues to increase, providing sufficient application scenarios of industrial big data.

In industrial production, data is generated in real time. The speed, energy consumption, temperature and humidity of the production machine tool, the combustion and combustion of the thermal generator set, the equipment data of the automobile, and the instrument are all data in the production process [3]. There are many data sources, and their characteristics are also diversified, including historical data, real time data of production processes, and monitoring data of the operating status of industrial equipment [4]. Industrial data analysis is the focus on industrial computing, and it is also the key value of industrial data analysis of instrumentation in this article. Through industrial big data analysis and other key technologies, the intelligent level of design, manufacturing, service and other links can be improved [5], meet the customized needs of users, improve production efficiency and reduce production costs, and create quantifiable value of enterprises; Data analysis is the process of converting
the data collected by a large number of organizations into understandable and insightful analysis [6]. This paper discusses the application and overview of instrumentation data in the industry, quantitatively analyzes the characteristics of instrument drift [7], and discusses the representation method of instrument data and the analysis method and process in practice. We have designed a set of drift data analysis tools and application systems to calibrate nuclear power long instruments The analysis provides a good reference value.

A. CHALLENGES OF INSTRUMENT DRIFT CHARACTERISTICS

1) Challenge 1: The level of intelligence of instrumentation is not high, and product applicability is not good. The accuracy of an instrument and meter is a quantity (expressed in engineering units or expressed as a percentage of the span), which defines the degree of conformity between the real output and the expected output when used under specified operating conditions [8]. With the development of information, automation, intelligence, and integration of instruments are indispensable conditions for the development of current instruments. It is also a good way to reduce errors, improve efficiency, improve accuracy, and expand uses [9]. Enterprises do not have a deep understanding of product applications, insufficient research on user applications, and insufficient product functional accessories, application software, and application operations. They often only focus on the host or only produce bare metal, resulting in a narrow product application range or application inconvenience. Affect the promotion and application of domestic instruments.

2) Challenge 2: Errors in the actual measurement process of instrumentation equipment.

In an ideal situation, there will be a perfect correlation between input and output, such as in the example shown in Figure 1, but in reality there will always be a certain error or uncertainty in each process measurement. The rated accuracy of the instrument consists of three instrument characteristics: repeatability, hysteresis, and linearity. These characteristics occur at the same time, and their cumulative effect is represented by the frequency band surrounding the true output. The manufacturer usually specifies this frequency band, to ensure that its comprehensive effect fully limits the performance of the instrument within the design life of the instrument.

3) Challenge 3: The influence of environment on the drift of instrument data.

Drift is usually described as an undesirable change in output over a period of time. The change will not have anything to do with input, environment, or load. As described here, drift is deemed to refer to any measurable change between calibrations. The instrument zero set is the most common type of drift. This drift can be described as the linear displacement of the instrument’s output within its operating range.

B. RELATED WORK

The goal of this paper is to use existing statistical analysis techniques and data analysis tools to calibrate the data, and to quantify the drift behavior of instrumentation safety-related increments, to provide other power plants with information about the typical drift performance of other nuclear power plants, and to evaluate through sample analysis The drift behavior of the meter over time is calibrated at the same time.

This paper outlines the method of historical data analysis of instrument calibration data. The purpose is to use statistical analysis techniques to calibrate or monitor data to quantify the drift characteristics of the instrument within a defined probability range. Traditional instrument data analysis or data abnormal point discrimination, most of them use it is the AFAL, which is the analysis method of instrument drift characteristics. This article proposes many innovative analysis methods of practical applications based on general data analysis, such as time correlation, abnormal data analysis, coverage analysis, outlier analysis, etc. And conducted in-depth research, discussion, and application of these methods. The main contributions of this paper are as follows:

1) A new analysis method and MatLab algorithm implementation are proposed for instrument data.

2) Design a set of application systems based on data drift analysis algorithm.

C. THE RESEARCHES CONTENT AND INNOVATION OF THIS PAPER

The first section of this paper gives an overview of the characteristics of instrumentation drifts, and proposes the challenges of industrial data analysis of power plant instruments and equipment, and at the same time summarizes the problems and objectives to be solved in this article. The second section analyze the data calibration of instrumentation And mathematical modeling, the database used in this article is summarized. Sections 3 and 4 respectively carry out in-depth analysis and discussion on the method of instrument drift data analysis and calculation process. Section 5 discusses the drift

FIGURE 1. Ideal instrument input/output curve.
of instrument data A brief introduction to the analysis tools of, and at the same time the research and comparison of the results of this industrial data analysis. Finally, the industrial data analysis is summarized and prospected.

The innovation of this article: In actual applications, due to the very high requirements for the time and data of instruments and meters, new time correlation and abnormal data analysis and algorithm methods are adopted, and W verification and D-prime data are used in the coverage analysis. Verification method of verification.

II. OVERVIEW OF INSTRUMENT DRIFT DATA
A. CALIBRATION METHOD OF INSTRUMENT DRIFT DATA
During the calibration process, the raw data is gained and recorded, which is an indication of the operating conditions of the equipment with established limits. The difference between the original data obtained from the current calibration and the original data obtained from the previous calibration, and the potential accuracy impact on these differences Calibrations is performed. These potential effects include instrument hysteresis and linearity errors during current and previous calibrations, instrument repeatability errors during current and previous calibrations, and measurement and test equipment errors during current and previous calibrations. Human-made or human-related errors occurred in the current and previous periods.

