Precision Qualification of Watthour Meters

John D. Ramboz, Charles P. Fenimore, *Member, IEEE*, and Susannah B. Schiller

***Abstract*—One of the NIST Measurement Assurance Programs [1] transfers the unit of the watthour using transport meters. For this application the response of these meters to variations in environmental conditions must be well-characterized. A statistically planned experiment is employed to determine corrections for the response of each meter to varying conditions of voltage, current, temperature, and power factor. This qualification procedure is designed to be efficient with the number of test points and to yield estimates of the model parameters describing the corrections.**

I. Introduction

ESTABLISHING "traceability" between a measuring laboratory and the National Institute of Standards and Technology, NIST, traditionally involves sending an instrument to NIST for calibration. Alternatively, NIST sends one of its watthour transport standards to the laboratory which performs a prescribed set of measurements and returns the meter with the accumulated data. The measurement results are analyzed by NIST, which reports an offset with which the laboratory corrects its standard to be in agreement with the National Standard. This process is known as a Measurement Assurance Program, MAP, for electric energy.

In such a measurement process, the absolute registration of the transport standard is relatively unimportant. The critical characteristics of the transport standard are its short-term stability and its relative insensitivity to changes in temperature, voltage, and current. Although these changes in instrumental response are small, they are usually statistically significant and may be corrected. Use of this correction provides the best possible calibration of a laboratory's watthour standard with an accuracy which is higher than that obtained by using traditional methods in the transfer of the watthour unit.

This paper describes the use of a statistically planned experiment to determine the effects of variations in temperature, voltage, and current at three power factors on a variety of commercially available watthour meters selected as transport standards. Previously, the qualification process had assumed the instrumental response to be linear [1], [2], and the test points to differ from reference conditions in only one variable at a time. The new set of measurements determines effects of nonlinearities and interactions between factors on the response. The test points are distributed over a multidimensional design region and are chosen to determine the statistically significant corrections with minimum variance and with little additional effort.

Because the qualification process is applied to several instruments, a two-stage approach may be used. The instruments are grouped by manufacturer and design, designated as Groups A, B, C, and D. Only one device from each group is qualified in the first stage of the process. The characteristics of the meter are represented by a mathematical model, which may be more complex than necessary. An experiment design is chosen to permit estimation of the parameters of the model.

Once the parameters of this complex model have been determined, it is possible to refine the model to include only the statistically significant terms. This refined model is used in the second stage when additional watthour meters of the same type are to be characterized. Assuming that the responses from the uncharacterized watthour meters will follow the reduced model, a smaller design can be chosen and the second set of watthour meters can be characterized with less effort. At both stages the design points are chosen to produce optimal estimates of the parameters of the particular model being studied.

II. Experimental Plan

To "qualify" a watthour meter for service as a MAP transport standard, NIST characterizes the meter's response, as defined by the percent registration, R, for the range of conditions over which it will be used. The variables which affect the response of the watthour meters are temperature, voltage, current, and power factor.

Power factor takes on three values: 0.5 lag, 0.5 lead, and unity. The voltage, current, and temperature vary continuously, defining three design volumes, one for each power factor (Fig. 1). The registration can be represented as a surface over the design volume. As shown in Fig. 2, a surface in three dimensions represents a section of the response surface with changing voltage (V) and temperature (T) and fixed current (I) and power factor (PF). The correction to be applied to the registration is the difference between R at the reference conditions and at the conditions actually encountered.

For each transport standard, the response surface should simply and adequately describe its behavior. To determine a response surface, the following process is used. First, a candidate model for the surface is proposed. Next, measurements are taken in the region described by the variables of interest at test points selected to provide good estimates of the parameters in the candidate model. Then, using least squares, a surface (which may be simpler than the candidate) is fitted to the data.

