Curve Fitting of Sensors' Characteristics

Boyanka Marinova Nikolova, Georgi Todorov Nikolov and Milen Hrabarov Todorov

Abstract - **More complex technical systems and higher levels of integration of electronic circuits lead to new requirements on the data treatment of modern signal processing systems, especially for sensing systems. This paper presents curve fitting methodology for analogue sensor's characteristics described by datasheet lookup tables. This is achieved by MATLAB's Curve Fitting Toolbox, which provides a library of linear, nonlinear, and nonparametric fitting models. A method for implement achieved polynomial equations as calibration template in transducer electronic data sheets is suggested in the end of presentation.**

Keywords – **Curve fitting, Goodness of fit statistics, MATLAB, Intelligent sensors, TEDS**

I. Introduction

At the core of any data acquisition system is interpretation of a voltage signal based on information about the analog sensor that makes the measurement intelligible. Typically, these are standard curves and equations specific to the type of transducer. Sensor calibration, however, takes this process one step farther by considering transducers on an individual basis. The sensor output voltage is mapped to a physical measurement based on metrics obtained from a specific sensor calibration. Although many sensors are linear over the limited range, these sensors exhibit a slight but progressively more nonlinear characteristic as the measurement range widens. Consequently, over an extended span, curve fitting is necessary if the system is to achieve a high level of precision.

With the facility of computation now available through digital computers and microprocessors, the problem of estimation of transducer's transfer characteristics is being increasingly tackled using software techniques. However, for inherent nonlinear sensors, a software solution depends upon the proper approach through mathematical modeling of the response curve [5, 9].

The purpose of this paper is to assist engineers and scientists to implement the newly released Curve Fitting Toolbox in order to achieve more precise results. The presented material facilitates users to create calibration equations from basic calibration data and use these equations to make accurate measurements.

II. Parametric Curve Fitting

Parametric fitting involves finding coefficients (parameters) for one or more models that fit to data [1, 2]. The model is a function of the independent variable and one or more coefficients. The appropriate mathematical model of sensors characteristics is can be obtained by Curve Fitting Toolbox™ software. This toolbox is a collection of graphical user interfaces and M-file functions for curve and surface fitting that operate in the MATLAB® technical computing environment. The toolbox includes a library of various parametric models, summarized in table 1. [3, 4]

TABLE 1. CURVE FITTING TOOLBOX LIBRARY MODELS

Model	Definition
Exponentials	$y = ae^{bx}$ and $y = ae^{bx} + ce^{dx}$, where *a*, *b*, *c* and *d* are model's parameters
Fourier Series	$y = a_0 + \sum_{i=1}^{n} a_i \cos(nwx) + b_i \sin(nwx)$, where a_0 models a intercept term in the data and is associated with the $i = 0$ cosine term, ω is the fundamental frequency of the signal, *n* is the number of terms (harmonics) in the series, and $1 \le n \le 8$.
Gaussian	$y = \sum_{i=1}^{n} a_i e^{\left[-\left(\frac{x-b_i}{c_i}\right)^2\right]}$, where a_i is the amplitude, b_i is the centroid (location), c_i is related to the peak width, *n* is the number of peaks to fit, and $1 \le n \le 8$
Polynomials	$y = \sum_{i=1}^{n+1} p_i x^{n+1-i}$, where p_i are model's parameters, $n + 1$ is the order of the polynomial, *n* is the degree of the polynomial, and $1 \le n \le 8$
Power Series	$y = ax^b$ and $y = a + bx^c$, where *a*, *b* and *c* are model's parameters
Rationals	$y = \frac{\sum_{i=1}^{n+1} p_i x^{n+1-i}}{x^m + \sum_{i=1}^{m} q_i x^{m-i}}$, where q_i and p_i are model's parameters, *n* and *m* are the degree of the numerator and denominator polynomials – $1 \le n \le 5$ and $1 \le m \le 5$
Sum of Sines	$y = \sum_{i=1}^{n} a_i \sin(b_i x + c_i)$, where a_i is the amplitude, b_i is the frequency, and c_i is the phase constant for each sine wave term, *n* is the number of terms in the series and $1 \le n \le 8$
Weibull Distribution	$y = abx^{b-1}e^{-ax^b}$, where *a* is the scale parameter and *b* is the shape parameter

Curve Fitting Toolbox software uses the method of least squares when fitting data. The supported types of least squares fitting include: linear, weighted linear, constrained, robust, and nonlinear [1, 2, 3, 4].

