Flomeko 2000 - IMEKO TC9 Conference
Salvador, Bahia, BRAZIL
4-8 June, 2000

GUIDE AND BAYESIAN APPROACH TO POLYNOMIAL REGRESSION MODELS IN FLOW METER CALIBRATION DATA

Olga S. Yoshida and Nilson M. Taira
Institute for Technological Research of Sao Paulo State
Av. Prof Almeida Prado 532 CEP 05508-901
São Paulo/SP Brasil

Márcia D'Elia Branco
Departamento de Estatística
Universidade de São Paulo
Caixa Postal 66281 CEP 05315-970
São Paulo/SP Brasil

Abstract

The polynomial regression model has been largely used in flow meter calibration. The document ISO 7006 - Part II describes how to adjust a polynomial expression to non linear calibration data. The Guide establishes general rules for evaluating and expressing uncertainty in measurement, but does not demonstrate methods of evaluating uncertainty in flow meter calibration. In this paper we concentrate on the uncertainty analysis for calibration of flow gas meters by using polynomial regression models. A detailed Guide analysis of this model will be given with selected examples of non linear calibration data. The Guide analysis of this model is dictated by assumptions made on the distribution of errors. Typically, it is assumed that the errors have normal distributions but it is not true in many situations. Alternatives to the normal distribution for regression errors are discussed such as Student-t with small number of degree of freedom using Bayesian statistics. The rules recommended in the Guide are viewed in the light of Bayesian concepts, and we conclude that the interpretation of the results are very natural. The Bayes calculations was done with the Bugs (Bayesian Using Gibbs Sampling) software.

Keywords: Flowmeter calibration, uncertainty, Guide, polynomial regression, Bayesian statistics.

1 Introduction

The results of a flowmeter calibration are usually presented in a table with a series of calibration points and the calibration uncertainty. The user has to decide how to use the indicated and the associated values to reported the uncertainty of a interpolated point.

To avoid this problem, the use of the calibration graph is normally indicated where is very important to know accurately the errors or "true" values of flowrates. This practice is most common in the calibration laboratories where the calibration of flowmeter is reported also through of a polynomial regression model.

So, the calibration will result in a graph which will subsequently be used to predict the flowrate. As this subsequent flowrate prediction has to have a uncertainty attached to it, then not only the functional relationship between the calibration coefficient and the flowrate but also the uncertainty in the calibration coefficient needs to be established during calibration.

Q_i is the indicated flowrate generated by the calibrated meter,

$Qcal_i$ is the "true flowrate" and

$D_i = (Q_i - Qcal_i)/Q_i * 100\ \%$ is the percentage deviation of indicated from "true flowrate".

In this paper we consider two set of calibration data, VT02 and VT05 data. Their calibration graph (Figure 1), the plot of percentage deviation (D_i) against the flowmeter indication (Q_i/Q_{max}) , show a trend over the range of meter operation. A curve than can be fitted to the trend, and the model

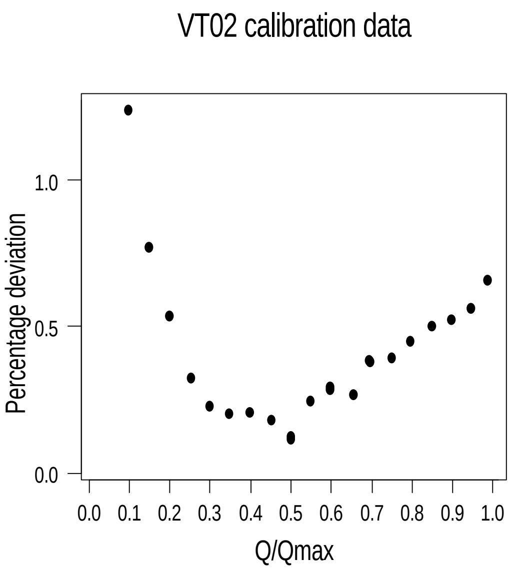
$$D_i = C_{-2}\left(\frac{Q_i}{Q_{\max}}\right)^{-2} + C_{-1}\left(\frac{Q_i}{Q_{\max}}\right)^{-1} + C_0 + C_1\left(\frac{Q_i}{Q_{\max}}\right)^{1} + C_2\left(\frac{Q_i}{Q_{\max}}\right)^{2} + e_i \tag{1}$$

with random errors terms is expected to be appropriated, the shape of the curve is expected from experience with similar types of flowmeter. This curve will subsequently be used to predict the test flowmeter deviation.

