MODELING AND TEST POINT SELECTION FOR A THERMAL TRANSFER STANDARD

A. D. Koffman, National Institute of Standards and Technology[1]
Gaithersburg, MD

H. L. Stott, John Fluke Manufacturing Co., Inc.
Everett, WA

Abstract: Full calibration support for multirange instruments can be costly and time consuming. This paper presents a case study in which a new empirical-model-based approach was used to substantially reduce the number of tests required to fully characterize an instrument. The Fluke 792A Thermal Transfer Standard was the subject instrument for the study. Test results showed that measurements made at 50 test points were sufficient to allow accurate predictions of the instrument's performance at all 255 test points specified by the manufacturer. An accurate model relating ac/dc difference to voltage and frequency for the instrument was formulated using complete test data from many devices collected by the manufacturer over several production runs. An empirical test point selection procedure was used to select an optimal set of test points and subsequently to predict the ac/dc differences of other 792As based on the limited set of measurements taken at the selected test points.

INTRODUCTION

Multirange instruments[6] can be expensive and time-consuming to test. The cost of testing at production time as well as periodically throughout the life cycle of an instrument can often exceed the cost of the instrument. A primary cause for this expense is the large number of test points necessary for instruments that have many independent modes and ranges, and allow wide variation of input variables, (e.g., frequency, voltage, current, etc.). If manual tests are required, the cost is driven still higher. A savings of time and money can be realized if the instrument's overall performance could be accurately deduced from measurements made at a smaller number of test points. Such an approach has been developed over the last several years at the National Institute of Standards and Technology for testing analog and mixed-signal devices[1-5]. The work presented here is a case study in which this approach was taken. It involved modeling and test point selection for the Fluke 792A Thermal Transfer Standard.[2]

---

[1]U.S. Department of Commerce, Technology Administration.

[2]In order to describe adequately the procedures and tests discussed in this paper, commercial instruments are identified by manufacturer's name and/or model number. In no

Description of the Thermal Transfer Standard

The Fluke 792A is a multirange thermal transfer standard or ac/dc difference standard. It contains nine switchable ranges from 22 millivolts to 1000 volts full-scale. It may be used (calibrated) for frequencies from 10 Hz to 1 MHz over most of the ranges. It incorporates a solid-state rms sensor as the thermal converter and contains input amplifiers on the 22 millivolt through 700 millivolt ranges. All other ranges have passive circuitry prior to the sensing element. The 1000 volt range makes use of a separate range resistor used in conjunction with the 2.2 volt range. The instrument is powered from a battery supply to improve electrical isolation during use.

Although the instrument is a general purpose ac/dc difference standard, it was developed primarily for use as a standard in calibrating a multifunction calibrator. The points at which the thermal transfer standard is calibrated were selected with this purpose in mind. Table 1 contains the points at which a production model is calibrated and provides the manufacturer's uncertainties for the ac/dc differences of those points.

Calibration Requirements and Present Cost

As shown in Table 1, there are 255 points at which calibrations are being performed by the manufacturer on each production instrument. (Subsequent to the time the data presented in this paper was collected, the manufacturer has added 59 new points and tightened the uncertainty on several of the original points.) To support the calibration of these 255 points, the manufacturer submits a reference standard 792A to NIST for calibration at 149 points. Of these 149 points, 120 are those shown preceded by an asterisk in Table 1. The other 29 points are at frequencies and levels not shown in the table and not specified to the end user of the 792A. The manufacturer calibrates the remainder of the 255 points by using cross-range calibrations and other means. For example, the 2 V calibration on the 2.2 V range of the reference standard is used to calibrate the 2 V level of the 7 V range of a working standard. A working standard is a Fluke 792A that is calibrated against the reference standard and used to calibrate other instruments, including other Fluke 792As.

Calibration of 149 points at NIST is quite costly; a full calibration of the thermal transfer standard is very expensive. Not only is the actual calibration costly, but the cost of transferring the calibration to the working standards is substantial. In addition to the high cost in dollars, the time required for the calibration is several months. Any method that could significantly reduces the number of test points and the associated cost without adversely affecting the manufacturer's specifications is worthy of investigation. These savings can be passed on to the user, making both the user and the manufacturer more competitive in the marketplace.

---

case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology.

Table 1 Fluke 792A Test Points Specified by the Manufacturer with Uncertainty in ppm. (Asterisk identifies points at which reference standard is calibrated at NIST.)

