Improving the Process of Certified and Witnessed Factory Testing for Chiller Procurement

Joseph P. Corcoran, P.E.

T. Agami Reddy, Ph.D., P.E.
Member ASHRAE

ABSTRACT

Currently, consulting engineers and/or design/build engineers involved in the process of performance-based chiller procurement perform four activities: (1) determine bid specifications, (2) identify a few different vendors and sales representatives of chiller manufacturers and request them to submit bids along with chiller performance maps in the form of tabular data, (3) select a particular chiller (or mix of chillers) from these bids, and (4) perform a certified and witnessed factory test on the performance of the selected chiller, which provides an insurance policy and a last chance to reject the equipment if the engineer deems that the chiller is not consistent with manufacturer's published performance data. Usually, the factory tests are performed based on ARI Standard 550/590 or on operating conditions that better reflect the actual load profile. This paper suggests a methodology to select the optimal set of test conditions that provide the most useful chiller performance data. It is shown that using the Gordon-Ng (GN) chiller model allows the engineer to accurately predict the complete chiller performance map with only four well-chosen operating points*. This paper justifies both the choice of the GN model and the experimental design proposed using monitored data from two actual chiller data sets (one from a laboratory chiller, and the other from a field-operated chiller using about five months of data). The methodology suggested in this paper should be of great benefit to chiller professionals, and enhance the quality assurance process of the entire chiller procurement process.*

BACKGROUND

A performance-based chiller procurement process involves four phases. The first phase is one during which the consulting and/or design/build engineer determines the bid specifications. These include such factors as total design load, anticipated load profile, design entering and leaving chiller temperatures, acceptable refrigerants, etc. (Taylor 2001). The second phase involves identifying a short list of vendors based on past experience and local representation and requesting them to submit bids based on performance specifications. The third phase is for the consulting engineer to select a particular chiller (or mix of chillers) from these bids that he deems most appropriate (based on criteria such as life-cycle costs, refrigerant type, client-imposed constraints such as first cost, space, etc.). This may involve performing several simulation runs using computer programs such as DOE-2 (Winklemann et al. 1993). The last phase involves a certified and witnessed factory test so that the consulting engineer can verify the performance of the selected chiller. This provides a reasonable assurance test, i.e., an insurance policy, and a last chance to reject the equipment if the engineer deems that the chiller is not up to manufacturer's published tabular data (Taylor 2001). How the consulting engineer chooses to specify the factory tests is discretionary. Usually, the engineer would base the selection on ARI Standard 550/590 (ARI 1998), or he/she may select operating conditions that better reflect the actual load profile.

OBJECTIVE

This objective of this paper is to suggest a methodology to rationalize and improve the last phase, namely: (1) to provide guidelines to the consulting engineer as to how to select the optimal set of test conditions that provide most useful chiller performance and (2) to illustrate how to analyze these data using a well-accepted chiller model. After a brief

Joe Corcoran is lead building mechanical engineer at CDI Engineering Group, Inc., Philadelphia, Penn. **Agami Reddy** is an associate professor in the Civil and Architectural Engineering Department, Drexel University, Philadelphia, Penn.

review of ARI standard 550/590, two of the most common chiller models are described. This is followed by a section describing pertinent theoretical concepts of model parameter estimation, experimental design, and statistical measures that allow quantification of the "information" contained in a data set. A methodology of how to select the optimal testing conditions during a certified and witnessed test is proposed and validated with actual chiller data from a lab. Finally, the accuracy of chiller models thus identified is compared to models identified from larger data sets using actual data from two single-stage vapor compression liquid chillers (one from a lab chiller, and the other from a field operated chiller).

ARI STANDARD 550/590

ARI standard 550/590 (ARI 1998) was originally developed to provide a rational means to compare the performance of chillers from different manufacturers under repeatable and controlled test conditions. It is intended for the guidance of the industry, including manufacturers, engineers, installers, contractors, and users. The standard establishes definitions, requirements for testing and rating, minimum data for requirements for published ratings, marking and nameplate data, and conformance conditions for water-chilling packages using the vapor compression cycle. Included in this standard are procedures for determining chiller performance under standard and application (nonstandard) rating conditions, at full- and part-load capacity, and the method for calculating the capacity tolerance the chiller must meet to comply with its stated rating. Published ARI ratings for all water-chilling packages must include the mandatory Standard Rating with 29.4°C (85°F) condenser water flowing at a rate of 0.054 L/s per kW (3.0 gpm per ton) and a fouling factor of 0.000044 m^2°C/W (0.00025 hr·Ft^2·°F/Btu), and 6.7°C (44°F) chilled water flowing at a rate of 0.043 L/s per kW (2.4 gpm per ton) with a fouling factor of 0.000018 m^2·°C/W (0.0001 hr·Ft^2·°F/Btu). Recommendations for application ratings that are intended for part-load conditions are also provided. These include measurements under varying conditions, within the operating range of the equipment, as follows: condenser water from 18.3°C (65°F) to 40.6°C (105°F) (at ≤2.78°C (5°F) increments) and chilled water from 4.44°C (40°F) to 8.89°C (48°F) (at ≤1.1°C (2°F) increments).

