

# Transformer Thermal Modeling: Improving Reliability Using Data Quality Control

Daniel J. Tylavsky, *Senior Member, IEEE*, Xiaolin Mao, *Student Member, IEEE*, and Gary A. McCulla, *Member, IEEE*

**Abstract**—Eventually, all large transformers will be dynamically loaded using models updated regularly from field-measured data. Models obtained from measured data give more accurate results than models based on transformer heat-run tests and can be easily generated using data already routinely monitored. The only significant challenge to use these models is to assess their reliability and improve their reliability as much as possible. In this work, we use data-quality control and data-set screening to show that model reliability can be increased by about 50% while decreasing model prediction error. These results are obtained for a linear model. We expect similar results for the nonlinear models currently being explored.

**Index Terms**—ANSI C57.91, top-oil temperature, transformer, transformer thermal modeling.

## I. INTRODUCTION

EVENTUALLY, all transformers greater than about 20 MVA will be loaded using dynamic thermal models and these models will be derived from measured field data rather than from the data contained in heat-run reports. The comparison of the top-oil temperature (TOT) performance of the traditional (Clause 7) American National Standards Institute (ANSI)/IEEE model [1] versus a model derived from field-measured data shows why this will be the case: even simple linear models derived from field-measured data are more accurate than the ANSI Clause 7 model using parameters taken from transformer test (heat-run) reports. (See Figs. 1 and 2 for this comparison. Also see the Appendix for a description of this transformer and thermal sensors used in generating these data.) Indeed, models derived from measured field data—data which utilities already routinely monitor and record—naturally account for many phenomena in operating transformers (operational faults, such as fouled heat exchangers, inoperative pumps/fans, etc.) that the nominal Clause 7 model does not. And it is not a simple matter to detect the type of operational fault that has occurred and then, for the Clause 7 model, adjust the model appropriately.

A recent flurry of activity is sorting out which model(s) will eventually be used to predict the TOT. The dust is a long way from settling. Predicting in the long run which model(s) will survive and which will perish is a risky business. But this much is

Manuscript received June 20, 2005; revised September 19, 2005. This work was supported in part by the Salt River Project, in part by Arizona Public Service, and in part by PSERC.

D. J. Tylavsky and X. Mao are with Arizona State University, Tempe AZ 85287-5706 USA (e-mail: tylavsky@asu.edu). Paper no. TPWRD-00359-2005.

G. A. McCulla is with the Salt River Project Apparatus Engineering Department, Tempe, AZ 85072.

Digital Object Identifier 10.1109/TPWRD.2005.864039

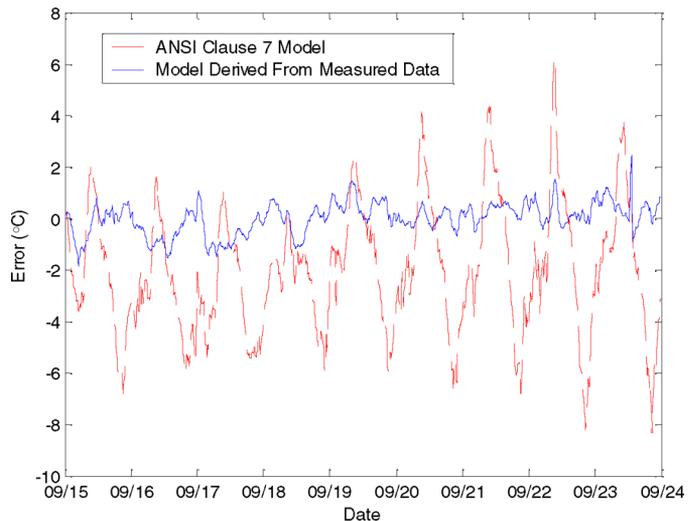


Fig. 1. Comparison of the ANSI Clause 7 model and a linear model derived from field-measured data. (Color version available online at <http://ieeexplore.ieee.org>.)

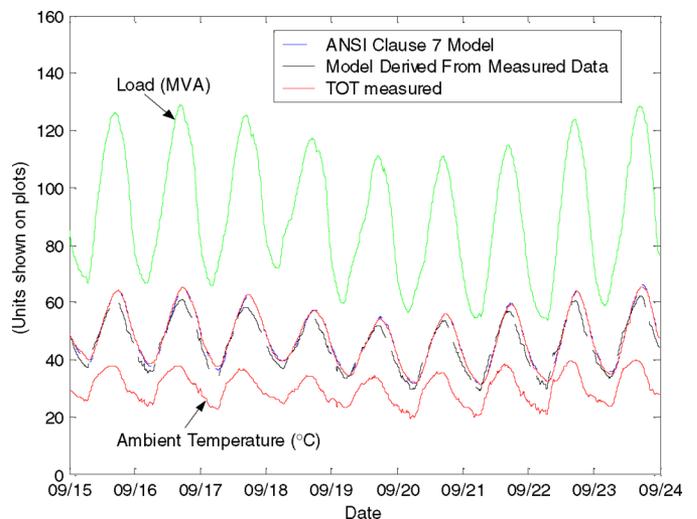


Fig. 2. Load, ambient temperature, measured TOT, and the TOT predicted by the ANSI Clause 7 model and a model derived from field-measured data. (Color version available online at <http://ieeexplore.ieee.org>.)

relatively certain: the traditional (Clause 7) ANSI model [1] (aka top-oil model) will not survive in its current form for two reasons: First, it does not accurately model the dynamic behavior of ambient temperature [2], [3]. Second, (and it is not yet clear how significant this error is), [4]–[6] have shown from first principles, that the placement of the exponent  $n$  used in the ANSI model [1] is, at best, suboptimal if its role is to account for the change in heat transfer under various cooling modes. (Although

[1] states that  $n$  is an “empirically derived exponent” selected “for each cooling mode to approximately account for changes in resistance with a change in load.” So the empirical  $n$  of [1] accounts for both changes in resistance and heat transfer under various cooling modes, and cannot be compared directly with the exponent of [4] and [5].)

In addition to these potentially viable first-order models [2]–[6], there are some higher order linear models [7]–[9], which may also prove eventually important; although, in order to determine the parameters for these higher order models, more quantities must be monitored (beyond load current, ambient temperature, and top oil temperature). While such monitoring may become routine in the future, it is rare today. Exactly which model and parameter estimation procedure will ultimately become standard in the industry is anyone’s guess at this point.

There is evidence that regardless of which model is used—whether first order linear [2], [3] or one of three different nonlinear models (Annex G of [1], [4], or [6])—determining parameters will involve an ill-conditioned process. By this, we mean that small changes in field-measured values can lead to large changes in model parameters. We have observed this when working with linear models and have observed a similar phenomenon when working with the nonlinear model of [4]. We also suspect such is the case with the Annex G model; although [10] does not report on issues of stability of the parameter estimation process. For linear modeling, ill-conditioning is well known, and referred to as multicollinearity.

The practical results of multicollinearity in linear models (ill conditioning in nonlinear models) are that a wide range of coefficients will have similar performance when measure against predicted values. That is, the prediction error of TOT, for example, may be small for a wide range of coefficients; however, this does not necessarily mean that all sets of coefficients will extrapolate to nearly identical values when the load or ambient temperature are significantly different from the conditions used to train the models.