Range Volts	Volts	0.01 kHz	0.02 kHz	0.04 kHz	0.1 kHz	1 kHz	10 kHz
0.022	0.002	*2200	*1600	*1600	*1600	*1600	*1600
	0.01	320	280	280	280	280	280
	0.02	*280	*180	*140	*140	*140	*140
0.22	0.02	350	230	190	190	190	190
	0.1	210	90	50	50	50	50
	0.2	*200	*80	*40	*40	*40	*40
0.7	0.2	220	80	50	50	50	50
	0.6	*200	*70	*30	*25	*25	*25
2.2	0.6	200	65	30	25	25	25
	1	190	60	30	*15	*15	*15
	2	*190	*60	*25	*10	*10	*10
7	2	190	65	30	*25	*25	*25
	6	*190	*60	*25	*10	*10	*10
22	6	190	65	30	25	25	25
	10	190	60	30	*20	*20	*20
	20	*190	*60	*25	*15	*15	*15
70	20	190	65	30	25	25	25
	60	*190	*60	*25	*20	*20	*20
220	60	190	65	40	40	40	40
	100	190	65	30	30	30	30
	200	*190	*60	*27	*27	*27	*27
1000	200	190	90	40	40	40	40
	1000			27	*27	*27	*27

Table 1(cont'd)

Range Volts	Volts	20 kHz	50 kHz	100 kHz	300 kHz	500 kHz	1000 kHz
0.022	0.002	*1600	*1600	*1900	*2500	*4500	*5000
	0.01	280	300	500	700	1200	1500
	0.02	*140	*200	*450	*650	*1100	*1300
0.22	0.02	190	220	450	650	1100	1900
	0.1	50	100	150	450	600	600
	0.2	*40	*100	*150	*450	*600	*600
0.7	0.2	50	100	150	450	600	600
	0.6	*25	*45	*60	*140	*430	*550
2.2	0.6	25	45	60	120	430	530
	1	*15	40	50	115	430	450
	2	*10	*40	*50	*115	*430	*450
7	2	*25	45	55	120	440	470
	6	*10	*40	*50	*120	*430	*460
22	6	25	45	55	120	440	480
	10	*20	40	50	120	430	480
	20	*15	*40	*50	*120	*430	*480
70	20	25	55	70	130		
	60	*20	*50	*65	130		
220	60	40	65	70			
	100	30	65	70			
	200	*27	*60	*70			
1000	200	40	65	70			
	1000	*27	*60	*70			

THE NIST APPROACH

The recently-developed NIST testing approach consists of establishing a linear coefficient matrix model describing the degrees of freedom of the manufacturing process. In matrix form, the modeling equation is given by

$$y = Ax \tag{1}$$

where y is an $m \times 1$ measurement response vector to be modeled, A is an $m \times n$ matrix, and x is an $n \times 1$ device parameter vector. Typically m is considerably larger than n. The rows of A correspond to different test conditions, i.e., test points.

Because $m > n$, their are more equations than unknowns, producing an overdetermined system. Assuming the model, A, is known, only n rows of y are actually needed to solve for the parameter vector, x. (In practice, it is useful to base the estimate of x on more than the minimum number of measurements to reduce the influence of random measurement noise, and to provide redundancy for detecting modeling error. For this purpose, a factor of 2-4 times n is usually considered sufficient.) Algebraic operations are performed on the model to select the set of rows, or test points, that gives the highest confidence for the estimates, $\hat{x}$. The algorithm developed for selecting the test points is described in reference [1]. The resulting matrix test equation typically contains many fewer test points than are required for exhaustive testing of an instrument:

$$y' = A'x \tag{2}$$

where y' and A' contain only the selected rows of **(1)**, corresponding to the selected test points. By making measurements, y', at the reduced set of test points, the estimated device parameters, $\hat{x}$, are then determined using the least-squares equation

$$\hat{x} = (A'^{T}A')^{-1}A'^{T}y' \tag{3}$$

Once estimates for the device parameters are obtained, the predicted response, $\hat{y}$, can be determined at all other candidate test points by substitution back into **(1)**:

$$\hat{y} = A\hat{x} \tag{4}$$

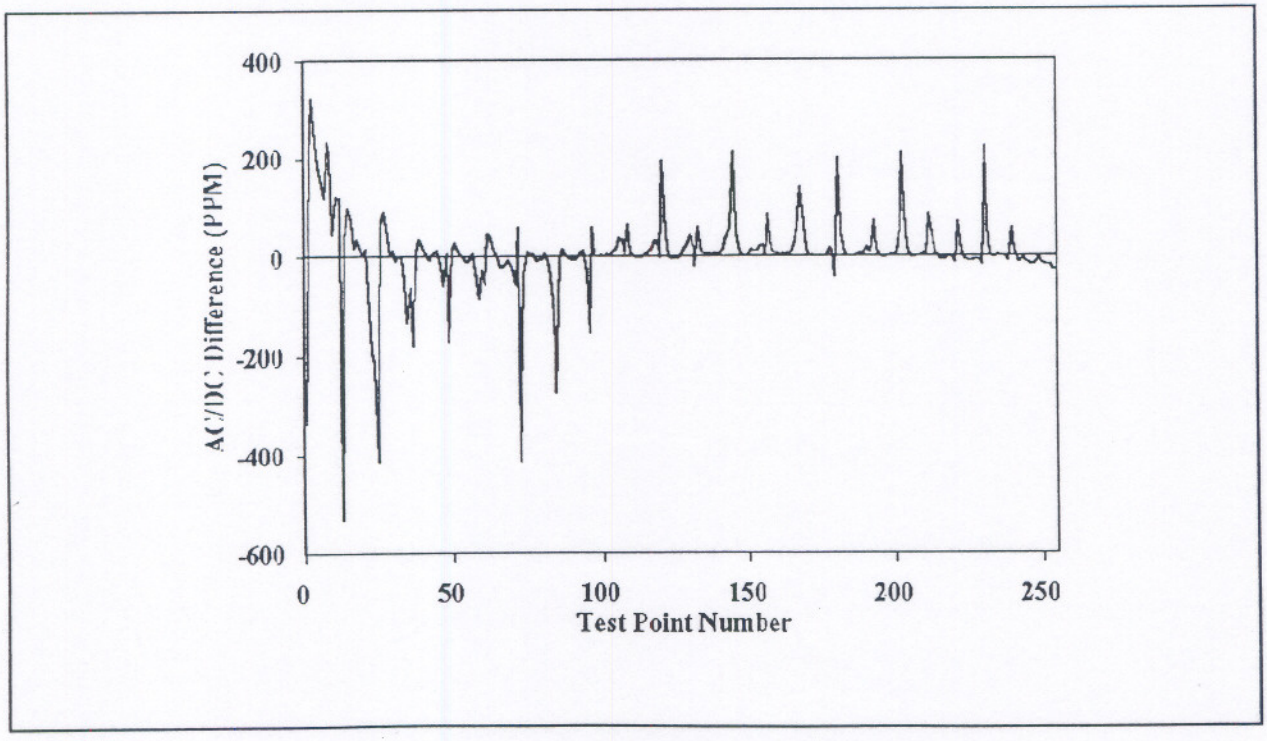
Comparison of the response prediction and the measured responses allows the assessment of the validity of the model. If behavior exists that cannot be accounted for by the model, a flag is raised on-line and exhaustive test point measurements should be taken to check model and instrument accuracy. Thus, the response prediction constantly monitors the manufacturing process so that the model can be updated if necessary. This modeling

approach also allows the estimation of the prediction errors at the unmeasured test points.

The success of this approach depends on the development of an accurate model, A. While several modeling approaches are possible, the most useful and robust approach for production testing is empirical modeling[3]. With this approach, the model is developed from exhaustive (i.e., all test points) measurement data collected on a large sample of k instruments of the type being tested. The k vectors of measurement data, y^1, y^2, ..., y^k, are collected as k columns of a matrix and analyzed to extract the linearly independent information that is expressed, using the singular value decomposition[5]. The final model matrix that is formed is able to express all of the degrees of freedom of the manufacturing process.

Building the Model

The construction and testing of empirical models was performed using 139 complete measurement data sets taken on the thermal transfer standard. Figure 1 shows a plot of a typical data set(vector). The data that were used for the model in this study were taken from final test data collected at the manufacturer's facility. The final test station is a fully automated station that compares the ac/dc difference of the product at each test point with the ac/dc difference of a production standard. The production standard is characterized in the manufacturer's primary standards laboratory by comparing it to two working standards. (In this paper the term characterized is used to identify the process of measuring a parameter and assigning a number to the resulting value. The 255 measured points each have an ac/dc difference value assigned.) The working standards, in turn, are characterized by comparing them to the reference standard characterized at NIST. In each step down the calibration chain, the ac/dc difference of the instrument under test is compared to the ac/dc difference of



the standard. The known difference of the standard is then used to calculate the difference for the instrument under test.

Many measurements were performed both at NIST and in the manufacturer's primary standards laboratory. Statistical analysis was performed on the results of the measurements to arrive at the assigned values for each of the 255 test points for the reference standard, working standards, and production standards.

