Acceptability of Four Transformer Top-Oil Thermal Models—Part I: Defining Metrics

Lida Jauregui-Rivera, *Member, IEEE*, and Daniel J. Tylavsky, *Senior Member, IEEE*

***Abstract*—Eventually, the prediction of transformer thermal performance for dynamic loading will be made using models distilled from measured data, rather than models derived from transformer heat-run tests. Which model(s) will be used for this purpose remains unclear. In this paper, we introduce metrics for measuring the acceptability of transformer thermal models. For a model to be acceptable, it must have the qualities of adequacy, accuracy, and consistency. We assess model adequacy using the metrics: prediction R^2 and plot of residuals against fitted values. To assess model consistency, we use the variance inflation factor (which measures multicollinearity), condition number, eigenstructure, parameter sensitivity, and the standard deviation of model parameters and predicted maximum steady-state load ($\mathbf{SSL}_{\max}$). To assess model accuracy, we use the comparison of model parameters with nominal values and error duration curves. Other metrics of interest are also introduced. In a companion paper, these metrics are applied to the four models defined in this paper and a relative ranking of the acceptability of these models is presented.**

***Index Terms*—ANSI C5791, top-oil temperature, transformer, transformer thermal modeling.**

I. Introduction

ALL transformers of distribution-class capacity or higher will eventually be dynamically loaded using thermal models derived from field data. These models are not only more accurate, but also support derivation of the statistics of the model's accuracy [4]. These statistics can be used in probabilistic transformer loading. The traditional ANSI/IEEE (Clause 7) top-oil rise model [1] is known to perform poorly compared to top-oil models trained with measured data [2], [3], [5]. (For a more detailed discussion of the limitations of the Clause 7 model, see [5].) Models trained using measured data are superior because they naturally account for many undetected phenomena in operating transformers, such as fouled heat exchangers, inoperative pumps/fans, etc., that the nominal Clause 7 model does not. The inability to account for undetected faults and to detect erroneous test report parameters is likewise a limitation of the IEEE/ANSI Appendix G model [1]. Using the Appendix G model also requires parameters that most utilities do not have.

There is a need in this field of research to define measures of acceptability of "model and method," that can be used by researchers and model users alike. When used by researchers, these measures should allow evaluation of the model's structure. For model users, these measures should give assessments of model consistency and accuracy.

One goal of this paper is to identify models—from the many available [1]–[13]—that require only parameters available from heat-run data and need only measurements utilities routinely monitor. We express these models using a common notation that allows their similarities and difference to be assessed easily. Our second goal is to define metrics to quantitatively measure adequacy, consistency, and accuracy, and then, in a companion paper, apply these metrics to rank models according to their acceptability.

Of the potential first and higher order, linear, and nonlinear models [1]–[13], we selected four top-oil models.

- The linearized Clause 7 model [2], [4].
- The (nonlinear) Clause 7 top-oil-rise model [1] corrected for ambient temperature variation.
- The (nonlinear) model of G. Swift *et al.* [6], [7].
- The (nonlinear) model of D. Susa *et al.* [8].

II. Model Description

A. ANSI/IEEE Clause 7 Top-Oil Rise Model Corrected for Ambient Temperature also Known as Nonlinear Top-Oil Model

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

$$T_o \frac{d\theta_o}{dt} = -\theta_o + \theta_u \tag{1}$$

where

$$\theta_u = \theta_{fl}\left(\frac{I^2 * R + 1}{R + 1}\right)^n \tag{2}$$

$$T_o = \frac{C\theta_{fl}}{P_{fl}} \tag{3}$$

and

θ_o top-oil rise over ambient temperature (°C);
θ_u ultimate top-oil rise for load L(°C);
θ_i initial top-oil rise for $t = 0$ (°C);
θ_{fl} top-oil rise over ambient temperature at rated load (°C);

T_o time constant (h);
C thermal capacity ($MWh/$°C);
P'_{fl} total loss at rated load (MW);
n oil exponent—(an empirically derived coefficient selected for each cooling mode to approximately account for change in resistance with load);
I ratio of load L to rated load;
R ratio of load loss to no-load loss at rated load.

Manuscript received July 26, 2006. This work was supported in part by the Salt River Project, in part by Arizona Public Service, and in part by PSERC. Paper no. TPWRD-00416-2006.

