

Acceptability of Four Transformer Top-Oil Thermal Models: Pt. 1: Defining Metrics

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

Abstract—Eventually, prediction of transformer thermal performance for dynamic loading will be made using models distilled from measure 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: variance inflation factor (VIP) (which measure multicollinearity), condition number, eigenstructure, parameter sensitivity and standard deviation (STD) of model parameters and of predicted maximum steady state load (SSL_{max}). To assess model accuracy we use comparison of model parameters with nominal values and error duration curves. Other metrics of interest are also introduced. In a companion paper, these metric 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 C57.91, transformer, top-oil temperature, 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 allow derivation of 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.

Manuscript received October 11, 2004. This work was supported in part by funding provided by Salt River Project and Arizona Public Service and by PSERC

D. J. Tylavsky is with Arizona State University, Tempe AZ 8287-5706 USA (phone: 480-965-3460; fax: 480-965-0745; e-mail: tylavsky@asu.edu). Lida Jauregui-Rivera is a Ph.D candidate at Arizona State University.

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 assess the model consistency and accuracy.

One goal of this paper is to identify models—from the many available [1-13]—that require 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 a.k.a Nonlinear Top-Oil Model

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

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

where,

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

$$T_o = \frac{C\theta_{fl}}{P_{fl}} \quad (3)$$

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.

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 ($\theta_u + \theta_{amb}$):

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

where

θ_{amb} ambient air temperature ($^{\circ}\text{C}$);

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

$$\frac{d\theta_{top}(k\Delta t)}{dt} = \frac{\theta_{top}(k\Delta t) - \theta_{top}((k-1)\Delta t)}{\Delta t}, \quad (5)$$

$$= f(\theta_{top}(k\Delta t), \theta_{amb}(k\Delta t), \theta_u(k\Delta t))$$

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

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

where

$$n = \begin{cases} 0.8 \text{ NONA} \\ 0.9 \text{ NOFA or FOFA Non-directed} \\ 1.0 \text{ FOFA Directed} \end{cases}, \quad (7)$$

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

Assuming $n \approx 1$, the above model simplifies to:

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

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

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

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

C. Swift et al. 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.

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

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 ultimate top-oil rise, but a modified value needed for equality in (10) at rated load. Swift et al., give no values for n since they obtain n through training of their model; however they do give typical values of n .

$$n = \begin{cases} 0.8 \text{ NONA (typical)} \\ 0.8 \leftrightarrow 1.0 \text{ FA (typical)} \end{cases}, \quad (12)$$

In [6], [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,

$$\begin{aligned} \theta_{top}(k) &= \theta_{top}(k-1) \\ &+ \frac{\Delta t}{T_o} \left[\left[\frac{I^2(k)R+1}{R+1} \right] \theta_{fl}^{1/n} - [\theta_{top}(k) - \theta_{amb}(k)]^{1/n} \right], \end{aligned} \quad (13)$$

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

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

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. model; however in this model, Susa et al. retain oil viscosity as a parameter. Their governing differential equation becomes:

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

where,

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

$$\begin{aligned} \mu &= \frac{\mu_{actual}}{\mu_{rated}} = \frac{1.3573 \bullet 10^{-6} \bullet e^{\left[\frac{2797.3}{\theta_{top} + 273} \right]}}{1.3573 \bullet 10^{-6} \bullet e^{\left[\frac{2797.3}{\theta_{fl} + 273} \right]}} \\ &= e^{\left[\frac{2797.3}{\theta_{top} + 273} - \frac{2797.3}{\theta_{fl} + 273} \right]} \end{aligned} \quad (16)$$

and,

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

After discretizing with backward Euler, (14) becomes,

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

As with the Swift et al. model, the (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 observer that when $n=1$ all of the models are identical; however, in general, there is no one cooling condition for which all of are the same. For NOFA conditions, we compare the time rate of change ($^{\circ}\text{C/hr}$) in θ_{top} , (in (6), (9),(13),(18)) due to a step increase in load to $I=0.9$ pu, assuming:

$$\begin{aligned} R &= 7; & T_0 &= 2(h); & \theta_{fl} &= 55(^{\circ}\text{C}); & \Delta t &= 0.25(h); \\ \theta_{top}(k-1) &= 30(^{\circ}\text{C}); & \theta_{amb}(k) &= 70(^{\circ}\text{F}) = 21(^{\circ}\text{C}); \\ \mu_{rated} &= 0.897 \cdot 10^{-3} (kg/ms); \end{aligned}$$