Part-load ARI ratings specified by the IPLV (integrated part-load value) or NPLV (non-standard part-load value) are based on the part-load energy efficiency at 100%, 75%, 50%, and 25% load points. To determine IPLV for water-cooled chillers, the following relation is used:

$$IPLV = 0.01A + 0.42B + 0.45C + 0.12D \qquad (1)$$

where

A = COP or EER at 100% and T_{cdi} = 29.4°C (85°F)

B = COP or EER at 75% and T_{cdi} = 23.9°C (75°F)

C = COP or EER at 50% and T_{cdi} = 18.3°C (65°F)

D = COP or EER at 25% and T_{cdi} = 18.3°C (65°F)

The chiller water outlet temperature T_{cho} is kept constant at 6.7°C (44°F). It is clear that four test conditions are required to obtain the relation given by Equation 1. The ARI standard also recommends that in case the chiller is to be operated over a narrower range of load variation than that needed for Equation 1, then the four data points should be selected uniformly over the operating range of chiller load. Note that Equation 1 presumes a certain annual or seasonal load distribution and should be interpreted accordingly.

DIFFERENT CHILLER MODELS

While ARI Standard 550/590 is targeted toward certification of published chiller ratings under rigidly controlled conditions, in-situ chiller tests under field conditions are not specifically addressed. Nevertheless, this standard is often used as the basis of the certified and witnessed factory tests, though the design engineer may often request more tests than the four test conditions needed to calculate IPLV. The consulting engineer will determine the selected chiller to be satisfactory if the chiller COP, under the certified and witnessed factory tests, is close to those given by the manufacturer. Hence, chiller performance is judged against the few tests actually performed. A more comprehensive evaluation process is to compare chiller performance against the entire chiller map, i.e., against the entire tabular data provided by the manufacturer. This can be done by identifying a mathematical model of chiller performance from the certified and witnessed test data.

Conceptually, one can distinguish two types of models: gray-box models, based on scientific understanding of the process, and black-box models, based on empirical and statistical considerations (ASHRAE 2001). In turn, two physical chiller model formulations that have been proposed in the literature are suitable for our purpose. The toolkit developed by Bourdouxhe et al. (1999) consists of a set of computational tools based on physical algorithms for primary HVAC equipment. These are closer to lumped inverse component models but the estimation of the parameters is nonlinear. The other formulation is the *Universal Thermodynamic Model* proposed by Gordon, Ng, and several collaborators, which is described in numerous publications, most of which have been succinctly described in a textbook by Gordon and Ng (2000). The GN model is a simple, analytical, universal (i.e, applies to all chiller types) model for chiller performance based on first principles of thermodynamics and linearized heat losses. The model predicts the dependent chiller COP (defined as the ratio of chiller (or evaporator) thermal cooling capacity by the electrical power P consumed by the chiller (or compressor) with specially chosen independent (and easily measurable) parameters such as the fluid (water or air) inlet temperature from the condenser T_{cdi}, fluid temperature entering the evaporator (or the chilled water supply temperature from the building) T_{chi}, and the thermal cooling capacity of the evaporator. The GN model is a three-parameter model, which, for parameter identification, takes the following form:

$$\left(\frac{1}{COP}+1\right)\frac{T_{chi}}{T_{cdi}}-1 = b_1\frac{T_{chi}}{Q_{ch}} + b_2\frac{(T_{cdi}-T_{chi})}{T_{cdi}Q_{ch}} + b_3\frac{(1/COP+1)Q_{ch}}{T_{cdi}} \tag{2}$$

where the temperatures are in absolute units, and the parameters of the model have the following physical meaning:

b_1 = the total internal entropy production rate in the chiller,

b_2 = the rate of heat losses (or gains) from (or into) the chiller,

b_3 = the total heat exchanger (evaporator and condenser) thermal resistance.

If we introduce

$$x_1 = \frac{T_{chi}}{Q_{ch}},\ x_2 = \frac{T_{cdi}-T_{chi}}{T_{cdi}Q_{ch}},\ x_3 = \frac{(1/COP+1)Q_{ch}}{T_{cdi}}$$
$$\text{and} \tag{3}$$
$$y = \left(\frac{1}{COP}+1\right)\frac{T_{chi}}{T_{cdi}}-1\ ,$$

Equation 2 becomes

$$y = b_1x_1 + b_2x_2 + b_3x_3\ . \tag{4}$$

Hence, the reformulation results in a model that is linear in the parameters but without an intercept term. Note that the model specified by Equation 4 has only three parameters, and so a minimum of three data points are needed to identify the model parameters.