If the output is in the time period between two calibrations. The difference between the initial sett and the left sett is data drifts. There are several ways to calibrate data drift [10]:

1) Vertical calibration: Periodically check or adjust instruments to ensure that they respond to 0 input signals within the required accuracy range. Because their operating accuracy is directly related to factory safety. Each calibration, calibration checks or Regular monitoring has acceptance standards set for the instrument. If the instrument is found to be within the acceptance standard, it may not be necessary or necessary to adjust the instrument. If the instrument is found to exceed the acceptable standard for a small extent, no further analysis methods are usually required. The instrument can be adjusted.

2) In-situ calibration: The condition found is to find a channel or part with a channel after a period of time, before any calibration.

3) Left calibration: The left side is the status of the remaining channels or part of the channels after calibration or calibration.

B. MATHEMATICAL REPRESENTATION OF INSTRUMENT DRIFT DATA
Taking into account the differences in the range of different instruments, we need to use a percentage to express the amount of instrument drift:

\[
d_n = \frac{D_n - D'_{n-1}}{L} \times 100\%
\]  

Among them, \(d_n\) is the drift of the meter between the \(n\)th calibration and the \(n - 1\)th calibration interval. \(D_n\) is the value of the \(n\) calibration. \(D'_{n-1}\) is the adjusted value of the \(n - 1\) calibration meter Value. \(L\) is the range of the meter. The drift of a single meter is not representative. After calculating the drift of a single meter, it is necessary to calculate the basic data of the group of meters, including the average value, standard deviation, and number of data points. Etc., to calculate the drift data of this type of instrument. Analyzing a set of data can get the calculated drift value of the set of data.

\[
AD = AD_{bias} \pm AD_{random}
\]  

C. DATABASE USED BY INSTRUMENTATION
At present, the sizewell B database [11] is mainly used in the database of power plant instruments and instruments, which is the UKs Sideswell B nuclear power plant database. And the EPRI main instrument database, which is the American Electric Power Research Institute database. This article uses the data of these two databases as analysis Foundation.

The data of the Sizewell B instrument are stored in a Microsoft Access database file. This file includes the original and the left sample calibration data and related instrument information. The data contained in the database cover the operating period from March 1995 to December 2002. The number is 140 data. All transmitters are manufactured by Bar- ton (models 752, 763, and 764). For each of the 9 checkpoints, the number of undiscovered data included in the analysis is 726, excluding outliers. For instrument replacement, all data before replacement is eliminated. Data lost at a given checkpoint, skip calibration data, and extend the calibration interval to the next recorded value. The calibration interval of all transmitters analyzed in this article, first One fuel cycle is 12 months, the second fuel cycle is 15 months, and all subsequent fuel cycles are 18 months.

The EPRI main instrument database [12], which was developed to support online monitoring applications. The instrument calibration data from 18 power plants were collected, entered into the database and analyzed in detail. The database contains data from May 1975 to May 1975. 1139 instruments, 6700 calibrations and nearly 34,000 individual calibration checkpoint values recorded in the time frame of November 1996. Data collection is concentrated on the main sensors as the main key equipment. Table 1 shows the EPRI database contains The instrument and calibration data of the instrument are classified by instrument manufacturer and model.

D. OVERVIEW OF AFAL ANALYSIS METHODS
The analysis method of the instrument drift characteristic (AFAL) is based on the statistical analysis of the recorded data, the study of the instrument characteristics, and then the demonstration result of whether the calibration cycle can be extended. The data recorded by the industrial power plant instrument includes: the actual measured value of each calibration point when the instrument is calibrated after one cycle of operation, that is, the as-found value. After the meter
TABLE 1. Instrument model and calibration data.

| Instrument Model | Number of Samples | Size Percentage | Calibration Model | Number of Samples | Size Percentage |
|------------------|------------------|-----------------|------------------|------------------|-----------------|
| Rosemount        | 1151             | 129             | 11.30%           | 1152             | 52              | 4.60%           |
|                  | 1153             | 466             | 40.60%           |                  | 368             | 5.50%           |
| Barton           | 384              | 30              | 2.60%            |                  | 752             | 27              | 2.10%           |
|                  | 763              | 7               | 0.60%            |                  | 764             | 109             | 9.60%           |
| Foxboro NE       | NE2              | 8               | 0.70%            | NE1              | 8               | 0.70%           |
|                  | NE13             | 41              | 3.60%            | 209.00%          | 3.10%           |
| Veritrak         | 59PM18           | 3               | 0.30%            | 16               | 0.20%           |
|                  | 70FH2            | 8               | 0.70%            | 42               | 0.60%           |
| Tobar            | 32D2P            | 24              | 2.10%            | 123              | 1.80%           |
|                  | 555              | 3               | 0.30%            | 24               | 0.40%           |
| Delaval          | XM-548S2         | 2               | 0.20%            | 16               | 0.20%           |
| Conax RTD        | 7Q77-10001       | 2               | 0.20%            | 13               | 0.20%           |
| Bailey RTD       | 6623690A2        | 4               | 0.40%            | 26               | 0.40%           |
| Unknown pressure | 30               | 2.60%           |                   | 209              | 3.10%           |
| Total            | 1139             | 100.00%         | 6700             | 100.00%          |

is calibrated and adjusted, the value of each calibration point is measured, namely the as-left value. The as-left values and as-found value are respectively defined as follows: as-found value is the state of the instrument before calibration after 1 operating cycle, expressed in AF; as-left is the state of the instrument after calibration, expressed in AL. Facts have shown that even the most stable instrument has a difference between its AF value and the AL value after the previous calibration, although the difference is small. This difference is the AF value minus the AL value, which is the drift of the instrument in the previous operating cycle [13], [14]. Analyzing, studying and mastering the drift value of the meter can enable the industrial power plant to clearly understand the characteristics of the meter and determine whether the calibration period of the meter can be extended. Therefore, the AF value is deducted from the AL value and calculated and analyzed, which is the AFAL analysis method [15].