We next consider test point and model selection. The model for the system response is specified in terms of deviations from reference conditions of voltage (ΔV), current (ΔI), and temperature (ΔT), respectively. The candidate model includes linear, quadratic, and cubic terms in ΔV, ΔI, ΔT to test for nonlinearities in the response. It also includes cross-products between pairs of factors to test for interactions.

Thus, the model to be fitted is

$$R = a_1 + a_2\Delta T + a_3\Delta I + a_4\Delta V + a_5\Delta T^2 + a_6\Delta T^3 + a_7\Delta V^2 + a_8\Delta V^3 + a_9\Delta I^2 + a_{10}\Delta I^3 + a_{11}\Delta V\Delta T + a_{12}\Delta I\Delta T + a_{13}\Delta I\Delta V + \text{error}. \tag{1}$$

Manuscript received June 14, 1990.

The authors are with the National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899.

IEEE Log Number 9042436.

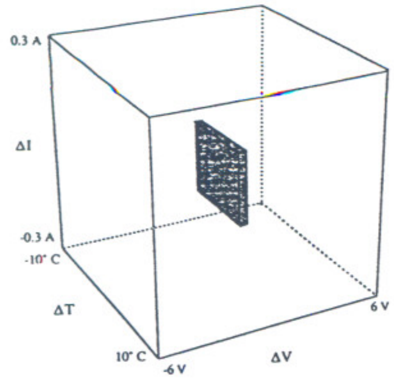

Fig. 1. The three-dimensional design region surrounds the typical-use region. This provides improved estimates of the model parameters. The use region for temperature, ΔT, is $[-1, +5]$°C, for current, ΔI, is $[-0.1, +0.1]$A, and for voltage, ΔV, is $[-0.1, +0.1]$V.

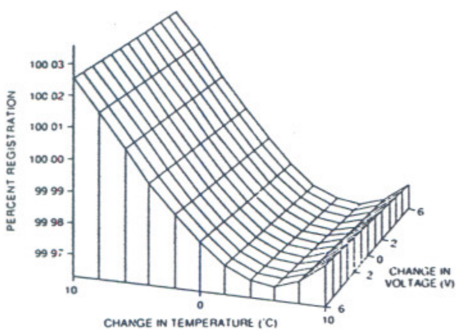

Fig. 2. The cubic model for percent registration displays a typical response surface.

The error in each measurement is assumed to be independent of the error in all other measurements, and to have a zero mean.

Varying one factor at a time, it is not possible to determine whether interaction effects are present. Therefore, a "box-star" experimental design [3] is employed which contains 15 points (Fig. 3). The design region is: $(-6 \le \Delta V \le 6)$V, $(-0.3 \le \Delta I \le 0.3)$A, and $(-10 \le \Delta T \le 10)$°C, where the reference values are 120 V, 5 A, and 25°C, respectively. Eight of the test points are placed at outer corners of the region: (± 6 V, ± 0.3 A, ± 10°C). One test point is placed at the center, and six more points are placed on the axes a little beyond the region of interest: (± 10 V, 0, 0), (0, ± 0.5 A, 0), and (0, 0, ± 15°C). Replicate measurements are made at each point. By spreading the design over this larger region, the expected error in the parameter estimates is reduced and the model is made applicable over the entire region.

Box-star designs are often employed in "response-surface" problems because they are balanced, symmetric, and permit estimation of higher-order terms to test for model bias. In this study, the percent registration response, R, is defined over the three-dimensional design region which permits estimation and tests of significance for two-way interactions between the variables and linear, quadratic, and cubic terms for each. Although this box-star design has more design points than parameters to be estimated, fifteen points were used to preserve balance in the design and to permit a test of lack-of-fit of the model.