The appropriate analysis of (1) is dictated by the assumptions made on the distribution of errors. Typically, it is assumed that the errors have mean zero and variance σ^2 and that errors associated with distinct calibration points are uncorrelated.

Based on *n* calibration points (d_i,q_i), $i=1,2,...,n,$ our objective is to estimate C_{-2} , C_{-1} , C_0 , C_1 , C_2 , σ^2 , and to indicate the uncertainty of a prediction obtained for a given indicated flowrate when estimatives of C_{-2} , C_{-1} , C_0 , C_1 , C_2 are used in (1). These inferences require further specification of the distribution of the errors. The classical results are developed assuming a normal distribution.

The document ISO 7066-Part II describes how to adjust a curve to a non-linear set of calibration data. The uncertainty calculation in this document is presented under the classical results assuming normality of the errors. In the next section, the classical least-squares analysis of VT02 and VT05 data are sumarized and the fitted curve is used to compose a uncertainty propagation model under the rules presented in the *Guide*. In the section "Bayesian polynomial regression model and the *Guide* uncertainty analysis" the Bayesian approach is proposed to cope with nonnormality of the errors. Wöger (1997) conclude that the rules recommended in the Guide viewed in the light of Bayesian concepts are very natural. In the last section "Discussion" we reinforce the Wöger conclusions under the Bayesian regression model.



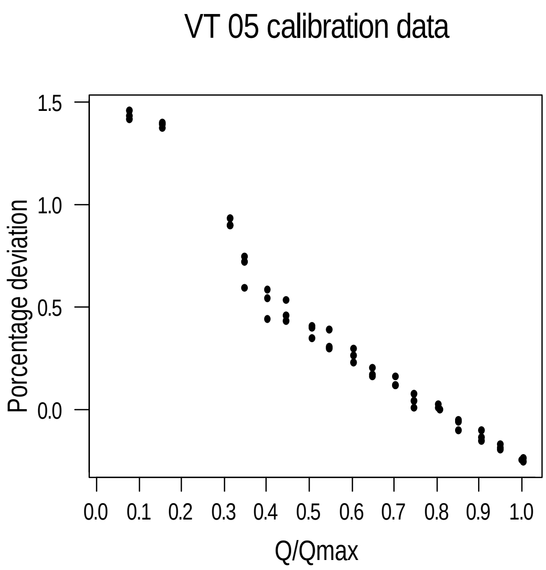


Figure 1 – Percentage deviation of indicated from calibrated flowrate for VT02 data and VT05 data.

2 Polynomial regression model and Guide uncertainty analysis

The normal error model is as follows:

$$D_i = C_{-2}\left(\frac{Q}{Q_{\max}}\right)_i^{-2} + C_{-1}\left(\frac{Q}{Q_{\max}}\right)_i^{-1} + C_0 + C_1\left(\frac{Q}{Q_{\max}}\right)_i^{1} + C_2\left(\frac{Q}{Q_{\max}}\right)_i^{2} + e_i \quad , \tag{2}$$

where:

- D_i is the observed deviation in the *i* th calibration point,
- $(Q/Q_{max})_i$ is the level of flowrate in the *i* th calibration point,
- C_{-2} , C_{-1} , C_0 , C_1 , and C_2 are the polynomial coefficients,
- ε_i are independent normally distributed, with mean 0 and variance σ_2.