Linear Clause 7 $n=(\text{not used})$ $(^{\circ}\text{C/h})$	Nonlinear Clause 7 $n=0.9$ $(^{\circ}\text{C})$	Swift et al. $n=1.0$ $(^{\circ}\text{C})$	Susa et al. $n=0.8$ $(^{\circ}\text{C})$
16.38	16.75	16.38	18.01

Except for the Susa et al. model, these models appear very similar in their time rate of change performance. Since our goal is to match measured data within a 2-3 $^{\circ}\text{C}$ 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've found test report data that does not match measure 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 developing 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 $\theta_{top}(m)$. A multiple regression model that describes the relationship between the process output, Y , and k regressor values is:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \quad (19)$$

where the parameter $\beta_j, j=1,2,\dots,k$ are called the true regression coefficients, x_j are regression variables (in the linear model case I^2 , θ_{amb} , and, as a lagged regressor, θ_{top}). In vector form the regression variables are written as \mathbf{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

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}_0 + \boldsymbol{\varepsilon} \quad (20)$$

where \mathbf{Y} , $Y_j, j=1,2,\dots,n$, 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 $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of random errors.

Because 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

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (21)$$

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

$$\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{e} \quad (22)$$

where \mathbf{e} is an $n \times 1$ vector of residuals, $\hat{\boldsymbol{\beta}}_0 = \bar{\mathbf{y}}$ is an $n \times 1$ vector of constant scalar values, and $\bar{\mathbf{y}}$ is the mean of the sampled values of \mathbf{y} . The predicted values, $\hat{\mathbf{y}}$, are given by,

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\beta}}_0 \quad (23)$$

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

B. Nonlinear Modeling

The nonlinear problem is:

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

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

The optimal values of $\boldsymbol{\beta}$ which minimize the square error in (24), (assuming all measurements have equal weights) obey:

$$\min_{\forall \boldsymbol{\beta}} (\mathbf{Y} - \mathbf{h}(\mathbf{x}, \boldsymbol{\beta}))^2 \quad (25)$$

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 \mathbf{x} in (25) gives,

$$\begin{aligned} \frac{\partial (\mathbf{Y} - \mathbf{h}(\boldsymbol{\beta}))^2}{\partial \boldsymbol{\beta}} &= 2 \mathbf{X}^T (\boldsymbol{\beta}) (\mathbf{Y} - \mathbf{h}(\boldsymbol{\beta})) = 0 \\ \therefore \mathbf{Y} - \mathbf{h}(\boldsymbol{\beta}) &= 0 \end{aligned} \quad (26)$$

where

$$\mathbf{X}(\boldsymbol{\beta}) = \frac{\partial \mathbf{h}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \quad (27)$$

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:

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

where \mathbf{y} is the vector the sampled outputs of the transformer thermal model, $\theta_{top}(j)$, $j=1\dots n$. To solve (28), we use a Gauss-Newton scheme and the Armijo rule [19]: the Newton step direction, $\Delta \hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{k+1} - \hat{\boldsymbol{\beta}}^k$, 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, $\mathbf{h}(\boldsymbol{\beta})$, with $\mathbf{y} - \mathbf{h}(\hat{\boldsymbol{\beta}}^k)$ 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 your 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 $\hat{y}_i(\theta_{top})$, capture the variation of measured values, \mathbf{y} .

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = \frac{SS_R}{SS_T} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (29)$$

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 match closely those measured.

A similar and more informative metric uses the prediction residual to measure the ability of a model to predict values not in the training data set. The prediction residual is the i th measured data point, e_i , which is the difference between the i th measured value, y_i , and the i th 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:

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2 \quad (30)$$

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:

$$R^2_{Pred} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n (y_i - \hat{y}_{(i)})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{PRESS}{SS_T} \quad (31)$$

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 inspection of the ordinary residuals, e_i , or the standardized residuals. A preferred form of scaled residuals are the studentized residuals, because the residual e_i , is divided by the exact standard deviation of the i th residual [15].

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

where

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

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

Fig. 1 shows typical layouts of residuals versus fitted values. If the plot looks like Fig. 1a (residuals are contained in a horizontal band) there are no obvious model defects. Plots resembling Fig. 1b indicate nonlinearity. This could mean that other regressor variables are needed, for instance, 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), plot of residuals against regressor. These will not be reported on 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.