The AFAL analysis method is based on the historical data of the instrument calibration of industrial power plants, that is, the instrument calibration record during the previous refueling overhaul. By collecting a certain amount of data samples and calculating, the statistical characteristic value of the instrument drift is obtained, and the instrument’s performance in the original calibration period is analyzed. Performance, evaluates whether the performance of the instrument meets the design requirements after the calibration period is extended, and demonstrates the feasibility of the extension of the calibration period.

III. DRIFT DATA ANALYSIS METHOD

A. TIME CORRELATION ANALYSIS

After confirming that the sample data obeys the normal distribution [16], analyze the correlation between the drift data of this group of meters and the time. The time correlation analysis is divided into the following three parts:

(1) Time correlation analysis of random drift.

The method of segment analysis is used to judge the time correlation between random drift. The verification data is grouped according to the length of the verification interval, such as 0−1C, 1−2C, 2−3C, etc., each group of data. The number should be as many as possible to make the analysis result more statistically significant. Find the standard deviation of each set of data and calculate the square of the ratio of the maximum standard deviation to the minimum standard deviation.

\[ F_{\text{calc}} = \frac{SD_{\text{max}}^2}{SD_{\text{min}}^2} \]  

(3)

Among them, \( SD_{\text{max}} \) = maximum standard deviation. \( SD_{\text{min}} \) = minimum standard deviation.

According to \( \nu_1 \) and \( \nu_2 \) and the corresponding confidence level [17](0.95), the corresponding critical value \( F_{\text{critical}} \) is found, and the \( F_{\text{critical}} \) and \( F_{\text{calc}} \) is compared, if \( F_{\text{calc}} > F_{\text{critical}} \), then time correlation is considered to exist. \( F_{\text{calc}} \) is the calculated value of \( F \).

\( \nu_1 = \) The maximum standard deviation corresponds to the number of samples in the group −1.
The minimum standard deviation corresponds to the number of samples in the group \(-1\).

2) Analysis of the strength of time correlation of random drift [18]. After analyzing the time correlation of random drift, continue to analyze the strength of time correlation of random drift. Group the data according to the check interval (such as 0-1C, 1-2C, 2-3C, etc.) calculate the average check interval (algebraic average) for each group of data, the ratio of its standard deviation to the standard deviation of the single-cycle data, and the difference between the average check interval and the single-cycle average check interval. The square root of the ratio, and compare.

(3) Time correlation judgment of deviation drift [19]. Based on the sample size and standard deviation, find the corresponding unbiased maximum mean value. If the deviation drift is greater than the unbiased maximum mean value, it is conservatively considered to have strong time correlation. Otherwise, it is considered that the deviation drift has no time correlation, and the deviation drift part can be ignored when calculating the drift amount after the period is extended.

### B. ANALYSIS OF ABNORMAL DATA

Test data recording errors, calibration errors, calibration equipment failures, scale or set value changes, instrument failures, design/manufacturing/installation defects may all cause abnormal data. Abnormal data can bring about the correctness of analysis results the impact is great. It is necessary to determine which abnormal data is through analysis, and remove or correct it from the sampled data. Use the \(T\) test to find possible abnormal data. The calculated \(T\) value is compared with the corresponding critical value. If the critical value is exceeded, the judgment The data may be abnormal data.

\[
T = \frac{|x - \mu|}{\sigma} \tag{4}
\]

The matlab pseudo code for abnormal data analysis is as following:

```matlab
function [] = findabnormaldata(afalvalueinput, effectivegrade, meansd)
criticalTtable = [10, 2.18, 2.29, 2.41, 2.48; 20, 2.56, 2.71, 2.88, 3.00; 30, 2.75, 2.91, 3.10, 3.24; 40, 2.87, 3.04, 3.24, 3.38; 50, 2.96, 3.13, 3.34, 3.48; 75, 3.10, 3.28, 3.50, 3.65; 100, 3.21, 3.38, 3.60, 3.75; 125, 3.28, 3.46, 3.68, 3.83; 150, 3.33, 3.51, 3.73, 3.88];
datanum = size(afalvalueinput,1);
if datanum < 11
t_critical = criticalTtable(1, effectivegrade+1);
else if datanum < 21
t_critical = criticalTtable(2, effectivegrade+1);
else if datanum < 31
t_critical = criticalTtable(3, effectivegrade+1);
else
if datanum < 41
t_critical = criticalTtable(4, effectivegrade+1);
else if datanum < 51
t_critical = criticalTtable(5, effectivegrade+1);
else if datanum < 76
t_critical = criticalTtable(6, effectivegrade+1);
else if datanum < 101
t_critical = criticalTtable(7, effectivegrade+1);
else if datanum < 126
t_critical = criticalTtable(8, effectivegrade+1);
else
    t_critical = criticalTtable(9, effectivegrade+1);
end
for j = 1 : 10
    for i = 1 : size(afalvalueinput,1)
        T = abs(afalvalueinput(i,j) - meansd(j,1))/meansd(j,2);
        if T > t_critical
            disp('abnormal data')
disp([i,j])
        end
    end
end
```