Ill-conditioning is a problem when the input (measured) data sets are noisy; hence, it is problem for all data sets and, therefore, all models. In this paper, we propose an algorithm for performing data-quality control on the field-measured data used for transformer thermal model creation. We also propose a method of data-set screening which can eliminate data sets with poor-quality data. We then assess the improvement these procedures have on the reliability of the linear model, as proposed and developed in [2] and [3].

Data-quality control (QC) and data-set screening have a similar effect on nonlinear and linear models. In this work, we limit our investigation to linear models for simplicity.

## II. MODEL DESCRIPTION

The traditional ANSI top-oil-rise (Clause 7) model [1], is governed by the differential equation

$$T_o \frac{d\theta_o}{dt} = -\theta_o + \theta_u \quad (1)$$

which has the solution

$$\theta_o = (\theta_u - \theta_i)(1 - e^{-(t/T_o)}) + \theta_i \quad (2)$$

where

$$\theta_u = \theta_{fl} \left( \frac{K^2 R + 1}{R + 1} \right)^n \quad (3)$$

$$T_o = \frac{C\theta_{fl}}{P_{fl}} \quad (4)$$

and where

- $\theta_o$  top-oil rise over ambient temperature (°C);
- $\theta_{fl}$  top-oil rise over ambient temperature at rated load (°C);
- $\theta_u$  ultimate top-oil rise for load L (°C);
- $\theta_i$  initial top-oil rise for  $t = 0$  (°C);
- $\theta_{amb}$  ambient air temperature (°C);
- $T_o$  time constant (h);
- $C$  thermal capacity (MWh°C);
- $P_{fl}$  total loss at rated load (in megawatts);
- $n$  (oil exponent) an empirically derived coefficient selected for each cooling mode to approximately account for a change in resistance with load;
- $K$  ratio of load L to rated load;
- $R$  ratio of load loss to no-load loss at rated load.

The TOT is then given by

$$\theta_{top} = \theta_o + \theta_{amb} = (\theta_u - \theta_i)(1 - e^{-(t/T_o)}) + \theta_i + \theta_{amb}. \quad (5)$$

The simplest of the linear models [2], [3], [7]–[9] and the top-oil model [2], [3] (cf. top-oil-rise model) corrects the dynamic limitations of the top-oil-rise model by including in (1) the dependence of the time-rate-of-change of  $\theta_{top}$  on ambient temperature  $\theta_{amb}$

$$T_o \frac{d\theta_{top}}{dt} = -\theta_{top} + \theta_{amb} + \theta_u. \quad (6)$$

To obtain a discrete-time model, we discretize (6) by applying the backward Euler discretization rule

$$\frac{d\theta_{top}[k\Delta t]}{dt} = \frac{\theta_{top}[k\Delta t] - \theta_{top}[(k-1)\Delta t]}{\Delta t} \quad (7)$$

to yield (with the assumption that  $n \approx 1$ )

$$\theta_{top}[k] = K_1 I[k]^2 + K_2 \theta_{amb}[k] + (1 - K_2) \theta_{top}[k-1] + K_3 \quad (8)$$

where  $K_1 - K_3$  are functions of the differential equation coefficients [3], and  $I[k]$  is the per-unit transformer current (based on the transformer’s rating) at time-step  $k$ . Equation (8) is a linear model with three coefficients. The least squares method can be used to obtain the model coefficients that best fit the measured data (rather than using the formulae for the  $K_x$ ’s from test report data).

## III. DATA-QUALITY CONTROL TECHNIQUES

It is impossible to remove all noise in the field data. For example, there is transducer/measurement noise, the presence of unmodeled or inaccurately modeled nonlinearities, the absence of unmeasured exogenous driving variables, etc; however, through the judicious use of data-quality control, some noise can be eliminated, improving the reliability of the model.

### A. Identifying Bad Data

The purpose of the data-quality control technique is to identify those data that degrade reliability and then remove them from the input data set. The data to be identified include

- 1) measurements flagged as erroneous in the input data set;
- 2) ambient temperature ( $\theta_{amb}$ ) out of range;
- 3) spikes in TOT;
- 4) rapid and large jumps in  $\theta_{amb}$ ;
- 5) large discrete load changes;
- 6) data with incorrect cooling mode information.

Of the above types of data, types 1, 2, and 3 represent erroneous measurements; type 4 and 6 are likely to be a valid measurement but indicate a change in the cooling mode of the transformer; and type 5 are usually good measurements but sampled at an insufficient sampling rate.

1) *Data Flagged as Erroneous in the Input Data Set:* Typically, many utilities do some rudimentary preprocessing of transformer measurements and allow for a quality-control (QC) flag in their data file (usually a 0 or -1) to identify data that they believe to be erroneous. Through experimentation, we have found that if there are less than 2.5 consecutive hours of erroneous measurements, our model will not be significantly affected if we use the linear interpolation to estimate the TOT or  $\theta_{amb}$  measurement during the time interval in question; for more than 2.5 consecutive hours, the bad measurements are discarded.

2) *Ambient Temperature Out of Range:* In some data sets we have received for transformers in Phoenix AZ, we have noticed  $\theta_{amb}$  significantly below 15 °F. Since the lowest recorded temperature locally was 17 °F, and the highest was 122 °F, we required all  $\theta_{amb}$  data to be between 15 °F and 125 °F. While these values are specific to Phoenix, such temperature ranges are easy to define for other locations.

3) *Spikes in TOT Data:* We have observed situations where TOT may change by 4 °C or more in 15 min. Since this is not physically possible, these measurements must be considered as errors. We correct these spikes by treating them as bad data and linearly interpolating between the adjacent TOT data points. In our data QC algorithm, we remove those spikes in which the TOT changes by  $\pm 0.5^\circ\text{C}$  or more in 15 min and then jumps back by 0.5 °C or more in the next 15 min. Figs. 3 and 4 give a comparison between the predicted results obtained before and after removing the TOT spikes. These figures use data taken from the Corbell transformer under FA cooling. (Note that the error plotted in these figures is multiplicatively scaled by a factor of 5, so a 5 °C error on the plot represents a 1 °C prediction error. Also, for the sake of convenience, the error plotted in these two figures is Predicted TOT–Measure TOT.) Fig. 5 gives a comparison between the error duration before (solid line) and after (dashed line) removing the TOT spikes. The error duration curve gives the amount of time (in per unit) that a given value of error exceeds.

4) *Rapid and Large Jumps in Changes in  $\theta_{amb}$ :* In many cases, we have found large prediction errors occur when there is a rapid change in ambient temperature, as shown in Fig. 6. This figure shows many curves; let us explain what these curves are as follows.

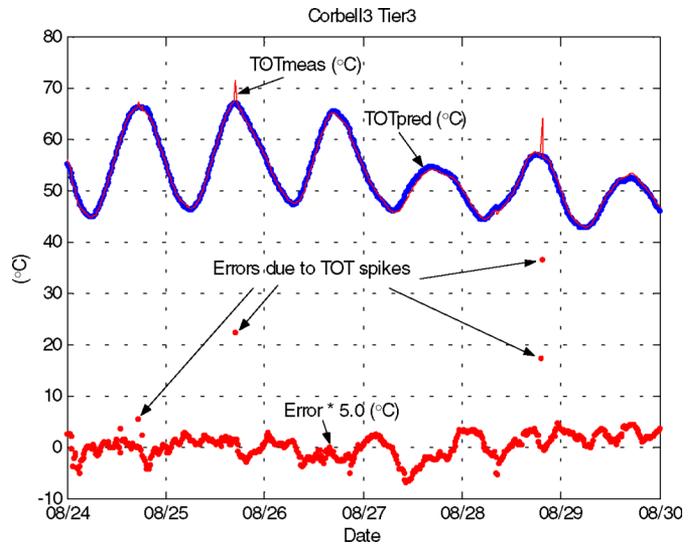


Fig. 3. Prediction errors before removing the spikes in TOT data. (Color version available online at <http://ieeexplore.ieee.org>.)