The authors are with Arizona State University, Tempe, AZ 85287-5706 USA (e-mail: Lida.Jauregui-Rivera@aps.com; tylavsky@asu.edu).

Digital Object Identifier 10.1109/TPWRD.2007.905555

To correct this for ambient temperature variation, recognize that the time rate of change in top-oil temperature is driven by the difference between existing top-oil temperature and ultimate top-oil temperature $(\theta_u + \theta_{\rm amb})$

$$T_o \frac{d\theta_{\rm top}}{dt} = -\theta_{\rm top} + \theta_{\rm amb} + \theta_u = f(\theta_{\rm top}, \theta_{\rm amb}, \theta_u) \quad (4)$$

where

$\theta_{\rm amb}$ ambient air temperature (°C).

Discretizing this model using the backward Euler rule because of its stability properties

$$\frac{d\theta_{\rm top}(k\Delta t)}{dt} = \frac{\theta_{\rm top}(k\Delta t) - \theta_{\rm top}((k-1)\Delta t)}{\Delta t} = f\left(\theta_{\rm top}(k\Delta t), \theta_{\rm amb}(k\Delta t), \theta_u(k\Delta t)\right) \quad (5)$$

and then suppressing the Δt in the argument of $\theta_{\rm top}$, substituting (5) and (2) into (4), and rearranging gives

$$\theta_{\rm top}(k) = \frac{T_0}{T_0 + \Delta t}\theta_{\rm top}(k-1) + \frac{\Delta t}{T_0 + \Delta t}\left[\left[\frac{I^2(k)R + 1}{R+1}\right]^n \theta_{fl} + \theta_{\rm amb}(k)\right] \quad (6)$$

where

$$n = \left\{ \begin{array}{c} 0.8 \text{ NONA} \\ 0.9 \text{ NOFAor FOFA Nondirected} \\ 1.0 \text{ FOFA Directed} \end{array} \right\}. \quad (7)$$

B. Linearized Top-Oil Model [3], [2]

Assuming $n \approx 1$, the above model is simplified to

$$\begin{aligned}\theta_{\rm top}(k) &= \frac{T_0}{T_0 + \Delta t}\theta_{\rm top}(k-1) + \frac{\Delta t}{T_0 + \Delta t}\left[\left[\frac{I^2(k)R + 1}{R+1}\right]\theta_{fl} + \theta_{\rm amb}(k)\right] \\ &= \frac{\Delta t R \theta_{fl}}{(T_0 + \Delta t)(R+1)} I^2(k) + \frac{T_0}{T_0 + \Delta t}\theta_{\rm top}(k-1) + \frac{\Delta t}{T_0 + \Delta t}\theta_{\rm amb}(k) + \frac{\Delta t \theta_{fl}}{(T_0 + \Delta t)(R+1)}. \end{aligned} \quad (8)$$

Since this is a linear model in terms of I^2, $\theta_{\rm top}$, and $\theta_{\rm amb}$, we rewrite this model substituting K's for the constant coefficients

$$\theta_{\rm top}(k) = K_1 I^2(k) + K_2 \theta_{\rm amb}(k) + (1 - K_2)\theta_{\rm top}(k-1) + K_3. \quad (9)$$

(Note the backward Euler method used in [2]–[5] is mislabeled as the forward Euler method.)

C. Swift Model (Modified-Convection Top-Oil Model [6], [7])

Swift *et al.* derive a model from first principals that changes placement of the ad-hoc oil exponent n in the Clause 7 model

$$T_o \frac{d\theta_{\rm top}}{dt} = -\left(\theta_{\rm top} - \theta_{\rm amb}\right)^{1/n} + \tilde{\theta}_u \quad (10)$$

where the revised value

$$\tilde{\theta}_u = \theta_{fl}^{1/n}\left(\frac{I^2 * R + 1}{R+1}\right) \quad (11)$$

is no longer an ultimate top-oil rise, but a modified value needed for equality in (10) at the rated load. Swift *et al.* do not give any values for n since they obtain n through training of their model; however, they do give typical values of n

$$n = \left\{ \begin{array}{c} 0.8 \text{ NONA (typical)} \\ 0.8 \leftrightarrow 1.0 \text{ FA (typical)} \end{array} \right\}. \quad (12)$$

In [6] and [7], it is not clear which discretization technique the authors use. We presume they use the implicit backward Euler method. (From our experimentation, using the explicit forward Euler method with their model yields similar results, but results that differ by 10%-15%.) After discretization, (10) and (11) become

$$\theta_{\rm top}(k) = \theta_{\rm top}(k-1) + \frac{\Delta t}{T_0} \times \left[\left[\frac{I^2(k)R + 1}{R+1}\right]\theta_{fl}^{1/n} - \left[\theta_{\rm top}(k) - \theta_{\rm amb}(k)\right]^{1/n}\right]. \quad (13)$$

Because of the form of the nonlinearity in (10), $\theta_{\rm top}(k)$ appears implicitly on both sides of the equation, which makes training the model more difficult.