### C. COVERAGE ANALYSIS

The normal distribution in mathematical statistics is one of the most commonly used distribution to describe random quantities. The analysis in this article is based on the assumption that the data obeys the normal distribution. The drift amount has a certain probability to fall within this interval, and the deviation drift and the time correlation are judged. The data is verified by W and D-prime [20]. W verification is used for samples with a sample size of \(<50\), through calculation The W value is compared with the actual critical value to verify whether it obeys the normal distribution. The steps are as following:

1) Arrange the samples from small to large.

2) Calculate \(S^2\).

\[
S^2 = (n - 1) \times s^2 \tag{5}
\]

Among them, \(n\) is the number of samples, \(s^2\) is the variance of the data.

3) Calculate \(b\) value.

\[
b = \sum [a_{n-i+1} \times (x_{n-i+1} - x_i)] \tag{6}
\]

Among them, \(i\) is from 1 to \(k\), if \(n\) is an even number, \(k = n/2\), if \(n\) is an odd number, \(k = (n - 1)/2\), \(b\) is the drift amount, \(a_n\) is the value of the \(n - 1\)th calibration, and \(x_n\) is the value of the \(n - 1\)th adjustment.

4) Calculate \(W\) value

\[
W = \frac{b^2}{S^2} \tag{7}
\]

5) Find the corresponding \(W\) critical value according to the required importance (1% or 5%) and the number of samples \(n\) and the actual value, and compare the calculated \(W\) value with the \(W\) critical value. If the calculated \(W\) value
is greater than the W critical value, The sample is considered to obey a normal distribution.

D-prime verification is used for data with a sample size of \( \geq 50 \), by comparing the calculated \( D \) value with the actual critical value to verify whether it obeys the normal distribution, the steps are as following:

1. Sort the samples from small to large.
2. Calculate \( S^2 \).

\[
S^2 = (n - 1) \times s^2
\]  

(8)

Among them, \( n \) is the number of samples, \( s^2 \) is the variance of the data.

3. Calculate T value.

\[
T = \sum \left[ (i - \frac{n + 1}{2}) \times x_i \right]
\]  

(9)

where \( i \) is from 1 to \( n \).

4. Calculate \( D \) value

\[
D = \frac{T}{S}
\]  

(10)

5. Find the acceptable interval according to the number of samples (interpolate according to \( n \)), and compare the calculated \( D \) value with the acceptable interval. If the \( D \) value is within the acceptable interval, the sample is considered to belong to a normal distribution.

If \( W \) verification and \( D-prime \) cannot verify that the sample belongs to a normal distribution, the coverage analysis method is used to conservatively treat the data as a normal sample. Among them, \( n \) is the number of samples, \( s^2 \) is the variance of the data.

6. Calculate \( AF \) so that the samples with \( (n - 1)/n \) or 97.5\% (the smaller value of both) are within the range of \( TFAF \).

7. Use the normal distribution with the mean value of and the standard deviation and the critical value of \( TFAF \) to verify whether it obeys the normal distribution.

The possible existence of external data is determined by the \( T \) test. In short, the test compares a single measurement value with sample statistics and calculates a parameter \( T \), called extreme student deviation, as showed below:

\[
T = \frac{|x_i - \bar{x}|}{S}
\]  

(11)

Among them, \( T \) is the calculated value of extreme student deviation and the critical value of \( T \) for sample size, \( \bar{x} \) is sample mean, \( x_i \) is single data point, \( S \) is sample standard deviation.

If the calculated value of \( T \) exceeds the critical value of the sample size and the required significance level, the evaluated data point is identified as an outlier.

\section*{IV. STATISTICAL ANALYSIS AND CALCULATION METHOD OF DRIFT DATA}

\subsection*{A. RANDOM BEHAVIOR}

When repeated measurements of a fixed parameter are performed, the measurement results are usually not completely consistent. Just as these measurement results are not completely consistent with each other, they also have some deviations from the true value. The uncertainty surrounding the fluctuation of the true value is considered to be Random, without any specific preference for a specific direction.

Figure 2 shows the expected nature of random, normally distributed data. The data will be more likely to be near the mean, and the standard deviation defines the change in the data about the mean.

\subsection*{B. DEVIATION BEHAVIOR}

Suppose the meter shows that the instrument is actually full of 50\%, but a poorly designed liquid level monitoring circuit shows that the instrument fluctuates randomly by about 60\%. 

G. Luo et al.: Analysis and Research Based on Instrument Drift Data
As mentioned above, some fluctuations in the central value (mean value) represent random uncertainty. However, this In this case, the fixed error of 10% is called systematic error or deviation uncertainty.

The offset error is a known constant value that can be calibrated outside the measurement circuit. In other cases, the known offset error will affect the measurement accuracy in a single direction, but the size of the error is not constant. Otherwise, the accurate measurement may become inaccurate due to the bias effect [21]. The measured value may have a small standard deviation (uncertainty), but because it is completely different from the actual value, it can be understood that the deviation will cause the measured value to change from the real value is offset by a certain amount. An example of a deviation is to set the watch 5 minutes slower; no matter what time of day, the watch will always be 5 minutes less than the actual time. Figure 3 illustrates the bias behavior.