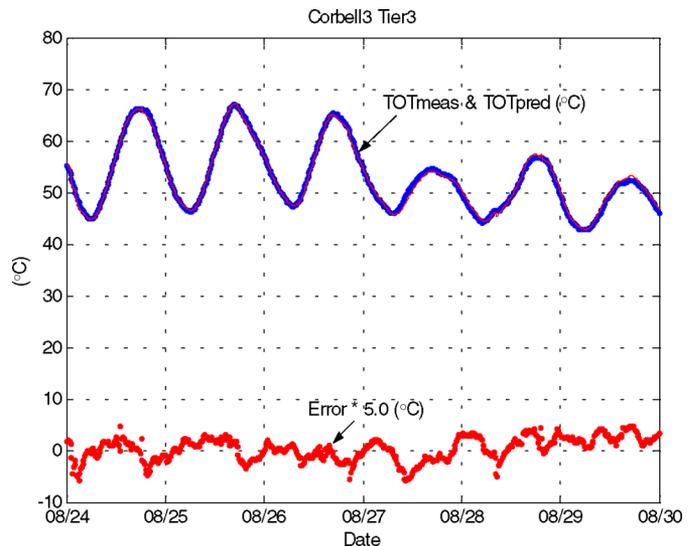


Fig. 4. Prediction errors after removing the spikes in TOT data. (Color version available online at <http://ieeexplore.ieee.org>.)

- The “QC flag” (labeled) indicates good data when equal to 70, and bad data when equal to 40. (Fig. 6 shows only good data.)
- Measured TOT is labeled “TOTmeas.”
- Calculated/predicted TOT is labeled “TOTpred.”
- TOT prediction error (multiplied by 5) is labeled “Error\*5.”
- Measured precipitation is labeled “Rain.”

In Fig. 6, there are two places (marked) where there is a quick drop in the ambient temperature. At both times, some amount of rain was recorded. Notice that in one location, there is a steep positive increase in the prediction error while at the other location there is none. (Error is defined as Predicted-TOT–Measured-TOT). The presence of rain indicates evaporative cooling, which means the transformer is experiencing a different thermodynamic cooling mode than the one we are modeling.

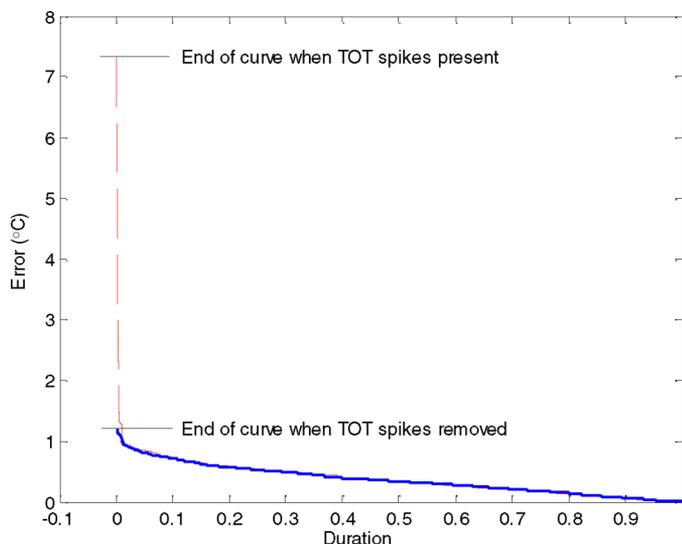


Fig. 5. Error duration curve before and after removing the spikes in TOT data. (Color version available online at <http://ieeexplore.ieee.org>.)

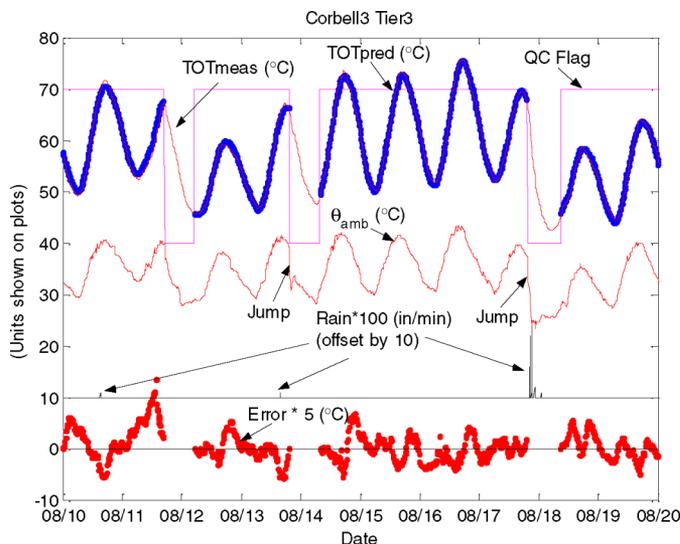


Fig. 7. Example of discarding data affected by quick changes of  $\theta_{amb}$ . (Color version available online at <http://ieeexplore.ieee.org>.)

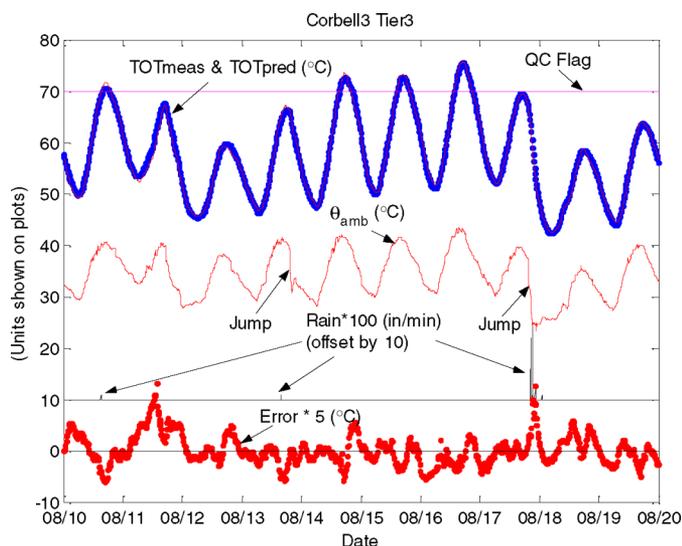


Fig. 6. Example of obtaining large errors due to a quick change of  $\theta_{amb}$ . (Color version available online at <http://ieeexplore.ieee.org>.)