It must be pointed out that Gordon and Ng (2000) proposed an earlier chiller model called the "quasi-empirical chiller model," which, though accurate for prediction purposes, is more restrictive and does not allow physical interpretation of the model parameters and, hence, has been superseded by the "universal thermodynamic model" given by Equation 2. The quasi-empirical model is also a three-parameter linear model that has been evaluated by Phelan et al. (1997) against the same field data used in this study.

Whereas the structure of a gray box model such as the GN model is determined from the underlying physics, the black box model is characterized as having no (or sparse) information about the physical problem incorporated in the model structure, and a statistical or empirical relationship between input and output is formulated. Using the same parameters as those for the physical model, previous research studies have adopted a second order linear polynomial model (termed *multivariate polynomial model* by Reddy and Andersen [2002]) which has the following structure:

$$COP = \alpha + \beta_1T_{cdi} + \beta_2T_{chi} + \beta_3Q_{ch} + \beta_4T_{edi}^2 + \beta_5T_{chi}^2 + \beta_6Q_{ch}^2 + \beta_7T_{cdi}T_{chi} + \beta_8T_{cdi}Q_{ch} + \beta_9T_{chi}Q_{ch} \tag{5}$$

TABLE 1
Experimental Uncertainty of the Various Performance Variables for the 10.5 kWt Chiller Tested in a Laboratory (Taken from Gordon and Ng 2000)

Variable	Uncertainty
T_{cdi}	±0.05 K
T_{chi}	±0.05 K
Q_{evap}	±0.2 kW
P	±0.01 kW

This model has 10 coefficients that need to be identified from monitored data. These coefficients, unlike the three coefficients appearing in the GN model, have no physical meaning and their magnitude cannot be interpreted in physical terms. Usually one needs to retain in the model only those parameters that are statistically significant, and this is done by step-wise regression (Chatterjee and Price 1981). Note, that one would need at least 10 data points in order to identify the model parameters of Equation 5, which is more than three times the number of data points needed for the GN model and more than the four data points provided by the ARI standard.

Hydeman and Gillespie (2002) and Hydeman et al. (2002) adopted a different version of the black-box chiller model based on the algorithm used in the commercially available DOE-2 simulation program (Winklemann et al. 1993). It consists of determining three models: (1) the available capacity as a function of evaporator and condenser temperatures, (2) the full-load efficiency as a function of evaporator and condenser temperatures, and (3) efficiency as a function of percentage loading. In total, there are 15 model parameters that need to be identified from data. The authors recommend that 20 to 30 data points be used for model identification covering the entire range of operating conditions.

DESCRIPTION OF THE DATA SETS

Laboratory Chiller Data

The data in this study, taken from Gordon and Ng (2000), are from a semi-hermatic reciprocating chiller with a nominal cooling capacity of 10.5 kW. It consists of 30 measurements of (T_{chi}, T_{cdi}, Q_{ch}, P) taken with a high accuracy instrument in a laboratory test loop. The units of these variables and their measurement accuracy are shown in Table 1. Scatter plots of capacity Q_{ch} and COP vs. T_{cdi} and T_{chi} are shown in Figure 1. These plots provide clear insight into how the tests were designed. Basically, they were performed with T_{chi} kept constant at six different temperature levels and varying T_{cdi} at five different levels. Since the chilled water flow rate and the chiller outlet temperature are closely controlled, capacity is a very strong function of T_{chi} (correlation coefficient of 0.94). Hence, there are essentially only two independent variables (T_{chi} and T_{cdi}). The chiller load was only varied down to 60% of full load, which is an important factor to note in view of the