**C. AFAL ANALYSIS PROCESS**

Meter drift AFAL calculation method. Statistical analysis of meter drift data is based on the meter calibration data performed during the previous refueling overhauls of the power plant [22]. By collecting appropriate data samples, performing statistical analysis and calculation, statistical characteristic parameters of meter drift error are obtained and evaluated. The instrument drift characteristics during the original calibration period are used to predict the instrument drift characteristics after prolonging the fuel cycle and demonstrate the feasibility of adjustment.

AFAL (As-Found Versus As-Left), the Chinese name is instrumental drift characteristics. Under normal circumstances, the instrument manufacturer usually assigns a drift margin to the instrument. Under ideal conditions, the margin will represent the expected time period however. AFAL analysis of calibration data may yield higher drift values than those listed in the manufacturers specifications. The reason is that AFAL values include multiple sources of uncertainty that exceeds the drift of the real instrument. Each of the following error sources may affect the size of the AFAL value.

1. The true instrument drift means that the changes in the instrument (time-related or opposite) are output during the time period between two calibrations.
2. There are instrument hysteresis and linearity errors in the current and previous calibration processes, and there are instrument repeatability errors in the current and previous calibration processes.
3. Measurement and test equipment errors, including any analog meter readability errors during the current and previous calibration periods.
4. Man-made or human-related errors in calibration occurred during the current and previous periods.
5. The influence of instrument temperature is attributed to two calibrations.
6. Other environmental influences that occur between two calibrations will cause the instrument to output.
7. Change with system operation (shutdown, cooling and decompression).
8. Improper application, improper installation or other operation impacts.

Some of the above effects are negligible, while others may be important. Most of these effects are not expected to have a significant impact on the measured in-situ and left-position settings; however, some effects such as changes in ambient temperature Will have a considerable impact. For example, sensor accuracy: 0.25% of the calibration range, sensor drift: 0.2% of the 30-month upper limit, sensor temperature influence: (per 100°F, 0.5% of the upper limit of the range +0.5% change of calibration range).

Assuming that the calibration span and the upper limit of the range are approximately the same, for temperature changes greater than 20°F, the drift value specified by the supplier is less than the specified accuracy or the specified temperature effect, the factory may observe an 18-month cycle.
In view of the above, anyone or all of the uncertainties will lead to the total change observed between two calibrations, which have the following potential effects on the analysis of instrument calibration data:

1. The calculated amount of change is conservative, so it may exceed the drift assumptions or the manufacturer’s prediction. When trying to verify the manufacturer’s performance requirements, the above-mentioned possible factors that may cause the calculated drift should be considered.

2. The magnitude of the calculated change including all the above sources of uncertainty may mask any real time-related drift. In other words, the analysis of AFAL data may not reveal any time dependence (the magnitude of the output change over time increasing trend). This does not mean that there is no time-related drift, but only that when all the above sources of uncertainty are combined, it may be small enough to accumulate instrument uncertainty during the assessment the impact is not easy to see. Figure 4 shows in the processing flow chart of the AFAL instrument analysis.

![Process flow chart of AFAL instrumental analysis](image)

**FIGURE 4.** Process flow chart of AFAL instrumental analysis.

**D. CALCULATION OF DRIFT**

\[ AD_E = AD_{E, bias} \pm AD_{E, random} \]  \hspace{1cm} (12)

Among them, \( AD_E \) is the extended drift value, and \( AD_{E, bias} \) is the extended deviation. The drift part, \( AD_{E, random} \) is the extended random drift part. The drift value is equal to the deviation drift part plus the random drift part value. The deviation value is compared with the standard value. If the drift value is greater than the deviation value, it is considered to have a time correlation, otherwise the task has no time correlation. After the calculation period is extended, the deviation drift can be ignored. Specifically, refer to the time correlation analysis content. The deviation drift part is calculated as following:

1. If the deviation drift is judged to be a strong time correlation, then

\[ AD_{E, bias} = AD_{bias} \times \frac{TI_E}{TI_0} \]  \hspace{1cm} (13)

Among them, \( AD_{bias} \) is the deviation drift before extension, \( TI_E \) is the extended calibration interval, \( TI_0 \) is the average check interval before extension.

2. If the deviation drift is judged to have no time correlation, then \( AD_{E, bias} = 0 \) The drift of the random part is calculated as follows, if the random drift is judged to be a strong time correlation, then

\[ AD_{E, random} = AD_{random} \times \frac{TI_B}{TI_0} \]  \hspace{1cm} (14)

Among them, \( TI_B \) is the extended verification interval, \( TI_0 \) is the average verification interval before the extension, \( AD_{random} \) is the random drift before extension

\[ AD_{random} = \sigma \times k \times NAF \]  \hspace{1cm} (15)

\( \sigma \) is the standard deviation before extension, \( k \) is the allowable factor and its value is 95%, \( NAF \) is the adjustment factor determined if the coverage analysis is performed. If the random drift is judged to be medium time correlation, then

\[ AD_{E, random} = AD_{random} \times \sqrt{\frac{TI_E}{TI_0}} \]  \hspace{1cm} (16)

**V. DRIFT DATA ANALYSIS TOOLS AND RESEARCH RESULTS**

**A. DRIFT DATA ANALYSIS APPLICATION DESIGN**

Aiming at the data analysis method and calculation method of instrumentation, this paper designs a set of application systems. Specifically, the PC uses a web page to establish a set of instrument data acquisition, analysis, and display tools. The main core business process is to extract from the power plant instrument software the data are sent to the Hadoop data platform. The application system of this paper reads data from Hadoop through the business center for analysis, calculation, and display. Business personnel access the application system of this article through the web page and call data analysis and calculation. The calculation results are visualized in a three-dimensional model. The business architecture diagram is shown in Figure 5.