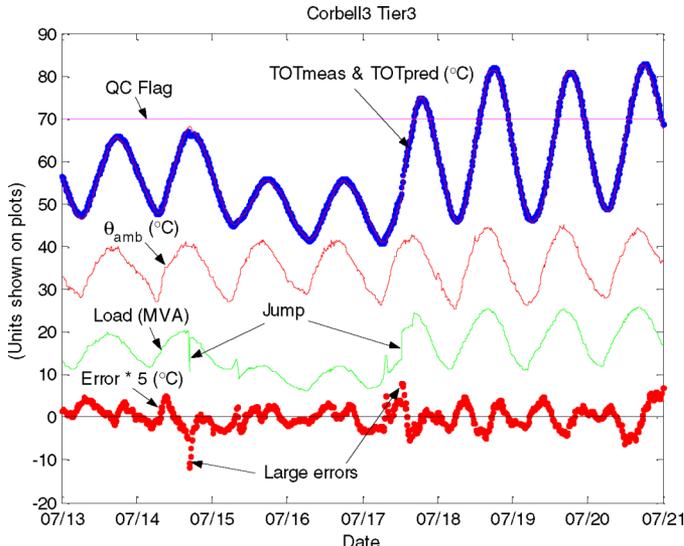


Fig. 8. Example of large errors due to a rapid change of load. (Color version available online at <http://ieeexplore.ieee.org>.)

The solution to this problem is to eliminate data during times of rain; however, we unfortunately do not usually have local rain data nor are the rain data we have reliable. (This is the reason a steep increase in error is not in evidence at the second location in Fig. 6.)

Because a quick drop in temperature is a product of the rain process—and since we have reliable temperature data—we use quick drops in ambient temperature as a marker for rain. We discard 12 h of data after each quick ambient-temperature change to allow the transformer to completely dry before we begin again to use the transformer’s data. Fig. 7 illustrates the prediction results after discarding the data affected by quick changes in ambient temperature (QC Flag = 40 indicates time intervals with discarded data). The large peak error in Fig. 6 has been eliminated in Fig. 7. Note that we have eliminated one interval in Fig. 7 where rain may not have occurred. Such unintended

consequences of this algorithm create no problems, other than shrinking the size of the data set.

5) *Large Discrete Load Changes:* We have observed that large discrete changes in load tend to produce large errors. Fig. 8 illustrates this point. Where the load profile is labeled “jump” on July 14, the load decreases from about 20 to 10 MVA in 15 min and jumps back to about 20 MVA at the next 15-min sample. (Note that we sample only every 15 min, a sampling rate we have found to be sufficient.) The prediction error here shows that the real TOT is much higher than the predicted TOT (i.e., the model under-predicts). In some other places where the load has an abrupt increase, the model over-predicts. When load drops by 10 MVA for one sample, we do not know whether the decrease occurred for a minute or for 29 min. We model this as a 15-min load drop, which may be too short for the specific outage shown in Fig. 8, or may be too long.

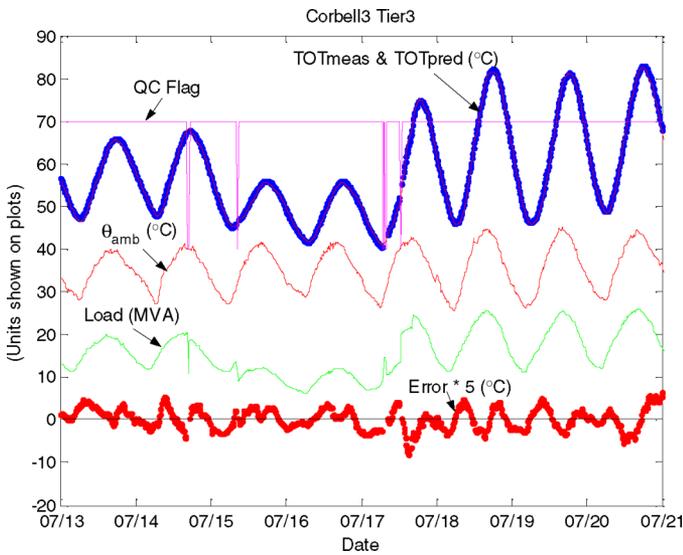


Fig. 9. Example of discarding data affected by large discrete load changes. (Color version available online at <http://ieeexplore.ieee.org>.)

Our solution to this problem is to discard the data that are affected by those abrupt load changes. For the above case, we would discard 30 min worth of data. Fig. 9 illustrates that, after discarding the data affected by large discrete load changes, the large peak errors in Fig. 8 have been eliminated.

6) *Problems Identifying the Cooling Mode:* Each mode of cooling, OA, FA, FOA, represents a different thermodynamic cooling condition and must be represented by a different transformer thermal model. Although the Salt River Project and Arizona Public Service are most interested in the top rating of their transformers, we have used our modeling procedure to predict performance under OA cooling as well. To develop models for each cooling mode, we divide the measured data into tiers. Tier 1 is defined as no fans on, tier 2 means some—but not all—fans are turned on, and tier 3 means all fans are on and/or oil-circulating pumps. We use the QC Flag to identify Tier 1, 2, and 3 data according to the following:

- QC Flag = 70: Tier 3 data;
- QC Flag = 60: Tier 2 data;
- QC Flag = 50: Tier 1 data;
- QC Flag = 40: discarded data.

Ideally, we have telemetered cooling fan/pump contactor data that indicates when fans/pumps turn on and off. Rarely is this the case. More often, we use design settings for the turn-on/turn-off temperatures (either TOT or HST) of the fans/pumps to determine the cooling mode of the transformer for any given TOT or HST temperature. Our experience with the field data shows that this information is sometimes unreliable.

Take Arcadia bay-2 for example. The Arcadia bay-2 transformer has two groups of fans for cooling purpose and, thus, can operate in three different cooling modes. The fans are set to turn on and turn off in accordance with the simulated hot-spot temperature (SHST). The first group of fans is designed to turn on when the SHST exceeds 65 °C, and turn off when the SHST drops below 59 °C. Likewise, the second group of fans is set to turn on when the SHST exceeds 75 °C, and turn off when the SHST drops below 69 °C.

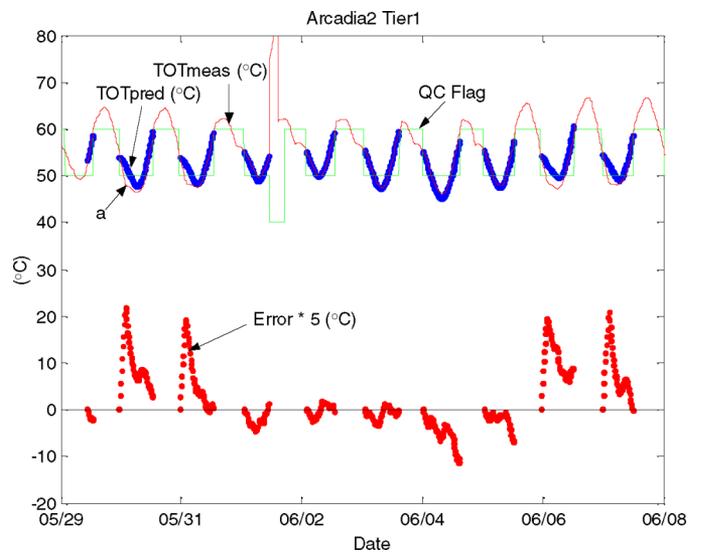


Fig. 10. Arcadia bay-2 TOT and TOT prediction error plot. (Color version available online at <http://ieeexplore.ieee.org>.)