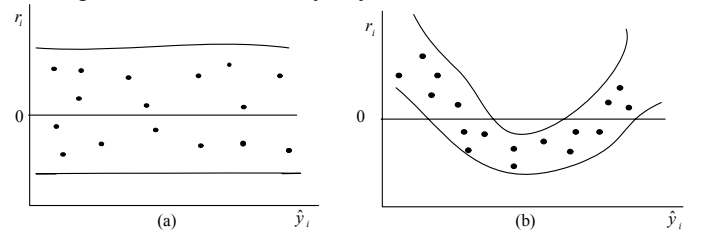


Fig. 1 Patterns for Residual Plots

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 both 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 STD of SSL_{max} and STD of model parameters.

1) Eigenvalues

One method of solving (28), is to invert the matrix $(\mathbf{X}^T\mathbf{X})$. The eigenvalues of this symmetric matrix have the form,

$$\mathbf{X}^T\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \quad (34)$$

Where $\mathbf{\Lambda}$ is a diagonal matrix with Eigenvalues, $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$, assuming $\mathbf{X}^T\mathbf{X}$ is a 4X4 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:

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

The condition number is based on the exact solutions 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):

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

For a particular parameter, $\hat{\beta}_i$, the sensitivity to a change in the input data has an L₂ norm:

$$\left\| \mathbf{S}_{\mathbf{y}}^{\hat{\beta}_i} \right\|_2 = \left\| \frac{\partial \hat{\beta}_i}{\partial \mathbf{y}} \right\|_2 = \left\| \mathbf{S}_{i1} \quad \cdots \quad \mathbf{S}_{im} \right\|_2 \quad (37)$$

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,

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma_{\hat{\epsilon}}^2}{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2} \quad (38)$$

where $\sigma_{\hat{\epsilon}}^2$ 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 the $\text{Var}(\hat{\beta}_j)$ is inflated:

$$\text{Var}(\hat{\beta}_j) = \frac{\sigma_{\hat{\epsilon}}^2}{\left[\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 \right] (1 - R_j^2)} \quad (39)$$

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

$$\text{VIF}_j = C_{jj} = (1 - R_j^2)^{-1} \quad (40)$$

where C_{jj} is the j th diagonal element of $(\mathbf{X}^T\mathbf{X})^{-1}$. If x_j is nearly orthogonal to the remaining regressors, R_j^2 is small and C_{jj} is close to unity, while if x_j is nearly dependent on some subset of the remaining regressors R_j^2 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 over-defining a model are other sources of large VIF's [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 also 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 also depend on the model's nonlinear form.

5) Variation of 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 (SSL_{max}) was highly correlated with consistency of dynamic load predicted by the model; hence we define the coefficient of variation of the SSL_{max} as a consistency metric. Given samples of SSL_{max} calculated from p independent but similar data sets, the SSL_{max} standard deviation is defined as:

$$s_p = \sqrt{\sum_{i=1}^p \left(SSL_{max}^i - \overline{SSL_{max}} \right)^2} \quad (41)$$

where $\overline{SSL_{max}}$ is the sample mean. SSL_{max} may be calculated for each model by setting $\theta_{top}(k) = \theta_{top}(k-1) = \theta_{top}^{max} (=110^\circ\text{C})$ and $\theta_{amb}(k) = \theta_{amb}^{max} (=117^\circ\text{F}$ for Phoenix) and solving each of the discretized models 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: Comparison of optimized parameters with nominal values, error duration curve, and extrapolated load 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 has an interpretation similar to the load duration curve. The residuals are arranged in a monotonically decreasing order, $e_{[1]} > e_{[2]} > \dots e_{[p]}$ 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 SSL_{max} obtained for light loads can be compared with the extrapolated SSL_{max} under more heavily loaded conditions. If the model and method are acceptable, the two values of 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 model. 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.

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.

ACKNOWLEDGEMENT

The authors wish to thank the following people at Arizona Public Service for their financial support and the generous contribution of their time to provide feedback and data for this work: Dr. Baj Agrawal, Charles Hoesel, D. Brad Larsen, Wayne Perry, Shaun Whitey and James Zhang. We also thank Gary McCulla and Kenneth Brown of Salt River Project for providing us with data. Also thanks go to Dr. Xiaolin Mao of Arizona State University who massaged the data we received into a clean and usable form.