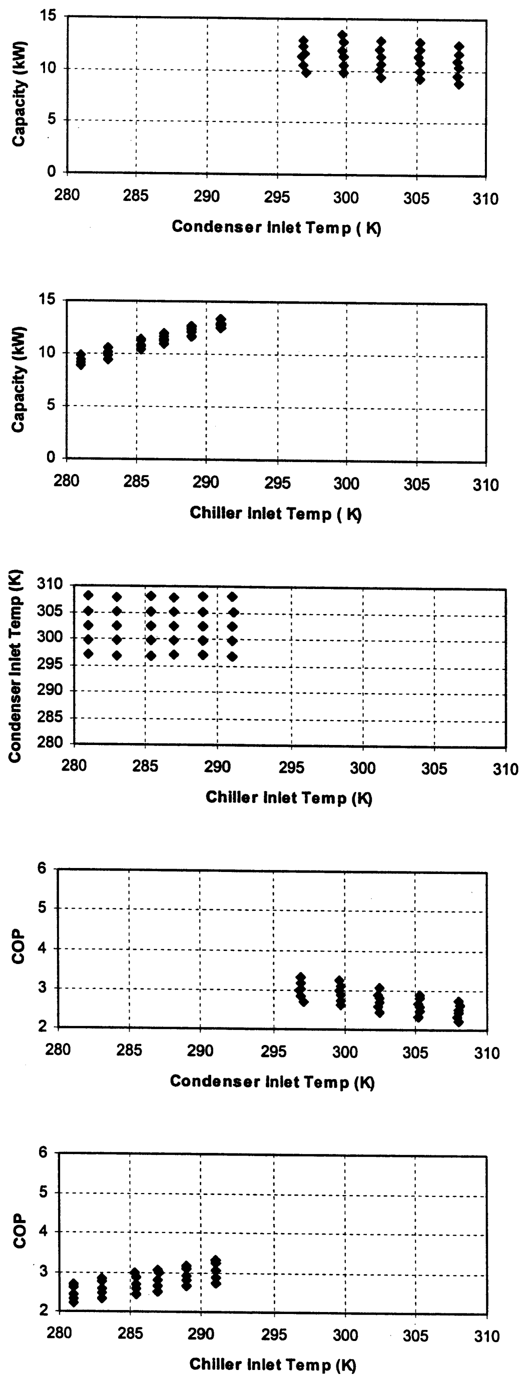

Figure 1 *Laboratory chiller data (30 data points).*

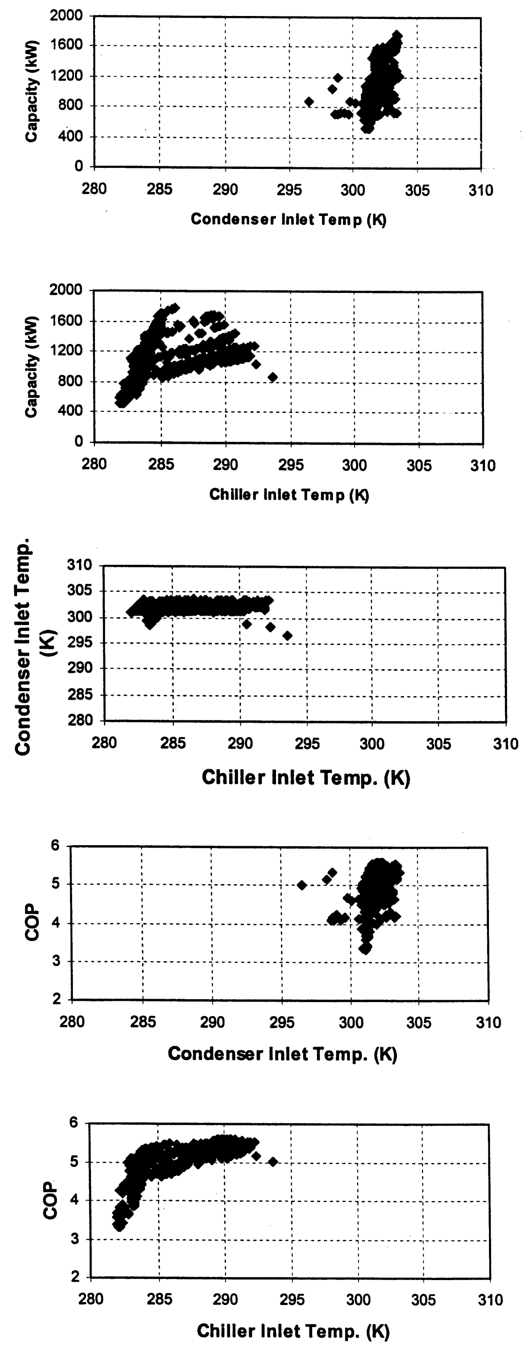

Figure 2 *Field-operated chiller data at Toronto (810 data points).*

analysis results discussed later in the paper. On the whole, and as one would expect from a well-defined lab testing protocol, there are clear and well-defined patterns in the data.

Field Chiller Data

Hourly field-monitored data from a centrifugal chiller located in Toronto, Ontario, Canada, from June to October 1997, consisting of the same four variables, are used in our analysis. The dataset contains 810 observations and is fully described in Reddy and Andersen (2002). The measurement accuracy of the sensors would typically be two to three times poorer than lab sensors (see Table 1). The same type of scatter plots as for the lab chiller have been generated and shown in Figure 2. The chiller load went down to 29% of full load. We note large scatter with some outlier points and no clear similarity between Figure 1 and Figure 2. For example, capacity vs. T_{chi} in Figure 2 show several linear clouds of variation rather than a clear single linear plot as in Figure 1. Also, T_{chi} and capacity are not as strongly correlated as for the lab chiller. Further, T_{cdi} exhibits small variation (only about 4°C if the few outliers are not considered), which is smaller than the relative variation in T_{chi} or Q_{ch}. Hence, in this case, it would seem that selecting T_{chi} and Q_{ch} as the two primary regressor variables would be better in terms of parameter estimation than selecting T_{chi} and T_{cdi}. The field data are useful in that they serve as a reality check on the proposed methodology; namely, can the same experimental design procedure involving just four data points provide a model accurate and robust enough to predict actual variation of COP under field conditions?