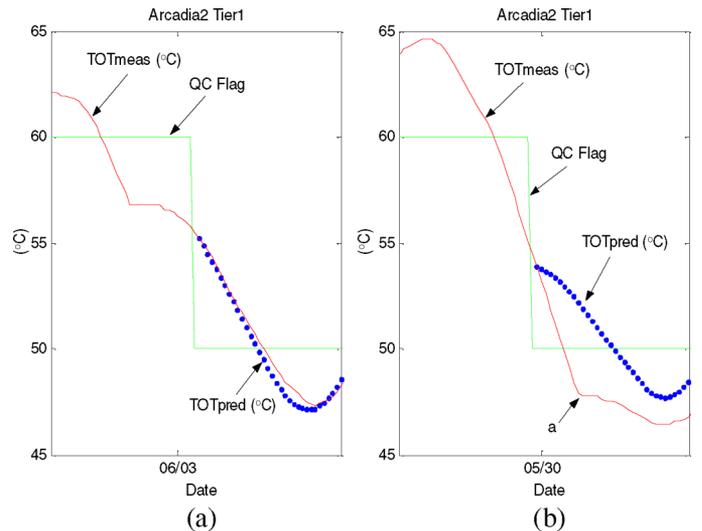


Fig. 11. Expanded view of TOT slope discontinuity. (Color version available online at <http://ieeexplore.ieee.org>.)

Given this turn on/off information, we divided the data into three tiers as described earlier, and applied our algorithm to the tier 1 data. The plots of the measured TOT, predicted TOT, and the prediction error are shown in Fig. 10. Also illustrated in this figure is the QC Flag, which indicates the boundaries of the tiers.

After carefully inspecting the curve of the measured TOT, we noticed that there are many points where the slope of the curve is apparently discontinuous. Normally, we expect a slope discontinuity at the fan turnoff point during a decreasing TOT as shown near the 6/03 day boundary in Fig. 10, and as amplified in Fig. 11(a). This slope discontinuity occurs when fans turn off because the amount of heat expelled per unit time by the transformer decreases, causing the cool-down to slow.

Sometimes we have observed the slope discontinuity (and, therefore, fan turn-off point) to be much lower than 59 °C, as found in the data of Fig. 10 near the 5/30 boundary (marked as point “a”) and as amplified in Fig. 11(b). This error in

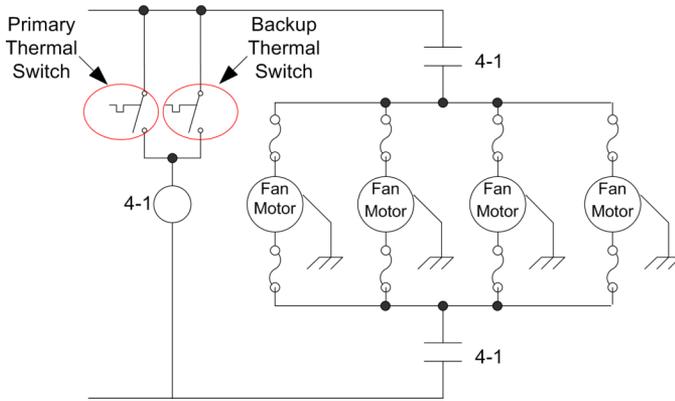


Fig. 12. Schematic of control circuitry for fan motor control. (Color version available online at <http://ieeexplore.ieee.org>.)

the cooling mode has two consequences. First, since we are assuming 59 °C SHST turn-off point in creating our data sets, this error causes us to include data gathering under Tier 2 cooling conditions in our Tier 1 data training-data set, skewing our model results. Second, when performing a simulation using our model to quantify the performance error, we see unusually large errors (Fig. 10) at the points where we are simulating Tier 2 performance with a model derived mostly from Tier 1 data.

We have observed similar (but not identical) unreliability in cooling mode switching in most other transformers whose data we have looked at closely and we suspect that the causes for the unreliable switching may be legion; however, for the situation presented in Fig. 10, we were able to determine the cause.

If you look carefully at Fig. 10, you will notice that the fan turn-off temperatures are unusually low only when the peak value of TOT and, consequently, SHST for a given 24-h cycle are very high, while if the peak value of TOT (and SHST) are rather low, the fan turn-off point is higher, closer to 59 °C. Fig. 12 is the electrical schematic that shows the first-stage cooling fan controls for the Arcadia bay 2 transformer. In this control scheme, the primary thermal switch (solid state) is set to close at 65 °C, and open at 59 °C. The secondary switch (electromechanical and less accurate/reliable) is redundant, and acts as a backup if the primary thermal switch fails to operate. As a backup switch, it picks up at a higher temperature and drops out at about 48 °C SHST. Therefore, the following is found:

- if the TOT/SHST peak is not large, only the primary switch picks up at 65 °C and drops out at approximately 59 °C;
- however, if the TOT/SHST peak is sufficiently high, then the secondary thermal switch is activated (along with the primary) and drops out at its design temperature, 48 °C.

The result of this control scheme is that the fan turn-off temperatures appears chaotic. We can develop an algorithm for classifying data according to the cooling mode to accommodate this control scheme; however, because of the uncertainty in the turn-on/turn-off performance, even this is problematic. We are aware of methods that may help to assign data to the proper cooling mode when the switching is unreliable; however, we currently do not have a working solution to this problem.

TABLE I  
EFFECT OF THE DATA-QUALITY CONTROL

	$CV(K_1)$	$CV(K_2)$	$CV(K_3)$	MAE	10% Error
Before DQC	6.61%	7.15%	7.09%	0.4202	0.8602
After DQC	5.51%	6.13%	5.20%	0.4140	0.8494
Improvement	16.72%	14.35%	26.75%	1.49%	1.26%

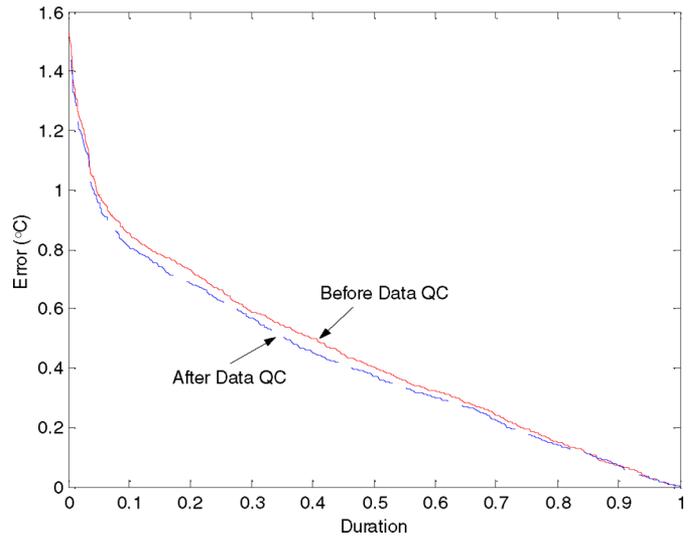


Fig. 13. Applying models derived before and after data QC to independent test data. (Color version available online at <http://ieeexplore.ieee.org>.)

### B. Effect of the Data-Quality Control

Selecting appropriate measures for assessing the reliability of the transformer thermal models is a larger issue than can be addressed here. We have chosen three metrics for assessment.