D. Susa et al. [8] Model (Modified-Convection Top-Oil Model With Nonlinear Permeability)

The authors of this model start with first principles to derive a top-oil model that shows a similar exponential behavior to the Swift *et al.* [6], [7] model; however, in this model, Susa *et al.* [8] retain oil viscosity as a parameter. Their governing differential equation becomes

$$T_o \frac{d\theta_{\rm top}}{dt} = -\frac{\left(\theta_{\rm top} - \theta_{\rm amb}\right)^{1/n}}{(\mu\theta_{fl})^{\left[\frac{1-n}{n}\right]}} + \hat{\theta}_u \quad (14)$$

where

$$\hat{\theta}_u = \theta_{fl}\left(\frac{I^2 * R + 1}{R+1}\right) \quad (15)$$

$$\mu = \frac{\mu_{actual}}{\mu_{\rm rated}} = \frac{1.3573 \bullet 10^{-6} \bullet e^{[2797.3/\theta_{\rm top}+273]}}{1.3573 \bullet 10^{-6} \bullet e^{[2797.3/\theta_{fl}+273]}} = e^{[(2797.3/\theta_{\rm top}+273) - (2797.3/\theta_{fl}+273)]} \quad (16)$$

The values of n for different cooling modes are shown in (17) at the bottom of the next page.

After discretizing with backward Euler, (14) becomes

$$\theta_{\rm top}(k) = \theta_{\rm top}(k-1) + \frac{\Delta t}{T_0}\left[\left[\frac{I^2(k)R + 1}{R+1}\right]\theta_{fl} - \frac{1}{[\mu(k)\theta_{fl}]^{[(1-n)/n]}}\left[\theta_{\rm top}(k) - \theta_{\rm amb}(k)\right]^{1/n}\right]. \quad (18)$$

As with the Swift *et al.* [6], [7] model, (18) must be solved implicitly to perform simulations.

III. Simple Model Comparison

One obvious question that arises is: Are these models really very different? First observe that when $n = 1$, all of the models

TABLE I
MODEL COMPARISONS

Linear Clause 7 n=(not used) (°C/h)	Nonlinear Clause 7 n=0.9 (°C)	Swift et al. n=1.0 (°C)	Susa et al. n=0.8 (°C)
16.38	16.75	16.38	18.01

are identical; however, in general, there is not one cooling condition for which all are the same. For NOFA conditions, we compare the time rate of change ($^\circ$C/hr) in $\theta_{\rm top}$, [in (6), (9), (13), and (18)] due to a step increase in load to $I = 0.9$ pu, assuming

$$\begin{aligned} R = 7; \quad T_0 = 2(h); \quad \theta_{fl} = 55(^\circ\mathrm{C}); \\ \Delta t = 0.25(h); \quad \theta_{\rm top}(k-1) = 30(^\circ\mathrm{C}); \\ \theta_{\rm amb}(k) = 70(^\circ F) = 21(^\circ\mathrm{C}); \\ \mu_{\rm rated} = 0.897 \bullet 10^{-3} \left(\frac{\rm kg}{\rm ms}\right). \end{aligned}$$

Except for the Susa *et al.* [8] model (Table I), these models appear very similar in their time rate of change performance. Since our goal is to match measured data within a 2–3 °C error bound, even the half-degree difference between the Swift *et al.* and the (nonlinear) Clause 7 model may accumulate through time and lead to errors that exceed our self-impose bound.

All, or possibly none, of these models may perform adequately when using parameters from the transformer test report. (We have found test report data that do not match measured transformer performance well.) However, all models will likely perform adequately when their parameters are selected to optimally fit measure data. Our goal is define a set of metrics for quantitatively assessing the acceptability of these four models.