The data for this case study were 139 sets, taken from several production runs and over a period of several months. It is critical that the model resulting from this work be an accurate model for those instruments which did not contribute data. The implication is that for the model to apply to future instruments (and it must for the method to be useful), there must be no design or parts changes that introduce new modes of response which cannot be described accurately by the model. If significant changes are made, the model should be rechecked and updated if needed.

This large data set was divided randomly into a set of 100 instruments used to build models and a validation set of 39 instruments used to test the models. The data at each test point was first normalized to the uncertainty specified by the manufacturer for that test condition, so that the model performance could be evaluated against specifications. The uncertainty specified for the 792A is shown in Figure 2, and in Figure 3 the data vector from Figure 1 is shown normalized to the uncertainty of Figure 2.

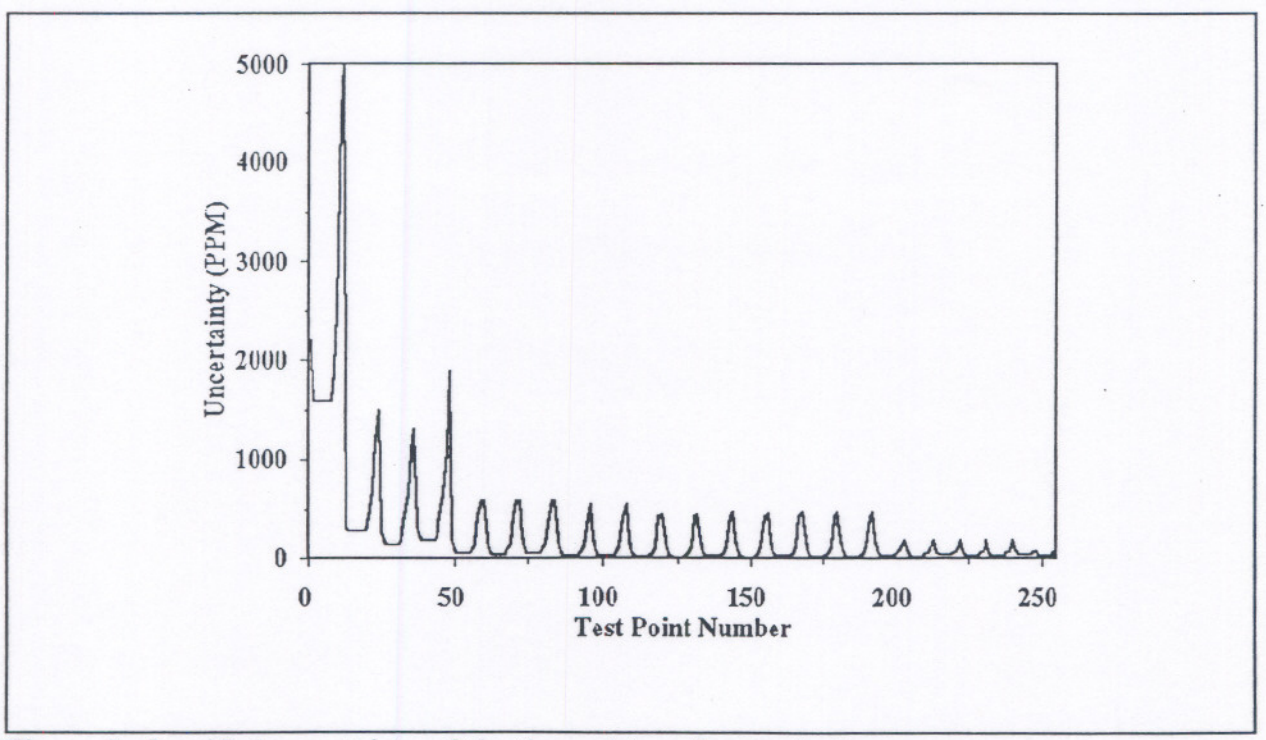


Figure 2 Specified uncertainty of the thermal transfer standard.