Boyanka Nikolova is with the Faculty of Telecommunications at Technical University of Sofia, Bulgaria. E-mail: bnikol@tu-sofia.bg.
Georgi Nikolov is with the Faculty of Electronic Engineering and Technologies at Technical University of Sofia, 8 Kl. Ohridski Blvd, Sofia 1000, Bulgaria, E-mail: gnikolov@tu-sofia.bg.
Milen Todorov is a student from the Faculty of Telecommunications at Technical University of Sofia, Bulgaria.

III. CURVE FITTING METHODOLOGY

The proposed methodology for curve fitting of sensors' characteristics is illustrated with the block diagram given on fig. 1. A particular application might dictate still other aspects of model fitting that are important to achieving a good fit, such as a simple model that is easy to interpret.

On the first step *Importing Data* the chosen characterristic must be represented as the predictor (*X*) data, response (*Y*) data, and weights. If the weights are not imported, then they are assumed to be 1 for all data points. The data may be taking either from the datasheets of sensor's manufacturer, or by measurements of the particular sensor.

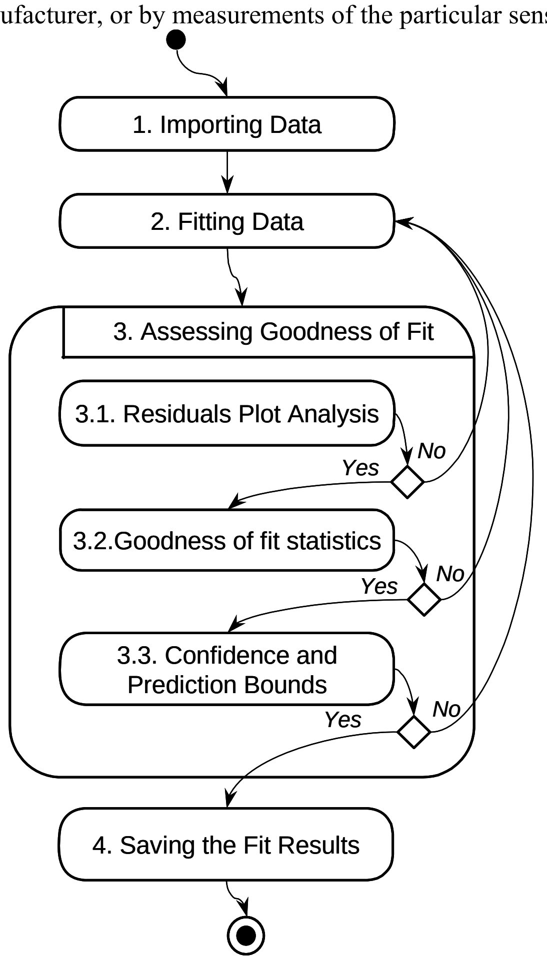

Fig. 1. The block diagram of the curve fitting methodology.

On the next step *Fitting Data* one model from the library of parametric models listed in table 1 is selected. According to the shape and specificity of the sensor characteristic the appropriate selection must be done. The fitting procedure can be successful in short time if the model is suitable and correspond to the real characteristic.

After fitting data with one or more models, the next important step *Assessing Goodness of Fit* must be implemented. The toolbox provides these methods to assess goodness of fit for both linear and nonlinear parametric fits [1, 2, 3, 4]: residual analysis, goodness of fit statistics and confidence and prediction bounds.

These methods can be divided into two types: graphical and numerical. Plotting residuals and prediction bounds are graphical methods that aid visual interpretation, while computing goodness of fit statistics and coefficient confidence bounds yield numerical measures, that aid statistical reasoning. Graphical measures allow viewing the entire data set at once, and they can easily display a wide range of relationships between the model and the data. The numerical measures are more narrowly focused on a particular aspect of the data and often try to compress that information into a single number. In practice, depending on sensor characteristic, number of data points and analysis requirements, often must be used both types to determine the best fit.