IV. EQUIVALENCE OF METRICS FOR THE LINEAR AND NONLINEAR MODELS

Much work has been done to develop metrics for linear regression. We show in this section that these metrics are also valid for assessing nonlinear model performance.

A. Linear Modeling

To explain the metrics that will be applied for model acceptability testing, we first establish our notation. For the sake of simplicity in notation, we define y_m to be the measured value of $\theta_{\rm top}(m)$. A multiple regression model that describes the relationship between the process output Y and k regressor values is

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_k x_k + \varepsilon, \tag{19}$$

where the parameter β_j and $j = 1, 2, \ldots k$ are called the true regression coefficients, x_j are regression variables (in the linear model case I^2, $\theta_{\rm amb}$, and, as a lagged regressor $\theta_{\rm top}$). In vector form, the regression variables are written as $\boldsymbol{x}$. The error term ε is assumed to have zero mean and variance σ^2. The residuals are assumed to be uncorrelated and normally distributed.

In vector form, this equation can be written

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}. \tag{20}$$

where $\mathbf{Y}$, Y_j, $j = 1, 2, .l \ldots n$, is an $n \times 1$ vector of outputs, $\mathbf{X}$ is an $n \times k$ matrix of sampled regressor variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of the regression coefficients (for the linear model, K_1, K_2), $\boldsymbol{\beta}_0$ is an $n \times k$ vector of constant scalar values, β_0, (for the linear model, $\beta_0 = K_3$) and ε is an $n \times 1$ vector of random errors.

Since we do not have outputs, $\mathbf{Y}$, but instead measured values, $\mathbf{y}$, we can never know the true value of $\boldsymbol{\beta}$, but can estimate the values $\hat{\boldsymbol{\beta}}$ as

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X^T X}\right)^{-1} \mathbf{X^T y}. \tag{21}$$

Using these estimated and measured values, we may write our best approximation to (20) as

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}_\mathbf{0} + \mathbf{e} \tag{22}$$

where $\mathbf{e}$ is an $n \times 1$ vector of residuals, $\hat{\boldsymbol{\beta}}_\mathbf{0} = \overline{\mathbf{y}}$ is an $n \times 1$ vector of constant scalar values, and $\overline{\mathbf{y}}$ is the mean of the sampled values of $\mathbf{y}$. The predicted values $\hat{\mathbf{y}}$ are given by

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}_\mathbf{0} \tag{23}$$

where $\hat{\mathbf{y}}$ is an $n \times 1$.

B. Nonlinear Modeling

The nonlinear problem is

$$\mathbf{Y} = \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\beta}) \tag{24}$$

where the vector function $\mathbf{h}$ defines the nonlinear relationship between regressors $\boldsymbol{x}$ (which are functions of measured driving variable I^2, $\theta_{\rm amb}$, and, as a lagged regressor, $\theta_{\rm top}$) and parameter $\boldsymbol{\beta} = [T_0 \quad \theta_{fl} \quad n \quad R]$.

The optimal values of β, which minimize the square error in (24), (assuming all measurements have equal weights) obey

$$\mathbf{min}_{\forall \boldsymbol{\beta}} \left(\mathbf{Y} - \boldsymbol{h}(\boldsymbol{x}, \boldsymbol{\beta})\right)^2. \tag{25}$$

$$n = \left\{ \begin{array}{c} 0.8, \text{ Init. Oil Speed} > 0, \text{(All Cooling Modes, Xfmr under load)} \\ 0.67 \text{ Init. Oil Speed} = 0, \text{NOFA FOFA(Cold Start)} \\ 1.0 \text{ Init. Oil Speed} = 0, \text{NONA(Cold Start)} \end{array} \right\}. \tag{17}$$

The optimal values of $\boldsymbol{\beta}$ may be found by taking the derivative of (25) with respect to $\boldsymbol{\beta}$, and setting it to zero. Suppressing the $\boldsymbol{x}$ in (25) gives

$$\frac{\partial(\mathbf{Y}-h(\boldsymbol{\beta}))^2}{\partial\boldsymbol{\beta}} = 2\boldsymbol{X}^T(\boldsymbol{\beta})(Y-h(\boldsymbol{\beta})) = 0$$
$$\therefore Y - h(\boldsymbol{\beta}) = 0 \tag{26}$$

where

$$\mathbf{X}(\boldsymbol{\beta}) = \frac{\partial h(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}.$$