**B. DATA ANALYSIS TOOL IPASS**

In addition to the 4V characteristics of industrial big data [23], it also has its own characteristics that distinguish big data from other industries: the proportion of unstructured data is large, the data relevance, especially the instrumentation data of nuclear power plants, is more professional, etc. It is mainly to solve the problem of intelligence In-depth analysis and application issues such as the full life cycle, fault detection, and health prediction of equipment and smart products [24], [25].
### TABLE 2. Drift tolerance interval statistics.

| Deviation tolerance interval by checkpoint number | average | maximum | minimum | standard deviation |
|---------------------------------------------------|---------|---------|---------|--------------------|
| 1.25                                              | 1.36    | 1.19    | 1.17    | 1.57               |
| 1.54                                              |         |         |         |                    |
| 1.35                                              | 1.06    | 0.75    | 0.59    | 0.66               |
| 1.66                                              | 1.05    | 0.72    | 0.69    | 0.74               |
| 1.14                                              | 1.16    | 1.35    | 1.31    | 1.32               |
| 2.24                                              | 2.36    | 2.62    | 2.72    | 2.63               |
| 2.16                                              | 2.28    | 2.8     | 2.41    | 3.11               |
| 2.05                                              | 2.18    | 2.34    | 2.26    | 2.14               |
| 3.2                                               | 2       | 1.6     | 1.1     | 1.1                |
| 1.9                                               | 0.9     | 0.8     | 0.6     | 0.7               |
| 2.4                                               | 2.8     | 2.8     | 2.7     | 2.6               |
| 1.74                                              | 0.77    | 0.85    | 0.58    | 0.66               |
| 2.46                                              | 1.15    | 1.08    | 0.83    | 0.65               |

**FIGURE 5. Drift data business architecture diagram.**

At present, data analysis tools commonly used in other industries include SAS cluster analysis, R-Programming free software programming language and software environment for statistical calculation and graphics, SPSS, python, etc. This article uses IPASS tools [26]–[28].

IPASS can store, evaluate, track and trend AFAL instrument calibration data and RTD cross-calibration data [29]. The general functions of IPASS include: creating a database of tools for trend and performance monitoring; defining its operations for each instrument in the database Features; calibration format and descriptive information; input and store historical AFAL calibration data of its instrument in chronological order; calibration data adopts the form of actual measurement and left measurement; input and store historical cross-calibration temperature measurements in chronological order for use in RTD; execution AFAL statistical analysis to characterize individual instrument drift instruments or user-defined instrument groups; evaluate RTD cross-calibration data to evaluate the user’s personal and overall performance; define RTD groups; determine the tolerance interval for instrument drift; determine RTD measurement deviations; Identify uncalibrated data points; filter the calibration data for abnormal data points, and exclude those points at the user’s discretion; perform analysis and graphic checks to confirm whether the data is normal or normal; use the normal distribution as the boundary; generate Related AFAL calibration data, instrument drift statistics; standard report screening and routine inspection of outliers; maintain database security by using login ID and password to control access; import and export calibration data files, etc.

The IPASS menu and toolbar contain commands for establishing program modes. IPASS has three modes: AFAL, Bistable and RTD. Each mode corresponds to a specific type of calibration data. After selecting the mode, IPASS will automatically configure its routines, menus and toolbars so that the calibration data corresponds to the instrument used in the selected mode. During the session, the user can switch between modes at any time. The active mode is displayed in the status bar at the bottom of the screen. IPASS data input functions are used to create, edit, delete and modify instrument records and calibration data. These functions are exclusively used to enter and view data; no analysis is performed in the data entry screen. However, it can generates related to instrument records and calibration data report. The three main functional areas related to data entry are instrument definition, calibration data entry and library.

Two windows are associated with the calibration data input. The “Calibration Data Input” window is used to input the calibration data of the instruments contained in the database in chronological order. The data input fields available to the user depend on the activation mode and previously established in the “Calibration Format” window.
C. RESEARCH RESULTS OF INSTRUMENT DATA ANALYSIS

Table 2 provides an overall summary of the calculated drift tolerance interval for each of the nine checkpoints for each instrument group, as well as basic statistics (minimum, maximum, average and standard deviation). Table 2 has made 14 records of 9 detection points. The average deviation of the first time are 1.57%, the maximum drift is 2.56% is the first detection point, the minimum 1.17% is the fourth detection point, and the standard is 0.44%. From Table 2, the average drift value is basically around 1%.

The result of the analysis is that the drift tolerance interval of the Barton 763 and 764 transmitters in the Sizewell B database is on average between 1% and 2%, while the interval error of the Barton 752 transmitter in the Sizewell B database is on average between 1% and 2%. Between 2.6%. For a given transmitter model, no consistent trend is determined, that is, whether the drift interval is greater near the low or high end of the interval. The drift tendency is random and depends on the situation.