MEASURES OF INFORMATION IN DATA

Parameter estimation is the science of determining the optimal numerical values of the model coefficients, such as $b = [b_1 \ b_2 \ b_3]$ in Equation 4. Let X be the matrix containing the regressor data points as row vectors. Adopting matrix notation, the ordinary least-squares (OLS) method can be formulated as follows:

$$Y = X b \tag{6}$$

where

$$X = [T_{chi} T_{cdi} Q_{ch}] \tag{7}$$

In fact, OLS is the best method to use when there are no errors in X, when there is no co-linearity between the regressors, when model residuals have constant variance and are not patterned, and when the errors are normally distributed. In this case, the parameter set b is determined such that the sum of error squares function is minimized. This leads to the system of normal equations, provided matrix X is not singular (Chatterjee and Price 1981):

$$b_{\mathrm{OLS}} = (X^T X)^{-1} X^T Y \tag{8}$$

Experimental design concerns itself with the determination of the optimal set of experiments one needs to perform in order to obtain sound and robust estimates of model parameter b (see, for example, Beck and Arnold [1977]; Box et al. [1978]). Consider a chiller operated in the field. Since the chiller operates in a fairly repetitive manner over the day and from day to day, collecting more data does not necessarily lead to a "richer" data set, i.e., one that provides more representative and accurate values of the model parameters. One needs to have a measure by which the "richness" in a particular data set can be evaluated. Equation 8 provides an indication of how to select such indices called "measures of information" (see Reddy et al. 2002). One simple measure of information is the **trace** of a matrix (see any pertinent textbook):

$$I_1 = trace((X^T X)^{-1}) \tag{9}$$

Since the trace of a matrix is equal to the sum of the diagonal elements, I_1 of Equation 9 can be interpreted as the sum of the parameter variances up to a scaling factor. There are pros and cons in using this measure of information: (a) the advantage is that we just have to monitor one single variable instead of the full set of parameters in order to determine the amount of data necessary to obtain consistent parameter estimates; (b) the drawback is that the measure does not take into account the correlation between the parameter estimates (the elements outside the diagonal of $(X^T X)$) and can be misleading if the parameters are seriously correlated (which is often the case as suggested by the case study presented in Reddy and Andersen (2002).

Another measure of information that is also widely used (Madsen and Holst 1997) is the **log of the mean of the determinant**:

$$I_2 = -\ln \ (det(X^T X)/m) \tag{10}$$

where m is the number of observations used to calculate ($X^T X$). In Equation 10, I_2 measures the average amount of information, and it is often used in experimental design to determine optimal input sequences with respect to a given model structure. Although we are usually unable to control the input sequences for field-monitored data, the measure is still useful since it allows the determination of the amount of information contained in different datasets. The measures of information I_1 or I_2 have been shown to be valid indices in that they provide the necessary insight into whether the data being collected are bringing in new information likely to lead to the identification of sounder and more representative model parameters. Measure I_2 is probably better than I_1 since, while satisfying (a) above, it also accounts for the covariance and variances between the parameters. Chiller performance data tend to exhibit such correlated regressor behavior (Reddy and Andersen 2002).

EXPERIMENTAL DESIGN PROPOSED

Experimental design theory (see Box et al. 1978) provides the following guidelines in selecting optimal data sets for identifying sound regression models:

a. if the model is known to be *strictly linear*, then it is best to select values of the regressor set X at either end of the variation range of the regressors;
b. if the model is only *approximately linear*, then it is best to select regressor values spread uniformly over their range of variation.

It is obvious that the ARI standard adopts the suggestion (b) above for its IPLV test (see Equation 1) in terms of its load variation. In order to keep the number of tests low, the standard specifies specific values of T_{chi} and T_{cdi}, which are probably practical and realistic values.

In order to determine the category to which the GN model applies, we shall use the measures of information I_1 and I_2 defined by Equations 9 and 10 along with the lab chiller data

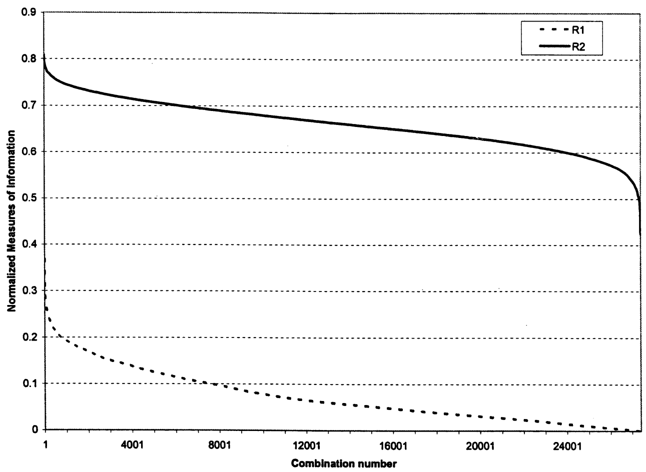

Figure 3 *Values of the ratios* R_1 *and* R_2 *defined by Equation 11, sorted by descending order, for the lab chiller data. The x-axis represents the number of combinations of 30 data points taken 4 at a time.*

consisting of 30 data points. Though only 3 data points are needed to identify the three GN model parameters, it is advisable from a statistical viewpoint to have additional degrees of freedom during parameter identification. Selecting four data points provides one degree of freedom and is not any more demanding experimentally than the four test data points proposed by the ARI standard during IPLV determination. We need to compute these measures I_1 and I_2 for all the combinations possible. The number of trials in such a case is:

$$^{n}C_r = {}^{30}C_4 = \frac{30 \times 29 \times 28 \times 27}{4 \times 3 \times 2 \times 1} = 27,405$$