(3)

The classical least-square analysis for VT02 data

The regression equation is

$$\hat{d}_i = -0.0162\left(\frac{Q}{Q_{max}}\right)_i^{-2} + 0.434\left(\frac{Q}{Q_{max}}\right)_i^{-1} - 1.76 + 2.62\left(\frac{Q}{Q_{max}}\right)_i^{1} - 0.632\left(\frac{Q}{Q_{max}}\right)_i^{2} \tag{4}$$

with coefficiente of multiple determination R^2=98.3%, thus,under this model, the variation in porcentage deviation is reduced by 98.3% which indicates a good fit.

Predictor	Coef	StDev	T	P
Constant	-1.7582	0.4087	-4.30	0.000
$(Q/Qmax)^{-2}$	-0.016207	0.004811	-3.37	0.004
$(Q/Qmax)^{-1}$	0.43375	0.08092	5.36	0.000
$(Q/Qmax)^{1}$	2.6177	0.7449	3.51	0.003
$(Q/Qmax)^{2}$	-0.6317	0.4330	-1.46	0.163

S = 0.03700 R-Sq = 98.3% R-Sq(adj) = 97.9%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	1.33780	0.33445	244.24	0.000
Error	17	0.02328	0.00137		
Total	21	1.36108			

Source	DF	Seq SS
$(Q/Qmax)^{-2}$	1	0.83431
$(Q/Qmax)^{-1}$	1	0.15572
$(Q/Qmax)^{1}$	1	0.34486
$(Q/Qmax)^{2}$	1	0.00291

The classical least-square analysis for VT05 data
The regression equation is

$$\hat{d}_i = -0.0166\left(\frac{Q}{Q_{\max}}\right)_i^{-2} + 0.315\left(\frac{Q}{Q_{\max}}\right)_i^{-1} + 0.158 - 0.577\left(\frac{Q}{Q_{\max}}\right)_i^{1} - 0.111\left(\frac{Q}{Q_{\max}}\right)_i^{2} \tag{5}$$

Predictor	Coef	StDev	T	P
Constant	0.1584	0.2768	0.57	0.570
$(Q/Qmax)^{-2}$	-0.016578	0.002560	-6.48	0.000
$(Q/Qmax)^{-1}$	0.31463	0.05039	6.24	0.000
$(Q/Qmax)^{1}$	-0.5774	0.5298	-1.09	0.281
$(Q/Qmax)^{2}$	-0.1114	0.3154	-0.35	0.726

S = 0.04806 R-Sq = 99.1% R-Sq(adj) = 99.1%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	4	12.0773	3.0193	1307.20	0.000
Error	46	0.1062	0.0023		
Total	50	12.1836			

Source	DF	Seq SS
$(Q/Qmax)^{-2}$	1	6.0023
$(Q/Qmax)^{-1}$	1	5.8022
$(Q/Qmax)^{1}$	1	0.2725
$(Q/Qmax)^{2}$	1	0.0003

Guide Uncertainty analysis

Assuming that the above curve models are correct we establish the following models for propagation of uncertainty:

Model I

$$E(D_i) = C_{-2}\left(\frac{Q_i}{Q_{\max}}\right)^{-2} + C_{-1}\left(\frac{Q_i}{Q_{\max}}\right)^{-1} + C_0 + C_1\left(\frac{Q_i}{Q_{\max}}\right)^{1} + C_2\left(\frac{Q_i}{Q_{\max}}\right)^{2} + d_i$$

δ_i is the systematic effect in the ith calibration point.

(6)

Table 1 - Uncertainty budget (expected deviation $E(D_i)$ corresponding to Q_i/Q_{max})

Quantity	estimate	Probability distribution	Standard Uncertainty (type of evaluation)	Degree of freedom	Sensi-tivity Coefficient	Contribution to the standard uncertainty	
						absolute	relative
curve	$\hat{d}_i$ (%)	Normal	$S(\hat{d}_i)$ (A)	(n-p)	1	$S(\hat{d}_i)$	$S^2(\hat{d}_i)/u^2(D_i)*100$
δ_i	0 (%)	Normal	$u(\delta)$ (B)	≥ 30	1	$u(\delta)$	$u^2(\delta)/u^2(D_i)*100$
$E(D_i)$	$\hat{d}_i$	Normal	-------	n_{eff}	-----	$u(\hat{d}_i)$	100%

u($\hat{d}_i$ (new)) = $\sqrt{S^2(\hat{d}_i) + u^2(d)}$ and $$v_{eff} = \frac{u^4(\hat{d}_i(new))}{\frac{S^4(\hat{d}_i)}{(n-p)} + \frac{u^4(d)}{30}}$$

(7)

Model II

$$D_{i(new)} = \hat{D}_i + e_i + d_i$$

where δ_i is the systematic effect in the i th calibration point.

