Research on Verification Error of Smart Electricity Meter based on Regression Analysis

Yanjie Dai\textsuperscript{1*}, Ding Xu\textsuperscript{1}, Qing Wang\textsuperscript{1}, Jingyuan Zhu\textsuperscript{2}, Tingting Wang\textsuperscript{3}, Yanxi Liu\textsuperscript{3}

\textsuperscript{1}State Grid ShanDong electric power research institute, Jinan, ShanDong 250002, China
\textsuperscript{2}School of Electrical and Electronic Engineering, North China Electric Power University International Education Institute, Beijing, 102206, China
\textsuperscript{3}Shandong Zhongshi Yitong Group Co., Ltd, Jinan, ShanDong, 250002, China

*Corresponding author’s e-mail: 85641088@qq.com

Abstract: In this paper, the errors of smart electricity meter are verified by using walk-in temperature and humidity box. First of all, this paper investigates the regression analysis method and analyzes the influence of temperature and humidity on the error of the smart electricity meter. The stepwise regression method is selected to model the verification error and the verification conditions of the smart electricity meter and the verification device. Finally, the accuracy of modeling is evaluated.

1. Introduction

In recent years, with the world's attention to environmental quality, the concept of low-carbon economy is becoming stronger and stronger. Intelligent management of smart electricity meters has become an inevitable trend of power development. The original data collected by the smart meter has the characteristics of large amount of data, high frequency of collection and diversification of data. Therefore, it is of great practical value for the development of smart grid to establish a relationship model between the verification error of the smart electricity meter and the verification environment conditions.

Scholars at home and abroad have also carried out research on regression analysis modeling. Reference [1] and [2] introduce the analysis of verification data of electric energy meter based on K-means algorithm. Reference [3-5] introduce the evaluation method based on regression analysis. Reference [6-9] introduce the influence of temperature on the error of electric energy meter. Reference [10-12] introduce the current common methods and application status of power data analysis. These researches provide reference for the verification error prediction of smart electricity meters, but there are also shortcomings. They use different modeling methods to analyze the verification error of smart electricity meters, but the impact of the environment on the verification device is not considered.

In this paper, the multiple regression analysis method is used to analyze and study the verification error of the smart electricity meter. The error of the verification device affected by temperature and humidity is introduced. Then the comprehensive verification error model of temperature and humidity impacting is established.
2. Stepwise regression modeling method

The basic idea of stepwise regression is to introduce variables into the model one by one. After introducing an explanatory variable, F-test should be carried out, and t-test should be carried out for the selected explanatory variables one by one. When the original explanatory variables are no longer significant due to the introduction of later explanatory variables, they will be deleted. To ensure that only significant variables are included in the regression equation before each new variable is introduced, until neither significant explanatory variables are selected into the regression equation, nor insignificant explanatory variables are removed from the regression equation.

According to the above thought, the specific steps of gradual regression are as follows:

1. Calculate the mean value \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \bar{y} \) and the sum of the squares of the differences of the variables \( L_{11}, L_{22}, \ldots, L_{pp}, L_{yy} \).

   Note that the standardized variables are:

2. Calculate the correlation coefficient matrix \( R^{(0)} \) of \( x_1, x_2, \ldots, x_p, y \).

3. Suppose \( K \) variables have been selected: \( x_{i_1}, x_{i_2}, \ldots, x_{i_K} \) and \( i_1, i_2, \ldots, i_K \) are different from each other, \( R^{(0)} \) is transformed into \( R^{(k)} = (r_{ij}^{(k)}) \) . Calculate the partial regression square sum of standardized variable \( u_{i_j} \) one by one for \( j = 1, 2, \ldots, k \):

   \[
   V_{i_j}^{(k)} = \frac{(r_{i_j(p+1)}^{(k)})^2}{r_{i_j}^{(k)}}
   \]

   F-test:

   \[
   F = \frac{V_{i_j}^{(k)}}{r_{i_j}^{(k)}}
   \]

   For a given significance level \( \alpha \), the rejection domain is \( F < F_{\alpha}(1, n-k-1) \).