To solve (26), we expand it into a linear form using a Taylor series. This defines the iterative scheme for finding $\boldsymbol{\beta}$ given by

$$\hat{\boldsymbol{\beta}}^{\mathbf{k+1}} = \left(\mathbf{X}^{\mathbf{T}}\left(\hat{\boldsymbol{\beta}}^{\mathbf{k}}\right)\mathbf{X}\left(\hat{\boldsymbol{\beta}}^{\mathbf{k}}\right)\right)^{-1}\mathbf{X}^{\mathbf{T}}\left(\hat{\boldsymbol{\beta}}^{\mathbf{k}}\right)\left(\mathbf{y}-h\left(\hat{\boldsymbol{\beta}}^{\mathbf{k}}\right)\right)+\hat{\boldsymbol{\beta}}^{\mathbf{k}} \tag{28}$$

where $\boldsymbol{y}$ is the vector the sampled outputs of the transformer thermal model $\theta_{\text{top}}(j)$, $j = 1\ldots n$. To solve (28), we use a Gauss–Newton scheme and the Armijo rule [19]: the Newton step direction $\Delta\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{\mathbf{k+1}} - \hat{\boldsymbol{\beta}}^{\mathbf{k}}$ is scaled by a scalar multiplier so that the step length is optimized.

Comparing (28) to (21), we recognize that at each iteration, we are solving the linear least-squares problem for a linearized version of $\boldsymbol{h}(\boldsymbol{\beta})$ with $\mathbf{y} - \boldsymbol{h}\left(\hat{\boldsymbol{\beta}}^{\mathbf{k}}\right)$, representing the measured values; hence, at the solution point—or any point—the metrics we derive for assessing acceptability of the linear model will be valid for the nonlinear model.

V. Defining Metrics

In this section, we define metrics for assessing adequacy, consistency, and accuracy.

A. Adequacy

Adequacy measures whether the model has the appropriate structure to capture the features of the process being modeled. Defining which metrics are appropriate for adequacy checking is neither well defined nor standard. Researchers may, in good faith, quibble with our selection for any of our categories. Regardless of one's persuasion, the metrics we choose—residual plots and prediction R^2—are important measures of acceptability.

1) The R^2 Pred Metric: A typical metric used for assessing model accuracy (not adequacy) is R^2. The R^2 metric measures how well the predicted values $\hat{\mathbf{y}}$ (i.e., θ_{top}) capture the variation of measured values $\boldsymbol{y}$

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2}{\sum_{i=1}^{n}(\hat{y}_i-\overline{y})^2} = \frac{SS_R}{SS_r} = \frac{\sum_{i=1}^{n}(y_i-\overline{y})^2 - \sum_{i=1}^{n}(y_i-\overline{y})^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2} \tag{29}$$

where SS_R is the sum of squares of the residuals and measures the explained variation of predicted values. The variable SS_T is the total variation of the measured variables. Values of R^2 close to 1.0 indicate the predicted values of the model closely match those that are measured.

A similar and more informative metric uses the prediction residual to measure the ability of a model to predict values that are not in the training data set. The prediction residual is the ith measured data point $e_{(i)}$, which is the difference between the ith measured value, y_i, and the ith predicted value based on a model omitting y_i and fitted to the remaining $n-1$ sample points $\hat{y}_i$. Then, the prediction error sum of squares (PRESS) is defined as

$$\text{PRESS} = \sum_{i=1}^{n}\left(y_i - \hat{y}_{(i)}\right)^2. \tag{30}$$

PRESS measures the ability of the model, absent one data point, to accurately predict that data point. And the R^2_{Pred} is then defined as

$$R^2_{\text{pred}} = \frac{\sum_{i=1}^{n}(y_i-\overline{y})^2 - \sum_{i=1}^{n}\left(y_i-\hat{y}_{(i)}\right)^2}{\sum_{i=1}^{n}(y_i-\overline{y})^2} = I - \frac{\text{PRESS}}{SS_r}. \tag{31}$$

2) Plot of Residuals Versus Fitted Values: A plot of residuals e_i versus the fitted values $\hat{y}_i$ is helpful for detecting common types of model inadequacies. Violations of model assumptions are more likely to occur at remote points, and the violation may be hard to detect from the inspection of the ordinary residuals e_i or the standardized residuals. A preferred form of scaled residuals is the studentized residuals, because the residual e_i is divided by the exact standard deviation of the ith residual [15]

$$r_i = \frac{e_i}{\sqrt{MS_{\text{Res}}(1-h_{ii})}}, \quad i = 1,2,\ldots n \tag{32}$$

where

$$MS_{\text{Res}} = \frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n-k-1} \tag{33}$$

where h_{ii} is the ith diagonal element of the hat matrix $\mathbf{H} = \mathbf{X}\left(\mathbf{X}^{\mathbf{T}}\mathbf{X}\right)^{-1}\mathbf{X}^T$ and k is the number of model parameters being determined and $n-k-1$ is the number of degrees of freedom in the model [15].