- Coefficient of variation (CV), defined as the standard deviation divided by the mean, provides a measure of the repeatability of the model parameters and, hence, a measure of the reliability of the model. We apply this measure to each of the model coefficients (i.e.,  $K_1$ ,  $K_2$ , and  $K_3$ ) obtained by building models from many independent training data sets.
- Mean absolute error (MAE) is the average of the absolute value of the TOT prediction errors when compared to measured TOT values.
- 10% error is the value (from the error duration curve) that the model error exceeds 10% of the time.

To measure the effect of the data QC scheme on reliability, we applied these three reliability measures to the models we derived before and after applying the QC scheme. Table I gives the raw data from this comparison. Depending on the coefficient you choose, the model reliability has increased by 15%–30%, while the prediction error, as measured by MAE and 10% error, has decreased by 1.5%; that is, we have increased the model reliability while improving model accuracy.

To measure the effect of data QC scheme on error performance, we also constructed Fig. 13, which shows the error duration curves for two models. The “Before Data QC” model was built using all data (including bad data). The “After Data QC” model was built after data QC eliminated the bad data. Fig. 13

shows that on the average, the “After Data QC” model has lower prediction errors.

#### IV. DATA-SET SCREENING TECHNIQUES

While data QC can eliminate some erroneous data, a data set screening technique can detect anomalous models and, therefore, by inference, anomalous data sets. In building a model, we normally only use part of the data for training (known as the training data set) and reserve the rest of the data to perform independent tests on the model’s performance. These latter data sets are known as testing data sets. A valid model is expected to perform well (we have chosen TOT prediction error as our gauge of performance) on both types of data sets: the training data set and testing data sets. If a model does not perform well on either or both of these classes of data sets, we suspect that the data of which the model was built from must be anomalous. By removing the anomalous data sets from our model building procedure, we improve the reliability of the model.

Our data screening technique involves the following steps.

- Divide the available data into 10-day data sets. (One could use smaller data sets, but using less than three-day data sets is unadvisable.)
- Build a model for each data set.
- Measure the performance of each models on its training data set. (This error is known as training error.)
- Measure the performance of each models on all of the other testing data sets. (This error is known as testing error.)
- Reject those data sets that produce anomalous models (i.e., models that do not perform “well” on these data sets).
- The accepted data sets are then used to create a thermal model of the transformer and assess the model’s reliability.

Since the way we define “good” performance is very different for training and testing errors, we describe each procedure separately.

##### A. Training-Error-Based Screening—Gaussian Distribution Based

To eliminate erroneous data based on training error, we build a model for each 10-day data set from available data and then use those models to calculate the training error, that is, the TOT prediction error on the respective data sets used to build each model. An error duration curve is constructed for each model’s training error and the error value on the error duration curve at the 10% point is used to measure the accuracy of the model. The 10% training-error-duration data points are shown in Table II for each of the 14 data sets for the Corbell bay-3 transformer.

We assume the training errors are Gaussian distributed with a zero mean, and reject a model if its training error exceeds a prescribed threshold value. The threshold value is based on

TABLE II  
10% ERROR DURATION OF TRAINING DATA SETS

Data Set/ Model Number	10% Training- Error Duration (°C)
1	- 0.8610
2	-0.7293
3	+0.7326
4	-0.7690
5	+0.8403
6	+0.7530
7	+0.9755
8	-0.6891
9	-0.8269
10	-0.8311
11	+0.7063
12	+1.1266
13	-1.1276
14	+0.9683
STD	0.89

the standard deviation (STD) of the 10% training-error-duration values in the following way:

Based on past experience, we have an idea of what the maximum STD should be, say  $STD^*$  (1.0°C), for good source data. We then compare the calculated STD, call it STD, of the 10% error-duration measures and compare that with  $STD^*$ .

- If  $STD \leq STD^*$ , then we have good quality data. This is the case for the data shown in Table II. We eliminate only the data sets in Table II that generated models whose 10% training-error-duration values are greater than  $STD^*$ . The eliminated models, shaded in black in Table II, are 12 and 13. For good quality data, we typically discard very few data sets. We use the retained data sets to build a thermal model and to assess reliability of the model.
- If  $STD > STD^*$ , that means the data quality is poorer. If there are much data, we can apply this same criteria for data-set rejection as in the bullet above; however, in many practical cases, available data are limited; consequently, in order to have enough data to assess the reliability of the model, we eliminate those data sets with a 10% training-error-duration value larger than the calculated STD.

Note that if this method of training, aka, diagonal training, is used alone to estimate the performance error, it will lead to an error estimate that is unrealistically low. In the approach we propose here, we do not use the absolute value of the diagonal/training error to qualify data sets, rather we use the relative value of the training error to disqualify data sets. Using this approach, it is possible to discard good models (those that may pass the test in the next section) but it is unlikely that we will retain bad models. Using the remaining qualified data sets, testing-error based screening (which does not suffer the bias of training-error-based screening) is then used as a second test to detect anomalous data sets.

##### B. Testing-Error-Based Screening— $\chi^2$ Distribution Based

The second step of our two-stage screening procedure is to eliminate models that do not perform well when using testing error performance as a measure.

TABLE III  
10% TESTING-ERROR DURATION VALUES

K1	K2	K3		Set 1	Set 2	Set 3	.....	Set 7	.....	Set 12	Set 13	Set 14	S <sup>2</sup>
2.1882	0.0827	0.6978	Set 1	-0.8160	-0.8196	-0.9962	.....	1.4033	.....	-1.1311	-1.5867	-1.4963	<b>0.95</b>
2.1991	0.0857	0.7651	Set 2	-0.8840	-0.7293	1.0238	.....	1.3704	.....	1.1069	-1.2802	-1.1592	<b>0.85</b>
2.3358	0.0891	0.7422	Set 3	-0.9932	-0.9490	0.7326	.....	1.1420	.....	-1.3014	-1.7882	-1.6411	<b>1.04</b>
2.2187	0.0857	0.7568	Set 4	0.8942	0.7299	-1.0102	.....	1.3624	.....	1.0775	-1.3208	-1.2145	<b>0.85</b>
2.2567	0.0873	0.7506	Set 5	0.8700	-0.7929	0.8320	.....	-1.2844	.....	-1.1645	-1.5468	-1.4343	<b>0.86</b>
2.3845	0.0895	0.7184	Set 6	-1.1584	-1.1153	0.8100	.....	1.0740	.....	-1.4713	-2.0198	-1.8870	<b>1.33</b>
2.5317	0.0945	0.7492	Set 7	-1.4865	-1.4502	-0.9021	.....	0.9755	.....	-1.7912	-2.2467	-2.0701	<b>2.00</b>
2.3300	0.0917	0.8230	Set 8	1.0517	-0.8216	0.8940	.....	1.1591	.....	-1.2818	-1.4124	-1.1847	<b>0.92</b>
2.2225	0.0838	0.7280	Set 9	-0.9682	0.8000	1.1843	.....	1.5409	.....	1.1139	-1.3221	-1.2508	<b>1.07</b>
2.1316	0.0837	0.7573	Set 10	0.8684	-0.7432	-1.1150	.....	1.4802	.....	1.1994	1.2481	-1.1250	<b>0.95</b>
2.2496	0.0838	0.7187	Set 11	1.0285	-0.8168	1.2122	.....	1.6137	.....	1.1321	-1.3496	-1.3062	<b>1.19</b>
2.1923	0.0817	0.6955	Set 12	0.9459	-0.8049	1.2133	.....	1.6077	.....	1.1266	-1.4016	-1.3481	<b>1.15</b>
2.0055	0.0743	0.6686	Set 13	1.7852	1.6304	2.0080	.....	2.4458	.....	1.6465	-1.1276	1.0957	<b>3.53</b>
2.1994	0.0786	0.7057	Set 14	2.0613	1.9938	2.3493	.....	2.8471	.....	1.9890	-1.0466	0.9683	<b>5.11</b>

In this step, we use each model created in step one and measure their testing-error performance (using the 10% error-duration value as a measure). For a data population with 14 data sets, we derive 13 testing error-duration measures as shown in Table III for each model derived from one 10-day data set. (The data sets already eliminated in step one of our procedure are shaded in black in Table III. The diagonal values that are darkly shaded in this table are training error values from Table II.)