4. Loop through part 3 until \( t \) variables are finally selected \( x_{i_1}, x_{i_2}, \ldots, x_{i_t} \) and \( i_1, i_2, \ldots, i_t \) is different from each other, \( R^{(0)} \) is transformed into \( R^{(t)} = (r_{ij}^{(t)}) \) . The corresponding regression equations:

   \[
   \hat{y} - \bar{y} = \frac{r_{i_k(p+1)}^{(k)}}{L_{yy}} \frac{x_{i_k} - \bar{x}_{i_k}}{L_{i_k}} + \cdots + \frac{r_{i_k(p+1)}^{(k)}}{L_{yy}} \frac{x_{i_k} - \bar{x}_{i_k}}{L_{i_k}}
   \]

   Through algebraic operation, we can get:

   \[
   \hat{y} = b_0 + b_{i_1} x_{i_1} + \cdots + b_{i_t} x_{i_t}
   \]

Stepwise regression method is simple and easy to operate, and the variables of the regression equation are less, and the most significant variables are retained. Moreover, in practice, this method has also been proved to be more effective, and the prediction accuracy is higher; at the same time, there are often interrelations between economic variables, and stepwise regression can modify multicollinearity to a certain extent.

3. Testing Condition

3.1 Testing environment and scheme

The comprehensive verification experiment of smart electricity meters is to put the verification device composed of standard meter, standard clock and power source together with the smart electricity meters to be tested into the experimental environment. The experiment needs to use the walk-in temperature and humidity box, and collect the comprehensive verification error of smart electricity meters under different temperature and humidity conditions.

1. The tested meters and verification device together are put in the walk-in temperature and humidity box. The verification device is disassembled, and the standard meter is out.
(2) Environment control system can adjust the temperature and humidity conditions of the smart electricity meter running at the same time. The system adopts the balanced temperature and humidity control system to adjust the equipment in real time to balance the overall temperature and humidity and form the required environmental conditions. The environment control system can carry out the comprehensive test of temperature and humidity. The range of the comprehensive environmental test of temperature and humidity is \(+10^\circ C\sim+30^\circ C\), and the relative humidity is \(15\%\sim60\%\).

### 3.2 Sample selection

There are 32 groups in this experiment. The smart electricity meter is placed in the temperature and humidity testing box, and the temperature and humidity are combined. The experimental environment temperature is set as \(10^\circ C\), \(15^\circ C\), \(18^\circ C\), \(20^\circ C\), \(23^\circ C\), \(25^\circ C\), \(27^\circ C\), \(30^\circ C\), and the humidity test points are \(15\%, 30\%, 45\%, 60\%\). The standard meter is placed outside the temperature and humidity box in a stable environment. From the experimental results, the comprehensive verification error is less affected by humidity, so it is necessary to further analyze the verification data by comprehensive modeling.

### 4. Comprehensive modeling of verification data

#### 4.1 Stepwise regression model

When the linear regression model is established, the F value and P value pass the reliability test, and the linear regression equation is established. When the F-value of the hypothesis test statistic far exceeds the critical value of F-test (obtained by looking up the F-distribution table), it proves the accuracy of the model, otherwise, it is considered that the linear correlation between the dependent variable and the independent variable is not significant. When the probability p corresponding to \(f\) is less than or equal to \(\alpha\), there is a significant linear correlation between the dependent variable and the independent variable.

According to the step-by-step regression method, the linear regression modeling process is as follows. Firstly, the temperature variable is introduced and the model 1 is established; then the humidity variable is introduced on the basis of the model 1 and the model 2 is established. Because variables aren’t removed during the process, The last regression equation is model 2, which includes temperature and humidity.

| Model | Non-standardized coefficient | Standardization Coefficient | t | Salience |
|-------|-----------------------------|-----------------------------|---|---------|
| Model 1 |                      |                             |   |         |
| Model 2 |                      |                             |   |         |

Figure 1 experimental schematic diagram

![Figure 1 experimental schematic diagram](image-url)
B          Standard error  Beta
1 (constant) -0.002 0.001  -2.088 0.037
  temperature 0.001 0.000  0.388  29.143 0.000
2 (constant) -0.003341 0.001  -3.094 0.002
  temperature 0.001217 0.000  0.388  29.159 0.000
  humidity  0.000038 0.000  0.033  2.489 0.013

From Table 1, the regression equation is as follows:

\[ y = -0.003341 + 0.001217 x_1 + 0.000038 x_2 \]  \hspace{1cm} (5)

Among them, ‘y’ represents comprehensive verification error, ‘x1’ represents temperature, ‘x2’ represents humidity.