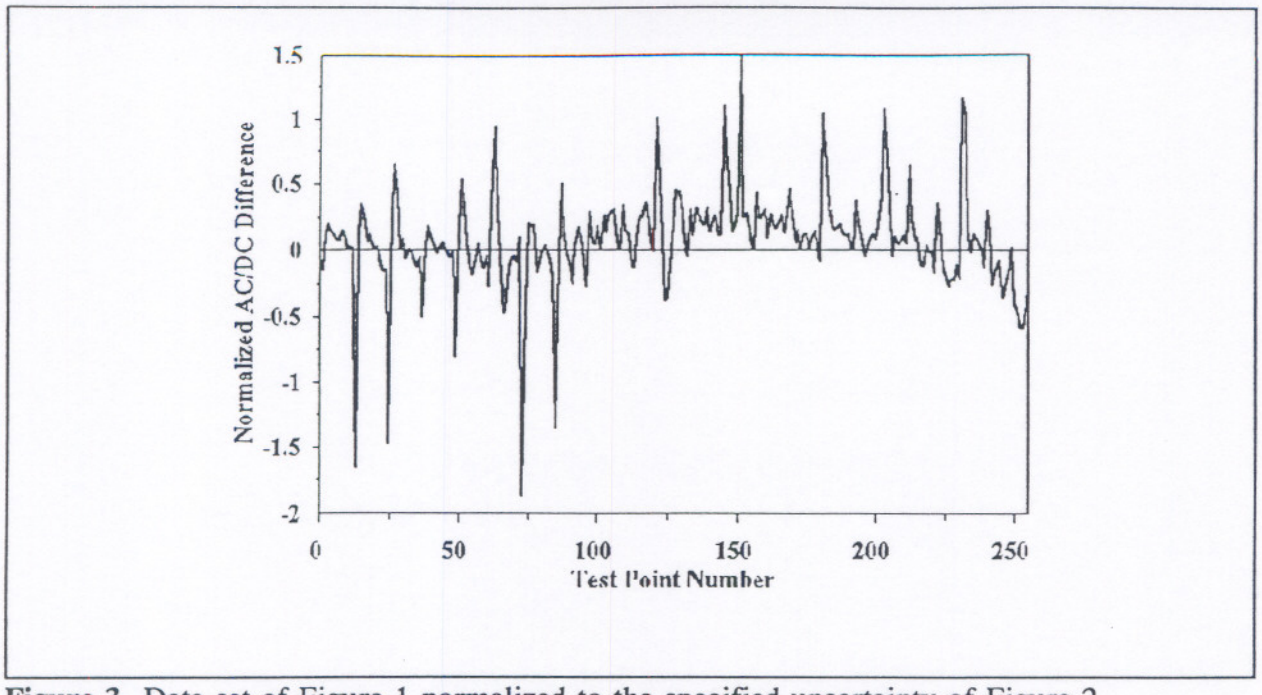


Figure 3 Data set of Figure 1 normalized to the specified uncertainty of Figure 2.

The formulation of the model began with the modeling set of data vectors, in this case, 100 vectors of length 255. Thus, the model matrix A was initially of size 255×100. To get an efficient model, the model's dimension must be minimized by eliminating redundant information, or information whose effects are negligible. Two matrix decomposition methods are suited to this purpose, the singular value decomposition(SVD) and the QR decomposition(QRD)[7].

The method found to be the most effective in performing the column reduction of the system model was the SVD. The SVD factors a matrix into a product of three new matrices

$$A = USV^T \tag{1}$$

where U and V are orthonormal matrices and S is a diagonal matrix whose elements are the singular values associated with the A matrix[7,8]. The singular values, and correspondingly the column vectors of U, are arranged in descending order of importance in describing the vectors in A. Because of random noise in the measurement data comprising A, the singular values never reach a value of zero; nevertheless, a relatively small number of vectors of U are generally sufficient to explain the significant structure of any vector of measurement data, y. Therefore, a reduced version of the U matrix can be an efficient model of the device characteristics.

A model constructed in this manner can be tested by fitting it to the data of the validation set, using **(3)** and **(4)** above. The predictions obtained with **(4)** are compared to the actual measurement data for each instrument in the validation set. In this case, the model is used

with full row dimension (before test point selection) to estimate the parameters. After testing differently sized models in this way, 20 columns of U were selected for the final model. The resulting model is of size 255×20 and the system equation is represented as

$$y = \hat{U}x \tag{2}$$

where $\hat{U}$ represents the column-reduced U matrix.

Once the model was selected, the test points were selected as described earlier. Fifty test points were selected in all, giving a final, reduced model, A', of size 50×20.