III. Model Selection and Reduction

In the first run, four transport standards—one from each group of instruments—are qualified using data gathered according to the experimental plan. The polynomial model in (1) is then fitted by least squares for each power factor for each watthour

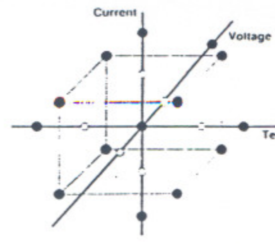

Fig. 3. The box-star design with 15 points permits estimation of interaction terms and terms up to third order. It also allows testing for lack-of-fit.

meter. As anticipated, this model is more complex than is necessary to describe the data because many of the parameter estimates are not significantly different from zero. Therefore, a stepwise linear regression procedure is used to determine the smallest subset of the complex model that adequately (in terms of statistical significance) describes the data. At each step, the procedure calculates an F-statistic for each term that reflects the contribution of that terms to the model if it is included [4]. The procedure adds the term that has the largest statistic and stops when none has a significance level greater than a preselected threshold. For two of the four meters tested, it is found that the simple linear model

$$R = a_1 + a_2\Delta T + a_3\Delta I + a_4\Delta V + \text{error} \tag{2}$$

is, in fact, adequate because no other terms are statistically significant.

For the other two meters, a variety of terms in (1) are found to be statistically significant. The terms which are important vary by meter and power factor. For example, for the Design-B meter, the following effects were found to be statistically significant: at lag power factor: ΔT, ΔV, ΔT^2, ΔT^3 and ΔV^2; at unity: ΔT, ΔV, ΔI, ΔT^2, ΔI^2, and ΔV^3; and at lead: ΔT, ΔV, ΔT^2, ΔI^2, ΔT^3, ΔV^3, and $\Delta V\Delta T$. However, as discussed earlier, the design region used to determine the parameters is larger than the typical region of use. Thus, even though a term might be statistically significant in the design region, in the use region it might be negligible when weighed against systematic error known to exist in the system.

This fact is exploited in developing a heuristic scheme to reduce the model. For each meter and power factor, the three models are fitted. The "fullest" model that could be fitted is given by (1). Only two reduced models are considered for simplicity's sake, a "linear" model (see (2)) and a "cubic" model (see (3)). The "cubic" model was selected because its terms appeared most frequently in the models found by the stepwise regression procedure.

$$R = a_1 + a_2\Delta T + a_3\Delta I + a_4\Delta V + a_5\Delta T^2 + a_6\Delta T^3 + \text{error}. \tag{3}$$

The maximum absolute difference between the corrections to reference conditions predicted by the full and reduced models is examined. If the difference between the corrections calculated from the full and reduced models is less than one-half the standard deviation of the random error in the measurements (estimated by the root-mean-square error (RMS) of the full model), then the reduced model is deemed "adequate." Although arbitrary, this criterion is an effective discriminator between the models.

Results for Design-B meter are in Table I. The "linear model" is inadequate for the lag and lead power factors. However, the relative bias introduced by the "cubic" model is at

TABLE I
THE MAXIMUM BIAS (ABSOLUTE VALUE) OF THE VARIOUS MODELS IS USED TO SELECT THE SIMPLEST, ADEQUATE MODEL. (A MODEL IS CONSIDERED ADEQUATE IF THE MAXIMUM BIAS IS LESS THAT 50% OF THE RMS ERROR FOR ALL THREE POWER FACTORS.)

		Model $\Delta T\ \Delta T^2\ \Delta T^3\ \Delta V\ \Delta I$		Model $\Delta T\ \Delta V\ \Delta I$	
Power Factor	RMS Error *ppm*	Bias *ppm*	Bias *relative*	Bias *ppm*	Bias *relative*
Lag	66	5	8%	37	56%
Unity	20	1	5%	8	40%
Lead	64	10	16%	37	58%

most 16% of the standard deviation of the random error. Therefore, the "cubic" model is chosen for all three power factors for the selected meter. None of the meters requires the full model.

IV. Fine-Tuning the Experiment

The box-star design probably provides much more information than is necessary. This conclusion may be drawn from the fact that the only terms in any of the models for the first four meters are ΔT, ΔI, ΔV, ΔT^2, and ΔT^3. Given any particular model, it is possible to select a design which is "optimal" for estimating model parameters, but this is only sensible if the form of the correct model is known. Four more watthour meters are also characterized in this program and the same reduced model is used since the additional meters are similar to the original meters qualified. This is an important assumption whose validity is necessary for the results below to make sense.