(8)

Table 2 - Uncertainty budget (prediction of new deviation $D_{i(new)}$ corresponding to Q_i/Q_{max})

Quantity	estimate	Probability distribution	Standard Uncertainty (type of evaluation)	Degree of freedom	Sensi-tivity Coeffi cient	Contribution to the standard uncertainty	
						absolute	relative
D_i	$\hat{d}_i$ (%)	Normal	S($\hat{d}_i$) (A)	(n-p)	1	S($\hat{d}_i$)	$S^2(\hat{d}_i)/u^2(D_i)*100$
ε_i	0 (%)	Normal	$\sqrt{MSE}$ (A)	(n-p)	1	$\sqrt{MSE}$	$MSE/u^2(D_i)*100$
δ_i	0 (%)	Normal	u(δ_i) (B)	≥ 30	1	u(δ_i)	$u^2(\delta)/u^2(D_i)*100$
$D_{i(new)}$	$\hat{d}_i$ (new)	Normal	-------	n_{eff}	-----	u($\hat{d}_i$ (new))	100%

u($\hat{d}_i$ (new)) = $\sqrt{S^2(\hat{d}_i) + MSE + u^2(d)}$ and $$v_{eff} = \frac{u^4(\hat{d}_i(new))}{\frac{S^4(\hat{d}_i)}{(n-p)} + \frac{MSE^2}{(n-p)} + \frac{u^4(d)}{30}}$$

(9)

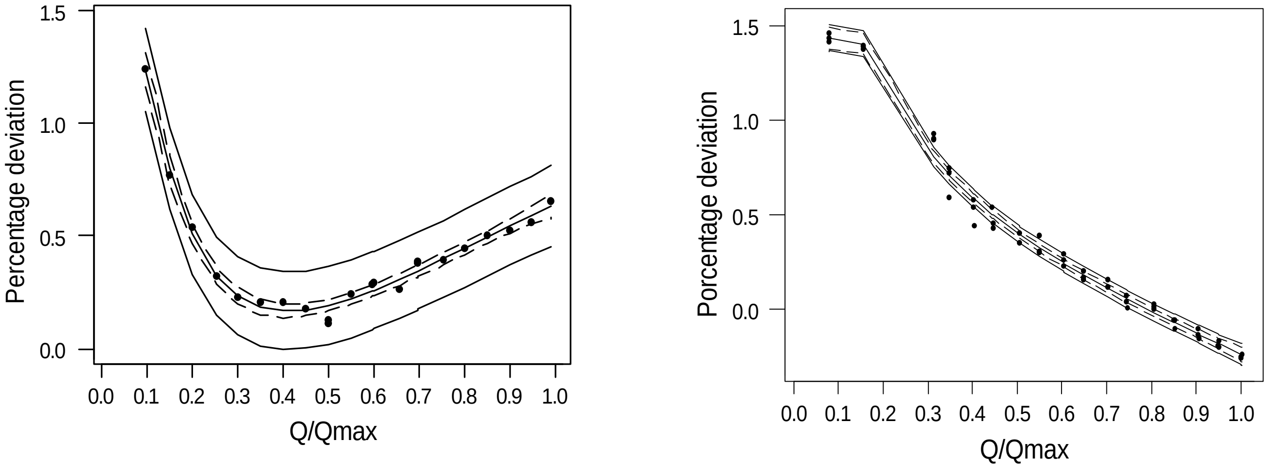


Figure 2 – The fit and the expanded uncertainty of prediction of new deviation $D_{i(new)}$ corresponding to Q_i/Q_{max}, i=1,2 ... n, for respectively VT02 and VT05 calibration data analysis.