After establishing the regression equation, the reliability of the regression equation is shown in Table 2.

| Model | Sum of squares | Freedom | mean square  | F         | Salience |
|-------|----------------|---------|--------------|-----------|----------|
| regression | 0.270       | 1       | 0.270        | 849.307   | 0.000    |
| residual          | 1.527       |         | 0.000        |           |          |
| Total                 | 1.797       | 4799    |              |           |          |
| regression | 0.272       | 2       | 0.136        | 428.210   | 0.000    |
| residual          | 1.525       |         | 0.000        |           |          |
| Total                 | 1.797       | 4799    |              |           |          |

From Table 2, the F=428.210 of model 2 is much larger than 3.313 (from the F distribution table of \( \alpha=0.05 \)). It is considered that the independent variable has a significant effect on the dependent variable. The significance of each variable in model 2 is less than 0.05, so that the dependent variable has a significant impact on the dependent variable.

The independent variables in the model are related to temperature and humidity. However, after the introduction of humidity, the coefficient of regression equation of temperature variable is 0.001217, and the coefficient of regression equation of humidity variable is 0.000038. The humidity coefficient is two orders of magnitude smaller than the temperature coefficient, so that the influence of humidity variable on the model establishment is very small.

4.2 Accuracy analysis of the model

![Figure 2. Normalized residual histogram](image1)

![Figure 3. Normal P-P diagram](image2)
Figure 2 shows the histogram of standardized residuals of data samples, with the mean value of $-3.31 \times 10^{-14}$ and the standard deviation of 1. It can be seen in Figure 2 that the residual conforms to the normal distribution. Figure 3 is a normal P-P diagram, which is a scatter diagram drawn according to the cumulative probability of variable. In Figure 3, the scattered points almost coincide with the straight lines, which further shows that the residual conforms to the normal distribution. Therefore the data used in the experiment is correct, reasonable and effective, which can be used for analysis and modeling.

Mean square error refers to the expected value of the square of the difference between the estimated value of the parameter and the true value of the parameter, which is recorded as MSE. MSE is a convenient method to measure the "average error". MSE can evaluate the change degree of data. The smaller the value of MSE is, the better the accuracy of the prediction model to describe the experimental data is.

$$MSE = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2$$ (6)

The paper uses MSE value of mean square error to analyze the prediction accuracy of linear regression model. MSE of the predicted value of linear regression and the mean error value of the validation set is calculated, as shown in Table 3. MSE is about $8.076 \times 10^{-6}$ which is small, so the modeling accuracy is high.

| Serial number | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
|---------------|-------|-------|-------|-------|-------|-------|-------|-------|
| Mean error    | 0.008008 | 0.01087 | 0.012795 | 0.01675 | 0.0222925 | 0.0279905 | 0.030493 | 0.0366005 |
| prediction    | 0.009399 | 0.015484 | 0.019135 | 0.021569 | 0.02522 | 0.027654 | 0.030088 | 0.033739 |
| Serial number | 9     | 10    | 11    | 12    | 13    | 14    | 15    | 16    |
| Mean error    | 0.00692 | 0.011545 | 0.017025 | 0.018076 | 0.022015 | 0.0279895 | 0.0305615 | 0.0365305 |
| prediction    | 0.009969 | 0.016054 | 0.019705 | 0.022139 | 0.02579 | 0.028224 | 0.030658 | 0.034309 |
| Serial number | 17    | 18    | 19    | 20    | 21    | 22    | 23    | 24    |
| Mean error    | 0.011557 | 0.012732 | 0.01814 | 0.0198715 | 0.022777 | 0.0275725 | 0.0309785 | 0.035558 |
| prediction    | 0.010539 | 0.016624 | 0.020275 | 0.022709 | 0.02636 | 0.028794 | 0.031228 | 0.034879 |
| Serial number | 25    | 26    | 27    | 28    | 29    | 30    | 31    | 32    |
| Mean error    | 0.01508 | 0.0189045 | 0.023197 | 0.0243765 | 0.0254865 | 0.0267355 | 0.0306955 | 0.0371565 |
| prediction    | 0.011109 | 0.017194 | 0.020845 | 0.023279 | 0.02693 | 0.029364 | 0.031798 | 0.035449 |
| MSE           | 8.07648E-06 |