The statistical literature offers a variety of criteria for selecting optimal designs for a specific model. The optimality criteria which we consider include minimizing the variance of parameter estimates ("D-optimality"), minimizing the maximum variance of the predicted responses over the design region ("G-optimality"), and minimizing the average variance of the predicted responses over the design region ("V-optimality"). Because it has been shown that D-optimal designs perform very well with respect to other criteria and because D-optimality has been studied extensively, we employ it as the criterion [5].

A D-optimal design consists of the set of n test points (here n is reduced from the original 15 points to 9 points) in the design space which will permit estimation of the parameters of a given model with minimum variance. Specifically, let X be the design matrix. Each row in X corresponds to the terms in the model (e.g., ΔT, ΔI, ΔV, ΔT^2, and ΔT^3) evaluated at a single design point. The variance-covariance matrix of the parameters estimates is proportional to the inverse of the matrix $X'X$. An n-point design which maximizes the determinant of this matrix, $|X'X|$, for all n-point designs is said to be D-optimal. To find a D-optimal design, an iterative computer search is used. Several commercially available experiment design programs include code for finding D-optimal designs.

A D-optimal experiment design is used in the design region for the second set of four watthour meters. The model in (3) is used for selecting the design because it is the largest model needed for the first four meters. The D-optimal design for this model is shown in Fig. 4. The center point of the region is included in the design, both because the goal of the model is to correct to nonimal conditions (those at the center of the region) and to detect potential systematic error in the model in the form of higher order effects. The heuristic described above is applied to determine the simplest adequate model for each of these meters as well.

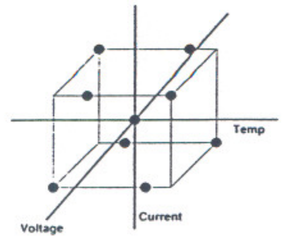

Fig. 4. The D-optimal design with an added ninth point at the center is used with the reduced models. The design permits estimation of cubic terms in temperature and linear terms in current and voltage.

V. Measurement and Analysis Results

Of the eight meters characterized, four are described adequately by the linear model (2), while the other four require the cubic model (3). Four of the meters are also characterized around the nominal voltage of 240 volts, for which the D-optimal design was used to gather data. Two surprising discoveries were made about the meters at higher voltage. First, the parameter estimates for the same model are statistically significantly different for the two nominal voltages for each meter. Second, the linear model is deemed adequate for all four meters around 240 volts, even though the "cubic" model had been necessary for two of these meters around 120 V. This is perhaps due to the fact that the "full" model, to which the simpler models are compared, is based on fewer measurements at 240 V than at 120 V.

Table II gives the results for the watthour meters tested at both 120 and 240 V. The parameter estimates given are the coefficients for the appropriate terms in (2) and (3), and are given in "ppm per unit." For example, for the voltage coefficient terms (coefficient a_4), the values are in ppm/volt. Likewise, for the current and temperature coefficients (a_3, and a_2, a_5, and a_6), the values are in ppm/A and ppm/°C, ppm/(°C)^2, and ppm/(°C)^3, respectively.

When used to refine the values in the calibration data in the MAP process, corrections less than a few ppm are meaningless because of systematic error in the system. All corrections are included in the table to demonstrate the range of values obtained for the meters tested. For example, the temperature influence is generally the largest of all factors. Examination of the data shows that the temperature coefficient for the Design-A meter has values in the range of 20 ppm/°C at 240 V, whereas the Design-D meter has coefficients of less than 5 ppm/°C.

Another interesting observation is that the coefficients for the current parameter can be quite different when the meter is being used on either 120 V or 240 V. One might initially think that the current coefficients would be independent of the input voltage, but for the large change from 120 to 240 V there is a voltage dependence. This may be due to nonlinearities in the multiplier portion or to interactions between the voltage and current portions of the meter.