There are thus 27,405 possible combinations of how we can select a set of 4 data points from the complete set of 30 values. It would be better to normalize the measures I_1 and I_2 so that their numerical values convey some sort of physical meaning. An obvious way to do so is as follows:

$$R_1 = \frac{I_1(30)}{I_1(4)} \qquad \text{and} \qquad R_2 = \frac{I_2(30)}{I_2(4)} \tag{11}$$

where $I_1(30)$ signifies that all 30 data points have been used for computing I_1. For the GN data set, $I_1(30) = 77{,}400$ and $I_2(3) = 10.585$. We have computed the ratios R_1 and R_2 for all 27,405 combinations, and plotted them in descending order in Figure 3. The best set of 4 data points would obviously correspond to that which has the highest value of either R_1 or R_2. We find that R_2 is a better measure than R_1 (consistent with the recommendation made by a previous study by Reddy et al. (2002) using field-monitored data) since it has less variability, indicating a more robust selection procedure. The data sets that have the highest R_2 values are those that correspond to the following scheme:

1. Low T_{chi} and low T_{cdi}
2. Low T_{chi} and high T_{cdi}
3. High T_{chi} and low T_{cdi}
4. High T_{chi} and high T_{cdi}

(12)

This is the exact same experimental design proposed by scheme (a) above. This clearly and unequivocally indicates that the GN model can be considered to be a strictly linear model and that adopting a selection procedure involving choosing values at either end of the variation of the two most influential regressors (i.e., those regressors that have the most variation) would result in the best model that a data set of four data points can provide. Recall, as noted earlier, that during the lab chiller tests, the load Q_{ch} was varied down to only 60% of full load. This factor along with strong correlation between Q_{ch} and T_{chi} resulted in $\{T_{chi}$ and $T_{cdi}\}$ being the more influential set of regressor variables. Now the final issue that still needs investigation is whether such a model is accurate enough for the procurement engineer to compare its predictions against data from the tabular chiller map data provided by the chiller manufacturer (and also serve as a chiller model for subsequent field commissioning and evaluation of chiller performance under routine operation).

The results of the analysis are summarized in Table 2 for both chiller data sets. The first row, denoted by "All data," pertains to the case when all data points are used to identify the GN model (30 in the case of lab data and 810 in the case of the field-operated chiller). The models are excellent, with very high R^2, and CV values of 1% and 1.5%, respectively. Note that the CV and NMBE values are calculated using Equations A2 and A4 with the COP values and not the y variable of Equation 4. When four data points are used ($n = 4$), the degrees of freedom (d.f.) = $n - p = 4 - 3 = 1$, while (d.f.) when 30 data points are used is equal to 27. It must also be pointed out that models of field-operated HVAC&R equipment with CV values less than 2% to 3% can be considered to be excellent models (Reddy and Andersen 2002). For the second set of runs, we have selected data points representing the full operating range of the chiller, i.e., distributed uniformly over the load variation (10 for the lab chiller, and 22 for the field chiller). The internal CV values (i.e., the CV values using these 10 and 22 data points, respectively) are poorer than those of the first run, as is to be expected. The external CV values are the normalized root mean square errors of the model applied to the remainder of the data points (20 and 788, respectively). The third run pertains to selecting the four data points in accordance with ARI guideline 550/590. In this case, we selected four equally spaced data points (load fraction of 66%, 75%, 90%, and 100%). The selection of the four data points in the last run followed our suggested protocol using $\{T_{chi}, T_{cdi}\}$ variables only. We note that the internal and external CV and the model prediction bias (NMBE) are lower than those of the ARI scheme and only slightly poorer than the second run when 10 data points were used for model identification. This demonstrates that the GN model identified using our selection protocol of 4 data points results in an excellent model during the model identification stage and that the same can be used with

TABLE 2
Performance Summary of the GN Chiller Model Identified Using Different Data Selection Schemes (The CV and NMBE Values Apply to the Chiller COP)