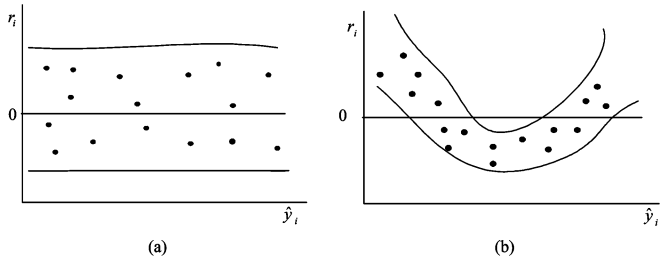

Fig. 1. Patterns for residual plots.

Fig. 1 shows typical layouts of residuals versus fitted values. If the plot looks similar to Fig. 1(a) (residuals are contained in a horizontal band), there are no obvious model defects. Plots resembling Fig. 1(b) indicate nonlinearity. This could mean that other regressor variables are needed, for instance, for a squared term.

3) Other Residual for Adequacy Checking: Other metrics we have used [16] are R^2, normal probability plots (to check that the normality assumption is not violated), a plot of residuals against regressor. These will not be reported here because

the conclusions drawn from these metrics support the conclusion drawn from the metrics R^2_{pred} and residuals-against-fitted-values plots. The normality assumption is not violated by any of our data.

B. Consistency

Consistency is a quantitative measure of the ability of a model and solution method (e.g., nonlinear least-squares regression) to produce the same model parameters when trained using similar data sets. (We define similar data sets as data sets with the same number of data points, taken from the same transformer on similar days (e.g., weekday, weekend) under the same cooling mode (e.g., NONA, NOFA, FOFA) and from the same season). Consistency measures the model and solution method. If the model's structure is inappropriate, then the parameter results will be inconsistent. This inconsistency is often an artifact of numerical problems in the solution procedure. The consistency metrics we use are eigenstructure, condition number, parameter sensitivity, and standard deviation (STD) of $\text{SSL}_{\max}$ and STD of model parameters.

1) Eigenvalues: One method of solving (28) is to invert the matrix $(X^T X)$. The eigenvalues of this symmetric matrix have the form

$$\mathbf{X^T X} = \mathbf{U \Lambda \Lambda U}^T \tag{34}$$

where $\mathbf{\Lambda}$ is a diagonal matrix with Eigenvalues $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$, assuming $\mathbf{X^T X}$ is a 4×4 matrix, which corresponds to a four-parameter model.

2) Condition Number: The condition number of a matrix is a measure of how close the matrix is to singular or, in the terminology of linear regression, how severe the multicollinearity problem is. For symmetric matrices (assuming an L2 norm), the condition number is given by

$$\boldsymbol{\kappa}(\mathbf{X^T X}) = \frac{|\lambda_1|}{|\lambda_4|}. \tag{35}$$

The condition number is based on the exact solution of two slightly different set of equations; hence, it is independent of the method used to arrive at the solution.

3) Parameter Sensitivity Values: For the linear regression problem, the sensitivity of parameters $\hat{\beta}$ to changes in the measure values $\mathbf{y}$ is given by [16] (assuming m parameters and n sampled outputs)

$$\frac{\partial \hat{\beta}}{\partial \mathbf{y}} = S^{\hat{\beta}}_{\mathbf{y}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T = \begin{bmatrix} S_{11} & \dots & S_{n1} \\ \vdots & \ddots & \vdots \\ S_{m1} & \dots & S_{mn} \end{bmatrix}. \tag{36}$$

For a particular parameter $\hat{\beta}_i$, the sensitivity to a change in the input data has an L_2 norm

$$\left\| \boldsymbol{S}^{\hat{\beta}_i}_{\mathbf{y}} \right\|_2 = \left\| \frac{\partial \hat{\beta}_i}{\partial \mathbf{y}} \right\|_2 = \| \boldsymbol{S_{i1}} \quad \dots \quad S_{im} \|_2. \tag{37}$$

For sensitive parameters, small changes in the input data can lead to large changes in solved parameter values. Since all measured data have noise, large sensitivities will generally lead to inconsistency between two models built from similar data.