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. on Power Delivery*, Vol. 15, No. 1, (Jan. 2000), pp. 178-185.
 - [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. On Power Delivery*, vol. 12, no. 1, Jan. 1997.
 - [4] D. J. Tylavsky, X. Mao, G. A. McCulla, "Transformer Thermal Modeling: Improving Reliability Using Data Quality Control," *IEEE Trans. On Power Delivery*, (In Press) vol. 21, no. 3, Jul 2006.
 - [5] X. Mao, D. J. Tylavsky, G. A. McCulla, "Assessing the Reliability of Linear Dynamic Transformer Thermal Modeling," *IEE Proc. On Generation, Transmission and Distribution* (In Press), vol. ??, no. ?, Month. 2007
 - [6] G. Swift, T. S. Molinski, and W. Lehn, "A Fundamental Approach to Transformer Thermal Modeling—Part I: Theory and Equivalent Circuit," *IEEE Trans. on Power Delivery*, Vol. 16, No. 2, (Apr. 2001), pp. 171-175.
 - [7] G. Swift, T. S. Molinski, R. Bray, and Robert Menzies, "A Fundamental Approach to Transformer Thermal Modeling—Part II: Field Verification," *IEEE Trans. on Power Delivery*, Vol. 16, No. 2, (Apr. 2001), pp. 176-180.
 - [8] D. Susa, M. Lehtonen, and H. Nordman, "Dynamic Thermal Modeling of Power Transformers," *IEEE Trans. on Power Delivery*, Vol. 20, No. 1, (Jan. 2005), pp. 197-204.
 - [9] W. H. Tang, Q.H. Wu, and Z.J. Richardson, "Equivalent heat circuit bases power transformer thermal model," *Electric Power Applications, IEE Proceedings*, Vol. 149, No. 3, (Mar. 2002), pp. 87-92.
 - [10] W. H. Tang, Q.H. Wu, and Z.J. Richardson, "A Simplified Transformer Thermal Model Based on Thermal-Electric Analogy," *IEEE Trans. on Power Delivery*, Vol. 19, No. 3, (Jul. 2004), pp. 1112-1119.
 - [11] A. P. S. Meliopoulos, "State Estimation Methods Applied to Transformer Monitoring," *Power Engineering Society Summer Meeting, 2001, IEEE* Vol. 1, 15-19, (Jul. 2001), 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. on Power Delivery*, Vol. 18, No. 1, (Jan. 2003), pp. 142-148.
 - [13] H. Nordman, N. Rafsback, D. Susa, "Temperature Responses to Step Changes in the Load Current of Power Transformers," *IEEE Trans. on Power Delivery*, Vol. 18, No. 4, (Oct. 2003), pp. 1110-1117.
 - [14] D. M. Levine, P.P. Ramsey, R. K. Smidt, *Applied Statistics For Engineers and Scientists*, New Jersey: Prentice Hall, 2001. ISBN 0-13-488801-4.
 - [15] D. C. Montgomery, E. A. Peck, G. G. Vining, "Introduction to Linear Regression Analysis," Third Edition, Wiley-Interscience, New York 2001.
 - [16] Lida Jauregui-Rivera, *Reliability Assessment of Transformer Thermal Models*, Ph.D. Dissertation, Arizona State University, 2006.
 - [17] A. Monticelli, *State Estimation in Electric Power Systems: A Generalized Approach*, Kluwer Academic Publishers, Boston, 1999.
 - [18] Xiaolin Mao, *Transformer Linear Thermal Models*, Ph.D. Dissertation, Arizona State University, 2005.
 - [19] .P. Bertsekas, *Nonlinear Programming*. Athena Scientific, Belmont, Massachusetts, 1995, 646 pp.
- Daniel J. Tylavsky** (SM 88) received the B.S. degree in Engineering Science in 1974, his M.S.E.E. in 1978 and the Ph.D. degree in Electrical Engineering in 1982, from the Pennsylvania State University.
- From 1974-76, he worked for Basic Technology Inc. and in 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 where he is currently Associate Professor of Engineering.
- He is a member of the IEEE PES and IAS societies, and is an RCA Fellow, NASA Fellow, and member of Phi Eta Sigma, Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi.
- Lida Jauregui-Rivera** Received the B.S. degree in Electrical Engineering from University of San Antonio Abad of Cusco, Peru, in 1998, and the M.S. degree in Electrical Engineering from University of Puerto Rico at Mayaguez, Mayaguez, in 2002. She is currently working toward the Ph.D. degree in electric power engineering at Arizona State University, Tempe. Her research interest includes modeling and simulation of power system and its components, system identification, parameter estimation, algorithms for the solution of parameter estimation in electrical energy processing systems.