Legend of Figure 2:

__________ Superior limit of expanded uncertainty of $D_{i(new)}$

--------------- Superior limit of expanded uncertainty of fit and errors

__________ Fit

--------------- Inferior limit of expanded uncertainty of fit and errors

__________ Inferior limit of expanded uncertainty of $D_{i(new)}$

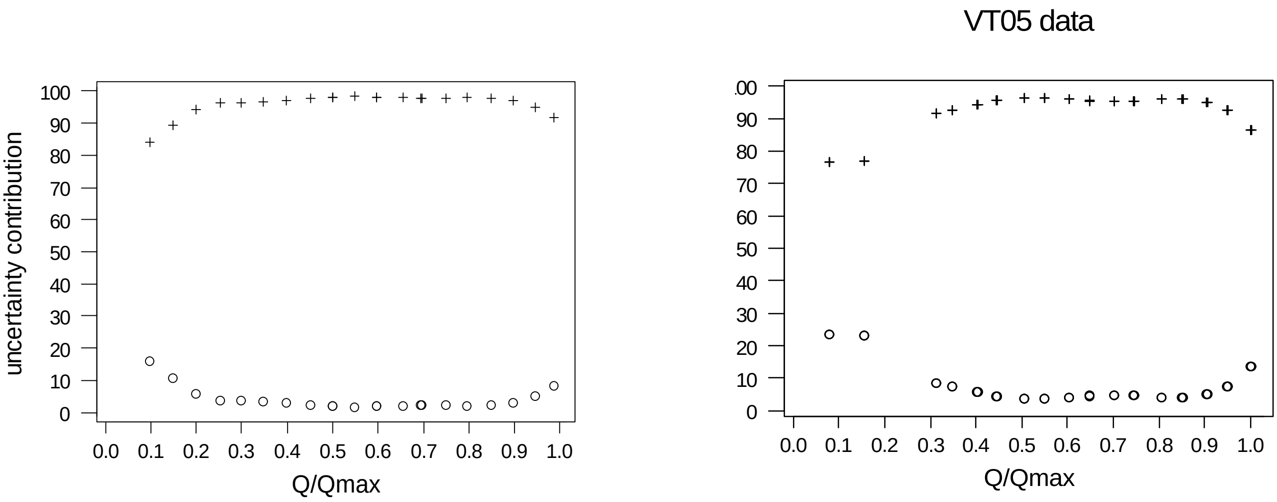


Figure 3 – Uncertainty contribution (%) to the combined uncertainty under Model II
+++++ : system contribution
ooooooo : fit and error contribution

3 Bayesian polynomial regression model and the Guide uncertainty analysis

There are many situations where the errors don't have normal distributions. The normal plot of the residuals for VT05 data suggests that the error distribution is not normal. The violation of errors normality may invalidate the uncertainty analysis, and beside that, the error normality doesn't accommodate the sensibility of the metrologisty that the error distribution should have, for instance, more probability concentrated in its tails. In this section we propose the following Student-t error model using Bayesian methodology:

- The errors ε_i , $i=1,2.....n$, given C_{-2} , C_{-1} , C_0 , C_1 ,C_2 , s^2 are independent and identically distributed according to the Student-t distribution with ν degrees of freedom and scale parameter s^2.
- The errors δ_i , $i=1,2...n$, given C_{-2} , C_{-1} , C_0 , C_1 ,C_2 , s^2 and $u^2(d)$ are independent of ε_i , $i=1,2.....n$, and identically distributed according t the normal distribution with mean zero and variance $u^2(d)$.
- The coefficients $C = (C_{-2}$, C_{-1} , C_0 , C_1 ,$C_2)$ and σ^2 are independent and have joint prior distribution $\pi(C,s^2)=\pi(C)\pi(\sigma^2)$.

The observed data calibration is written as $d=(d_1,d_2, ..., d_n)$, $q=(q_1,q_2, ..., q_n)$. We also consider a new flowrate $q_{i(new)}$ to predict a new deviation d_{inew}.