RESULTS

Many model sizes were tested against the validation set prior to the selection of the 50×20 model size. In each case, the parameter values for each instrument of the validation set were estimated using **(3)**, based on the data at the selected test points. Then, the entire response at all test points was predicted using **(4)** for each instrument. The errors, ε, of the predictions for each of the instruments were then calculated from the measured responses, y_m, at all test points using

$$\varepsilon = y_m - \hat{U}\hat{x} \tag{7}$$

Figure 4 illustrates the tradeoffs between the model dimensions (number of selected test points and number of column vectors or parameters) and the prediction accuracy determined

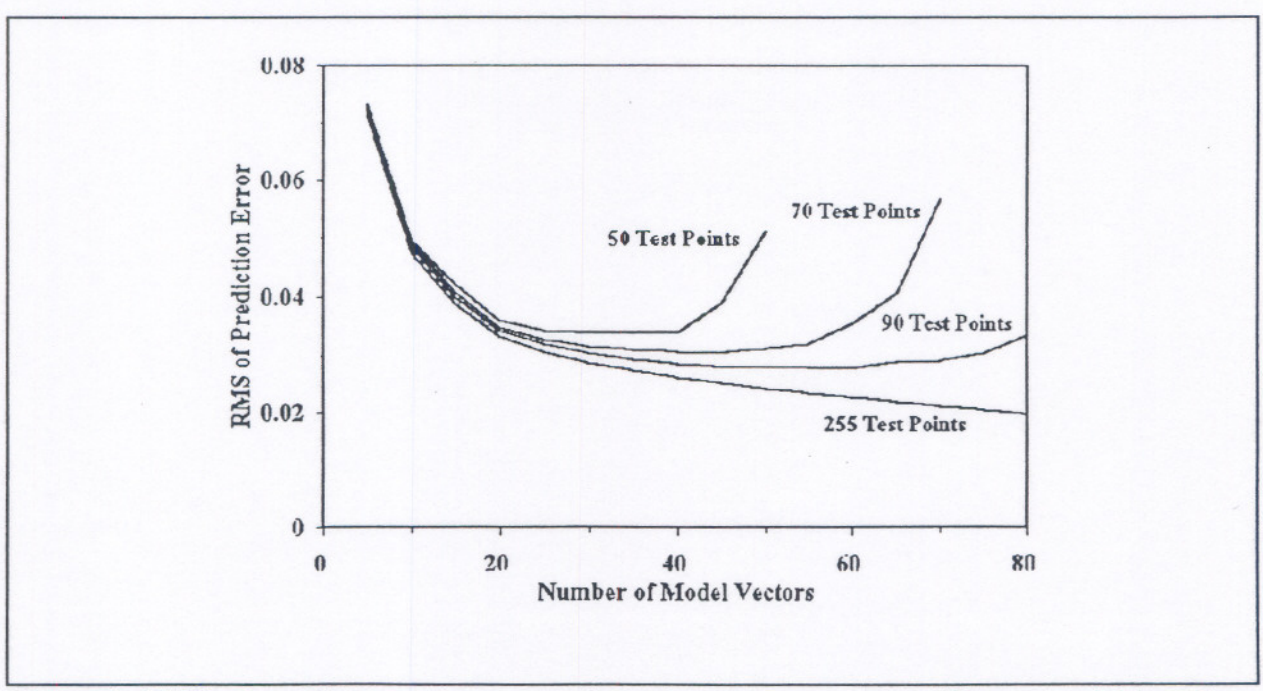


Figure 4 RMS of prediction error vs. number of model vectors for different model lengths.

by (7). Here, the plotted errors are the rms values of all 255 elements of all 39 vectors, ε, computed for each model size. This plot represents the accuracy of the various models measured against the entire validation set of devices. The four curves give the results for four different numbers of selected test points. Along each curve the matrix model size is increasing in column size and maintaining the same row size (constant number of test points). The prediction errors on the vertical axis are normalized to units of the manufacturer's specified uncertainties for the instrument.

These curves show that the 50×20 model gives an overall, rms error of about 3.6 percent of the manufacturer's specified uncertainties. Furthermore, they show that many more model vectors and test points are required to gain a substantial improvement over this performance.

While Figure 4 gives the overall performance of models of various sizes, Figures 5 and 6 give a more detailed picture of the performance obtained with the selected 50×20 model. In Figure 5, each of the 39 vectors of prediction error, ε, calculated from (7), are overlaid, giving a total of 39×255 or 9945 data points plotted. A histogram of the prediction errors is graphed in Figure 6. Comparison of the prediction errors for the 39 validation device data sets normalized to the uncertainty shows that more than 98 percent of the errors are less than 10 percent of the uncertainty values. All but two of the 9945 errors were less than 20 percent. The standard deviation of the errors was 3.6 percent and the mean was 0.14 percent of the uncertainty.