Residual Plot Analysis. The residual for the i^{th} data point r_i is defined as the difference between the observed response value y_i and the fitted response value $\hat{y}_i$, and is identified as the error associated with the data:

$$r_i = y_i - \hat{y}_i . \quad (1)$$

Assuming the fitting model is correct, the residuals approximate the random errors. Therefore, if the residuals appear to behave randomly, it suggests that the model fits the data well. However, if the residuals display a systematic pattern, it is a clear sign that the model fits the data poorly. It must be notice that many results of model fitting, such as confidence bounds, will be invalid should the model be grossly inappropriate for the data.

Goodness of fit statistics. After using graphical methods to evaluate the goodness of fit, it must be examine the goodness of fit statistics. For parametric models Curve Fitting Toolbox supports following types of statistics [3, 4]: the sum of squares due to error (*SSE*); coefficient of determination (*R*-square); adjusted *R*-square and root mean squared error (*RMSE*).

The sum of squares due to error measures the total deviation of the response values from the fit to the response values. It is also called the summed square of residuals and is given whit equation:

$$SSE = \sum_{i=1}^{n} \omega_i (y_i - \hat{y}_i)^2 \quad (2)$$

where ω_i are the weights. The weights determine how much each response value influences the final parameter estimates. If a value of *SSE* is closer to 0 indicates that the model has a smaller random error component, and that the fit will be more useful for prediction.

Coefficient of determination (*R*-square) determining how successful the fit is in explaining the variation of the data. *R*-square is the square of the correlation between the response values and the predicted response values:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \omega_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} \omega_i (y_i - \bar{y})^2} \quad (3)$$

where $\bar{y}$ is mean value of y_i. *R*-square with a value closer to 1 indicating that the model accounts for a greater proportion of variance.

If the number of fitted coefficients increases in the model, R^2 will increase although the fit may not improve in a practical sense. To avoid this situation, should be used the degrees of freedom *adjusted* R^2 statistic. This statistic is based on the residual degrees of freedom v defined as the number of response values n minus the number of fitted coefficients m estimated from the response values:

$$v = n - m \,, \qquad (4)$$

where v indicates the number of independent pieces of information involving the n data points that are required to calculate the sum of squares. Note that if parameters are bounded and one or more of the estimates are at their bounds, then those estimates are regarded as fixed. The number of such parameters increases the degrees of freedom. The *adjusted* R^2 statistic is given:

$$\text{Adjusted } R^2 = 1 - \frac{\sum_{i=1}^{n-1} \omega_i (y_i - \hat{y}_i)^2}{\sum_{i=1}^{v-1} \omega_i (y_i - \bar{y})^2} \,. \qquad (5)$$

The adjusted R-square statistic has a value closer to 1 indicating a better fit. Negative values can occur when the model contains terms that do not help to predict the response.

Root mean squared error (*RMSE*) is also known as the fit standard error and the standard error of the regression. It is an estimate of the standard deviation of the random component in the data, and is defined as

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} \omega_i (y_i - \hat{y}_i)^2}{v}} \,. \qquad (6)$$

Just as *RMSE* value closer to 0 indicates a fit that is more useful for prediction.

Confidence and prediction bounds define the lower and upper values of the associated interval, and define the width of the interval. The width of the interval indicates how uncertain have about the fitted coefficients, the predicted observation, or the predicted fit. A very wide interval for the fitted coefficients indicate that should be used more data when fitting.

The saved information on the last step *Saving the Fit Results*, can be used for documentation purposes, or to extend data exploration and analysis.

IV. APPLICATION OF SENSORS CURVE FITTING

IEEE 1451.4 is a standard that defines a relatively simple, straightforward mechanism for adding intelligent, plug and play capabilities to traditional analog sensors. Without adding any new hardware to the system, these plug and play sensors can bring real, immediate benefits in ease of use and productivity to any measurement and automation system that uses sensors [6, 7]. The underlying mechanism for plug and play identification is the standardization of a Transducer Electronic Data Sheet (TEDS). The information stored in TEDS enables the system to identify, characterize, interface, and properly use the signal from the analog sensor. In addition to such configuration information TEDS can store calibration data specific to an individual sensor. Sensor calibration considers transducers on an individual basis. In other words, the sensor output voltage is mapped to a physical measurement based on metrics obtained from a specific sensor calibration.