Chiller	Data selection	Data Points	R^2	Internal CV	External CV	NMBE
Lab (GN chiller)	All data	30	0.945	0.99%	-	-
	Distributed uniformly over load variation	10	0.975	2.62%	1.18%	0.02%
	According to ARI 550/590	4	0.995	0.88%	1.37%	0.71%
	According to proposed selection procedure[1]	4	0.999	0.38%	1.24%	-0.59%
Field (Toronto chiller)	All data	810	0.974	1.50%	-	-
	Distributed uniformly over load variation	22	0.967	1.87%	1.60%	-0.15%
	According to ARI 550/590	4	0.989	4.99%	1.96%	-1.24%
	According to proposed selection procedure[1]	4	0.956	3.70%	2.16%	0.37%
	According to proposed selection procedure[2]	4	0.995	2.95%	2.01%	-1.31%

[1] Extreme points of $\{T_{chi}, T_{cdi}\}$.
[2] Extreme points of $\{T_{chi}, Q_{ch}\}$

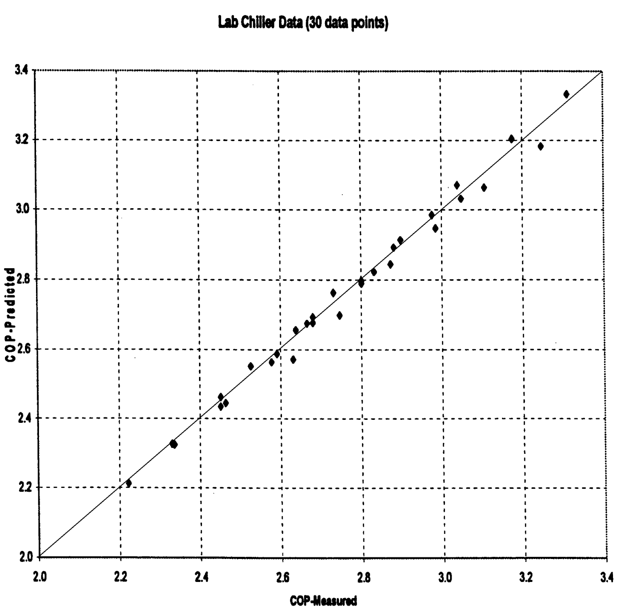

Figure 4 *Prediction accuracy of the GN chiller model identified from four data points following the selection procedure proposed in this study.*

great confidence to predict chiller performance over the entire range of chiller operation (see Figure 4).

Similar conclusions can be drawn from Table 2 for the field chiller. As noted earlier, load was varied down to 29% of full load while the variation in T_{cdi} was relatively small. Hence, we have performed two runs (namely, we chose two sets of four data points), one set based on extreme values of $\{T_{chi}, T_{cdi}\}$, and the other on extreme values of $\{T_{chi}, Q_{ch}\}$. The third run (labeled ARI 550/590) pertains to selecting our data uniformly over the range of load variation (load fractions of 29%, 61%, 75%, and 100%). We find that the external CV of the model based on the four ARI points is very slightly better (CV = 1.96%) than those from our suggested procedure (CV = 2.01% and 2.16%). However, internal CV and the NMBE are much poorer. Thus, whether the procedure of selecting four data points results in a better model than one based on ARI selection is inconclusive under field conditions. However, what is clear is that models using only four data points are only marginally poorer than those using much larger data points (provided of course, that the four data points are selected according to procedure and that the GN chiller model is used). Finally, we also note that the models identified from the four-point data set as in our selection schemes based on either $\{T_{chi}, T_{cdi}\}$ or $\{T_{chi}, Q_{ch}\}$ provide extremely accurate predictions and either scheme is equally good, though the latter is preferable, due to physical reasons noted earlier.

Note also from Table 2 that, in some cases, external CV is less than the internal CV values, which may seem implausible to data analysts. This is not a limitation of the model; rather, it is a statistical consequence due to the different values of (d.f.) used in both cases. For example, for the field chiller data, (d.f.) = 1 when four data points are used for model identification (i.e., for internal CV determination), while (d.f.) = 807 when determining external CV.

SUMMARY

The objective of this paper is to suggest a rational methodology by which chiller procurement engineers can improve the certified and witnessed factory testing phase. It is shown that using the Gordon and Ng (GN) chiller model allows the engineer to accurately predict the complete chiller performance map with *only four well-chosen operating points*. Once a model has been chosen, it is straightforward to use it to verify that the measured chiller performance is consistent (within experimental error) with the chiller manufacturer's published data. This paper justifies both the choice of the GN model and the experimental design proposed using monitored data from two actual chiller data sets (one from a laboratory chiller and the other from a field-operated chiller). The procedure of selecting the four data points proposed in this paper entails identifying which two of the three regressor variables $\{T_{chi}, T_{cdi}, Q_{ch}\}$ are likely to exhibit large variation in their operating range and perform the factory tests under their extreme operating conditions. While the proposed procedure was superior to the ARI suggestion of picking the four data points uniformly over the range of variation of Q_{ch} for the lab chiller, the advantage was less obvious for the field-operated chiller. The methodology suggested in this paper should be of great benefit to chiller professionals and enhance the quality assurance process of the entire chiller procurement process.