Comparing the average results obtained by different transmitter types will provide a frame of reference for performance metrics based on transmitter behavior of other nuclear power plants. Therefore, the total meter drift tolerance interval from the EPRI database can be compared with the transmitter for each analysis the drift tolerance interval obtained by the type is compared.

Figure 6, Figure 7 respectively shows the comparison of the overall results of the 359 Barton 763 and 764 transmitter meters in the EPRI database and the 84 Barton 763 and 764 transmitter meters analyzed from Sizewell B. Checkpoint values due to different applications Different, therefore showing a common calibration checkpoint. Compared with transmitters included in the EPRI database, the Sizewell B Barton 763 and 764 transmitters have a lower drift tolerance interval.

In the study of this paper, the drift performance of the Sizewell B Barton 752 transmitter is similar to the drift performance obtained for the Barton 763 and 764 transmitters in other nuclear power plants. The drift performance of the Sizewell B Barton 752 transmitter causes its value to be consistently higher than those obtained from other nuclear power plants. The value obtained is larger. A variety of factors may cause this result. The most obvious are the deviation from the EPRI database or Sizewell B data. Because compared with the sample size of other transmitter types, Figure 8 Barton 752 transmitter sample size On the low side. In addition, the given application of the transmitter is not considered,
so the environmental impact may be a problem. There is not enough data information about the EPRI database to draw important conclusions about the performance comparison of the Barton752 transmitter.

VI. SUMMARY AND OUTLOOK

This paper discusses the data drift of nuclear power plant instrumentation based on the characteristics of domain knowledge analysis equipment data, and points out some calibration methods and mathematical formula expressions for data drift, which provides a valuable reference to other nuclear power plants or industrial big data. In addition, this paper also analyzes some professional analysis methods and analysis processes of instrumentation, which provides a theoretical guidance of future development. The actual analysis results of data analysis tools show that it has a certain practical value of nuclear power plants or corporate industries. In addition to the analytical methods mentioned in this article, the data analysis of industrial scenarios actually includes probabilistic latent semantic analysis (PLSA) and potential Dirichlet allocation (LDA) [30]. There will be many applications for industrial scenarios of the future. The main body model of this, which will continue to be studied in future work.

At present, industrial big data have gradually become the foothold on the deep integration of traditional manufacturing and a new generation of information technology [31]. Industrial big data will become an important foundation for promoting the innovation and development of manufacturing, and inject strong impetus into Chinas industrial upgrading and transformation [32]. There are many types of industrial equipment, the integration of multi-source data and the processing of massive data, as well as security event detection, multi-device data fusion modeling, knowledge acquisition, etc, are all directions worth studying [33], [34].

REFERENCES

[1] J. Li et al., “Research on the status Quo, constraints and improvement countermeasures of the key technologies, products and industrial ecology of the integration of industrialization and industrialization in China,” China Comput. Integr. Manuf. Syst., vol. 25, no. 9, 2019.

[2] T. Ju et al., “Research on the development strategy of industrial Internet to promote intelligent manufacturing,” China Guangxi Qual. Supervision Herald, to be published.

[3] T. Reus, “Eliminate the danger of radio sparks-temperature, pressure and humidity testing instruments with explosion-proof protection function,” China Process Ind., vol. 5, pp. 50–51, 2006.

[4] Q. Qi and F. Tao, “Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison,” IEEE Access, vol. 6, pp. 3585–3593, 2018.

[5] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, “Digital twin-driven product design, manufacturing and service with big data,” Int. J. Adv. Manuf. Technol., vol. 94, nos. 9–12, pp. 3563–3576, Feb. 2018.

[6] P. Liamputtong, “Qualitative data analysis: Conceptual and practical considerations,” Health Promotion J. Aust., vol. 20, no. 2, pp. 133–139, 2009.

[7] N. Blackie, Analyzing Quantitative Data: From Description to Explanation, Newbury Park, CA, USA: Sage, 2003.

[8] U. A. Bakshi and A. V. Bakshi, Electrical Measurements and Instrumentation. Technical Publications, 2009.

[9] M. M. W. Yan, “Accurate detecting concept drift in evolving data streams,” ICT Exp., vol. 6, no. 4, pp. 332–338, Dec. 2020.

[10] S. Delefortrie, P. De Smedt, T. Saey, E. Van De Vijver, and M. Van Meirvenne, “An efficient calibration procedure for correction of drift in EMI survey data,” J. Appl. Geophys., vol. 110, pp. 115–125, Nov. 2014.

[11] R. E. Sharp, “Radiation tolerance assessment of in-reactor equipment (at Sizewell B),” in Proc. Int. Conf. Elect. Control Aspects Sizewell B PWR (IET), Sep. 1992, pp. 38–43.

[12] R. S. Senior, T. D. Bracken, and R. F. Rankin, “EPRI EMF exposure database: University of North Carolina (UNC) electrical utility worker data set: Final report,” Tech. Rep., 1998, pp. 60–67.

[13] W. Liu, H. Zhang, Z. Ding, Q. Liu, and C. Zhu, “A comprehensive active learning method for multiclass imbalanced data streams with concept drift,” Knowl.-Based Syst., vol. 215, Mar. 2021, Art. no. 106778.