The data stored in the sensor's EEPROM or a virtual TEDS file is compressed to save space [7]. There are sixteen separate standard templates for each type of sensor. In addition to these standard templates, there are three calibration templates: a calibration table, frequency response table and polynomial calibration curve.

The calibration table template describes a lookup table with the electrical-to-physical transfer function of a transducer. The calibration table template provides a simple means of recording a few data points for calibration purposes. This is particularly useful when calibrating around a very narrow range of values.

The frequency response table template specifies the frequency response transfer function of a transducer with a lookup table of frequency-amplitude pairs.

The aim of presented investigation is to precisely characterize a transducer over its entire range. Most appropriate for such initiative is polynomial calibration curve. This may prove difficult, especially in the case of a transducer with a highly nonlinear transfer function, such as a thermocouple or thermistor. To make effective use of the calibration curve template, several calibration points must be taken and then fitted to a polynomial curve. This may take significantly more time than collecting a few data points for a calibration table. However, the tradeoff is greater accuracy and a wider range of calibrated values.

Using the proposed approach the transfer characteristic of thermistor PR222J2 is fitted with appropriate parametric model. For input data of temperature dependence is used the detailed look up table given in datasheet of thermistor's manufacturer U.S. SENSOR Inc. [8].

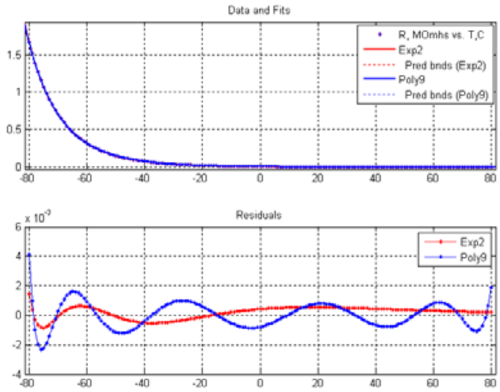

Fig. 2. The data, fit, prediction bounds and residuals for transfer characteristic of thermistor PR222J2.

The data, fit, prediction bounds and residuals for fitting of full transfer characteristic are shown in figure 2. In table 2 are given the fitting numerical results for models' parameters and goodness of fit statistics. For a first fitting model four parametric exponentials (Exp2) is chosen, because it is well know that the transfer functions of thermistors are exponential. On the other hand the TEDS standard support only polynomial functions, therefore like a second model is proposed 9th degree polynomial (Poly9). It can be seen that the residuals for the Exp2 model appear randomly scattered around zero and indicating that this model describes the data well. The same conclusion can be done comparing numerical values of *SSE* and *RMSE* (see Table 2).

TABLE 2. RESULTS FROM CURVE FITTING OF THERMISTOR'S TRANSFER CHARACTERISTIC

		Full Characteristic				Divided Characteristic		
						T = [−80 ÷−68] °C	T = [−67 ÷ 4] °C	T = [5 ÷ 80] °C
Model		Exponential (Exp2)		9th Degree Polynomial (Poly9)				
Equation		$f(x) = ae^{bx} + ce^{dx}$		$f(x) = p_1x^9 + p_2x^8 + p_3x^7 + p_4x^6 + p_5x^5 + p_6x^4 + p_7x^3 + p_8x^2 + p_9x + p_{10}$				
Parameters		a	0.0002613	p_1	$-5.912.10^{-18}$	$1.324.10^{-10}$	$-4.737.10^{-16}$	$4.922.10^{-19}$
		b	−0.1048	p_2	$5.691.10^{-16}$	$8.816.10^{-8}$	$-9.46.10^{-14}$	$-1.736.10^{-16}$
		c	0.006674	p_3	$2.925.10^{-14}$	$2.607.10^{-5}$	$-8.671.10^{-12}$	$2.18.10^{-14}$
		d	−0.0545	p_4	$-2.463.10^{-12}$	0.004496	$-4.071.10^{-10}$	$-5.428.10^{-13}$
				p_5	$-1.741.10^{-10}$	0.4982	$-1.188.10^{-8}$	$-1.7.10^{-10}$
				P_6	$1.181.10^{-8}$	36.79	$-1.266.10^{-7}$	$2.553.10^{-8}$
				p_7	$-9.112.10^{-8}$	1810	$-3.317.10^{-6}$	$-1.987.10^{-6}$
				p_8	$4.382.10^{-6}$	$5.725.10^{4}$	0.0001078	0.0001042
				p_9	−0.0004184	$1.056.10^{6}$	−0.003739	−0.003732
				p_{10}	0.008147	$8.648.10^{6}$	0.07354	0.07347
Goodness of fit	*SSE*	$2.936.10^{-5}$		$1.209.10^{-4}$		$6.211.10^{-11}$	$3.92.10^{-8}$	$3.519.10^{-10}$
	R^2	1		1		1	1	1
	Adj. R^2	1		1		1	1	1
	RMSE	$4.325.10^{-4}$		$0.8947.10^{-3}$		$4.55.10^{-6}$	$2.514.10^{-5}$	$2.309.10^{-6}$