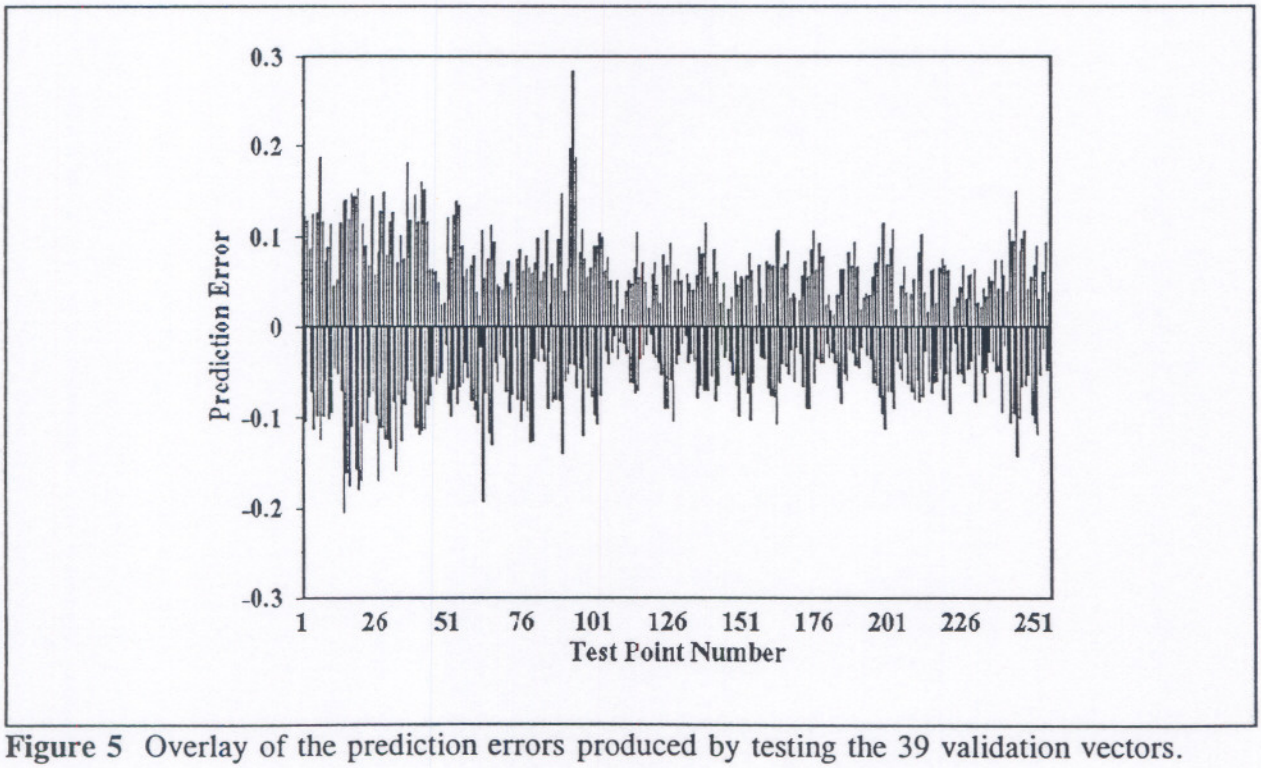


Figure 5 Overlay of the prediction errors produced by testing the 39 validation vectors.