[14] S. Ancy and D. Paulraj, “Handling imbalanced data with concept drift by applying dynamic sampling and ensemble classification model,” Comput. Commun., vol. 153, pp. 553–560, Mar. 2020.

[15] P. Zyblyewski, R. Sabourin, and M. Wodniak, “Preprocessed dynamic classifier ensemble selection for highly imbalanced drifted data streams,” Inf. Fusion, vol. 66, pp. 138–154, Feb. 2021.

[16] Y. L. Tong, The Multivariate Normal Distribution. Springer, 2012.

[17] K. H. Teigen and M. Jørgensen, “When 90% confidence intervals are 50% certain: On the credibility of credible intervals,” Appl. Cognit. Psychol., vol. 19, no. 4, pp. 455–473, May 2005.

[18] X. Wang, J. Li, and F. Zhang, “Modeling and identifying of fiber optic gyro random drift data based on time series analysis method,” Torpedo Technol., to be published.

[19] S. Sikström and F. Jönsson, “A model for stochastic drift in memory strength to account for judgments of learning,” Psychol. Rev., vol. 112, no. 4, p. 932, 2005.

[20] N. Almutairi, F. Coenen, and K. Dares, “Third party data clustering over encrypted data without data owner participation: Introducing the encrypted distance matrix,” in Proc. Int. Conf. Big Data Analytics Knowl. Discovery, Cham, Switzerland: Springer, 2018, pp. 163–173.

[21] T. X. Barber and M. J. Silver, “Fact, fiction, and the experimenter bias effect,” Psychol. Bull., vol. 70, no. 6, Pt.2, pp. 1–29, 1968.

[22] P. Zhou, X. Xu, and C. Tu, “The application of AFAL methodology in substantiation of instrument calibration intervals extension in nuclear power plant,” in Proc. Fuel Cycle, Radioactive Waste Manage. Decommissioning: React. Phys. Transp. Theory; Nucl. Educ., Public Acceptance Rel. Issues; Instrom. Controls; Fusion Eng., vol. 5, Jul. 2013, Art. no. V005T13A004.

[23] J. M. Wang, “Survey on industrial big data,” Big Data Res., vol. 3, no. 6, pp. 3–14, 2017.

[24] Q. Liu and S. Z. Qin, “Perspectives on Big Data modeling of process industries,” Acta Automatica Sinica, vol. 42, no. 2, pp. 161–171, 2016.

[25] Z. Gang, Industrial Big Data Technology and Application Practice (2017), Beijing, China: Electronic Industry Press, 2017.

[26] M. E. Stokes, C. S. Davis, and G. G. Koch, Categorical Data Analysis Using SAS Cary, NC, USA: SAS Institute, 2012.

[27] I. P. Verma, Data Analysis in Management With SPSS Software. Springer, 2012.

[28] F. Milano, “A python-based software tool for power system analysis,” in Proc. IEEE Power Energy Soc. Gen. Meeting, Jul. 2013, pp. 1–5.

[29] G. W. Morton, B. D. Shumaker, and H. M. Hashemian, “Cross-calibration in substantiation of instrument calibration intervals extension in nuclear power plant,” in Proc. Fuel Cycle, Radioactive Waste Manage. Decommissioning: React. Phys. Transp. Theory; Nucl. Educ., Public Acceptance Rel. Issues; Instrom. Controls; Fusion Eng., vol. 5, Jul. 2013, Art. no. V005T13A004.

[30] D. Jiang, Y. Song, R. Lian, S. Bao, J. Peng, H. He, and H. Wu, “Familia: A configurable topic modeling framework for industrial text engineering,” 2018, arXiv:1808.03733. [Online]. Available: http://arxiv.org/abs/1808.03733.

[31] F. Xuejun, “Development status and prospects of industrial big data,” Modern Telecomun. Technol., vol. 47, no. 4, pp. 30–33, 2017.

[32] C. Fulin, “Promoting high-quality development requires accelerating power change,” China Small Medium Enterprises, vol. 2018, pp. 68–69, Apr. 2018.

[33] J. Lin and L. Liu, “Research on security detection and data analysis for industrial Internet,” in Proc. IEEE 19th Int. Conf. Softw. Qual. Rel. Secur. Companion (QRS-C), Jul. 2019, pp. 466–470.

[34] Z. Xu, “Research on enterprise knowledge unified retrieval based on industrial big data,” in Proc. 6th Int. Conf. Adv. Cloud Big Data (CBD), Aug. 2018, pp. 130–134.
GUANGSHENG LUO (Student Member, IEEE) received the bachelor’s degree from the Department of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, China, in 2005, the master’s degree from the School of Software, Fudan University, in 2015, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Technology. He is currently a Lecturer with Shanghai International Studies University. His research interests include machine learning, data analysis, and big data application.

CHUNYANG RUAN received the Ph.D. degree from the School of Computer Science, Fudan University, Shanghai, China. He is currently a Lecturer with the Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University, Shanghai. His research interests include graph embedding, graph neural networks, heterogeneous information networks, data analysis, and knowledge discovery.

YU YANG is currently the Vice President with Aishu Information Technology Company Ltd., Shanghai, China. His research interests include machine learning, deep learning, and smart finance.

WENWEI LI is currently pursuing the Ph.D. degree with the School of Computer Science, Fudan University. His research interests include cloud computing, artificial intelligence, and distributed computing.

* * *