In TEDS template the overall function is described in a piecewise manner, each segment defined by an array of coefficients and powers. For example, a polynomial curve from table 2 may be described by [(9, −5.91E−18), (8, 5.69E−16), (7, 2.92E−14), (6, −2.46E−12), (5, −1.74E−10), (4, 1.18E−8), (3, −9.11E−8), (2, 4.38E−6), (1, −4.18E−4), (0, 8.14E−3)]. [(3,1),(2,5),(1,−2),(0,1)], where the numbers in parentheses are interpreted as (Curve Power, Curve Coefficient).

In order to achieve better accuracy, especially for more non-linear sensors, TEDS provides an alternative means of calibration and linearization. The calibration curve template specifies a multi-segment polynomial curve and each segment is bounded by sequential values of, closed below and open above [6, 7]. The implementation of this capability in the case of treated thermistor is shown in last three columns in Table 2. As can be seen from the numerical fit results that the *SSE* and *RMSE* statistics have values very close to zero. Therefore the parametric models of divided transfer characteristic of the thermistor much better fits the input data.

V. CONCLUSION

The present paper suggests an systematic approach for implementing curve fitting models and methods in order to achieve equation that precisely describe sensor's transfer function. The aim of such equation is to implement it in calibration TEDS template and in such a way to improve measurement accuracy. To proof usability of presented approach a number of equation for thermocouples, RTDs and thermistors have been achieved. As illustration experimental results from curve fitting of thermistor's transfer characteristic is presented. The obtained parametric models are suitable for calibration and transfer functions in measurement system of non-electric quantities or for behavior modeling of various sensors.

This investigation has been carried out in the framework of the research projects № 091 НИ 109-07 and № Д002-126/2008.

REFERENCES

[1] Riley, F., Hobson, P., Bence, J. *Mathematical Methods for Physics and Engineering*, Cambridge University Press, New York, 2006.

[2] Karris, Steven T., *Mathematics for Business, Science, and Technology with MATLAB® and Excel® Computations*, Orchard Publications, California, U.S.A., 2007.

[3] Weisberg, S. et., *Applied Linear Regression*, John Wiley & Sons, Inc., New Jersey, U.S.A., 2005.

[4] *Curve Fitting Toolbox 2 User's Guide*. The MathWorks, Inc., 2001-2009.

[5] Dumčius, A., T. Žizniakovas, *Approximation T-U Characteristics of Analog Semiconductors Temperature Sensors*, Electronics and Electrical Engineering, No. 1, 2009.

[6] National Instruments, *Sensor Calibration with TEDS Technology*, Tutorial, Sep 6, 2006.

[7] National Instruments, *Upgrading Your System for Virtual TEDS*, Tutorial, Jul 12, 2007.

[8] U.S. SENSOR CORP., *Ultra Precision Interchangeable Thermistors*. www.ussensor.com/rt charts/PR222J2.htm.

[9] Nikolov, G., B. Nikolova, *Design, Development and Calibration of Virtual System for Relative Humidity Measurement*, Sensors & Transducers Journal, ISSN 1726-5479, Vol. 93, Issue 6, pp. 1-14, June 2008.