4) Variance Inflation Factor (VIF): In linear regression, the variance of a parameter $\hat{\beta}_j$ is a measure of uncertainty in the parameter and is given by

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2_\varepsilon}{\sum_{i=1}^{n} (x_{ji} - \bar{x}_j)^2} \tag{38}$$

where σ^2_ε is the variance of the residuals. This definition is true provided all of the regressors are orthogonal. Orthogonality occurs if regressing each regressor x_j against all other regressors $x_1, x_2, \dots x_{j-1}, x_{j+1}, \dots x_n$ produces a value of zero. If this regression yields a nonzero value, then $\hat{\beta}_j$ is more uncertain and $\text{Var}(\hat{\beta}_j)$ is inflated

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma^2_e}{\left[\sum_{i=1}^{n} (x_{ji} - \overline{x}_j)^2\right] \left(1 - R^2_j\right)} \tag{39}$$

where R^2_j is the R^2 coefficient of the x_j regression. The corresponding variance inflation factor is

$$\text{VIF}_j = C_{jj} = \left(1 - R^2_j\right)^{-1} \tag{40}$$

where C_{jj} is the jth diagonal element of $(\mathbf{X^T X})^{-1}$. If x_j is nearly orthogonal to the remaining regressors, R^2_j is small and C_{jj} is close to unity, while if x_j is nearly dependent on some subset of the remaining regressors, R^2_j is near to unity and C_{jj} is large. One or more VIFs larger than 10 indicates that the associated regression coefficients are poorly estimated. Model misspecification and overdefining a model are sources of large VIFs [15].

Multicollinearity—another term used for the linear dependence among the regressors—occurs in the linear case because summer load I^2 is nearly in phase with the ambient temperature θ_{amb} and nearly in phase with the lagged regressor θ_{top}. There is little that can be done about this. For the nonlinear model, the derivatives are functions of these same variables and depend on the model's nonlinear form.

5) Variation of $\text{SSL}_{\max}$ *and Model Coefficients:* Performing dynamic loading is one important use of transformer thermal models. It was observed, for the case of a linear model [18], that the consistency of the maximum steady-state load ($\text{SSL}_{\max}$) highly correlated with the consistency of the dynamic load predicted by the model; hence, we define the coefficient of variation of the $\text{SSL}_{\max}$ as a consistency metric. Given samples of $\text{SSL}_{\max}$, calculated from p-independent but similar data sets, the $\text{SSL}_{\max}$ standard deviation is defined as

$$s_p = \sqrt{\sum_{i=1}^{p} \left(\text{SSL}^i_{\max} - \overline{\text{SSL}_{\max}}\right)^2} \tag{41}$$

where $\overline{\text{SSL}_{\max}}$ is the sample mean. $\text{SSL}_{\max}$ may be calculated for each model by setting $\theta_{\text{top}}(k) = \theta_{\text{top}}(k-1) = \theta^{\max}_{\text{top}} (= 110°\text{C})$ and $\theta_{\text{amb}}(k) = \theta^{\max}_{\text{amb}}$ ($= 117°\text{F}$ for Phoenix) and solving each discretized model for I^2. Checking the variation of coefficients that make up the models is also important. The STD of these is given by an equation similar to (41).

C. Accuracy

We define three metrics for testing accuracy: 1) comparison of optimized parameters with nominal values; 2) error duration curve; and 3) extrapolated load $\mathrm{SSL}_{\max}$ values.

1) Parameter Comparison With Nominal Values: A rough, but useful metric is the comparison of parameters identified from measured data with those measured during heat-run tests.

2) Residual Duration Curve: The residual duration curve functions as an interpretation similar to the load duration curve. The residuals are arranged in monotonically decreasing order $e_{[1]} > e_{[2]} > \cdots e_{[p]}$ and plotted against the normalized index. Unlike error-versus-time plots, this allows us to succinctly visualize the residuals for a larger time span.