VI. Conclusions

The qualification of watthour meters for the MAP program requires that the response of these devices be found with high precision over a range of conditions of temperature, voltage,

TABLE II
CORRECTION COEFFICIENTS FOR 4 WATTHOUR METERS TESTED AT 120 AND 240 V. (THE CONDITION OF TYPICAL USE ARE $-1°C \leq \Delta T \leq +5°C$, $\Delta I \sim \pm 0.1A$, and $\Delta V \sim \pm 0.1$ V

		Parameter Estimates for (2) or (3) (in ppm/unit)									
		ΔT (a_2)		ΔI (a_3)		ΔV (a_4)		ΔT^2 (a_5)		ΔT^3 (a_6)	
Meter Design	Power Factor	Nominal Test Voltage									
		120	240	120	240	120	240	120	240	120	240
A	0.5 Lag	+11.0	+19.0	+29	110	+4.8	+0.2	—	—	—	—
	1.0	+13.0	+24.0	+130	+100	−1.6	−1.2	—	—	—	—
	0.5 Lead	+15.0	+26.0	+400	+300	−11.0	−6.2	—	—	—	—
B	0.5 Lag	−28.0	−7.0	−19	+20	−3.3	+0.2	+0.37	—	+0.060	—
	1.0	−7.3	−17.0	+65	+19	−0.1	−1.7	+0.42	—	−0.003	—
	0.5 Lead	−27.0	−4.4	+7	+29	−0.3	−0.3	+0.78	—	+0.027	—
C	0.5 Lag	+7.6	+8.6	−36	+43	+2.4	+2.6	0.00	—	+0.003	—
	1.0	+11.0	+12.0	−36	−7	−1.5	−0.1	−0.20	—	+0.011	—
	0.5 Lead	+16.0	+18.0	+61	+130	−1.2	+0.7	−0.26	—	+0.004	—
D	0.5 Lag	+0.2	+0.2	+23	+31	0.0	+0.3	—	—	—	—
	1.0	−2.1	−2.7	+14	−0.4	0.0	−0.1	—	—	—	—
	0.5 Lead	−4.6	−4.7	+8	+21	−0.3	−0.4	—	—	—	—

and current. Using a statistically planned experiment, a significant increase in the information previously available about each meter is achieved without a large increase in the effort expended. The statistical design and analysis provide a systematic method for characterizing the varied response characteristics of individual meters.

Values of the parameter estimates (coefficients) are determined and found to be quite different for separate meters of the same model, and also significantly different when operated on the 120- or 240-V ranges. Differences are also noted for the three power factors used in this investigation.

In practice, overall corrections to reported MAP data of 30 to 50 ppm or more are common. Although these corrections are not large, they are a "trim adjustment" to the data and generally represent a worthwhile refinement. There is pressure to continually improve the calibration process from the electric utility and meter communities. Applying these techniques is one such improvement.

ACKNOWLEDGMENT

The authors wish to thank Ms. Dawn McBrien and Tom Nelson who collected data for this investigation.

REFERENCES

[1] N. M. Oldham, *A Measurement Assurance Program for Electric Energy*, Nat. Bur. Stand. (U.S.), Tech. Note 930 (Sept. 1976).

[2] J. D. Ramboz and R. C. McAuliff, *A Calibration Service for Wattmeters and Watthour Meters*, Nat. Bur. Stand. (U.S.), Tech. Note 1179 (July 1983).

[3] D. C. Montgomery, *Design and Analysis of Experiments*. New York: Wiley, pp. 462ff, 1984.

[4] S. Weisberg, *Applied Linear Regression*, 2nd ed. New York: Wiley, pp. 211–215, 1985.

[5] T. J. Mitchell, "An algorithm for the construction of D-optimal experimental designs," *Technometrics*, vol. 16, pp. 203–210, 1974.