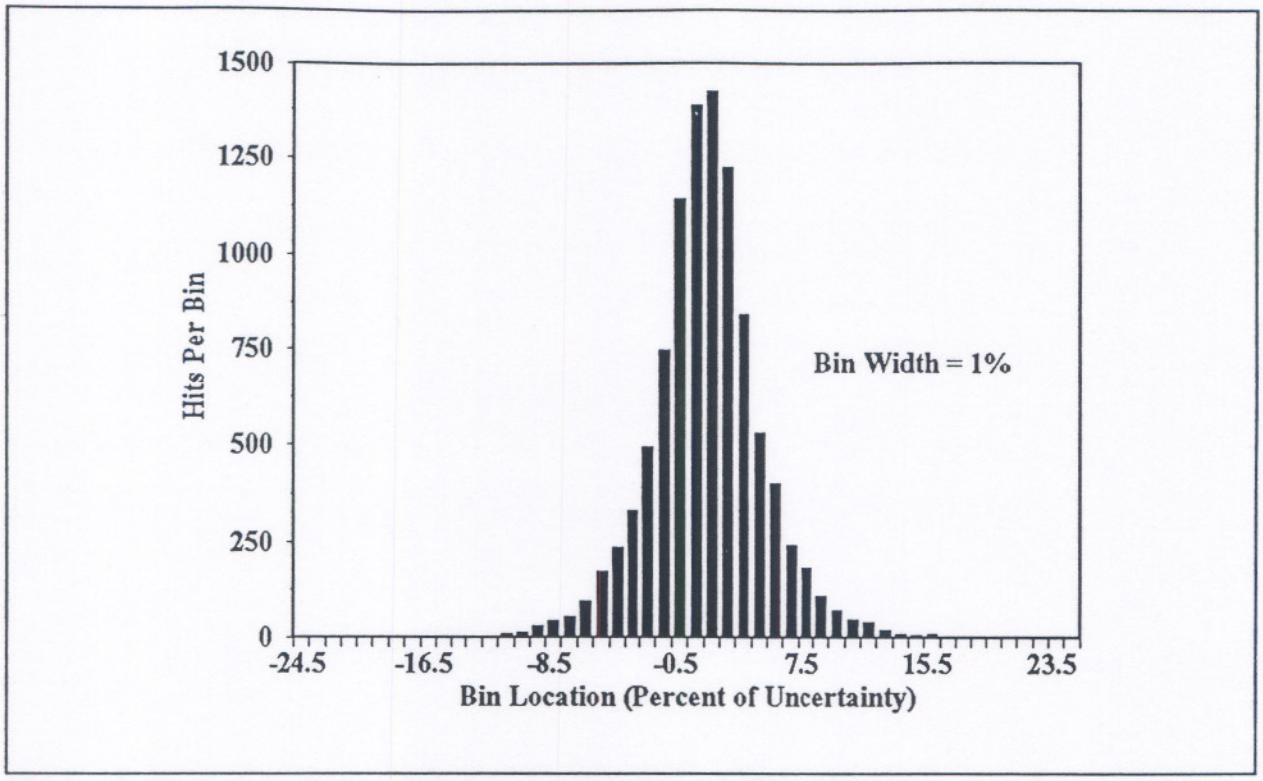


Figure 6 Histogram of the prediction errors plotted in Figure 5.

CONCLUSIONS

Relative to testing the complete set of test points, the modeling technique described in this paper has been shown to significantly reduce test time and test cost for calibrating a multirange thermal transfer standard. In general, this new testing strategy can likely be applied to other types of multirange instruments. Since the development of the test model would be done once (with periodic checks and updates), the greatest cost of testing would likely be the taking of the device measurements. Therefore, the SVD model is very promising. In this case only 50 of 255 test point measurements provide accuracy well within the uncertainty specified by the manufacturer.

In today's world of smart instruments, the model of the instrument response based on a limited set of test points could be included in a software program built into many types of future instruments. The methodology can be applied to both sources and measuring devices. The user would then need only to calibrate at the limited set of test points and let the instrument software predict the response at the remaining points. There would, of course, have to be checks built into the process that would ensure that the model still fit the instrument's performance. The obvious benefit is reduced calibration time with attendant reduced cost and delivery time.

ACKNOWLEDGEMENTS

The authors wish to thank T.M. Souders and G. N. Stenbakken of the National Institute of Standards and Technology for many valuable discussions concerning the theory and techniques involved in this work. Thanks go also to R. Henderson and D. Deaver of John Fluke Mfg. Co., Inc. for providing the data for analysis.

REFERENCES

[1] Stenbakken, G. N. and Souders, T. M., "Test Point Selection and Testability Measures Via QR Factorization of Linear Models," *IEEE Trans. I&M*, Vol. IM-36, No. 2, June 1987, pp. 406-410.

[2] Souders, T. M. and Stenbakken, G. N., "A Comprehensive Approach for Modeling and Testing Analog and Mixed-Signal Devices," *1990 Proc. ITC*, Washington, DC, IEEE Computer Society Press, September 1990, pp. 169-176.

[3] Stenbakken, G. N. and Souders, T. M., "Linear Error Modeling of Analog and Mixed Signal Devices," *1991 Proc. ITC*, Nashville, TN, IEEE Computer Society Press, September 1991, pp. 573-581.

[4] Souders, T. M. and Stenbakken, G. N., "Cutting the High Cost of Testing," *IEEE Spectrum*, March 1991, pp. 48-51.

[5] Stenbakken, G. N. and Souders, T. M., "Developing Linear Error Models for Analog Devices," *IMTC 1993 Conference Record*, Irvine, CA, May 1993.

[6] Koeman, H., "Multifunction, Multirange Electronic Instrument Performance Prediction," *IEEE Trans. I&M*, Vol. 41, No. 6, December 1992, pp. 985-990.

[7] Leon, S. J., Linear Algebra with Applications, Macmillan Publishing Co., New York, 1980, pp. 283-284, 308-310.

[8] Schilling, R. J. and Lee, H. L., ENGINEERING ANALYSIS A Vector Space Approach, John Wiley & Sons, New York, 1988, pp. 311-312.