3) Other Residual for Adequacy Checking: To test the ability of a model to extrapolated (rather than interpolate), the $\mathrm{SSL}_{\max}$ obtained for light loads can be compared with the extrapolated $\mathrm{SSL}_{\max}$ under more heavily loaded conditions. If the model and method are acceptable, the two values of $\mathrm{SSL}_{\max}$ should be comparable. While this is an important metric, the transformer loads in our dataset were not near their maximum values; hence, we could not compare the maximum load predicted with the actual maximum loads.

VI. Summary

Four top-oil models found in the literature have been defined using a common notation. All but one of these are nonlinear models. In the companion paper, these metrics will be applied to the models introduced here, and a relative ranking of the acceptability of these models will be presented. Algorithms for identifying the parameters of these models have been proposed. Metrics for assessing model adequacy, consistency, and accuracy have been proposed. It has been shown that metrics defined for linear regression may also be used in the nonlinear case.

Acknowledgment

The authors would like to thank the following people at the Arizona Public Service (APS) utility company for their financial support and the generous contribution of their time to provide feedback and data for this work: Dr. B. Agrawal, C. Hoesel, D. B. Larsen, W. Perry, S. Whitey, and J. Zhang. The authors would also like to thank G. McCulla and K. Brown of the Salt River Project for providing data.

References

[1] *IEEE Guide for Loading Mineral-Oil-Immersed Power Transformers*, IEEE C57.91-1995.

[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] D. J. Tylavsky, X. Mao, and G. A. McCulla, "Transformer thermal modeling: Improving reliability using data quality control," *IEEE Trans. Power Del.*, vol. 21, no. 3, pp. 1357–1366, Jul. 2006.

[5] X. Mao, D. J. Tylavsky, and G. A. McCulla, "Assessing the reliability of linear dynamic transformer thermal modeling," *Proc. Inst. Elect. Eng., Gen., Transm. Distrib.*, vol. 153, no. 4, pp. 414–422, Jul. 2006.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] W. H. Tang, Q. H. Wu, and Z. J. Richardson, "A simplified transformer thermal model based on thermal-electric analogy," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1112–1119, Jul. 2004.

[11] A. P. S. Meliopoulus, "State estimation methods applied to transformer monitoring," in *Proc. IEEE Power Eng. Soc. Summer Meeting,*, Jul. 2001, vol. 1, pp. 419–423.

[12] 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.

[13] H. Nordman, N. Rafsback, and D. Susa, "Temperature responses to step changes in the load current of power transformers," *IEEE Trans. Power Del.*, vol. 18, no. 4, pp. 1110–1117, Oct. 2003.

[14] D. M. Levine, P. P. Ramsey, and R. K. Smidt, *Applied Statistics for Engineers and Scientists*. Upper Saddle River, NJ: Prentice-Hall, 2001.

[15] D. C. Montgomery, E. A. Peck, and G. G. Vining, *Introduction to Linear Regression Analysis*, 3rd ed. New York: Wiley, 2001.

[16] L. Jauregui-Rivera, "Reliability assessment of transformer thermal models," Ph.D. dissertation, Arizona State Univ., Tempe, 2006.

[17] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*. Norwell, MA: Kluwer, 1999.

[18] X. Mao, "Transformer linear thermal models," Ph.D. dissertation, Dept. Elect. Eng., Arizona State Univ., Tempe, 2005.

[19] P. Bertsekas, *Nonlinear Programming*. Belmont, MA: Athena Scientific, 1995, pp. 646–646.

Lida Jauregui-Rivera (M'08) received the B.S. degree in electrical engineering from the University of San Antonio Abad of Cusco, Cusco, Perú, in 1998, the M.S. degree in electrical engineering from the University of Puerto Rico at Mayaguez, Mayaguez, Puerto Rico, in 2002, and the Ph.D. degree in electrical engineering from the Arizona State University, Tempe, in 2006.

From 1998 to 1999, she worked in the areas of generation, transmission, and distribution of energy utility industries in Peru. In 2007, she joined the marketing and trading group of the Arizona Public Service (APS) utility company. Her research interests include modeling and simulation of power systems and components, parameter estimation strategies, algorithms for the solution of parameter estimation in electrical energy processing systems, and stability and control of power systems.

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, PA, 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 Penn 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 an RCA Fellow, NASA Fellow, and member of Phi Eta Sigma, Eta Kappa Nu, Tau Beta Pi, and Phi Kappa Phi.