We assume the prediction-error measures of each row of Table III (including, for simplicity, the diagonal training measure) are a sample taken from an independent Gaussian-distributed whole population. (Note that for a small number of data sets, including the training-error measure can lead to a bias in the error estimate; hence, the training-error measure should be discarded in this instance.) We assign the following variables for the subsequent discussion:

- $S^2$  is the variance of the sample/row, (shown in the last column of Table III);
- $\sigma^2$  is the variance of the corresponding whole Gaussian-distributed population (unknown);
- $n$  is the size of the sample (14).

The sample variance  $S^2$  is an approximation of the population variance  $\sigma^2$  but for different samples from one population, we may get a wide range of sample variances  $S^2$ . Since  $S^2$  is a random variable, it is easy to image that we could get a low  $S^2$  (with a lucky sample) from a population of poor quality data and a large  $S^2$  value (from an unlucky sample) from a population of high-quality data. By only knowing  $S^2$ , what we wish to do is determine whether the sample/row comes from a good population or flawed population; that is, by knowing  $S^2$  in Table III, determining whether our model is good or bad and, hence, whether the data that generated it are good or bad. If the population (model) is bad, than we reject the data set that created the model; otherwise, we keep the data set.

The problem is that we cannot know with certainty for any given  $S^2$  whether it comes from a good or bad population. Fortunately, we can apply the  $\chi^2$  hypothesis test for variance (or standard deviation) to these samples (rows) to determine the probability of accepting a bad model (population) as good in the following way.

With the assumptions we are using, it can be shown that the calculated value of  $(n-1)S^2/\sigma^2$  for each sample (row) of Table III is a  $\chi^2$  distributed random variable with  $(n-1)$  degrees of freedom, or  $\chi^2(n-1)$  [11]. In the  $\chi^2$  hypothesis test for variance, if we want to know the probability of accepting a model (population) whose  $\sigma$  is indicative of poor quality data, that is, greater than some acceptable value, say  $\gamma\sigma^*$ , and we need to do this solely based on the variance of one sample whose variance is  $S^2$ , then the value of the integral of the  $\chi^2(n-1)$  density curve, from 0 to the point  $(n-1)S^2/\gamma^2\sigma^{*2}$  gives us an upper-bound on that probability. We can write this probability as

$$P \left\{ \frac{(n-1)S^2}{\sigma^2} < \frac{(n-1)S^2}{\gamma^2\sigma^{*2}} \right\}.$$

From past experience with good models/populations for transformer thermal modeling, we have gained an idea of what the  $\sigma$  value of the testing errors are: say  $\sigma^* = 1.5$ . And, again based on past experience, we think that the models can be considered as bad with much confidence, if their corresponding  $\sigma$  is larger than  $\sigma^*$  by a specified factor of  $\gamma = 1.25$  (i.e.,  $\sigma = \gamma\sigma^*$ ). Assume also that we wish to limit the probability of accepting a bad model (population) as good to some value  $\beta$ . If we take a sample with variance  $S^2$  from this bad population with  $\sigma = \gamma\sigma^*$ , then, according to the discussion above,  $(n-1)S^2/\sigma^2 = (n-1)S^2/\gamma^2\sigma^{*2}$  is  $\chi^2(n-1)$ . Therefore, there exists some value of  $\chi^2(n-1)$ , let us call it  $\chi_{\beta}^2(n-1)$ , such that

$$P \left\{ \frac{(n-1)S^2}{\gamma^2\sigma^{*2}} < \chi_{\beta}^2(n-1) \right\} = \beta \quad (9)$$

where  $P$  stands for probability (see Fig. 14 for a graphical interpretation of  $\chi_{\beta}^2(n-1)$ ). This means that if we use  $(n-1)S^2/\gamma^2\sigma^{*2} < \chi_{\beta}^2(n-1)$  as the testing condition for accepting a model as good, the probability of accepting a bad model with  $\sigma = \gamma\sigma^*$  is  $\beta$ . For a bad model with  $\sigma > \gamma\sigma^*$ , we have

$$\beta = P \left\{ \frac{(n-1)S^2}{\sigma^2} < \chi_{\beta}^2(n-1) \right\} > P \left\{ \frac{(n-1)S^2}{\gamma^2\sigma^{*2}} < \chi_{\beta}^2(n-1) \right\} \quad (10)$$

indicating that  $\beta$  is the upper limit of the probability of accepting a bad model as good, which is what we desire.

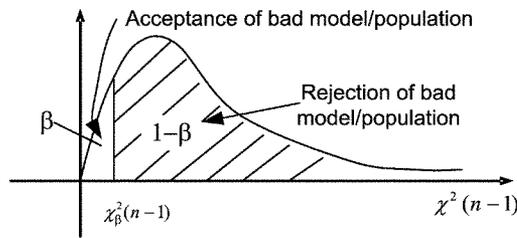


Fig. 14.  $\chi^2$  distribution of the entire population.

TABLE IV  
EFFECT OF THE DATA-SET SCREENING

	CV( $K_1$ )	CV( $K_2$ )	CV( $K_3$ )	MAE	10% Error
Before screening	5.51%	6.13%	5.20%	0.4140	0.8494
After screening	3.44%	3.47%	4.63%	0.3929	0.7694
Improvement	37.50%	43.30%	10.89%	5.09%	9.42%

TABLE V  
EFFECTS OF THE DATA QC AND DATA-SET SCREENING

	CV( $K_1$ )	CV( $K_2$ )	CV( $K_3$ )	MAE	10% Error
Before	6.61%	7.15%	7.09%	0.4202	0.8602
After	3.44%	3.47%	4.63%	0.3929	0.7694
Improvement	47.95%	51.43%	34.73%	6.51%	10.56%

Therefore, we use  $(n-1)S^2/\gamma^2\sigma^{*2} < \chi_{\beta}^2(n-1)$  (i.e.,  $S^2 < (\chi_{\beta}^2(n-1) \cdot \gamma^2\sigma^{*2})/(n-1)$ ), as the testing condition for judging a bad model as good, based on its testing errors.

### C. Results of Data Screening

In Table III with data sets 7 and 14, the lightly shaded rows are screened out based on testing errors. These models perform well on their training data sets as seen in this table, but they perform poorly on the other 13 testing data sets, as evidenced by their variance values—which is what the data-screening procedure measures. Table IV gives a comparison between the coefficients of variation of the coefficients as well as the error measures for TOT prediction, obtained before and after applying data-set screening. The CVs of linear model coefficients  $K_x$  improve by 11%–38%, and the error measures have improved by 5%–9%.