Bayesian version of Model I:

Under the Bayesian methodology the prediction of D_i is given by the mean of posterior distribution of D_i given $d=(d_1,d_2, ..., d_n)$, $q=(q_1,q_2, ..., q_n)$, n and $u^2(d)$, *that is*

$$E(D_i / (d_1,d_2, ..., d_n), (q_1,q_2, ..., q_n)), \tag{10}$$

and the standard uncertainty of posterior distribution of D_i given $d=(d_1,d_2, ..., d_n)$, $q=(q_1,q_2, ..., q_n)$, n and $u^2(d)$, is

$$\begin{aligned} u(D_i / (d_1,d_2, ..., d_n), (q_1,q_2, ..., q_n), n , u^2(d)) &= \sqrt{\mathrm{Var}(D_i / d_1d_2...d_n, q_1q_2...q_n, \nu, u^2(\delta))} \\ &= \sqrt{\mathrm{Var}(D_i / d_1d_2...d_n, q_1q_2...q_n, \nu)} + u(d). \end{aligned} \tag{11}$$

Bayesian version of Model II

The prediction of a new deviation $D_{i(new)}$ is the mean of posterior distribution of $D_{i(new)}$ given $d=(d_1,d_2, ..., d_n)$, $q=(q_1,q_2, ..., q_n)$, $q_{i(new)}$, n and $u^2(d)$, that is

$$E(D_{i(new)} / (d_1,d_2, ..., d_n), (q_1,q_2, ..., q_n), q_{i(new)}), \tag{12}$$

and the standard uncertainty of posterior distribution of $D_{i(new)}$ given $d=(d_1,d_2, ..., d_n)$, $q=(q_1,q_2, ..., q_n)$, $q_{i(new)}$, n and $u^2(d)$, is

$$u(D_{i(new)} / (d_1,d_2, ..., d_n), (q_1,q_2, ..., q_n), q_{i(new)} , n , u^2(d)) = \sqrt{\mathrm{Var}(D_{i(new)} / d_1d_2...d_n, q_1q_2...q_n, q_{i(new)}, \nu)} + u(d). \tag{13}$$

4 Discussion

The integral calculations of (10) (11) (12) and (13) must be performed on a computer and the Bugs software may be used. These calculations are, in fact, recommended by the *Guide* to describe the uncertainty propagation under the models I and II. These recommendation appear as a simple consequence of probability theory under the Bayesian view.

The uncertainties in (11) and (13) are mixtures of ν/ν-1 * σ^2. Than these uncertainties are affected by the freedom degree off Student–t the errors distribution, the bigger is ν the smaller will be the uncertainties in (11) and (13). Another important observation is about the role played by the prior information and the sample information as a Type B and Type A information. Under the Bayesian view, these informations are treated in the same way, as conditioning information.

Finally we couldn't have Student-t distribution for the errors under the classical methodology, for a while, only the Bayesian methodology allow this distribution for the errors.

References

[1] Branco, M., Bolfarine, H. and Iglesias, P. (1998) Bayesian calibration under a Student-t model. *Computational Statistics*, 13:319-338.
[2] Gelman A. Carlin J.B. Stern H.S. and Rubin D.B. (1995). *Bayesian Data Analysis*. New York: Chapman & Hall
[3] Guide to expression of uncertainty in measurement. ISO international Organisation for Standardization, Geneva 1995
[4] International Organization for Standardization. Measurement of fluid flow: assessment of uncertainty in the calibration and use of flow measurement devices, Part II, Non-linear calibration relationships, Geneva, 7066/2, 1988.
[5] Spiegelhalter D., Thomas A. and Best N, (1999) WinBUGS user manual. Version 1.1.1.
[6] Weise K., Wöger W.: A Bayesian theory of measurement uncertainty. *Meas. Sci. Technol*. 4; 1-11; 1993
[7] W. Wöger: Bayesian Statistics and the International Guidelines for Expressing Uncertainty in Measurement, Advanced School of Metrology: Evaluation of Uncertainty in Measurement, Angra dos Reis, RJ, Brazil, March 1997