ACKNOWLEDGMENTS

The authors would like to thank Prof. Jeff Haberl of Texas A&M University for kindly supplying us with the Toronto chiller data used in this research and for his many comments on the draft of this paper. The authors are also grateful for comments and insights by Prof. Jeff Gordon during this research. Also, the assistance of Chanin Panjapompon for performing some of the analysis is appreciated. Finally, the authors are indebted to Steven Taylor and Mark Hydeman for useful comments as well as for generously sharing their slide presentation, which triggered this research.

NOMENCLATURE

COP	=	chiller coefficient of performance (dimensionless)
CV	=	coefficient of variation of the root mean square error (Equation A2)
I_1, I_2	=	measures of information defined by Equations 9 and 10
n	=	number of data points
$NMBE$	=	normalized mean bias error (Equation 4)
P	=	power input to the compressor (kWe)
p	=	number of model parameters
Q_{ch}	=	chiller (or evaporator) cooling capacity (kWt)
R^2	=	coefficient of determination
R_1 and R_2	=	normalized measures of information defined by Equation 11
T	=	temperature (K)
X	=	regressor variable or regressor matrix
Y	=	response variable

Subscripts

cdi	=	water inlet to condenser
chi	=	water inlet to chiller or evaporator

REFERENCES

ARI. 1998. *ARI Standard 550/590, 1998 Standard for Water-Chilling Packages Using the Vapor Compression Cycle.* Arlington, VA: American Refrigeration Institute.

ASHRAE. 2001. *2001 ASHRAE Handbook—Fundamentals*. Atlanta: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc.

Beck, J.V., and K.J. Arnold. 1977. *Parameter Estimation in Engineering and Science*. New York: John Wiley and Sons.

Bourdouxhe, J-P., M. Grodent, and J. Lebrun. 1999. *HVAC1 toolkit: Algorithms and subroutines for primary HVAC system energy calculations*. Atlanta: American Society of Heating Refrigerating and Air-Conditioning Engineers, Inc.

Box, G.E.P., W.G. Hunter, and J.S. Hunter. 1978. *Statistics for Experimenters*. New York: Wiley Interscience.

Chatterjee, S., and B. Price. 1981. *Regression Analysis by Example*. New York: John Wiley & Sons.

Gordon, J.M., and K.C. Ng. 2000. *Cool Thermodynamics*. Cambridge, UK: Cambridge International Science Publishing.

Hydeman, M., and K. Gillespie. 2002. Tools and techniques to calibrate electric chiller component models. *ASHRAE Transactions* 108(1).

Hydeman, M., N. Webb, P. Sreedharan, and S. Blanc. 2002. Development and testing of a reformulated regression based electric chiller model. *ASHRAE Transactions* 108(2).

Madsen, H., and J. Holst. 1997. Non-linear and non-stationary time series analysis, lecture notes. Denmark: Technical University of Denmark (available for download at <www.imm.dtu.dk/~ps>).

Phelan, J., M.J. Brandemuehl, and M. Krarti. 1997. In-situ performance testing of chillers for energy analysis. *ASHRAE Transactions* 103(1).

Reddy, T.A., and K.K. Andersen. 2002. An evaluation of classical steady-state off-line linear parameter estimation methods applied to chiller performance data. *HVAC&R Research Journal* 8(1): 101.

Reddy, T.A., K.K. Andersen, and D. Niebur. 2002. Information content of incoming data during field monitoring: Application to online chiller modeling. Submitted to *HVAC&R Research Journal*, January.

Taylor, S.T. 2001. Application of performance based bids for procurement of chilled water plant equipment. ASHRAE seminar presentation, Cincinnati, Ohio, June.

Winklemann, F.C., et al. 1993. DOE-2 Supplement, version 2.1E. Berkeley, Calif.: Lawrence Berkeley National Laboratories.

APPENDIX A

The **root mean square error** (RMSE) of a model identified from data is defined as follows:

$$RMSE = \left[\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n-p} \right]^{1/2} \tag{A1}$$

The RMSE is an absolute measure and its range is $0 \leq \mathrm{RMSE} \leq \infty$. A normalized measure is often more appropriate: the coefficient of variation of the RMSE (CVRMSE) or simply CV, defined as

$$CVRMSE = CV = RMSE/\bar{y}\,. \tag{A2}$$

Hence, a CVRMSE value of, say, 2% implies that the root mean value of the unexplained variation (i.e., the errors) is 2% of the mean value of the dependent variable.

The **mean bias error** (MBE) is defined as the mean difference between the actual data values and model predicted values:

$$MBE = \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)}{n-p} \tag{A3}$$

Note that when a model is identified by least squares regression, the model MBE using the original set of regressor variables is zero. Only when, say, the model identified from a first set of observations is used to predict under a second set of circumstances will MBE be greater than zero. Under such circumstances, the MBE is also called the mean simulation error. A normalized MBE (NMBE) can also be defined as

$$NMBE = MBE/\bar{y}\,. \tag{A4}$$