## V. CONCLUSION

By applying both data-quality control and data-set screening procedures, we have been able to improve the reliability of the transformer thermal model coefficients by 35% to 50% while decreasing the prediction error 7%–11% as shown in Table V; that is, we have increased the model reliability substantially while improving the model accuracy. While we have applied these techniques to a linear model, we expect similar gains when applying these techniques to other nonlinear models.

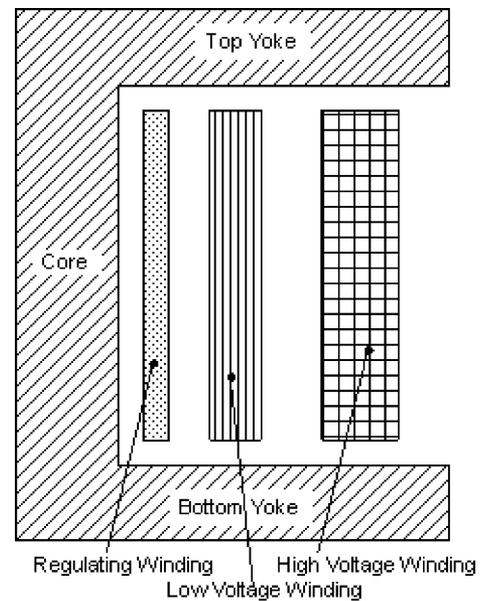


Fig. 15. Cross-section diagram of the core form transformer with winding arrangement.

TABLE VI  
CORBELL TRANSFORMER HEAT-RUN DATA

Quantity (Units)	Value
No Load Loss (W)	1,400
Load Loss (W)	200,000
Top Oil Rise (°C)	50.9
Hot Spot Rise (°C)	30.6
Core/Coil Weight (lbs)	60,000
Tank Weight (lbs)	25,000
Oil Volume (gal)	9,000

## APPENDIX TRANSFORMER DESCRIPTION

The Corbel transformer used in this work is a conventional three-phase, three-leg core-form (Fig. 15), built for step-down operation. The transformer is of “sealed tank” design. There is a gas space above the oil inside the transformer. The high-voltage winding contains a de-energized full capacity tap changer. The reduced capacity load tap changer (LTC) is electrically located in the low-voltage winding and is rated at 1500 A; hence, no series transformer is necessary.

Each phase has three windings. From inner to outer they are as follows:

- regulating or tap winding;
- low-voltage helical winding;
- high-voltage disk-type winding.

The heat-run data for this transformer are contained in Table VI and the basic data for this transformer are contained in Table VII.

The temperature sensor consists of a 100- $\Omega$  platinum resistance temperature detector (RTD) installed in a dry well in segment number one near the top of the transformer. The RTD is connected to a commercially available transformer temperature monitor, the Barrington TTM, which feeds the signal to the SCADA system. The top-oil temperature measurements are recorded by the SCADA computer every 15 min. The top-oil

TABLE VII  
BASIC TRANSFORMER DATA

Capacity	15/20/25 MVA @ 45°C 28 MVA @ 55 °C
Frequency	60 Hz
H.V winding	67 kV delta
L.V winding	12,470 Y/ 7200 volt
BIL, H.V winding	350 kV
BIL, L.V winding	110 kV
Impedance	7.5 % @ 15 MVA
LTC	+ / - 10 %, 16 steps 5/8 % each
DETC tap voltages	70.6, 68.8, 67.0, 65.2, and 63.4 kV

probe is located just below the gas blanket and above the lowest oil level.

#### REFERENCES

- [1] *IEEE Guide for Loading Mineral-Oil-Immersed Power Transformers*, IEEE C57.91-1995, 1995.
- [2] D. J. Tylavsky, Q. He, G. A. McCulla, and J. R. Hunt, "Sources of error in substation distribution-transformer dynamic thermal modeling," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 178–185, Jan. 2000.
- [3] B. C. Lesieutre, W. H. Hagman, and J. L. Kirtley Jr., "An improved transformer top oil temperature model for use in an on-line monitoring and diagnostic system," *IEEE Trans. Power Del.*, vol. 12, no. 1, pp. 249–256, Jan. 1997.
- [4] G. Swift, T. S. Molinski, and W. Lehn, "A fundamental approach to transformer thermal modeling—Part I: Theory and equivalent circuit," *IEEE Trans. Power Del.*, vol. 16, no. 2, pp. 171–175, Apr. 2001.
- [5] G. Swift, T. S. Molinski, R. Bray, and R. Menzies, "A fundamental approach to transformer thermal modeling—Part II: Field verification," *IEEE Trans. Power Del.*, vol. 16, no. 2, pp. 176–180, Apr. 2001.
- [6] D. Susa, M. Lehtonen, and H. Nordman, "Dynamic thermal modeling of power transformers," *IEEE Trans. Power Del.*, vol. 20, no. 1, pp. 197–204, Jan. 2005.
- [7] W. H. Tang, Q. H. Wu, and Z. J. Richardson, "Equivalent heat circuit bases power transformer thermal model," *Proc. Inst. Elect. Eng., Elect. Power Appl.*, vol. 149, no. 3, pp. 87–92, Mar. 2002.
- [8] —, "A simplified transformer thermal model based on thermal-electric analogy," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1112–1119, Jul. 2004.
- [9] A. P. S. Meliopoulos, "State estimation methods applied to transformer monitoring," in *Proc. Power Engineering Soc. Summer Meeting*, vol. 1, Jul. 2001, 15–19, pp. 419–423.
- [10] M. F. Lachman, P. J. Griffin, W. Walter, and A. Wilson, "Real-time dynamic loading and thermal diagnostic of power transformers," *IEEE Trans. Power Del.*, vol. 18, no. 1, pp. 142–148, Jan. 2003.
- [11] D. M. Levine, P. P. Ramsey, and R. K. Smidt, *Proc. Applied Statistics for Engineers and Scientists*. Upper Saddle River, NJ: Prentice-Hall, 2001.

**Daniel J. Tylavsky** (SM'88) received the B.S., M.S.E.E., and Ph.D. degrees in engineering science from the Pennsylvania State University, University Park, in 1974, 1978, and 1982, respectively.

From 1974 to 1976, he was with Basic Technology, Inc., Pittsburgh, PA, and from 1978 to 1980, he was an Instructor of electrical engineering at Pennsylvania State. In 1982, he joined the Faculty in the Electrical Engineering Department of Arizona State University, Tempe, where he is currently Associate Professor of engineering.

Dr. Tylavsky is a member of the IEEE Power Engineering Society and Industry Applications Society, and is an RCA Fellow, NASA Fellow, and member of Phi Eta Sigma, Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi.

**Xiaolin Mao** (S'04) was born in Jiangxi, China. He received the B.S. and M.S. degrees in electrical engineering from Tsinghua University, Beijing, China, in 1997 and 2000, respectively, and is currently pursuing the Ph.D. degree in electric power engineering with Arizona State University, Tempe.

**Gary A. McCulla** (M'86) received the B.S.E. degree from Arizona State University, Tempe, in 1979.

Currently, he is a Senior Principal Engineer with the Salt River Project Apparatus Engineering Department, Tempe. He has been involved in preparing specifications, planning, and maintenance of large and medium power transformers for more than 20 years. Recently, he was an Industry Advisor with EPRI for the development of PT loads.

Mr. McCulla is a member of the local chapter of the IEEE Power Engineering Society.