5. Summary
Under the analysis of a large number of test data, the comprehensive error modeling method of smart electricity meter and verification device is studied. The comprehensive modeling method based on stepwise regression is proposed, which lays a solid foundation for the next work. In the follow-up research, we can use the function relationship of multi-dimensional variables to study the decoupling method of the comprehensive error, and propose a decoupling method suitable for the comprehensive error of smart electricity meter under the random disturbance of multiple parameters.

Acknowledgments
This work was supported by the Science and Technology Project of State Grid Corporation under Grant No. 520626180046 and the Science and Technology Project of State Grid Shandong Electric Power Company under Grand No. 520626190043. The paper is completed with the support and help of my colleagues. They have put forward some valuable comments on my research and thesis. I would like to express my gratitude to them.
References

[1] He, A., Zuo, L.B., Wang, X., et al. (2018) Analysis and Research of Verification Error of Electric Energy Meter Based on K-means Algorithm. COMPUTER ENGINEERING & SOFTWARE, 39(6): 64-73.

[2] Xia, W.Z., Teng, H., Cao, M. (2019) Research on evaluation model of three-phase electric energy meter based on K-means++ algorithm. http://kns.cnki.net/kcms/detail/23.1202.TH.20190709.1431.006.html.

[3] Zhou, Y. (2015) reliability life assessment method of intelligent meters based on regression analysis. J. Industrial technology innovation, 02(02):191-195.

[4] Liang, J., Liang, G.M. (2019) short term prediction of low voltage carrier meter reading delay based on multiple regression. J. Industrial measurement. 29(02):16-18.

[5] Xu, Q., Shen, Q.Y., X Guo.X.,Wang, W.P.,Jin, P., Gong, D. (2014) Study on the Lifetime Prediction and Evaluation of Metering Assets Based on the Multivariate Linear Regression Analysis. J. Electrical Measurement & Instrumentation, 51(11):13-17.

[6] Yin, X., Lu, Y.B., Han, D., et al. (2017) The Error Model of Smart Meter under the Influence of Temperature. J. Electrical Measurement & Instrumentation, 54(8): 85-88.

[7] Zheng, J., Chen, L., Yuan, W. (2019) study on the error of electric energy meter and environmental impact factors in typical high altitude environment. J. Electrical measurement & instrument, 56(09):142-147.

[8] Li, F.H., Xia, X.H. (2001) The Influence of Ambient Temperature on Power Meter Verification Results and Their Compensation Measures. J. Metering Technology, 2001(07): 38-39.

[9] Sun, Y., Han, D., Liu, H., et al. (2014) Analysis of Metering Performance of Electronic Energy Meter in the Environment of Low Temperature. J. Heilongjiang Electric Power, 2014(6): 521-523, 528.

[10] Archana, C., Preeti, M. (2019) A bibliometric survey on incremental clustering algorithm for electricity smart meter data analysis. Iran Journal of Computer Science, Vol.2 (4), pp.197-206.

[11] Cai, W.J., Li, H.B., Gong, H. (2019) Supplier evaluation method of smart meter based on analytic hierarchy process. J. Electrical Measurement & Instrumentation, 56(1): 129-135, 143.

[12] Xie, J.J. (2016) Research on the management and data analysis of electric energy meter based on hierarchical model. D. Nanjing University of Posts and Telecommunications.